

A First Course in
**Abstract
Algebra**

8th EDITION

John B. **FRALEIGH**
Neal E. **BRAND**

This page is intentionally left blank

A First Course in Abstract Algebra

Eighth Edition

John B. Fraleigh
University of Rhode Island

Neal Brand
University of North Texas

Historical Notes by Victor Katz
University of District of Columbia



Copyright © 2021, 2003, 1994 by Pearson Education, Inc. or its affiliates, 221 River Street, Hoboken, NJ 07030. All Rights Reserved. Manufactured in the United States of America. This publication is protected by copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise. For information regarding permissions, request forms, and the appropriate contacts within the Pearson Education Global Rights and Permissions department, please visit www.pearsoned.com/permissions/.

Acknowledgments of third-party content appear on the appropriate page within the text.

Cover image credit: Edmund Sumner/AGE Fotostock

PEARSON, ALWAYS LEARNING, and MYLAB are exclusive trademarks owned by Pearson Education, Inc. or its affiliates in the U.S. and/or other countries.

Unless otherwise indicated herein, any third-party trademarks, logos, or icons that may appear in this work are the property of their respective owners, and any references to third-party trademarks, logos, icons, or other trade dress are for demonstrative or descriptive purposes only. Such references are not intended to imply any sponsorship, endorsement, authorization, or promotion of Pearson's products by the owners of such marks, or any relationship between the owner and Pearson Education, Inc., or its affiliates, authors, licensees, or distributors.

Library of Congress Cataloging-in-Publication Data

Names: Fraleigh, John B., author. | Katz, Victor J., writer of added commentary.

Title: A first course in abstract algebra / John B. Fraleigh ; historical notes by Victor Katz.

Description: Eighth edition. | [Hoboken, New Jersey] : Pearson, [2021] |

Series: World student series | Includes bibliographical references and index.

Identifiers: LCCN 2019038536 | ISBN 9780135758168 (paperback) | ISBN 9780321390363 (ebook)

Subjects: LCSH: Algebra, Abstract.

Classification: LCC QA162 .F7 2020 | DDC 512.02–dc23

LC record available at <https://lccn.loc.gov/2019038536>

ScoutAutomatedPrintCode



Rental
ISBN-10: 0-13-673162-7
ISBN-13: 978-0-13-673162-7

Loose Leaf Version
ISBN-10: 0-13-575816-5
ISBN-13: 978-0-13-575816-8

Contents

Instructor's Preface vii

Dependence Chart xii

Student's Preface xv

- 0** Sets and Relations 1

I

GROUPS AND SUBGROUPS

11

- 1** Binary Operations 11
2 Groups 19
3 Abelian Examples 32
4 Nonabelian Examples 39
5 Subgroups 52
6 Cyclic Groups 61
7 Generating Sets and Cayley Digraphs 70

II

STRUCTURE OF GROUPS

77

- 8** Groups of Permutations 77
9 Finitely Generated Abelian Groups 88
10 Cosets and the Theorem of Lagrange 97
[†]11 Plane Isometries 105

III

HOMOMORPHISMS AND FACTOR GROUPS

113

- 12** Factor Groups 113
13 Factor-Group Computations and Simple Groups 121

[‡] 14	Group Action on a Set	132
[†] 15	Applications of G -Sets to Counting	140

IV**ADVANCED GROUP THEORY**

16	Isomorphism Theorems	145
17	Sylow Theorems	149
18	Series of Groups	157
19	Free Abelian Groups	166
20	Free Groups	172
21	Group Presentations	177

145

V**RINGS AND FIELDS**

22	Rings and Fields	185
23	Integral Domains	194
24	Fermat's and Euler's Theorems	200
25	Encryption	205

185

VI**CONSTRUCTING RINGS AND FIELDS**

26	The Field of Quotients of an Integral Domain	211
27	Rings of Polynomials	218
28	Factorization of Polynomials over a Field	228
[†] 29	Algebraic Coding Theory	237
30	Homomorphisms and Factor Rings	243
31	Prime and Maximal Ideals	250
[†] 32	Noncommutative Examples	258

211

VII**COMMUTATIVE ALGEBRA**

33	Vector Spaces	267
34	Unique Factorization Domains	275
35	Euclidean Domains	286
36	Number Theory	292
[†] 37	Algebraic Geometry	297
[†] 38	Gröbner Bases for Ideals	303

267

VIII**EXTENSION FIELDS**

39	Introduction to Extension Fields	311
40	Algebraic Extensions	319
[†] 41	Geometric Constructions	328
42	Finite Fields	335

311

IX**GALOIS THEORY****341**

- 43** Introduction to Galois Theory 341
44 Splitting Fields 349
45 Separable Extensions 357
46 Galois Theory 364
47 Illustrations of Galois Theory 372
48 Cyclotomic Extensions 378
49 Insolvability of the Quintic 384

Appendix: Matrix Algebra 391

Bibliography 395

Notations 397

Answers to Odd-Numbered Exercises Not Asking for Definitions or Proofs 401

Index 419

[†] Not required for the remainder of the text.

[‡] This section is a prerequisite for Sections 17 and 36 only.

This page is intentionally left blank

Instructor's Preface

This is an introduction to abstract algebra. It is anticipated that the students have studied calculus and probably linear algebra. However, these are primarily *mathematical maturity* prerequisites; subject matter from calculus and linear algebra appears mostly in illustrative examples and exercises.

As in previous editions of the text, our aim remains to teach students as much about groups, rings, and fields as we can in a first course. For many students, abstract algebra is their first extended exposure to an axiomatic treatment of mathematics. Recognizing this, we have included extensive explanations concerning what we are trying to accomplish, how we are trying to do it, and why we choose these methods. Mastery of this text constitutes a firm foundation for more specialized work in algebra and also provides valuable experience for any further axiomatic study of mathematics.

New to This Edition

[Editor's Note: You may have noticed something new on the cover of the book. Another author! I am thrilled that Neal Brand agreed to update this classic text. He has done so carefully and thoughtfully, staying true to the spirit in which it was written. Neal's years of experience teaching the course with this text at the University of North Texas have helped him produce a meaningful and worthwhile update to John Fraleigh's work.]

Updates for the eText

A focus of this revision was transforming it from a primarily print-based learning tool to a digital learning tool. The eText is therefore filled with content and tools that will help bring the content of the course to life for students in new ways and help you improve instruction. Specifically,

- **Mini lectures.** These brief author-created videos for each section of the text give an overview to the section but not every example or proof. Some sections will have two videos. I have used these videos effectively with my students, who were assigned to watch them ahead of the lecture on that topic. Students came to class with a basic overview of the topic of the day, which had the effect of reducing lecture time and increasing the class time used for discussion and student

presentations. Students reported that the videos were helpful in giving an overview of the topics and a better understanding of the concepts and proofs. Students were also encouraged to view the videos after the topic was covered in class to reinforce what they learned. Many students also used the videos to review topics while preparing for exams. Although I have not attempted to flip the classroom, my intention was to provide sufficient resources in the eText to make it feasible without requiring other resources.

- **Key idea quizzes.** A database of definitions and named theorems will allow students to quiz themselves on these key ideas. The database can be used in the way that flash cards were traditionally used.
- **Self-assessments.** Occasional questions interspersed in the narrative allow students to check their understanding of new ideas.
- **Interactive figures and utilities.** I have added a number of opportunities for students to interact with content in a dynamic manner in order to build or enhance understanding. Interactive figures allow students to explore concepts geometrically or computationally in ways that are not possible without technology.
- **Notes, Labels, and Highlights.** Notes allow instructors to add their personal teaching style to important topics, call out need-to-know information, or clarify difficult concepts. Students can make their eText their own by creating highlights with meaningful labels and notes, helping them focus on what they need to study. The customizable Notebook allows students to filter, arrange, and group their notes in a way that makes sense to them.
- **Dashboard.** Instructors can create reading assignments and see the time spent in the eText so that they can plan more effective instruction.
- **Portability.** Portable access lets students read their eText whenever they have a moment in their day, on Android and iOS mobile phones and tablets. Even without an Internet connection, offline reading ensures students never miss a chance to learn.
- **Ease-of-Use.** Straightforward setup makes it easy for instructors to get their class up and reading quickly on the first day of class. In addition, Learning Management System (LMS) integration provides institutions, instructors, and students with single sign-on access to the eText via many popular LMSs.

Exercises

Many exercises in the text have been updated, and many are new. In order to prevent students from using solutions from the previous edition, I purposefully replaced or reworded some exercises.

I created an Instructor Solutions Manual, which is available online at www.pearson.com to instructors only. Solutions to exercises involving proofs are often sketches or hints, which would not be in the proper form to turn in.

Text Organization Modifications

For each part of the text, I provide an overview of the changes followed by significant changes to sections. In cases where changes to parts or sections were minor, I have not included a list of changes.

Part I: Groups and Subgroups

- Overview of changes: My main goals were to define groups and to introduce the symmetric and dihedral groups as early as possible. The early introduction of these

groups provides students with examples of finite groups that are consistently used throughout the book.

- Section 1 (Binary Operations). Former Section 2. Added definition of an identity for a binary operation.
- Section 2 (Groups). Former Section 4. Included the formal definition of a group isomorphism.
- Section 3 (Abelian Examples). Former Section 1. Included definition of circle group, R_a , and Z_n . Used circle group to show associativity of Z_n and R_a .
- Section 4 (Nonabelian Examples). Based on parts of former Sections 5, 8, and 9. Defined dihedral group and symmetric group. Gave a standardized notation for the dihedral group that is used consistently throughout the book. Introduced both two-row and cycle notation for the symmetric group
- Section 5 (Subgroups). Former Section 5. Included statement of two other conditions that imply a subset is a subgroup and kept the proofs in the exercise section. Made minor modifications using examples from new Section 4.
- Section 6 (Cyclic Groups). Former Section 6. Added examples using dihedral group and symmetric group.
- Section 7 (Generating Sets and Cayley Digraphs). Minor modification of former Section 7.

Part II: Structure of Groups

- Overview of changes: The main goal was to give the formal definition of homomorphism earlier in order to simplify the proofs of Cayley's and Lagrange's theorems.
- Section 8 (Groups of Permutations). Included formal definition of homomorphism. Based on parts of former Sections 8, 9, and 13. Used two-row permutation notation to motivate Cayley's theorem before proof. Deleted first part of section 13 (covered in Section 4). Omitted determinant proof of even/odd permutations since definition of determinant usually uses sign of a permutation. Kept orbit counting proof. Put determinant proof and inversion counting proof in exercises.
- Section 9 (Finitely Generated Abelian Groups). Former Section 11. Added the invariant factor version of the theorem. Showed how to go back and forth between the two versions of the fundamental theorem.
- Section 10 (Cosets and the Theorem of Lagrange). Former Section 10. Changed the order by putting Lagrange's Theorem first, motivating G/H later in the section.
- Section 11 (Plane Isometries). Minor modification of former Section 12.

Part III: Homomorphisms and Factor Groups

- Overview of changes: My main goal was to include a few more examples to motivate the theory and give an introduction to using group actions to prove properties of groups.
- Sections 12-15 are based on former Sections 14-17, respectively.
- Section 12 (Factor Groups). Started section with Z/nZ example to motivate general construction. Defined factor groups from normal subgroups first instead of from homomorphisms. After developing factor groups, showed how they are formed from homomorphisms.
- Section 13 (Factor-Group Computations and Simple Groups). Added a few more examples of computing factor groups. Explicitly used the fundamental homomorphism theorem in computation examples.

Instructor's Preface

- Section 14 (Group Action on a Set). Expanded examples of the general linear group and the dihedral group acting on sets. Added some applications of group actions to finite groups in anticipation of the Sylow Theorems, including Cauchy's Theorem and that fact that p -groups have a nontrivial center.
- Section 15 (Applications of G -sets to Counting). Minor modifications.

Part IV: Advanced Group Theory

- Overview of changes: I moved this part to be closer to the rest of the group theory sections. More examples were included to help clarify the concepts.
- Section 16 (Isomorphism Theorems). Former Section 3. Added two examples and rewrote proofs of two theorems.
- Section 17 (Sylow Theorems). Former Sections 36 and 37. Since Cauchy's Theorem and a few other theorems leading to the Sylow Theorems were covered in new Section 14, this material was removed and the old Sections 36 and 37 were combined. A few examples and exercises were added and a proof was rewritten.
- Section 18 (Series of Groups). Former Section 35. The proof of the Zassenhaus Lemma was placed after the theorem instead of making the argument before stating the theorem. One example added.
- Sections 19 (Free Abelian Groups), 20 (Free Groups), and 21 (Group Presentations). Minor modifications of former Sections 38–40.

Part V: Rings and Fields

- Overview of changes: The previous Part IV was split into two parts, one giving an introduction and the second giving methods of constructing rings and fields.
- Section 22 (Rings and Fields). Minor modification of former Section 18.
- Section 23 (Integral Domains). Former Section 19. Changed former Theorem 19.3 to classify all elements in Z_n . Added corollary that Z_p is a field, anticipating the theorem that all finite integral domains are fields.
- Section 24 (Fermat's and Euler's Theorems). Former Section 20. Simplified proof of Euler's generalization using classification of elements in Z_n .
- Section 25 (Encryption). New section outlining how RSA encryption works. This provides a nice application of the material in Section 24.

Part VI: Constructing Rings and Fields

- Overview of changes: Part VI includes sections from the previous Parts IV and V. The change emphasizes construction techniques used to form rings and fields.
- Section 26 (The Field of Quotients of an Integral Domain). Former Section 21. Rewrote the introduction to include two examples of integral domains and their field of quotients to motivate the general construction.
- Section 27 (Rings of Polynomials). Minor modification of former Section 22.
- Section 28 (Factorization of Polynomials over a Field). Former Section 23. Rewrote former Theorem 23.1 by making a lemma showing how to reduce degree of polynomials in set S . Included proof of former 23.11 in the exercises.
- Section 29 (Algebraic Coding Theory). New section introducing coding theory, focusing on polynomial codes. This gives an application of polynomial computation over a finite field.
- Section 30 (Homomorphisms and Factor Rings). Former Section 26. Motivated why you need the usual conditions for an ideal by starting the section with the example of Z/nZ . Rearranged the order by showing that I an ideal of R gives rise

to the factor ring R/I , then included the material on homomorphisms and factor rings from the kernel. Expanded the statement of former Theorem 26.3 to make it easier to read and more approachable.

- Section 31 (Prime and Maximal Ideals). Minor modification of former Section 27.
- Section 32 (Noncommutative Examples). Minor modification of former Section 24.

Part VII: Commutative Algebra

- Overview of changes: This part includes sections that fit under the general heading of commutative algebra.
- Section 33 (Vector Spaces). Former Section 30. Added two examples and a brief introduction to R -modules over a ring motivated by vector spaces and abelian groups. Moved Former Theorem 30.23 to Section 45 on field extensions.
- Section 34 (Unique Factorization Domains). Former Section 45. Included definition of a Noetherian ring and made other minor changes.
- Section 35 (Euclidean Domains) and Section 36 (Number Theory) are minor modifications of Sections 46 and 47, respectively.
- Section 37 (Algebraic Geometry). Based on the first half of former Section 28. Added a proof of the Hilbert Basis Theorem.
- Section 38 (Gröbner Bases for Ideals). Based on the second half of former Section 28. Added two applications of Gröbner Bases: deriving the formulas for conic sections and determining if a graph can be colored with k colors.

Part VIII: Extension Fields

- Overview of changes: Part VIII consists of minor changes from former Part VI.
- Section 39 (Introduction to Extension Fields). Former Section 29. Divided former Theorem 29.13 into a theorem and a corollary. Rewrote former Theorem 29.18 and its proof to make it easier to follow. Included example moved from former Section 30.
- Section 40 (Algebraic Extensions), Section 41 (Geometric Constructions), and Section 42 (Finite Fields) are minor modifications of former Sections 31–33, respectively.

Part IX: Galois Theory

- Overview of changes: The previous Part X was rewritten to form Part IX. The goal was to improve the readability of the material while maintaining a rigorous development of the theory.
- Section 43 (Introduction to Galois Theory). New section. Uses the field extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ throughout to motivate and illustrate basic definitions and theorems including field automorphism, field fixed by an automorphism, group of automorphisms fixing a subfield, conjugates, and the conjugate isomorphism theorem. By using an easy-to-understand example consistently throughout, the concepts become more concrete.
- Section 44 (Splitting Fields). Includes the contents of former Sections 49 and 50, but it is completely rewritten. Less emphasis is given to the algebraic closure of a field and more emphasis is given to subfields of splitting fields.
- Section 45 (Separable Extensions). Contents include most of former Section 51 and a little from former Section 53, but material has been rewritten. The notation $\{E:F\}$ was omitted and definition of separable was given in terms of multiplicity of zeros. Emphasized subfields of the complex numbers.

- Former Section 52 on totally inseparable extensions was omitted since it was not used elsewhere and it detracts from the flow of the rest of Part IX.
- Section 46 (Galois Theory). Former Section 53. Separated the parts of Galois Theory into separate theorems. Continued the same example throughout the section to motivate and illustrate the theorems. By the end of the section, the continued example illustrates how Galois Theory can be used.
- Section 47 (Illustrations of Galois Theory). Minor modification of former Section 54.
- Section 48 (Cyclotomic Extensions). Former Section 55. In order to make the text more readable, restricted the field extensions to subfields of the complex numbers over the rational numbers since this is the only case that is used in the book.
- Section 49 (Insolvability of the Quintic). Former Section 56. Replaced construction of a polynomial that is not solvable by radicals with a specific concrete polynomial. The previous construction of a nonsolvable polynomial was moved to the exercises.

Part X: Groups in Topology (Online at bit.ly/2VBCiej)

- Sections 50-53 are minor modifications of former sections 41-44.

Some Features Retained

I continue to break down most exercise sets into parts consisting of computations, concepts, and theory. Answers to most odd-numbered exercises not requesting a proof again appear at the back of the text. I am supplying the answers to parts a, c, e, g, and i only of our 10-part true-false exercises. The excellent historical notes by Victor Katz are, of course, retained.

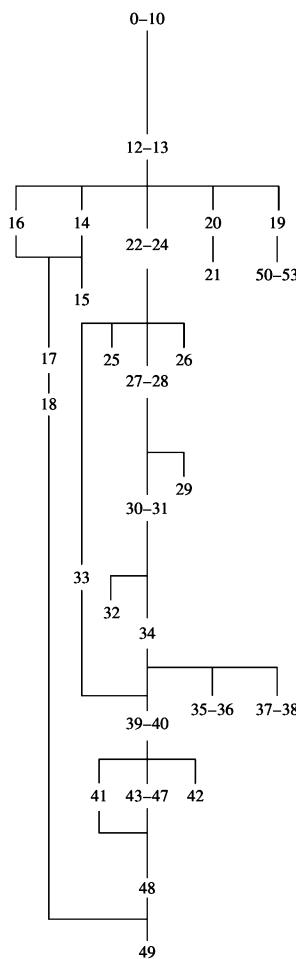
Suggestions for New Instructors of Algebra

Those who have taught algebra several times have discovered the difficulties and developed their own solutions. The comments we make here are not for them.

This course is an abrupt change from the typical undergraduate calculus for the students. A graduate-style lecture presentation, writing out definitions and proofs on the board for most of the class time, will not work with most students. We have found it best to spend at least the first half of each class period answering questions on homework, trying to get a volunteer to give a proof requested in an exercise, and generally checking to see if they seem to understand the material assigned for that class. Typically, we spent only about the last 20 minutes of our 50-minute time talking about new ideas for the next class, and giving at least one proof. The videos for each section can effectively be used to supplement or replace lectures. From a practical point of view, it is a waste of time to try to write on the board all the definitions and proofs. They are in the text.

We suggest that at least half of the assigned exercises consist of the computational ones. Students are used to doing computations in calculus. Although there are many exercises asking for proofs that we would love to assign, we recommend that you assign at most two or three such exercises and try to get someone to explain how each proof is performed in the next class. We do think students should be asked to do at least one proof in each assignment.

Students face a barrage of definitions and theorems, something they have never encountered before. They are not used to mastering this type of material. Grades on tests that seem reasonable to us, requesting a few definitions and proofs, are apt to be low and depressing for most students and instructors. To encourage students to keep up

**Dependence Chart**

with the basic material, I give approximately ten pop quizzes per semester that typically involve stating a definition, giving an example, or stating a major theorem.

At the University of North Texas, abstract algebra is a two-semester sequence. The first semester is required of all math majors and the second semester is optional. Because most students opt not to continue with the second semester, it is not offered every year. When I teach either class, I give three 50-minute in-class exams. With exam reviews and going over completed exams, this leaves approximately 36 class periods for new material.

In the first-semester class, the base material I always cover includes Sections 0-6, 8, 9, 12, 13, and 22-25. I average approximately two class periods per section, so I can usually cover a few more sections. Options I have used for the remaining time include

Sections 14 and 15, Sections 26–28, Section 17, or Sections 30 and 31. One semester I attempted to cover enough field extension material in order to cover Section 41. This required me to carefully select material in Sections 27, 28, 39, and 40 in order to prepare the students for Section 41.

For the second semester, I usually have as goals proving the impossibility of bisecting an angle using compass and straightedge and the insolubility of quintic polynomials. Assuming that students have seen the basic material in the first semester as described above, these goals require covering material from Sections 16, 18, 27, 28, 30, 31, 33, 34, and 39–49. This turns out to be an ambitious undertaking, but the purpose of rewriting Part IX was to make the material more accessible to students, and therefore make the goal of covering Galois Theory in a second-semester class more feasible.

Acknowledgments

I am very grateful to those who have reviewed the text or who have sent suggestions and corrections. Below is a list of faculty who contributed their thoughts on improving the text.

- Deb Bergstrand, Swarthmore College
- Anthony E. Clement, Brooklyn College
- Richard M. Green, University of Colorado
- Cheryl Grood, Swarthmore College
- Gary Gordon, Lafayette College
- John Harding, New Mexico State University
- Timothy Kohl, Boston University
- Cristian Lenart, University at Albany, SUNY
- Mariana Montiel, Georgia Southern University
- Anne Shiu, Texas A&M University
- Mark Stankus, California Polytechnic State University
- Janet Vassilev, University of New Mexico
- Cassie L. Williams, James Madison University
- T. E. Williamson, Montclair State University
- Michael Zuker, Massachusetts Institute of Technology

I also wish to express appreciation to Jeff Weidenaar, Tara Corpuz, and Jon Krebs at Pearson for their help with this project.

Neal Brand
University of North Texas

Student's Preface

This course may well require a different approach than those you used in previous mathematics courses. You may have become accustomed to working a homework problem by turning back in the text to find a similar problem, and then just changing some numbers. That may work with a few problems in this text, but it will not work for most of them. This is a subject in which understanding is all-important, and where problems should not be tackled without first studying the text.

Let us make some suggestions on studying the text. Notice that the text bristles with definitions, theorems, corollaries, and examples. The definitions are crucial. We must agree on terminology to make any progress. Sometimes a definition is followed by an example that illustrates the concept. Examples are probably the most important aids in studying the text. *Pay attention to the examples.*

Before reading a section, it may be helpful to watch the video associated with the section. I have two general pieces of advice for watching a video or reading the text. First, minimize your distractions. It takes a good deal of concentration for most of us to learn new technical information. Second, have paper and pen (or the electronic equivalent) at hand to take notes and to occasionally work out computations on your own.

I suggest you skip the proofs of the theorems on your first reading of a section, unless you are really "gung-ho" on proofs. You should read the statement of the theorem and try to understand just what it means. Often, a theorem is followed or preceded by an example that illustrates it, which is a great aid in really understanding what the theorem says. Pay particular attention to the summary at the end of each video to get an overview of the topics covered.

In summary, on your first viewing and reading of a section, I suggest you concentrate on what information the section gives and on gaining a real understanding of it. If you do not understand what the statement of a theorem means, it will probably be meaningless for you to read the proof.

Proofs are basic to mathematics. After you feel you understand the information given in a section, you should read and try to understand at least some of the proofs. In the videos you will find a few proofs. Watching the videos a second time after you have a better understanding of the definitions and the statements of the theorems will help to clarify these proofs. Proofs of corollaries are usually the easiest ones, for they often follow directly from the theorem. Many of the exercises under the "Theory" heading

ask for a proof. Try not to be discouraged at the outset. It takes a bit of practice and experience. Proofs in algebra can be more difficult than proofs in geometry and calculus, for there are usually no suggestive pictures that you can draw. Often, a proof falls out easily if you happen to look at just the right expression. Of course, it is hopeless to devise a proof if you do not really understand what it is that you are trying to prove. For example, if an exercise asks you to show that a given thing is a member of a certain set, you must *know* the defining criterion for a thing to be a member of that set, and then show that your given thing satisfies that criterion.

There are several aids for your study at the back of the text. Of course, you will discover the answers to odd-numbered problems that do not involve a proof. If you run into a notation such as Z_n that you do not understand, look in the list of notations that appears after the bibliography. If you run into terminology like *inner automorphism* that you do not understand, look in the index for the first page where the term occurs.

In summary, although an understanding of the subject is important in every mathematics course, it is crucial to your performance in this course. May you find it a rewarding experience.

SECTION 0**SETS AND RELATIONS****On Definitions, and the Notion of a Set**

Many students do not realize the great importance of definitions to mathematics. This importance stems from the need for mathematicians to communicate with each other. If two people are trying to communicate about some subject, they must have the same understanding of its technical terms. However, there is an important structural weakness.

It is impossible to define every concept.

Suppose, for example, we define the term *set* as “A **set** is a well-defined collection of objects.” One naturally asks what is meant by a *collection*. We could define it as “A collection is an aggregate of things.” What, then, is an *aggregate*? Now our language is finite, so after some time we will run out of new words to use and have to repeat some words already examined. The definition is then circular and obviously worthless. Mathematicians realize that there must be some undefined or primitive concept with which to start. At the moment, they have agreed that *set* shall be such a primitive concept. We shall not define *set*, but shall just hope that when such expressions as “the set of all real numbers” or “the set of all members of the United States Senate” are used, people’s various ideas of what is meant are sufficiently similar to make communication feasible.

We summarize briefly some of the things we shall simply assume about sets.

1. A set S is made up of **elements**, and if a is one of these elements, we shall denote this fact by $a \in S$.
2. There is exactly one set with no elements. It is the **empty set** and is denoted by \emptyset .
3. We may describe a set either by giving a characterizing property of the elements, such as “the set of all members of the United States Senate,” or by listing the elements. The standard way to describe a set by listing elements is to enclose the designations of the elements, separated by commas, in braces, for example, $\{1, 2, 15\}$. If a set is described by a characterizing property $P(x)$ of its elements x , the brace notation $\{x | P(x)\}$ is also often used, and is read “the set of all x such that the statement $P(x)$ about x is true.” Thus

$$\begin{aligned}\{2, 4, 6, 8\} &= \{x | x \text{ is an even whole positive number } \leq 8\} \\ &= \{2x | x = 1, 2, 3, 4\}.\end{aligned}$$

The notation $\{x | P(x)\}$ is often called “set-builder notation.”

4. A set is **well defined**, meaning that if S is a set and a is some object, then either a is definitely in S , denoted by $a \in S$, or a is definitely not in S , denoted by $a \notin S$. Thus, we should never say, “Consider the set S of some positive numbers,” for it is not definite whether $2 \in S$ or $2 \notin S$. On the other hand, we can consider the set T of all prime positive integers. Every positive integer is definitely either prime or not prime. Thus $5 \in T$ and $14 \notin T$. It may be hard to actually determine whether an object is in a set. For example, as this book goes to press it is probably unknown whether $2^{(2^{65})} + 1$ is in T . However, $2^{(2^{65})} + 1$ is certainly either prime or not prime.

It is not feasible for this text to push the definition of everything we use all the way back to the concept of a set. For example, we will never define the number π in terms of a set.

Every definition is an *if and only if* type of statement.

With this understanding, definitions are often stated with the *only if* suppressed, but it is always to be understood as part of the definition. Thus we may define an isosceles triangle as follows: “A triangle is **isosceles** if it has two congruent sides” when we really mean that a triangle is isosceles *if and only if* it has two congruent sides.

In our text, we have to define many terms. We use specifically labeled and numbered definitions for the main algebraic concepts with which we are concerned. To avoid an overwhelming quantity of such labels and numberings, we define many terms within the body of the text and exercises using boldface type.

Boldface Convention

A term printed in **boldface** in a sentence is being defined by that sentence.

Do not feel that you have to memorize a definition word for word. The important thing is to *understand* the concept, so that you can define precisely the same concept in your own words. Thus the definition “An **isosceles** triangle is one having two sides of equal length” is perfectly correct. Of course, we had to delay stating our boldface convention until we had finished using boldface in the preceding discussion of sets, because we do not define a set!

In this section, we do define some familiar concepts as sets, both for illustration and for review of the concepts. First we give a few definitions and some notation.

0.1 Definition A set B is a **subset of a set A** , denoted by $B \subseteq A$ or $A \supseteq B$, if every element of B is in A . The notations $B \subset A$ or $A \supset B$ will be used for $B \subseteq A$ but $B \neq A$. ■

Note that according to this definition, for any set A , A itself and \emptyset are both subsets of A .

0.2 Definition If A is any set, then A is the **improper subset of A** . Any other subset of A is a **proper subset of A** . ■

0.3 Example Let $S = \{1, 2, 3\}$. This set S has a total of eight subsets, namely \emptyset , $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{1, 2, 3\}$. ▲

0.4 Definition Let A and B be sets. The set $A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$ is the **Cartesian product of A and B** . ■

0.5 Example If $A = \{1, 2, 3\}$ and $B = \{3, 4\}$, then we have

$$A \times B = \{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4)\}.$$



Throughout this text, much work will be done involving familiar sets of numbers. Let us take care of notation for these sets once and for all.

\mathbb{Z} is the set of all integers (that is, whole numbers: positive, negative, and zero).

\mathbb{Q} is the set of all rational numbers (that is, numbers that can be expressed as quotients m/n of integers, where $n \neq 0$).

\mathbb{R} is the set of all real numbers.

\mathbb{Z}^+ , \mathbb{Q}^+ , and \mathbb{R}^+ are the sets of positive members of \mathbb{Z} , \mathbb{Q} , and \mathbb{R} , respectively.

\mathbb{C} is the set of all complex numbers.

\mathbb{Z}^* , \mathbb{Q}^* , \mathbb{R}^* , and \mathbb{C}^* are the sets of nonzero members of \mathbb{Z} , \mathbb{Q} , \mathbb{R} , and \mathbb{C} , respectively.

- 0.6 Example** The set $\mathbb{R} \times \mathbb{R}$ is the familiar Euclidean plane that we use in first-semester calculus to draw graphs of functions. ▲

Relations Between Sets

We introduce the notion of an element a of set A being *related* to an element b of set B , which we might denote by $a \mathcal{R} b$. The notation $a \mathcal{R} b$ exhibits the elements a and b in left-to-right order, just as the notation (a, b) for an element in $A \times B$. This leads us to the following definition of a relation \mathcal{R} as a *set*.

- 0.7 Definition** A **relation** between sets A and B is a subset \mathcal{R} of $A \times B$. We read $(a, b) \in \mathcal{R}$ as “ a is related to b ” and write $a \mathcal{R} b$. ■

- 0.8 Example** Let S be any set. We can define an **Equality Relation** = between S and itself as the subset $\{(x, x) | x \in S\}$. Of course, this is nothing new. It is simply the usual idea of what it means for two “things” to be equal. So if $x, y \in S$ are different elements, then they are not related by the equality relation and we write $x \neq y$, but if x and y are the same then we write $x = y$. ▲

We will refer to any relation between a set S and itself, as in the preceding example, as a **relation on S** .

- 0.9 Example** The graph of the function f where $f(x) = x^3$ for all $x \in \mathbb{R}$, is the subset $\{(x, x^3) | x \in \mathbb{R}\}$ of $\mathbb{R} \times \mathbb{R}$. Thus it is a relation on \mathbb{R} . The function is completely determined by its graph. ▲

The preceding example suggests that rather than define a “function” $y = f(x)$ to be a “rule” that assigns to each $x \in \mathbb{R}$ exactly one $y \in \mathbb{R}$, we can easily describe it as a certain type of subset of $\mathbb{R} \times \mathbb{R}$, that is, as a type of relation. We free ourselves from \mathbb{R} and deal with any sets X and Y .

- 0.10 Definition** A **function** ϕ mapping X into Y is a relation between X and Y with the property that each $x \in X$ appears as the first member of exactly one ordered pair (x, y) in ϕ . Such a function is also called a **map or mapping** of X into Y . We write $\phi : X \rightarrow Y$ and express $(x, y) \in \phi$ by $\phi(x) = y$. The **domain** of ϕ is the set X and the set Y is the **codomain** of ϕ . The **range** of ϕ is $\phi[X] = \{\phi(x) | x \in X\}$. ■

- 0.11 Example** We can view the addition of real numbers as a function $+$: $(\mathbb{R} \times \mathbb{R}) \rightarrow \mathbb{R}$, that is, as a mapping of $\mathbb{R} \times \mathbb{R}$ into \mathbb{R} . For example, the action of $+$ on $(2, 3) \in \mathbb{R} \times \mathbb{R}$ is given in function notation by $+(2, 3) = 5$. In set notation we write $((2, 3), 5) \in +$. Of course, our familiar notation is $2 + 3 = 5$. ▲

Cardinality

The number of elements in a set X is the **cardinality** of X and is often denoted by $|X|$. For example, we have $|\{2, 5, 7\}| = 3$. It will be important for us to know whether two sets have the same cardinality. If both sets are finite, there is no problem; we can simply count the elements in each set. But do \mathbb{Z} , \mathbb{Q} , and \mathbb{R} have the same cardinality?

To convince ourselves that two sets X and Y have the same cardinality, we try to exhibit a pairing of each x in X with only one y in Y in such a way that each element of Y is also used only once in this pairing. For the sets $X = \{2, 5, 7\}$ and $Y = \{?, !, \#\}$, the pairing

$$2 \leftrightarrow ?, \quad 5 \leftrightarrow \#, \quad 7 \leftrightarrow !$$

shows they have the same cardinality. Notice that we could also exhibit this pairing as $\{(2, ?), (5, \#), (7, !)\}$ which, as a subset of $X \times Y$, is a *relation* between X and Y . The pairing

1	2	3	4	5	6	7	8	9	10	...
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	...
0	-1	1	-2	2	-3	3	-4	4	-5	...

shows that the sets \mathbb{Z} and \mathbb{Z}^+ have the same cardinality. Such a pairing, showing that sets X and Y have the same cardinality, is a special type of relation \leftrightarrow between X and Y called a **one-to-one correspondence**. Since each element x of X appears precisely once in this relation, we can regard this one-to-one correspondence as a *function* with domain X . The range of the function is Y because each y in Y also appears in some pairing $x \leftrightarrow y$. We formalize this discussion in a definition.

0.12 Definition

*A function $\phi : X \rightarrow Y$ is **one-to-one** or **injective** if $\phi(x_1) = \phi(x_2)$ only when $x_1 = x_2$. The function ϕ is **onto** or **surjective** if the range of ϕ is Y . If ϕ is both injective and surjective, ϕ is said to be **bijective**. ■

If a subset of $X \times Y$ is a *one-to-one* function ϕ mapping X onto Y , then each $x \in X$ appears as the first member of exactly one ordered pair in ϕ and also each $y \in Y$ appears as the second member of exactly one ordered pair in ϕ . Thus if we interchange the first and second members of all ordered pairs (x, y) in ϕ to obtain a set of ordered pairs (y, x) , we get a subset of $Y \times X$, which gives a one-to-one function mapping Y onto X . This function is called the **inverse function** of ϕ , and is denoted by ϕ^{-1} . Summarizing, if ϕ maps X one-to-one onto Y and $\phi(x) = y$, then ϕ^{-1} maps Y one-to-one onto X , and $\phi^{-1}(y) = x$.

0.13 Definition

Two sets X and Y have the **same cardinality** if there exists a one-to-one function mapping X onto Y , that is, if there exists a one-to-one correspondence between X and Y . ■

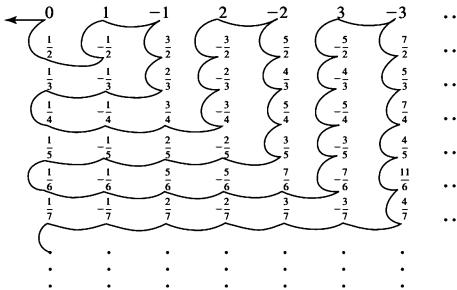
0.14 Example

The function $f : \mathbb{R} \rightarrow \mathbb{R}$ where $f(x) = x^2$ is not one-to-one because $f(2) = f(-2) = 4$ but $2 \neq -2$. Also, it is not onto \mathbb{R} because the range is the proper subset of all nonnegative numbers in \mathbb{R} . However, $g : \mathbb{R} \rightarrow \mathbb{R}$ defined by $g(x) = x^3$ is both one-to-one and onto \mathbb{R} . ▲

We showed that \mathbb{Z} and \mathbb{Z}^+ have the same cardinality. We denote this cardinal number by \aleph_0 , so that $|\mathbb{Z}| = |\mathbb{Z}^+| = \aleph_0$. It is fascinating that a proper subset of an infinite set may have the same number of elements as the whole set; an **infinite set** can be defined as a set having this property.

We naturally wonder whether all infinite sets have the same cardinality as the set \mathbb{Z} . A set has cardinality \aleph_0 if and only if *all* of its elements could be listed in an infinite row, so that we could “number them” using \mathbb{Z}^+ . Figure 0.15 indicates that this is possible for the set \mathbb{Q} . The square array of fractions extends infinitely to the right and infinitely

* We should mention another terminology, used by the disciples of N. Bourbaki, in case you encounter it elsewhere. In Bourbaki’s terminology, a one-to-one map is an **injection**, an onto map is a **surjection**, and a map that is both one-to-one and onto is a **bijection**.



0.15 Figure

downward, and contains all members of \mathbb{Q} . We have shown a string winding its way through this array. Imagine the fractions to be glued to this string. Taking the beginning of the string and pulling to the left in the direction of the arrow, the string strengthens out and all elements of \mathbb{Q} appear on it in an infinite row as $0, \frac{1}{2}, -\frac{1}{2}, 1, -1, \frac{3}{2}, \dots$. Thus $|\mathbb{Q}| = \aleph_0$ also.

If the set $S = \{x \in \mathbb{R} \mid 0 < x < 1\}$ has cardinality \aleph_0 , all its elements could be listed as unending decimals in a column extending infinitely downward, perhaps as

$$\begin{aligned} & 0.3659663426 \dots \\ & 0.7103958453 \dots \\ & 0.0358493553 \dots \\ & 0.9968452214 \dots \\ & \vdots \end{aligned}$$

We now argue that any such array must omit some number in S . Surely S contains a number r having as its n th digit after the decimal point a number different from 0, from 9, and from the n th digit of the n th number in this list. For example, r might start $.5637\dots$. The 5 rather than 3 after the decimal point shows r cannot be the first number in S listed in the array shown. The 6 rather than 1 in the second digit shows r cannot be the second number listed, and so on. Because we could make this argument with *any list*, we see that S has too many elements to be paired with those in \mathbb{Z}^+ . Exercise 15 indicates that \mathbb{R} has the same number of elements as S . We just denote the cardinality of \mathbb{R} by $|\mathbb{R}|$. Exercise 19 indicates that there are infinitely many different cardinal numbers even greater than $|\mathbb{R}|$.

Partitions and Equivalence Relations

Sets are **disjoint** if no two of them share a common element. In Example 0.17 we break up the integers into subsets. Eventually we will see how to define an algebraic structure on these subsets of \mathbb{Z} . That is, we will be able to “add” two of these subsets to get another subset. We will find that breaking a set into subsets is a valuable tool in a number of settings, so we conclude this section with a brief study of *partitions* of sets.

0.16 Definition A **partition** of a set S is a collection of nonempty subsets of S such that every element of S is in exactly one of the subsets. The subsets are the **cells** of the partition. ■

When discussing a partition of a set S , we denote by \bar{x} the cell containing the element x of S .

0.17 Example Splitting \mathbb{Z} into the subset of even integers and the subset of odd integers, we obtain a partition of \mathbb{Z} into the two cells listed below.

$$\bar{0} = \{\dots, -8, -6, -4, -2, 0, 2, 4, \dots\}$$

$$\bar{1} = \{\dots, -7, -5, -3, -1, 1, 3, 5, \dots\}$$

We can think of $\bar{0}$ as being the integers that are divisible by 2 and $\bar{1}$ as the integers that when divided by 2 yield a remainder of 1. This idea can be used for positive integers other than 2. For example, we can partition \mathbb{Z} into three cells:

$$\bar{0} = \{x \in \mathbb{Z} \mid x \text{ is a multiple of } 3\},$$

$$\bar{1} = \{x \in \mathbb{Z} \mid \text{the remainder of } x \text{ divided by } 3 \text{ is } 1\}, \text{ and}$$

$$\bar{2} = \{x \in \mathbb{Z} \mid \text{the remainder of } x \text{ divided by } 3 \text{ is } 2\}.$$

Note that when dividing a negative number by 3, we still obtain a non-negative remainder. For example, $-5 \div 3$ is -2 with remainder 1, which says that $\overline{-5} = \bar{1}$.

Generalizing, for each $n \in \mathbb{Z}^+$, we obtain a partition of \mathbb{Z} consisting of n cells, $\bar{0}, \bar{1}, \bar{2}, \dots, \bar{n-1}$. For each $0 \leq r \leq n-1$, an integer x is in the cell \bar{r} exactly when the remainder of $x \div n$ is r . These cells are the **residue classes modulo n** in \mathbb{Z} and n is called the **modulus**. We define the set $\mathbb{Z}/n\mathbb{Z}$ as the set containing the cells in this partition. So, for example, $\mathbb{Z}/3\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}\}$. As we can see, $\mathbb{Z}/n\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}, \dots, \bar{n-1}\}$ has exactly n elements. \blacktriangle

Each partition of a set S yields a relation \mathcal{R} on S in a natural way: namely, for $x, y \in S$, let $x \mathcal{R} y$ if and only if x and y are in the same cell of the partition. In set notation, we would write $x \mathcal{R} y$ as $(x, y) \in \mathcal{R}$ (see Definition 0.7). A bit of thought shows that this relation \mathcal{R} on S satisfies the three properties of an *equivalence relation* in the following definition.

0.18 Definition An **equivalence relation** \mathcal{R} on a set S is one that satisfies these three properties for all $x, y, z \in S$.

1. (Reflexive) $x \mathcal{R} x$.
2. (Symmetric) If $x \mathcal{R} y$, then $y \mathcal{R} x$.
3. (Transitive) If $x \mathcal{R} y$ and $y \mathcal{R} z$ then $x \mathcal{R} z$. \blacksquare

To illustrate why the relation \mathcal{R} corresponding to a partition of S satisfies the symmetric condition in the definition, we need only observe that if y is in the same cell as x (that is, if $x \mathcal{R} y$), then x is in the same cell as y (that is, $y \mathcal{R} x$). We leave the similar observations to verify the reflexive and transitive properties to Exercise 28.

0.19 Example For any nonempty set S , the equality relation $=$ defined by the subset $\{(x, x) \mid x \in S\}$ of $S \times S$ is an equivalence relation. \blacktriangle

0.20 Example (**Congruence Modulo n**) Let $n \in \mathbb{Z}^+$. The equivalence relation on \mathbb{Z} corresponding to the partition of \mathbb{Z} into residue classes modulo n , discussed in Example 0.17, is **congruence modulo n** . It is sometimes denoted by \equiv_n . Rather than write $a \equiv_n b$, we usually write $a \equiv b \pmod{n}$, read, “ a is congruent to b modulo n .” For example, we have $15 \equiv 27 \pmod{4}$ because both 15 and 27 have remainder 3 when divided by 4. \blacktriangle

0.21 Example Let a relation \mathcal{R} on the set \mathbb{Z} be defined by $n \mathcal{R} m$ if and only if $nm \geq 0$, and let us determine whether \mathcal{R} is an equivalence relation.

Reflexive $a \mathcal{R} a$, because $a^2 \geq 0$ for all $a \in \mathbb{Z}$.

Symmetric If $a \mathcal{R} b$, then $ab \geq 0$, so $ba \geq 0$ and $b \mathcal{R} a$.

Transitive If $a \mathcal{R} b$ and $b \mathcal{R} c$, then $ab \geq 0$ and $bc \geq 0$. Thus $ab^2c = acb^2 \geq 0$.

If we knew $b^2 > 0$, we could deduce $ac \geq 0$ whence $a \mathcal{R} c$. We have to examine the case $b = 0$ separately. A moment of thought shows that $-3 \mathcal{R} 0$ and $0 \mathcal{R} 5$, but we do not have $-3 \mathcal{R} 5$. Thus the relation \mathcal{R} is not transitive, and hence is not an equivalence relation. \blacktriangleleft

We observed above that a partition yields a natural equivalence relation. We now show that an equivalence relation on a set yields a natural partition of the set. The theorem that follows states both results for reference.

0.22 Theorem (Equivalence Relations and Partitions) Let S be a nonempty set and let \sim be an equivalence relation on S . Then \sim yields a partition of S , where

$$\bar{a} = \{x \in S \mid x \sim a\}.$$

Also, each partition of S gives rise to an equivalence relation \sim on S where $a \sim b$ if and only if a and b are in the same cell of the partition.

Proof We must show that the different cells $\bar{a} = \{x \in S \mid x \sim a\}$ for $a \in S$ do give a partition of S , so that every element of S is in some cell and so that if $a \in \bar{b}$, then $\bar{a} = \bar{b}$. Let $a \in S$. Then $a \in \bar{a}$ by the reflexive condition (1), so a is in at least one cell.

Suppose now that $a \in \bar{b}$. We need to show that $\bar{a} = \bar{b}$ as sets; this will show that a cannot be in more than one cell. There is a standard way to show that two sets are the same:

Show that each set is a subset of the other.

We show that $\bar{a} \subseteq \bar{b}$. Let $x \in \bar{a}$. Then $x \sim a$. But $a \in \bar{b}$, so $a \sim b$. Then, by the transitive condition (3), $x \sim b$, so $x \in \bar{b}$. Thus $\bar{a} \subseteq \bar{b}$. Now we show that $\bar{b} \subseteq \bar{a}$. Let $y \in \bar{b}$. Then $y \sim b$. But $a \in \bar{b}$, so $a \sim b$ and, by symmetry (2), $b \sim a$. Then by transitivity (3), $y \sim a$, so $y \in \bar{a}$. Hence $\bar{b} \subseteq \bar{a}$ also, so $\bar{b} = \bar{a}$ and our proof is complete. \blacklozenge

Each cell in the partition arising from an equivalence relation is an **equivalence class**.

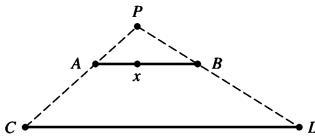
■ EXERCISES 0

In Exercises 1 through 4, describe the set by listing its elements.

- | | |
|---|--|
| 1. $\{x \in \mathbb{R} \mid x^2 = 3\}$ | 2. $\{m \in \mathbb{Z} \mid m^2 + m = 6\}$ |
| 3. $\{m \in \mathbb{Z} \mid mn = 60 \text{ for some } n \in \mathbb{Z}\}$ | 4. $\{x \in \mathbb{Z} \mid x^2 - 10x + 16 \leq 0\}$ |

In Exercises 5 through 10, decide whether the object described is indeed a set (is well defined). Give an alternate description of each set.

- | | |
|---|--|
| 5. $\{n \in \mathbb{Z}^+ \mid n \text{ is a large number}\}$ | 6. $\{n \in \mathbb{Z} \mid n^2 < 0\}$ |
| 7. $\{n \in \mathbb{Z} \mid 39 < n^3 < 57\}$ | |
| 8. $\{r \in \mathbb{Q} \mid \text{When } r \text{ is multiplied by a sufficiently large power of 2, one obtains a whole number.}\}$ | |
| 9. $\{x \in \mathbb{Z}^+ \mid x \text{ is an easy number to factor}\}$ | |
| 10. $\{x \in \mathbb{Q} \mid x \text{ may be written with positive denominator less than 4}\}$ | |
| 11. List the elements in $\{a, b, c\} \times \{1, 2, c\}$. | |



0.23 Figure

12. Let $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$. For each relation between A and B given as a subset of $A \times B$, decide whether it is a function mapping A into B . If it is a function, decide whether it is one-to-one and whether it is onto B .
- $\{(1, 2), (2, 6), (3, 4)\}$
 - $\{[1, 3] \text{ and } [5, 7]\}$
 - $\{(1, 6), (1, 2), (1, 4)\}$
 - $\{[2, 2], \{3, 6\}, \{1, 6\}\}$
 - $\{(1, 6), (2, 6), (3, 6)\}$
 - $\{(1, 2), \{2, 6\}\}$
13. Illustrate geometrically that two line segments AB and CD of different lengths have the same number of points by indicating in Fig. 0.23 what point y of CD might be paired with point x of AB .
14. Recall that for $a, b \in \mathbb{R}$ and $a < b$, the **closed interval** $[a, b]$ in \mathbb{R} is defined by $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$. Show that the given intervals have the same cardinality by giving a formula for a one-to-one function f mapping the first interval onto the second.
- $[0, 1]$ and $[0, 2]$
 - $[1, 3]$ and $[5, 7]$
 - $[a, b]$ and $[c, d]$
15. Show that $S = \{x \in \mathbb{R} \mid 0 < x < 1\}$ has the same cardinality as \mathbb{R} . [Hint: Find an elementary function of calculus that maps an interval one-to-one onto \mathbb{R} , and then translate and scale appropriately to make the domain the set S .]
- For any set A , we denote by $\mathcal{P}(A)$ the collection of all subsets of A . For example, if $A = \{a, b, c, d\}$, then $\{a, b, d\} \in \mathcal{P}(A)$. The set $\mathcal{P}(A)$ is the **power set** of A . Exercises 16 through 19 deal with the notion of the power set of a set A .
16. List the elements of the power set of the given set and give the cardinality of the power set.
- \emptyset
 - $\{a\}$
 - $\{a, b\}$
 - $\{a, b, c\}$
17. Let A be a finite set, and let $|A| = s$. Based on the preceding exercise, make a conjecture about the value of $|\mathcal{P}(A)|$. Then try to prove your conjecture.
18. For any set A , finite or infinite, let B^A be the set of all functions mapping A into the set $B = \{0, 1\}$. Show that the cardinality of B^A is the same as the cardinality of the set $\mathcal{P}(A)$. [Hint: Each element of B^A determines a subset of A in a natural way.]
19. Show that the power set of a set A , finite or infinite, has too many elements to be able to be put in a one-to-one correspondence with A . Explain why this intuitively means that there are an infinite number of infinite cardinal numbers. [Hint: Imagine a one-to-one function ϕ mapping A into $\mathcal{P}(A)$ to be given. Show that ϕ cannot be onto $\mathcal{P}(A)$ by considering, for each $x \in A$, whether $x \in \phi(x)$ and using this idea to define a subset S of A that is not in the range of ϕ . Is the set of everything a logically acceptable concept? Why or why not?]
20. Let $A = \{1, 2\}$ and let $B = \{3, 4, 5\}$.
- Illustrate, using A and B , why we consider that $2 + 3 = 5$. Use similar reasoning with sets of your own choice to decide what you would consider to be the value of
 - $3 + \aleph_0$
 - $\aleph_0 + \aleph_0$. - Illustrate why we consider that $2 \cdot 3 = 6$ by plotting the points of $A \times B$ in the plane $\mathbb{R} \times \mathbb{R}$. Use similar reasoning with a figure in the text to decide what you would consider to be the value of $\aleph_0 \cdot \aleph_0$.
 - How many numbers in the interval $0 \leq x \leq 1$ can be expressed in the form $\#\#$, where each $\#$ is a digit $0, 1, 2, 3, \dots, 9$? How many are there of the form $\#\#\#\#\#$? Following this idea, and Exercise 15, decide what you would consider to be the value of 10^{\aleph_0} . How about 12^{\aleph_0} and 2^{\aleph_0} ?

- 22.** Continuing the idea in the preceding exercise and using Exercises 18 and 19, use exponential notation to fill in the three blanks to give a list of five cardinal numbers, each of which is greater than the preceding one.

$$\aleph_0, |\mathbb{R}|, __, __, __.$$

In Exercises 23 through 27, find the number of different partitions of a set having the given number of elements.

23. 1 element

24. 2 elements

25. 3 elements

26. 4 elements

27. 5 elements

- 28.** Consider a partition of a set S . The paragraph following Definition 0.18 explained why the relation

$$x \mathcal{R} y \text{ if and only if } x \text{ and } y \text{ are in the same cell}$$

satisfies the symmetric condition for an equivalence relation. Write similar explanations of why the reflexive and transitive properties are also satisfied.

In Exercises 29 through 34, determine whether the given relation is an equivalence relation on the set. Describe the partition arising from each equivalence relation.

29. $n \mathcal{R} m$ in \mathbb{Z} if $nm > 0$

30. $x \mathcal{R} y$ in \mathbb{R} if $x \geq y$

31. $x \mathcal{R} y$ in \mathbb{Z}^+ if the greatest common divisor of x and y is greater than 1

32. $(x_1, y_1) \mathcal{R} (x_2, y_2)$ in $\mathbb{R} \times \mathbb{R}$ if $x_1^2 + y_1^2 = x_2^2 + y_2^2$

33. $n \mathcal{R} m$ in \mathbb{Z}^+ if n and m have the same number of digits in the usual base ten notation

34. $n \mathcal{R} m$ in \mathbb{Z}^+ if n and m have the same final digit in the usual base ten notation

35. Using set notation of the form $\{\dots, \#, \#, \#\dots\}$, write the residue classes modulo n in \mathbb{Z} as discussed in Example 0.17 for the indicated values of n .

a. 3

b. 4

c. 5

- 36.** Write each set by listing its elements.

a. $\mathbb{Z}/3\mathbb{Z}$

b. $\mathbb{Z}/4\mathbb{Z}$

c. $\mathbb{Z}/5\mathbb{Z}$

37. When discussing residue classes, $\bar{1}$ is not well defined until the modulus n is given. Explain.

38. Let $n \in \mathbb{Z}^+$ and let \sim be defined on \mathbb{Z} by $r \sim s$ if and only if $r - s$ is divisible by n , that is, if and only if $r - s = nq$ for some $q \in \mathbb{Z}$.

a. Show that \sim is an equivalence relation on \mathbb{Z} .

b. Show that this \sim is the equivalence relation, *congruence modulo n* , of Example 0.20.

39. Let $n \in \mathbb{Z}^+$. Using the relation from Exercise 38, show that if $a_1 \sim a_2$ and $b_1 \sim b_2$, then $(a_1 + b_1) \sim (a_2 + b_2)$.

40. Let $n \in \mathbb{Z}^+$. Using the relation from Exercise 38, show that if $a_1 \sim a_2$ and $b_1 \sim b_2$, then $(a_1 b_1) \sim (a_2 b_2)$.

41. Students often misunderstand the concept of a one-to-one function (mapping). I think I know the reason. You see, a mapping $\phi : A \rightarrow B$ has a *direction* associated with it, from A to B . It seems reasonable to expect a one-to-one mapping simply to be a mapping that carries one point of A into one point of B , in the direction indicated by the arrow. But of course, *every* mapping of A into B does this, and Definition 0.12 did not say that at all. With this unfortunate situation in mind, make as good a pedagogical case as you can for calling the functions described in Definition 0.12 *two-to-two functions* instead. (Unfortunately, it is almost impossible to get widely used terminology changed.)

This page is intentionally left blank

Groups and Subgroups

-
- Section 1** Binary Operations
 - Section 2** Groups
 - Section 3** Abelian Examples
 - Section 4** Nonabelian Examples
 - Section 5** Subgroups
 - Section 6** Cyclic Groups
 - Section 7** Generating Sets and Cayley Digraphs

SECTION 1

BINARY OPERATIONS

The transition from elementary school arithmetic to high school algebra involves using letters to represent unknown numbers and studying the basic properties of equations and expressions. The two main binary operations used in high school algebra are addition and multiplication. Abstract algebra takes the next step in abstraction. Not only are the variables unknown, but the actual operations involved may be unknown! We will study sets that have binary operations with properties similar to those of addition and multiplication of numbers. In Part I, our goal will be to develop some of the basic properties of a group. In this section we start our investigation of groups by defining binary operations, naming properties possessed by some binary operations, and giving examples.

Definitions and Examples

The first step in our journey to understand groups is to give a precise mathematical definition of a binary operation that generalizes the familiar addition and multiplication of numbers. Recall that for any set S , Definition 0.4 defines the set $S \times S$ to contain all ordered pairs (a, b) with $a, b \in S$.

1.1 Definition A **binary operation** $*$ on a set S is a function mapping $S \times S$ into S . For each $(a, b) \in S \times S$, we will denote the element $*((a, b))$ of S by $a * b$. ■

Intuitively, we may regard a binary operation $*$ on S as assigning, to each ordered pair (a, b) of elements of S , an element $a * b$ of S .

Binary refers to the fact that we are mapping *pairs* of elements from S into S . We could also define a ternary operation as a function mapping triples of elements of S to S , but we will have no need for this type of operation. Throughout this book we will often drop the term binary and use the term operation to mean binary operation.

1.2 Example Our usual addition $+$ is an operation on the set \mathbb{R} . Our usual multiplication \cdot is a different operation on \mathbb{R} . In this example, we could replace \mathbb{R} by any of the sets \mathbb{C} , \mathbb{Z} , \mathbb{R}^+ , or \mathbb{Z}^+ . ▲

Note that we require an operation on a set S to be defined for *every* ordered pair (a, b) of elements from S .

1.3 Example Let $M(\mathbb{R})$ be the set of all matrices[†] with real entries. The usual matrix addition $+$ is *not* an operation on this set since $A + B$ is not defined for an ordered pair (A, B) of matrices having different numbers of rows or of columns. ▲

Sometimes an operation on S provides an operation on a subset H of S also. We make a formal definition.

1.4 Definition Let $*$ be an operation on S and let H be a subset of S . The subset H is **closed under $*$** if for all $a, b \in H$ we also have $a * b \in H$. In this case, the operation on H given by restricting $*$ to H is the **induced operation** of $*$ on H . ■

By our very definition of an operation $*$ on S , the set S is closed under $*$, but a subset may not be, as the following example shows.

1.5 Example Our usual addition $+$ on the set \mathbb{R} of real numbers does not induce an operation on the set \mathbb{R}^* of nonzero real numbers because $2 \in \mathbb{R}^*$ and $-2 \in \mathbb{R}^*$, but $2 + (-2) = 0$ and $0 \notin \mathbb{R}^*$. Thus \mathbb{R}^* is not closed under $*$. ▲

In our text, we will often have occasion to decide whether a subset H of S is closed under a binary operation $*$ on S . To arrive at a correct conclusion, *we have to know what it means for an element to be in H* , and to use this fact. Students often have trouble here. Be sure you understand the next example.

1.6 Example Let $+$ and \cdot be the usual operations of addition and multiplication on the set \mathbb{Z} , and let $H = \{n^2 | n \in \mathbb{Z}^+\}$. Determine whether H is closed under (a) addition and (b) multiplication.

For part (a), we need only observe that $1^2 = 1$ and $2^2 = 4$ are in H , but that $1 + 4 = 5$ and $5 \notin H$. Thus H is not closed under addition.

For part (b), suppose that $r \in H$ and $s \in H$. Using what it means for r and s to be in H , we see that there must be integers n and m in \mathbb{Z}^+ such that $r = n^2$ and $s = m^2$. Consequently, $rs = n^2m^2 = (nm)^2$. By the characterization of elements in H and the fact that $nm \in \mathbb{Z}^+$, this means that $rs \in H$, so H is closed under multiplication. ▲

1.7 Example Let F be the set of all real-valued functions f having as domain the set \mathbb{R} of real numbers. We are familiar from calculus with the operations $+, -, \cdot$, and \circ on F . Namely, for each ordered pair (f, g) of functions in F , we define for each $x \in \mathbb{R}$

$$\begin{aligned} f + g \text{ by } (f + g)(x) &= f(x) + g(x) && \text{addition,} \\ f - g \text{ by } (f - g)(x) &= f(x) - g(x) && \text{subtraction,} \\ f \cdot g \text{ by } (f \cdot g)(x) &= f(x)g(x) && \text{multiplication, and} \\ f \circ g \text{ by } (f \circ g)(x) &= f(g(x)) && \text{composition.} \end{aligned}$$

All four of these functions are again real valued with domain \mathbb{R} , so F is closed under all four operations $+, -, \cdot$, and \circ . ▲

The operations described in the examples above are very familiar to you. In this text, we want to *abstract* basic structural concepts from our familiar algebra. To empha-

[†] Most students of abstract algebra have studied linear algebra and are familiar with matrices and matrix operations. For the benefit of those students, examples involving matrices are often given. The reader who is not familiar with matrices can either skip all references to them or turn to the Appendix at the back of the text, where there is a short summary.

size this concept of *abstraction* from the familiar, we should illustrate these structural concepts with unfamiliar examples.

The most important method of describing a particular binary operation $*$ on a given set is to characterize the element $a * b$ assigned to each pair (a, b) by some property defined in terms of a and b .

1.8 Example On \mathbb{Z}^+ , we define an operation $*$ by $a * b$ equals the smaller of a and b , or the common value if $a = b$. Thus $2 * 11 = 2$; $15 * 10 = 10$; and $3 * 3 = 3$. \blacktriangle

1.9 Example On \mathbb{Z}^+ , we define an operation $*'$ by $a *' b = a$. Thus $2 *' 3 = 2$; $25 *' 10 = 25$; and $5 *' 5 = 5$. \blacktriangle

1.10 Example On \mathbb{Z}^+ , we define an operation $*''$ by $a *'' b = (a * b) + 2$, where $*$ is defined in Example 1.8. Thus $4 *'' 7 = 6$; $25 *'' 9 = 11$; and $6 *'' 6 = 8$. \blacktriangle

It may seem that these examples are of no importance, but in fact they are used millions of times each day. For example, consider the GPS navigational system installed in most cars and cell phones. It searches alternative driving routes, computes the travel time, and determines which route takes less time. The operation in Example 1.8 is programmed into modern GPS systems and it plays an essential role.

Examples 1.8 and 1.9 were chosen to demonstrate that an operation may or may not depend on the order of the given pair. Thus in Example 1.8, $a * b = b * a$ for all $a, b \in \mathbb{Z}^+$, and in Example 1.9 this is not the case, for $5 *' 7 = 5$ but $7 *' 5 = 7$.

1.11 Definition An operation $*$ on a set S is **commutative** if (and only if) $a * b = b * a$ for all $a, b \in S$. \blacksquare

As was pointed out in Section 0, it is customary in mathematics to omit the words *and only if* from a definition. Definitions are always understood to be if and only if statements. *Theorems are not always if and only if statements, and no such convention is ever used for theorems.*

Now suppose we wish to consider an expression of the form $a * b * c$. A binary operation $*$ enables us to combine only two elements, and here we have three. The obvious attempts to combine the three elements are to form either $(a * b) * c$ or $a * (b * c)$. With $*$ defined as in Example 1.8, $(2 * 5) * 9$ is computed by $2 * 5 = 2$ and then $2 * 9 = 2$. Likewise, $2 * (5 * 9)$ is computed by $5 * 9 = 5$ and then $2 * 5 = 2$. Hence $(2 * 5) * 9 = 2 * (5 * 9)$, and it is not hard to see that for this $*$,

$$(a * b) * c = a * (b * c),$$

so there is no ambiguity in writing $a * b * c$. But for $*''$ of Example 1.10,

$$(2 *'' 5) *'' 9 = 4 *'' 9 = 6,$$

while

$$2 *'' (5 *'' 9) = 2 *'' 7 = 4.$$

Thus $(a *'' b) *'' c$ need not equal $a *'' (b *'' c)$, and the expression $a *'' b *'' c$ is ambiguous.

1.12 Definition An operation on a set S is **associative** if $(a * b) * c = a * (b * c)$ for all $a, b, c \in S$. \blacksquare

It can be shown that if $*$ is associative, then longer expressions such as $a * b * c * d$ are not ambiguous. Parentheses may be inserted in any fashion for purposes of computation; the final results of two such computations will be the same.

Composition of functions mapping \mathbb{R} into \mathbb{R} was reviewed in Example 1.7. For any set S and any functions f and g mapping S into S , we similarly define the composition $f \circ g$ of g followed by f as the function mapping S into S such that $(f \circ g)(x) = f(g(x))$ for all $x \in S$. Some of the most important binary operations we consider are defined using composition of functions. It is important to know that function composition is always associative whenever it is defined.

1.13 Theorem (Associativity of Composition) Let S be a set and let f , g , and h be functions mapping S into S . Then $f \circ (g \circ h) = (f \circ g) \circ h$.

Proof To show these two functions are equal, we must show that they give the same assignment to each $x \in S$. Computing we find that

$$(f \circ (g \circ h))(x) = f((g \circ h)(x)) = f(g(h(x)))$$

and

$$((f \circ g) \circ h)(x) = (f \circ g)(h(x)) = f(g(h(x))),$$

so the same element $f(g(h(x)))$ of S is indeed obtained. ◆

As an example of using Theorem 1.13 to save work, recall that it is a fairly painful exercise in summation notation to show that multiplication of $n \times n$ matrices is an associative operation. If, in a linear algebra course, we first show that there is a one-to-one correspondence between matrices and linear transformations and that multiplication of matrices corresponds to the composition of the linear transformations (functions), we obtain this associativity at once from Theorem 1.13.

There is another property that an operation on a set may have that is of particular importance in algebra. The numbers 0 and 1 play special roles as real numbers because for any real number a , $a + 0 = a$ and $a \times 1 = a$. Because of these properties, 0 is called the *additive identity* in \mathbb{R} and 1 is called the *multiplicative identity* in \mathbb{R} . In general we have the following definition of an identity.

1.14 Definition Let S be a set with binary operation $*$. If $e \in S$ has the property that for all $a \in S$, $a * e = e * a = a$, then e is called an **identity element for $*$** . ■

We included both conditions $a * e = a$ and $e * a = a$ in the definition of an identity because we are not assuming that the operation on S is commutative. Of course, if the operation is commutative, such as $+$ and \times on the real numbers, then we would only have to check one of the conditions and the other follows from commutativity.

1.15 Theorem (Uniqueness of Identity) A set with binary operation $*$ has at most one identity element.

Proof We need to show that there cannot be two different identity elements. To do this, we assume that both e and e' are identities and show that $e = e'$. Consider the element $e * e'$. Since e is an identity, $e * e' = e'$. But $e * e' = e$ because e' is also an identity. Therefore $e = e'$. ◆

1.16 Example Continuing with Example 1.7, let F be the set of all functions mapping the real numbers to the real numbers. We verify that the function defined by $\iota(x) = x$ is the identity for the operation function composition. Let $f \in F$. Then $f \circ \iota(x) = f(\iota(x)) = f(x)$ and $\iota \circ f(x) = \iota(f(x)) = f(x)$.

The function $m(x) = 1$ is the identity for the operation function multiplication, $a(x) = 0$ is the identity for function addition, but function subtraction has no identity element. ▲

The last property that we consider in this section is the existence of inverse elements. For addition, the inverse of a real number a is $-a$. Using multiplication, the inverse of a nonzero real number a is $\frac{1}{a}$. We now give the formal definition of an inverse for an element x .

1.17 Definition If $*$ is an operation on the set S and S has an identity e , then for any $x \in S$, the inverse of x is an element x' such that $x * x' = x' * x = e$. ■

1.18 Example Continuing Example 1.16, let F be the set of functions mapping the real numbers to the real numbers with operation function composition. We have two definitions for the inverse of a function $f \in F$, the usual definition of an inverse function and Definition 1.17. The two definitions match since both say that an inverse for f is a function f' such that $f \circ f' = f' \circ f = \iota$. So $f \in F$ has an inverse if and only if f is one-to-one and onto. ▲

Tables

For a finite set, a binary operation on the set can be defined by means of a table in which the elements of the set are listed across the top as heads of columns and at the left side as heads of rows. We always require that the elements of the set be listed as heads across the top in the same order as heads down the left side. The next example illustrates the use of a table to define a binary operation.

1.19 Example Table 1.20 defines the binary operation $*$ on $S = \{a, b, c\}$ by the following rule:

$$(i\text{th entry on the left}) * (j\text{th entry on the top})$$

$$= (\text{entry in the } i\text{th row and } j\text{th column of the table body}).$$

1.20 Table

*	a	b	c
a	b	c	b
b	a	c	b
c	c	b	a

Thus $a * b = c$ and $b * a = a$, so $*$ is not commutative. ▲

We can easily see that a binary operation defined by a table is commutative if and only if the entries in the table are symmetric with respect to the diagonal that starts at the upper left corner of the table and terminates at the lower right corner.

1.21 Example Complete Table 1.22 so that $*$ is a commutative operation on the set $S = \{a, b, c, d\}$.

Solution From Table 1.22, we see that $b * a = d$. For $*$ to be commutative, we must have $a * b = d$ also. Thus we place d in the appropriate square defining $a * b$, which is located symmetrically across the diagonal in Table 1.23 from the square defining $b * a$. We obtain the rest of Table 1.23 in this fashion to give our solution. ▲

1.22 Table

*	a	b	c	d
a	b			
b	d	a		
c	a	c	d	
d	a	b	b	c

1.23 Table

*	a	b	c	d
a	b	d	a	a
b	d	a	c	b
c	a	c	d	b
d	a	b	b	c

- 1.24 Example** When an operation has an identity element, it is customary to put the identity first in the list of heads. This makes the first row and the first column match the head row and head column as seen in Table 1.25. ▲

1.25 Table

*	e	a	b	c
e	e	a	b	c
a	a	c	c	a
b	b	a	e	c
c	c	e	a	b

Some Words of Warning

Classroom experience shows the chaos that may result if a student is given a set and asked to define some binary operation on it. Remember that in an attempt to define a binary operation $*$ on a set S we must be sure that

1. exactly one element is assigned to each possible ordered pair of elements of S ,
2. for each ordered pair of elements of S , the element assigned to it is again in S .

Regarding Condition 1, a student will often make an attempt that assigns an element of S to “most” ordered pairs, but for a few pairs, determines no element. In this event, $*$ is **not everywhere defined** on S . It may also happen that for some pairs, the attempt could assign any of several elements of S , that is, there is ambiguity. In any case of ambiguity, $*$ is **not well defined**. If Condition 2 is violated, then S is **not closed under $*$** .

- 1.26 Example** On which of the sets \mathbb{Q} , \mathbb{Q}^* , and \mathbb{Z}^+ does the formula $a * b = a/b$ define an operation? Note that this formula does not make sense in the case that $b = 0$. So for example, $2 * 0 = 2/0$ is not defined, which means Condition 1 is not satisfied. So $*$ is not an operation on \mathbb{Q} .

If we throw out 0, we do have an operation on \mathbb{Q}^* since both Conditions 1 and 2 are satisfied. That is, for any $a, b \in \mathbb{Q}^*$, $a * b = a/b$ is a nonzero rational number.

The set \mathbb{Z}^+ does not include 0, but there is another issue. For example, $1 * 2 = 1/2 \notin \mathbb{Z}^+$, which means that Condition 2 is violated and $*$ is not an operation on \mathbb{Z}^+ . ▲

Following are several illustrations of attempts to define operations on sets. Some of them need some work! The symbol $*$ is used for the attempted operation in all these examples.

- 1.27 Example** Let F be the set of all real-valued functions with domain \mathbb{R} as in Example 1.7. Suppose we “define” $*$ to give the usual quotient of f by g , that is, $f * g = h$, where $h(x) = f(x)/g(x)$. Here Condition 2 is violated, for the functions in F are defined for *all* real numbers, and for some $g \in F$, $g(x)$ will be zero for some values of x in \mathbb{R} and $h(x)$ would not be defined at those numbers in \mathbb{R} . For example, if $f(x) = \cos x$ and $g(x) = x^2$, then $h(0)$ is undefined, so $h \notin F$. ▲

- 1.28 Example** Let F be as in Example 1.27 and let $f * g = h$, where h is the function greater than both f and g . This “definition” is extremely vague. In the first place, we have not defined what it means for one function to be greater than another. Even if we had, any sensible definition would result in there being many functions greater than both f and g , and $*$ would still be *not well defined*. ▲

- 1.29 Example** Let S be a set consisting of 20 people, no two of whom are of the same height. Define $*$ by $a * b = c$, where c is the tallest person among the 20 in S . This is a perfectly good binary operation on the set, although not a particularly interesting one. ▲

- 1.30 Example** Let S be as in Example 1.29 and let $a * b = c$, where c is the shortest person in S who is taller than both a and b . This $*$ is *not everywhere defined*, since if either a or b is the tallest person in the set, $a * b$ is not determined. ▲

■ EXERCISES 1

Computations

Exercises 1 through 4 concern the binary operation $*$ defined on $S = \{a, b, c, d, e\}$ by means of Table 1.31.

1. Compute $b * d, c * c$, and $[(a * c) * e] * a$.
2. Compute $(a * b) * c$ and $a * (b * c)$. Can you say on the basis of this computation whether $*$ is associative?
3. Compute $(b * d) * c$ and $b * (d * c)$. Can you say on the basis of this computation whether $*$ is associative?
4. Is $*$ commutative? Why?
5. Complete Table 1.32 so as to define a commutative binary operation $*$ on $S = \{a, b, c, d\}$.
6. Table 1.33 can be completed to define an associative binary operation $*$ on $S = \{a, b, c, d\}$. Assume this is possible and compute the missing entries. Does S have an identity element?

1.31 Table

*	a	b	c	d	e
a	a	b	c	b	d
b	b	c	a	e	c
c	c	a	b	b	a
d	b	e	b	e	d
e	d	b	a	d	c

1.32 Table

*	a	b	c	d
a	a	b	c	
b	b	d		c
c	c	a	d	b
d	d			a

1.33 Table

*	a	b	c	d
a	a	b	c	d
b	b	a	c	d
c	c	d	c	d
d				

In Exercises 7 through 11, determine whether the operation $*$ is associative, whether the operation is commutative, and whether the set has an identity element.

7. $*$ defined on \mathbb{Z} by letting $a * b = a - b$
8. $*$ defined on \mathbb{Q} by letting $a * b = 2ab + 3$
9. $*$ defined on \mathbb{Z} by letting $a * b = ab + a + b$
10. $*$ defined on \mathbb{Z}^+ by letting $a * b = 2^{ab}$
11. $*$ defined on \mathbb{Z}^+ by letting $a * b = a^b$
12. Let S be a set having exactly one element. How many different binary operations can be defined on S ? Answer the question if S has exactly 2 elements; exactly 3 elements; exactly n elements.
13. How many different commutative binary operations can be defined on a set of 2 elements? on a set of 3 elements? on a set of n elements?
14. How many different binary operations on a set S with n elements have the property that for all $x \in S$, $x * x = x$?
15. How many different binary operations on a set S with n elements have an identity element?

Concepts

In Exercises 16 through 19, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

16. A binary operation $*$ is *commutative* if and only if $a * b = b * a$.
17. A binary operation $*$ on a set S is *associative* if and only if, for all $a, b, c \in S$, we have $(b * c) * a = b * (c * a)$.
18. A subset H of a set S is *closed* under a binary operation $*$ on S if and only if $(a * b) \in H$ for all $a, b \in S$.
19. An identity in the set S with operation $*$ is an element $e \in S$ such that $a * e = e * a = a$.
20. Is there an example of a set S , a binary operation on S , and two different elements $e_1, e_2 \in S$ such that for all $a \in S$, $e_1 * a = a$ and $a * e_2 = a$? If so, give an example and if not, prove there is not one.

In Exercises 21 through 26, determine whether the definition of $*$ does give a binary operation on the set. In the event that $*$ is not a binary operation, state whether Condition 1, Condition 2, or both conditions regarding defining binary operations are violated.

21. On \mathbb{Z}^+ , define $a * b = b^a$.
22. On \mathbb{R}^+ , define $*$ by letting $a * b = 2a - b$.
23. On \mathbb{R}^+ , define $*$ by $a * b$ to be the minimum of a and $b - 1$ if they are different and their common value if they are the same.
24. On \mathbb{R} , define $a * b$ to be the number c so that $c^b = a$.
25. On \mathbb{Z}^+ , define $*$ by letting $a * b = c$, where c is at least 5 more than $a + b$.
26. On \mathbb{Z}^+ , define $*$ by letting $a * b = c$, where c is the largest integer less than the product of a and b .
27. Let H be the subset of $M_2(\mathbb{R})$ consisting of all matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ for $a, b \in \mathbb{R}$. Is H closed under
 - a. matrix addition?
 - b. matrix multiplication?
28. Determine whether each of the following is true or false.
 - a. If $*$ is any binary operation on any set S , then $a * a = a$ for all $a \in S$.
 - b. If $*$ is any commutative binary operation on any set S , then $a * (b * c) = (b * c) * a$ for all $a, b, c \in S$.
 - c. If $*$ is any associative binary operation on any set S , then $a * (b * c) = (b * c) * a$ for all $a, b, c \in S$.
 - d. The only binary operations of any importance are those defined on sets of numbers.
 - e. A binary operation $*$ on a set S is commutative if there exist $a, b \in S$ such that $a * b = b * a$.
 - f. Every binary operation defined on a set having exactly one element is both commutative and associative.
 - g. A binary operation on a set S assigns at least one element of S to each ordered pair of elements of S .
 - h. A binary operation on a set S assigns at most one element of S to each ordered pair of elements of S .
 - i. A binary operation on a set S assigns exactly one element of S to each ordered pair of elements of S .
 - j. A binary operation on a set S may assign more than one element of S to some ordered pair of elements of S .
 - k. For any binary operation $*$ on the set S , if $a, b, c \in S$ and $a * b = a * c$, then $b = c$.
 - l. For any binary operation $*$ on the set S , there is an element $e \in S$ such that for all $x \in S$, $x * e = x$.
 - m. There is an operation on the set $S = \{e_1, e_2, a\}$ so that for all $x \in S$, $e_1 * x = e_2 * x = x$.
 - n. Identity elements are always called e .
29. Give a set different from any of those described in the examples of the text and not a set of numbers. Define two different binary operations $*$ and $*'$ on this set. Be sure that your set is *well defined*.

Theory

30. Prove that if $*$ is an associative and commutative binary operation on a set S , then

$$(a * b) * (c * d) = [(d * c) * a] * b$$

for all $a, b, c, d \in S$. Assume the associative law only for triples as in the definition, that is, assume only

$$(x * y) * z = x * (y * z)$$

for all $x, y, z \in S$.

In Exercises 31 and 32, either prove the statement or give a counterexample.

31. Every binary operation on a set consisting of a single element is both commutative and associative.
32. Every commutative binary operation on a set having just two elements is associative.

Let F be the set of all real-valued functions having as domain the set \mathbb{R} of all real numbers. Example 1.7 defined the binary operations $+$, $-$, \cdot , and \circ on F . In Exercises 33 through 41, either prove the given statement or give a counterexample.

33. Function addition $+$ on F is associative.
34. Function subtraction $-$ on F is commutative.

35. Function subtraction – on F is associative.
 36. Under function subtraction – F has an identity.
 37. Under function multiplication \cdot F has an identity.
 38. Function multiplication \cdot on F is commutative.
 39. Function multiplication \cdot on F is associative.
 40. Function composition \circ on F is commutative.
 41. If $*$ and $*'$ are any two binary operations on a set S , then

$$a * (b *' c) = (a * b) *' (a * c) \quad \text{for all } a, b, c \in S.$$

42. Suppose that $*$ is an *associative binary* operation on a set S . Let $H = \{a \in S \mid a * x = x * a \text{ for all } x \in S\}$. Show that H is closed under $*$. (We think of H as consisting of all elements of S that *commute* with every element in S .)
 43. Suppose that $*$ is an associative and commutative binary operation on a set S . Show that $H = \{a \in S \mid a * a = a\}$ is closed under $*$. (The elements of H are **idempotents** of the binary operation $*$.)
 44. Let S be a set and let $*$ be a binary operation on S satisfying the two laws

- $x * x = x$ for all $x \in S$, and
- $(x * y) * z = (y * z) * x$ for all $x, y, z \in S$.

Show that $*$ is associative and commutative. (This is problem B-1 on the 1971 Putnam Competition.)

SECTION 2 GROUPS

In high school algebra, one of the key objectives is to learn how to solve equations. Even before learning algebra, students in elementary school are given problems like $5 + \square = 2$ or $2 \times \square = 3$, which become $5 + x = 2$ and $2x = 3$ in high school algebra. Let us closely examine the steps we use to solve these equations:

$$\begin{aligned} 5 + x &= 2, && \text{given,} \\ -5 + (5 + x) &= -5 + 2, && \text{adding } -5, \\ (-5 + 5) + x &= -5 + 2, && \text{associative law,} \\ 0 + x &= -5 + 2, && \text{computing } -5 + 5, \\ x &= -5 + 2, && \text{property of 0,} \\ x &= -3, && \text{computing } -5 + 2. \end{aligned}$$

Strictly speaking, we have not shown here that -3 is a solution, but rather that it is the only possibility for a solution. To show that -3 is a solution, one merely computes $5 + (-3)$. A similar analysis could be made for the equation $2x = 3$ in the rational numbers with the operation of multiplication:

$$\begin{aligned} 2x &= 3, && \text{given,} \\ \frac{1}{2}(2x) &= \frac{1}{2}(3), && \text{multiplying by } \frac{1}{2}, \\ (\frac{1}{2} \cdot 2)x &= \frac{1}{2}3, && \text{associative law,} \\ 1 \cdot x &= \frac{1}{2}3, && \text{computing } \frac{1}{2}2, \\ x &= \frac{1}{2}3, && \text{property of 1,} \\ x &= \frac{3}{2}, && \text{computing } \frac{1}{2}3. \end{aligned}$$

Now suppose that we have a set with a binary operation $*$. What properties does the operation need to have in order to solve an equation of the form $a * x = b$ where a and b are fixed elements of S ? Both equations $5 + x = 2$ and $2x = 3$ have this form; the

first uses the operation $+$, and the second uses the operation \times . By examining the steps used we can see what properties of the operation $*$ are required as summarized in the table below.

Property	$+$	\times
Associative Property	$-5 + (5 + x) = (-5 + 5) + x$	$\frac{1}{2}(2x) = (\frac{1}{2} \cdot 2)x$
Identity Element	$0: 0 + x = x$	$1: 1 \cdot x = x$
Inverse Element	$-5: -5 + 5 = 0$	$\frac{1}{2}: \frac{1}{2} \cdot 2 = 1$

If S is a set with an operation $*$ satisfying these three properties, then an equation of the form $a * x = b$ could be solved for x using exactly the same steps used to solve $5 + x = 2$ or $2x = 3$. These three essential properties are all that is required in order to have a group. We are now ready to present the precise definition.

Definition and Examples

2.1 Definition A **group** $\langle G, * \rangle$ is a set G , closed under a binary operation $*$, such that the following axioms are satisfied:

\mathcal{G}_1 : For all $a, b, c \in G$, we have

$$(a * b) * c = a * (b * c). \quad \text{associativity of } *$$

\mathcal{G}_2 : There is an element e in G such that for all $x \in G$,

$$e * x = x * e = x. \quad \text{identity element } e \text{ for } *$$

\mathcal{G}_3 : Corresponding to each $a \in G$, there is an element a' in G such that

$$a * a' = a' * a = e. \quad \text{inverse } a' \text{ of } a$$

2.2 Example $\langle \mathbb{R}, + \rangle$ is a group with identity element 0 and the inverse of any real number a is $-a$. However, $\langle \mathbb{R}, \cdot \rangle$ is not a group since 0 has no multiplicative inverse. We were still able to solve $2x = 3$ in the example above because $\langle \mathbb{R}^*, \cdot \rangle$ is a group since multiplication of real numbers is associative, 1 is an identity, and every real number except 0 has an inverse. ▲

It is often convenient to say that G is a group under the operation $*$ rather than write $\langle G, * \rangle$ is a group. At times, there is only one obvious operation that makes $\langle G, * \rangle$ a group. In this case, we may abuse notation and say that G is a group. For example, if we say that \mathbb{R} is a group, we mean that \mathbb{R} is a group under addition.

2.3 Definition A group G is **abelian** if its binary operation is commutative. ■

Let us give some examples of some sets with binary operations that give groups and also of some that do not give groups.

2.4 Example The set \mathbb{Z}^+ under addition is *not* a group. There is no identity element for $+$ in \mathbb{Z}^+ . ▲

2.5 Example The set of all nonnegative integers (including 0) under addition is still *not* a group. There is an identity element 0, but no inverse for 2. ▲

■ HISTORICAL NOTE

There are three historical roots of the development of abstract group theory evident in the mathematical literature of the nineteenth century: the theory of algebraic equations, number theory, and geometry. All three of these areas used group-theoretic methods of reasoning, although the methods were considerably more explicit in the first area than in the other two.

One of the central themes of geometry in the nineteenth century was the search for invariants under various types of geometric transformations. Gradually attention became focused on the transformations themselves, which in many cases can be thought of as elements of groups.

In number theory, already in the eighteenth century Leonhard Euler had considered the remainders on division of powers a^n by a fixed prime p . These remainders have “group” properties.

Similarly, Carl F. Gauss, in his *Disquisitiones Arithmeticae* (1800), dealt extensively with quadratic forms $ax^2 + 2bxy + cy^2$, and in particular showed that equivalence classes of these forms under composition possessed what amounted to group properties.

Finally, the theory of algebraic equations provided the most explicit prefiguring of the group concept. Joseph-Louis Lagrange (1736–1813) in fact initiated the study of permutations of the roots of an equation as a tool for solving it. These permutations, of course, were ultimately considered as elements of a group.

It was Walther von Dyck (1856–1934) and Heinrich Weber (1842–1913) who in 1882 were able independently to combine the three historical roots and give clear definitions of the notion of an abstract group.

2.6 Example The familiar additive properties of integers and of rational, real, and complex numbers show that \mathbb{Z} , \mathbb{Q} , \mathbb{R} , and \mathbb{C} under addition are abelian groups. ▲

2.7 Example The set \mathbb{Z}^+ under multiplication is *not* a group. There is an identity 1, but no inverse of 3. ▲

■ HISTORICAL NOTE

Commutative groups are called *abelian* in honor of the Norwegian mathematician Niels Henrik Abel (1802–1829). Abel was interested in the question of solvability of polynomial equations. In a paper written in 1828, he proved that if all the roots of such an equation can be expressed as rational functions f, g, \dots, h of one of them, say x , and if for any two of these roots, $f(x)$ and $g(x)$, the relation $f(g(x)) = g(f(x))$ always holds, then the equation is solvable by radicals. Abel showed that each of these functions in fact permutes the roots of the equation; hence, these functions are elements of the group of permutations of the roots. It was this property of commutativity in these permutation groups associated with solvable equations that led Camille Jordan in his 1870 treatise on algebra to name such groups *abelian*; the name since

then has been applied to commutative groups in general.

Abel was attracted to mathematics as a teenager and soon surpassed all his teachers in Norway. He finally received a government travel grant to study elsewhere in 1825 and proceeded to Berlin, where he befriended August Crelle, the founder of the most influential German mathematical journal. Abel contributed numerous papers to Crelle’s *Journal* during the next several years, including many in the field of elliptic functions, whose theory he created virtually single-handedly. Abel returned to Norway in 1827 with no position and an abundance of debts. He nevertheless continued to write brilliant papers, but died of tuberculosis at the age of 26, two days before Crelle succeeded in finding a university position for him in Berlin.

2.8 Example The familiar multiplicative properties of rational, real, and complex numbers show that the sets \mathbb{Q}^+ and \mathbb{R}^+ of positive numbers and the sets \mathbb{Q}^* , \mathbb{R}^* , and \mathbb{C}^* of nonzero numbers under multiplication are abelian groups. \blacktriangle

2.9 Example The set of all real-valued functions with domain \mathbb{R} under function addition is a group. This group is abelian. \blacktriangle

2.10 Example (**Linear Algebra**) Those who have studied vector spaces should note that the axioms for a vector space V pertaining just to vector addition can be summarized by asserting that V under vector addition is an abelian group. \blacktriangle

2.11 Example The set $M_{m \times n}(\mathbb{R})$ of all $m \times n$ matrices under matrix addition is a group. The $m \times n$ matrix with all entries 0 is the identity matrix. This group is abelian. \blacktriangle

2.12 Example The set $M_n(\mathbb{R})$ of all $n \times n$ matrices under matrix multiplication is *not* a group. The $n \times n$ matrix with all entries 0 has no inverse. \blacktriangle

Each of the groups we have seen in the above examples is an abelian group. There are many examples of groups which are not abelian, two of which we now present.

2.13 Example Here we give an example of a group that is not abelian. We let T be the set of all isometries of the plane. An **isometry of the plane** is a function mapping the plane to the plane which preserves distance. So if ϕ is an isometry of the plane and P, Q are points in the plane, then the distance between P and Q is the same as the distance between $\phi(P)$ and $\phi(Q)$. Isometries of the plane map the plane one-to-one and onto itself. Examples of isometries include translations and rotations of the plane. The set T under the operation of composition forms a group. To verify this we first must check that function composition is an operation. Certainly, the composition of two isometries is an isometry since each preserves distance. So function composition gives an operation on T . Theorem 1.13 states that function composition is associative, so \mathcal{G}_1 is satisfied. The identity function ι that maps each point P in the plane to itself gives an identity element in T , which means that \mathcal{G}_2 is satisfied. Finally, for any isometry ϕ , the inverse function ϕ^{-1} is also an isometry and it serves as an inverse as defined in \mathcal{G}_3 . Therefore T is a group under function composition.

To show that T is not abelian, we only need to find two isometries ϕ and θ such that $\phi \circ \theta \neq \theta \circ \phi$. The functions $\phi(x, y) = (-x, y)$ (reflection across the y -axis) and $\theta(x, y) = (-y, x)$ (rotation by $\pi/2$ about the origin) foot the bill. Note that $\phi \circ \theta(1, 0) = \phi(\theta(1, 0)) = \phi(0, 1) = (0, 1)$ and $\theta \circ \phi(1, 0) = \theta(\phi(1, 0)) = \theta(-1, 0) = (0, -1)$, which implies that $\phi \circ \theta \neq \theta \circ \phi$ and T is not an abelian group under function composition. \blacktriangle

2.14 Example Show that the subset S of $M_n(\mathbb{R})$ consisting of all *invertible* $n \times n$ matrices under matrix multiplication is a group.

Solution We start by showing that S is closed under matrix multiplication. Let A and B be in S , so that both A^{-1} and B^{-1} exist and $AA^{-1} = BB^{-1} = I_n$. Then

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = I_n,$$

so that AB is invertible and consequently is also in S .

Since matrix multiplication is associative and I_n acts as the identity element, and since each element of S has an inverse by definition of S , we see that S is indeed a group. This group is *not* commutative. \blacktriangle

The group of invertible $n \times n$ matrices described in the preceding example is of fundamental importance in linear algebra. It is the **general linear group of degree n** ,

and is usually denoted by $GL(n, \mathbb{R})$. Those of you who have studied linear algebra know that a matrix A in $GL(n, \mathbb{R})$ gives rise to an invertible linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$, defined by $T(\mathbf{x}) = A\mathbf{x}$, and that conversely, every invertible linear transformation of \mathbb{R}^n into itself is defined in this fashion by some matrix in $GL(n, \mathbb{R})$. Also, matrix multiplication corresponds to composition of linear transformations. Thus all invertible linear transformations of \mathbb{R}^n into itself form a group under function composition; this group is usually denoted by $GL(\mathbb{R}^n)$. Since the sets $GL(\mathbb{R}^n)$ and $GL(n, \mathbb{R})$ and their operations are essentially the same, we say that the two groups are *isomorphic*. We give a formal definition later in this section.

We conclude our list of examples of groups with one that may seem a bit contrived. We include it to show that there are many ways to define groups and to illustrate the steps needed to verify that a given set with an operation is a group.

2.15 Example Let $*$ be defined on \mathbb{Q}^+ by $a * b = ab/2$. Then

$$(a * b) * c = \frac{ab}{2} * c = \frac{abc}{4},$$

and likewise

$$a * (b * c) = a * \frac{bc}{2} = \frac{abc}{4}.$$

Thus $*$ is associative. Computation shows that

$$2 * a = a * 2 = a$$

for all $a \in \mathbb{Q}^+$, so 2 is an identity element for $*$. Finally,

$$a * \frac{4}{a} = \frac{4}{a} * a = 2,$$

so $a' = 4/a$ is an inverse for a . Hence \mathbb{Q}^+ with the operation $*$ is a group. ▲

Elementary Properties of Groups

As we proceed to prove our first theorem about groups, we must use Definition 2.1, which is the only thing we know about groups at the moment. The proof of a second theorem can employ both Definition 2.1 and the first theorem; the proof of a third theorem can use the definition and the first two theorems, and so on.

Our first theorem will establish cancellation laws. In real number arithmetic, we know that $2a = 2b$ implies that $a = b$. We need only divide both sides of the equation $2a = 2b$ by 2, or equivalently, multiply both sides by $\frac{1}{2}$, which is the multiplicative inverse of 2. We parrot this proof to establish cancellation laws for any group. Note that we will also use the associative law.

2.16 Theorem If G is a group with binary operation $*$, then the **left and right cancellation laws** hold in G , that is, $a * b = a * c$ implies $b = c$, and $b * a = c * a$ implies $b = c$ for all $a, b, c \in G$.

Proof Suppose $a * b = a * c$. Then by \mathcal{G}_3 , there exists a' , and

$$a' * (a * b) = a' * (a * c).$$

By the associative law,

$$(a' * a) * b = (a' * a) * c.$$

By the definition of a' in \mathcal{G}_3 , $a' * a = e$, so

$$e * b = e * c.$$

By the definition of e in \mathcal{G}_2 ,

$$b = c.$$

Similarly, from $b * a = c * a$ one can deduce that $b = c$ upon multiplication on the right by a' and use of the axioms for a group. \blacklozenge

Our next proof can make use of Theorem 2.16. We show that a “linear equation” in a group has a *unique* solution. Recall that we chose our group properties to allow us to find solutions of such equations.

2.17 Theorem If G is a group with binary operation $*$, and if a and b are any elements of G , then the linear equations $a * x = b$ and $y * a = b$ have unique solutions x and y in G .

Proof First we show the existence of *at least* one solution by just computing that $a' * b$ is a solution of $a * x = b$. Note that

$$\begin{aligned} a * (a' * b) &= (a * a') * b, && \text{associative law,} \\ &= e * b, && \text{definition of } a', \\ &= b, && \text{property of } e. \end{aligned}$$

Thus $x = a' * b$ is a solution of $a * x = b$. In a similar fashion, $y = b * a'$ is a solution of $y * a = b$.

To show uniqueness of y , we use the standard method of assuming that we have two solutions, y_1 and y_2 , so that $y_1 * a = b$ and $y_2 * a = b$. Then $y_1 * a = y_2 * a$, and by Theorem 2.16, $y_1 = y_2$. The uniqueness of x follows similarly. \blacklozenge

Of course, to prove the uniqueness in the last theorem, we could have followed the procedure we used in motivating the definition of a group, showing that if $a * x = b$, then $x = a' * b$. However, we chose to illustrate the standard way to prove an object is unique; namely, suppose you have two such objects, and then prove they must be the same. Note that the solutions $x = a' * b$ and $y = b * a'$ need not be the same unless $*$ is commutative.

Because a group has a binary operation, we know from Theorem 1.15 that the identity e in a group is unique. We state this again as part of the next theorem for easy reference.

2.18 Theorem In a group G with binary operation $*$, there is only one element e in G such that

$$e * x = x * e = x$$

for all $x \in G$. Likewise for each $a \in G$, there is only one element a' in G such that

$$a' * a = a * a' = e.$$

In summary, the identity element and inverse of each element are unique in a group.

Proof Theorem 1.15 shows that an identity element for any binary operation is unique. No use of the other group axioms was required to show this.

Turning to the uniqueness of an inverse, suppose that $a \in G$ has inverses a' and a'' so that $a' * a = a * a' = e$ and $a'' * a = a * a'' = e$. Then

$$a * a'' = a * a' = e$$

and, by Theorem 2.16,

$$a'' = a',$$

so the inverse of a in a group is unique. \blacklozenge

Note that in a group G , we have

$$(a * b) * (b' * a') = a * (b * b') * a' = (a * e) * a' = a * a' = e.$$

This equation and Theorem 2.18 show that $b' * a'$ is the unique inverse of $a * b$. That is, $(a * b)' = b' * a'$. We state this as a corollary.

2.19 Corollary Let G be a group. For all $a, b \in G$, we have $(a * b)' = b' * a'$. ◆

For your information, we remark that binary algebraic structures with weaker axioms than those for a group have also been studied quite extensively. Of these weaker structures, the **semigroup**, a set with an associative binary operation, has perhaps had the most attention. A **monoid** is a semigroup that has an identity element for the binary operation. Note that every group is both a semigroup and a monoid.

Finally, it is possible to give axioms for a group $\langle G, *\rangle$ that seem at first glance to be weaker, namely:

1. The binary operation $*$ on G is associative.
2. There exists a **left identity element** e in G such that $e * x = x$ for all $x \in G$.
3. For each $a \in G$, there exists a **left inverse** a' in G such that $a' * a = e$.

From this *one-sided definition*, one can prove that the left identity element is also a right identity element, and a left inverse is also a right inverse for the same element. Thus these axioms should not be called *weaker*, since they result in exactly the same structures being called groups. It is conceivable that it might be easier in some cases to check these *left axioms* than to check our *two-sided axioms*. Of course, by symmetry it is clear that there are also *right axioms* for a group.

Group Isomorphisms

All our examples have been of infinite groups, that is, groups where the set G has an infinite number of elements. We turn to finite groups, starting with the smallest finite sets.

Since a group has to have at least one element, namely, the identity, a minimal set that might give rise to a group is a one-element set $\{e\}$. The only possible binary operation $*$ on $\{e\}$ is defined by $e * e = e$. The three group axioms hold. The identity element is always its own inverse in every group.

There is a group with only two elements, namely $G = \{1, -1\}$ with operation the usual multiplication. It is clear that G is closed under multiplication and we know that multiplication is associative. Furthermore, 1 is the identity, the inverse of 1 is 1, and the inverse of -1 is -1 . Table 2.20 is the group table for G .

Is this the only group with exactly two elements? To see, let us try to put a group structure on a set with two elements. Since one of the elements must be the identity, we will label the identity element e and we will label the other element a . Following tradition, we place the identity first both on the top and to the left as in the following table.

*	e	a
e		
a		

Since e is to be the identity,

$$e * x = x * e = x$$

2.20 Table

*	1	-1
1	1	-1
-1	-1	1

for all $x \in \{e, a\}$. We are forced to fill in the table as follows, if $*$ is to give a group:

*	e	a
e	e	a
a	a	

Also, a must have an inverse a' such that

$$a * a' = a' * a = e.$$

2.21 Table

*	e	a
e	e	a
a	a	e

In our case, a' must be either e or a . Since $a' = e$ obviously does not work, we must have $a' = a$, so we have to complete the table as shown in Table 2.21.

All the group axioms are now satisfied, except possibly associativity. But if we relabel 1 as e and -1 as a in Table 2.20 we obtain Table 2.21. Therefore, the table we constructed for $\{e, a\}$ must also satisfy \mathcal{G}_1 , the associative property. The table also shows clearly that properties \mathcal{G}_2 and \mathcal{G}_3 are satisfied, so $(\{e, a\}, *)$ is a group. The groups $\{1, -1\}$ and $\{e, a\}$ are not the same, but they are essentially the same since by relabeling elements of one with the names of the other, the operations match. When the elements of one group can be matched with another in such a way that the operations are the same, we say that the groups are **isomorphic** and the matching is called a **group isomorphism**. We showed that any group with two elements is isomorphic with $\{1, -1\}$ under multiplication. The notation used to indicate isomorphism is \cong , so we could write $(\{1, -1\}, \times) \cong (\{e, a\}, *)$. Of course the matching is a one-to-one function from one group onto the other. If we were only interested in groups whose tables are easy to compute, then we would not need a more precise definition for isomorphism. We would simply see if we can relabel one group table to make it look like the other. However, in the case of infinite groups or even groups with more than a few elements, we need a better way to verify that groups are isomorphic. We now give a more precise definition of a group isomorphism.

2.22 Definition Let $\langle G_1, *_1 \rangle$ and $\langle G_2, *_2 \rangle$ be groups and $f : G_1 \rightarrow G_2$. We say that f is a **group isomorphism** if the following two conditions are satisfied.

1. The function f is one-to-one and maps onto G_2 .
2. For all $a, b \in G_1$, $f(a *_1 b) = f(a) *_2 f(b)$. ■

Note that Condition 1 simply gives a way to relabel the elements of G_1 with elements in G_2 . Condition 2, which we will refer to as the **homomorphism property**, says that with this relabeling, the operations $*_1$ on G_1 and $*_2$ on G_2 match. If we are in the context of groups, we will often use the term isomorphism to mean group isomorphism. If there is an isomorphism from a group G_1 to G_2 , we say that G_1 is **isomorphic** with (or to) G_2 . In Exercise 44, you are asked to show that if $f : G_1 \rightarrow G_2$ is an isomorphism, then $f^{-1} : G_2 \rightarrow G_1$, the inverse function, is also an isomorphism. So if G_1 is isomorphic with G_2 , then G_2 is isomorphic with G_1 . If you wish to verify that two groups, G_1 and G_2 , are isomorphic, you can either construct an isomorphism mapping G_1 to G_2 or one mapping G_2 to G_1 .

2.23 Example In Exercise 10 you will be asked to show that $2\mathbb{Z}$, the even integers, forms a group under addition. Here we show \mathbb{Z} and $2\mathbb{Z}$ are isomorphic groups. In this case, the operations on the groups are both addition. We need a function $f : \mathbb{Z} \rightarrow 2\mathbb{Z}$ that is both one-to-one and onto $2\mathbb{Z}$. Let $f : \mathbb{Z} \rightarrow 2\mathbb{Z}$ be given by $f(m) = 2m$. We need to verify Condition 1 for an isomorphism, which says that f is one-to-one and onto. Suppose that $a, b \in \mathbb{Z}$ and $f(a) = f(b)$. Then $2a = 2b$, which implies that $a = b$, so f is one-to-one. We now show f

is onto. Let $y \in 2\mathbb{Z}$. Since y is even, $y = 2c$ for some $c \in \mathbb{Z}$. Therefore, $y = 2c = f(c)$, so f maps onto $2\mathbb{Z}$. We now turn our attention to the homomorphism property and consider arbitrary $a, b \in \mathbb{Z}$. Then

$$f(a + b) = 2(a + b) = 2a + 2b = f(a) + f(b),$$

which verifies Condition 2. Therefore f is a group isomorphism and \mathbb{Z} and $2\mathbb{Z}$ are isomorphic groups.

As noted above, we could have defined an isomorphism by using the inverse function $f^{-1} : 2\mathbb{Z} \rightarrow \mathbb{Z}$, which is defined by $f^{-1}(x) = x/2$. \blacktriangle

Properties of Group Tables

With Table 2.21 as background, we should be able to list some necessary conditions that a table giving a binary operation on a finite set must satisfy for the operation to give a group structure on the set. There must be one element of the set, which we may as well denote by e , that acts as the identity element. The condition $e * x = x$ means that the row of the table opposite e at the extreme left must contain exactly the elements appearing across the very top of the table in the same order. Similarly, the condition $x * e = x$ means that the column of the table under e at the very top must contain exactly the elements appearing at the extreme left in the same order. The fact that every element a has a right and a left inverse means that in the row having a at the extreme left, the element e must appear, and in the column under a at the very top, the e must appear. Thus e must appear in each row and in each column. We can do even better than this, however. By Theorem 2.17, not only do the equations $a * x = e$ and $y * a = e$ have unique solutions, but also the equations $a * x = b$ and $y * a = b$. By a similar argument, this means that *each element b of the group must appear once and only once in each row and each column of the table*.

Suppose conversely that a table for a binary operation on a finite set is such that there is an element acting as identity and that in each row and each column, each element of the set appears exactly once. Then it can be seen that the structure is a group structure if and only if the associative law holds. If a binary operation $*$ is given by a table, the associative law is usually messy to check. If the operation $*$ is defined by some characterizing property of $a * b$, the associative law is often easy to check. Fortunately, this second case turns out to be the one usually encountered.

We saw that there was essentially only one group of two elements in the sense that if the elements are denoted by e and a with the identity element e appearing first, the table must be as shown in Table 2.21. Suppose that a set has three elements. As before, we may as well let the set be $\{e, a, b\}$. For e to be an identity element, a binary operation $*$ on this set has to have a table of the form shown in Table 2.24. This leaves four places to be filled in. You can quickly see that Table 2.24 must be completed as shown in Table 2.25 if each row and each column are to contain each element exactly once. We find a group whose table is the same as Table 2.25. The elements of the group are the three matrices

$$e = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, a = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}, \text{ and } b = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}.$$

We let $G = \{e, a, b\}$. In Exercise 18 you will show that G is a group under matrix multiplication. By computing matrix products it is easy to check that the group table for G is identical with Table 2.25. Therefore Table 2.25 gives a group.

Now suppose that G' is any other group of three elements and imagine a table for G' with identity element appearing first. Since our filling out of the table for $G = \{e, a, b\}$ could be done in only one way, we see that if we take the table for G' and rename the identity e , the next element listed a , and the last element b , the resulting table for G' must be the same as the one we had for G . As explained above, this renaming gives an isomorphism of the group G' with the group G . Thus our work above can be summarized

2.24 Table

*	e	a	b
e	e	a	b
a	a		
b	b		

2.25 Table

*	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

by saying that all groups with a single element are isomorphic, all groups with just two elements are isomorphic, and all groups with just three elements are isomorphic. We use the phrase *up to isomorphism* to express this identification. Thus we may say, “There is only one group of three elements, up to isomorphism.”

An interesting problem in group theory is to determine up to isomorphism all the groups with a given number of elements n . In Exercise 20, you will be asked to show that there are up to isomorphism exactly two groups of order 4. It is beyond the scope of this book to give a thorough investigation of this problem, but we will solve the problem for some other special values of n in later sections.

■ EXERCISES 2

Computations

In Exercises 1 through 9, determine whether the binary operation $*$ gives a group structure on the given set. If no group results, give the first axiom in the order $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ from Definition 2.1 that does not hold.

1. Let $*$ be defined on \mathbb{Z} by letting $a * b = ab$.
2. Let $*$ be defined on $2\mathbb{Z} = \{2n \mid n \in \mathbb{Z}\}$ by letting $a * b = a + b$.
3. Let $*$ be defined on \mathbb{R}^+ by letting $a * b = \sqrt{ab}$.
4. Let $*$ be defined on \mathbb{Q} by letting $a * b = ab$.
5. Let $*$ be defined on the set \mathbb{R}^* of nonzero real numbers by letting $a * b = a/b$.
6. Let $*$ be defined on \mathbb{C} by letting $a * b = |ab|$.
7. Let $*$ be defined on the set $\{a, b\}$ by Table 2.26.
8. Let $*$ be defined on the set $\{a, b\}$ by Table 2.27.
9. Let $*$ be defined on the set $\{e, a, b\}$ by Table 2.28.

2.26 Table

$*$	a	b
a	a	b
b	b	b

2.27 Table

$*$	a	b
a	a	b
b	a	b

2.28 Table

$*$	e	a	b
e	e	a	b
a	a	e	b
b	b	b	e

10. Let n be a positive integer and let $n\mathbb{Z} = \{nm \mid m \in \mathbb{Z}\}$.

- a. Show that $(n\mathbb{Z}, +)$ is a group.
- b. Show that $(n\mathbb{Z}, +) \cong (\mathbb{Z}, +)$.

In Exercises 11 through 18, determine whether the given set of matrices under the specified operation, matrix addition or multiplication, is a group. Recall that a **diagonal matrix** is a square matrix whose only nonzero entries lie on the **main diagonal**, from the upper left to the lower right corner. An **upper-triangular matrix** is a square matrix with only zero entries below the main diagonal. Associated with each $n \times n$ matrix A is a number called the determinant of A , denoted by $\det(A)$. If A and B are both $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$. Also, $\det(I_n) = 1$ and A is invertible if and only if $\det(A) \neq 0$.

11. All $n \times n$ diagonal matrices under matrix addition.
12. All $n \times n$ diagonal matrices under matrix multiplication.
13. All $n \times n$ diagonal matrices with no zero diagonal entry under matrix multiplication.
14. All $n \times n$ diagonal matrices with all diagonal entries 1 or -1 under matrix multiplication.
15. All $n \times n$ upper-triangular matrices under matrix multiplication.
16. All $n \times n$ upper-triangular matrices under matrix addition.
17. All $n \times n$ upper-triangular matrices with determinant 1 under matrix multiplication.

18. The set of 2×2 matrices $G = \{e, a, b\}$ where $e = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $a = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$, and $b = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$ under matrix multiplication.

19. Let S be the set of all real numbers except -1 . Define $*$ on S by

$$a * b = a + b + ab.$$

- a. Show that $*$ gives a binary operation on S .
- b. Show that $(S, *)$ is a group.
- c. Find the solution of the equation $2 * x * 3 = 7$ in S .

20. This exercise shows that there are two nonisomorphic group structures on a set of 4 elements.

Let the set be $\{e, a, b, c\}$, with e the identity element for the group operation. A group table would then have to start in the manner shown in Table 2.29. The square indicated by the question mark cannot be filled in with a . It must be filled in either with the identity element e or with an element different from both e and a . In this latter case, it is no loss of generality to assume that this element is b . If this square is filled in with e , the table can then be completed in two ways to give a group. Find these two tables. (You need not check the associative law.) If this square is filled in with b , then the table can only be completed in one way to give a group. Find this table. (Again, you need not check the associative law.) Of the three tables you now have, two give isomorphic groups. Determine which two tables these are, and give the one-to-one onto relabeling function which is an isomorphism.

- a. Are all groups of 4 elements commutative?
- b. Find a way to relabel the four matrices

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$$

so the matrix multiplication table is identical to one you constructed. This shows that the table you constructed defines an associative operation and therefore gives a group.

- c. Show that for a particular value of n , the group elements given in Exercise 14 can be relabeled so their group table is identical to one you constructed. This implies the operation in the table is also associative.

21. According to Exercise 12 of Section 1, there are 16 possible binary operations on a set of 2 elements. How many of these give a structure of a group? How many of the 19,683 possible binary operations on a set of 3 elements give a group structure?

Concepts

22. Consider our axioms G_1 , G_2 , and G_3 for a group. We gave them in the order $G_1 G_2 G_3$. Conceivable other orders to state the axioms are $G_1 G_3 G_2$, $G_2 G_1 G_3$, $G_2 G_3 G_1$, $G_3 G_1 G_2$, and $G_3 G_2 G_1$. Of these six possible orders, exactly three are acceptable for a definition. Which orders are not acceptable, and why? (Remember this. Most instructors ask the student to define a group on at least one test.)

2.29 Table

*	e	a	b	c
e	e	a	b	c
a	a	$?$		
b	b			
c	c			

23. The following “definitions” of a group are taken verbatim, including spelling and punctuation, from papers of students who wrote a bit too quickly and carelessly. Criticize them.

- a. A group G is a set of elements together with a binary operation $*$ such that the following conditions are satisfied

* is associative

There exists $e \in G$ such that

$$e * x = x * e = x = \text{identity}.$$

For every $a \in G$ there exists an a' (inverse) such that

$$a \cdot a' = a' \cdot a = e$$

- b.** A group is a set G such that

The operation on G is associative.

there is an identity element (e) in G .

for every $a \in G$, there is an a' (inverse for each element)

- c.** A group is a set with a binary operation such

the binary operation is defined

an inverse exists

an identity element exists

- d.** A set G is called a group over the binary operation * such that for all $a, b \in G$

Binary operation * is associative under addition

there exist an element $\{e\}$ such that

$$a * e = e * a = e$$

For every element a there exists an element a' such that

$$a * a' = a' * a = e$$

- 24.** Give a table defining an operation satisfying axioms \mathcal{G}_2 and \mathcal{G}_3 in the definition of a group, but not satisfying axiom \mathcal{G}_1 for the set

- a.** $\{e, a, b\}$
- b.** $\{e, a, b, c\}$

- 25.** Mark each of the following true or false.

- _____ **a.** A group may have more than one identity element.
- _____ **b.** Any two groups of three elements are isomorphic.
- _____ **c.** In a group, each linear equation has a solution.
- _____ **d.** The proper attitude toward a definition is to memorize it so that you can reproduce it word for word as in the text.
- _____ **e.** Any definition a person gives for a group is correct provided that everything that is a group by that person's definition is also a group by the definition in the text.
- _____ **f.** Any definition a person gives for a group is correct provided he or she can show that everything that satisfies the definition satisfies the one in the text and conversely.
- _____ **g.** Every finite group of at most three elements is abelian.
- _____ **h.** An equation of the form $a * x * b = c$ always has a unique solution in a group.
- _____ **i.** The empty set can be considered a group.
- _____ **j.** Every group is a binary algebraic structure.

Proof synopsis

We give an example of a proof synopsis. Here is a one-sentence synopsis of the proof that the inverse of an element a in a group $(G, *)$ is unique.

Assuming that $a * a' = e$ and $a * a'' = e$, apply the left cancellation law to the equation $a * a' = a * a''$.

Note that we said “the left cancellation law” and not “Theorem 2.16.” We always suppose that our synopsis was given as an explanation given during a conversation at lunch, with no reference to text numbering and as little notation as is practical.

26. Give a one-sentence synopsis of the proof of the left cancellation law in Theorem 2.16.
27. Give at most a two-sentence synopsis of the proof in Theorem 2.17 that an equation $ax = b$ has a unique solution in a group.

Theory

28. An element $a \neq e$ in a group is said to have order 2 if $a * a = e$. Prove that if G is a group and $a \in G$ has order 2, then for any $b \in G$, $b' * a * b$ also has order 2.
29. Show that if G is a finite group with identity e and with an even number of elements, then there is $a \neq e$ in G such that $a * a = e$.
30. Let \mathbb{R}^* be the set of all real numbers except 0. Define $*$ on \mathbb{R}^* by letting $a * b = |a|b$.
- Show that $*$ gives an associative binary operation on \mathbb{R}^* .
 - Show that there is a left identity for $*$ and a right inverse for each element in \mathbb{R}^* .
 - Is \mathbb{R}^* with this binary operation a group?
 - Explain the significance of this exercise.
31. If $*$ is a binary operation on a set S , an element x of S is an **idempotent for $*$** if $x * x = x$. Prove that a group has exactly one idempotent element. (You may use any theorems proved so far in the text.)
32. Show that every group G with identity e and such that $x * x = e$ for all $x \in G$ is abelian. [Hint: Consider $(a * b) * (a * b)$.]
33. Let G be an abelian group and let $c^n = c * c * \dots * c$ for n factors c , where $c \in G$ and $n \in \mathbb{Z}^+$. Give a mathematical induction proof that $(a * b)^n = (a^n) * (b^n)$ for all $a, b \in G$.
34. Suppose that G is a group and $a, b \in G$ satisfy $a * b = b * a'$ where as usual, a' is the inverse for a . Prove that $b * a = a' * b$.
35. Suppose that G is a group and a and b are elements of G that satisfy $a * b = b * a^3$. Rewrite the element $(a * b)^2$ in the form $b^k a^l$. (See Exercise 33 for power notation.)
36. Let G be a group with a finite number of elements. Show that for any $a \in G$, there exists an $n \in \mathbb{Z}^+$ such that $a^n = e$. See Exercise 33 for the meaning of a^n . [Hint: Consider $e, a, a^2, a^3, \dots, a^m$, where m is the number of elements in G , and use the cancellation laws.]
37. Show that if $(a * b)^2 = a^2 * b^2$ for a and b in a group G , then $a * b = b * a$. See Exercise 33 for the meaning of a^2 .
38. Let G be a group and let $a, b \in G$. Show that $(a * b)' = a' * b'$ if and only if $a * b = b * a$.
39. Let G be a group and suppose that $a * b * c = e$ for $a, b, c \in G$. Show that $b * c * a = e$ also.
40. Prove that a set G , together with a binary operation $*$ on G satisfying the left axioms 1, 2, and 3 given after Corollary 2.19, is a group.
41. Prove that a nonempty set G , together with an associative binary operation $*$ on G such that

$$a * x = b \text{ and } y * a = b \text{ have solutions in } G \text{ for all } a, b \in G,$$

- is a group. [Hint: Use Exercise 40.]
42. Let G be a group. Prove that $(a')' = a$.
43. Let $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be an isometry of the plane.
- Prove that ϕ is a one-to-one function.
 - Prove that ϕ maps onto \mathbb{R}^2 .
44. Prove that if $f : G_1 \rightarrow G_2$ is a group isomorphism from the group $\langle G_1, *_1 \rangle$ to the group $\langle G_2, *_2 \rangle$, then $f^{-1} : G_2 \rightarrow G_1$ is also a group isomorphism.
45. Suppose that G is a group with n elements and $A \subseteq G$ has more than $\frac{n}{2}$ elements. Prove that for every $g \in G$, there exists $a, b \in A$ such that $a * b = g$. (This was Problem B-2 on the 1968 Putnam exam.)

SECTION 3

ABELIAN EXAMPLES

In this section we introduce two families of abelian groups and one special abelian group. These groups will be very useful in our study of groups in that they provide examples we can use to help understand concepts and test conjectures. Furthermore, we will see that some of them arise frequently in the study of groups.

We start by defining the set $\mathbb{Z}_n = \{0, 1, 2, 3, \dots, n - 1\}$, the first $n - 1$ positive integers together with 0, which makes a total of n elements. To define an operation $+_n$ on \mathbb{Z}_n , we let $a, b \in \mathbb{Z}_n$. Then

$$a +_n b = \begin{cases} a + b & \text{if } a + b < n \\ a + b - n & \text{if } a + b \geq n \end{cases}.$$

Note that for any $a, b \in \mathbb{Z}_n$, $0 \leq a + b \leq 2n - 2$, so $0 \leq a +_n b \leq n - 1$ is an operation which we call **addition modulo n** . Addition modulo n is clearly commutative: $a +_n b = b +_n a$ for any $a, b \in \mathbb{Z}_n$. The number 0 is an identity, the inverse of $a \in \mathbb{Z}_n$ is $n - a$ for $a \neq 0$, and the inverse of 0 is 0. To show that $(\mathbb{Z}_n, +_n)$ is an abelian group, it only remains to show that $+_n$ is associative. Although it is not difficult to show directly that $+_n$ is associative, it is a little tedious, so we defer the proof until we develop the circle group and then use properties of that group to conclude that $(\mathbb{Z}_n, +_n)$ is an abelian group.

3.1 Example For $n = 1$, $\mathbb{Z}_1 = \{0\}$, which is the trivial group with just one element. For $n = 2$, $\mathbb{Z}_2 = \{0, 1\}$, which as we saw in Section 2 is isomorphic with $\{1, -1\}$ under multiplication. It is important to note that completely different operations on sets can still define isomorphic groups. We also saw in Section 2 that any group with exactly three elements is isomorphic with any other group with exactly three elements. Therefore \mathbb{Z}_3 under addition modulo 3 is isomorphic with the group consisting of the three matrices

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \right\}$$

under matrix multiplication. Again we see that two groups can be isomorphic, but have completely different sets and operations. ▲

3.2 Example

Let us look more closely at the group table for \mathbb{Z}_4 , Table 3.3. We see that the inverse for 0 is 0, the inverse for 1 is $4 - 1 = 3$, and the inverse for 2 is $4 - 2 = 2$. In Exercise 20 in Section 2, you were asked to show that there are two groups with exactly four elements. The other group is the **Klein 4-group** denoted V , which stands for Vier, German for “four.” The group table for V is displayed as Table 3.4. How can we tell that the two groups \mathbb{Z}_4 and V are not isomorphic? We could try all possible one-to-one functions from \mathbb{Z}_4 onto K_4 to see if any of them make the table for \mathbb{Z}_4 look like the table for K_4 . This is tedious, so instead we look for a sneaky method. Notice that the diagonal entries of the table for K_4 are all the identity. No matter how we relabel

3.3 Table

\mathbb{Z}_4 :	$+_4$	0	1	2	3
	0	0	1	2	3
	1	1	2	3	0
	2	2	3	0	1
	3	3	0	1	2

3.4 Table

V :	*	e	a	b	c
	e	e	a	b	c
	a	a	e	c	b
	b	b	c	e	a
	c	c	b	a	e

the entries in the table for \mathbb{Z}_4 , only two entries along the diagonal will be the same. Therefore \mathbb{Z}_4 and K_4 are not isomorphic. \blacktriangle

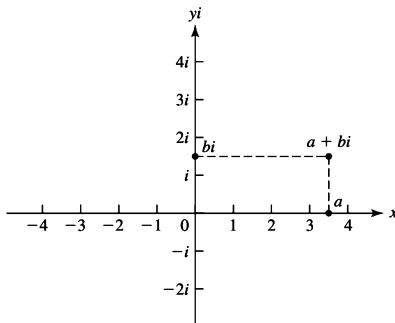
Looking back at the definition of $+_n$ there is no reason we had to restrict our set to integers a with $0 \leq a < n$. In fact, the same formula defines an operation on all real numbers a with $0 \leq a < n$. In general, let c be any positive real number and $a, b \in [0, c)$. We define $+_c$ by

$$a +_c b = \begin{cases} a + b & \text{if } a + b < c \\ a + b - c & \text{if } a + b \geq c \end{cases}.$$

This operation is called **addition modulo c** . It is easy to see that addition modulo c is an operation on $[0, c)$, it is commutative, 0 is an identity, the inverse of 0 is 0, and the inverse of any $a \in (0, c)$ is $c - a$. Instead of writing $[0, c)$ we will denote this set as \mathbb{R}_c . In order to show that $(\mathbb{R}_c, +_c)$ is an abelian group, it remains to show that $+_c$ is associative. Again, we defer the proof until after we develop the circle group.

3.5 Example Let $c = 2\pi$. Then $\frac{2}{5}\pi +_{2\pi} \frac{6}{5}\pi = \frac{8}{5}\pi$ and $\frac{7}{5}\pi +_{2\pi} \frac{6}{5}\pi = \frac{3}{5}\pi$. The inverse of $\frac{\pi}{2}$ is $2\pi - \frac{\pi}{2} = \frac{3}{2}\pi$. \blacktriangle

In the group $(\mathbb{R}_{2\pi}, +_{2\pi})$, we are essentially equating 0 with 2π in the sense that if a and b add to give 2π , we know that $a +_{2\pi} b = 0$. Intuitively, we can think of this geometrically as taking a string of length 2π and attaching the ends together to form a circle of radius 1. Our next goal is to make this idea more precise by defining a group on the unit circle in the plane and showing that this group is isomorphic with $\mathbb{R}_{2\pi}$. To do this, we first review some facts about complex numbers.



3.6 Figure

Complex Numbers

A real number can be visualized geometrically as a point on a line that we often regard as an x -axis. A complex number can be regarded as a point in the Euclidean plane, as shown in Fig. 3.6. Note that we label the vertical axis as the yi -axis rather than just the y -axis, and label the point one unit above the origin with i rather than 1. The point with Cartesian coordinates (a, b) is labeled $a + bi$ in Fig. 3.6. The set \mathbb{C} of **complex numbers** is defined by

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}.$$

We consider \mathbb{R} to be a subset of the complex numbers by identifying a real number r with the complex number $r + 0i$. For example, we write $3 + 0i$ as 3 and $-\pi + 0i$ as $-\pi$ and $0 + 0i$ as 0. Similarly, we write $0 + 1i$ as i and $0 + si$ as si .

Complex numbers were developed after the development of real numbers. The complex number i was *invented* to provide a solution to the quadratic equation $x^2 = -1$, so we require that

$$i^2 = -1. \quad (1)$$

Unfortunately, i has been called an **imaginary number**, and this terminology has led generations of students to view the complex numbers with more skepticism than the real numbers. Actually, *all* numbers, such as $1, 3, \pi, -\sqrt{3}$, and i are inventions of our minds. There is no physical entity that *is* the number 1. If there were, it would surely be in a place of honor in some great scientific museum, and past it would file a steady stream of mathematicians, gazing at 1 in wonder and awe. A basic goal of this text is to show how we can invent solutions of polynomial equations when the coefficients of the polynomial may not even be real numbers!

Multiplication of Complex Numbers

The product $(a + bi)(c + di)$ is defined in the way it must be if we are to enjoy the familiar properties of real arithmetic and require that $i^2 = -1$, in accord with Eq. (1). Namely, we see that we want to have

$$\begin{aligned} (a + bi)(c + di) &= ac + adi + bci + bdi^2 \\ &= ac + adi + bci + bd(-1) \\ &= (ac - bd) + (ad + bc)i. \end{aligned}$$

Consequently, we define multiplication of $z_1 = a + bi$ and $z_2 = c + di$ as

$$z_1 z_2 = (a + bi)(c + di) = (ac - bd) + (ad + bc)i, \quad (2)$$

which is of the form $r + si$ with $r = ac - bd$ and $s = ad + bc$. It is routine to check that the usual properties $z_1 z_2 = z_2 z_1$ (commutative), $z_1(z_2 z_3) = (z_1 z_2) z_3$ (associative), and $z_1(z_2 + z_3) = z_1 z_2 + z_1 z_3$ (distributive) all hold for all $z_1, z_2, z_3 \in \mathbb{C}$.

3.7 Example Compute $(2 - 5i)(8 + 3i)$.

Solution We don't memorize Eq. (2), but rather we compute the product as we did to motivate that equation. We have

$$(2 - 5i)(8 + 3i) = 16 + 6i - 40i + 15 = 31 - 34i. \quad \blacktriangle$$

To establish the geometric meaning of complex multiplication, we first define the **absolute value** $|a + bi|$ of $a + bi$ by

$$|a + bi| = \sqrt{a^2 + b^2}. \quad (3)$$

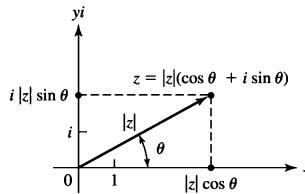
This absolute value is a nonnegative real number and is the distance from $a + bi$ to the origin in Fig. 3.6. We can now describe a complex number z in the polar-coordinate form

$$z = |z|(\cos \theta + i \sin \theta), \quad (4)$$

where θ is the angle measured counterclockwise from the positive x -axis to the vector from 0 to z , as shown in Fig. 3.8. A famous formula due to Leonard Euler states that

$$e^{i\theta} = \cos \theta + i \sin \theta.$$

Euler's Formula



3.8 Figure

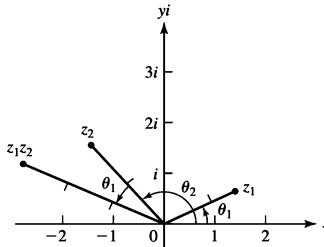
We ask you to derive Euler's formula formally from the power series expansions for e^θ , $\cos \theta$, and $\sin \theta$ in Exercise 43. Using this formula, we can express z in Eq. (4) as $z = |z|e^{i\theta}$. Let us set

$$z_1 = |z_1|e^{i\theta_1} \quad \text{and} \quad z_2 = |z_2|e^{i\theta_2}$$

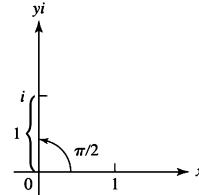
and compute their product in this form, assuming that the usual laws of exponentiation hold with complex number exponents. We obtain

$$\begin{aligned} z_1 z_2 &= |z_1|e^{i\theta_1}|z_2|e^{i\theta_2} = |z_1||z_2|e^{i(\theta_1+\theta_2)} \\ &= |z_1||z_2|[\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)]. \end{aligned} \quad (5)$$

Note that Eq. 5 concludes in the polar form of Eq. 4 where $|z_1 z_2| = |z_1||z_2|$ and the polar angle θ for $z_1 z_2$ is the sum $\theta = \theta_1 + \theta_2$. Thus, geometrically, we multiply complex numbers by multiplying their absolute values and adding their polar angles, as shown in Fig. 3.9. Exercise 41 indicates how this can be derived via trigonometric identities without recourse to Euler's formula and assumptions about complex exponentiation.



3.9 Figure



3.10 Figure

Note that i has polar angle $\pi/2$ and absolute value 1, as shown in Fig. 3.10. Thus i^2 has polar angle $2(\pi/2) = \pi$ and $|1 \cdot 1| = 1$, so that $i^2 = -1$.

3.11 Example Find all solutions in \mathbb{C} of the equation $z^2 = i$.

Solution Writing the equation $z^2 = i$ in polar form and using Eq. (5), we obtain

$$|z|^2(\cos 2\theta + i \sin 2\theta) = 1(0 + i).$$

Thus $|z|^2 = 1$, so $|z| = 1$. The angle θ for z must satisfy $\cos 2\theta = 0$ and $\sin 2\theta = 1$. Consequently, $2\theta = (\pi/2) + n(2\pi)$, so $\theta = (\pi/4) + n\pi$ for an integer n . The values of

n yielding values θ where $0 \leq \theta < 2\pi$ are 0 and 1, yielding $\theta = \pi/4$ or $\theta = 5\pi/4$. Our solutions are

$$z_1 = 1 \left(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4} \right) \quad \text{and} \quad z_2 = 1 \left(\cos \frac{5\pi}{4} + i \sin \frac{5\pi}{4} \right)$$

or

$$z_1 = \frac{1}{\sqrt{2}}(1+i) \quad \text{and} \quad z_2 = \frac{-1}{\sqrt{2}}(1+i).$$

▲

3.12 Example Find all solutions of $z^4 = -16$.

Solution As in Example 3.11 we write the equation in polar form, obtaining

$$|z|^4(\cos 4\theta + i \sin 4\theta) = 16(-1 + 0i).$$

Consequently, $|z|^4 = 16$, so $|z| = 2$ while $\cos 4\theta = -1$ and $\sin 4\theta = 0$. We find that $4\theta = \pi + n(2\pi)$, so $\theta = (\pi/4) + n(\pi/2)$ for integers n . The different values of θ obtained where $0 \leq \theta < 2\pi$ are $\pi/4, 3\pi/4, 5\pi/4$, and $7\pi/4$. Thus one solution of $z^4 = -16$ is

$$2 \left(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4} \right) = 2 \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i \right) = \sqrt{2}(1+i).$$

In a similar way, we find three more solutions,

$$\sqrt{2}(-1+i), \quad \sqrt{2}(-1-i), \quad \text{and} \quad \sqrt{2}(1-i).$$

▲

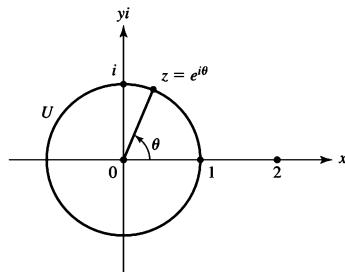
The last two examples illustrate that we can find solutions of an equation $z^n = a + bi$ by writing the equation in polar form. There will always be n solutions, provided that $a + bi \neq 0$. Exercises 16 through 21 ask you to solve equations of this type.

We will not use addition or division of complex numbers, but we probably should mention that addition is given by

$$(a+bi) + (c+di) = (a+c) + (b+d)i. \quad (6)$$

and division of $a+bi$ by nonzero $c+di$ can be performed using only division of real numbers as follows:

$$\begin{aligned} \frac{a+bi}{c+di} &= \frac{a+bi}{c+di} \cdot \frac{c-di}{c-di} = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} \\ &= \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}i. \end{aligned} \quad (7)$$



3.13 Figure

Algebra on the Unit Circle

Let $U = \{z \in \mathbb{C} \mid |z| = 1\}$, so that U is the circle in the Euclidean plane with center at the origin and radius 1, as shown in Fig. 3.13.

3.14 Theorem $\langle U, \cdot \rangle$ is an abelian group.

Proof We first check that U is closed under multiplication. Let $z_1, z_2 \in U$. Then $|z_1| = |z_2| = 1$, which implies that $|z_1 z_2| = 1$, showing $z_1 z_2 \in U$.

Since multiplication of complex numbers is associative and commutative in general, multiplication in U is also associative and commutative, which verifies \mathcal{G}_1 and the condition for abelian.

The number 1 $\in U$ is the identity, verifying condition \mathcal{G}_2 .

For each $a + bi \in U$,

$$(a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2 = |a + bi|^2 = 1.$$

So the inverse of $a + bi$ is $a - bi$, which verifies condition \mathcal{G}_3 . Thus U is an abelian group under multiplication. \blacklozenge

Figure 3.13 gives us a way of relabeling points in U as points in $\mathbb{R}_{2\pi}$. We simply relabel z as θ where $0 \leq \theta < 2\pi$. Let $f : U \rightarrow \mathbb{R}_{2\pi}$ be given by $f(z) = \theta$ according to this relabeling. Then for $z_1, z_2 \in U$, $f(z_1 z_2) = f(z_1) +_{2\pi} f(z_2)$ since multiplying in U simply adds the corresponding angles:

$$\text{if } z_1 \leftrightarrow \theta_1 \text{ and } z_2 \leftrightarrow \theta_2, \text{ then } z_1 \cdot z_2 \leftrightarrow (\theta_1 +_{2\pi} \theta_2). \quad (8)$$

Recall that all that remains to show that $\mathbb{R}_{2\pi}$ is a group is to show that $+_{2\pi}$ is associative. Since the operations of multiplication in U and addition modulo 2π in $\mathbb{R}_{2\pi}$ are the same using the above relabeling and multiplication in U is associative, addition modulo 2π is also associative. This completes the proof that $(\mathbb{R}_{2\pi}, +_{2\pi})$ is a group. Furthermore, the relabeling (8) shows that the two groups $\langle U, \cdot \rangle$ and $\langle \mathbb{R}_{2\pi}, +_{2\pi} \rangle$ are isomorphic. In Exercise 45, you will be asked to prove that for any $b > 0$ and $c > 0$, $\langle \mathbb{R}_b, +_b \rangle$ is an abelian group and $\langle \mathbb{R}_b, +_b \rangle \cong \langle \mathbb{R}_c, +_c \rangle$. Since $(\mathbb{R}_{2\pi}, +_{2\pi})$ is isomorphic with $\langle U, \cdot \rangle$, for every $c > 0$, $\langle \mathbb{R}_c, +_c \rangle$ is also isomorphic with $\langle U, \cdot \rangle$, meaning they have the same algebraic properties.

3.15 Example The equation $z \cdot z \cdot z \cdot z = 1$ in U has exactly four solutions, namely, 1, i , -1 , and $-i$. Now $1 \in U$ and $0 \in \mathbb{R}_{2\pi}$ correspond, and the equation $x +_{2\pi} x +_{2\pi} x +_{2\pi} x = 0$ in $\mathbb{R}_{2\pi}$ has exactly four solutions, namely, $0, \pi/2, \pi$, and $3\pi/2$, which, of course, correspond to $1, i, -1$, and $-i$, respectively. \blacktriangle

Roots of Unity

The elements of the set $U_n = \{z \in \mathbb{C} \mid z^n = 1\}$ are called the n^{th} **roots of unity**. In Exercise 46 you are asked to prove that U_n is a group under multiplication. Using the techniques from Examples 3.11 and 3.12, we see that the elements of this set are the numbers

$$e^{(m \frac{2\pi}{n})i} = \cos\left(m \frac{2\pi}{n}\right) + i \sin\left(m \frac{2\pi}{n}\right) \quad \text{for } m = 0, 1, 2, \dots, n-1.$$

They all have absolute value 1, so $U_n \subset U$. If we let $\zeta = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$, then these n^{th} roots of unity can be written as

$$1 = \zeta^0, \zeta^1, \zeta^2, \zeta^3, \dots, \zeta^{n-1}. \quad (9)$$

Because $\zeta^n = 1$, these n powers of ζ are closed under multiplication. For example, with $n = 10$, we have

$$\zeta^6 \zeta^8 = \zeta^{14} = \zeta^{10} \zeta^4 = 1 \cdot \zeta^4 = \zeta^4.$$

Thus we see that we can compute $\zeta^i \zeta^j$ by computing $i +_n j$, viewing i and j as elements of \mathbb{Z}_n .

By relabeling an element $\zeta^m \in U_n$ to $m \in \mathbb{Z}_n$ we can see that addition modulo n in \mathbb{Z}_n is also associative, which completes the proof that $(\mathbb{Z}_n, +_n)$ is an abelian group.

3.16 Example We solve the equation $x +_8 x +_8 x = 1$ in \mathbb{Z}_8 using trial and error. We note that neither 0, 1, nor 2 is a solution simply by substitution. However, substituting $x = 3$ gives $3 +_8 3 +_8 3 = 6 +_8 3 = 1$, which shows $x = 3$ is a solution. We can also check by substituting that neither 4, 5, 6, nor 7 are solutions. So the only solution is $x = 3$. Because \mathbb{Z}_8 is isomorphic with U_8 by the correspondence $k \in \mathbb{Z}_8$ corresponds with ζ^k , the corresponding equation in U_8 is $z \cdot z \cdot z = \zeta = e^{\frac{2\pi}{8}i}$. Without further calculations we know that there is only one solution to $z \cdot z \cdot z = \zeta$ in U_8 and that solution is $z = \zeta^3 = e^{3\frac{2\pi}{8}i} = \cos(6\pi/8) + i \sin(6\pi/8) = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$ since this is the corresponding solution in \mathbb{Z}_8 .

There are three solutions to $z^3 = \zeta$ in U . We leave it to the reader to find the solutions and check that only one of them, ζ^3 , is in U_8 . \blacktriangle

We summarize the results of this section.

1. For any $n \in \mathbb{Z}^+$, \mathbb{Z}_n is an abelian group under addition modulo n .
2. For any $n \in \mathbb{Z}^+$, \mathbb{Z}_n is isomorphic with U_n , an abelian group under complex number multiplication.
3. For any $c > 0$, R_c under addition modulo c is a group.
4. U under multiplication is a group.
5. For any $c \in \mathbb{R}^+$, \mathbb{R}_c under addition modulo c is isomorphic with U under multiplication.

■ EXERCISES 3

In Exercises 1 through 9 compute the given arithmetic expression and give the answer in the form $a + bi$ for $a, b \in \mathbb{R}$.

- | | | |
|---------------------------------|-----------------------|--|
| 1. i^3 | 2. i^4 | 3. i^{26} |
| 4. $(-i)^{39}$ | 5. $(3 - 2i)(6 + i)$ | 6. $(8 + 2i)(3 - i)$ |
| 7. $(2 - 3i)(4 + i) + (6 - 5i)$ | 8. $(1 + i)^3$ | 9. $(1 - i)^5$ (Use the binomial theorem.) |
| 10. Find $ 5 - 12i $. | 11. Find $ p + ei $. | |

In Exercises 12 through 15 write the given complex number z in the polar form $|z|(p + qi)$ where $|p + qi| = 1$.

- | | | | |
|--------------|--------------|---------------|---------------|
| 12. $3 - 4i$ | 13. $-1 - i$ | 14. $12 + 5i$ | 15. $-3 + 5i$ |
|--------------|--------------|---------------|---------------|

In Exercises 16 through 21, find all solutions in C of the given equation.

- | | | | |
|---------------|-----------------|------------------|------------------|
| 16. $z^4 = 1$ | 17. $z^4 = -1$ | 18. $z^3 = -125$ | 19. $z^3 = -27i$ |
| 20. $z^6 = 1$ | 21. $z^6 = -64$ | | |

In Exercises 22 through 27, compute the given expression using the indicated modular addition.

- | | | |
|--|--|---|
| 22. $10 +_{17} 16$ | 23. $14 +_{99} 92$ | 24. $3.141 +_4 2.718$ |
| 25. $\frac{1}{2} +_1 \frac{7}{8}$ | 26. $\frac{3\pi}{4} +_{2\pi} \frac{3\pi}{2}$ | 27. $2\sqrt{2} +_{\sqrt{32}} 3\sqrt{2}$ |
| 28. Explain why the expression $5 +_6 8$ in \mathbb{R}_6 makes no sense. | | |

In Exercises 29 through 34, find *all* solutions x of the given equation.

29. $x +_{10} 7 = 3$ in \mathbb{Z}_{10}
30. $x +_{2\pi} \pi = \frac{\pi}{2}$ in $\mathbb{R}_{2\pi}$
31. $x +_7 x = 3$ in \mathbb{Z}_7
32. $x +_{13} x +_{13} x = 5$ in \mathbb{Z}_{13}
33. $x +_{12} x = 2$ in \mathbb{Z}_{12}
34. $x +_8 x +_8 x +_8 x = 4$ in \mathbb{Z}_8
35. Prove or give a counterexample to the statement that for any $n \in \mathbb{Z}^+$ and $a \in \mathbb{Z}_n$, the equation $x +_n x = a$ has at most two solutions in \mathbb{Z}_n .
36. Prove or give a counterexample to the statement that for any $n \in \mathbb{Z}^+$ and $a \in \mathbb{Z}_n$, if n is not a multiple of 3, then the equation $x +_n x +_n x = a$ has exactly one solution in \mathbb{Z}_n .
37. There is an isomorphism of U_8 with \mathbb{Z}_8 in which $\zeta = e^{i(\pi/4)} \leftrightarrow 5$ and $\zeta^2 \leftrightarrow 2$. Find the element of \mathbb{Z}_8 that corresponds to each of the remaining six elements ζ^m in U_8 for $m = 0, 3, 4, 5, 6$, and 7.
38. There is an isomorphism of U_7 with \mathbb{Z}_7 in which $\zeta = e^{i(2\pi/7)} \leftrightarrow 4$. Find the element in \mathbb{Z}_7 to which ζ^m must correspond for $m = 0, 2, 3, 4, 5$, and 6.
39. Why can there be no isomorphism of U_6 with \mathbb{Z}_6 in which $\zeta = e^{i(\pi/3)}$ corresponds to 4?
40. Derive the formulas

$$\sin(a + b) = \sin a \cos b + \cos a \sin b$$

and

$$\cos(a + b) = \cos a \cos b - \sin a \sin b$$

by using Euler's formula and computing $e^{ia}e^{ib}$.

41. Let $z_1 = |z_1|(\cos \theta_1 + i \sin \theta_1)$ and $z_2 = |z_2|(\cos \theta_2 + i \sin \theta_2)$. Use the trigonometric identities in Exercise 40 to derive $z_1 z_2 = |z_1||z_2|[\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)]$.
42. a. Derive a formula for $\cos 3\theta$ in terms of $\sin \theta$ and $\cos \theta$ using Euler's formula.
b. Derive the formula $\cos 3\theta = 4\cos^3 \theta - 3\cos \theta$ from part (a) and the identity $\sin^2 \theta + \cos^2 \theta = 1$. (We will have use for this identity in Section 41.)
43. Recall the power series expansions

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots + \frac{x^n}{n!} + \cdots,$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} + \cdots, \text{ and}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots$$

from calculus. Derive Euler's formula $e^{i\theta} = \cos \theta + i \sin \theta$ formally from these three series expansions.

44. Prove that for any $n \in \mathbb{Z}^+$, $(\mathbb{Z}_n, +_n)$ is associative without using the fact that U_n is associative.
45. Let $b, c \in \mathbb{R}^+$. Find a one-to-one and onto function $f : \mathbb{R}_b \rightarrow \mathbb{R}_c$ that has the homomorphism property. Conclude that \mathbb{R}_c is an abelian group that is isomorphic with U .
46. Prove that for any $n \geq 1$, U_n is a group.

SECTION 4 NONABELIAN EXAMPLES

Notation and Terminology

It is time to explain some conventional notation and terminology used in group theory. Algebraists as a rule do not use a special symbol $*$ to denote a binary operation different from the usual addition and multiplication. They stick with the conventional additive or multiplicative notation and even call the operation *addition* or *multiplication*, depending

on the symbol used. The symbol for addition is, of course, $+$, and usually multiplication is denoted by juxtaposition without a dot, if no confusion results. Thus in place of the notation $a * b$, we shall be using either $a + b$ to be read “the sum of a and b ,” or ab to be read “the product of a and b . ” There is a sort of unwritten agreement that the symbol $+$ should be used only to designate commutative operations. Algebraists feel very uncomfortable when they see $a + b \neq b + a$. For this reason, when developing our theory in a general situation where the operation may or may not be commutative, we shall always use multiplicative notation.

Algebraists frequently use the symbol 0 to denote an additive identity element and the symbol 1 to denote a multiplicative identity element, even though they may not be actually denoting the integers 0 and 1 . Of course, if they are also talking about numbers at the same time, so that confusion would result, symbols such as e or u are used as identity elements. Thus a table for a group of three elements might be one like Table 4.1 or, since such a group is commutative, the table might look like Table 4.2. In general situations we shall continue to use e to denote the identity element of a group.

It is customary to denote the inverse of an element a in a group by a^{-1} in multiplicative notation and by $-a$ in additive notation. From now on, we shall use these notations in place of the symbol a' .

Let n be a positive integer. If a is an element of a group G , written multiplicatively, we denote the product $aaa\dots a$ for n factors a by a^n . We let a^0 be the identity element e , and denote the product $a^{-1}a^{-1}a^{-1}\dots a^{-1}$ for n factors by a^{-n} . It is easy to see that our usual law of exponents, $a^m a^n = a^{m+n}$ for $m, n \in \mathbb{Z}$, holds. For $m, n \in \mathbb{Z}^+$, it is clear. We illustrate another type of case by an example:

$$\begin{aligned} a^{-2}a^5 &= a^{-1}a^{-1}aaaaa = a^{-1}(a^{-1}a)aaaa = a^{-1}aaaa = a^{-1}(ea)aaa \\ &= a^{-1}aaaa = (a^{-1}a)aaa = eaaa = (ea)aa = aaa = a^3. \end{aligned}$$

In additive notation, we denote $a + a + a + \dots + a$ for n summands by na , denote $(-a) + (-a) + (-a) + \dots + (-a)$ for n summands by $-na$, and let $0a$ be the identity element. Be careful: In the notation na , the number n is in \mathbb{Z} , not in G . One reason we prefer to present group theory using multiplicative notation, even if G is abelian, is the confusion caused by regarding n as being in G in this notation na . No one ever misinterprets the n when it appears in an exponent.

The following table summarizes basic notations and facts using both additive and multiplicative notation. We assume that a is an element of a group, n, m are integers, and k is a positive integer.

* Notation	+ Notation	· Notation
May or may not be abelian	Abelian	May or may not be abelian
e	0	1
a'	$-a$	a^{-1}
$a * b$	$a + b$	ab
$\underbrace{a * a * \dots * a}_k$	ka	a^k
$\underbrace{(a' * a' * \dots * a')}_k$	$-ka$	a^{-k}
	$0a = 0$	$a^0 = 1$
	$(n+m)a = na + ma$	$a^{n+m} = a^n a^m$
	$n(ma) = (nm)a$	$(a^n)^m = a^{nm}$

Typically when stating a theorem we will use multiplicative notation, but the theorem also applies when using additive notation by using the above table to translate.

We often refer to the number of elements in a group, so we have a term for this number.

4.3 Definition If G is a group, then the **order** of G is the number of elements or cardinality of G . The order of G is denoted $|G|$. ■

Permutations

We have seen examples of groups of numbers, like the groups \mathbb{Z} , \mathbb{Q} , and \mathbb{R} under addition. We have also introduced groups of matrices, like the group $GL(2, \mathbb{R})$. Each element A of $GL(2, \mathbb{R})$ yields a transformation of the plane \mathbb{R}^2 into itself; namely, if we regard \mathbf{x} as a 2-component column vector, then $A\mathbf{x}$ is also a 2-component column vector. The group $GL(2, \mathbb{R})$ is typical of many of the most useful groups in that its elements *act on things* to transform them. Often, an action produced by a group element can be regarded as a *function*, and the binary operation of the group can be regarded as *function composition*. In this section, we construct some finite groups whose elements, called *permutations*, act on finite sets. These groups will provide us with examples of finite nonabelian groups.

You may be familiar with the notion of a permutation of a set as a rearrangement of the elements of the set. Thus for the set $\{1, 2, 3, 4, 5\}$, a rearrangement of the elements could be given schematically as in Fig. 4.4, resulting in the new arrangement $\{4, 2, 5, 3, 1\}$. Let us think of this schematic diagram in Fig. 4.4 as a function mapping each element listed in the left column into a single (not necessarily different) element from the same set listed at the right. Thus 1 is carried into 4, 2 is mapped into 2, and so on. Furthermore, to be a permutation of the set, this mapping must be such that each element appears in the right column once and only once. For example, the diagram in Fig. 4.5 does *not* give a permutation, for 3 appears twice while 1 does not appear at all in the right column. We now define a permutation to be such a mapping.

$$\begin{array}{ll} 1 \rightarrow 4 & 1 \rightarrow 3 \\ 2 \rightarrow 2 & 2 \rightarrow 2 \\ 3 \rightarrow 5 & 3 \rightarrow 4 \\ 4 \rightarrow 3 & 4 \rightarrow 5 \\ 5 \rightarrow 1 & 5 \rightarrow 3 \end{array}$$

4.4 Figure 4.5 Figure

4.6 Definition A **permutation of a set A** is a function $\phi : A \rightarrow A$ that is both one-to-one and onto. ■

Permutation Groups

We now show that function composition \circ is a binary operation on the collection of all permutations of a set A . We call this operation *permutation multiplication*. Let A be a set, and let σ and τ be permutations of A so that σ and τ are both one-to-one functions mapping A onto A . The composite function $\sigma \circ \tau$ defined schematically by

$$A \xrightarrow{\tau} A \xrightarrow{\sigma} A,$$

gives a mapping of A into A . Rather than keep the symbol \circ for permutation multiplication, we will denote $\sigma \circ \tau$ by the juxtaposition $\sigma\tau$. Now $\sigma\tau$ will be a permutation if it is one-to-one and onto A . Remember that the action of $\sigma\tau$ on A must be read in right-to-left order: first apply τ and then σ . Let us show that $\sigma\tau$ is one-to-one. If

$$(\sigma\tau)(a_1) = (\sigma\tau)(a_2),$$

then

$$\sigma(\tau(a_1)) = \sigma(\tau(a_2)),$$

and since σ is given to be one-to-one, we know that $\tau(a_1) = \tau(a_2)$. But then, since τ is one-to-one, this gives $a_1 = a_2$. Hence $\sigma\tau$ is one-to-one. To show that $\sigma\tau$ is onto A , let $a \in A$. Since σ is onto A , there exists $a' \in A$ such that $\sigma(a') = a$. Since τ is onto A , there exists $a'' \in A$ such that $\tau(a'') = a'$. Thus

$$a = \sigma(a') = \sigma(\tau(a'')) = (\sigma\tau)(a''),$$

so $\sigma\tau$ is onto A .

4.7 Example Suppose that

$$A = \{1, 2, 3, 4, 5\}$$

and that σ is the permutation given by Fig. 4.4. We write σ in a more standard notation, changing the columns to rows in parentheses and omitting the arrows, as

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

so that $\sigma(1) = 4$, $\sigma(2) = 2$, and so on. Let

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 2 & 1 \end{pmatrix}.$$

Then

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 5 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 3 & 2 & 4 \end{pmatrix}.$$

For example, multiplying in right-to-left order,

$$(\sigma\tau)(1) = \sigma(\tau(1)) = \sigma(3) = 5.$$



■ HISTORICAL NOTE

One of the earliest recorded studies of permutations occurs in the *Sefer Yetzirah*, or *Book of Creation*, written by an unknown Jewish author sometime before the eighth century. The author was interested in counting the various ways in which the letters of the Hebrew alphabet can be arranged. The question was in some sense a mystical one. It was believed that the letters had magical powers; therefore, suitable arrangements could subjugate the forces of nature. The actual text of the *Sefer Yetzirah* is very sparse: “Two letters build two words, three build six words, four build 24 words, five build 120, six build 720, seven build 5040.” Interestingly enough, the idea of counting the arrangements of the letters of the alphabet also occurred in Islamic mathematics in the eighth and ninth centuries. By the thirteenth century, in both the Islamic and Hebrew cultures, the abstract idea

of a permutation had taken root so that both Abu-l’ Abbas ibn al-Banna (1256–1321), a mathematician from Marrakech in what is now Morocco, and Levi ben Gerson, a French rabbi, philosopher, and mathematician, were able to give rigorous proofs that the number of permutations of any set of n elements is $n!$, as well as prove various results about counting combinations.

Levi and his predecessors, however, were concerned with permutations as simply arrangements of a given finite set. It was the search for solutions of polynomial equations that led Lagrange and others in the late eighteenth century to think of permutations as functions from a finite set to itself, the set being that of the roots of a given equation. And it was Augustin-Louis Cauchy (1789–1857) who developed in detail the basic theorems of permutation theory and who introduced the standard notation used in this text.

We now show that the collection of all permutations of a nonempty set A forms a group under this permutation multiplication.

4.8 Theorem Let A be a nonempty set, and let S_A be the collection of all permutations of A . Then S_A is a group under permutation multiplication.

Proof We have shown that composition of two permutations of A yields a permutation of A , so S_A is closed under permutation multiplication.

Now permutation multiplication is defined as function composition, and in Section 1, we showed that *function composition is associative*. Hence \mathcal{G}_1 is satisfied.

The permutation ι such that $\iota(a) = a$, for all $a \in A$ acts as identity. Therefore \mathcal{G}_2 is satisfied.

For a permutation σ , the inverse function, σ^{-1} , is the permutation that reverses the direction of the mapping σ , that is, $\sigma^{-1}(a)$ is the element a' of A such that $a = \sigma(a')$. The existence of exactly one such element a' is a consequence of the fact that, as a function, σ is both one-to-one and onto. For each $a \in A$ we have

$$\iota(a) = a = \sigma(a') = \sigma(\sigma^{-1}(a)) = (\sigma\sigma^{-1})(a)$$

and also

$$\iota(a') = a' = \sigma^{-1}(a) = \sigma^{-1}(\sigma(a')) = (\sigma^{-1}\sigma)(a'),$$

so that $\sigma^{-1}\sigma$ and $\sigma\sigma^{-1}$ are both the permutation ι . Thus \mathcal{G}_3 is satisfied. ◆

Warning: Some texts compute a product $\sigma\mu$ of permutations in left-to-right order, so that $(\sigma\mu)(a) = \mu(\sigma(a))$. Thus the permutation they get for $\sigma\mu$ is the one we would get by computing $\mu\sigma$. Exercise 34 asks us to check in two ways that we still get a group. If you refer to another text on this material, be sure to check its order for permutation multiplication.

There was nothing in our definition of a permutation to require that the set A be finite. However, most of our examples of permutation groups will be concerned with permutations of finite sets. Note that the *structure* of the group S_A is concerned only with the number of elements in the set A , and not what the elements in A are. If sets A and B have the same cardinality, then $S_A \cong S_B$. To define an isomorphism $\phi : S_A \rightarrow S_B$, we let $f : A \rightarrow B$ be a one-to-one function mapping A onto B , which establishes that A and B have the same cardinality. For $\sigma \in S_A$, we let $\phi(\sigma)$ be the permutation $\bar{\sigma} \in S_B$ such that $\bar{\sigma}(f(a)) = f(\sigma(a))$ for all $a \in A$. To illustrate this for $A = \{1, 2, 3\}$ and $B = \{\#, \$, \%\}$ and the function $f : A \rightarrow B$ defined as

$$f(1) = \#, \quad f(2) = \$, \quad f(3) = \%,$$

ϕ maps

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \text{ into } \begin{pmatrix} \# & \$ & \% \\ \% & \$ & \# \end{pmatrix}.$$

We simply rename the elements of A in our two-row notation by elements in B using the renaming function f , thus renaming elements of S_A to be those of S_B . We can take $\{1, 2, 3, \dots, n\}$ to be a prototype for a finite set A of n elements.

4.9 Definition Let A be the finite set $\{1, 2, \dots, n\}$. The group of all permutations of A is the **symmetric group on n letters**, and is denoted by S_n . ■

Note that S_n has $n!$ elements, where

$$n! = n(n - 1)(n - 2) \cdots (3)(2)(1).$$

Let $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$. Then

$$\sigma\tau(1) = \sigma(1) = 2$$

and

$$\tau\sigma(1) = 3$$

which says that $\sigma\tau \neq \tau\sigma$. Therefore S_3 is not abelian. We have seen that any group with at most four elements is abelian. Furthermore we will see later that up to isomorphism, the abelian group \mathbb{Z}_5 is the only group of order 5. Thus S_3 is the smallest group which is not abelian.

4.10 Example Suppose that $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 6 & 1 & 4 & 5 \end{pmatrix}$. We find the inverse σ^{-1} . We saw in the proof of Theorem 4.8 that the inverse function of a permutation is the group inverse. So it is easy to find inverses for permutations, we simply turn the tables! That is, we switch the top and bottom rows and sort the columns so the top row is in order:

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 2 & 1 & 5 & 6 & 3 \end{pmatrix}. \quad \blacktriangleleft$$

Disjoint Cycles

There is a more efficient way of specifying the action of a permutation. In the two-row notation that we have been using, we list each number 1 through n twice, once in the top row and once in the bottom row. Disjoint cycle notation allows us to write the permutation using each number only once. We illustrate with an example. Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 6 & 2 & 5 & 1 \end{pmatrix}$. To write in disjoint cycle notation we start by writing

(1

We see that $\sigma(1) = 3$, so we place 3 just to the right of 1:

(1, 3

Now we see that σ maps 3 to 6, so we write:

(1, 3, 6

Our permutation maps 6 to 1, but there is no reason to write 1 again, so we just place a parenthesis after the 6 to indicate that 6 maps back to the first element listed:

(1, 3, 6)

This is called a **cycle** because when we apply σ repeatedly, we cycle through the numbers 1, 3, and 6. A cycle containing exactly k numbers is called a **k -cycle**. So the cycle (1, 3, 6) is a 3-cycle. This is not the end of the story for σ because we have not indicated that 2 maps to 4. So we start another cycle and write

(1, 3, 6)(2,

to indicate that σ maps 2 to 4. Since 4 maps back to 2, we obtain a 2-cycle:

(1, 3, 6)(2, 4)

We still have not indicated what σ does to 5. We can write (1, 3, 6)(2, 4)(5) to indicate that 5 maps to itself, but usually we will simply leave out 1-cycles with the understanding that any number not listed maps to itself. So in disjoint cycle notation

$\sigma = (1, 3, 6)(2, 4).$

We see that σ is a product of a 3-cycle and a 2-cycle. Sometimes we refer to a 2-cycle as a **transposition**.

A collection of cycles is said to be **disjoint** if no entry is in more than one cycle. Note that σ could also be written as $(3, 6, 1)(4, 2), (2, 4)(1, 3, 6)$, or in a number of other ways. In general it doesn't matter which order we write the disjoint cycles, and inside each cycle we can start with any number as long as we keep the cyclic order the same. It is clear that any permutation in S_n can be written in disjoint cycle notation and that the representation is unique up to the order the cycles are written and the cyclic order within each cycle.

- 4.11 Example** In disjoint cycle notation, $\sigma \in S_9$ is written as $(1, 5, 2, 7)(3, 4, 9)$. Let us rewrite σ in two-row notation. Reading off the disjoint cycle notation we see that $\sigma(1) = 5$, $\sigma(5) = 2$, $\sigma(2) = 7$, $\sigma(7) = 1$, $\sigma(3) = 4$, $\sigma(4) = 9$, and $\sigma(9) = 3$. Since 6 and 8 do not appear in either cycle, we know that $\sigma(6) = 6$ and $\sigma(8) = 8$. Therefore,

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

▲

The operation that makes S_n a group is composition of functions. Keeping this in mind, we can see how to multiply permutations written in disjoint cycle notation.

- 4.12 Example** Let $\sigma = (1, 5, 3, 2, 6)$ and $\tau = (1, 2, 4, 3, 6)$ in S_6 . Let us find $\sigma\tau$ in disjoint cycle notation without resorting to using two-row notation. So

$$\sigma\tau = (1, 5, 3, 2, 6)(1, 2, 4, 3, 6).$$

We need to rewrite this product in disjoint cycles. So we ask where 1 is mapped. Since the operation is function composition, we see that the cycle τ on the right sends 1 to 2 and then the cycle on the left sends 2 to 6. So $\sigma\tau(1) = 6$ and we start our cycle by writing

$$(1, 6$$

Now we see that τ maps 6 to 1 and σ maps 1 to 5, so we write

$$(1, 6, 5$$

We note that 5 is not in the cycle $(1, 2, 4, 3, 6)$, so $\tau(5) = 5$ and $\sigma\tau(5) = \sigma(5) = 3$. So we write

$$(1, 6, 5, 3$$

Continuing in the same manner, we see that 3 maps to 1 and we complete the first cycle:

$$(1, 6, 5, 3)$$

We are now ready to start the second cycle. We note that we have still not seen where 2 maps, so we start the next cycle with 2 and we write

$$\sigma\tau = (1, 5, 3, 2, 6)(1, 2, 4, 3, 6) = (1, 6, 5, 3)(2, 4)$$

using the same method we used for the first cycle. We know we are through since we have used every number 1 through 6. ▲

Example 4.12 illustrates the process of multiplying permutations in general. We move from right to left between the cycles, and within the cycles we move from left to right.

- 4.13 Example** We compute the product of the permutations

$$\sigma = (1, 5)(2, 4)(1, 4, 3)(2, 5)(4, 2, 1)$$

using disjoint cycle notation.

We start by seeing where 1 is mapped. The first cycle on the right maps 1 to 4. We are using function composition, so we next check what (2, 5) does to 4, which is nothing. So we move to the cycle (1, 4, 3) and note that 4 is mapped to 3. Next, 3 is not in the cycle (2, 4) and so (2, 4) does not move 3. Finally, (1, 5) also does not move 3 and we conclude that $\sigma(1) = 3$. We next need to determine where 3 is mapped by σ and continue until we arrive at

$$\sigma = (1, 3, 5, 4)(2) = (1, 3, 5, 4). \quad \blacktriangle$$

It is interesting to note that in Example 4.13 the group was never specified. The same calculation is valid whether the group is S_5 , S_6 , or S_n for any $n \geq 5$.

4.14 Example We compute the inverse of $\sigma = (1, 5, 7)(3, 8, 2, 4, 6)$. We first note that in general for a group $(ab)^{-1} = b^{-1}a^{-1}$, so

$$\sigma^{-1} = (3, 8, 2, 4, 6)^{-1}(1, 5, 7)^{-1}.$$

The inverse of a cycle is simply the cycle written backward:

$$\sigma^{-1} = (6, 4, 2, 8, 3)(7, 5, 1).$$

This is a perfectly good way of writing σ^{-1} , but since disjoint cycles commute and we can start each cycle with any entry in the cycle, we could write

$$\sigma^{-1} = (1, 7, 5)(2, 8, 3, 6, 4). \quad \blacktriangle$$

With a little practice, computing products of permutations in disjoint cycle notation becomes routine. We give the table for S_3 .

4.15 Table

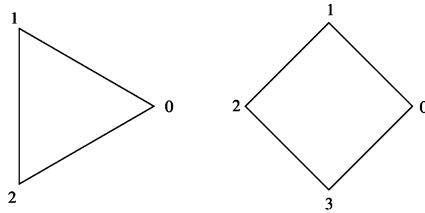
S_3	\circ	ι	(1, 2, 3)	(1, 3, 2)	(1, 2)	(1, 3)	(2, 3)
ι	ι	ι	(1, 2, 3)	(1, 3, 2)	(1, 2)	(1, 3)	(2, 3)
(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	ι	ι	(1, 3)	(2, 3)	(1, 2)
(1, 3, 2)	(1, 3, 2)	ι	(1, 2, 3)	ι	(2, 3)	(1, 2)	(1, 3)
(1, 2)	(1, 2)	(2, 3)	(1, 3)	ι	(1, 3, 2)	(1, 2, 3)	ι
(1, 3)	(1, 3)	(1, 2)	(2, 3)	(1, 2, 3)	ι	(1, 3, 2)	ι
(2, 3)	(2, 3)	(1, 3)	(1, 2)	(1, 3, 2)	(1, 2, 3)	ι	ι

Again we can see that S_3 is not abelian since the table is not symmetric about the main diagonal. We also notice that although disjoint cycles commute, the same cannot be said for cycles that are not disjoint. For example we see in Table 4.15 that $(1, 2)(2, 3) = (1, 2, 3) \neq (1, 3, 2) = (2, 3)(1, 2)$.

The Dihedral Group

We next define a collection of finite groups based on the symmetries of regular n -gons. To be specific, we use as our standard regular n -gon the one whose points are U_n . Recall that U_n includes the point $(1, 0)$ and the other points are spaced uniformly around the unit circle to form the vertices of a regular n -gon, which we denote by P_n . We label the points starting at $(1, 0)$ with 0 and continue labeling them $1, 2, 3, \dots, n - 1$ around the circle counterclockwise. Note that this is the same labeling as the isomorphism between U_n and \mathbb{Z}_n that we saw in Section 3. When we refer to a vertex we will reference it by its label. So vertex 0 is the point $(1, 0)$. Note that the edges of P_n consist of the line segments between vertices k and $k +_n 1$ for $0 \leq k \leq n - 1$.

4.16 Definition Let $n \geq 3$. Then D_n is the set of all one-to-one functions $\phi : \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ that map onto \mathbb{Z}_n with the property that the line segment between vertices i and j is an edge in P_n if and only if the line segment between $\phi(i)$ and $\phi(j)$ is an edge of P_n . The n^{th} **dihedral group** is the set D_n with binary operation function composition. ■



We justify calling $\langle D_n, \circ \rangle$ a group with Theorem 4.17.

4.17 Theorem For any $n \geq 3$, $\langle D_n, \circ \rangle$ is a group.

Proof We first show that function composition is an operation on D_n . Let $\phi, \theta \in D_n$ and suppose that the line between vertices i and j is an edge in P_n . Since $\theta \in D_n$, the line between $\theta(i)$ and $\theta(j)$ is an edge of P_n . Because $\phi \in D_n$, and the line between $\theta(i)$ and $\theta(j)$ is an edge, the line between $\phi(\theta(i)) = \phi \circ \theta(i)$ and $\phi(\theta(j)) = \phi \circ \theta(j)$ is an edge of P_n .

We leave it to the reader to check that if the line segment between $\phi(\theta(i)) = \phi \circ \theta(i)$ and $\phi(\theta(j)) = \phi \circ \theta(j)$ is an edge of P_n , then the line segment between i and j is an edge of P_n .

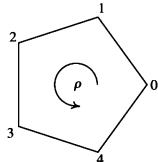
We also know that the composition of one-to-one and onto functions is one-to-one and onto, so $\phi \circ \theta \in D_n$. Therefore, function composition is an operation on D_n .

The operation of composition of functions is associative, so \mathcal{G}_1 is satisfied. The function $\iota : \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ defined by $\iota(k) = k$ is an identity in D_n , so \mathcal{G}_2 is satisfied. Finally, if $\phi \in D_n$, then $\phi^{-1} \in D_n$; the inverse function for f acts as the inverse in the group sense, so \mathcal{G}_3 is satisfied. Therefore, $\langle D_n, \circ \rangle$ is a group. ◆

Following tradition, we will use multiplicative notation in the dihedral groups instead of using \circ . If the operation on D_n were abelian, we could use additive notation, but in Example 4.18 we find that D_n is not abelian.

4.18 Example Let $n \geq 3$ and $\rho : \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ be given by rotating the n -gon P_n by $\frac{2\pi}{n}$, which just rotates each vertex to the next one. That is,

4.19 Figure



$$\rho(k) = k +_n 1$$

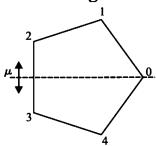
for each $k \in \mathbb{Z}_n$, as can be visualized in Figure 4.19. The function ρ matches edges to edges and it is one-to-one and onto. So $\rho \in D_n$.

A second element in D_n is reflection about the x -axis, which we call μ . By glancing at Figure 4.20 we see that in D_5 , $\mu(0) = 0$, $\mu(1) = 4$, $\mu(2) = 3$, $\mu(3) = 2$, and $\mu(4) = 1$. For any $n \geq 3$ in general, if $k \in \mathbb{Z}_n$, then

$$\mu(k) = -k.$$

(Recall that in \mathbb{Z}_n , $-k$ is the additive inverse of k , which is $n - k$ for $k > 0$ and $-0 = 0$.)

4.20 Figure



Let us check if $\mu\rho = \rho\mu$. We start by checking what each function does to 0.

$$\begin{aligned}\mu(\rho(0)) &= \mu(1) \\ &= n - 1\end{aligned}$$

$$\begin{aligned}\rho(\mu(0)) &= \rho(0) \\ &= 1\end{aligned}$$

Since $n \geq 3$, $n - 1 \neq 1$, which implies that $\mu\rho \neq \rho\mu$. Thus for all $n \geq 3$, D_n is not abelian. \blacktriangle

4.21 Theorem Let $n \geq 3$. The order of the dihedral group D_n is $2n$ and

$$D_n = \{\iota, \rho, \rho^2, \rho^3, \dots, \rho^{n-1}, \mu, \mu\rho, \mu\rho^2, \mu\rho^3, \dots, \mu\rho^{n-1}\}.$$

Proof We first show there can be at most $2n$ elements of D_n . If we map the vertices \mathbb{Z}_n to the vertices \mathbb{Z}_n , vertex 0 has n possible images. Let y be the image of vertex 0. Since y is connected by an edge to just two vertices, 1 must map to one of these two vertices. So after the image of vertex 0 is determined, there are only two choices for the image of 1. After the images of vertices 0 and 1 are determined, the rest are fixed. This means that there are at most $2n$ elements of D_n .

To show that $|D_n| = 2n$ we only need to show that no two of the functions $\iota = \rho^0, \rho, \rho^2, \rho^3, \dots, \rho^{n-1}, \mu, \mu\rho, \mu\rho^2, \mu\rho^3, \dots, \mu\rho^{n-1}$ are the same. We first suppose that $\rho^k = \rho^r$ for some integers $0 \leq k \leq n - 1$ and $0 \leq r \leq n - 1$. Then:

$$\begin{aligned}\rho^k(0) &= \rho^r(0) \\ k +_n 0 &= r +_n 0 \\ k &= r\end{aligned}$$

This shows that no two of $\iota = \rho^0, \rho, \rho^2, \rho^3, \dots, \rho^{n-1}$ are the same.

We next show that no two of $\mu = \mu\rho^0, \mu\rho, \mu\rho^2, \mu\rho^3, \dots, \mu\rho^{n-1}$ are the same. As before we assume that $\mu\rho^k = \mu\rho^r$ where $0 \leq k \leq n - 1$ and $0 \leq r \leq n - 1$ are integers. By cancellation, we have $\rho^k = \rho^r$. But then $k = r$ as shown above. Therefore no two of $\mu = \mu\rho^0, \mu\rho, \mu\rho^2, \mu\rho^3, \dots, \mu\rho^{n-1}$ are the same.

It now only remains to show that there are no values for k and r with $\rho^k = \mu\rho^r$. Note that traversing the n -gon in the order

$$\rho^k(0), \rho^k(1), \rho^k(2), \dots, \rho^k(n-1)$$

progresses in a counterclockwise manner regardless of which k we use. On the other hand,

$$\mu\rho^k(0), \mu\rho^k(1), \mu\rho^k(2), \dots, \mu\rho^k(n)$$

traverses the n -gon in a clockwise manner. This shows that there are no values of k and r for which $\rho^k = \mu\rho^r$. Therefore, D_n has at least $2n$ elements. Combining this with the fact that D_n has at most $2n$ elements shows that $|D_n| = 2n$ and

$$D_n = \{\iota, \rho, \rho^2, \rho^3, \dots, \rho^{n-1}, \mu, \mu\rho, \mu\rho^2, \mu\rho^3, \dots, \mu\rho^{n-1}\}. \quad \blacklozenge$$

Theorem 4.21 says that if $\phi \in D_n$, then there is an integer $0 \leq k \leq n - 1$ such that either $\phi = \rho^k$ or else $\phi = \mu\rho^k$. We refer to this representation of ϕ as the **standard form**. We notice that each application of μ reverses the direction traversed by the images of $0, 1, 2, 3, \dots, n$. We use this fact in the following example.

4.22 Example Let $n \geq 3$. We know $\rho\mu \neq \mu\rho$ from Example 4.18, so let us determine $\rho\mu \in D_n$ in standard form. Each time we apply μ we reverse the clock direction of the images of $0, 1, 2, 3, \dots, n - 1$. This means that $\mu\rho\mu$ reverses direction twice, so the rotation is

back to counterclockwise. Thus $\mu\rho\mu = \rho^k$ for some k . We determine the value of k by determining where 0 is sent:

$$k = \rho^k(0) = \mu\rho\mu(0) = \mu\rho(0) = \mu(1) = n - 1$$

Therefore,

$$\mu\rho\mu = \rho^{n-1}.$$

Multiplying both sides on the left by μ yields:

$$\mu\mu\rho\mu = \mu\rho^{n-1}$$

Since $\mu\mu = \iota$, we conclude that

$$\rho\mu = \mu\rho^{n-1}. \quad \blacktriangle$$

When computing products in D_n we normally want our answer in standard form. This is not difficult if we keep in mind a few basic facts about the group D_n . We have shown some of the properties listed below, and the rest you will be asked to verify in the exercises.

1. $\rho^n = \iota$ (Rotation by 2π is the identity map.)
2. $(\rho^k)^{-1} = \rho^{n-k}$
3. $\mu^2 = \iota$, which implies $\mu^{-1} = \mu$ (Reflect across a line twice is the identity map.)
4. $\rho^k\mu = \mu\rho^{n-k}$ (Example 4.22 for $k = 1$ and Exercise 30 for any k .)

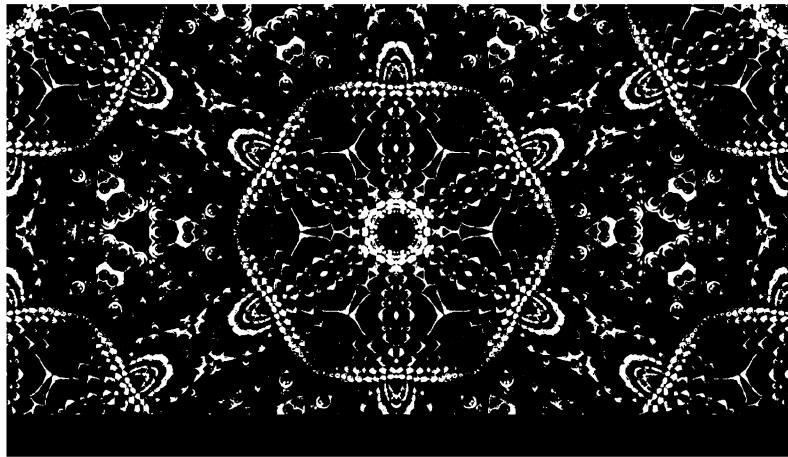
4.23 Example In the group D_5 compute $(\mu\rho^2)(\mu\rho)$. We see that

$$\begin{aligned} (\mu\rho^2)(\mu\rho) &= \mu\rho^2\mu\rho \\ &= \mu(\rho^2\mu)\rho \\ &= \mu(\mu\rho^{5-2})\rho \\ &= \mu^2\rho^4 \\ &= \rho^4 \end{aligned} \quad \blacktriangle$$

4.24 Example In the dihedral group D_n compute $(\mu\rho^k)^{-1}$.

$$\begin{aligned} (\mu\rho^k)^{-1} &= (\rho^k)^{-1}\mu^{-1} \\ &= \rho^{n-k}\mu \\ &= \mu\rho^{n-(n-k)} \\ &= \mu\rho^k \end{aligned} \quad \blacktriangle$$

In Example 4.24 we determined that the inverse of $\mu\rho^k$ is itself, which suggests that $\mu\rho^k$ could be reflection across a line of symmetry. In Exercise 37, you will be asked to show this is the case. Geometrically, we can see that each of the elements of the form $\mu\rho^k$ is reflection across a line. Placing one mirror along the line of reflection for μ and another mirror along the line of reflection for $\mu\rho$ is the basis for designing a kaleidoscope. Any element in D_n can be written as a product using only the elements μ and $\mu\rho$ since we can write $\rho = \mu\mu\rho$. In a kaleidoscope successive reflections across the mirrors correspond to taking products involving μ and $\mu\rho$. So the image you see in the kaleidoscope has all the symmetries in D_n . That is, you can rotate the image by $\frac{360^\circ}{n}$ or reflect it across any one of the lines of reflection for the elements $\mu\rho^k$. Figure 4.25 is a typical image from a kaleidoscope with dihedral group D_{16} symmetries.



ZI_creation/Shutterstock

4.25 Figure

■ EXERCISES 4

Computation

In Exercises 1 through 5, compute the indicated product involving the following permutations in S_6 :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 1 & 4 & 5 & 6 & 2 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 1 & 3 & 6 & 5 \end{pmatrix}, \quad \mu = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 4 & 3 & 1 & 6 \end{pmatrix}.$$

1. $\tau\sigma$

2. $\tau^2\sigma$

3. $\mu\sigma^2$

4. $\sigma^{-2}\tau$

5. $\sigma^{-1}\tau\sigma$

In Exercises 6 through 9, compute the expressions shown for the permutations σ , τ , and μ defined prior to Exercise 1.

6. σ^6

7. μ^2

8. σ^{100}

9. μ^{100}

10. Convert the permutations σ , τ , and μ defined prior to Exercise 1 to disjoint cycle notation.

11. Convert the following permutations in S_8 from disjoint cycle notation to two-row notation.

a. $(1, 4, 5)(2, 3)$

b. $(1, 8, 5)(2, 6, 7, 3, 4)$

c. $(1, 2, 3)(4, 5)(6, 7, 8)$

12. Compute the permutation products.

a. $(1, 5, 2, 4)(1, 5, 2, 3)$

b. $(1, 5, 3)(1, 2, 3, 4, 5, 6)(1, 5, 3)^{-1}$

c. $[(1, 6, 7, 2)^2(4, 5, 2, 6)^{-1}(1, 7, 3)]^{-1}$

d. $(1, 6)(1, 5)(1, 4)(1, 3)(1, 2)$

13. Compute the following elements of D_{12} . Write your answer in standard form.

a. $\mu\rho^2\mu\rho^8$

b. $\mu\rho^{10}\mu\rho^{-1}$

c. $\rho\mu\rho^{-1}$

d. $(\mu\rho^3\mu^{-1}\rho^{-1})^{-1}$

14. Write the group table for D_3 . Compare the group tables for D_3 and S_3 . Are the groups isomorphic?

Let A be a set and let $\sigma \in S_A$. For a fixed $a \in A$, the set

$$\mathcal{O}_{a,\sigma} = \{\sigma^n(a) \mid n \in \mathbb{Z}\}$$

is the **orbit of a under σ** . In Exercises 15 through 17, find the orbit of 1 under the permutation defined prior to Exercise 1.

15. σ

16. τ

17. μ

18. Verify that $H = \{\iota, \mu, \rho^2, \mu\rho^2\} \subseteq D_4$ is a group using the operation function composition.

19. a. Verify that the six matrices

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

form a group under matrix multiplication. [Hint: Don't try to compute all products of these matrices. In-

stead, think how the column vector $\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$ is transformed by multiplying it on the left by each of the matrices.]

- b. What group discussed in this section is isomorphic to this group of six matrices?

20. After working Exercise 18, write down eight matrices that form a group under matrix multiplication that is isomorphic to D_4 .

Concepts

In Exercises 21 through 23, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

21. The *dihedral group* D_n is the set of all functions $\phi : \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ such that the line segment between vertex i and vertex j of U_n is an edge of P_n if and only if the line segment between vertices $\phi(i)$ and $\phi(j)$ in U_n is an edge of P_n .

22. A *permutation* of a set S is a one-to-one map from S to S .

23. The *order* of a group is the number of elements in the group.

In Exercises 24 through 28, determine whether the given function is a permutation of \mathbb{R} .

24. $f_1 : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f_1(x) = x + 1$

25. $f_2 : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f_2(x) = x^2$

26. $f_3 : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f_3(x) = -x^3$

27. $f_4 : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f_4(x) = e^x$

28. $f_5 : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f_5(x) = x^3 - x^2 - 2x$

29. Determine whether each of the following is true or false.

- a. Every permutation is a one-to-one function.
- b. Every function is a permutation if and only if it is one-to-one.
- c. Every function from a finite set onto itself must be one-to-one.
- d. Every subset of an abelian group G that is also a group using the same operation as G is abelian.
- e. The symmetric group S_{10} has 10 elements.
- f. If $\phi \in D_n$, then ϕ is a permutation on the set \mathbb{Z}_n .
- g. The group D_n has exactly n elements.
- h. D_3 is a subset of D_4 .

Theory

30. Let $n \geq 3$ and $k \in \mathbb{Z}_n$. Prove that in D_n , $\rho^k \mu = \mu \rho^{n-k}$.

31. Show that S_n is a nonabelian group for $n \geq 3$.

32. Strengthening Exercise 31, show that if $n \geq 3$, then the only element of σ of S_n satisfying $\sigma\gamma = \gamma\sigma$ for all $\gamma \in S_n$ is $\sigma = \iota$, the identity permutation.
33. Orbit were defined before Exercise 15. Let $a, b \in A$ and $\sigma \in S_A$. Show that if $\mathcal{O}_{a,\sigma}$ and $\mathcal{O}_{b,\sigma}$ have an element in common, then $\mathcal{O}_{a,\sigma} = \mathcal{O}_{b,\sigma}$.
34. (See the warning following Theorem 4.8.) Let G be a group with binary operation $*$. Let G' be the same set as G , and define a binary operation $*'$ on G' by $x *' y = y * x$ for all $x, y \in G'$.
- (Intuitive argument that G' under $*'$ is a group.) Suppose the front wall of your classroom were made of transparent glass, and that all possible products $a * b = c$ and all possible instances $a * (b * c) = (a * b) * c$ of the associative property for G under $*$ were written on the wall with a magic marker. What would a person see when looking at the other side of the wall from the next room in front of yours?
 - Show from the mathematical definition of $*'$ that G' is a group under $*'$.
35. Give a careful proof using the definition of isomorphism that if G and G' are both groups with G abelian and G' not abelian, then G and G' are not isomorphic.
36. Prove that for any integer $n \geq 2$, there are at least two nonisomorphic groups with exactly $2n$ elements.
37. Let $n \geq 3$ and $0 \leq k \leq n - 1$. Prove that the map $\mu\rho^k \in D_n$ is reflection about the line through the origin that makes an angle of $-\frac{\pi k}{n}$ with the x -axis.
38. Let $n \geq 3$ and $k, r \in \mathbb{Z}_n$. Based on Exercise 37, determine the element of D_n that corresponds to first reflecting across the line through the origin at an angle of $-\frac{2\pi k}{n}$ and then reflection across the line through the origin making an angle of $-\frac{2\pi r}{n}$. Prove your answer.

SECTION 5 SUBGROUPS

Subsets and Subgroups

You may have noticed that we sometimes have had groups contained within larger groups. For example, the group \mathbb{Z} under addition is contained within the group \mathbb{Q} under addition, which in turn is contained in the group \mathbb{R} under addition. When we view the group $\langle \mathbb{Z}, + \rangle$ as contained in the group $\langle \mathbb{R}, + \rangle$, it is very important to notice that the operation $+$ on integers n and m as elements of $\langle \mathbb{Z}, + \rangle$ produces the same element $n + m$ as would result if you were to think of n and m as elements in $\langle \mathbb{R}, + \rangle$. Thus we should *not* regard the group $\langle \mathbb{Q}^+, \cdot \rangle$ as contained in $\langle \mathbb{R}, + \rangle$, even though \mathbb{Q}^+ is contained in \mathbb{R} as a set. In this instance, $2 \cdot 3 = 6$ in $\langle \mathbb{Q}^+, \cdot \rangle$, while $2 + 3 = 5$ in $\langle \mathbb{R}, + \rangle$. We are requiring not only that the set of one group be a subset of the set of the other, but also that the group operation on the subset be the *induced operation* that assigns the same element to each ordered pair from this subset as is assigned by the group operation on the whole set.

5.1 Definition

If a subset H of a group G is closed under the binary operation of G and if H with the induced operation from G is itself a group, then H is a **subgroup of G** . We shall let $H \leq G$ or $G \geq H$ denote that H is a subgroup of G , and $H < G$ or $G > H$ shall mean $H \leq G$ but $H \neq G$. ■

Thus $\langle \mathbb{Z}, + \rangle < \langle \mathbb{R}, + \rangle$ but $\langle \mathbb{Q}^+, \cdot \rangle$ is *not* a subgroup of $\langle \mathbb{R}, + \rangle$, even though as sets, $\mathbb{Q}^+ \subset \mathbb{R}$. Every group G has as subgroups G itself and $\{e\}$, where e is the identity element of G .

5.2 Definition

If G is a group, then the subgroup consisting of G itself is the **improper subgroup** of G . All other subgroups are **proper subgroups**. The subgroup $\{e\}$ is the **trivial subgroup** of G . All other subgroups are **nontrivial**. ■

We turn to some illustrations.

5.3 Example Let \mathbb{R}^n be the additive group of all n -component row vectors with real number entries. The subset consisting of all of these vectors having 0 as entry in the first component is a subgroup of \mathbb{R}^n . \blacktriangle

5.4 Example \mathbb{Q}^+ under multiplication is a proper subgroup of \mathbb{R}^+ under multiplication. \blacktriangle

5.5 Example The n^{th} roots of unity in \mathbb{C} , U_n , form a subgroup of U , the complex numbers whose absolute value is 1, which in turn is a subgroup of \mathbb{C}^* , the nonzero complex numbers under multiplication. \blacktriangle

5.6 Example Recall that $S_{\mathbb{Z}_n}$ is the set of all one-to-one functions mapping \mathbb{Z}_n onto \mathbb{Z}_n and D_n is the set of all one-to-one functions ϕ mapping \mathbb{Z}_n onto \mathbb{Z}_n with the further property that the line segment between i and j is an edge of the regular n -gon P_n if and only if the line segment between $\phi(i)$ and $\phi(j)$ is an edge. $D_n \subseteq S_{\mathbb{Z}_n}$. Since both D_n and $S_{\mathbb{Z}_n}$ are groups under composition of functions, $D_n \leq S_{\mathbb{Z}_n}$. \blacktriangle

5.7 Example There are two different types of group structures of order 4 (see Exercise 20 of Section 2). We describe them by their group tables (Tables 5.8 and 5.9). The group V is the Klein 4-group.

The only nontrivial proper subgroup of \mathbb{Z}_4 is $\{0, 2\}$. Note that $\{0, 3\}$ is *not* a subgroup of \mathbb{Z}_4 , since $\{0, 3\}$ is *not closed* under $+$. For example, $3 + 3 = 2$, and $2 \notin \{0, 3\}$. However, the group V has three nontrivial proper subgroups, $\{e, a\}$, $\{e, b\}$, and $\{e, c\}$. Here $\{e, a, b\}$ is *not* a subgroup, since $\{e, a, b\}$ is *not closed* under the operation of V because $ab = c$, and $c \notin \{e, a, b\}$. \blacktriangle

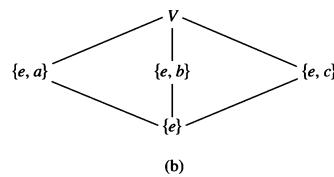
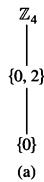
5.8 Table

\mathbb{Z}_4 :	+	0	1	2	3
0	0	1	2	3	
1	1	2	3	0	
2	2	3	0	1	
3	3	0	1	2	

5.9 Table

V :	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

It is often useful to draw a *subgroup diagram* of the subgroups of a group. In such a diagram, a line running downward from a group G to a group H means that H is a subgroup of G . Thus the larger group is placed nearer the top of the diagram. Figure 5.10 contains the subgroup diagrams for the groups \mathbb{Z}_4 and V of Example 5.7.

5.10 Figure (a) Subgroup diagram for \mathbb{Z}_4 . (b) Subgroup diagram for V .

Note that if $H \leq G$ and $a \in H$, then by Theorem 2.17, the equation $ax = a$ must have a unique solution, namely the identity element of H . But this equation can also

be viewed as one in G , and we see that this unique solution must also be the identity element e of G . A similar argument then applied to the equation $ax = e$, viewed in both H and G , shows that the inverse a^{-1} of a in G is also the inverse of a in the subgroup H .

5.11 Example Let F be the group of all real-valued functions with domain \mathbb{R} under addition. The subset of F consisting of those functions that are continuous is a subgroup of F , for the sum of continuous functions is continuous, the function f where $f(x) = 0$ for all x is continuous and is the additive identity element, and if f is continuous, then $-f$ is continuous. ▲

It is convenient to have routine steps for determining whether a subset of a group G is a subgroup of G . Example 5.11 indicates such a routine, and in the next theorem, we demonstrate carefully its validity.

5.12 Theorem A subset H of a group G is a subgroup of G if and only if

1. H is closed under the binary operation of G ,
2. the identity element e of G is in H , and
3. for all $a \in H$, $a^{-1} \in H$ also.

Proof The fact that if $H \leq G$ then Conditions 1, 2, and 3 must hold follows at once from the definition of a subgroup and from the remarks preceding Example 5.11.

Conversely, suppose H is a subset of a group G such that Conditions 1, 2, and 3 hold. By 2 we have at once that \mathcal{G}_2 is satisfied. Also \mathcal{G}_3 is satisfied by 3. It remains to check the associative axiom, \mathcal{G}_1 . But surely for all $a, b, c \in H$ it is true that $(ab)c = a(bc)$ in H , for we may actually view this as an equation in G , where the associative law holds. Hence $H \leq G$. ◆

5.13 Example Let F be as in Example 5.11. The subset of F consisting of those functions that are differentiable is a subgroup of F , for the sum of differentiable functions is differentiable, the constant function 0 is differentiable, and if f is differentiable, then $-f$ is differentiable. ▲

5.14 Example Recall from linear algebra that every square matrix A has associated with it a number $\det(A)$ called its determinant, and that A is invertible if and only if $\det(A) \neq 0$. If A and B are square matrices of the same size, then it can be shown that $\det(AB) = \det(A) \cdot \det(B)$. Let G be the multiplicative group of all invertible $n \times n$ matrices with entries in \mathbb{C} and let T be the subset of G consisting of those matrices with determinant 1. The equation $\det(AB) = \det(A) \cdot \det(B)$ shows that T is closed under matrix multiplication. Recall that the identity matrix I_n has determinant 1. From the equation $\det(A) \cdot \det(A^{-1}) = \det(AA^{-1}) = \det(I_n) = 1$, we see that if $\det(A) = 1$, then $\det(A^{-1}) = 1$. Theorem 5.12 then shows that T is a subgroup of G . ▲

Theorem 5.15 provides an alternate way of checking that a subset of a group is a subgroup.

5.15 Theorem A nonempty subset H of the group G is a subgroup of G if and only if for all $a, b \in G$, $ab^{-1} \in G$.

Proof We leave the proof as Exercise 51. ◆

On the surface Theorem 5.15 may seem simpler than Theorem 5.12 since we only need to show that H is not empty and one other condition. In practice, it is usually just as efficient to use Theorem 5.12. On the other hand, Theorem 5.16 can often be used efficiently.

5.16 Theorem Let H be a finite nonempty subset of the group G . Then H is a subgroup of G if and only if H is closed under the operation of G .

Proof We leave the proof as Exercise 57. ◆

5.17 Example Recall that $U_n = \{z \in \mathbb{C} \mid z^n = 1\}$. We could use Theorem 5.16 to verify that U_n is a subgroup of \mathbb{C}^* by noting that U_n has exactly n elements, so U_n is a finite nonempty subset of \mathbb{C}^* and if $z_1, z_2 \in U_n$, then $(z_1 z_2)^n = 1$, which implies that U_n is closed under multiplication. ▲

5.18 Example We verify that the subset $H = \{\iota = \rho^0, \rho, \rho^2, \dots, \rho^{n-1}\} \subset D_n$ is a subgroup of D_n . By Theorem 5.16, we only need to check that H is closed under the operation of D_n . Let $k, r \in \mathbb{Z}_n$. Then $\rho^k \rho^r = \rho^{k+r} \in H$. Therefore $H \leq D_n$. ▲

Cyclic Subgroups

Let us see how large a subgroup H of \mathbb{Z}_{12} would have to be if it contains 3. It would have to contain the identity element 0 and $3 + 3$, which is 6. Then it has to contain $6 + 3$, which is 9. Note that the inverse of 3 is 9 and the inverse of 6 is 6. It is easily checked that $H = \{0, 3, 6, 9\}$ is a subgroup of \mathbb{Z}_{12} , and it is the smallest subgroup containing 3.

Let us imitate this reasoning in a general situation. As we remarked before, for a general argument we always use multiplicative notation. Let G be a group and let $a \in G$. A subgroup of G containing a must, by Theorem 5.12, contain a^n , the result of computing products of a and itself for n factors for every positive integer n . These positive integral powers of a do give a set closed under multiplication. It is possible, however, that the inverse of a is not in this set. Of course, a subgroup containing a must also contain a^{-1} , and, in general, it must contain a^{-m} for all $m \in \mathbb{Z}^+$. It must contain the identity element $e = a^0$. Summarizing, a subgroup of G containing the element a must contain all elements a^n (or na for additive groups) for all $n \in \mathbb{Z}$. That is, a subgroup containing a must contain $\{a^n \mid n \in \mathbb{Z}\}$. Observe that these powers a^n of a need not be distinct. For example, in the group V of Example 5.7,

$$a^2 = e, \quad a^3 = a, \quad a^4 = e, \quad a^{-1} = a, \quad \text{and so on.}$$

We have almost proved the next theorem.

5.19 Theorem Let G be a group and let $a \in G$. Then

$$H = \{a^n \mid n \in \mathbb{Z}\}$$

is a subgroup of G and is the smallest[†] subgroup of G that contains a , that is, every subgroup containing a contains H .

Proof We check the three conditions given in Theorem 5.12 for a subset of a group to give a subgroup. Since $a^r a^s = a^{r+s}$ for $r, s \in \mathbb{Z}$, we see that the product in G of two elements of H is again in H . Thus H is closed under the group operation of G . Also $a^0 = e$, so $e \in H$, and for $a^r \in H$, $a^{-r} \in H$ and $a^{-r} a^r = e$. Hence all the conditions are satisfied, and $H \leq G$.

[†] We may find occasion to distinguish between the terms *minimal* and *smallest* as applied to subsets of a set S that have some property. A subset H of S is minimal with respect to the property if H has the property, and no subset $K \subset H$, $K \neq H$, has the property. If H has the property and $H \subseteq K$ for every subset K with the property, then H is the smallest subset with the property. There may be many minimal subsets, but there can be only one smallest subset. To illustrate, $\{e, a\}$, $\{e, b\}$, and $\{e, c\}$ are all minimal nontrivial subgroups of the group V . (See Fig. 5.10.) However, V contains no smallest nontrivial subgroup.

Our arguments prior to the statement of the theorem showed that any subgroup of G containing a must contain H , so H is the smallest subgroup of G containing a . \blacklozenge

5.20 Definition Let G be a group and let $a \in G$. Then the subgroup $\{a^n \mid n \in \mathbb{Z}\}$ of G , characterized in Theorem 5.19, is called the **cyclic subgroup of G generated by a** , and denoted by $\langle a \rangle$. \blacksquare

5.21 Example Let us find two of the cyclic subgroups to D_{10} . We first consider $\langle \mu\rho^k \rangle$ for $k \in \mathbb{Z}_{10}$. Since $(\mu\rho^k)^2 = \iota$ and $(\mu\rho^k)^{-1} = \mu\rho^k$, for any integer r , $(\mu\rho^k)^r$ is either $\mu\rho^k$ or ι . Thus

$$\langle \mu\rho^k \rangle = \{\iota, \mu\rho^k\}.$$

Since $\rho^{-1} = \rho^9$, every negative power of ρ is also a positive power of ρ and $\rho^{10} = \iota$,

$$\langle \rho \rangle = \{\iota, \rho, \rho^2, \dots, \rho^9\}.$$



5.22 Definition An element a of a group G **generates** G and is a **generator for G** if $\langle a \rangle = G$. A group G is **cyclic** if there is some element a in G that generates G . \blacksquare

5.23 Example Let \mathbb{Z}_4 and V be the groups of Example 5.7. Then \mathbb{Z}_4 is cyclic and both 1 and 3 are generators, that is,

$$\langle 1 \rangle = \langle 3 \rangle = \mathbb{Z}_4.$$

However, V is *not* cyclic, for $\langle a \rangle$, $\langle b \rangle$, and $\langle c \rangle$ are proper subgroups of two elements. Of course, $\langle e \rangle$ is the trivial subgroup of one element. \blacktriangle

5.24 Example The group \mathbb{Z} under addition is a cyclic group. Both 1 and -1 are generators for this group, and they are the only generators. Also, for $n \in \mathbb{Z}^+$, the group \mathbb{Z}_n under addition modulo n is cyclic. If $n > 1$, then both 1 and $n - 1$ are generators, but there may be others. \blacktriangle

5.25 Example Consider the group \mathbb{Z} under addition. Let us find $\langle 3 \rangle$. Here the notation is additive, and $\langle 3 \rangle$ must contain

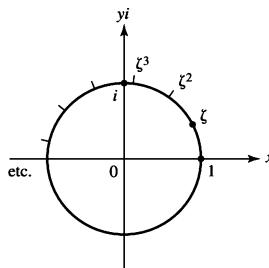
$$\begin{aligned} 3, \quad 3 + 3 = 6, \quad 3 + 3 + 3 = 9, & \quad \text{and so on,} \\ 0, \quad -3, \quad -3 + -3 = -6, \quad -3 + -3 + -3 = -9, & \quad \text{and so on.} \end{aligned}$$

In other words, the cyclic subgroup generated by 3 consists of all multiples of 3, positive, negative, and zero. We denote this subgroup by $3\mathbb{Z}$ as well as $\langle 3 \rangle$. In a similar way, we shall let $n\mathbb{Z}$ be the cyclic subgroup $\langle n \rangle$ of \mathbb{Z} . Note that $6\mathbb{Z} < 3\mathbb{Z}$. \blacktriangle

5.26 Example For each positive integer n , U_n is the multiplicative group of the n th roots of unity in \mathbb{C} . These elements of U_n can be represented geometrically by equally spaced points on a circle about the origin, as illustrated in Fig. 5.27. The point labeled represents the number

$$\zeta = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}.$$

The geometric interpretation of multiplication of complex numbers, explained in Section 3, shows at once that as ζ is raised to powers, it works its way counterclockwise around the circle, landing on each of the elements of U_n in turn. Thus U_n under multiplication is a cyclic group, and ζ is a generator. The group U_n is the cyclic subgroup $\langle \zeta \rangle$ of the group U of all complex numbers z , where $|z| = 1$, under multiplication. \blacktriangle



5.27 Figure

■ EXERCISES 5

Computations

In Exercises 1 through 6, determine whether the given subset of the complex numbers is a subgroup of the group \mathbb{C} of complex numbers under addition.

1. \mathbb{R}
2. \mathbb{Q}^+
3. $7\mathbb{Z}$
4. The set $i\mathbb{R}$ of pure imaginary numbers including 0
5. The set $\pi\mathbb{Q}$ of rational multiples of π
6. The set $\{\pi^n \mid n \in \mathbb{Z}\}$
7. Which of the sets in Exercises 1 through 6 are subgroups of the group \mathbb{C}^* of nonzero complex numbers under multiplication?

In Exercises 8 through 13, determine whether the given set of invertible $n \times n$ matrices with real number entries is a subgroup of $GL(n, \mathbb{R})$.

8. The $n \times n$ matrices with determinant greater than or equal to 1
9. The diagonal $n \times n$ matrices with no zeros on the diagonal
10. The $n \times n$ matrices with determinant 2^k for some integer k
11. The $n \times n$ matrices with determinant -1
12. The $n \times n$ matrices with determinant -1 or 1
13. The set of all $n \times n$ matrices A such that $(A^T)A = I_n$. [These matrices are called **orthogonal**. Recall that A^T , the transpose of A , is the matrix whose j th column is the j th row of A for $1 \leq j \leq n$, and that the transpose operation has the property $(AB)^T = (B^T)(A^T)$.]

Let F be the set of all real-valued functions with domain \mathbb{R} and let \tilde{F} be the subset of F consisting of those functions that have a nonzero value at every point in \mathbb{R} . In Exercises 14 through 19, determine whether the given subset of F with the induced operation is (a) a subgroup of the group F under addition, (b) a subgroup of the group \tilde{F} under multiplication.

14. The subset \tilde{F}
15. The subset of all $f \in F$ such that $f(1) = 0$
16. The subset of all $f \in \tilde{F}$ such that $f(1) = 1$
17. The subset of all $f \in \tilde{F}$ such that $f(0) = 1$
18. The subset of all $f \in \tilde{F}$ such that $f(0) = -1$
19. The subset of all constant functions in F .

20. Nine groups are given below. Give a *complete* list of all subgroup relations, of the form $G_i \leq G_j$, that exist between these given groups G_1, G_2, \dots, G_9 .

$G_1 = \mathbb{Z}$ under addition
 $G_2 = 12\mathbb{Z}$ under addition
 $G_3 = \mathbb{Q}^+$ under multiplication
 $G_4 = \mathbb{R}$ under addition
 $G_5 = \mathbb{R}^+$ under multiplication
 $G_6 = \{\pi^n \mid n \in \mathbb{Z}\}$ under multiplication
 $G_7 = 3\mathbb{Z}$ under addition
 $G_8 = \text{the set of all integral multiples of 6}$ under addition
 $G_9 = \{6^n \mid n \in \mathbb{Z}\}$ under multiplication

21. Write at least 5 elements of each of the following cyclic groups.

- a. $25\mathbb{Z}$ under addition
b. $\{(\frac{1}{2})^n \mid n \in \mathbb{Z}\}$ under multiplication
c. $\{\pi^n \mid n \in \mathbb{Z}\}$ under multiplication
d. $\langle \rho^3 \rangle$ in the group D_{18}
e. $\langle (1, 2, 3)(5, 6) \rangle$ in the group S_6

In Exercises 22 through 25, describe all the elements in the cyclic subgroup of $GL(2, \mathbb{R})$ generated by the given 2×2 matrix.

22. $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ 23. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ 24. $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ 25. $\begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix}$

26. Which of the following groups are cyclic? For each cyclic group, list all the generators of the group.

$G_1 = \langle \mathbb{Z}, + \rangle$ $G_2 = \langle \mathbb{Q}, + \rangle$ $G_3 = \langle \mathbb{Q}^+, \cdot \rangle$ $G_4 = \langle 6\mathbb{Z}, + \rangle$
 $G_5 = \{6^n \mid n \in \mathbb{Z}\}$ under multiplication
 $G_6 = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$ under addition

In Exercises 27 through 35, find the order of the cyclic subgroup of the given group generated by the indicated element.

27. The subgroup of \mathbb{Z}_4 generated by 3
28. The subgroup of V generated by c (see Table 5.9)
29. The subgroup of U_6 generated by $\cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3}$
30. The subgroup of \mathbb{Z}_{10} generated by 8
31. The subgroup of \mathbb{Z}_{16} generated by 12
32. The subgroup of the symmetric group S_8 generated by $(2, 4, 6, 9)(3, 5, 7)$
33. The subgroup of the symmetric group S_{10} generated by $(1, 10)(2, 9)(3, 8)(4, 7)(5, 6)$
34. The subgroup of the multiplicative group G of invertible 4×4 matrices generated by

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

35. The subgroup of the multiplicative group G of invertible 4×4 matrices generated by

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

36. a. Complete Table 5.28 to give the group \mathbb{Z}_6 of 6 elements.
b. Compute the subgroups $\langle 0 \rangle, \langle 1 \rangle, \langle 2 \rangle, \langle 3 \rangle, \langle 4 \rangle$, and $\langle 5 \rangle$ of the group \mathbb{Z}_6 given in part (a).

- c. Which elements are generators for the group \mathbb{Z}_6 of part (a)?
d. Give the subgroup diagram for the part (b) subgroups of \mathbb{Z}_6 . (We will see later that these are all the subgroups of \mathbb{Z}_6 .)

5.28 Table

$\mathbb{Z}_6:$	$+$	0	1	2	3	4	5
0	0	1	2	3	4	5	0
1	1	2	3	4	5	0	
2	2						
3	3						
4	4						
5	5						

Concepts

In Exercises 37 and 38, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

37. A *subgroup* of a group G is a subset H of G that contains the identity element e of G and also contains the inverse of each of its elements.
38. A group G is *cyclic* if and only if there exists $a \in G$ such that $G = \{a^n \mid n \in \mathbb{Z}\}$.
39. Determine whether each of the following is true or false.
- a. The associative law holds in every group.
 - b. There may be a group in which the cancellation law fails.
 - c. Every group is a subgroup of itself.
 - d. Every group has exactly two improper subgroups.
 - e. In every cyclic group, every element is a generator.
 - f. A cyclic group has a unique generator.
 - g. Every set of numbers that is a group under addition is also a group under multiplication.
 - h. A subgroup may be defined as a subset of a group.
 - i. \mathbb{Z}_4 is a cyclic group.
 - j. Every subset of every group is a subgroup under the induced operation.
 - k. For any $n \geq 3$, the dihedral group D_n has at least $n + 2$ cyclic subgroups.
40. Show by means of an example that it is possible for the quadratic equation $x^2 = e$ to have more than two solutions in some group G with identity e .

In Exercises 41 through 44 let B be a subset of A , and let b be a particular element of B . Determine whether the given set is a subgroup of the symmetric group S_A under the induced operation. Here $\sigma[B] = \{\sigma(x) \mid x \in B\}$.

41. $\{\sigma \in S_A \mid \sigma(b) = b\}$
42. $\{\sigma \in S_A \mid \sigma(b) \in B\}$
43. $\{\sigma \in S_A \mid \sigma[B] \subseteq B\}$
44. $\{\sigma \in S_A \mid \sigma[B] = B\}$

Theory

In Exercises 45 and 46, let $\phi : G \rightarrow G'$ be an isomorphism of a group $\langle G, *\rangle$ with a group $\langle G', *' \rangle$. Write out a proof to convince a skeptic of the intuitively clear statement.

45. If H is a subgroup of G , then $\phi[H] = \{\phi(h) \mid h \in H\}$ is a subgroup of G' . That is, an isomorphism carries subgroups into subgroups.

46. If there is an $a \in G$ such that $\langle a \rangle = G$, then G' is cyclic.

47. Show that if H and K are subgroups of an abelian group G , then

$$\{hk \mid h \in H \text{ and } k \in K\}$$

is a subgroup of G .

48. Find an example of a group G and two subgroups H and K such that the set in Exercise 47 is not a subgroup of G .

49. Prove that for any integer $n \geq 3$, S_n has a subgroup isomorphic with D_n .

50. Find the flaw in the following argument: “Condition 2 of Theorem 5.12 is redundant, since it can be derived from 1 and 3, for let $a \in H$. Then $a^{-1} \in H$ by 3, and by 1, $aa^{-1} = e$ is an element of H , proving 2.”

51. Prove Theorem 5.15.

52. Prove that if G is a cyclic group and $|G| \geq 3$, then G has at least 2 generators.

53. Prove that if G is an abelian group, written multiplicatively, with identity element e , then all elements x of G satisfying the equation $x^2 = e$ form a subgroup H of G .

54. Repeat Exercise 53 for the general situation of the set H of all solutions x of the equation $x^n = e$ for a fixed integer $n \geq 1$ in an abelian group G with identity e .

55. Find a counterexample to Exercise 53 if the assumption of abelian is dropped.

56. Show that if $a \in G$, where G is a finite group with identity e , then there exists $n \in \mathbb{Z}^+$ such that $a^n = e$.

57. Prove Theorem 5.16.

58. Let G be a group and let a be one fixed element of G . Show that

$$H_a = \{x \in G \mid xa = ax\}$$

is a subgroup of G .

59. Generalizing Exercise 58, let S be any subset of a group G .

a. Show that $H_S = \{x \in G \mid xs = sx \text{ for all } s \in S\}$ is a subgroup of G .

b. In reference to part (a), the subgroup H_G is the **center of G** . Show that H_G is an abelian group.

60. Let H be a subgroup of a group G . For $a, b \in G$, let $a \sim b$ if and only if $ab^{-1} \in H$. Show that \sim is an equivalence relation on G .

61. For sets H and K , we define the **intersection** $H \cap K$ by

$$H \cap K = \{x \mid x \in H \text{ and } x \in K\}.$$

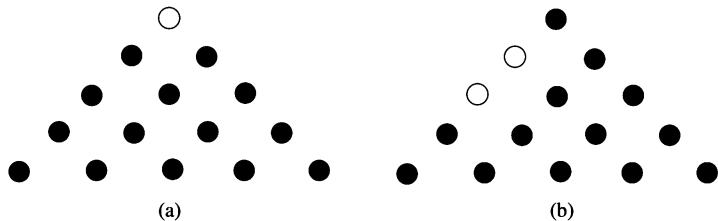
Show that if $H \leq G$ and $K \leq G$, then $H \cap K \leq G$. (Remember: \leq denotes “is a subgroup of,” not “is a subset of.”)

62. Prove that every cyclic group is abelian.

63. Let G be a group and let $G_n = \{g^n \mid g \in G\}$. Under what hypothesis about G can we show that G_n is a subgroup of G ?

64. Show that a group with no proper nontrivial subgroups is cyclic.

65. Cracker Barrel Restaurants place a puzzle called “Jump All But One Game” at each table. The puzzle starts with golf tees arranged in a triangle as in Figure 5.29a where the presence of a tee is noted with a solid dot and the absence is noted with a hollow dot. A move can be made if a tee can jump over one adjacent tee and land on an empty space. When a move is made, the tee that is jumped over is removed. A possible first move is shown in Figure 5.29b. The goal is to have just one remaining tee. Use the Klein 4-group to show that no matter what sequence of (legal) moves you make, the last remaining tee cannot be in a bottom corner position.



5.29 Figure

SECTION 6**CYCLIC GROUPS**

Recall the following facts and notations from Section 5. If G is a group and $a \in G$, then

$$H = \{a^n \mid n \in \mathbb{Z}\}$$

is a subgroup of G (Theorem 5.19). This group is the **cyclic subgroup $\langle a \rangle$ of G generated by a** . Also, given a group G and an element a in G , if

$$G = \{a^n \mid n \in \mathbb{Z}\},$$

then a is a **generator of G** and the group $G = \langle a \rangle$ is **cyclic**. We introduce one new bit of terminology. Let a be an element of a group G . If the cyclic subgroup $\langle a \rangle$ of G is finite, then the **order of a** is the order $|\langle a \rangle|$ of this cyclic subgroup. Otherwise, we say that a is of **infinite order**. We will see in this section that if $a \in G$ is of finite order m , then m is the smallest positive integer such that $a^m = e$.

The first goal of this section is to describe all cyclic groups and all subgroups of cyclic groups. This is not an idle exercise. We will see later that cyclic groups serve as building blocks for a significant class of abelian groups, in particular, for all finite abelian groups. Cyclic groups are fundamental to the understanding of groups.

Elementary Properties of Cyclic Groups

We start with a demonstration that cyclic groups are abelian.

6.1 Theorem Every cyclic group is abelian.

Proof Let G be a cyclic group and let a be a generator of G so that

$$G = \langle a \rangle = \{a^n \mid n \in \mathbb{Z}\}.$$

If g_1 and g_2 are any two elements of G , there exist integers r and s such that $g_1 = a^r$ and $g_2 = a^s$. Then

$$g_1 g_2 = a^r a^s = a^{r+s} = a^{s+r} = a^s a^r = g_2 g_1,$$

so G is abelian. ◆

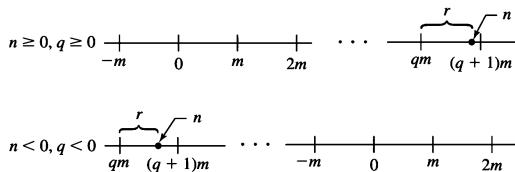
We shall continue to use multiplicative notation for our general work on cyclic groups, even though they are abelian.

The *division algorithm* that follows is well known and seems pretty simple. In fact, this algorithm is taught in elementary school. If you divide an integer n by a positive integer m , you get an integer quotient q with a remainder r where $0 \leq r < m$. You might write this as $n \div m = q \text{ R } r$, which of course means $\frac{n}{m} = q + \frac{r}{m}$. Multiplying both sides by m gives the form of the division algorithm that is a fundamental tool for the study of cyclic groups.

6.2 Division Algorithm for \mathbb{Z} If m is a positive integer and n is any integer, then there exist unique integers q and r such that

$$n = mq + r \quad \text{and} \quad 0 \leq r < m.$$

Proof We give an intuitive diagrammatic explanation, using Fig. 6.3. On the number line, mark off the multiples of m and the position of n . Now n falls either on a multiple qm of m and r can be taken as 0, or n falls between two multiples of m . If the latter is the case, let qm be the first multiple of m to the left of n . Then r is as shown in Fig. 6.3. Note that $0 \leq r < m$. Uniqueness of q and r follows since if n is not a multiple of m so that we can take $r = 0$, then there is a unique multiple qm of m to the left of n and at distance less than m from n , as illustrated in Fig. 6.3. \blacklozenge



6.3 Figure

In the notation of the division algorithm, we regard q as the **quotient** and r as the nonnegative **remainder** when n is divided by m .

6.4 Example Find the quotient q and remainder r when 38 is divided by 7 according to the division algorithm.

Solution The positive multiples of 7 are 7, 14, 21, 28, 35, 42, Choosing the multiple to leave a nonnegative remainder less than 7, we write

$$38 = 35 + 3 = 7(5) + 3$$

so the quotient is $q = 5$ and the remainder is $r = 3$. \blacktriangle

6.5 Example Find the quotient q and remainder r when -38 is divided by 7 according to the division algorithm.

Solution The negative multiples of 7 are $-7, -14, -21, -28, -35, -42, \dots$. Choosing the multiple to leave a nonnegative remainder less than 7, we write

$$-38 = -42 + 4 = 7(-6) + 4$$

so the quotient is $q = -6$ and the remainder is $r = 4$. \blacktriangle

We will use the division algorithm to show that a subgroup H of a cyclic group G is also cyclic. Think for a moment what we will have to do to prove this. We will have to use the *definition* of a cyclic group since we have proved little about cyclic groups yet. That is, we will have to use the fact that G has a generating element a . We must then exhibit, in terms of this generator a , some generator $c = a^m$ for H in order to show that H is cyclic. There is really only one natural choice for the power m of a to try. Can you guess what it is before you read the proof of the theorem?

6.6 Theorem A subgroup of a cyclic group is cyclic.

Proof Let G be a cyclic group generated by a and let H be a subgroup of G . If $H = \{e\}$, then $H = \langle e \rangle$ is cyclic. If $H \neq \{e\}$, then $a^n \in H$ for some $n \in \mathbb{Z}^+$. Let m be the smallest integer in \mathbb{Z}^+ such that $a^m \in H$.

We claim that $c = a^m$ generates H ; that is,

$$H = \langle a^m \rangle = \langle c \rangle.$$

We must show that every $b \in H$ is a power of c . Since $b \in H$ and $H \leq G$, we have $b = a^n$ for some n . Find q and r such that

$$n = mq + r \quad \text{for} \quad 0 \leq r < m$$

in accord with the division algorithm. Then

$$a^n = a^{mq+r} = (a^m)^q a^r,$$

so

$$a^r = (a^m)^{-q} a^n.$$

Now since $a^n \in H$, $a^m \in H$, and H is a group, both $(a^m)^{-q}$ and a^n are in H . Thus

$$(a^m)^{-q} a^n \in H; \quad \text{that is,} \quad a^r \in H.$$

Since m was the smallest positive integer such that $a^m \in H$ and $0 \leq r < m$, we must have $r = 0$. Thus $n = mq$ and

$$b = a^n = (a^m)^q = c^q,$$

so b is a power of c . ◆

As noted in Examples 5.24 and 5.25, \mathbb{Z} under addition is cyclic and for a positive integer n , the set $n\mathbb{Z}$ of all multiples of n is a subgroup of \mathbb{Z} under addition, the cyclic subgroup generated by n . Theorem 6.6 shows that these cyclic subgroups are the only subgroups of \mathbb{Z} under addition. We state this as a corollary.

6.7 Corollary The subgroups of \mathbb{Z} under addition are precisely the groups $n\mathbb{Z}$ under addition for $n \in \mathbb{Z}$. ◆

This corollary gives us an elegant way to define the *greatest common divisor* of two positive integers r and s . Exercise 54 shows that $H = \{nr + ms \mid n, m \in \mathbb{Z}\}$ is a subgroup of the group \mathbb{Z} under addition. Thus H must be cyclic and have a generator d , which we may choose to be positive.

6.8 Definition Let r be a positive integer and s be a non-negative integer. The positive generator d of the cyclic group

$$H = \{nr + ms \mid n, m \in \mathbb{Z}\}$$

under addition is the **greatest common divisor** (abbreviated gcd) of r and s . We write $d = \gcd(r, s)$. ■

Note that $d\mathbb{Z} = H$, $r = 1r + 0s \in H$, and $s = 0r + 1s \in H$. This implies that $r, s \in d\mathbb{Z}$, which says that d is a divisor of both r and s . Since $d \in H$, we can write

$$d = nr + ms$$

for some integers n and m . We see that every integer dividing both r and s divides the right-hand side of the equation, and hence must be a divisor of d also. Thus d must

be the largest number dividing both r and s ; this accounts for the name given to d in Definition 6.8.

The fact that the greatest common divisor d of r and s can be written in the form $d = nr + ms$ for some integers n and m is called Bézout's identity. Bézout's identity is very useful in number theory, as we will see in studying cyclic groups.

6.9 Example Find the gcd of 42 and 72.

Solution The positive divisors of 42 are 1, 2, 3, 6, 7, 14, 21, and 42. The positive divisors of 72 are 1, 2, 3, 4, 6, 8, 9, 12, 18, 24, 36, and 72. The greatest common divisor is 6. Note that $6 = (3)(72) + (-5)(42)$. There is an algorithm for expressing the greatest common divisor d of r and s in the form $d = nr + ms$, but we will not need to make use of it here. The interested reader can find the algorithm by searching the Internet for the Euclidean algorithm and Bézout's identity. ▲

Two positive integers are **relatively prime** if their gcd is 1. For example, 12 and 25 are relatively prime. Note that they have no prime factors in common. In our discussion of subgroups of cyclic groups, we will need to know the following:

$$\text{If } r \text{ and } s \text{ are relatively prime and if } r \text{ divides } sm, \text{ then } r \text{ must divide } m. \quad (1)$$

Let's prove this. If r and s are relatively prime, then we may write

$$1 = ar + bs \quad \text{for some } a, b \in \mathbb{Z}.$$

Multiplying by m , we obtain

$$m = arm + bsm.$$

Now r divides both arm and bsm since r divides sm . Thus r is a divisor of the right-hand side of this equation, so r must divide m .

The Structure of Cyclic Groups

We can now describe all cyclic groups, up to an isomorphism.

6.10 Theorem Let G be a cyclic group with generator a . If the order of G is infinite, then G is isomorphic to $(\mathbb{Z}, +)$. If G has finite order n , then G is isomorphic to $(\mathbb{Z}_n, +_n)$.

Proof

Case I For all positive integers $m, a^m \neq e$. In this case we claim that no two distinct exponents h and k can give equal elements a^h and a^k of G . Suppose that $a^h = a^k$ and say $h > k$. Then

$$a^h a^{-k} = a^{h-k} = e,$$

contrary to our Case I assumption. Hence every element of G can be expressed as a^m for a unique $m \in \mathbb{Z}$. The map $\phi : G \rightarrow \mathbb{Z}$ given by $\phi(a^i) = i$ is thus well defined, one-to-one, and onto \mathbb{Z} . Also,

$$\phi(a^i a^j) = \phi(a^{i+j}) = i + j = \phi(a^i) + \phi(a^j),$$

so the homomorphism property is satisfied and ϕ is an isomorphism.

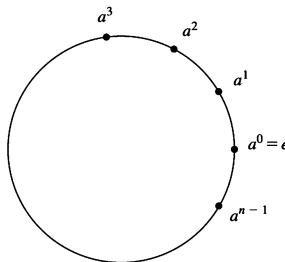
Case II $a^m = e$ for some positive integer m . Let n be the smallest positive integer such that $a^n = e$. If $s \in \mathbb{Z}$ and $s = nq + r$ for $0 \leq r < n$, then $a^s = a^{nq+r} = (a^n)^q a^r = e^q a^r = a^r$. As in Case 1, if $0 \leq k < h < n$ and $a^h = a^k$, then $a^{h-k} = e$ and $0 < h - k < n$, contradicting our choice of n . Thus the elements

$$a^0 = e, a, a^2, a^3, \dots, a^{n-1}$$

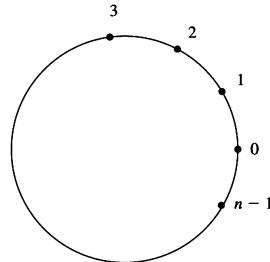
are all distinct and comprise all elements of G . The map $\psi : G \rightarrow \mathbb{Z}_n$ given by $\psi(a^i) = i$ for $i = 0, 1, 2, \dots, n - 1$ is thus well defined, one-to-one, and onto \mathbb{Z}_n . Because $a^n = e$, we see that $a^i a^j = a^k$ where $k = i +_n j$. Thus

$$\psi(a^i a^j) = i +_n j = \psi(a^i) +_n \psi(a^j),$$

so the homomorphism property is satisfied and ψ is an isomorphism. \blacklozenge



6.11 Figure



6.12 Figure

6.13 Example Motivated by our work with U_n , it is nice to visualize the elements $e = a^0, a^1, a^2, \dots, a^{n-1}$ of a cyclic group of order n as being distributed evenly on a circle (see Fig. 6.11). The element a^h is located h of these equal units counterclockwise along the circle, measured from the right where $e = a^0$ is located. To multiply a^h and a^k diagrammatically, we start from a^h and go k additional units around counterclockwise. To see arithmetically where we end up, find q and r such that

$$h + k = nq + r \quad \text{for} \quad 0 \leq r < n.$$

The nq takes us all the way around the circle q times, and we then wind up at a^r . \blacktriangle

Figure 6.12 is essentially the same as Fig. 6.11 but with the points labeled with the exponents on the generator. The operation on these exponents is *addition modulo n* .

This is simply the isomorphism between $\langle a \rangle$ and \mathbb{Z}_n . Of course this is the same isomorphism we saw when we defined \mathbb{Z}_n from U_n , but using a instead of ζ .

As promised at the beginning of this section, we can see now that the order of an element a in a group G is simply the smallest positive number n such that $a^n = e$. \blacktriangle

6.14 Example Let us find the order of the k -cycle, $\sigma = (a_1, a_2, a_3, \dots, a_k)$, in the symmetric group. The order of σ is the smallest positive power of σ that is ι . Note that applying σ just maps each number to the next one in the cyclic order. So after k applications of σ , each number maps back to itself, but not before k applications of σ . Therefore, the order of a k -cycle is k . \blacktriangle

Subgroups of Finite Cyclic Groups

We have completed our description of cyclic groups and turn to their subgroups. Corollary 6.7 gives us complete information about subgroups of infinite cyclic groups. Let us give the basic theorem regarding generators of subgroups for the finite cyclic groups.

6.15 Theorem Let G be a cyclic group with n elements and generated by a . Let $b \in G$ and let $b = a^s$. Then b generates a cyclic subgroup H of G containing n/d elements, where d is the greatest common divisor of n and s . Also, $\langle a^s \rangle = \langle a^t \rangle$ if and only if $\gcd(s, n) = \gcd(t, n)$.

Proof That b generates a cyclic subgroup H of G is known from Theorem 5.19. We need show only that H has n/d elements. Following the argument of Case II of Theorem 6.10, we see that H has as many elements as the smallest positive power m of b that gives the identity. Now $b = a^s$, and $b^m = e$ if and only if $(a^s)^m = e$, or if and only if n divides ms . What is the smallest positive integer m such that n divides ms ? Let d be the gcd of n and s . Then there exist integers u and v such that

$$d = un + vs.$$

Since d divides both n and s , we may write

$$1 = u(n/d) + v(s/d)$$

where both n/d and s/d are integers. This last equation shows that n/d and s/d are relatively prime, for any integer dividing both of them must also divide 1. We wish to find the smallest positive m such that

$$\frac{ms}{n} = \frac{m(s/d)}{(n/d)} \text{ is an integer.}$$

From the division property (1) following Example 6.9, we conclude that n/d must divide m , so the smallest such m is n/d . Thus the order of H is n/d .

Taking for the moment \mathbb{Z}_n as a model for a cyclic group of order n , we see that if d is a divisor of n , then the cyclic subgroup $\langle d \rangle$ of \mathbb{Z}_n has n/d elements, and contains all the positive integers m less than n such that $\gcd(m, n) = d$. Thus there is only one subgroup of \mathbb{Z}_n of order n/d . Taken with the preceding paragraph, this shows at once that if a is a generator of the cyclic group G , then $\langle a^s \rangle = \langle a^t \rangle$ if and only if $\gcd(s, n) = \gcd(t, n)$. ◆

6.16 Example For an example using additive notation, consider \mathbb{Z}_{12} , with the generator $a = 1$. Since the greatest common divisor of 3 and 12 is 3, $3 = 3 \cdot 1$ generates a subgroup of $\frac{12}{3} = 4$ elements, namely

$$\langle 3 \rangle = \{0, 3, 6, 9\}.$$

Since the gcd of 8 and 12 is 4, 8 generates a subgroup of $\frac{12}{4} = 3$ elements, namely,

$$\langle 8 \rangle = \{0, 4, 8\}.$$

Since the gcd of 12 and 5 is 1, 5 generates a subgroup of $\frac{12}{1} = 12$ elements; that is, 5 is a generator of the whole group \mathbb{Z}_{12} . ▲

The following corollary follows immediately from Theorem 6.15.

6.17 Corollary If a is a generator of a finite cyclic group G of order n , then the other generators of G are the elements of the form a^r , where r is relatively prime to n .

6.18 Example Let us find all subgroups of \mathbb{Z}_{18} and give their subgroup diagram. All subgroups are cyclic. By Corollary 6.17, the elements 1, 5, 7, 11, 13, and 17 are all generators of \mathbb{Z}_{18} . Starting with 2,

$$\langle 2 \rangle = \{0, 2, 4, 6, 8, 10, 12, 14, 16\}.$$

is of order 9 and has as generators elements of the form $h2$, where h is relatively prime to 9, namely, $h = 1, 2, 4, 5, 7$, and 8, so $h2 = 2, 4, 8, 10, 14$, and 16. The element 6 of $\langle 2 \rangle$ generates $\{0, 6, 12\}$, and 12 also is a generator of this subgroup.

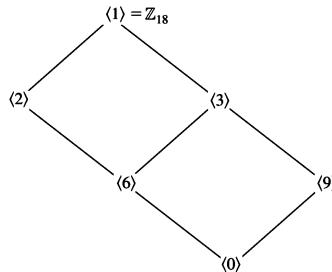
We have thus far found all subgroups generated by 0, 1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 16, and 17. This leaves just 3, 9, and 15 to consider.

$$\langle 3 \rangle = \{0, 3, 6, 9, 12, 15\},$$

and 15 also generates this group of order 6, since $15 = 5 \cdot 3$, and the gcd of 5 and 6 is 1. Finally,

$$\langle 9 \rangle = \{0, 9\}.$$

The subgroup diagram for these subgroups of \mathbb{Z}_{18} is given in Fig. 6.19.



6.19 Figure Subgroup diagram for \mathbb{Z}_{18} .

This example is straightforward; we are afraid we wrote it out in such detail that it may look complicated. The exercises give some practice along these lines. ▲

6.20 Corollary Let G be a finite cyclic group and $H \leq G$. Then $|H|$ divides $|G|$. That is, $|G|$ is a multiple of $|H|$.

Proof Let g be a generator for G and let $n = |G|$. By Theorem 6.6, H is cyclic, so there is an element in $h \in H$ such that h generates H . Since $h \in H \leq G$, $h = g^s$ for some s . Theorem 6.15 states that

$$|H| = \frac{n}{\gcd(n, s)}$$

which is a divisor of n . ◆

6.21 Example We find all orders of the subgroups of \mathbb{Z}_{28} . Factoring gives $28 = 2^2 \cdot 7$, so the possible orders of subgroups of the cyclic group \mathbb{Z}_{28} are 1, 2, 4, 7, 14, and 28. We note that $|\langle 0 \rangle| = 1$, $|\langle 14 \rangle| = 2$, $|\langle 7 \rangle| = 4$, $|\langle 4 \rangle| = 7$, $|\langle 2 \rangle| = 14$, $|\langle 1 \rangle| = |\mathbb{Z}_{28}| = 28$. So there are subgroups of order 1, 2, 4, 7, 14, and 28. ▲

Actually, Corollary 6.20 can be strengthened considerably. The assumption that G is cyclic is completely unnecessary. As we will see in Section 10, Lagrange's Theorem states that for any finite group, the order of a subgroup divides the order of the group.

■ EXERCISES 6

Computations

In Exercises 1 through 4, find the quotient and remainder, according to the division algorithm, when n is divided by m .

1. $n = 42, m = 9$ 2. $n = -42, m = 9$
 3. $n = -37, m = 8$ 4. $n = 37, m = 8$

In Exercises 5 through 7, find the greatest common divisor of the two integers.

5. 32 and 24 6. 48 and 88 7. 360 and 420

In Exercises 8 through 11, find the number of generators of a cyclic group having the given order.

8. 5 9. 8 10. 24 11. 84

An isomorphism of a group with itself is an **automorphism of the group**. In Exercises 12 through 16, find the number of automorphisms of the given group.

[Hint: You may use Exercise 53. What must be the image of a generator under an automorphism?]

12. \mathbb{Z}_2 13. \mathbb{Z}_6 14. \mathbb{Z}_8 15. \mathbb{Z} 16. \mathbb{Z}_{84}

In Exercises 17 through 23, find the number of elements in the indicated cyclic group.

17. The cyclic subgroup of \mathbb{Z}_{30} generated by 25
 18. The cyclic subgroup of \mathbb{Z}_{42} generated by 30
 19. The cyclic subgroup $\langle i \rangle$ of the group \mathbb{C}^* of nonzero complex numbers under multiplication
 20. The cyclic subgroup of the group \mathbb{C}^* of Exercise 19 generated by $(1+i)/\sqrt{2}$
 21. The cyclic subgroup of the group \mathbb{C}^* of Exercise 19 generated by $1+i$
 22. The cyclic subgroup $\langle \rho^{10} \rangle$ of D_{24}
 23. The cyclic subgroup $\langle \rho^{35} \rangle$ of D_{375}
 24. Consider the group S_{10}
 a. What is the order of the cycle $(2, 4, 6, 7)$?
 b. What is the order of $(1, 4)(2, 3, 5)$? Of $(1, 3)(2, 4, 6, 7, 8)$?
 c. What is the order of $(1, 5, 9)(2, 6, 7)$? Of $(1, 3)(2, 5, 6, 8)$?
 d. What is the order of $(1, 2)(3, 4, 5, 6, 7, 8)$? Of $(1, 2, 3)(4, 5, 6, 7, 8, 9)$?
 e. State a theorem suggested by parts (c) and (d). [Hint: The important words you are looking for are *least common multiple*.]

In Exercises 25 through 30, find the maximum possible order for an element of S_n for a given value of n .

25. $n = 5$ 26. $n = 6$ 27. $n = 7$
 28. $n = 8$ 29. $n = 10$ 30. $n = 15$

In Exercises 31 through 33, find all subgroups of the given group, and draw the subgroup diagram for the subgroups.

31. \mathbb{Z}_{12} 32. \mathbb{Z}_{36} 33. \mathbb{Z}_8

In Exercises 34 through 38, find all orders of subgroups of the given group.

34. \mathbb{Z}_6 35. \mathbb{Z}_8 36. \mathbb{Z}_{12} 37. \mathbb{Z}_{20} 38. \mathbb{Z}_{17}

Concepts

In Exercises 39 and 40, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

39. An element a of a group G has *order* $n \in \mathbb{Z}^+$ if and only if $a^n = e$.
 40. The *greatest common divisor* of two positive integers is the largest positive integer that divides both of them.

41. Determine whether each of the following is true or false.

- a. Every cyclic group is abelian.
- b. Every abelian group is cyclic.
- c. \mathbb{Q} under addition is a cyclic group.
- d. Every element of every cyclic group generates the group.
- e. There is at least one abelian group of every finite order >0 .
- f. Every group of order ≤ 4 is cyclic.
- g. All generators of \mathbb{Z}_{20} are prime numbers.
- h. If G and G' are groups, then $G \cap G'$ is a group.
- i. If H and K are subgroups of a group G , then $H \cap K$ is a group.
- j. Every cyclic group of order >2 has at least two distinct generators.

In Exercises 42 through 46, either give an example of a group with the property described, or explain why no example exists.

42. A finite abelian group that is not cyclic

43. An infinite group that is not cyclic

44. A cyclic group having only one generator

45. An infinite cyclic group having four generators

46. A finite cyclic group having four generators

The generators of the cyclic multiplicative group U_n of all n th roots of unity in \mathbb{C} are the **primitive n th roots of unity**. In Exercises 47 through 50, find the primitive n th roots of unity for the given value of n .

47. $n = 4$

48. $n = 6$

49. $n = 8$

50. $n = 12$

Proof Synopsis

51. Give a one-sentence synopsis of the proof of Theorem 6.1.

52. Give at most a three-sentence synopsis of the proof of Theorem 6.6.

Theory

53. Let G be a cyclic group with generator a , and let G' be a group isomorphic to G . If $\phi : G \rightarrow G'$ is an isomorphism, show that, for every $x \in G$, $\phi(x)$ is completely determined by the value $\phi(a)$. That is, if $\phi : G \rightarrow G'$ and $\psi : G \rightarrow G'$ are two isomorphisms such that $\phi(a) = \psi(a)$, then $\phi(x) = \psi(x)$ for all $x \in G$.

54. Let r and s be integers. Show that $\{nr + ms \mid n, m \in \mathbb{Z}\}$ is a subgroup of \mathbb{Z} .

55. Prove that if G is a finite cyclic group, H and K are subgroups of G , and $H \neq K$, then $|H| \neq |K|$.

56. Let a and b be elements of a group G . Show that if ab has finite order n , then ba also has order n .

57. Let r and s be positive integers.

a. Define the **least common multiple** of r and s as a generator of a certain cyclic group.

b. Under what condition is the least common multiple of r and s their product, rs ?

c. Generalizing part (b), show that the product of the greatest common divisor and of the least common multiple of r and s is rs .

58. Show that a group that has only a finite number of subgroups must be a finite group.
59. Show by a counterexample that the following “converse” of Theorem 6.6 is not a theorem: “If a group G is such that every proper subgroup is cyclic, then G is cyclic.”
60. Let G be a group and suppose $a \in G$ generates a cyclic subgroup of order 2 and is the *unique* such element. Show that $ax = xa$ for all $x \in G$. [Hint: Consider $(xax^{-1})^2$.]
61. Prove that if G is a cyclic group with an odd number of generators, then G has two elements.
62. Let p and q be distinct prime numbers. Find the number of generators of the cyclic group \mathbb{Z}_{pq} .
63. Let p be a prime number. Find the number of generators of the cyclic group \mathbb{Z}_{p^r} , where r is an integer ≥ 1 .
64. Show that in a finite cyclic group G of order n , written multiplicatively, the equation $x^m = e$ has exactly m solutions x in G for each positive integer m that divides n .
65. With reference to Exercise 64, what is the situation if $1 < m < n$ and m does not divide n ?
66. Show that \mathbb{Z}_p has no proper nontrivial subgroups if p is a prime number.
67. Let G be an abelian group and let H and K be finite cyclic subgroups with $|H| = r$ and $|K| = s$.
 - Show that if r and s are relatively prime, then G contains a cyclic subgroup of order rs .
 - Generalizing part (a), show that G contains a cyclic subgroup of order the least common multiple of r and s .

SECTION 7

GENERATING SETS AND CAYLEY DIGRAPHS

Let G be a group, and let $a \in G$. We have described the cyclic subgroup $\langle a \rangle$ of G , which is the smallest subgroup of G that contains the element a . Suppose we want to find as small a subgroup as possible that contains both a and b for another element b in G . By Theorem 5.19, we see that any subgroup containing a and b must contain a^n and b^m for all $m, n \in \mathbb{Z}$, and consequently must contain all finite products of such powers of a and b . For example, such an expression might be $a^2b^4a^{-3}b^2a^5$. Note that we cannot “simplify” this expression by writing first all powers of a followed by the powers of b , since G may not be abelian. However, products of such expressions are again expressions of the same type. Furthermore, $e = a^0$ and the inverse of such an expression is again of the same type. For example, the inverse of $a^2b^4a^{-3}b^2a^5$ is $a^{-5}b^{-2}a^3b^{-4}a^{-2}$. By Theorem 5.12, this shows that all such products of integral powers of a and b form a subgroup of G , which surely must be the smallest subgroup containing both a and b . We call a and b **generators** of this subgroup. If this subgroup should be all of G , then we say that $\{a, b\}$ **generates** G . Of course, there is nothing sacred about taking just two elements $a, b \in G$. We could have made similar arguments for three, four, or any number of elements of G , as long as we take only finite products of their integral powers.

7.1 Example As we have seen, the dihedral group is generated by $\{\mu, \rho\}$ since every element in D_n can be written in the form ρ^k or $\mu\rho^k$ for $0 \leq k < n$. Also, $\{\mu, \mu\rho\}$ generates D_n since $\rho = \mu(\mu\rho)$, so any element in the dihedral group can also be written as a product of copies of μ and $\mu\rho$. It is interesting to note that both μ and $\mu\rho$ have order 2, while in the generating set $\{\mu, \rho\}$ one element has order 2, but the other has order n . ▲

7.2 Example The Klein 4-group $V = \{e, a, b, c\}$ of Example 5.7 is generated by $\{a, b\}$ since $ab = c$. It is also generated by $\{a, c\}$, $\{b, c\}$, and $\{a, b, c\}$. If a group G is generated by a subset S , then every subset of G containing S generates G . ▲

7.3 Example The group \mathbb{Z}_6 is generated by $\{1\}$ and $\{5\}$. It is also generated by $\{2, 3\}$ since $2 + 3 = 5$, so that any subgroup containing 2 and 3 must contain 5 and must therefore be \mathbb{Z}_6 . It is also generated by $\{3, 4\}$, $\{2, 3, 4\}$, $\{1, 3\}$, and $\{3, 5\}$, but it is not generated by $\{2, 4\}$ since $\langle 2 \rangle = \{0, 2, 4\}$ contains 2 and 4. ▲

We have given an intuitive explanation of the subgroup of a group G generated by a subset of G . What follows is a detailed exposition of the same idea approached in another way, namely via intersections of subgroups. After we get an intuitive grasp of a concept, it is nice to try to write it up as neatly as possible. We give a set-theoretic definition and generalize a theorem that was in Exercise 61 of Section 5.

7.4 Definition Let $\{S_i \mid i \in I\}$ be a collection of sets. Here I may be any set of indices. The **intersection** $\cap_{i \in I} S_i$ of the sets S_i is the set of all elements that are in all the sets S_i ; that is,

$$\cap_{i \in I} S_i = \{x \mid x \in S_i \text{ for all } i \in I\}.$$

If I is finite, $I = \{1, 2, \dots, n\}$, we may denote $\cap_{i \in I} S_i$ by

$$S_1 \cap S_2 \cap \dots \cap S_n.$$

■

7.5 Theorem For any group G and any nonempty collection of subgroups $\{H_i \leq G \mid i \in I\}$, the intersection of all the subgroups H_i , $\cap_{i \in I} H_i$, is also a subgroup of G .

Proof Let us show closure. Let $a \in \cap_{i \in I} H_i$ and $b \in \cap_{i \in I} H_i$, so that $a \in H_i$ for all $i \in I$ and $b \in H_i$ for all $i \in I$. Then $ab \in H_i$ for all $i \in I$, since H_i is a group. Thus $ab \in \cap_{i \in I} H_i$.

Since H_i is a subgroup for all $i \in I$, we have $a^{-1} \in H_i$ for all $i \in I$, and hence $a^{-1} \in \cap_{i \in I} H_i$.

Finally, for $a \in \cap_{i \in I} H_i$, we have $a \in H_i$ for all $i \in I$, so $a^{-1} \in H_i$ for all $i \in I$, which implies that $a^{-1} \in \cap_{i \in I} H_i$. ◆

Let G be a group and let $a_i \in G$ for $i \in I$. There is at least one subgroup of G containing all the elements a_i for $i \in I$, namely G itself. Theorem 7.5 assures us that if we take the intersection of all subgroups of G containing all a_i for $i \in I$, we will obtain a subgroup H of G . This subgroup H is the smallest subgroup of G containing all the a_i for $i \in I$.

7.6 Definition Let G be a group and let $a_i \in G$ for $i \in I$. The smallest subgroup of G containing $\{a_i \mid i \in I\}$ is the **subgroup generated** by $\{a_i \mid i \in I\}$. If this subgroup is all of G , then $\{a_i \mid i \in I\}$ generates G and the a_i are **generators** of G . If there is a finite set $\{a_i \mid i \in I\}$ that generates G , then G is **finitely generated**. ■

Note that this definition is consistent with our previous definition of a generator for a cyclic group. Note also that the statement a is a generator of G may mean either that $G = \langle a \rangle$ or that a is a member of a subset of G that generates G . The context in which the statement is made should indicate which is intended. Our next theorem gives the structural insight into the subgroup of G generated by $\{a_i \mid i \in I\}$ that we discussed for two generators before Example 7.1.

7.7 Theorem If G is a group and $a_i \in G$ for $i \in I \neq \phi$, then the subgroup H of G generated by $\{a_i \mid i \in I\}$ has as elements precisely those elements of G that are finite products of integral powers of the a_i , where powers of a fixed a_i may occur several times in the product.

Proof Let K denote the set of all finite products of integral powers of the a_i . Then $K \subseteq H$. We need only observe that K is a subgroup and then, since H is the smallest subgroup containing a_i for $i \in I$, we will be done. Observe that a product of elements in K is again in K . Since $(a_i)^0 = e$, we have $e \in K$. For every element k in K , if we form from the product giving k a new product with the order of the a_i reversed and the opposite sign on all exponents, we have k^{-1} , which is thus in K . For example,

$$[(a_1)^3(a_2)^2(a_1)^{-7}]^{-1} = (a_1)^7(a_2)^{-2}(a_1)^{-3},$$

which is again in K . ◆

7.8 Example Recall that the dihedral group D_n consists of permutations of \mathbb{Z}_n that map edges to edges in the regular n -gon P_n . In disjoint cycle notation, $\rho = (0, 1, 2, 3, \dots, n-1)$ and $\mu = (1, n-1)(2, n-2) \cdots (\frac{n-1}{2}, \frac{n+1}{2})$ if n is odd, and $\mu = (1, n-1)(2, n-2) \cdots (\frac{n-2}{2}, \frac{n+2}{2})$ if n is even. Since $\mu^2 = \iota$ and $\rho^n = \iota$ any product of integer powers of μ and ρ can be rewritten to only have powers of 0 or 1 for μ and powers of 0, 1, 2, 3, ..., $n-1$ for ρ . Furthermore, the relation $\rho\mu = \mu\rho^{n-1}$ allows us to move all the powers of μ to the left and all the powers of ρ to the right, being careful to replace ρ with ρ^{n-1} each time we move a μ past a ρ . So in the case of $n = 6$,

$$\rho^8\mu^9 = \rho^2\mu = \rho\mu\rho^5 = \mu\rho^5\rho^5 = \mu\rho^4.$$

Thus the subgroup of $S_{\mathbb{Z}_n}$ generated by μ and ρ is the set

$$\{\iota, \rho, \rho^2, \dots, \rho^{n-1}, \mu, \mu\rho, \mu\rho^2, \dots, \mu\rho^{n-1}\}$$

which is the dihedral group. ▲

Cayley Digraphs

For each generating set S of a finite group G , there is a directed graph representing the group in terms of the generators in S . The term *directed graph* is usually abbreviated as *digraph*. These visual representations of groups were devised by Cayley, and are also referred to as *Cayley diagrams* in the literature.

Intuitively, a **digraph** consists of a finite number of points, called **vertices** of the digraph, and some **arcs** (each with a direction denoted by an arrowhead) joining vertices. In a digraph for a group G using a generating set S we have one vertex, represented by a dot, for each element of G . Each generator in S is denoted by one type of arc. We could use different colors for different arc types in pencil and paperwork. Since different colors are not available in our text, we use different style arcs, like solid, dashed, and dotted, to denote different generators. Thus if $S = \{a, b, c\}$ we might denote

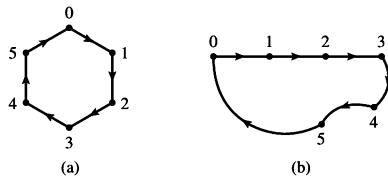
$$a \text{ by } \longrightarrow, \quad b \text{ by } \dashrightarrow\dashleftarrow, \quad \text{and} \quad c \text{ by } \cdots\cdots\rightarrow\cdots\cdots.$$

With this notation, an occurrence of $x \xrightarrow{\hspace{1cm}} y$ in a Cayley digraph means that $xa = y$. That is, traveling an arc in the direction of the arrow indicates that multiplication of the group element at the start of the arc *on the right* by the generator corresponding to that type of arc yields the group element at the end of the arc. Of course, since we are in a group, we know immediately that $ya^{-1} = x$. Thus traveling an arc in the direction opposite to the arrow corresponds to multiplication on the right by the inverse of the corresponding generator. If a generator in S is its own inverse, it is customary to denote this by omitting the arrowhead from the arc, rather than using a double arrow. For example, if $b^2 = e$, we might denote b by $\dash\dash\dash$.

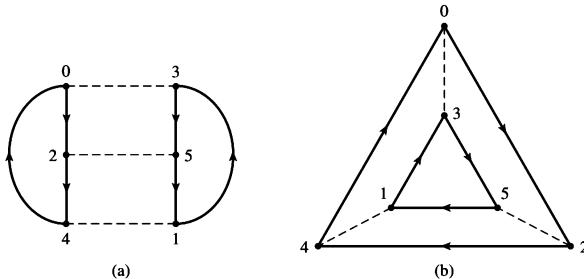
7.9 Example Both of the digraphs shown in Fig. 7.10 represent the group \mathbb{Z}_6 with generating set $S = \{1\}$. Neither the length and shape of an arc nor the angle between arcs has any significance. ▲

7.12 Example Both of the digraphs shown in Fig. 7.11 represent the group \mathbb{Z}_6 with generating set $S = \{2, 3\}$. Since 3 is its own inverse, there is no arrowhead on the dashed arcs representing 3. Notice how different these Cayley diagrams look from those in Fig. 7.10 for the same group. The difference is due to the different choice for the set of generators. ▲

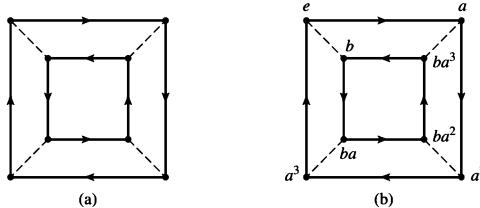
Every digraph for a group must satisfy these four properties for the reasons indicated.



7.10 Figure Two digraphs for \mathbb{Z}_6 with $S = \{1\}$ using $\xrightarrow{1}$.



7.11 Figure Two digraphs for \mathbb{Z}_6 with $S = \{2, 3\}$ using $\xrightarrow{2}$ and $\xrightarrow{3}$.



7.13 Figure

Property

1. The digraph is connected, that is, we can get from any vertex g to any vertex h by traveling along consecutive arcs, starting at g and ending at h .
2. At most one arc goes from a vertex g to a vertex h .
3. Each vertex g has exactly one arc of each type starting at g , and one of each type ending at g .
4. If two different sequences of arc types starting from vertex g lead to the same vertex h , then those same sequences of arc types starting from any vertex u will lead to the same vertex v .

Reason

Every equation $gx = h$ has a solution in a group.

The solution of $gx = h$ is unique.

For $g \in G$ and each generator b we can compute gb , and $(gb^{-1})b = g$.

If $gq = h$ and $gr = h$, then $ug = ug^{-1}h = ur$.

It can be shown that, conversely, every digraph satisfying these four properties is a Cayley digraph for some group. Due to the symmetry of such a digraph, we can choose labels like a, b, c for the various arc types, name any vertex e to represent the identity, and name each other vertex by a product of arc labels and their inverses that we can travel to attain that vertex starting from the one that we named e . Some finite groups were first constructed (found) using digraphs.

- 7.14 Example** A digraph satisfying the four properties given above is shown in Fig. 7.13 (a). To obtain Fig. 7.13 (b), we selected the labels

\xrightarrow{a} and \xleftarrow{b} ,

named a vertex e , and then named the other vertices as shown. We have a group

$$\{e, a, a^2, a^3, b, ba, ba^2, ba^3\}$$

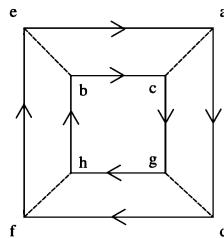
of eight elements. From the diagram we could compute any product. For example, to compute ba^2ba^3 we start at the vertex labeled ba^2 , follow a dotted edge, and then follow three solid edges to arrive at a . Note that the way we labeled the vertices is not unique. For example, the vertex labeled ba^3 could have been labeled ab simply by going along a different path starting at e . This says that $ab = ba^3$. We also see that $a^4 = e$ and $b^2 = e$. We hope that this example is starting to look familiar. In fact, Figure 7.13 is a Cayley digraph of the dihedral group D_4 . We simply relabel a with ρ and b with μ ! \blacktriangle

■ EXERCISES 7

Computations

In Exercises 1 through 8, list the elements of the subgroup generated by the given subset.

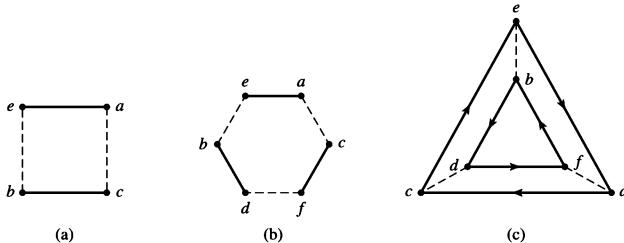
1. The subset $\{2, 3\}$ of \mathbb{Z}_{12}
2. The subset $\{4, 6\}$ of \mathbb{Z}_{12}
3. The subset $\{4, 6\}$ in \mathbb{Z}_{25}
4. The subset $\{12, 30\}$ of \mathbb{Z}_{36}
5. The subset $\{12, 42\}$ of \mathbb{Z}
6. The subset $\{18, 24, 39\}$ of \mathbb{Z}
7. The subset $\{\mu, \mu\rho^2\}$ in D_8
8. The subset $\{\rho^8, \rho^{10}\}$ in D_{18}
9. Use the Cayley digraph in Figure 7.15 to compute these products. Note that the solid edges represent the generator a and the dashed lines represent b .
 - a. $(ba^2)a^3$
 - b. $(ba)(ba^3)$
 - c. $b(a^2b)$



7.15 Figure

In Exercises 10 through 12, give the table for the group having the indicated digraph. In each digraph, take e as identity element. List the identity e first in your table, and list the remaining elements alphabetically, so that your answers will be easy to check.

10. The digraph in Fig. 7.16(a)
11. The digraph in Fig. 7.16(b)
12. The digraph in Fig. 7.16(c)



7.16 Figure

Concepts

13. How can we tell from a Cayley digraph whether or not the corresponding group is commutative?
14. Using the condition found in Exercise 13, show that the group corresponding to the Cayley digraph in Figure 7.13 is not commutative.
15. Is it obvious from a Cayley digraph of a group whether or not the group is cyclic? [Hint: Look at Fig. 7.9(b).]
16. The large outside triangle in Fig. 7.11(b) exhibits the cyclic subgroup $\{0, 2, 4\}$ of \mathbb{Z}_6 . Does the smaller inside triangle similarly exhibit a cyclic subgroup of \mathbb{Z}_6 ? Why or why not?
17. The generating set $S = \{1, 2\}$ for \mathbb{Z}_6 contains more generators than necessary, since 1 is a generator for the group. Nevertheless, we can draw a Cayley digraph for \mathbb{Z}_6 with this generating set S . Draw such a Cayley digraph.
18. Draw a Cayley digraph for \mathbb{Z}_8 with generating set $S = \{2, 5\}$.
19. A **relation** on a set S of generators of a group G is an equation that equates some product of generators and their inverses to the identity e of G . For example, if $S = \{a, b\}$ and G is commutative so that $ab = ba$, then one relation is $aba^{-1}b^{-1} = e$. If, moreover, b is its own inverse, then another relation is $b^2 = e$.
 - a. Explain how we can find some relations on S from a Cayley digraph of G .
 - b. Find three relations on the set $S = \{a, b\}$ of generators for the group described by Fig. 7.13(b).
20. Draw digraphs of the two possible structurally different groups of order 4, taking as small a generating set as possible in each case. You need not label vertices.

Theory

21. Use Cayley digraphs to show that for $n \geq 3$, there exists a nonabelian group with $2n$ elements that is generated by two elements of order 2.
22. Prove that there are at least three different abelian groups of order 8. [Hint: Find a Cayley digraph for a group of order 8 having one generator of order 4 and another of order 2. Find a second Cayley digraph for a group of order 8 having three generators each with order 2.]

This page is intentionally left blank

Structure of Groups

-
- Section 8** Groups of Permutations
 - Section 9** Finitely Generated Abelian Groups
 - Section 10** Cosets and the Theorem of Lagrange
 - Section 11** Plane Isometries

SECTION 8

GROUPS OF PERMUTATIONS

Let $\phi : G \rightarrow G'$ be a function mapping the group G to G' . Recall that the homomorphism property of an isomorphism states that for all $a, b \in G$, $\phi(ab) = \phi(a)\phi(b)$. Whenever a function has this property whether or not the function is one-to-one or onto, we say that ϕ is a **group homomorphism**. Of course any group isomorphism is a group homomorphism, but the reverse is not necessarily true.

8.1 Definition Let G and G' be groups with $\phi : G \rightarrow G'$. The map ϕ is a **homomorphism** if the homomorphism property

$$\phi(ab) = \phi(a)\phi(b)$$

holds for all $a, b \in G$. ■

8.2 Example Let $\phi : \mathbb{R} \rightarrow U$ (the circle group) be defined by the formula

$$\phi(x) = \cos(2\pi x) + i \sin(2\pi x) = e^{2\pi ix}.$$

Then

$$\phi(a + b) = \cos(2\pi(a + b)) + i \sin(2\pi(a + b)) = e^{2\pi i(a+b)}.$$

Using either the usual properties of the exponential function or the formulas from trigonometry involving the sum of two angles, we see that

$$\phi(a + b) = (\cos(2\pi a) + i \sin(2\pi a))(\cos(2\pi b) + i \sin(2\pi b)) = e^{2\pi ai}e^{2\pi bi},$$

so

$$\phi(a + b) = \phi(a)\phi(b),$$

which says that ϕ is a group homomorphism. Although ϕ maps onto U , it is not one-to-one, so ϕ is not an isomorphism.

The identity $0 \in \mathbb{R}$ maps to 1, the identity in U . Furthermore, for any $x \in \mathbb{R}$,

$$\phi(-x) = e^{-2\pi ix} = \frac{1}{e^{2\pi ix}} = (\phi(x))^{-1}.$$



8.3 Example Recall that $U_n = \{z \in \mathbb{C} \mid z^n = 1\}$. Let $\phi : U_{28} \rightarrow U_4$ be given by $\phi(z) = z^7$. To check that ϕ is well defined, we see that if $z \in U_{28}$, then $z^{28} = 1$. Therefore, $(z^7)^4 = 1$, which implies that $z^7 \in U_4$. We check that ϕ is a homomorphism.

$$\phi(z_1 z_2) = (z_1 z_2)^7 = z_1^7 z_2^7 = \phi(z_1) \phi(z_2).$$

As in the previous example, ϕ maps the identity in U_{28} , in this case 1, to the identity 1 in U_4 . Furthermore,

$$\phi(z^{-1}) = z^{-7} = (z^7)^{-1} = (\phi(z))^{-1}. \quad \blacktriangle$$

8.4 Definition Let $\phi : X \rightarrow Y$ and suppose that $A \subseteq X$ and $B \subseteq Y$. The set $\phi[A] = \{\phi(a) \mid a \in A\}$ is called the **image of A in Y under the mapping ϕ** . The set $\phi^{-1}[B] = \{a \in A \mid \phi(a) \in B\}$ is called the **inverse image of B under the mapping ϕ** . ■

The four properties of a homomorphism given in the theorem that follows are obvious in the case of an isomorphism since we think of an isomorphism as simply relabeling the elements of a group. However, it is not obvious that these properties hold for all homomorphisms whether or not they are one-to-one or onto maps. Consequently, we give careful proofs of all four properties.

8.5 Theorem Let ϕ be a homomorphism of a group G into a group G' .

1. If e is the identity element in G , then $\phi(e)$ is the identity element e' in G' .
2. If $a \in G$, then $\phi(a^{-1}) = \phi(a)^{-1}$.
3. If H is a subgroup of G , then $\phi[H]$ is a subgroup of G' .
4. If K' is a subgroup of G' , then $\phi^{-1}[K']$ is a subgroup of G .

Loosely speaking, ϕ preserves the identity element, inverses, and subgroups.

Proof Let ϕ be a homomorphism of G into G' . Then

$$\phi(e) = \phi(ee) = \phi(e)\phi(e).$$

Multiplying on the left by $\phi(e)^{-1}$, we see that $e' = \phi(e)$. Thus $\phi(e)$ must be the identity element e' in G' . The equation

$$e' = \phi(e) = \phi(aa^{-1}) = \phi(a)\phi(a^{-1})$$

shows that $\phi(a^{-1}) = \phi(a)^{-1}$ for all $a \in G$.

Turning to Statement (3), let H be a subgroup of G , and let $\phi(a)$ and $\phi(b)$ be any two elements in $\phi[H]$. Then $\phi(a)\phi(b) = \phi(ab)$, so we see that $\phi(a)\phi(b) \in \phi[H]$; thus, $\phi[H]$ is closed under the operation of G' . The fact that $e' = \phi(e)$ and $\phi(a^{-1}) = \phi(a)^{-1}$ completes the proof that $\phi[H]$ is a subgroup of G' .

Going the other way for Statement (4), let K' be a subgroup of G' . Suppose a and b are in $\phi^{-1}[K']$. Then $\phi(a)\phi(b) \in K'$ since K' is a subgroup. The equation $\phi(ab) = \phi(a)\phi(b)$ shows that $ab \in \phi^{-1}[K']$. Thus $\phi^{-1}[K']$ is closed under the binary operation in G . Also, K' must contain the identity element $e' = \phi(e)$, so $e \in \phi^{-1}[K']$. If $a \in \phi^{-1}[K']$, then $\phi(a) \in K'$, so $\phi(a)^{-1} \in K'$. But $\phi(a)^{-1} = \phi(a^{-1})$, so we must have $a^{-1} \in \phi^{-1}[K']$. Hence $\phi^{-1}[K']$ is a subgroup of G . ◆

Let $\phi : G \rightarrow G'$ be a homomorphism and let e' be the identity element of G' . Now $\{e'\}$ is a subgroup of G' , so $\phi^{-1}[\{e'\}]$ is a subgroup H of G by Statement (4) in Theorem 8.5. This subgroup is critical to the study of homomorphisms.

8.6 Definition Let $\phi : G \rightarrow G'$ be a homomorphism of groups. The subgroup $\phi^{-1}[\{e'\}] = \{x \in G \mid \phi(x) = e'\}$ is the **kernel of ϕ** , denoted by $\text{Ker}(\phi)$. ■

We will use the kernel of a homomorphism when we define the alternating group later in this section.

Another extreme is to let $H = G$ in Statement (3) of Theorem 8.5. In this case, the theorem says that $\phi[G]$ is a subgroup of G' . We use this in the proof of Cayley's Theorem.

8.7 Example In Example 8.2, the homomorphism $\phi : \mathbb{R} \rightarrow U$ is defined by $\phi(x) = \cos(2\pi x) + i \sin(2\pi x) = e^{2\pi ix}$. The kernel of ϕ is the set of integers since $\cos(2\pi x) + i \sin(2\pi x) = 1$ if and only if x is an integer.

Let n be a positive integer. Then $\left\langle \frac{1}{n} \right\rangle$ is a subgroup of \mathbb{R} and

$$\begin{aligned}\phi \left[\left\langle \frac{1}{n} \right\rangle \right] &= \phi \left[\left\{ \frac{m}{n} \mid m \in \mathbb{Z} \right\} \right] \\ &= \{ \cos(2\pi m/n) + i \sin(2\pi m/n) \mid m \in \mathbb{Z} \} \\ &= U_n.\end{aligned}$$

▲

8.8 Example Let $\phi : \mathbb{Z}_n \rightarrow D_n$ be given by $\phi(k) = \rho^k$. We check that ϕ is a homomorphism. Let $a, b \in \mathbb{Z}_n$. If $a + b < n$, then $a +_n b = a + b$, so $\phi(a +_n b) = \phi(a + b) = \rho^{a+b} = \rho^a \rho^b = \phi(a)\phi(b)$. If $a + b \geq n$, then $\phi(a +_n b) = \phi(a + b - n) = \rho^{a+b-n} = \rho^a \rho^b \rho^{-n} = \rho^a \rho^b = \phi(a)\phi(b)$. The image $\phi[\mathbb{Z}_n]$ is $\langle \rho \rangle$.

▲

Cayley's Theorem

Each of the groups we have seen so far is isomorphic to a subgroup of permutations on some set. For example, \mathbb{Z}_n is isomorphic with the cyclic group $\langle (1, 2, 3, \dots, n) \rangle \leq S_{\mathbb{Z}_n}$. The dihedral group D_n is defined to be the permutations in $S_{\mathbb{Z}_n}$ with the property that the line segment between vertices i and j is an edge in P_n , a regular n -gon, if and only if the line segment between the images of i and j is also an edge. The infinite group $GL(n, \mathbb{R})$ can be thought of as invertible linear transformations of \mathbb{R}^n . Each element of $GL(n, \mathbb{R})$ permutes the vectors in \mathbb{R}^n , which makes $GL(n, \mathbb{R})$ isomorphic with a permutation group on vectors in \mathbb{R}^n . We refer to a subgroup of a permutation group as a **group of permutations**. Cayley's Theorem states that any group is isomorphic with a group of permutations.

At first Cayley's Theorem seems like a remarkable result that could be used to understand all groups. In fact, this is a nice and intriguing classic result. Unfortunately, approaching group theory by trying to determine all possible permutation groups is not feasible. On the other hand, Cayley's theorem does show the strength and generality of permutation groups and it deserves a special place in group theory for that reason. For example, if we wish to find a counterexample to a conjecture about groups, provided that there is one, it will occur in a permutation group.

It may seem a mystery how we could start with an arbitrary group and come up with a permutation group that is isomorphic with the given group. The key is to think about the group table. Each row contains each element of the group exactly once. So each row defines a permutation of the elements of the group by placing the table head as the top row in the two-row representation of a permutation and placing the row corresponding to an element a in the group as the bottom row. Table 8.9 is the group table for D_3 . Note that the permutation obtained using the row $\mu\rho$ is

$$\begin{pmatrix} \iota & \rho & \rho^2 & \mu & \mu\rho & \mu\rho^2 \\ \mu\rho & \mu\rho^2 & \mu & \rho^2 & \iota & \rho \end{pmatrix}.$$

8.9 Table

D_3	ι	ρ	ρ^2	μ	$\mu\rho$	$\mu\rho^2$
ι	ι	ρ	ρ^2	μ	$\mu\rho$	$\mu\rho^2$
ρ	ρ	ρ^2	ι	$\mu\rho^2$	μ	$\mu\rho$
ρ^2	ρ^2	ι	ρ	$\mu\rho$	$\mu\rho^2$	μ
μ	μ	$\mu\rho$	$\mu\rho^2$	ι	ρ	ρ^2
$\mu\rho$	$\mu\rho$	$\mu\rho^2$	μ	ρ^2	ι	ρ
$\mu\rho^2$	$\mu\rho^2$	μ	$\mu\rho$	ρ	ρ^2	ι

All that remains to prove Cayley's Theorem, at least when the group is finite, is to check that the permutations obtained from the group table form a group isomorphism with the original group. Let λ_x be the permutation of the elements of G given by the x row of the table for G . Then for any $g \in G$, $\lambda_x(g)$ is the entry in the x row and g column of the group table. In other words, $\lambda_x(g) = xg$, which is perfectly valid in the case of an infinite as well as a finite group. We formalize this connection between G and permutations on G in Definition 8.10.

8.10 Definition Let G be a group. The function $\phi : G \rightarrow S_G$ given by $\phi(x) = \lambda_x$ where $\lambda_x(g) = xg$ for all $g \in G$ is called the **left regular representation** of G . ■

In order to be sure that λ_x is a permutation, it should be verified that λ_x is both one-to-one and onto. We see that λ_x is one-to-one since if $\lambda_x(a) = \lambda_x(b)$, $xa = xb$ and cancellation gives $a = b$. Also, λ_x maps onto G because for any $b \in G$, $\lambda_x(x^{-1}b) = b$. We are now ready to prove Cayley's Theorem.

8.11 Theorem (**Cayley's Theorem**) Every group is isomorphic to a group of permutations.

Proof Let G be a group. The left regular representation provides a map $\phi : G \rightarrow S_G$ defined by $\phi(x) = \lambda_x$. We must verify that ϕ is a group homomorphism and that ϕ is one-to-one. Then $\phi[G]$ is a subgroup of S_G by Theorem 8.5 and $\phi : G \rightarrow \phi[G]$ is an isomorphism.

We first show that ϕ is one-to-one. Suppose that $a, b \in G$ and $\phi(a) = \phi(b)$. Then the permutations λ_a and λ_b are the same, so $\lambda_a(e) = \lambda_b(e)$. Thus $ae = be$ and $a = b$. So ϕ is one-to-one.

We now need to show that ϕ is a group homomorphism. Let $a, b \in G$. Then $\phi(ab) = \lambda_{ab}$ and $\phi_a\phi_b = \lambda_a\lambda_b$. We must show that the two permutations λ_{ab} and $\lambda_a\lambda_b$ are the same. Let $g \in G$.

$$\lambda_{ab}(g) = (ab)g = a(bg) = \lambda_a(bg) = \lambda_a(\lambda_b(g)) = (\lambda_a\lambda_b)(g).$$

Thus $\lambda_{ab} = \lambda_a\lambda_b$, which implies that $\phi(ab) = \phi(a)\phi(b)$. So ϕ is a one-to-one homomorphism, which completes the proof. ◆

8.12 Example The proof of Cayley's Theorem shows that any group G is isomorphic with a subgroup of S_G , but this is typically not the smallest symmetric group that has a subgroup isomorphic with G . For example, D_n is isomorphic with a subgroup of $S_{\mathbb{Z}_n}$ while the proof of Cayley's Theorem gives a subgroup of S_{D_n} and D_n has $2n$ elements while \mathbb{Z}_n has only n elements. On the surface, it may seem that \mathbb{Z}_6 cannot be isomorphic with a subgroup of S_n for $n < 6$, but $(1, 2, 3)(4, 5) \in S_5$ generates a subgroup isomorphic with \mathbb{Z}_6 . ▲

We defined the left regular representation in Definition 8.10. We now define the right regular representation. Instead of λ_x representing the row for x in the group table, we use σ_x to represent the column with head x . Instead of using ϕ for the function that sends x to λ_x , we use τ , which sends x to $\sigma_{x^{-1}}$.

■ HISTORICAL NOTE

Arthur Cayley (1821–1895) gave an abstract-sounding definition of a group in a paper of 1854: “A set of symbols, $1, \alpha, \beta, \dots$, all of them different and such that the product of any two of them (no matter in what order) or the product of any one of them into itself, belongs to the set, is said to be a group.” He then proceeded to define a group table and note that every line and column of the table “will contain all the symbols $1, \alpha, \beta, \dots$.” Cayley’s symbols, however, always represented operations on sets; it does not seem that he was aware of any other kind of group. He noted, for instance, that the four matrix operations $1, \alpha = \text{inversion}$, $\beta = \text{transposition}$, and $\gamma = \alpha\beta$, form, abstractly, the non-cyclic group of four elements. In any case, his definition went unnoticed for a quarter of a century.

This paper of 1854 was one of about 300 written during the 14 years Cayley was practicing law,

being unable to find a suitable teaching post. In 1863, he finally became a professor at Cambridge. In 1878, he returned to the theory of groups by publishing four papers, in one of which he stated Theorem 8.11 of this text; his “proof” was simply to notice from the group table that multiplication by any group element permuted the group elements. However, he wrote, “this does not in any wise show that the best or the easiest mode of treating the general problem [of finding all groups of a given order] is thus to regard it as a problem of [permutations]. It seems clear that the better course is to consider the general problem in itself.”

The papers of 1878, unlike the earlier one, found a receptive audience; in fact, they were an important influence on Walther von Dyck’s 1882 axiomatic definition of an abstract group, the definition that led to the development of abstract group theory.

8.13 Definition Let G be a group. The map $\tau : G \rightarrow S_G$ given by $\tau(x) = \sigma_{x^{-1}}$ where $\sigma_x(g) = gx$ is called the **right regular representation** of G . ■

We could have used the right regular representation to prove Cayley’s Theorem instead of the left regular representation. Exercise 54 asks for the details of the proof.

Even and Odd Permutations

It seems reasonable that every reordering of the sequence $1, 2, \dots, n$ can be achieved by repeated interchange of positions of pairs of numbers. We discuss this a bit more formally.

8.14 Definition A cycle of length 2 is a **transposition**. ■

Thus a transposition leaves all elements but two fixed, and maps each of these onto the other. A computation shows that

$$(a_1, a_2, \dots, a_n) = (a_1, a_n)(a_1, a_{n-1}) \cdots (a_1, a_3)(a_1, a_2).$$

Therefore any cycle of length n can be written as a product of $n - 1$ transpositions. Since any permutation of a finite set can be written as a product of cycles, we have the following.

8.15 Theorem Any permutation of a finite set containing at least two elements is a product of transpositions. ♦

Naively, this theorem just states that any rearrangement of n objects can be achieved by successively interchanging pairs of them.

8.16 Example Following the remarks prior to the theorem, we see that $(1, 6)(2, 5, 3)$ is the product $(1, 6)(2, 3)(2, 5)$ of transpositions. ▲

8.17 Example In S_n for $n \geq 2$, the identity permutation is the product $(1, 2)(1, 2)$ of transpositions. ▲

We have seen that every permutation of a finite set with at least two elements is a product of transpositions. The transpositions may not be disjoint, and a representation of the permutation in this way is not unique. For example, we can always insert at the beginning the transposition $(1, 2)$ twice, because $(1, 2)(1, 2)$ is the identity permutation. What is true is that the number of transpositions used to represent a given permutation must either always be even or always be odd. This is an important fact. The proof involves counting orbits and was suggested by David M. Bloom.

Let $\sigma \in S_A$ and $a \in A$. We let the **orbit** of a be the set $\{\sigma^k(a) \mid k \in \mathbb{Z}\}$. In the case of $\sigma \in S_n$, a simple way to think of the orbit of a is to think of the elements in the cycle containing a in the disjoint cycle representation of σ .

8.18 Example Let $\sigma = (1, 2, 6)(3, 5) \in S_6$. Then the orbit of 1 is the set $\{1, 2, 6\}$, which is also the orbit of 2 and the orbit of 6. The set $\{3, 5\}$ is the orbit of 3 and the orbit of 5. What about the orbit of 4? Recall that if we include 1-cycles, $\sigma = (1, 2, 6)(3, 5)(4)$, which says the orbit of 4 is $\{4\}$. ▲

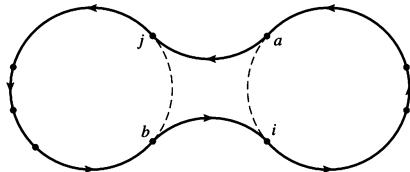
8.19 Theorem No permutation in S_n can be expressed both as a product of an even number of transpositions and as a product of an odd number of transpositions.

Proof Let $\sigma \in S_n$ and let $\tau = (i, j)$ be a transposition in S_n . We claim that the number of orbits of σ and of $\tau\sigma$ differ by 1.

Case I Suppose i and j are in different orbits of σ . Write σ as a product of disjoint cycles, the first of which contains j and the second of which contains i , symbolized by the two circles in Fig. 8.20. We may write the product of these two cycles symbolically as

$$(b, j, \times, \times, \times)(a, i, \times, \times, \times)$$

where the symbols \times denote possible other elements in these orbits.



8.20 Figure

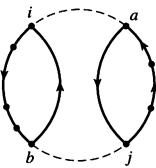
Computing the product of the first three cycles in $\tau\sigma = (i, j)\sigma$, we obtain

$$(i, j)(b, j, \times, \times, \times)(a, i, \times, \times, \times) = (a, j, \times, \times, \times, b, i, \times, \times).$$

The original 2 orbits have been joined to form just one in $\tau\sigma$ as symbolized in Fig. 8.20. Exercise 42 asks us to repeat the computation to show that the same thing happens if either one or both of i and j should be the only element of their orbit in σ .

Case II Suppose i and j are in the same orbit of σ . We can then write σ as a product of disjoint cycles with the first cycle of the form

$$(a, i, \times, \times, \times, b, j, \times, \times)$$



8.21 Figure

shown symbolically by the circle in Fig. 8.20. Computing the product of the first two cycles in $\tau\sigma = (i,j)\sigma$, we obtain

$$(i,j)(a, i, \times, \times, \times, b, j, \times, \times) = (a, j, \times, \times)(b, i, \times, \times, \times).$$

The original single orbit has been split into two as symbolized in Fig. 8.21.

We have shown that the number of orbits of $\tau\sigma$ differs from the number of orbits of σ by 1. The identity permutation ι has n orbits, because each element is the only member of its orbit. Now the number of orbits of a given permutation $\sigma \in S_n$ differs from n by either an even or an odd number, but not both. Thus it is impossible to write

$$\sigma = \tau_1 \tau_2 \tau_3 \cdots \tau_m \iota$$

where the τ_k are transpositions in two ways, once with m even and once with m odd. ◆

8.22 Definition A permutation of a finite set is **even** or **odd** according to whether it can be expressed as a product of an even number of transpositions or the product of an odd number of transpositions, respectively. ■

8.23 Example The identity permutation ι in S_n is an even permutation since we have $\iota = (1,2)(1,2)$. If $n = 1$ so that we cannot form this product, we define ι to be even. On the other hand, the permutation $(1, 4, 5, 6)(2, 1, 5)$ in S_6 can be written as

$$(1, 4, 5, 6)(2, 1, 5) = (1, 6)(1, 5)(1, 4)(2, 5)(2, 1)$$

which has five transpositions, so this is an odd permutation. ▲

The Alternating Groups

We claim that for $n \geq 2$, the number of even permutations in S_n is the same as the number of odd permutations; that is, S_n is split equally and both numbers are $(n!)/2$. To show this, let A_n be the set of even permutations in S_n and let B_n be the set of odd permutations for $n \geq 2$. We proceed to define a one-to-one function from A_n onto B_n . This is exactly what is needed to show that A_n and B_n have the same number of elements.

Let τ be any fixed transposition in S_n ; it exists since $n \geq 2$. We may as well suppose that $\tau = (1,2)$. We define a function

$$\lambda_\tau : A_n \rightarrow B_n$$

by

$$\lambda_\tau(\sigma) = \tau\sigma,$$

that is, $\sigma \in A_n$ is mapped into $(1,2)\sigma$ by λ_τ . Observe that since σ is even, the permutation $(1,2)\sigma$ can be expressed as a product of a (1 + even number), or odd number, of transpositions, so $(1,2)\sigma$ is indeed in B_n . If for σ and μ in A_n it is true that $\lambda_\tau(\sigma) = \lambda_\tau(\mu)$, then

$$(1,2)\sigma = (1,2)\mu,$$

and since S_n is a group, we have $\sigma = \mu$. Thus λ_τ is a one-to-one function. Finally,

$$\tau = (1,2) = \tau^{-1},$$

so if $\rho \in B_n$, then

$$\tau^{-1}\rho \in A_n,$$

and

$$\lambda_\tau(\tau^{-1}\rho) = \tau(\tau^{-1}\rho) = \rho.$$

Thus λ_τ maps onto B_n . Hence the number of elements in A_n is the same as the number in B_n since there is a one-to-one correspondence between the elements of the sets.

Note that the product of two even permutations is again even. Also since $n \geq 2$, S_n has the transposition $(1, 2)$ and $\iota = (1, 2)(1, 2)$ is an even permutation. Finally, note that if σ is expressed as a product of transpositions, the product of the same transpositions taken in just the opposite order is σ^{-1} . Thus if σ is an even permutation, σ^{-1} must also be even. Referring to Theorem 5.12, we see that we have proved the following statement.

8.24 Theorem If $n \geq 2$, then the collection of all even permutations of $\{1, 2, 3, \dots, n\}$ forms a subgroup of order $n!/2$ of the symmetric group S_n .

We can define a function called the **sign of a permutation**, $\text{sgn} : S_n \rightarrow \{1, -1\}$ by the formula

$$\text{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

Thinking of $\{1, -1\}$ as a group under multiplication, it is easy to see that sgn is a homomorphism. Since 1 is the identity in the group $\{1, -1\}$, $\text{Ker}(\text{sgn}) = \text{sgn}^{-1}[\{1\}]$ is a subgroup of S_n consisting of all the even permutations. The homomorphism sgn is used in the standard way of defining the determinant of a square matrix. Exercise 52 asks you to prove some of the standard facts about determinants using this definition.

8.25 Definition The subgroup of S_n consisting of the even permutations of n letters is the **alternating group A_n on n letters**. ■

Both S_n and A_n are very important groups. Cayley's theorem shows that every finite group G is structurally identical to some subgroup of S_n for $n = |G|$. It can be shown that there are no formulas involving just radicals for solution of polynomial equations of degree n for $n \geq 5$. This fact is actually due to the structure of A_n , surprising as that may seem!

■ EXERCISES 8

Computations

In Exercises 1 through 10 determine whether the given map is a group homomorphism. [Hint: To verify that a map is a homomorphism, you must check the homomorphism property. To check that a map is not a homomorphism you could either find a and b such that $\phi(ab) \neq \phi(a)\phi(b)$, or else you could determine that any of the properties in Theorem 8.5 fail.]

1. Let $\phi : \mathbb{Z}_{10} \rightarrow \mathbb{Z}_2$ be given by $\phi(x) =$ the remainder when x is divided by 2.
2. Let $\phi : \mathbb{Z}_9 \rightarrow \mathbb{Z}_2$ be given by $\phi(x) =$ the remainder when x is divided by 2.
3. Let $\phi : \mathbb{Q}^* \rightarrow \mathbb{Q}^*$ be given by $\phi(x) = |x|$.
4. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}^+$ be given by $\phi(x) = 2^x$.
5. Let $\phi : D_4 \rightarrow \mathbb{Z}_4$ be given by $\phi(\rho^i) = \phi(\mu\rho^i) = i$ for $0 \leq i \leq 3$.

6. Let F be the additive group of all functions mapping \mathbb{R} to \mathbb{R} . Let $\phi : F \rightarrow F$ be given by $\phi(f) = g$ where $g(x) = f(x) + x$.
7. Let F be as in Exercise 6 and $\phi : F \rightarrow F$ be defined by $\phi(f) = 5f$.
8. Let F be the additive group of all continuous functions mapping \mathbb{R} to \mathbb{R} . Let $\phi : F \rightarrow \mathbb{R}$ be defined by $\phi(g) = \int_0^1 g(x) dx$.
9. Let M_n be the additive group of $n \times n$ matrices with real entries. Let $\phi : M_n \rightarrow \mathbb{R}$ be given by $\phi(A) = \det(A)$, the determinant of A .
10. Let M_n be as in Exercise 9 and $\phi : M_n \rightarrow \mathbb{R}$ be defined by $\phi(A) = \text{tr}(A)$ where $\text{tr}(A)$ is the trace of A , which is the sum of the entries on the diagonal.

In Exercises 11 through 16, compute the kernel for the given homomorphism ϕ .

11. $\phi : \mathbb{Z} \rightarrow \mathbb{Z}_8$ such that $\phi(1) = 6$.
12. $\phi : \mathbb{Z} \rightarrow \mathbb{Z}$ such that $\phi(1) = 12$.
13. $\phi : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ where $\phi(1, 0) = 3$ and $\phi(0, 1) = -5$.
14. $\phi : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ where $\phi(1, 0) = 6$ and $\phi(0, 1) = 9$.
15. $\phi : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z} \times \mathbb{Z}$ where $\phi(1, 0) = (2, 5)$ and $\phi(0, 1) = (-3, 2)$.
16. Let D be the additive group of all differentiable functions mapping \mathbb{R} to \mathbb{R} and F the additive group of all functions from \mathbb{R} to \mathbb{R} . $\phi : D \rightarrow F$ is given by $\phi(f) = f'$, the derivative of f .

In Exercises 17 through 22, find all orbits of the given permutation.

17. $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 1 & 3 & 6 & 2 & 4 \end{pmatrix}$

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

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

20. $\sigma : \mathbb{Z} \rightarrow \mathbb{Z}$ where $\sigma(n) = n + 1$

21. $\sigma : \mathbb{Z} \rightarrow \mathbb{Z}$ where $\sigma(n) = n + 2$

22. $\sigma : \mathbb{Z} \rightarrow \mathbb{Z}$ where $\sigma(n) = n - 3$

In Exercises 23 through 25, express the permutation of $\{1, 2, 3, 4, 5, 6, 7, 8\}$ as a product of disjoint cycles, and then as a product of transpositions.

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

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

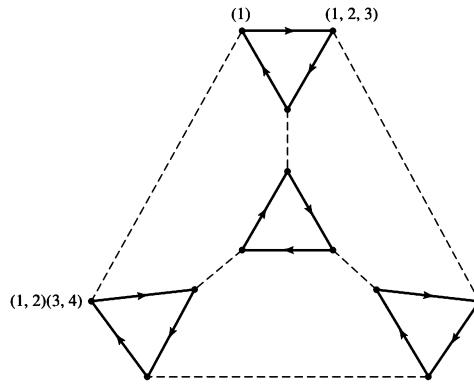
25. $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 3 & 2 & 8 & 4 & 7 & 6 & 1 \end{pmatrix}$

26. Figure 8.26 shows a Cayley digraph for the alternating group A_4 using the generating set $S = \{(1, 2, 3), (1, 2)(3, 4)\}$. Continue labeling the other nine vertices with the elements of A_4 , expressed as a product of disjoint cycles.

Concepts

In Exercises 27 through 29, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

27. For a permutation σ of a set A , an *orbit* of σ is a nonempty minimal subset of A that is mapped onto itself by σ .
28. The left regular representation of a group G is the map of G into S_G whose value at $g \in G$ is the permutation of G that carries each $x \in G$ into gx .
29. The *alternating group* is the group of all even permutations.



8.26 Figure

30. Before the proof of Cayley's Theorem, it is shown that λ_x is one-to-one. In the proof, one-to-one is shown again. Is it necessary to show one-to-one twice? Explain.
31. Determine whether each of the following is true or false.
- Every permutation is a cycle.
 - Every cycle is a permutation.
 - The definition of even and odd permutations could have been given equally well before Theorem 8.19.
 - Every nontrivial subgroup H of S_9 containing some odd permutation contains a transposition.
 - A_5 has 120 elements.
 - S_n is not cyclic for any $n \geq 1$.
 - A_3 is a commutative group.
 - S_7 is isomorphic to the subgroup of all those elements of S_8 that leave the number 8 fixed.
 - S_7 is isomorphic to the subgroup of all those elements of S_8 that leave the number 5 fixed.
 - The odd permutations in S_8 form a subgroup of S_8 .
 - Every group G is isomorphic with a subgroup of S_G .
32. The dihedral group is defined to be permutations with certain properties. Use the usual notation involving μ and ρ for elements in D_n .
- Identify which elements in D_3 are even. Do the even elements form a cyclic group?
 - Identify which of elements of D_4 are even. Do the even elements form a cyclic group?
 - For which values of n do the even permutations of D_n form a cyclic group?

Proof Synopsis

33. Give a two-sentence synopsis of the proof of Cayley's Theorem.
 34. Give a two-sentence synopsis of the proof of Theorem 8.19.

Theory

- Suppose that $\phi : G \rightarrow G'$ is a group homomorphism and $a \in \text{Ker}\phi$. Show that for any $g \in G$, $gag^{-1} \in \text{Ker}\phi$.
- Prove that a homomorphism $\phi : G \rightarrow G'$ is one-to-one if and only if $\text{Ker}(\phi)$ is the trivial subgroup of G .
- Let $\phi : G \rightarrow G'$ be a group homomorphism. Show that $\phi(a) = \phi(b)$ if and only if $a^{-1}b \in \text{Ker}\phi$.
- Use Exercise 37 to prove that if $\phi : G \rightarrow G'$ is a group homomorphism mapping onto G' and G is a finite group, then for any $b, c \in G'$, $|\phi^{-1}[\{b\}]| = |\phi^{-1}[\{c\}]|$. Conclude that if $|G|$ is a prime number, then either ϕ is an isomorphism or else G' is the trivial group.

39. Show that if $\phi : G \rightarrow G'$ and $\gamma : G' \rightarrow G''$ are group homomorphisms, then $\gamma \circ \phi : G \rightarrow G''$ is also a group homomorphism.
40. Let $\phi : G \rightarrow G'$ be a group homomorphism. Show that $\phi[G]$ is abelian if and only if $xyx^{-1}y^{-1} \in \text{Ker}(\phi)$ for all $x, y \in G$.
41. Prove the following about S_n if $n \geq 3$.
- Every permutation in S_n can be written as a product of at most $n - 1$ transpositions.
 - Every permutation in S_n that is not a cycle can be written as a product of at most $n - 2$ transpositions.
 - Every odd permutation in S_n can be written as a product of $2n + 3$ transpositions, and every even permutation as a product of $2n + 8$ transpositions.
42. a. Draw a figure like Fig. 8.20 to illustrate that if i and j are in different orbits of σ and $\sigma(i) = i$, then the number of orbits of $(i, j)\sigma$ is one less than the number of orbits of σ .
b. Repeat part (a) if $\sigma(j) = j$ also.
43. Show that for every subgroup H of S_n for $n \geq 2$, either all the permutations in H are even or exactly half of them are even.
44. Let σ be a permutation of a set A . We shall say " σ moves $a \in A$ " if $\sigma(a) \neq a$. If A is a finite set, how many elements are moved by a cycle $\sigma \in S_A$ of length n ?
45. Let A be an infinite set. Let H be the set of all $\sigma \in S_A$ such that the number of elements moved by σ (see Exercise 44) is finite. Show that H is a subgroup of S_A .
46. Let A be an infinite set. Let K be the set of all $\sigma \in S_A$ that move (see Exercise 44) at most 50 elements of A . Is K a subgroup of S_A ? Why?
47. Consider S_n for a fixed $n \geq 2$ and let σ be a fixed odd permutation. Show that every odd permutation in S_n is a product of σ and some permutation in A_n .
48. Show that if σ is a cycle of odd length, then σ^2 is a cycle.
49. Following the line of thought opened by Exercise 48, complete the following with a condition involving n and r so that the resulting statement is a theorem:

If σ is a cycle of length n , then σ^r is also a cycle if and only if ...

50. Show that S_n is generated by $\{(1, 2), (1, 2, 3, \dots, n)\}$. [Hint: Show that as r varies, $(1, 2, 3, \dots, n)^r(1, 2)(1, 2, 3, \dots, n)^{n-r}$ gives all the transpositions $(1, 2), (2, 3), (3, 4), \dots, (n-1, n), (n, 1)$. Then show that any transposition is a product of some of these transpositions and use Theorem 8.15.]

51. Let $\sigma \in S_n$ and define a relation on $\{1, 2, 3, \dots, n\}$ by $i \sim j$ if and only if $j = \sigma^k(i)$ for some $k \in \mathbb{Z}$.

- Prove that \sim is an equivalence relation.
- Prove that for any $1 \leq i \leq n$, the equivalence class of i is the orbit of i .

52. The usual definition for the determinant of an $n \times n$ matrix $A = (a_{ij})$ is

$$\det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,\sigma(3)} \cdots a_{n,\sigma(n)}$$

where $\text{sgn}(\sigma)$ is the sign of σ . Using this definition, prove the following properties of determinants.

- If a row of matrix A has all zero entries, then $\det(A) = 0$.
 - If two different rows of A are switched to obtain B , then $\det(B) = -\det(A)$.
 - If r times one row of A is added to another row of A to obtain a matrix B , then $\det(A) = \det(B)$.
 - If a row of A is multiplied by r to obtain the matrix B , then $\det(B) = r \det(A)$.
53. Prove that any finite group G is isomorphic with a subgroup of $\text{GL}(n, \mathbb{R})$ for some n . [Hint: For each $\sigma \in S_n$, find a matrix in $\text{GL}(n, \mathbb{R})$ that sends each basis vector e_i to $e_{\sigma(i)}$. Use this to show that S_n is isomorphic with a subgroup of $\text{GL}(n, \mathbb{R})$.]
54. Prove Cayley's Theorem using the right regular representation rather than the left regular representation.
55. Let $\sigma \in S_n$. An inversion is a pair (i, j) such that $i < j$ and $\sigma(i) > \sigma(j)$. Prove Theorem 8.19 by showing that multiplying a permutation by a transposition changes the number of inversions by an odd number.

56. The sixteen puzzle consists of 15 tiles numbered 1 through 15 arranged in a four-by-four grid with one position left blank. A move is sliding a tile adjacent to the blank position into the blank position. The goal is to arrange the numbers in order by a sequence of moves. Is it possible to start with the configuration pictured in Figure 8.27(a) and solve the puzzle as indicated in Figure 8.27(b)? Prove your answer by finding a sequence of moves to solve the puzzle or by proving that it is impossible to solve.

1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	

a.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

b.

8.27 Figure

SECTION 9**FINITELY GENERATED ABELIAN GROUPS****Direct Products**

Let us take a moment to review our present stockpile of groups. Starting with finite groups, we have the cyclic group \mathbb{Z}_n , the symmetric group S_n , and the alternating group A_n for each positive integer n . We also have the dihedral groups D_n and the Klein 4-group V . Of course we know that subgroups of these groups exist. Turning to infinite groups, we have groups consisting of sets of numbers under the usual addition or multiplication, as, for example, \mathbb{Z} , \mathbb{R} , and \mathbb{C} under addition, and their nonzero elements under multiplication. We have the group U of complex numbers of magnitude 1 under multiplication, which is isomorphic to each of the groups \mathbb{R}_c under addition modulo c , where $c \in \mathbb{R}^+$. We also have the group S_A of all permutations of an infinite set A , as well as various groups formed from matrices such as $GL(n, \mathbb{R})$.

One purpose of this section is to show a way to use known groups as building blocks to form more groups. The Klein 4-group will be recovered in this way from the cyclic groups. Employing this procedure with the cyclic groups gives us a large class of abelian groups that can be shown to include all possible structure types for a finite abelian group. We start by generalizing Definition 0.4.

9.1 Definition

The **Cartesian product of sets** B_1, B_2, \dots, B_n is the set of all ordered n -tuples (b_1, b_2, \dots, b_n) , where $b_i \in B_i$ for $i = 1, 2, \dots, n$. The Cartesian product is denoted by either

$$B_1 \times B_2 \times \cdots \times B_n$$

or by

$$\prod_{i=1}^n B_i.$$

■

We could also define the Cartesian product of an infinite number of sets, but the definition is considerably more sophisticated and we shall not need it.

Now let G_1, G_2, \dots, G_n be groups, and let us use multiplicative notation for all the group operations. Regarding the G_i as sets, we can form $\prod_{i=1}^n G_i$. Let us show that we can make $\prod_{i=1}^n G_i$ into a group by means of a binary operation of *multiplication by components*. Note again that we are being sloppy when we use the same notation for a group as for the set of elements of the group.

9.2 Theorem Let G_1, G_2, \dots, G_n be groups. For (a_1, a_2, \dots, a_n) and (b_1, b_2, \dots, b_n) in $\prod_{i=1}^n G_i$, define $(a_1, a_2, \dots, a_n)(b_1, b_2, \dots, b_n)$ to be the element $(a_1b_1, a_2b_2, \dots, a_nb_n)$. Then $\prod_{i=1}^n G_i$ is a group, the **direct product of the groups** G_i , under this binary operation.

Proof Note that since $a_i \in G_i$, $b_i \in G_i$, and G_i is a group, we have $a_i b_i \in G_i$. Thus the definition of the binary operation on $\prod_{i=1}^n G_i$ given in the statement of the theorem makes sense; that is, $\prod_{i=1}^n G_i$ is closed under the binary operation.

The associative law in $\prod_{i=1}^n G_i$ is thrown back onto the associative law in each component as follows:

$$\begin{aligned} & (a_1, a_2, \dots, a_n)[(b_1, b_2, \dots, b_n)(c_1, c_2, \dots, c_n)] \\ &= (a_1, a_2, \dots, a_n)(b_1c_1, b_2c_2, \dots, b_nc_n) \\ &= (a_1(b_1c_1), a_2(b_2c_2), \dots, a_n(b_nc_n)) \\ &= ((a_1b_1)c_1, (a_2b_2)c_2, \dots, (a_nb_n)c_n) \\ &= (a_1b_1, a_2b_2, \dots, a_nb_n)(c_1, c_2, \dots, c_n) \\ &= [(a_1, a_2, \dots, a_n)(b_1, b_2, \dots, b_n)](c_1, c_2, \dots, c_n). \end{aligned}$$

If e_i is the identity element in G_i , then clearly, with multiplication by components, (e_1, e_2, \dots, e_n) is an identity in $\prod_{i=1}^n G_i$. Finally, an inverse of (a_1, a_2, \dots, a_n) is $(a_1^{-1}, a_2^{-1}, \dots, a_n^{-1})$; compute the product by components. Hence $\prod_{i=1}^n G_i$ is a group. \blacklozenge

In the event that the operation of each G_i is commutative, we sometimes use additive notation in $\prod_{i=1}^n G_i$ and refer to $\prod_{i=1}^n G_i$ as the **direct sum of the groups** G_i . The notation $\bigoplus_{i=1}^n G_i$ is sometimes used in this case in place of $\prod_{i=1}^n G_i$, especially with abelian groups with operation $+$. The direct sum of abelian groups G_1, G_2, \dots, G_n may be written $G_1 \oplus G_2 \oplus \dots \oplus G_n$. We leave to Exercise 46 the proof that a direct product of abelian groups is again abelian.

It is quickly seen that if B_i has r_i elements for $i = 1, \dots, n$, then $\prod_{i=1}^n B_i$ has $r_1 r_2 \cdots r_n$ elements, for in an n -tuple, there are r_1 choices for the first component from B_1 , and for each of these there are r_2 choices for the next component from B_2 , and so on.

9.3 Example Consider the group $\mathbb{Z}_2 \times \mathbb{Z}_3$, which has $2 \cdot 3 = 6$ elements, namely $(0, 0), (0, 1), (0, 2), (1, 0), (1, 1)$, and $(1, 2)$. We claim that $\mathbb{Z}_2 \times \mathbb{Z}_3$ is cyclic. It is only necessary to find a generator. Let us try $(1, 1)$. Here the operations in \mathbb{Z}_2 and \mathbb{Z}_3 are written additively, so we do the same in the direct product $\mathbb{Z}_2 \times \mathbb{Z}_3$.

$$\begin{aligned} (1, 1) &= (1, 1) \\ 2(1, 1) &= (1, 1) + (1, 1) = (0, 2) \\ 3(1, 1) &= (1, 1) + (1, 1) + (1, 1) = (1, 0) \\ 4(1, 1) &= 3(1, 1) + (1, 1) = (1, 0) + (1, 1) = (0, 1) \\ 5(1, 1) &= 4(1, 1) + (1, 1) = (0, 1) + (1, 1) = (1, 2) \\ 6(1, 1) &= 5(1, 1) + (1, 1) = (1, 2) + (1, 1) = (0, 0) \end{aligned}$$

Thus $(1, 1)$ generates all of $\mathbb{Z}_2 \times \mathbb{Z}_3$. Since there is, up to isomorphism, only one cyclic group structure of a given order, we see that $\mathbb{Z}_2 \times \mathbb{Z}_3$ is isomorphic to \mathbb{Z}_6 . \blacktriangle

9.4 Example Consider $\mathbb{Z}_3 \times \mathbb{Z}_3$. This is a group of nine elements. We claim that $\mathbb{Z}_3 \times \mathbb{Z}_3$ is *not* cyclic. Since the addition is by components, and since in \mathbb{Z}_3 every element added to itself three times gives the identity, the same is true in $\mathbb{Z}_3 \times \mathbb{Z}_3$. Thus no element can generate the group, for a generator added to itself successively could only give the identity after nine

summands. We have found another group structure of order 9. A similar argument shows that $\mathbb{Z}_2 \times \mathbb{Z}_2$ is not cyclic. Thus $\mathbb{Z}_2 \times \mathbb{Z}_2$ must be isomorphic to the Klein 4-group. \blacktriangle

The preceding examples illustrate the following theorem:

9.5 Theorem The group $\mathbb{Z}_m \times \mathbb{Z}_n$ is cyclic and is isomorphic to \mathbb{Z}_{mn} if and only if m and n are relatively prime, that is, the gcd of m and n is 1.

Proof Consider the cyclic subgroup of $\mathbb{Z}_m \times \mathbb{Z}_n$ generated by $(1, 1)$ as described by Theorem 5.19. As our previous work has shown, the order of this cyclic subgroup is the smallest power of $(1, 1)$ that gives the identity $(0, 0)$. Here taking a power of $(1, 1)$ in our additive notation will involve adding $(1, 1)$ to itself repeatedly. Under addition by components, the first component $1 \in \mathbb{Z}_m$ yields 0 only after m summands, $2m$ summands, and so on, and the second component $1 \in \mathbb{Z}_n$ yields 0 only after n summands, $2n$ summands, and so on. For them to yield 0 simultaneously, the number of summands must be a multiple of both m and n . The smallest number that is a multiple of both m and n will be mn if and only if the gcd of m and n is 1; in this case, $(1, 1)$ generates a cyclic subgroup of order mn , which is the order of the whole group. This shows that $\mathbb{Z}_m \times \mathbb{Z}_n$ is cyclic of order mn , and hence isomorphic to \mathbb{Z}_{mn} if m and n are relatively prime.

For the converse, suppose that the gcd of m and n is $d > 1$. Then mn/d is divisible by both m and n . Consequently, for any (r, s) in $\mathbb{Z}_m \times \mathbb{Z}_n$, we have

$$\underbrace{(r, s) + (r, s) + \cdots + (r, s)}_{mn/d \text{ summands}} = (0, 0).$$

Hence no element (r, s) in $\mathbb{Z}_m \times \mathbb{Z}_n$ can generate the entire group, so $\mathbb{Z}_m \times \mathbb{Z}_n$ is not cyclic and therefore not isomorphic to \mathbb{Z}_{mn} . \blacklozenge

This theorem can be extended to a product of more than two factors by similar arguments. We state this as a corollary without going through the details of the proof.

9.6 Corollary The group $\prod_{i=1}^n \mathbb{Z}_{m_i}$ is cyclic and isomorphic to $\mathbb{Z}_{m_1 m_2 \cdots m_n}$ if and only if the numbers m_i for $i = 1, \dots, n$ are such that the gcd of any two of them is 1.

9.7 Example The preceding corollary shows that if n is written as a product of powers of distinct prime numbers, as in

$$n = (p_1)^{n_1} (p_2)^{n_2} \cdots (p_r)^{n_r},$$

then \mathbb{Z}_n is isomorphic to

$$\mathbb{Z}_{(p_1)^{n_1}} \times \mathbb{Z}_{(p_2)^{n_2}} \times \cdots \times \mathbb{Z}_{(p_r)^{n_r}}.$$

In particular, \mathbb{Z}_{72} is isomorphic to $\mathbb{Z}_8 \times \mathbb{Z}_9$. \blacktriangle

We remark that changing the order of the factors in a direct product yields a group isomorphic to the original one. The names of elements have simply been changed via a permutation of the components in the n -tuples.

Exercise 57 of Section 6 asked you to define the least common multiple of two positive integers r and s as a generator of a certain cyclic group. It is straightforward to prove that the subset of \mathbb{Z} consisting of all integers that are multiples of both r and s is a subgroup of \mathbb{Z} , and hence is a cyclic group. Likewise, the set of all common multiples of n positive integers r_1, r_2, \dots, r_n is a subgroup of \mathbb{Z} , and hence is cyclic.

9.8 Definition Let r_1, r_2, \dots, r_n be positive integers. Their **least common multiple** (abbreviated lcm) is the positive generator of the cyclic group of all common multiples of the r_i , that is, the cyclic group of all integers divisible by each r_i for $i = 1, 2, \dots, n$. \blacksquare

From Definition 9.8 and our work on cyclic groups, we see that the lcm of r_1, r_2, \dots, r_n is the smallest positive integer that is a multiple of each r_i for $i = 1, 2, \dots, n$, hence the name *least common multiple*.

9.9 Theorem Let $(a_1, a_2, \dots, a_n) \in \prod_{i=1}^n G_i$. If a_i is of finite order r_i in G_i , then the order of (a_1, a_2, \dots, a_n) in $\prod_{i=1}^n G_i$ is equal to the least common multiple of all the r_i .

Proof This follows by a repetition of the argument used in the proof of Theorem 9.5. For a power of (a_1, a_2, \dots, a_n) to give (e_1, e_2, \dots, e_n) , the power must simultaneously be a multiple of r_1 so that this power of the first component a_1 will yield e_1 , a multiple of r_2 , so that this power of the second component a_2 will yield e_2 , and so on. ◆

9.10 Example Find the order of $(8, 4, 10)$ in the group $\mathbb{Z}_{12} \times \mathbb{Z}_{60} \times \mathbb{Z}_{24}$.

Solution Since the gcd of 8 and 12 is 4, we see that 8 is of order $\frac{12}{4} = 3$ in \mathbb{Z}_{12} . (See Theorem 6.15.) Similarly, we find that 4 is of order 15 in \mathbb{Z}_{60} and 10 is of order 12 in \mathbb{Z}_{24} . The lcm of 3, 15, and 12 is $3 \cdot 5 \cdot 4 = 60$, so $(8, 4, 10)$ is of order 60 in the group $\mathbb{Z}_{12} \times \mathbb{Z}_{60} \times \mathbb{Z}_{24}$. ▲

9.11 Example The group $\mathbb{Z} \times \mathbb{Z}_2$ is generated by the elements $(1, 0)$ and $(0, 1)$. More generally, the direct product of n cyclic groups, each of which is either \mathbb{Z} or \mathbb{Z}_m for some positive integer m , is generated by the n n -tuples

$$(1, 0, 0, \dots, 0), \quad (0, 1, 0, \dots, 0), \quad (0, 0, 1, \dots, 0), \quad \dots, \quad (0, 0, 0, \dots, 1).$$

Such a direct product might also be generated by fewer elements. For example, $\mathbb{Z}_3 \times \mathbb{Z}_4 \times \mathbb{Z}_{35}$ is generated by the single element $(1, 1, 1)$. ▲

Note that if $\prod_{i=1}^n G_i$ is the direct product of groups G_i , then the subset

$$\bar{G}_i = \{(e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) \mid a_i \in G_i\},$$

that is, the set of all n -tuples with the identity elements in all places but the i th, is a subgroup of $\prod_{i=1}^n G_i$. It is also clear that this subgroup \bar{G}_i is naturally isomorphic to G_i ; just rename

$$(e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) \text{ by } a_i.$$

The group G_i is mirrored in the i th component of the elements of \bar{G}_i , and the e_j in the other components just ride along. We consider $\prod_{i=1}^n G_i$ to be the *internal direct product* of these subgroups \bar{G}_i . The direct product given by Theorem 9.2 is called the *external direct product* of the groups G_i . The terms *internal* and *external*, as applied to a direct product of groups, just reflect whether or not (respectively) we are regarding the component groups as subgroups of the product group. We shall usually omit the words *external* and *internal* and just say *direct product*. Which term we mean will be clear from the context.

The Structure of Finitely Generated Abelian Groups

Some theorems of abstract algebra are easy to understand and use, although their proofs may be quite technical and time-consuming to present. This is one section in the text where we explain the meaning and significance of a theorem but omit its proof. The

■ HISTORICAL NOTE

In his *Disquisitiones Arithmeticae*, Carl Gauss demonstrated various results in what is today the theory of abelian groups in the context of number theory. Not only did he deal extensively with equivalence classes of quadratic forms, but he also considered residue classes modulo a given integer. Although he noted that results in these two areas were similar, he did not attempt to develop an abstract theory of abelian groups.

In the 1840s, Ernst Kummer in dealing with ideal complex numbers noted that his results were in many respects analogous to those of Gauss. (See the Historical Note in Section 30.) But it was Kummer's student Leopold Kronecker (see the Historical Note in Section 39) who finally realized that an abstract theory could be developed out of

the analogies. As he wrote in 1870, "these principles [from the work of Gauss and Kummer] belong to a more general, abstract realm of ideas. It is therefore appropriate to free their development from all unimportant restrictions, so that one can spare oneself from the necessity of repeating the same argument in different cases. This advantage already appears in the development itself, and the presentation gains in simplicity, if it is given in the most general admissible manner, since the most important features stand out with clarity." Kronecker then proceeded to develop the basic principles of the theory of finite abelian groups and was able to state and prove a version of Theorem 9.12 restricted to finite groups.

meaning of any theorem whose proof we omit is well within our understanding, and we feel we should be acquainted with it. It would be impossible for us to meet some of these fascinating facts in a one-semester course if we were to insist on wading through complete proofs of all theorems. The theorem that we now state gives us complete structural information about many abelian groups, in particular, about all finite abelian groups.

9.12 Theorem (Primary Factor Version of the Fundamental Theorem of Finitely Generated Abelian Groups) Every finitely generated abelian group G is isomorphic to a direct product of cyclic groups in the form

$$\mathbb{Z}_{(p_1)^{r_1}} \times \mathbb{Z}_{(p_2)^{r_2}} \times \cdots \times \mathbb{Z}_{(p_n)^{r_n}} \times \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z},$$

where the p_i are primes, not necessarily distinct, and the r_i are positive integers. The direct product is unique except for possible rearrangement of the factors; that is, the number (**Betti number** of G) of factors \mathbb{Z} is unique and the prime powers $(p_i)^{r_i}$ are unique.

Proof The proof is omitted here. ◆

9.13 Example Find all abelian groups, up to isomorphism, of order 360. The phrase *up to isomorphism* signifies that any abelian group of order 360 should be structurally identical (isomorphic) to one of the groups of order 360 exhibited.

Solution We make use of Theorem 9.12. Since our groups are to be of the finite order 360, no factors \mathbb{Z} will appear in the direct product shown in the statement of the theorem.

First we express 360 as a product of prime powers $2^3 3^2 5$. Then using Theorem 9.12, we get as possibilities

1. $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5$
2. $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_9 \times \mathbb{Z}_5$
3. $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5$

4. $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_9 \times \mathbb{Z}_5$

5. $\mathbb{Z}_8 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5$

6. $\mathbb{Z}_8 \times \mathbb{Z}_9 \times \mathbb{Z}_5$

Thus there are six different abelian groups (up to isomorphism) of order 360. \blacktriangle

There is another version of the Fundamental Theorem of Finitely Generated Abelian Groups. Each version can be proven from the other, so technically, if one version is used to prove something, the other version could also be used. However, it is sometimes more convenient to use one version rather than the other for a particular problem.

9.14 Theorem (Invariant Factor Version of the Fundamental Theorem of Finitely Generated Abelian Groups) Every finitely generated abelian group is isomorphic to a direct product of cyclic groups of the form

$$\mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \mathbb{Z}_{d_3} \times \cdots \times \mathbb{Z}_{d_k} \times \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z},$$

where each of the $d_i \geq 2$ is an integer and d_i divides d_{i+1} for $1 \leq i \leq k-1$. Furthermore, the representation is unique. \blacklozenge

The Betti number of a group is the number of factors of \mathbb{Z} in both Theorem 9.12 and 9.14. The numbers d_i are called the **invariant factors** or the **torsion coefficients**. Theorem 9.12 implies Theorem 9.14 and the other way around. Here we show with an example how to start with a finite group that is in the form specified in Theorem 9.12 and find its representation in the form of Theorem 9.14.

9.15 Example Let us find the invariant factor form of the abelian group $G = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_8 \times \mathbb{Z}_3 \times \mathbb{Z}_9 \times \mathbb{Z}_7$, which is in primary factor form. We make a table, one row for each prime number involved in G : 2, 3, and 7. We list the powers of each prime in the primary factor form starting with the highest power to the lowest power, filling the ends of the short rows with $1 = p^0$. Table 9.16 is the table for G . The group G is the direct product of cyclic groups of the orders listed in the table. The products of the entries in the columns give the invariant factors. For G , the invariant factors are $d_4 = 8 \cdot 9 \cdot 7 = 504$, $d_3 = 4 \cdot 3 \cdot 1 = 12$, $d_2 = 2 \cdot 1 \cdot 1 = 2$, and $d_1 = 2 \cdot 1 \cdot 1 = 2$. The construction of the table insures that d_1 divides d_2 , d_2 divides d_3 , d_3 divides d_4 , and G is isomorphic with $\mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \mathbb{Z}_{d_3} \times \mathbb{Z}_{d_4} = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{12} \times \mathbb{Z}_{504}$. \blacktriangle

9.16 Table

8	4	2	2
9	3	1	1
7	1	1	1

Example 9.15 shows how to create a table from a finitely generated abelian group that is in primary factor form. From the table we can find the invariant form of the group. This process can easily be reversed by factoring the invariants to find the primary factors.

Applications

Because of Theorems 9.12 and 9.14, there is a plethora of theorems regarding finitely generated abelian groups that are fairly easily proven. We present a few examples.

9.17 Definition A group G is **decomposable** if it is isomorphic to a direct product of two proper non-trivial subgroups. Otherwise G is **indecomposable**. \blacksquare

9.18 Theorem The finite indecomposable abelian groups are exactly the cyclic groups with order a power of a prime.

Proof Let G be a finite indecomposable abelian group. Then by Theorem 9.12, G is isomorphic to a direct product of cyclic groups of prime power order. Since G is indecomposable, this direct product must consist of just one cyclic group whose order is a power of a prime number.

Conversely, let p be a prime. Then \mathbb{Z}_{p^r} is indecomposable, for if \mathbb{Z}_{p^r} were isomorphic to $\mathbb{Z}_{p^i} \times \mathbb{Z}_{p^j}$, where $i + j = r$, then every element would have an order at most $p^{\max(i,j)} < p^r$. \blacklozenge

9.19 Theorem If m divides the order of a finite abelian group G , then G has a subgroup of order m .

Proof By Theorem 9.12, we can think of G as being

$$\mathbb{Z}_{(p_1)^{r_1}} \times \mathbb{Z}_{(p_2)^{r_2}} \times \cdots \times \mathbb{Z}_{(p_n)^{r_n}},$$

where not all primes p_i need be distinct. Since $(p_1)^{r_1}(p_2)^{r_2} \cdots (p_n)^{r_n}$ is the order of G , then m must be of the form $(p_1)^{s_1}(p_2)^{s_2} \cdots (p_n)^{s_n}$, where $0 \leq s_i \leq r_i$. By Theorem 6.15, $(p_i)^{r_i-s_i}$ generates a cyclic subgroup of $\mathbb{Z}_{(p_i)^{r_i}}$ of order equal to the quotient of $(p_i)^{r_i}$ by the gcd of $(p_i)^{r_i}$ and $(p_i)^{r_i-s_i}$. But the gcd of $(p_i)^{r_i}$ and $(p_i)^{r_i-s_i}$ is $(p_i)^{r_i-s_i}$. Thus $(p_i)^{r_i-s_i}$ generates a cyclic subgroup of $\mathbb{Z}_{(p_i)^{r_i}}$ of order

$$[(p_i)^{r_i}] / [(p_i)^{r_i-s_i}] = (p_i)^{s_i}.$$

Recalling that $\langle a \rangle$ denotes the cyclic subgroup generated by a , we see that

$$\langle (p_1)^{r_1-s_1} \rangle \times \langle (p_2)^{r_2-s_2} \rangle \times \cdots \times \langle (p_n)^{r_n-s_n} \rangle$$

is the required subgroup of order m . \blacklozenge

9.20 Theorem If m is a square-free integer, that is, m is not divisible by the square of any integer $n \geq 2$ then every abelian group of order m is cyclic.

Proof Let G be a finite abelian group of square-free order m . Then by Theorem 9.14, G is isomorphic to

$$\mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \cdots \times \mathbb{Z}_{d_k},$$

where each $d_i \geq 2$ divides d_{i+1} for $1 \leq i \leq k - 1$. The order of G is $m = d_1 \cdot d_2 \cdots d_k$. If $k \geq 2$, then d_1^2 divides m , which is a contradiction. Thus $k = 1$ and G is cyclic. \blacklozenge

■ EXERCISES 9

Computations

1. List the elements of $\mathbb{Z}_2 \times \mathbb{Z}_4$. Find the order of each of the elements. Is this group cyclic?
2. Repeat Exercise 1 for the group $\mathbb{Z}_3 \times \mathbb{Z}_4$.
- In Exercises 3 through 7, find the order of the given element of the direct product.

3. $(2, 6)$ in $\mathbb{Z}_4 \times \mathbb{Z}_{12}$	4. $(3, 4)$ in $\mathbb{Z}_{21} \times \mathbb{Z}_{12}$	5. $(40, 12)$ in $\mathbb{Z}_{45} \times \mathbb{Z}_{18}$
6. $(3, 10, 9)$ in $\mathbb{Z}_4 \times \mathbb{Z}_{12} \times \mathbb{Z}_{15}$	7. $(3, 6, 12, 16)$ in $\mathbb{Z}_4 \times \mathbb{Z}_{12} \times \mathbb{Z}_{20} \times \mathbb{Z}_{24}$	
8. What is the largest order among the orders of all the cyclic subgroups of $\mathbb{Z}_6 \times \mathbb{Z}_8$? of $\mathbb{Z}_{12} \times \mathbb{Z}_{15}$?
9. Find all proper nontrivial subgroups of $\mathbb{Z}_2 \times \mathbb{Z}_2$.
10. Find all proper nontrivial subgroups of $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.
11. Find all subgroups of $\mathbb{Z}_2 \times \mathbb{Z}_4$ of order 4.
12. Find all subgroups of $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4$ that are isomorphic to the Klein 4-group.
13. Disregarding the order of the factors, write direct products of two or more groups of the form \mathbb{Z}_n so that the resulting product is isomorphic to \mathbb{Z}_{60} in as many ways as possible.
14. Fill in the blanks.
 - a. The cyclic subgroup of \mathbb{Z}_{24} generated by 18 has order ____.
 - b. $\mathbb{Z}_3 \times \mathbb{Z}_4$ is of order ____.

- c. The element $(4, 2)$ of $\mathbb{Z}_{12} \times \mathbb{Z}_8$ has order ____.
 - d. The Klein 4-group is isomorphic to $\mathbb{Z}_{\text{ }} \times \mathbb{Z}_{\text{ }}$.
 - e. $\mathbb{Z}_2 \times \mathbb{Z} \times \mathbb{Z}_4$ has ____ elements of finite order.
15. Find the maximum possible order for some element of $\mathbb{Z}_4 \times \mathbb{Z}_6$.
16. Are the groups $\mathbb{Z}_2 \times \mathbb{Z}_{12}$ and $\mathbb{Z}_4 \times \mathbb{Z}_6$ isomorphic? Why or why not?
17. Find the maximum possible order for some element of $\mathbb{Z}_8 \times \mathbb{Z}_{28} \times \mathbb{Z}_{24}$.
18. Are the groups $\mathbb{Z}_8 \times \mathbb{Z}_{10} \times \mathbb{Z}_{24}$ and $\mathbb{Z}_4 \times \mathbb{Z}_{12} \times \mathbb{Z}_{40}$ isomorphic? Why or why not?
19. Find the maximum possible order for some element of $\mathbb{Z}_4 \times \mathbb{Z}_{18} \times \mathbb{Z}_{15}$.
20. Are the groups $\mathbb{Z}_4 \times \mathbb{Z}_{18} \times \mathbb{Z}_{15}$ and $\mathbb{Z}_3 \times \mathbb{Z}_{36} \times \mathbb{Z}_{10}$ isomorphic? Why or why not?

In Exercises 21 through 25, proceed as in Example 9.13 to find all abelian groups, up to isomorphism, of the given order. For each group, find the invariant factors and find an isomorphic group of the form indicated in Theorem 9.14.

- | | | |
|---------------|----------------|--------------|
| 21. Order 8 | 22. Order 16 | 23. Order 32 |
| 24. Order 720 | 25. Order 1089 | |
26. How many abelian groups (up to isomorphism) are there of order 24? of order 25? of order (24)(25)?
27. Following the idea suggested in Exercise 26, let m and n be relatively prime positive integers. Show that if there are (up to isomorphism) r abelian groups of order m and s of order n , then there are (up to isomorphism) rs abelian groups of order mn .
28. Use Exercise 27 to determine the number of abelian groups (up to isomorphism) of order $(10)^5$.
29. a. Let p be a prime number. Fill in the second row of the table to give the number of abelian groups of order p^n , up to isomorphism.

n	2	3	4	5	6	7	8
number of groups							

- b. Let p, q , and r be distinct prime numbers. Use the table you created to find the number of abelian groups, up to isomorphism, of the given order.
- i. $p^3q^4r^7$ ii. $(qr)^7$ iii. $q^5r^4q^3$
30. Indicate schematically a Cayley digraph for $\mathbb{Z}_m \times \mathbb{Z}_n$ for the generating set $S = \{(1, 0), (0, 1)\}$.
31. Consider Cayley digraphs with two arc types, a solid one with an arrow and a dashed one with no arrow, and consisting of two regular n -gons, for $n \geq 3$, with solid arc sides, one inside the other, with dashed arcs joining the vertices of the outer n -gon to the inner one. Figure 7.11(b) shows such a Cayley digraph with $n = 3$, and Figure 7.13(b) shows one with $n = 4$. The arrows on the outer n -gon may have the same (clockwise or counterclockwise) direction as those on the inner n -gon, or they may have the opposite direction. Let G be a group with such a Cayley digraph.
- a. Under what circumstances will G be abelian?
 - b. If G is abelian, to what familiar group is it isomorphic?
 - c. If G is abelian, under what circumstances is it cyclic?
 - d. If G is not abelian, to what group we have discussed is it isomorphic?

Concepts

32. Determine whether each of the following is true or false.
- a. If G_1 and G_2 are any groups, then $G_1 \times G_2$ is always isomorphic to $G_2 \times G_1$.
 - b. Computation in an external direct product of groups is easy if you know how to compute in each component group.
 - c. Groups of finite order must be used to form an external direct product.
 - d. A group of prime order could not be the internal direct product of two proper nontrivial subgroups.

- e. $\mathbb{Z}_2 \times \mathbb{Z}_4$ is isomorphic to \mathbb{Z}_8 .
 f. $\mathbb{Z}_2 \times \mathbb{Z}_4$ is isomorphic to S_8 .
 g. $\mathbb{Z}_3 \times \mathbb{Z}_8$ is isomorphic to S_4 .
 h. Every element in $\mathbb{Z}_4 \times \mathbb{Z}_8$ has order 8.
 i. The order of $\mathbb{Z}_{12} \times \mathbb{Z}_{15}$ is 60.
 j. $\mathbb{Z}_m \times \mathbb{Z}_n$ has mn elements whether m and n are relatively prime or not.
33. Give an example illustrating that not every nontrivial abelian group is the internal direct product of two proper nontrivial subgroups.
34. a. How many subgroups of $\mathbb{Z}_5 \times \mathbb{Z}_6$ are isomorphic to $\mathbb{Z}_5 \times \mathbb{Z}_6$?
 b. How many subgroups of $\mathbb{Z} \times \mathbb{Z}$ are isomorphic to $\mathbb{Z} \times \mathbb{Z}$?
35. Give an example of a nontrivial group that is not of prime order and is not the internal direct product of two nontrivial subgroups.
36. Determine whether each of the following is true or false.
- a. Every abelian group of prime order is cyclic.
 - b. Every abelian group of prime power order is cyclic.
 - c. \mathbb{Z}_8 is generated by $\{4, 6\}$.
 - d. \mathbb{Z}_8 is generated by $\{4, 5, 6\}$.
 - e. All finite abelian groups are classified up to isomorphism by Theorem 9.12.
 - f. Any two finitely generated abelian groups with the same Betti number are isomorphic.
 - g. Every abelian group of order divisible by 5 contains a cyclic subgroup of order 5.
 - h. Every abelian group of order divisible by 4 contains a cyclic subgroup of order 4.
 - i. Every abelian group of order divisible by 6 contains a cyclic subgroup of order 6.
 - j. Every finite abelian group has a Betti number of 0.
37. Let p and q be distinct prime numbers. How does the number (up to isomorphism) of abelian groups of order p^r compare with the number (up to isomorphism) of abelian groups of order q^r ?
38. Let G be an abelian group of order 72.
- a. Can you say how many subgroups of order 8 G has? Why, or why not?
 - b. Can you say how many subgroups of order 4 G has? Why, or why not?
39. Let G be an abelian group. Show that the elements of finite order in G form a subgroup. This subgroup is called the **torsion subgroup** of G .

Exercises 40 through 43 deal with the concept of the torsion subgroup just defined.

40. Find the order of the torsion subgroup of $\mathbb{Z}_4 \times \mathbb{Z} \times \mathbb{Z}_3$; of $\mathbb{Z}_{12} \times \mathbb{Z} \times \mathbb{Z}_{12}$.
41. Find the torsion subgroup of the multiplicative group \mathbb{R}^* of nonzero real numbers.
42. Find the torsion subgroup T of the multiplicative group \mathbb{C}^* of nonzero complex numbers.
43. An abelian group is **torsion free** if e is the only element of finite order. Use Theorem 9.12 to show that every finitely generated abelian group is the internal direct product of its torsion subgroup and of a torsion-free subgroup. (Note that $\{e\}$ may be the torsion subgroup, and is also torsion free.)
44. Find the torsion coefficients for each of the following groups.
- | | |
|---|---|
| a. $\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4$
b. $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_8 \times \mathbb{Z}_3 \times \mathbb{Z}_{27}$ | c. $\mathbb{Z}_8 \times \mathbb{Z}_2 \times \mathbb{Z}_{49} \times \mathbb{Z}_7$
d. $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_9 \times \mathbb{Z}_5$ |
|---|---|

Proof Synopsis

45. Give a two-sentence synopsis of the proof of Theorem 9.5.

Theory

46. Prove that a direct product of abelian groups is abelian.
47. Let G be an abelian group. Let H be the subset of G consisting of the identity e together with all elements of G of order 2. Show that H is a subgroup of G .
48. Following up the idea of Exercise 47 determine whether H will always be a subgroup for every abelian group G if H consists of the identity e together with all elements of G of order 3; of order 4. For what positive integers n will H always be a subgroup for every abelian group G , if H consists of the identity e together with all elements of G of order n ? Compare with Exercise 54 of Section 5.
49. Find a counterexample of Exercise 47 with the hypothesis that G is abelian omitted.

Let H and K be subgroups of a group G . Exercises 50 and 51 ask you to establish necessary and sufficient criteria for G to appear as the internal direct product of H and K .

50. Let H and K be groups and let $G = H \times K$. Recall that both H and K appear as subgroups of G in a natural way. Show that these subgroups H (actually $H \times \{e\}$) and K (actually $\{e\} \times K$) have the following properties.
- a. Every element of G is of the form hk for some $h \in H$ and $k \in K$.
 - b. $hk = kh$ for all $h \in H$ and $k \in K$.
 - c. $H \cap K = \{e\}$.
51. Let H and K be subgroups of a group G satisfying the three properties listed in the preceding exercise. Show that for each $g \in G$, the expression $g = hk$ for $h \in H$ and $k \in K$ is unique. Then let each g be renamed (h, k) . Show that, under this renaming, G becomes structurally identical (isomorphic) to $H \times K$.
52. Show that a finite abelian group is not cyclic if and only if it contains a subgroup isomorphic to $\mathbb{Z}_p \times \mathbb{Z}_p$ for some prime p .
53. Prove that if a finite abelian group has order a power of a prime p , then the order of every element in the group is a power of p .
54. Let G, H , and K be finitely generated abelian groups. Show that if $G \times K$ is isomorphic to $H \times K$, then $G \simeq H$.
55. Using the notation of Theorem 9.14, prove that for any finite abelian group G , every cyclic subgroup of G has order no more than d_k , the largest invariant factor for G .

SECTION 10**COSETS AND THE THEOREM OF LAGRANGE**

You may have noticed that the order of a subgroup H of a finite group G seems always to be a divisor of the order of G . This is the theorem of Lagrange. We shall prove it by exhibiting a partition of G into cells, all having the same size as H . Thus if there are r such cells, we will have

$$r(\text{order of } H) = (\text{order of } G)$$

from which the theorem follows immediately. The cells in the partition will be called *cosets of H* , and they are important in their own right. In Section 12, we will see that if H satisfies a certain property, then each coset can be regarded as an element of a group in a very natural way. We give some indication of such *coset groups* in this section to help you develop a feel for the topic.

Cosets

Let H be a subgroup of a group G , which may be of finite or infinite order. We exhibit a partition of G by defining an equivalence relation, \sim_L on G .

10.1 Theorem Let H be a subgroup of G . Let the relation \sim_L be defined on G by

$$a \sim_L b \quad \text{if and only if} \quad a^{-1}b \in H.$$

Then \sim_L is an equivalence relation on G .

Proof When reading the proof, notice how we must constantly make use of the fact that H is a subgroup of G .

Reflexive Let $a \in G$. Then $a^{-1}a = e$ and $e \in H$ since H is a subgroup. Thus $a \sim_L a$.

Symmetric Suppose $a \sim_L b$. Then $a^{-1}b \in H$. Since H is a subgroup, $(a^{-1}b)^{-1}$ is in H and $(a^{-1}b)^{-1} = b^{-1}a$, so $b^{-1}a$ is in H and $b \sim_L a$.

Transitive Let $a \sim_L b$ and $b \sim_L c$. Then $a^{-1}b \in H$ and $b^{-1}c \in H$. Since H is a subgroup, $(a^{-1}b)(b^{-1}c) = a^{-1}c$ is in H , so $a \sim_L c$. \blacklozenge

The equivalence relation \sim_L in Theorem 10.1 defines a partition of G , as described in Theorem 0.22. Let's see what the cells in this partition look like. Suppose $a \in G$. The cell containing a consists of all $x \in G$ such that $a \sim_L x$, which means all $x \in G$ such that $a^{-1}x \in H$. Now $a^{-1}x \in H$ if and only if $a^{-1}x = h$ for some $h \in H$, or equivalently, if and only if $x = ah$ for some $h \in H$. Therefore the cell containing a is $\{ah \mid h \in H\}$, which we denote by aH .

10.2 Definition Let H be a subgroup of a group G . The subset $aH = \{ah \mid h \in H\}$ of G is the **left coset** of H containing a . \blacksquare

10.3 Example Exhibit the left coset of the subgroup $3\mathbb{Z}$ of \mathbb{Z} .

Solution Our notation here is additive, so the left coset of $3\mathbb{Z}$ containing m is $m + 3\mathbb{Z}$. Taking $m = 0$, we see that

$$3\mathbb{Z} = \{\dots, -9, -6, -3, 0, 3, 6, 9, \dots\}$$

is itself one of its left cosets, the coset containing 0. To find another left coset, we select an element of \mathbb{Z} not in $3\mathbb{Z}$, say 1, and find the left coset containing it. We have

$$1 + 3\mathbb{Z} = \{\dots, -8, -5, -2, 1, 4, 7, 10, \dots\}.$$

These two left cosets, $3\mathbb{Z}$ and $1 + 3\mathbb{Z}$, do not yet exhaust \mathbb{Z} . For example, 2 is in neither of them. The left coset containing 2 is

$$2 + 3\mathbb{Z} = \{\dots, -7, -4, -1, 2, 5, 8, 11, \dots\}.$$

It is clear that these three left cosets we have found do exhaust \mathbb{Z} , so they constitute the partition of \mathbb{Z} into left cosets of $3\mathbb{Z}$. \blacktriangle

10.4 Example We find the partition of \mathbb{Z}_{12} into left cosets of $H = \langle 3 \rangle$. One coset is always the subgroup itself, so $0 + H = \{0, 3, 6, 9\}$. We next find $1 + H = \{1, 4, 7, 10\}$. We are still not done since we have not included every element of \mathbb{Z}_{12} in the two cosets we listed so far. Finally, $2 + H = \{2, 5, 8, 11\}$ and we have computed all the left cosets of H in \mathbb{Z}_{12} . \blacktriangle

10.5 Example We now list the left cosets of the subgroup $H = \langle \mu \rangle = \{\iota, \mu\}$ of the nonabelian group $D_4 = \{\iota, \rho, \rho^2, \rho^3, \mu, \mu\rho, \mu\rho^2, \mu\rho^3\}$.

$$\begin{aligned}\iota\{\iota, \mu\} &= \{\iota, \mu\} \\ \rho\{\iota, \mu\} &= \{\rho, \mu\rho^3\} \\ \rho^2\{\iota, \mu\} &= \{\rho^2, \mu\rho^2\} \\ \rho^3\{\iota, \mu\} &= \{\rho^3, \mu\rho\}\end{aligned}$$

We know this is a complete list of the left cosets since every element of D_4 appears in exactly one of the listed sets. \blacktriangle

The Theorem of Lagrange

In Example 10.4 each left coset of $\langle 3 \rangle \leq \mathbb{Z}_{12}$ has four elements. In Example 10.5, each left coset has two elements. From the computation of the left cosets, it is no surprise that all left cosets of a subgroup have the same number of elements. Theorem 10.6 confirms this is what happens in general.

10.6 Theorem Let H be a subgroup of G . Then for any $a \in G$, the coset aH has the same cardinality as H .

Proof Let $f : H \rightarrow aH$ be defined by the formula $f(h) = ah$. To show f is one-to-one, we suppose that $b, c \in H$ and $f(b) = f(c)$. Then $ab = ac$ and left cancellation gives $b = c$. So f is one-to-one. Now suppose that $y \in aH$. Then there is an $h \in H$ such that $y = ah$ by definition of the left coset aH . Thus $y = f(h)$ and f is surjective. Since there is a one-to-one function mapping H onto aH , H and aH have the same cardinality. ◆

In the case of a finite subgroup H , Theorem 10.6 says that H and aH have the same number of elements for any a in the group G . This is precisely what we were seeking in order to prove Lagrange's Theorem.

10.7 Theorem (Theorem of Lagrange) Let H be a subgroup of a finite group G . Then the order of H is a divisor of the order of G .

Proof Let n be the order of G , and let H have order m . Theorem 10.6 shows that every coset of H also has m elements. Let r be the number of cells in the partition of G into left cosets of H . Then $n = rm$, so m is indeed a divisor of n . ◆

Note that this elegant and important theorem comes from the simple counting of cosets and the number of elements in each coset. We continue to derive consequences of Theorem 10.7, which should be regarded as a counting theorem.

10.8 Corollary Every group of prime order is cyclic.

Proof Let G be of prime order p , and let a be an element of G different from the identity. Then the cyclic subgroup $\langle a \rangle$ of G generated by a has at least two elements, a and e . But by Theorem 10.7, the order $m \geq 2$ of $\langle a \rangle$ must divide the prime p . Thus we must have $m = p$ and $\langle a \rangle = G$, so G is cyclic. ◆

Since every cyclic group of order p is isomorphic to \mathbb{Z}_p , we see that *there is only one group structure, up to isomorphism, of a given prime order p* . Now doesn't this elegant result follow easily from the theorem of Lagrange, a *counting theorem*? *Never underestimate a theorem that counts something*. Proving the preceding corollary is a favorite examination question.

10.9 Theorem The order of an element of a finite group divides the order of the group.

Proof Remembering that the order of an element is the same as the order of the cyclic subgroup generated by the element, we see that this theorem follows directly from Lagrange's Theorem. ◆

10.10 Definition Let H be a subgroup of a group G . The number of left cosets of H in G is the **index** $(G : H)$ of H in G . ■

The index $(G : H)$ just defined may be finite or infinite. If G is finite, then obviously $(G : H)$ is finite and $(G : H) = |G|/|H|$, since every coset of H contains $|H|$ elements. We state a basic theorem concerning indices of subgroups, and leave the proof to the exercises (see Exercise 40).

10.11 Theorem Suppose H and K are subgroups of a group G such that $K \leq H \leq G$, and suppose $(H : K)$ and $(G : H)$ are both finite. Then $(G : K) = (G : H)(H : K)$.

Lagrange's Theorem says that for any subgroup H of a finite group G , the order of H divides the order of G . But if d is a divisor of the order of G , does G necessarily have a subgroup with exactly d elements? We will show in Section 13 that the answer is no for some groups. This suggests a new question: Under what conditions does G have a subgroup of every order d that is a divisor of G ? We saw in Section 9 that for every divisor of the order of an abelian group, there is a subgroup of that order. The complete answer to this question is beyond the scope of this book, but we will come back to the question later.

Cosets Left and Right!

It is possible to do everything we have done in this section using right cosets instead of left cosets. All it takes is some minor and straightforward modifications to the definitions and proofs. We briefly give the corresponding definitions that lead to right cosets and point out some of their properties.

Let H be a subgroup of G . To start with, instead of \sim_L we could have used \sim_R defined by

$$a \sim_R b \quad \text{if and only if} \quad ab^{-1} \in H.$$

With this definition, \sim_R is an equivalence relation and the equivalence classes are the **right cosets**. The right coset of H containing the element $a \in G$ is

$$Ha = \{ha \mid h \in H\}.$$

Just like left cosets, each right coset of a subgroup H has the same cardinality as H . So left cosets and right cosets have the same cardinality. In abelian groups, the right and left cosets are the same, but there is no reason to think they would be the same in general for nonabelian groups. If the right and left cosets are the same, we can drop left or right and just refer to cosets.

10.12 Example In Example 10.5 we computed the left cosets of the subgroup $H = \langle \mu \rangle = \{\iota, \mu\}$ of the group $D_4 = \{\iota, \rho, \rho^2, \rho^3, \mu, \mu\rho, \mu\rho^2, \mu\rho^3\}$. We now compute the right cosets.

$$\begin{aligned}\{\iota, \mu\}\iota &= \{\iota, \mu\} \\ \{\iota, \mu\}\rho &= \{\rho, \mu\rho\} \\ \{\iota, \mu\}\rho^2 &= \{\rho^2, \mu\rho^2\} \\ \{\iota, \mu\}\rho^3 &= \{\rho^3, \mu\rho^3\}\end{aligned}$$

The right cosets and the left cosets are not the same. For example, $\rho H = \{\rho, \mu\rho^3\}$ while $H\rho = \{\rho, \mu\rho\}$. ▲

If this were the whole story of left and right cosets, there would be no reason to even mention right cosets. We could just use left coset, prove Lagrange's Theorem, and call it a day. However, as we shall see in Part III, a curious thing happens when the left and right cosets are the same. We illustrate with an example.

10.13 Example The group \mathbb{Z}_6 is abelian. Find the partition of \mathbb{Z}_6 into cosets of the subgroup $H = \{0, 3\}$.

Solution One coset is $\{0, 3\}$ itself. The coset containing 1 is $1 + \{0, 3\} = \{1, 4\}$. The coset containing 2 is $2 + \{0, 3\} = \{2, 5\}$. Since $\{0, 3\}$, $\{1, 4\}$, and $\{2, 5\}$ exhaust all of \mathbb{Z}_6 , these are all the cosets. ▲

We point out a fascinating thing that we will develop in detail in Section 12. Referring back to Example 10.13, Table 10.14 gives the binary operation for \mathbb{Z}_6 but with elements listed in the order they appear in the cosets $\{0, 3\}, \{1, 4\}, \{2, 5\}$. We shaded the table according to these cosets.

10.14 Table

$+_6$	0	3	1	4	2	5
0	0	3	1	4	2	5
3	3	0	4	1	5	2
1	1	4	2	5	3	0
4	4	1	5	2	0	3
2	2	5	3	0	4	1
5	5	2	0	3	1	4

10.15 Table

	LT	MD	DK
LT	LT	MD	DK
MD	MD	DK	LT
DK	DK	LT	MD

Suppose we denote these cosets by LT(light), MD(medium), and DK(dark) according to their shading. Table 10.14 then defines a binary operation on these shadings, as shown in Table 10.15. Note that if we replace LT by 0, MD by 1, and DK by 2 in Table 10.15, we obtain the table for \mathbb{Z}_3 . Thus the table of shadings forms a group!

We will see in Section 12 that when left cosets and right cosets are the same, then the cosets form a group as in Example 10.13. If right and left cosets are different, the construction fails.

10.16 Example Let $H = \{\iota, \mu\} \leq D_3$. The group table for D_3 is given below with the elements arranged so that left cosets are together. The double lines divide the cosets.

	ι	μ	ρ	$\mu\rho^2$	ρ^2	$\mu\rho$
ι	ι	μ	ρ	$\mu\rho^2$	ρ^2	$\mu\rho$
μ	μ	ι	$\mu\rho$	ρ^2	$\mu\rho^2$	ρ
ρ	ρ	$\mu\rho^2$	ρ^2	$\mu\rho$	ι	μ
$\mu\rho^2$	$\mu\rho^2$	ρ	μ	ι	$\mu\rho$	ρ^2
ρ^2	ρ^2	$\mu\rho$	ι	μ	ρ	$\mu\rho^2$
$\mu\rho$	$\mu\rho$	ρ^2	$\mu\rho^2$	ρ	μ	ι

The situation here is much different from the situation in Example 10.13. In Table 10.14 the two-by-two blocks in the table each contain only elements of a left coset. In the present example, most blocks do not contain elements from only one left coset. Furthermore, even if we tried to use the two-by-two blocks of elements to form a three-by-three group table, the second row of blocks contains two blocks, both having the same elements, $\{\rho^2, \mu\rho, \mu, \iota\}$. So the table of blocks would have a row with the same element listed twice. In this case, there is no natural way of making the left cosets a group. \blacktriangle

If G is an abelian group, then the left and right cosets are the same. Theorem 10.17 gives another condition when left and right cosets are the same. Recall that if $\phi : G \rightarrow G'$ is a group homomorphism, then $\text{Ker}(\phi) = \phi^{-1}[\{e\}] \leq G$ is the kernel of ϕ .

10.17 Theorem Let $\phi : G \rightarrow G'$ be a group homomorphism. Then the left and right cosets of $\text{Ker}(\phi)$ are identical. Furthermore, $a, b \in G$ are in the same coset of $\text{Ker}(\phi)$ if and only if $\phi(a) = \phi(b)$.

Proof We first assume that a and b are in the same left cosets of $\text{Ker}(\phi)$ and show they are also in the same right cosets. Then $a^{-1}b \in \text{Ker}(\phi)$. So $\phi(a^{-1}b) = e$, the identity element. Because ϕ is a homomorphism, $\phi(a)^{-1}\phi(b) = e$, which implies that $\phi(a) = \phi(b)$. Therefore, $\phi(ab^{-1}) = \phi(a)\phi(b)^{-1} = \phi(a)\phi(a)^{-1} = e$. Thus $ab^{-1} \in \text{Ker}(\phi)$, which says that a and b are in the same right coset. Note that in the process we showed that if a and b are in the same left coset of $\text{Ker}(\phi)$, then $\phi(a) = \phi(b)$.

Now suppose that $\phi(a) = \phi(b)$. Then $\phi(b^{-1}a) = \phi(b)^{-1}\phi(a) = e$. Thus $b^{-1}a \in \text{Ker}(\phi)$, which implies that a and b are in the same left coset.

To complete the proof, we need to show that if a and b are in the same right coset, then they are also in the same left coset. The proof is essentially the same as above, so we leave this detail to the reader. \blacklozenge

10.18 Example Consider the determinant map $\det : \text{GL}(2, \mathbb{R}) \rightarrow \mathbb{R}^*$. In linear algebra you learn that $\det(AB) = \det(A)\det(B)$, so the determinant is a group homomorphism. The kernel of \det is the set of all 2×2 matrices with determinant 1. Two matrices $A, B \in \text{GL}(2, \mathbb{R})$ are in the same left coset of $\text{Ker}(\det)$ if and only if they are in the same right coset of $\text{Ker}(\det)$ if and only if $\det(A) = \det(B)$. In particular, the two matrices

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}$$

each have determinant 2, so they are in the same left (and right) cosets of $\text{Ker}(\det)$. \blacktriangle

10.19 Corollary A homomorphism $\phi : G \rightarrow G'$ is one-to-one if and only if $\text{Ker}(\phi)$ is the trivial subgroup of G .

Proof We first assume that $\text{Ker}(\phi) = \{e\}$. Every coset of $\text{Ker}(\phi)$ has only one element. Suppose that $\phi(a) = \phi(b)$. Then a and b are in the same coset of $\text{Ker}(\phi)$ by Theorem 10.17. Thus $a = b$.

Now suppose that ϕ is one-to-one. Then only the identity e is mapped to the identity in G' . So $\text{Ker}(\phi) = \{e\}$. \blacklozenge

Corollary 10.19 says that to check if a homomorphism $\phi : G \rightarrow G'$ is one-to-one one merely needs to check that $\text{Ker}(\phi)$ is the trivial subgroup. In other words, show that the only solution to $\phi(x) = e'$ is e , where e and e' are the identities in G and G' , respectively.

10.20 Example Let $\phi : \mathbb{R} \rightarrow \mathbb{R}^+$ be defined by $\phi(x) = 2^x$. Since ϕ is a homomorphism, we can check that ϕ is one-to-one by solving $\phi(x) = 1$. The equation $2^x = \phi(x) = 1$ has only the solution 0 since for $x > 0$, $2^x > 1$ and for $x < 0$, $2^x < 1$. Thus ϕ is one-to-one. \blacktriangle

■ EXERCISES 10

Computations

1. Find all cosets of the subgroup $4\mathbb{Z}$ of \mathbb{Z} .
2. Find all cosets of the subgroup $4\mathbb{Z}$ of $2\mathbb{Z}$.
3. Find all cosets of the subgroup $\langle 3 \rangle$ in \mathbb{Z}_{18} .
4. Find all cosets of the subgroup $\langle 6 \rangle$ in \mathbb{Z}_{18} .
5. Find all cosets of the subgroup $\langle 18 \rangle$ of \mathbb{Z}_{36} .
6. Find all left cosets of $\langle \mu\rho \rangle$ in D_4 .
7. Repeat the preceding exercise, but find the right cosets this time. Are they the same as the left cosets?

8. Are the left and right cosets the same for the subgroup $\{\iota, \rho^4, \mu, \mu\rho^4\}$ of D_8 ? If so, display the cosets. If not, find a left coset that is not the same as any right coset.
9. Find all the left cosets of $\langle \rho^2 \rangle \leq D_4$.
10. Repeat the previous exercise, but find the right cosets. Are the left and right cosets the same? If so, make the group table for D_4 , ordering the elements so that the cosets are in blocks, see if the blocks form a group with four elements, and determine what group of order 4 the blocks form.
11. Find the index of $\langle \rho^2 \rangle$ in the group D_6 .
12. Find the index of $\langle 3 \rangle$ in the group \mathbb{Z}_{24} .
13. Find the index of $12\mathbb{Z}$ in \mathbb{Z} .
14. Find the index of $12\mathbb{Z}$ in $3\mathbb{Z}$.
15. Let $\sigma = (1, 2, 5, 4)(2, 3)$ in S_5 . Find the index of $\langle \sigma \rangle$ in S_5 .
16. Let $\mu = (1, 2, 4, 5)(3, 6)$ in S_6 . Find the index of $\langle \mu \rangle$ in S_6 .

Concepts

In Exercises 17 through 19, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

17. Let G be a group and let $H \subseteq G$. The *left coset of H containing a* is $aH = \{ah \mid h \in H\}$.
18. Let G be a group and let $H \leq G$. The *index of H in G* is the number of right cosets of H in G .
19. Let $\phi : G \rightarrow G'$. Then the *kernel of ϕ* is $\text{Ker}(\phi) = \{g \in G \mid \phi(g) = e\}$.
20. Determine whether each of the following is true or false.
 - a. Every subgroup of every group has left cosets.
 - b. The number of left cosets of a subgroup of a finite group divides the order of the group.
 - c. Every group of prime order is abelian.
 - d. One cannot have left cosets of a finite subgroup of an infinite group.
 - e. A subgroup of a group is a left coset of itself.
 - f. Only subgroups of finite groups can have left cosets.
 - g. A_n is of index 2 in S_n for $n > 1$.
 - h. The theorem of Lagrange is a nice result.
 - i. Every finite group contains an element of every order that divides the order of the group.
 - j. Every finite cyclic group contains an element of every order that divides the order of the group.
 - k. The kernel of a homomorphism is a subgroup of the range of the homomorphism.
 - l. Left cosets and right cosets of the kernel of a homomorphism are the same.

In Exercises 21 through 26, give an example of the desired subgroup and group if possible. If impossible, say why it is impossible.

21. A subgroup $H \leq G$ with G infinite and H having only a finite number of left cosets in G .
22. A subgroup of an abelian group G whose left cosets and right cosets give different partitions of G .
23. A subgroup of a group G whose left cosets give a partition of G into just one cell.
24. A subgroup of a group of order 6 whose left cosets give a partition of the group into 6 cells.
25. A subgroup of a group of order 6 whose left cosets give a partition of the group into 12 cells.
26. A subgroup of a group of order 6 whose left cosets give a partition of the group into 4 cells.

Proof Synopsis

27. Give a one-sentence synopsis of the proof of the Theorem of Lagrange.

Theory

28. Prove that the relation \sim_R that is used to define right cosets is an equivalence relation.
29. Let H be a subgroup of a group G and let $g \in G$. Define a one-to-one map of H onto Hg . Prove that your map is one-to-one and is onto Hg .
30. Let H be a subgroup of a group G such that $g^{-1}hg \in H$ for all $g \in G$ and all $h \in H$. Show that every left coset gH is the same as the right coset Hg .
31. Let H be a subgroup of a group G . Prove that if the partition of G into left cosets of H is the same as the partition into right cosets of H , then $g^{-1}hg \in H$ for all $g \in G$ and all $h \in H$. (Note that this is the converse of Exercise 30.)

Let H be a subgroup of a group G and let $a, b \in G$. In Exercises 32 through 35 prove the statement or give a counterexample.

32. If $aH = bH$, then $Ha = Hb$.
33. If $Ha = Hb$, then $b \in Ha$.
34. If $aH = bH$, then $Ha^{-1} = Hb^{-1}$.
35. If $aH = bH$, then $a^2H = b^2H$.
36. Let G be a group of order pq , where p and q are prime numbers. Show that every proper subgroup of G is cyclic.
37. Show that there are the same number of left as right cosets of a subgroup H of a group G ; that is, exhibit a one-to-one map of the collection of left cosets onto the collection of right cosets. (Note that this result is obvious by counting for finite groups. Your proof must hold for any group.)
38. Exercise 29 of Section 2 showed that every finite group of even order $2n$ contains an element of order 2. Using the theorem of Lagrange, show that if n is odd, then an abelian group of order $2n$ contains precisely one element of order 2.
39. Show that a group with at least two elements but with no proper nontrivial subgroups must be finite and of prime order.
40. Prove Theorem 10.11 [Hint: Let $\{a_iH \mid i = 1, \dots, r\}$ be the collection of distinct left cosets of H in G and $\{b_jK \mid j = 1, \dots, s\}$ be the collection of distinct left cosets of K in H . Show that

$$\{(a_i b_j)K \mid i = 1, \dots, r; j = 1, \dots, s\}$$

is the collection of distinct left cosets of K in G .]

41. Show that if H is a subgroup of index 2 in a finite group G , then every left coset of H is also a right coset of H .
42. Show that if a group G with identity e has finite order n , then $a^n = e$ for all $a \in G$.
43. Show that every left coset of the subgroup \mathbb{Z} of the additive group of real numbers contains exactly one element x such that $0 \leq x < 1$.
44. Show that the function *sine* assigns the same value to each element of any fixed left coset of the subgroup $\langle 2\pi \rangle$ of the additive group \mathbb{R} of real numbers. (Thus *sine* induces a well-defined function on the set of cosets; the value of the function on a coset is obtained when we choose an element x of the coset and compute $\sin x$.)
45. Let H and K be subgroups of a group G . Define \sim on G by $a \sim b$ if and only if $a = hbk$ for some $h \in H$ and some $k \in K$.
- Prove that \sim is an equivalence relation on G .
 - Describe the elements in the equivalence class containing $a \in G$. (These equivalence classes are called **double cosets**.)
46. Let S_A be the group of all permutations of the set A , and let c be one particular element of A .
- Show that $\{\sigma \in S_A \mid \sigma(c) = c\}$ is a subgroup $S_{c,c}$ of S_A .
 - Let $d \neq c$ be another particular element of A . Is $S_{c,d} = \{\sigma \in S_A \mid \sigma(c) = d\}$ a subgroup of S_A ? Why or why not?
 - Characterize the set $S_{c,d}$ of part (b) in terms of the subgroup $S_{c,c}$ of part (a).

47. Show that a finite cyclic group of order n has exactly one subgroup of each order d dividing n , and that these are all the subgroups it has.
48. The **Euler phi-function** is defined for positive integers n by $\varphi(n) = s$, where s is the number of positive integers less than or equal to n that are relatively prime to n . Use Exercise 47 to show that

$$n = \sum_{d|n} \varphi(d),$$

the sum being taken over all positive integers d dividing n . [Hint: Note that the number of generators of \mathbb{Z}_d is $\varphi(d)$ by Corollary 6.17.]

49. Let G be a finite group. Show that if for each positive integer m the number of solutions x of the equation $x^m = e$ in G is at most m , then G is cyclic. [Hint: Use Theorem 10.9 and Exercise 48 to show that G must contain an element of order $n = |G|$.]
50. Show that a finite group cannot be written as the union of two of its proper subgroups. Does the statement remain true if “two” is replaced by “three”? (This was problem B-2 on the 1969 Putnam Exam.)

SECTION 11

†PLANE ISOMETRIES

Consider the Euclidean plane \mathbb{R}^2 . An **isometry** of \mathbb{R}^2 is a permutation $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that preserves distance, so that the distance between points P and Q is the same as the distance between the points $\phi(P)$ and $\phi(Q)$ for all points P and Q in \mathbb{R}^2 . If ψ is also an isometry of \mathbb{R}^2 , then the distance between $\psi(\phi(P))$ and $\psi(\phi(Q))$ must be the same as the distance between $\phi(P)$ and $\phi(Q)$, which in turn is the distance between P and Q , showing that the composition of two isometries is again an isometry. Since the identity map is an isometry and the inverse of an isometry is an isometry, we see that the isometries of \mathbb{R}^2 form a subgroup of the group of all permutations of \mathbb{R}^2 .

Given any subset S of \mathbb{R}^2 , the isometries of \mathbb{R}^2 that carry S onto itself form a subgroup of the group of isometries. This subgroup is the **group of symmetries** of S in \mathbb{R}^2 . Although we defined the dihedral group D_n as one-to-one maps from the vertices of a regular n -gon onto itself that preserves edges, we can extend each map in D_n to an isometry of the whole plane; μ is reflection across the x -axis and ρ is rotation about the origin by $\frac{2\pi}{n}$. So we can think of D_n as the group of isometries of a regular n -gon in \mathbb{R}^2 .

Everything we have defined in the two preceding paragraphs could equally well have been done for n -dimensional Euclidean space \mathbb{R}^n , but we will concern ourselves chiefly with plane isometries here.

It can be proved that every isometry of the plane is one of just four types (see Artin [5]). We will list the types and show, for each type, a labeled figure that can be carried into itself by an isometry of that type. In each of Figs. 11.1, 11.3, and 11.4, consider the line with spikes shown to be extended infinitely to the left and to the right. We also give an example of each type in terms of coordinates.

translation τ : Slide every point the same distance in the same direction. See Fig. 11.1. (Example: $\tau(x, y) = (x, y) + (2, -3) = (x + 2, y - 3)$.)

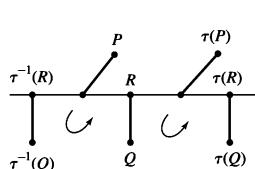
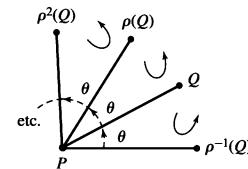
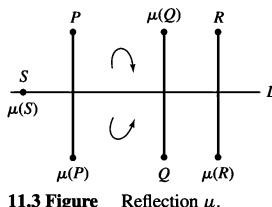
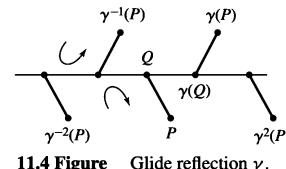
rotation ρ : Rotate the plane about a point P through an angle θ . See Fig. 11.2. (Example: $\rho(x, y) = (-y, x)$ is a rotation through 90° counterclockwise about the origin $(0, 0)$.)

reflection μ : Map each point into its mirror image (μ for mirror) across a line L , each point of which is left fixed by μ . See Fig. 11.3. The line L is the *axis of reflection*. (Example: $\mu(x, y) = (y, x)$ is a reflection across the line $y = x$.)

† This section is not used in the remainder of the text.

glide reflection γ : The product of a translation and a reflection across a line mapped into itself by the translation. See Fig. 11.4. (Example: $\gamma(x, y) = (x + 4, -y)$ is a glide reflection along the x -axis.)

Notice the little curved arrow that is carried into another curved arrow in each of Figs. 11.1 through 11.4. For the translation and rotation, the counterclockwise directions of the curved arrows remain the same, but for the reflection and glide reflection, the counterclockwise arrow is mapped into a clockwise arrow. We say that translations and rotations *preserve orientation*, while the reflection and glide reflection *reverse orientation*. We do not classify the identity isometry as any definite one of the four types listed; it could equally well be considered to be a translation by the zero vector or a rotation about any point through an angle of 0° . We always consider a glide reflection to be the product of a reflection and a translation that is different from the identity isometry.

11.1 Figure Translation τ .11.2 Figure Rotation ρ .11.3 Figure Reflection μ .11.4 Figure Glide reflection γ .

The theorem that follows describes the possible structures of finite subgroups of the full isometry group.

11.5 Theorem Every finite group G of isometries of the plane is isomorphic to either the Klein 4-group, \mathbb{Z}_n for $n \geq 1$, or D_n for some $n \geq 3$.

Proof (Outline) First we show that there is a point in the plane that is fixed by every element of G . We let $G = \{\phi_1, \phi_2, \phi_3, \dots, \phi_m\}$ and $(x_i, y_i) = \phi_i(0, 0)$. Then the point

$$P = (\bar{x}, \bar{y}) = \left(\frac{x_1 + x_2 + x_3 + \dots + x_m}{m}, \frac{y_1 + y_2 + y_3 + \dots + y_m}{m} \right)$$

is the center of mass of the set $S = \{(x_i, y_i) \mid 1 \leq i \leq m\}$ where each point is weighted by the number of ϕ_i that map $(0, 0)$ to that point. It is easy to see that the isometries in G permute the points in S since for each i and j , $\phi_i \circ \phi_j = \phi_k$ for some k . Thus $\phi_i(x_j, y_j) = (x_k, y_k)$. This implies the center of mass of $\phi(S)$ is the same as the center of mass of S . It can be shown that given the distances from the center of mass to the points of the set S , the center of mass is the only point having these distances from the points of S . This says that (\bar{x}, \bar{y}) is fixed by every isometry in G .

The orientation preserving isometries of G form a subgroup H of G which is either all of G or else of order $m/2$. You are asked to prove this in Exercise 22. Of course H consists of the identity and possibly rotations about the point (\bar{x}, \bar{y}) . If H has only one element, then G has one or two elements and is therefore isomorphic with \mathbb{Z}_1 or \mathbb{Z}_2 . If H has two elements, then G has two or four elements and is therefore isomorphic with either the Klein 4-group, \mathbb{Z}_4 , or \mathbb{Z}_2 . So we can assume that H has at least three elements.

If we choose a rotation ρ in H that rotates through the smallest positive angle θ among all the elements of H , ρ generates H . The proof of this fact is similar to the proof that a subgroup of a cyclic group is cyclic and you are asked to provide the details of the proof in Exercise 23. If $G = H$, then G is isomorphic with \mathbb{Z}_n . So we now assume that G contains a reflection, say μ . Then the coset μH contains only isometries of G that reverse orientation. Each coset H and μH contains half the elements of G , so $G = H \cup \mu H$.

Consider now a regular n -gon (recall that we are assuming that $n \geq 3$) with center the point (\bar{x}, \bar{y}) and having a vertex v_0 on the line fixed by μ . Each element of G permutes the vertices of the n -gon and preserves edges. Furthermore, no two elements of G permute the vertices in the same way. Thus G is isomorphic with a subgroup of the dihedral group D_n . Since $|G| = |D_n|$, G is isomorphic with D_n . \blacklozenge

In Theorem 11.5 the Klein 4-group, V , seems like an exception. However, V fits into the family of dihedral groups since V has two elements of order 2, a and b , with the property that $ab = ba^{-1}$. Sometimes V is denoted D_2 and considered a dihedral group. The isometries of the plane that map a line segment to itself are isomorphic with V .

The precedings theorem gives the complete story about finite plane isometry groups. We turn now to some infinite groups of plane isometries that arise naturally in decorating and art. Among these are the *discrete frieze groups*. A discrete frieze consists of a pattern of finite width and height that is repeated endlessly in both directions along its baseline to form a strip of infinite length but finite height; think of it as a decorative border strip that goes around a room next to the ceiling on wallpaper. We consider those isometries that carry each basic pattern onto itself or onto another instance of the pattern in the frieze. The set of all such isometries is called the "**frieze group**." All discrete frieze groups are infinite and have a subgroup isomorphic to \mathbb{Z} generated by the translation that slides the frieze lengthwise until the basic pattern is superimposed on the position of its next neighbor pattern in that direction. As a simple example of a discrete frieze, consider integral signs spaced equal distances apart and continuing infinitely to the left and right, indicated schematically like this.

Let us consider the integral signs to be one unit apart. The symmetry group of this frieze is generated by a translation τ sliding the plane one unit to the right, and by a rotation ρ of 180° about a point in the center of some integral sign. There are no horizontal or vertical reflections, and no glide reflections. This frieze group is nonabelian; we can check that $\tau\rho = \rho\tau^{-1}$. This relation between τ and ρ looks very familiar. The dihedral group D_n is also generated by two elements ρ and μ that satisfy the relation $\rho\mu = \mu\rho^{-1}$. If τ and ρ in the frieze group are replaced by ρ and μ , respectively, we have the same relation. In D_n , μ has order 2, as does ρ in the frieze group, but the element ρ in D_n has order n while τ has infinite order. Thus it is natural to use the notation D_∞ for this nonabelian frieze group.

As another example, consider the frieze given by an infinite string of D's.

…RRRRRRRRRRRRR…

Its group is generated by a translation τ one step to the right and by a vertical reflection μ across a horizontal line cutting through the middle of all the D's. We can check that these group generators commute this time, that is, $\tau\mu = \mu\tau$, so this frieze group is abelian and is isomorphic to $\mathbb{Z} \times \mathbb{Z}_2$.

It can be shown that if we classify such discrete friezes only by whether or not their groups contain a

rotation	horizontal axis reflection
vertical axis reflection	nontrivial glide reflection

then there are a total of seven possibilities. A *nontrivial glide reflection* in a symmetry group is one that is not equal to a product of a translation in that group and a reflection in that group. The group for the string of D's above contains glide reflections across the horizontal line through the centers of the D's, but the translation component of each glide reflection is also in the group so they are all considered trivial glide reflections in that group. The frieze group for

$$\begin{array}{ccccccccccccc} \dots & \mathbf{D} & & \dots \\ \dots & \mathbf{D} & & \dots \end{array}$$

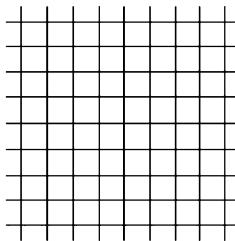
contains a nontrivial glide reflection whose translation component is not an element of the group. The exercises exhibit the seven possible cases, and ask you to tell, for each case, which of the four types of isometries displayed above appear in the symmetry group. We do not obtain seven different group structures. Each of the groups obtained can be shown to be isomorphic to one of

$$\mathbb{Z}, \quad D_\infty, \quad \mathbb{Z} \times \mathbb{Z}_2, \quad \text{or} \quad D_\infty \times \mathbb{Z}_2.$$

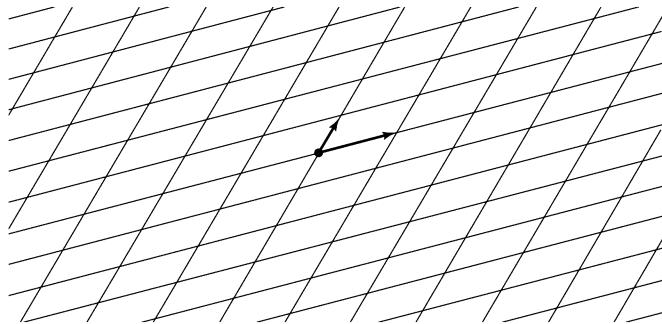
Equally interesting is the study of symmetries when a pattern in the shape of a square, parallelogram, rhombus, or hexagon is repeated by translations along *two non-parallel vector directions* to fill the entire plane, like patterns that appear on wallpaper. These groups are called the *wallpaper groups* or the *plane crystallographic groups*. While a frieze could not be carried into itself by a rotation through a positive angle less than 180° , it is possible to have rotations of 60° , 90° , 120° , and 180° for some of these plane-filling patterns. Figure 11.6 provides an illustration where the pattern consists of a square. We are interested in the group of plane isometries that carry this square onto itself or onto another square. Generators for this group are given by two translations (one sliding a square to the next neighbor to the right and one to the next above), by a rotation through 90° about the center of a square, and by a reflection in a vertical (or horizontal) line along the edges of the square. The one reflection is all that is needed to “turn the plane over”; a diagonal reflection can also be used. After being turned over, the translations and rotations can be used again. The isometry group for this *periodic pattern* in the plane surely contains a subgroup isomorphic to $\mathbb{Z} \times \mathbb{Z}$ generated by the unit translations to the right and upward, and a subgroup isomorphic to D_4 generated by those isometries that carry one square (it can be any square) into itself.

If we consider the plane to be filled with parallelograms as in Fig. 11.7, we do not get all the types of isometries that we did for Fig. 11.6. The symmetry group this time is generated by the translations indicated by the arrows and a rotation through 180° about any vertex of a parallelogram.

It can be shown that there are 17 different types of wallpaper patterns when they are classified according to the types of rotations, reflections, and nontrivial glide reflections that they admit. We refer you to Gallian [8] for pictures of these 17 possibilities and a chart to help you identify them. The exercises illustrate a few of them. The situation



11.6 Figure



11.7 Figure

in space is more complicated; it can be shown that there are 230 three-dimensional crystallographic groups. The final exercise we give involves rotations in space.

M. C. Escher (1898–1973) was an artist whose work included plane-filling patterns. In the exercises you are asked to analyze two of his works of this type.

■ EXERCISES 11

1. This exercise shows that the group of symmetries of a certain type of geometric figure may depend on the dimension of the space in which we consider the figure to lie.
 - a. Describe all symmetries of a point in the real line \mathbb{R} ; that is, describe all isometries of \mathbb{R} that leave one point fixed.
 - b. Describe all symmetries (translations, reflections, etc.) of a point in the plane \mathbb{R}^2 .
 - c. Describe all symmetries of a line segment in \mathbb{R} .
 - d. Describe all symmetries of a line segment in \mathbb{R}^2 .
 - e. Describe some symmetries of a line segment in \mathbb{R}^3 .
2. Let P stand for an orientation preserving plane isometry and R for an orientation reversing one. Fill in the table with P or R to denote the orientation preserving or reversing property of a product.

	P	R
P		
R		

3. Fill in the table to give *all* possible types of plane isometries given by a product of two types as indicated in Tables 11.1 through 11.4. For example, a product of two rotations may be a rotation, or it may be another type. Fill in the box corresponding to $\rho\rho$ with both letters. Use your answer to Exercise 2 to eliminate some types. Eliminate the identity from consideration.

	τ	ρ	μ	γ
τ				
ρ				
μ				
γ				

4. Draw a plane figure that has a one-element group as its group of symmetries in \mathbb{R}^2 .
5. Draw a plane figure that has a two-element group as its group of symmetries in \mathbb{R}^2 .
6. Draw a plane figure that has a three-element group as its group of symmetries in \mathbb{R}^2 .
7. Draw a plane figure that has a four-element group isomorphic to \mathbb{Z}_4 as its group of symmetries in \mathbb{R}^2 .
8. Draw a plane figure that has a four-element group isomorphic to the Klein 4-group V as its group of symmetries in \mathbb{R}^2 .
9. For each of the four types of plane isometries (other than the identity), give the possibilities for the order of an isometry of that type in the group of plane isometries.
10. A plane isometry ϕ has a *fixed point* if there exists a point P in the plane such that $\phi(P) = P$. Which of the four types of plane isometries (other than the identity) can have a fixed point?
11. Referring to Exercise 10, which types of plane isometries, if any, have exactly one fixed point?
12. Referring to Exercise 10, which types of plane isometries, if any, have exactly two fixed points?
13. Referring to Exercise 10, which types of plane isometries, if any, have an infinite number of fixed points?
14. Argue geometrically that a plane isometry that leaves three noncolinear points fixed must be the identity map.
15. Using Exercise 14, show algebraically that if two plane isometries ϕ and ψ agree on three noncolinear points, that is, if $\phi(P_i) = \psi(P_i)$ for noncolinear points P_1, P_2 , and P_3 , then ϕ and ψ are the same map.
16. Do the rotations, together with the identity map, form a subgroup of the group of plane isometries? Why or why not?
17. Do the translations, together with the identity map, form a subgroup of the group of plane isometries? Why or why not?
18. Do the rotations about one particular point P , together with the identity map, form a subgroup of the group of plane isometries? Why or why not?
19. Does the reflection across one particular line L , together with the identity map, form a subgroup of the group of plane isometries? Why or why not?
20. Do the glide reflections, together with the identity map, form a subgroup of the group of plane isometries? Why or why not?
21. Which of the four types of plane isometries can be elements of a *finite* subgroup of the group of plane isometries?
22. Completing a detail of the proof of Theorem 11.5, let G be a finite group of plane isometries. Show that the rotations in G , together with the identity isometry, form a subgroup H of G , and that either $H = G$ or $|G| = 2|H|$. [Hint: Use the same method that we used to show that $|S_n| = 2|A_n|$.]

23. Completing a detail in the proof of Theorem 11.5, let G be a finite group consisting of the identity isometry and rotations about one point P in the plane. Show that G is cyclic, generated by the rotation in G that turns the plane counterclockwise about P through the smallest angle $\theta > 0$. [Hint: Follow the idea of the proof that a subgroup of a cyclic group is cyclic.]

Exercises 24 through 30 illustrate the seven different types of friezes when they are classified according to their symmetries. Imagine the figure shown to be continued infinitely to the right and left. The symmetry group of a frieze always contains translations. For each of these exercises answer these questions about the symmetry group of the frieze.

- a. Does the group contain a rotation?
 - b. Does the group contain a reflection across a horizontal line?
 - c. Does the group contain a reflection across a vertical line?
 - d. Does the group contain a nontrivial glide reflection?
 - e. To which of the possible groups \mathbb{Z} , D_∞ , $\mathbb{Z} \times \mathbb{Z}_2$, or $D_\infty \times \mathbb{Z}_2$ do you think the symmetry group of the frieze is isomorphic?

24. F F F F F F F F F F F F

25. T T T T T T T T T

26 EEEEEE EEEEEE

27 Z Z Z Z Z Z Z Z Z Z Z Z Z

27. 222222222222
28. ННННННННННННН

28. **H H H H H H H H H H**

29.]]]]]]

30. $\cap \cup \cap \cup \cap \cup \cap \cup$

Exercises 31 through 37 describe a pattern to be used to fill the plane by translation in the two directions given by the specified vectors. Answer these questions in each case.

- a. Does the symmetry group contain any rotations? If so, through what possible angles θ where $0 < \theta \leq 180^\circ$?
 - b. Does the symmetry group contain any reflections?
 - c. Does the symmetry group contain any nontrivial glide reflections?

31. A square with horizontal and vertical edges using translation directions given by vectors $(1, 0)$ and $(0, 1)$.

32. A square as in Exercise 31 using translation directions given by vectors $(1, 1/2)$ and $(0, 1)$.

33. A square as in Exercise 31 with the letter L at its center using translation directions given by vectors $(1, 0)$ and $(0, 1)$.

34. A square as in Exercise 31 with the letter E at its center using translation directions given by vectors $(1, 0)$ and $(0, 1)$.

35. A square as in Exercise 31 with the letter H at its center using translation directions given by vectors $(1, 0)$ and $(0, 1)$.

36. A regular hexagon with a vertex at the top using translation directions given by vectors $(1, 0)$ and $(1, \sqrt{3})$.

37. A regular hexagon with a vertex at the top containing an equilateral triangle with vertex at the top and centroid at the center of the hexagon, using translation directions given by vectors $(1, 0)$ and $(1, \sqrt{3})$.

Exercises 38 and 39 are concerned with art works of M. C. Escher. Find images of the indicated art by searching on the internet. Neglect the shading and colors in the figures and assume the markings in each human figure, reptile,

or horseman are the same, even though they may be invisible due to shading. Answer the same questions (a), (b), and (c) that were asked for Exercises 31 through 36, and also answer this part (d).

- d. Assuming horizontal and vertical coordinate axes with equal scales as usual, give vectors in the two nonparallel directions of vectors that generate the translation subgroup. Do not concern yourself with the length of these vectors.
38. *The Study of Regular Division of the Plane with Horsemen.*
39. *The Study of Regular Division of the Plane with Reptiles.*
40. Let $\phi : \mathbb{R} \rightarrow U$ be given by $\phi(\theta) = \cos(\theta) + i \sin(\theta)$ and $S = \phi[\mathbb{Z}]$.
- Show that any rotation mapping S to S is a rotation by an angle $n \in \mathbb{Z}$ where angles are measured in radians.
 - Show that reflection across the x -axis maps S to S .
 - What is the group of symmetries of S ?
41. Show that the rotations of a cube in space form a group isomorphic to S_4 . [Hint: A rotation of the cube permutes the diagonals through the center of the cube.]

Homomorphisms and Factor Groups

Section 12 Factor Groups

Section 13 Factor-Group Computations and Simple Groups

Section 14 Group Action on a Set

Section 15 Applications of G -Sets to Counting

SECTION 12

FACTOR GROUPS

Recall from Section 10 that for some group tables we can arrange the head on top and on the left so that the elements are grouped into left cosets of a subgroup in such a way that the coset blocks form a group table. We start this section by looking more closely at why the cosets of $\{0, 3\} \leq \mathbb{Z}_6$ form a group and why the cosets of the subgroup $\{\iota, \mu\} \leq D_3$ do not. Table 12.1 is the group table for \mathbb{Z}_6 with the heads at the top and left sorted by cosets of $\{0, 3\}$.

12.1 Table

$+_6$	0	3	1	4	2	5
0	0	3	1	4	2	5
3	3	0	4	1	5	2
1	1	4	2	5	3	0
4	4	1	5	2	0	3
2	2	5	3	0	4	1
5	5	2	0	3	1	4

According to Table 12.1 the coset $\{1, 4\}$ plus the coset $\{2, 5\}$ is the coset $\{0, 3\}$. This means that if we add either 1 or 4 to either 2 or 5 in \mathbb{Z}_6 , we should get either 0 or 3. This is easily checked by adding the four possibilities.

$$1 +_6 2 = 3$$

$$1 +_6 5 = 0$$

$$4 +_6 2 = 0$$

$$4 +_6 5 = 3$$

We observe that if we wish to break up a group into its left cosets so the group table shows an operation on the left cosets, we need to be sure that if a_1, a_2 are in the same

left coset and b_1, b_2 are in the same left coset, then a_1b_1 and a_2b_2 are in the same left coset. If this condition is satisfied for a subgroup $H \leq G$, we say that the operation on the left cosets of H is **induced** by the operation of G or that the operation of G **induces** an operation on the left cosets of H . In this case for any $a, b \in G$ we write

$$(aH)(bH) = (ab)H$$

to mean that the product of any element in aH multiplied by any element in bH must be in the left coset $(ab)H$.

12.2 Example We show that the operation $+$ in the group \mathbb{Z} induces an operation on the cosets of $5\mathbb{Z} \leq \mathbb{Z}$. We first list the left cosets.

$$\begin{aligned} 5\mathbb{Z} &= \{\dots -10, -5, 0, 5, 10, \dots\} \\ 1 + 5\mathbb{Z} &= \{\dots -9, -4, 1, 6, 11, \dots\} \\ 2 + 5\mathbb{Z} &= \{\dots -8, -3, 2, 7, 12, \dots\} \\ 3 + 5\mathbb{Z} &= \{\dots -7, -2, 3, 8, 13, \dots\} \\ 4 + 5\mathbb{Z} &= \{\dots -6, -1, 4, 9, 14, \dots\} \end{aligned}$$

Let a_1 and a_2 be in the same left coset of $5\mathbb{Z}$. Then $a_2 = a_1 + 5r$ for some $r \in \mathbb{Z}$. We also let b_1, b_2 be in the same left coset of $5\mathbb{Z}$. Then $b_2 = b_1 + 5s$ for some $s \in \mathbb{Z}$. We compute $a_2 + b_2$.

$$\begin{aligned} a_2 + b_2 &= (a_1 + 5r) + (b_1 + 5s) \\ &= a_1 + 5r + b_1 + 5s \\ &= a_1 + b_1 + 5r + 5s \\ &= (a_1 + b_1) + 5(r + s) \\ &\in (a_1 + b_1) + 5\mathbb{Z} \end{aligned} \tag{1}$$

$$\tag{2}$$

So $a_2 + b_2$ is in the same coset as $a_1 + b_1$, which says that addition in \mathbb{Z} induces an operation on the five left cosets $5\mathbb{Z}, 1 + 5\mathbb{Z}, 2 + 5\mathbb{Z}, 3 + 5\mathbb{Z}, 4 + 5\mathbb{Z}$. Looking back at the calculations, we see that only properties shared by all groups were used in each step except in line (1) where we used the fact that \mathbb{Z} is abelian. Furthermore, line (2) is not necessary since $5\mathbb{Z}$ is a subgroup of \mathbb{Z} so we know that $5\mathbb{Z}$ is closed under addition. From this example, it appears that as long as G is an abelian group, the operation of G induces an operation on the left cosets of any subgroup of G . \blacktriangle

In Equation (1) of Example 12.2 we used the fact that $5r + b_1 = b_1 + 5r$. If we were doing the same computation in multiplicative notation and using any group G and subgroup H of G , this would correspond to $hb_1 = b_1h$. If the group G is not abelian, then this computation fails. However, we can weaken the abelian condition slightly and still get an induced operation on the left cosets. All we really need is that $hb_1 = b_1h'$ for some $h' \in H$. This happens when the left coset b_1H is the same set as the right coset Hb_1 .

12.3 Definition Let H be a subgroup of G . We say that H is a **normal** subgroup of G if for all $g \in G$, $gH = Hg$. If H is a normal subgroup of G , we write $H \trianglelefteq G$. \blacksquare

Recall that Theorem 10.17 states that if $\phi : G \rightarrow G'$ is a group homomorphism and e' is the identity element in G' , then $\text{Ker}(\phi) = \{g \in G \mid \phi(g) = e'\}$ has the property that left and right cosets of $\text{Ker}(\phi)$ are the same. So the kernel of any homomorphism is a normal subgroup.

12.4 Example The subgroup of even permutations $A_n \leq S_n$ is normal since A_n is the kernel of the homomorphism $\text{sgn} : S_n \rightarrow \{1, -1\}$. \blacktriangle

12.5 Example If $H \leq G$ and G is an abelian group, then H is a normal subgroup of G . ▲

12.6 Example Let $H = \{A \in \text{GL}(n, \mathbb{R}) \mid \det(A) = 1\}$. The determinant map satisfies $\det(AB) = \det(A)\det(B)$, which means that the determinant map is a homomorphism, $\det : \text{GL}(n, \mathbb{R}) \rightarrow \mathbb{R}^*$. Thus $H = \text{Ker}(\det)$, which says that $H \trianglelefteq \text{GL}(n, \mathbb{R})$. This subgroup H is called the **special linear group** and it is denoted by $\text{SL}(n, \mathbb{R})$. ▲

12.7 Theorem Let H be a subgroup of a group G . Then left coset multiplication is well defined by the equation

$$(aH)(bH) = (ab)H$$

if and only if H is a normal subgroup of G .

Proof Suppose first that $(aH)(bH) = (ab)H$ does give a well-defined binary operation on left cosets. Let $a \in G$. We want to show that aH and Ha are the same set. We use the standard technique of showing that each is a subset of the other.

Let $x \in aH$. Choosing representatives $x \in aH$ and $a^{-1} \in a^{-1}H$, we have $(xH)(a^{-1}H) = (xa^{-1})H$. On the other hand, choosing representatives $a \in aH$ and $a^{-1} \in a^{-1}H$, we see that $(aH)(a^{-1}H) = eH = H$. Using our assumption that left coset multiplication by representatives is well defined, we must have $xa^{-1} = h \in H$. Then $x = ha$, so $x \in Ha$ and $aH \subseteq Ha$. We leave the symmetric proof that $Ha \subseteq aH$ to Exercise 26.

We turn now to the converse: If H is a normal subgroup, then left coset multiplication by representatives is well-defined. Due to our hypothesis, we can simply say *cosets*, omitting *left* and *right*. Suppose we wish to compute $(aH)(bH)$. Choosing $a \in aH$ and $b \in bH$, we obtain the coset $(ab)H$. Choosing different representatives $ah_1 \in aH$ and $bh_2 \in bH$, we obtain the coset ah_1bh_2H . We must show that these are the same cosets. Now $h_1b \in Hb = bH$, so $h_1b = bh_3$ for some $h_3 \in H$. Thus

$$(ah_1)(bh_2) = a(h_1b)h_2 = a(bh_3)h_2 = (ab)(h_3h_2)$$

and $(ab)(h_3h_2) \in (ab)H$. Therefore, ah_1bh_2 is in $(ab)H$. ◆

Theorem 12.7 shows that we have an operation on the left cosets of $H \leq G$ induced by the operation on G if and only if H is a normal subgroup of G . We next verify that this operation makes G/H , the cosets of H in G , a group.

12.8 Corollary Let H be a normal subgroup of G . Then the cosets of H form a group G/H under the binary operation $(aH)(bH) = (ab)H$. ▲

Proof Computing, $(aH)[(bH)(cH)] = (aH)[(bc)H] = [a(bc)]H$, and similarly, we have $[(aH)(bH)](cH) = [(ab)c]H$, so associativity in G/H follows from associativity in G . Because $(aH)(eH) = (ae)H = aH = (ea)H = (eH)(aH)$, we see that $eH = H$ is the identity element in G/H . Finally, $(a^{-1}H)(aH) = (a^{-1}a)H = eH = (aa^{-1})H = (aH)(a^{-1}H)$ shows that $a^{-1}H = (aH)^{-1}$. ◆

12.9 Definition The group G/H in the preceding corollary is the **factor group** (or **quotient group**) of G by H . ■

12.10 Example Since \mathbb{Z} is an abelian group, $n\mathbb{Z}$ is a normal subgroup. Corollary 12.8 allows us to construct the factor group $\mathbb{Z}/n\mathbb{Z}$. For any integer m , the division algorithm says that $m = nq + r$ for some $0 \leq r < n$. Therefore, $m \in r + n\mathbb{Z}$. So $\mathbb{Z}/n\mathbb{Z} = \{k + n\mathbb{Z} \mid 0 \leq k < n\}$. Thus $\langle 1 + n\mathbb{Z} \rangle = \mathbb{Z}/n\mathbb{Z}$, which implies that $\mathbb{Z}/n\mathbb{Z}$ is cyclic and isomorphic with \mathbb{Z}_n . ▲

- 12.11 Example** Consider the abelian group \mathbb{R} under addition, and let $c \in \mathbb{R}^+$. The cyclic subgroup $\langle c \rangle$ of \mathbb{R} contains as elements

$$\dots -3c, -2c, -c, 0, c, 2c, 3c, \dots$$

Every coset of $\langle c \rangle$ contains just one element x such that $0 \leq x < c$. If we choose these elements as representatives of the cosets when computing in $\mathbb{R}/\langle c \rangle$, we find that we are computing their sum modulo c as discussed for the computation in \mathbb{R}_c in Section 3. For example, if $c = 5.37$, then the sum of the cosets $4.65 + \langle 5.37 \rangle$ and $3.42 + \langle 5.37 \rangle$ is the coset $8.07 + \langle 5.37 \rangle$, which contains $8.07 - 5.37 = 2.7$, which is $4.65 +_{5.37} 3.42$. Working with these coset elements x where $0 \leq x < c$, we thus see that the group \mathbb{R}_c of Section 3 is isomorphic to $\mathbb{R}/\langle c \rangle$ under an isomorphism ψ where $\psi(x) = x + \langle c \rangle$ for all $x \in \mathbb{R}_c$. Of course, $\mathbb{R}/\langle c \rangle$ is then also isomorphic to the circle group U of complex numbers of magnitude 1 under multiplication. \blacktriangleleft

We have seen that the group $\mathbb{Z}/\langle n \rangle$ is isomorphic to the group \mathbb{Z}_n , and as a set, $\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$, the set of nonnegative integers less than n . Example 12.11 shows that the group $\mathbb{R}/\langle c \rangle$ is isomorphic to the group \mathbb{R}_c . In Section 3, we choose the notation \mathbb{R}_c rather than the conventional $[0, c)$ for the half-open interval of nonnegative real numbers less than c . We did that to bring out now the comparison of these factor groups of \mathbb{Z} with these factor groups of \mathbb{R} .

Homomorphisms and Factor Groups

We learned that the kernel of any homomorphism $\phi : G \rightarrow G'$ is a normal subgroup of G . Do all normal subgroups arise in this way? That is, for any normal subgroup $H \trianglelefteq G$, is there a group homomorphism $\phi : G \rightarrow G'$ for some group G' such that H is the kernel of G ? The answer to the question is yes as we see in Theorem 12.12.

- 12.12 Theorem** Let H be a normal subgroup of G . Then $\gamma : G \rightarrow G/H$ given by $\gamma(x) = xH$ is a homomorphism with kernel H .

Proof Let $x, y \in G$. Then

$$\gamma(xy) = (xy)H = (xH)(yH) = \gamma(x)\gamma(y),$$

so γ is a homomorphism. Since $xH = H$ if and only if $x \in H$, we see that the kernel of γ is indeed H . \blacklozenge

Since the kernel of any homomorphism $\phi : G \rightarrow G'$ is a normal subgroup, it is natural to ask how the factor group $G/\text{Ker}(\phi)$ is related to G' . Theorem 12.12 and the next example illustrate that there is a very strong connection.

- 12.13 Example (Reduction Modulo n)** Let $\phi : \mathbb{Z} \rightarrow \mathbb{Z}_n$ be defined by letting $\phi(m)$ be the remainder when m is divided by n . We check that ϕ is a group homomorphism. Let $m_1, m_2 \in \mathbb{Z}$ and suppose that the division algorithm gives us

$$m_1 = nq_1 + r_1 \quad \text{and}$$

$$m_2 = nq_2 + r_2.$$

Then $m_1 + m_2 = n(q_1 + q_2) + r_1 + r_2$. If $r_1 + r_2 < n$, then

$$\phi(m_1 + m_2) = r_1 + r_2 = \phi(m_1) +_n \phi(m_2).$$

On the other hand, if $r_1 + r_2 \geq n$, then $m_1 + m_2 = n(q_1 + q_2 + 1) + (r_1 + r_2 - n)$ and $0 \leq r_1 + r_2 - n < n$, which implies

$$\phi(m_1 + m_2) = r_1 + r_2 - n = \phi(m_1) +_n \phi(m_2).$$

The kernel of ϕ is the set of all the multiples of n , $n\mathbb{Z}$. So $\mathbb{Z}/\text{Ker}(\phi) = \mathbb{Z}/n\mathbb{Z}$, which is isomorphic to \mathbb{Z}_n . \blacktriangle

The previous example is a special case of the Fundamental Homomorphism Theorem.

12.14 Theorem (The Fundamental Homomorphism Theorem) Let $\phi : G \rightarrow G'$ be a group homomorphism with kernel H . Then $\phi[G]$ is a group, and $\mu : G/H \rightarrow \phi[G]$ given by $\mu(gH) = \phi(g)$ is an isomorphism. If $\gamma : G \rightarrow G/H$ is the homomorphism given by $\gamma(g) = gH$, then $\phi(g) = \mu \circ \gamma(g)$ for each $g \in G$.

Proof Theorem 8.5 says that $\phi[G]$ is a subgroup of G' . Theorem 10.17 shows that the map $\mu : G/H \rightarrow \phi[G]$ is well defined. We show μ is a homomorphism. Let $aH, bH \in G/H$. Then $\mu((aH)(bH)) = \mu((ab)H) = \phi(ab) = \phi(a)\phi(b) = \mu(aH)\mu(bH)$. Since ϕ maps G onto $\phi[G]$, μ maps G/H onto $\phi[G]$. To show that μ is one-to-one, we compute the kernel of μ . Since $\mu(aH) = \phi(a)$, the kernel of μ is $\{aH \mid \phi(a) = e'\}$. But $\phi(a) = e'$ if and only if $a \in \text{Ker}(\phi) = H$. So $\text{Ker}(\mu) = \{H\}$ which is the trivial subgroup of G/H . By Corollary 10.19 μ is one-to-one, which completes the proof that μ is an isomorphism.

We next turn to the final statement of the theorem. Let $g \in G$. Then

$$\phi(g) = \mu(gH) = \mu(\gamma(g)) = \mu \circ \gamma(g).$$

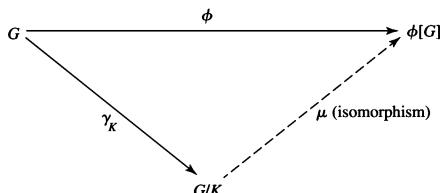
◆

The Fundamental Homomorphism Theorem is sometimes called the First Isomorphism Theorem. As the name suggests, there are other related theorems. In fact we will prove two others, the Second Isomorphism Theorem and the Third Isomorphism Theorem, in Section 16.

Theorem 12.14 states that $\phi(g) = \mu \circ \gamma(g)$. This can be visualized in Figure 12.15. If we start with an element $g \in G$, and map it to $\phi(g)$, we get the same result as first mapping g to $\gamma(g)$ and then mapping $\gamma(g)$ to $\mu \circ \gamma(g)$. When we have a situation like this, we say that the map ϕ can be *factored* as $\phi = \mu \circ \gamma$.

The isomorphism μ in Theorem 12.14 is referred to as a *natural* or *canonical* isomorphism, and the same adjectives are used to describe the homomorphism γ . There may be other isomorphisms and homomorphisms for these same groups, but the maps μ and γ have a special status with ϕ and are uniquely determined by Theorem 12.14.

In summary, every homomorphism with domain G gives rise to a factor group G/H , and every factor group G/H gives rise to a homomorphism mapping G into G/H . Homomorphisms and factor groups are closely related. We give an example indicating how useful this relationship can be.



12.15 Figure

12.16 Example Classify the group $(\mathbb{Z}_4 \times \mathbb{Z}_2)/(\{0\} \times \mathbb{Z}_2)$ according to the fundamental theorem of finitely generated abelian groups (Theorem 9.12).

Solution The projection map $\pi_1 : \mathbb{Z}_4 \times \mathbb{Z}_2 \rightarrow \mathbb{Z}_4$ given by $\pi_1(x, y) = x$ is a homomorphism of $\mathbb{Z}_4 \times \mathbb{Z}_2$ onto \mathbb{Z}_4 with kernel $\{0\} \times \mathbb{Z}_2$. By Theorem 12.14, we know that the given factor group is isomorphic to \mathbb{Z}_4 . \blacktriangle

Normal Subgroups and Inner Automorphisms

We derive some alternative characterizations of normal subgroups, which often provide us with an easier way to check normality than finding both the left and the right coset decompositions.

Suppose that H is a subgroup of G such that $ghg^{-1} \in H$ for all $g \in G$ and all $h \in H$. Then $gHg^{-1} = \{ghg^{-1} \mid h \in H\} \subseteq H$ for all $g \in G$. We claim that actually $gHg^{-1} = H$. We must show that $H \subseteq gHg^{-1}$ for all $g \in G$. Let $h \in H$. Replacing g by g^{-1} in the relation $ghg^{-1} \in H$, we obtain $g^{-1}h(g^{-1})^{-1} = g^{-1}hg = h_1$ where $h_1 \in H$. Consequently, $h = gh_1g^{-1} \in gHg^{-1}$, and we are done.

Suppose that $gH = Hg$ for all $g \in G$. Then $gh = h_1g$, so $ghg^{-1} \in H$ for all $g \in G$ and all $h \in H$. By the preceding paragraph, this means that $gHg^{-1} = H$ for all $g \in G$. Conversely, if $gHg^{-1} = H$ for all $g \in G$, then $ghg^{-1} = h_1$ so $gh = h_1g \in Hg$, and $gH \subseteq Hg$. But also, $g^{-1}Hg = H$ giving $g^{-1}hg = h_2$, so that $hg = h_2g$ and $Hg \subseteq gH$.

The comments after Definition 12.3 show that the kernel of any homomorphism is a normal subgroup of the domain. Also, Theorem 12.12 says that any normal subgroup is the kernel of some homomorphism.

We summarize our work as a theorem.

12.17 Theorem The following are four equivalent conditions for a subgroup H of a group G to be a *normal* subgroup of G .

1. $ghg^{-1} \in H$ for all $g \in G$ and $h \in H$.
2. $gHg^{-1} = H$ for all $g \in G$.
3. There is a group homomorphism $\phi : G \rightarrow G'$ such that $\text{Ker}(\phi) = H$.
4. $gH = Hg$ for all $g \in G$.

Condition (2) of Theorem 12.17 is often taken as the definition of a normal subgroup H of a group G . \blacklozenge

12.18 Example Every subgroup H of an abelian group G is normal. We need only note that $gh = hg$ for all $h \in H$ and all $g \in G$, so, of course, $ghg^{-1} = h \in H$ for all $g \in G$ and all $h \in H$. \blacktriangle

If G is a group and $g \in G$, then the map $i_g : G \rightarrow G$ defined by $i_g(x) = gxg^{-1}$ is a group homomorphism since $i_g(xy) = gxg^{-1}xyg^{-1} = gxg^{-1}gyg^{-1} = i_g(x)i_g(y)$. We see that $gag^{-1} = gbg^{-1}$ if and only if $a = b$, so i_g is one-to-one. Since $g(g^{-1}yg)g^{-1} = y$, we see that i_g is onto G , so it is an isomorphism of G with itself.

12.19 Definition An isomorphism $\phi : G \rightarrow G$ of a group G with itself is an **automorphism** of G . The automorphism $i_g : G \rightarrow G$, where $i_g(x) = gxg^{-1}$ for all $x \in G$, is the **inner automorphism of G by g** . Performing i_g on x is called **conjugation of x by g** . \blacksquare

The equivalence of conditions (1) and (2) in Theorem 12.17 shows that $gH = Hg$ for all $g \in G$ if and only if $i_g[H] = H$ for all $g \in G$, that is, if and only if H is **invariant** under all inner automorphisms of G . It is important to realize that $i_g[H] = H$ is an

equation in *sets*; we need not have $i_g(h) = h$ for all $h \in H$. That is i_g may perform a nontrivial *permutation* of the set H . We see that the normal subgroups of a group G are precisely those that are invariant under all inner automorphisms. A subgroup K of G is a **conjugate subgroup** of H if $K = i_g[H] = gHg^{-1}$ for some $g \in G$.

■ EXERCISES 12

Computations

In Exercises 1 through 8, find the order of the given factor group.

1. $\mathbb{Z}_6/\langle 3 \rangle$
2. $(\mathbb{Z}_4 \times \mathbb{Z}_{12})/(\langle 2 \rangle \times \langle 2 \rangle)$
3. $(\mathbb{Z}_4 \times \mathbb{Z}_2)/(\langle 2, 1 \rangle)$
4. $(\mathbb{Z}_3 \times \mathbb{Z}_5)/(\{0\} \times \mathbb{Z}_5)$
5. $(\mathbb{Z}_3 \times \mathbb{Z}_6)/(\langle 1, 1 \rangle)$
6. $(\mathbb{Z}_{50} \times \mathbb{Z}_{75})/(\langle 15, 15 \rangle)$
7. $(\mathbb{Z}_{26} \times \mathbb{Z}_{15})/(\langle 1, 1 \rangle)$
8. $(\mathbb{Z}_8 \times S_3)/(\langle (2, (1, 2, 3)) \rangle)$

In Exercises 9 through 15, give the order of the element in the factor group.

9. $5 + \langle 4 \rangle$ in $\mathbb{Z}_{12}/\langle 4 \rangle$
10. $26 + \langle 12 \rangle$ in $\mathbb{Z}_{60}/\langle 12 \rangle$
11. $(2, 1) + \langle (1, 1) \rangle$ in $(\mathbb{Z}_3 \times \mathbb{Z}_6)/(\langle 1, 1 \rangle)$
12. $(3, 1) + \langle (1, 1) \rangle$ in $(\mathbb{Z}_4 \times \mathbb{Z}_4)/(\langle 1, 1 \rangle)$
13. $(2, 3) + \langle (0, 3) \rangle$ in $(\mathbb{Z}_{10} \times \mathbb{Z}_4)/(\langle 0, 3 \rangle)$
14. $(2, 5) + \langle (1, 2) \rangle$ in $(\mathbb{Z}_3 \times \mathbb{Z}_6)/(\langle 1, 2 \rangle)$
15. $(2, 0) + \langle (4, 4) \rangle$ in $(\mathbb{Z}_6 \times \mathbb{Z}_8)/(\langle 4, 4 \rangle)$
16. Compute $i_\rho[H]$ for the subgroup $H = \{\iota, \mu\}$ of the dihedral group D_3 .

Concepts

In Exercises 17 through 19, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

17. A *normal subgroup* H of G is one satisfying $hG = Gh$ for all $h \in H$.
18. A *normal subgroup* H of G is one satisfying $g^{-1}hg \in H$ for all $h \in H$ and all $g \in G$.
19. An *automorphism* of a group G is a homomorphism mapping G into G .
20. What is the importance of a *normal* subgroup of a group G ?

Students often write nonsense when first proving theorems about factor groups. The next two exercises are designed to call attention to one basic type of error.

21. A student is asked to show that if H is a normal subgroup of an abelian group G , then G/H is abelian. The student's proof starts as follows:

We must show that G/H is abelian. Let a and b be two elements of G/H .

- a. Why does the instructor reading this proof expect to find nonsense from here on in the student's paper?
- b. What should the student have written?
- c. Complete the proof.

22. A **torsion group** is a group all of whose elements have finite order. A group is **torsion free** if the identity is the only element of finite order. A student is asked to prove that if G is a torsion group, then so is G/H for every normal subgroup H of G . The student writes

We must show that each element of G/H is of finite order. Let $x \in G/H$.

Answer the same questions as in Exercise 21.

23. Determine whether each of the following is true or false.

- a. It makes sense to speak of the factor group G/N if and only if N is a normal subgroup of the group G .
- b. Every subgroup of an abelian group G is a normal subgroup of G .
- c. The only automorphism of an abelian group is the identity map.

- d. Every factor group of a finite group is again of finite order.
- e. Every factor group of a torsion group is a torsion group. (See Exercise 22.)
- f. Every factor group of a torsion-free group is torsion free. (See Exercise 22.)
- g. Every factor group of an abelian group is abelian.
- h. Every factor group of a nonabelian group is nonabelian.
- i. $\mathbb{Z}/n\mathbb{Z}$ is cyclic of order n .
- j. $\mathbb{R}/n\mathbb{R}$ is cyclic of order n , where $n\mathbb{R} = \{nr \mid r \in \mathbb{R}\}$ and \mathbb{R} is under addition.

Theory

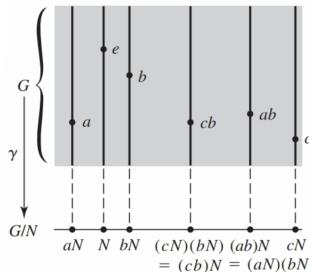
24. Let G_1 and G_2 be groups and $\pi_1 : G_1 \times G_2 \rightarrow G_1$ be the function defined by $\pi_1(a, b) = a$. Prove that π_1 is a homomorphism, find $\text{Ker}(\pi_1)$, and prove $(G_1 \times G_2)/\text{Ker}(\pi_1)$ is isomorphic to G_1 .
25. Let G_1 and G_2 be groups and $\phi : G_1 \times G_2 \rightarrow G_1 \times G_2$ be the function defined by $\phi(a, b) = (a, e_2)$ where e_2 is the identity in G_2 . Prove that ϕ is a homomorphism, find $\text{Ker}(\phi)$, and prove $(G_1 \times G_2)/\text{Ker}(\phi)$ is isomorphic to G_1 .
26. Complete the proof of Theorem 12.7 by showing that if H is a subgroup of a group G and if left coset multiplication $(aH)(bH) = (ab)H$ is well defined, then $Ha \subseteq aH$.
27. Prove that the torsion subgroup T of an abelian group G is a normal subgroup of G , and that G/T is torsion free. (See Exercise 22.)
28. A subgroup H is **conjugate to a subgroup K** of a group G if there exists an inner automorphism i_g of G such that $i_g[H] = K$. Show that conjugacy is an equivalence relation on the collection of subgroups of G .
29. Characterize the normal subgroups of a group G in terms of the cells where they appear in the partition given by the conjugacy relation in the preceding exercise.
30. Find all subgroups of D_3 that are conjugate to $H = \{\iota, \mu\}$. (See Exercise 28.)
31. (**Evaluation Homomorphism**) Let F be the set of all functions mapping the real numbers to the real numbers and let $c \in \mathbb{R}$. The sum of two functions $f + g$ is the function defined by $(f + g)(x) = f(x) + g(x)$. Function addition makes F a group. Let $\phi_c : F \rightarrow \mathbb{R}$ be defined by $\phi_c(f) = f(c)$.
 - a. Show that ϕ_c is a group homomorphism.
 - b. Find $\text{Ker}(\phi_c)$.
 - c. Identify the coset of $\text{Ker}(\phi_c)$ that contains the constant function $f(x) = 1$.
 - d. Find a well-known group that is isomorphic with $F/\text{Ker}(\phi_c)$. Use the Fundamental Homomorphism Theorem to prove your answer.
32. Let H be a normal subgroup of a group G , and let $m = (G : H)$. Show that $a^m \in H$ for every $a \in G$.
33. Show that an intersection of normal subgroups of a group G is again a normal subgroup of G .
34. Given any subset S of a group G , show that it makes sense to speak of the smallest normal subgroup that contains S . [Hint: Use Exercise 33.]
35. Let G be a group. An element of G that can be expressed in the form $aba^{-1}b^{-1}$ for some $a, b \in G$ is a **commutator** in G . The preceding exercise shows that there is a smallest normal subgroup C of a group G containing all commutators in G ; the subgroup C is the **commutator subgroup** of G . Show that G/C is an abelian group.
36. Show that if a finite group G has exactly one subgroup H of a given order, then H is a normal subgroup of G .
37. Show that if H and N are subgroups of a group G , and N is normal in G , then $H \cap N$ is normal in H . Show by an example that $H \cap N$ need not be normal in G .
38. Let G be a group containing at least one subgroup of a fixed finite order s . Show that the intersection of all subgroups of G of order s is a normal subgroup of G . [Hint: Use the fact that if H has order s , then so does $x^{-1}Hx$ for all $x \in G$.]

39. a. Show that all automorphisms of a group G form a group under function composition.
 b. Show that the inner automorphisms of a group G form a normal subgroup of the group of all automorphisms of G under function composition. [Warning: Be sure to show that the inner automorphisms do form a subgroup.]
40. Show that the set of all $g \in G$ such that $i_g : G \rightarrow G$ is the identity inner automorphism i_e is a normal subgroup of a group G .
41. Let G and G' be groups, and let H and H' be normal subgroups of G and G' , respectively. Let ϕ be a homomorphism of G into G' . Show that ϕ induces a natural homomorphism $\phi_* : (G/H) \rightarrow (G'/H')$ if $\phi[H] \subseteq H'$. (This fact is used constantly in algebraic topology.)
42. Use the properties $\det(AB) = \det(A) \cdot \det(B)$ and $\det(I_n) = 1$ for $n \times n$ matrices to show the $n \times n$ matrices with determinant ± 1 form a normal subgroup of $\mathrm{GL}(n, \mathbb{R})$.
43. Let G be a group, and let $\mathcal{P}(G)$ be the set of all subsets of G . For any $A, B \in \mathcal{P}(G)$, let us define the product subset $AB = \{ab \mid a \in A, b \in B\}$.
- Show that this multiplication of subsets is associative and has an identity element, but that $\mathcal{P}(G)$ is not a group under this operation.
 - Show that if N is a normal subgroup of G , then the set of cosets of N is closed under the above operation on $\mathcal{P}(G)$, and that this operation agrees with the multiplication given by the formula in Corollary 12.8.
 - Show (without using Corollary 12.8) that the cosets of N in G form a group under the above operation. Is its identity element the same as the identity element of $\mathcal{P}(G)$?

SECTION 13**FACTOR-GROUP COMPUTATIONS AND SIMPLE GROUPS**

Factor groups can be a tough topic for students to grasp. There is nothing like a bit of computation to strengthen understanding in mathematics. We start by attempting to improve our intuition concerning factor groups. Since we will be dealing with normal subgroups throughout this section, we often denote a subgroup of a group G by N rather than H .

Let N be a normal subgroup of G . In the factor group G/N , the subgroup N acts as identity element. We may regard N as being *collapsed* to a single element, either to 0 in additive notation or to e in multiplicative notation. This collapsing of N together with the algebraic structure of G require that other subsets of G , namely, the cosets of N , also each collapse into a single element in the factor group. A visualization of this collapsing is provided by Fig. 13.1. Recall from Theorem 12.12 that $\gamma : G \rightarrow G/N$ defined by $\gamma(a) = aN$ for $a \in G$ is a homomorphism of G onto G/N . We can view the “line” G/N at the bottom of Figure 13.1 as obtained by collapsing to a point each coset of N in a copy of G . Each point of G/N thus corresponds to a whole vertical line



13.1 Figure

segment in the shaded portion, representing a coset of N in G . It is crucial to remember that multiplication of cosets in G/N can be computed by multiplying in G , using any representative elements of the cosets as shown in the figure.

Additively, two elements of G will collapse into the same element of G/N if they differ by an element of N . Multiplicatively, a and b collapse together if ab^{-1} is in N . The degree of collapsing can vary from nonexistent to catastrophic. We illustrate the two extreme cases by examples.

13.2 Example The trivial subgroup $N = \{0\}$ of \mathbb{Z} is, of course, a normal subgroup. Compute $\mathbb{Z}/\{0\}$.

Solution Since $N = \{0\}$ has only one element, every coset of N has only one element. That is, the cosets are of the form $\{m\}$ for $m \in \mathbb{Z}$. There is no collapsing at all, and consequently, $\mathbb{Z}/\{0\} \cong \mathbb{Z}$. Each $m \in \mathbb{Z}$ is simply renamed $\{m\}$ in $\mathbb{Z}/\{0\}$. \blacktriangle

13.3 Example Let n be a positive integer. The set $n\mathbb{R} = \{nr \mid r \in \mathbb{R}\}$ is a subgroup of \mathbb{R} under addition, and it is normal since \mathbb{R} is abelian. Compute $\mathbb{R}/n\mathbb{R}$.

Solution A bit of thought shows that actually $n\mathbb{R} = \mathbb{R}$, because each $x \in \mathbb{R}$ is of the form $n(x/n)$ and $x/n \in \mathbb{R}$. Thus $\mathbb{R}/n\mathbb{R}$ has only one element, the subgroup $n\mathbb{R}$. The factor group is a trivial group consisting only of the identity element. \blacktriangle

As illustrated in Examples 13.2 and 13.3 for any group G , we have $G/\{e\} \cong G$ and $G/G \cong \{e\}$, where $\{e\}$ is the trivial group consisting only of the identity element e . These two extremes of factor groups are of little importance. We would like knowledge of a factor group G/N to give some information about the structure of G . If $N = \{e\}$, the factor group has the same structure as G and we might as well have tried to study G directly. If $N = G$, the factor group has no significant structure to supply information about G . If G is a finite group and $N \neq \{e\}$ is a normal subgroup of G , then G/N is a smaller group than G , and consequently may have a more simple structure than G . The multiplication of cosets in G/N reflects the multiplication in G , since products of cosets can be computed by multiplying in G representative elements of the cosets.

We give two examples showing that even when G/N has order 2, we may be able to deduce some useful results. If G is a finite group and G/N has just two elements, then we must have $|G| = 2|N|$. Note that every subgroup H containing just half the elements of a finite group G must be a normal subgroup, since for each element a in G but not in H , both the left coset aH and the right coset Ha must consist of all elements in G that are not in H . Thus the left and right cosets of H coincide and H is a normal subgroup of G .

13.4 Example Because $|S_n| = 2|A_n|$, we see that A_n is a normal subgroup of S_n , and S_n/A_n has order 2. Let σ be an odd permutation in S_n , so that $S_n/A_n = \{A_n, \sigma A_n\}$. Renaming the element A_n “even” and the element σA_n “odd,” the multiplication in S_n/A_n shown in Table 13.5 becomes

13.5 Table

	A_n	σA_n
A_n	A_n	σA_n
σA_n	σA_n	A_n

$$\begin{array}{ll} (\text{even})(\text{even}) = \text{even} & (\text{odd})(\text{even}) = \text{odd} \\ (\text{even})(\text{odd}) = \text{odd} & (\text{odd})(\text{odd}) = \text{even}. \end{array}$$

Thus the factor group reflects these multiplicative properties for all the permutations in S_n . \blacktriangle

Example 13.4 illustrates that while knowing the product of two cosets in G/N does not tell us what the product of two elements of G is, it may tell us that the product in G of two types of elements is itself of a certain type.

13.6 Example (The Converse of the Theorem of Lagrange is False) Recall that the Theorem of Lagrange states that the order of a subgroup of a finite group G must divide the order of G . We are now in a position to demonstrate that although the group A_4 has 12 elements and 6 divides 12, A_4 has no subgroup of order 6.

Suppose that H were a subgroup of A_4 having order 6. As observed before in Example 13.4, it would follow that H would be a normal subgroup of A_4 . Then A_4/H would have only two elements, H and σH for some $\sigma \in A_4$ not in H . Since in a group of order 2, the square of each element is the identity, we would have $HH = H$ and $(\sigma H)(\sigma H) = H$. Now computation in a factor group can be achieved by computing with representatives in the original group. Thus, computing in A_4 , we find that for each $\alpha \in H$ we must have $\alpha^2 \in H$ and for each $\beta \in \sigma H$ we must have $\beta^2 \in H$. That is, the square of every element in A_4 must be in H . But in A_4 , we have

$$(1, 2, 3) = (1, 3, 2)^2 \quad \text{and} \quad (1, 3, 2) = (1, 2, 3)^2$$

so $(1, 2, 3)$ and $(1, 3, 2)$ are in H . A similar computation shows that $(1, 2, 4), (1, 4, 2), (1, 3, 4), (1, 4, 3), (2, 3, 4)$, and $(2, 4, 3)$ are all in H . This shows that there must be at least 8 elements in H , contradicting the fact that H was supposed to have order 6. \blacktriangle

We now turn to several examples that *compute* factor groups. If the group we start with is finitely generated and abelian, then its factor group will be also. Computing such a factor group means classifying it according to the fundamental theorem (Theorem 9.12 or Theorem 9.14).

13.7 Example Let us compute the factor group $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(0, 1)\rangle$. Here $\langle(0, 1)\rangle$ is the cyclic subgroup H of $\mathbb{Z}_4 \times \mathbb{Z}_6$ generated by $(0, 1)$. Thus

$$H = \{(0, 0), (0, 1), (0, 2), (0, 3), (0, 4), (0, 5)\}.$$

Since $\mathbb{Z}_4 \times \mathbb{Z}_6$ has 24 elements and H has 6 elements, all cosets of H must have 6 elements, and $(\mathbb{Z}_4 \times \mathbb{Z}_6)/H$ must have order 4. Since $\mathbb{Z}_4 \times \mathbb{Z}_6$ is abelian, so is $(\mathbb{Z}_4 \times \mathbb{Z}_6)/H$ (remember, we compute in a factor group by means of representatives from the original group). In additive notation, the cosets are

$$H = (0, 0) + H, \quad (1, 0) + H, \quad (2, 0) + H, \quad (3, 0) + H.$$

Since we can compute by choosing the representatives $(0, 0), (1, 0), (2, 0)$, and $(3, 0)$, it is clear that $(\mathbb{Z}_4 \times \mathbb{Z}_6)/H$ is isomorphic to \mathbb{Z}_4 . Note that this is what we would expect, since in a factor group modulo H , everything in H becomes the identity element; that is, we are essentially setting everything in H equal to zero. Thus the whole second factor \mathbb{Z}_6 of $\mathbb{Z}_4 \times \mathbb{Z}_6$ is collapsed, leaving just the first factor \mathbb{Z}_4 . \blacktriangle

Example 13.7 is a special case of a general theorem that we now state and prove. We should acquire an intuitive feeling for this theorem in terms of *collapsing one of the factors to the identity element*.

13.8 Theorem Let $G = H \times K$ be the direct product of groups H and K . Then $\bar{H} = \{(h, e) \mid h \in H\}$ is a normal subgroup of G . Also G/\bar{H} is isomorphic to K in a natural way. Similarly, $G/\bar{K} \simeq H$ in a natural way.

Proof Consider the homomorphism $\pi_2 : H \times K \rightarrow K$, where $\pi_2(h, k) = k$. Because $\text{Ker}(\pi_2) = \bar{H}$, we see that \bar{H} is a normal subgroup of $H \times K$. Because π_2 is onto K , Theorem 12.14 tells us that $(H \times K)/\bar{H} \simeq K$. \blacklozenge

We continue with additional computations of abelian factor groups. To illustrate how easy it is to compute in a factor group if we can compute in the whole group, we prove the following theorem.

13.9 Theorem If G is a cyclic group and N is a subgroup of G , then G/N is cyclic.

Proof Let G be a cyclic group, so $\langle a \rangle = G$ for some $a \in G$. Let N be any subgroup of G . Since G is abelian, N is a normal subgroup of G . We compute the cyclic subgroup of G/N generated by aN .

$$\langle aN \rangle = \{(aN)^n \mid n \in \mathbb{Z}\} = \{a^n N \mid n \in \mathbb{Z}\}$$

Since $\{a^n \mid n \in \mathbb{Z}\} = G$,

$$\{a^n N \mid n \in \mathbb{Z}\} = \{gN \mid g \in G\}.$$

So $\langle aN \rangle$ contains every coset of G and we see that G/N is cyclic with generator $\langle aN \rangle$. \blacklozenge

13.10 Example Let us compute the factor group $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(0, 2)\rangle$. Now $(0, 2)$ generates the subgroup

$$H = \{(0, 0), (0, 2), (0, 4)\}$$

of $\mathbb{Z}_4 \times \mathbb{Z}_6$ of order 3. Here the first factor \mathbb{Z}_4 of $\mathbb{Z}_4 \times \mathbb{Z}_6$ is left alone. The \mathbb{Z}_6 factor, on the other hand, is essentially collapsed by a subgroup of order 3, giving a factor group in the second factor of order 2 that must be isomorphic to \mathbb{Z}_2 . Thus $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(0, 2)\rangle$ is isomorphic to $\mathbb{Z}_4 \times \mathbb{Z}_2$.

We can verify that $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(0, 2)\rangle$ is isomorphic to $\mathbb{Z}_4 \times \mathbb{Z}_2$ by using Theorem 12.14. We need a homomorphism $\phi : \mathbb{Z}_4 \times \mathbb{Z}_6 \rightarrow \mathbb{Z}_4 \times \mathbb{Z}_2$ that is onto, with kernel $\langle(0, 2)\rangle$. Defining ϕ by $\phi(a, b) = (0, r)$ where r is the remainder when b is divided by 2 does the trick. \blacktriangle

13.11 Example Let us compute the factor group $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(2, 3)\rangle$. Be careful! There is a great temptation to say that we are setting the 2 of \mathbb{Z}_4 and the 3 of \mathbb{Z}_6 both equal to zero, so that \mathbb{Z}_4 is collapsed to a factor group isomorphic to \mathbb{Z}_2 and \mathbb{Z}_6 to one isomorphic to \mathbb{Z}_3 , giving a total factor group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_3$. This is wrong! Note that

$$H = \langle(2, 3)\rangle = \{(0, 0), (2, 3)\}$$

is of order 2, so $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(2, 3)\rangle$ has order 12, not 6. Setting $(2, 3)$ equal to zero does not make $(2, 0)$ and $(0, 3)$ equal to zero individually, so the factors do not collapse separately.

The possible abelian groups of order 12 are $\mathbb{Z}_4 \times \mathbb{Z}_3$ and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$, and we must decide to which one our factor group is isomorphic. These two groups are most easily distinguished in that $\mathbb{Z}_4 \times \mathbb{Z}_3$ has an element of order 4, and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$ does not. We claim that the coset $(1, 0) + H$ is of order 4 in the factor group $(\mathbb{Z}_4 \times \mathbb{Z}_6)/H$. To find the smallest power of a coset giving the identity in a factor group modulo H , we must, by choosing representatives, find the smallest power of a representative that is in the subgroup H . Now,

$$4(1, 0) = (1, 0) + (1, 0) + (1, 0) + (1, 0) = (0, 0)$$

is the first time that $(1, 0)$ added to itself gives an element of H . Thus $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(2, 3)\rangle$ has an element of order 4 and is isomorphic to $\mathbb{Z}_4 \times \mathbb{Z}_3$ or \mathbb{Z}_{12} .

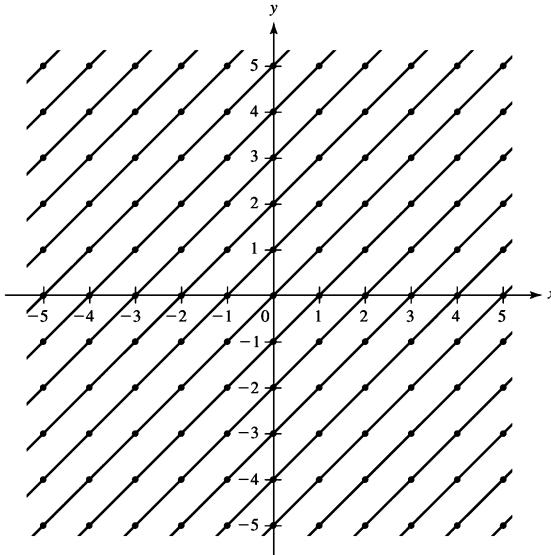
We can use Theorem 12.14 to verify that $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(2, 3)\rangle$ is isomorphic to \mathbb{Z}_{12} , although it is a little challenging to see what the homomorphism $\phi : \mathbb{Z}_4 \times \mathbb{Z}_6 \rightarrow \mathbb{Z}_{12}$ should be. We define $\phi : \mathbb{Z}_4 \times \mathbb{Z}_6 \rightarrow \mathbb{Z}_{12}$ by setting $\phi(a, b) = 3a +_{12} (12 - 2b)$. Here we interpret $3a$ and $2b$ as integer multiplication, so $0 \leq 3a < 12$ and $0 \leq 2b < 12$. The map ϕ is a homomorphism, but this takes some checking, which we leave to the reader. Also, $\text{Ker}(\phi) = \{(a, b) \in \mathbb{Z}_4 \times \mathbb{Z}_6 \mid 3a = 2b\} = \{(0, 0), (2, 3)\} = \langle(2, 3)\rangle$. We also see that $\phi(1, 1) = 1$, which implies that ϕ maps onto \mathbb{Z}_{12} . By the Fundamental Homomorphism Theorem, $(\mathbb{Z}_4 \times \mathbb{Z}_6)/\langle(2, 3)\rangle$ is isomorphic to \mathbb{Z}_{12} . \blacktriangle

13.12 Example Let us compute (that is, classify as in Theorem 9.12) the group $(\mathbb{Z} \times \mathbb{Z})/\langle(1, 1)\rangle$. We may visualize $\mathbb{Z} \times \mathbb{Z}$ as the points in the plane with both coordinates integers, as indicated by the dots in Fig. 13.13. The subgroup $\langle(1, 1)\rangle$ consists of those points that lie on the 45° line through the origin, indicated in the figure. The coset $(1, 0) + \langle(1, 1)\rangle$ consists of those dots on the 45° line through the point $(1, 0)$, also shown in the figure. Continuing, we see that each coset consists of those dots lying on one of the 45° lines in the figure. We may choose the representatives

$$\dots, (-3, 0), (-2, 0), (-1, 0), (0, 0), (1, 0), (2, 0), (3, 0), \dots$$

of these cosets to compute in the factor group. Since these representatives correspond precisely to the points of \mathbb{Z} on the x -axis, we see that the factor group $(\mathbb{Z} \times \mathbb{Z})/\langle(1, 1)\rangle$ is isomorphic to \mathbb{Z} .

Again, we can use the Fundamental Homomorphism Theorem as another method of computing this group. We let $\phi : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ be defined by $\phi(n, m) = n - m$. It is easy to verify that ϕ is a homomorphism, ϕ maps onto \mathbb{Z} , and $\text{Ker}(\phi) = \{(n, m) \in \mathbb{Z} \times \mathbb{Z} \mid n = m\} = \langle(1, 1)\rangle$. So by the Fundamental Homomorphism Theorem, $(\mathbb{Z} \times \mathbb{Z})/\langle(1, 1)\rangle$ is isomorphic to \mathbb{Z} . Furthermore, an isomorphism is given by $\mu((n, m) + \langle(1, 1)\rangle) = n - m$. This is the same isomorphism that we saw above. \blacktriangle



13.13 Figure

13.14 Example We now compute $(\mathbb{Z} \times \mathbb{Z})/\langle(2, 4)\rangle$. This is similar to Example 13.12, but there is a little twist to this one. In this example, we know that the factor group has an element with order 2, since $(1, 2) \notin \langle(2, 4)\rangle$, but $(1, 2) + (1, 2) \in \langle(2, 4)\rangle$. Furthermore, $(\mathbb{Z} \times \mathbb{Z})/\langle(2, 4)\rangle$ has an element $(1, 0) + \langle(2, 4)\rangle$ with infinite order since $(n, 0) \notin \langle(2, 4)\rangle$ for any $n \in \mathbb{Z}^+$. Figure 13.15 illustrates the situation. Along the line $y = 2x$ only every other lattice point is in $\langle(2, 4)\rangle$. These points are filled dots in the figure. Each line with slope two contains

two cosets, one indicated with solid dots and one with hollow dots. Adding $(1, 2)$ moves the solid dot cosets to the hollow dot cosets and the hollow dot cosets to the solid dot cosets while staying on the same line. Adding $(0, 1)$ moves a coset from one line to the next. We may choose coset representatives

$$\dots, (0, -3), (0, -2), (0, -1), (0, 0), (0, 1), (0, 2), (0, 3), \dots$$

for the solid dot cosets and

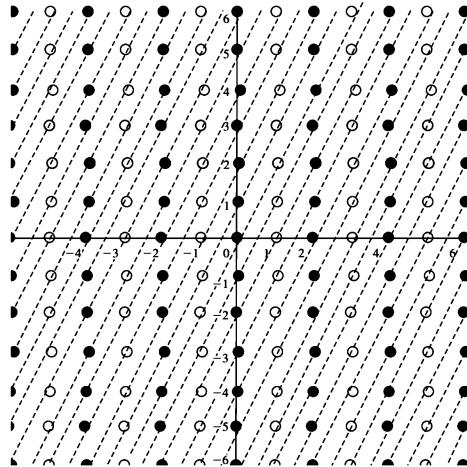
$$\dots, (1, -3), (1, -2), (1, -1), (1, 0), (1, 1), (1, 2), (1, 3), \dots$$

for the hollow dot cosets. So it seems that we have two copies of the integers, one with a zero in the first coordinate and one with a one in the first coordinate. This leads us to guess that $(\mathbb{Z} \times \mathbb{Z})/\langle(2, 4)\rangle$ is isomorphic with $\mathbb{Z}_2 \times \mathbb{Z}$.

To verify that our guess is correct, we seek a homomorphism $\phi : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}_2 \times \mathbb{Z}$ that maps onto $\mathbb{Z}_2 \times \mathbb{Z}$ and whose kernel is $\langle(2, 4)\rangle$. We let $\phi(a, b) = (r, 2a - b)$ where r is the remainder when a is divided by 2. It is easy to check that ϕ is a homomorphism. Furthermore, $\phi(0, -1) = (0, 1)$ and $\phi(1, 2) = (1, 0)$, which implies that ϕ maps onto $\mathbb{Z} \times \mathbb{Z}_2$. It remains to compute $\text{Ker}(\phi)$.

$$\text{Ker}(\phi) = \{(a, b) \mid b = 2a \text{ and } a \text{ is even}\} = \{(2n, 4n) \mid n \in \mathbb{Z}\} = \langle(2, 4)\rangle.$$

Thus $(\mathbb{Z} \times \mathbb{Z})/\langle(2, 4)\rangle$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}_2$ by the Fundamental Homomorphism Theorem. Furthermore, an isomorphism $\mu : (\mathbb{Z} \times \mathbb{Z})/\langle(2, 4)\rangle \rightarrow \mathbb{Z}_2 \times \mathbb{Z}$ is defined by the formula $\mu((a, b) + \langle(2, 4)\rangle) = (r, 2a - b)$ where r is the remainder when a is divided by 2. \blacktriangle



13.15 Figure

Simple Groups

As we mentioned in the preceding section, one feature of a factor group is that it gives crude information about the structure of the whole group. Of course, sometimes there may be no nontrivial proper normal subgroup. For example, Lagrange's Theorem shows that a group of prime order can have no nontrivial proper subgroup of any sort.

13.16 Definition A group is **simple** if it is nontrivial and has no proper nontrivial normal subgroup. ■

13.17 Theorem The alternating group A_n is simple for $n \geq 5$.

Proof See Exercise 41. ◆

There are many simple groups other than those given above. For example, A_5 is of order 60 and A_6 is of order 360, and there is a simple group of nonprime order, namely 168, between these orders.

The complete determination and classification of all finite simple groups is one of the mathematical triumphs of the twentieth century. Hundreds of mathematicians worked on this task from 1950 to 1980. It can be shown that a finite group has a sort of factorization into simple groups, where the factors are unique up to order. The situation is similar to the factorization of positive integers into primes. The knowledge of all finite simple groups can be used to solve some problems of finite group theory and combinatorics.

We have seen in this text that a finite simple abelian group is isomorphic to \mathbb{Z}_p for some prime p . In 1963, Thompson and Feit [21] published their proof of a long-standing conjecture of Burnside, showing that every finite nonabelian simple group is of even order. Further great strides toward the complete classification were made by Aschbacher in the 1970s. Early in 1980, Griess announced that he had constructed a predicted “monster” simple group of order

$$\begin{aligned} 808,017,424,794,512,875,886,459,904,961,710,757,005,754,368, \\ 000,000,000. \end{aligned}$$

Aschbacher added the final details of the classification in August 1980. The research papers contributing to the entire classification fill roughly 5000 journal pages.

We turn to the characterization of those normal subgroups N of a group G for which G/N is a simple group. First we state an addendum to Theorem 8.5 on properties of a group homomorphism. The proof is left to Exercises 37 and 38.

13.18 Theorem Let $\phi : G \rightarrow G'$ be a group homomorphism. If N is a normal subgroup of G , then $\phi[N]$ is a normal subgroup of $\phi[G]$. Also, if N' is a normal subgroup of $\phi[G]$, then $\phi^{-1}[N']$ is a normal subgroup of G . ◆

Theorem 13.18 should be viewed as saying that a homomorphism $\phi : G \rightarrow G'$ preserves normal subgroups between G and $\phi[G]$. It is important to note that $\phi[N]$ may not be normal in G' , even though N is normal in G . For example, $\phi : \mathbb{Z}_2 \rightarrow S_3$, where $\phi(0) = \iota$ and $\phi(1) = (1, 2)$ is a homomorphism, and \mathbb{Z}_2 is a normal subgroup of itself, but $\{\iota, (1, 2)\}$ is not a normal subgroup of S_3 .

We can now characterize when G/N is a simple group.

13.19 Definition A **maximal normal subgroup of a group G** is a normal subgroup M not equal to G such that there is no proper normal subgroup N of G properly containing M . ■

13.20 Theorem M is a maximal normal subgroup of G if and only if G/M is simple.

Proof Let M be a maximal normal subgroup of G . Consider the canonical homomorphism $\gamma : G \rightarrow G/M$ given by Theorem 12.12. Now γ^{-1} of any nontrivial proper normal subgroup of G/M is a proper normal subgroup of G properly containing M . But M is maximal, so this cannot happen. Thus G/M is simple.

Conversely, Theorem 13.18 shows that if N is a normal subgroup of G properly containing M , then $\gamma[N]$ is normal in G/M . If also $N \neq G$, then

$$\gamma[N] \neq G/M \quad \text{and} \quad \gamma[N] \neq \{M\}.$$

Thus, if G/M is simple so that no such $\gamma[N]$ can exist, no such N can exist, and M is maximal. \blacklozenge

The Center and Commutator Subgroups

Every nonabelian group G has two important normal subgroups, the *center* $Z(G)$ of G and the *commutator subgroup* C of G . (The letter Z comes from the German word *zentrum*, meaning center.) The center $Z(G)$ is defined by

$$Z(G) = \{z \in G \mid zg = gz \text{ for all } g \in G\}.$$

Exercise 59 of Section 5 shows that $Z(G)$ is an abelian subgroup of G . Since for each $g \in G$ and $z \in Z(G)$ we have $gzg^{-1} = zgg^{-1} = ze = z$, we see at once that $Z(G)$ is a normal subgroup of G . If G is abelian, then $Z(G) = G$; in this case, the center is not useful.

13.21 Example The center of a group G always contains the identity element e . It may be that $Z(G) = \{e\}$, in which case we say that the **center of G is trivial**. For example, examination of Table 4.15 for the group S_3 shows us that $Z(S_3) = \{e\}$, so the center of S_3 is trivial. (This is a special case of Exercise 40, which shows that the center of every nonabelian group of order pq for primes p and q is trivial.) Consequently, the center of $S_3 \times \mathbb{Z}_5$ must be $\{e\} \times \mathbb{Z}_5$, which is isomorphic to \mathbb{Z}_5 . \blacktriangle

Turning to the commutator subgroup, recall that in forming a factor group of G modulo a normal subgroup N , we are essentially putting every element in G that is in N equal to e , for N forms our new identity in the factor group. This indicates another use for factor groups. Suppose, for example, that we are studying the structure of a non-abelian group G . Since Theorem 9.12 gives complete information about the structure of all finitely generated abelian groups, it might be of interest to try to form an abelian group as much like G as possible, an *abelianized version* of G , by starting with G and then requiring that $ab = ba$ for all a and b in our new group structure. To require that $ab = ba$ is to say that $aba^{-1}b^{-1} = e$ in our new group. An element $aba^{-1}b^{-1}$ in a group is a **commutator of the group**. Thus we wish to attempt to form an abelianized version of G by replacing every commutator of G by e . By the first observation of this paragraph, we should then attempt to form the factor group of G modulo the smallest normal subgroup we can find that contains all commutators of G .

13.22 Theorem Let G be a group. The set of all commutators $aba^{-1}b^{-1}$ for $a, b \in G$ generates a subgroup C (the **commutator subgroup**) of G . This subgroup C is a normal subgroup of G . Furthermore, if N is a normal subgroup of G , then G/N is abelian if and only if $C \leq N$.

Proof The commutators certainly generate a subgroup C ; we must show that it is normal in G . Note that the inverse $(aba^{-1}b^{-1})^{-1}$ of a commutator is again a commutator, namely, $bab^{-1}a^{-1}$. Also $e = eee^{-1}e^{-1}$ is a commutator. Theorem 7.7 then shows that C consists precisely of all finite products of commutators. For $x \in C$, we must show that $g^{-1}xg \in C$ for all $g \in G$, or that if x is a product of commutators, so is $g^{-1}xg$ for all $g \in G$. By inserting $e = gg^{-1}$ between each product of commutators occurring in x , we see that it is sufficient to show for each commutator $cdc^{-1}d^{-1}$ that $g^{-1}(cdc^{-1}d^{-1})g$ is in C . But

$$\begin{aligned} g^{-1}(cdc^{-1}d^{-1})g &= (g^{-1}cdc^{-1})(e)(d^{-1}g) \\ &= (g^{-1}cdc^{-1})(gd^{-1}dg^{-1})(d^{-1}g) \\ &= [(g^{-1}c)d(g^{-1}c)^{-1}d^{-1}][dg^{-1}d^{-1}g], \end{aligned}$$

which is in C . Thus C is normal in G .

The rest of the theorem is obvious if we have acquired the proper feeling for factor groups. One doesn't visualize in this way, but writing out that G/C is abelian follows from

$$\begin{aligned}(aC)(bC) &= abC = ab(b^{-1}a^{-1}ba)C \\ &= (abb^{-1}a^{-1})baC = baC = (bC)(aC).\end{aligned}$$

Furthermore, if N is a normal subgroup of G and G/N is abelian, then $(a^{-1}N)(b^{-1}N) = (b^{-1}N)(a^{-1}N)$; that is, $aba^{-1}b^{-1}N = N$, so $aba^{-1}b^{-1} \in N$, and $C \leq N$. Finally, if $C \leq N$, then

$$\begin{aligned}(aN)(bN) &= abN = ab(b^{-1}a^{-1}ba)N \\ &= (abb^{-1}a^{-1})baN = baN = (bN)(aN).\end{aligned}\quad \blacklozenge$$

13.23 Example Using cycle notation in the symmetric group S_3 , one commutator is

$$(1, 2, 3)(2, 3)(1, 2, 3)^{-1}(2, 3)^{-1} = (1, 2, 3)(2, 3)(1, 3, 2)(2, 3) = (1, 3, 2).$$

So the commutator subgroup C contains $\langle(1, 3, 2)\rangle = A_3$, the alternating group. Since S_3/A_3 is abelian (isomorphic with \mathbb{Z}_2), Theorem 13.22 says that $C \leq A_3$. Therefore, A_3 is the commutator subgroup. \blacktriangle

■ EXERCISES 13

Computations

In Exercises 1 through 14, classify the given group according to the fundamental theorem of finitely generated abelian groups.

1. $(\mathbb{Z}_2 \times \mathbb{Z}_4)/\langle(0, 1)\rangle$
2. $(\mathbb{Z}_2 \times \mathbb{Z}_4)/\langle(0, 2)\rangle$
3. $(\mathbb{Z}_2 \times \mathbb{Z}_4)/\langle(1, 2)\rangle$
4. $(\mathbb{Z}_4 \times \mathbb{Z}_8)/\langle(1, 2)\rangle$
5. $(\mathbb{Z}_4 \times \mathbb{Z}_4 \times \mathbb{Z}_8)/\langle(1, 2, 4)\rangle$
6. $(\mathbb{Z} \times \mathbb{Z})/\langle(0, 1)\rangle$
7. $(\mathbb{Z} \times \mathbb{Z})/\langle(0, 2)\rangle$
8. $(\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z})/\langle(1, 1, 1)\rangle$
9. $(\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}_4)/\langle(3, 0, 0)\rangle$
10. $(\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}_8)/\langle(0, 4, 0)\rangle$
11. $(\mathbb{Z} \times \mathbb{Z})/\langle(2, 2)\rangle$
12. $(\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z})/\langle(3, 3, 3)\rangle$
13. $(\mathbb{Z} \times \mathbb{Z})/\langle(2, 6)\rangle$
14. $(\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}_2)/\langle(1, 1, 1)\rangle$
15. Find both the center and the commutator subgroup of D_4 .
16. Find both the center and the commutator subgroup of $\mathbb{Z}_3 \times S_3$.
17. Find both the center and the commutator subgroup of $S_3 \times D_4$.
18. Describe all subgroups of order ≤ 4 of $\mathbb{Z}_4 \times \mathbb{Z}_4$, and in each case classify the factor group of $\mathbb{Z}_4 \times \mathbb{Z}_4$ modulo the subgroup by Theorem 9.12. That is, describe the subgroup and say that the factor group of $\mathbb{Z}_4 \times \mathbb{Z}_4$ modulo the subgroup is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_4$, or whatever the case may be. [Hint: $\mathbb{Z}_4 \times \mathbb{Z}_4$ has six different cyclic subgroups of order 4. Describe them by giving a generator, such as the subgroup $\langle(1, 0)\rangle$. There is one subgroup of order 4 that is isomorphic to the Klein 4-group. There are three subgroups of order 2.]

Concepts

In Exercises 19 and 20, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

19. The *center* of a group G contains all elements of G that commute with every element of G .
20. The *commutator subgroup* of a group G is $\{a^{-1}b^{-1}ab \mid a, b \in G\}$.

21. Determine whether each of following is true or false.

- a. Every factor group of a cyclic group is cyclic.
- b. A factor group of a noncyclic group is again noncyclic.
- c. \mathbb{R}/\mathbb{Z} under addition has no element of order 3.
- d. \mathbb{R}/\mathbb{Q} under addition has no element of order 2.
- e. \mathbb{R}/\mathbb{Z} under addition has an infinite number of elements of order 4.
- f. If the commutator subgroup C of a group G is $\{e\}$, then G is abelian.
- g. If G/H is abelian, then the commutator subgroup C of G contains H .
- h. The commutator subgroup of a simple group G must be G itself.
- i. The commutator subgroup of a nonabelian simple group G must be G itself.
- j. All nontrivial finite simple groups have prime order.

In Exercises 22 through 25, let F be the additive group of all functions mapping \mathbb{R} into \mathbb{R} , and let F^* be the multiplicative group of all elements of F that do not assume the value 0 at any point of \mathbb{R} .

22. Let K be the subgroup of F consisting of the constant functions. Find a subgroup of F to which F/K is isomorphic.
23. Let K^* be the subgroup of F^* consisting of the nonzero constant functions. Find a subgroup of F^* to which F^*/K^* is isomorphic.
24. Let K be the subgroup of continuous functions in F . Can you find an element of F/K having order 2? Why or why not?
25. Let K^* be the subgroup of F^* consisting of the continuous functions in F^* . Can you find an element of F^*/K^* having order 2? Why or why not?

In Exercises 26 through 28, let U be the multiplicative group $\{z \in \mathbb{C} \mid |z| = 1\}$.

26. Let $z_0 \in U$. Show that $z_0U = \{z_0z \mid z \in U\}$ is a subgroup of U , and compute U/z_0U .
27. To what group we have mentioned in the text is $U/\langle -1 \rangle$ isomorphic?
28. Let $\zeta_n = \cos(2\pi/n) + i \sin(2\pi/n)$ where $n \in \mathbb{Z}^+$. To what group we have mentioned is $U/\langle \zeta_n \rangle$ isomorphic?
29. To what group mentioned in the text is the additive group \mathbb{R}/\mathbb{Z} isomorphic?
30. Give an example of a group G having no elements of finite order greater than 1 and a normal subgroup $H \trianglelefteq G$, $H \neq G$, so that in G/H every element has finite order.
31. Let H and K be normal subgroups of a group G . Give an example showing that we may have $H \cong K$ while G/H is not isomorphic to G/K .
32. Describe the center of every simple
 - a. abelian group
 - b. nonabelian group.
33. Describe the commutator subgroup of every simple
 - a. abelian group
 - b. nonabelian group.

Proof Synopsis

34. Give a one-sentence synopsis of the proof of Theorem 13.9.
35. Give at most a two-sentence synopsis of the proof of Theorem 13.20.

Theory

36. Show that if a finite group G contains a nontrivial subgroup of index 2 in G , then G is not simple.
37. Let $\phi : G \rightarrow G'$ be a group homomorphism, and let N be a normal subgroup of G . Show that $\phi[N]$ is a normal subgroup of $\phi[G]$.

38. Let $\phi : G \rightarrow G'$ be a group homomorphism, and let N' be a normal subgroup of G' . Show that $\phi^{-1}[N']$ is a normal subgroup of G .
39. Show that if G is nonabelian, then the factor group $G/Z(G)$ is not cyclic. [Hint: Show the equivalent contrapositive, namely, that if $G/Z(G)$ is cyclic then G is abelian (and hence $Z(G) = G$).]
40. Using Exercise 39, show that a nonabelian group G of order pq where p and q are primes has a trivial center.
41. Prove that A_n is simple for $n \geq 5$, following the steps and hints given.
- Show A_n contains every 3-cycle if $n \geq 3$.
 - Show A_n is generated by the 3-cycles for $n \geq 3$. [Hint: Note that $(a, b)(c, d) = (a, c, b)(a, c, d)$ and $(a, c)(a, b) = (a, b, c)$.]
 - Let r and s be fixed elements of $\{1, 2, \dots, n\}$ for $n \geq 3$. Show that A_n is generated by the n “special” 3-cycles of the form (r, s, i) for $1 \leq i \leq n$ [Hint: Show every 3-cycle is the product of “special” 3-cycles by computing

$$(r, s, i)^2, \quad (r, s, j)(r, s, i)^2, \quad (r, s, j)^2(r, s, i),$$

and

$$(r, s, i)^2(r, s, k)(r, s, j)^2(r, s, i).$$

Observe that these products give all possible types of 3-cycles.]

- d. Let N be a normal subgroup of A_n for $n \geq 3$. Show that if N contains a 3-cycle, then $N = A_n$. [Hint: Show that $(r, s, i) \in N$ implies that $(r, s, j) \in N$ for $j = 1, 2, \dots, n$ by computing

$$((r, s)(i, j))(r, s, i)^2((r, s)(i, j))^{-1}.$$

- e. Let N be a nontrivial normal subgroup of A_n for $n \geq 5$. Show that one of the following cases must hold, and conclude in each case that $N = A_n$.

Case I N contains a 3-cycle.

Case II N contains a product of disjoint cycles, at least one of which has length greater than 3. [Hint: Suppose N contains the disjoint product $\sigma = \mu(a_1, a_2, \dots, a_r)$. Show $\sigma^{-1}(a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1}$ is in N , and compute it.]

Case III N contains a disjoint product of the form $\sigma = \mu(a_4, a_5, a_6)(a_1, a_2, a_3)$. [Hint: Show $\sigma^{-1}(a_1, a_2, a_4)\sigma(a_1, a_2, a_4)^{-1}$ is in N , and compute it.]

Case IV N contains a disjoint product of the form $\sigma = \mu(a_1, a_2, a_3)$ where μ is a product of disjoint 2-cycles. [Hint: Show $\sigma^2 \in N$ and compute it.]

Case V N contains a disjoint product σ of the form $\sigma = \mu(a_3, a_4)(a_1, a_2)$, where μ is a product of an even number of disjoint 2-cycles. [Hint: Show that $\sigma^{-1}(a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1}$ is in N , and compute it to deduce that $\alpha = (a_2, a_4)(a_1, a_3)$ is in N . Using $n \geq 5$ for the first time, find $i \neq a_1, a_2, a_3, a_4$ in $\{1, 2, \dots, n\}$. Let $\beta = (a_1, a_3, i)$. Show that $\beta^{-1}\alpha\beta\alpha \in N$, and compute it.]

42. Let N be a normal subgroup of G and let H be any subgroup of G . Let $HN = \{hn \mid h \in H, n \in N\}$. Show that HN is a subgroup of G , and is the smallest subgroup containing both N and H .
43. With reference to the preceding exercise, let M also be a normal subgroup of G . Show that NM is again a normal subgroup of G .
44. Show that if H and K are normal subgroups of a group G such that $H \cap K = \{e\}$, then $hk = kh$ for all $h \in H$ and $k \in K$. [Hint: Consider the commutator $hkh^{-1}k^{-1} = (hkh^{-1})k^{-1} = h(kh^{-1}k^{-1})$.]
45. With reference to the three preceding exercises, let H and K be normal subgroups of a group G such that $HK = G$ and $H \cap K = \{e\}$. Prove that G is isomorphic with $H \times G$.

SECTION 14

† GROUP ACTION ON A SET

We have seen examples of how groups may *act on things*, like the group of symmetries of a triangle or of a square, the group of rotations of a cube, the general linear group acting on \mathbb{R}^n , and so on. In this section we give the general notion of group action and apply it to learn more about finite groups. The next section will give applications to counting.

The Notion of a Group Action

Definition 1.1 defines a binary operation $*$ on a set S to be a function mapping $S \times S$ into S . The function $*$ gives us a rule for “multiplying” an element s_1 in S and an element s_2 in S to yield an element $s_1 * s_2$ in S .

More generally, for any sets A , B , and C , we can view a map $* : A \times B \rightarrow C$ as defining a “multiplication,” where any element a of A times any element b of B has as value some element c of C . Of course, we write $a * b = c$, or simply $ab = c$. In this section, we will be concerned with the case where X is a set, G is a group, and we have a map $* : G \times X \rightarrow X$. We shall write $*(g, x)$ as $g * x$ or gx .

14.1 Example

Let $G = \text{GL}(n, \mathbb{R})$ and X the set of all column vectors in \mathbb{R}^n . Then for any matrix $A \in G$ and vector $v \in X$, Av is a vector in X . So multiplying is an operation $* : G \times X \rightarrow X$. From linear algebra, we know that if B is also a matrix in G , then $(AB)v = A(Bv)$. Furthermore, for the identity matrix I , $Iv = v$. \blacktriangle

14.2 Example

Let G be the dihedral group D_n . Then elements of D_n permute the set $\mathbb{Z}_n = \{0, 1, 2, 3, \dots, n-1\}$. For example, $\rho(k) = k +_n 1$. Thus we have an operation $* : D_n \times \mathbb{Z}_n \rightarrow \mathbb{Z}_n$. Furthermore, if $\alpha, \gamma \in D_n$ and $k \in \mathbb{Z}_n$, then $(\alpha\gamma)(k) = \alpha(\gamma(k))$ and $i(k) = k$. \blacktriangle

The two previous examples share the same properties, which we formalize in Definition 14.3.

14.3 Definition

Let X be a set and G a group. An **action of G on X** is a map $* : G \times X \rightarrow X$ such that

1. $ex = x$ for all $x \in X$,
2. $(g_1g_2)x = g_1(g_2x)$ for all $x \in X$ and all $g_1, g_2 \in G$.

Under these conditions, X is a *G-set*. \blacksquare

14.4 Example

Let X be any set, and let H be a subgroup of the group S_X of all permutations of X . Then X is an *H-set*, where the action of $\sigma \in H$ on X is its action as an element of S_X , so that $\sigma x = \sigma(x)$ for all $x \in X$. Condition 2 is a consequence of the definition of permutation multiplication as function composition, and Condition 1 is immediate from the definition of the identity permutation as the identity function. Note that, in particular, $\{1, 2, 3, \dots, n\}$ is an S_n -set. \blacktriangle

Our next theorem will show that for every *G-set* X and each $g \in G$, the map $\sigma_g : X \rightarrow X$ defined by $\sigma_g(x) = gx$ is a permutation of X , and that there is a homomorphism $\phi : G \rightarrow S_X$ such that the action of G on X is essentially the Example 14.4 action of the image subgroup $H = \phi[G]$ of S_X on X . So actions of subgroups of S_X on X describe all possible group actions on X . When studying the set X , actions using subgroups of S_X suffice. However, sometimes a set X is used to study G via a group action of G on X . Thus we need the more general concept given by Definition 14.3.

[†] This section is a prerequisite only for Sections 15 and 17.

14.5 Theorem Let X be a G -set. For each $g \in G$, the function $\sigma_g : X \rightarrow X$ defined by $\sigma_g(x) = gx$ for $x \in X$ is a permutation of X . Also, the map $\phi : G \rightarrow S_X$ defined by $\phi(g) = \sigma_g$ is a homomorphism with the property that $\phi(g)(x) = gx$.

Proof To show that σ_g is a permutation of X , we must show that σ_g is a one-to-one map of X onto itself. Suppose that $\sigma_g(x_1) = \sigma_g(x_2)$ for $x_1, x_2 \in X$. Then $gx_1 = gx_2$. Consequently, $g^{-1}(gx_1) = g^{-1}(gx_2)$. Using Condition 2 in Definition 14.3, we see that $(g^{-1}g)x_1 = (g^{-1}g)x_2$, so $x_1 = x_2$. Condition 1 of the definition then yields $x_1 = x_2$, so σ_g is one-to-one. The two conditions of the definition show that for $x \in X$, we have $\sigma_g(g^{-1}x) = g(g^{-1}x) = (gg^{-1})x = ex = x$, so σ_g maps X onto X . Thus σ_g is indeed a permutation.

To show that $\phi : G \rightarrow S_X$ defined by $\phi(g) = \sigma_g$ is a homomorphism, we must show that $\phi(g_1g_2) = \phi(g_1)\phi(g_2)$ for all $g_1, g_2 \in G$. We show the equality of these two permutations in S_X by showing they both carry an $x \in X$ into the same element. Using the two conditions in Definition 14.3 and the rule for function composition, we obtain

$$\begin{aligned}\phi(g_1g_2)(x) &= \sigma_{g_1g_2}(x) = (g_1g_2)x = g_1(g_2x) = g_1\sigma_{g_2}(x) = \sigma_{g_1}(\sigma_{g_2}(x)) \\ &= (\sigma_{g_1} \circ \sigma_{g_2})(x) = (\sigma_{g_1}, \sigma_{g_2})(x) = (\phi(g_1)\phi(g_2))(x).\end{aligned}$$

Thus ϕ is a homomorphism. The stated property of ϕ follows at once since by our definitions, we have $\phi(g)(x) = \sigma_g(x) = gx$. \blacklozenge

It follows from the preceding theorem and Theorem 12.17 that if X is a G -set, then the subset of G leaving every element of X fixed is a normal subgroup N of G , and we can regard X as a G/N -set where the action of a coset gN on X is given by $(gN)x = gx$ for each $x \in X$. If $N = \{e\}$, then the identity element of G is the only element that leaves every $x \in X$ fixed; we then say that G acts faithfully on X . A group G is transitive on a G -set X if for each $x_1, x_2 \in X$, there exists $g \in G$ such that $gx_1 = x_2$.

We continue with more examples of G -sets.

14.6 Example Every group G is itself a G -set, where the action on $g_2 \in G$ by $g_1 \in G$ is given by left multiplication. That is, $*(g_1, g_2) = g_1g_2$. If H is a subgroup of G , we can also regard G as an H -set, where $*(h, g) = hg$. \blacktriangle

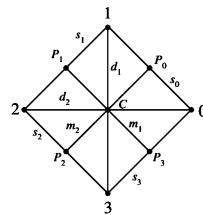
14.7 Example Let H be a subgroup of G . Then G is an H -set under conjugation where $*(h, g) = hgh^{-1}$ for $g \in G$ and $h \in H$. Condition 1 is obvious, and for Condition 2 note that

$$*(h_1h_2, g) = (h_1h_2)g(h_1h_2)^{-1} = h_1(h_2gh_2^{-1})h_1^{-1} = *(h_1, *(h_2, g)).$$

We always write this action of H on G by conjugation as hgh^{-1} . The abbreviation hg described before the definition would cause terrible confusion with the group operation of G . \blacktriangle

14.8 Example Let H be a subgroup of G , and let L_H be the set of all left cosets of H . Then L_H is a G -set, where the action of $g \in G$ on the left coset xH is given by $g(xH) = (gx)H$. Observe that this action is well defined: if $yH = xH$, then $y = xh$ for some $h \in H$, and $g(yH) = (gy)H = (gxh)H = (gx)(hH) = (gx)H = g(xH)$. A series of exercises shows that every G -set is isomorphic to one that may be formed using these left coset G -sets as building blocks. (See Exercises 22 through 25.) \blacktriangle

14.9 Example Let us look closer at the dihedral group D_4 , which permutes the vertices of the square as labeled in Figure 14.10. As indicated in the figure, we label the vertices 0, 1, 2, 3 as usual; the sides s_0, s_1, s_2, s_3 ; the midpoints of the sides P_0, P_1, P_2, P_3 ; the diagonals d_1, d_2 ; the lines joining opposite side midpoints m_1, m_2 ; and we label the intersection of the lines d_1, d_2, m_1, m_2 with C .



14.10 Figure

We can think of the set

$$X = \{0, 1, 2, 3, s_0, s_1, s_2, s_3, m_1, m_2, d_1, d_2, C, P_0, P_1, P_2, P_3\}$$

as a D_4 -set in a natural way. Table 14.11 shows the action of D_4 on X . Recall that ι is the identity, ρ^k is rotation by $k\pi/2$, and μ is reflection across the line d_2 . We can see from the table that $\mu\rho$ is reflection across the line m_1 , $\mu\rho^2$ is reflection across the line d_1 , and $\mu\rho^3$ is reflection across the line m_2 . It is worthwhile to spend a little time to understand how Table 14.11 was constructed before continuing. ▲

14.11 Table

	0	1	2	3	s_0	s_1	s_2	s_3	m_1	m_2	d_1	d_2	C	P_0	P_1	P_2	P_3
ι	0	1	2	3	s_0	s_1	s_2	s_3	m_1	m_2	d_1	d_2	C	P_0	P_1	P_2	P_3
ρ	1	2	3	0	s_1	s_2	s_3	s_0	m_2	m_1	d_2	d_1	C	P_1	P_2	P_3	P_0
ρ^2	2	3	0	1	s_2	s_3	s_0	s_1	m_1	m_2	d_1	d_2	C	P_2	P_3	P_0	P_1
ρ^3	3	0	1	2	s_3	s_0	s_1	s_2	m_2	m_1	d_2	d_1	C	P_3	P_0	P_1	P_2
μ	0	3	2	1	s_3	s_2	s_1	s_0	m_2	m_1	d_1	d_2	C	P_3	P_2	P_1	P_0
$\mu\rho$	3	2	1	0	s_2	s_1	s_0	s_3	m_1	m_2	d_2	d_1	C	P_2	P_1	P_0	P_3
$\mu\rho^2$	2	1	0	3	s_1	s_0	s_3	s_2	m_2	m_1	d_1	d_2	C	P_1	P_0	P_3	P_2
$\mu\rho^3$	1	0	3	2	s_0	s_3	s_2	s_1	m_1	m_2	d_2	d_1	C	P_0	P_3	P_2	P_1

Isotropy Subgroups

Let X be a G -set. Let $x \in X$ and $g \in G$. It will be important to know when $gx = x$. We let

$$X_g = \{x \in X \mid gx = x\} \quad \text{and} \quad G_x = \{g \in G \mid gx = x\}.$$

14.12 Example For the D_4 -set X in Example 14.9, we have

$$X_\iota = X, \quad X_\rho = \{C\}, \quad X_\mu = \{0, 2, d_1, d_2, C\}.$$

Also, using the same D_4 action on X ,

$$G_0 = \{\iota, \mu\}, \quad G_{s_2} = \{\iota, \mu\rho^3\}, \quad G_{d_1} = \{\iota, \rho^2, \mu, \mu\rho^2\}.$$

We leave the computations of the other sets of the form X_σ and G_x to Exercises 1 and 2. ▲

Note that the subsets G_x given in the preceding example were, in each case, subgroups of G . This is true in general.

14.13 Theorem Let X be a G -set. Then G_x is a subgroup of G for each $x \in X$.

Proof Let $x \in X$ and let $g_1, g_2 \in G_x$. Then $g_1x = x$ and $g_2x = x$. Consequently, $(g_1g_2)x = g_1(g_2x) = g_1x = x$, so $g_1g_2 \in G_x$, and G_x is closed under the induced operation of G . Of course, $ex = x$, so $e \in G_x$. If $g \in G_x$, then $gx = x$, so $x = ex = (g^{-1}g)x = g^{-1}(gx) = g^{-1}x$, and consequently $g^{-1} \in G_x$. Thus G_x is a subgroup of G . \blacklozenge

14.14 Definition Let X be a G -set and let $x \in X$. The subgroup G_x is the **isotropy subgroup** of x . \blacksquare

Orbits

For the D_4 -set X of Example 14.9 with action table in Table 14.11, the elements in the subset $\{0, 1, 2, 3\}$ are carried into elements of this same subset under action by D_4 . Furthermore, each of the elements 0, 1, 2, and 3 is carried into all the other elements of the subset by the various elements of D_4 . We proceed to show that every G -set X can be partitioned into subsets of this type.

14.15 Theorem Let X be a G -set. For $x_1, x_2 \in X$, let $x_1 \sim x_2$ if and only if there exists $g \in G$ such that $gx_1 = x_2$. Then \sim is an equivalence relation on X .

Proof For each $x \in X$, we have $ex = x$, so $x \sim x$ and \sim is reflexive.

Suppose $x_1 \sim x_2$, so $gx_1 = x_2$ for some $g \in G$. Then $g^{-1}x_2 = g^{-1}(gx_1) = (g^{-1}g)x_1 = ex_1 = x_1$, so $x_2 \sim x_1$, and \sim is symmetric.

Finally, if $x_1 \sim x_2$ and $x_2 \sim x_3$, then $g_1x_1 = x_2$ and $g_2x_2 = x_3$ for some $g_1, g_2 \in G$. Then $(g_2g_1)x_1 = g_2(g_1x_1) = g_2x_2 = x_3$, so $x_1 \sim x_3$ and \sim is transitive. \blacklozenge

14.16 Definition Let X be a G -set. Each cell in the partition of the equivalence relation described in Theorem 14.15 is an **orbit in X under G** . If $x \in X$, the cell containing x is the **orbit of x** . We let this cell be Gx . \blacksquare

The relationship between the orbits in X and the group structure of G lies at the heart of many applications. The following theorem gives this relationship. Recall that for a set X , we use $|X|$ for the number of elements in X , and $(G : H)$ is the index of a subgroup H in a group G .

14.17 Theorem Let X be a G -set and let $x \in X$. Then $|Gx| = (G : G_x)$. If $|G|$ is finite, then $|Gx|$ is a divisor of $|G|$.

Proof We define a one-to-one map ψ from Gx onto the collection of left cosets of G_x in G . Let $x_1 \in Gx$. Then there exists $g_1 \in G$ such that $g_1x = x_1$. We define $\psi(x_1)$ to be the left coset g_1G_x of G_x . We must show that this map ψ is well defined, independent of the choice of $g_1 \in G$ such that $g_1x = x_1$. Suppose also that $g_1'x = x_1$. Then, $g_1x = g_1'x$, so $g_1^{-1}(g_1x) = g_1^{-1}(g_1'x)$, from which we deduce $x = (g_1^{-1}g_1')x$. Therefore $g_1^{-1}g_1' \in G_x$, so $g_1' \in g_1G_x$, and $g_1G_x = g_1'G_x$. Thus the map ψ is well defined.

To show the map ψ is one-to-one, suppose $x_1, x_2 \in Gx$, and $\psi(x_1) = \psi(x_2)$. Then there exist $g_1, g_2 \in G$ such that $x_1 = g_1x$, $x_2 = g_2x$, and $g_2 \in g_1G_x$. Then $g_2 = g_1g$ for some $g \in G_x$, so $x_2 = g_2x = g_1(gx) = g_1x = x_1$. Thus ψ is one-to-one.

Finally, we show that each left coset of G_x in G is of the form $\psi(x_1)$ for some $x_1 \in Gx$. Let g_1G_x be a left coset. Then if $g_1x = x_1$, we have $g_1G_x = \psi(x_1)$. Thus ψ maps Gx one-to-one onto the collection of left cosets so $|Gx| = (G : G_x)$.

If $|G|$ is finite, then the equation $|G| = |G_x|(G : G_x)$ shows that $|Gx| = (G : G_x)$ is a divisor of $|G|$. \blacklozenge

14.18 Example Let X be the D_4 -set in Example 14.9, with action table given by Table 14.11. With $G = D_4$, we have $G_0 = \{\iota, \mu\}$. Since $|G| = 8$, we have $|G_0| = (G : G_0) = 4$. From Table 14.11, we see that $G_0 = \{0, 1, 2, 3\}$, which indeed has four elements. \blacktriangle

We should remember not only the cardinality equation in Theorem 14.17 but also that the *elements of G carrying x into g_1x are precisely the elements of the left coset g_1G_x* . Namely, if $g \in G_x$, then $(g_1g)x = g_1(gx) = g_1x$. On the other hand, if $g_2x = g_1x$, then $g_1^{-1}(g_2x) = x$ so $(g_1^{-1}g_2)x = x$. Thus $g_1^{-1}g_2 \in G_x$ so $g_2 \in g_1G_x$.

Applications of G -Sets to Finite Groups

Theorem 14.17 is a very useful theorem in the study of finite groups. Suppose that X is a G -set for a finite group G and we pick out one element from each orbit of X to make the set $S = \{x_1, x_2, \dots, x_r\}$ where we indexed the elements of X so that if $i \leq j$, then $|Gx_i| \geq |Gx_j|$. That is, we arrange by orbit size, largest first and smallest last. Every element in X is in precisely one orbit, so

$$|X| = \sum_{i=1}^r |Gx_i|. \quad (1)$$

We let $X_G = \{x \in X \mid gx = x \text{ for all } g \in G\}$. That is, X_G is the set of all elements of X whose orbit size is 1. So by equation (1),

$$|X| = |X_G| + \sum_{i=1}^s |Gx_i| \quad (2)$$

where we simply place all the orbits with one element into X_G and we are left with s orbits each containing at least two elements. Although Equation (2) is simply saying that if you add up the sizes of all the orbits you account for all the elements of X , when coupled with Theorem 14.17, it gives some very interesting results. We give a few in the remainder of this section. In Section 17 we will use Equation 2 extensively to prove the Sylow Theorems.

For the remainder of this section, we assume that p is a prime number.

14.19 Theorem Let G be a group with p^n elements. If X is a G -set, then $|X| \equiv |X_G| \pmod{p}$.

Proof Using Equation 2,

$$|X| = |X_G| + \sum_{i=1}^s |Gx_i|.$$

Since for each $i \leq s$, $|Gx_i| \geq 2$ and $|Gx_i| = (G : G_{x_i})$ is a divisor of $|G| = p^n$, by Theorem 14.17 p divides each term in the sum $\sum_{i=1}^s |Gx_i|$. Thus $|X| \equiv |X_G| \pmod{p}$. \blacklozenge

Knowing that k divides the order of a group is not sufficient information to assume that the group has a subgroup of order k . For example, we saw that A_4 has no subgroup of order 6 and that in general, A_n has no subgroup of index 2 if $n \geq 4$. On the positive side, in Exercise 29 in Section 2, you were asked to show that if a group has an even number of elements, then it has an element of order two. Theorem 14.20 generalizes this result to show that if a prime number p divides the order of a group, then the group has an element of order p . The proof of this theorem relies on Theorem 14.19.

14.20 Theorem (Cauchy's Theorem) Let G be a group such that p divides the order of G . Then G has an element of order p and therefore a subgroup of order p .

Proof We let

$$X = \{(g_0, g_1, g_2, \dots, g_{p-1}) \mid g_0, g_1, \dots, g_{p-1} \in G \text{ and } g_0 g_1 g_2 \dots g_{p-1} = e\}.$$

That is, X is the set of all p -tuples with entries in G so that when the entries are multiplied together (in order) their product is the identity e . Since the product is e , $g_0 = (g_1 g_2 \dots g_{p-1})^{-1}$ and given any $g_1, g_2, \dots, g_{p-1} \in G$, by picking $g_0 = (g_1 g_2 \dots g_{p-1})^{-1}$ we have an element in X . Thus $|X| = |G|^{p-1}$ and in particular, p divides the order of X since p divides the order of G .

Suppose that $(g_0, g_1, g_2, \dots, g_{p-1}) \in X$. Since $g_0 = (g_1 g_2 \dots g_{p-1})^{-1}$, it follows that $(g_1, g_2, \dots, g_{p-1}, g_0)$ is in X . Repeating this process, noting that $g_1 = (g_2 g_3 \dots g_{p-1} g_0)^{-1}$ we conclude that $(g_2, g_3, g_4, \dots, g_{p-1}, g_0, g_1) \in X$. Continuing in this manner we have that for any $k \in \mathbb{Z}_p$,

$$(g_k, g_{k+p}, g_{k+2p}, \dots, g_{k+(p-1)p}) \in X.$$

We check that this gives a group action of \mathbb{Z}_p on X . Let $k \in \mathbb{Z}_p$ and $(g_0, g_1, g_2, \dots, g_{p-1}) \in X$. Then

$$k(g_0, g_1, g_2, \dots, g_{p-1}) = (g_k, g_{k+p}, g_{k+2p}, \dots, g_{k+(p-1)p}) \in X.$$

Since

$$0(g_0, g_1, g_2, \dots, g_{p-1}) = (g_0, g_1, g_2, \dots, g_{p-1}) \text{ and}$$

$$\begin{aligned} k(l(g_0, g_1, g_2, \dots, g_{p-1})) &= k(g_l, g_{l+p}, g_{l+2p}, \dots, g_{l+(p-1)p}) \\ &= (g_{k+p}, g_{k+p+l}, \dots, g_{k+p+l+(p-1)p}) \\ &= (k + p)l(g_0, g_1, g_2, \dots, g_{p-1}) \end{aligned}$$

this is indeed a group action.

By Theorem 14.19, $0 \equiv |X| \equiv |X_{\mathbb{Z}_p}| \pmod{p}$. The p -tuple (e, e, e, \dots, e) is in $X_{\mathbb{Z}_p}$ because rearranging the entries does not change the p -tuple. Since $X_{\mathbb{Z}_p}$ contains at least one element and p divides $|X_{\mathbb{Z}_p}|$, $X_{\mathbb{Z}_p}$ must contain at least one element other than (e, e, e, \dots, e) . That element must have the form (a, a, a, \dots, a) with $a \neq e$ and $a^p = e$. So a has order p and the subgroup it generates is a subgroup of G with order p . \blacklozenge

14.21 Definition A p -group is a group such that each element in the group has order a power of p . A p -subgroup of a group is a subgroup that is a p -group. \blacksquare

14.22 Example The group D_{16} is a 2-group since the order of any element of D_{16} divides $|D_{16}| = 32$. \blacktriangle

14.23 Example Using the Fundamental Theorem of Finitely Generated Abelian Groups, a finite abelian group is a p -group if and only if it is isomorphic to

$$\mathbb{Z}_{p^{r_1}} \times \mathbb{Z}_{p^{r_2}} \times \mathbb{Z}_{p^{r_3}} \times \cdots \times \mathbb{Z}_{p^{r_n}}.$$

This is because if there were a factor of the form \mathbb{Z}_{q^s} where $q \neq p$ is a prime number and $s \geq 1$, then there would be an element in G with order q^s which is not a power of p .

In Exercise 30, you are asked to show that for G a finite group, G is a p -group if and only if the order of G is a power of p .

The next theorem assures us that any finite p -group has a nontrivial normal subgroup, namely the center of the group. \blacktriangle

14.24 Theorem Let G be a finite p -group. Then the center of G , $Z(G)$, is not the trivial group.

Proof We let $X = G$ and we make X into a G -set using conjugation. That is, $*(g, a) = gag^{-1}$. Equation 2 states that $0 \equiv |X| \equiv |X_G| \pmod{p}$. For all $g \in G$, $gag^{-1} = e$. So X_G has at

least one element, namely e . Since the number of elements in X_G must be at least p , there is an element $a \in X$ such that $a \neq e$ and $gag^{-1} = a$ for all $g \in G$. Thus $ga = ag$ for all $g \in G$, which says that $a \in Z(G)$. So $Z(G)$ is not the trivial subgroup. \blacklozenge

When studying p -groups, the fact that the center is nontrivial is often very helpful. We conclude this section with a theorem that illustrates the utility of Theorem 14.24.

14.25 Theorem Every group of order p^2 is abelian.

Proof Let G be a group of order p^2 with center $Z(G)$. By Theorem 14.24, $Z(G)$ is not the trivial group so it is either all of G or else it has order p . We wish to show that $Z(G) = G$ using proof by contradiction. So we assume that $Z(G)$ has p elements. Since $Z(G)$ is a normal subgroup of G , we can form $G/Z(G)$. The group $G/Z(G)$ also has p elements and so both $Z(G)$ and $G/Z(G)$ are cyclic. Let $\langle a \rangle = Z(G)$ and $\langle bZ(G) \rangle = G/Z(G)$. Let $x, y \in G$. Then $x = b^i a^j$ and $y = b^r a^s$ for some integers i, j, r, s since the cosets of $Z(G)$ partition G . Then

$$xy = b^i a^j b^r a^s = b^r b^i a^s a^j$$

since $\langle a \rangle$ is the center of G . So

$$xy = b^{i+r} a^{j+s} = b^r b^i a^s a^j = b^r a^s b^i a^j = yx.$$

Since every element in G commutes with every other element, $Z(G) = G$, which contradicts our assumption that the center has only p elements. So the center of G must be G , which means that G is abelian. \blacklozenge

14.26 Example Since every group of order p^2 is abelian, the Fundamental Homomorphism Theorem says that every group with p^2 elements is isomorphic to either \mathbb{Z}_{p^2} or $\mathbb{Z}_p \times \mathbb{Z}_p$. The two groups of order 4 are \mathbb{Z}_4 and the Klein 4-group. The two groups of order 9 are \mathbb{Z}_9 and $\mathbb{Z}_3 \times \mathbb{Z}_3$. \blacktriangle

■ EXERCISES 14

Computations

In Exercises 1 through 3, let

$$X = \{0, 1, 2, 3, s_0, s_1, s_2, s_3, m_1, m_2, d_1, d_2, C, P_0, P_1, P_2, P_3\}$$

be the D_4 -set of Example 14.9. Find the following, where $G = D_4$.

1. The fixed sets X_σ for each $\sigma \in D_4$.
2. The isotropy subgroups G_x for each $x \in X$, that is, $G_0, G_1, \dots, G_{P_2}, G_{P_3}$.
3. The orbits in X under D_4 .
4. Theorem 14.24 states that every p -group has nontrivial center. Find the center of D_8 .
5. Find the center of D_7 .
6. Let $G = X = S_3$ and make X a G -set using conjugation. That is, $*(\sigma, \tau) = \sigma \tau \sigma^{-1}$. Find all the orbits of X using this action. (Write permutations in disjoint cycle notation.)
7. Let $G = D_4$ and X be the set of all subgroups of D_4 with order two. The set X is a G -set using conjugation, $*(\sigma, H) = \sigma H \sigma^{-1}$. Find all the orbits of this group action.
8. Let $G = U = \{z \in \mathbb{C} \mid |z| = 1\}$ be the circle group. Then $X = \mathbb{C}$, the set of complex numbers, is a G -set with group action given by complex number multiplication. That is, if $z \in U$ and $w \in \mathbb{C}$, $*(z, w) = zw$. Find all the orbits of this action. Also, find X_G .

9. Let G be a group of order 3 and suppose that $|X| = 6$. For each possible action of G on X , give a list of the orbit sizes. List the orbit sizes from largest to smallest. (Recall that the orbits partition the set X .)
10. Let G be a group of order 9 and suppose that $|X| = 10$. For each possible action of G on X , give a list of the orbit sizes. List the orbit sizes from largest to smallest.
11. Let G be a group of order 8 and suppose that $|X| = 10$. For each possible way to make X a G -set the orbits partition X . For each possible action of G on X , give a list of the orbit sizes. List the orbit sizes from largest to smallest.

Concepts

In Exercises 12 and 13, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

12. A group G acts *faithfully* on X if and only if $gx = x$ implies that $g = e$.
13. A group G is *transitive* on a G -set X if and only if, for some $g \in G$, gx can be every other x .
14. Let X be a G -set and let $S \subseteq X$. If $Gs \subseteq S$ for all $s \in S$, then S is a **sub- G -set**. Characterize a sub- G -set of a G -set X in terms of orbits in X under G .
15. Characterize a transitive G -set in terms of its orbits.
16. Determine whether each of the following is true or false.
 - a. Every G -set is also a group.
 - b. Each element of a G -set is fixed by the identity of G .
 - c. If every element of a G -set is fixed by the same element g of G , then g must be the identity e .
 - d. Let X be a G -set with $x_1, x_2 \in X$ and $g \in G$. If $gx_1 = gx_2$, then $x_1 = x_2$.
 - e. Let X be a G -set with $x \in X$ and $g_1, g_2 \in G$. If $g_1x = g_2x$, then $g_1 = g_2$.
 - f. Each orbit of a G -set X is a transitive sub- G -set. (See Exercise 14.)
 - g. Let X be a G -set and let $H \leq G$. Then X can be regarded in a natural way as an H -set.
 - h. With reference to (g), the orbits in X under H are the same as the orbits in X under G .
 - i. If X is a G -set, then each element of G acts as a permutation of X .
 - j. Let X be a G -set and let $x \in X$. If G is finite, then $|G| = |Gx| \cdot |G_x|$.
17. Let X and Y be G -sets with the *same* group G . An **isomorphism** between G -sets X and Y is a map $\phi : X \rightarrow Y$ that is one-to-one, onto Y , and satisfies $g\phi(x) = \phi(gx)$ for all $x \in X$ and $g \in G$. Two G -sets are **isomorphic** if such an isomorphism between them exists. Let X be the D_4 -set of Example 14.9.
 - a. Find two distinct orbits of X that are isomorphic sub- D_4 -sets. (See Exercise 14.)
 - b. Show that the orbits $\{0, 1, 2, 3\}$ and $\{s_0, s_1, s_2, s_3\}$ are not isomorphic sub- D_4 -sets. [Hint: Find an element of G that acts in an essentially different fashion on the two orbits.]
 - c. Are the orbits you gave for your answer to part (a) the only two different isomorphic sub- D_4 -sets of X ?
18. Let X be the D_4 -set in Example 14.9.
 - a. Does D_4 act faithfully on X ?
 - b. Find all orbits in X on which D_4 acts faithfully as a sub- D_4 -set. (See Exercise 14.)

Theory

19. Let X be a G -set. Show that G acts faithfully on X if and only if no two distinct elements of G have the same action on each element of X .
20. Let X be a G -set and let $Y \subseteq X$. Let $G_Y = \{g \in G \mid gy = y \text{ for all } y \in Y\}$. Show G_Y is a subgroup of G , generalizing Theorem 14.13.
21. Let G be the additive group of real numbers. Let the action of $\theta \in G$ on the real plane \mathbb{R}^2 be given by rotating the plane counterclockwise about the origin through θ radians. Let P be a point other than the origin in the plane.
 - a. Show \mathbb{R}^2 is a G -set.
 - b. Describe geometrically the orbit containing P .
 - c. Find the group G_P .

Exercises 22 through 25 show how all possible G -sets, up to isomorphism (see Exercise 17), can be formed from the group G .

22. Let $\{X_i \mid i \in I\}$ be a disjoint collection of sets, so $X_i \cap X_j = \emptyset$ for $i \neq j$. Let each X_i be a G -set for the same group G .
 - a. Show that $\bigcup_{i \in I} X_i$ can be viewed in a natural way as a G -set, the **union** of the G -sets X_i .
 - b. Show that every G -set X is the union of its orbits.
23. Let X be a transitive G -set, and let $x_0 \in X$. Show that X is isomorphic (see Exercise 17) to the G -set L of all left cosets of G_{x_0} , described in Example 14.8. [Hint: For $x \in X$, suppose $x = gx_0$, and define $\phi : X \rightarrow L$ by $\phi(x) = gG_{x_0}$. Be sure to show ϕ is well defined!]
24. Let X_i for $i \in I$ be G -sets for the same group G , and suppose the sets X_i are not necessarily disjoint. Let $X'_i = \{(x, i) \mid x \in X_i\}$ for each $i \in I$. Then the sets X'_i are disjoint, and each can still be regarded as a G -set in an obvious way. (The elements of X_i have simply been tagged by i to distinguish them from the elements of X_j for $i \neq j$.) The G -set $\bigcup_{i \in I} X'_i$ is the **disjoint union** of the G -sets X_i . Using Exercises 22 and 23, show that every G -set is isomorphic to a disjoint union of left coset G -sets, as described in Example 14.12.
25. The preceding exercises show that every G -set X is isomorphic to a disjoint union of left coset G -sets. The question then arises whether left coset G -sets of distinct subgroups H and K of G can themselves be isomorphic. Note that the map defined in the hint of Exercise 23 depends on the choice of x_0 as “base point.” If x_0 is replaced by g_0x_0 and if $G_{x_0} \neq G_{g_0x_0}$, then the collections L_H of left cosets of $H = G_{x_0}$ and L_K of left cosets of $K = G_{g_0x_0}$ form distinct G -sets that must be isomorphic, since both L_H and L_K are isomorphic to X .
 - a. Let X be a transitive G -set and let $x_0 \in X$ and $g_0 \in G$. If $H = G_{x_0}$, describe $K = G_{g_0x_0}$ in terms of H and g_0 .
 - b. Based on part (a), conjecture conditions on subgroups H and K of G such that the left coset G -sets of H and K are isomorphic.
 - c. Prove your conjecture in part (b).
26. Up to isomorphism, how many transitive \mathbb{Z}_4 -sets X are there? (Use the preceding exercises.) Give an example of each isomorphism type, listing an action table of each as in Table 14.11. Take lowercase names a, b, c , and so on for the elements in the set X .
27. Repeat Exercise 26 for the group \mathbb{Z}_6 .
28. Repeat Exercise 26 for the group S_3 . List the elements of S_3 in the order $\iota, (1, 2, 3), (1, 3, 2), (2, 3), (1, 3), (1, 2)$.
29. Prove that if G is a group of order p^3 , where p is a prime number, then $|Z(G)|$ is either p or p^3 . Give an example where $|Z(G)| = p$ and an example where $|Z(G)| = p^3$.
30. Let p be a prime number. Prove that a finite group G is a p -group if and only if $|G| = p^n$ for some integer $n \geq 0$.
31. Let G be a group that acts on $X = \{H \mid H \leq G\}$ by conjugation. That is, $g * H = gHg^{-1}$. State and prove an equivalent condition for a subgroup $H \leq G$ to be a normal subgroup of G in terms of
 - a. G_H , the isotropy subgroup of H .
 - b. GH , the orbit of H .

SECTION 15

[†] APPLICATIONS OF G -SETS TO COUNTING

This section presents an application of our work with G -sets to counting. Suppose, for example, we wish to count how many distinguishable ways the six faces of a cube can be marked with from one to six dots to form a die. The standard die is marked so that when placed on a table with the 1 on the bottom and the 2 toward the front, the 6 is on top, the 3 on the left, the 4 on the right, and the 5 on the back. Of course, other ways of marking the cube to give a distinguishably different die are possible.

[†] This section is not used in the remainder of the text.

Let us distinguish between the faces of the cube for the moment and call them the bottom, top, left, right, front, and back. Then the bottom can have any one of six marks from one dot to six dots, the top any one of the five remaining marks, and so on. There are $6! = 720$ ways the cube faces can be marked in all. Some markings yield the same die as others, in the sense that one marking can be carried into another by a rotation of the marked cube. For example, if the standard die described above is rotated 90° counterclockwise as we look down on it, then 3 will be on the front face rather than 2, but it is the same die.

There are 24 possible positions of a cube on a table, for any one of six faces can be placed down, and then any one of four to the front, giving $6 \cdot 4 = 24$ possible positions. Any position can be achieved from any other by a rotation of the die. These rotations form a group G , which is isomorphic to a subgroup of S_8 . We let X be the 720 possible ways of marking the cube and let G act on X by rotation of the cube. We consider two markings to give the same die if one can be carried into the other under action by an element of G , that is, by rotating the cube. In other words, we consider each *orbit* in X under G to correspond to a single die, and different orbits to give different dice. The determination of the number of distinguishable dice thus leads to the question of determining the number of orbits under G in a G -set X .

The following theorem gives a tool for determining the number of orbits in a G -set X under G . Recall that for each $g \in G$ we let X_g be the set of elements of X fixed by g , so that $X_g = \{x \in X \mid gx = x\}$. Recall also that for each $x \in X$, we let $G_x = \{g \in G \mid gx = x\}$, and Gx is the orbit of x under G .

15.1 Theorem (Burnside's Formula) Let G be a finite group and X a finite G -set. If r is the number of orbits in X under G ,

$$r \cdot |G| = \sum_{g \in G} |X_g|. \quad (1)$$

Proof We consider all pairs (g, x) where $gx = x$, and let N be the number of such pairs. For each $g \in G$ there are $|X_g|$ pairs having g as first member. Thus,

$$N = \sum_{g \in G} |X_g|. \quad (2)$$

On the other hand, for each $x \in X$ there are $|G_x|$ pairs having x as second member. Thus we also have

$$N = \sum_{x \in X} |G_x|.$$

By Theorem 14.17 we have $|Gx| = (G : G_x)$. But we know that $(G : G_x) = |G|/|G_x|$, so we obtain $|G_x| = |G|/|Gx|$. Then

$$N = \sum_{x \in X} \frac{|G|}{|Gx|} = |G| \left(\sum_{x \in X} \frac{1}{|Gx|} \right). \quad (3)$$

Now $1/|Gx|$ has the same value for all x in the same orbit, and if we let \mathcal{O} be any orbit, then

$$\sum_{x \in \mathcal{O}} \frac{1}{|Gx|} = \sum_{x \in \mathcal{O}} \frac{1}{|\mathcal{O}|} = 1. \quad (4)$$

Substituting (4) in (3), we obtain

$$N = |G| (\text{number of orbits in } X \text{ under } G) = |G| \cdot r. \quad (5)$$

Comparison of Eq. 2 and Eq. 5 gives Eq. 1. ◆

15.2 Corollary If G is a finite group and X is a finite G -set, then

$$(\text{number of orbits in } X \text{ under } G) = \frac{1}{|G|} \cdot \sum_{g \in G} |X_g|.$$

Proof The proof of this corollary follows immediately from the preceding theorem. ◆

Let us continue our computation of the number of distinguishable dice as our first example.

15.3 Example We let X be the set of 720 different markings of faces of a cube using from one to six dots. Let G be the group of 24 rotations of the cube as discussed above. We saw that the number of distinguishable dice is the number of orbits in X under G . Now $|G| = 24$. For $g \in G$ where $g \neq e$, we have $|X_g| = 0$, because any rotation other than the identity element changes any one of the 720 markings into a different one. However, $|X_e| = 720$ since the identity element leaves all 720 markings fixed. Then by Corollary 15.2,

$$(\text{number of orbits}) = \frac{1}{24} \cdot 720 = 30,$$

so there are 30 distinguishable dice. ▲

Of course, the number of distinguishable dice could be counted without using the machinery of the preceding corollary, but by using elementary combinatorics as often taught in a freshman finite math course. In marking a cube to make a die, we can, by rotation if necessary, assume the face marked 1 is down. There are five choices for the top (opposite) face. By rotating the die as we look down on it, any one of the remaining four faces could be brought to the front position, so there are no different choices involved for the front face. But with respect to the number on the front face, there are $3 \cdot 2 \cdot 1$ possibilities for the remaining three side faces. Thus there are $5 \cdot 3 \cdot 2 \cdot 1 = 30$ possibilities in all.

The next two examples appear in some finite math texts and are easy to solve by elementary means. We use Corollary 15.2 so that we have more practice thinking in terms of orbits.

15.4 Example How many distinguishable ways can seven people be seated at a round table, where there is no distinguishable “head” to the table? Of course there are $7!$ ways to assign people to the different chairs. We take X to be the $7!$ possible assignments. A rotation of people achieved by asking each person to move one place to the right results in the same arrangement. Such a rotation generates a cyclic group G of order 7, which we consider to act on X in the obvious way. Again, only the identity e leaves any arrangement fixed, and it leaves all $7!$ arrangements fixed. By Corollary 15.2

$$(\text{number of orbits}) = \frac{1}{7} \cdot 7! = 6! = 720.$$

15.5 Example How many distinguishable necklaces (with no clasp) can be made using seven different-colored beads of the same size? Unlike the table in Example 15.4, the necklace can be turned over as well as rotated. Thus we consider the full dihedral group D_7 of order $2 \cdot 7 = 14$ as acting on the set X of $7!$ possibilities. Then the number of distinguishable necklaces is

$$(\text{number of orbits}) = \frac{1}{14} \cdot 7! = 360.$$

In using Corollary 15.2, we have to compute $|G|$ and $|X_g|$ for each $g \in G$. In the examples and the exercises, $|G|$ will pose no real problem. Let us give an example

where $|X_g|$ is not as trivial to compute as in the preceding examples. We will continue to assume knowledge of very elementary combinatorics.

15.6 Example Let us find the number of distinguishable ways the edges of an equilateral triangle can be painted if four different colors of paint are available, assuming only one color is used on each edge, and the same color may be used on different edges.

Of course there are $4^3 = 64$ ways of painting the edges in all, since each of the three edges may be any one of four colors. We consider X to be the set of these 64 possible painted triangles. The group G acting on X is the group of symmetries of the triangle, which is isomorphic to S_3 and which we consider to be S_3 . We need to compute $|X_g|$ for each of the six elements g in S_3 .

$$|X_i| = 64$$

Every painted triangle is fixed by i .

$$|X_{(1,2,3)}| = 4$$

To be invariant under $(1,2,3)$ all edges must be the same color, and there are 4 possible colors.

$$|X_{(1,3,2)}| = 4$$

Same reason as for $(1,2,3)$.

$$|X_{(1,2)}| = 16$$

The edges that are interchanged must be the same color (4 possibilities) and the other edge may also be any of the colors (times 4 possibilities).

$$|X_{(2,3)}| = |X_{(1,3)}| = 16$$

Same reason as for $(1,2)$.

Then

$$\sum_{g \in S_3} |X_g| = 64 + 4 + 4 + 16 + 16 + 16 = 120.$$

Thus

$$(\text{number of orbits}) = \frac{1}{6} \cdot 120 = 20,$$

and there are 20 distinguishable painted triangles. ▲

15.7 Example We repeat Example 15.6 with the assumption that a different color is used on each edge. The number of possible ways of painting the edges is then $4 \cdot 3 \cdot 2 = 24$, and we let X be the set of 24 possible painted triangles. Again, the group acting on X can be considered to be S_3 . Since all edges are a different color, we see $|X_i| = 24$ while $|X_g| = 0$ for $g \neq i$. Thus

$$(\text{number of orbits}) = \frac{1}{6} \cdot 24 = 4,$$

so there are four distinguishable triangles. ▲

We will use group actions in Section 17 to develop the Sylow Theorems, which give a tremendous amount of information about finite groups. In this section, we barely scratch the surface of how to count using Burnside's Formula. To explore this fascinating topic further, search the Internet using key words such as "cycle index" and "Pólya's Enumeration Theorem." Given a group action on a set, the cycle index is a polynomial that can be computed by hand for small groups and by computer for larger groups. Pólya's Enumeration Theorem then says that the number of different ways to color an object can be computed by simply substituting certain values into the polynomial. It is remarkable that counting the number of different colorings of geometric objects can be elegantly reduced to algebra!

■ EXERCISES 15**Computations**

In each of the following exercises use Corollary 15.2, even though the answer might be obtained by more elementary methods.

1. Find the number of orbits in $\{1, 2, 3, 4, 5, 6, 7, 8\}$ under the cyclic subgroup $((1, 3, 5, 6))$ of S_8 .
2. Find the number of orbits in $\{1, 2, 3, 4, 5, 6, 7, 8\}$ under the subgroup of S_8 generated by $(1, 3)$ and $(2, 4, 7)$.
3. Find the number of distinguishable tetrahedral dice that can be made using one, two, three, and four dots on the faces of a regular tetrahedron, rather than a cube.
4. Wooden cubes of the same size are to be painted a different color on each face to make children's blocks. How many distinguishable blocks can be made if eight colors of paint are available?
5. Answer Exercise 4 if colors may be repeated on different faces at will. [Hint: The 24 rotations of a cube consist of the identity, 9 that leave a pair of opposite faces invariant, 8 that leave a pair of opposite vertices invariant, and 6 leaving a pair of opposite edges invariant.]
6. Each of the eight corners of a cube is to be tipped with one of four colors, each of which may be used on from one to all eight corners. Find the number of distinguishable markings possible. (See the hint in Exercise 5.)
7. Find the number of distinguishable ways the edges of a square of cardboard can be painted if six colors of paint are available and
 - a. no color is used more than once.
 - b. the same color can be used on any number of edges.
8. Consider six straight wires of equal lengths with ends soldered together to form edges of a regular tetrahedron. Either a 50-ohm or 100-ohm resistor is to be inserted in the middle of each wire. Assume there are at least six of each type of resistor available. How many essentially different wirings are possible?
9. A rectangular prism 2 ft long with 1-ft square ends is to have each of its six faces painted with one of six possible colors. How many distinguishable painted prisms are possible if
 - a. no color is to be repeated on different faces,
 - b. each color may be used on any number of faces?

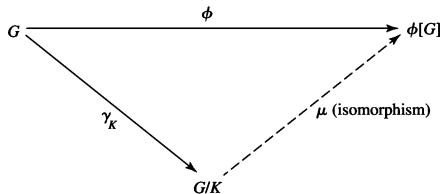
Advanced Group Theory

- Section 16** Isomorphism Theorems
- Section 17** Sylow Theorems
- Section 18** Series of Groups
- Section 19** Free Abelian Groups
- Section 20** Free Groups
- Section 21** Group Presentations

SECTION 16

ISOMORPHISM THEOREMS

There are several theorems concerning isomorphic factor groups that are known as the *isomorphism theorems* of group theory. The first of these is Theorem 12.14, which we restate for easy reference. The theorem is diagrammed in Fig. 16.1.



16.1 Figure

16.2 Theorem (First Isomorphism Theorem) Let $\phi : G \rightarrow G'$ be a homomorphism with kernel K , and let $\gamma_K : G \rightarrow G/K$ be the canonical homomorphism. There is a unique isomorphism $\mu : G/K \rightarrow \phi[G]$ such that $\phi(x) = \mu(\gamma_K(x))$ for each $x \in G$. \blacklozenge

The lemma that follows will be of great aid in our proof and intuitive understanding of the other two isomorphism theorems.

16.3 Lemma Let N be a normal subgroup of a group G and let $\gamma : G \rightarrow G/N$ be the canonical homomorphism. Then the map ϕ from the set of normal subgroups of G containing N to the set of normal subgroups of G/N given by $\phi(L) = \gamma[L]$ is one-to-one and onto.

Proof Theorem 13.18 shows that if L is a normal subgroup of G containing N , then $\phi(L) = \gamma[L]$ is a normal subgroup of G/N . Because $N \leq L$, for each $x \in L$ the entire coset xN in G is contained in L . Thus by Theorem 10.17, $\gamma^{-1}[\phi(L)] = L$. Consequently, if L and M are normal subgroups of G , both containing N , and if $\phi(L) = \phi(M) = H$, then $L = \gamma^{-1}[H] = M$. Therefore ϕ is one-to-one.

If H is a normal subgroup of G/N , then $\gamma^{-1}[H]$ is a normal subgroup of G by Theorem 13.18. Because $N \in H$ and $\gamma^{-1}[N] = N$, we see that $N \subseteq \gamma^{-1}[H]$. Then $\phi(\gamma^{-1}[H]) = \gamma[\gamma^{-1}[H]] = H$. This shows that ϕ is onto the set of normal subgroups of G/N . \blacklozenge

If H and N are subgroups of a group G , then we let

$$HN = \{hn \mid h \in H, n \in N\}.$$

We define the **join** $H \vee N$ of H and N as the intersection of all subgroups of G that contain HN ; thus $H \vee N$ is the smallest subgroup of G containing HN . Of course $H \vee N$ is also the smallest subgroup of G containing both H and N , since any such subgroup must contain HN . In general, HN need not be a subgroup of G . However, we have the following lemma.

16.4 Lemma If N is a normal subgroup of G , and if H is any subgroup of G , then $H \vee N = HN = NH$. Furthermore, if H is also normal in G , then HN is normal in G .

Proof We show that HN is a subgroup of G , from which $H \vee N = HN$ follows at once. Let $h_1, h_2 \in H$ and $n_1, n_2 \in N$. Since N is a normal subgroup, we have $n_1 h_2 = h_2 n_3$ for some $n_3 \in N$. Then $(h_1 n_1)(h_2 n_2) = h_1(n_1 h_2)n_2 = h_1(h_2 n_3)n_2 = (h_1 h_2)(n_3 n_2) \in HN$, so HN is closed under the induced operation in G . Clearly $e = ee$ is in HN . For $h \in H$ and $n \in N$, we have $(hn)^{-1} = n^{-1}h^{-1} = h^{-1}n_4$ for some $n_4 \in N$, since N is a normal subgroup. Thus $(hn)^{-1} \in HN$, so $HN \leq G$. A similar argument shows that NH is a subgroup, so $NH = H \vee N = HN$.

Now suppose that H is also normal in G , and let $h \in H, n \in N$, and $g \in G$. Then $ghng^{-1} = (ghg^{-1})(gng^{-1}) \in HN$, so HN is indeed normal in G . \blacklozenge

We are now ready for the second isomorphism theorem.

16.5 Theorem (Second Isomorphism Theorem) Let H be a subgroup of G and let N be a normal subgroup of G . Then $(HN)/N \cong H/(H \cap N)$.

Proof Since $N \leq HN \leq G$ and N is a normal subgroup of G , N is a normal subgroup of HN , which allows us to form the group HN/N . We define a map $\phi : H \rightarrow HN/N$ by $\phi(h) = hN$. The map ϕ is a homomorphism since for any $h_1, h_2 \in H$,

$$\phi(h_1 h_2) = (h_1 h_2)N = (h_1 N)(h_2 N) = \phi(h_1)\phi(h_2).$$

The map ϕ maps onto HN/N since any element of HN/N can be written as hnN for some $h \in H$ and $n \in N$ and $hnN = hN = \phi(h)$. We now compute the kernel of ϕ .

$$\text{Ker}(\phi) = \{h \in H \mid hN = N\} = \{h \in H \mid h \in N\} = H \cap N$$

By the First Isomorphism Theorem, Theorem 16.2, the map $\mu : H/(H \cap N) \rightarrow HN/N$ defined by $\mu(h(H \cap N)) = hN$ is an isomorphism. \blacklozenge

16.6 Example Let $G = \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$, $H = \mathbb{Z} \times \mathbb{Z} \times \{0\}$, and $N = \{0\} \times \mathbb{Z} \times \mathbb{Z}$. Then clearly $HN = \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ and $H \cap N = \{0\} \times \mathbb{Z} \times \{0\}$. We have $(HN)/N \cong \mathbb{Z}$ and we also have $H/(H \cap N) \cong \mathbb{Z}$. \blacktriangle

16.7 Example Let $G = \mathbb{Z}$, $N = \langle n \rangle$, and $H = \langle h \rangle$ where n and h are positive. The group \mathbb{Z} is abelian, so N is a normal subgroup of \mathbb{Z} . Since \mathbb{Z} is an additive group, we write $N + H$ instead of NH to avoid confusion. The group $N + H = \langle \gcd(n, h) \rangle$ since anything in $N + H$ is a multiple of $\gcd(n, h)$ and $\gcd(n, h) = xn + yh \in N + H$ for some integers x and y . Also, $N \cap H = \langle \text{lcm}(n, h) \rangle$ since $a \in N \cap H$ if and only if a is a multiple of both n and h . The Second Isomorphism Theorem states that $(N + H)/N = \langle \gcd(n, h) \rangle / \langle n \rangle$

is isomorphic with $H/(N \cap H) = \langle h \rangle /(\text{lcm}(n, h))$. In Exercise 10, you will be asked to prove that if $a, b \in \mathbb{Z}^+$ and a divides b , then $|\langle a \rangle / \langle b \rangle| = b/a$. Since $\langle \text{gcd}(n, h) \rangle / \langle n \rangle \cong \langle h \rangle /(\text{lcm}(n, h))$, we have

$$\frac{n}{\text{gcd}(n, h)} = \frac{\text{lcm}(n, h)}{h}$$

$$nh = \text{gcd}(n, h)\text{lcm}(n, h),$$

which provides a complicated way of proving a basic number theory fact! We conclude that

$$\langle \text{gcd}(n, h) \rangle / \langle n \rangle \cong \langle h \rangle /(\text{lcm}(n, h)) \cong \mathbb{Z}_{\frac{n}{\text{gcd}(n, h)}}$$

since a factor group of a cyclic group is cyclic. \blacktriangle

If H and K are two normal subgroups of G and $K \leq H$, then H/K is a normal subgroup of G/K . The third isomorphism theorem concerns these groups.

16.8 Theorem (Third Isomorphism Theorem) Let H and K be normal subgroups of a group G with $K \leq H$. Then $G/H \cong (G/K)/(H/K)$.

Proof Since K is a subgroup of H , for any $g \in G$, $gK \subseteq gH$. That is, each left coset of K is completely contained in one coset of H . We define $\phi : G/K \rightarrow G/H$ by $\phi(gK) = gH$. That is, we map a coset of K to the coset of H that contains it. Again, our strategy is to use the First Isomorphism Theorem. The map ϕ is a homomorphism since for any $g_1, g_2 \in G$,

$$\begin{aligned}\phi((g_1K)(g_2K)) &= \phi((g_1g_2)K) = (g_1g_2)H \\ &= (g_1H)(g_2H) \\ &= \phi(g_1K)\phi(g_2K).\end{aligned}$$

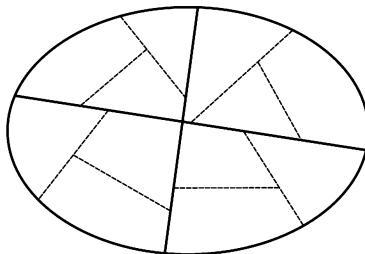
The map ϕ maps onto G/H since for any coset $gH \in G/H$, $\phi(gK) = gH$. We now compute the kernel of ϕ .

$$\begin{aligned}\text{Ker}(\phi) &= \{gK \in G/K \mid gH = H\} \\ &= \{gK \in G/K \mid g \in H\} \\ &= H/K.\end{aligned}$$

By the First Isomorphism Theorem $(G/K)/(H/K)$ is isomorphic with G/H and a formula for an isomorphism $\mu : (G/K)/(H/K) \rightarrow G/H$ is $\mu((gK)H/K) = gH$. \blacklozenge

The formula for the isomorphism in the previous proof says that if we collapse the subgroup K in G to form G/K and then collapse all the K cosets inside of H we have the same group as collapsing the subgroup H . Figure 16.9 illustrates the situation. Think of the large ellipse as being the group G . The cosets of H are the sets bounded by the thick solid lines. The cosets of K are the smaller sets inside the cosets of H . The set G/H consists of the cosets of H , which are represented by the four larger areas, each of which is one point in G/H . On the other hand, G/K is represented by the twelve smaller sets each collapsed to a point. Then $(G/K)/(H/K)$ collapses each of the three small areas in the same H coset to their common H coset. Either way, we end up collapsing each H coset to a point.

It is sometimes difficult to think about what groups look like when they contain more than one factor group in their definition. For example, what does an element of $(G/K)/(H/K)$ really look like? Keep in mind that a factor group has cosets of the subgroup as elements. Example 16.10 is intended to clarify what the group $(G/K)/(H/K)$ looks like and to explicitly show what the isomorphism of Theorem 16.8 looks like.



16.9 Figure

16.10 Example Let $G = \mathbb{Z}_8$, $H = \langle 2 \rangle = \{0, 2, 4, 6\} < G$, and $K = \langle 4 \rangle = \{0, 4\} < H$. We list the elements of each of the factor groups used in Theorem 16.8.

$$G/K = \{\{0, 4\}, \{1, 5\}, \{2, 6\}, \{3, 7\}\}$$

$$G/H = \{\{0, 2, 4, 6\}, \{1, 3, 5, 7\}\}$$

$$H/K = \{\{0, 4\}, \{2, 6\}\}$$

Before listing the elements of $(G/K)/(H/K)$ we note that any element of this group would be a coset of the subgroup $H/K < G/K$. So each element is a set whose elements are sets.

$$(G/K)/(H/K) = \{\{\{0, 4\}, \{2, 6\}\}, \{\{1, 5\}, \{3, 7\}\}\}$$

Comparing G/H and $(G/K)/(H/K)$, we see that each element in G/H is the union of two cosets of K in G/K . Also, the isomorphism in Theorem 16.8 is the map $\phi : (G/K)/(H/K) \rightarrow G/H$ defined by

$$\phi(\{\{0, 4\}, \{2, 6\}\}) = \{0, 4, 2, 6\} \quad \text{and}$$

$$\phi(\{\{1, 5\}, \{3, 7\}\}) = \{1, 5, 3, 7\}.$$

So as illustrated in Figure 16.9, collapsing H to form G/H can be accomplished by first collapsing K and then collapsing H/K . ▲

■ EXERCISES 16

Computations

In using the three isomorphism theorems, it is often necessary to know the actual correspondence given by the isomorphism and not just the fact that the groups are isomorphic. The first six exercises give us training for this.

1. Let $\phi : \mathbb{Z}_{12} \rightarrow \mathbb{Z}_3$ be the homomorphism such that $\phi(1) = 2$.
 - a. Find the kernel K of ϕ .
 - b. List the cosets in \mathbb{Z}_{12}/K , showing the elements in each coset.
 - c. Give the correspondence between \mathbb{Z}_{12}/K and \mathbb{Z}_3 given by the map μ described in Theorem 16.2.
2. Let $\phi : \mathbb{Z}_{18} \rightarrow \mathbb{Z}_{12}$ be the homomorphism where $\phi(1) = 10$.
 - a. Find the kernel K of ϕ .
 - b. List the cosets in \mathbb{Z}_{18}/K , showing the elements in each coset.
 - c. Find the group $\phi[\mathbb{Z}_{18}]$.
 - d. Give the correspondence between \mathbb{Z}_{18}/K and $\phi[\mathbb{Z}_{18}]$ given by the map μ described in Theorem 16.2.

3. In the group \mathbb{Z}_{24} , let $H = \langle 4 \rangle$ and $N = \langle 6 \rangle$.
 - a. List the elements in HN (which we might write $H + N$ for these additive groups) and in $H \cap N$.
 - b. List the cosets in HN/N , showing the elements in each coset.
 - c. List the cosets in $H/(H \cap N)$, showing the elements in each coset.
 - d. Give the correspondence between $H/(H \cap N)$ and HN/N described in the proof of Theorem 16.5.
4. Repeat Exercise 3 for the dihedral group D_6 with $N = \{\iota, \mu, \rho^2, \mu\rho^2, \rho^4, \mu\rho^4\}$ and $H = \langle \rho \rangle$.
5. In the group $G = \mathbb{Z}_{24}$, let $H = \langle 4 \rangle$ and $K = \langle 8 \rangle$.
 - a. List the cosets in G/H , showing the elements in each coset.
 - b. List the cosets in G/K , showing the elements in each coset.
 - c. List the cosets in H/K , showing the elements in each coset.
 - d. List the cosets in $(G/K)/(H/K)$, showing the elements in each coset.
 - e. Give the correspondence between $(G/K)/(H/K)$ and G/H described in the proof of Theorem 16.8.
6. Repeat Exercise 5 for the dihedral group $G = D_8$, $H = \langle \rho^2 \rangle = \{\iota, \rho^2, \rho^4, \rho^6\}$, and $K = \langle \rho^4 \rangle = \{\iota, \rho^4\}$.

Theory

7. Show directly from the definition of a normal subgroup that if H and N are subgroups of a group G , and N is normal in G , then $H \cap N$ is normal in H .
8. Let H, K , and L be normal subgroups of G with $H < K < L$. Let $A = G/H, B = K/H$, and $C = L/H$.
 - a. Show that B and C are normal subgroups of A , and $B < C$.
 - b. To what factor group of G is $(A/B)/(C/B)$ isomorphic?
9. Let K and L be normal subgroups of G with $K \vee L = G$, and $K \cap L = \{e\}$. Show that $G/K \cong L$ and $G/L \cong K$.
10. Use one of the Isomorphism Theorems to prove that if $a, b \in \mathbb{Z}^+$ and a divides b , then $|a\mathbb{Z}/b\mathbb{Z}| = b/a$.
11. Let G be a group with subgroups $H^* \leq H \leq G$ and $K^* \leq K \leq G$. Prove that the sets $H^*(H \cap K^*) \cap (H \cap K)$ and $(H^* \cap K)(H \cap K^*)$ are equal.

SECTION 17 SYLOW THEOREMS

The Fundamental Theorem for Finitely Generated Abelian Groups (Theorems 9.12 and 9.14) give us complete information about all finite abelian groups. The study of finite nonabelian groups is much more complicated. The Sylow theorems give us some important information about them.

The Theorem of Lagrange says that if H is a subgroup of a finite group G , then the order of H divides the order of G . The Fundamental Theorem for Finitely Generated Abelian Groups implies that if k divides the order of a finite abelian group G , then G has a subgroup of order k . The situation is different for nonabelian groups. Example 13.6 shows that although A_4 has 12 elements, it has no subgroup of order 6. Furthermore, for $n \geq 5$, A_n can have no subgroup of index 2 since A_n is simple and any subgroup of index 2 is a normal subgroup. On the positive side, Cauchy's theorem (14.20) says that if the prime number p divides the order of a group G , then G has a subgroup of order p . The Sylow theorems generalize Cauchy's theorem to show that if p^n divides the order of G , then G has a subgroup of order p^n as long as p is a prime number. Furthermore, Sylow's theorems gives us information about relationships between these subgroups of G . As we shall see, this information is very useful in the study of finite nonabelian groups.

Proofs of the Sylow theorems give us another application of action of a group on a set described in Section 14. This time, the set itself is formed from the group; in some instances the set is the group itself, sometimes it is a collection of cosets of a subgroup, and sometimes it is a collection of subgroups.

The Sylow Theorems

Let G be a group, and let \mathcal{S} be the collection of all subgroups of G . We make \mathcal{S} into a G -set by letting G act on \mathcal{S} by conjugation. That is, if $H \in \mathcal{S}$ so $H \leq G$ and $g \in G$, then g acting on H yields the conjugate subgroup gHg^{-1} . (To avoid confusion, we will never write this action as gH .) By Theorem 14.13 $G_H = \{g \in G \mid gHg^{-1} = H\}$ is a subgroup of G , which is called an isotropy subgroup. In Exercise 14 you will be asked to show directly that G_H is a subgroup of G . Since G_H consists of *all* elements of G that leave H invariant under conjugation, G_H is the largest subgroup of G having H as a normal subgroup.

17.1 Definition The subgroup G_H just discussed is the **normalizer of H in G** and will be denoted $N[H]$ from now on. ■

In the proof of the lemma that follows, we will use the fact that if H is a *finite* subgroup of a group G , then $g \in N[H]$ if $ghg^{-1} \in H$ for all $h \in H$. To see this, note that if $gh_1g^{-1} = gh_2g^{-1}$, then $h_1 = h_2$ by cancellation in the group G . Thus the conjugation map $i_g : H \rightarrow H$ given by $i_g(h) = ghg^{-1}$ is one-to-one. Because $|H|$ is finite, i_g must then map H onto H , so $gHg^{-1} = H$ and $g \in N[H]$.

17.2 Lemma Let H be a p -subgroup of a finite group G . Then

$$(N[H] : H) \equiv (G : H) \pmod{p}.$$

Proof Let \mathcal{L} be the set of left cosets of H in G , and let H act on \mathcal{L} by left translation, so that $h(xH) = (hx)H$. Then \mathcal{L} becomes an H -set. Note that $|\mathcal{L}| = (G : H)$.

Let us determine \mathcal{L}_H , that is, those left cosets that are fixed under action by all elements of H . Now $xH = h(xH)$ if and only if $H = x^{-1}hxH$, or if and only if $x^{-1}hx \in H$. Thus $xH = h(xH)$ for all $h \in H$ if and only if $x^{-1}hx = x^{-1}h(x^{-1})^{-1} \in H$ for all $h \in H$, or if and only if $x^{-1} \in N[H]$ (see the comment before the lemma), or if and only if $x \in N[H]$. Thus the left cosets in \mathcal{L}_H are those contained in $N[H]$. The number of such cosets is $(N[H] : H)$, so $|\mathcal{L}_H| = (N[H] : H)$.

Since H is a p -group, it has order a power of p . Theorem 14.9 then tells us that $|\mathcal{L}| \equiv |\mathcal{L}_H| \pmod{p}$, that is, that $(G : H) \equiv (N[H] : H) \pmod{p}$. ◆

17.3 Corollary Let H be a p -subgroup of a finite group G . If p divides $(G : H)$, then $N[H] \neq H$.

Proof It follows from Lemma 17.2 that p divides $(N[H] : H)$, which must then be different from 1. Thus $H \neq N[H]$. ◆

■ HISTORICAL NOTE

The Sylow theorems are due to the Norwegian mathematician Peter Ludvig Mejdell Sylow (1832–1918), who published them in a brief paper in 1872. Sylow stated the theorems in terms of permutation groups (since the abstract definition of a group had not yet been given). Georg Frobenius re-proved the theorems for abstract groups in 1887, even though he noted that in fact every group can be considered as a permutation group (Cayley's theorem [Theorem 8.11]). Sylow himself

immediately applied the theorems to the question of solving algebraic equations and showed that any equation whose Galois group has order a power of a prime p is solvable by radicals.

Sylow spent most of his professional life as a high school teacher in Halden, Norway, and was only appointed to a position at Christiania University in 1898. He devoted eight years of his life to the project of editing the mathematical works of his countryman Niels Henrik Abel.

We are now ready for the first of the Sylow theorems, which asserts the existence of prime-power subgroups of G for any prime power dividing $|G|$.

17.4 Theorem (First Sylow Theorem) Let G be a finite group and let $|G| = p^n m$ where $n \geq 1$ and where p does not divide m . Then

1. G contains a subgroup of order p^i for each i where $1 \leq i \leq n$,
2. Every subgroup H of G of order p^i is a normal subgroup of a subgroup of order p^{i+1} for $1 \leq i < n$.

Proof

1. We know G contains a subgroup of order p by Cauchy's theorem (Theorem 14.20). We use an induction argument and show that the existence of a subgroup of order p^i for $i < n$ implies the existence of a subgroup of order p^{i+1} . Let H be a subgroup of order p^i . Since $i < n$, we see p divides $(G : H)$. By Lemma 17.2, we then know p divides $(N[H] : H)$. Since H is a normal subgroup of $N[H]$, we can form $N[H]/H$, and we see that p divides $|N[H]/H|$. By Cauchy's theorem, the factor group $N[H]/H$ has a subgroup K , which is of order p . If $\gamma : N[H] \rightarrow N[H]/H$ is the canonical homomorphism, then $\gamma^{-1}[K] = \{x \in N[H] \mid \gamma(x) \in K\}$ is a subgroup of $N[H]$ and hence of G . This subgroup contains H and is of order p^{i+1} .
2. We repeat the construction in part 1 and note that $H < \gamma^{-1}[K] \leq N[H]$ where $|\gamma^{-1}[K]| = p^{i+1}$. Since H is normal in $N[H]$, it is of course normal in the possibly smaller group $\gamma^{-1}[K]$. ◆

17.5 Definition A Sylow p -subgroup P of a group G is a maximal p -subgroup of G , that is, a p -subgroup contained in no larger p -subgroup. ■

Let G be a finite group, where $|G| = p^n m$ as in Theorem 17.4. The theorem shows that the Sylow p -subgroups of G are precisely those subgroups of order p^n . If P is a Sylow p -subgroup, every conjugate gPg^{-1} of P is also a Sylow p -subgroup. The second Sylow theorem states that every Sylow p -subgroup can be obtained from P in this fashion; that is, any two Sylow p -subgroups are conjugate.

17.6 Theorem (Second Sylow Theorem) Let P_1 and P_2 be Sylow p -subgroups of a finite group G . Then P_1 and P_2 are conjugate subgroups of G .

Proof Here we will let one of the subgroups act on left cosets of the other, and use Theorem 14.19. Let \mathcal{L} be the collection of left cosets of P_1 , and let P_2 act on \mathcal{L} by $y(xP_1) = (yx)P_1$ for $y \in P_2$. Then \mathcal{L} is a P_2 -set. By Theorem 14.19, $|\mathcal{L}_{P_2}| \equiv |\mathcal{L}| \pmod{p}$, and $|\mathcal{L}| = (G : P_1)$ is not divisible by p , so $|\mathcal{L}_{P_2}| \neq 0$. Let $xP_1 \in \mathcal{L}_{P_2}$. Then $yxP_1 = xP_1$ for all $y \in P_2$, so $x^{-1}yxP_1 = P_1$ for all $y \in P_2$. Thus $x^{-1}yx \in P_1$ for all $y \in P_2$, so $x^{-1}P_2x \subseteq P_1$. Since $|P_1| = |P_2|$, we must have $P_1 = x^{-1}P_2x$, so P_1 and P_2 are indeed conjugate subgroups. ◆

The final Sylow theorem gives information on the number of Sylow p -subgroups.

17.7 Theorem (Third Sylow Theorem) If G is a finite group and p divides $|G|$, then the number of Sylow p -subgroups is congruent to 1 modulo p and divides $|G|$.

Proof Let P be one Sylow p -subgroup of G . Let \mathcal{S} be the set of all Sylow p -subgroups and let P act on \mathcal{S} by conjugation, so that $x \in P$ carries $T \in \mathcal{S}$ into xTx^{-1} . By Theorem 14.19, $|\mathcal{S}| \equiv |\mathcal{S}_P| \pmod{p}$. Let us find \mathcal{S}_P . If $T \in \mathcal{S}_P$, then $xTx^{-1} = T$ for all $x \in P$. Thus $P \leq N[T]$. Of course, $T \leq N[T]$ also. Since P and T are both Sylow p -subgroups of G ,

they are also Sylow p -subgroups of $N[T]$. But then they are conjugate in $N[T]$ by Theorem 17.6. Since T is a normal subgroup of $N[T]$, it is its only conjugate in $N[T]$. Thus $T = P$. Then $\mathcal{S} = \{P\}$. Since $|\mathcal{S}| \equiv |\mathcal{S}| = 1 \pmod{p}$, we see the number of Sylow p -subgroups is congruent to 1 modulo p .

Now let G act on \mathcal{S} by conjugation. Since all Sylow p -subgroups are conjugate, there is only one orbit in \mathcal{S} under G . If $P \in \mathcal{S}$, then $|\mathcal{S}| = |\text{orbit of } P| = (G : G_P)$ by Theorem 14.17. (G_P is, in fact, the normalizer of P .) But $(G : G_P)$ is a divisor of $|G|$, so the number of Sylow p -subgroups divides $|G|$. \blacklozenge

Theorem 17.7 is really a bit better than it sounds. Let $|G| = p^n m$ where the prime number p does not divide m and suppose that G contains k Sylow p -subgroups. Then Theorem 17.7 says that k is equivalent to 1 modulo p and k divides $|G|$. Since $\gcd(k, p) = 1$, k must divide m .

Applications of the Sylow Theorems

17.8 Example The Sylow 2-subgroups of D_3 have order 2. Three Sylow 2-subgroups are

$$\{\iota, \mu\}, \quad \{\iota, \mu\rho\}, \quad \{\iota, \mu\rho^2\}$$

Notice that Theorem 17.7 says that the number k of Sylow 2-subgroups must be odd and k must divide 6. However, by the observation above, k must divide 3. So in fact, the three subgroups listed are all three of the subgroups of D_3 having order 2. \blacktriangle

17.9 Lemma Let G be a group containing normal subgroups H and K such that $H \cap K = \{e\}$ and $H \vee K = G$. Then G is isomorphic to $H \times K$. \blacklozenge

Proof We start by showing that $hk = kh$ for $k \in K$ and $h \in H$. Consider the commutator $hkh^{-1}k^{-1} = (hkh^{-1})k^{-1} = h(kh^{-1}k^{-1})$. Since H and K are normal subgroups of G , the two groupings with parentheses show that $hkh^{-1}k^{-1}$ is in both K and H . Since $K \cap H = \{e\}$, we see that $hkh^{-1}k^{-1} = e$, so $hk = kh$.

Let $\phi : H \times K \rightarrow G$ be defined by $\phi(h, k) = hk$. Then

$$\begin{aligned} \phi((h, k)(h', k')) &= \phi(hh', kk') = hh'kk' \\ &= hkh'h'k' = \phi(h, k)\phi(h', k'), \end{aligned}$$

so ϕ is a homomorphism.

If $\phi(h, k) = e$, then $hk = e$, so $h = k^{-1}$, and both h and k are in $H \cap K$. Thus $h = e$, so $\text{Ker}(\phi) = \{(e, e)\}$ and ϕ is one-to-one.

By Lemma 16.4, we know that $HK = H \vee K$, and $H \vee K = G$ by hypothesis. Thus ϕ is onto G , and $H \times K \cong G$. \blacklozenge

We turn now to a discussion of whether there exist simple groups of certain orders. We have seen that every group of prime order is simple. We also asserted that A_n is simple for $n \geq 5$ and that A_5 is the smallest simple group that is not of prime order. There was a famous conjecture of Burnside that every finite simple group of nonprime order must be of even order. It was a triumph when this was proved by Thompson and Feit [21].

17.10 Theorem If p and q are distinct primes with $p < q$, then every group G of order pq has a single subgroup of order q and this subgroup is normal in G . Hence G is not simple. If q is not congruent to 1 modulo p , then G is abelian and cyclic.

Proof Theorems 17.4 and 17.7 tell us that G has a Sylow q -subgroup and that the number of such subgroups is congruent to 1 modulo q and divides pq , and therefore must divide p . Since $p < q$, the only possibility is the number 1. Thus there is only one Sylow

q -subgroup Q of G . This group Q must be normal in G , for under an inner automorphism it would be carried into a group of the same order, hence itself. Thus G is not simple.

Likewise, there is a Sylow p -subgroup P of G , and the number of these divides pq and is congruent to 1 modulo p . This number must be either 1 or q . If q is not congruent to 1 modulo p , then the number must be 1 and P is normal in G . Let us assume that $q \not\equiv 1 \pmod{p}$. Since every element in Q other than e is of order q and every element in P other than e is of order p , we have $Q \cap P = \{e\}$. Also $Q \vee P$ must be a subgroup of G properly containing Q and of order dividing pq . Hence $Q \vee P = G$ and by Lemma 17.9 is isomorphic to $Q \times P$ or $\mathbb{Z}_q \times \mathbb{Z}_p$. Thus G is abelian and cyclic. \blacklozenge

- 17.11 Example** Recall that if p is a prime number, then up to isomorphism there is only one group of order p and it is cyclic. Theorem 17.10 shows that there are many nonprime numbers n such that every group of order n is cyclic. Since 5 is not equivalent to 1 modulo 3, by Theorem 17.10, every group of order 15 is cyclic. Exercise 33 shows that 15 is the smallest composite number with this property. \blacktriangle

We need another lemma for some of the counting arguments that follow.

- 17.12 Lemma** If H and K are finite subgroups of a group G , then

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

Proof Let

$$h_1(H \cap K), h_2(H \cap K), h_3(H \cap K), \dots, h_r(H \cap K)$$

be the left cosets of $H \cap K$ in H with each coset listed exactly once. We let

$$S = \{h_1, h_2, h_3, \dots, h_r\},$$

which includes exactly one element from each left coset of $H \cap K$ in H . So

$$|S| = \frac{|H|}{|H \cap K|}.$$

Let $f : S \times K \rightarrow HK$ be defined by $f(h_i, k) = h_i k$. We show that f is one-to-one and onto.

Suppose that $hk \in HK$. Then $h \in H$ is in some left coset of $H \cap K$, so $h \in h_i(H \cap K)$ for some $h_i \in S$. We have that $h = h_i x$ for some $x \in H \cap K$. Let $k_1 = xk$. Then $(h_i, k_1) \in S \times K$ and

$$f(h_i, k_1) = h_i k_1 = h_i xk = hk.$$

Thus f is onto.

We now show that f is one-to-one. Suppose that $f(h_i, k) = f(h_j, k_1)$. So $h_i k = h_j k_1$. Then $h_j^{-1} h_i = k_1 k^{-1} \in H \cap K$. But this implies that h_i and h_j are in the same left coset of $H \cap K$, so $h_i = h_j$. By cancellation, $k = k_1$ and f is one-to-one.

Since there is a one-to-one and onto function $f : S \times K \rightarrow HK$, we have

$$\begin{aligned} |HK| &= |S||K| \\ &= \frac{|H|}{|H \cap K|} \cdot |K| \\ &= \frac{(|H|)(|K|)}{|H \cap K|}. \end{aligned}$$



Lemma 17.12 is another result that counts something, so do not underestimate it. The lemma will be used in the following way: A finite group G cannot have subgroups H and K that are too large with intersections that are too small, or the order of HK would have to exceed the order of G , which is impossible. For example, a group of order 24 cannot have two subgroups of orders 12 and 8 with an intersection of order 2.

The remainder of this section consists of several examples illustrating techniques of proving that all groups of certain orders are abelian or that they have nontrivial proper normal subgroups, that is, that they are not simple. We recall that a subgroup H of index 2 in a finite group G is a normal subgroup. This is because the two left cosets of H in G are H and the set of all elements in G that are not in H . But these are also the right cosets, which says that H is a normal subgroup of G .

17.13 Example No group of order p^r for $r > 1$ is simple, where p is a prime. For by Theorem 17.4 such a group G contains a subgroup of order p^{r-1} normal in a subgroup of order p^r , which must be all of G . Thus a group of order 16 is not simple; it has a normal subgroup of order 8. ▲

17.14 Example No group of order 20 is simple, for such a group G contains Sylow 5-subgroups in number congruent to 1 modulo 5 and a divisor of 4, hence only 1. This Sylow 5-subgroup is then normal, since all conjugates of it must be itself. ▲

17.15 Example No group of order 30 is simple. We have seen that if there is only one Sylow p -subgroup for some prime p dividing 30, we are done. By Theorem 17.7 the possibilities for the number of Sylow 5-subgroups are 1 or 6, and those for Sylow 3-subgroups are 1 or 10. But if G has six Sylow 5-subgroups, then the intersection of any two is a subgroup of each of order dividing 5, and hence just $\{e\}$. Thus each contains 4 elements of order 5 that are in none of the others. Hence G must contain 24 elements of order 5. Similarly, if G has 10 Sylow 3-subgroups, it has at least 20 elements of order 3. The two types of Sylow subgroups together would require at least 44 elements in G . Thus there is a normal subgroup either of order 5 or of order 3. ▲

17.16 Example No group of order 48 is simple. Indeed, we shall show that a group G of order 48 has a normal subgroup of either order 16 or order 8. By Theorem 17.7 G has either one or three Sylow 2-subgroups of order 16. If there is only one subgroup of order 16, it is normal in G , by now a familiar argument.

Suppose that there are three subgroups of order 16, and let H and K be two of them. Then $H \cap K$ must be of order 8, for if $H \cap K$ were of order ≤ 4 , then by Lemma 17.12 HK would have at least $(16)(16)/4 = 64$ elements, contradicting the fact that G has only 48 elements. Therefore, $H \cap K$ is normal in both H and K (being of index 2, or by Theorem 17.4). Hence the normalizer of $H \cap K$ contains both H and K and must have order a multiple > 1 of 16 and a divisor of 48, therefore 48. Thus $H \cap K$ must be normal in G . ▲

17.17 Example No group of order 36 is simple. Such a group G has either one or four subgroups of order 9. If there is only one such subgroup, it is normal in G . If there are four such subgroups, let H and K be two of them. As in Example 17.16, $H \cap K$ must have at least 3 elements, or HK would have to have 81 elements, which is impossible. Thus the normalizer of $H \cap K$ has as order a multiple of > 1 of 9 and a divisor of 36; hence the order must be either 18 or 36. If the order is 18, the normalizer is then of index 2 and therefore is normal in G . If the order is 36, then $H \cap K$ is normal in G . ▲

17.18 Example We show every group of order $255 = (3)(5)(17)$ is abelian (hence cyclic by the Fundamental Theorem 9.12 and not simple, since 255 is not a prime). By Theorem 17.7 such a group G has only one subgroup H of order 17. Then G/H has order 15 and is abelian

by Theorem 17.10. By Theorem 13.22, we see that the commutator subgroup C of G is contained in H . Thus as a subgroup of H , C has either order 1 or 17. Theorem 17.7 also shows that G has either 1 or 85 subgroups of order 3 and either 1 or 51 subgroups of order 5. However, 85 subgroups of order 3 would require 170 elements of order 3, and 51 subgroups of order 5 would require 204 elements of order 5 in G ; both together would then require 375 elements in G , which is impossible. Hence there is a subgroup K having either order 3 or order 5 and normal in G . Then G/K has either order $(5)(17)$ or order $(3)(17)$, and in either case Theorem 17.10 shows that G/K is abelian. Thus $C \leq K$ and has order either 3, 5, or 1. Since $C \leq H$ showed that C has order 17 or 1, we conclude that C has order 1. Hence $C = \{e\}$, and $G/C \cong G$ is abelian. The Fundamental Theorem 9.12 then shows that G is cyclic. \blacktriangle

■ EXERCISES 17

Computations

In Exercises 1 through 4, determine the values of n_i that make each statement true.

1. A Sylow 3-subgroup of a group of order 12 has order n_1 .
2. A Sylow 3-subgroup of a group of order 54 has order n_1 .
3. A group of order 24 must have either n_1 or n_2 Sylow 2-subgroups. (Use only the information given in Theorem 17.7.)
4. A group of order $255 = (3)(5)(17)$ must have either n_1 or n_2 Sylow 3-subgroups and n_3 or n_4 Sylow 5-subgroups. (Use only the information given in Theorem 17.7.)
5. Find all Sylow 3-subgroups of S_4 and demonstrate that they are all conjugate.
6. Find two Sylow 2-subgroups of S_4 and show that they are conjugate.
7. Determine for which $n \leq 20$ any group of order n is abelian.
8. Determine for which $n \leq 20$ any group of order n is cyclic.

Concepts

In Exercises 9 through 11, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

9. Let p be a prime. A *p -group* is a group with the property that every element has order p .
10. The *normalizer* $N[H]$ of a subgroup H of a group G is the set of all inner automorphisms that carry H onto itself.
11. Let G be a group whose order is divisible by a prime p . The *Sylow p -subgroup* of a group is the largest subgroup P of G with the property that P has some power of p as its order.
12. Determine whether each of the following is true or false.
 - a. Any two Sylow p -subgroups of a finite group are conjugate.
 - b. Theorem 17.7 shows that a group of order 15 has only one Sylow 5-subgroup.
 - c. Every Sylow p -subgroup of a finite group has order a power of p .
 - d. Every p -subgroup of every finite group is a Sylow p -subgroup.
 - e. Every finite abelian group has exactly one Sylow p -subgroup for each prime p dividing the order of G .
 - f. The normalizer in G of a subgroup H of G is always a normal subgroup of G .
 - g. If H is a subgroup of G , then H is always a normal subgroup of $N[H]$.
 - h. A Sylow p -subgroup of a finite group G is normal in G if and only if it is the only Sylow p -subgroup of G .
 - i. If G is an abelian group and H is a subgroup of G , then $N[H] = H$.
 - j. A group of prime-power order p^n has no Sylow p -subgroup.

13. Determine whether each of the following is true or false.
- Every group of order 159 is cyclic.
 - Every group of order 102 has a nontrivial proper normal subgroup.
 - Every group of order p^3 is abelian, assuming that p is a prime number.
 - There is a simple group of order 1128.
 - It would be quite tedious to show that no group of nonprime order between 60 and 168 is simple by the methods illustrated in the text.
 - No group of order 21 is simple.
 - Every group of 125 elements has at least 5 elements that commute with every element in the group.
 - Every group of order 42 has a normal subgroup of order 7.
 - Every group of order 42 has a normal subgroup of order 3.
 - The only simple groups are the groups \mathbb{Z}_p and A_n , where p is a prime and $n > 4$.

Theory

- Let H be a subgroup of a group G . Show that $G_H = \{g \in G \mid gHg^{-1} = H\}$ is a subgroup of G without using Theorem 14.13.
- Let G be a finite group and let primes p and $q \neq p$ divide $|G|$. Prove that if G has precisely one proper Sylow p -subgroup, it is a normal subgroup, so G is not simple.
- Show that every group of order 45 has a normal subgroup of order 9.
- Let G be a finite group and let p be a prime dividing $|G|$. Let P be a Sylow p -subgroup of G . Show that $N[N[P]] = N[P]$. [Hint: Argue that P is the only Sylow p -subgroup of $N[N[P]]$, and use Theorem 17.6.]
- Let G be a finite group and let a prime p divide $|G|$. Let P be a Sylow p -subgroup of G and let H be any p -subgroup of G . Show there exists $g \in G$ such that $gHg^{-1} \leq P$.
- Show that every group of $(35)^3$ has a normal subgroup of order 125.
- Show that there are no simple groups of order $255 = (3)(5)(17)$.
- Show that there are no simple groups of order $p^r m$, where p is a prime, r is a positive integer, and $m < p$.
- Prove that all simple groups of order at most 20 are cyclic.
- Let p be a prime. Show that a finite group of order p^n contains *normal* subgroups H_i for $0 \leq i \leq n$ such that $|H_i| = p^i$ and $H_i < H_{i+1}$ for $0 \leq i < n$. [Hint: See Theorem 14.24.]
- Let G be a finite group and let P be a normal p -subgroup of G . Show that P is contained in every Sylow p -subgroup of G .
- Prove that if $p \geq 3$ is a prime number and $k \geq 1$, then any group G of order $2p^k$ is not simple.
- Prove that every group of order $(5)(7)(47)$ is abelian and cyclic.
- Prove that no group of order 96 is simple.
- Show that every group of order 30 contains a subgroup of order 15. [Hint: Use the last sentence in Example 17.15 and go to the factor group.]
- Prove that no group of order 160 is simple.
- Let G be a finite group and suppose that for each k that divides $|G|$, G has at most one subgroup of order k . Prove that G is cyclic.
- Let G be a finite group. Use the group action of G on G given by conjugation, $g * x = gxg^{-1}$, to prove the formula $|G| = |Z(G)| + n_1 + n_2 + \dots + n_k$, where $Z(G)$ is the center of G and n_1, n_2, \dots, n_k are the orbit sizes for the orbits containing at least two elements. This formula is called the **class equation**.
- By arguments similar to those used in the examples of this section, convince yourself that the only simple groups of order less than 60 are cyclic. You need not write out all the details.
- Show that for every positive integer $n < 15$, if every group of order n is cyclic, then n is prime.

SECTION 18**SERIES OF GROUPS****Subnormal and Normal Series**

This section is concerned with the notion of a *series* of a group G , which gives insight into the structure of G . The results hold for both abelian and nonabelian groups. They are not too important for finitely generated abelian groups because of the Fundamental Theorem of Finitely Generated Abelian Groups. Many of our illustrations will be taken from abelian groups, however, for ease of computation.

- 18.1 Definition** A **subnormal (or subinvariant) series of a group** G is a finite sequence H_0, H_1, \dots, H_n of subgroups of G such that $H_i < H_{i+1}$ and H_i is a normal subgroup of H_{i+1} with $H_0 = \{e\}$ and $H_n = G$. A **normal (or invariant) series** of G is a finite sequence H_0, H_1, \dots, H_n of normal subgroups of G such that $H_i < H_{i+1}$, $H_0 = \{e\}$, and $H_n = G$. ■

Note that for abelian groups the notions of subnormal and normal series coincide, since every subgroup is normal. A normal series is always subnormal, but the converse need not be true. We defined a subnormal series before a normal series, since the concept of a subnormal series is more important for our work.

- 18.2 Example** Two examples of normal series of \mathbb{Z} under addition are

$$\{0\} < 8\mathbb{Z} < 4\mathbb{Z} < \mathbb{Z}$$

and

$$\{0\} < 9\mathbb{Z} < \mathbb{Z}. \quad \blacktriangle$$

- 18.3 Example** We let $G = D_4$, the dihedral group. The series

$$\{\iota\} < \{\iota, \mu\} < \{\iota, \mu, \rho^2, \mu\rho^2\} < D_4$$

is a subnormal series since each subgroup is normal in the one to its right. The subgroup $\{\iota, \mu\}$ is not a normal subgroup of D_4 since $\rho\mu\rho^{-1} = \mu\rho^2 \notin \{\iota, \mu\}$. So this series is a subnormal series, but not a normal series. ■

- 18.4 Definition** A subnormal (normal) series $\{K_j\}$ is a **refinement of a subnormal (normal) series** $\{H_i\}$ of a group G if $\{H_i\} \subseteq \{K_j\}$, that is, if each H_i is one of the K_j . ■

- 18.5 Example** The series

$$\{0\} < 72\mathbb{Z} < 24\mathbb{Z} < 8\mathbb{Z} < 4\mathbb{Z} < \mathbb{Z}$$

is a refinement of the series

$$\{0\} < 72\mathbb{Z} < 8\mathbb{Z} < \mathbb{Z}.$$

Two new terms, $4\mathbb{Z}$ and $24\mathbb{Z}$, have been inserted. ■

Of interest in studying the structure of G are the factor groups H_{i+1}/H_i . These are defined for both normal and subnormal series, since H_i is normal in H_{i+1} in either case.

- 18.6 Definition** Two subnormal (normal) series $\{H_i\}$ and $\{K_j\}$ of the same group G are **isomorphic** if there is a one-to-one correspondence between the collections of factor groups $\{H_{i+1}/H_i\}$ and $\{K_{j+1}/K_j\}$ such that corresponding factor groups are isomorphic. ■

Clearly, two isomorphic subnormal (normal) series must have the same number of groups.

18.7 Example The two series of \mathbb{Z}_{15} ,

$$\{0\} < \langle 5 \rangle < \mathbb{Z}_{15}$$

and

$$\{0\} < \langle 3 \rangle < \mathbb{Z}_{15},$$

are isomorphic. Both $\mathbb{Z}_{15}/\langle 5 \rangle$ and $\langle 3 \rangle/\{0\}$ are isomorphic to \mathbb{Z}_5 , and $\mathbb{Z}_{15}/\langle 3 \rangle$ is isomorphic to $\langle 5 \rangle/\{0\}$, or to \mathbb{Z}_3 . ▲

The Schreier Theorem

We proceed to prove that two subnormal series of a group G have isomorphic refinements. This is a fundamental result in the theory of series. Although the proof is a little technical, it is broken up into smaller pieces that make it easier to follow. Before starting the proof, we give an example to illustrate the goal of our investigation.

18.8 Example Let us try to find isomorphic refinements of the series

$$\{0\} < 8\mathbb{Z} < 4\mathbb{Z} < \mathbb{Z}$$

and

$$\{0\} < 9\mathbb{Z} < \mathbb{Z}$$

given in Example 18.2. Consider the refinement

$$\{0\} < 72\mathbb{Z} < 8\mathbb{Z} < 4\mathbb{Z} < \mathbb{Z}$$

of $\{0\} < 8\mathbb{Z} < 4\mathbb{Z} < \mathbb{Z}$ and the refinement

$$\{0\} < 72\mathbb{Z} < 18\mathbb{Z} < 9\mathbb{Z} < \mathbb{Z}$$

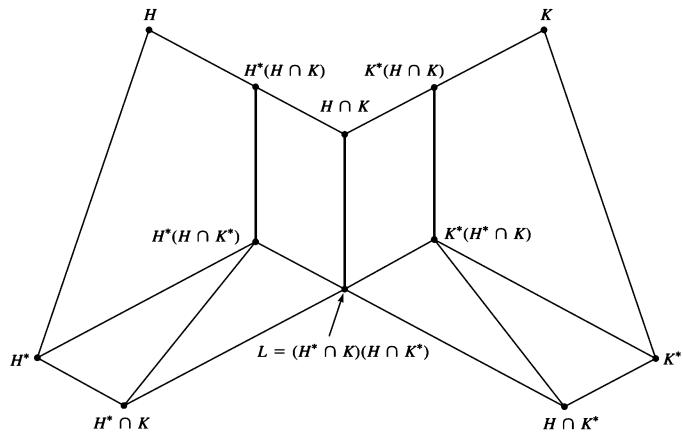
of $\{0\} < 9\mathbb{Z} < \mathbb{Z}$. In both cases the refinements have four factor groups isomorphic to $\mathbb{Z}_4, \mathbb{Z}_2, \mathbb{Z}_9$, and $72\mathbb{Z}$ or \mathbb{Z} . The *order* in which the factor groups occur is different to be sure. ▲

We start with a rather technical lemma developed by Zassenhaus. This lemma is sometimes called the *butterfly lemma*, since Fig. 18.9, which accompanies the lemma, has a butterfly shape.

18.10 Lemma (Zassenhaus Lemma) Let $H, K \leq G$ be subgroups and $H^* \trianglelefteq H$, $K^* \trianglelefteq K$ be normal subgroups of H and K , respectively. Then

1. $H^*(H \cap K^*)$ is a normal subgroup of $H^*(H \cap K)$,
2. $K^*(H^* \cap K)$ is a normal subgroup of $K^*(H \cap K)$, and
3. The factor groups $H^*(H \cap K)/H^*(H \cap K^*)$, $K^*(H \cap K)/K^*(H^* \cap K)$, and $(H \cap K)/[(H^* \cap K)(H \cap K^*)]$ are all isomorphic.

Proof It may be helpful to follow along with Figure 18.9 to visualize the subgroups that we refer to in the proof. Before beginning the proof of normality, we need to verify that the sets involved are in fact subgroups of G . All three sets H^* , $H \cap K^*$, and $H \cap K$ are subgroups of H . Furthermore, H^* is a normal subgroup of H , so by Lemma 16.4, $H^*(H \cap K^*)$ and $H^*(H \cap K)$ are subgroups of H . Thus $H^*(H \cap K^*)$ and $H^*(H \cap K)$ are also subgroups of G . Clearly, $H^*(H \cap K^*)$ is a subgroup of $H^*(H \cap K)$.



18.9 Figure

We now show that $H^*(H \cap K^*)$ is a normal subgroup of $H^*(H \cap K)$. We let $w \in H^*(H \cap K^*)$ and $y \in H^*(H \cap K)$. We need to verify that $ywy^{-1} \in H^*(H \cap K^*)$. By definition, $w = h_1x$ and $y = h_2g$ for some $h_1, h_2 \in H^*$, $x \in H \cap K^*$, and $g \in H \cap K$. We write

$$\begin{aligned} ywy^{-1} &= h_2gh_1xg^{-1}h_2^{-1} \\ &= h_2(gh_1g^{-1})gxg^{-1}h_2^{-1} \\ &= h_2h_3gxg^{-1}h_2^{-1} \end{aligned}$$

for some $h_3 \in H^*$ since H^* is a normal subgroup of H . We note that h_2^{-1} and h_2h_3 are both elements of $H^* \leq H^*(H \cap K^*)$. Furthermore $gxg^{-1} \in K^*$ since $g \in K$ and K^* is a normal subgroup of K . Also $gxg^{-1} \in H$ since both g and x are elements of H . Thus, $gxg^{-1} \in H \cap K^* \leq K^*(H \cap K^*)$. So, ywy^{-1} is the product of elements in the group $H^*(H \cap K^*)$, which implies that $ywy^{-1} \in H^*(H \cap K^*)$. Thus we have shown Part 1 of the Theorem.

We use the Second Isomorphism Theorem (16.5) to prove the third part of the Lemma. Let $N' = H^*(H \cap K^*)$ and $H' = H \cap K$. So N' is a normal subgroup of $H^*(H \cap K)$ and H' is a subgroup of $H^*(H \cap K)$. By Lemma 16.4 $N'H'$ is a group and

$$\begin{aligned} N'H' &= H^*(H \cap K^*)(H \cap K) \\ &= H^*(H \cap K). \end{aligned}$$

The Second Isomorphism Theorem says that $N'/N' \cong H'/(H' \cap N')$. We have

$$N'/N' = H^*(H \cap K)/H^*(H \cap K^*)$$

and

$$H'/(H' \cap N') = (H \cap K)/(H^*(H \cap K^*) \cap (H \cap K)).$$

Exercise 11 in Section 16 shows that

$$H^*(H \cap K^*) \cap (H \cap K) = (H^* \cap K)(H \cap K^*).$$

Thus

$$H^*(H \cap K)/(H^*(H \cap K^*)) \cong (H \cap K)/((H^* \cap K)(H \cap K^*)).$$

By reversing the roles of H and K (as well as H^* and K^*), the proof given above proves Part 2 as well as the other half of Part 3 of the Zassenhaus Lemma. \diamond

18.11 Theorem (Schreier Theorem) Two subnormal (normal) series of a group G have isomorphic refinements.

Proof Let G be a group and let

$$\{e\} = H_0 < H_1 < H_2 < \cdots < H_n = G \quad (1)$$

and

$$\{e\} = K_0 < K_1 < K_2 < \cdots < K_m = G \quad (2)$$

be two subnormal series for G . For i where $0 \leq i \leq n - 1$, form the chain of groups

$$H_i = H_i(H_{i+1} \cap K_0) \leq H_i(H_{i+1} \cap K_1) \leq \cdots \leq H_i(H_{i+1} \cap K_m) = H_{i+1}.$$

This inserts $m - 1$ not necessarily distinct groups between H_i and H_{i+1} . If we do this for each i where $0 \leq i \leq n - 1$ and let $H_{i,j} = H_i(H_{i+1} \cap K_j)$, then we obtain the chain of groups

$$\begin{aligned} \{e\} &= H_{0,0} \leq H_{0,1} \leq H_{0,2} \leq \cdots \leq H_{0,m-1} \leq H_{1,0} \\ &\leq H_{1,1} \leq H_{1,2} \leq \cdots \leq H_{1,m-1} \leq H_{2,0} \\ &\leq H_{2,1} \leq H_{2,2} \leq \cdots \leq H_{2,m-1} \leq H_{3,0} \\ &\leq \cdots \\ &\leq H_{n-1,1} \leq H_{n-1,2} \leq \cdots \leq H_{n-1,m-1} \leq H_{n-1,m} \\ &= G. \end{aligned} \quad (3)$$

This chain (3) contains $nm + 1$ not necessarily distinct groups, and $H_{i,0} = H_i$ for each i . By the Zassenhaus Lemma, chain (3) is a subnormal chain, that is, each group is normal in the following group. This chain refines the series (1).

In a symmetric fashion, we set $K_{j,i} = K_j(K_{j+1} \cap H_i)$ for $0 \leq j \leq m - 1$ and $0 \leq i \leq n$. This gives a subnormal chain

$$\begin{aligned} \{e\} &= K_{0,0} \leq K_{0,1} \leq K_{0,2} \leq \cdots \leq K_{0,n-1} \leq K_{1,0} \\ &\leq K_{1,1} \leq K_{1,2} \leq \cdots \leq K_{1,n-1} \leq K_{2,0} \\ &\leq K_{2,1} \leq K_{2,2} \leq \cdots \leq K_{2,n-1} \leq K_{3,0} \\ &\leq \cdots \\ &\leq K_{m-1,1} \leq K_{m-1,2} \leq \cdots \leq K_{m-1,n-1} \leq K_{m-1,n} \\ &= G. \end{aligned} \quad (4)$$

This chain (4) contains $mn + 1$ not necessarily distinct groups, and $K_{j,0} = K_j$ for each j . This chain refines the series (2).

By the Zassenhaus Lemma 18.10, we have

$$H_i(H_{i+1} \cap K_{j+1})/H_i(H_{i+1} \cap K_j) \cong K_j(K_{j+1} \cap H_{i+1})/K_j(K_{j+1} \cap H_i),$$

or

$$H_{i,j+1}/H_{i,j} \cong K_{j+1}/K_j \quad (5)$$

for $0 \leq i \leq n - 1$ and $0 \leq j \leq m - 1$. The isomorphisms of relation (5) give a one-to-one correspondence of isomorphic factor groups between the subnormal chains (3) and (4). To verify this correspondence, note that $H_{i,0} = H_i$ and $H_{i,m} = H_{i+1}$, while $K_{j,0} = K_j$ and $K_{j,n} = K_{j+1}$. Each chain in (3) and (4) contains a rectangular array of mn symbols \leq 's. Each \leq gives rise to a factor group. The factor groups arising from the r th row of \leq 's

in chain (3) correspond to the factor groups arising from the *rth column* of \leq 's in chain (4). Deleting repeated groups from the chains in (3) and (4), we obtain subnormal series of distinct groups that are isomorphic refinements of chains (1) and (2). This establishes the theorem for subnormal series.

For normal series, where all H_i and K_j are normal in G , we merely observe that all the groups $H_{i,j}$ and $K_{j,i}$ formed above are also normal in G , so the same proof applies. This normality of $H_{i,j}$ and $K_{j,i}$ follows at once from the second assertion in Lemma 16.4 and from the fact that intersections of normal subgroups of a group yield normal subgroups. \blacklozenge

The Jordan–Hölder Theorem

We now come to the real meat of the theory.

18.12 Definition A subnormal series $\{H_i\}$ of a group G is a **composition series** if all the factor groups H_{i+1}/H_i are simple. A normal series $\{H_i\}$ of G is a **principal** or **chief series** if all the factor groups H_{i+1}/H_i are simple. \blacksquare

Note that for abelian groups the concepts of composition and principal series coincide. Also, since every normal series is subnormal, every principal series is a composition series for any group, abelian or not.

18.13 Example We claim that \mathbb{Z} has no composition (and also no principal) series. For if

$$\{0\} = H_0 < H_1 < \cdots < H_{n-1} < H_n = \mathbb{Z}$$

is a subnormal series, H_1 must be of the form $r\mathbb{Z}$ for some $r \in \mathbb{Z}^+$. But then H_1/H_0 is isomorphic to $r\mathbb{Z}$, which is infinite cyclic with many nontrivial proper normal subgroups, for example, $2r\mathbb{Z}$. Thus \mathbb{Z} has no composition (and also no principal) series. \blacktriangle

18.14 Example The series

$$\{e\} < A_n < S_n$$

for $n \geq 5$ is a composition series (and also a principal series) of S_n , because $A_n/\{e\}$ is isomorphic to A_n , which is simple for $n \geq 5$, and S_n/A_n is isomorphic to \mathbb{Z}_2 , which is simple. Likewise, the two series given in Example 18.7 are composition series (and also principal series) of \mathbb{Z}_{15} . They are isomorphic, as shown in that example. This illustrates our main theorem, which will be stated shortly. \blacktriangle

Observe that by Theorem 13.20, H_{i+1}/H_i is simple if and only if H_i is a maximal normal subgroup of H_{i+1} . Thus for a composition series, each H_i must be a maximal normal subgroup of H_{i+1} . To form a composition series of a group G , we just hunt for a maximal normal subgroup H_{n-1} of G , then for a maximal normal subgroup H_{n-2} of H_{n-1} , and so on. If this process terminates in a finite number of steps, we have a composition series. Note that by Theorem 13.20, a composition series cannot have any further refinement. To form a principal series, we have to hunt for a maximal normal subgroup H_{n-1} of G , then for a maximal normal subgroup H_{n-2} of H_{n-1} that is also normal in G , and so on. The main theorem is as follows.

18.15 Theorem (Jordan–Hölder Theorem) Any two composition (principal) series of a group G are isomorphic.

Proof Let $\{H_i\}$ and $\{K_i\}$ be two composition (principal) series of G . By Theorem 18.11, they have isomorphic refinements. But since all factor groups are already simple,

Theorem 13.20 shows that neither series has any further refinement. Thus $\{H_i\}$ and $\{K_i\}$ must already be isomorphic. \blacklozenge

For a finite group, we should regard a composition series as a type of factorization of the group into simple factor groups, analogous to the factorization of a positive integer into primes. In both cases, the factorization is unique, up to the order of the factors.

18.16 Example We illustrate the analogy between factoring integers and composition series with an example. Let $n \in \mathbb{Z}^+$. We factor n into its prime factors $n = p_1 p_2 p_3 \dots p_k$, where the prime factors may be repeated and they are in any order. The series

$$\{0\} < \langle p_1 p_2 p_3 \dots p_{k-1} \rangle < \langle p_1 p_2 p_3 \dots p_{k-2} \rangle < \langle p_1 p_2 p_3 \dots p_{k-3} \rangle < \dots < \langle p_1 \rangle < \mathbb{Z}_n$$

is a composition series since the factor groups are isomorphic with $\mathbb{Z}_{p_k}, \mathbb{Z}_{p_{k-1}}, \mathbb{Z}_{p_{k-2}}, \dots, \mathbb{Z}_{p_1}$, which are all simple. For each choice of ordering the prime numbers p_1, p_2, \dots, p_k we get a different composition series, but they are all isomorphic since the factor groups are $\mathbb{Z}_{p_1}, \mathbb{Z}_{p_2}, \dots, \mathbb{Z}_{p_k}$ in some order. \blacktriangle

■ HISTORICAL NOTE

This first appearance of what became the Jordan–Hölder theorem occurred in 1869 in a commentary on the work of Galois by the brilliant French algebraist Camille Jordan (1838–1922). The context of its appearance is the study of permutation groups associated with the roots of polynomial equations. Jordan asserted that even though the sequence of normal subgroups G, I, J, \dots of the group of the equation is not necessarily unique, nevertheless the sequence of indices of this composition series is unique. Jordan gave a proof in his monumental 1870 *Treatise on Substitutions and Algebraic Equations*. This latter work, though restricted to

what we now call permutation groups, remained the standard treatise on group theory for many years.

The Hölder part of the theorem, that the sequence of factor groups in a composition series is unique up to order, was due to Otto Hölder (1859–1937), who played a very important role in the development of group theory once the completely abstract definition of a group had been given. Among his other contributions, he gave the first abstract definition of a “factor group” and determined the structure of all finite groups of square-free order.

18.17 Theorem If G has a composition (principal) series, and if N is a proper normal subgroup of G , then there exists a composition (principal) series containing N .

Proof The series

$$\{e\} < N < G$$

is both a subnormal and a normal series. Since G has a composition series $\{H_i\}$, then by Theorem 18.11 there is a refinement of $\{e\} < N < G$ to a subnormal series isomorphic to a refinement of $\{H_i\}$. But as a composition series, $\{H_i\}$ can have no further refinement. Thus $\{e\} < N < G$ can be refined to a subnormal series all of whose factor groups are simple, that is, to a composition series. A similar argument holds if we start with a principal series $\{K_j\}$ of G . \blacklozenge

18.18 Example A composition (and also a principal) series of $\mathbb{Z}_4 \times \mathbb{Z}_9$ containing $\langle(0, 1)\rangle$ is

$$\{(0, 0)\} < \langle(0, 3)\rangle < \langle(0, 1)\rangle < \langle 2 \rangle \times \langle 1 \rangle < \langle 1 \rangle \times \langle 1 \rangle = \mathbb{Z}_4 \times \mathbb{Z}_9. \quad \blacktriangle$$

The next definition is basic to the characterization of those polynomial equations whose solutions can be expressed in terms of radicals.

18.19 Definition A group G is **solvable** if it has a composition series $\{H_i\}$ such that all factor groups H_{i+1}/H_i are abelian. ■

By the Jordan–Hölder theorem, we see that for a solvable group, *every* composition series $\{H_i\}$ must have abelian factor groups H_{i+1}/H_i .

18.20 Example The group S_3 is solvable, because the composition series

$$\{e\} < A_3 < S_3$$

has factor groups isomorphic to \mathbb{Z}_3 and \mathbb{Z}_2 , which are abelian. The group S_5 is not solvable, for since A_5 is simple, the series

$$\{e\} < A_5 < S_5$$

is a composition series, and $A_5/\{e\}$, which is isomorphic to A_5 , is not abelian. *This group A_5 of order 60 can be shown to be the smallest group that is not solvable.* This fact is closely connected with the fact that a polynomial equation of degree 5 is not in general solvable by radicals, but a polynomial equation of degree ≤ 4 is. ▲

The Ascending Central Series

We mention one subnormal series for a group G that can be formed using centers of groups. Recall from Section 13 that the center $Z(G)$ of a group G is defined by

$$Z(G) = \{z \in G \mid zg = gz \text{ for all } g \in G\},$$

and that $Z(G)$ is a normal subgroup of G . If we have the table for a finite group G , it is easy to find the center. An element a is in the center of G if and only if the row with header a and the column with header a list the elements of G in the same order.

Now let G be a group, and let $Z(G)$ be the center of G . Since $Z(G)$ is normal in G , we can form the factor group $G/Z(G)$ and find the center $Z(G/Z(G))$ of this factor group. Since $Z(G/Z(G))$ is normal in $G/Z(G)$, if $\gamma : G \rightarrow G/Z(G)$ is the canonical map, then by Theorem 13.18, $\gamma^{-1}[Z(G/Z(G))]$ is a normal subgroup $Z_1(G)$ of G . We can then form the factor group $G/Z_1(G)$ and find its center, take $(\gamma_1)^{-1}$ of it to get $Z_2(G)$, and so on.

18.21 Definition The series

$$\{e\} \leq Z(G) \leq Z_1(G) \leq Z_2(G) \leq \dots$$

described in the preceding discussion is the **ascending central series of the group G** . ■

18.22 Example For $n \geq 3$, the center of S_n is just the identity ι . Thus the ascending central series of S_n is

$$\{\iota\} \leq \{\iota\} \leq \{\iota\} \leq \dots$$

The center of the dihedral group D_4 is $\{\iota, \rho^2\}$. The factor group $D_4/\{\iota, \rho^2\}$ has order 4, and each element has order 1 or 2, so $D_4/\{\iota, \rho^2\}$ is isomorphic with the Klein 4-group, which is abelian. Therefore the center of $D_4/\{\iota, \rho^2\}$ is the whole group, and the central series for D_4 is

$$\{\iota\} \leq \{\iota, \rho^2\} \leq D_4 \leq D_4 \leq D_4 \leq \dots$$



■ EXERCISES 18

Computations

In Exercises 1 through 5, give isomorphic refinements of the two series.

1. $\{0\} < 10\mathbb{Z} < \mathbb{Z}$ and $\{0\} < 25\mathbb{Z} < \mathbb{Z}$
2. $\{0\} < 60\mathbb{Z} < 20\mathbb{Z} < \mathbb{Z}$ and $\{0\} < 245\mathbb{Z} < 49\mathbb{Z} < \mathbb{Z}$
3. $\{0\} < \langle 9 \rangle < \mathbb{Z}_{54}$ and $\{0\} < \langle 2 \rangle < \mathbb{Z}_{54}$
4. $\{0\} < \langle 9 \rangle < \langle 3 \rangle < \mathbb{Z}_{72}$ and $\{0\} < \langle 36 \rangle < \langle 12 \rangle < \mathbb{Z}_{72}$
5. $\{(0, 0)\} < (60\mathbb{Z}) \times \mathbb{Z} < (10\mathbb{Z}) \times \mathbb{Z} < \mathbb{Z} \times \mathbb{Z}$ and $\{(0, 0)\} < \mathbb{Z} \times (80\mathbb{Z}) < \mathbb{Z} \times (20\mathbb{Z}) < \mathbb{Z} \times \mathbb{Z}$
6. Find all composition series of \mathbb{Z}_{90} and show that they are isomorphic.
7. Find all composition series of \mathbb{Z}_{48} and show that they are isomorphic.
8. Find all composition series of $\mathbb{Z}_5 \times \mathbb{Z}_5$.
9. Find all composition series of $S_3 \times \mathbb{Z}_2$.
10. Find all composition series of $\mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_7$.
11. Find the center of $S_3 \times \mathbb{Z}_4$.
12. Find the center of $S_3 \times D_4$.
13. Find the ascending central series of $S_3 \times \mathbb{Z}_4$.
14. Find the ascending central series of $S_3 \times D_4$.

Concepts

In Exercises 15 and 16, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

15. A *composition series* of a group G is a finite sequence

$$\{e\} = H_0 < H_1 < H_2 < \cdots < H_{n-1} < H_n = G$$
 of subgroups of G such that H_i is a maximal normal subgroup of H_{i+1} for $i = 0, 1, 2, \dots, n - 1$.
16. A *solvable group* is one that has a composition series of abelian groups.
17. Determine whether each of the following is true or false.
 - a. Every normal series is also subnormal.
 - b. Every subnormal series is also normal.
 - c. Every principal series is a composition series.
 - d. Every composition series is a principal series.
 - e. Every abelian group has exactly one composition series.
 - f. Every finite group has a composition series.
 - g. A group is solvable if and only if it has a composition series with simple factor groups.
 - h. S_7 is a solvable group.
 - i. The Jordan–Hölder theorem has some similarity with the Fundamental Theorem of Arithmetic, which states that every positive integer greater than 1 can be factored into a product of primes uniquely up to order.
 - j. Every finite group of prime order is solvable.
18. Find a composition series of $S_3 \times S_3$. Is $S_3 \times S_3$ solvable?
19. Is the dihedral group D_4 solvable?
20. Let G be \mathbb{Z}_{36} . Refer to the proof of Theorem 18.11. Let the subnormal series (1) be

$$\{0\} < \langle 12 \rangle < \langle 3 \rangle < \mathbb{Z}_{36}$$
 and let the subnormal series (2) be

$$\{0\} < \langle 18 \rangle < \mathbb{Z}_{36}.$$

Find chains (3) and (4) and exhibit the isomorphic factor groups as described in the proof. Write chains (3) and (4) in the rectangular array shown in the text.

21. Repeat Exercise 20 for the group \mathbb{Z}_{24} with the subnormal series (1)

$$\{0\} < \langle 12 \rangle < \langle 4 \rangle < \mathbb{Z}_{24}$$

and (2)

$$\{0\} < \langle 6 \rangle < \langle 3 \rangle < \mathbb{Z}_{24}.$$

Theory

22. Let H^* , H , and K be subgroups of G with H^* normal in H . Show that $H^* \cap K$ is normal in $H \cap K$.

23. Show that if

$$H_0 = \{e\} < H_1 < H_2 < \cdots < H_n = G$$

is a subnormal (normal) series for a group G , and if H_{i+1}/H_i is of finite order s_{i+1} , then G is of finite order $s_1 s_2 \cdots s_n$.

24. Show that an infinite abelian group can have no composition series. [Hint: Use Exercise 23, together with the fact that an infinite abelian group always has a proper nontrivial subgroup.]

25. Show that a finite direct product of solvable groups is solvable.

26. Show that if $H \trianglelefteq G$ is a normal subgroup, H is solvable, and G/H is solvable, then G is solvable.

27. Show that for $n \geq 3$, D_n is solvable.

28. Show that a subgroup K of a solvable group G is solvable. [Hint: Let $H_0 = \{e\} < H_1 < \cdots < H_n = G$ be a composition series for G . Show that the distinct groups among $K \cap H_i$ for $i = 0, \dots, n$ form a composition series for K . Observe that

$$(K \cap H_i)/(K \cap H_{i-1}) \cong [H_{i-1}(K \cap H_i)]/[H_{i-1}],$$

by Theorem 16.5, with $H = K \cap H_i$ and $N = H_{i-1}$, and that $H_{i-1}(K \cap H_i) \leq H_i$.]

29. Let $H_0 = \{e\} < H_1 < \cdots < H_n = G$ be a composition series for a group G . Let N be a normal subgroup of G , and suppose that N is a simple group. Show that the distinct groups among $H_0, H_i N$ for $i = 0, \dots, n$ also form a composition series for G . [Hint: $H_i N$ is a group by Lemma 16.4. Show that $H_{i-1} N$ is normal in $H_i N$. By Theorem 16.5

$$(H_i N)/(H_{i-1} N) \cong H_i/[H_i \cap (H_{i-1} N)],$$

and the latter group is isomorphic to

$$[H_i/H_{i-1}]/[(H_i \cap (H_{i-1} N))/H_{i-1}],$$

by Theorem 16.8. But H_i/H_{i-1} is simple.]

30. Let G be a group, and let $H_0 = \{e\} < H_1 < \cdots < H_n = G$ be a composition series for G . Let N be a normal subgroup of G , and let $\gamma : G \rightarrow G/N$ be the canonical map. Show that the distinct groups among $\gamma[H_i]$ for $i = 0, \dots, n$, form a composition series for G/N . [Hint: Observe that the map

$$\psi : H_i N \rightarrow \gamma[H_i]/\gamma[H_{i-1}]$$

defined by

$$\psi(h_i n) = \gamma(h_i n)\gamma[H_{i-1}]$$

is a homomorphism with kernel $H_{i-1} N$. By Theorem 16.2,

$$\gamma[H_i]/\gamma[H_{i-1}] \cong (H_i N)/(H_{i-1} N).$$

Proceed via Theorem 16.5, as shown in the hint for Exercise 29.]

31. Prove that a homomorphic image of a solvable group is solvable. [Hint: Apply Exercise 30 to get a composition series for the homomorphic image. The hints for Exercises 29 and 30 then show how the factor groups of this composition series in the image look.]

32. Prove that a finite p -group is solvable.

33. Prove that a group G with $2^n p^k$ elements is solvable if $p > 2^n$ is a prime.

SECTION 19

FREE ABELIAN GROUPS

In this section we introduce the concept of free abelian groups and prove some results concerning them. The section concludes with a demonstration of the Fundamental Theorem of Finitely Generated Abelian Groups (Theorem 9.12).

Free Abelian Groups

We should review the notions of a generating set for a group G and a finitely generated group, as given in Section 7. In this section we shall deal exclusively with abelian groups and use the standard additive notations as follows:

0 for the identity, + for the operation,

$$\left. \begin{aligned} na &= \underbrace{a + a + \cdots + a}_{n \text{ summands}} \\ -na &= \underbrace{(-a) + (-a) + \cdots + (-a)}_{n \text{ summands}} \end{aligned} \right\} \text{for } n \in \mathbb{Z}^+ \text{ and } a \in G.$$

$0a = 0$ for the first 0 in \mathbb{Z} and the second in G .

We shall continue to use the symbol \times for direct product of groups rather than change to direct sum notation.

Notice that $\{(1, 0), (0, 1)\}$ is a generating set for the group $\mathbb{Z} \times \mathbb{Z}$ since $(n, m) = n(1, 0) + m(0, 1)$ for any (n, m) in $\mathbb{Z} \times \mathbb{Z}$. This generating set has the property that each element of $\mathbb{Z} \times \mathbb{Z}$ can be *uniquely* expressed in the form $n(1, 0) + m(0, 1)$. That is, the coefficients n and m in \mathbb{Z} are unique.

19.1 Theorem Let X be a subset of a nonzero abelian group G . The following conditions on X are equivalent.

1. Each nonzero element a in G can be expressed *uniquely* (up to order of summands) in the form $a = n_1x_1 + n_2x_2 + \cdots + n_rx_r$, for $n_i \neq 0$ in \mathbb{Z} and distinct x_i in X .
2. X generates G , and $n_1x_1 + n_2x_2 + \cdots + n_rx_r = 0$ for $n_i \in \mathbb{Z}$ and distinct $x_i \in X$ if and only if $n_1 = n_2 = \cdots = n_r = 0$.

Proof Suppose Condition 1 is true. Since $G \neq \{0\}$, we have $X \neq \{0\}$. It follows from 1 that $0 \notin X$, for if $x_i = 0$ and $x_j \neq 0$, then $x_j = x_i + x_j$, which would contradict the uniqueness of the expression for x_j . From Condition 1, X generates G , and $n_1x_1 + n_2x_2 + \cdots + n_rx_r = 0$ if $n_1 = n_2 = \cdots = n_r = 0$. Suppose that $n_1x_1 + n_2x_2 + \cdots + n_rx_r = 0$ with some $n_i \neq 0$; by dropping terms with zero coefficients and renumbering, we can assume all $n_i \neq 0$. Then

$$\begin{aligned} x_1 &= x_1 + (n_1x_1 + n_2x_2 + \cdots + n_rx_r) \\ &= (n_1 + 1)x_1 + n_2x_2 + \cdots + n_rx_r, \end{aligned}$$

which gives two ways of writing $x_1 \neq 0$, contradicting the uniqueness assumption in Condition 1. Thus Condition 1 implies Condition 2.

We now show that Condition 2 implies Condition 1. Let $a \in G$. Since X generates G , we see a can be written in the form $a = n_1x_1 + n_2x_2 + \cdots + n_rx_r$. Suppose a has another such expression in terms of elements of X . By using some zero coefficients in the two expressions, we can assume they involve the same elements in X and are of the form

$$\begin{aligned} a &= n_1x_1 + n_2x_2 + \cdots + n_rx_r \\ a &= m_1x_1 + m_2x_2 + \cdots + m_rx_r. \end{aligned}$$

Subtracting, we obtain

$$0 = (n_1 - m_1)x_1 + (n_2 - m_2)x_2 + \cdots + (n_r - m_r)x_r,$$

so $n_i - m_i = 0$ by Condition 2, and $n_i = m_i$ for $i = 1, 2, \dots, r$. Thus the coefficients are unique. \blacklozenge

19.2 Definition An abelian group having a generating set X satisfying the conditions described in Theorem 19.1 is a **free abelian group**, and X is a **basis** for the group. \blacksquare

19.3 Example The group $\mathbb{Z} \times \mathbb{Z}$ is free abelian and $\{(1, 0), (0, 1)\}$ is a basis. Similarly, a basis for the free abelian group $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ is $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and so on. Thus finite direct products of the group \mathbb{Z} with itself are free abelian groups. \blacktriangle

19.4 Example The group \mathbb{Z}_n is not free abelian, for $nx = 0$ for every $x \in \mathbb{Z}_n$, and $n \neq 0$, which would contradict Condition 2. \blacktriangle

From Example 19.4 it seems reasonable that if G is an abelian group with a nonzero element of finite order, then G is not a free abelian group. Exercise 10 asks you to provide a proof of this fact. However, there are other obstacles that prevent an abelian group from being free. For example, no rational number other than 0 has finite order, but Exercise 13 asks for a proof that \mathbb{Q} is not a free abelian group.

Suppose a free abelian group G has a finite basis $X = \{x_1, x_2, \dots, x_r\}$. If $a \in G$ and $a \neq 0$, then a has a *unique* expression of the form

$$a = n_1x_1 + n_2x_2 + \cdots + n_rx_r \quad \text{for } n_i \in \mathbb{Z}.$$

(Note that in the preceding expression for a , we included all elements x_i of our finite basis X , as opposed to the expression for a in Condition 1 of Theorem 19.1 where the basis may be infinite. Thus in the preceding expression for a we must allow the possibility that some of the coefficients n_i are zero, whereas in Condition 1 of Theorem 19.1, we specified that each $n_i \neq 0$.)

We define

$$\phi : G \rightarrow \underbrace{\mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}}_{r \text{ factors}}$$

by $\phi(a) = (n_1, n_2, \dots, n_r)$ and $\phi(0) = (0, 0, \dots, 0)$. It is straightforward to check that ϕ is an isomorphism. We leave the details to the exercises (see Exercise 9) and state the result as a theorem.

19.5 Theorem If G is a nonzero free abelian group with a basis of r elements, then G is isomorphic to $\mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$ for r factors.

It is a fact that any two bases of a free abelian group G contain the same number of elements. We shall prove this only if G has a finite basis, although it is also true if every basis of G is infinite. The proof is really lovely; it gives an easy characterization of the number of elements in a basis in terms of the size of a factor group.

19.6 Theorem Let $G \neq \{0\}$ be a free abelian group with a finite basis. Then every basis of G is finite, and all bases of G have the same number of elements.

Proof Let G have a basis $\{x_1, x_2, \dots, x_r\}$. Then G is isomorphic to $\mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z}$ for r factors. Let $2G = \{2g \mid g \in G\}$. It is readily checked that $2G$ is a subgroup of G . Since $G \cong \mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z}$ for r factors, we have

$$\begin{aligned} G/2G &\cong (\mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z})/(2\mathbb{Z} \times 2\mathbb{Z} \times \dots \times 2\mathbb{Z}) \\ &\cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \dots \times \mathbb{Z}_2 \end{aligned}$$

for r factors. Thus $|G/2G| = 2^r$, so the number of elements in any finite basis X is $\log_2 |G/2G|$. Thus any two finite bases have the same number of elements.

It remains to show that G cannot also have an infinite basis. Let Y be any basis for G , and let $\{y_1, y_2, \dots, y_s\}$ be distinct elements in Y . Let H be the subgroup of G generated by $\{y_1, y_2, \dots, y_s\}$, and let K be the subgroup of G generated by the remaining elements of Y . It is readily checked that $G \cong H \times K$, so

$$G/2G \cong (H \times K)/(2H \times 2K) \cong (H/2H) \times (K/2K).$$

Since $|H/2H| = 2^s$, we see $|G/2G| \geq 2^s$. Since we have $|G/2G| = 2^r$, we see that $s \leq r$. Then Y cannot be an infinite set, for we could take $s > r$. \blacklozenge

19.7 Definition If G is a free abelian group, the **rank** of G is the number of elements in a basis for G . (All bases have the same number of elements.) \blacksquare

Proof of the Fundamental Theorem

We shall prove the Invariant Factor version of the Fundamental Theorem (Theorem 9.14) by showing that any finitely generated abelian group is isomorphic to a factor group of the form

$$(\mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z})/(d_1\mathbb{Z} \times d_2\mathbb{Z} \times \dots \times d_s\mathbb{Z} \times \{0\} \times \dots \times \{0\}),$$

where both “numerator” and “denominator” have n factors, and d_1 divides d_2 , which divides $d_3 \dots$, which divides d_s . The Prime Factor version, Theorem 9.12, will then follow.

To show that G is isomorphic to such a factor group, we will show that there is a homomorphism of $\mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z}$ onto G with kernel of the form $d_1\mathbb{Z} \times d_2\mathbb{Z} \times \dots \times d_s\mathbb{Z} \times \{0\} \times \dots \times \{0\}$. The result will then follow by Theorem 12.14. The theorems that follow give the details of the argument. Our purpose in these introductory paragraphs is to let us see where we are going as we read what follows.

19.8 Theorem Let G be a finitely generated abelian group with generating set $\{a_1, a_2, \dots, a_n\}$. Let

$$\phi : \underbrace{\mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z}}_{n \text{ factors}} \rightarrow G$$

be defined by $\phi(h_1, h_2, \dots, h_n) = h_1a_1 + h_2a_2 + \dots + h_na_n$. Then ϕ is a homomorphism onto G .

Proof From the meaning of $h_i a_i$ for $h_i \in \mathbb{Z}$ and $a_i \in G$, we see at once that

$$\begin{aligned} \phi[(h_1, \dots, h_n) + (k_1, \dots, k_n)] &= \phi(h_1 + k_1, \dots, h_n + k_n) \\ &= (h_1 + k_1)a_1 + \dots + (h_n + k_n)a_n \\ &= (h_1a_1 + k_1a_1) + \dots + (h_n a_n + k_n a_n) \\ &= (h_1a_1 + \dots + h_n a_n) + (k_1a_1 + \dots + k_n a_n) \\ &= \phi(k_1, \dots, k_n) + \phi(h_1, \dots, h_n). \end{aligned}$$

Since $\{a_1, \dots, a_n\}$ generates G , clearly the homomorphism ϕ is onto G . \blacklozenge

We now prove a “replacement property” that makes it possible for us to adjust a basis.

19.9 Theorem If $X = \{x_1, \dots, x_r\}$ is a basis for a free abelian group G and $t \in \mathbb{Z}$, then for $i \neq j$, the set

$$Y = \{x_1, \dots, x_{j-1}, x_j + tx_i, x_{j+1}, \dots, x_r\}$$

is also a basis for G .

Proof Since $x_j = (-t)x_i + (1)(x_j + tx_i)$, we see that x_j can be recovered from Y , which thus also generates G . Suppose

$$n_1x_1 + \dots + n_{j-1}x_{j-1} + n_j(x_j + tx_i) + n_{j+1}x_{j+1} + \dots + n_rx_r = 0.$$

Then

$$n_1x_1 + \dots + (n_i + n_jt)x_i + \dots + n_jx_j + \dots + n_rx_r = 0.$$

and since X is a basis, $n_1 = \dots = n_i = n_jt = \dots = n_j = \dots = n_r = 0$. From $n_j = 0$ and $n_i + n_jt = 0$, it follows that $n_i = 0$ also, so $n_1 = \dots = n_i = n_j = \dots = n_r = 0$, and Condition 2 of Theorem 19.1 is satisfied. Thus Y is a basis. \blacklozenge

19.10 Example A basis for $\mathbb{Z} \times \mathbb{Z}$ is $\{(1, 0), (0, 1)\}$. Another basis is $\{(1, 0), (4, 1)\}$ for $(4, 1) = 4(1, 0) + (0, 1)$. However, $\{(3, 0), (0, 1)\}$ is not a basis. For example, we cannot express $(2, 0)$ in the form $n_1(3, 0) + n_2(0, 1)$, for $n_1, n_2 \in \mathbb{Z}$. Here $(3, 0) = (1, 0) + 2(1, 0)$, and a multiple of a basis element was added to *itself*, rather than to a *different* basis element. \blacktriangle

A free abelian group G of finite rank may have many bases. We show that if $K \leq G$, then K is also free abelian with rank not exceeding that of G . Equally important, there exist bases of G and K nicely related to each other.

19.11 Theorem Let G be a nonzero free abelian group of finite rank n , and let K be a nonzero subgroup of G . Then K is free abelian of rank $s \leq n$. Furthermore, there exists a basis $\{x_1, x_2, \dots, x_n\}$ for G and positive integers, d_1, d_2, \dots, d_s where d_i divides d_{i+1} for $i = 1, \dots, s-1$, such that $\{d_1x_1, d_2x_2, \dots, d_sx_s\}$ is a basis for K .

Proof We show that K has a basis of the described form, which will show that K is free abelian of rank at most n . Suppose $Y = \{y_1, \dots, y_n\}$ is a basis for G . All nonzero elements in K can be expressed in the form

$$k_1y_1 + \dots + k_ny_n,$$

where some $|k_i|$ is nonzero. Among all bases Y for G , select one Y_1 that yields the minimal such nonzero value $|k_i|$ as all nonzero elements of K are written in terms of the basis elements in Y_1 . By renumbering the elements of Y_1 if necessary, we can assume there is $w_1 \in K$ such that

$$w_1 = d_1y_1 + k_2y_2 + \dots + k_ny_n$$

where $d_1 > 0$ and d_1 is the minimal attainable coefficient as just described. Using the division algorithm, we write $k_j = d_1q_j + r_j$ where $0 \leq r_j < d_1$ for $j = 2, \dots, n$. Then

$$w_1 = d_1(y_1 + q_2y_2 + \dots + q_ny_n) + r_2y_2 + \dots + r_ny_n. \quad (1)$$

Now let $x_1 = y_1 + q_2y_2 + \dots + q_ny_n$. By Theorem 19.9 $\{x_1, y_2, \dots, y_n\}$ is also a basis for G . From Eq. (1) and our choice of Y_1 for minimal coefficient d_1 , we see that $r_2 = \dots = r_n = 0$. Thus $d_1x_1 \in K$.

We now consider bases for G of the form $\{x_1, y_2, \dots, y_n\}$. Each element of K can be expressed in the form

$$h_1x_1 + k_2y_2 + \dots + k_ny_n.$$

Since $d_1x_1 \in K$, we can subtract a suitable multiple of d_1x_1 and then using the minimality of d_1 to see that h_1 is a multiple of d_1 , we see we actually have $k_2y_2 + \cdots + k_ny_n \in K$. Among all such bases $\{x_1, y_2, \dots, y_n\}$, we choose one Y_2 that leads to some $k_i \neq 0$ of minimal magnitude. (It is possible all k_i are always zero. In this case, K is generated by d_1x_1 and we are done.) By renumbering the elements of Y_2 we can assume that there is $w_2 \in K$ such that

$$w_2 = d_2y_2 + \cdots + k_ny_n$$

where $d_2 > 0$ and d_2 is minimal as just described. Exactly as in the preceding paragraph, we can modify our basis from $Y_2 = \{x_1, y_2, \dots, y_n\}$ to a basis $\{x_1, x_2, y_3, \dots, y_n\}$ for G where $d_1x_1 \in K$ and $d_2x_2 \in K$. Writing $d_2 = d_1q + r$ for $0 \leq r < d_1$, we see that $\{x_1 + qx_2, x_2, y_3, \dots, y_n\}$ is a basis for G , and $d_1x_1 + d_2x_2 = d_1(x_1 + qx_2) + rx_2 \in K$. By our minimal choice of d_1 , we see $r = 0$, so d_1 divides d_2 .

We now consider all bases of the form $\{x_1, x_2, y_3, \dots, y_n\}$ for G and examine elements of K of the form $k_3y_3 + \cdots + k_ny_n$. The pattern is clear. The process continues until we obtain a basis $\{x_1, x_2, \dots, x_s, y_{s+1}, \dots, y_n\}$ where the only element of K of the form $k_{s+1}y_{s+1} + \cdots + k_ny_n$ is zero, that is, all k_i are zero. We then let $x_{s+1} = y_{s+1}, \dots, x_n = y_n$ and obtain a basis for G of the form described in the statement of Theorem 19.11. ◆

We now prove the Invariant Factor version of the Fundamental Theorem, Theorem 9.11. We restate it here for easy reference.

19.12 Theorem Every finitely generated abelian group is isomorphic to a group of the form

$$\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_r} \times \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z},$$

where m_i divides m_{i+1} for $i = 1, \dots, r - 1$.

Furthermore, this representation is unique up to order of the factors.

Proof For the purposes of this proof, it will be convenient to use as notations $\mathbb{Z}/1\mathbb{Z} = \mathbb{Z}/\mathbb{Z} \simeq \mathbb{Z}_1 = \{0\}$. Let G be finitely generated by n elements. Let $F = \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$ for n factors. Consider the homomorphism $\phi : F \rightarrow G$ of Theorem 19.8, and let K be the kernel of this homomorphism. Then there is a basis for F of the form $\{x_1, \dots, x_n\}$, where $\{d_1x_1, \dots, d_nx_n\}$ is a basis for K and d_i divides d_{i+1} for $i = 1, \dots, s - 1$. By Theorem 12.14, G is isomorphic to F/K . But

$$\begin{aligned} F/K &\simeq (\mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z})/(d_1\mathbb{Z} \times d_2\mathbb{Z} \times \cdots \times d_s\mathbb{Z} \times \{0\} \times \cdots \times \{0\}) \\ &\simeq \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \cdots \times \mathbb{Z}_{d_s} \times \mathbb{Z} \times \cdots \times \mathbb{Z}. \end{aligned}$$

It is possible that $d_1 = 1$, in which case $\mathbb{Z}_{d_1} = \{0\}$ and can be dropped (up to isomorphism) from this product. Similarly, d_2 may be 1, and so on. We let m_1 be the first $d_i > 1$, m_2 be the next d_i , and so on, and our theorem follows at once.

We have demonstrated the toughest part of the Fundamental Theorem. Of course, a prime-power decomposition exists since we can break the groups \mathbb{Z}_{m_i} into prime-power factors. The only remaining part of Theorem 9.12 concerns the uniqueness of the Betti number, of the torsion coefficients, and of the prime powers. The Betti number appears as the rank of the free abelian group G/T , where T is the torsion subgroup of G . This rank is invariant by Theorem 19.6, which shows the uniqueness of the Betti number. The uniqueness of the torsion coefficients and of prime powers is a bit more difficult to show. We give some exercises that indicate their uniqueness (see Exercises 14 through 22). ◆

■ EXERCISES 19

Computations

1. Find a basis $\{(a_1, a_2, a_3), (b_1, b_2, b_3), (c_1, c_2, c_3)\}$ for $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ with all $a_i \neq 0$, all $b_i \neq 0$, and all $c_i \neq 0$. (Many answers are possible.)
2. Is $\{(2, 1), (3, 1)\}$ a basis for $\mathbb{Z} \times \mathbb{Z}$? Prove your assertion.
3. Is $\{(2, 1), (4, 1)\}$ a basis for $\mathbb{Z} \times \mathbb{Z}$? Prove your assertion.
4. Find conditions on $a, b, c, d \in \mathbb{Z}$ for $\{(a, b), (c, d)\}$ to be a basis for $\mathbb{Z} \times \mathbb{Z}$. [Hint: Solve $x(a, b) + y(c, d) = (e, f)$ in \mathbb{R} , and see when the x and y lie in \mathbb{Z} .]

Concepts

In Exercises 5 and 6, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

5. The *rank* of a free abelian group G is the number of elements in a generating set for G .
6. A *basis* for a nonzero abelian group G is a generating set $X \subseteq G$ such that $n_1x_1 + n_2x_2 + \cdots + n_mx_m = 0$ for distinct $x_i \in X$ and $n_i \in \mathbb{Z}$ only if $n_1 = n_2 = \cdots = n_m = 0$.
7. Show by example that it is possible for a proper subgroup of a free abelian group of finite rank r also to have rank r .
8. Determine whether each of the following is true or false.
 - a. Every free abelian group is torsion free.
 - b. Every finitely generated torsion-free abelian group is a free abelian group.
 - c. There exists a free abelian group of every positive integer rank.
 - d. A finitely generated abelian group is free abelian if its Betti number equals the number of elements in some generating set.
 - e. If X generates a free abelian group G and $X \subseteq Y \subseteq G$, then Y generates G .
 - f. If X is a basis for a free abelian group G and $X \subseteq Y \subseteq G$, then Y is a basis for G .
 - g. Every nonzero free abelian group has an infinite number of bases.
 - h. Every free abelian group of rank at least 2 has an infinite number of bases.
 - i. If K is a nonzero subgroup of a finitely generated free abelian group, then K is free abelian.
 - j. If K is a nonzero subgroup of a finitely generated free abelian group, then G/K is free abelian.

Theory

9. Complete the proof of Theorem 19.5 (See the two sentences preceding the theorem).
10. Show that a free abelian group contains no nonzero elements of finite order.
11. Show that if G and G' are free abelian groups, then $G \times G'$ is free abelian.
12. Show that free abelian groups of finite rank are precisely the finitely generated abelian groups containing no nonzero elements of finite order.
13. Show that \mathbb{Q} under addition is not a free abelian group.

Exercises 14 through 19 deal with showing the uniqueness of the prime powers appearing in the prime-power decomposition of the torsion subgroup T of a finitely generated abelian group.

14. Let p be a fixed prime. Show that the elements of T having as order some power of p , together with zero, form a subgroup T_p of T .
15. Show that in any prime-power decomposition of T , the subgroup T_p in the preceding exercise is isomorphic to the direct product of those cyclic factors of order some power of the prime p . [This reduces our problem to showing that the group T_p cannot have essentially different decompositions into products of cyclic groups.]
16. Let G be any abelian group and let n be any positive integer. Show that $G[n] = \{x \in G \mid nx = 0\}$ is a subgroup of G . (In multiplicative notation, $G[n] = \{x \in G \mid x^n = e\}$.)

17. Referring to Exercise 16, show that $\mathbb{Z}_{p^r}[p] \cong \mathbb{Z}_p$ for any $r \geq 1$ and prime p .
 18. Using Exercise 17, show that

$$(\mathbb{Z}_{p^{r_1}} \times \mathbb{Z}_{p^{r_2}} \times \cdots \times \mathbb{Z}_{p^{r_m}})[p] \cong \underbrace{\mathbb{Z}_p \times \mathbb{Z}_p \times \cdots \times \mathbb{Z}_p}_{m \text{ factors}}$$

provided each $r_i \geq 1$.

19. Let G be a finitely generated abelian group and T_p the subgroup defined in Exercise 14. Suppose $T_p \cong \mathbb{Z}_{p^{r_1}} \times \mathbb{Z}_{p^{r_2}} \times \cdots \times \mathbb{Z}_{p^{r_m}} \cong \mathbb{Z}_{p^{s_1}} \times \mathbb{Z}_{p^{s_2}} \times \cdots \times \mathbb{Z}_{p^{s_n}}$, where $1 \leq r_1 \leq r_2 \leq \cdots \leq r_m$ and $1 \leq s_1 \leq s_2 \leq \cdots \leq s_n$. We need to show that $m = n$ and $r_i = s_i$ for $i = 1, \dots, n$ to complete the demonstration of uniqueness of the prime-power decomposition.
- a. Use Exercise 18 to show that $n = m$.
 - b. Show $r_1 = s_1$. Suppose $r_i = s_i$ for all $i < j$. Show $r_j = s_j$, which will complete the proof. [Hint: Suppose $r_j < s_j$. Consider the subgroup $p^{r_j}T_p = \{p^{r_j}x \mid x \in T_p\}$, and show that this subgroup would then have two prime-power decompositions involving different numbers of nonzero factors. Then argue that this is impossible by part (a) of this exercise.]
- Let T be the torsion subgroup of a finitely generated abelian group. Suppose $T \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_r} \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_s}$, where m_i divides m_{i+1} for $i = 1, \dots, r-1$, and n_j divides n_{j+1} for $j = 1, \dots, s-1$, and $m_1 > 1$ and $n_1 > 1$. We wish to show that $r = s$ and $m_k = n_k$ for $k = 1, \dots, r$, demonstrating the uniqueness of the torsion coefficients. This is done in Exercises 20 through 22.
20. Indicate how a prime-power decomposition can be obtained from a torsion-coefficient decomposition. (Observe that the preceding exercises show the prime powers obtained are unique.)
21. Argue from Exercise 20 that m_r and n_s can both be characterized as follows. Let p_1, \dots, p_t be the distinct primes dividing $|T|$, and let $p_1^{h_1}, \dots, p_t^{h_t}$ be the highest powers of these primes appearing in the (unique) prime-power decomposition. Then $m_r = n_s = p_1^{h_1} p_2^{h_2} \cdots p_t^{h_t}$.
22. Characterize m_{r-1} and n_{s-1} , showing that they are equal, and continue to show $m_{r-i} = n_{s-i}$ for $i = 1, \dots, r-1$, and then $r = s$.

SECTION 20 FREE GROUPS

For any group with elements a and b we have certain relations that a and b must satisfy simply because they are elements of a group. For example, $a^n a^m = a^{n+m}$ and $(ab)^{-1} = b^{-1}a^{-1}$. For most of the groups we have studied so far there are relations among the elements other than the relations that all groups possess. For example, the elements μ and ρ in the dihedral group D_n satisfy relations $\mu\rho = \mu\rho^{-1}$ and $\mu^2 = \rho^n = \iota$. In this section, we construct free groups that have only the relations that are required in the definition of a group. These groups and their factor groups as described in Section 21 are of great interest in the study of algebra and topology.

Words and Reduced Words

Let $A \neq \emptyset$ be any (not necessarily finite) set of elements a_i for $i \in I$. We think of A as an **alphabet** and of the a_i as **letters** in the alphabet. Any symbol of the form a_i^n with $n \in \mathbb{Z}$ is a **syllable** and a finite string w of syllables written in juxtaposition is a **word**. We also introduce the **empty word** 1 , which has no syllables.

20.1 Example Let $A = \{a_1, a_2, a_3\}$. Then

$$a_1 a_3^{-4} a_2^2 a_3, \quad a_2^3 a_2^{-1} a_3 a_1^2 a_1^{-7}, \quad \text{and} \quad a_3^2$$

are all words, if we follow the convention of understanding that a_i^1 is the same as a_i . ▲

There are two natural types of modifications of certain words, the **elementary contractions**. The first type consists of replacing an occurrence of $a_i^m a_i^n$ in a word by a_i^{m+n} . The second type consists of replacing an occurrence of a_i^0 in a word by 1, that is, dropping it out of the word. By means of a finite number of elementary contractions, every word can be changed to a **reduced word**, one for which no more elementary contractions are possible. Note that these elementary contractions formally amount to the usual manipulations of integer exponents and would have to be satisfied if we wish for the letters to be elements of a group.

20.2 Example The reduced form of the word $a_2^3 a_2^{-1} a_3 a_1^2 a_1^{-7}$ of Example 20.1 is $a_2^2 a_3 a_1^{-5}$. ▲

It should be said here once and for all that we are going to gloss over several points that some books spend pages proving, usually by complicated induction arguments broken down into many cases. For example, suppose we are given a word and wish to find its reduced form. There may be a variety of elementary contractions that could be performed first. How do we know that the reduced word we end up with is the same no matter in what order we perform the elementary contractions? The student will probably say this is obvious. Some authors spend considerable effort proving this. The authors agree here with the student. Proofs of this sort we regard as tedious, and they have never made us more comfortable about the situation. However, the authors are the first to acknowledge that we are not great mathematicians. In deference to the fact that many mathematicians feel that these things do need considerable discussion, we shall mark an occasion when we just state such facts by the phrase, “It would seem obvious that,” keeping the quotation marks.

Free Groups

Let the set of all reduced words formed from our alphabet A be $F[A]$. We now make $F[A]$ into a group in a natural way. For w_1 and w_2 in $F[A]$, define $w_1 \cdot w_2$ to be the reduced form of the word obtained by the juxtaposition $w_1 w_2$ of the two words.

20.3 Example If

$$w_1 = a_2^3 a_1^{-5} a_3^2$$

and

$$w_2 = a_3^{-2} a_1^2 a_3 a_2^{-2},$$

$$\text{then } w_1 \cdot w_2 = a_2^3 a_1^{-3} a_3 a_2^{-2}.$$

▲

“It would seem obvious that” this operation of multiplication on $F[A]$ is well defined and associative. The empty word 1 acts as an identity element. “It would seem obvious that” given a reduced word $w \in F[A]$, if we form the word obtained by first writing the syllables of w in the opposite order and second by replacing each a_i^n by a_i^{-n} , then the resulting word w^{-1} is a reduced word also, and

$$w \cdot w^{-1} = w^{-1} \cdot w = 1.$$

20.4 Definition The group $F[A]$ just described is the **free group generated by A** . ■

Look back at Theorem 7.7 and the definition preceding it to see that the present use of the term *generated* is consistent with the earlier use.

Starting with a group G and a generating set $\{a_i \mid i \in I\}$, which we will abbreviate by $\{a_i\}$, we might ask if G is *free* on $\{a_i\}$, that is, if G is essentially the free group generated by $\{a_i\}$. We define precisely what this is to mean.

20.5 Definition If G is a group with a set $A = \{a_i\}$ of generators, and if G is isomorphic to $F[A]$ under a map $\phi : G \rightarrow F[A]$ such that $\phi(a_i) = a_i$, then G is **free on A** , and the a_i are **free generators of G** . A group is **free** if it is free on some nonempty set A . ■

20.6 Example The only example of a free group that has occurred before is \mathbb{Z} , which is free on one generator. Note that every free group is infinite, for it contains a subgroup isomorphic with \mathbb{Z} . ▲

Refer to the literature for proofs of the next three theorems. We will not use these results. They are stated simply to inform us of these interesting facts.

20.7 Theorem If a group G is free on A and also on B , then the sets A and B have the same number of elements; that is, any two sets of free generators of a free group have the same cardinality.

20.8 Definition If G is free on A , the number of elements in A is the **rank of the free group G** . ■

Actually, the next theorem is quite evident from Theorem 20.7.

20.9 Theorem Two free groups are isomorphic if and only if they have the same rank.

20.10 Theorem A nontrivial subgroup of a free group is free.

20.11 Example Let $F[\{x, y\}]$ be the free group on $\{x, y\}$. Let

$$y_k = x^k y x^{-k}$$

for $k \geq 0$. The y_k for $k \geq 0$ are free generators for the subgroup of $F[\{x, y\}]$ that they generate. This illustrates the bizarre fact that although a subgroup of a free group is free, the rank of the subgroup may be much greater than the rank of the whole group! ▲

Homomorphisms of Free Groups

Our work in this section will be concerned primarily with homomorphisms defined on a free group. The results here are simple and elegant.

20.12 Theorem Let G be generated by $A = \{a_i \mid i \in I\}$ and let G' be any group. If a'_i for $i \in I$ are any elements in G' , not necessarily distinct, then there is at most one homomorphism $\phi : G \rightarrow G'$ such that $\phi(a_i) = a'_i$. If G is free on A , then there is exactly one such homomorphism.

Proof Let ϕ be a homomorphism from G into G' such that $\phi(a_i) = a'_i$. Now by Theorem 7.7, for any $x \in G$ we have

$$x = \prod_j a_i^{n_j}$$

for some finite product of the generators a_i , where the a_i appearing in the product need not be distinct. Then since ϕ is a homomorphism, we must have

$$\phi(x) = \prod_j \phi(a_i^{n_j}) = \prod_j (a_i')^{n_j}.$$

Thus a homomorphism is completely determined by its values on elements of a generating set. This shows that there is at most one homomorphism such that $\phi(a_i) = a'_i$.

Now suppose G is free on A ; that is, $G = F[A]$. For

$$x = \prod_j a_i^{\eta_j}$$

in G , define $\psi : G \rightarrow G'$ by

$$\psi(x) = \prod_j (a_i')^{\eta_j}.$$

The map is well defined, since $F[A]$ consists precisely of reduced words; no two different formal products in $F[A]$ are equal. Since the rules for computation involving exponents in G' are formally the same as those involving exponents in G , it is clear that $\psi(xy) = \psi(x)\psi(y)$ for any elements x and y in G , so ψ is indeed a homomorphism. ◆

Perhaps we should have proved the first part of Theorem 20.12 earlier, rather than having relegated it to the exercises. Note that the theorem states that *a homomorphism of a group is completely determined if we know its value on each element of a generating set*. In particular, a homomorphism of a cyclic group is completely determined by its value on any single generator of the group.

20.13 Theorem Every group G' is a homomorphic image of a free group G .

Proof Let $G' = \{a'_i \mid i \in I\}$, and let $A = \{a_i \mid i \in I\}$ be a set with the same number of elements as G' . Let $G = F[A]$. Then by Theorem 20.12 there exists a homomorphism ψ mapping G into G' such that $\psi(a_i) = a'_i$. Clearly the image of G under ψ is all of G' . ◆

Another Look at Free Abelian Groups

It is important that we do not confuse the notion of a free group with the notion of a free abelian group. A free group on more than one generator is not abelian. In the preceding section, we defined a free abelian group as an abelian group that has a basis, that is, a generating set satisfying properties described in Theorem 19.1. There is another approach, via free groups, to free abelian groups. We now describe this approach.

Let $F[A]$ be the free group on the generating set A . We shall write F in place of $F[A]$ for the moment. Note that F is not abelian if A contains more than one element. Let C be the commutator subgroup of F . Then F/C is an abelian group, and it is not hard to show that F/C is free abelian with basis $\{aC \mid a \in A\}$. If aC is renamed a , we can view F/C as a free abelian group with basis A . This indicates how a free abelian group having a given set as basis can be constructed. Every free abelian group can be constructed in this fashion, up to isomorphism. That is, if G is free abelian with basis X , form the free group $F[X]$, form the factor group of $F[X]$ modulo its commutator subgroup, and we have a group isomorphic to G .

Theorems 20.7, 20.9, and 20.10 hold for free abelian groups as well as for free groups. In fact, the abelian version of Theorem 20.10 was proved for the finite rank case in Theorem 19.11. In contrast to Example 20.11 for free groups, it is true that for a free abelian group the rank of a subgroup is at most the rank of the entire group. Theorem 19.11 also showed this for the finite rank case.

■ EXERCISES 20

Computations

1. Find the reduced form and the inverse of the reduced form of each of the following words.
 - a. $a^2b^{-1}b^3c^3c^{-1}c^4b^{-2}$
 - b. $a^2a^{-3}b^3a^4c^4c^2a^{-1}$
2. Compute the products given in parts (a) and (b) of Exercise 1 in the case that $\{a, b, c\}$ is a set of generators forming a basis for a free abelian group. Find the inverse of these products.
3. How many different homomorphisms are there of a free group of rank 2 into
 - a. \mathbb{Z}_4 ?
 - b. \mathbb{Z}_6 ?
 - c. S_3 ?
4. How many different homomorphisms are there of a free group of rank 2 onto
 - a. \mathbb{Z}_4 ?
 - b. \mathbb{Z}_6 ?
 - c. S_3 ?
5. How many different homomorphisms are there of a free abelian group of rank 2 into
 - a. \mathbb{Z}_4 ?
 - b. \mathbb{Z}_6 ?
 - c. S_3 ?
6. How many different homomorphisms are there of a free abelian group of rank 2 onto
 - a. \mathbb{Z}_4 ?
 - b. \mathbb{Z}_6 ?
 - c. S_3 ?

Concepts

In Exercises 7 and 8, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

7. A *reduced word* is one in which there are no appearances in juxtaposition of two syllables having the same letter and also no appearances of a syllable with exponent 0.
8. The *rank of a free group* is the number of elements in a set of generators for the group.
9. Take one of the instances in this section in which the phrase “It would seem obvious that” was used and discuss your reaction in that instance.
10. Determine whether each of the following is true or false.
 - a. Every proper subgroup of a free group is a free group.
 - b. Every proper subgroup of every free abelian group is a free group.
 - c. A homomorphic image of a free group is a free group.
 - d. Every free abelian group has a basis.
 - e. The free abelian groups of finite rank are precisely the finitely generated abelian groups.
 - f. No free group is free.
 - g. No free abelian group is free.
 - h. No free abelian group of rank > 1 is free.
 - i. Any two free groups are isomorphic.
 - j. Any two free abelian groups of the same rank are isomorphic.

Theory

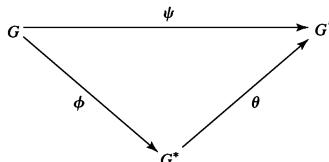
11. Let G be a finitely generated abelian group with identity 0. A finite set $\{b_1, \dots, b_n\}$, where $b_i \in G$, is a **basis** for G if $\{b_1, \dots, b_n\}$ generates G and $\sum_{i=1}^n m_i b_i = 0$ if and only if each $m_i b_i = 0$, where $m_i \in \mathbb{Z}$.
 - a. Show that $\{2, 3\}$ is not a basis for \mathbb{Z}_4 . Find a basis for \mathbb{Z}_4 .
 - b. Show that both $\{1\}$ and $\{2, 3\}$ are bases for \mathbb{Z}_6 . (This shows that for a finitely generated abelian group G with torsion, the number of elements in a basis may vary; that is, it need not be an *invariant* of the group G .)
 - c. Is a basis for a free abelian group as we defined it in Section 19 a basis in the sense in which it is used in this exercise?
 - d. Show that every finite abelian group has a basis $\{b_1, \dots, b_n\}$, where the order of b_i divides the order of b_{i+1} .

In present-day expositions of algebra, a frequently used technique (particularly by the disciples of N. Bourbaki) for introducing a new algebraic entity is the following:

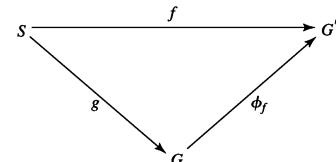
1. Describe algebraic properties that this algebraic entity is to possess.
2. Prove that any two algebraic entities with these properties are isomorphic, that is, that these properties characterize the entity.
3. Show that at least one such entity exists.

The next three exercises illustrate this technique for three algebraic entities, each of which we have met before. So that we do not give away their identities, we use fictitious names for them in the first two exercises. The last part of these first two exercises asks us to give the usual name for the entity.

12. Let G be any group. An abelian group G^* is a **blip group of G** if there exists a fixed homomorphism ϕ of G onto G^* such that each homomorphism ψ of G into an abelian group G' can be factored as $\psi = \theta\phi$, where θ is a homomorphism of G^* into G' (see Fig. 20.14).
- a. Show that any two blip groups of G are isomorphic. [Hint: Let G_1^* and G_2^* be two blip groups of G . Then each of the fixed homomorphisms $\phi_1 : G \rightarrow G_1^*$ and $\phi_2 : G \rightarrow G_2^*$ can be factored via the other blip group according to the definition of a blip group; that is, $\phi_1 = \theta_1\phi_2$ and $\phi_2 = \theta_2\phi_1$. Show that θ_1 is an isomorphism of G_2^* onto G_1^* by showing that both $\theta_1\theta_2$ and $\theta_2\theta_1$ are identity maps.]
 - b. Show for every group G that a blip group G^* of G exists.
 - c. What concept that we have introduced before corresponds to this idea of a blip group of G ?



20.14 Figure



20.15 Figure

13. Let S be any set. A group G together with a fixed function $g : S \rightarrow G$ constitutes a **blop group on S** if for each group G' and map $f : S \rightarrow G'$ there exists a *unique* homomorphism ϕ_f of G into G' such that $f = \phi_f g$ (see Fig. 20.15).
- a. Let S be a fixed set. Show that if both G_1 , together with $g_1 : S \rightarrow G_1$, and G_2 , together with $g_2 : S \rightarrow G_2$, are blop groups on S , then G_1 and G_2 are isomorphic. [Hint: Show that g_1 and g_2 are one-to-one maps and that g_1S and g_2S generate G_1 and G_2 , respectively. Then proceed in a way analogous to that given by the hint for Exercise 12.]
 - b. Let S be a set. Show that a blop group on S exists. You may use any theorems of the text.
 - c. What concept that we have introduced before corresponds to this idea of a blop group on S ?
14. Characterize a free abelian group by properties in a fashion similar to that used in Exercise 13.

SECTION 21 GROUP PRESENTATIONS

Definition

Following most of the literature on group presentations, in this section we let 1 be the identity of a group. The idea of a **group presentation** is to form a group by giving a set of generators for the group and certain equations or relations that we want the generators to satisfy. We want the group to be as free as it possibly can be on the generators, subject to these relations.

21.1 Example Suppose G is free with generation x and y except for the relation $xy = yx$, which we may express as $xyx^{-1}y^{-1} = 1$. Note that the condition $xy = yx$ is exactly what is needed to make G abelian, even though $xyx^{-1}y^{-1}$ is just one of the many possible commutators of $F[\{x, y\}]$. Thus G is free abelian on two generators and is isomorphic to $F[\{x, y\}]$ modulo its commutator subgroup. This commutator subgroup of $F[\{x, y\}]$ is the smallest normal subgroup containing $xyx^{-1}y^{-1}$, since any normal subgroup containing $xyx^{-1}y^{-1}$ gives rise to a factor group that is abelian and thus contains the commutator subgroup by Theorem 13.22. \blacktriangle

The preceding example illustrates the general situation. Let $F[A]$ be a free group and suppose that we want to form a new group as much like $F[A]$ as it can be, subject to certain equations that we want satisfied. Any equation can be written in a form in which the right-hand side is 1. Thus we can consider the equations to be $r_i = 1$ for $i \in I$, where $r_i \in F[A]$. If we require that $r_i = 1$, then we will have to have

$$x(r_i^n)x^{-1} = 1$$

for any $x \in F[A]$ and $n \in \mathbb{Z}$. Also any product of elements equal to 1 will again have to equal 1. Thus any finite product of the form

$$\prod_j x_j(r_{i_j}^{n_j})x_j^{-1},$$

where the r_{i_j} need not be distinct, will have to equal 1 in the new group. It is readily checked that the set of all these finite products is a normal subgroup R of $F[A]$. Thus any group looking as much as possible like $F[A]$, subject to the requirements $r_i = 1$, also has $r = 1$ for every $r \in R$. But $F[A]/R$ looks like $F[A]$ (remember that we multiply cosets by choosing representatives), except that R has been collapsed to form the identity 1. Hence the group we are after is (at least isomorphic to) $F[A]/R$. We can view this group as described by the generating set A and the set $\{r_i \mid i \in I\}$, which we will abbreviate $\{r_i\}$.

■ HISTORICAL NOTE

The idea of a group presentation already appears in Arthur Cayley's 1859 paper, "On the Theory of Groups as Depending on the Symbolic Equation $\theta^n = 1$. Third Part." In this article, Cayley gives a complete enumeration of the five groups of order 8, both by listing all the elements of each and by giving for each a presentation. For example, his third example is what is here called the *dihedral group* D_4 ; Cayley notes that this group is generated by the two elements α, β with the relations $\alpha^4 = 1, \beta^2 = 1, \alpha\beta = \beta\alpha^3$. He also shows more generally that a group of order mn is generated by α, β with the relations $\alpha^m = 1, \beta^n = 1, \alpha\beta = \beta\alpha^s$ if and only if $s^n \equiv 1 \pmod{m}$ (see Exercise 13).

In 1878, Cayley returned to the theory of groups and noted that a central problem in that

theory is the determination of all groups of a given order n . In the early 1890s, Otto Hölder published several papers attempting to solve Cayley's problem. Using techniques similar to those discussed in Sections 17 and 21, Hölder determined all simple groups of order up to 200 and characterized all the groups of orders p^3, pq^2, pqr , and p^4 , where p, q, r are distinct prime numbers. Furthermore, he developed techniques for determining the possible structures of a group G , if one is given the structure of a normal subgroup H and the structure of the factor group G/H . Interestingly, since the notion of an abstract group was still fairly new at this time, Hölder typically began his papers with the definition of a group and also emphasized that isomorphic groups are essentially one and the same object.

21.2 Definition Let A be a set and let $\{r_i\} \subseteq F[A]$. Let R be the least normal subgroup of $F[A]$ containing the r_i . An isomorphism ϕ of $F[A]/R$ onto a group G is a **presentation of G** . The

sets A and $\{r_i\}$ give a **group presentation**. The set A is the set of **generators for the presentation** and each r_i is a **relator**. Each $r \in R$ is a **consequence** of $\{r_i\}$. An equation $r_i = 1$ is a **relation**. A **finite presentation** is one in which both A and $\{r_i\}$ are finite sets. ■

This definition may seem complicated, but it really is not. In Example 21.1, $\{x, y\}$ is our set of generators and $xyx^{-1}y^{-1}$ is the only relator. The equation $xyx^{-1}y^{-1} = 1$, or $xy = yx$, is a relation. This was an example of a finite presentation.

If a group presentation has generators x_j and relators r_i , we shall use the notations

$$(x_j : r_i) \quad \text{or} \quad (x_j : r_i = 1)$$

to denote the group presentation. We may refer to $F[\{x_j\}]/R$ as *the group with presentation* $(x_j : r_i)$.

Isomorphic Presentations

21.3 Example Consider the group presentation with

$$A = \{a\} \quad \text{and} \quad \{r_i\} = \{a^6\},$$

that is, the presentation

$$(a : a^6 = 1).$$

This group defined by one generator a , with the relation $a^6 = 1$, is isomorphic to \mathbb{Z}_6 .

Now consider the group defined by two generators a and b , with $a^2 = 1, b^3 = 1$, and $ab = ba$, that is, the group with presentation

$$(a, b : a^2, b^3, aba^{-1}b^{-1}).$$

The condition $a^2 = 1$ gives $a^{-1} = a$. Also $b^3 = 1$ gives $b^{-1} = b^2$. Thus every element in this group can be written as a product of nonnegative powers of a and b . The relation $aba^{-1}b^{-1} = 1$, that is, $ab = ba$, allows us to write first all the factors involving a and then the factors involving b . Hence every element of the group is equal to some $a^m b^n$. But then $a^2 = 1$ and $b^3 = 1$ show that there are just six distinct elements,

$$1, b, b^2, a, ab, ab^2.$$

The subgroup $\langle ab \rangle$ contains the elements 1, ab , and the powers of ab :

$$(ab)^2 = a^2b^2 = b^2$$

$$(ab)^3 = abb^2 = a$$

$$(ab)^4 = a(ab) = b$$

$$(ab)^5 = (ab)b = ab^2.$$

So this group is also a cyclic group of order 6 isomorphic with \mathbb{Z}_6 . ▲

21.4 Example The dihedral group has presentation

$$D_n : (a, b | a^n, b^2, abab)$$

since if we let $a = \mu$ and $b = \rho$ the three relations are exactly the defining relations for D_n . (The last relation $abab = 1$ is equivalent to $ab = ba^{-1}$.) The element $abab$ is in R if and only if $b(abab)b^{-1}$ is in R since R is a normal subgroup. We have $b(abab)b^{-1} = baba$. So in any presentation with generators a and b and a relator $abab$, we can replace $abab$ with $baba$ and get the same subgroup R and therefore the same factor group. Hence the dihedral group also has presentation

$$D_n : (a, b : a^n, b^2, baba).$$

Setting the relators to 1 gives the equivalent presentation

$$(a, b : a^n = 1, b^2 = 1, ba = a^{-1}b^{-1})$$

which can also be rewritten as

$$(a, b : a^n = 1, b^2 = 1, ba = a^{n-1}b).$$



The preceding examples illustrate that different presentations may give isomorphic groups. When this happens, we have **isomorphic presentations**. To determine whether two presentations are isomorphic may be very hard. It has been shown (see Rabin [22]) that a number of such problems connected with this theory are not generally solvable; that is, there is no *routine* and well-defined way of discovering a solution in all cases. These unsolvable problems include the problem of deciding whether two presentations are isomorphic, whether a group given by a presentation is finite, free, abelian, or trivial, and the famous *word problem* of determining whether a given word w is a consequence of a given set of relations $\{r_i\}$.

The importance of this material is indicated by our Theorem 20.13, which guarantees that *every group has a presentation*.

21.5 Example

Let us show that

$$(x, y : y^2x = y, yx^2y = x)$$

is a presentation of the trivial group of one element. We need only show that x and y are consequences of the relators y^2xy^{-1} and yx^2yx^{-1} , or that $x = 1$ and $y = 1$ can be deduced from $y^2x = y$ and $yx^2y = x$. We illustrate both techniques.

As a consequence of y^2xy^{-1} , we get yx upon conjugation by y^{-1} . From yx we deduce $x^{-1}y^{-1}$, and then $(x^{-1}y^{-1})(yx^2yx^{-1})$ gives xyx^{-1} . Conjugating xyx^{-1} by x^{-1} , we get y . From y we get y^{-1} , and $y^{-1}(yx)$ is x .

Working with relations instead of relators, from $y^2x = y$ we deduce $yx = 1$ upon multiplication by y^{-1} on the left. Then substituting $yx = 1$ into $yx^2y = x$, that is, $(yx)(xy) = x$, we get $xy = x$. Then multiplying by x^{-1} on the left, we have $y = 1$. Substituting this in $yx = 1$, we get $x = 1$.

Both techniques amount to the same work, but it somehow seems more natural to most of us to work with relations.



Applications

We conclude this chapter with two applications.

21.6 Example

Let us determine all groups of order 10 up to isomorphism. We know from the Fundamental Theorem 9.12 that every abelian group of order 10 is isomorphic to \mathbb{Z}_{10} . Suppose that G is nonabelian of order 10. By Sylow theory, G contains a normal subgroup H of order 5, and H must be cyclic. Let a be a generator of H . Then G/H is of order 2 and thus isomorphic to \mathbb{Z}_2 . If $b \in G$ and $b \notin H$, we must then have $b^2 \in H$. Since every element of H except 1 has order 5, if b^2 were not equal to 1, then b^2 would have order 5, so b would have order 10. This would mean that G would be cyclic, contradicting our assumption that G is not abelian. Thus $b^2 = 1$. Finally, since H is a normal subgroup of G , $bHb^{-1} = H$, so in particular, $bab^{-1} \in H$. Since conjugation by b is an automorphism of H , bab^{-1} must be another element of H of order 5, hence bab^{-1} equals a, a^2, a^3 , or a^4 . But $bab^{-1} = a$ would give $ba = ab$, and then G would be abelian, since a and b generate G . Thus the possibilities for presentations of G are:

1. $(a, b : a^5 = 1, b^2 = 1, ba = a^2b),$
2. $(a, b : a^5 = 1, b^2 = 1, ba = a^3b),$
3. $(a, b : a^5 = 1, b^2 = 1, ba = a^4b).$

Note that all three of these presentations can give groups of order at most 10, since the last relation $ba = a^b$ enables us to express every product of a 's and b 's in G in the form $a^b b^t$. Then $a^5 = 1$ and $b^2 = 1$ show that the set

$$S = \{a^0 b^0, a^1 b^0, a^2 b^0, a^3 b^0, a^4 b^0, a^0 b^1, a^1 b^1, a^2 b^1, a^3 b^1, a^4 b^1\}$$

includes all elements of G .

It is not yet clear that all these elements in S are distinct, so that we have in all three cases a group of order 10. For example, the group presentation

$$(a, b : a^5 = 1, b^2 = 1, ba = a^2b)$$

gives a group in which, using the associative law, we have

$$\begin{aligned} a &= b^2 a = (bb)a = b(ba) = b(a^2b) = (ba)(ab) \\ &= (a^2b)(ab) = a^2(ba)b = a^2(a^2b)b = a^4b^2 = a^4 \end{aligned}$$

Thus in this group, $a = a^4$, so $a^3 = 1$, which, together with $a^5 = 1$, yields $a^2 = 1$. But $a^2 = 1$, together with $a^3 = 1$, means that $a = 1$. Hence every element in the group with presentation

$$(a, b : a^5 = 1, b^2 = 1, ba = a^2b)$$

is equal to either 1 or b ; that is, this group is isomorphic to \mathbb{Z}_2 . A similar study of

$$(bb)a = b(ba)$$

for

$$(a, b : a^5 = 1, b^2 = 1, ba = a^3b)$$

shows that $a = a^4$ again, so this also yields a group isomorphic to \mathbb{Z}_2 .

This leaves just

$$(a, b : a^5 = 1, b^2 = 1, ba = a^4b)$$

as a candidate for a nonabelian group of order 10. As in Example 21.4 this is a presentation of the dihedral group D_5 .

If we were unaware of the dihedral group, how would we show that the presentation gives a group with 10 elements? One attack is as follows. Let us try to make S into a group by defining $(a^x b^y)(a^u b^v)$ to be $a^x b^y$, where x is the remainder of $s + u(4^t)$ when divided by 5, and y is the remainder of $t + v$ when divided by 2, in the sense of the division algorithm (Theorem 6.2). The formula $s + u(4^t)$ is counting what the power of a should be after moving u copies of a by t copies of b . In other words, we use the relation $ba = a^4b$ as a guide in defining the product $(a^x b^y)(a^u b^v)$ of two elements of S . We see that $a^0 b^0$ acts as identity, and that given $a^u b^v$, we can determine t and s successively by letting

$$t \equiv -v \pmod{2}$$

and then

$$s \equiv -u(4^t) \pmod{5},$$

giving $a^s b^t$, which is a left inverse for $a^u b^v$. We will then have a group structure on S if and only if the associative law holds. Exercise 13 asks us to carry out the straightforward computation for the associative law and to discover a condition for S to be a group under such a definition of multiplication. The criterion of the exercise in this case amounts to the valid congruence

$$4^2 \equiv 1 \pmod{5}.$$

Thus we do get a group of order 10. Note that

$$2^2 \not\equiv 1 \pmod{5}$$

and

$$3^2 \not\equiv 1 \pmod{5},$$

so Exercise 13 also shows that

$$(a, b : a^5 = 1, b^2 = 1, ba = a^2b)$$

and

$$(a, b : a^5 = 1, b^2 = 1, ba = a^3b)$$

do not give groups of order 10. ▲

21.7 Example Let us determine all groups of order 8 up to isomorphism. We know the three abelian ones:

$$\mathbb{Z}_8, \quad \mathbb{Z}_2 \times \mathbb{Z}_4, \quad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2.$$

Using generators and relations, we shall give presentations of the nonabelian groups.

Let G be nonabelian of order 8. Since G is nonabelian, it has no elements of order 8, so each element but the identity is of order either 2 or 4. If every element were of order 2, then for $a, b \in G$, we would have $(ab)^2 = 1$, that is, $abab = 1$. Then since $a^2 = 1$ and $b^2 = 1$ also, we would have

$$ba = a^2bab^2 = a(ab)^2b = ab,$$

contrary to our assumption that G is not abelian. Thus G must have an element of order 4.

Let $\langle a \rangle$ be a subgroup of G of order 4. If $b \notin \langle a \rangle$, the cosets $\langle a \rangle$ and $b\langle a \rangle$ exhaust all of G . Hence a and b are generators for G and $a^4 = 1$. Since $\langle a \rangle$ is normal in G (by Sylow theory, or because it is of index 2), $G/\langle a \rangle$ is isomorphic to \mathbb{Z}_2 and we have $b^2 \in \langle a \rangle$. If $b^2 = a$ or $b^2 = a^3$, then b would be of order 8. Hence $b^2 = 1$ or $b^2 = a^2$. Finally, since $\langle a \rangle$ is normal, we have $bab^{-1} \in \langle a \rangle$, and since $b\langle a \rangle b^{-1}$ is a subgroup conjugate to $\langle a \rangle$ and hence isomorphic to $\langle a \rangle$, we see that bab^{-1} must be an element of order 4. Thus $bab^{-1} = a$ or $bab^{-1} = a^3$. If bab^{-1} were equal to a , then ba would equal ab , which would make G abelian. Hence $bab^{-1} = a^3$, so $ba = a^3b$. Thus we have two possibilities for G , namely,

$$G_1 : (a, b : a^4 = 1, b^2 = 1, ba = a^3b)$$

and

$$G_2 : (a, b : a^4 = 1, b^2 = a^2, ba = a^3b).$$

Note that $a^{-1} = a^3$, and that b^{-1} is b in G_1 and b^3 in G_2 . These facts, along with the relation $ba = a^3b$, enable us to express every element in G_i in the form $a^m b^n$, as in Examples 21.3 and 21.6. Since $a^4 = 1$ and either $b^2 = 1$ or $b^2 = a^2$, the possible elements in each group are

$$1, \quad a, \quad a^2, \quad a^3, \quad b, \quad ab, \quad a^2b, \quad a^3b.$$

Thus G_1 and G_2 each have order at most 8. The first group G_1 is sometimes called the **octic group**, but as we saw in Example 21.4 it is isomorphic with our old friend D_4 , the dihedral group. For the second we can make S into a group by defining $(a^i b^j)(a^s b^t)$ to be $a^x b^y$ where y is $j + s$ modulo 2 and if $j + s < 2$, then x is the remainder of $i + r(2j + 1)$ when divided by 4 and if $j + s = 2$, then x is the remainder when $i + 2 + r(2j + 1)$ is divided by 4. We leave it as an exercise to show that this operation makes S a group, which shows that G_2 is a presentation of a group of order 8.

Since $ba = a^3b \neq ab$, we see that both G_1 and G_2 are nonabelian. That the two groups are not isomorphic follows from the fact that a computation shows that G_1 has only two elements of order 4, namely, a and a^3 . On the other hand, in G_2 all elements but 1 and a^2 are of order 4. We leave the computations of the tables for these groups

to Exercise 3. To illustrate suppose we wish to compute $(a^2b)(a^3b)$. Using $ba = a^3b$ repeatedly, we get

$$(a^2b)(a^3b) = a^2(ba)a^2b = a^5(ba)ab = a^8(ba)b = a^{11}b^2.$$

Then for G_1 , we have

$$a^{11}b^2 = a^{11} = a^3,$$

but if we are in G_2 , we get

$$a^{11}b^2 = a^{13} = a.$$

The group G_2 is called the **quaternion group**. We shall encounter the quaternion group again in Section 32. ▲

■ EXERCISES 21

Computations

1. Give a presentation of \mathbb{Z}_4 involving one generator; involving two generators; involving three generators.
2. Give a presentation of S_3 involving three generators.
3. Give the tables for both the octic group

$$(a, b : a^4 = 1, b^2 = 1, ba = a^3b)$$

and the quaternion group

$$(a, b : a^4 = 1, b^2 = a^2, ba = a^3b).$$

In both cases, write the elements in the order $1, a, a^2, a^3, b, ab, a^2b, a^3b$. (Note that we do not have to compute *every* product. We know that these presentations give groups of order 8, and once we have computed enough products the rest are forced so that each row and each column of the table has each element exactly once.)

4. Determine all groups of order 14 up to isomorphism. [Hint: Follow the outline of Example 21.6 and use Exercise 13, part (b).]
5. Determine all groups of order 21 up to isomorphism. [Hint: Follow the outline of Example 21.6 and use Exercise 13, part (b). It may seem that there are two presentations giving nonabelian groups. Show that they are isomorphic.]

Concepts

In Exercises 6 and 7, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

6. A *consequence* of the set of relators is any finite product of relators raised to powers.
7. Two group presentations are *isomorphic* if and only if there is a one-to-one correspondence of the generators of the first presentation with the generators of the second that yields, by renaming generators, a one-to-one correspondence of the relators of the first presentation with those of the second.
8. Determine whether each of the following is true or false.
 - a. Every group has a presentation.
 - b. Every group has many different presentations.
 - c. Every group has two presentations that are not isomorphic.
 - d. Every group has a finite presentation.
 - e. Every group with a finite presentation is of finite order.
 - f. Every cyclic group has a presentation with just one generator.
 - g. Every conjugate of a relator is a consequence of the relator.

- h.** Two presentations with the same number of generators are always isomorphic.
- i.** In a presentation of an abelian group, the set of consequences of the relators contains the commutator subgroup of the free group on the generators.
- j.** Every presentation of a free group has 1 as the only relator.

Theory

- 9.** Use the methods of this section and Exercise 13, part (b), to show that there are no nonabelian groups of order 15.
- 10.** Show, using Exercise 13, that

$$(a, b : a^3 = 1, b^2 = 1, ba = a^2b)$$

gives a group of order 6. Show that it is nonabelian.

- 11.** Show that the presentation

$$(a, b : a^3 = 1, b^2 = 1, ba = a^2b)$$

of Exercise 10 gives (up to isomorphism) the only nonabelian group of order 6, and hence gives a group isomorphic to S_3 .

- 12.** We showed in Example 13.6 that A_4 has no subgroup of order 6. The preceding exercise shows that such a subgroup of A_4 would have to be isomorphic to either \mathbb{Z}_6 or S_3 . Show again that this is impossible by considering orders of elements.

- 13.** Let

$$S = \{a^i b^j \mid 0 \leq i < m, 0 \leq j < n\},$$

that is, S consists of all formal products $a^i b^j$ starting with $a^0 b^0$ and ending with $a^{m-1} b^{n-1}$. Let r be a positive integer, and define multiplication on S by

$$(a^s b^t)(a^u b^v) = a^x b^y,$$

where x is the remainder of $s + u(r^t)$ when divided by m , and y is the remainder of $t + v$ when divided by n , in the sense of the division algorithm (Theorem 6.2).

- a.** Show that a necessary and sufficient condition for the associative law to hold and for S to be a group under this multiplication is that $r^n \equiv 1 \pmod{m}$.
- b.** Deduce from part (a) that the group presentation

$$(a, b : a^m = 1, b^n = 1, ba = a^r b)$$

gives a group of order mn if and only if $r^n \equiv 1 \pmod{m}$. (See the Historical Note in this section.)

- 14.** Without using Exercise 13, prove that $(a, b : a^5 = 1, b^2 = 1, ba = a^3b)$ is a presentation for the group \mathbb{Z}_2 .
- 15.** Is the group obtained from the group presentation with the letters a through z as generators and the words in a standard English dictionary as relators the trivial group? Prove your answer.

Rings and Fields

Section 22 Rings and Fields

Section 23 Integral Domains

Section 24 Fermat's and Euler's Theorems

Section 25 Encryption

SECTION 22

RINGS AND FIELDS

All our work thus far has been concerned with sets on which a single binary operation has been defined. Our years of work with the integers and real numbers show that a study of sets on which two binary operations have been defined should be of great importance. Algebraic structures of this type are introduced in this section. In one sense, this section seems more intuitive than those that precede it, for the structures studied are closely related to those we have worked with for many years. However, we will be continuing with our axiomatic approach. So, from another viewpoint this study is more complicated than group theory, for we now have two binary operations and more axioms to deal with.

Definitions and Basic Properties

The most general algebraic structure with two binary operations that we shall study is called a *ring*. As Example 22.2 following Definition 22.1 indicates, we have all worked with rings since elementary school.

22.1 Definition A ring $\langle R, +, \cdot \rangle$ is a set R together with two binary operations $+$ and \cdot , which we call *addition* and *multiplication*, defined on R such that the following axioms are satisfied:

\mathcal{R}_1 . $\langle R, + \rangle$ is an abelian group.

\mathcal{R}_2 . Multiplication is associative.

\mathcal{R}_3 . For all $a, b, c \in R$, the **left distributive law**, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and the **right distributive law** $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$ hold. ■

22.2 Example We are well aware that axioms \mathcal{R}_1 , \mathcal{R}_2 , and \mathcal{R}_3 for a ring hold in any subset of the complex numbers that is a group under addition and that is closed under multiplication. For example, $\langle \mathbb{Z}, +, \cdot \rangle$, $\langle \mathbb{Q}, +, \cdot \rangle$, $\langle \mathbb{R}, +, \cdot \rangle$, and $\langle \mathbb{C}, +, \cdot \rangle$ are rings. ▲

■ HISTORICAL NOTE

The theory of rings grew out of the study of two particular classes of rings, polynomial rings in n variables over the real or complex numbers (Section 27) and the “integers” of an algebraic number field. It was David Hilbert (1862–1943) who first introduced the term *ring*, in connection with the latter example, but it was not until the second decade of the twentieth century that a fully abstract definition appeared. The theory of commutative rings was given a firm axiomatic foundation by Emmy Noether (1882–1935) in her monumental paper “Ideal Theory in Rings,” which appeared in 1921. A major concept of this paper is the ascending chain condition for ideals. Noether proved that in any ring in which every ascending chain of ideals has a maximal element, every ideal is finitely generated.

Emmy Noether received her doctorate from the University of Erlangen, Germany, in 1907. Hilbert invited her to Göttingen in 1915, but his efforts to secure her a paid position were blocked because of her sex. Hilbert complained, “I do not see that the sex of the candidate is an argument against her admission [to the faculty]. After all, we are a university, not a bathing establishment.” Noether was, however, able to lecture under Hilbert’s name. Ultimately, after the political changes accompanying the end of the First World War reached Göttingen, she was given in 1923 a paid position at the University. For the next decade, she was very influential in the development of the basic concepts of modern algebra. Along with other Jewish faculty members, however, she was forced to leave Göttingen in 1933. She spent the final two years of her life at Bryn Mawr College near Philadelphia.

It is customary to denote multiplication in a ring by juxtaposition, using ab in place of $a \cdot b$. We shall also observe the usual convention that multiplication is performed before addition in the absence of parentheses, so the left distributive law, for example, becomes

$$a(b + c) = ab + ac,$$

without the parentheses on the right side of the equation. Also, as a convenience analogous to our notation in group theory, we shall somewhat incorrectly refer to a *ring* R in place of a *ring* $\langle R, +, \cdot \rangle$, provided that no confusion will result. In particular, from now on \mathbb{Z} will always be $\langle \mathbb{Z}, +, \cdot \rangle$, and \mathbb{Q}, \mathbb{R} , and \mathbb{C} will also be the rings in Example 22.2. We may on occasion refer to $\langle R, + \rangle$ as the *additive group of the ring* R .

22.3 Example Let R be any ring and let $M_n(R)$ be the collection of all $n \times n$ matrices having elements of R as entries. The operations of addition and multiplication in R allow us to add and multiply matrices in the usual fashion, explained in the appendix. We can quickly check that $\langle M_n(R), + \rangle$ is an abelian group. The associativity of matrix multiplication and the two distributive laws in $M_n(R)$ are more tedious to demonstrate, but straightforward calculations indicate that they follow from the same properties in R . We will assume from now on that we know that $M_n(R)$ is a ring. In particular, we have the rings $M_n(\mathbb{Z}), M_n(\mathbb{Q}), M_n(\mathbb{R})$, and $M_n(\mathbb{C})$. Note that multiplication is not a commutative operation in any of these rings for $n \geq 2$. ▲

22.4 Example Let F be the set of all functions $f : \mathbb{R} \rightarrow \mathbb{R}$. We know that $\langle F, + \rangle$ is an abelian group under the usual function addition,

$$(f + g)(x) = f(x) + g(x).$$

We define multiplication on F by

$$(fg)(x) = f(x)g(x).$$

That is, fg is the function whose value at x is $f(x)g(x)$. It is readily checked that F is a ring; we leave the demonstration to Exercise 36. We have used this juxtaposition

notation $\sigma\mu$ for the composite function $\sigma(\mu(x))$ when discussing permutation multiplication. If we were to use both function multiplication and function composition in F , we would use the notation $f \circ g$ for the composite function. However, we will use composition of functions almost exclusively with homomorphisms, which we will denote by Greek letters, and the usual product defined in this example chiefly when multiplying polynomial function's $f(x)g(x)$, so no confusion should result. ▲

22.5 Example Recall that in group theory, $n\mathbb{Z}$ is the cyclic subgroup of \mathbb{Z} under addition consisting of all integer multiples of the integer n . Since $(nr)(ns) = n(ns)$, we see that $n\mathbb{Z}$ is closed under multiplication. The associative and distributive laws that hold in \mathbb{Z} then assure us that $(n\mathbb{Z}, +, \cdot)$ is a ring. From now on in the text, we will consider $n\mathbb{Z}$ to be this ring. ▲

22.6 Example Consider the cyclic group $(\mathbb{Z}_n, +)$. If we define for $a, b \in \mathbb{Z}_n$ the product ab as the remainder of the usual product of integers when divided by n , it can be shown that $(\mathbb{Z}_n, +, \cdot)$ is a ring. We shall feel free to use this fact. For example, in \mathbb{Z}_{10} we have $(3)(7) = 1$. This operation on \mathbb{Z}_n is **multiplication modulo n** . We do not check the ring axioms here, for they will follow in Section 30 from some of the theory we develop there. From now on, \mathbb{Z}_n will always be the ring $(\mathbb{Z}_n, +, \cdot)$. ▲

22.7 Example If R_1, R_2, \dots, R_n are rings, we can form the set $R_1 \times R_2 \times \dots \times R_n$ of all ordered n -tuples (r_1, r_2, \dots, r_n) , where $r_i \in R_i$. Defining addition and multiplication of n -tuples by components (just as for groups), we see at once from the ring axioms in each component that the set of all these n -tuples forms a ring under addition and multiplication by components. The ring $R_1 \times R_2 \times \dots \times R_n$ is the **direct product** of the rings R_i . ▲

Continuing matters of notation, we shall always let 0 be the additive identity of a ring. The additive inverse of an element a of a ring is $-a$. We shall frequently have occasion to refer to a sum

$$a + a + \dots + a$$

having n summands. We shall let this sum be $n \cdot a$, always using the dot. However, $n \cdot a$ is not to be interpreted as a multiplication of n and a in the ring, for the integer n may not be in the ring at all. If $n < 0$, we let

$$n \cdot a = (-a) + (-a) + \dots + (-a)$$

for $|n|$ summands. Finally, we define

$$0 \cdot a = 0$$

for $0 \in \mathbb{Z}$ on the left side of the equations and $0 \in R$ on the right side. Actually, the equation $0a = 0$ holds also for $0 \in R$ on both sides. The following theorem proves this and various other elementary but important facts. Note the strong use of the distributive laws in the proof of this theorem. Axiom \mathcal{R}_1 for a ring concerns only addition, and axiom \mathcal{R}_2 concerns only multiplication. This shows that in order to prove anything that gives a relationship between these two operations, we are going to have to use axiom \mathcal{R}_3 . For example, the first thing that we will show in Theorem 22.8 is that $0a = 0$ for any element a in a ring R . Now this relation involves both addition and multiplication. The multiplication $0a$ stares us in the face, and 0 is an *additive* concept. Thus we will have to come up with an argument that uses a distributive law to prove this.

22.8 Theorem If R is a ring with additive identity 0 , then for any $a, b \in R$ we have

1. $0a = a0 = 0$,
2. $a(-b) = (-a)b = -(ab)$,
3. $(-a)(-b) = ab$.

Proof For Property 1, note that by axioms \mathcal{R}_1 and \mathcal{R}_2 ,

$$a0 + a0 = a(0 + 0) = a0 = 0 + a0.$$

Then by the cancellation law for the additive group $\langle R, + \rangle$, we have $a0 = 0$. Likewise,

$$0a + 0a = (0 + 0)a = 0a = 0 + 0a$$

implies that $0a = 0$. This proves Property 1.

In order to understand the proof of Property 2, we must remember that, by definition, $-(ab)$ is the element that when added to ab gives 0. Thus to show that $a(-b) = -(ab)$, we must show precisely that $a(-b) + ab = 0$. By the left distributive law,

$$a(-b) + ab = a(-b + b) = a0 = 0,$$

since $a0 = 0$ by Property 1. Likewise,

$$(-a)b + ab = (-a + a)b = 0b = 0.$$

For Property 3, note that

$$(-a)(-b) = -(a(-b))$$

by Property 2. Again by Property 2,

$$-(a(-b)) = -(-(ab)),$$

and $-(-(ab))$ is the element that when added to $-(ab)$ gives 0. This is ab by definition of $-(ab)$ and by the uniqueness of an inverse in a group. Thus, $(-a)(-b) = ab$. ◆

Based on high school algebra it seems natural to begin a proof of Property 2 in Theorem 22.8 by writing $(-a)b = ((-1)a)b$. In Exercise 30 you will be asked to find an error in a “proof” of this sort.

It is important that you *understand* the preceding proof. The theorem allows us to use our usual rules for signs.

Homomorphisms and Isomorphisms

From our work in group theory, it is quite clear how a structure-relating map of a ring R into a ring R' should be defined.

22.9 Definition For rings R and R' , a map $\phi : R \rightarrow R'$ is a **homomorphism** if the following two conditions are satisfied for all $a, b \in R$:

1. $\phi(a + b) = \phi(a) + \phi(b),$
2. $\phi(ab) = \phi(a)\phi(b).$

■

In the preceding definition, Condition 1 is the statement that ϕ is a group homomorphism mapping the abelian group $\langle R, + \rangle$ into $\langle R', + \rangle$. Condition 2 requires that ϕ relate the multiplicative structures of the rings R and R' in the same way. Since ϕ is also a group homomorphism, all the results concerning group homomorphisms are valid for the additive structure of the rings. In particular, ϕ is one-to-one if and only if its **kernel** $\text{Ker}(\phi) = \{a \in R \mid \phi(a) = 0'\}$ is just the subset $\{0\}$ of R . The homomorphism ϕ of the group $\langle R, + \rangle$ gives rise to a factor group. We expect that a ring homomorphism will give rise to a factor ring. This is indeed the case. We delay discussion of this to Section 30, where the treatment will parallel our treatment of factor groups in Section 12.

22.10 Example Let F be the ring of all functions mapping \mathbb{R} into \mathbb{R} defined in Example 22.4. For each $a \in \mathbb{R}$, we have the **evaluation homomorphism** $\phi_a : F \rightarrow \mathbb{R}$, where $\phi_a(f) = f(a)$ for $f \in F$. We will work a great deal with this homomorphism in the rest of this text, for finding a real solution of a polynomial equation $p(x) = 0$ amounts precisely to finding $a \in \mathbb{R}$ such that $\phi_a(p) = 0$. Much of the remainder of this text deals with solving polynomial equations. We leave the demonstration of the homomorphism properties for ϕ_a to Exercise 37. ▲

22.11 Example The map $\phi : \mathbb{Z} \rightarrow \mathbb{Z}_n$ where $\phi(a)$ is the remainder of a modulo n is a ring homomorphism for each positive integer n . We know $\phi(a+b) = \phi(a) + \phi(b)$ by group theory. To show the multiplicative property, write $a = q_1n + r_1$ and $b = q_2n + r_2$ according to the division algorithm. Then $ab = n(q_1q_2n + r_1q_2 + q_1r_2) + r_1r_2$. Thus $\phi(ab)$ is the remainder of r_1r_2 when divided by n . Since $\phi(a) = r_1$ and $\phi(b) = r_2$, Example 22.6 indicates that $\phi(a)\phi(b)$ is also this same remainder, so $\phi(ab) = \phi(a)\phi(b)$. From group theory, we anticipate that the ring \mathbb{Z}_n might be isomorphic to a factor ring $\mathbb{Z}/n\mathbb{Z}$. This is indeed the case; factor rings will be discussed in Section 30. ▲

We realize that in the study of any sort of mathematical structure, an idea of basic importance is the concept of two systems being *structurally identical*, that is, one being just like the other except for names. In algebra this concept is always called *isomorphism*. The concept of two things being just alike except for names of elements leads us, just as it did for groups, to the following definition.

22.12 Definition An **isomorphism** $\phi : R \rightarrow R'$ from a ring R to a ring R' is a homomorphism that is one-to-one and onto R' . The rings R and R' are then **isomorphic**. ■

From our work in group theory, we expect that isomorphism gives an equivalence relation on any collection of rings. We need to check that the multiplicative property of an isomorphism is satisfied for the inverse map $\phi^{-1} : R' \rightarrow R$ (to complete the symmetry argument). Similarly, we check that if $\mu : R' \rightarrow R''$ is also a ring isomorphism, then the multiplicative requirement holds for the composite map $\mu\phi : R \rightarrow R''$ (to complete the transitivity argument). We ask you to do this in Exercise 38.

22.13 Example As abelian groups, $(\mathbb{Z}, +)$ and $(2\mathbb{Z}, +)$ are isomorphic under the map $\phi : \mathbb{Z} \rightarrow 2\mathbb{Z}$, with $\phi(x) = 2x$ for $x \in \mathbb{Z}$. Here ϕ is *not* a ring isomorphism, for $\phi(xy) = 2xy$, while $\phi(x)\phi(y) = 2x2y = 4xy$. ▲

Multiplicative Questions: Fields

Many of the rings we have mentioned, such as \mathbb{Z} , \mathbb{Q} , and \mathbb{R} , have a multiplicative identity element 1. However, $2\mathbb{Z}$ does not have an identity element for multiplication. Note also that multiplication is not commutative in the matrix rings described in Example 22.3.

It is evident that $\{0\}$, with $0 + 0 = 0$ and $(0)(0) = 0$, gives a ring, the **zero ring**. Here 0 acts as multiplicative as well as additive identity element. By Theorem 22.8, this is the only case in which 0 could act as a multiplicative identity element, for from $0a = 0$, we can then deduce that $a = 0$. Theorem 1.15 shows that if a ring has a multiplicative identity element, it is unique. We denote a multiplicative identity element in a ring by 1.

22.14 Definition A ring in which the multiplication is commutative is a **commutative ring**. A ring with a multiplicative identity element is a **ring with unity**; the multiplicative identity element 1 is called “**unity**.” ■

In a ring with unity 1 the distributive laws show that

$$(1 + 1 + \cdots + 1) (1 + 1 + \cdots + 1) = (1 + 1 + \cdots + 1), \\ n \text{ summands} \quad m \text{ summands} \quad nm \text{ summands}$$

that is, $(n \cdot 1)(m \cdot 1) = (nm) \cdot 1$. The next example gives an application of this observation.

22.15 Example We claim that for integers r and s where $\gcd(r, s) = 1$, the rings \mathbb{Z}_{rs} and $\mathbb{Z}_r \times \mathbb{Z}_s$ are isomorphic. Additively, they are both cyclic abelian groups of order rs with generators 1 and $(1, 1)$ respectively. Thus $\phi : \mathbb{Z}_{rs} \rightarrow \mathbb{Z}_r \times \mathbb{Z}_s$, defined by $\phi(n \cdot 1) = n \cdot (1, 1)$ is an additive group isomorphism. To check the multiplicative Condition 2 of Definition 22.9, we use the observation preceding this example for the unity $(1, 1)$ in the ring $\mathbb{Z}_r \times \mathbb{Z}_s$, and compute.

$$\phi(nm) = (nm) \cdot (1, 1) = [n \cdot (1, 1)][m \cdot (1, 1)] = \phi(n)\phi(m). \quad \blacktriangle$$

Note that a direct product $R = R_1 \times R_2 \times \cdots \times R_n$ of rings is commutative if and only if each ring R_i is commutative. Furthermore, R has a unity if and only if each R_i has a unity.

The set \mathbb{R}^* of nonzero real numbers forms a group under multiplication. However, the nonzero integers do not form a group under multiplication since only the integers 1 and -1 have multiplicative inverses in \mathbb{Z} . In general, a **multiplicative inverse** of an element a in a ring R with unity $1 \neq 0$ is an element $a^{-1} \in R$ such that $aa^{-1} = a^{-1}a = 1$. Precisely as for groups, a multiplicative inverse for an element a in R is unique, if it exists at all (see Exercise 45). Theorem 22.8 shows that it would be hopeless to have a multiplicative inverse for 0 except for the ring $\{0\}$, where $0 + 0 = 0$ and $(0)(0) = 0$, with 0 as both additive and multiplicative identity element. We are thus led to discuss the existence of multiplicative inverses for nonzero elements in a ring with nonzero unity. There is unavoidably a lot of terminology to be defined in this introductory section on rings. We are almost done.

22.16 Definition Let R be a ring with unity $1 \neq 0$. An element u in R is a **unit** of R if it has a multiplicative inverse in R . If every nonzero element of R is a unit, then R is a **division ring** (or **skew field**). A **field** is a commutative division ring. A noncommutative division ring is called a “**strictly skew field**.” ■

22.17 Example Let us find the units in \mathbb{Z}_{14} . Of course, 1 and $-1 = 13$ are units. Since $(3)(5) = 1$ we see that 3 and 5 are units; therefore $-3 = 11$ and $-5 = 9$ are also units. None of the remaining elements of \mathbb{Z}_{14} can be units, since no multiple of $2, 4, 6, 7, 8$, or 10 can be one more than a multiple of 14 ; they all have a common factor, either 2 or 7 , with 14 . Section 24 will show that the units in \mathbb{Z}_n are precisely those $m \in \mathbb{Z}_n$ such that $\gcd(m, n) = 1$. ▲

22.18 Example \mathbb{Z} is not a field, because 2 , for example, has no multiplicative inverse, so 2 is not a unit in \mathbb{Z} . The only units in \mathbb{Z} are 1 and -1 . However, \mathbb{Q} and \mathbb{R} are fields. An example of a strictly skew field is given in Section 32. ▲

We have the natural concepts of a subring of a ring and a subfield of a field. A **subring of a ring** is a subset of the ring that is a ring under induced operations from the whole ring; a **subfield** is defined similarly for a subset of a field. In fact, let us say here once and for all that if we have a set, together with a certain specified type of *algebraic structure* (group, ring, field, integral domain, vector space, and so on), then any subset of this set, together with a natural induced algebraic structure *that yields an algebraic structure of the same type*, is a **substructure**. If K and L are both structures, we shall let $K \leq L$ denote that K is a substructure of L and $K < L$ denote that $K \leq L$ but $K \neq L$. Exercise 50 gives criteria for a subset S of a ring R to form a subring of R .

■ HISTORICAL NOTE

Although fields were implicit in the early work on the solvability of equations by Abel and Galois, it was Leopold Kronecker (1823–1891) who in connection with his own work on this subject first published in 1881 a definition of what he called a “domain of rationality”: “The domain of rationality (R', R'', R''', \dots) contains \dots every one of those quantities which are rational functions of the quantities R', R'', R''', \dots with integral coefficients.” Kronecker, however, who insisted that any mathematical subject must be constructible in finitely many steps, did not view the domain of rationality as a complete entity, but merely as a region in which took place various operations on its elements.

Richard Dedekind (1831–1916), the inventor of the Dedekind cut definition of a real number,

considered a field as a completed entity. In 1871, he published the following definition in his supplement to the second edition of Dirichlet's text on number theory: “By a field we mean any system of infinitely many real or complex numbers, which in itself is so closed and complete, that the addition, subtraction, multiplication, and division of any two numbers always produces a number of the same system.” Both Kronecker and Dedekind had, however, dealt with their varying ideas of this notion as early as the 1850s in their university lectures.

A more abstract definition of a field, similar to the one in the text, was given by Heinrich Weber (1842–1913) in a paper of 1893. Weber's definition, unlike that of Dedekind, specifically included fields with finitely many elements as well as other fields, such as function fields, which were not subfields of the field of complex numbers.

Finally, be careful not to confuse our use of the words *unit* and *unity*. *Unity* is the multiplicative identity element, while a *unit* is any element having a multiplicative inverse. Thus the multiplicative identity element or unity is a unit, but not every unit is unity. For example, -1 is a unit in \mathbb{Z} , but -1 is not unity, that is, $-1 \neq 1$.

■ EXERCISES 22

Computations

In Exercises 1 through 6, compute the product in the given ring.

- | | |
|---|---|
| 1. $(12)(16)$ in \mathbb{Z}_{24}
3. $(11)(-4)$ in \mathbb{Z}_{15}
5. $(2,3)(3,5)$ in $\mathbb{Z}_5 \times \mathbb{Z}_9$ | 2. $(16)(3)$ in \mathbb{Z}_{32}
4. $(20)(-8)$ in \mathbb{Z}_{26}
6. $(-3,5)(2,-4)$ in $\mathbb{Z}_4 \times \mathbb{Z}_{11}$ |
|---|---|

In Exercises 7 through 13, decide whether the indicated operations of addition and multiplication are defined (closed) on the set, and give a ring structure. If a ring is not formed, tell why this is the case. If a ring is formed, state whether the ring is commutative, whether it has unity, and whether it is a field.

7. $n\mathbb{Z}$ with the usual addition and multiplication
8. \mathbb{Z}^+ with the usual addition and multiplication
9. $\mathbb{Z} \times \mathbb{Z}$ with addition and multiplication by components
10. $2\mathbb{Z} \times \mathbb{Z}$ with addition and multiplication by components
11. $\{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$ with the usual addition and multiplication
12. $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ with the usual addition and multiplication
13. The set of all pure imaginary complex numbers ri for $r \in \mathbb{R}$ with the usual addition and multiplication

In Exercises 14 through 19, describe all units in the given ring

14. \mathbb{Z}

15. $\mathbb{Z} \times \mathbb{Z}$

16. \mathbb{Z}_5

17. \mathbb{Q}

18. $\mathbb{Z} \times \mathbb{Q} \times \mathbb{Z}$

19. \mathbb{Z}_4

20. Consider the matrix ring $M_2(\mathbb{Z}_2)$.

- a. Find the **order** of the ring, that is, the number of elements in it.
- b. List all units in the ring.

21. If possible, give an example of a homomorphism $\phi : R \rightarrow R'$ where R and R' are rings with unity $1 \neq 0$ and $1' \neq 0'$, and where $\phi(1) \neq 0'$ and $\phi(1) \neq 1'$.

22. (Linear algebra) Consider the map $\det : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ where $\det(A)$ is the determinant of the matrix A for $A \in M_n(\mathbb{R})$. Is \det a ring homomorphism? Why or why not?

23. Describe all ring homomorphisms of \mathbb{Z} into \mathbb{Z} .

24. Describe all ring homomorphisms of \mathbb{Z} into $\mathbb{Z} \times \mathbb{Z}$.

25. Describe all ring homomorphisms of $\mathbb{Z} \times \mathbb{Z}$ into \mathbb{Z} .

26. How many homomorphisms are there of $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ into \mathbb{Z} ?

27. Consider this solution of the equation $X^2 = I_3$ in the ring $M_3(\mathbb{R})$.

$X^2 = I_3$ implies $X^2 - I_3 = 0$, the zero matrix, so factoring, we have $(X - I_3)(X + I_3) = 0$
whence either $X = I_3$ or $X = -I_3$.

Is this reasoning correct? If not, point out the error, and if possible, give a counterexample to the conclusion.

28. Find all solutions of the equation $x^2 + x - 6 = 0$ in the ring \mathbb{Z}_{14} by factoring the quadratic polynomial. Compare with Exercise 27.

29. Find all solutions to the equations $x^2 + x - 6 = 0$ in the ring \mathbb{Z}_{13} by factoring the quadratic polynomial. Why are there not the same number of solutions in Exercise 28?

30. What is wrong with the following attempt at a proof of Property 2 in Theorem 22.8?

$$(-a)b = ((-1)a)b = (-1)(ab) = -(ab).$$

Concepts

In Exercises 31 and 32, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

31. A **field** F is a ring with nonzero unity such that the set of nonzero elements of F is a group under multiplication.

32. A **unit** in a ring is an element of magnitude 1.

33. Give an example of a ring having two elements a and b such that $ab = 0$ but neither a nor b is zero.

34. Give an example of a ring with unity $1 \neq 0$ that has a subring with nonzero unity $1' \neq 1$. [Hint: Consider a direct product, or a subring of \mathbb{Z}_6 .]

35. Determine whether each of the following is true or false.

- a. Every field is also a ring.
- b. Every ring has a multiplicative identity.
- c. Every ring with unity has at least two units.
- d. Every ring with unity has at most two units.
- e. It is possible for a subset of some field to be a ring but not a subfield, under the induced operations.
- f. The distributive laws for a ring are not very important.
- g. Multiplication in a field is commutative.
- h. The nonzero elements of a field form a group under the multiplication in the field.
- i. Addition in every ring is commutative.
- j. Every element in a ring has an additive inverse.

Theory

36. Show that the multiplication defined on the set F of functions in Example 22.4 satisfies axioms \mathcal{R}_2 and \mathcal{R}_3 for a ring.
37. Show that the evaluation map ϕ_a of Example 22.10 is a ring homomorphism.
38. Complete the argument outlined after Definitions 22.12 to show that isomorphism gives an equivalence relation on a collection of rings.
39. Show that if U is the collection of all units in a ring $\langle R, +, \cdot \rangle$ with unity, then $\langle U, \cdot \rangle$ is a group. [Warning: Be sure to show that U is closed under multiplication.]
40. Show that $a^2 - b^2 = (a + b)(a - b)$ for all a and b in a ring R if and only if R is commutative.
41. Let $\langle R, + \rangle$ be an abelian group. Show that $\langle R, +, \cdot \rangle$ is a ring if we define $ab = 0$ for all $a, b \in R$.
42. Show that the rings $2\mathbb{Z}$ and $3\mathbb{Z}$ are not isomorphic. Show that the fields \mathbb{R} and \mathbb{C} are not isomorphic.
43. (Freshman exponentiation) Let p be a prime. Show that in the ring \mathbb{Z}_p we have $(a + b)^p = a^p + b^p$ for all $a, b \in \mathbb{Z}_p$. [Hint: Observe that the usual binomial expansion for $(a + b)^p$ is valid in a *commutative* ring.]
44. Show that the unity element in a subfield of a field must be the unity of the whole field, in contrast to Exercise 34 for rings.
45. Show that the multiplicative inverse of a unit in a ring with unity is unique.
46. An element a of a ring R is **idempotent** if $a^2 = a$.
 - Show that the set of all idempotent elements of a commutative ring is closed under multiplication.
 - Find all idempotents in the ring $\mathbb{Z}_6 \times \mathbb{Z}_{12}$.
47. (Linear algebra) Recall that for an $m \times n$ matrix A , the *transpose* A^T of A is the matrix whose j th column is the j th row of A . Show that if A is an $m \times n$ matrix such that $A^T A$ is invertible, then the *projection matrix* $P = A(A^T A)^{-1} A^T$ is an idempotent in the ring of $n \times n$ matrices.
48. An element a of a ring R is **nilpotent** if $a^n = 0$ for some $n \in \mathbb{Z}^+$. Show that if a and b are nilpotent elements of a *commutative* ring, then $a + b$ is also nilpotent.
49. Show that a ring R has no nonzero nilpotent element if and only if 0 is the only solution of $x^2 = 0$ in R .
50. Show that a subset S of a ring R gives a subring of R if and only if the following hold:

$$0 \in S;$$

$$(a - b) \in S \text{ for all } a, b \in S;$$

$$ab \in S \text{ for all } a, b \in S.$$

51. a. Show that an intersection of subrings of a ring R is again a subring of R .
 b. Show that an intersection of subfields of a field F is again a subfield of F .
52. Let R be a ring, and let a be a fixed element of R . Let $I_a = \{x \in R \mid ax = 0\}$. Show that I_a is a subring of R .
53. Let R be a ring, and let a be a fixed element of R . Let R_a be the subring of R that is the intersection of all subrings of R containing a (see Exercise 51). The ring R_a is the **subring of R generated by a** . Show that the abelian group $\langle R_a, + \rangle$ is generated (in the sense of Section 7) by $\{a^n \mid n \in \mathbb{Z}^+\}$.
54. (Chinese Remainder Theorem for two congruences) Let r and s be positive integers such that $\gcd(r, s) = 1$. Use the isomorphism in Example 22.15 to show that for $m, n \in \mathbb{Z}$, there exists an integer x such that $x \equiv m \pmod{r}$ and $x \equiv n \pmod{s}$.
 - State and prove the generalization of Example 22.15 for a direct product with n factors.
 - Prove the Chinese Remainder Theorem: Let $a_i, b_i \in \mathbb{Z}^+$ for $i = 1, 2, \dots, n$ and let $\gcd(b_i, b_j) = 1$ for $i \neq j$. Then there exists $x \in \mathbb{Z}^+$ such that $x \equiv a_i \pmod{b_i}$ for $i = 1, 2, \dots, n$.
55. Consider $\langle S, +, \cdot \rangle$, where S is a set and $+$ and \cdot are binary operations on S such that
 - $\langle S, + \rangle$ is a group,
 - $\langle S^*, \cdot \rangle$ is a group where S^* consists of all elements of S except the additive identity element,
 - $a(b + c) = (ab) + (ac)$ and $(a + b)c = (ac) + (bc)$ for all $a, b, c \in S$.

Show that $\langle S, +, \cdot \rangle$ is a division ring. [Hint: Apply the distributive laws to $(1+1)(a+b)$ to prove the commutativity of addition.]

57. A ring R is a **Boolean ring** if $a^2 = a$ for all $a \in R$, so that every element is idempotent. Show that every Boolean ring is commutative.
58. (For students having some knowledge of the laws of set theory) For a set S , let $\mathcal{P}(S)$ be the collection of all subsets of S . Let binary operations $+$ and \cdot on $\mathcal{P}(S)$ be defined by

$$A + B = (A \cup B) - (A \cap B) = \{x \mid x \in A \text{ or } x \in B \text{ but } x \notin (A \cap B)\}$$

and

$$A \cdot B = A \cap B$$

for $A, B \in \mathcal{P}(S)$.

- a. Give the tables for $+$ and \cdot for $\mathcal{P}(S)$, where $S = \{a, b\}$. [Hint: $\mathcal{P}(S)$ has four elements.]
 b. Show that for any set S , $\langle \mathcal{P}(S), +, \cdot \rangle$ is a Boolean ring (see Exercise 57).

SECTION 23

INTEGRAL DOMAINS

While a careful treatment of polynomials is not given until Section 27, for purposes of motivation we shall make intuitive use of them in this section.

Divisors of Zero and Cancellation

One of the most important algebraic properties of our usual number system is that a product of two numbers can be 0 only if at least one of the factors is 0. We have used this fact many times in solving equations, perhaps without realizing that we were using it. Suppose, for example, we are asked to solve the equation

$$x^2 - 5x + 6 = 0.$$

The first thing we do is factor the left side:

$$x^2 - 5x + 6 = (x - 2)(x - 3).$$

Then we conclude that the only possible values for x are 2 and 3. Why? The reason is that if x is replaced by any number a , the product $(a - 2)(a - 3)$ of the resulting numbers is 0 if and only if either $a - 2 = 0$ or $a - 3 = 0$.

23.1 Example Solve the equation $x^2 - 5x + 6 = 0$ in \mathbb{Z}_{12} .

Solution The factorization $x^2 - 5x + 6 = (x - 2)(x - 3)$ is still valid if we think of x as standing for any number in \mathbb{Z}_{12} . But in \mathbb{Z}_{12} , not only is $0a = a0 = 0$ for all $a \in \mathbb{Z}_{12}$, but also

$$\begin{aligned} (2)(6) &= (6)(2) = (3)(4) = (4)(3) = (3)(8) = (8)(3) \\ &= (4)(6) = (6)(4) = (4)(9) = (9)(4) = (6)(6) = (6)(8) \\ &= (8)(6) = (6)(10) = (10)(6) = (8)(9) = (9)(8) = 0. \end{aligned}$$

We find, in fact, that our equation has not only 2 and 3 as solutions, but also 6 and 11, for $(6 - 2)(6 - 3) = (4)(3) = 0$ and $(11 - 2)(11 - 3) = (9)(8) = 0$ in \mathbb{Z}_{12} . ▲

These ideas are of such importance that we formalize them in a definition.

23.2 Definition If a and b are two nonzero elements of a ring R such that $ab = 0$, then a and b are **divisors of 0** (or **0 divisors**). ■

Example 23.1 shows that in \mathbb{Z}_{12} the elements 2, 3, 4, 6, 8, 9, and 10 are divisors of 0. Note that these are exactly the numbers in \mathbb{Z}_{12} that are not relatively prime to 12, that is, whose gcd with 12 is not 1.

If R is a ring with unity and a is a unit in R , then a is not a divisor of 0. To see this, note that if $ab = 0$, then $a^{-1}ab = 0$, so $b = 0$. Similarly, if $ba = 0$, then $baa^{-1} = 0$, so $b = 0$. Theorem 23.3 shows that in the ring \mathbb{Z}_n every element is either 0, a unit, or a 0 divisor.

23.3 Theorem Let $m \in \mathbb{Z}_n$. Either $m = 0$, m is relatively prime to n , in which case m is a unit in \mathbb{Z}_n , or m is not relatively prime to n , in which case m is a 0 divisor in \mathbb{Z}_n .

Proof We first suppose that $m \neq 0$ and $\gcd(m, n) = d \neq 1$. Then, using integer multiplication

$$m\left(\frac{n}{d}\right) = \left(\frac{m}{d}\right)n$$

is a multiple of n , so in \mathbb{Z}_n ,

$$m\left(\frac{n}{d}\right) = 0 \in \mathbb{Z}_n.$$

Neither m nor n/d is 0 in \mathbb{Z}_n . Thus m is a divisor of 0.

Now suppose that $\gcd(m, n) = 1$. Then there are integers a and b such that $an + bm = 1$. By the division algorithm, there are integers q and r such that $0 \leq r \leq n - 1$ and $b = nq + r$. We can write

$$rm = (b - nq)m = bm - nqm = (1 - an) - nqm = 1 - n(a + qm).$$

So in \mathbb{Z}_n , $rm = mr = 1$ and m is a unit. ◆

23.4 Example Classify each nonzero element of \mathbb{Z}_{20} as a unit or a 0 divisor.

Solution The greatest common divisor of m and 20 is 1 if $m = 1, 3, 7, 9, 11, 13, 17, 19$, so these are all units. For $m = 2, 4, 5, 6, 8, 10, 12, 14, 15, 16, 18$, $\gcd(m, 20) > 1$, so these are all 0 divisors. We see that

$$1 \cdot 1 = 3 \cdot 7 = 9 \cdot 9 = 11 \cdot 11 = 13 \cdot 17 = 19 \cdot 19 = 1 \in \mathbb{Z}_{20}$$

which verifies that each is a unit. We also see that

$$2 \cdot 10 = 4 \cdot 5 = 6 \cdot 10 = 8 \cdot 15 = 12 \cdot 5 = 14 \cdot 10 = 16 \cdot 5 = 18 \cdot 10 = 0 \in \mathbb{Z}_{20}$$

which verifies that each of these is a 0 divisor in \mathbb{Z}_{20} . ▲

23.5 Corollary If p is a prime number, then every nonzero element of \mathbb{Z}_p is a unit, which means that \mathbb{Z}_p is a field and it has no divisors of 0.

Proof For any $0 < m \leq p - 1$, $\gcd(m, p) = 1$. So m is a unit in \mathbb{Z}_p by Theorem 23.3. ◆

The preceding corollary shows that when we consider the ring $M_n(\mathbb{Z}_p)$, we are talking about a ring of matrices over a *field*. In the typical undergraduate linear algebra course, only the field properties of the real or complex numbers are used in much of the work. Such notions as matrix reduction to solve linear systems, determinants, Cramer's rule, eigenvalues and eigenvectors, and similarity transformations to try to diagonalize a matrix are valid using matrices over any field; they depend only on the arithmetic properties of a field. Considerations of linear algebra involving notions of magnitude, such

as least-squares approximate solutions or orthonormal bases, make sense only when using fields where we have an idea of magnitude. The relation

$$p \cdot 1 = 1 + 1 + \cdots + 1 = 0$$

p summands

indicates that there can be no very natural notion of magnitude in the field \mathbb{Z}_p .

Another indication of the importance of the concept of 0 divisors is shown in the following theorem. Let R be a ring, and let $a, b, c \in R$. The **cancellation laws** hold in R if $ab = ac$ with $a \neq 0$ implies $b = c$, and $ba = ca$ with $a \neq 0$ implies $b = c$. These are multiplicative cancellation laws. Of course, the additive cancellation laws hold in R , since $(R, +)$ is a group.

23.6 Theorem The cancellation laws hold in a ring R if and only if R has no divisors of 0.

Proof Let R be a ring in which the cancellation laws hold, and suppose $ab = 0$ for some $a, b \in R$. We must show that either a or b is 0. If $a \neq 0$, then $ab = a0$ implies that $b = 0$ by cancellation laws. Therefore, either $a = 0$ or $b = 0$.

Conversely, suppose that R has no divisors of 0, and suppose that $ab = ac$ with $a \neq 0$. Then

$$ab - ac = a(b - c) = 0.$$

Since $a \neq 0$, and since R has no divisors of 0, we must have $b - c = 0$, so $b = c$. A similar argument shows that $ba = ca$ with $a \neq 0$ implies $b = c$. \blacklozenge

Suppose that R is a ring with no divisors of 0. Then an equation $ax = b$, with $a \neq 0$, in R can have at most one solution x in R , for if $ax_1 = b$ and $ax_2 = b$, then $ax_1 = ax_2$, and by Theorem 23.6 $x_1 = x_2$, since R has no divisors of 0. If R has unity $1 \neq 0$ and a is a unit in R with multiplicative inverse a^{-1} , then the solution x of $ax = b$ is $a^{-1}b$. In the case that R is commutative, in particular if R is a field, it is customary to denote $a^{-1}b$ and ba^{-1} (they are equal by commutativity) by the formal quotient b/a . This quotient notation must not be used in the event that R is not commutative, for then we do not know whether b/a denotes $a^{-1}b$ or ba^{-1} . In particular, the multiplicative inverse a^{-1} of a nonzero element a of a field may be written $1/a$.

Integral Domains

The integers are really our most familiar number system. In terms of the algebraic properties we are discussing, \mathbb{Z} is a commutative ring with unity and no divisors of 0. Surely this is responsible for the name that the next definition gives to such a structure.

23.7 Definition An **integral domain** D is a commutative ring with unity $1 \neq 0$ that contains no divisors of 0. \blacksquare

Thus, if the coefficients of a polynomial are from an integral domain, one can solve a polynomial equation in which the polynomial can be factored into linear factors in the usual fashion by setting each factor equal to 0.

In our hierarchy of algebraic structures, an integral domain belongs between a commutative ring with unity and a field, as we shall show. Theorem 23.6 shows that the cancellation laws for multiplication hold in an integral domain.

23.8 Example We have seen that \mathbb{Z} and \mathbb{Z}_p for any prime p are integral domains, but \mathbb{Z}_n is not an integral domain if n is not prime. A moment of thought shows that the direct product $R \times S$ of two nonzero rings R and S is not an integral domain. Just observe that for $r \in R$ and $s \in S$ both nonzero, we have $(r, 0)(0, s) = (0, 0)$. \blacktriangle

23.9 Example Show that although \mathbb{Z}_2 is an integral domain, the matrix ring $M_2(\mathbb{Z}_2)$ has divisors of zero.

Solution We need only observe that

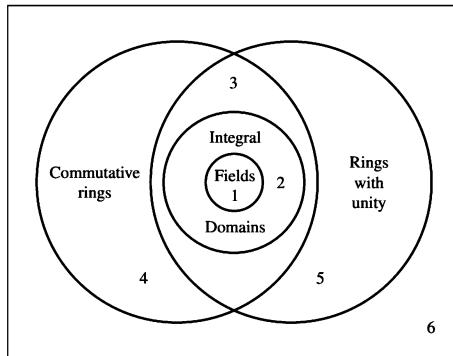
$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$



In a field, every nonzero element is a unit. We saw that units cannot be divisors of 0, so in a field there are no divisors of 0. Since multiplication in a field is commutative, every field is an integral domain.

Figure 23.10 gives a Venn diagram view of containment for the algebraic structures having two binary operations with which we will be chiefly concerned. In Exercise 26 we ask you to redraw this figure to include strictly skew fields as well.

We have seen that \mathbb{Q} , \mathbb{R} , \mathbb{C} , and \mathbb{Z}_p for p a prime number are all fields. Theorem 23.3 implies that if \mathbb{Z}_n is an integral domain, then \mathbb{Z}_n is a field. In fact, the next theorem says that any finite integral domain is a field. The proof of this theorem is a personal favorite of both authors. It is done by counting, one of the most powerful techniques in mathematics.



23.10 Figure A collection of rings.

23.11 Theorem Every finite integral domain is a field.

Proof Let R be a finite integral domain and a a nonzero element of R . We wish to show there is an element $b \in R$ such that $ab = 1$. To this end, we define a function $f : R \rightarrow R$ by

$$f(x) = ax.$$

We first show that f is a one-to-one function. Suppose that $f(x_1) = f(x_2)$, then

$$ax_1 = ax_2$$

$$x_1 = x_2$$

since $a \neq 0$ and cancellation holds in an integral domain. Thus f is one-to-one. Since R is finite and $f : R \rightarrow R$ is one-to-one, f must also map onto R . Therefore, there is a $b \in R$ such that

$$1 = f(a) = ab = ba$$

which verifies that a is a unit. ◆

The finite condition in Theorem 23.11 is necessary since \mathbb{Z} is an infinite integral domain, which is not a field. The counting argument fails in the case where the integral domain is infinite since there are one-to-one functions from an infinite set to itself that are not onto. For example, multiplication by 2 is a one-to-one function mapping \mathbb{Z} to \mathbb{Z} , but 1 is not in the range of the function.

In Section 39 we will see that other than \mathbb{Z}_p there are many finite integral domains and therefore fields.

The Characteristic of a Ring

Let R be any ring. We might ask whether there is a positive integer n such that $n \cdot a = 0$ for all $a \in R$, where $n \cdot a$ means $a + a + \cdots + a$ for n summands, as explained in Section 22. For example, the integer m has this property for the ring \mathbb{Z}_m .

23.12 Definition If for a ring R a positive integer n exists such that $n \cdot a = 0$ for all $a \in R$, then the least such positive integer is the **characteristic of the ring R** . If no such positive integer exists, then R is of **characteristic 0**. ■

We shall use the concept of a characteristic chiefly for fields. Exercise 35 asks us to show that the characteristic of an integral domain is either 0 or a prime p .

23.13 Example The ring \mathbb{Z}_n is of characteristic n , while $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, and \mathbb{C} all have characteristic 0. ▲

At first glance, determination of the characteristic of a ring seems to be a tough job, unless the ring is obviously of characteristic 0. Do we have to examine *every* element a of the ring in accordance with Definition 23.12? Our final theorem of this section shows that if the ring has unity, it suffices to examine only $a = 1$.

23.14 Theorem Let R be a ring with unity. If $n \cdot 1 \neq 0$ for all $n \in \mathbb{Z}^+$, then R has characteristic 0. If $n \cdot 1 = 0$ for some $n \in \mathbb{Z}^+$, then the smallest such integer n is the characteristic of R .

Proof If $n \cdot 1 \neq 0$ for all $n \in \mathbb{Z}^+$, then surely we cannot have $n \cdot a = 0$ for all $a \in R$ for some positive integer n , so by Definition 23.12, R has characteristic 0.

Suppose that n is a positive integer such that $n \cdot 1 = 0$. Then for any $a \in R$, we have

$$n \cdot a = a + a + \cdots + a = a(1 + 1 + \cdots + 1) = a(n \cdot 1) = a0 = 0.$$

Our theorem follows directly. ◆

■ EXERCISES 23

Computations

1. Find all solutions of the equation $x^3 - 2x^2 - 3x = 0$ in \mathbb{Z}_{12} .
2. Solve the equation $3x = 2$ in the field \mathbb{Z}_7 ; in the field \mathbb{Z}_{23} .
3. Find all solutions of the equation $x^2 + 2x + 2 = 0$ in \mathbb{Z}_6 .
4. Find all solutions of $x^2 + 2x + 4 = 0$ in \mathbb{Z}_6 .

In Exercises 5 through 10, find the characteristic of the given ring.

- | | | |
|--|--|--|
| 5. $2\mathbb{Z}$ | 6. $\mathbb{Z} \times \mathbb{Z}$ | 7. $\mathbb{Z}_3 \times 3\mathbb{Z}$ |
| 8. $\mathbb{Z}_3 \times \mathbb{Z}_3$ | 9. $\mathbb{Z}_3 \times \mathbb{Z}_4$ | 10. $\mathbb{Z}_6 \times \mathbb{Z}_{15}$ |

In Exercises 11 through 16, classify each nonzero element of the ring as a unit, a divisor of 0, or neither.

11. \mathbb{Z}_6

12. \mathbb{Z}_8

13. \mathbb{Z}_{15}

14. \mathbb{Z}

15. $\mathbb{Z}_3 \times \mathbb{Z}_3$

16. $\mathbb{Z}_4 \times \mathbb{Z}_5$

17. Let R be a commutative ring with unity of characteristic 4. Compute and simplify $(a + b)^4$ for $a, b \in R$.

18. Let R be a commutative ring with unity of characteristic 3. Compute and simplify $(a + b)^9$ for $a, b \in R$.

19. Let R be a commutative ring with unity of characteristic 3. Compute and simplify $(a + b)^6$ for $a, b \in R$.

20. Show that the matrix $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ is a divisor of zero in $M_2(\mathbb{Z})$.

Concepts

In Exercises 21 and 22, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

21. If $ab = 0$, then a and b are *divisors of zero*.

22. If $n \cdot a = 0$ for all elements a in a ring R , then n is the *characteristic of R* .

23. Determine whether each of the following is true or false.

- a. $n\mathbb{Z}$ has zero divisors if n is not prime.
- b. Every field is an integral domain.
- c. The characteristic of $n\mathbb{Z}$ is n .
- d. As a ring, \mathbb{Z} is isomorphic to $n\mathbb{Z}$ for all $n \geq 1$.
- e. The cancellation law holds in any ring that is isomorphic to an integral domain.
- f. Every integral domain of characteristic 0 is infinite.
- g. The direct product of two integral domains is again an integral domain.
- h. A divisor of zero in a commutative ring with unity can have no multiplicative inverse.
- i. $n\mathbb{Z}$ is a subdomain of \mathbb{Z} .
- j. \mathbb{Z} is a subfield of \mathbb{Q} .

24. Each of the six numbered regions in Fig. 23.10 corresponds to a certain type of a ring. Give an example of a ring in each of the six cells. For example, a ring in the region numbered 3 must be commutative (it is inside the commutative circle), have unity, but not be an integral domain.

25. (For students who have had a semester of linear algebra) Let F be a field. Give five different characterizations of the elements A of $M_n(F)$ that are divisors of 0.

26. Redraw Fig. 23.10 to include a subset corresponding to strictly skew fields.

Proof Synopsis

27. Give a one-sentence synopsis of the proof of the “if” part of Theorem 23.6.

28. Give a two-sentence synopsis of the proof of Theorem 23.11.

Theory

- 29. An element a of a ring R is **idempotent** if $a^2 = a$. Show that a division ring contains exactly two idempotent elements.
- 30. Show that an intersection of subdomains of an integral domain D is again a subdomain of D .
- 31. Show that a finite ring R with unity $1 \neq 0$ and no divisors of 0 is a division ring. (It is actually a field, although commutativity is not easy to prove. See Theorem 32.10.) [Note: In your proof, to show that $a \neq 0$ is a unit, you must show that a “left multiplicative inverse” of $a \neq 0$ in R is also a “right multiplicative inverse.”]

32. Let R be a ring that contains at least two elements. Suppose for each nonzero $a \in R$, there exists a unique $b \in R$ such that $aba = a$.
- Show that R has no divisors of 0.
 - Show that $bab = b$.
 - Show that R has unity.
 - Show that R is a division ring.
33. Show that the characteristic of a subdomain of an integral domain D is equal to the characteristic of D .
34. Show that if D is an integral domain, then $\{n \cdot 1 \mid n \in \mathbb{Z}\}$ is a subdomain of D contained in every subdomain of D .
35. Show that the characteristic of an integral domain D must be either 0 or a prime p . [Hint: If the characteristic of D is mn , consider $(m \cdot 1)(n \cdot 1)$ in D .]
36. This exercise shows that every ring R can be enlarged (if necessary) to a ring S with unity, having the same characteristic as R . Let $S = R \times \mathbb{Z}$ if R has characteristic 0, and $R \times \mathbb{Z}_n$ if R has characteristic n . Let addition in S be the usual addition by components, and let multiplication be defined by

$$(r_1, n_1)(r_2, n_2) = (r_1 r_2 + n_1 \cdot r_2 + n_2 \cdot r_1, n_1 n_2)$$

where $n \cdot r$ has the meaning explained in Section 22.

- Show that S is a ring.
- Show that S has unity.
- Show that S and R have the same characteristic.
- Show that the map $\phi : R \rightarrow S$ given by $\phi(r) = (r, 0)$ for $r \in R$ maps R isomorphically onto a subring of S .

SECTION 24

FERMAT'S AND EULER'S THEOREMS

Fermat's Theorem

We know that as additive groups, \mathbb{Z}_n and $\mathbb{Z}/n\mathbb{Z}$ are naturally isomorphic, with the coset $a + n\mathbb{Z}$ corresponding to a for each $a \in \mathbb{Z}_n$. Furthermore, addition of cosets in $\mathbb{Z}/n\mathbb{Z}$ may be performed by choosing any representatives, adding them in \mathbb{Z} , and finding the coset of $n\mathbb{Z}$ containing their sum. It is easy to see that $\mathbb{Z}/n\mathbb{Z}$ can be made into a ring by multiplying cosets in the same fashion, that is, by multiplying any chosen representatives. While we will be showing this later in a more general situation, we do this special case now. We need only show that such coset multiplication is well defined, because the associativity of multiplication and the distributive laws will follow immediately from those properties of the chosen representatives in \mathbb{Z} . To this end, choose representatives $a + rn$ and $b + sn$, rather than a and b , from the cosets $a + n\mathbb{Z}$ and $b + n\mathbb{Z}$. Then

$$(a + rn)(b + sn) = ab + (as + rb + rs)n,$$

which is also an element of $ab + n\mathbb{Z}$. Thus the multiplication is well-defined, and our cosets form a ring isomorphic to the ring \mathbb{Z}_n .

Exercise 39 in Section 22 asks us to show that the units in a ring form a group under the multiplication operation of the ring. This is a very useful fact that we will use to provide simple proofs for both Fermat's Little Theorem and Euler's generalization. We start with Fermat's Theorem.

24.1 Theorem (Little Theorem of Fermat) If $a \in \mathbb{Z}$ and p is a prime not dividing a , then p divides $a^{p-1} - 1$, that is, $a^{p-1} \equiv 1 \pmod{p}$ for $a \not\equiv 0 \pmod{p}$.

Proof The ring \mathbb{Z}_p is a field, which implies that all the nonzero elements are units. Thus $\langle \mathbb{Z}_p^*, \cdot \rangle$ is a group with $p - 1$ elements. Any b in the group \mathbb{Z}_p^* has order a divisor of $|\mathbb{Z}_p^*| = p - 1$. Therefore

$$b^{p-1} = 1 \in \mathbb{Z}_p.$$

The rings \mathbb{Z}_p and $\mathbb{Z}/p\mathbb{Z}$ are isomorphic where the element $b \in \mathbb{Z}_p$ corresponds to the coset $b + p\mathbb{Z}$. For any integer a that is not a multiple of p , $a + p\mathbb{Z} = b + p\mathbb{Z}$ for some $0 \leq b \leq p - 1$. Thus

$$(a + p\mathbb{Z})^{p-1} = (b + p\mathbb{Z})^{p-1} = 1 + p\mathbb{Z} \in \mathbb{Z}/p\mathbb{Z}.$$

In other words,

$$a^{p-1} \equiv 1 \pmod{p}. \quad \blacklozenge$$

24.2 Corollary If $a \in \mathbb{Z}$, then $a^p \equiv a \pmod{p}$ for any prime p .

Proof The corollary follows from Theorem 24.1 if $a \not\equiv 0 \pmod{p}$. If $a \equiv 0 \pmod{p}$, then both sides reduce to 0 modulo p . \blacklozenge

24.3 Example Let us compute the remainder of 8^{103} when divided by 13. Using Fermat's theorem, we have

$$\begin{aligned} 8^{103} &\equiv (8^{12})^8(8^7) \equiv (1^8)(8^7) \equiv 8^7 \equiv (-5)^7 \\ &\equiv (25)^3(-5) \equiv (-1)^3(-5) \equiv 5 \pmod{13}. \end{aligned} \quad \blacktriangle$$

■ HISTORICAL NOTE

The statement of Theorem 24.1 occurs in a letter from Pierre de Fermat (1601–1665) to Bernard Frenicle de Bessy, dated 18 October 1640. Fermat's version of the theorem was that for any prime p and any geometric progression $a, a^2, \dots, a^t, \dots$, there is a least number a^T of the progression such that p divides $a^T - 1$. Furthermore, T divides $p - 1$ and p also divides all numbers of the form $a^{KT} - 1$. (It is curious that Fermat failed to note the condition that p not divide a ; perhaps he felt that it was obvious that the result fails in that case.)

Fermat did not in the letter or elsewhere indicate a proof of the result and, in fact, never mentioned it again. But we can infer from other parts

of this correspondence that Fermat's interest in this result came from his study of perfect numbers. (A perfect number is a positive integer m that is the sum of all of its divisors less than m ; for example, $6 = 1 + 2 + 3$ is a perfect number.) Euclid had shown that $2^{n-1}(2^n - 1)$ is perfect if $2^n - 1$ is prime. The question then was to find methods for determining whether $2^n - 1$ was prime. Fermat noted that $2^n - 1$ was composite if n is composite, and then derived from his theorem the result that if n is prime, the only possible divisors of $2^n - 1$ are those of the form $2kn + 1$. From this result he was able quickly to show, for example, that $2^{37} - 1$ was divisible by $223 = 2 \cdot 3 \cdot 37 + 1$.

24.4 Example Show that $2^{11,213} - 1$ is not divisible by 11.

Solution By Fermat's theorem, $2^{10} \equiv 1 \pmod{11}$, so

$$\begin{aligned} 2^{11,213} - 1 &\equiv [(2^{10})^{1,121} \cdot 2^3] - 1 \equiv [1^{1,121} \cdot 2^3] - 1 \\ &\equiv 2^3 - 1 \equiv 8 - 1 \equiv 7 \pmod{11}. \end{aligned}$$

Thus the remainder of $2^{11,213} - 1$ when divided by 11 is 7, not 0. (The number 11,213 is prime, and it has been shown that $2^{11,213} - 1$ is a prime number. Primes of the form $2^p - 1$ where p is prime are known as **Mersenne primes**.) \blacktriangle

24.5 Example Show that for every integer n , the number $n^{33} - n$ is divisible by 15.

Solution This seems like an incredible result. It means that 15 divides $2^{33} - 2, 3^{33} - 3, 4^{33} - 4$, etc.

Now $15 = 3 \cdot 5$, and we shall use Fermat's theorem to show that $n^{33} - n$ is divisible by both 3 and 5 for every n . Note that $n^{33} - n = n(n^{32} - 1)$.

If 3 divides n , then surely 3 divides $n(n^{32} - 1)$. If 3 does not divide n , then by Fermat's theorem, $n^2 \equiv 1 \pmod{3}$ so

$$n^{32} - 1 \equiv (n^2)^{16} - 1 \equiv 1^{16} - 1 \equiv 0 \pmod{3},$$

and hence 3 divides $n^{32} - 1$.

If $n \equiv 0 \pmod{5}$, then $n^{33} - n \equiv 0 \pmod{5}$. If $n \not\equiv 0 \pmod{5}$, then by Fermat's theorem, $n^4 \equiv 1 \pmod{5}$, so

$$n^{32} - 1 \equiv (n^4)^8 - 1 \equiv 1^8 - 1 \equiv 0 \pmod{5}.$$

Thus $n^{33} - n \equiv 0 \pmod{5}$ for every n also. ▲

Euler's Generalization

Theorem 23.3 classifies all the elements in \mathbb{Z}_n into three categories. An element k in \mathbb{Z}_n is either 0, a unit if the $\gcd(n, k) = 1$, or else a divisor of 0 if $\gcd(n, k) > 1$. Exercise 39 in Section 22 shows that the units in a ring form a group under multiplication. Therefore, the set of nonzero elements in \mathbb{Z}_n , which are relatively prime to n , form a multiplicative group. Euler's generalization of Fermat's theorem is based on the number of units in \mathbb{Z}_n .

Let n be a positive integer. Let $\varphi(n)$ be defined as the number of positive integers less than or equal to n and relatively prime to n . Note that $\varphi(1) = 1$.

24.6 Example Let $n = 12$. The positive integers less than or equal to 12 and relatively prime to 12 are 1, 5, 7, and 11, so $\varphi(12) = 4$. ▲

By Theorem 23.3, $\varphi(n)$ is the number of nonzero elements of \mathbb{Z}_n that are not divisors of 0. This function $\varphi : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ is the **Euler phi-function**. We can now describe Euler's generalization of Fermat's theorem.

24.7 Theorem (Euler's Theorem) If a is an integer relatively prime to n , then $a^{\varphi(n)} - 1$ is divisible by n , that is, $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Proof If a is relatively prime to n , then the coset $a + n\mathbb{Z}$ of $n\mathbb{Z}$ containing a contains an integer $b < n$ and relatively prime to n . Using the fact that multiplication of these cosets by multiplication modulo n of representatives is well-defined, we have

$$a^{\varphi(n)} \equiv b^{\varphi(n)} \pmod{n}.$$

But by Theorem 23.3, b can be viewed as an element of the multiplicative group G_n of order $\varphi(n)$ consisting of the $\varphi(n)$ elements of \mathbb{Z}_n relatively prime to n . Thus

$$b^{\varphi(n)} \equiv 1 \pmod{n},$$

and our theorem follows. ◆

24.8 Example Let $n = 12$. We saw in Example 24.6 that $\varphi(12) = 4$. Thus if we take any integer a relatively prime to 12, then $a^4 \equiv 1 \pmod{12}$. For example, with $a = 7$, we have $7^4 = (49)^2 = 2,401 = 12(200) + 1$, so $7^4 \equiv 1 \pmod{12}$. Of course, the easy way to compute $7^4 \pmod{12}$, without using Euler's theorem, is to compute it in \mathbb{Z}_{12} . In \mathbb{Z}_{12} , we have $7 = -5$ so

$$7^2 = (-5)^2 = (5)^2 = 1 \quad \text{and} \quad 7^4 = 1^2 = 1.$$
▲

Application to $ax \equiv b \pmod{m}$

We can find all solutions of a linear congruence $ax \equiv b \pmod{m}$. We prefer to work with an equation in \mathbb{Z}_m and interpret the results for congruences.

24.9 Theorem Let m be a positive integer and let $a \in \mathbb{Z}_m$ be relatively prime to m . For each $b \in \mathbb{Z}_m$, the equation $ax = b$ has a unique solution in \mathbb{Z}_m .

Proof By Theorem 23.3, a is a unit in \mathbb{Z}_m and $s = a^{-1}b$ is certainly a solution of the equation. Multiplying both sides of $ax = b$ on the left by a^{-1} , we see this is the only solution. ◆

Interpreting this theorem for congruences, we obtain at once the following corollary.

24.10 Corollary If a and m are relatively prime integers, then for any integer b , the congruence $ax \equiv b \pmod{m}$ has as solutions all integers in precisely one residue class modulo m . ◆

Theorem 24.9 serves as a lemma for the general case.

24.11 Theorem Let m be a positive integer and let $a, b \in \mathbb{Z}_m$. Let d be the gcd of a and m . The equation $ax = b$ has a solution in \mathbb{Z}_m if and only if d divides b . When d divides b , the equation has exactly d solutions in \mathbb{Z}_m .

Proof First we show there is no solution of $ax = b$ in \mathbb{Z}_m unless d divides b . Suppose $s \in \mathbb{Z}_m$ is a solution. Then $as - b = qm$ in \mathbb{Z} , so $b = as - qm$. Since d divides both a and m , we see that d divides the right-hand side of the equation $b = as - qm$, and hence divides b . Thus a solution s can exist only if d divides b .

Suppose now that d does divide b . Let

$$a = a_1d, \quad b = b_1d, \quad \text{and} \quad m = m_1d.$$

Then the equation $as - b = qm$ in \mathbb{Z} can be rewritten as $d(a_1s - b_1) = dqm_1$. We see that $a_1s - b_1$ is a multiple of m_1 if and only if $a_1s - b_1$ is a multiple of m_1 . Thus the solutions s of $ax = b$ in \mathbb{Z}_m are precisely the elements that, read modulo m_1 , yield solutions of $a_1x = b_1$ in \mathbb{Z}_{m_1} . Now let $s \in \mathbb{Z}_{m_1}$ be the unique solution of $a_1x = b_1$ in \mathbb{Z}_{m_1} given by Theorem 24.9. The numbers in \mathbb{Z}_m that reduce to s modulo m_1 are precisely those that can be computed in \mathbb{Z}_m as

$$s, s + m_1, s + 2m_1, s + 3m_1, \dots, s + (d-1)m_1.$$

Thus there are exactly d solutions of the equation in \mathbb{Z}_m . ◆

Theorem 24.11 gives us at once this classical result on the solutions of a linear congruence.

24.12 Corollary Let d be the gcd of positive integers a and m . The congruence $ax \equiv b \pmod{m}$ has a solution if and only if d divides b . When this is the case, the solutions are the integers in exactly d distinct residue classes modulo m . ◆

Actually, our proof of Theorem 24.11 shows a bit more about the solutions of $ax \equiv b \pmod{m}$ than we stated in this corollary; namely, it shows that if any solution s is found, then the solutions are precisely all elements of the residue classes $(s + km_1) + (m\mathbb{Z})$ where $m_1 = m/d$ and k runs through the integers from 0 to $d-1$. It also tells us that we can find such an s by finding $a_1 = a/d$ and $b_1 = b/d$, and solving $a_1x \equiv b_1 \pmod{m_1}$. To solve this congruence, we may consider a_1 and b_1 to be replaced by their remainders modulo m_1 and solve the equation $a_1x = b_1$ in \mathbb{Z}_{m_1} .

24.13 Example Find all solutions of the congruence $12x \equiv 27 \pmod{18}$.

Solution The gcd of 12 and 18 is 6, and 6 is not a divisor of 27. Thus by the preceding corollary, there are no solutions. \blacktriangleleft

24.14 Example Find all solutions of the congruence $15x \equiv 27 \pmod{18}$.

Solution The gcd of 15 and 18 is 3, and 3 does divide 27. Proceeding as explained before Example 24.13, we divide everything by 3 and consider the congruence $5x \equiv 9 \pmod{6}$, which amounts to solving the equation $5x = 3$ in \mathbb{Z}_6 . Now the units in \mathbb{Z}_6 are 1 and 5, and 5 is clearly its own inverse in this group of units. Thus the solution in \mathbb{Z}_6 is $x = (5^{-1})(3) = (5)(3) = 3$. Consequently, the solutions of $15x \equiv 27 \pmod{18}$ are the integers in the three residue classes

$$3 + 18\mathbb{Z} = \{\dots, -33, -15, 3, 21, 39, \dots\},$$

$$9 + 18\mathbb{Z} = \{\dots, -27, -9, 9, 27, 45, \dots\}.$$

$$15 + 18\mathbb{Z} = \{\dots, -21, -3, 15, 33, 51, \dots\},$$

illustrating Corollary 24.12. Note the $d = 3$ solutions 3, 9, and 15 in \mathbb{Z}_{18} . All the solutions in the three displayed residue classes modulo 18 can be collected in the one residue class $3 + 6\mathbb{Z}$ modulo 6, for they came from the solution $x = 3$ of $5x = 3$ in \mathbb{Z}_6 . \blacktriangleleft

■ EXERCISES 24

Computations

We will see later that the multiplicative group of nonzero elements of a finite field is cyclic. Illustrate this by finding a generator for this group for the given finite field.

1. \mathbb{Z}_7
 2. \mathbb{Z}_{11}
 3. \mathbb{Z}_{17}
 4. Using Fermat's theorem, find the remainder of 3^{47} when it is divided by 23.
 5. Use Fermat's theorem to find the remainder of 37^{49} when it is divided by 7.
 6. Compute the remainder of $2^{(2^{17})} + 1$ when divided by 19. [Hint: You will need to compute the remainder of 2^{17} modulo 18.]
 7. Make a table of values of $\varphi(n)$ for $n \leq 30$.
 8. Compute $\varphi(p^2)$ where p is a prime.
 9. Compute $\varphi(pq)$ where both p and q are primes.
 10. Use Euler's generalization of Fermat's theorem to find the remainder of 7^{1000} when divided by 24.
- In Exercises 11 through 18, describe all solutions of the given congruence, as we did in Examples 24.13 and 24.14.
11. $2x \equiv 6 \pmod{4}$
 12. $22x \equiv 5 \pmod{15}$
 13. $36x \equiv 15 \pmod{24}$
 14. $45x \equiv 15 \pmod{24}$
 15. $39x \equiv 125 \pmod{9}$
 16. $41x \equiv 125 \pmod{9}$
 17. $155x \equiv 75 \pmod{65}$
 18. $39x \equiv 52 \pmod{130}$
 19. Let p be a prime ≥ 3 . Use Exercise 28 below to find the remainder of $(p - 2)!$ modulo p .
 20. Using Exercise 28 below, find the remainder of $34!$ modulo 37.
 21. Using Exercise 28 below, find the remainder of $49!$ modulo 53.
 22. Using Exercise 28 below, find the remainder of $24!$ modulo 29.

Concepts

23. Determine whether each of the following is true or false.
- $a^{p-1} \equiv 1 \pmod{p}$ for all integers a and primes p .
 - $a^{p-1} \equiv 1 \pmod{p}$ for all integers a such that $a \not\equiv 0 \pmod{p}$ for a prime p .
 - $\varphi(n) \leq n$ for all $n \in \mathbb{Z}^+$.
 - $\varphi(n) \leq n - 1$ for all $n \in \mathbb{Z}^+$.
 - The units in \mathbb{Z}_n are the positive integers less than n and relatively prime to n .
 - The product of two units in \mathbb{Z}_n is always a unit.
 - The product of two nonunits in \mathbb{Z}_n may be a unit.
 - The product of a unit and a nonunit in \mathbb{Z}_n is never a unit.
 - Every congruence $ax \equiv b \pmod{p}$, where p is a prime, has a solution.
 - Let d be the gcd of positive integers a and m . If d divides b , then the congruence $ax \equiv b \pmod{m}$ has exactly d incongruent solutions.
24. Give the group multiplication table for the multiplicative group of units in \mathbb{Z}_{12} . To which group of order 4 is it isomorphic?

Proof Synopsis

25. Give a one-sentence synopsis of the proof of Theorem 24.1.
 26. Give a one-sentence synopsis of the proof of Theorem 24.7.

Theory

27. Show that 1 and $p - 1$ are the only elements of the field \mathbb{Z}_p that are their own multiplicative inverse. [Hint: Consider the equation $x^2 - 1 = 0$.]
28. Using Exercise 27, deduce the half of Wilson's theorem that states that if p is a prime, then $(p - 1)! \equiv -1 \pmod{p}$. [The other half states that if n is an integer > 1 such that $(n - 1)! \equiv -1 \pmod{n}$, then n is a prime. Just think what the remainder of $(n - 1)!$ would be modulo n if n is not a prime.]
29. Use Fermat's theorem to show that for any positive integer n , the integer $n^{37} - n$ is divisible by 383838. [Hint: $383838 = (37)(19)(13)(7)(3)(2)$.]
30. Referring to Exercise 29, find a number larger than 383838 that divides $n^{37} - n$ for all positive integers n .

SECTION 25 ENCRYPTION

An encryption scheme is a method to disguise a message so that it is extremely difficult for anyone other than the intended receiver to read. The sender **encrypts** the message and the receiver **decrypts** the message. One method, called cypher encryption, requires the sender to use a permutation of the letters in the alphabet to replace each letter with a different letter. The receiver then uses the inverse of the permutation to recover the original message. This method has two major weaknesses. First, both the sender and the receiver need to know the permutation, but no one else should know the permutation or else the message is not secure. It would be difficult to implement a cypher for a transaction when a company wishes to receive many orders each day, each using a different permutation that only the customer and the company know. Furthermore, cyphers are generally not difficult to crack. In fact, some newspapers carry a daily puzzle, which is essentially decrypting an encrypted message.

Researchers in the second half of the twentieth century sought a method that allows the receiver to publish public information that any sender could use to encrypt a message, yet only the receiver could decrypt it. This means that knowing how a message was encrypted is little help in decryption. This method relies on a function that is easy for computers to compute, but whose inverse is virtually impossible to compute without

more information. Functions of this type are called **trap door functions**. Most commercial online transactions are communicated with trap door functions. This allows anyone to make a secure credit card purchase with little risk of a third party gaining private information.

RSA Public and Private Keys

Euler's generalization of Fermat's Theorem is the basis of a very common trap door encryption scheme referred to as **RSA encryption**. RSA comes from the names of the three inventors of the system, Ron Rivest, Adi Shamir, and Leonard Adleman. The trap door function relies on the fact that it is easy to multiply two large prime numbers, but if you are only given their product, it is very difficult to factor the number to recover the two prime numbers. The following theorem is the key to this encryption scheme.

25.1 Theorem Let $n = pq$ where p and q are distinct prime numbers. If $a \in \mathbb{Z}$ with $\gcd(a, pq) = 1$ and $w \equiv 1 \pmod{(p-1)(q-1)}$, then $a^w \equiv a \pmod{n}$.

Proof Since $w \equiv 1 \pmod{(p-1)(q-1)}$, we can write

$$w = k(p-1)(q-1) + 1$$

for some integer k . Recall that the Euler phi-function $\phi(n)$ counts the number of positive integers less than or equal to n that are relatively prime to n . Since $n = pq$, we can compute $\phi(pq)$ by subtracting the number of integers less than n that are divisible by either p or q from $n - 1$. There are $p - 1$ multiples of q and $q - 1$ multiples of p that are less than pq . Furthermore, the least common multiple of p and q is pq since p and q are distinct primes. Thus

$$\begin{aligned}\phi(pq) &= (pq - 1) - (p - 1) - (q - 1) \\ &= pq - p - q + 1 \\ &= (p - 1)(q - 1).\end{aligned}$$

By Euler's Theorem (Theorem 24.7),

$$\begin{aligned}a^w &= a^{k(p-1)(q-1)+1} \\ &= a(a^{(p-1)(q-1)})^k \\ &= a(a^{\phi(n)})^k \\ &\equiv a(1^k) \\ &\equiv a \pmod{n}.\end{aligned}$$
◆

The RSA encryption scheme requires two sets of positive integers called the **private key** and the **public key**. The private key is known only by the person who will receive the message, and the public key is available to anyone who wishes to send a message to the receiver.

The **private key** consists of

- Two prime numbers p and q with $p \neq q$.
- The product $n = pq$.
- An integer $1 < r < (p-1)(q-1) - 1$ that is relatively prime to $(p-1)(q-1)$.

We know that r has an inverse in $\mathbb{Z}_{(p-1)(q-1)}$ since r is relatively prime to $(p-1)(q-1)$.

The public key consists of

- The integer s where $1 < s < (p - 1)(q - 1)$ and s is the inverse of r in $\mathbb{Z}_{(p-1)(q-1)}$.
- The product $n = pq$.

The public key does not include p , q , r , or $(p - 1)(q - 1)$. Knowing any of these numbers and the numbers in the public key would make it relatively easy to decrypt any encrypted message.

We can now give the encryption and decryption algorithms. The sender wishes to send a message to the receiver. We will assume the message is simply a number between 2 and $n - 1$. To send a text message, the sender would use a standard way of representing the text as a number, such as the ASCII code. A long text would be broken up into smaller texts so that each would be coded as a number in the allowable range 2 to $n - 1$ and each would be sent separately. Let $2 \leq m \leq n - 1$ be the message to be sent.

Encryption Using the public key, the sender encrypts the message as a number $0 \leq y \leq n - 1$ to be sent to the receiver where

$$y \equiv m^s \pmod{n}.$$

That is, the sender computes y to be the remainder when m^s is divided by n and sends y to the receiver.

Decryption Using the private key, the receiver decrypts y , the message received from the sender, by computing

$$y^r \pmod{n},$$

the remainder when y^r is divided by n . Since $rs \equiv 1 \pmod{(p-1)(q-1)}$, Theorem 25.1 says,

$$y^r = (m^s)^r = m^{rs} \equiv m \pmod{n}.$$

Thus the receiver reconstructs the original message m .

Of course, in practice the prime numbers p and q are very large. As of the writing of this book it is thought that prime numbers requiring 4096 bits or approximately 1200 digits are sufficient to make the RSA scheme secure. To illustrate how the process works, we will use small primes.

25.2 Example Let $p = 17$ and $q = 11$. The private key consists of

- $p = 17$, $q = 11$,
- $n = pq = 187$ and
- a number r relatively prime to $(p - 1)(q - 1) = 160$. For this example we take $r = 23$.

The public key consists of

- $n = 187$ and
- $s = 7$. A little calculation shows that $23 \cdot 7 = 161 = 160 + 1 \equiv 1 \pmod{160}$, which implies that $s = 7$. Since the public key consists of only n and s , $(p - 1)(q - 1)$ is unknown to all but the receiver. Without knowing $(p - 1)(q - 1)$, the value of r cannot be determined from the value of s .

Suppose the sender wishes to send the message $m = 2$ to the receiver. The message is encrypted by computing

$$y = 2^7 \equiv 128 \pmod{187}.$$

The receiver recovers the original message by computing

$$128^{23} \equiv 2 \pmod{187}. \quad \blacktriangle$$

In Example 25.2 some of the computations would be long and tedious without the use of a computer. For large primes p and q , it is essential to have an efficient algorithm to compute $m^s \pmod{n}$ and $y^r \pmod{n}$. This can be accomplished by using base 2. We illustrate with the following example.

- 25.3 Example** In Example 25.2 we needed to compute $128^{23} \pmod{187}$. We can compute this value by expressing 23 in base 2, $23 = 16 + 4 + 2 + 1$, and then computing the following:

$$\begin{aligned} 128^1 &= 128 \\ 128^2 &= 1638 \equiv 115 \pmod{187} \\ 128^4 &= (128^2)^2 \equiv 115^2 \equiv 135 \pmod{187} \\ 128^8 &= (128^4)^2 \equiv 135^2 \equiv 86 \pmod{187} \\ 128^{16} &= (128^8)^2 \equiv 86^2 \equiv 103 \pmod{187}, \end{aligned}$$

Thus

$$\begin{aligned} 128^{23} &\equiv 128^{16+4+2+1} \\ &\equiv (128^{16}128^4)(128^2128^1) \\ &\equiv (103 \cdot 135)(115 \cdot 128) \\ &\equiv 67 \cdot 134 \\ &\equiv 2 \pmod{187}. \quad \blacktriangle \end{aligned}$$

As illustrated in the above example, this method gives a more efficient computation of $a^k \pmod{n}$.

The Euclidean algorithm is a simple and efficient way to compute the inverse of a unit in $\mathbb{Z}_{(p-1)(q-1)}$. It involves the repeated use of the division algorithm. However, we will not discuss the Euclidean algorithm here.

The reader may have noticed a potential flaw in the RSA encryption scheme. It is possible that m is a multiple of either p or q . In that case, $m^{(p-1)(q-1)} \not\equiv 1 \pmod{n}$, which means that m^r may not be equivalent to m modulo n . In this case RSA encryption fails. However, when using large prime numbers the probability that the message is a multiple of p or q is extremely low. If one is concerned about this issue, the algorithm could be modified slightly to be sure that the message is smaller than both p and q .

How are the large prime numbers p and q in RSA encryption found? Basically, the process is to guess a value and check that it is prime. Unfortunately, there is no known fast method to test for primality, but it is possible to do a fast probabilistic test. One simple probabilistic test uses Fermat's Theorem (Theorem 24.1). The idea is to generate a random positive integer less than p and check if $a^{p-1} \equiv 1 \pmod{p}$. If p is prime, then $a^{p-1} \equiv 1 \pmod{p}$, so if $a^{p-1} \not\equiv 1 \pmod{p}$, then p is not a prime number and the number p is rejected. On the other hand, if $a^{p-1} \equiv 1 \pmod{p}$, then p passes the test and p could be a prime. If p passes the test, we repeat the process for a different random value of a . The probability that a composite number p is picked given that p passes the test several times is low enough to safely assume that p is prime.

■ EXERCISES 25

In Exercises 1 through 8, the notation is consistent with the notation used in the text for RSA encryption. It may be helpful to use a calculator or computer.

1. Let $p = 3$ and $q = 5$. Find n , and all possible pairs (r, s) .
2. Let $p = 3$ and $q = 7$. Find n and all possible pairs (r, s) .
3. Let $p = 3$ and $q = 11$. Find n and all possible pairs (r, s) .
4. Let $p = 5$ and $q = 7$. Find n and all possible pairs (r, s) .
5. Let $p = 13$, $q = 17$, and $r = 5$. Find the value of s .
6. For RSA encryption it is assumed that the message m is at least 2. Why should m not be 1?
7. The public key is $n = 143$ and $s = 37$.
 - a. Compute the value of y if the message is $m = 25$.
 - b. Find r . (Computer Algebra Systems have built-in functions to compute in \mathbb{Z}_m .)
 - c. Use your answers to parts a) and b) to decrypt y .
8. The public key is $n = 1457$ and $s = 239$.
 - a. Compute the value of y if the message is $m = 999$.
 - b. Find r . (Computer Algebra Systems have built-in functions to compute in \mathbb{Z}_m .)
 - c. Use your answers to parts a) and b) to decrypt y .
9. For $p = 257$, $q = 359$, and $r = 1493$ identify the private and public keys.

This page is intentionally left blank

Constructing Rings and Fields

-
- Section 26 The Field of Quotients of an Integral Domain
 - Section 27 Rings of Polynomials
 - Section 28 Factorization of Polynomials over a Field
 - Section 29 Algebraic Coding Theory
 - Section 30 Homomorphisms and Factor Rings
 - Section 31 Prime and Maximal Ideals
 - Section 32 Noncommutative Examples

SECTION 26

THE FIELD OF QUOTIENTS OF AN INTEGRAL DOMAIN

Let L be a field and D a subring of L that contains the unity. The ring D is an integral domain since it has no zero divisors. Also F , the set of all quotients of the form $\frac{a}{b}$ with a and $b \neq 0$ both in D , forms a subfield of L . The field F is called a *field of quotients of the integral domain D* .

26.1 Example Let $L = \mathbb{R}$. If $D = \mathbb{Z}$, then

$$F = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0 \right\} = \mathbb{Q}$$

which is a field.

If $D = \{x + y\sqrt{2} \mid x, y \in \mathbb{Z}\}$, then

$$F = \left\{ \frac{a}{b} \mid a, b \in D, b \neq 0 \right\} = \left\{ \frac{x + y\sqrt{2}}{z + w\sqrt{2}} \mid x, y, z, w \in \mathbb{Z}, z + w\sqrt{2} \neq 0 \right\}.$$

By rationalizing the denominator we see that

$$F = \left\{ r + s\sqrt{2} \mid r, s \in \mathbb{Q} \right\}$$

which is a field by Exercise 12 in Section 22. ▲

In this section, we start with an integral domain D and construct a field F . We then show that D is isomorphic with a subring D' of F and that F consists of all quotients $\frac{a}{b}$ with $a, b \in D', b \neq 0$. Thus we can think of any integral domain as being a subring of a field and every element of the field is the quotient of elements from the integral domain.

The Construction

Let D be an integral domain that we desire to enlarge to a field of quotients F . A coarse outline of the steps we take is as follows:

1. Define what the elements of F are to be.
2. Define the binary operations of addition and multiplication on F .

3. Check all the field axioms to show that F is a field under these operations.
4. Show that F can be viewed as containing D as an integral subdomain.

Steps 1, 2, and 4 are very interesting, and Step 3 is largely a mechanical chore. We proceed with the construction.

Step 1 Let D be a given integral domain, and form the Cartesian product

$$D \times D = \{(a, b) \mid a, b \in D\}$$

We are going to think of an ordered pair (a, b) as representing a *formal quotient* a/b , that is, if $D = \mathbb{Z}$, the pair $(2, 3)$ will eventually represent the number $\frac{2}{3}$ for us. The pair $(2, 0)$ represents no element of \mathbb{Q} and suggests that we cut the set $D \times D$ down a bit. Let S be the subset of $D \times D$ given by

$$S = \{(a, b) \mid a, b \in D, b \neq 0\}.$$

Now S is still not going to be our field as is indicated by the fact that, with $D = \mathbb{Z}$, different pairs of integers such as $(2, 3)$ and $(4, 6)$ can represent the same rational number. We next define when two elements of S will eventually represent the same element of F , or, as we shall say, when two elements of S are *equivalent*.

26.2 Definition Two elements (a, b) and (c, d) in S are **equivalent**, denoted by $(a, b) \sim (c, d)$, if and only if $ad = bc$. ■

Observe that this definition is reasonable, since the criterion for $(a, b) \sim (c, d)$ is an equation $ad = bc$ involving elements in D and concerning the known multiplication in D . Note also that for $D = \mathbb{Z}$, the criterion gives us our usual definition of *equality* of $\frac{a}{b}$ with $\frac{c}{d}$, for example, $\frac{2}{3} = \frac{4}{6}$ since $(2)(6) = (3)(4)$. The rational number that we usually denote by $\frac{2}{3}$ can be thought of as the collection of *all* quotients of integers that reduce to, or are equivalent to, $\frac{2}{3}$.

26.3 Lemma The relation \sim between elements of the set S as just described is an equivalence relation.

Proof We must check the three properties of an equivalence relation.

Reflexive $(a, b) \sim (a, b)$ since $ab = ba$, for multiplication in D is commutative.

Symmetric If $(a, b) \sim (c, d)$, then $ad = bc$. Since multiplication in D is commutative, we deduce that $cb = da$, and consequently $(c, d) \sim (a, b)$.

Transitive If $(a, b) \sim (c, d)$ and $(c, d) \sim (r, s)$, then $ad = bc$ and $cs = dr$. Using these relations and the fact that multiplication in D is commutative, we have

$$asd = sad = sbc = bcs = bdr = brd.$$

Now $d \neq 0$, and D is an integral domain, so cancellation is valid; this is a crucial step in the argument. Hence from $asd = brd$ we obtain $as = br$, so that $(a, b) \sim (r, s)$. ◆

We now know, in view of Theorem 0.22, that \sim gives a partition of S into equivalence classes. To avoid long bars over extended expressions, we shall let $[(a, b)]$, rather than (a, b) , be the equivalence class of (a, b) in S under the relation \sim . We now finish Step 1 by defining F to be the set of all equivalence classes $[(a, b)]$ for $(a, b) \in S$.

Step 2 The next lemma serves to define addition and multiplication in F . Observe that if $D = \mathbb{Z}$ and $[(a, b)]$ is viewed as $(a/b) \in \mathbb{Q}$, these definitions applied to \mathbb{Q} give the usual operations.

26.4 Lemma For $[(a, b)]$ and $[(c, d)]$ in F , the equations

$$[(a, b)] + [(c, d)] = [(ad + bc, bd)]$$

and

$$[(a, b)][(c, d)] = [(ac, bd)]$$

give well-defined operations of addition and multiplication on F .

Proof Observe first that if $[(a, b)]$ and $[(c, d)]$ are in F , then (a, b) and (c, d) are in S , so $b \neq 0$ and $d \neq 0$. Because D is an integral domain, $bd \neq 0$, so both $(ad + bc, bd)$ and (ac, bd) are in S . (Note the crucial use here of the fact that D has no divisors of 0.) This shows that the right-hand sides of the defining equations are at least in F .

It remains for us to show that these operations of addition and multiplication are well defined. That is, they were defined by means of representatives in S of elements of F , so we must show that if different representatives in S are chosen, the same element of F will result. To this end, suppose that $(a_1, b_1) \in [(a, b)]$ and $(c_1, d_1) \in [(c, d)]$. We must show that

$$(a_1d_1 + b_1c_1, b_1d_1) \in [(ad + bc, bd)]$$

and

$$(a_1c_1, b_1d_1) \in [(ac, bd)].$$

Now $(a_1, b_1) \in [(a, b)]$ means that $(a_1, b_1) \sim (a, b)$; that is,

$$a_1b = b_1a.$$

Similarly, $(c_1, d_1) \in [(c, d)]$ implies that

$$c_1d = d_1c.$$

To get a “common denominator” (common second member) for the four pairs (a, b) , (a_1, b_1) , (c, d) , and (c_1, d_1) , we multiply the first equation by d_1d and the second equation by b_1b . Adding the resulting equations, we obtain the following equation in D :

$$a_1bd_1d + c_1db_1b = b_1ad_1d + d_1cb_1b.$$

Using various axioms for an integral domain, we see that

$$(a_1d_1 + b_1c_1)bd = b_1d_1(ad + bc),$$

so

$$(a_1d_1 + b_1c_1, b_1d_1) \sim (ad + bc, bd),$$

giving $(a_1d_1 + b_1c_1, b_1d_1) \in [(ad + bc, bd)]$. This takes care of addition in F . For multiplication in F , on multiplying the equations $a_1b = b_1a$ and $c_1d = d_1c$, we obtain

$$a_1bc_1d = b_1ad_1c,$$

so, using axioms of D , we get

$$a_1c_1bd = b_1d_1ac,$$

which implies that

$$(a_1c_1, b_1d_1) \sim (ac, bd).$$

Thus $(a_1c_1, b_1d_1) \in [(ac, bd)]$, which completes the proof. \blacklozenge

It is important to *understand* the meaning of the last lemma and the necessity for proving it. This completes our Step 2.

Step 3 Step 3 is routine, but it is good for us to work through a few of these details. The reason for this is that we cannot work through them unless we *understand* what we have done. Thus working through them will contribute to our understanding of this construction. We list the things that must be proved and prove a couple of them. The rest are left to the exercises.

1. Addition in F is commutative.

Proof Now $[(a, b)] + [(c, d)]$ is by definition $[(ad + bc, bd)]$. Also $[(c, d)] + [(a, b)]$ is by definition $[(cb + da, db)]$. We need to show that $(ad + bc, bd) \sim (cb + da, db)$. This is true, since $ad + bc = cb + da$ and $bd = db$, by the axioms of D . \blacklozenge

2. Addition is associative.
3. $[(0, 1)]$ is an identity element for addition in F .
4. $[(−a, b)]$ is an additive inverse for $[(a, b)]$ in F .
5. Multiplication in F is associative.
6. Multiplication in F is commutative.
7. The distributive laws hold in F .
8. $[(1, 1)]$ is a multiplicative identity element in F .
9. If $[(a, b)] \in F$ is not the additive identity element, then $a \neq 0$ in D and $[(b, a)]$ is a multiplicative inverse for $[(a, b)]$.

Proof Let $[(a, b)] \in F$. If $a = 0$, then

$$a1 = b0 = 0,$$

so

$$(a, b) \sim (0, 1),$$

that is, $[(a, b)] = [(0, 1)]$. But $[(0, 1)]$ is the additive identity by Part 3. Thus if $[(a, b)]$ is not the additive identity in F , we have $a \neq 0$, so it makes sense to talk about $[(b, a)]$ in F . Now $[(a, b)][(b, a)] = [(ab, ba)]$. But in D we have $ab = ba$, so $(ab)1 = (ba)1$, and

$$(ab, ba) \sim (1, 1).$$

Thus

$$[(a, b)][(b, a)] = [(1, 1)],$$

and $[(1, 1)]$ is the multiplicative identity by Part 8. \blacklozenge

This completes Step 3.

Step 4 It remains for us to show that F can be regarded as containing D . To do this, we show that there is an isomorphism i of D with a subdomain of F . Then if we rename the image of D under i using the names of the elements of D , we will be done. The next lemma gives us this isomorphism. We use the letter i for this isomorphism to suggest *injection*; we will inject D into F .

26.5 Lemma The map $i : D \rightarrow F$ given by $i(a) = [(a, 1)]$ is an isomorphism of D with a subring D' of F .

Proof For a and b in D , we have

$$i(a + b) = [(a + b, 1)].$$

Also,

$$i(a) + i(b) = [(a, 1)] + [(b, 1)] = [(a1 + 1b, 1)] = [(a + b, 1)]$$

so $i(a + b) = i(a) + i(b)$. Furthermore,

$$i(ab) = [(ab, 1)],$$

while

$$i(a)i(b) = [(a, 1)][(b, 1)] = [(ab, 1)],$$

so $i(ab) = i(a)i(b)$.

It remains for us to show only that i is one-to-one. If $i(a) = i(b)$, then

$$[(a, 1)] = [(b, 1)],$$

so $(a, 1) \sim (b, 1)$ giving $a1 = b1$; that is,

$$a = b.$$

Thus i is an isomorphism of D with $i[D] = D'$, and, of course, D' is then a subdomain of F . \blacklozenge

Since $[(a, b)] = [(a, 1)][(1, b)] = [(a, 1)]/[(b, 1)] = i(a)/i(b)$ clearly holds in F , we have now proved the following theorem.

26.6 Theorem Any integral domain D can be enlarged to (or embedded in) a field F such that every element of F can be expressed as a quotient of two elements of D . (Such a field F is a **field of quotients of D** .)

Uniqueness

The field F can be regarded as a minimal field containing D . This is intuitively evident, since every field containing D must contain all elements a/b for every $a, b \in D$ with $b \neq 0$. The next theorem will show that every field containing D contains a subfield that is a field of quotients of D , and that any two fields of quotients of D are isomorphic.

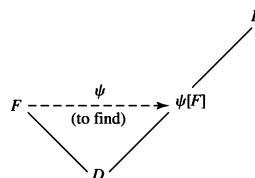
26.7 Theorem Let F be a field of quotients of D and let L be any field containing D . Then there exists a map $\psi : F \rightarrow L$ that gives an isomorphism of F with a subfield of L such that $\psi(a) = a$ for $a \in D$.

Proof The subring and mapping diagram in Fig. 26.8 may help you to visualize the situation for this theorem.

An element of F is of the form $a/f b$ where $/f$ denotes the quotient of $a \in D$ by $b \in D$ regarded as elements of F . We of course want to map $a/f b$ onto $a/L b$ where $/L$ denotes the quotient of elements in L . The main job will be to show that such a map is well defined.

We must define $\psi : F \rightarrow L$, and we start by defining

$$\psi(a) = a \quad \text{for } a \in D.$$



26.8 Figure

Every $x \in F$ is a quotient $a /_F b$ of some two elements a and $b, b \neq 0$, of D . Let us attempt to define ψ by

$$\psi(a /_F b) = \psi(a) /_L \psi(b).$$

We must first show that this map ψ is sensible and well-defined. Since ψ is the identity on D , for $b \neq 0$ we have $\psi(b) \neq 0$, so our definition of $\psi(a /_F b)$ as $\psi(a) /_L \psi(b)$ makes sense. If $a /_F b = c /_F d$ in F , then $ad = bc$ in D , so $\psi(ad) = \psi(bc)$. But since ψ is the identity on D ,

$$\psi(ad) = \psi(a)\psi(d) \quad \text{and} \quad \psi(bc) = \psi(b)\psi(c).$$

Thus

$$\psi(a) /_L \psi(b) = \psi(c) /_L \psi(d)$$

in L , so ψ is well-defined.

The equations

$$\psi(xy) = \psi(x)\psi(y)$$

and

$$\psi(x + y) = \psi(x) + \psi(y)$$

follow easily from the definition of ψ on F and from the fact that ψ is the identity on D .

If $\psi(a /_F b) = \psi(c /_F d)$, we have

$$\psi(a) /_L \psi(b) = \psi(c) /_L \psi(d)$$

so

$$\psi(a)\psi(d) = \psi(b)\psi(c).$$

Since ψ is the identity on D , we then deduce that $ad = bc$, so $a /_F b = c /_F d$. Thus ψ is one-to-one.

By definition, $\psi(a) = a$ for $a \in D$. ◆

26.9 Corollary Every field L containing an integral domain D contains a field of quotients of D .

Proof In the proof of Theorem 26.7 every element of the subfield $\psi[F]$ of L is a quotient in L of elements of D . ◆

26.10 Corollary Any two fields of quotients of an integral domain D are isomorphic.

Proof Suppose in Theorem 26.7 that L is a field of quotients of D , so that every element x of L can be expressed in the form $a /_L b$ for $a, b \in D$. Then L is the field $\psi[F]$ of the proof of Theorem 26.7 and is thus isomorphic to F . ◆

■ EXERCISES 26

Computations

1. Describe the field F of quotients of the integral subdomain

$$D = \{n + mi \mid n, m \in \mathbb{Z}\}$$

of \mathbb{C} . “Describe” means give the elements of \mathbb{C} that make up the field of quotients of D in \mathbb{C} . (The elements of D are the **Gaussian integers**.)

2. Describe (in the sense of Exercise 1) the field F of quotients of the integral subdomain $D = \{n + m\sqrt{3} \mid n, m \in \mathbb{Z}\}$ of \mathbb{R} .

Concepts

3. Correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.
- A *field of quotients* of an integral domain D is a field F in which D can be embedded so that every nonzero element of D is a unit in F .
4. Determine whether each of the following is true or false.
- \mathbb{Q} is a field of quotients of \mathbb{Z} .
 - \mathbb{R} is a field of quotients of \mathbb{Z} .
 - \mathbb{R} is a field of quotients of \mathbb{R} .
 - \mathbb{C} is a field of quotients of \mathbb{R} .
 - If D is a field, then any field of quotients of D is isomorphic to D .
 - The fact that D has no divisors of 0 was used strongly several times in the construction of a field F of quotients of the integral domain D .
 - Every element of an integral domain D is a unit in a field F of quotients of D .
 - Every nonzero element of an integral domain D is a unit in a field F of quotients of D .
 - A field of quotients F' of a subdomain D' of an integral domain D can be regarded as a subfield of some field of quotients of D .
 - Every field of quotients of \mathbb{Z} is isomorphic to \mathbb{Q} .
5. Show by an example that a field F' of quotients of a proper subdomain D' of an integral domain D may also be a field of quotients for D .

Theory

- Prove Part 2 of Step 3. You may assume any preceding part of Step 3.
- Prove Part 3 of Step 3. You may assume any preceding part of Step 3.
- Prove Part 4 of Step 3. You may assume any preceding part of Step 3.
- Prove Part 5 of Step 3. You may assume any preceding part of Step 3.
- Prove Part 6 of Step 3. You may assume any preceding part of Step 3.
- Prove Part 7 of Step 3. You may assume any preceding part of Step 3.
- Let R be a nonzero commutative ring, and let T be a nonempty subset of R closed under multiplication and containing neither 0 nor divisors of 0. Starting with $R \times T$ and otherwise exactly following the construction in this section, we can show that the ring R can be enlarged to a *partial ring of quotients* $Q(R, T)$. Think about this for 15 minutes or so; look back over the construction and see why things still work. In particular, show the following:
 - $Q(R, T)$ has unity even if R does not.
 - In $Q(R, T)$, every nonzero element of T is a unit.
- Prove from Exercise 12 that every nonzero commutative ring containing an element a that is not a divisor of 0 can be enlarged to a commutative ring with unity. Compare with Exercise 36 of Section 23.
- With reference to Exercise 12, how many elements are there in the ring $Q(\mathbb{Z}_4, \{1, 3\})$?
- With reference to Exercise 12, describe the ring $Q(\mathbb{Z}, \{2^n \mid n \in \mathbb{Z}^+\})$, by describing a subring of \mathbb{R} to which it is isomorphic.
- With reference to Exercise 12, describe the ring $Q(3\mathbb{Z}, \{6^n \mid n \in \mathbb{Z}^+\})$ by describing a subring of \mathbb{R} to which it is isomorphic.
- With reference to Exercise 12, suppose we drop the condition that T have no divisors of zero and just require that nonempty T not containing 0 be closed under multiplication. The attempt to enlarge R to a commutative ring with unity in which every nonzero element of T is a unit must fail if T contains an element a that is a divisor of 0, for a divisor of 0 cannot also be a unit. Try to discover where a construction parallel to that in the text but starting with $R \times T$ first runs into trouble. In particular, for $R = \mathbb{Z}_6$ and $T = \{1, 2, 4\}$, illustrate the first difficulty encountered. [Hint: It is in Step 1.]

SECTION 27

RINGS OF POLYNOMIALS

Polynomials in an Indeterminate

We all have a pretty workable idea of what constitutes a *polynomial in x with coefficients in a ring R* . We can guess how to add and multiply such polynomials and know what is meant by the *degree* of a polynomial. We expect that the set $R[x]$ of all polynomials with coefficients in the ring R is itself a ring with the usual operations of polynomial addition and multiplication, and that R is a subring of $R[x]$. However, we will be working with polynomials from a slightly different viewpoint than the approach in high school algebra or calculus, and there are a few things that we want to say.

In the first place, we will call x an **indeterminate** rather than a variable. Suppose, for example that our ring of coefficients is \mathbb{Z} . One of the polynomials in the ring $\mathbb{Z}[x]$ is $1x$, which we shall write simply as x . Now x is not 1 or 2 or any of the other elements of $\mathbb{Z}[x]$. Thus from now on we will never write such things as " $x = 1$ " or " $x = 2$," as we have done in other courses. We call x an indeterminate rather than a variable to emphasize this change. Also, we will never write an expression such as " $x^2 - 4 = 0$," simply because $x^2 - 4$ is not the zero polynomial in our ring $\mathbb{Z}[x]$. We are accustomed to speaking of "solving a polynomial equation," and will be spending a lot of time in the remainder of our text discussing this, but we will always refer to it as "finding a zero of a polynomial." In summary, we try to be careful in our discussion of algebraic structures not to say in one context that things are equal and in another context that they are not equal.

If a person knows nothing about polynomials, it is not an easy task to describe precisely the nature of a polynomial in x with coefficients in a ring R . If we just define such a polynomial to be a *finite formal sum*

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

■ HISTORICAL NOTE

The use of x and other letters near the end of the alphabet to represent an "indeterminate" is due to René Descartes (1596–1650). Earlier, François Viète (1540–1603) had used vowels for indeterminates and consonants for known quantities. Descartes is also responsible for the first publication of the factor theorem (Corollary 28.4) in his work *The Geometry*, which appeared as an appendix to his *Discourse on Method* (1637). This work also contained the first publication of the basic concepts of analytic geometry; Descartes showed how geometric curves can be described algebraically.

Descartes was born to a wealthy family in La Haye, France; since he was always of delicate health, he formed the habit of spending his mornings in bed. It was at these times that he accomplished his most productive work. The *Discourse on Method* was Descartes' attempt to show the proper procedures for "searching for truth in the sciences." The first step in this process was

to reject as absolutely false everything of which he had the least doubt; but, since it was necessary that he who was thinking was "something," he conceived his first principle of philosophy: "I think, therefore I am." The most enlightening parts of the *Discourse on Method*, however, are the three appendices: *The Optics*, *The Geometry*, and *The Meteorology*. It was here that Descartes provided examples of how he actually applied his method. Among the important ideas Descartes discovered and published in these works were the sine law of refraction of light, the basics of the theory of equations, and a geometric explanation of the rainbow.

In 1649, Descartes was invited by Queen Christina of Sweden to come to Stockholm to tutor her. Unfortunately, the Queen required him, contrary to his long-established habits, to rise at an early hour. He soon contracted a lung disease and died in 1650.

where $a_i \in R$, we get ourselves into a bit of trouble. For surely $0 + a_1x$ and $0 + a_1x + 0x^2$ are different as formal sums, but we want to regard them as the same polynomial. A practical solution to this problem is to define a polynomial as an *infinite formal sum*

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

where $a_i = 0$ for all but a finite number of values of i . Now there is no problem of having more than one finite formal sum represent what we wish to consider a single polynomial.

27.1 Definition Let R be a ring. A **polynomial $f(x)$ with coefficients in R** is an infinite formal sum

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

where $a_i \in R$ and for all but a finite number of values of i , $a_i = 0$. The a_i are **coefficients** of $f(x)$. If for some $i \geq 0$ it is true that $a_i \neq 0$, the largest such value of i is the **degree** of $f(x)$. If all $a_i = 0$, then the degree of $f(x)$ is undefined.[†] ■

To simplify working with polynomials, let us agree that if $f(x) = a_0 + a_1 x + \cdots + a_n x^n + \cdots$ has $a_i = 0$ for $i > n$, then we may denote $f(x)$ by $a_0 + a_1 x + \cdots + a_n x^n$. Also, if R has unity $1 \neq 0$, we will write a term $1x^k$ in such a sum as x^k . For example, in $\mathbb{Z}[x]$, we will write the polynomial $2 + 1x$ as $2 + x$. Finally, we shall agree that we may omit altogether from the formal sum any term $0x^i$, or a_0 if $a_0 = 0$ but not all $a_i = 0$. Thus 0 , 2 , x , and $2 + x^2$ are polynomials with coefficients in \mathbb{Z} . An element of R is a **constant polynomial**.

Addition and multiplication of polynomials with coefficients in a ring R are defined in a way familiar to us. If

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n + \cdots$$

and

$$g(x) = b_0 + b_1 x + \cdots + b_n x^n + \cdots,$$

then for polynomial addition, we have

$$f(x) + g(x) = c_0 + c_1 x + \cdots + c_n x^n + \cdots \text{ where } c_n = a_n + b_n,$$

and for polynomial multiplication, we have

$$f(x)g(x) = d_0 + d_1 x + \cdots + d_n x^n + \cdots \text{ where } d_n = \sum_{i=0}^n a_i b_{n-i}$$

Observe that both c_i and d_i are 0 for all but a finite number of values of i , so these definitions make sense. Note that $\sum_{i=0}^n a_i b_{n-i}$ need not equal $\sum_{i=0}^n b_i a_{n-i}$ if R is not commutative. With these definitions of addition and multiplication, we have the following theorem.

27.2 Theorem The set $R[x]$ of all polynomials in an indeterminate x with coefficients in a ring R is a ring under polynomial addition and multiplication. If R is commutative, then so is $R[x]$, and if R has unity $1 \neq 0$, then 1 is also unity for $R[x]$.

Proof That $\langle R[x], + \rangle$ is an abelian group is apparent. The associative law for multiplication and the distributive laws are straightforward, but slightly cumbersome, computations. We illustrate by proving the associative law.

[†] The degree of the zero polynomial is sometimes defined to be -1 , which is the first integer less than 0, or defined to be $-\infty$ so that the degree of $f(x)g(x)$ will be the sum of the degrees of $f(x)$ and $g(x)$ if one of them is zero.

Applying ring axioms to $a_i, b_j, c_k \in R$, we obtain

$$\begin{aligned}
& \left[\left(\sum_{i=0}^{\infty} a_i x^i \right) \left(\sum_{j=0}^{\infty} b_j x^j \right) \right] \left(\sum_{k=0}^{\infty} c_k x^k \right) = \left[\sum_{n=0}^{\infty} \left(\sum_{i=0}^n a_i b_{n-i} \right) x^n \right] \left(\sum_{k=0}^{\infty} c_k x^k \right) \\
& = \sum_{s=0}^{\infty} \left[\sum_{n=0}^s \left(\sum_{i=0}^n a_i b_{n-i} \right) c_{s-n} \right] x^s \\
& = \sum_{s=0}^{\infty} \left(\sum_{i+j+k=s} a_i b_j c_k \right) x^s \\
& = \sum_{s=0}^{\infty} \left[\sum_{m=0}^s a_{s-m} \left(\sum_{j=0}^m b_j c_{m-j} \right) \right] x^s \\
& = \left(\sum_{i=0}^{\infty} a_i x^i \right) \left[\sum_{m=0}^{\infty} \left(\sum_{j=0}^m b_j c_{m-j} \right) x^m \right] \\
& = \left(\sum_{i=0}^{\infty} a_i x^i \right) \left[\left(\sum_{j=0}^{\infty} b_j x^j \right) \left(\sum_{k=0}^{\infty} c_k x^k \right) \right].
\end{aligned}$$

Whew! In this computation, the fourth expression, having just two summation signs, should be viewed as the value of the triple product $f(x)g(x)h(x)$ of these polynomials under this associative multiplication. (In a similar fashion, we view $f(g(h(x)))$ as the value of the associative composition $(f \circ g \circ h)(x)$ of three functions f , g , and h .)

The distributive laws are similarly proved. (See Exercise 26.)

The comments prior to the statement of the theorem show that $R[x]$ is a commutative ring if R is commutative, and a unity $1 \neq 0$ in R is also unity for $R[x]$, in view of the definition of multiplication in $R[x]$. ◆

Thus $\mathbb{Z}[x]$ is the ring of polynomials in the indeterminate x with integral coefficients, $\mathbb{Q}[x]$ the ring of polynomials in x with rational coefficients, and so on.

27.3 Example In $\mathbb{Z}_2[x]$, we have

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

Still working in $\mathbb{Z}_2[x]$, we obtain

$$(x+1) + (x+1) = (1+1)x + (1+1) = 0x + 0 = 0. \quad \blacktriangle$$

If R is a ring and x and y are two indeterminates, then we can form the ring $(R[x])[y]$, that is, the ring of polynomials in y with coefficients that are polynomials in x . Every polynomial in y with coefficients that are polynomials in x can be rewritten in a natural way as a polynomial in x with coefficients that are polynomials in y as illustrated by Exercise 20. This indicates that $(R[x])[y]$ is naturally isomorphic to $(R[y])[x]$, although a careful proof is tedious. We shall identify these rings by means of this natural isomorphism, and shall consider this ring $R[x, y]$ the **ring of polynomials in two indeterminates x and y with coefficients in R** . The **ring $R[x_1, \dots, x_n]$ of polynomials in the n indeterminates x_i with coefficients in R** is similarly defined.

We leave as Exercise 24 the proof that if D is an integral domain, then so is $D[x]$. In particular, if F is a field, then $F[x]$ is an integral domain. Note that $F[x]$ is not a field, for x is not a unit in $F[x]$. That is, there is no polynomial $f(x) \in F[x]$ such that $xf(x) = 1$. By Theorem 26.6, one can construct the field of quotients $F(x)$ of $F[x]$. Any element in $F(x)$ can be represented as a quotient $f(x)/g(x)$ of two polynomials in $F[x]$ with

$g(x) \neq 0$. We similarly define $F(x_1, \dots, x_n)$ to be the field of quotients of $F[x_1, \dots, x_n]$. This field $F(x_1, \dots, x_n)$ is the **field of rational functions in n indeterminates over F** . These fields play a very important role in algebraic geometry.

The Evaluation Homomorphisms

We are now ready to proceed to show how homomorphisms can be used to study what we have always referred to as “solving a polynomial equation.” Let E and F be fields, with F a subfield of E , that is, $F \leq E$. The next theorem asserts the existence of very important homomorphisms of $F[x]$ into E . *These homomorphisms will be the fundamental tools for much of the rest of our work.*

27.4 Theorem (The Evaluation Homomorphisms for Field Theory) Let F be a subfield of a field E , let α be any element of E , and let x be an indeterminate. The map $\phi_\alpha : F[x] \rightarrow E$ defined by

$$\phi_\alpha(a_0 + a_1x + \dots + a_nx^n) = a_0 + a_1\alpha + \dots + a_n\alpha^n$$

for $(a_0 + a_1x + \dots + a_nx^n) \in F[x]$ is a homomorphism of $F[x]$ into E . Also, $\phi_\alpha(x) = \alpha$, and ϕ_α maps F isomorphically by the identity map; that is, $\phi_\alpha(a) = a$ for $a \in F$. The homomorphism ϕ_α is **evaluation at α** .

Proof The subfield and mapping diagram in Fig. 27.5 may help us to visualize this situation. The dashed lines indicate an element of the set. The theorem is really an immediate consequence of our definitions of addition and multiplication in $F[x]$. The map ϕ_α is well defined, that is, independent of our representation of $f(x) \in F[x]$ as a finite sum

$$a_0 + a_1x + \dots + a_nx^n,$$

since such a finite sum representing $f(x)$ can be changed only by insertion or deletion of terms $0x^i$, which does not affect the value of $\phi_\alpha(f(x))$.

If $f(x) = a_0 + a_1x + \dots + a_nx^n$, $g(x) = b_0 + b_1x + \dots + b_mx^m$, and $h(x) = f(x) + g(x) = c_0 + c_1x + \dots + c_rx^r$, then

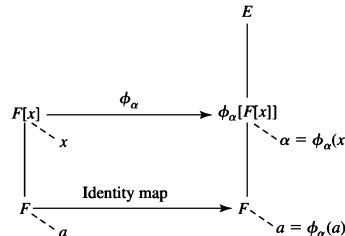
$$\phi_\alpha(f(x) + g(x)) = \phi_\alpha(h(x)) = c_0 + c_1\alpha + \dots + c_r\alpha^r,$$

while

$$\phi_\alpha(f(x)) + \phi_\alpha(g(x)) = (a_0 + a_1\alpha + \dots + a_n\alpha^n) + (b_0 + b_1\alpha + \dots + b_m\alpha^m).$$

Since by definition of polynomial addition we have $c_i = a_i + b_i$, we see that

$$\phi_\alpha(f(x) + g(x)) = \phi_\alpha(f(x)) + \phi_\alpha(g(x)).$$



27.5 Figure

Turning to multiplication, we see that if

$$f(x)g(x) = d_0 + d_1x + \cdots + d_sx^s,$$

then

$$\phi_\alpha(f(x)g(x)) = d_0 + d_1\alpha + \cdots + d_s\alpha^s,$$

while

$$[\phi_\alpha(f(x))][\phi_\alpha(g(x))] = (a_0 + a_1\alpha + \cdots + a_n\alpha^n)(b_0 + b_1\alpha + \cdots + b_m\alpha^m).$$

Since by definition of polynomial multiplication $d_j = \sum_{i=0}^j a_i b_{j-i}$, we see that

$$\phi_\alpha(f(x)g(x)) = [\phi_\alpha(f(x))][\phi_\alpha(g(x))].$$

Thus ϕ_α is a homomorphism.

The very definition of ϕ_α applied to a constant polynomial $a \in F[x]$, where $a \in F$, gives $\phi_\alpha(a) = a$, so ϕ_α maps F isomorphically by the identity map. Again by definition of ϕ_α , we have $\phi_\alpha(x) = \phi_\alpha(1x) = 1\alpha = \alpha$. \blacklozenge

We point out that this theorem is valid with the identical proof if F and E are merely commutative rings with unity rather than fields. However, we shall be interested primarily in the case in which they are fields.

It is hard to overemphasize the importance of this simple theorem for us. It is the very foundation for all of our further work in field theory. It is so simple that it could justifiably be called an *observation* rather than a theorem. It was perhaps a little misleading to write out the proof because the polynomial notation makes it look so complicated that you may be fooled into thinking it is a difficult theorem.

27.6 Example Let F be \mathbb{Q} and E be \mathbb{R} in Theorem 27.4, and consider the evaluation homomorphism $\phi_0 : \mathbb{Q}[x] \rightarrow \mathbb{R}$. Here

$$\phi_0(a_0 + a_1x + \cdots + a_nx^n) = a_0 + a_10 + \cdots + a_n0^n = a_0.$$

Thus every polynomial is mapped onto its constant term. \blacktriangle

27.7 Example Let F be \mathbb{Q} and E be \mathbb{R} in Theorem 27.4 and consider the evaluation homomorphism $\phi_2 : \mathbb{Q}[x] \rightarrow \mathbb{R}$. Here

$$\phi_2(a_0 + a_1x + \cdots + a_nx^n) = a_0 + a_12 + \cdots + a_n2^n.$$

Note that

$$\phi_2(x^2 + x - 6) = 2^2 + 2 - 6 = 0.$$

Thus $x^2 + x - 6$ is in the kernel N of ϕ_2 . Of course,

$$x^2 + x - 6 = (x - 2)(x + 3),$$

and the reason that $\phi_2(x^2 + x - 6) = 0$ is that $\phi_2(x - 2) = 2 - 2 = 0$. \blacktriangle

27.8 Example Let F be \mathbb{Q} and E be \mathbb{C} in Theorem 27.4 and consider the evaluation homomorphism $\phi_i : \mathbb{Q}[x] \rightarrow \mathbb{C}$. Here

$$\phi_i(a_0 + a_1x + \cdots + a_nx^n) = a_0 + a_1i + \cdots + a_ni^n$$

and $\phi_i(x) = i$. Note that

$$\phi_i(x^2 + 1) = i^2 + 1 = 0,$$

so $x^2 + 1$ is in the kernel N of ϕ_i . \blacktriangle

27.9 Example Let F be \mathbb{Q} and let E be \mathbb{R} in Theorem 27.4 and consider the evaluation homomorphism $\phi_\pi : \mathbb{Q}[x] \rightarrow \mathbb{R}$. Here

$$\phi_\pi(a_0 + a_1x + \cdots + a_nx^n) = a_0 + a_1\pi + \cdots + a_n\pi^n.$$

It can be proved that $a_0 + a_1\pi + \cdots + a_n\pi^n = 0$ if and only if $a_i = 0$ for $i = 0, 1, \dots, n$. Thus the kernel of ϕ_π is $\{0\}$, and ϕ_π is a one-to-one map. This shows that all *formal polynomials in π with rational coefficients* form a ring isomorphic to $\mathbb{Q}[x]$ in a natural way with $\phi_\pi(x) = \pi$. \blacktriangle

The New Approach

We now complete the connection between our new ideas and the classical concept of solving a polynomial equation. Rather than speak of *solving a polynomial equation*, we shall refer to *finding a zero of a polynomial*.

27.10 Definition Let F be a subfield of a field E , and let α be an element of E . Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be in $F[x]$, and let $\phi_\alpha : F[x] \rightarrow E$ be the evaluation homomorphism of Theorem 27.4. Let $f(\alpha)$ denote

$$\phi_\alpha(f(x)) = a_0 + a_1\alpha + \cdots + a_n\alpha^n.$$

If $f(\alpha) = 0$, then α is a **zero of $f(x)$** . \blacksquare

In terms of this definition, we can rephrase the classical problem of finding all real numbers r such that $r^2 + r - 6 = 0$ by letting $F = \mathbb{Q}$ and $E = \mathbb{R}$ and *finding all $\alpha \in \mathbb{R}$ such that*

$$\phi_\alpha(x^2 + x - 6) = 0,$$

that is, *finding all zeros of $x^2 + x - 6$ in \mathbb{R}* . Both problems have the same answer, since

$$\{\alpha \in \mathbb{R} \mid \phi_\alpha(x^2 + x - 6) = 0\} = \{r \in \mathbb{R} \mid r^2 + r - 6 = 0\} = \{2, -3\}.$$

It may seem that we have merely succeeded in making a simple problem seem quite complicated. In fact, *what we have done is to phrase the problem in the language of mappings, and we can now use all the mapping machinery that we have developed and will continue to develop for its solution*.

Our Basic Goal

We continue to attempt to put our future work in perspective. Sections 30 and 31 are concerned with topics in ring theory that are analogous to the material on factor groups and homomorphisms for group theory. However, our aim in developing these analogous concepts for rings will be quite different from our aims in group theory. In group theory we used the concepts of factor groups and homomorphisms to study the structure of a given group and to determine the types of group structures of certain orders that could exist. We will be talking about homomorphisms and factor rings in Section 30 with an eye to finding zeros of polynomials, which is one of the oldest and most fundamental problems in algebra. Let us take a moment to talk about this aim in the light of mathematical history, using the language of “solving polynomial equations” to which we are accustomed.

We start with the Pythagorean school of mathematics of about 525 B.C. The Pythagoreans worked with the assumption that all distances are **commensurable**; that is, given distances a and b , there should exist a unit of distance u and integers n and m such that $a = (n)(u)$ and $b = (m)(u)$. In terms of numbers, then, thinking of u as being

one unit of distance, they maintained that all numbers are integers. This idea of commensurability can be rephrased according to our ideas as an assertion that all numbers are rational, for if a and b are rational numbers, then each is an integral multiple of the reciprocal of the least common multiple of their denominators. For example, if $a = \frac{7}{12}$ and $b = \frac{19}{15}$, then $a = (35)(\frac{1}{60})$ and $b = (76)(\frac{1}{60})$.

The Pythagoreans knew, of course, what is now called the *Pythagorean theorem*; that is, for a right triangle with legs of lengths a and b and a hypotenuse of length c ,

$$a^2 + b^2 = c^2.$$

They also had to grant the existence of a hypotenuse of a right triangle having two legs of equal length, say one unit each. The hypotenuse of such a right triangle would, as we know, have to have a length of $\sqrt{2}$. Imagine then their consternation and dismay when one of their society—according to some stories it was Pythagoras himself—came up with the embarrassing fact that is stated in our terminology in the following theorem.

27.11 Theorem The polynomial $x^2 - 2$ has no zeros in the rational numbers. Thus $\sqrt{2}$ is not a rational number.

Proof Suppose that m/n for $m, n \in \mathbb{Z}$ is a rational number such that $(m/n)^2 = 2$. We assume that we have canceled any factors common to m and n , so that the fraction m/n is in lowest terms with $\gcd(m, n) = 1$. Then

$$m^2 = 2n^2,$$

where both m^2 and $2n^2$ are integers. Since m^2 and $2n^2$ are the same integer, and since 2 is a factor of $2n^2$, we see that 2 must be one of the factors of m^2 . But as a square, m^2 has as factors the factors of m repeated twice. Thus m^2 must have two factors 2. Then $2n^2$ must have two factors 2, so n^2 must have 2 as a factor, and consequently n has 2 as a factor. We have deduced from $m^2 = 2n^2$ that both m and n must be divisible by 2, contradicting the fact that the fraction m/n is in lowest terms. Thus we have $2 \neq (m/n)^2$ for any $m, n \in \mathbb{Z}$. ◆

■ HISTORICAL NOTE

The solution of polynomial equations has been a goal of mathematics for nearly 4000 years. The Babylonians developed versions of the quadratic formula to solve quadratic equations. For example, to solve $x^2 - x = 870$, the Babylonian scribe instructed his students to take half of 1 ($\frac{1}{2}$), square it ($\frac{1}{4}$), and add that to 870. The square root of $870\frac{1}{4}$, namely $29\frac{1}{2}$, is then added to $\frac{1}{2}$ to give 30 as the answer. What the scribes did not discuss, however, was what to do if the square root in this process was not a rational number. Chinese mathematicians, however, from about 200 B.C., discovered a method similar to what is now called *Horner's method* to solve quadratic equations numerically; since they used a decimal system, they were able in principle to

carry out the computation to as many places as necessary and could therefore ignore the distinction between rational and irrational solutions. The Chinese, in fact, extended their numerical techniques to polynomial equations of higher degree. In the Arab world, the Persian poet-mathematician Omar Khayyam (1048–1131) developed methods for solving cubic equations geometrically by finding the point(s) of intersection of appropriately chosen conic sections, while Sharaf al-Din al-Tusi (died 1213) used, in effect, techniques of calculus to determine whether or not a cubic equation had a real positive root. It was the Italian Girolamo Cardano (1501–1576) who first published a procedure for solving cubic equations algebraically.

Thus the Pythagoreans ran right into the question of a solution of a polynomial equation, $x^2 - 2 = 0$. We refer the student to Shanks [36, Chapter 3], for a lively and totally delightful account of this Pythagorean dilemma and its significance in mathematics.

In our motivation of the definition of a group, we commented on the necessity of having negative numbers, so that equations such as $x + 2 = 0$ might have solutions. The introduction of negative numbers caused a certain amount of consternation in some philosophical circles. We can visualize 1 apple, 2 apples, and even $\frac{13}{11}$ apples, but how can we point to anything and say that it is -17 apples? Finally, consideration of the equation $x^2 + 1 = 0$ led to the introduction of the number i . The very name of an “imaginary number” given to i shows how this number was regarded. Even today, many students are led by this name to regard i with some degree of suspicion. The negative numbers were introduced to us at such an early stage in our mathematical development that we accepted them without question.

We first met polynomials in high school freshman algebra. The first problem there was to learn how to add, multiply, and factor polynomials. Then, in both freshman algebra and in the second course in algebra in high school, considerable emphasis was placed on solving polynomial equations. These topics are exactly those with which we shall be concerned. The difference is that while in high school, only polynomials with real number coefficients were considered, *we shall be doing our work for polynomials with coefficients from any field.*

Once we have developed the machinery of homomorphisms and factor rings in Section 30, we will proceed with our **basic goal**: to show that given any polynomial of degree ≥ 1 , where the coefficients of the polynomial may be from any field, we can find a zero of this polynomial in some field containing the given field. After the machinery is developed in Sections 30 and 31, the achievement of this goal will be very easy, and is really a very elegant piece of mathematics.

All this fuss may seem ridiculous, but just think back in history. This is the *culmination of more than 2000 years of mathematical endeavor in working with polynomial equations*. After achieving our *basic goal*, we shall spend the rest of our time studying the nature of these solutions of polynomial equations. We need have no fear in approaching this material. *We shall be dealing with familiar topics of high school algebra. This work should seem much more natural than group theory.*

In conclusion, we remark that the machinery of factor rings and ring homomorphisms is not really necessary in order for us to achieve our *basic goal*. For a direct demonstration, see Artin [27, p. 29]. However, factor rings and ring homomorphisms are fundamental ideas that we should grasp, and our *basic goal* will follow very easily once we have mastered them.

■ EXERCISES 27

Computations

In Exercises 1 through 4, find the sum and the product of the given polynomials in the given polynomial ring.

1. $f(x) = 4x - 5, g(x) = 2x^2 - 4x + 2$ in $\mathbb{Z}_8[x]$.
2. $f(x) = x + 1, g(x) = x + 1$ in $\mathbb{Z}_2[x]$.
3. $f(x) = 2x^2 + 3x + 4, g(x) = 3x^2 + 2x + 3$ in $\mathbb{Z}_6[x]$.
4. $f(x) = 2x^3 + 4x^2 + 3x + 2, g(x) = 3x^4 + 2x + 4$ in $\mathbb{Z}_5[x]$.
5. How many polynomials are there of degree ≤ 3 in $\mathbb{Z}_2[x]$? (Include 0.)
6. How many polynomials are there of degree ≤ 2 in $\mathbb{Z}_5[x]$? (Include 0.)

In Exercises 7 and 8, $F = E = \mathbb{C}$ in Theorem 27.4. Compute for the indicated evaluation homomorphism.

7. $\phi_2(x^2 + 3)$

8. $\phi_1(2x^3 - x^2 + 3x + 2)$

In Exercises 9 through 11, $F = E = \mathbb{Z}_7$ in Theorem 27.4. Compute for the indicated evaluation homomorphism.

9. $\phi_3[(x^4 + 2x)(x^3 - 3x^2 + 3)]$

10. $\phi_5[(x^3 + 2)(4x^2 + 3)(x^7 + 3x^2 + 1)]$

11. $\phi_4(3x^{106} + 5x^{99} + 2x^{53})$ [Hint: Use Fermat's theorem.]

In Exercises 12 through 15, find all zeros in the indicated finite field of the given polynomial with coefficients in that field. [Hint: One way is simply to try all candidates!]

12. $x^2 + 1$ in \mathbb{Z}_2

13. $x^3 + 2x + 2$ in \mathbb{Z}_7

14. $x^5 + 3x^3 + x^2 + 2x$ in \mathbb{Z}_5

15. $f(x)g(x)$ where $f(x) = x^3 + 2x^2 + 5$ and $g(x) = 3x^2 + 2x$ in \mathbb{Z}_7

16. Let $\phi_a : \mathbb{Z}_5[x] \rightarrow \mathbb{Z}_5$ be an evaluation homomorphism as in Theorem 27.4. Use Fermat's theorem to evaluate $\phi_3(x^{231} + 3x^{117} - 2x^{53} + 1)$.

17. Use Fermat's theorem to find all zeros in \mathbb{Z}_5 of $2x^{219} + 3x^{74} + 2x^{57} + 3x^{44}$.

Concepts

In Exercises 18 and 19, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

18. A polynomial with coefficients in a ring R is an infinite formal sum

$$\sum_{i=0}^{\infty} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots$$

where $a_i \in R$ for $i = 0, 1, 2, \dots$.

19. Let F be a field and let $f(x) \in F[x]$. A zero of $f(x)$ is an $\alpha \in F$ such that $\phi_\alpha(f(x)) = 0$, where $\phi_\alpha : F(x) \rightarrow F$ is the evaluation homomorphism mapping x into α .

20. Consider the element

$$f(x, y) = (3x^3 + 2x)y^3 + (x^2 - 6x + 1)y^2 + (x^4 - 2x)y + (x^4 - 3x^2 + 2)$$

of $(\mathbb{Q}[x])[y]$. Write $f(x, y)$ as it would appear if viewed as an element of $(\mathbb{Q}[y])[x]$.

21. Consider the evaluation homomorphism $\phi_5 : \mathbb{Q}[x] \rightarrow \mathbb{R}$. Find six elements in the kernel of the homomorphism ϕ_5 .

22. Find a polynomial of degree > 0 in $\mathbb{Z}_4[x]$ that is a unit.

23. Determine whether each of the following is true or false.

- a. The polynomial $(a_n x^n + \cdots + a_1 x + a_0) \in R[x]$ is 0 if and only if $a_i = 0$, for $i = 0, 1, \dots, n$.
- b. If R is a commutative ring, then $R[x]$ is commutative.
- c. If D is an integral domain, then $D[x]$ is an integral domain.
- d. If R is a ring containing divisors of 0, then $R[x]$ has divisors of 0.
- e. If R is a ring and $f(x)$ and $g(x)$ in $R[x]$ are of degrees 3 and 4, respectively, then $f(x)g(x)$ may be of degree 8 in $R[x]$.
- f. If R is any ring and $f(x)$ and $g(x)$ in $R[x]$ are of degrees 3 and 4, respectively, then $f(x)g(x)$ is always of degree 7.
- g. If F is a subfield of E and $\alpha \in E$ is a zero of $f(x) \in F[x]$, then α is a zero of $h(x) = f(x)g(x)$ for all $g(x) \in F[x]$.
- h. If F is a field, then the units in $F[x]$ are precisely the units in F .
- i. If R is a ring with unity, then x is never a divisor of 0 in $R[x]$.
- j. If R is a ring, then the zero divisors in $R[x]$ are precisely the zero divisors in R .

Theory

24. Prove that if D is an integral domain, then $D[x]$ is an integral domain.
25. Let D be an integral domain and x an indeterminate.
- Describe the units in $D[x]$.
 - Find the units in $\mathbb{Z}[x]$.
 - Find the units in $\mathbb{Z}_7[x]$.
26. Prove the left distributive law for $R[x]$, where R is a ring and x is an indeterminate.
27. Let F be a field of characteristic zero and let D be the formal polynomial differentiation map, so that
- $$D(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + 2 \cdot a_2x + \cdots + n \cdot a_nx^{n-1}.$$
- Show that $D : F[x] \rightarrow F[x]$ is a group homomorphism of $\langle F[x], + \rangle$ into itself. Is D a ring homomorphism?
 - Find the kernel of D .
 - Find the image of $F[x]$ under D .
28. Let F be a subfield of a field E .
- Define an *evaluation homomorphism*
- $$\phi_{\alpha_1, \dots, \alpha_n} : F[x_1, \dots, x_n] \rightarrow E \quad \text{for } \alpha_i \in E,$$
- stating the analog of Theorem 27.4.
- With $E = F = \mathbb{Q}$, compute $\phi_{-3,2}(x_1^2x_2^3 + 3x_1^4x_2)$.
 - Define the concept of a *zero of a polynomial* $f(x_1, \dots, x_n) \in F[x_1, \dots, x_n]$ in a way analogous to the definition in the text of a zero of $f(x)$.
29. Let R be a ring, and let R^R be the set of all functions mapping R into R . For $\phi, \psi \in R^R$, define the sum $\phi + \psi$ by
- $$(\phi + \psi)(r) = \phi(r) + \psi(r)$$
- and the product $\phi \cdot \psi$ by
- $$(\phi \cdot \psi)(r) = \phi(r)\psi(r)$$
- for $r \in R$. Note that \cdot is *not* function composition. Show that $\langle R^R, +, \cdot \rangle$ is a ring.
30. Referring to Exercise 29, let F be a field. An element ϕ of F^F is a **polynomial function** on F , if there exists $f(x) \in F[x]$ such that $\phi(a) = f(a)$ for all $a \in F$.
- Show that the set P_F of all polynomial functions on F forms a subring of F^F .
 - Show that the ring P_F is not necessarily isomorphic to $F[x]$. [Hint: Show that if F is a finite field, P_F and $F[x]$ don't even have the same number of elements.]
31. Refer to Exercises 29 and 30 for the following questions.
- How many elements are there in $\mathbb{Z}_2^{\mathbb{Z}_2}$? in $\mathbb{Z}_3^{\mathbb{Z}_3}$?
 - Classify $(\mathbb{Z}_2^{\mathbb{Z}_2}, +)$ and $(\mathbb{Z}_3^{\mathbb{Z}_3}, +)$ by Theorem 9.12, the Fundamental Theorem of finitely generated abelian groups.
 - Show that if F is a finite field, then $F^F = P_F$. [Hint: Of course, $P_F \subseteq F^F$. Let F have as elements a_1, \dots, a_n . Note that if
- $$f_i(x) = c(x - a_1) \cdots (x - a_{i-1})(x - a_{i+1}) \cdots (x - a_n),$$
- then $f_i(a_j) = 0$ for $j \neq i$, and the value $f_i(a_i)$ can be controlled by the choice of $c \in F$. Use this to show that every function on F is a polynomial function.]
32. Let $\phi : R_1 \rightarrow R_2$ be a ring homomorphism. Show that there is a unique ring homomorphism $\psi : R_1[x] \rightarrow R_2[x]$ such that $\psi(a) = \phi(a)$ for any $a \in R_1$ and $\psi(x) = x$.

SECTION 28**FACTORIZATION OF POLYNOMIALS OVER A FIELD**

Recall that we are concerned with finding zeros of polynomials. Let E and F be fields, with $F \leq E$. Suppose that $f(x) \in F[x]$ factors in $F[x]$, so that $f(x) = g(x)h(x)$ for $g(x), h(x) \in F[x]$ and let $\alpha \in E$. Now for the evaluation homomorphism ϕ_α , we have

$$f(\alpha) = \phi_\alpha(f(x)) = \phi_\alpha(g(x)h(x)) = \phi_\alpha(g(x))\phi_\alpha(h(x)) = g(\alpha)h(\alpha).$$

Thus if $\alpha \in E$, then $f(\alpha) = 0$ if and only if either $g(\alpha) = 0$ or $h(\alpha) = 0$. The attempt to find a zero of $f(x)$ is reduced to the problem of finding a zero of a factor of $f(x)$. This is one reason why it is useful to study factorization of polynomials.

The Division Algorithm in $F[x]$

The following theorem is the basic tool for our work in this section. Note the similarity with the division algorithm for \mathbb{Z} given in Theorem 6.2, the importance of which has been amply demonstrated.

We prove the following lemma, which is used in our proof of the division algorithm.

28.1 Lemma Let F be a field and $f(x), g(x), s(x) \in F[x]$ with $g(x) \neq 0$. If

$$\deg(f(x) - g(x)s(x)) \geq \deg(g(x)),$$

then there is a polynomial $s_1(x) \in F[x]$ such that either

$$\deg(f(x) - g(x)s_1(x)) < \deg(f(x) - g(x)s(x))$$

or

$$f(x) - g(x)s_1(x) = 0.$$

Proof Let $n = \deg(f(x) - g(x)s(x))$. We can write $(f(x) - g(x)s(x)) = a_n x^n + r(x)$ where $a_n \neq 0$ and either $r(x) = 0$ or $\deg(r(x)) < n$. Similarly, since $g(x) \neq 0$, we can write $g(x) = b_k x^k + g_1(x)$ where $b_k \neq 0$ and either $g_1(x) = 0$ or $\deg(g_1(x)) < k$.

We let $s_1(x) = s(x) + \frac{a_n}{b_k} x^{n-k}$. Then

$$\begin{aligned} f(x) - g(x)s_1(x) &= f(x) - g(x)s(x) - g(x)\frac{a_n}{b_k}x^{n-k} \\ &= a_n x^n + r(x) - b_k x^k \frac{a_n}{b_k} x^{n-k} - g_1(x) \frac{a_n}{b_k} x^{n-k} \\ &= r(x) - g_1(x) \frac{a_n}{b_k} x^{n-k}. \end{aligned}$$

Each polynomial $r(x)$ and $g_1(x) \frac{a_n}{b_k} x^{n-k}$ is either 0 or has degree less than n . Thus $r(x) - g_1(x) \frac{a_n}{b_k} x^{n-k} = 0$ or $\deg(r(x) - g_1(x) \frac{a_n}{b_k} x^{n-k}) < n = \deg(f(x) - g(x)s(x))$, which completes the proof. \blacklozenge

28.2 Theorem (Division Algorithm for $F[x]$) Let

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$$

and

$$g(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_0$$

be two elements of $F[x]$, with a_n and b_m both nonzero elements of F and $m > 0$. Then there are unique polynomials $q(x)$ and $r(x)$ in $F[x]$ such that $f(x) = g(x)q(x) + r(x)$, where either $r(x) = 0$ or the degree of $r(x)$ is less than the degree m of $g(x)$.

Proof Consider the set $S = \{f(x) - g(x)s(x) \mid s(x) \in F[x]\}$. If $0 \in S$ then there exists an $s(x)$ such that $f(x) - g(x)s(x) = 0$, so $f(x) = g(x)s(x)$. Taking $q(x) = s(x)$ and $r(x) = 0$, we are done. Otherwise, let $r(x)$ be an element of minimal degree in S . Then

$$f(x) = g(x)q(x) + r(x)$$

for some $q(x) \in F[x]$. By Lemma 28.1, the degree of $r(x)$ is less than the degree of $g(x)$ since if the degree of $r(x)$ were at least as large as the degree of $g(x)$, then $r(x)$ would not have minimal degree in S .

For uniqueness, if

$$f(x) = g(x)q_1(x) + r_1(x)$$

and

$$f(x) = g(x)q_2(x) + r_2(x),$$

then subtracting we have

$$g(x)[q_1(x) - q_2(x)] = r_2(x) - r_1(x).$$

Because either $r_2(x) - r_1(x) = 0$ or the degree of $r_2(x) - r_1(x)$ is less than the degree of $g(x)$, this can hold only if $q_1(x) - q_2(x) = 0$ so $q_1(x) = q_2(x)$. Then we must also have $r_2(x) - r_1(x) = 0$ so $r_1(x) = r_2(x)$. \blacklozenge

We can compute the polynomials $q(x)$ and $r(x)$ of Theorem 28.2 by long division just as we divided polynomials in $\mathbb{R}[x]$ in high school.

28.3 Example Let us work with polynomials in $\mathbb{Z}_5[x]$ and divide

$$f(x) = x^4 - 3x^3 + 2x^2 + 4x - 1$$

by $g(x) = x^2 - 2x + 3$ to find $q(x)$ and $r(x)$ of Theorem 28.2. The long division should be easy to follow, but remember that we are in $\mathbb{Z}_5[x]$, so, for example, $4x - (-3x) = 2x$.

$$\begin{array}{r} x^2 - x - 3 \\ \hline x^2 - 2x + 3 \end{array} \overline{\left| \begin{array}{r} x^4 - 3x^3 + 2x^2 + 4x - 1 \\ x^4 - 2x^3 + 3x^2 \\ \hline -x^3 - x^2 + 4x \\ -x^3 + 2x^2 - 3x \\ \hline -3x^2 + 2x - 1 \\ -3x^2 + x - 4 \\ \hline x + 3 \end{array} \right.}$$

Thus

$$q(x) = x^2 - x - 3, \quad \text{and} \quad r(x) = x + 3. \quad \blacktriangle$$

We give three important corollaries of Theorem 28.2. The first one appears in high school algebra for the special case $F[x] = \mathbb{R}[x]$. We phrase our proof in terms of the mapping (homomorphism) approach described in Section 27.

28.4 Corollary (Factor Theorem) An element $a \in F$ is a zero of $f(x) \in F[x]$ if and only if $x - a$ is a factor of $f(x)$ in $F[x]$.

Proof Suppose that for $a \in F$ we have $f(a) = 0$. By Theorem 28.2, there exist $q(x)$, $r(x) \in F[x]$ such that

$$f(x) = (x - a)q(x) + r(x),$$

where either $r(x) = 0$ or the degree of $r(x)$ is less than 1. Thus we must have $r(x) = c$ for $c \in F$, so

$$f(x) = (x - a)q(x) + c.$$

Applying our evaluation homomorphism, $\phi_a : F[x] \rightarrow F$ of Theorem 27.4, we find

$$0 = f(a) = 0q(a) + c,$$

so it must be that $c = 0$. Then $f(x) = (x - a)q(x)$, so $x - a$ is a factor of $f(x)$.

Conversely, if $x - a$ is a factor of $f(x)$ in $F[x]$, where $a \in F$, then applying our evaluation homomorphism ϕ_a to $f(x) = (x - a)q(x)$, we have $f(a) = 0q(a) = 0$. \blacklozenge

28.5 Example

Working again in $\mathbb{Z}_5[x]$, note that 1 is a zero of

$$(x^4 + 3x^3 + 2x + 4) \in \mathbb{Z}_5[x].$$

Thus by Corollary 28.4, we should be able to factor $x^4 + 3x^3 + 2x + 4$ into $(x - 1)q(x)$ in $\mathbb{Z}_5[x]$. Let us find the factorization by long division.

$$\begin{array}{r} x^3 + 4x^2 + 4x + 1 \\ x - 1 \overline{)x^4 + 3x^3 + 2x + 4} \\ x^4 - x^3 \\ \hline 4x^3 \\ 4x^3 - 4x^2 \\ \hline 4x^2 + 2x \\ 4x^2 - 4x \\ \hline x + 4 \\ x - 1 \\ \hline 0 \end{array}$$

Thus $x^4 + 3x^3 + 2x + 4 = (x - 1)(x^3 + 4x^2 + 4x + 1)$ in $\mathbb{Z}_5[x]$. Since 1 is seen to be a zero of $x^3 + 4x^2 + 4x + 1$ also, we can divide this polynomial by $x - 1$ and get

$$\begin{array}{r} x^2 + 4 \\ x - 1 \overline{x^3 + 4x^2 + 4x + 1} \\ x^3 - x^2 \\ \hline 0 + 4x + 1 \\ 4x - 4 \\ \hline 0 \end{array}$$

Since $x^2 + 4$ still has 1 as a zero, we can divide again by $x - 1$ and get

$$\begin{array}{r} x + 1 \\ x - 1 \overline{x^2 + 4} \\ x^2 - x \\ \hline x + 4 \\ x - 1 \\ \hline 0 \end{array}$$

Thus $x^4 + 3x^3 + 2x + 4 = (x - 1)^3(x + 1)$ in $\mathbb{Z}_5[x]$. \blacktriangle

The next corollary should also look familiar.

28.6 Corollary A nonzero polynomial $f(x) \in F[x]$ of degree n can have at most n zeros in a field F .

Proof The preceding corollary shows that if $a_1 \in F$ is a zero of $f(x)$, then

$$f(x) = (x - a_1)q_1(x),$$

where, of course, the degree of $q_1(x)$ is $n - 1$. A zero $a_2 \in F$ of $q_1(x)$ then results in a factorization

$$f(x) = (x - a_1)(x - a_2)q_2(x).$$

Continuing this process, we arrive at

$$f(x) = (x - a_1) \cdots (x - a_r)q_r(x),$$

where $q_r(x)$ has no further zeros in F . Since the degree of $f(x)$ is n , at most n factors $(x - a_i)$ can appear on the right-hand side of the preceding equation, so $r \leq n$. Also, if $b \neq a_i$ for $i = 1, \dots, r$ and $b \in F$, then

$$f(b) = (b - a_1) \cdots (b - a_r)q_r(b) \neq 0,$$

since F has no divisors of 0 and none of $b - a_i$ or $q_r(b)$ are 0 by construction. Hence the a_i for $i = 1, \dots, r \leq n$ are all the zeros in F of $f(x)$. \blacklozenge

Our final corollary is concerned with the structure of the multiplicative group F^* of nonzero elements of a field F , rather than with factorization in $F[x]$. It may at first seem surprising that such a result follows from the division algorithm in $F[x]$, but recall that the result that a subgroup of a cyclic group is cyclic follows from the division algorithm in \mathbb{Z} .

28.7 Corollary If G is a finite subgroup of the multiplicative group $\langle F^*, \cdot \rangle$ of a field F , then G is cyclic. In particular, the multiplicative group of all nonzero elements of a finite field is cyclic.

Proof By Theorem 9.12 as a finite abelian group, G is isomorphic to a direct product $\mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \cdots \times \mathbb{Z}_{d_r}$, where each d_i is a power of a prime. Let us think of each of the \mathbb{Z}_{d_i} as a cyclic group of order d_i in multiplicative notation. Let m be the least common multiple of all the d_i for $i = 1, 2, \dots, r$; note that $m \leq d_1 d_2 \cdots d_r$. If $a_i \in \mathbb{Z}_{d_i}$, then $a_i^{d_i} = 1$, so $a_i^m = 1$ since d_i divides m . Thus for all $\alpha \in G$, we have $\alpha^m = 1$, so every element of G is zero of $x^m - 1$. But G has $d_1 d_2 \cdots d_r$ elements, while $x^m - 1$ can have at most m zeros in the field F by Corollary 28.6, so $m \geq d_1 d_2 \cdots d_r$. Hence $m = d_1 d_2 \cdots d_r$, so the primes involved in the prime powers d_1, d_2, \dots, d_r are distinct, and the group G is isomorphic to the cyclic group \mathbb{Z}_m . \blacklozenge

Exercises 5 through 8 ask us to find all generators of the cyclic groups of units for some finite fields. The fact that the multiplicative group of units of a finite field is cyclic has been applied in algebraic coding and combinatorial designs.

Irreducible Polynomials

Our next definition singles out a type of polynomial in $F[x]$ that will be of utmost importance to us. The concept is probably already familiar. We really are doing high school algebra in a more general setting.

28.8 Definition A nonconstant polynomial $f(x) \in F[x]$ is **irreducible over F** or is an **irreducible polynomial in $F[x]$** if $f(x)$ cannot be expressed as a product $g(x)h(x)$ of two polynomials $g(x)$ and $h(x)$ in $F[x]$ both of lower degree than the degree of $f(x)$. If $f(x) \in F[x]$ is a nonconstant polynomial that is not irreducible over F , then $f(x)$ is **reducible over F** . \blacksquare

Note that the preceding definition concerns the concept *irreducible over F* and not just the concept *irreducible*. A polynomial $f(x)$ may be irreducible over F , but may not be irreducible if viewed over a larger field E containing F . We illustrate this.

- 28.9 Example** Theorem 27.11 shows that $x^2 - 2$ viewed in $\mathbb{Q}[x]$ has no zeros in \mathbb{Q} . This shows that $x^2 - 2$ is irreducible over \mathbb{Q} , for a factorization $x^2 - 2 = (ax + b)(cx + d)$ for $a, b, c, d \in \mathbb{Q}$ would give rise to zeros of $x^2 - 2$ in \mathbb{Q} . However, $x^2 - 2$ viewed in $\mathbb{R}[x]$ is not irreducible over \mathbb{R} , because $x^2 - 2$ factors in $\mathbb{R}[x]$ into $(x - \sqrt{2})(x + \sqrt{2})$. \blacktriangle

It is worthwhile to remember that *the units in $F[x]$ are precisely the nonzero elements of F* . Thus we could have defined an irreducible polynomial $f(x)$ as a nonconstant polynomial such that in any factorization $f(x) = g(x)h(x)$ in $F[x]$, either $g(x)$ or $h(x)$ is a unit.

- 28.10 Example** Let us show that $f(x) = x^3 + 3x + 2$ viewed in $\mathbb{Z}_5[x]$ is irreducible over \mathbb{Z}_5 . If $x^3 + 3x + 2$ factored in $\mathbb{Z}_5[x]$ into polynomials of lower degree then there would exist at least one linear factor of $f(x)$ of the form $x - a$ for some $a \in \mathbb{Z}_5$. But then $f(a)$ would be 0, by Corollary 28.4. However, $f(0) = 2, f(1) = 1, f(-1) = -2, f(2) = 1$, and $f(-2) = -2$, showing that $f(x)$ has no zeros in \mathbb{Z}_5 . Thus $f(x)$ is irreducible over \mathbb{Z}_5 . This test for irreducibility by finding zeros works nicely for quadratic and cubic polynomials over a finite field with a small number of elements. \blacktriangle

Irreducible polynomials will play a very important role in our work from now on. The problem of determining whether a given $f(x) \in F[x]$ is irreducible over F may be difficult. We now give some criteria for irreducibility that are useful in certain cases. One technique for determining irreducibility of quadratic and cubic polynomials was illustrated in Examples 28.9 and 28.10. We formalize it in a theorem.

- 28.11 Theorem** Let $f(x) \in F[x]$, and let $f(x)$ be of degree 2 or 3. Then $f(x)$ is reducible over F if and only if it has a zero in F .

Proof If $f(x)$ is reducible so that $f(x) = g(x)h(x)$, where the degree of $g(x)$ and the degree of $h(x)$ are both less than the degree of $f(x)$, then since $f(x)$ is either quadratic or cubic, either $g(x)$ or $h(x)$ is of degree 1. If, say, $g(x)$ is of degree 1, then except for a possible factor in F , $g(x)$ is of the form $x - a$. Then $g(a) = 0$, which implies that $f(a) = 0$, so $f(x)$ has a zero in F .

Conversely, Corollary 28.4 shows that if $f(a) = 0$ for $a \in F$, then $x - a$ is a factor of $f(x)$, so $f(x)$ is reducible. \blacklozenge

We turn to some conditions for irreducibility over \mathbb{Q} of polynomials in $\mathbb{Q}[x]$. The most important condition that we shall give is contained in the next theorem. The proof is to be worked out in Exercises 38–40.

- 28.12 Theorem** If $f(x) \in \mathbb{Z}[x]$, then $f(x)$ factors into a product of two polynomials of lower degrees r and s in $\mathbb{Q}[x]$ if and only if it has such a factorization with polynomials of the same degrees r and s in $\mathbb{Z}[x]$. \blacklozenge

- 28.13 Corollary** If $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ is in $\mathbb{Z}[x]$ with $a_0 \neq 0$, and if $f(x)$ has a zero in \mathbb{Q} , then it has a zero m in \mathbb{Z} , and m must divide a_0 .

Proof If $f(x)$ has a zero a in \mathbb{Q} , then $f(x)$ has a linear factor $x - a$ in $\mathbb{Q}[x]$ by Corollary 28.4. But then by Theorem 28.12, $f(x)$ has a factorization with a linear factor in $\mathbb{Z}[x]$, so for some $m \in \mathbb{Z}$ we must have

$$f(x) = (x - m)(x^{n-1} + \cdots - a_0/m).$$

Thus a_0/m is in \mathbb{Z} , so m divides a_0 . \blacklozenge

28.14 Example Corollary 28.13 gives us another proof of the irreducibility of $x^2 - 2$ over \mathbb{Q} , for $x^2 - 2$ factors nontrivially in $\mathbb{Q}[x]$ if and only if it has a zero in \mathbb{Q} by Theorem 28.11. By Corollary 28.13, it has a zero in \mathbb{Q} if and only if it has a zero in \mathbb{Z} , and moreover the only possibilities are the divisors ± 1 and ± 2 . A quick check shows that none of these numbers is a zero of $x^2 - 2$. \blacktriangle

28.15 Example Let us use Theorem 28.12 to show that

$$f(x) = x^4 - 2x^2 + 8x + 1$$

viewed in $\mathbb{Q}[x]$ is irreducible over \mathbb{Q} . If $f(x)$ has a linear factor in $\mathbb{Q}[x]$, then it has a zero in \mathbb{Z} , and by Corollary 28.13, this zero would have to be a divisor in \mathbb{Z} of 1, that is, either ± 1 . But $f(1) = 8$, and $f(-1) = -8$, so such a factorization is impossible.

If $f(x)$ factors into two quadratic factors in $\mathbb{Q}[x]$, then by Theorem 28.12, it has a factorization.

$$(x^2 + ax + b)(x^2 + cx + d)$$

in $\mathbb{Z}[x]$. Equating coefficients of powers of x , we find that we must have

$$bd = 1, \quad ad + bc = 8, \quad ac + b + d = -2, \quad \text{and} \quad a + c = 0$$

for integers $a, b, c, d \in \mathbb{Z}$. From $bd = 1$, we see that either $b = d = 1$ or $b = d = -1$. In any case, $b = d$ and from $ad + bc = 8$, we deduce that $d(a + c) = 8$. But this is impossible since $a + c = 0$. Thus a factorization into two quadratic polynomials is also impossible and $f(x)$ is irreducible over \mathbb{Q} . \blacktriangle

We conclude our irreducibility criteria with the famous Eisenstein criterion for irreducibility. An additional very useful criterion is given in Exercise 37.

28.16 Theorem (Eisenstein Criterion) Let $p \in \mathbb{Z}$ be a prime. Suppose that $f(x) = a_nx^n + \cdots + a_0$ is in $\mathbb{Z}[x]$, and $a_n \not\equiv 0 \pmod{p}$, but $a_i \equiv 0 \pmod{p}$ for all $i < n$, with $a_0 \not\equiv 0 \pmod{p^2}$. Then $f(x)$ is irreducible over \mathbb{Q} .

Proof By Theorem 28.12 we need only show that $f(x)$ does not factor into polynomials of lower degree in $\mathbb{Z}[x]$. If

$$f(x) = (b_r x^r + \cdots + b_0)(c_s x^s + \cdots + c_0)$$

is a factorization in $\mathbb{Z}[x]$, with $b_r \neq 0, c_s \neq 0$ and $r, s < n$, then $a_0 \not\equiv 0 \pmod{p^2}$ implies that b_0 and c_0 are not both congruent to 0 modulo p . Suppose that $b_0 \not\equiv 0 \pmod{p}$ and $c_0 \equiv 0 \pmod{p}$. Now $a_n \not\equiv 0 \pmod{p}$ implies that $b_r, c_s \not\equiv 0 \pmod{p}$, since $a_n = b_r c_s$. Let m be the smallest value of k such that $c_k \not\equiv 0 \pmod{p}$. Then

$$a_m = b_0 c_m + b_1 c_{m-1} + \cdots + \begin{cases} b_m c_0 & \text{if } r \geq m, \\ b_r c_{m-r} & \text{if } r < m. \end{cases}$$

The fact that neither b_0 nor c_m are congruent to 0 modulo p while c_{m-1}, \dots, c_0 are all congruent to 0 modulo p implies that $a_m \not\equiv 0 \pmod{p}$, so $m = n$. Consequently, $s = n$, contradicting our assumption that $s < n$; that is, that our factorization was nontrivial. \blacklozenge

Note that if we take $p = 2$, the Eisenstein criterion gives us still another proof of the irreducibility of $x^2 - 2$ over \mathbb{Q} .

28.17 Example Taking $p = 3$, we see by Theorem 28.16 that

$$25x^5 - 9x^4 - 3x^2 - 12$$

is irreducible over \mathbb{Q} . \blacktriangle

28.18 Corollary The polynomial

$$\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1$$

is irreducible over \mathbb{Q} for any prime p .

Proof Again by Theorem 28.12, we need only consider factorizations in $\mathbb{Z}[x]$. We remarked following Theorem 27.4 that its proof actually shows that evaluation homomorphisms can be used for commutative rings. Here we want to use the evaluation homomorphism $\phi_{x+1} : \mathbb{Q}[x] \rightarrow \mathbb{Q}[x]$. It is natural for us to denote $\phi_{x+1}(f(x))$ by $f(x+1)$ for $f(x) \in \mathbb{Q}[x]$. Let

$$g(x) = \Phi_p(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = \frac{x^p + \binom{p}{1}x^{p-1} + \cdots + px}{x}.$$

The coefficient of x^{p-r} for $0 < r < p$ is the binomial coefficient $p!/[r!(p-r)!]$, which is divisible by p because p divides $p!$ but does not divide either $r!$ or $(p-r)!$ when $0 < r < p$. Thus

$$g(x) = x^{p-1} + \binom{p}{1}x^{p-2} + \cdots + p$$

satisfies the Eisenstein criterion for the prime p and is thus irreducible over \mathbb{Q} . But if $\Phi_p(x) = h(x)r(x)$ were a nontrivial factorization of $\Phi_p(x)$ in $\mathbb{Z}[x]$, then

$$\Phi_p(x+1) = g(x) = h(x+1)r(x+1)$$

would give a nontrivial factorization of $g(x)$ in $\mathbb{Z}[x]$. Thus $\Phi_p(x)$ must also be irreducible over \mathbb{Q} . \blacklozenge

The polynomial $\Phi_p(x)$ in Corollary 28.18 is the p^{th} **cyclotomic polynomial**.

Uniqueness of Factorization in $F[x]$

Polynomials in $F[x]$ can be factored into a product of irreducible polynomials in $F[x]$ in an essentially unique way. For $f(x), g(x) \in F[x]$ we say that $g(x)$ **divides** $f(x)$ in $F[x]$ if there exists $q(x) \in F[x]$ such that $f(x) = g(x)q(x)$. Note the similarity of the theorem that follows with Property (1) for \mathbb{Z} following Example 6.9.

28.19 Theorem Let $p(x)$ be an irreducible polynomial in $F[x]$. If $p(x)$ divides $r(x)s(x)$ for $r(x), s(x) \in F[x]$, then either $p(x)$ divides $r(x)$ or $p(x)$ divides $s(x)$.

Proof We delay the proof of this theorem to Section 31. (See Theorem 31.27.) \blacklozenge

28.20 Corollary If $p(x)$ is irreducible in $F[x]$ and $p(x)$ divides the product $r_1(x) \cdots r_n(x)$ for $r_i(x) \in F[x]$, then $p(x)$ divides $r_i(x)$ for at least one i .

Proof Using mathematical induction, we find that this is immediate from Theorem 28.19. \blacklozenge

28.21 Theorem If F is a field, then every nonconstant polynomial $f(x) \in F[x]$ can be factored in $F[x]$ into a product of irreducible polynomials, the irreducible polynomials being unique except for order and for unit (that is, nonzero constant) factors in F .

Proof Let $f(x) \in F[x]$ be a nonconstant polynomial. If $f(x)$ is not irreducible, then $f(x) = g(x)h(x)$, with the degree of $g(x)$ and the degree of $h(x)$ both less than the degree of $f(x)$.

If $g(x)$ and $h(x)$ are both irreducible, we stop here. If not, at least one of them factors into polynomials of lower degree. Continuing this process, we arrive at a factorization

$$f(x) = p_1(x)p_2(x) \cdots p_r(x),$$

where $p_i(x)$ is irreducible for $i = 1, 2, \dots, r$.

It remains for us to show uniqueness. Suppose that

$$f(x) = p_1(x)p_2(x) \cdots p_r(x) = q_1(x)q_2(x) \cdots q_s(x)$$

are two factorizations of $f(x)$ into irreducible polynomials. Then by Corollary 28.20, $p_1(x)$ divides some $q_j(x)$, let us assume $q_1(x)$. Since $q_1(x)$ is irreducible,

$$q_1(x) = u_1 p_1(x),$$

where $u_1 \neq 0$, but u_1 is in F and thus is a unit. Then substituting $u_1 p_1(x)$ for $q_1(x)$ and canceling, we get

$$p_2(x) \cdots p_r(x) = u_1 q_2(x) \cdots q_s(x).$$

By a similar argument, say $q_2(x) = u_2 p_2(x)$, so

$$p_3(x) \cdots p_r(x) = u_1 u_2 q_3(x) \cdots q_s(x).$$

Continuing in this manner, we eventually arrive at

$$1 = u_1 u_2 \cdots u_r q_{r+1}(x) \cdots q_s(x).$$

This is only possible if $s = r$, so that this equation is actually $1 = u_1 u_2 \cdots u_r$. Thus the irreducible factors $p_i(x)$ and $q_j(x)$ were the same except possibly for order and unit factors. \blacklozenge

28.22 Example Example 28.5 shows a factorization of $x^4 + 3x^3 + 2x + 4$ in $\mathbb{Z}_5[x]$ is $(x - 1)^3(x + 1)$. These irreducible factors in $\mathbb{Z}_5[x]$ are only unique up to units in $\mathbb{Z}_5[x]$, that is, nonzero constants in \mathbb{Z}_5 . For example, $(x - 1)^3(x + 1) = (x - 1)^2(2x - 2)(3x + 3)$. \blacktriangle

■ EXERCISES 28

Computations

In Exercises 1 through 4, find $q(x)$ and $r(x)$ as described by the division algorithm so that $f(x) = g(x)q(x) + r(x)$ with $r(x) = 0$ or of degree less than the degree of $g(x)$.

1. $f(x) = x^6 + 3x^5 + 4x^2 - 3x + 2$ and $g(x) = x^2 + 2x - 3$ in $\mathbb{Z}_7[x]$.
2. $f(x) = x^6 + 3x^5 + 4x^2 - 3x + 2$ and $g(x) = 3x^2 + 2x - 3$ in $\mathbb{Z}_7[x]$.
3. $f(x) = x^5 - 2x^4 + 3x - 5$ and $g(x) = 2x + 1$ in $\mathbb{Z}_{11}[x]$.
4. $f(x) = x^4 + 5x^3 - 3x^2$ and $g(x) = 5x^2 - x + 2$ in $\mathbb{Z}_{11}[x]$.

In Exercises 5 through 8, find all generators of the cyclic multiplicative group of units of the given finite field. (Review Corollary 6.17.)

- | | | | |
|--------------------------|--------------------------|-----------------------------|-----------------------------|
| 5. \mathbb{Z}_5 | 6. \mathbb{Z}_7 | 7. \mathbb{Z}_{17} | 8. \mathbb{Z}_{23} |
|--------------------------|--------------------------|-----------------------------|-----------------------------|
9. The polynomial $x^4 + 4$ can be factored into linear factors in $\mathbb{Z}_5[x]$. Find this factorization.
 10. The polynomial $x^3 + 2x^2 + 2x + 1$ can be factored into linear factors in $\mathbb{Z}_7[x]$. Find this factorization.
 11. The polynomial $2x^3 + 3x^2 - 7x - 5$ can be factored into linear factors in $\mathbb{Z}_{11}[x]$. Find this factorization.
 12. Is $x^3 + 2x + 3$ an irreducible polynomial of $\mathbb{Z}_5[x]$? Why? Express it as a product of irreducible polynomials of $\mathbb{Z}_5[x]$.

13. Is $2x^3 + x^2 + 2x + 2$ an irreducible polynomial in $\mathbb{Z}_5[x]$? Why? Express it as a product of irreducible polynomials in $\mathbb{Z}_5[x]$.
14. Show that $f(x) = x^2 + 8x - 2$ is irreducible over \mathbb{Q} . Is $f(x)$ irreducible over \mathbb{R} ? Over \mathbb{C} ?
15. Repeat Exercise 14 with $g(x) = x^2 + 6x + 12$ in place of $f(x)$.
16. Demonstrate that $x^3 + 3x^2 - 8$ is irreducible over \mathbb{Q} .
17. Demonstrate that $x^4 - 22x^2 + 1$ is irreducible over \mathbb{Q} .

In Exercises 18 through 21, determine whether the polynomial in $\mathbb{Z}[x]$ satisfies an Eisenstein criterion for irreducibility over \mathbb{Q} .

18. $x^2 - 12$
19. $8x^3 + 6x^2 - 9x + 24$
20. $4x^{10} - 9x^3 + 24x - 18$
21. $2x^{10} - 25x^3 + 10x^2 - 30$
22. Find all zeros of $6x^4 + 17x^3 + 7x^2 + x - 10$ in \mathbb{Q} . (This is a tedious high school algebra problem. You might use a bit of analytic geometry and calculus and make a graph, or use Newton's method to see which are the best candidates for zeros.)

Concepts

In Exercises 23 and 24, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

23. A polynomial $f(x) \in F[x]$ is *irreducible over the field F* if and only if $f(x) \neq g(x)h(x)$ for any polynomials $g(x), h(x) \in F[x]$.
24. A nonconstant polynomial $f(x) \in F[x]$ is *irreducible over the field F* if and only if in any factorization of it in $F[x]$, one of the factors is in F .
25. Determine whether each of the following is true or false.
- $x - 2$ is irreducible over \mathbb{Q} .
 - $3x - 6$ is irreducible over \mathbb{Q} .
 - $x^2 - 3$ is irreducible over \mathbb{Q} .
 - $x^2 + 3$ is irreducible over \mathbb{Z}_7 .
 - If F is a field, the units of $F[x]$ are precisely the nonzero elements of F .
 - If F is a field, the units of $F(x)$ are precisely the nonzero elements of F .
 - A polynomial $f(x)$ of degree n with coefficients in a field F can have at most n zeros in F .
 - A polynomial $f(x)$ of degree n with coefficients in a field F can have at most n zeros in any given field E such that $F \leq E$.
 - Every polynomial of degree 1 in $F[x]$ has at least one zero in the field F .
 - Each polynomial in $F[x]$ can have at most a finite number of zeros in the field F .
26. Find all prime numbers p such that $x + 2$ is a factor of $x^4 + x^3 + x^2 - x + 1$ in $\mathbb{Z}_p[x]$.

In Exercises 27 through 30, find all irreducible polynomials of the indicated degree in the given ring.

27. Degree 2 in $\mathbb{Z}_2[x]$
28. Degree 3 in $\mathbb{Z}_2[x]$
29. Degree 2 in $\mathbb{Z}_3[x]$
30. Degree 3 in $\mathbb{Z}_3[x]$
31. Find the number of irreducible quadratic polynomials in $\mathbb{Z}_p[x]$, where p is a prime. [Hint: Find the number of reducible polynomials of the form $x^2 + ax + b$, then the number of reducible quadratics, and subtract this from the total number of quadratics.]

Proof Synopsis

32. Give a synopsis of the proof of Corollary 28.6.
33. Give a synopsis of the proof of Corollary 28.7.

Theory

34. Show that for p a prime, the polynomial $x^p + a$ in $\mathbb{Z}_p[x]$ is not irreducible for any $a \in \mathbb{Z}_p$.
35. If F is a field and $a \neq 0$ is a zero of $f(x) = a_0 + a_1x + \cdots + a_nx^n$ in $F[x]$, show that $1/a$ is a zero of $a_n + a_{n-1}x + \cdots + a_0x^n$.
36. (Remainder Theorem) Let $f(x) \in F[x]$ where F is a field, and let $\alpha \in F$. Show that the remainder $r(x)$ when $f(x)$ is divided by $x - \alpha$, in accordance with the division algorithm, is $f(\alpha)$.
37. Let $\sigma_m : \mathbb{Z} \rightarrow \mathbb{Z}_m$ be the natural homomorphism given by $\sigma_m(a) = (\text{the remainder of } a \text{ when divided by } m)$ for $a \in \mathbb{Z}$.
- Show that $\overline{\sigma_m} : \mathbb{Z}[x] \rightarrow \mathbb{Z}_m[x]$ given by

$$\overline{\sigma_m}(a_0 + a_1x + \cdots + a_nx^n) = \sigma_m(a_0) + \sigma_m(a_1)x + \cdots + \sigma_m(a_n)x^n$$
is a homomorphism of $\mathbb{Z}[x]$ onto $\mathbb{Z}_m[x]$.
 - Show that if $f(x) \in \mathbb{Z}[x]$ and $\overline{\sigma_m}(f(x))$ both have degree n and $\overline{\sigma_m}(f(x))$ does not factor in $\mathbb{Z}_m[x]$ into two polynomials of degree less than n , then $f(x)$ is irreducible in $\mathbb{Q}[x]$.
 - Use part (b) to show that $x^3 + 17x + 36$ is irreducible in $\mathbb{Q}[x]$. [Hint: Try a prime value of m that simplifies the coefficients.]

The goal of Exercises 38 through 40 is to prove Theorem 28.12.

38. Let $f(x) \in \mathbb{Z}[x]$. We say that $f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_0$ is **primitive** if the greatest common divisor of the coefficients a_0, a_1, \dots, a_n is 1. Prove the product of two primitive polynomials is primitive.
39. Let $f(x) \in \mathbb{Z}[x]$. The **content** of $f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_0$ is defined to be the greatest common divisor of a_0, a_1, \dots, a_n and it is denoted $\text{cont}(f(x))$. Prove that $\text{cont}(f(x)g(x)) = \text{cont}(f(x)) \cdot \text{cont}(g(x))$ for any $f(x), g(x) \in \mathbb{Z}[x]$. (Hint: Use Exercise 38.)
40. Prove Theorem 28.12. (Hint: Use Exercise 39.)

SECTION 29**[†]ALGEBRAIC CODING THEORY**

Suppose you wish to send a message, but occasionally the transmission line makes an error. When an error occurs, it would be nice if the receiver could detect that there is an error and ask you to resend the message. In other situations, such as a space probe transmitting images back to earth, it may be impossible to resend the data. In this case, it would be desirable if the receiving earthling could not only detect, but also correct a transmission error.

We will think of a message as an element in $\mathbb{Z}_2^k = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2$. Each message consists of a string of zeros and ones of length k . Each of the \mathbb{Z}_2 entries will be referred to as a **bit**. Coding theory in general allows transmitted messages to be in F^n for any finite field F , but for our introduction to the subject we will restrict our attention to $F = \mathbb{Z}_2$.

- 29.1 Example** A common way to detect a single-bit error is to use a parity check bit. Instead of transmitting a byte consisting of eight bits, that is, an element in \mathbb{Z}_2^8 , nine bits are transmitted with the last bit being the sum in \mathbb{Z}_2 of the first eight bits. The message

$$(1, 1, 0, 1, 0, 0, 1, 1)$$

would be transmitted as

$$(1, 1, 0, 1, 0, 0, 1, 1, 1).$$

Regardless of whether the 8-bit message has an even or odd number of ones, the transmitted string of 9 bits has an even number of ones. If the sum of the nine bits of the received message is not zero, then a transmission error must have occurred. ▲

[†] This section is not used in the remainder of the text.

29.2 Example An inefficient, but possible method of correcting transmission errors is to send a message three times. If two of the received messages agree, then that common message is accepted as the most likely correct message. In this case, if there is only one error, the original message will be retrieved. ▲

29.3 Definition A **code** is a subset $C \subseteq \mathbb{Z}_2^n$. An element of C is a **code word**. The **length** of a code word in $C \subseteq \mathbb{Z}_2^n$ is n . ■

In practice, when a message is to be sent, it is broken into shorter pieces consisting of k bits. A predetermined one-to-one function $f : \mathbb{Z}_2^k \rightarrow C$ mapping all possible k bit messages to code words is then applied to the message pieces and transmitted. The receiver then checks that each received message piece is in the range of f . If so, the sent code word is most likely the received code word and the message corresponding to the received code word can be computed since f is one-to-one. If the received message is not a code word, then a transmission error occurred. We will not concern ourselves with the function f . Instead, we will investigate certain types of codes. We restrict our attention to linear codes as defined below.

29.4 Definition A **linear code** is a subgroup C of \mathbb{Z}_2^n . Since C is a subgroup of \mathbb{Z}_2^n , the order of C is 2^k for some integer k . The **information rate** or **rate** of the linear code is the ratio $\frac{k}{n}$. A linear code is **cyclic** if for any code word $(a_0, a_1, \dots, a_{n-1}), (a_{n-1}, a_0, a_1, \dots, a_{n-2})$ is also a code word. That is, a linear code is cyclic if a cyclic shift of any code word is a code word. ■

An information rate of $\frac{k}{n}$ means that in order to transmit a message of length k , n bits are required. It is clearly desirable to make the information rate as large as possible subject to the desired number of bit errors that can be detected or corrected.

29.5 Example Let $C \subseteq \mathbb{Z}_2^9$ be the set of all strings of length 9 such that the sum of the bits is 0 modulo 2 as in Example 29.1. Note that C is the kernel of the group homomorphism

$$\phi : \mathbb{Z}_2^9 \rightarrow \mathbb{Z}_2$$

given by

$$\phi(a_0, a_1, \dots, a_8) = a_0 + a_1 + \dots + a_8 \pmod{2}.$$

Thus C is a subgroup of \mathbb{Z}_2^9 and therefore C is a linear code. In this example, $n = 9$ and $k = 8$ since C is a subgroup of \mathbb{Z}_2^9 with index 2. Thus C has an information rate of $\frac{8}{9}$. Furthermore, the code is cyclic since any cyclic shift of a code word does not change the number of ones. ▲

If two code words differ in only one position, then it would not be possible to detect every error that occurs in just one bit. If any pair of code words differ in two or more positions, then any error of just one bit could be detected, that is, it could be determined that there is an error, but it may not be possible to reconstruct the original code word. Furthermore, if any pair of code words differ at three or more positions, then an error of just one bit could not only be detected, but it could be corrected since only one code word would differ from the erroneous word at one position.

29.6 Definition The **Hamming weight** or **weight** of a string in \mathbb{Z}_2^n is the number of ones in the string. The **Hamming distance** or **distance** between two strings in \mathbb{Z}_2^n is the number of bits where the two strings differ. ■

29.7 Example The Hamming weight of the string $(1, 0, 0, 1, 1, 0, 1, 1)$ is 5. The Hamming distance between the code words $(1, 0, 0, 0, 1, 1, 0, 1)$ and $(1, 1, 0, 1, 0, 0, 0, 1)$ is 4. ▲

29.8 Theorem For a linear code C , the minimum weight among the nonzero code words of C is the same as the minimum distance between two different code words.

Proof For any two code words $w, u \in \mathbb{Z}_2^n$, the distance between w and u is the number of bits where the words differ. That is, the weight of $w - u$ is the distance between w and u . Since C is a subgroup of \mathbb{Z}_2^n , $w - u \in C$. Thus the minimum weight of nonzero code words is less than or equal to the minimum distance between two different code words. We also notice that $0 \in \mathbb{Z}_2^n$ is a code word, so the weight of a code word w is the distance between 0 and w , which implies that the minimum distance between two different code words is less than or equal to the minimum weight among the nonzero code words. \blacklozenge

If the Hamming distance between any two different code words in a code C is at least d , then we say that C **detects $d - 1$ bit errors** since any change to a code word in at most $d - 1$ bits is not a code word. If C is a code in \mathbb{Z}_2^n and for any string $v \in \mathbb{Z}_2^n$, there is at most one code word whose Hamming distance from v is d or less, then we say that C **corrects d bit errors**. The idea is that if a string is received that is not a code word, then the best guess for the sent code word is the code word that is closest to the received string. For a code that corrects d bit errors, by taking the closest code to a received string we reconstruct the sent code word as long as the number of errors is at most d .

29.9 Example Let $C = \{(0, 0, 0, 0), (1, 0, 1, 0), (1, 1, 0, 1), (0, 1, 1, 1)\} \subseteq \mathbb{Z}_2^4$. It is not difficult to check that C is a subgroup of \mathbb{Z}_2^4 , so C is a linear code. The code word $(1, 0, 1, 0)$ has weight 2 and the other two nonzero code words have weight 3. By Theorem 29.8, the minimum distance between any two code words is 2. Thus C detects one-bit errors, but it cannot correct a one-bit error. A received message of $m = (1, 0, 0, 0)$ differs from both $(0, 0, 0, 0)$ and $(1, 0, 1, 0)$ by only one bit, so even if we know m is only incorrect in one position, we would not know if the sent code word was $(0, 0, 0, 0)$ or $(1, 0, 1, 0)$. \blacktriangle

There are many schemes to generate codes having various properties, but we will focus on just one method. We can think of an element $(a_0, a_1, a_2, \dots, a_{n-1}) \in \mathbb{Z}_2^n$ as corresponding to the coefficients of the polynomial $a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} \in \mathbb{Z}_2[x]$. In this way, instead of thinking of code words as strings of zeros and ones of length n , we can think of them as polynomials in $\mathbb{Z}_2[x]$ of degree at most $n - 1$. We note that this correspondence is a group isomorphism ϕ mapping \mathbb{Z}_2^n onto the additive group of polynomials in $\mathbb{Z}_2[x]$ of degree at most $n - 1$.

29.10 Example Let $n = 5$ and $g(x) = x^2 + x + 1$. We define C to be all the multiples of $x^2 + x + 1$, including 0, whose degree is less than 5.

$$\begin{aligned} C &= \{f(x)g(x) \mid f(x) \in \mathbb{Z}_2[x] \text{ and either } \deg(f(x)) \leq 2 \text{ or } f(x) = 0\} \\ &= \{0 \cdot g(x), 1 \cdot g(x), x \cdot g(x), (x+1) \cdot g(x), \\ &\quad x^2 \cdot g(x), (x^2+1) \cdot g(x), (x^2+x) \cdot g(x), (x^2+x+1) \cdot g(x)\} \\ &= \{0, x^2+x+1, x^3+x^2+x, x^3+1, \\ &\quad x^4+x^3+x^2, x^4+x^3+x+1, x^4+x, x^4+x^2+1\}. \end{aligned}$$

By reading off the coefficients of these polynomials we determine the code words to be

$$\begin{aligned} &(0, 0, 0, 0, 0) (0, 0, 1, 1, 1) (0, 1, 1, 1, 0) (0, 1, 0, 0, 1) \\ &(1, 1, 1, 0, 0) (1, 1, 0, 1, 1) (1, 0, 0, 1, 0) (1, 0, 1, 0, 1). \end{aligned}$$

It is not difficult to check that this collection of elements in \mathbb{Z}_2^5 is a subgroup of \mathbb{Z}_2^5 and therefore gives a linear code. The code is not cyclic since $(0, 1, 0, 0, 1)$ is a code word,

but $(1, 0, 1, 0, 0)$ is not a code word. We see that the minimum weight among all the nonzero code words is 2. By Theorem 29.8, the minimum Hamming distance between any two code words is also 2, which implies that the code detects a one-bit error, but it does not correct a one-bit error. \blacktriangle

In Example 29.10, we simply read off the coefficients of the polynomials in C to construct a linear code. For the rest of this section we will abuse notation slightly by referring to a set C of polynomials in $\mathbb{Z}_2[x]$ as a linear code if C is a subgroup of $\mathbb{Z}_2[x]$ containing no polynomial of degree n or larger. The fact that ϕ mapping \mathbb{Z}_2^n to the polynomials of degree at most $n - 1$ is a group isomorphism assures us that any subgroup $C \leq \mathbb{Z}_2[x]$ having no polynomial of degree n or larger provides a linear code by simply reading off the coefficients of the polynomials in C .

29.11 Theorem Let $g(x)$ be a polynomial in $\mathbb{Z}_2[x]$ of degree less than n . Then $C = \{f(x)g(x) \mid f(x) \in \mathbb{Z}_2[x] \text{ and either } f(x) = 0 \text{ or } \deg(f(x)) < n - \deg(g(x))\}$ is a linear code. Furthermore, if the polynomial $g(x)$ is a factor of $x^n + 1$ in $\mathbb{Z}_2[x]$, then C is a cyclic code.

Proof We first show that C is closed under addition. Let $f(x), h(x) \in \mathbb{Z}_2[x]$ so that each either has degree less than $n - \deg(g(x))$ or is the 0 polynomial. Then $f(x) + h(x)$ is either the zero polynomial or else its degree is less than $n - \deg(g(x))$. Therefore

$$f(x)g(x) + h(x)g(x) = (f(x) + h(x))g(x),$$

which implies that C is closed under addition. Also C contains the 0 polynomial and if $f(x)g(x) \in C$, then $-(f(x)g(x)) = f(x)g(c) \in C$. Thus C is a subgroup of the additive group $G = \{w(x) \in \mathbb{Z}_2[x] \mid w(x) = 0 \text{ or } \deg(w(x)) < n\}$, which means that C is a linear code.

Now we assume that $g(x)$ is a factor of $x^n + 1$ in $\mathbb{Z}_2[x]$, that is, there is a polynomial $h(x) \in \mathbb{Z}_2[x]$ with

$$h(x)g(x) = x^n + 1.$$

Apparently,

$$\deg(h(x)) = n - \deg(g(x)).$$

Let $f(x)g(x) \in C$. If

$$\deg(f(x)g(x)) < n - 1,$$

then

$$(xf(x))g(x) \in C$$

and $xf(x)g(x)$ simply increases by one the degree of each term in the polynomial $f(x)g(x)$. This implies that $xf(x)g(x) \in C$ is a cyclic shift of $f(x)g(x)$. On the other hand, if

$$\deg(f(x)g(x)) = n - 1,$$

then a cyclic shift of the code word $f(x)g(x)$ is

$$p(x) = xf(x)g(x) + (x^n + 1).$$

We have

$$\begin{aligned} xf(x)g(x) + (x^n + 1) &= xf(x)g(x) + h(x)g(x) \\ &= (xf(x) + h(x))g(x) \end{aligned}$$

Since $xf(x)$ and $h(x)$ each have degree $n - \deg(g(x))$, the coefficient of $x^{n-\deg(g(x))}$ in their sum is 0. So either $xf(x) + h(x) = 0$ or $\deg(xf(x) + h(x)) < n - \deg(g(x))$. In either case, the cyclic shift $xf(x)g(x) + (x^n + 1)$ is a code word in C . Therefore, C is a cyclic code. \blacklozenge

29.12 Definition The code C in Theorem 29.11 is called the **polynomial code of length n generated by $g(x)$** . \blacksquare

29.13 Example Find the code words for C , the polynomial code of length 7 generated by the polynomial $g(x) = x^3 + x^2 + 1$. What is the information rate for C ? Determine if C detects a one-bit error and if so, can C correct a one-bit error? What about detecting and correcting two-bit errors?

Solution As in Example 29.10, one method of finding all the code words is to multiply every polynomial of degree 3 or less by $g(x)$, but there is a much simpler method if the code is cyclic. The polynomial $x^7 + 1$ can be seen to factor in $\mathbb{Z}_2[x]$ as

$$x^7 + 1 = (x^3 + x^2 + 1)(x^4 + x^3 + x^2 + 1)$$

simply by using long division of polynomials. Therefore C is a cyclic code by Theorem 29.11. Since $1 \cdot g(x) = g(x) \in C$ and C contains all cyclic shifts of $g(x)$, we have all the polynomials in the first column of Figure 29.14 as code words in C . Since C is a group, $(x^3 + x^2 + 1) + (x^4 + x^3 + x) = x^4 + x^2 + x + 1 \in C$. The fact that C is cyclic implies the second column of Figure 29.14 is contained in C . There are $2^4 = 16$ polynomials of degree less than 4 (including the zero polynomial) with coefficients in \mathbb{Z}_2 . Thus C contains 16 elements. Since C is a subgroup, the zero polynomial is in C , leaving only one more polynomial to complete the list. This polynomial must remain the same when a cyclic shift is applied. Other than the polynomial 0, the only polynomial that remains the same when a cyclic shift is applied is

$$x^6 + x^5 + x^4 + x^3 + x^2 + x + 1.$$

Thus Figure 29.14 gives the code C as polynomials. Figure 29.15 gives the code as elements in \mathbb{Z}_2^7 .

Since $|C| = 2^4$ and the code word length is 7, the information rate is $\frac{4}{7}$.

It is easy to see that the minimum weight among all the nonzero code words is 3. By Theorem 29.8, the minimum distance between code words is 3. So not only can a single-bit error be detected, it can be corrected. Since the distance between any two code words is at least 3, the code detects two-bit errors. However, the code does not correct two-bit errors since a two-bit error could produce a word with Hamming distance of one from another code word. For example, $(0, 0, 0, 0, 0, 1)$ differs from the code word $(0, 0, 0, 1, 1, 0, 1)$ in two bits, but it differs from the code word $(0, 0, 0, 0, 0, 0)$ in only one bit. \blacktriangle

$$\begin{array}{llll} x^3 + x^2 + 1 & x^4 + x^2 + x + 1 & 0 & x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\ x^4 + x^3 + x & x^5 + x^3 + x^2 + x & & \\ x^5 + x^4 + x^2 & x^6 + x^4 + x^3 + x^2 & & \\ x^6 + x^5 + x^3 & x^5 + x^4 + x^3 + 1 & & \\ x^6 + x^4 + 1 & x^6 + x^5 + x^4 + x & & \\ x^5 + x + 1 & x^6 + x^5 + x^2 + 1 & & \\ x^6 + x^2 + x & x^6 + x^3 + x + 1 & & \end{array}$$

29.14 Figure

(0,0,0,1,1,0,1)	(0,0,1,0,1,1,1)	(0,0,0,0,0,0,0)	(1,1,1,1,1,1,1)
(0,0,1,1,0,1,0)	(0,1,0,1,1,1,0)		
(0,1,1,0,1,0,0)	(1,0,1,1,1,0,0)		
(1,1,0,1,0,0,0)	(0,1,1,1,0,0,1)		
(1,0,1,0,0,0,1)	(1,1,1,0,0,1,0)		
(0,1,0,0,0,1,1)	(1,1,0,0,1,0,1)		
(1,0,0,0,1,1,0)	(1,0,0,1,0,1,1)		

29.15 Figure

Examples 29.2 and 29.13 each provide a code that can correct a one-bit error. Example 29.2 requires sending 24 bits to transmit a message of length 8. That is, the information rate is $\frac{1}{3}$. In Example 29.13, in order to transmit a message of length 8, 14 bits are required and the information rate is $\frac{4}{7}$. Clearly the code in Example 29.13 is a much more efficient way of coding data for transmission.

■ EXERCISES 29

1. If a code has word length 10 and transmission rate of $\frac{1}{2}$, how many code words are in the code?
2. If a linear code contains exactly 16 code words and the transmission rate is $\frac{2}{3}$, find the length of code words.
3. Find the smallest cyclic linear code C that contains $(1, 0, 0, 0, 0)$.
4. Find all cyclic linear codes C in \mathbb{Z}_2^5 that have a transmission rate of $\frac{2}{5}$.
5. Find all cyclic linear codes of length n for
 - a. $n = 2$
 - b. $n = 3$
 - c. $n = 4$
6. Determine whether each of the following is true or false.
 - a. A code is a subset of \mathbb{Z}_2^n for some positive integer n .
 - b. The length of a code word in \mathbb{Z}_2^n is n .
 - c. Every code is a linear code.
 - d. If the Hamming distance between any two different code words is at least 4, then the code corrects two-bit errors.
 - e. If C is a linear code in \mathbb{Z}_2^n , then the information rate is the number of elements in C divided by the number of elements in \mathbb{Z}_2^n .
 - f. Every linear code contains the code word consisting of all zeros.
 - g. If the Hamming distance between two code words in a linear code is d , then there is a code word with Hamming weight d .
 - h. The set $\{f(x)g(x) \mid f(x) \in \mathbb{Z}_2[x]\}$ is the polynomial code of length n generated by $g(x)$ if $g(x) \in \mathbb{Z}_2[x]$ and $g(x)$ has degree n .
 - i. Not every polynomial code is cyclic.
 - j. Every cyclic linear code contains at most two code words that remain the same when a cyclic shift is applied.
7. Let $g(x) = x^3 + x + 1 \in \mathbb{Z}_2[x]$.
 - a. Verify that $g(x)$ is a factor of $x^7 + 1$ in $\mathbb{Z}_2[x]$.
 - b. Find all the code words in the polynomial code C of length 7 generated by $g(x)$.
 - c. Determine if C detects single-bit errors and if so, determine if it corrects single-bit errors.
 - d. Determine if C detects two-bit errors and if so, determine if it corrects two-bit errors.

8. The transmission of a code word from the previous exercise produced the polynomial $p(x) = x^6 + x^5 + x^4 + x^3$. Was there a transmission error? If so, find the closest code word from C as measured by the Hamming distance.
9. Let $g(x) = x^6 + x^3 + 1 \in \mathbb{Z}_2[x]$.
 - a. Verify that $g(x)$ is a factor of $x^9 + 1$ in $\mathbb{Z}_2[x]$.
 - b. Find all the code words in the polynomial code C of length 9 generated by $g(x)$.
 - c. Determine if C detects single-bit errors and if so, determine if it corrects single-bit errors.
 - d. Determine if C detects two-bit errors and if so, determine if it corrects two-bit errors.
10. Let $g(x) = x^4 + x^3 + x + 1 \in \mathbb{Z}_2[x]$ and let C be the code generated by $g(x)$ with code word length 7.
 - a. Is C cyclic?
 - b. Find all the code words in the polynomial code C of length 7 generated by $g(x)$.
 - c. Can C detect one-bit errors and if so, can C correct one-bit errors?
 - d. Can C detect two-bit errors and if so, can C correct two-bit errors?
11. Find six polynomials $g(x) \in \mathbb{Z}_2[x]$ so that the code generated by $g(x)$ with code words of length 9 is a cyclic code.
12. If the minimal weight among all nonzero code words in a cyclic linear code $C \subseteq \mathbb{Z}_2^n$ is 1, prove that $C = \mathbb{Z}_2^n$.
13. Let $g(x)$ be a polynomial in $\mathbb{Z}_2[x]$. Prove that if the polynomial code C generated by $g(x)$ with length n is cyclic, then $g(x)$ is a factor of $x^n + 1$ in $\mathbb{Z}_2[x]$.
14. Let $C \subseteq \mathbb{Z}_2^n$ be a linear code with d the minimal weight among the nonzero code words. Determine necessary and sufficient conditions on d for C to correct k -bit errors.
15. Let $C \subseteq \mathbb{Z}_2^n$ be a linear code. Show that as a group, C is isomorphic with \mathbb{Z}_2^k for some k .
16. Is there a polynomial $g(x) \in \mathbb{Z}_2[x]$ such that the code generated by $g(x)$ of length 9 is the same code as in Example 29.5? Prove your answer.

SECTION 30**HOMOMORPHISMS AND FACTOR RINGS****Factor Rings**

In Section 12 we investigated which subgroups of a given groups could be used to form a factor group. In this section we wish to do an analogous construction on a ring to form a factor ring. We start with an example.

30.1 Example For any $n \in \mathbb{Z}$, $n\mathbb{Z}$ is a subring of \mathbb{Z} . Thinking of \mathbb{Z} as an abelian group, we know that $n\mathbb{Z}$ is a normal subgroup of \mathbb{Z} . As we have seen, $\mathbb{Z}/n\mathbb{Z} = \{a + n\mathbb{Z} \mid a \in \mathbb{Z}\}$ forms a group using addition defined by adding coset representatives. Furthermore, $\mathbb{Z}/n\mathbb{Z}$ is a ring where multiplication is defined by

$$(a + n\mathbb{Z})(b + n\mathbb{Z}) = ab + n\mathbb{Z}.$$

We check that this multiplication is well defined. Let $a' \in a + n\mathbb{Z}$ and $b' \in b + n\mathbb{Z}$. Then $a' = a + nk$ and $b' = b + nr$ for some integers k and r . Thus

$$\begin{aligned} a'b' &= (a + nk)(b + nr) \\ &= ab + n(kb + knr) + anr \\ &= ab + n(kb + knr + ar) \\ &\in ab + n\mathbb{Z}. \end{aligned}$$

From this calculation we see that regardless of which representatives from $a + n\mathbb{Z}$ and $b + n\mathbb{Z}$ we pick, our product is in the coset $ab + n\mathbb{Z}$. So we have a well-defined multiplication on the cosets of $n\mathbb{Z}$. ▲

Looking at the second line of the above computation, we can see that what was needed to verify $a'b' \in ab + n\mathbb{Z}$ is that $n(kb + knr) + anr \in n\mathbb{Z}$. The key to make this computation work is that when an element of \mathbb{Z} is multiplied by an element of $n\mathbb{Z}$, the product is in $n\mathbb{Z}$. This observation is the reason for the following definition.

30.2 Definition An additive subgroup N of the ring R is an **ideal** if

$$aN = \{an \mid n \in N\} \subseteq N \quad \text{and} \quad Na = \{na \mid n \in N\} \subseteq N \quad \text{for all } a \in R. \quad \blacksquare$$

30.3 Example We see that $n\mathbb{Z}$ is an ideal in the ring \mathbb{Z} since we know it is a subring, and $s(nm) = (nm)s = n(ms) \in n\mathbb{Z}$ for all $s \in \mathbb{Z}$. \blacktriangle

30.4 Example Let F be the ring of all functions mapping \mathbb{R} into \mathbb{R} , and let C be the subring of F consisting of all the constant functions in F . Is C an ideal in F ? Why?

Solution It is not true that the product of a constant function with every function is again a constant function. For example, the product of $\sin x$ and 2 is the function $2 \sin x$. Thus C is not an ideal of F . \blacktriangle

■ HISTORICAL NOTE

It was Ernst Eduard Kummer (1810–1893) who introduced the concept of an “ideal complex number” in 1847 in order to preserve the notion of unique factorization in certain rings of algebraic integers. In particular, Kummer wanted to be able to factor into primes numbers of the form $a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{p-1}\alpha^{p-1}$, where α is a complex root of $x^p = 1$ (p prime) and the a_i are ordinary integers. Kummer had noticed that the naive definition of primes as “unfactorable numbers” does not lead to the expected results; the product of two such “unfactorable” numbers may well be divisible by other “unfactorable” numbers. Kummer defined “ideal prime factors” and “ideal numbers” in terms of certain congruence relationships; these “ideal factors” were then used as the divisors

necessary to preserve unique factorization. By use of these, Kummer was in fact able to prove certain cases of Fermat’s Last Theorem, which states that $x^n + y^n = z^n$ has no solutions $x, y, z \in \mathbb{Z}^+$ if $n > 2$.

It turned out that an “ideal number,” which was in general not a “number” at all, was uniquely determined by the set of integers it “divided.” Richard Dedekind took advantage of this fact to identify the ideal factor with this set; he therefore called the set itself an ideal, and proceeded to show that it satisfied the definition given in the text. Dedekind was then able to define the notions of prime ideal and product of two ideals and show that any ideal in the ring of integers of any algebraic number field could be written uniquely as a product of prime ideals.

30.5 Example Let F be as in the preceding example, and let N be the subring of all functions f such that $f(2) = 0$. Is N an ideal in F ? Why or why not?

Solution Let $f \in N$ and let $g \in F$. Then $(fg)(2) = f(2)g(2) = 0g(2) = 0$, so $fg \in N$. Similarly, we find that $gf \in N$. Therefore N is an ideal of F . \blacktriangle

30.6 Theorem **(Analogue of Theorem 12.7)** Let H be an additive subgroup of the ring R . Multiplication of additive cosets of H is well defined by the equation

$$(a + H)(b + H) = ab + H$$

if and only if H is an ideal in R .

Proof Suppose first that H is an ideal in R . Let $a, b \in R$, $a' \in a + H$, and $b' \in b + H$. There are elements $h_1, h_2 \in H$ with $a' = a + h_1$ and $b' = b + h_2$. We have

$$\begin{aligned} a'b' &= (a + h_1)(b + h_2) \\ &= ab + ah_2 + h_1b + h_1h_2 \\ &\in ab + H \quad \text{since } H \text{ is an ideal.} \end{aligned}$$

We now suppose that $(a + H)(b + H) = ab + H$ defines a binary operation on cosets of H in R . We let $a \in R$ and $h \in H$ with the goal of showing that $aH \subseteq H$ and $Ha \subseteq H$. Since $h + H = 0 + H$,

$$H = 0a + H = (0 + H)(a + H) = (h + H)(a + H) = ha + H.$$

This shows $ha \in H$, which implies $Ha \subseteq H$. Similarly,

$$H = a0 + H = (a + H)(0 + H) = (a + H)(h + H) = ah + H.$$

This shows $ah \in H$ and therefore $aH \subseteq H$. Thus H is an ideal in R . \diamond

Once we know that multiplication by choosing representatives is well defined on additive cosets of a subring N of R , the associative law for multiplication and the distributive laws for these cosets follow immediately from the same properties in R . We have at once this corollary of Theorem 30.6.

30.7 Corollary (**Analogue of Corollary 12.8**) Let N be an ideal of a ring R . Then the additive cosets of N form a ring R/N with the binary operations defined by

$$(a + N) + (b + N) = (a + b) + N$$

and

$$(a + N)(b + N) = ab + N. \quad \diamond$$

30.8 Definition The ring R/N in the preceding corollary is the **factor ring** (or **quotient ring**) of R by N . \blacksquare

If we use the term *quotient ring*, be sure not to confuse it with the notion of the *field of quotients* of an integral domain, discussed in Section 26.

Homomorphisms

We defined the concepts of *homomorphism* and *isomorphism* for rings in Section 22, since we wished to talk about evaluation homomorphisms for polynomials and about isomorphic rings. We repeat some definitions here for easy reference. Recall that a homomorphism is a *structure-relating map*. A homomorphism for rings must relate both their additive structure and their multiplicative structure.

30.9 Definition A map ϕ of a ring R into a ring R' is a **homomorphism** if

$$\phi(a + b) = \phi(a) + \phi(b)$$

and

$$\phi(ab) = \phi(a)\phi(b)$$

for all elements a and b in R . \blacksquare

In Example 22.10 we defined evaluation homomorphisms, and Example 22.11 showed that the map $\phi : \mathbb{Z} \rightarrow \mathbb{Z}_n$, where $\phi(m)$ is the remainder of m when divided

by n , is a homomorphism. We give another simple but very fundamental example of a homomorphism.

30.10 Example (Projection Homomorphisms) Let R_1, R_2, \dots, R_n be rings. For each i , the map $\pi_i : R_1 \times R_2 \times \dots \times R_n \rightarrow R_i$ defined by $\pi_i(r_1, r_2, \dots, r_n) = r_i$ is a homomorphism, *projection onto the i th component*. The two required properties of a homomorphism hold for π_i since both addition and multiplication in the direct product are computed by addition and multiplication in each individual component. \blacktriangle

Properties of Homomorphisms

We continue to parallel our development of ring homomorphisms and factor rings with the analogous material for group homomorphisms and factor groups.

30.11 Theorem Let $\phi : R \rightarrow R'$ be a ring homomorphism.

1. If 0 is the additive identity in R , then $\phi(0) = 0'$ is the additive identity in R' .
2. If $a \in R$, then $\phi(-a) = -\phi(a)$.
3. If S is a subring of R , then $\phi[S]$ is a subring of R' .
4. If S' is a subring of R' , then $\phi^{-1}[S']$ is a subring of R .
5. If R has unity 1 , then $\phi(1)$ is unity for $\phi[R]$.
6. If N is an ideal in R , then $\phi[N]$ is an ideal in $\phi[R]$.
7. If N' is an ideal in either R' or $\phi[R]$, then $\phi^{-1}[N']$ is an ideal in R .

Proof Let ϕ be a homomorphism of a ring R into a ring R' . Since, in particular, ϕ can be viewed as a group homomorphism of $\langle R, + \rangle$ into $\langle R', +' \rangle$, Theorem 8.5 tells us that $\phi(0) = 0'$ is the additive identity element of R' and that $\phi(-a) = -\phi(a)$.

Theorem 8.5 also tells us that if S is a subring of R , then, considering the additive group $\langle S, + \rangle$, the set $\langle \phi[S], +' \rangle$ gives a subgroup of $\langle R', +' \rangle$. If $\phi(s_1)$ and $\phi(s_2)$ are two elements of $\phi[S]$, then

$$\phi(s_1)\phi(s_2) = \phi(s_1s_2)$$

and $\phi(s_1s_2) \in \phi[S]$. Thus $\phi(s_1)\phi(s_2) \in \phi[S]$, so $\phi[S]$ is closed under multiplication. Consequently, $\phi[S]$ is a subring of R' .

Going the other way, Theorem 8.5 also shows that if S' is a subring of R' , then $\langle \phi^{-1}[S'], + \rangle$ is a subgroup of $\langle R, + \rangle$. Let $a, b \in \phi^{-1}[S']$, so that $\phi(a) \in S'$ and $\phi(b) \in S'$. Then

$$\phi(ab) = \phi(a)\phi(b).$$

Since $\phi(a)\phi(b) \in S'$, we see that $ab \in \phi^{-1}[S']$, so $\phi^{-1}[S']$ is closed under multiplication and thus is a subring of R .

If R has unity 1, then for all $r \in R$,

$$\phi(r) = \phi(1r) = \phi(r1) = \phi(1)\phi(r) = \phi(r)\phi(1),$$

so $\phi(1)$ is unity for $\phi[R]$.

The proof of the remainder of the theorem is Exercise 22. \blacklozenge

Note in Theorem 30.11 that $\phi(1)$ is unity for $\phi[R]$, but not necessarily for R' as we ask you to illustrate in Exercise 9. Furthermore, although $\phi[N]$ is an ideal in $\phi[R]$, it may not be an ideal in R' as verified in Exercise 22.

30.12 Definition Let a map $\phi : R \rightarrow R'$ be a homomorphism of rings. The subring

$$\phi^{-1}[0'] = \{r \in R \mid \phi(r) = 0'\}$$

is the **kernel** of ϕ , denoted by $\text{Ker}(\phi)$. ■

If we forget about the multiplicative part of a ring, we see that the kernel of a ring homomorphism is the same as the kernel of the underlying group homomorphism. Any property of a group homomorphism must also hold for a ring homomorphism.

30.13 Theorem **Analogue of Theorem 10.17** Let $\phi : R_1 \rightarrow R_2$ be a ring homomorphism. The elements $a, b \in R_1$ are in the same additive coset of $\text{Ker}(\phi)$ if and only if $\phi(a) = \phi(b)$ ♦

30.14 Theorem **Analogue of Corollary 10.19** A ring homomorphism $\phi : R_1 \rightarrow R_2$ is one-to-one if and only if $\text{Ker}(\phi) = \{0\}$. ♦

The kernel of a group homomorphism $\phi : G_1 \rightarrow G_2$ is a normal subgroup of G_1 and normality is what is needed in order to construct a factor group from a subgroup. The situation is similar in rings. We need a subring to be an ideal in order to construct a factor ring. The following theorem states that in fact the kernel of a ring homomorphism is an ideal.

30.15 Theorem Let $\phi : R_1 \rightarrow R_2$ be a ring homomorphism. Then $\text{Ker}(\phi)$ is an ideal in R_1 .

Proof Since $\{0\} \subset R_2$ is an ideal in R_2 , $\text{Ker}(\phi) = \phi^{-1}[\{0\}]$ is an ideal in R_1 by Property 7 of Theorem 30.11. ♦

Fundamental Homomorphism Theorem

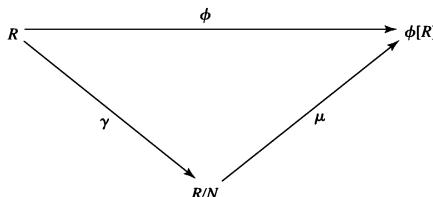
To complete our analogy with groups, we give the analogues of Theorems 12.12 and 12.14.

30.16 Theorem **(Analogue of Theorem 12.12)** Let N be an ideal of a ring R . Then $\gamma : R \rightarrow R/N$ given by $\gamma(x) = x + N$ is a ring homomorphism with kernel N .

Proof The additive part is done in Theorem 12.12. Turning to the multiplicative question, we see that

$$\gamma(xy) = (xy) + N = (x + N)(y + N) = \gamma(x)\gamma(y).$$

30.17 Theorem **(Fundamental Homomorphism Theorem; Analogue of Theorem 12.14)** Let $\phi : R \rightarrow R'$ be a ring homomorphism with kernel N . Then $\phi[R]$ is a ring, and the map $\mu : R/N \rightarrow \phi[R]$ given by $\mu(x + N) = \phi(x)$ is an isomorphism. If $\gamma : R \rightarrow R/N$ is the homomorphism given by $\gamma(x) = x + N$, then for each $x \in R$, we have $\phi(x) = \mu \circ \gamma(x)$.



30.18 Figure

Proof This follows at once from Theorems 30.15 and 30.16. Figure 30.18 is the analogue of Fig. 12.15. \blacklozenge

30.19 Example Example 30.3 shows that $n\mathbb{Z}$ is an ideal of \mathbb{Z} , so we can form the factor ring $\mathbb{Z}/n\mathbb{Z}$. Example 22.11 shows that $\phi : \mathbb{Z} \rightarrow \mathbb{Z}_n$ where $\phi(m)$ is the remainder of m modulo n is a homomorphism, and we see that $\text{Ker}(\phi) = n\mathbb{Z}$. Theorem 30.17 then shows that the map $\mu : \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}_n$ where $\mu(m + n\mathbb{Z})$ is the remainder of m modulo n is well defined and is an isomorphism. \blacktriangle

30.20 Example Continuing Example 30.5, let F be the ring of all functions mapping \mathbb{R} into \mathbb{R} and let N be the subset of F consisting of all functions f with $f(2) = 0$. The set N is an ideal in F , so F/N is a ring. Furthermore, N is the kernel of the evaluation homomorphism $\phi_2 : F \rightarrow \mathbb{R}$ defined by $\phi_2(f) = f(2)$. Since ϕ_2 maps onto \mathbb{R} , F/N is isomorphic with \mathbb{R} by Theorem 30.17. The function $\mu : F/N \rightarrow \mathbb{R}$ given by $\mu(f) = f(2)$ is an isomorphism. \blacktriangle

In summary, every ring homomorphism with domain R gives rise to a factor ring R/N , and every factor ring R/N gives rise to a homomorphism mapping R into R/N . An *ideal* in ring theory is analogous to a *normal subgroup* in the group theory. Both are the type of substructure needed to form a factor structure.

■ EXERCISES 30

Computations

- Describe all ring homomorphisms of $\mathbb{Z} \times \mathbb{Z}$ into $\mathbb{Z} \times \mathbb{Z}$. [Hint: Note that if ϕ is such a homomorphism, then $\phi((1, 0)) = \phi((1, 0))\phi((1, 0))$ and $\phi((0, 1)) = \phi((0, 1))\phi((0, 1))$. Consider also $\phi((1, 0)(0, 1))$.]
- Find all positive integers n such that \mathbb{Z}_n contains a subring isomorphic to \mathbb{Z}_2 .
- Find all ideals N of \mathbb{Z}_{12} . In each case compute \mathbb{Z}_{12}/N ; that is, find a known ring to which the quotient ring is isomorphic.
- Give addition and multiplication tables for $2\mathbb{Z}/8\mathbb{Z}$. Are $2\mathbb{Z}/8\mathbb{Z}$ and \mathbb{Z}_4 isomorphic rings?

Concepts

In Exercises 5 through 7, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

- An *isomorphism* of a ring R with a ring R' is a homomorphism $\phi : R \rightarrow R'$ such that $\text{Ker}(\phi) = \{0\}$.
- An *ideal* N of a ring R is an additive subgroup of $(R, +)$ such that for all $r \in R$ and all $n \in N$, we have $rn \in N$ and $nr \in N$.
- The *kernel* of a homomorphism ϕ mapping a ring R into a ring R' is $\{\phi(r) = 0' \mid r \in R\}$.
- Let F be the ring of all functions mapping \mathbb{R} into \mathbb{R} and having derivatives of all orders. Differentiation gives a map $\delta : F \rightarrow F$ where $\delta(f(x)) = f'(x)$. Is δ a homomorphism? Why? Give the connection between this exercise and Example 30.4.
- Give an example of a ring homomorphism $\phi : R \rightarrow R'$ where R has unity 1 and $\phi(1) \neq 0'$, but $\phi(1)$ is not unity for R' .
- Determine whether each of the following is true or false.
 - The concept of a ring homomorphism is closely connected with the idea of a factor ring.
 - A ring homomorphism $\phi : R \rightarrow R'$ carries ideals of R into ideals of R' .
 - A ring homomorphism is one-to-one if and only if the kernel is $\{0\}$.
 - \mathbb{Q} is an ideal in \mathbb{R} .

- e. Every ideal in a ring is a subring of the ring.
 - f. Every subring of every ring is an ideal of the ring.
 - g. Every quotient ring of every commutative ring is again a commutative ring.
 - h. The rings $\mathbb{Z}/4\mathbb{Z}$ and \mathbb{Z}_4 are isomorphic.
 - i. An ideal N in a ring R with unity 1 is all of R if and only if $1 \in N$.
 - j. The concept of an ideal is to the concept of a ring as the concept of a normal subgroup is to the concept of a group.
11. Let R be a ring. Observe that $\{0\}$ and R are both ideals of R . Are the factor rings R/R and $R/\{0\}$ of real interest? Why?
12. Give an example to show that a factor ring of an integral domain may be a field.
13. Give an example to show that a factor ring of an integral domain may have divisors of 0.
14. Give an example to show that a factor ring of a ring with divisors of 0 may be an integral domain.
15. Find a subring of the ring $\mathbb{Z} \times \mathbb{Z}$ that is not an ideal of $\mathbb{Z} \times \mathbb{Z}$.
16. A student is asked to prove that a quotient ring of a ring R modulo an ideal N is commutative if and only if $(rs - sr) \in N$ for all $r, s \in R$. The student starts out:
Assume R/N is commutative. Then $rs = sr$ for all $r, s \in R/N$.
 - a. Why does the instructor reading this expect an incorrect proof?
 - b. What should the student have written?
 - c. Prove the assertion. (Note the “if and only if.”)

Theory

17. Let $R = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$ and let R' consist of all 2×2 matrices of the form $\begin{bmatrix} a & 2b \\ b & a \end{bmatrix}$ for $a, b \in \mathbb{Z}$. Show that R is a subring of \mathbb{R} and that R' is a subring of $M_2(\mathbb{Z})$. Then show that $\phi : R \rightarrow R'$, where $\phi(a + b\sqrt{2}) = \begin{bmatrix} a & 2b \\ b & a \end{bmatrix}$ is an isomorphism.
18. Show that each homomorphism from a field to a ring is either one-to-one or maps everything onto 0.
19. Show that if R, R' , and R'' are rings, and if $\phi : R \rightarrow R'$ and $\psi : R' \rightarrow R''$ are homomorphisms, then the composite function $\psi\phi : R \rightarrow R''$ is a homomorphism. (See Exercise 39 of Section 8.)
20. Let R be a commutative ring with unity of prime characteristic p . Show that the map $\phi_p : R \rightarrow R$ given by $\phi_p(a) = a^p$ is a homomorphism (the **Frobenius homomorphism**).
21. Let R and R' be rings and let $\phi : R \rightarrow R'$ be a ring homomorphism such that $\phi[R] \neq \{0'\}$. Show that if R has unity 1 and R' has no 0 divisors, then $\phi(1)$ is unity for R' .
22. Let $\phi : R \rightarrow R'$ be a ring homomorphism and let N be an ideal of R .
 - a. Show that $\phi[N]$ is an ideal of $\phi[R]$.
 - b. Give an example to show that $\phi[N]$ need not be an ideal of R' .
 - c. Let N' be an ideal either of $\phi[R]$ or of R' . Show that $\phi^{-1}[N']$ is an ideal of R .
23. Let F be a field, and let S be any subset of $F \times F \times \cdots \times F$ for n factors. Show that the set N_S of all $f(x_1, \dots, x_n) \in F[x_1, \dots, x_n]$ that have every element (a_1, \dots, a_n) of S as a zero (see Exercise 28 of Section 27) is an ideal in $F[x_1, \dots, x_n]$. This is of importance in algebraic geometry.
24. Show that a factor ring of a field is either the trivial (zero) ring of one element or is isomorphic to the field.
25. Show that if R is a ring with unity and N is an ideal of R such that $N \neq R$, then R/N is a ring with unity.
26. Let R be a commutative ring and let $a \in R$. Show that $I_a = \{x \in R \mid ax = 0\}$ is an ideal of R .
27. Show that an intersection of ideals of a ring R is again an ideal of R .
28. Let R and R' be rings and let N and N' be ideals of R and R' , respectively. Let ϕ be a homomorphism of R into R' . Show that ϕ induces a natural homomorphism $\phi_* : R/N \rightarrow R'/N'$ if $\phi[N] \subseteq N'$. (Use Exercise 41 of Section 12.)

29. Let ϕ be a homomorphism of a ring R with unity onto a nonzero ring R' . Let u be a unit in R . Show that $\phi(u)$ is a unit in R' .
30. An element a of a ring R is **nilpotent** if $a^n = 0$ for some $n \in \mathbb{Z}^+$. Show that the collection of all nilpotent elements in a commutative ring R is an ideal, the **nilradical** of R .
31. Referring to the definition given in Exercise 30, find the nilradical of the ring \mathbb{Z}_{12} and observe that it is one of the ideals of \mathbb{Z}_{12} found in Exercise 3. What is the nilradical of \mathbb{Z}' of \mathbb{Z}_{32} ?
32. Referring to Exercise 30, show that if N is the nilradical of a commutative ring R , then R/N has as nilradical the trivial ideal $\{0 + N\}$.
33. Let R be a commutative ring and N an ideal of R . Referring to Exercise 30, show that if every element of N is nilpotent and the nilradical of R/N is R/N , then the nilradical of R is R .
34. Let R be a commutative ring and N an ideal of R . Show that the set \sqrt{N} of all $a \in R$, such that $a^n \in N$ for some $n \in \mathbb{Z}^+$, is an ideal of R , the **radical** of N .
35. Referring to Exercise 34, show by examples that for proper ideals N of a commutative ring R ,
- \sqrt{N} need not equal N
 - \sqrt{N} may equal N .
36. What is the relationship of the ideal \sqrt{N} of Exercise 34 to the nilradical of R/N (see Exercise 30)? Word your answer carefully.
37. Show that $\phi : \mathbb{C} \rightarrow M_2(\mathbb{R})$ given by

$$\phi(a + bi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

for $a, b \in \mathbb{R}$ gives an isomorphism of \mathbb{C} with the subring $\phi[\mathbb{C}]$ of $M_2(\mathbb{R})$.

SECTION 31

PRIME AND MAXIMAL IDEALS

Exercises 12 through 14 of the preceding section asked us to provide examples of factor rings R/N where R and R/N have very different structural properties. We start with some examples of this situation, and in the process, provide solutions to those exercises.

31.1 Example As was shown in Corollary 23.5, the ring \mathbb{Z}_p , which is isomorphic to $\mathbb{Z}/p\mathbb{Z}$, is a field for p a prime. *Thus a factor ring of an integral domain may be a field.* ▲

31.2 Example The ring $\mathbb{Z} \times \mathbb{Z}$ is not an integral domain, for

$$(0, 1)(1, 0) = (0, 0),$$

showing that $(0, 1)$ and $(1, 0)$ are 0 divisors. Let $N = \{(0, n) \mid n \in \mathbb{Z}\}$. Now N is an ideal of $\mathbb{Z} \times \mathbb{Z}$, and $(\mathbb{Z} \times \mathbb{Z})/N$ is isomorphic to \mathbb{Z} under the correspondence $[(m, 0) + N] \leftrightarrow m$, where $m \in \mathbb{Z}$. Thus a *factor ring* of a ring may be an integral domain, even though the original ring is not. ▲

31.3 Example The subset $N = \{0, 3\}$ of \mathbb{Z}_6 is easily seen to be an ideal of \mathbb{Z}_6 , and \mathbb{Z}_6/N has three elements, $0 + N$, $1 + N$, and $2 + N$. These add and multiply in such a fashion as to show that $\mathbb{Z}_6/N \cong \mathbb{Z}_3$ under the correspondence

$$(0 + N) \leftrightarrow 0, \quad (1 + N) \leftrightarrow 1, \quad (2 + N) \leftrightarrow 2.$$

This example shows that *if R is not even an integral domain, that is, if R has zero divisors, it is still possible for R/N to be a field.* ▲

31.4 Example Note that \mathbb{Z} is an integral domain, but $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}_6$ is not. The preceding examples showed that a factor ring may have a structure that seems *better* than the original ring.

This example indicates that the structure of a factor ring may seem *worse* than that of the original ring. ▲

Every nonzero ring R has at least two ideals, the **improper ideal** R and the **trivial ideal** $\{0\}$. For these ideals, the factor rings are R/R , which has only one element, and $R/\{0\}$, which is isomorphic to R . These are uninteresting cases. Just as for a subgroup of a group, a **proper nontrivial ideal** of a ring R is an ideal N of R such that $N \neq R$ and $N \neq \{0\}$.

While factor rings of rings and integral domains may be of great interest, as the above examples indicate, Corollary 31.6, which follows our next theorem, shows that a factor ring of a field is really not useful to us.

31.5 Theorem If R is a ring with unity, and N is an ideal of R containing a unit, then $N = R$.

Proof Let N be an ideal of R , and suppose that $u \in N$ for some unit u in R . Then the condition $rN \subseteq N$ for all $r \in R$ implies, if we take $r = u^{-1}$ and $u \in N$, that $1 = u^{-1}u$ is in N . But then $rN \subseteq N$ for all $r \in R$ implies that $r1 = r$ is in N for all $r \in R$, so $N = R$. ◆

31.6 Corollary A field contains no proper nontrivial ideals.

Proof Since every nonzero element of a field is a unit, it follows at once from Theorem 31.5 that an ideal of a field F is either $\{0\}$ or all of F . ◆

Maximal and Prime Ideals

We now consider the questions of when a factor ring is a field and when it is an integral domain. In our analogy between groups and rings, we noticed that ideals in rings correspond to normal subgroups. Corollary 31.6 states that a field contains no proper nontrivial ideals. In group theory, this corresponds to a group having no proper nontrivial normal subgroups, that is, a simple group. Theorem 13.20 states that a factor group G/H is simple if and only if H is a maximal normal subgroup of G . The following definition is analogous to maximal normal subgroups.

31.7 Definition A **maximal ideal of a ring** R is an ideal M different from R such that there is no proper ideal N of R properly containing M . ■

31.8 Example Let p be a prime positive integer. We know that $\mathbb{Z}/p\mathbb{Z}$ is isomorphic to \mathbb{Z}_p . Forgetting about multiplication for the moment and regarding $\mathbb{Z}/p\mathbb{Z}$ and \mathbb{Z}_p as additive groups, we know that \mathbb{Z}_p is a simple group, and consequently $p\mathbb{Z}$ must be a maximal normal subgroup of \mathbb{Z} by Theorem 13.20. Since \mathbb{Z} is an abelian group and every subgroup is a normal subgroup, we see that $p\mathbb{Z}$ is a maximal proper subgroup of \mathbb{Z} . Since $p\mathbb{Z}$ is an ideal of the ring \mathbb{Z} , it follows that $p\mathbb{Z}$ is a maximal ideal of \mathbb{Z} . We know that $\mathbb{Z}/p\mathbb{Z}$ is isomorphic to the ring \mathbb{Z}_p , and that \mathbb{Z}_p is actually a field. Thus $\mathbb{Z}/p\mathbb{Z}$ is a field. This illustrates the next theorem. ▲

31.9 Theorem (Analogue of Theorem 13.20) Let R be a commutative ring with unity. Then M is a maximal ideal of R if and only if R/M is a field.

Proof We first assume that M is a maximal ideal in R . Since R is a commutative ring with unity, so is R/M . Furthermore, since $M \neq R$, $0 + M \neq 1 + M$ and R/M is a nonzero ring. Let $(a + M) \in R/M$, with $a \notin M$, so that $a + M$ is not the additive identity element of R/M . Suppose $a + M$ has no multiplicative inverse in R/M . Then the set $(R/M)(a + M) = \{(r + M)(a + M) \mid (r + M) \in R/M\}$ does not contain $1 + M$. We easily see that $(R/M)(a + M)$ is an ideal of R/M . It is nontrivial because $a \notin M$, and it is a

proper ideal because it does not contain $1 + M$. By Theorem 30.11, if $\gamma : R \rightarrow R/M$ is the canonical homomorphism, then $\gamma^{-1}[(R/M)(a + M)]$ is a proper ideal of R properly containing M . But this contradicts our assumption that M is a maximal ideal, so $a + M$ must have a multiplicative inverse in R/M .

Conversely, suppose that R/M is a field. By Theorem 30.11, if N is any ideal of R such that $M \subset N \subset R$ and γ is the canonical homomorphism of R onto R/M , then $\gamma[N]$ is an ideal of R/M with $\{(0 + M)\} \subset \gamma[N] \subset R/M$. But this is contrary to Corollary 31.6, which states that the field R/M contains no proper nontrivial ideals. Hence if R/M is a field, then M is maximal. \blacklozenge

31.10 Example Since $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to \mathbb{Z}_n , and \mathbb{Z}_n is a field if and only if n is a prime, we see that the maximal ideals of \mathbb{Z} are precisely the ideals $p\mathbb{Z}$ for prime positive integers p . \blacktriangle

31.11 Corollary A commutative ring with unity is a field if and only if it has no proper nontrivial ideals.

Proof Corollary 31.6 shows that a field has no proper nontrivial ideals.

Conversely, if a commutative ring R with unity has no proper nontrivial ideals, then $\{0\}$ is a maximal ideal and $R/\{0\}$, which is isomorphic to R , is a field by Theorem 31.9. \blacklozenge

We now turn to the question of characterizing, for a commutative ring R with unity, the ideals $N \neq R$ such that R/N is an integral domain. The answer here is rather obvious. The factor ring R/N will be an integral domain if and only if $(a + N)(b + N) = N$ implies that either

$$a + N = N \quad \text{or} \quad b + N = N.$$

This is exactly the statement that R/N has no divisors of 0, since the coset N plays the role of 0 in R/N . Looking at representatives, we see that this condition amounts to saying that $ab \in N$ implies that either $a \in N$ or $b \in N$.

31.12 Example All ideals of \mathbb{Z} are of the form $n\mathbb{Z}$. For $n = 0$, we have $n\mathbb{Z} = \{0\}$, and $\mathbb{Z}/\{0\} \cong \mathbb{Z}$, which is an integral domain. For $n > 0$, we have $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$ and \mathbb{Z}_n is an integral domain if and only if n is a prime. Thus the nonzero ideals $n\mathbb{Z}$ such that $\mathbb{Z}/n\mathbb{Z}$ is an integral domain are of the form $p\mathbb{Z}$, where p is a prime. Of course, $\mathbb{Z}/p\mathbb{Z}$ is actually a field, so that $p\mathbb{Z}$ is a maximal ideal of \mathbb{Z} . Note that for a product rs of integers to be in $p\mathbb{Z}$, the prime p must divide either r or s . The role of prime integers in this example makes the use of the word *prime* in the next definition more reasonable. \blacktriangle

31.13 Definition An ideal $N \neq R$ in a commutative ring R is a **prime ideal** if $ab \in N$ implies that either $a \in N$ or $b \in N$ for $a, b \in R$. \blacksquare

Note that $\{0\}$ is a prime ideal in \mathbb{Z} , and indeed, in any integral domain.

31.14 Example Note that $\mathbb{Z} \times \{0\}$ is a prime ideal of $\mathbb{Z} \times \mathbb{Z}$, for if $(a, b)(c, d) \in \mathbb{Z} \times \{0\}$, then we must have $bd = 0$ in \mathbb{Z} . This implies that either $b = 0$ so $(a, b) \in \mathbb{Z} \times \{0\}$ or $d = 0$ so $(c, d) \in \mathbb{Z} \times \{0\}$. Note that $(\mathbb{Z} \times \mathbb{Z})/(\mathbb{Z} \times \{0\})$ is isomorphic to \mathbb{Z} , which is an integral domain. \blacktriangle

Our remarks preceding Example 31.12 constitute a proof of the following theorem, which is illustrated by Example 31.14.

31.15 Theorem Let R be a commutative ring with unity, and let $N \neq R$ be an ideal in R . Then R/N is an integral domain if and only if N is a prime ideal in R .

31.16 Corollary Every maximal ideal in a commutative ring R with unity is a prime ideal.

Proof If M is maximal in R , then R/M is a field, hence an integral domain, and therefore M is a prime ideal by Theorem 31.15. \blacklozenge

The material that has just been presented regarding maximal and prime ideals is very important and we shall be using it quite a lot. We should keep the main ideas well in mind. We must know and understand the definitions of maximal and prime ideals and must remember the following facts that we have demonstrated.

For a commutative ring R with unity:

1. An ideal M of R is maximal if and only if R/M is a field.
2. An ideal N of R is prime if and only if R/N is an integral domain.
3. Every maximal ideal of R is a prime ideal.

Prime Fields

We now proceed to show that the rings \mathbb{Z} and \mathbb{Z}_n form foundations upon which all rings with unity rest, and that \mathbb{Q} and \mathbb{Z}_p perform a similar service for all fields. Let R be any ring with unity 1. Recall that by $n \cdot 1$ we mean $1 + 1 + \dots + 1$ for n summands for $n > 0$, and $(-1) + (-1) + \dots + (-1)$ for $|n|$ summands for $n < 0$, while $n \cdot 1 = 0$ for $n = 0$.

31.17 Theorem If R is a ring with unity 1, then the map $\phi : \mathbb{Z} \rightarrow R$ given by

$$\phi(n) = n \cdot 1$$

for $n \in \mathbb{Z}$ is a homomorphism of \mathbb{Z} into R .

Proof Observe that

$$\phi(n+m) = (n+m) \cdot 1 = (n \cdot 1) + (m \cdot 1) = \phi(n) + \phi(m).$$

The distributive laws in R show that

$$\underbrace{(1+1+\dots+1)}_{n \text{ summands}} \underbrace{(1+1+\dots+1)}_{m \text{ summands}} = \underbrace{(1+1+\dots+1)}_{nm \text{ summands}}.$$

Thus $(n \cdot 1)(m \cdot 1) = (nm) \cdot 1$ for $n, m > 0$. Similar arguments with the distributive laws show that for all $n, m \in \mathbb{Z}$, we have

$$(n \cdot 1)(m \cdot 1) = (nm) \cdot 1.$$

Thus

$$\phi(nm) = (nm) \cdot 1 = (n \cdot 1)(m \cdot 1) = \phi(n)\phi(m). \quad \blacklozenge$$

31.18 Corollary Let R be a ring with unity. If R has characteristic $n > 1$, then R contains a subring isomorphic to \mathbb{Z}_n . If R has characteristic 0, then R contains a subring isomorphic to \mathbb{Z} .

Proof The map $\phi : \mathbb{Z} \rightarrow R$ given by $\phi(m) = m \cdot 1$ for $m \in \mathbb{Z}$ is a homomorphism by Theorem 31.17. The kernel must be an ideal in \mathbb{Z} . All ideals in \mathbb{Z} are of the form $s\mathbb{Z}$ for some $s \in \mathbb{Z}$. By Theorem 23.14 we see that if R has characteristic $n > 0$, then the kernel of ϕ is $n\mathbb{Z}$. Then the image $\phi[\mathbb{Z}] \leq R$ is isomorphic to $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$. If the characteristic of R is 0, then $m \cdot 1 \neq 0$ for all $m \neq 0$, so the kernel of ϕ is $\{0\}$. Thus, the image $\phi[\mathbb{Z}] \leq R$ is isomorphic to \mathbb{Z} . \blacklozenge

31.19 Theorem A field F is either of prime characteristic p and contains a subfield isomorphic to \mathbb{Z}_p or of characteristic 0 and contains a subfield isomorphic to \mathbb{Q} .

Proof If the characteristic of F is not 0, the above corollary shows that F contains a subring isomorphic to \mathbb{Z}_n . Then n must be a prime p , or F would have 0 divisors. If F is of characteristic 0, then F must contain a subring isomorphic to \mathbb{Z} . In this case Corollaries 26.9 and 26.10 show that F must contain a field of quotients of this subring and that this field of quotients must be isomorphic to \mathbb{Q} . \blacklozenge

Thus every field contains either a subfield isomorphic to \mathbb{Z}_p for some prime p or a subfield isomorphic to \mathbb{Q} . These fields \mathbb{Z}_p and \mathbb{Q} are the fundamental building blocks on which all fields rest.

31.20 Definition The fields \mathbb{Z}_p and \mathbb{Q} are **prime fields**. \blacksquare

Ideal Structure in $F[x]$

Throughout the rest of this section, we assume that F is a field. We give the next definition for a general commutative ring R with unity, although we are only interested in the case $R = F[x]$. Note that for a commutative ring R with unity and $a \in R$, the set $\{ra \mid r \in R\}$ is an ideal in R that contains the element a .

31.21 Definition If R is a commutative ring with unity and $a \in R$, the ideal $\{ra \mid r \in R\}$ of all multiples of a is the **principal ideal generated by a** and is denoted by $\langle a \rangle$. An ideal N of R is a **principal ideal** if $N = \langle a \rangle$ for some $a \in R$. \blacksquare

31.22 Example Every ideal of the ring \mathbb{Z} is of the form $n\mathbb{Z}$, which is generated by n , so every ideal of \mathbb{Z} is a principal ideal. \blacktriangle

31.23 Example The principal ideal $\langle x \rangle$ in $F[x]$ consists of all polynomials in $F[x]$ having zero constant term. \blacktriangle

The next theorem is another simple but very important application of the division algorithm for $F[x]$. (See Theorem 28.2.) The proof of this theorem is to the division algorithm in $F[x]$ as the proof that a subgroup of a cyclic group is cyclic is to the division algorithm in \mathbb{Z} .

31.24 Theorem If F is a field, every ideal in $F[x]$ is principal.

Proof Let N be an ideal of $F[x]$. If $N = \{0\}$, then $N = \langle 0 \rangle$. Suppose that $N \neq \{0\}$, and let $g(x)$ be a nonzero element of N of minimal degree. If the degree of $g(x)$ is 0, then $g(x) \in F$ and is a unit, so $N = F[x] = \langle 1 \rangle$ by Theorem 31.5, so N is principal. If the degree of $g(x)$ is ≥ 1 , let $f(x)$ be any element of N . Then by Theorem 28.2, $f(x) = g(x)q(x) + r(x)$, where $r(x) = 0$ or $(\text{degree } r(x)) < (\text{degree } g(x))$. Now $f(x) \in N$ and $g(x) \in N$ imply that $f(x) - g(x)q(x) = r(x)$ is in N by definition of an ideal. Since $g(x)$ is a nonzero element of minimal degree in N , we must have $r(x) = 0$. Thus $f(x) = g(x)q(x)$ and $N = \langle g(x) \rangle$. \blacklozenge

We can now characterize the maximal ideals of $F[x]$. This is a crucial step in achieving our **basic goal**: to show that any nonconstant polynomial $f(x)$ in $F[x]$ has a zero in some field E containing F .

31.25 Theorem An ideal $\langle p(x) \rangle \neq \{0\}$ of $F[x]$ is maximal if and only if $p(x)$ is irreducible over F .

Proof Suppose that $\langle p(x) \rangle \neq \{0\}$ is a maximal ideal of $F[x]$. Then $\langle p(x) \rangle \neq F[x]$, so $p(x) \notin F$. Let $p(x) = f(x)g(x)$ be a factorization of $p(x)$ in $F[x]$. Since $\langle p(x) \rangle$ is a maximal ideal and hence also a prime ideal, $(f(x)g(x)) \in \langle p(x) \rangle$ implies that $f(x) \in \langle p(x) \rangle$ or $g(x) \in \langle p(x) \rangle$; that is, either $f(x)$ or $g(x)$ has $p(x)$ as a factor. But then we can't have the degrees of both $f(x)$ and $g(x)$ less than the degree of $p(x)$. This shows that $p(x)$ is irreducible over F .

Conversely, if $p(x)$ is irreducible over F , suppose that N is an ideal such that $\langle p(x) \rangle \subseteq N \subseteq F[x]$. Now N is a principal ideal by Theorem 31.24, so $N = \langle g(x) \rangle$ for some $g(x) \in N$. Then $p(x) \in N$ implies that $p(x) = g(x)q(x)$ for some $q(x) \in F[x]$. But $p(x)$ is irreducible, which implies that either $g(x)$ or $q(x)$ is of degree 0. If $g(x)$ is of degree 0, that is, a nonzero constant in F , then $g(x)$ is a unit in $F[x]$, so $\langle g(x) \rangle = N = F[x]$. If $q(x)$ is of degree 0, then $q(x) = c$, where $c \in F$, and $g(x) = (1/c)p(x)$ is in $\langle p(x) \rangle$, so $N = \langle p(x) \rangle$. Thus $\langle p(x) \rangle \subset N \subset F[x]$ is impossible, so $\langle p(x) \rangle$ is maximal. \blacklozenge

31.26 Example Example 28.10 shows that $x^3 + 3x + 2$ is irreducible in $\mathbb{Z}_5[x]$, so $\mathbb{Z}_5[x]/(x^3 + 3x + 2)$ is a field. Similarly, Theorem 27.11 shows that $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$, so $\mathbb{Q}[x]/(x^2 - 2)$ is a field. We shall examine such fields in more detail later. \blacktriangle

Application to Unique Factorization in $F[x]$

In Section 28, we stated without proof Theorem 31.27, which follows. (See Theorem 28.19.) Assuming this theorem, we proved in Section 28 that factorization of polynomials in $F[x]$ into irreducible polynomials is unique, except for order of factors and units in F . We delayed the proof of Theorem 31.27 until now since the machinery we have developed enables us to give such a simple proof. This proof fills the gap in our proof of unique factorization in $F[x]$.

31.27 Theorem Let $p(x)$ be an irreducible polynomial in $F[x]$. If $p(x)$ divides $r(x)s(x)$ for $r(x), s(x) \in F[x]$, then either $p(x)$ divides $r(x)$ or $p(x)$ divides $s(x)$.

Proof Suppose $p(x)$ divides $r(x)s(x)$. Then $r(x)s(x) \in \langle p(x) \rangle$, which is maximal by Theorem 31.25. Therefore, $\langle p(x) \rangle$ is a prime ideal by Corollary 31.16. Hence $r(x)s(x) \in \langle p(x) \rangle$ implies that either $r(x) \in \langle p(x) \rangle$, giving $p(x)$ divides $r(x)$, or that $s(x) \in \langle p(x) \rangle$, giving $p(x)$ divides $s(x)$. \blacklozenge

A Preview of Our Basic Goal

We close this section with an outline of the demonstration in Section 39 of our basic goal. We have all the ideas for the proof at hand now; perhaps you can fill in the details from this outline.

Basic goal: Let F be a field and let $f(x)$ be a nonconstant polynomial in $F[x]$. Show that there exists a field E containing F and containing a zero α of $f(x)$.

Outline of the Proof

1. Let $p(x)$ be an irreducible factor of $f(x)$ in $F[x]$.
2. Let E be the field $F[x]/\langle p(x) \rangle$. (See Theorems 31.25 and 31.9.)
3. Show that no two different elements of E are in the same coset of $F[x]/\langle p(x) \rangle$, and deduce that we may consider E to be (isomorphic to) a subfield of E .
4. Let α be the coset $x + \langle p(x) \rangle$ in E . Show that for the evaluation homomorphism $\phi_\alpha : F[x] \rightarrow E$, we have $\phi_\alpha(f(x)) = 0$. That is, α is a zero of $f(x)$ in E .

An example of a field constructed according to this outline is given in Section 39. There, we give addition and multiplication tables for the field $\mathbb{Z}_2[x]/\langle x^2 + x + 1 \rangle$. We show there that this field has just four elements, the cosets

$$0 + \langle x^2 + x + 1 \rangle, \quad 1 + \langle x^2 + x + 1 \rangle, \quad x + \langle x^2 + x + 1 \rangle,$$

and

$$(x + 1) + \langle x^2 + x + 1 \rangle.$$

We rename these four cosets 0, 1, α , and $\alpha + 1$ respectively, and obtain Tables 39.21 and 39.22 for addition and multiplication in this 4-element field. To see how these tables are constructed, remember that we are in a field of characteristic 2, so that $\alpha + \alpha = \alpha(1 + 1) = \alpha 0 = 0$. Remember also that α is a zero of $x^2 + x + 1$, so that $\alpha^2 + \alpha + 1 = 0$ and consequently $\alpha^2 = -\alpha - 1 = \alpha + 1$.

■ EXERCISES 31

Computations

1. Find all prime ideals and all maximal ideals of \mathbb{Z}_6 .
2. Find all prime ideals and all maximal ideals of \mathbb{Z}_{12} .
3. Find all prime ideals and all maximal ideals of $\mathbb{Z}_2 \times \mathbb{Z}_2$.
4. Find all prime ideals and all maximal ideals of $\mathbb{Z}_2 \times \mathbb{Z}_4$.
5. Find all $c \in \mathbb{Z}_3$ such that $\mathbb{Z}_3[x]/\langle x^2 + c \rangle$ is a field.
6. Find all $c \in \mathbb{Z}_3$ such that $\mathbb{Z}_3[x]/\langle x^3 + x^2 + c \rangle$ is a field.
7. Find all $c \in \mathbb{Z}_3$ such that $\mathbb{Z}_3[x]/\langle x^3 + cx^2 + 1 \rangle$ is a field.
8. Find all $c \in \mathbb{Z}_5$ such that $\mathbb{Z}_5[x]/\langle x^2 + x + c \rangle$ is a field.
9. Find all $c \in \mathbb{Z}_5$ such that $\mathbb{Z}_5[x]/\langle x^2 + cx + 1 \rangle$ is a field.

Concepts

In Exercises 10 through 13, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

10. A *maximal ideal* of a ring R is an ideal that is not contained in any other ideal of R .
11. A *prime ideal* of a commutative ring R is an ideal of the form $pR = \{pr \mid r \in R\}$ for some prime p .
12. A *prime field* is a field that has no proper subfields.
13. A *principal ideal* of a commutative ring with unity is an ideal N with the property that there exists $a \in N$ such that N is the smallest ideal that contains a .
14. Determine whether each of the following is true or false.
 - a. Every prime ideal of every commutative ring with unity is a maximal ideal.
 - b. Every maximal ideal of every commutative ring with unity is a prime ideal.
 - c. \mathbb{Q} is its own prime subfield.
 - d. The prime subfield of \mathbb{C} is \mathbb{R} .
 - e. Every field contains a subfield isomorphic to a prime field.
 - f. A ring with zero divisors may contain one of the prime fields as a subring.
 - g. Every field of characteristic zero contains a subfield isomorphic to \mathbb{Q} .
 - h. Let F be a field. Since $F[x]$ has no divisors of 0, every ideal of $F[x]$ is a prime ideal.
 - i. Let F be a field. Every ideal of $F[x]$ is a principal ideal.
 - j. Let F be a field. Every principal ideal of $F[x]$ is a maximal ideal.
15. Find a maximal ideal of $\mathbb{Z} \times \mathbb{Z}$.

16. Find a prime ideal of $\mathbb{Z} \times \mathbb{Z}$ that is not maximal.
 17. Find a nontrivial proper ideal of $\mathbb{Z} \times \mathbb{Z}$ that is not prime.
 18. Is $\mathbb{Q}[x]/\langle x^2 - 5x + 6 \rangle$ a field? Why?
 19. Is $\mathbb{Q}[x]/\langle x^2 - 6x + 6 \rangle$ a field? Why?

Proof Synopsis

20. Give a one- or two-sentence synopsis of “only if” part of Theorem 31.9.
 21. Give a one- or two-sentence synopsis of “if” part of Theorem 31.9.
 22. Give a one- or two-sentence synopsis of Theorem 31.24.
 23. Give a one- or two-sentence synopsis of the “only if” part of Theorem 31.25.

Theory

24. Give an example of an ideal in $\mathbb{Q}[x, y]$ that is not a principal ideal. Conclude that if R is an integral domain with the property that every ideal in R is principal, it does not follow that every ideal in $R[x]$ is a principal ideal.
 25. Prove that if R is a commutative ring with unity and $a \in R$, then $\langle a \rangle = \{ra \mid r \in R\}$ is an ideal in R .
 26. Let R be a finite commutative ring with unity. Show that every prime ideal in R is a maximal ideal.
 27. Corollary 31.18 tells us that every ring with unity contains a subring isomorphic to either \mathbb{Z} or some \mathbb{Z}_n . Is it possible that a ring with unity may simultaneously contain two subrings isomorphic to \mathbb{Z}_n and \mathbb{Z}_m for $n \neq m$? If it is possible, give an example. If it is impossible, prove it.
 28. Continuing Exercise 27, is it possible that a ring with unity may simultaneously contain two subrings isomorphic to the fields \mathbb{Z}_p and \mathbb{Z}_q for two different primes p and q ? Give an example or prove it is impossible.
 29. Following the idea of Exercise 28, is it possible for an integral domain to contain two subrings isomorphic to \mathbb{Z}_p and \mathbb{Z}_q for $p \neq q$ and p and q both prime? Give reasons or an illustration.
 30. Prove directly from the definitions of maximal and prime ideals that every maximal ideal of a commutative ring R with unity is a prime ideal. [Hint: Suppose M is maximal in R , $ab \in M$, and $a \notin M$. Argue that the smallest ideal $\{ra + m \mid r \in R, m \in M\}$ containing a and M must contain 1. Express 1 as $ra + m$ and multiply by b .]
 31. Show that N is a maximal ideal in a ring R if and only if R/N is a simple ring, that is, it is nontrivial and has no proper nontrivial ideals. (Compare with Theorem 13.20.)
 32. Prove that if F is a field, every proper nontrivial prime ideal of $F[x]$ is maximal.
 33. Let F be a field and $f(x), g(x) \in F[x]$. Show that $f(x)$ divides $g(x)$ if and only if $g(x) \in \langle f(x) \rangle$.
 34. Let F be a field and let $f(x), g(x) \in F[x]$. Show that

$$N = \{r(x)f(x) + s(x)g(x) \mid r(x), s(x) \in F[x]\}$$

is an ideal of $F[x]$. Show that if $f(x)$ and $g(x)$ have different degrees and $N \neq F[x]$, then $f(x)$ and $g(x)$ cannot both be irreducible over F .

35. Use Theorem 31.24 to prove the *equivalence* of these two theorems:

Fundamental Theorem of Algebra: Every nonconstant polynomial in $\mathbb{C}[x]$ has a zero in \mathbb{C} .

Nullstellensatz for $\mathbb{C}[x]$: Let $f_1(x), \dots, f_r(x) \in \mathbb{C}[x]$ and suppose that every $\alpha \in \mathbb{C}$ that is a zero of all r of these polynomials is also a zero of a polynomial $g(x)$ in $\mathbb{C}[x]$. Then some power of $g(x)$ is in the smallest ideal of $\mathbb{C}[x]$ that contains the r polynomials $f_1(x), \dots, f_r(x)$.

There is a sort of arithmetic of ideals in a ring. The next three exercises define sum, product, and quotient of ideals.

36. If A and B are ideals of a ring R , the **sum** $A + B$ of A and B is defined by

$$A + B = \{a + b \mid a \in A, b \in B\}.$$

- a. Show that $A + B$ is an ideal.
 b. Show that $A \subseteq A + B$ and $B \subseteq A + B$.

37. Let A and B be ideals of a ring R . The **product AB of A and B** is defined by

$$AB = \left\{ \sum_{i=1}^n a_i b_i \mid a_i \in A, b_i \in B, n \in \mathbb{Z}^+ \right\}.$$

- a. Show that AB is an ideal in R .
 b. Show that $AB \subseteq (A \cap B)$.
38. Let A and B be ideals of a *commutative* ring R . The **quotient $A : B$ of A by B** is defined by

$$A : B = \{r \in R \mid rb \in A \text{ for all } b \in B\}.$$

- Show that $A : B$ is an ideal of R .
 39. Show that for a field F , the set S of all matrices of the form
- $$\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$$
- for $a, b \in F$ is a **right ideal** but not a **left ideal** of $M_2(F)$. That is, show that S is a subring closed under multiplication on the *right* by any element of $M_2(F)$, but is not closed under *left* multiplication.
 40. Show that the matrix ring $M_2(\mathbb{Z}_2)$ is a simple ring; that is, $M_2(\mathbb{Z}_2)$ has no proper nontrivial ideals.

SECTION 32

[†]NONCOMMUTATIVE EXAMPLES

Thus far, the only example we have presented of a ring that is not commutative is the ring $M_n(F)$ of all $n \times n$ matrices with entries in a field F . We shall do almost nothing with noncommutative rings and strictly skew fields. To show that there are other important noncommutative rings occurring very naturally in algebra, we give several examples of such rings.

Rings of Endomorphisms

Let A be any abelian group. A homomorphism of A into itself is an **endomorphism of A** . Let the set of all endomorphisms of A be $\text{End}(A)$. Since the composition of two homomorphisms of A into itself is again such a homomorphism, we define multiplication on $\text{End}(A)$ by function composition, and thus multiplication is associative.

To define addition, for $\phi, \psi \in \text{End}(A)$, we have to describe the value of $(\phi + \psi)$ on each $a \in A$. Define

$$(\phi + \psi)(a) = \phi(a) + \psi(a).$$

Since

$$\begin{aligned} (\phi + \psi)(a + b) &= \phi(a + b) + \psi(a + b) \\ &= [\phi(a) + \phi(b)] + [\psi(a) + \psi(b)] \\ &= [\phi(a) + \psi(a)] + [\phi(b) + \psi(b)] \\ &= (\phi + \psi)(a) + (\phi + \psi)(b) \end{aligned}$$

we see that $\phi + \psi$ is again in $\text{End}(A)$.

Since A is commutative, we have

$$(\phi + \psi)(a) = \phi(a) + \psi(a) = \psi(a) + \phi(a) = (\psi + \phi)(a)$$

[†]This section is not used in the remainder of the text.

for all $a \in A$, so $\phi + \psi = \psi + \phi$ and addition in $\text{End}(A)$ is commutative. The associativity of addition follows from

$$\begin{aligned} [\phi + (\psi + \theta)](a) &= \phi(a) + [(\psi + \theta)(a)] \\ &= \phi(a) + [\psi(a) + \theta(a)] \\ &= [\phi(a) + \psi(a)] + \theta(a) \\ &= (\phi + \psi)(a) + \theta(a) \\ &= [(\phi + \psi) + \theta](a). \end{aligned}$$

If e is the additive identity of A , then the homomorphism 0 defined by

$$0(a) = e$$

for $a \in A$ is an additive identity in $\text{End}(A)$. Finally, for

$$\phi \in \text{End}(A),$$

$-\phi$ defined by

$$(-\phi)(a) = -\phi(a)$$

is in $\text{End}(A)$, since

$$\begin{aligned} (-\phi)(a + b) &= -\phi(a + b) = -[\phi(a) + \phi(b)] \\ &= -\phi(a) - \phi(b) = (-\phi)(a) + (-\phi)(b), \end{aligned}$$

and $\phi + (-\phi) = 0$. Thus $\langle \text{End}(A), + \rangle$ is an abelian group.

Note that we have not yet used the fact that our functions are *homomorphisms* except to show that $\phi + \psi$ and $-\phi$ are again *homomorphisms*. Thus the set A^A of *all functions* from A into A is an abelian group under exactly the same definition of addition, and, of course, function composition again gives a nice associative multiplication in A^A . However, we do need the fact that these functions in $\text{End}(A)$ are homomorphisms now to prove the left distributive law in $\text{End}(A)$. Except for this left distributive law, $\langle A^A, +, \cdot \rangle$ satisfies all the axioms for a ring. Let ϕ, ψ , and θ be in $\text{End}(A)$, and let $a \in A$. Then

$$(\theta(\phi + \psi))(a) = \theta((\phi + \psi)(a)) = \theta(\phi(a) + \psi(a)).$$

Since θ is a *homomorphism*,

$$\begin{aligned} \theta(\phi(a) + \psi(a)) &= \theta(\phi(a)) + \theta(\psi(a)) \\ &= (\theta\phi)(a) + (\theta\psi)(a) \\ &= (\theta\phi + \theta\psi)(a). \end{aligned}$$

Thus $\theta(\phi + \psi) = \theta\phi + \theta\psi$. The right distributive law causes no trouble, even in A^A , and follows from

$$\begin{aligned} ((\psi + \theta)\phi)(a) &= (\psi + \theta)(\phi(a)) = \psi(\phi(a)) + \theta(\phi(a)) \\ &= (\psi\phi)(a) + (\theta\phi)(a) = (\psi\phi + \theta\phi)(a). \end{aligned}$$

Thus we have proved the following theorem.

32.1 Theorem The set $\text{End}(A)$ of all endomorphisms of an abelian group A forms a ring under homomorphism addition and homomorphism multiplication (function composition).

Again, to show relevance to this section, we should give an example showing that $\text{End}(A)$ need not be commutative. Since function composition is in general not commutative, this seems reasonable to expect. However, $\text{End}(A)$ may be commutative in some cases. Indeed, Exercise 15 asks us to show that $\text{End}(\mathbb{Z}, +)$ is commutative.

32.2 Example Consider the abelian group $\langle \mathbb{Z} \times \mathbb{Z}, + \rangle$ discussed in Section 9. It is straightforward to verify that two elements of $\text{End}(\langle \mathbb{Z} \times \mathbb{Z}, + \rangle)$ are ϕ and ψ defined by

$$\phi((m, n)) = (m + n, 0) \quad \text{and} \quad \psi((m, n)) = (0, n).$$

Note that ϕ maps everything onto the first factor of $\mathbb{Z} \times \mathbb{Z}$, and ψ collapses the first factor. Thus

$$(\psi\phi)(m, n) = \psi(m + n, 0) = (0, 0)$$

while

$$(\phi\psi)(m, n) = \phi(0, n) = (n, 0).$$

Hence $\phi\psi \neq \psi\phi$. ▲

32.3 Example Let F be a field of characteristic zero, and let $\langle F[x], + \rangle$ be the additive group of the ring $F[x]$ of polynomials with coefficients in F . For this example, let us denote this additive group by $F[x]$, to simplify this notation. We can consider $\text{End}(F[x])$. One element of $\text{End}(F[x])$ acts on each polynomial in $F[x]$ by multiplying it by x . Let this endomorphism be X , so

$$X(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_0x + a_1x^2 + a_2x^3 + \cdots + a_nx^{n+1}.$$

Another element of $\text{End}(F[x])$ is formal differentiation with respect to x . (The familiar formula “the derivation of a sum is the sum of the derivatives” guarantees that differentiation is an endomorphism of $F[x]$.) Let Y be this endomorphism, so

$$Y(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1}.$$

Exercise 17 asks us to show that $YX - XY = 1$, where 1 is unity (the identity map) in $\text{End}(F[x])$. Thus $XY \neq YX$. Multiplication of polynomials in $F[x]$ by any element of F also gives an element of $\text{End}(F[x])$. The subring of $\text{End}(F[x])$ generated by X and Y and multiplications by elements of F is the **Weyl algebra** and is important in quantum mechanics. ▲

Group Rings and Group Algebras

Let $G = \{g_i \mid i \in I\}$ be any group written multiplicatively and let R be any commutative ring with nonzero unity. Let RG be the set of all *formal sums*.

$$\sum_{i \in I} a_i g_i$$

for $a_i \in R$ and $g_i \in G$, where all but a finite number of the a_i are 0. Define the sum of two elements of RG by

$$\left(\sum_{i \in I} a_i g_i \right) + \left(\sum_{i \in I} b_i g_i \right) = \sum_{i \in I} (a_i + b_i) g_i.$$

Observe that $(a_i + b_i) = 0$ except for a finite number of indices i , so $\sum_{i \in I} (a_i + b_i) g_i$ is again in RG . It is immediate that $\langle RG, + \rangle$ is an abelian group with additive identity $\sum_{i \in I} 0 g_i$.

Multiplication of two elements of RG is defined by the use of the multiplications in G and R as follows:

$$\left(\sum_{i \in I} a_i g_i \right) \left(\sum_{j \in I} b_j g_j \right) = \sum_{i \in I} \left(\sum_{g_j g_k = g_i} a_j b_k \right) g_i.$$

Naively, we formally distribute the sum $\sum_{i \in I} a_i g_i$ over the sum $\sum_{j \in I} b_j g_j$ and rename a term $a_j b_k g_k$ by $a_j b_k g_i$ where $g_j g_k = g_i$ in G . Since a_i and b_i are 0 for all but a finite

number of i , the sum $\sum_{g_j g_k = g_i} a_j b_k$ contains only a finite number of nonzero summands $a_j b_k \in R$ and may thus be viewed as an element of R . Again, at most a finite number of such sums $\sum_{g_j g_k = g_i} a_j b_k$ are nonzero. Thus multiplication is closed on RG .

The distributive laws follow at once from the definition of addition and the formal way we used distributivity to define multiplication. For the associativity of multiplication

$$\begin{aligned} \left(\sum_{i \in I} a_i g_i \right) \left[\left(\sum_{i \in I} b_i g_i \right) \left(\sum_{i \in I} c_i g_i \right) \right] &= \left(\sum_{i \in I} a_i g_i \right) \left[\sum_{i \in I} \left(\sum_{g_j g_k = g_i} b_j c_k \right) g_i \right] \\ &= \sum_{i \in I} \left(\sum_{g_h g_j g_k = g_i} a_h b_j c_k \right) g_i \\ &= \left[\sum_{i \in I} \left(\sum_{g_h g_j = g_i} a_h b_j \right) g_i \right] \left(\sum_{i \in I} c_i g_i \right) \\ &= \left[\left(\sum_{i \in I} a_i g_i \right) \left(\sum_{i \in I} b_i g_i \right) \right] \left(\sum_{i \in I} c_i g_i \right). \end{aligned}$$

Thus we have proved the following theorem.

32.4 Theorem If G is any group written multiplicatively and R is a commutative ring with nonzero unity, then $\langle RG, +, \cdot \rangle$ is a ring.

Corresponding to each $g \in G$, we have an element 1_g in RG . If we identify (rename) 1_g with g , we see that $\langle RG, \cdot \rangle$ can be considered to contain G naturally as a multiplicative subsystem. Thus, if G is not abelian, RG is not a commutative ring.

32.5 Definition The ring RG defined above is the **group ring of G over R** . If F is a field, then FG is the **group algebra of G over F** . ■

32.6 Example Let us give the addition and multiplication tables for the group algebra \mathbb{Z}_2G , where $G = \{e, a\}$ is cyclic of order 2. The elements of Z_2G are

$$0e + 0a, \quad 0e + 1a, \quad 1e + 0a, \quad \text{and} \quad 1e + 1a.$$

If we denote these elements in the obvious, natural way by

$$0, \quad a, \quad e, \quad \text{and} \quad e + a,$$

32.7 Table

+	0	a	e	$e + a$
0	0	a	e	$e + a$
a	a	0	$e + a$	e
e	e	$e + a$	0	a
$e + a$	$e + a$	e	a	0

32.8 Table

	0	a	e	$e + a$
0	0	0	0	0
a	0	e	a	$e + a$
e	0	a	e	$e + a$
$e + a$	0	$e + a$	$e + a$	0

respectively, we get Tables 32.7 and 32.8. For example, to see that $(e + a)(e + a) = 0$, we have

$$(1e + 1a)(1e + 1a) = (1 + 1)e + (1 + 1)a = 0e + 0a.$$

This example shows that a group algebra may have 0 divisors. Indeed, this is usually the case. ▲

The Quaternions

We have not yet given an example of a noncommutative division ring. The *quaternions* of Hamilton are the standard example of a strictly skew field; let us describe them.

■ HISTORICAL NOTE

Sir William Rowan Hamilton (1805–1865) discovered quaternions in 1843 while he was searching for a way to multiply number triplets (vectors in \mathbb{R}^3). Six years earlier he had developed the complex numbers abstractly as pairs (a, b) of real numbers with addition $(a, b) + (a', b') = (a + a', b + b')$ and multiplication $(a, b)(a', b') = (aa' - bb', ab' + a'b)$; he was then looking for an analogous multiplication for 3-vectors that was distributive and such that the length of the product vector was the product of the lengths of the factors. After many unsuccessful attempts to multiply vectors of the form $a + bi + cj$ (where $1, i, j$ are mutually perpendicular), he realized while walking

along the Royal Canal in Dublin on October 16, 1843, that he needed a new “imaginary symbol” k to be perpendicular to the other three elements. He could not “resist the impulse … to cut with a knife on a stone of Brougham Bridge” the fundamental defining formulas for multiplying these quaternions.

The quaternions were the first known example of a strictly skew field. Though many others were subsequently discovered, it was eventually noted that none were finite. In 1909 Joseph Henry Maclagan Wedderburn (1882–1948), then a preceptor at Princeton University, gave the first proof of Theorem 32.10.

Let the set \mathbb{H} , for Hamilton, be $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Now $(\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}, +)$ is a group under addition by components, the direct product of \mathbb{R} under addition with itself four times. This gives the operation of addition on \mathbb{H} . Let us rename certain elements of \mathbb{H} . We shall let

$$\begin{aligned} 1 &= (1, 0, 0, 0), & i &= (0, 1, 0, 0), \\ j &= (0, 0, 1, 0), & \text{and } k &= (0, 0, 0, 1). \end{aligned}$$

We furthermore agree to let

$$\begin{aligned} a_1 &= (a_1, 0, 0, 0), & a_2i &= (0, a_2, 0, 0), \\ a_3j &= (0, 0, a_3, 0) & \text{and } a_4k &= (0, 0, 0, a_4). \end{aligned}$$

In view of our definition of addition, we then have

$$(a_1, a_2, a_3, a_4) = a_1 + a_2i + a_3j + a_4k.$$

Thus

$$\begin{aligned} (a_1 + a_2i + a_3j + a_4k) + (b_1 + b_2i + b_3j + b_4k) \\ = (a_1 + b_1) + (a_2 + b_2)i + (a_3 + b_3)j + (a_4 + b_4)k. \end{aligned}$$

We now give Hamilton’s fundamental formulas for multiplication in \mathbb{H} . We start by defining

$$\begin{aligned} 1a &= a1 = a & \text{for } a \in \mathbb{H}, \\ i^2 &= j^2 = k^2 = -1, \end{aligned}$$

and

$$ij = k, \quad jk = i, \quad ki = j, \quad ji = -k, \quad kj = -i, \quad \text{and } ik = -j.$$

Note the similarity with the so-called cross product of vectors. These formulas are easy to remember if we think of the sequence

$$i, j, k, i, j, k.$$

The product from left to right of two adjacent elements is the next one to the right. The product from right to left of two adjacent elements is the negative of the next one to the left. We then define a product to be what it must be to make the distributive laws hold, namely,

$$\begin{aligned} & (a_1 + a_2i + a_3j + a_4k)(b_1 + b_2i + b_3j + b_4k) \\ &= (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4) + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)i \\ &\quad + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)j \\ &\quad + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)k. \end{aligned}$$

Exercise 19 shows that the quaternions are isomorphic to a subring of $M_2(\mathbb{C})$, so multiplication is associative. Since $ij = k$ and $ji = -k$, we see that multiplication is not commutative, so \mathbb{H} is definitely not a field. Turning to the existence of multiplicative inverses, let $a = a_1 + a_2i + a_3j + a_4k$, with not all $a_i = 0$. Computation shows that

$$(a_1 + a_2i + a_3j + a_4k)(a_1 - a_2i - a_3j - a_4k) = a_1^2 + a_2^2 + a_3^2 + a_4^2.$$

If we let

$$|a|^2 = a_1^2 + a_2^2 + a_3^2 + a_4^2 \quad \text{and} \quad \bar{a} = a_1 - a_2i - a_3j - a_4k,$$

we see that

$$\frac{\bar{a}}{|a|^2} = \frac{a_1}{|a|^2} - \left(\frac{a_2}{|a|^2} \right)i - \left(\frac{a_3}{|a|^2} \right)j - \left(\frac{a_4}{|a|^2} \right)k$$

is a multiplicative inverse for a . We have demonstrated the following theorem.

32.9 Theorem The quaternions \mathbb{H} form a strictly skew field under addition and multiplication. ◆

Note that $G = \{\pm 1, \pm i, \pm j, \pm k\}$ is a group of order 8 under quaternion multiplication. This group is generated by i and j , where

$$i^4 = 1, \quad j^2 = i^2 \quad \text{and} \quad ji = i^3j.$$

The group G is not cyclic. Recall that Corollary 28.7 states that if F is a field and H is a finite subgroup of the multiplicative group F^* , then H is cyclic. This example shows that Corollary 28.7 cannot be generalized to skew fields.

There are no finite strictly skew fields. This is the content of a famous theorem of Wedderburn, which we state without proof.

32.10 Theorem (Wedderburn's Theorem) Every finite division ring is a field.

Proof See Artin, Nesbitt, and Thrall [24] for a proof of Wedderburn's theorem. ◆

■ EXERCISES 32

Computations

In Exercises 1 through 3, let $G = \{e, a, b\}$ be a cyclic group of order 3 with identity element e . Write the element in the group algebra \mathbb{Z}_5G in the form

$$re + sa + tb \quad \text{for} \quad r, s, t \in \mathbb{Z}_5.$$

1. $(2e + 3a + 0b) + (4e + 2a + 3b)$ 2. $(2e + 3a + 0b)(4e + 2a + 3b)$ 3. $(3e + 3a + 3b)^4$

In Exercises 4 through 7, write the element of \mathbb{H} in the form $a_1 + a_2i + a_3j + a_4k$ for $a_i \in \mathbb{R}$.

4. $(i + 3j)(4 + 2j - k)$ 5. $i^2 j^3 k j i^5$
 6. $(i + j)^{-1}$ 7. $[(1 + 3i)(4j + 3k)]^{-1}$
 8. Referring to the dihedral group $D_3 = \{\iota, \rho, \rho^2, \mu, \mu\rho, \mu\rho^2\}$ as defined in Section 4, compute the product

$$(0\iota + 1\rho + 0(\rho^2) + 0\mu + 1(\mu\rho) + 1(\mu\rho^2))(1\iota + 1\rho + 0(\rho^2) + 1\mu + 0(\mu\rho) + 1(\mu\rho^2))$$

in the group ring $\mathbb{Z}_2 D_3$.

9. Find the center of the group $\langle \mathbb{H}^*, \cdot \rangle$, where \mathbb{H}^* is the set of nonzero quaternions.

Concepts

10. Find two subsets of \mathbb{H} different from \mathbb{C} and from each other, each of which is a field isomorphic to \mathbb{C} under the induced addition and multiplication from \mathbb{H} .
11. Determine whether each of the following is true or false.
- $M_n(F)$ has no divisors of 0 for any n and any field F .
 - Every nonzero element of $M_2(\mathbb{Z}_2)$ is a unit.
 - $\text{End}(A)$ is always a ring with unity $\neq 0$ for every abelian group A .
 - $\text{End}(A)$ is never a ring with unity $\neq 0$ for any abelian group A .
 - The subset $\text{Iso}(A)$ of $\text{End}(A)$, consisting of the isomorphisms of A onto A , forms a subring of $\text{End}(A)$ for every abelian group A .
 - $R(\mathbb{Z}, +)$ is isomorphic to $\langle \mathbb{Z}, +, \cdot \rangle$ for every commutative ring R with unity.
 - The group ring RG of an abelian group G is a commutative ring for any commutative ring R with unity.
 - The quaternions are a field.
 - $\langle \mathbb{H}^*, \cdot \rangle$ is a group where \mathbb{H}^* is the set of nonzero quaternions.
 - No subring of \mathbb{H} is a field.
12. Show each of the following by giving an example.
- A polynomial of degree n with coefficients in a strictly skew field may have more than n zeros in the skew field.
 - A finite multiplicative subgroup of a strictly skew field need not be cyclic.

Theory

13. Let ϕ be the element of $\text{End}(\langle \mathbb{Z} \times \mathbb{Z}, + \rangle)$ given in Example 32.2. That example showed that ϕ is a right divisor of 0. Show that ϕ is also a left divisor of 0.
14. Show that $M_2(F)$ has at least six units for every field F . Exhibit these units. [Hint: F has at least two elements, 0 and 1.]
15. Show that $\text{End}(\langle \mathbb{Z}, + \rangle)$ is naturally isomorphic to $\langle \mathbb{Z}, +, \cdot \rangle$ and that $\text{End}(\langle \mathbb{Z}_n, + \rangle)$ is naturally isomorphic to $\langle \mathbb{Z}_n, +, \cdot \rangle$.
16. Show that $\text{End}(\langle \mathbb{Z}_2 \times \mathbb{Z}_2, + \rangle)$ is not isomorphic to $\langle \mathbb{Z}_2 \times \mathbb{Z}_2, +, \cdot \rangle$.
17. Referring to Example 32.3, show that $YX - XY = 1$.
18. If $G = \{e\}$, the group of one element, show that RG is isomorphic to R for any ring R .
19. There exists a matrix $K \in M_2(\mathbb{C})$ such that $\phi : \mathbb{H} \rightarrow M_2(\mathbb{C})$ defined by

$$\phi(a + bi + cj + dk) = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} + dK,$$

for all $a, b, c, d \in \mathbb{R}$, gives an isomorphism of \mathbb{H} with $\phi[\mathbb{H}]$.

- a. Find the matrix K .
 - b. What 8 equations should you check to see that ϕ really is a homomorphism?
 - c. What other thing should you check to show that ϕ gives an isomorphism of \mathbb{H} with $\phi[\mathbb{H}]$?
20. Let R be a ring with unity, let $a \in R$, and let $\lambda_a : R \rightarrow R$ be given by
- $$\lambda_a(x) = ax$$
- for $x \in R$.
- a. Show that λ_a is an endomorphism of $(R, +)$.
 - b. Show that $R' = \{\lambda_a \mid a \in R\}$ is a subring of $\text{End}(R, +)$.
 - c. Prove the analogue of Cayley's theorem for R by showing that R' of (b) is isomorphic to R .

This page is intentionally left blank

Commutative Algebra

-
- Section 33** Vector Spaces
 - Section 34** Unique Factorization Domains
 - Section 35** Euclidean Domains
 - Section 36** Number Theory
 - Section 37** Algebraic Geometry
 - Section 38** Gröbner Bases for Ideals

SECTION 33

VECTOR SPACES

The notions of a vector space, scalars, independent vectors, and bases may be familiar. In this section, we present these ideas where the scalars may be elements of any field. We use Greek letters like α and β for vectors. In our application, the vectors will be elements of a field E containing field F . The proofs are all identical with those often given in a first course in linear algebra.

Definition and Elementary Properties

The topic of vector spaces is the cornerstone of linear algebra. Since linear algebra is not the subject for study in this text, our treatment of vector spaces will be brief, designed to develop only the concepts of linear independence and dimension that we need for our development of field theory.

The terms *vector* and *scalar* are probably familiar from calculus. Here we allow scalars to be elements of any field, not just the real numbers, and develop the theory by axioms just as for the other algebraic structures we have studied.

33.1 Definition Let F be a field. A **vector space over F** (or F -vector space) consists of an abelian group V under addition together with an operation of scalar multiplication of each element of V by each element of F on the left, such that for all $a, b \in F$ and $\alpha, \beta \in V$ the following conditions are satisfied:

- \mathcal{V}_1 . $a\alpha \in V$.
- \mathcal{V}_2 . $a(b\alpha) = (ab)\alpha$.
- \mathcal{V}_3 . $(a + b)\alpha = (a\alpha) + (b\alpha)$.
- \mathcal{V}_4 . $a(\alpha + \beta) = (a\alpha) + (a\beta)$.
- \mathcal{V}_5 . $1\alpha = \alpha$.

The elements of V are **vectors** and the elements of F are **scalars**. When only one field F is under discussion, we drop the reference to F and refer to a *vector space*. ■

Note that scalar multiplication for a vector space is not a binary operation on one set in the sense we defined it in Section 1. It associates an element $a\alpha$ of V with each

ordered pair (a, α) , consisting of an element a of F and an element α of V . Thus scalar multiplication is a *function* mapping $F \times V$ into V . Both the additive identity for V , the 0-vector, and the additive identity for F , the 0-scalar, will be denoted by 0.

33.2 Example Consider the abelian group $(\mathbb{R}^n, +) = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ for n factors, which consists of ordered n -tuples under addition by components. Define scalar multiplication for scalars in \mathbb{R} by

$$r\alpha = (ra_1, \dots, ra_n)$$

for $r \in \mathbb{R}$ and $\alpha = (a_1, \dots, a_n) \in \mathbb{R}^n$. With these operations, \mathbb{R}^n becomes a vector space over \mathbb{R} . The axioms for a vector space are readily checked. In particular, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ as a vector space over \mathbb{R} can be viewed as all “vectors whose starting points are the origin of the Euclidean plane” in the sense often studied in calculus courses. ▲

33.3 Example For any field F , $F[x]$ can be viewed as a vector space over F , where addition of vectors is ordinary addition of polynomials in $F[x]$ and scalar multiplication $a\alpha$ of an element of $F[x]$ by an element of F is ordinary multiplication in $F[x]$. The axioms \mathcal{V}_1 through \mathcal{V}_6 for a vector space then follow immediately from the fact that $F[x]$ is a ring with unity. ▲

33.4 Example Let F be a subfield of the field E . Then E can be regarded as a vector space over F , where addition of vectors is the usual addition in E and scalar multiplication $a\alpha$ is the usual field multiplication in E with $a \in F$ and $\alpha \in E$. The axioms follow at once from the field axioms for E . Here our field of scalars is actually a subset of our space of vectors. *It is this example that is the important one for us.* ▲

We are assuming nothing about vector spaces from previous work and shall prove everything we need from the definition, even though the results may be familiar from calculus.

■ HISTORICAL NOTE

The ideas behind the abstract notion of a vector space occurred in many concrete examples during the nineteenth century and earlier. For example, William Rowan Hamilton dealt with complex numbers explicitly as pairs of real numbers and, as noted in Section 32, also dealt with triples and eventually quadruples of real numbers in his invention of the quaternions. In these cases, the “vectors” turned out to be objects that could both be added and multiplied by scalars, using “reasonable” rules for both of these operations. Other examples of such objects included differential forms (expressions under integral signs) and algebraic integers.

Although Hermann Grassmann (1809–1877) succeeded in working out a detailed theory of n dimensional spaces in his *Die Lineale Ausdehnungslehre* of 1844 and 1862, the first mathematician to give an abstract definition of a vector

space equivalent to Definition 33.1 was Giuseppe Peano (1858–1932) in his *Calcolo Geometrico* of 1888. Peano’s aim in the book, as the title indicates, was to develop a geometric calculus. According to Peano, such a calculus “consists of a system of operations analogous to those of algebraic calculus, but in which the objects with which the calculations are performed are, instead of numbers, geometrical objects.” Curiously, Peano’s work had no immediate effect on the mathematical scene. Although Hermann Weyl (1885–1955) essentially repeated Peano’s definition in his *Space-Time-Matter* of 1918, the definition of a vector space did not enter the mathematical mainstream until it was announced for a third time by Stefan Banach (1892–1945) in the 1922 publication of his dissertation dealing with what we now call *Banach spaces*, complete normed vector spaces.

33.5 Theorem If V is a vector space over F , then $0\alpha = 0$, $a0 = 0$, and $(-a)\alpha = a(-\alpha) = -(a\alpha)$ for all $a \in F$ and $\alpha \in V$.

Proof The equation $0\alpha = 0$ is to be read “(0-scalar) α = 0-vector.” Likewise, $a0 = 0$ is to be read “ a (0-vector) = 0-vector.” The proofs here are very similar to those in Theorem 22.8 for a ring and again depend heavily on the distributive laws \mathcal{D}_3 and \mathcal{D}_4 . Now

$$(0\alpha) = (0 + 0)\alpha = (0\alpha) + (0\alpha)$$

is an equation in the abelian group $(V, +)$, so by the group cancellation law, $0 = 0\alpha$. Likewise, from

$$a0 = a(0 + 0) = a0 + a0,$$

we conclude that $a0 = 0$. Then

$$0 = 0\alpha = (a + (-a))\alpha = a\alpha + (-a)\alpha,$$

so $(-a)\alpha = -(a\alpha)$. Likewise, from

$$0 = a0 = a(\alpha + (-\alpha)) = a\alpha + a(-\alpha),$$

we conclude that $a(-\alpha) = -(a\alpha)$ also. ◆

Linear Independence and Bases

33.6 Definition Let V be a vector space over F . The vectors in a subset $S = \{\alpha_i \mid i \in I\}$ of V **span** (or **generate**) V if for every $\beta \in V$, we have

$$\beta = a_1\alpha_{i_1} + a_2\alpha_{i_2} + \cdots + a_n\alpha_{i_n}$$

for some $a_j \in F$ and $\alpha_{i_j} \in S, j = 1, \dots, n$. A vector $\sum_{j=1}^n a_j\alpha_{i_j}$ is a **linear combination** of the α_{i_j} . ■

33.7 Example In the vector space \mathbb{R}^n over \mathbb{R} of Example 33.2, the vectors

$$(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)$$

clearly span \mathbb{R}^n , for

$$(a_1, a_2, \dots, a_n) = a_1(1, 0, \dots, 0) + a_2(0, 1, \dots, 0) + \cdots + a_n(0, 0, \dots, 1).$$

Also, the monomials x^m for $m \geq 0$ span $F[x]$ over F , the vector space of Example 33.3. ▲

33.8 Definition A vector space V over a field F is **finite dimensional** if there is a finite subset of V whose vectors span V . ■

33.9 Example Example 33.7 shows that \mathbb{R}^n is finite dimensional. The vector space $F[x]$ over F is *not* finite dimensional, since polynomials of arbitrarily large degree could not be linear combinations of elements of any *finite* set of polynomials. ▲

The next definition contains the most important idea in this section.

33.10 Definition The vectors in a subset $S = \{\alpha_i \mid i \in I\}$ of a vector space V over a field F are **linearly independent over F** if, for any distinct vectors $\alpha_{i_j} \in S$, coefficients $a_j \in F$, and $n \in \mathbb{Z}^+$, we have $\sum_{j=1}^n a_j\alpha_{i_j} = 0$ in V only if $a_j = 0$ for $j = 1, \dots, n$. If the vectors are not linearly independent over F , they are **linearly dependent over F** . ■

Thus the vectors in $\{\alpha_i \mid i \in I\}$ are linearly independent over F if the only way the 0-vector can be expressed as a linear combination of the vectors α_i is to have all scalar coefficients equal to 0. If the vectors are linearly dependent over F , then there exist $a_j \in F$ for $j = 1, \dots, n$ such that $\sum_{j=1}^n a_j \alpha_j = 0$, where not all $a_j = 0$.

- 33.11 Example** Observe that the vectors spanning the space \mathbb{R}^n that are given in Example 33.7 are linearly independent over \mathbb{R} . Likewise, the vectors in $\{x^m \mid m \geq 0\}$ are linearly independent vectors of $F[x]$ over F . Note that $(1, -1)$, $(2, 1)$, and $(-3, 2)$ are linearly dependent in \mathbb{R}^2 over \mathbb{R} , since

$$7(1, -1) + (2, 1) + 3(-3, 2) = (0, 0). \quad \blacktriangle$$

- 33.12 Definition** If V is a vector space over a field F , the vectors in a subset $B = \{\beta_i \mid i \in I\}$ of V form a **basis for V over F** if they span V and are linearly independent. ■

- 33.13 Example** As seen from Examples 33.7 and 33.11,

$$\{(1, 0, 0, \dots, 0), (0, 1, 0, 0, \dots, 0), \dots, (0, 0, \dots, 0, 1)\}$$

is a basis for \mathbb{R}^n and

$$\{1, x, x^2, \dots\}$$

is a basis for $F[x]$ where F is a field. ▲

- 33.14 Example** Let F be a field and $p(x) \in F[x]$ be a degree $n \geq 1$ irreducible polynomial over F . Theorems 31.9 and 31.25 imply that the factor ring

$$E = F[x]/\langle p(x) \rangle$$

is a field. We can think of F as a subfield of E by identifying $a \in F$ with $a + \langle p(x) \rangle \in E$. Example 33.4 shows that E is a vector space over F . The vectors

$$\alpha_j = x^j + \langle p(x) \rangle \quad \text{for } 0 \leq j \leq n-1$$

are linearly independent since if

$$a_0\alpha_0 + a_1\alpha_1 + a_2\alpha_2 + \dots + a_{n-1}\alpha_{n-1} = 0 \in F[X]/\langle p(x) \rangle,$$

then

$$a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} \in \langle p(x) \rangle.$$

Every polynomial in $\langle p(x) \rangle$ except the zero polynomial has degree at least n . Thus each coefficient a_j is zero and the vectors $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ form an independent set. On the other hand, given any polynomial $f(x) \in F[x]$, the division algorithm implies that $f(x) + \langle p(x) \rangle = g(x) + \langle p(x) \rangle$ for some polynomial $g(x)$ where either $g(x) = 0$ or the degree of $g(x)$ is less than n . It follows that the vectors

$$\alpha_0, \alpha_1, \dots, \alpha_{n-1}$$

span E and therefore form a basis for E .

Looking back at this example, it is not necessary for $p(x)$ to be an irreducible polynomial. The field F can be thought of as a subring of the commutative ring with unity $E = F[x]/\langle p(x) \rangle$ and all the axioms of a vector space follow from the properties of a ring and the fact that the unity in F and the unity in E are the same. ▲

Dimension

The only other results we wish to prove about vector spaces are that every finite-dimensional vector space has a basis, and that any two bases of a finite-dimensional

vector space have the same number of elements. Both these facts are true without the assumption that the vector space is finite dimensional, but the proofs require more knowledge of set theory than we are assuming, and the finite-dimensional case is all we need. First we give an easy lemma.

33.15 Lemma Let V be a vector space over a field F , and let $\alpha \in V$. If α is a linear combination of vectors β_i in V for $i = 1, \dots, m$ and each β_i is a linear combination of vectors γ_j in V for $j = 1, \dots, n$, then α is a linear combination of the γ_j .

Proof Let $\alpha = \sum_{i=1}^m a_i \beta_i$, and let $\beta_i = \sum_{j=1}^n b_{ij} \gamma_j$, where a_i and b_{ij} are in F . Then

$$\alpha = \sum_{i=1}^m a_i \left(\sum_{j=1}^n b_{ij} \gamma_j \right) = \sum_{j=1}^n \left(\sum_{i=1}^m a_i b_{ij} \right) \gamma_j,$$

and $(\sum_{i=1}^m a_i b_{ij}) \in F$. \diamond

33.16 Theorem In a finite-dimensional vector space, every finite set of vectors spanning the space contains a subset that is a basis.

Proof Let V be finite dimensional over F , and let vectors $\alpha_1, \dots, \alpha_n$ in V span V . Let us list the α_i in a row. Examine each α_i in succession, starting at the left with $i = 1$, and discard the first α_j that is some linear combination of the preceding α_i for $i < j$. Then continue, starting with the following α_{j+1} , and discard the next α_k that is some linear combination of its remaining predecessors, and so on. When we reach α_n after a finite number of steps, those α_i remaining in our list are such that none is a linear combination of the preceding α_i in this reduced list. Lemma 33.15 shows that any vector that is a linear combination of the original collection of α_i is still a linear combination of our reduced, and possibly smaller, set in which no α_i is a linear combination of its predecessors. Thus the vectors in the reduced set of α_i again span V .

For the reduced set, suppose that

$$a_1 \alpha_{i_1} + \dots + a_r \alpha_{i_r} = 0$$

for $i_1 < i_2 < \dots < i_r$ and that some $a_j \neq 0$. We may assume from Theorem 33.5 that $a_r \neq 0$, or we could drop $a_r \alpha_{i_r}$ from the left side of the equation. Then, using Theorem 33.5 again, we obtain

$$\alpha_{i_r} = \left(-\frac{a_1}{a_r} \right) \alpha_{i_1} + \dots + \left(-\frac{a_{r-1}}{a_r} \right) \alpha_{i_{r-1}},$$

which shows that α_{i_r} is a linear combination of its predecessors, contradicting our construction. Thus the vectors α_i in the reduced set both span V and are linearly independent, so they form a basis for V over F . \diamond

33.17 Corollary A finite-dimensional vector space has a finite basis.

Proof By definition, a finite-dimensional vector space has a finite set of vectors that span the space. Theorem 33.16 completes the proof. \diamond

The next theorem is the culmination of our work on vector spaces.

33.18 Theorem Let $S = \{\alpha_1, \dots, \alpha_r\}$ be a finite set of linearly independent vectors of a finite-dimensional vector space V over a field F . Then S can be enlarged to a basis for V over F . Furthermore, if $B = \{\beta_1, \dots, \beta_n\}$ is any basis for V over F , then $r \leq n$.

Proof By Corollary 33.17, there is a basis $B = \{\beta_1, \dots, \beta_n\}$ for V over F . Consider the finite sequence of vectors

$$\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_n.$$

These vectors span V , since B is a basis. Following the technique, used in Theorem 33.16, of discarding in turn each vector that is a linear combination of its remaining predecessors, working from left to right, we arrive at a basis for V . Observe that no α_i is cast out, since the α_i are linearly independent. Thus S can be enlarged to a basis for V over F .

For the second part of the conclusion, consider the sequence

$$\alpha_1, \beta_1, \dots, \beta_n.$$

These vectors are not linearly independent over F , because α_1 is a linear combination

$$\alpha_1 = b_1\beta_1 + \dots + b_n\beta_n,$$

since the β_i form a basis. Thus

$$\alpha_1 + (-b_1)\beta_1 + \dots + (-b_n)\beta_n = 0.$$

The vectors in the sequence do span V , and if we form a basis by the technique of working from left to right and casting out in turn each vector that is a linear combination of its remaining predecessors, at least one β_i must be cast out, giving a basis

$$\{\alpha_1, \beta_1^{(1)}, \dots, \beta_m^{(1)}\},$$

where $m \leq n - 1$. Applying the same technique to the sequence of vectors

$$\alpha_1, \alpha_2, \beta_1^{(1)}, \dots, \beta_m^{(1)},$$

we arrive at a new basis

$$\{\alpha_1, \alpha_2, \beta_1^{(2)}, \dots, \beta_s^{(2)}\},$$

with $s \leq n - 2$. Continuing, we arrive finally at a basis

$$\{\alpha_1, \dots, \alpha_r, \beta_1^{(r)}, \dots, \beta_t^{(r)}\},$$

where $0 \leq t \leq n - r$. Thus $r \leq n$. ◆

33.19 Corollary Any two bases of a finite-dimensional vector space V over F have the same number of elements.

Proof Let $B = \{\beta_1, \dots, \beta_n\}$ and $B' = \{\beta'_1, \dots, \beta'_m\}$ be two bases. Then by Theorem 33.18, regarding B as an independent set of vectors and B' as a basis, we see that $n \leq m$. A symmetric argument gives $m \leq n$, so $m = n$. ◆

33.20 Definition If V is a finite-dimensional vector space over a field F , the number of elements in a basis (independent of the choice of basis, as just shown) is the **dimension of V over F** . ■

33.21 Example Let F be a field and $V \subseteq F[x]$ be the set of all polynomials of degree less than n including 0. The monomials $1, x, x^2, \dots, x^{n-1}$ span V and they are independent. Consequently, the dimension of V over F is n . From this we can conclude that any set of fewer than n polynomials in V does not span V and any set of more than n polynomials in V is not an independent set. Of course, an arbitrary set of n polynomials in V may or may not form a basis. ▲

Modules over a Ring

When studying abelian groups using additive notation we defined what it means to multiply an integer times an element in a group. For example, if g is an element of an abelian group, then $2g = g + g$. The table at the beginning of Section 4 looks similar to the definition of a vector space. The difference is that instead of a field, in the case of abelian groups we used the ring of integers.

33.22 Definition Let R be a ring with unity. A **left R -module** is an abelian group M under addition together with an operation of scalar multiplication of each element of M by each element

of R on the left, such that for all $a, b \in R$ and $\alpha, \beta \in M$ the following conditions are satisfied:

- $M_1 : a\alpha \in M$
- $M_2 : a(b\alpha) = (ab)\alpha$
- $M_3 : (a + b)\alpha = (a\alpha) + (b\alpha)$
- $M_4 : a(\alpha + \beta) = (a\alpha) + (a\beta)$
- $M_5 : 1\alpha = \alpha.$

■

A **right R -module** differs from a left R -module simply by multiplying module element by an element of R on the right with the obvious changes in the five conditions for a left module. Here we consider only left R -modules, so we will use the term R -module to mean left R -module.

33.23 Example For any abelian group G , G is a \mathbb{Z} -module using the usual notation for an integer times an element of G .

If R is a ring with unity and $I \subseteq R$ is an ideal, then I is an additive abelian group and for any $r \in R$ and $\alpha \in I$, $r\alpha \in I$. The defining properties of a ring with unity give the remaining properties of an R -module. Thus I is an R -module. ▲

33.24 Example Elements of \mathbb{R}^n (written as column vectors) can be multiplied on the left by elements in the ring $M_n(\mathbb{R})$ of $n \times n$ matrices with real number entries. The five properties defining an R -module are all satisfied, which implies \mathbb{R}^n is an $M_n(\mathbb{R})$ -module. ▲

The key properties of vector spaces are Corollary 33.17 and Corollary 33.19 (and their generalizations to vector spaces that are not finitely generated). They say any vector space has a basis, and any two bases of a given vector space have the same number of elements. The definitions of independent, spanning, and basis vectors are the same in R -module as in vector spaces. However, in general an R -module need not have a basis, and even if it does, in some cases two bases may have a different number of elements.

33.25 Example The abelian group \mathbb{Z}_3 is a \mathbb{Z} -module. There is no nonempty subset of \mathbb{Z}_3 that is independent since for any $\alpha \in \mathbb{Z}_3$, $3\alpha = 0$ and 3 is a nonzero integer. A similar argument shows that for any finite abelian group G , as a \mathbb{Z} -module G has no nonempty independent set. We conclude that as a \mathbb{Z} -module a finite abelian group does not have a basis. ▲

■ EXERCISES 33

Computations

- Find three bases for \mathbb{R}^2 over \mathbb{R} , no two of which have a vector in common.

In Exercises 2 and 3, determine whether the given set of vectors is a basis for \mathbb{R}^3 over \mathbb{R} .

- $\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$
- $\{(-1, 1, 2), (2, -3, 1), (10, -14, 0)\}$

Determine if the indicated vector space is finite dimensional over the field. If it is, find a basis.
In Exercises 4 through 9, give a basis for the indicated vector space over the field.

- \mathbb{Z}_{13} over \mathbb{Z}_{13}
- $\left\{ \begin{array}{l} \frac{a+b\sqrt{3}}{c+d\sqrt{3}} \\ \end{array} \mid a, b, c, d \in \mathbb{Q} \text{ and } c + d\sqrt{3} \neq 0 \right\}$ over \mathbb{Q}
- \mathbb{C} over \mathbb{R}
- $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ over \mathbb{Q}
- \mathbb{R} over $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$
- \mathbb{R} over \mathbb{Q}

10. There is a field E with 32 elements. Determine which prime field is isomorphic with a subfield of E and determine the dimension of E over its prime field.

Concepts

In Exercises 11 through 14, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

11. The vectors in a subset S of a vector space V over a field F *span* V if and only if each $\beta \in V$ can be expressed uniquely as a linear combination of the vectors in S .
12. The vectors in a subset S of a vector space V over a field F are *linearly independent over* F if and only if the zero vector cannot be expressed as a linear combination of vectors in S .
13. The *dimension over* F of a finite-dimensional vector space V over a field F is the minimum number of vectors required to span V .
14. A *basis* for a vector space V over a field F is a set of vectors in V that span V and are linearly dependent.
15. Determine whether each of the following is true or false.
 - a. The sum of two vectors is a vector.
 - b. The sum of two scalars is a vector.
 - c. The product of two scalars is a scalar.
 - d. The product of a scalar and a vector is a vector.
 - e. Every vector space has a finite basis.
 - f. The vectors in a basis are linearly dependent.
 - g. The 0-vector may be part of a basis.
 - h. A vector space over a field F is an F -module.
 - i. If R is a commutative ring with unity and M is an R module, then M has a basis over R .
 - j. Every vector space has a basis.

Exercises 16–27 deal with the further study of vector spaces. In many cases, we are asked to define for vector spaces some concept that is analogous to one we have studied for other algebraic structures. These exercises should improve our ability to recognize parallel and related situations in algebra. Any of these exercises may assume knowledge of concepts defined in the preceding exercises.

16. Let V be a vector space over a field F .
 - a. Define a *subspace of the vector space V over F* .
 - b. Prove that an intersection of subspaces of V is again a subspace of V over F .
17. Let V be a vector space over a field F , and let $S = \{\alpha_i \mid i \in I\}$ be a nonempty collection of vectors in V .
 - a. Using Exercise 16(b), define the *subspace of V generated by S* .
 - b. Prove that the vectors in the subspace of V generated by S are precisely the (finite) linear combinations of vectors in S . (Compare with Theorem 7.7.)
18. Let V_1, \dots, V_n be vector spaces over the same field F . Define the *direct sum* $V_1 \oplus \dots \oplus V_n$ of the vector spaces V_i for $i = 1, \dots, n$, and show that the direct sum is again a vector space over F .
19. Generalize Example 33.2 to obtain the vector space F^n of ordered n -tuples of elements of F over the field F , for any field F . What is a basis for F^n ?
20. Define an *isomorphism* of a vector space V over a field F with a vector space V' over the same field F .

Theory

21. Prove that if V is a finite-dimensional vector space over a field F , then a subset $\{\beta_1, \beta_2, \dots, \beta_n\}$ of V is a basis for V over F if and only if every vector in V can be expressed *uniquely* as a linear combination of the β_i .

22. Let F be any field. Consider the “system of m simultaneous linear equations in n unknowns”

$$\begin{aligned} a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n &= b_1, \\ a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n &= b_2, \\ &\vdots \\ a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n &= b_m, \end{aligned}$$

where $a_{ij}, b_i \in F$.

- a. Show that the “system has a solution,” that is, there exist $X_1, \dots, X_n \in F$ that satisfy all m equations, if and only if the vector $\beta = (b_1, \dots, b_m)$ of F^m is a linear combination of the vectors $\alpha_j = (a_{1j}, \dots, a_{mj})$. (This result is straightforward to prove, being practically the definition of a solution, but should really be regarded as the *fundamental existence theorem for a simultaneous solution of a system of linear equations*.)
 - b. From part (a), show that if $n = m$ and $\{\alpha_j \mid j = 1, \dots, n\}$ is a basis for F^n , then the system always has a unique solution.
23. Prove that every finite-dimensional vector space V of dimension n over a field F is isomorphic to the vector space F^n of Exercise 19.
24. Let V and V' be vector spaces over the same field F . A function $\phi : V \rightarrow V'$ is a **linear transformation of V into V'** if the following conditions are satisfied for all $\alpha, \beta \in V$, and $a \in F$:

$$\begin{aligned} \phi(\alpha + \beta) &= \phi(\alpha) + \phi(\beta). \\ \phi(a\alpha) &= a(\phi(\alpha)). \end{aligned}$$

- a. If $\{\beta_i \mid i \in I\}$ is a basis for V over F , show that a linear transformation $\phi : V \rightarrow V'$ is completely determined by the vectors $\phi(\beta_i) \in V'$.
 - b. Let $\{\beta_i \mid i \in I\}$ be a basis for V , and let $\{\beta'_i \mid i \in I\}$ be any set of vectors, not necessarily distinct, of V' . Show that there exists exactly one linear transformation $\phi : V \rightarrow V'$ such that $\phi(\beta_i) = \beta'_i$.
25. Let V and V' be vector spaces over the same field F , and let $\phi : V \rightarrow V'$ be a linear transformation.
- a. To what concept that we have studied for the algebraic structures of groups and rings does the concept of a *linear transformation* correspond?
 - b. Define the *kernel* (or *nullspace*) of ϕ , and show that it is a subspace of V .
 - c. Describe when ϕ is an isomorphism of V with V' .
26. Let V be a vector space over a field F , and let S be a subspace of V . Define the *quotient space* V/S , and show that it is a vector space over F .
27. Let V and V' be vector spaces over the same field F , and let V be finite dimensional over F . Let $\dim(V)$ be the dimension of the vector space V over F . Let $\phi : V \rightarrow V'$ be a linear transformation.
- a. Show that $\phi[V]$ is a subspace of V' .
 - b. Show that $\dim(\phi[V]) = \dim(V) - \dim(\text{Ker}(\phi))$. [Hint: Choose a convenient basis for V , using Theorem 33.18. For example, enlarge a basis for $\text{Ker}(\phi)$ to a basis for V .]
28. Let R be a commutative ring with unity and F a subring of R that is a field. Think of F as the scalars and R as the set of vectors with scalar multiplication given by multiplication in the ring R .
- a. Give an example to show that R need not be a vector space over F .
 - b. Show that if the unity of R and the unity of F are the same, then R is a vector space over F .

SECTION 34 UNIQUE FACTORIZATION DOMAINS

The integral domain \mathbb{Z} is our standard example of an integral domain in which there is unique factorization into primes (irreducibles). Section 28 showed that for a field F , $F[x]$ is also such an integral domain with unique factorization. In order to discuss analogous

ideas in an arbitrary integral domain, we shall give several definitions, some of which are repetitions of earlier ones. It is nice to have them all in one place for reference.

34.1 Definition Let R be a commutative ring with unity and let $a, b \in R$. If there exists $c \in R$ such that $b = ac$, then a divides b (or a is a factor of b), denoted by $a|b$. We read $a \nmid b$ as “ a does not divide b .” ■

34.2 Definition An element u of a commutative ring with unity R is a **unit of R** if u divides 1, that is, if u has a multiplicative inverse in R . Two elements $a, b \in R$ are **associates in R** if $a = bu$, where u is a unit in R .

Exercise 27 asks us to show that this criterion for a and b to be associates is an equivalence relation on R . ■

34.3 Example The only units in \mathbb{Z} are 1 and -1 . Thus the only associates of 26 in \mathbb{Z} are 26 and -26 . ▲

34.4 Definition A nonzero element p that is not a unit of an integral domain D is an **irreducible of D** if every factorization $p = ab$ in D has the property that either a or b is a unit. ■

Note that an associate of an irreducible p is again an irreducible, for if $p = uc$ for a unit u , then any factorization of c provides a factorization of p .

34.5 Definition An integral domain D is a **unique factorization domain** (abbreviated UFD) if the following conditions are satisfied:

1. Every element of D that is neither 0 nor a unit can be factored into a product of a finite number of irreducibles.
2. If $p_1 \cdots p_r$ and $q_1 \cdots q_s$ are two factorizations of the same element of D into irreducibles, then $r = s$ and the q_j can be renumbered so that p_i and q_i are associates. ■

34.6 Example Theorem 28.21 shows that for a field F , $F[x]$ is a UFD. Also we know that \mathbb{Z} is a UFD; we have made frequent use of this fact, although we have never proved it. For example, in \mathbb{Z} we have

$$24 = (2)(2)(3)(2) = (-2)(-3)(2)(2).$$

Here 2 and -2 are associates, as are 3 and -3 . Thus except for order and associates, the irreducible factors in these two factorizations of 24 are the same. ▲

Recall that the *principal ideal* $\langle a \rangle$ of D consists of all multiples of the element a . After just one more definition we can describe what we wish to achieve in this section.

34.7 Definition An integral domain D is a **principal ideal domain** (abbreviated PID) if every ideal in D is a principal ideal. ■

We know that \mathbb{Z} is a PID because every ideal is of the form $n\mathbb{Z}$, generated by some integer n . Theorem 31.24 shows that if F is a field, then $F[x]$ is a PID. Our purpose in this section is to prove two exceedingly important theorems:

1. Every PID is a UFD. (Theorem 34.18)
2. If D is a UFD, then $D[x]$ is a UFD. (Theorem 34.30)

■ HISTORICAL NOTE

The question of unique factorization in an integral domain other than the integers was first raised in public in connection with the attempted proof by Gabriel Lamé (1795–1870) of Fermat's Last Theorem, the conjecture that $x^n + y^n = z^n$ has no nontrivial integral solutions for $n > 2$. It is not hard to show that the conjecture is true if it can be proved for all odd primes p . At a meeting of the Paris Academy on March 1, 1847, Lamé announced that he had proved the theorem and presented a sketch of the proof. Lamé's idea was first to factor $x^p + y^p$ over the complex numbers as

$$\begin{aligned}x^p + y^p &= \\(x + y)(x + \alpha y)(x + \alpha^2 y) \cdots (x + \alpha^{p-1} y)\end{aligned}$$

where α is a primitive p th root of unity. He next proposed to show that if the factors in this expression are relatively prime and if $x^p + y^p = z^p$, then each of the p factors must be a p th power. He could then demonstrate that this Fermat equation would be true for a triple x', y', z' , each number smaller than the corresponding number in the original triple. This would lead to an infinite descending sequence of positive integers, an impossibility that would prove the theorem.

After Lamé finished his announcement, however, Joseph Liouville (1809–1882) cast serious doubts on the purported proof, noting that the conclusion that each of the relatively prime factors was a p th power because their product was a p th power depended on the result that any integer can be uniquely factored into a product of primes. It was by no means clear that “integers” of the

form $x + \alpha^k y$ had this unique factorization property. Although Lamé attempted to overcome Liouville's objections, the matter was settled on May 24, when Liouville produced a letter from Ernst Kummer noting that in 1844 he had already proved that unique factorization failed in the domain $\mathbb{Z}[\alpha]$, where α is a 23rd root of unity.

It was not until 1994 that Fermat's Last Theorem was proved, and by techniques of algebraic geometry unknown to Lamé and Kummer. In the late 1950s, Yutaka Taniyama and Goro Shimura noticed a curious relationship between two seemingly disparate fields of mathematics, elliptic curves and modular forms. A few years after Taniyama's tragic death at age 31, Shimura clarified this idea and eventually formulated what became known as the Taniyama–Shimura Conjecture. In 1984, Gerhard Frey asserted and in 1986 Ken Ribet proved that the Taniyama–Shimura Conjecture would imply the truth of Fermat's Last Theorem. But it was finally Andrew Wiles of Princeton University who, after secretly working on this problem for seven years, gave a series of lectures at Cambridge University in June 1993 in which he announced a proof of enough of the Taniyama–Shimura Conjecture to derive Fermat's Last Theorem. Unfortunately, a gap in the proof was soon discovered, and Wiles went back to work. It took him more than a year, but with the assistance of his student Richard Taylor, he finally was able to fill the gap. The result was published in the *Annals of Mathematics* in May 1995, and this 350-year-old problem was now solved.

The fact that $F[x]$ is a UFD, where F is a field (by Theorem 28.21), illustrates both theorems. For by Theorem 31.24, $F[x]$ is a PID. Also, since F has no nonzero elements that are not units, F satisfies our definition for a UFD. Thus Theorem 34.30 would give another proof that $F[x]$ is a UFD, except for the fact that we shall actually use Theorem 28.21 in proving Theorem 34.30. In the following section we shall study properties of a certain special class of UFDs, the *Euclidean domains*.

Let us proceed to prove the two theorems.

Every PID Is a UFD

The steps leading up to Theorem 28.21 and its proof indicate the way for our proof of Theorem 34.18. Much of the material will be repetitive. We inefficiently handled the special case of $F[x]$ separately in Theorem 28.21, since it was easy and was the only case we needed for our field theory in general.

To prove that an integral domain D is a UFD, it is necessary to show that both Conditions 1 and 2 of the definition of a UFD are satisfied. For our special case of $F[x]$ in Theorem 28.21, Condition 1 was very easy and resulted from an argument that in a factorization of a polynomial of degree > 0 into a product of two nonconstant polynomials, the degree of each factor was less than the degree of the original polynomial. Thus we couldn't keep on factoring indefinitely without running into unit factors, that is, polynomials of degree 0. For the general case of a PID, it is harder to show that this is so. We now turn to this problem. We shall need the definition of the union of an arbitrary collection of sets. The definition must include the possibility that the collection of sets is infinite.

34.8 Definition If $\{A_i \mid i \in I\}$ is a collection of sets, then the **union** $\cup_{i \in I} A_i$ of the sets A_i is the set of all x such that $x \in A_i$ for at least one $i \in I$. ■

34.9 Lemma Let R be a commutative ring and let $N_1 \subseteq N_2 \subseteq \dots$ be an ascending chain of ideals N_i in R . Then $N = \cup_i N_i$ is an ideal of R .

Proof Let $a, b \in N$. Then there are ideals N_i and N_j in the chain, with $a \in N_i$ and $b \in N_j$. Now either $N_i \subseteq N_j$ or $N_j \subseteq N_i$; let us assume that $N_i \subseteq N_j$, so both a and b are in N_j . This implies that $a \pm b$ and ab are in N_j , so $a \pm b$ and ab are in N . Taking $a = 0$, we see that $b \in N$ implies $-b \in N$, and $0 \in N$ since $0 \in N_i$. Thus N is a subring of D . For $a \in N$ and $d \in D$, we must have $a \in N_i$ for some N_i . Then since N_i is an ideal, $da = ad$ is in N_i . Therefore, $da \in \cup_i N_i$; that is, $da \in N$. Hence N is an ideal. ◆

34.10 Lemma (**Ascending Chain Condition for a PID**) Let D be a PID. If $N_1 \subseteq N_2 \subseteq \dots$ is an ascending chain of ideals N_i , then there exists a positive integer r such that $N_r = N_s$ for all $s \geq r$. Equivalently, every strictly ascending chain of ideals (all inclusions proper) in a PID is of finite length. We express this by saying that the **ascending chain condition** (ACC) holds for ideals in a PID.

Proof By Lemma 34.9, we know that $N = \cup_i N_i$ is an ideal of D . Now as an ideal in D , which is a PID, $N = \langle c \rangle$ for some $c \in D$. Since $N = \cup_i N_i$, we must have $c \in N_r$, for some $r \in \mathbb{Z}^+$. For $s \geq r$, we have

$$\langle c \rangle \subseteq N_r \subseteq N_s \subseteq N = \langle c \rangle.$$

Thus $N_r = N_s$ for $s \geq r$.

The equivalence with the ACC is immediate. ◆

34.11 Definition A commutative ring with unity R that satisfies the ascending chain condition is a **Noetherian ring**. That is, a commutative ring with unity R is Noetherian if for every chain of ideals $N_1 \subseteq N_2 \subseteq N_3 \subseteq \dots$ in R , there is an integer r such that if $s \geq r$, then $N_r = N_s$. ■

Lemma 34.10 states that every PID is a Noetherian ring. In Section 37 we will see that if R is a Noetherian ring, then $R[x]$ is also a Noetherian ring.

In what follows, it will be useful to remember that for elements a and b of a domain D ,

$$\langle a \rangle \subseteq \langle b \rangle \text{ if and only if } b \text{ divides } a, \text{ and}$$

$$\langle a \rangle = \langle b \rangle \text{ if and only if } a \text{ and } b \text{ are associates.}$$

For the first property, note that $\langle a \rangle \subseteq \langle b \rangle$ if and only if $a \in \langle b \rangle$, which is true if and only if $a = bd$ for some $d \in D$, so that b divides a . Using this first property, we see that $\langle a \rangle = \langle b \rangle$ if and only if $a = bc$ and $b = ad$ for some $c, d \in D$. But then $a = adc$ and by canceling, we obtain $1 = dc$. Thus d and c are units, so a and b are associates.

We can now prove Condition 1 of the definition of a UFD for an integral domain that is a PID.

34.12 Theorem Let D be a PID. Every element that is neither 0 nor a unit in D is a product of irreducibles.

Proof Let $a \in D$, where a is neither 0 nor a unit. We first show that a has at least one irreducible factor. If a is an irreducible, we are done. If a is not an irreducible, then $a = a_1 b_1$, where neither a_1 nor b_1 is a unit. Now

$$\langle a \rangle \subset \langle a_1 \rangle,$$

for $\langle a \rangle \subseteq \langle a_1 \rangle$ follows from $a = a_1 b_1$, and if $\langle a \rangle = \langle a_1 \rangle$, then a and a_1 would be associates and b_1 would be a unit, contrary to construction. Continuing this procedure then, starting now with a_1 , we arrive at a strictly ascending chain of ideals

$$\langle a \rangle \subset \langle a_1 \rangle \subset \langle a_2 \rangle \subset \dots$$

By the ACC in Lemma 34.10, this chain terminates with some $\langle a_r \rangle$, and a_r must then be irreducible. Thus a has an irreducible factor a_r .

By what we have just proved, for an element a that is neither 0 nor a unit in D , either a is irreducible or $a = p_1 c_1$ for p_1 an irreducible and c_1 not a unit. By an argument similar to the one just made, in the latter case we can conclude that $\langle a \rangle \subset \langle c_1 \rangle$. If c_1 is not irreducible, then $c_1 = p_2 c_2$ for an irreducible p_2 with c_2 not a unit. Continuing, we get a strictly ascending chain of ideals

$$\langle a \rangle \subset \langle c_1 \rangle \subset \langle c_2 \rangle \subset \dots$$

This chain must terminate, by the ACC in Lemma 34.10, with some $c_r = q_r$, that is an irreducible. Then $a = p_1 p_2 \cdots p_r q_r$. \blacklozenge

This completes our demonstration of Condition 1 of the definition of a UFD. Let us turn to Condition 2. Our arguments here are parallel to those leading to Theorem 28.21. The results we encounter along the way are of some interest in themselves.

34.13 Lemma (Generalization of Theorem 31.25) An ideal $\langle p \rangle$ in a PID is maximal if and only if p is an irreducible.

Proof Let $\langle p \rangle$ be a maximal ideal of D , a PID. Suppose that $p = ab$ in D . Then $\langle p \rangle \subseteq \langle a \rangle$. Suppose that $\langle a \rangle = \langle p \rangle$. Then a and p would be associates, so b must be a unit. If $\langle a \rangle \neq \langle p \rangle$, then we must have $\langle a \rangle = \langle 1 \rangle = D$, since $\langle p \rangle$ is maximal. But then a and 1 are associates, so a is a unit. Thus, if $p = ab$, either a or b must be a unit. Hence p is an irreducible of D .

Conversely, suppose that p is an irreducible in D . Then if $\langle p \rangle \subseteq \langle a \rangle$, we must have $p = ab$. Now if a is a unit, then $\langle a \rangle = \langle 1 \rangle = D$. If a is not a unit, then b must be a unit, so there exists $u \in D$ such that $bu = 1$. Then $pu = abu = a$, so $\langle a \rangle \subseteq \langle p \rangle$, and we have $\langle a \rangle = \langle p \rangle$. Thus $\langle p \rangle \subseteq \langle a \rangle$ implies that either $\langle a \rangle = D$ or $\langle a \rangle = \langle p \rangle$, and $\langle p \rangle \neq D$ or p would be a unit. Hence $\langle p \rangle$ is a maximal ideal. \blacklozenge

34.14 Lemma (Generalization of Theorem 31.27) In a PID, if an irreducible p divides ab , then either $p \mid a$ or $p \mid b$.

Proof Let D be a PID and suppose that for an irreducible p in D we have $p \mid ab$. Then $(ab) \in \langle p \rangle$. Since every maximal ideal in D is a prime ideal by Corollary 31.16, $(ab) \in \langle p \rangle$ implies that either $a \in \langle p \rangle$ or $b \in \langle p \rangle$, giving either $p \mid a$ or $p \mid b$. \blacklozenge

34.15 Corollary If p is an irreducible in a PID and p divides the product $a_1 a_2 \cdots a_n$ for $a_i \in D$, then $p \mid a_i$ for at least one i .

Proof Proof of this corollary is immediate from Lemma 34.14 if we use mathematical induction. \diamond

34.16 Definition A nonzero nonunit element p of an integral domain D is a **prime** if, for all $a, b \in D$, $p | ab$ implies either $p | a$ or $p | b$. \blacksquare

Lemma 34.14 focused our attention on the defining property of a prime. In Exercises 25 and 26, we ask you to show that a prime in an integral domain is always an irreducible and that in a UFD an irreducible is also a prime. Thus the concepts of prime and irreducible coincide in a UFD. Example 34.17 will exhibit an integral domain containing some irreducibles that are not primes, so the concepts do not coincide in every domain.

34.17 Example Let F be a field and let D be the subdomain $F[x^3, xy, y^3]$ of $F[x, y]$. Then x^3, xy , and y^3 are irreducibles in D , but

$$(x^3)(y^3) = (xy)(xy)(xy).$$

Since xy divides x^3y^3 but not x^3 or y^3 , we see that xy is not a prime. Similar arguments show that neither x^3 nor y^3 is a prime. \blacktriangle

The defining property of a prime is precisely what is needed to establish uniqueness of factorization, Condition 2 in the definition of a UFD. We now complete the proof of Theorem 34.18 by demonstrating the uniqueness of factorization in a PID.

34.18 Theorem (Generalization of Theorem 28.21) Every PID is a UFD.

Proof Theorem 34.12 shows that if D is a PID, then each $a \in D$, where a is neither 0 nor a unit, has a factorization

$$a = p_1 p_2 \cdots p_r$$

into irreducibles. It remains for us to show uniqueness. Let

$$a = q_1 q_2 \cdots q_s$$

be another such factorization into irreducibles. Then we have $p_1 | (q_1 q_2 \cdots q_s)$, which implies that $p_1 | q_j$ for some j by Corollary 34.15. By changing the order of the q_j if necessary, we can assume that $j = 1$ so $p_1 | q_1$. Then $q_1 = p_1 u_1$, and since q_1 is an irreducible, u_1 is a unit, so p_1 and q_1 are associates. We have then

$$p_1 p_2 \cdots p_r = p_1 u_1 q_2 \cdots q_s,$$

so by the cancellation law in D ,

$$p_2 \cdots p_r = u_1 q_2 \cdots q_s.$$

Continuing this process, starting with p_2 and so on, we finally arrive at

$$1 = u_1 u_2 \cdots u_r q_{r+1} \cdots q_s.$$

Since the q_j are irreducibles, we must have $r = s$. \diamond

Example 34.32 at the end of this section will show that the converse to Theorem 34.18 is false. That is, a UFD need not be a PID.

Many algebra texts start by proving the following corollary of Theorem 34.18. We have assumed that you were familiar with this corollary and used it freely in our other work.

34.19 Corollary (Fundamental Theorem of Arithmetic) The integral domain \mathbb{Z} is a UFD.

Proof We have seen that all ideals in \mathbb{Z} are of the form $n\mathbb{Z} = \langle n \rangle$ for $n \in \mathbb{Z}$. Thus \mathbb{Z} is a PID, and Theorem 34.18 applies. \blacklozenge

It is worth noting that the proof that \mathbb{Z} is a PID was really way back in Corollary 6.7. We proved Theorem 6.6 by using the division algorithm for \mathbb{Z} exactly as we proved, in Theorem 31.24, that $F[x]$ is a PID by using the division algorithm for $F[x]$. In Section 35, we shall examine this parallel more closely.

If D Is a UFD, Then $D[x]$ Is a UFD

We now start the proof of Theorem 34.30, our second main result for this section. The idea of the argument is as follows. Let D be a UFD. We can form a field of quotients F of D . Then $F[x]$ is a UFD by Theorem 28.21, and we shall show that we can recover a factorization for $f(x) \in D[x]$ from its factorization in $F[x]$. It will be necessary to compare the irreducibles in $F[x]$ with those in $D[x]$, of course. This approach, which we prefer as more intuitive than some more efficient modern ones, is essentially due to Gauss.

34.20 Definition Let D be a UFD and let a_1, a_2, \dots, a_n be nonzero elements of D . An element d of D is a **greatest common divisor** (abbreviated gcd) of all of the a_i if $d | a_i$ for $i = 1, \dots, n$ and any other $d' \in D$ that divides all the a_i also divides d . \blacksquare

In this definition, we called d “a” gcd rather than “the” gcd because gcd’s are only defined up to units. Suppose that d and d' are two gcd’s of a_i for $i = 1, \dots, n$. Then $d | d'$ and $d' | d$ by our definition. Thus $d = q'd'$ and $d' = qd$ for some $q, q' \in D$, so $1d = q'qd$. By cancellation in D , we see that $q'q = 1$ so q and q' are indeed units.

The technique in the example that follows shows that gcd’s exist in a UFD.

34.21 Example Let us find a gcd of 420, -168, and 252 in the UFD \mathbb{Z} . Factoring, we obtain $420 = 2^2 \cdot 3 \cdot 5 \cdot 7$, $-168 = 2^3 \cdot (-3) \cdot 7$, and $252 = 2^2 \cdot 3^2 \cdot 7$. We choose one of these numbers, say 420, and find the highest power of each of its irreducible factors (up to associates) that divides all the numbers, 420, -168, and 252 in our case. We take as gcd the product of these highest powers of irreducibles. For our example, these powers of irreducible factors of 420 are $2^2, 3^1, 5^0$, and 7^1 so we take as gcd $d = 4 \cdot 3 \cdot 1 \cdot 7 = 84$. The only other gcd of these numbers in \mathbb{Z} is -84, because 1 and -1 are the only units. \blacktriangle

Execution of the technique in Example 34.21 depends on being able to factor an element of a UFD into a product of irreducibles. This can be a tough job, even in \mathbb{Z} . Section 35 will exhibit a technique, the Euclidean Algorithm, that will allow us to find gcd’s without factoring in a class of UFD’s that includes \mathbb{Z} and $F[x]$ for a field F .

34.22 Definition Let D be a UFD. A nonconstant polynomial

$$f(x) = a_0 + a_1x + \cdots + a_nx^n$$

in $D[x]$ is **primitive** if 1 is a gcd of the a_i for $i = 0, 1, \dots, n$. \blacksquare

34.23 Example In $\mathbb{Z}[x]$, $4x^2 + 3x + 2$ is primitive, but $4x^2 + 6x + 2$ is not, since 2, a nonunit in \mathbb{Z} , is a common divisor of 4, 6, and 2. \blacktriangle

Observe that every nonconstant irreducible in $D[x]$ must be a primitive polynomial.

34.24 Lemma If D is a UFD, then for every nonconstant $f(x) \in D[x]$ we have $f(x) = (c)g(x)$, where $c \in D$, $g(x) \in D[x]$, and $g(x)$ is primitive. The element c is unique up to a unit factor in D and is the **content** of $f(x)$. Also $g(x)$ is unique up to a unit factor in D .

Proof Let $f(x) \in D[x]$ be given where $f(x)$ is a nonconstant polynomial with coefficients a_0, a_1, \dots, a_n . Let c be a gcd of the a_i for $i = 0, 1, \dots, n$. Then for each i , we have $a_i = cq_i$ for some $q_i \in D$. By the distributive law, we have $f(x) = (c)g(x)$, where no irreducible in D divides all of the coefficients q_0, q_1, \dots, q_n of $g(x)$. Thus $g(x)$ is a primitive polynomial.

For uniqueness, if also $f(x) = (d)h(x)$ for $d \in D$, $h(x) \in D[x]$, and $h(x)$ primitive, then each irreducible factor of c must divide d and conversely. By setting $(c)g(x) = (d)h(x)$ and canceling irreducible factors of c into d , we arrive at $(u)g(x) = (v)h(x)$ for a unit $u \in D$. But then v must be a unit of D or we would be able to cancel irreducible factors of v into u . Thus u and v are both units, so c is unique up to a unit factor. From $f(x) = (c)g(x)$, we see that the primitive polynomial $g(x)$ is also unique up to a unit factor. \diamond

34.25 Example In $\mathbb{Z}[x]$,

$$4x^2 + 6x - 8 = (2)(2x^2 + 3x - 4),$$

where $2x^2 + 3x - 4$ is primitive. \blacktriangle

34.26 Lemma (Gauss's Lemma) If D is a UFD, then a product of two primitive polynomials in $D[x]$ is again primitive.

Proof Let

$$f(x) = a_0 + a_1x + \dots + a_nx^n$$

and

$$g(x) = b_0 + b_1x + \dots + b_mx^m$$

be primitive in $D[x]$, and let $h(x) = f(x)g(x)$. Let p be an irreducible in D . Then p does not divide all a_i and p does not divide all b_j , since $f(x)$ and $g(x)$ are primitive. Let a_r be the first coefficient of $f(x)$ not divisible by p ; that is, $p \nmid a_i$ for $i < r$, but $p \nmid a_r$ (that is, p does not divide a_r). Similarly, let $p \mid b_j$ for $j < s$, but $p \nmid b_s$. The coefficient of x^{r+s} in $h(x) = f(x)g(x)$ is

$$c_{r+s} = (a_0b_{r+s} + \dots + a_{r-1}b_{s+1}) + a_rb_s + (a_{r+1}b_{s-1} + \dots + a_{r+s}b_0).$$

Now $p \mid a_i$ for $i < r$ implies that

$$p \mid (a_0b_{r+s} + \dots + a_{r-1}b_{s+1}),$$

and also $p \mid b_j$ for $j < s$ implies that

$$p \mid (a_{r+1}b_{s-1} + \dots + a_{r+s}b_0).$$

But p does not divide a_r or b_s , so p does not divide a_rb_s , and consequently p does not divide c_{r+s} . This shows that given an irreducible $p \in D$, there is some coefficient of $f(x)g(x)$ not divisible by p . Thus $f(x)g(x)$ is primitive. \diamond

34.27 Corollary If D is a UFD, then a finite product of primitive polynomials in $D[x]$ is again primitive.

Proof This corollary follows from Lemma 34.26 by induction. \diamond

Now let D be a UFD and let F be a field of quotients of D . By Theorem 28.21, $F[x]$ is a UFD. As we said earlier, we shall show that $D[x]$ is a UFD by carrying a factorization in $F[x]$ of $f(x) \in D[x]$ back into one in $D[x]$. The next lemma relates the nonconstant irreducibles of $D[x]$ to those of $F[x]$. This is the last important step.

34.28 Lemma Let D be a UFD and let F be a field of quotients of D . Let $f(x) \in D[x]$, where $(\text{degree } f(x)) > 0$. If $f(x)$ is an irreducible in $D[x]$, then $f(x)$ is also an irreducible in $F[x]$. Also, if $f(x)$ is primitive in $D[x]$ and irreducible in $F[x]$, then $f(x)$ is irreducible in $D[x]$.

Proof Suppose that a nonconstant $f(x) \in D[x]$ factors into polynomials of lower degree in $F[x]$, that is,

$$f(x) = r(x)s(x)$$

for $r(x), s(x) \in F[x]$. Then since F is a field of quotients of D , each coefficient in $r(x)$ and $s(x)$ is of the form a/b for some $a, b \in D$. By clearing denominators, we can get

$$(d)f(x) = r_1(x)s_1(x)$$

for $d \in D$, and $r_1(x), s_1(x) \in D[x]$, where the degrees of $r_1(x)$ and $s_1(x)$ are the degrees of $r(x)$ and $s(x)$, respectively. By Lemma 34.24, $f(x) = (c)g(x)$, $r_1(x) = (c_1)r_2(x)$, and $s_1(x) = (c_2)s_2(x)$ for primitive polynomials $g(x)$, $r_2(x)$, and $s_2(x)$, and $c, c_1, c_2 \in D$. Then

$$(dc)g(x) = (c_1c_2)r_2(x)s_2(x),$$

and by Lemma 34.26, $r_2(x)s_2(x)$ is primitive. By the uniqueness part of Lemma 34.24, $c_1c_2 = dcu$ for some unit u in D . But then

$$(dc)g(x) = (dcu)r_2(x)s_2(x),$$

so

$$f(x) = (c)g(x) = (cu)r_2(x)s_2(x).$$

We have shown that if $f(x)$ factors nontrivially in $F[x]$, then $f(x)$ factors nontrivially into polynomials of the same degrees in $D[x]$. Thus if $f(x) \in D[x]$ is irreducible in $D[x]$, it must be irreducible in $F[x]$.

A nonconstant $f(x) \in D[x]$ that is primitive in $D[x]$ and irreducible in $F[x]$ is also irreducible in $D[x]$, since $D[x] \subseteq F[x]$. ◆

Lemma 34.28 shows that if D is a UFD, the irreducibles in $D[x]$ are precisely the irreducibles in D , together with the nonconstant primitive polynomials that are irreducible in $F[x]$, where F is a field of quotients of $D[x]$.

The preceding lemma is very important in its own right. This is indicated by the following corollary, a special case of which was our Theorem 28.12. (We admit that it does not seem very sensible to call a special case of a corollary of a lemma a theorem. The label assigned to a result depends somewhat on the context in which it appears.)

34.29 Corollary If D is a UFD and F is a field of quotients of D , then a nonconstant $f(x) \in D[x]$ factors into a product of two polynomials of lower degrees r and s in $F[x]$ if and only if it has a factorization into polynomials of the same degrees r and s in $D[x]$.

Proof It was shown in the proof of Lemma 34.28 that if $f(x)$ factors into a product of two polynomials of lower degree in $F[x]$, then it has a factorization into polynomials of the same degrees in $D[x]$ (see the next-to-last sentence of the first paragraph of the proof).

The converse holds since $D[x] \subseteq F[x]$. ◆

We are now prepared to prove our main theorem.

34.30 Theorem If D is a UFD, then $D[x]$ is a UFD.

Proof Let $f(x) \in D[x]$, where $f(x)$ is neither 0 nor a unit. If $f(x)$ is of degree 0, we are done, since D is a UFD. Suppose that $(\deg f(x)) > 0$. Let

$$f(x) = g_1(x)g_2(x) \cdots g_r(x)$$

be a factorization of $f(x)$ in $D[x]$ having the greatest number r of factors of positive degree. (There is such a greatest number of such factors because r cannot exceed the degree of $f(x)$.) Now factor each $g_i(x)$ in the form $g_i(x) = c_i h_i(x)$ where c_i is the content

of $g_i(x)$ and $h_i(x)$ is a primitive polynomial. Each of the $h_i(x)$ is irreducible, because if it could be factored, none of the factors could lie in D , hence all would have positive degree leading to a corresponding factorization of $g_i(x)$, and then to a factorization of $f(x)$ with more than r factors of positive degree, contradicting our choice of r . Thus we now have

$$f(x) = c_1 h_1(x) c_2 h_2(x) \cdots c_r h_r(x)$$

where the $h_i(x)$ are irreducible in $D[x]$. If we now factor the c_i into irreducibles in D , we obtain a factorization of $f(x)$ into a product of irreducibles in $D[x]$.

The factorization of $f(x) \in D[x]$, where $f(x)$ has degree 0, is unique since D is a UFD; see the comment following Lemma 34.28. If $f(x)$ has degree greater than 0, we can view any factorization of $f(x)$ into irreducibles in $D[x]$ as a factorization in $F[x]$ into units (that is, the factors in D) and irreducible polynomials in $F[x]$ by Lemma 34.28. By Theorem 28.21, these polynomials are unique, except for possible constant factors in F . But as an irreducible in $D[x]$, each polynomial of degree >0 appearing in the factorization of $f(x)$ in $D[x]$ is primitive. By the uniqueness part of Lemma 34.24, this shows that these polynomials are unique in $D[x]$ up to unit factors, that is, associates. The product of the irreducibles in D in the factorization of $f(x)$ is the content of $f(x)$, which is again unique up to a unit factor by Lemma 34.24. Thus all irreducibles in $D[x]$ appearing in the factorization are unique up to order and associates. \blacklozenge

34.31 Corollary If F is a field and x_1, \dots, x_n are indeterminates, then $F[x_1, \dots, x_n]$ is a UFD.

Proof By Theorem 28.21, $F[x_1]$ is a UFD. By Theorem 34.30, so is $(F[x_1])[x_2] = F[x_1, x_2]$. Continuing in this procedure, we see (by induction) that $F[x_1, \dots, x_n]$ is a UFD. \blacklozenge

We have seen that a PID is a UFD. Corollary 34.31 makes it easy for us to give an example that shows that *not every UFD is a PID*.

34.32 Example Let F be a field and let x and y be indeterminates. Then $F[x, y]$ is a UFD by Corollary 34.30. Consider the set N of all polynomials in x and y in $F[x, y]$ having constant term 0. Then N is an ideal, but not a principal ideal. Thus $F[x, y]$ is not a PID. \blacktriangle

Another example of a UFD that is not a PID is $\mathbb{Z}[x]$, as shown in Exercise 12, Section 35.

■ EXERCISES 34

Computations

In Exercises 1 through 8, determine whether the element is an irreducible of the indicated domain.

- 1. 5 in \mathbb{Z}
- 2. -17 in \mathbb{Z}
- 3. 14 in \mathbb{Z}
- 4. $2x - 3$ in $\mathbb{Z}[x]$
- 5. $2x - 10$ in $\mathbb{Z}[x]$
- 6. $2x - 3$ in $\mathbb{Q}[x]$
- 7. $2x - 10$ in $\mathbb{Q}[x]$
- 8. $2x - 10$ in $\mathbb{Z}_{11}[x]$
- 9. If possible, give four different associates of $2x - 7$ viewed as an element of $\mathbb{Z}[x]$; of $\mathbb{Q}[x]$; of $\mathbb{Z}_{11}[x]$.
- 10. Factor the polynomial $4x^2 - 4x + 8$ into a product of irreducibles viewing it as an element of the integral domain $\mathbb{Z}[x]$; of the integral domain $\mathbb{Q}[x]$; of the integral domain $\mathbb{Z}_{11}[x]$.

In Exercises 11 through 13, find all gcd's of the given elements of \mathbb{Z} .

- 11. 234, 3250, 1690
- 12. 784, -1960 , 448
- 13. 2178, 396, 792, 594

In Exercises 14 through 17, express the given polynomial as the product of its content with a primitive polynomial in the indicated UFD.

14. $18x^2 - 12x + 48$ in $\mathbb{Z}[x]$

15. $18x^2 - 12x + 48$ in $\mathbb{Q}[x]$

16. $2x^2 - 3x + 6$ in $\mathbb{Z}[x]$

17. $2x^2 - 3x + 6$ in $\mathbb{Z}_7[x]$

Concepts

In Exercises 18 through 20, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

18. Two elements a and b in an integral domain D are *associates* in D if and only if their quotient a/b in D is a unit.
19. An element of an integral domain D is an *irreducible* of D if and only if it cannot be factored into a product of two elements of D .
20. An element of an integral domain D is a *prime* of D if and only if it cannot be factored into a product of two smaller elements of D .

21. Determine whether each of the following is true or false.

- a. Every field is a UFD.
 - b. Every field is a PID.
 - c. Every PID is a UFD.
 - d. Every UFD is a PID.
 - e. $\mathbb{Z}[x]$ is a UFD.
 - f. Any two irreducibles in any UFD are associates.
 - g. If D is a PID, then $D[x]$ is a PID.
 - h. If D is a UFD, then $D[x]$ is a UFD.
 - i. In any UFD, if $p \mid a$ for an irreducible p , then p itself appears in every factorization of a .
 - j. A UFD has no divisors of 0.
22. Let D be a UFD. Describe the irreducibles in $D[x]$ in terms of the irreducibles in D and the irreducibles in $F[x]$, where F is a field of quotients of D .
 23. Lemma 34.28 states that if D is a UFD with a field of quotients F , then a nonconstant irreducible $f(x)$ of $D[x]$ is also an irreducible of $F[x]$. Show by an example that a $g(x) \in D[x]$ that is an irreducible of $F[x]$ need not be an irreducible of $D[x]$.
 24. All our work in this section was restricted to integral domains. Taking the same definition in this section but for a commutative ring with unity, consider factorizations into irreducibles in $\mathbb{Z} \times \mathbb{Z}$. What can happen? Consider in particular $(1, 0)$.

Theory

25. Prove that if p is a prime in an integral domain D , then p is an irreducible.
26. Prove that if p is an irreducible in a UFD, then p is a prime.
27. For a commutative ring R with unity show that the relation $a \sim b$ if a is an associate of b (that is, if $a = bu$ for u a unit in R) is an equivalence relation on R .
28. Let D be an integral domain. Exercise 39, Section 22 showed that $\langle U, \cdot \rangle$ is a group where U is the set of units of D . Show that the set $D^* - U$ of nonunits of D excluding 0 is closed under multiplication. Is this set a group under the multiplication of D ?
29. Let D be a UFD. Show that a nonconstant divisor of a primitive polynomial in $D[x]$ is again a primitive polynomial.
30. Show that in a PID, every proper ideal is contained in a maximal ideal. [Hint: Use Lemma 34.10.]
31. Factor $x^3 - y^3$ into irreducibles in $\mathbb{Q}[x, y]$ and prove that each of the factors is irreducible.

There are several other concepts often considered that are similar in character to the ascending chain condition on ideals in a ring. The following three exercises concern some of these concepts.

32. Let R be any ring. The **ascending chain condition** (ACC) for ideals holds in R if every strictly increasing sequence $N_1 \subset N_2 \subset N_3 \subset \dots$ of ideals in R is of finite length. The **maximum condition** (MC) for ideals holds in R if every nonempty set S of ideals in R contains an ideal not properly contained in any other ideal of the set S . The **finite basis condition** (FBC) for ideals holds in R if for each ideal N in R , there is a finite set $B_N = \{b_1, \dots, b_n\} \subseteq N$ such that N is the intersection of all ideals of R containing B_N . The set B_N is a **finite generating set for N** .

Show that for every ring R , the conditions ACC, MC, and FBC are equivalent.

33. Let R be any ring. The **descending chain condition** (DCC) for ideals holds in R if every strictly decreasing sequence $N_1 \supset N_2 \supset N_3 \supset \dots$ of ideals in R is of finite length. The **minimum condition** (mC) for ideals holds in R if given any set S of ideals of R , there is an ideal of S that does not properly contain any other ideal in the set S .

Show that for every ring, the conditions DCC and mC are equivalent.

34. Give an example of a ring in which ACC holds but DCC does not hold. (See Exercises 32 and 33.)

SECTION 35

EUCLIDEAN DOMAINS

We have remarked several times on the importance of division algorithms. Our first contact with them was the *division algorithm for \mathbb{Z}* in Section 6. This algorithm was used to prove the important theorem that a subgroup of a cyclic group is cyclic, that is, has a single generator. Of course, this shows at once that \mathbb{Z} is a PID. The *division algorithm for $F[x]$* appeared in Theorem 28.2 and was used in a completely analogous way to show that $F[x]$ is a PID. A technique of mathematics is to take some clearly related situations and to try to bring them under one roof by abstracting the important ideas common to them. The following definition is an illustration of this technique, as is this whole text! Let us see what we can develop by starting with the existence of a fairly general division algorithm in an integral domain.

35.1 Definition

A **Euclidean norm** on an integral domain D is a function ν mapping the nonzero elements of D into the nonnegative integers such that the following conditions are satisfied:

1. For all $a, b \in D$ with $b \neq 0$, there exist q and r in D such that $a = bq + r$, where either $r = 0$ or $\nu(r) < \nu(b)$.
2. For all $a, b \in D$, where neither a nor b is 0, $\nu(a) \leq \nu(ab)$.

An integral domain D is a **Euclidean domain** if there exists a Euclidean norm on D . ■

The importance of Condition 1 is clear from our discussion. The importance of Condition 2 is that it will enable us to characterize the units of a Euclidean domain D .

35.2 Example

The integral domain \mathbb{Z} is a Euclidean domain, for the function ν defined by $\nu(n) = |n|$ for $n \neq 0$ in \mathbb{Z} is a Euclidean norm on \mathbb{Z} . Condition 1 holds by the division algorithm for \mathbb{Z} . Condition 2 follows from $|ab| = |a||b|$ and $|a| \geq 1$ for $a \neq 0$ in \mathbb{Z} . ▲

35.3 Example

If F is a field, then $F[x]$ is a Euclidean domain, for the function ν defined by $\nu(f(x)) = (\text{degree } f(x))$ for $f(x) \in F[x]$, and $f(x) \neq 0$ is a Euclidean norm. Condition 1 holds by Theorem 28.2, and Condition 2 holds since the degree of the product of two polynomials is the sum of their degrees. ▲

Of course, we should give some examples of Euclidean domains other than these familiar ones that motivated the definition. We shall do this in Section 36. In view of the opening remarks, we anticipate the following theorem.

35.4 Theorem

Every Euclidean domain is a PID.

Proof Let D be a Euclidean domain with a Euclidean norm v , and let N be an ideal in D . If $N = \{0\}$, then $N = \langle 0 \rangle$ and N is principal. Suppose that $N \neq \{0\}$. Then there exists $b \neq 0$ in N . Let us choose b such that $v(b)$ is minimal among all $v(n)$ for $n \in N$. We claim that $N = \langle b \rangle$. Let $a \in N$. Then by Condition 1 for a Euclidean domain, there exist q and r in D such that

$$a = bq + r,$$

where either $r = 0$ or $v(r) < v(b)$. Now $r = a - bq$ and $a, b \in N$, so that $r \in N$ since N is an ideal. Thus $v(r) < v(b)$ is impossible by our choice of b . Hence $r = 0$, so $a = bq$. Since a was any element of N , we see that $N = \langle b \rangle$. \blacklozenge

35.5 Corollary A Euclidean domain is a UFD.

Proof By Theorem 35.4, a Euclidean domain is a PID and by Theorem 34.18, a PID is a UFD. \blacklozenge

Finally, we should mention that while a Euclidean domain is a PID by Theorem 35.4, not every PID is a Euclidean domain. Examples of PIDs that are not Euclidean are not easily found, however.

Arithmetic in Euclidean Domains

We shall now investigate some properties of Euclidean domains related to their multiplicative structure. We emphasize that the arithmetic structure of a Euclidean domain is not affected in any way by a Euclidean norm v on the domain. A Euclidean norm is merely a useful tool for possibly throwing some light on this arithmetic structure of the domain. The arithmetic structure of a domain D is completely determined by the set D and the two binary operations $+$ and \cdot on D .

Let D be a Euclidean domain with a Euclidean norm v . We can use Condition 2 of a Euclidean norm to characterize the units of D .

35.6 Theorem For a Euclidean domain with a Euclidean norm v , $v(1)$ is minimal among all $v(a)$ for nonzero $a \in D$, and $u \in D$ is a unit if and only if $v(u) = v(1)$.

Proof Condition 2 for v tells us at once that for $a \neq 0$,

$$v(1) \leq v(1a) = v(a).$$

On the other hand, if u is a unit in D , then

$$v(u) \leq v(uu^{-1}) = v(1).$$

Thus

$$v(u) = v(1)$$

for a unit u in D .

Conversely, suppose that a nonzero $u \in D$ is such that $v(u) = v(1)$. Then by the division algorithm, there exist q and r in D such that

$$1 = uq + r,$$

where either $r = 0$ or $v(r) < v(u)$. But since $v(u) = v(1)$ is minimal over all $v(d)$ for nonzero $d \in D$, $v(r) < v(u)$ is impossible. Hence $r = 0$ and $1 = uq$, so u is a unit. \blacklozenge

■ HISTORICAL NOTE

The Euclidean algorithm appears in Euclid's *Elements* as propositions 1 and 2 of Book VII, where it is used as here to find the greatest common divisor of two integers. Euclid uses it again in Book X (propositions 2 and 3) to find the greatest common measure of two magnitudes (if it exists) and to determine whether two magnitudes are incommensurable.

The algorithm appears again in the *Brahmesphutasiddhanta* (Correct Astronomical System of Brahma) (628) of the seventh-century Indian mathematician and astronomer Brahmagupta. To solve the indeterminate equation $rx + c = sy$ in integers, Brahmagupta uses Euclid's procedure to "reciprocally divide" r by s until he reaches the final nonzero remainder. By then using, in effect, a substitution procedure based on the various quotients and remainders, he produces a straightforward algorithm for finding the smallest positive solution to his equation.

The thirteenth-century Chinese algebraist Qin Jiushao also used the Euclidean algorithm in his solution of the so-called Chinese Remainder problem published in the *Shushu jiuzhang* (Mathematical Treatise in Nine Sections) (1247). Qin's goal was to display a method for solving the system of congruences $N \equiv r_i \pmod{m_i}$. As part of that method he needed to solve congruences of the form $Nx \equiv 1 \pmod{m}$, where N and m are relatively prime. The solution to a congruence of this form is again found by a substitution procedure, different from the Indian one, using the quotients and remainders from the Euclidean algorithm applied to N and m . It is not known whether the common element in the Indian and Chinese algorithms, the Euclidean algorithm itself, was discovered independently in these cultures or was learned from Greek sources.

35.7 Example For \mathbb{Z} with $\nu(n) = |n|$, the minimum of $\nu(n)$ for nonzero $n \in \mathbb{Z}$ is 1, and 1 and -1 are the only elements of \mathbb{Z} with $\nu(n) = 1$. Of course, 1 and -1 are exactly the units of \mathbb{Z} . \blacktriangle

35.8 Example For $F[x]$ with $\nu(f(x)) = (\text{degree } f(x))$ for $f(x) \neq 0$, the minimum value of $\nu(f(x))$ for all nonzero $f(x) \in F[x]$ is 0. The nonzero polynomials of degree 0 are exactly the nonzero elements of F , and these are precisely the units of $F[x]$. \blacktriangle

We emphasize that everything we prove here holds in *every* Euclidean domain, in particular in \mathbb{Z} and $F[x]$. As indicated in Example 34.21, we can show that any a and b in a UFD have a gcd and actually compute one by factoring a and b into irreducibles, but such factorizations can be very tough to find. However, if a UFD is actually Euclidean, and we know an easily computed Euclidean norm, there is an easy constructive way to find gcd's, as the next theorem shows.

35.9 Theorem (Euclidean Algorithm) Let D be a Euclidean domain with a Euclidean norm ν , and let a and b be nonzero elements of D . Let r_1 be as in Condition 1 for a Euclidean norm, that is,

$$a = bq_1 + r_1,$$

where either $r_1 = 0$ or $\nu(r_1) < \nu(b)$. If $r_1 \neq 0$, let r_2 be such that

$$b = r_1q_2 + r_2,$$

where either $r_2 = 0$ or $\nu(r_2) < \nu(r_1)$. In general, let r_{i+1} be such that

$$r_{i-1} = r_iq_{i+1} + r_{i+1},$$

where either $r_{i+1} = 0$ or $\nu(r_{i+1}) < \nu(r_i)$. Then the sequence r_i, r_2, \dots must terminate with some $r_s = 0$. If $r_1 = 0$, then b is a gcd of a and b . If $r_1 \neq 0$ and r_s is the first $r_i = 0$, then a gcd of a and b is r_{s-1} .

Furthermore, if d is a gcd of a and b , then there exist λ and μ in D such that $d = \lambda a + \mu b$.

Proof Since $v(r_i) < v(r_{i-1})$ and $v(r_i)$ is a nonnegative integer, it follows that after some finite number of steps we must arrive at some $r_s = 0$.

If $r_1 = 0$, then $a = bq_1$, and b is a gcd of a and b . Suppose $r_1 \neq 0$. Then if $d \mid a$ and $d \mid b$, we have

$$d \mid (a - bq_1),$$

so $d \mid r_1$. However, if $d_1 \mid r_1$ and $d_1 \mid b$, then

$$d_1 \mid (bq_1 + r_1),$$

so $d_1 \mid a$. Thus the set of common divisors of a and b is the same set as the set of common divisors of b and r_1 . By a similar argument, if $r_2 \neq 0$, the set of common divisors of b and r_1 is the same set as the set of common divisors of r_1 and r_2 . Continuing this process, we see finally that the set of common divisors of a and b is the same set as the set of common divisors of r_{s-2} and r_{s-1} , where r_s is the first r_i equal to 0. Thus a gcd of r_{s-2} and r_{s-1} is also a gcd of a and b . But the equation

$$r_{s-2} = q_s r_{s-1} + r_s = q_s r_{s-1}$$

shows that a gcd of r_{s-2} and r_{s-1} is r_{s-1} .

It remains to show that we can express a gcd d of a and b as $d = \lambda a + \mu b$. In terms of the construction just given, if $d = b$, then $d = 0a + 1b$ and we are done. If $d = r_{s-1}$, then, working backward through our equations, we can express each r_i in the form $\lambda_i r_{i-1} + \mu_i r_{i-2}$ for some $\lambda_i, \mu_i \in D$. To illustrate using the first step, from the equation

$$r_{s-3} = q_{s-1} r_{s-2} + r_{s-1}$$

we obtain

$$d = r_{s-1} = r_{s-3} - q_{s-1} r_{s-2}. \quad (1)$$

We then express r_{s-2} in terms of r_{s-3} and r_{s-4} and substitute in Eq. (1) to express d in terms of r_{s-3} and r_{s-4} . Eventually, we will have

$$\begin{aligned} d &= \lambda_3 r_2 + \mu_3 r_1 = \lambda_3(b - r_1 q_2) + \mu_3 r_1 = \lambda_3 b + (\mu_3 - \lambda_3 q_2) r_1 \\ &= \lambda_3 b + (\mu_3 - \lambda_3 q_2)(a - bq_1) \end{aligned}$$

which can be expressed in the form $d = \lambda a + \mu b$. If d' is any other gcd of a and b , then $d' = ud$ for some unit u , so $d' = (\lambda u)a + (\mu u)b$. ◆

The nice thing about Theorem 35.9 is that it can be implemented on a computer. Of course, we anticipate that of anything that is labeled an “algorithm.”

35.10 Example Let us illustrate the Euclidean algorithm for the Euclidean norm $\|\cdot\|$ on \mathbb{Z} by computing a gcd of 22,471 and 3,266. We just apply the division algorithm over and over again, and the last nonzero remainder is a gcd. We label the numbers obtained as in Theorem 35.9 to further illustrate the statement and proof of the theorem. The computations are easily checked.

$$\begin{array}{lll} a = 22,471 & & \\ b = 3,266 & & \\ 22,471 = (3,266)6 + 2,875 & r_1 = 2,875 & \\ 3,266 = (2,875)1 + 391 & r_2 = 391 & \\ 2,875 = (391)7 + 138 & r_3 = 138 & \\ 391 = (138)2 + 115 & r_4 = 115 & \\ 138 = (115)1 + 23 & r_5 = 23 & \\ 115 = (23)5 + 0 & r_6 = 0 & \end{array}$$

Thus $r_5 = 23$ is a gcd of 22,471 and 3,266. We found a gcd without factoring! This is important, for sometimes it is very difficult to find a factorization of an integer into primes. \blacktriangle

35.11 Example

Note that the division algorithm Condition 1 in the definition of a Euclidean norm says nothing about r being “positive.” In computing a gcd in \mathbb{Z} by the Euclidean algorithm for $| |$, as in Example 35.10, it is surely in our interest to make $|r_i|$ as small as possible in each division. Thus, repeating Example 35.10, it would be more efficient to write

$$\begin{array}{ll} a = 22,471 & \\ b = 3,266 & \\ 22,471 = (3,266)7 - 391 & r_1 = -391 \\ 3,266 = (391)8 + 138 & r_2 = 138 \\ 391 = (138)3 - 23 & r_3 = -23 \\ 138 = (23)6 + 0 & r_4 = 0 \end{array}$$

We can change the sign of r_i from negative to positive when we wish since the divisors of r_i and $-r_i$ are the same. \blacktriangle

■ EXERCISES 35

Computations

In Exercises 1 through 5, state whether the given function v is a Euclidean norm for the given integral domain.

1. The function v for \mathbb{Z} given by $v(n) = n^2$ for nonzero $n \in \mathbb{Z}$
2. The function v for $\mathbb{Z}[x]$ given by $v(f(x)) = (\text{degree of } f(x))$ for $f(x) \in \mathbb{Z}[x], f(x) \neq 0$
3. The function v for $\mathbb{Z}[x]$ given by $v(f(x)) = (\text{the absolute value of the coefficient of the highest degree nonzero term of } f(x))$ for nonzero $f(x) \in \mathbb{Z}[x]$
4. The function v for \mathbb{Q} given by $v(a) = a^2$ for nonzero $a \in \mathbb{Q}$
5. The function v for \mathbb{Q} given by $v(a) = 50$ for nonzero $a \in \mathbb{Q}$
6. By referring to Example 35.11, actually express the gcd 23 in the form $\lambda(22,471) + \mu(3,266)$ for $\lambda, \mu \in \mathbb{Z}$. [Hint: From the next-to-last line of the computation in Example 35.11, $23 = (138)3 - 391$. From the line before that, $138 = 3,266 - (391)8$, so substituting, you get $23 = [3,266 - (391)8]3 - 391$, and so on. That is, work your way back up to actually find values for λ and μ .]
7. Find a gcd of 49,349 and 15,555 in \mathbb{Z} .
8. Following the idea of Exercise 6 and referring to Exercise 7, express the positive gcd of 49,349 and 15,555 in \mathbb{Z} in the form $\lambda(49,349) + \mu(15,555)$ for $\lambda, \mu \in \mathbb{Z}$.
9. Find a gcd of

$$x^{10} - 3x^9 + 3x^8 - 11x^7 + 11x^6 - 11x^5 + 19x^4 - 13x^3 + 8x^2 - 9x + 3$$

and

$$x^6 - 3x^5 + 3x^4 - 9x^3 + 5x^2 - 5x + 2$$

in $\mathbb{Q}[x]$.

10. Describe how the Euclidean Algorithm can be used to find the gcd of n members a_1, a_2, \dots, a_n of a Euclidean domain.
11. Using your method devised in Exercise 10, find the gcd of 2178, 396, 792, and 726.

Concepts

12. Let us consider $\mathbb{Z}[x]$.

- a. Is $\mathbb{Z}[x]$ a UFD? Why?
 - b. Show that $\{a + xf(x) \mid a \in 2\mathbb{Z}, f(x) \in \mathbb{Z}[x]\}$ is an ideal in $\mathbb{Z}[x]$.
 - c. Is $\mathbb{Z}[x]$ a PID? (Consider part (b).)
 - d. Is $\mathbb{Z}[x]$ a Euclidean domain? Why?
13. Determine whether each of the following is true or false.
- a. Every Euclidean domain is a PID.
 - b. Every PID is a Euclidean domain.
 - c. Every Euclidean domain is a UFD.
 - d. Every UFD is a Euclidean domain.
 - e. A gcd of 2 and 3 in \mathbb{Q} is $\frac{1}{2}$.
 - f. The Euclidean algorithm gives a constructive method for finding a gcd of two integers.
 - g. If v is a Euclidean norm on a Euclidean domain D , then $v(1) \leq v(a)$ for all nonzero $a \in D$.
 - h. If v is a Euclidean norm on a Euclidean domain D , then $v(1) < v(a)$ for all nonzero $a \in D, a \neq 1$.
 - i. If v is a Euclidean norm on a Euclidean domain D , then $v(1) < v(a)$ for all nonzero nonunits $a \in D$.
 - j. For any field F , $F[x]$ is a Euclidean domain.
14. Does the choice of a particular Euclidean norm v on a Euclidean domain D influence the arithmetic structure of D in any way? Explain.

Theory

15. Let D be a Euclidean domain and let v be a Euclidean norm on D . Show that if a and b are associates in D , then $v(a) = v(b)$.
16. Let D be a Euclidean domain and let v be a Euclidean norm on D . Show that for nonzero $a, b \in D$, one has $v(a) < v(ab)$ if and only if b is not a unit of D . [Hint: Argue from Exercise 15 that $v(a) < v(ab)$ implies that b is not a unit of D . Using the Euclidean algorithm, show that $v(a) = v(ab)$ implies $\langle a \rangle = \langle ab \rangle$. Conclude that if b is not a unit, then $v(a) < v(ab)$.]
17. Prove or disprove the following statement: If v is a Euclidean norm on Euclidean domain D , then $\{a \in D \mid v(a) > v(1)\} \cup \{0\}$ is an ideal of D .
18. Show that every field is a Euclidean domain.
19. Let v be a Euclidean norm on a Euclidean domain D .
 - a. Show that if $s \in \mathbb{Z}$ such that $s + v(1) > 0$, then $\eta : D^* \rightarrow \mathbb{Z}$ defined by $\eta(a) = v(a) + s$ for nonzero $a \in D$ is a Euclidean norm on D . As usual, D^* is the set of nonzero elements of D .
 - b. Show that for $t \in \mathbb{Z}^+$, $\lambda : D^* \rightarrow \mathbb{Z}$ given by $\lambda(a) = t \cdot v(a)$ for nonzero $a \in D$ is a Euclidean norm on D .
 - c. Show that there exists a Euclidean norm μ on D such that $\mu(1) = 1$ and $\mu(a) > 100$ for all nonzero nonunits $a \in D$.
20. Let D be a UFD. An element c in D is a **least common multiple** (abbreviated lcm) of two elements a and b in D if $a \mid c, b \mid c$ and if c divides every element of D that is divisible by both a and b . Show that every two nonzero elements a and b of a Euclidean domain D have an lcm in D . [Hint: Show that all common multiples, in the obvious sense, of both a and b form an ideal of D .]
21. Use the last statement in Theorem 35.9 to show that two nonzero elements $r, s \in \mathbb{Z}$ generate the group $(\mathbb{Z}, +)$ if and only if r and s , viewed as integers in the domain \mathbb{Z} , are **relatively prime**, that is, have a gcd of 1.
22. Using the last statement in Theorem 35.9, show that for nonzero $a, b, n \in \mathbb{Z}$, the congruence $ax \equiv b \pmod{n}$ has a solution in \mathbb{Z} if a and n are relatively prime.
23. Generalize Exercise 22 by showing that for nonzero $a, b, n \in \mathbb{Z}$, the congruence $ax \equiv b \pmod{n}$ has a solution in \mathbb{Z} if and only if the positive gcd of a and n in \mathbb{Z} divides b . Interpret this result in the ring \mathbb{Z}_n .
24. Following the idea of Exercises 6 and 23, outline a constructive method for finding a solution in \mathbb{Z} of the congruence $ax \equiv b \pmod{n}$ for nonzero $a, b, n \in \mathbb{Z}$, if the congruence does have a solution. Use this method to find a solution of the congruence $22x \equiv 18 \pmod{42}$.

SECTION 36**NUMBER THEORY**

In this section we will show how the ideas in Section 35 can be used to derive some interesting results in number theory. We usually think of number theory as a study of properties of the integers, but Gauss expanded the study of numbers to include what is now called the Gaussian integers. The Gaussian integers form a subring of the complex numbers that, like the integers, form a Euclidean domain, but not a field. After studying the Gaussian integers, we will prove that for any prime number $p \in \mathbb{Z}^+$ that is equivalent to 1 modulo 4, p can be written as the sum of two squares.

Gaussian Integers

36.1 Definition A Gaussian integer is a complex number $a + bi$, where $a, b \in \mathbb{Z}$. For a Gaussian integer $\alpha = a + bi$, the norm $N(\alpha)$ of α is $a^2 + b^2$. ■

Although we defined $N(\alpha)$ for a Gaussian integer α , we can also think of N as defined on any complex number using the same formula $N(a + bi) = a^2 + b^2$. This norm can also be written as $N(\alpha) = |\alpha|^2$.

We shall let $\mathbb{Z}[i]$ be the set of all Gaussian integers. The following lemma gives some basic properties of the norm function N on $\mathbb{Z}[i]$ and leads to a demonstration that the function v defined by $v(\alpha) = N(\alpha)$ for nonzero $\alpha \in \mathbb{Z}[i]$ is a Euclidean norm on $\mathbb{Z}[i]$. Note that the Gaussian integers include all the **rational integers**, that is, all the elements of \mathbb{Z} .

■ HISTORICAL NOTE

In his *Disquisitiones Arithmeticae*, Gauss studied in detail the theory of quadratic residues, that is, the theory of solutions to the congruence $x^2 \equiv p \pmod{q}$ and proved the famous quadratic reciprocity theorem showing the relationship between the solutions of the congruences $x^2 \equiv p \pmod{q}$ and $x^2 \equiv q \pmod{p}$ where p and q are primes. In attempting to generalize his results to theories of quartic residues, however, Gauss realized that it was much more natural to consider the Gaussian integers rather than the ordinary integers.

Gauss's investigations of the Gaussian integers are contained in a long paper published in 1832 in which he proved various analogies between them and the ordinary integers. For example, after noting that there are four units (invertible elements)

among the Gaussian integers, namely $1, -1, i$, and $-i$, and defining the norm as in Definition 36.1, he generalized the notion of a prime integer by defining a prime Gaussian integer to be one that cannot be expressed as the product of two other integers, neither of them units. He was then able to determine which Gaussian integers are prime: A Gaussian integer that is not real is prime if and only if its norm is a real prime, which can only be 2 or of the form $4n + 1$. The real prime $2 = (1+i)(1-i)$ and real primes congruent to 1 modulo 4 like $13 = (2+3i)(2-3i)$ factor as the product of two Gaussian primes. Real primes of the form $4n + 3$ like 7 and 11 are still prime in the domain of Gaussian integers. See Exercise 10.

36.2 Lemma In $\mathbb{Z}[i]$, the following properties of the norm function N hold for all $\alpha, \beta \in \mathbb{Z}[i]$:

1. $N(\alpha) \geq 0$.
2. $N(\alpha) = 0$ if and only if $\alpha = 0$.
3. $N(\alpha\beta) = N(\alpha)N(\beta)$.

Proof If we let $\alpha = a_1 + a_2i$ and $\beta = b_1 + b_2i$, these results are all straightforward computations. We leave the proof of these properties as an exercise (see Exercise 11). ◆

The proof of Lemma 36.2 does not depend on the complex numbers α and β being Gaussian integers. In fact the three properties listed in the lemma are true for all complex numbers.

36.3 Lemma $\mathbb{Z}[i]$ is an integral domain.

Proof It is obvious that $\mathbb{Z}[i]$ is a commutative ring with unity. We show that there are no divisors of 0. Let $\alpha, \beta \in \mathbb{Z}[i]$. Using Lemma 36.2, if $\alpha\beta = 0$ then

$$N(\alpha)N(\beta) = N(\alpha\beta) = N(0) = 0.$$

Thus $\alpha\beta = 0$ implies that $N(\alpha) = 0$ or $N(\beta) = 0$. By Lemma 36.2 again, this implies that either $\alpha = 0$ or $\beta = 0$. Thus $\mathbb{Z}[i]$ has no divisors of 0, so $\mathbb{Z}[i]$ is an integral domain. \blacklozenge

Of course, since $\mathbb{Z}[i]$ is a subring of \mathbb{C} , where \mathbb{C} is the field of complex numbers, it is really obvious that $\mathbb{Z}[i]$ has no 0 divisors. We gave the argument of Lemma 36.3 to illustrate the use of the multiplicative property 3 of the norm function N and to avoid going outside of $\mathbb{Z}[i]$ in our argument.

However, in the proof of Theorem 36.4, we will use property 3 for complex numbers that are not Gaussian integers and therefore we will stay outside the Gaussian integers.

36.4 Theorem The function v given by $v(\alpha) = N(\alpha)$ for nonzero $\alpha \in \mathbb{Z}[i]$ is a Euclidean norm on $\mathbb{Z}[i]$. Thus $\mathbb{Z}[i]$ is a Euclidean domain.

Proof Note that for $\beta = b_1 + b_2i \neq 0$, $N(b_1 + b_2i) = b_1^2 + b_2^2$, so $N(\beta) \geq 1$. Then for all $\alpha, \beta \neq 0$ in $\mathbb{Z}[i]$, $N(\alpha) \leq N(\alpha)N(\beta) = N(\alpha\beta)$. This proves Condition 2 for a Euclidean norm in Definition 35.1.

It remains to prove the division algorithm, Condition 1, for N . Let $\alpha, \beta \in \mathbb{Z}[i]$, with $\alpha = a_1 + a_2i$ and $\beta = b_1 + b_2i$, where $\beta \neq 0$. We must find σ and ρ in $\mathbb{Z}[i]$ such that $\alpha = \beta\sigma + \rho$, where either $\rho = 0$ or $N(\rho) < N(\beta) = b_1^2 + b_2^2$. Let $\alpha/\beta = r + si$ for $r, s \in \mathbb{Q}$. Let q_1 and q_2 be integers in \mathbb{Z} as close as possible to the rational numbers r and s , respectively. Let $\sigma = q_1 + q_2i$ and $\rho = \alpha - \beta\sigma$. If $\rho = 0$, we are done. Otherwise, by construction of σ , we see that $|r - q_1| \leq \frac{1}{2}$ and $|s - q_2| \leq \frac{1}{2}$. Therefore

$$\begin{aligned} N\left(\frac{\alpha}{\beta} - \sigma\right) &= N((r + si) - (q_1 + q_2i)) \\ &= N((r - q_1) + (s - q_2)i) \leq \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}. \end{aligned}$$

Thus we obtain

$$N(\rho) = N(\alpha - \beta\sigma) = N\left(\beta\left(\frac{\alpha}{\beta} - \sigma\right)\right) = N(\beta)N\left(\frac{\alpha}{\beta} - \sigma\right) \leq N(\beta)\frac{1}{2},$$

so we do indeed have $N(\rho) < N(\beta)$ as desired. \blacklozenge

36.5 Example We can now apply all our results of Section 35 to $\mathbb{Z}[i]$. In particular, since $N(1) = 1$, the units of $\mathbb{Z}[i]$ are exactly the $\alpha = a_1 + a_2i$ with $N(\alpha) = a_1^2 + a_2^2 = 1$. From the fact that a_1 and a_2 are integers, it follows that the only possibilities are $a_1 = \pm 1$ with $a_2 = 0$, or $a_1 = 0$ with $a_2 = \pm 1$. Thus the units of $\mathbb{Z}[i]$ are ± 1 and $\pm i$. One can also use the Euclidean Algorithm to compute a gcd of two nonzero elements. We leave such computations to the exercises. Finally, note that while 5 is an irreducible in \mathbb{Z} , 5 is no longer an irreducible in $\mathbb{Z}[i]$, for $5 = (1 + 2i)(1 - 2i)$, and neither $1 + 2i$ nor $1 - 2i$ is a unit. \blacktriangle

Multiplicative Norms

Let us point out again that for an integral domain D , the arithmetic concepts of irreducibles and units are not affected in any way by a norm that may be defined on the domain. However, as the preceding section and our work thus far in this section show, a suitably defined norm may be of help in determining the arithmetic structure of D . This is strikingly illustrated in *algebraic number theory*, where for a domain of *algebraic integers* we consider many different norms of the domain, each doing its part in helping to uncover the arithmetic structure of the domain. In a domain of algebraic integers, we have essentially one norm for each irreducible (up to associates), and each such norm gives information concerning the behavior in the integral domain of the irreducible to which it corresponds. This is an example of the importance of studying properties of elements in an algebraic structure by means of mappings associated with them.

Let us study integral domains that have a multiplicative norm satisfying Properties 2 and 3 of N on $\mathbb{Z}[i]$ given in Lemma 36.2.

36.6 Definition Let D be an integral domain. A **multiplicative norm N on D** is a function mapping D into the integers \mathbb{Z} such that the following conditions are satisfied:

1. $N(\alpha) = 0$ if and only if $\alpha = 0$.
2. $N(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in D$.

■

36.7 Theorem If D is an integral domain with a multiplicative norm N , then $N(1) = 1$ and $|N(u)| = 1$ for every unit u in D . If, furthermore, every α such that $|N(\alpha)| = 1$ is a unit in D , then an element π in D , with $|N(\pi)| = p$ for a prime $p \in \mathbb{Z}$, is an irreducible of D .

Proof Let D be an integral domain with a multiplicative norm N . Then

$$N(1) = N((1)(1)) = N(1)N(1)$$

shows that $N(1) = 1$. Also, if u is a unit in D , then

$$1 = N(1) = N(uu^{-1}) = N(u)N(u^{-1}).$$

Since $N(u)$ is an integer, this implies that $|N(u)| = 1$.

Now suppose that the units of D are exactly the elements of norm ± 1 . Let $\pi \in D$ be such that $|N(\pi)| = p$, where p is a prime in \mathbb{Z} . Then if $\pi = \alpha\beta$, we have

$$p = |N(\pi)| = |N(\alpha)N(\beta)| = |N(\alpha)||N(\beta)|,$$

so either $|N(\alpha)| = 1$ or $|N(\beta)| = 1$. By assumption, this means that either α or β is a unit of D . Thus π is an irreducible of D . ◆

36.8 Example On $\mathbb{Z}[i]$, the function N defined by $N(a + bi) = a^2 + b^2$ gives a multiplicative norm in the sense of our definition. We saw that the function v given by $v(\alpha) = N(\alpha)$ for nonzero $\alpha \in \mathbb{Z}[i]$ is a Euclidean norm on $\mathbb{Z}[i]$, so the units are precisely the elements α of $\mathbb{Z}[i]$ with $N(\alpha) = N(1) = 1$. Thus the second part of Theorem 36.7 applies in $\mathbb{Z}[i]$. We saw in Example 36.5 that 5 is not an irreducible in $\mathbb{Z}[i]$, for $5 = (1 + 2i)(1 - 2i)$. Since $N(1 + 2i) = N(1 - 2i) = 1^2 + 2^2 = 5$ and 5 is a prime in \mathbb{Z} , we see from Theorem 36.7 that $1 + 2i$ and $1 - 2i$ are both irreducibles in $\mathbb{Z}[i]$. ▲

As an application of multiplicative norms, we shall now give another example of an integral domain that is *not* a UFD. We saw one example in Example 34.17. The following is the standard illustration.

36.9 Example Let $\mathbb{Z}[\sqrt{-5}] = \{a + ib\sqrt{-5} \mid a, b \in \mathbb{Z}\}$. As a subset of the complex numbers closed under addition, subtraction, and multiplication, and containing 0 and 1, $\mathbb{Z}[\sqrt{-5}]$ is an integral

domain. Define N on $\mathbb{Z}[\sqrt{-5}]$ by

$$N(a + b\sqrt{-5}) = a^2 + 5b^2.$$

(Here $\sqrt{-5} = i\sqrt{5}$.) Clearly, $N(\alpha) = 0$ if and only if $\alpha = a + b\sqrt{-5} = 0$. That $N(\alpha\beta) = N(\alpha)N(\beta)$ is a straightforward computation that we leave to the exercises (see Exercise 12). Let us find all candidates for units in $\mathbb{Z}[\sqrt{-5}]$ by finding all elements α in $\mathbb{Z}[\sqrt{-5}]$ with $N(\alpha) = 1$. If $\alpha = a + b\sqrt{-5}$, and $N(\alpha) = 1$, we must have $a^2 + 5b^2 = 1$ for integers a and b . This is possible only if $b = 0$ and $a = \pm 1$. Hence ± 1 are the only candidates for units. Since ± 1 are units, they are then precisely the units in $\mathbb{Z}[\sqrt{-5}]$.

Now in $\mathbb{Z}[\sqrt{-5}]$, we have $21 = (3)(7)$ and also

$$21 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5}).$$

If we can show that 3 , 7 , $1 + 2\sqrt{-5}$, and $1 - 2\sqrt{-5}$ are all irreducibles in $\mathbb{Z}[\sqrt{-5}]$, we will then know that $\mathbb{Z}[\sqrt{-5}]$ cannot be a UFD, since neither 3 nor 7 is $\pm(1 + 2\sqrt{-5})$.

Suppose that $3 = \alpha\beta$. Then

$$9 = N(3) = N(\alpha)N(\beta)$$

shows that we must have $N(\alpha) = 1$, 3 , or 9 . If $N(\alpha) = 1$, then α is a unit. If $\alpha = a + b\sqrt{-5}$, then $N(\alpha) = a^2 + 5b^2$, and for no choice of integers a and b is $N(\alpha) = 3$. If $N(\alpha) = 9$, then $N(\beta) = 1$, so β is a unit. Thus from $3 = \alpha\beta$, we can conclude that either α or β is a unit. Therefore, 3 is an irreducible in $\mathbb{Z}[\sqrt{-5}]$. A similar argument shows that 7 is also an irreducible in $\mathbb{Z}[\sqrt{-5}]$.

If $1 + 2\sqrt{-5} = \gamma\delta$, we have

$$21 = N(1 + 2\sqrt{-5}) = N(\gamma)N(\delta).$$

so $N(\gamma) = 1$, 3 , 7 , or 21 . We have seen that there is no element of $\mathbb{Z}[\sqrt{-5}]$ of norm 3 or 7 . Thus either $N(\gamma) = 1$, and γ is a unit, or $N(\gamma) = 21$, so $N(\delta) = 1$, and δ is a unit. Therefore, $1 + 2\sqrt{-5}$ is an irreducible in $\mathbb{Z}[\sqrt{-5}]$. A parallel argument shows that $1 - 2\sqrt{-5}$ is also an irreducible in $\mathbb{Z}[\sqrt{-5}]$.

In summary, we have shown that

$$\mathbb{Z}[\sqrt{-5}] = \{a + ib\sqrt{5} \mid a, b \in \mathbb{Z}\}$$

is an integral domain but not a UFD. In particular, there are two different factorizations

$$21 = 3 \cdot 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$$

of 21 into irreducibles. These irreducibles cannot be primes, for the property of a prime enables us to prove uniqueness of factorization (see the proof of Theorem 34.18). \blacktriangle

We conclude with a classical application, determining which primes p in \mathbb{Z} are equal to a sum of squares of two integers in \mathbb{Z} . For example, $2 = 1^2 + 1^2$, $5 = 1^2 + 2^2$, and $13 = 2^2 + 3^2$ are sums of squares. Since we have now answered this question for the only even prime number, 2 , we can restrict ourselves to odd primes.

36.10 Theorem (Fermat's $p = a^2 + b^2$ Theorem) Let p be an odd prime in \mathbb{Z} . Then $p = a^2 + b^2$ for integers a and b in \mathbb{Z} if and only if $p \equiv 1 \pmod{4}$.

Proof First, suppose that $p = a^2 + b^2$. Now a and b cannot both be even or both be odd since p is an odd number. If $a = 2r$ and $b = 2s + 1$, then $a^2 + b^2 = 4r^2 + 4(s^2 + s) + 1$, so $p \equiv 1 \pmod{4}$. This takes care of one direction for this "if and only if" theorem.

For the other direction, we assume that $p \equiv 1 \pmod{4}$. Now the multiplicative group of nonzero elements of the finite field \mathbb{Z}_p is cyclic, and has order $p - 1$. Since

4 is a divisor of $p - 1$, we see that \mathbb{Z}_p contains an element n of multiplicative order 4. It follows that n^2 has multiplicative order 2, so $n^2 = -1$ in \mathbb{Z}_p . Thus in \mathbb{Z} , we have $n^2 \equiv -1 \pmod{p}$, so p divides $n^2 + 1$ in \mathbb{Z} .

Viewing p and $n^2 + 1$ in $\mathbb{Z}[i]$, we see that p divides $n^2 + 1 = (n+i)(n-i)$. Suppose that p is irreducible in $\mathbb{Z}[i]$; then p would have to divide $n+i$ or $n-i$. If p divides $n+i$, then $n+i = p(a+bi)$ for some $a, b \in \mathbb{Z}$. Equating coefficients of i , we obtain $1 = pb$, which is impossible. Similarly, p divides $n-i$ would lead to an impossible equation $-1 = pb$. Thus our assumption that p is irreducible in $\mathbb{Z}[i]$ must be false.

Since p is not irreducible in $\mathbb{Z}[i]$, we have $p = (a+bi)(c+di)$ where neither $a+bi$ nor $c+di$ is a unit. Taking norms, we have $p^2 = (a^2+b^2)(c^2+d^2)$ where neither $a^2+b^2 = 1$ nor $c^2+d^2 = 1$. Consequently, we have $p = a^2+b^2$, which completes our proof. [Since $a^2+b^2 = (a+bi)(a-bi)$, we see that this is the factorization of p , that is, $c+di = a-bi$.] \blacklozenge

Exercise 10 asks you to determine which primes p in \mathbb{Z} remain irreducible in $\mathbb{Z}[i]$.

■ EXERCISES 36

Computations

In Exercises 1 through 4, factor the Gaussian integer into a product of irreducibles in $\mathbb{Z}[i]$. [Hint: Since an irreducible factor of $\alpha \in \mathbb{Z}[i]$ must have norm > 1 and dividing $N(\alpha)$, there are only a finite number of Gaussian integers $a+bi$ to consider as possible irreducible factors of a given α . Divide α by each of them in \mathbb{C} , and see for which ones the quotient is again in $\mathbb{Z}[i]$.]

1. 5

2. 7

3. $4 + 3i$

4. $6 - 7i$

5. Show that 6 does not factor uniquely (up to associates) into irreducibles in $\mathbb{Z}[\sqrt{-5}]$. Exhibit two different factorizations.

6. Consider $\alpha = 7 + 2i$ and $\beta = 3 - 4i$ in $\mathbb{Z}[i]$. Find σ and ρ in $\mathbb{Z}[i]$ such that

$$\alpha = \beta\sigma + \rho \quad \text{with} \quad N(\rho) < N(\beta).$$

[Hint: Use the construction in the proof of Theorem 36.4.]

7. Use a Euclidean algorithm in $\mathbb{Z}[i]$ to find a gcd of $8 + 6i$ and $5 - 15i$ in $\mathbb{Z}[i]$. [Hint: Use the construction in the proof of Theorem 36.4.]

Concepts

8. Determine whether each of the following is true or false.

- a. $\mathbb{Z}[i]$ is a PID.
 - b. $\mathbb{Z}[i]$ is a Euclidean domain.
 - c. Every integer in \mathbb{Z} is a Gaussian integer.
 - d. Every complex number is a Gaussian integer.
 - e. A Euclidean algorithm holds in $\mathbb{Z}[i]$.
 - f. A multiplicative norm on an integral domain is sometimes an aid in finding irreducibles of the domain.
 - g. If N is a multiplicative norm on an integral domain D , then $|N(u)| = 1$ for every unit u of D .
 - h. If F is a field, then the function N defined by $N(f(x)) = (\text{degree of } f(x))$ is a multiplicative norm on $F[x]$.
 - i. If F is a field, then the function defined by $N(f(x)) = 2^{(\text{degree of } f(x))}$ for $f(x) \neq 0$ and $N(0) = 0$ is a multiplicative norm on $F[x]$ according to our definition.
 - j. $\mathbb{Z}[\sqrt{-5}]$ is an integral domain but not a UFD.
9. Let D be an integral domain with a multiplicative norm N such that $|N(\alpha)| = 1$ for $\alpha \in D$ if and only if α is a unit of D . Let π be such that $|N(\pi)|$ is minimal among all $|\mathcal{N}(\beta)| > 1$ for $\beta \in D$. Show that π is an irreducible of D .

- 10.** **a.** Show that 2 is equal to the product of a unit and the square of an irreducible in $\mathbb{Z}[i]$.
b. Show that an odd prime p in \mathbb{Z} is irreducible in $\mathbb{Z}[i]$ if and only if $p \equiv 3 \pmod{4}$. (Use Theorem 36.10.)
- 11.** Prove Lemma 36.2.
- 12.** Prove that N of Example 36.9 is multiplicative, that is, that $N(\alpha\beta) = N(\alpha)N(\beta)$ for $\alpha, \beta \in \mathbb{Z}[\sqrt{-5}]$.
- 13.** Let D be an integral domain with a multiplicative norm N such that $|N(\alpha)| = 1$ for $\alpha \in D$ if and only if α is a unit of D . Show that every nonzero nonunit of D has a factorization into irreducibles in D .
- 14.** Use a Euclidean algorithm in $\mathbb{Z}[i]$ to find a gcd of $16 + 7i$ and $10 - 5i$ in $\mathbb{Z}[i]$. [Hint: Use the construction in the proof of Theorem 36.4.]
- 15.** Let $\langle \alpha \rangle$ be a nonzero principal ideal in $\mathbb{Z}[i]$.
a. Show that $\mathbb{Z}[i]/\langle \alpha \rangle$ is a finite ring. [Hint: Use the division algorithm.]
b. Show that if σ is an irreducible of $\mathbb{Z}[i]$, then $\mathbb{Z}[i]/\langle \sigma \rangle$ is a field.
c. Referring to part (b), find the order and characteristic of each of the following fields.
i. $\mathbb{Z}[i]/(3)$ **ii.** $\mathbb{Z}[i]/(1+i)$ **iii.** $\mathbb{Z}[i]/(1+2i)$
- 16.** Let $n \in \mathbb{Z}^+$ be square free, that is, not divisible by the square of any prime integer. Let $\mathbb{Z}[\sqrt{-n}] = \{a + ib\sqrt{n} \mid a, b \in \mathbb{Z}\}$.
a. Show that the norm N , defined by $N(\alpha) = a^2 + nb^2$ for $\alpha = a + ib\sqrt{n}$, is a multiplicative norm on $\mathbb{Z}[\sqrt{-n}]$.
b. Show that $N(\alpha) = 1$ for $\alpha \in \mathbb{Z}[\sqrt{-n}]$ if and only if α is a unit of $\mathbb{Z}[\sqrt{-n}]$.
c. Show that every nonzero $\alpha \in \mathbb{Z}[\sqrt{-n}]$ that is not a unit has a factorization into irreducibles in $\mathbb{Z}[\sqrt{-n}]$. [Hint: Use part (b).]
- 17.** Repeat Exercise 16 for $\mathbb{Z}[\sqrt{n}] = \{a + b\sqrt{n} \mid a, b \in \mathbb{Z}\}$ for square free $n > 1$, with N defined by $N(\alpha) = a^2 - nb^2$ for $\alpha = a + b\sqrt{n}$ in $\mathbb{Z}[\sqrt{n}]$. For part b show $|N(\alpha)| = 1$.
- 18.** Show by a construction analogous to that given in the proof of Theorem 36.4 that the division algorithm holds in the integral domain $\mathbb{Z}[\sqrt{-2}]$ for $v(\alpha) = N(\alpha)$ for nonzero α in this domain (see Exercise 16). (Thus this domain is Euclidean. See Hardy and Wright [29] for a discussion of which domains $\mathbb{Z}[\sqrt{n}]$ and $\mathbb{Z}[\sqrt{-n}]$ are Euclidean.)

SECTION 37**† ALGEBRAIC GEOMETRY**

This section gives a brief introduction to algebraic geometry. Algebraic geometry is the study of the common zeros of a finite collection of polynomials. For example, the zeros of the set of polynomials $\{x^2 + y^2 - 25, (x - 6)^2 + y^2 - 25\}$ consist of just two points in \mathbb{R}^2 , $(3, 4)$ and $(3, -4)$. In Section 38 we will develop a very useful algorithm that reduces a finite set of polynomials to a simpler set of polynomials whose zeros are identical to the zeros of the original set. In the example $\{x^2 + y^2 - 25, (x - 6)^2 + y^2 - 25\}$, the algorithm yields $\{x - 3, y^2 - 16\}$ making it much easier to see the two zeros.

Algebraic Varieties and Ideals

Let F be a field. Recall that $F[x_1, x_2, \dots, x_n]$ is the ring of polynomials in n indeterminants x_1, x_2, \dots, x_n with coefficients in F . We let F^n be the Cartesian product $F \times F \times \dots \times F$ for n factors. For ease in writing, we denote an element (a_1, a_2, \dots, a_n) of F^n by \mathbf{a} , in bold type. Using similar economy, we let $F[\mathbf{x}] = F[x_1, x_2, \dots, x_n]$. For each $\mathbf{a} \in F^n$, we have an evaluation homomorphism $\phi_{\mathbf{a}}: F[\mathbf{x}] \rightarrow F$ just as in Theorem 27.4. That is, for $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) \in F[\mathbf{x}]$, we define $\phi_{\mathbf{a}}(f(\mathbf{x})) = f(\mathbf{a}) = f(a_1, a_2, \dots, a_n)$.

† This section is used only in Section 38.

The proof that ϕ_a is indeed a homomorphism follows from the associative, commutative, and distributive properties of the operations in $F[x]$ and F . Just as for the one-indeterminate case, an element a of F^n is a **zero** of $f(x) \in F[x]$ if $f(a) = 0$. In what follows, we further abbreviate a polynomial $f(x)$ by “ f .”

In this section and the following we discuss the problem of finding common zeros in F^n of a finite number of polynomials f_1, f_2, \dots, f_r in $F[x]$. Finding and studying geometric properties of the set of all these common zeros is the subject of algebraic geometry.

37.1 Definition Let S be a finite subset of $F[x]$. The **algebraic variety** $V(S)$ in F^n is the set of all common zeros in F^n of the polynomials in S . ■

In our illustrative examples, which usually involve at most three indeterminates, we use x, y, z in place of x_1, x_2 , and x_3 .

37.2 Example Let $S = \{2x + y - 2\} \subset \mathbb{R}[x, y]$. The algebraic variety $V(S)$ in \mathbb{R}^2 is the line with x -intercept 1 and y -intercept 2. ▲

We leave to Exercise 14 the straightforward proof that for r elements f_1, f_2, \dots, f_r in a commutative ring R with unity, the set

$$I = \{c_1f_1 + c_2f_2 + \dots + c_rf_r \mid c_i \in R \text{ for } i = 1, \dots, r\}$$

is an ideal of R . We denote this ideal by $\langle f_1, f_2, \dots, f_r \rangle$. We are interested in the case $R = F[x]$ where all the c_i and all the f_i are polynomials in $F[x]$. We regard the c_i as “coefficient polynomials.” By its construction, this ideal I is the smallest ideal containing the polynomials f_1, f_2, \dots, f_r ; it can also be described as the intersection of all ideals containing these r polynomials.

37.3 Definition Let I be an ideal in a commutative ring R with unity. A subset $\{b_1, b_2, \dots, b_r\}$ of I is a **basis** for I if $I = \langle b_1, b_2, \dots, b_r \rangle$. ■

Unlike the situation in linear algebra, there is no requirement of independence for elements of a basis, or of unique representation of an ideal member in terms of a basis.

37.4 Theorem Let $f_1, f_2, \dots, f_r \in F[x]$. The set of common zeros in F^n of the polynomials f_i for $i = 1, 2, \dots, r$ is the same as the set of common zeros in F^n of all the polynomials in the entire ideal $I = \langle f_1, f_2, \dots, f_r \rangle$.

Proof

Let

$$f = c_1f_1 + c_2f_2 + \dots + c_rf_r \quad (1)$$

be any element of I , and let $a \in F^n$ be a common zero of f_1, f_2, \dots, f_r . Applying the evaluation homomorphism ϕ_a to Eq. (1), we obtain

$$\begin{aligned} f(a) &= c_1(a)f_1(a) + c_2(a)f_2(a) + \dots + c_r(a)f_r(a) \\ &= c_1(a)0 + c_2(a)0 + \dots + c_r(a)0 = 0, \end{aligned}$$

showing that a is also a zero of every polynomial f in I . Of course, a zero of every polynomial in I will be a zero of each f_i because each $f_i \in I$. ◆

For an ideal I in $F[x]$, we let $V(I)$ be the set of all common zeros of all elements of I . We can summarize Theorem 37.4 as

$$V(\{f_1, f_2, \dots, f_r\}) = V(\langle f_1, f_2, \dots, f_r \rangle).$$

Recall that a commutative ring with unity is Noetherian if for every chain $N_1 \subseteq N_2 \subseteq N_3 \subseteq \dots$ of ideals in R , there is an integer r such that for $s \geq r$, $N_r = N_s$. Lemma 34.10

states that if R is a PID, then R is a Noetherian ring. Theorem 37.5 states that if R is a Noetherian ring, then polynomials with coefficients in R also form a Noetherian ring.

37.5 Example If R is a Noetherian ring, then $R[x]$ is also a Noetherian ring.

Proof Let $I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$ be a chain of ideals in $R[x]$. As we saw in Lemma 34.9, $I = \bigcup_{n=1}^{\infty} I_n$ is an ideal in R .

We will show by contradiction that I has a finite basis. So we suppose that no finite set of polynomials is a basis for I . We let f_1 be a polynomial in I of minimal degree. We then let f_2 be a polynomial of minimum degree that is in I , but not in $\langle f_1 \rangle$. Continuing in this manner, we let f_n be a polynomial of minimal degree in I , but not in $\langle f_1, f_2, \dots, f_{n-1} \rangle$. This defines an infinite sequence of polynomials since by our assumption no finite set of polynomials is a basis for I . It is clear that $\deg(f_1) \leq \deg(f_2) \leq \dots$ since otherwise we should have picked f_2 instead of f_1 as the first polynomial in the sequence. In general, $\deg(f_1) \leq \deg(f_2) \leq \deg(f_3) \leq \dots$

We let a_j be the leading coefficient of f_j . Then in R , we have a chain of ideals

$$\langle a_1 \rangle \subseteq \langle a_1, a_2 \rangle \subseteq \langle a_1, a_2, a_3 \rangle \subseteq \dots$$

By the ascending chain condition in R , there is some integer N so that for any $s \geq N$,

$$\langle a_1, a_2, \dots, a_N \rangle = \langle a_1, a_2, \dots, a_s \rangle.$$

In particular,

$$a_{N+1} = \sum_{j=1}^N c_j a_j$$

for some elements c_j in R . The polynomial

g(x) = \sum_{j=1}^N c_j x^{\deg(f_{N+1}) - \deg(f_j)} f_j

is in the ideal $\langle f_1, f_2, f_3, \dots, f_N \rangle$. Thus $f_{N+1} - g \notin \langle f_1, f_2, \dots, f_N \rangle = J$. The degrees of g and f_{N+1} are equal and they have the same leading coefficient a_{N+1} . Thus the degree of $f_{N+1} - g$ is less than the degree of f_{N+1} . But this contradicts the choice of f_{N+1} since $f_{N+1} - g \notin J$. $f_N - g$ has lower degree than f_{N+1} , and f_{N+1} has lowest degree among all the polynomials in I that are not in $\langle f_1(x), f_2(x), \dots, f_N(x) \rangle$. We conclude that there is a finite set of polynomials with $I = \langle f_1, f_2, \dots, f_n \rangle$.

Since every polynomial f_j is in some I_k , and $I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$, it follows that there is an integer r such that f_j is in I_r for each j . Therefore, $I = \langle f_1, f_2, f_3, \dots, f_n \rangle = I_r$. ◆

By repeated application of Theorem 37.5, it is immediate that $F[x_1, x_2, \dots, x_n]$ is a Noetherian ring for any field F . Exercise 21 shows that if R is a commutative ring with unity, then R is a Noetherian ring if and only if every ideal I in R has a finite basis. These observations prove the following significant theorem.

37.6 Theorem (Hilbert Basis Theorem) Every ideal I in $F[x_1, x_2, \dots, x_n]$ has a finite basis. ◆

Our objective: Given a basis for an ideal I in $F[\mathbf{x}]$, modify it if possible to become a basis that better exhibits the structure of I and the geometry of the associated algebraic variety $V(I)$.

The theorem that follows provides a tool for this task. You should notice that the theorem gives information about the division algorithm that we did not mention in Theorem 28.2. We use the same notation here as in Theorem 28.2, but with x rather than x . If $f(x) = g(x)h(x)$ in $F[x]$, then $g(x)$ and $h[x]$ are called “**divisors**” or “**factors**” of $f(x)$.

37.7 Theorem (Property of the Division Algorithm) Let $f(x), g(x), q(x)$, and $r(x)$ be polynomials in $F[x]$ such that $f(x) = g(x)q(x) + r(x)$. The common zeros in F^n of $f(x)$ and $g(x)$ are the same as the common zeros of $g(x)$ and $r(x)$. Also the common divisors in $F[x]$ of $f(x)$ and $g(x)$ are the same as the common divisors of $g(x)$ and $r(x)$.

If $f(x)$ and $g(x)$ are two members of a basis for an ideal I of $F[x]$, then replacement of $f(x)$ by $r(x)$ in the basis still yields a basis for I .

Proof If $a \in F^n$ is a common zero of $g(x)$ and $r(x)$, then applying ϕ_a to both sides of the equation $f(x) = g(x)q(x) + r(x)$, we obtain $f(a) = g(a)q(a) + r(a) = 0q(a) + 0 = 0$, so a is a zero of both $f(x)$ and $g(x)$. If $b \in F[x]$ is a common zero of $f(x)$ and $g(x)$, then applying ϕ_b yields $f(b) = g(b)q(b) + r(b)$ so $0 = 0q(b) + r(b)$ and we see that $r(b) = 0$ and $g(b) = 0$.

The proof concerning common divisors is essentially the same, and is left as Exercise 15.

Finally, let B be a basis for an ideal I , let $f(x), g(x) \in B$, and let $f(x) = g(x)q(x) + r(x)$. Let B' be the set obtained by replacing $f(x)$ by $r(x)$ in B , and let I' be the ideal having B' as a basis. Let S be the set obtained from B by adjoining $r(x)$ to B . Note that S can also be obtained by adjoining $f(x)$ to B' . The equation $f(x) = g(x)q(x) + r(x)$ shows that $f(x) \in I'$, so we have $B' \subseteq S \subseteq I'$. Thus S is a basis for I' . The equation $r(x) = f(x) - q(x)g(x)$ shows that $r(x) \in I$, so we have $B \subseteq S \subseteq I$. Thus S is basis for I . Therefore $I = I'$ and B' is a basis for I . ◆

A Familiar Linear Illustration

A basic technique for problem solving in linear algebra is finding all common solutions of a finite number of linear equations. For the moment we abandon our practice of never writing “ $f(x) = 0$ ” for a nonzero polynomial, and work a typical problem as we do in a linear algebra course.

37.8 Example (Solution as in a Linear Algebra Course) Find all solutions in \mathbb{R}^3 of the linear system

$$\begin{aligned} x + y - 3z &= 8 \\ 2x + y + z &= -5. \end{aligned}$$

Solution We multiply the first equation by -2 and add it to the second, obtaining the new system

$$\begin{aligned} x + y - 3z &= 8 \\ -y + 7z &= -21 \end{aligned}$$

which has the same solution set in \mathbb{R}^3 as the preceding one. For any value z , we can find the corresponding y -value from the second equation and then determine x from the first equation. Keeping z as parameter, we obtain $\{(-4z - 13, 7z + 21, z) \mid z \in \mathbb{R}\}$ as solution set, which is a line in Euclidean 3-space through the point $(-13, 21, 0)$. ▲

In the notation of this section, the problem in the preceding example can be phrased as follows:

$$\text{Describe } V(\langle x + y - 3z - 8, 2x + y + z + 5 \rangle) \text{ in } \mathbb{R}^3.$$

We solved it by finding a more useful basis, namely

$$\{x + y - 3z - 8, -y + 7z + 21\}.$$

Notice that the second member, $-y + 7z + 21$, of this new basis can be obtained from the original two basis polynomials as a remainder $r(x, y, z)$ in a division process, namely

$$\begin{array}{r} 2 \\ \boxed{x + y - 3z - 8} \end{array} \overline{\left[\begin{array}{r} 2x + y + z + 5 \\ 2x + 2y - 6z - 16 \\ \hline -y + 7z + 21 \end{array} \right]}$$

Thus $2x + y + z + 5 = (x + y - 3z - 8)(2) + (-y + 7z + 21)$, an expression of the form $f(x, y, z) = g(x, y, z)q(x, y, z) + r(x, y, z)$. We replaced the polynomial f by the polynomial r , as in Theorem 37.7, which assures us that $V(\langle f, g \rangle) = V(\langle g, r \rangle)$ and that $\langle f, g \rangle = \langle g, r \rangle$. We chose a very simple, 1-step problem in Example 37.8. However, it is clear that the method introduced in a linear algebra course for solving a linear system can be phrased in terms of applying a division algorithm process repeatedly to change a given ideal basis into one that better illuminates the geometry of the associated algebraic variety.

A Single Indeterminate Illustration

Suppose now that we want to find the variety $V(I)$ in \mathbb{R} associated with an ideal I in $F[x]$, the ring of polynomials in the single indeterminate x . By Theorem 31.24, every ideal in $F[x]$ is principal, so there exists $f(x) \in F[x]$ such that $I = \langle f(x) \rangle$. Thus $V(I)$ consists of the zeros of a single polynomial, and $\{f(x)\}$ is probably as simple a basis for I as we could desire. We give an example illustrating computation of such a single generator $f(x)$ for I in a case where the given basis for I contains more than one polynomial. Because a polynomial in $\mathbb{R}[x]$ has only a finite number of zeros in \mathbb{R} , we expect two or more randomly selected polynomials in $\mathbb{R}[x]$ to have no common zeros, but we constructed the basis in our example carefully!

37.9 Example Let us describe the algebraic variety V in \mathbb{R} consisting of common zeros of

$$f(x) = x^4 + x^3 - 3x^2 - 5x - 2 \quad \text{and} \quad g(x) = x^3 + 3x^2 - 6x - 8.$$

We want to find a new basis for $\langle f, g \rangle$ having polynomials of as small degree as possible, so we use the division algorithm $f(x) = g(x)q(x) + r(x)$ in Theorem 28.2, where $r(x)$ will have degree at most 2. We then replace the basis $\{f, g\}$ by the basis $\{g, r\}$.

$$\begin{array}{r} x - 2 \\ \boxed{x^3 + 3x^2 - 6x - 8} \end{array} \overline{\left[\begin{array}{r} x^4 + x^3 - 3x^2 - 5x - 2 \\ x^4 + 3x^3 - 6x^2 - 8x \\ \hline -2x^3 + 3x^2 + 3x - 2 \\ -2x^3 - 6x^2 + 12x + 16 \\ \hline 9x^2 - 9x - 18 \end{array} \right]}$$

Because zeros of $9x^2 - 9x - 18$ are the same as zeros of $x^2 - x - 2$, we let $r(x) = x^2 - x - 2$, and take as new basis

$$\{g, r\} = (x^3 + 3x^2 - 6x - 8, x^2 - x - 2).$$

By dividing $g(x)$ by $r(x)$ to obtain a remainder $r_1(x)$, we will now be able to find a basis $\{r(x), r_1(x)\}$ consisting of polynomials of degree at most 2.

$$\begin{array}{r} x+4 \\ \hline x^2-x-2 \left| \begin{array}{r} x^3+3x^2-6x-8 \\ x^3-x^2-2x \\ \hline 4x^2-4x-8 \\ 4x^2-4x-8 \\ \hline 0 \end{array} \right. \end{array}$$

Our new basis $\{r(x), r_1(x)\}$ now becomes $\{x^2 - x - 2\}$. Thus $I = \langle f(x), g(x) \rangle = \langle x^2 - x - 2 \rangle = \langle (x-2)(x+1) \rangle$, and we see that $V = \{-1, 2\}$. \blacktriangle

Theorem 37.7 tells us that the common divisors of $f(x)$ and $g(x)$ in the preceding example are the same as the common divisors of $r(x)$ and $r_1(x)$. Because $0 = (0)r(x)$, we see that $r(x)$ itself divides 0, so the common divisors of $f(x)$ and $g(x)$ are just those of $r(x)$, which, of course, include $r(x)$ itself. Thus $r(x)$ is called a “*greatest common divisor*” (abbreviated gcd) of $f(x)$ and $g(x)$.

■ EXERCISES 37

In Exercises 1–4 find a basis for the given ideals in $\mathbb{R}[x, y]$.

1. The set of polynomials with constant 0.
2. The kernel of the evaluation homomorphism $\phi_{(2,3)} : \mathbb{R}[x, y] \rightarrow \mathbb{R}$.
3. The kernel of the evaluation homomorphism $\phi_{(-4,5)} : \mathbb{R}[x, y] \rightarrow \mathbb{R}$.
4. The set of all polynomials with zeros on the circle centered at the origin with radius 1.

In Exercises 5–8, use the techniques from Examples 37.8 and 37.9 to find a simpler basis for the ideal where the field is \mathbb{R} . Describe the algebraic variety associated with the ideal.

5. $I = \langle x + y + z, 2x + y + 3z - 4 \rangle$
6. $I = \langle 3x + 4y + 7z - 10, 2x + 3y - 2z + 1 \rangle$
7. $I = \langle x^4 + 5x^3 + 3x^2 - 7x - 2, x^3 + 6x^2 + 3x - 10 \rangle$
8. $I = \langle x^6 - x^5 - 6x^4 + 3x^3 - 8x^2 - 4x + 3, x^3 - 2x^2 - 9 \rangle$
9. Describe the algebraic variety for the ideal $\{0\}$ in $F[x, y]$.
10. Describe the algebraic variety for the ideal $\{1\}$ in $F[x, y]$.
11. Describe the algebraic variety in F for the ideal $\langle x^2 + 1 \rangle$ a) for $F = \mathbb{R}$ and b) for $F = \mathbb{C}$.
12. Compare the algebraic varieties for the ideals $I = \langle x^2 + 4xy + 4y^2 \rangle$ and $J = \langle x + 2y \rangle$.

Concepts

13. Determine whether each of the following is true or false.
 - a. Every ideal in $F[x]$ has a finite basis.
 - b. Every subset of \mathbb{R}^2 is an algebraic variety.
 - c. The empty subset of \mathbb{R}^2 is an algebraic variety.
 - d. Every finite subset of \mathbb{R}^2 is an algebraic variety.
 - e. Every line in \mathbb{R}^2 is an algebraic variety.
 - f. Every finite collection of lines in \mathbb{R}^2 is an algebraic variety.
 - g. A greatest common divisor of a finite number of polynomials in $\mathbb{R}[x]$ (one indeterminate) can be computed using the division algorithm repeatedly.

- h.** In the context of ideals in a commutative ring with unity, elements in a basis are independent.
- i.** If R is a Noetherian ring, then so is $R[x]$.
- j.** The ideals $\langle x, y \rangle$ and $\langle x^2, y^2 \rangle$ are equal because they both yield the same algebraic variety, namely $\{(0, 0)\}$, in \mathbb{R}^2 .

Theory

- 14.** Show that if f_1, f_2, \dots, f_r are elements of a commutative ring R with unity, then $I = \{c_1f_1 + c_2f_2 + \dots + c_rf_r \mid c_i \in R \text{ for } i = 1, \dots, r\}$ is an ideal of R .
- 15.** Show that if $f(x) = g(x)q(x) + r(x)$ in $F[x]$, then the common divisors in $F[x]$ of $f(x)$ and $g(x)$ are the same as the common divisors in $F[x]$ of $g(x)$ and $r(x)$.
- 16.** Let F be a field. Show that if S is a nonempty subset of F^n , then

$$I(S) = \{f(x) \in F[x] \mid f(s) = 0 \text{ for all } s \in S\}$$

is an ideal of $F[x]$.

- 17.** Referring to Exercise 16, show that $S \subseteq V(I(S))$.
- 18.** Referring to Exercise 16, give an example of a subset S of \mathbb{R}^2 such that $V(I(S)) \neq S$.
- 19.** Referring to Exercise 16, show that if N is an ideal of $F[x]$, then $N \subseteq I(V(N))$.
- 20.** Referring to Exercise 16, give an example of an ideal N in $\mathbb{R}[x, y]$ such that $I(V(N)) \neq N$.
- 21.** Prove for R a commutative ring with unity, R is a Noetherian ring if and only if every ideal in R has a finite basis.

SECTION 38 [†]GRÖBNER BASES FOR IDEALS

We tackle the problem of finding a nice basis for an ideal I in $F[x] = F[x_1, x_2, \dots, x_n]$. In view of our Section 37 illustrations for the linear and single indeterminant cases, it seems reasonable to try to replace polynomials in a basis by polynomials of lower degree, or containing fewer indeterminates. It is crucial to have a systematic way to accomplish this. As you probably learned in linear algebra, when row reducing matrices, it is important to follow the standard order for which matrix entries you make zero. If you make an entry in the second column zero before dealing with the first column, you may have wasted your time. As a first step in our goal, we tackle this problem of specifying an order for polynomials in a basis.

Ordering Power Products

Our polynomials in $F[x]$ have terms of the form $ax_1^{m_1}x_2^{m_2}\cdots x_n^{m_n}$ where $a \in F$. Let us consider a **power product** in $F[x]$ to be an expression

$$P = x_1^{m_1}x_2^{m_2}\cdots x_n^{m_n} \text{ where all the } m_i \geq 0 \text{ in } \mathbb{Z}.$$

Notice that all x_i are present, perhaps some with exponent 0. Thus in $F[x, y, z]$, we must write xz^2 as xy^0z^2 to be a power product. We want to describe a *total ordering* $<$ on the set of all power products so that we know just what it means to say that $P_i < P_j$ for two power products, providing us with a notion of relative size for power products. We can then try to change an ideal basis in a systematic way to create one with polynomials having terms a_iP_i with as “small” power products P_i as possible. We denote by 1 the power product with all exponents 0, and require that an ordering of the power products has the properties listed below. Suppose that such an ordering has been described and that $P_i \neq P_j$ and P_i divides P_j so that $P_j = PP_i$ where $1 < P$. From Property 4, we then have $1P_i < PP_i = P_j$, so $P_i < P_j$. Thus P_i divides P_j implies that

[†] This section is not used in the remainder of the text.

$P_i < P_j$. In Exercise 28, we ask you to show by a counterexample that $P_i < P_j$ does not imply that P_i divides P_j .

Properties for an Ordering of Power Products

1. $1 < P$ for all power products $P \neq 1$.
2. For any two power products P_i and P_j , exactly one of $P_i < P_j, P_i = P_j, P_j < P_i$ holds.
3. If $P_i < P_j$ and $P_j < P_k$, then $P_i < P_k$.
4. If $P_i < P_j$, then $PP_i < PP_j$ for any power product P .

It can also be shown that these properties guarantee that any step-by-step process for modifying a finite ideal basis that does not increase the size of any maximal power product in a basis element and replaces at least one by something smaller at each step will terminate in a finite number of steps.

In $F[x]$ with x the only indeterminate, there is only one power product ordering, for by Property 1, we must have $1 < x$. Multiplying repeatedly by x and using Property 4, we have $x < x^2, x^2 < x^3$, etc. Property 3 then shows that $1 < x < x^2 < x^3 < \dots$ is the only possible order. Notice that in Example 37.9, we modified a basis by replacing basis polynomials by polynomials containing smaller power products.

There are a number of possible orderings for power products in $F[x]$ with n indeterminates. We present just one, the *lexicographical order* (denoted by “lex”). In lex, we define

$$x_1^{s_1} x_2^{s_2} \cdots x_n^{s_n} < x_1^{t_1} x_2^{t_2} \cdots x_n^{t_n} \quad (2)$$

if and only if $s_i < t_i$ for the first subscript i , reading from left to right, such that $s_i \neq t_i$. Thus in $F[x, y]$, if we write power products in the order $x^n y^m$, we have $y = x^0 y^1 < x^1 y^0 = x$ and $xy < xy^2$. Using lex, the order of n indeterminates is given by $1 < x_n < x_{n-1} < \dots < x_2 < x_1$. Our reduction in Example 37.8, where we first got rid of all “big” x ’s that we could and then the “smaller” y ’s, corresponded to the lex order $z < y < x$, that is, to writing all power products in the $x^m y^n z^s$ order. For the two-indeterminate case with $y < x$, the total lex term order schematically is

$$1 < y < y^2 < y^3 \cdots < x < xy < xy^2 < xy^3 < \cdots < x^2 < x^2 y < x^2 y^2 < \cdots .$$

In all the examples that follow we will use lexicographic ordering using a specified ordering of the indeterminates.

An ordering of power products P induces an obvious ordering of terms aP of a polynomial in $F[x]$, which we will refer to as a **term order**. From now on, given an ordering of power products, we consider every polynomial f in $F[x]$ to be written in decreasing order of terms, so that the leading (first) term has the highest order. We denote by $1t(f)$ the leading term of f and by $1p(f)$ the power product of the leading term. If f and g are polynomials in $F[x]$ such that $1p(g)$ divides $1p(f)$, then we can execute a division of f by g , as illustrated by the linear and one-indeterminate cases, in Section 37 to obtain $f(x) = g(x)q(x) + r(x)$ where $1p(r) < 1p(f)$. Note that we did not say that $1p(r) < 1p(g)$. We illustrate with an example.

38.1 Example By division, reduce the basis $\{xy^2, y^2 - y\}$ for the ideal $I = \langle xy^2, y^2 - y \rangle$ in $\mathbb{R}[x, y]$ to one with smaller maximum term size, assuming the order lex with $y < x$.

Solution We see that y^2 divides xy^2 and compute

$$\begin{array}{r} x \\ y^2 - y \sqrt{xy^2} \\ \hline xy^2 - xy \\ \hline xy \end{array}$$

Because y^2 does not divide xy , we cannot continue the division. Note that $\text{lp}(xy) = xy$ is not less than $\text{lp}(y^2 - y) = y^2$. However, we do have $\text{lp}(xy) < \text{lp}(xy^2)$. Our new basis for I is $\{xy, y^2 - y\}$. \blacktriangle

When dealing with more than one indeterminate, it is often easier to perform basis reduction by multiplying a basis polynomial $g(\mathbf{x})$ by a polynomial $-q(\mathbf{x})$ and adding it to a polynomial $f(\mathbf{x})$ to obtain $r(\mathbf{x})$, as we perform matrix reduction in linear algebra, rather than writing out the division display as we did in the preceding example. Starting with basis polynomials xy^2 and $y^2 - y$, we can reduce the xy^2 by multiplying $y^2 - y$ by $-x$ and adding the resulting $-xy^2 + xy$ to xy^2 , obtaining the replacement xy for xy^2 . We can do that in our head, and write down the result directly.

Referring again to Example 38.1, it will follow from what we state later that given any polynomial $f(x, y) = c_1(x, y)(xy) + c_2(x, y)(y^2 - y)$ in $\langle xy, y^2 - y \rangle$, either xy or $y^2 - y$ will divide $\text{lp}(f)$. (See Exercise 32.) This illustrates the defining property of a *Gröbner basis*.

38.2 Definition A set $\{g_1, g_2, \dots, g_r\}$ of nonzero polynomials in $F[x_1, x_2, \dots, x_n]$, with term ordering $<$, is a **Gröbner basis** for the ideal $I = \langle g_1, g_2, \dots, g_r \rangle$ if and only if, for each nonzero $f \in I$, there exists some i where $1 \leq i \leq r$ such that $\text{lp}(g_i)$ divides $\text{lp}(f)$. \blacksquare

While we have illustrated the computation of a Gröbner basis from a given basis for an ideal in Examples 37.8, 37.9, and 38.1, we have not given a specific algorithm. We refer the reader to Adams and Loustaunau [23]. The method consists of multiplying some polynomial in the basis by any polynomial in $F[\mathbf{x}]$ and adding the result to another polynomial in the basis in a manner that reduces the size of power products. In our illustrations, we have treated the case involving division of $f(\mathbf{x})$ by $g(\mathbf{x})$ where $\text{lp}(g)$ divides $\text{lp}(f)$, but we can also use the process if $\text{lp}(g)$ only divides some other power product in f . For example, if two elements in a basis are $xy - y^3$ and $y^2 - 1$, we can multiply $y^2 - 1$ by y and add it to $xy - y^3$, reducing $xy - y^3$ to $xy - y$. Theorem 37.7 shows that this is a valid computation.

You may wonder how any basis $\{g_1, g_2, \dots, g_r\}$ can fail to be a Gröbner basis for $I = \langle g_1, g_2, \dots, g_r \rangle$ because, when we form an element $c_1g_1 + c_2g_2 + \dots + c_rg_r$ in I , we see that $\text{lp}(g_i)$ is a divisor of $\text{lp}(c_ig_i)$ for $i = 1, 2, \dots, r$. However, cancellation of power products can occur in the addition. We illustrate with an example.

38.3 Example Consider the ideal $I = \langle x^2y - 2, xy^2 - y \rangle$ in $\mathbb{R}[x, y]$. The polynomials in the basis shown cannot be reduced further. However, the ideal I contains $y(x^2y - 2) - x(xy^2 - y) = xy - 2y$, whose leading power product xy is not divisible by either of the leading power products x^2y or xy^2 of the given basis. Thus $\{x^2y - 2, xy^2 - y\}$ is not a Gröbner basis for I , according to Definition 38.2. \blacktriangle

When we run into a situation like that in Example 38.3, we realize that a Gröbner basis must contain some polynomial with a smaller leading power product than those in the given basis. Let f and g be polynomials in the given basis. Just as we did in Example 38.3, we can multiply f and g by as small power products as possible so that the resulting two leading power products will be the same, the *least common multiple*

(lcm) of $1p(f)$ and $1p(g)$, and then subtract or add with suitable coefficients from F so cancellation results. We denote a polynomial formed in this fashion by $S(f, g)$. We state without proof a theorem that can be used to test whether a basis is a Gröbner basis.

38.4 Theorem A basis $G = \{g_1, g_2, \dots, g_r\}$ is a Gröbner basis for the ideal $\langle g_1, g_2, \dots, g_r \rangle$ if and only if, for all $i \neq j$, the polynomial $S(g_i, g_j)$ can be reduced to zero by repeatedly dividing remainders by elements of G , as in the division algorithm.

As we mentioned before, we may prefer to think of reducing $S(g_i, g_j)$ by a sequence of operations consisting of adding (or subtracting) multiples of polynomials in G , rather than writing out division.

We can now indicate how we can obtain a Gröbner basis from a given basis. First, reduce the polynomials in the basis as far as possible among themselves. Then choose polynomials g_i and g_j in the basis, and form the polynomial $S(g_i, g_j)$. See if $S(g_i, g_j)$ can be reduced to zero as just described. If so, choose a different pair of polynomials, and repeat the procedure with them. If $S(g_i, g_j)$ cannot be reduced to zero as described above, augment the given basis with this $S(g_i, g_j)$, and start all over, reducing this basis as much as possible. By Theorem 38.4, when every polynomial $S(g_i, g_j)$ for all $i \neq j$ can be reduced to zero using polynomials from the latest basis, we have arrived at a Gröbner basis. We conclude with a continuation of Example 38.3.

38.5 Example Continuing Example 38.3, let $g_1 = x^2y - 2$, $g_2 = xy^2 - y$, and $I = \langle g_1, g_2 \rangle$ in \mathbb{R}^2 . In Example 38.3, we obtained the polynomial $S(g_1, g_2) = xy - 2y$, which cannot be reduced to zero using g_1 and g_2 . We now reduce the basis $\{x^2y - 2, xy^2 - y, xy - 2y\}$, indicating each step.

$\{x^2y - 2, xy^2 - y, xy - 2y\}$	augmented basis
$\{2xy - 2, xy^2 - y, xy - 2y\}$	by adding $(-x)$ (third) to first
$\{2xy - 2, 2y^2 - y, xy - 2y\}$	by adding $(-y)$ (third) to second
$\{4y - 2, 2y^2 - y, xy - 2y\}$	by adding (-2) (third) to first
$\{4y - 2, 0, xy - 2y\}$	by adding $(-\frac{y}{2})$ (first) to second
$\{4y - 2, 0, \frac{1}{2}x - 2y\}$	by adding $(-\frac{x}{4})$ (first) to third
$\{4y - 2, 0, \frac{1}{2}x - 1\}$	by adding $(\frac{1}{2})$ (first) to third

Clearly, $\{y - \frac{1}{2}, x - 2\}$ is a Gröbner basis. Note that if $f = y - \frac{1}{2}$ and $g = x - 2$, then $S(f, g) = xf - yg = (xy - \frac{x}{2}) - (xy - 2y) = -\frac{x}{2} + 2y$, which can readily be reduced to zero by adding $\frac{1}{2}(x - 2)$ and $-2(y - \frac{1}{2})$.

From the Gröbner basis, we see that the algebraic variety $V(I)$ contains only one point, $(2, \frac{1}{2})$, in \mathbb{R}^2 . ▲

Applications

Here we give a simple example of how a Gröbner basis can be used to derive a geometric formula.

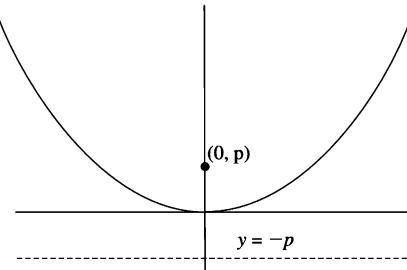
38.6 Example Using a Gröbner basis, derive the standard formula for a parabola.

Solution Recall that a parabola is the set of all points in the plane that are equidistant from a fixed line (directrix) and a fixed point (focus). In standard form, the directrix is the line $y = -p$ and the focus is the point $(0, p)$ where $p > 0$. The algebraic variety defined by the ideal $\langle x^2 + (y - p)^2 - d^2, y + p - d \rangle$ gives the set of all points $(x_0, y_0, p_0, d_0) \in \mathbb{R}^4$

such that the point (x_0, y_0) is a distance of d_0 from the point $(0, p_0)$ and a distance d_0 from the line $y = -p$. We order the indeterminates by $p < x < y < d$ and seek a Gröbner basis.

$$\begin{array}{ll} \{-d^2 + y^2 - 2yp + x^2 + p^2, d - y - p\} & \text{initial basis} \\ \{-dy - dp + y^2 - 2yp + x^2 + p^2, d - y - p\} & \text{by adding (d)(second) to first} \\ \{-dp - 3yp + x^2 + p^2, d - y - p\} & \text{by adding (y)(second) to first} \\ \{-4yp + x^2, d - y - p\} & \text{by adding (p)(second) to first} \end{array}$$

In Exercise 33 you are asked to check that $\{-4yp + x^2, d - y - p\}$ is a Gröbner basis. Note that for any $(x_0, y_0, p_0) \in \mathbb{R}^3$ that is a zero of the first polynomial $-4yp + x^2$, there is a real number d_0 so that (x_0, y_0, p_0, d_0) is a zero of the second polynomial. This verifies the standard form of the equation of a parabola, $4py = x^2$. \blacktriangle



Care must be taken in deciding the order for the indeterminate. In Example 38.6, had we taken the order of the indeterminates to be $d < p < x < y$, the resulting Gröbner basis would have been $\{-4dp + 4p^2 + x^2, d - p - y\}$, which does not directly give the standard form for the equation of a parabola. The reason for making the indeterminate d the largest is that the algorithm attempts to eliminate d from polynomials in the basis since the reduction process reduces the maximum-sized term.

The algorithm for computing Gröbner bases can be programmed. In fact, it is a built-in function on many common mathematical packages including Mathematica, Maple, Wolfram Alpha, and others. By carefully picking the order of the indeterminates, it is possible to solve a large variety of problems using automated computations of Gröbner bases.

38.7 Example We use a software package to compute a Gröbner basis that yields the equation of an ellipse in standard position. An ellipse in standard position is the set of all points in the plane so that the sum of the distances to the two foci $(c, 0)$ and $(-c, 0)$ is a fixed distance $2a$. We start with the basis

$$\{(x - c)^2 + y^2 - d_1^2, (x + c)^2 + y^2 - d_2^2, d_1 + d_2 - 2a\}$$

where d_1 represents the distance to focus $(c, 0)$ and d_2 represents the distance to focus $(-c, 0)$. We wish to arrive at a polynomial that includes x and y as well as the parameters a and c that describe the shape and size of the ellipse. Therefore, we take d_1 and d_2 as the largest indeterminates. Using the order $a < c < y < x < d_2 < d_1$, a mathematical software package quickly computes the Gröbner basis

$$\begin{aligned} & \{a^4 - a^2c^2 - a^2x^2 + c^2x^2 - a^2y^2, -a^2 + ad_2 - cx, a^3 - ac^2 - acs + cd_2x - ax^2 - ay^2, \\ & \quad -c^2 + d_2^2 - 2cx - x^2 - y^2, -2a + d_1 + d_2\}. \end{aligned}$$

The first polynomial can be written as

$$a^2(a^2 - c^2) - (a^2 - c^2)x^2 - a^2y^2.$$

We think of a and c as parameters and recall that in an ellipse, the parameter $b > 0$ is defined by $b^2 = a^2 - c^2$. We see that the variety in \mathbb{R}^2 defined from the ideal $\langle a^2b^2 - b^2x^2 - a^2y^2 \rangle$ has equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad \blacktriangle$$

A **graph** is a finite set with elements called **vertices** together with a collection of pairs of vertices called **edges** or **arcs**. Graphs are often drawn as in Figure 38.10, which shows a graph with four vertices x_1, x_2, x_3, x_4 and four edges $\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_1\}, \{x_3, x_4\}$. A **coloring** of a graph is an assignment of colors to the vertices in such a way that no two vertices in the same edge have the same color. The problem of determining if a graph can be colored with n colors can be restated as a question about an ideal, which can (at least theoretically) be answered using Gröbner bases.

38.8 Theorem Let G be a graph with vertex set $\{x_1, x_2, \dots, x_k\}$. We identify each vertex with an indeterminate and form the ideal $I \subseteq \mathbb{C}[x_1, x_2, \dots, x_k]$ with basis consisting of polynomials $x_1 - 1, x_2 - 1, x_3 - 1, \dots, x_k - 1$ together with a polynomial $x_i^{n-1} + x_i^{n-2}x_j + x_i^{n-3}x_j^2 + \dots + x_i x_j^{n-2} + x_j^{n-1}$ for each edge in the graph. The graph G can be colored with n colors if and only if the algebraic variety $V(I)$ is nonempty.

Proof The proof is left for Exercise 34. ◆

The variety $V(I)$ in Theorem 38.8 gives all possible colorings of the graph G using the n^{th} roots of unity for the colors and coloring vertex x_1 with the color 1. For each point in $V(I)$, the j^{th} coordinate gives the color assigned to x_j .

38.9 Example Using the graph in Figure 38.10, the basis described in Theorem 38.8 for $n = 2$ is

$$\{x_1 - 1, x_2^2 - 1, x_3^2 - 1, x_4^2 - 1, x_1 + x_2, x_2 + x_3, x_3 + x_1, x_3 + x_4\}.$$

A Gröbner basis is $\{1\}$. This can be determined electronically, or by noticing that

$$(x_1 + x_2) - (x_2 + x_3) + (x_3 + x_1) = 2x_1 \in I.$$

Therefore, $x_1 \in I$, so $x_1 - (x_1 - 1) = 1 \in I$. Thus the ideal is all of $\mathbb{C}[x_1, x_2, x_3, x_4]$, which has $\{1\}$ as a basis. Since the algebraic variety for $\{1\}$ is the empty set, the graph cannot be colored with two colors.

We now compute a Gröbner basis when $n = 3$. Theorem 38.8 tells us to start with the basis

$$\begin{aligned} \{x_1 - 1, x_2^3 - 1, x_3^3 - 1, x_4^3 - 1, x_1^2 + x_1x_2 + x_2^2, x_2^2 + x_2x_3 + x_3^2, x_3^2 + x_3x_1 \\ + x_1^2, x_3^2 + x_3x_4 + x_4^2\}. \end{aligned}$$

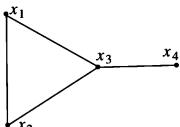
Using a software package, a Gröbner basis with indeterminate order $x_1 < x_2 < x_3 < x_4$ is

$$\{x_1 - 1, x_2^2 + x_2 + 1, 1 + x_2 + x_3, x_2 - x_4 - x_2x_4 + x_4^2\}.$$

Let $\zeta = -\frac{\sqrt{3}}{2} + \frac{1}{2}i$ be a third root of unity. We find an element $(a_1, a_2, a_3, a_4) \in V(I)$:

$$\begin{aligned} a_1 - 1 &= 0 \quad a_1 = 1 \\ a_2^2 + a_2 + 1 &= 0 \quad a_2 \text{ is a third root of unity. Let } a_2 = \zeta \\ 1 + \zeta + a_3 &= 0 \quad a_3 = -1 - \zeta = \zeta^2 \\ \zeta - a_4 - \zeta a_4 + a_4^2 &= 0 \quad \text{factoring, } (\zeta - a_4)(1 - a_4) = 0. \end{aligned}$$

38.10 Figure



We see that we can color vertex 1 the color 1, vertex 2 ζ , vertex 3 ζ^2 , and then we can color vertex 4 either the color 1 or ζ . \blacktriangle

There are many applications of Gröbner bases in mathematics, statistics, engineering, and computer science. These include controlling robot arms and automated theorem proving. The algorithm to compute a Gröbner basis is generally too long and tedious to be done by hand, so computers are essential. However, some larger problems are not feasible even on modern high-speed computers.

■ EXERCISES 38

In Exercises 1 through 4, write the polynomials in $\mathbb{R}[x, y, z]$ in decreasing term order, using the order lex for power products $x^m y^n z^s$ where $z < y < x$.

1. $2xy^3z^5 - 5x^2yz^3 + 7x^2y^2z - 3x^3$
 3. $3y - 7x + 10z^3 - 2xy^2z^2 + 2x^2yz^2$

2. $3y^2z^5 - 4x + 5y^3z^3 - 8z^7$
 4. $38 - 4xz + 2yz - 8xy + 3yz^3$

In Exercises 5 through 8, write the polynomials in $\mathbb{R}[x, y, z]$ in decreasing term order, using the order lex for power products $z^m y^n x^s$ where $x < y < z$.

5. The polynomial in Exercise 1.
 7. The polynomial in Exercise 3.

6. The polynomial in Exercise 2.
 8. The polynomial in Exercise 4.

Another ordering, deglex, for power products in $F[x]$ is defined as follows:

$$x_1^{s_1} x_2^{s_2} \cdots x_n^{s_n} < x_1^{t_1} x_2^{t_2} \cdots x_n^{t_n}$$

if and only if either $\sum_{i=1}^n s_i < \sum_{i=1}^n t_i$, or these two sums are equal and $s_i < t_i$ for the smallest value of i such that $s_i \neq t_i$. Exercises 9 through 13 are concerned with the order deglex.

9. List, in increasing order, the smallest 20 power products in $\mathbb{R}[x, y, z]$ for the order deglex with power products $x^m y^n z^s$ where $z < y < x$.

In Exercises 10 through 13, write the polynomials in order of decreasing terms using the order deglex with power products $x^m y^n z^s$ where $z < y < x$.

10. The polynomial in Exercise 1.
 11. The polynomial in Exercise 2.
 12. The polynomial in Exercise 3.
 13. The polynomial in Exercise 4.

For Exercises 14 through 17, let power products in $\mathbb{R}[x, y, z]$ have order lex where $z < y < x$. If possible, perform a single-step division algorithm reduction that changes the given ideal basis to one having smaller maximum term order.

14. $\langle xy^2 - 2x, x^2y + 4xy, xy - y^2 \rangle$
 15. $\langle xy + y^3, y^3 + z, x - y^4 \rangle$
 16. $\langle xyz - 3z^2, x^3 + y^2z^3, x^2yz^3 + 4 \rangle$
 17. $\langle y^2z^3 + 3, y^3z^2 - 2z, y^2z^2 + 3 \rangle$

In Exercises 18 and 19, let the order of power products in $\mathbb{R}[w, x, y, z]$ be lex with $z < y < x < w$. Find a Gröbner basis for the given ideal.

18. $\langle w + x - y + 4z - 3, 2w + x + y - 2z + 4, w + 3x - 3y + z - 5 \rangle$
 19. $\langle w - 4x + 3y - z + 2, 2w - 2x + y - 2z + 5, w - 10x + 8y - z - 1 \rangle$

In Exercises 20 through 22, find a Gröbner basis for the indicated ideal in $\mathbb{R}[x]$.

20. $\langle x^4 + x^3 - 3x^2 - 4x - 4, x^3 + x^2 - 4x - 4 \rangle$
 21. $\langle x^4 - 4x^3 + 5x^2 - 2x, x^3 - x^2 - 4x + 4, x^3 - 3x + 2 \rangle$
 22. $\langle x^5 + x^2 + 2x - 5, x^3 - x^2 + x - 1 \rangle$

In Exercises 23 through 26, find a Gröbner basis for the given ideal in $\mathbb{R}[x, y]$. Consider the order of power products to be lex with $y < x$. If you can, describe the corresponding algebraic variety in $\mathbb{R}[x, y]$.

23. $\langle x^2y - x - 2, xy + 2y - 9 \rangle$

25. $\langle x^2y + x + 1, xy^2 + y - 1 \rangle$

24. $\langle x^2y + x, xy^2 - y \rangle$

26. $\langle x^2y + xy^2, xy - x \rangle$

Concepts

27. Determine whether each of the following is true or false.

- a. Polynomials in a Gröbner basis are linearly independent.
 - b. The set $\{1\}$ is a Gröbner basis.
 - c. The set $\{0\}$ is a Gröbner basis.
 - d. The order one picks for the power products does not affect the resulting Gröbner basis.
 - e. For any total ordering of all the power products, $x_1^2 > x_1$.
 - f. A Gröbner basis can be used to determine if a graph can be colored with n colors starting with a basis consisting of polynomials each of degree at most n .
 - g. A Gröbner basis can be used to determine if a graph can be colored with n colors starting with a basis consisting of $r+s$ polynomials where r is the number of vertices in the graph and s is the number of edges in the graph.
 - h. I have computed Gröbner bases before I knew what they were.
 - i. Any ideal in $F[\mathbf{x}]$ has a unique Gröbner basis.
 - j. A basis for an ideal I in $F[x_1, x_2, \dots, x_n]$ is a Gröbner basis if and only if each polynomial in the basis cannot be reduced further using the division algorithm.
28. Let $\mathbb{R}[x, y]$ be ordered by lex. Give an example to show that $P_i < P_j$ does not imply that P_i divides P_j .
29. What other orders of the indeterminate a, c, x, y, d_1, d_2 would you expect the equation of an ellipse to result from computing a Gröbner basis for the ideal in Example 38.7?
30. Use a Gröbner basis to derive the formula for a hyperbola in standard position. Recall that a hyperbola in standard position is the set of all points in the plane whose difference in distances from $(c, 0)$ and $(-c, 0)$ is $\pm 2a$. You may use a computer to compute the Gröbner basis.
31. Use a Gröbner basis to show that the graph with vertex set $\{x_1, x_2, x_3, x_4, x_5\}$ and edge set $\{\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_1, x_3\}, \{x_1, x_5\}, \{x_5, x_4\}\}$ cannot be colored with three colors, but it can be colored with four colors. You may use a computer to compute the Gröbner basis.

Theory

32. Show that $\{xy, y^2 - y\}$ is a Gröbner basis for $\langle xy, y^2 - y \rangle$, as asserted after Example 38.1.
33. Show that $\{-4yp + x^2, d - y - p\}$ is a Gröbner basis for the ideal $\langle -4yp + x^2, d - y - p \rangle$ as asserted in Example 38.6.
34. Prove Theorem 38.8. [Hint: Think about coloring a graph with the n^{th} roots of unity.]

Extension Fields

Section 39 Introduction to Extension Fields

Section 40 Algebraic Extensions

Section 41 [†]Geometric Constructions

Section 42 Finite Fields

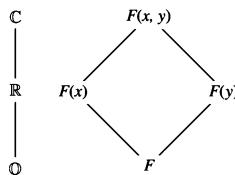
SECTION 39

INTRODUCTION TO EXTENSION FIELDS

Our Basic Goal Achieved

We are now in a position to achieve our **basic goal**, which, loosely stated, is to show that every nonconstant polynomial has a zero. This will be stated more precisely and proved in Theorem 39.3. We first introduce some new terminology for some old ideas.

39.1 Definition A field E is an **extension field of a field F** if $F \leq E$. ■



39.2 Figure

Thus \mathbb{R} is an extension field of \mathbb{Q} , and \mathbb{C} is an extension field of both \mathbb{R} and \mathbb{Q} . As in the study of groups, it will often be convenient to use subfield diagrams to picture extension fields, the larger field being on top. We illustrate this in Fig. 39.2. (Recall that $F(x)$ is the field of quotients constructed from $F[x]$.) A configuration where there is just one single column of fields, as at the left-hand side of Fig. 39.2, is often referred to, without any precise definition, as a **tower of fields**.

Now for our *basic goal!* This great and important result follows quickly and elegantly from the techniques we now have at our disposal.

[†] Section 41 is not required for the remainder of the text.

39.3 Theorem (Kronecker's Theorem) (Basic Goal) Let F be a field and let $f(x)$ be a nonconstant polynomial in $F[x]$. Then there exists an extension field E of F and an $\alpha \in E$ such that $f(\alpha) = 0$.

Proof By Theorem 28.21, $f(x)$ has a factorization in $F[x]$ into polynomials that are irreducible over F . Let $p(x)$ be an irreducible polynomial in such a factorization. It is clearly sufficient to find an extension field E of F containing an element α such that $p(\alpha) = 0$.

By Theorem 31.25, $\langle p(x) \rangle$ is a maximal ideal in $F[x]$, so $F[x]/\langle p(x) \rangle$ is a field. We claim that F can be identified with a subfield of $F[x]/\langle p(x) \rangle$ in a natural way by use of the map $\psi : F \rightarrow F[x]/\langle p(x) \rangle$ given by

$$\psi(a) = a + \langle p(x) \rangle$$

for $a \in F$. This map is one-to-one, for if $\psi(a) = \psi(b)$, that is, if $a + \langle p(x) \rangle = b + \langle p(x) \rangle$ for some $a, b \in F$, then $(a - b) \in \langle p(x) \rangle$, so $a - b$ must be a multiple of the polynomial $p(x)$, which is of degree ≥ 1 . Now $a, b \in F$ implies that $a - b$ is in F . Thus we must have $a - b = 0$, so $a = b$. We defined addition and multiplication in $F[x]/\langle p(x) \rangle$ by choosing any representatives, so we may choose $a \in (a + \langle p(x) \rangle)$. Thus ψ is a homomorphism that maps F one-to-one onto a subfield of $F[x]/\langle p(x) \rangle$. We identify F with $\{a + \langle p(x) \rangle \mid a \in F\}$ by means of this map ψ . Thus we shall view $E = F[x]/\langle p(x) \rangle$ as an extension field of F . We have now manufactured our desired extension field E of F . It remains for us to show that E contains a zero of $p(x)$.

■ HISTORICAL NOTE

Léopold Kronecker is known for his insistence on constructibility of mathematical objects. As he noted, “God made the integers; all else is the work of man.” Thus, he wanted to be able to construct new “domains of rationality” (fields) by using only the existence of integers and indeterminates. He did not believe in starting with the real or complex numbers, because as far as he was concerned, those fields could not be determined in a constructive way. Hence in an 1881 paper, Kronecker created an extension field by simply adjoining to a given field a root α of an irreducible n th degree polynomial $p(x)$; that is, his new field consisted of expressions

rational in the original field elements and his new root α with the condition that $p(\alpha) = 0$. The proof of the theorem presented in the text (Theorem 39.3) dates from the twentieth century.

Kronecker completed his dissertation in 1845 at the University of Berlin. For many years thereafter, he managed the family business, ultimately becoming financially independent. He then returned to Berlin, where he was elected to the Academy of Sciences and thus permitted to lecture at the university. On the retirement of Kummer, he became a professor at Berlin, and with Karl Weierstrass (1815–1897) directed the influential mathematics seminar.

Let us set

$$\alpha = x + \langle p(x) \rangle,$$

so $\alpha \in E$. Consider the evaluation homomorphism $\phi_\alpha : F[x] \rightarrow E$, given by Theorem 27.4. If $p(x) = a_0 + a_1x + \cdots + a_nx^n$, where $a_i \in F$, then we have

$$\phi_\alpha(p(x)) = a_0 + a_1(x + \langle p(x) \rangle) + \cdots + a_n(x + \langle p(x) \rangle)^n$$

in $E = F[x]/\langle p(x) \rangle$. But we can compute in $F[x]/\langle p(x) \rangle$ by choosing representatives, and x is a representative of the coset $\alpha = x + \langle p(x) \rangle$. Therefore,

$$\begin{aligned} p(\alpha) &= (a_0 + a_1x + \cdots + a_nx^n) + \langle p(x) \rangle \\ &= p(x) + \langle p(x) \rangle = \langle p(x) \rangle = 0 \end{aligned}$$

in $F[x]/\langle p(x) \rangle$. We have found an element α in $E = F[x]/\langle p(x) \rangle$ such that $p(\alpha) = 0$, and therefore $f(\alpha) = 0$. \blacklozenge

We illustrate the construction involved in the proof of Theorem 39.3 by two examples.

39.4 Example Let $F = \mathbb{R}$, and let $f(x) = x^2 + 1$, which is well known to have no zeros in \mathbb{R} and thus is irreducible over \mathbb{R} by Theorem 28.11. Then $\langle x^2 + 1 \rangle$ is a maximal ideal in $\mathbb{R}[x]$, so $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ is a field. Identifying $r \in \mathbb{R}$ with $r + \langle x^2 + 1 \rangle$ in $\mathbb{R}[x]/\langle x^2 + 1 \rangle$, we can view \mathbb{R} as a subfield of $E = \mathbb{R}[x]/\langle x^2 + 1 \rangle$. Let

$$\alpha = x + \langle x^2 + 1 \rangle.$$

Computing in $\mathbb{R}[x]/\langle x^2 + 1 \rangle$, we find

$$\begin{aligned} \alpha^2 + 1 &= (x + \langle x^2 + 1 \rangle)^2 + (1 + \langle x^2 + 1 \rangle) \\ &= (x^2 + 1) + \langle x^2 + 1 \rangle = 0. \end{aligned}$$

Thus α is a zero of $x^2 + 1$. We shall identify $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ with \mathbb{C} near the close of this section. \blacktriangle

39.5 Example Let $F = \mathbb{Q}$, and consider $f(x) = x^4 - 5x^2 + 6$. This time $f(x)$ factors in $\mathbb{Q}[x]$ into $(x^2 - 2)(x^2 - 3)$, both factors being irreducible over \mathbb{Q} , as we have seen. We can start with $x^2 - 2$ and construct an extension field E of \mathbb{Q} containing α such that $\alpha^2 - 2 = 0$, or we can construct an extension field K of \mathbb{Q} containing an element β such that $\beta^2 - 3 = 0$. The construction in either case is just as in Example 39.4. \blacktriangle

Algebraic and Transcendental Elements

As we said before, most of the rest of this text is devoted to the study of zeros of polynomials. We commence this study by putting an element of an extension field E of a field F into one of two categories.

39.6 Definition An element α of an extension field E of a field F is **algebraic over F** if $f(\alpha) = 0$ for some nonzero $f(x) \in F[x]$. If α is not algebraic over F , then α is **transcendental over F** . \blacksquare

39.7 Example \mathbb{C} is an extension field of \mathbb{Q} . Since $\sqrt{2}$ is a zero of $x^2 - 2$, we see that $\sqrt{2}$ is an algebraic element over \mathbb{Q} . Also, i is an algebraic element over \mathbb{Q} , being a zero of $x^2 + 1$. \blacktriangle

39.8 Example It is well known (but not easy to prove) that the real numbers π and e are transcendental over \mathbb{Q} . Here e is the base for the natural logarithms. \blacktriangle

Just as we do not speak simply of an *irreducible polynomial*, but rather of an *irreducible polynomial over F* , similarly we don't speak simply of an *algebraic element*, but rather of an *element algebraic over F* . The following illustration shows the reason for this.

39.9 Example The real number π is transcendental over \mathbb{Q} , as we stated in Example 39.8. However, π is algebraic over \mathbb{R} , for it is a zero of $(x - \pi) \in \mathbb{R}[x]$. \blacktriangle

39.10 Example It is easy to see that the real number $\sqrt{1 + \sqrt{3}}$ is algebraic over \mathbb{Q} . For if $\alpha = \sqrt{1 + \sqrt{3}}$, then $\alpha^2 = 1 + \sqrt{3}$, so $\alpha^2 - 1 = \sqrt{3}$ and $(\alpha^2 - 1)^2 = 3$. Therefore $\alpha^4 - 2\alpha^2 - 2 = 0$, so α is a zero of $x^4 - 2x^2 - 2$, which is in $\mathbb{Q}[x]$. \blacktriangle

To connect these ideas with those of number theory, we give the following definition.

39.11 Definition An element of \mathbb{C} that is algebraic over \mathbb{Q} is an **algebraic number**. A **transcendental number** is an element of \mathbb{C} that is transcendental over \mathbb{Q} . ■

There is an extensive and elegant theory of algebraic numbers. (See the Bibliography.)

The next theorem gives a useful characterization of algebraic and transcendental elements over F in an extension field E of F . It also illustrates the importance of our evaluation homomorphisms ϕ_α . Note that once more we are describing our concepts in terms of mappings.

39.12 Theorem Let E be an extension field of a field F and let $\alpha \in E$. Let $\phi_\alpha : F[x] \rightarrow E$ be the evaluation homomorphism of $F[x]$ into E such that $\phi_\alpha(a) = a$ for $a \in F$ and $\phi_\alpha(x) = \alpha$. Then α is transcendental over F if and only if ϕ_α gives an isomorphism of $F[x]$ with a subdomain of E , that is, if and only if ϕ_α is a one-to-one map.

Proof The element α is transcendental over F if and only if $f(\alpha) \neq 0$ for all nonzero $f(x) \in F[x]$, which is true (by definition) if and only if $\phi_\alpha(f(x)) \neq 0$ for all nonzero $f(x) \in F[x]$, which is true if and only if the kernel of ϕ_α is $\{0\}$, that is, if and only if ϕ_α is a one-to-one map. ◆

The Irreducible Polynomial for α over F

Consider the extension field \mathbb{R} of \mathbb{Q} . We know that $\sqrt{2}$ is algebraic over \mathbb{Q} , being a zero of $x^2 - 2$. Of course, $\sqrt{2}$ is also a zero of $x^3 - 2x$ and of $x^4 - 3x^2 + 2 = (x^2 - 2)(x^2 - 1)$. Both these other polynomials having $\sqrt{2}$ as a zero were multiples of $x^2 - 2$. The next theorem shows that this is an illustration of a general situation. This theorem plays a central role in our later work.

39.13 Theorem Let E be a field extension of F , and let $\alpha \in E$ be algebraic over F . Then $\{f(x) \in F[x] | f(\alpha) = 0\} = \langle p(x) \rangle$ for some polynomial $p(x) \in F[x]$. Furthermore, $p(x)$ is irreducible over F .

Proof Let $I = \{f(x) \in F[x] | f(\alpha) = 0\}$. Then I is the kernel of the evaluation homomorphism $\phi_\alpha : F[x] \rightarrow E$, which implies that I is an ideal in $F[x]$. By Theorem 31.24, I is a principal ideal generated by a polynomial $p(x) \in F[x]$. Thus $I = \langle p(x) \rangle$.

It remains to show that $p(x)$ is irreducible over F . The degree of $p(x)$ is at least 1, which implies that $p(x)$ is neither 0 nor a unit in $F[x]$. Suppose that $p(x) = r(x)s(x)$ is a factorization of $p(x)$ over the field F . Then applying the evaluation homomorphism, either $r(\alpha) = 0$ or $s(\alpha) = 0$, since E is a field. We relabel r and s , if necessary, so that $r(\alpha) = 0$. Then $r(x) \in I = \langle p(x) \rangle$. Thus

$$p(x) = r(x)s(x) = p(x)r_1(x)s(x)$$

for some $r_1(x) \in F[x]$. Canceling $p(x)$ shows that $s(x)$ is a unit. Therefore, $p(x)$ is irreducible. ◆

By multiplying by a suitable constant in F , we can assume that the coefficient of the highest power of x appearing in $p(x)$ of Theorem 39.13 is 1. Such a polynomial having 1 as the coefficient of the highest power of x appearing is a **monic polynomial**.

39.14 Corollary Let E be an extension field of F , and let $\alpha \in E$ be algebraic over F . Then there is a unique irreducible polynomial $p(x) \in F[x]$ such that $p(x)$ is monic, $p(\alpha) = 0$, and for any polynomial $f(x) \in F[x]$ with $f(\alpha) = 0$, $p(x)$ divides $f(x)$.

Proof Let $p(x)$ be the polynomial of Theorem 39.13. By multiplying $p(x)$ with an appropriate element of F , we can assume that $p(x)$ is monic. Since $\{f(x) | f(\alpha) = 0\} = \langle p(x) \rangle$, $p(\alpha) = 0$ and for any $f(x) \in F[x]$ with $f(\alpha) = 0$, $p(x)$ divides $f(x)$.

You are asked to prove uniqueness in Exercise 38. ◆

39.15 Definition Let E be a field extension of a field F , and let $\alpha \in E$ be algebraic over F . The unique monic polynomial $p(x)$ in Corollary 39.14 is called the **irreducible polynomial for α over F** or the **minimal polynomial for α over F** , and it is denoted $\text{irr}(\alpha, F)$. The degree of the polynomial $\text{irr}(\alpha, F)$ is called the **degree of α over F** and this number is denoted by $\deg(\alpha, F)$. ■

39.16 Example We know that $\text{irr}(\sqrt{2}, \mathbb{Q}) = x^2 - 2$. Referring to Example 39.10, we see that for $\alpha = \sqrt{1 + \sqrt{3}}$ in \mathbb{R} , α is a zero of $x^4 - 2x^2 - 2$, which is in $\mathbb{Q}[x]$. Since $x^4 - 2x^2 - 2$ is irreducible over \mathbb{Q} (by Eisenstein with $p = 2$, or by application of the technique of Example 28.15), we see that

$$\text{irr}(\sqrt{1 + \sqrt{3}}, \mathbb{Q}) = x^4 - 2x^2 - 2.$$

Thus $\sqrt{1 + \sqrt{3}}$ is algebraic of degree 4 over \mathbb{Q} . ▲

Just as we must speak of an element α as *algebraic over F* rather than simply as *algebraic*, we must speak of the *degree of α over F* rather than the *degree of α* . To take a trivial illustration, $\sqrt{2} \in \mathbb{R}$ is algebraic of degree 2 over \mathbb{Q} but algebraic of degree 1 over \mathbb{R} , for $\text{irr}(\sqrt{2}, \mathbb{R}) = x - \sqrt{2}$.

The quick development of the theory here is due to the machinery of homomorphisms and ideal theory that we now have at our disposal. Note especially our constant use of the evaluation homomorphisms ϕ_α .

Simple Extensions

Let E be an extension field of a field F , and let $\alpha \in E$. Let ϕ_α be the evaluation homomorphism of $F[x]$ into E with $\phi_\alpha(a) = a$ for $a \in F$ and $\phi_\alpha(x) = \alpha$, as in Theorem 27.4. We use two cases to define the field $F(\alpha)$.

Case I Suppose α is algebraic over F . Then as in Corollary 39.14, the kernel of ϕ_α is $\langle \text{irr}(\alpha, F) \rangle$ and by Theorem 31.25, $\langle \text{irr}(\alpha, F) \rangle$ is a maximal ideal of $F[x]$. Therefore, $F[x]/\langle \text{irr}(\alpha, F) \rangle$ is a field and is isomorphic to the image $\phi_\alpha[F[x]]$ in E . This subfield $\phi_\alpha[F[x]]$ of E is then the smallest subfield of E containing F and α . We shall denote this field by $F(\alpha)$.

Case II Suppose α is transcendental over F . Then by Theorem 39.12, ϕ_α gives an isomorphism of $F[x]$ with a subdomain of E . Thus in this case $\phi_\alpha[F[x]]$ is not a field but an integral domain that we shall denote by $F[\alpha]$. By Corollary 26.9, E contains a field of quotients of $F[\alpha]$, which is thus the smallest subfield of E containing F and α . As in Case I, we denote this field by $F(\alpha)$.

39.17 Example Since π is transcendental over \mathbb{Q} , the field $\mathbb{Q}(\pi)$ is isomorphic to the field $\mathbb{Q}(x)$ of rational functions over \mathbb{Q} in the indeterminate x . Thus from a structural viewpoint, an element that is transcendental over a field F behaves as though it were an indeterminate over F . ▲

39.18 Definition An extension field E of a field F is a **simple extension of F** if $E = F(\alpha)$ for some $\alpha \in E$. ■

Many important results appear throughout this section. We have now developed so much machinery that results are starting to pour out of our efficient plant at an alarming

rate. The next theorem gives us insight into the nature of the field $F(\alpha)$ in the case where α is algebraic over F .

39.19 Theorem Let $E = F(\alpha)$ be a simple extension of a field F with α algebraic over F . Let $n = \deg(\alpha, F)$. Then every $\beta \in F(\alpha)$ can be uniquely expressed in the form

$$\beta = b_0 + b_1\alpha + b_2\alpha^2 + \cdots + b_{n-1}\alpha^{n-1},$$

where the b_i are in F .

Proof Let $\beta \in F(\alpha)$. Then $\beta = f(\alpha)$ for some polynomial $f(x) \in F[x]$ by the definition of $F[\alpha]$. The division algorithm says that there are unique polynomials $q(x), r(x) \in F[x]$ such that either $r(x) = 0$ or the degree of $r(x)$ is less than n , and

$$f(x) = \text{irr}(\alpha, F)q(x) + r(x).$$

Applying the evaluation homomorphism ϕ_α , we see that $f(\alpha) = r(\alpha)$. Thus

$$\beta = f(\alpha) = r(\alpha) = b_0 + b_1\alpha + b_2\alpha^2 + \cdots + b_{n-1}\alpha^{n-1}$$

for some elements b_i in F .

To show uniqueness, we assume that $s(x) \in F[x]$ is any polynomial with $r(\alpha) = s(\alpha)$, and $s(x)$ is either zero or else its degree is less than n . Let $d(x) = r(x) - s(x)$. Then $d(\alpha) = 0$, and either $d(x) = 0$ or $\deg(d(x)) < n$. Since the degree of the minimal polynomial for α over F is n , $d(x)$ is the zero polynomial and $r(x) = s(x)$. Thus the representation of β as

$$\beta = b_0 + b_1\alpha + b_2\alpha^2 + \cdots + b_{n-1}\alpha^{n-1},$$

where the b_i are in F , is unique. \blacklozenge

We give an impressive example illustrating Theorem 39.19.

39.20 Example The polynomial $p(x) = x^2 + x + 1$ in $\mathbb{Z}_2[x]$ is irreducible over \mathbb{Z}_2 by Theorem 28.11, since neither element 0 nor element 1 of \mathbb{Z}_2 is a zero of $p(x)$. By Theorem 39.3, we know that there is an extension field E of \mathbb{Z}_2 containing a zero α of $x^2 + x + 1$. By Theorem 39.19, $\mathbb{Z}_2(\alpha)$ has as elements $0 + 0\alpha, 1 + 0\alpha, 0 + 1\alpha$, and $1 + 1\alpha$, that is, 0, 1, α , and $1 + \alpha$. This gives us a new finite field, of four elements! The addition and multiplication tables for this field are shown in Tables 39.21 and 39.22. For example, to compute $(1 + \alpha)(1 + \alpha)$ in $\mathbb{Z}_2(\alpha)$, we observe that since $p(\alpha) = \alpha^2 + \alpha + 1 = 0$, then

$$\alpha^2 = -\alpha - 1 = \alpha + 1.$$

Therefore,

$$(1 + \alpha)(1 + \alpha) = 1 + \alpha + \alpha + \alpha^2 = 1 + \alpha^2 = 1 + \alpha + 1 = \alpha. \quad \blacktriangle$$

We can use Theorem 39.19 to fulfill our promise of Example 39.4 and show that $\mathbb{R}[x]/(x^2 + 1)$ is isomorphic to the field \mathbb{C} of complex numbers. We saw in Example 39.4 that we can view $\mathbb{R}[x]/(x^2 + 1)$ as an extension field of \mathbb{R} . Let

39.21 Table

+	0	1	α	$1 + \alpha$
0	0	1	α	$1 + \alpha$
1	1	0	$1 + \alpha$	α
α	α	$1 + \alpha$	0	1
$1 + \alpha$	$1 + \alpha$	α	1	0

39.22 Table

	0	1	α	$1 + \alpha$
0	0	0	0	0
1	0	1	α	$1 + \alpha$
α	0	α	$1 + \alpha$	1
$1 + \alpha$	0	$1 + \alpha$	1	α

$$\alpha = x + \langle x^2 + 1 \rangle.$$

Then $\mathbb{R}(\alpha) = \mathbb{R}[x]/\langle x^2 + 1 \rangle$ and consists of all elements of the form $a + b\alpha$ for $a, b \in \mathbb{R}$, by Theorem 39.19. But since $\alpha^2 + 1 = 0$, we see that α plays the role of $i \in \mathbb{C}$, and $a + b\alpha$ plays the role of $(a + bi) \in \mathbb{C}$. Thus $\mathbb{R}(\alpha) \cong \mathbb{C}$. *This is the elegant algebraic way to construct \mathbb{C} from \mathbb{R} .*

39.23 Corollary Let E be an extension field of F and let $\alpha \in E$ be algebraic over F . If $\deg(\alpha, F) = n$, then $F(\alpha)$ is a vector space over F with dimension n and basis $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$. Furthermore, every element β of $F(\alpha)$ is algebraic over F and $\deg(\beta, F) \leq \deg(\alpha, F)$.

Proof Since F is a subfield of $F(\alpha)$, $F(\alpha)$ is a vector space over F . Theorem 39.19 shows that the set $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ spans $F(\alpha)$. If

$$0 = b_0(1) + b_1\alpha + b_2\alpha^2 + \cdots + b_{n-1}\alpha^{n-1},$$

by uniqueness of the coefficients in Theorem 39.19, each b_i is 0. We have that $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ is linearly independent over F , and therefore a basis of $F(\alpha)$ over F . Thus the dimension of $F(\alpha)$ over F is $n = \deg(\alpha, F)$.

For any $\beta \in F(\alpha)$, $F \leq F(\beta) \leq F(\alpha)$, so any set of more than n vectors in $F(\beta)$ is not linearly independent over F . The set $\{1, \beta, \beta^2, \dots, \beta^n\}$ either has fewer than $n+1$ elements or else it is not linearly independent over F . In the first case, $\beta^r = \beta^s$ for some $r \neq s$ and in the second case, there are elements $b_i \in F$, not all zero, such that

$$b_0(1) + b_1\beta + b_2\beta^2 + \cdots + b_n\beta^n = 0.$$

In either case, we see that β is algebraic over F . Furthermore the dimension of $F(\beta)$ over F , k , is at most n and we have

$$\deg(\beta, F) = k \leq n = \deg(\alpha, F).$$



39.24 Example The number $i \in \mathbb{C}$ has minimal polynomial $x^2 + 1$ over \mathbb{R} and $\mathbb{C} = \mathbb{R}(i)$. By Corollary 39.23, for every complex number β , $\deg(\beta, \mathbb{R}) \leq 2$. This implies that every complex number that is not a real number is a zero of some irreducible polynomial of degree two in $\mathbb{R}[x]$. Of course, this fact can also be verified using the techniques of Example 39.10.



■ EXERCISES 39

Computations

In Exercises 1 through 5, show that the given number $\alpha \in \mathbb{C}$ is algebraic over \mathbb{Q} by finding $f(x) \in \mathbb{Q}[x]$ such that $f(\alpha) = 0$.

1. $1 + \sqrt{2}$

2. $\sqrt{2} + \sqrt{3}$

3. $1 + i$

4. $\sqrt[3]{1 + \sqrt{2}}$

5. $\sqrt[3]{2 - i}$

In Exercises 6 through 8, find $\text{irr}(\alpha, \mathbb{Q})$ and $\deg(\alpha, \mathbb{Q})$ for the given algebraic number $\alpha \in \mathbb{C}$. Be prepared to prove that your polynomials are irreducible over \mathbb{Q} if challenged to do so.

6. $\sqrt{3 - \sqrt{6}}$

7. $\sqrt{\left(\frac{1}{3}\right) + \sqrt{7}}$

8. $\sqrt{2} + i$

In Exercises 9 through 16, classify the given $\alpha \in \mathbb{C}$ as algebraic or transcendental over the given field F . If α is algebraic over F , find $\deg(\alpha, F)$.

9. $\alpha = i, F = \mathbb{Q}$
 10. $\alpha = 1 + i, F = \mathbb{R}$
 11. $\alpha = \sqrt{\pi}, F = \mathbb{Q}$
 12. $\alpha = \sqrt{\pi}, F = \mathbb{R}$
 13. $\alpha = \sqrt{\pi}, F = \mathbb{Q}(\pi)$
 14. $\alpha = \pi^2, F = \mathbb{Q}$
 15. $\alpha = \pi^2, F = \mathbb{Q}(\pi)$
 16. $\alpha = \pi^2, F = \mathbb{Q}(\pi^3)$
17. Refer to Example 39.20 of the text. The polynomial $x^2 + x + 1$ has a zero α in $\mathbb{Z}_2(\alpha)$ and thus must factor into a product of linear factors in $(\mathbb{Z}_2(\alpha))[x]$. Find this factorization. [Hint: Divide $x^2 + x + 1$ by $x - \alpha$ by long division, using the fact that $\alpha^2 = \alpha + 1$.]
18. a. Show that the polynomial $x^2 + 1$ is irreducible in $\mathbb{Z}_3[x]$.
 b. Let α be a zero of $x^2 + 1$ in an extension field of \mathbb{Z}_3 . As in Example 39.20, give the multiplication and addition tables for the nine elements of $\mathbb{Z}_3(\alpha)$, written in the order 0, 1, 2, α , 2α , 1 + α , 1 + 2α , 2 + α , and 2 + 2α .

Concepts

In Exercises 19 through 22, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

19. An element α of an extension field E of a field F is *algebraic over F* if and only if α is a zero of some polynomial.
20. An element β of an extension field E of a field F is *transcendental over F* if and only if β is not a zero of any polynomial in $F[x]$.
21. A *monic polynomial* in $F[x]$ is one having all coefficients equal to 1.
22. A field E is a *simple extension* of a subfield F if and only if there exists some $\alpha \in E$ such that no proper subfield of E contains α .
23. Determine whether each of the following is true or false.
 - a. The number π is transcendental over \mathbb{Q} .
 - b. \mathbb{C} is a simple extension of \mathbb{R} .
 - c. Every element of a field F is algebraic over F .
 - d. \mathbb{R} is an extension field of \mathbb{Q} .
 - e. \mathbb{Q} is an extension field of \mathbb{Z}_2 .
 - f. Let $\alpha \in \mathbb{C}$ be algebraic over \mathbb{Q} of degree n . If $f(\alpha) = 0$ for nonzero $f(x) \in \mathbb{Q}[x]$, then $\deg(f(x)) \geq n$.
 - g. Let $\alpha \in \mathbb{C}$ be algebraic over \mathbb{Q} of degree n . If $f(\alpha) = 0$ for nonzero $f(x) \in \mathbb{R}[x]$, then $\deg(f(x)) \geq n$.
 - h. Every nonconstant polynomial in $F[x]$ has a zero in some extension field of F .
 - i. Every nonconstant polynomial in $F[x]$ has a zero in every extension field of F .
 - j. If x is an indeterminate, $\mathbb{Q}[\pi] \cong \mathbb{Q}[x]$.
24. We have stated without proof that π and e are transcendental over \mathbb{Q} .
 - a. Find a subfield F of \mathbb{R} such that π is algebraic of degree 3 over F .
 - b. Find a subfield E of \mathbb{R} such that e^2 is algebraic of degree 5 over E .
25. a. Show that $x^3 + x^2 + 1$ is irreducible over \mathbb{Z}_2 .
 b. Let α be a zero of $x^3 + x^2 + 1$ in an extension field of \mathbb{Z}_2 . Show that $x^3 + x^2 + 1$ factors into three linear factors in $(\mathbb{Z}_2(\alpha))[x]$ by actually finding this factorization. [Hint: Every element of $\mathbb{Z}_2(\alpha)$ is of the form $a_0 + a_1\alpha + a_2\alpha^2$ for $a_i = 0, 1$.]

Divide $x^3 + x^2 + 1$ by $x - \alpha$ by long division. Show that the quotient also has a zero in $\mathbb{Z}_2(\alpha)$ by simply trying the eight possible elements. Then complete the factorization.]

26. Let E be an extension field of \mathbb{Z}_2 and let $\alpha \in E$ be algebraic of degree 3 over \mathbb{Z}_2 . Classify the groups $(\mathbb{Z}_2(\alpha), +)$ and $((\mathbb{Z}_2(\alpha))^*, \cdot)$ according to the Fundamental Theorem of finitely generated abelian groups. As usual, $(\mathbb{Z}_2(\alpha))^*$ is the set of nonzero elements of $\mathbb{Z}_2(\alpha)$.
27. Definition 39.15 defined the terms **irreducible polynomial for α over F** and **minimal polynomial for α over F** to mean the same polynomial. Why are both designations appropriate?

Proof Synopsis

28. Give a two- or three-sentence synopsis of Theorem 39.3.

Theory

29. Let E be an extension field of F , and let $\alpha, \beta \in E$. Suppose α is transcendental over F but algebraic over $F(\beta)$. Show that β is algebraic over $F(\alpha)$.
30. Let E be an extension field of a finite field F , where F has q elements. Let $\alpha \in E$ be algebraic over F of degree n . Prove that $F(\alpha)$ has q^n elements.
31. a. Show that there exists an irreducible polynomial of degree 3 in $\mathbb{Z}_3[x]$.
 b. Show from part (a) that there exists a finite field of 27 elements. [Hint: Use Exercise 30.]
32. Consider the prime field \mathbb{Z}_p of characteristic $p \neq 0$.
 a. Show that, for $p \neq 2$, not every element in \mathbb{Z}_p is a square of an element of \mathbb{Z}_p . [Hint: $1^2 = (p-1)^2 = 1$ in \mathbb{Z}_p . Deduce the desired conclusion by counting.]
 b. Using part (a), show that there exist finite fields of p^2 elements for every prime p in \mathbb{Z}^+ .
33. Let E be an extension field of a field F and let $\alpha \in E$ be transcendental over F . Show that every element of $F(\alpha)$ that is not in F is also transcendental over F .
34. Show that $\{a + b(\sqrt[3]{2}) + c(\sqrt[3]{2})^2 \mid a, b, c \in \mathbb{Q}\}$ is a subfield of \mathbb{R} by using the ideas of this section, rather than by a formal verification of the field axioms. [Hint: Use Theorem 39.19.]
35. Following the idea of Exercise 31, show that there exists a field of 8 elements; of 16 elements; of 25 elements.
36. Let F be a finite field of characteristic p . Show that every element of F is algebraic over the prime field $\mathbb{Z}_p \leq F$. [Hint: Let F^* be the set of nonzero elements of F . Apply group theory to the group $\langle F^*, \cdot \rangle$ to show that every $\alpha \in F^*$ is a zero of some polynomial in $\mathbb{Z}_p[x]$ of the form $x^n - 1$.]
37. Use Exercises 30 and 36 to show that every finite field is of prime-power order, that is, it has a prime-power number of elements.
38. Prove the uniqueness of the polynomial in Corollary 39.14.

SECTION 40 ALGEBRAIC EXTENSIONS

Finite Extensions

In Corollary 39.23 we saw that if E is an extension field of a field F and $\alpha \in E$ is algebraic over F , then every element of $F(\alpha)$ is algebraic over F . In studying zeros of polynomials in $F[x]$, we shall be interested almost exclusively in extensions of F containing only elements algebraic over F .

40.1 Definition An extension field E of a field F is an **algebraic extension of F** if every element in E is algebraic over F . ■

40.2 Definition If an extension field E of a field F is of finite dimension n as a vector space over F , then E is a **finite extension of degree n over F** . We shall let $[E : F]$ be the degree n of E over F . ■

To say that a field E is a finite extension of a field F does *not* mean that E is a finite field. It just asserts that E is a finite-dimensional vector space over F , that is, that $[E : F]$ is finite.

We shall often use the fact that if E is a finite extension of F , then, $[E : F] = 1$ if and only if $E = F$. We need only observe that by Theorem 33.18, $\{1\}$ can always be enlarged to a basis for E over F . Thus $[E : F] = 1$ if and only if $E = F(1) = F$.

We show that a finite extension E of a field F must be an algebraic extension of F .

40.3 Theorem A finite extension field E of a field F is an algebraic extension of F .

Proof We must show that for $\alpha \in E$, α is algebraic over F . By Theorem 33.18 if $[E : F] = n$, then

$$1, \alpha, \dots, \alpha^n$$

cannot be linearly independent elements, so there exist $a_i \in F$ such that

$$a_n\alpha^n + \dots + a_1\alpha + a_0 = 0,$$

and not all $a_i = 0$. Then $f(x) = a_nx^n + \dots + a_1x + a_0$ is a nonzero polynomial in $F[x]$, and $f(\alpha) = 0$. Therefore, α is algebraic over F . \blacklozenge

We cannot overemphasize the importance of our next theorem. It plays a role in field theory analogous to the role of the theorem of Lagrange in group theory. While its proof follows easily from our brief work with vector spaces, it is a tool of incredible power. An elegant application of it in the section that follows shows the impossibility of performing certain geometric constructions with a straightedge and a compass. *Never underestimate a theorem that counts something.*

40.4 Theorem If E is a finite extension field of a field F , and K is a finite extension field of E , then K is a finite extension of F , and

$$[K : F] = [K : E][E : F].$$

Proof Let $\{\alpha_i \mid i = 1, \dots, n\}$ be a basis for E as a vector space over F , and let the set $\{\beta_j \mid j = 1, \dots, m\}$ be a basis for K as a vector space over E . The theorem will be proved if we can show that the mn elements $\alpha_i\beta_j$ form a basis for K , viewed as a vector space over F . (See Fig. 40.5.)

Let γ be any element of K . Since the β_j form a basis for K over E , we have

$$\gamma = \sum_{j=1}^m b_j \beta_j$$

for some $b_j \in E$. Since the α_i form a basis for E over F , we have

$$b_j = \sum_{i=1}^n a_{ij} \alpha_i$$

for some $a_{ij} \in F$. Then

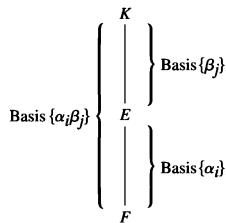
$$\gamma = \sum_{j=1}^m \left(\sum_{i=1}^n a_{ij} \alpha_i \right) \beta_j = \sum_{i,j} a_{ij} (\alpha_i \beta_j),$$

so the mn vectors $\alpha_i\beta_j$ span K over F .

It remains for us to show that the mn elements $\alpha_i\beta_j$ are independent over F . Suppose that $\sum_{i,j} c_{ij} (\alpha_i \beta_j) = 0$, with $c_{ij} \in F$. Then

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

and $(\sum_{i=1}^n c_{ij} \alpha_i) \in E$. Since the elements β_j are independent over E , we must have



40.5 Figure

$$\sum_{i=1}^n c_{ij}\alpha_i = 0$$

for all j . But now the α_i are independent over F , so $\sum_{i=1}^n c_{ij}\alpha_i = 0$ implies that $c_{ij} = 0$ for all i and j . Thus the $\alpha_i\beta_j$ not only span K over F but also are independent over F . Thus they form a basis for K over F . \blacklozenge

Note that we proved this theorem by actually exhibiting a basis. It is worth remembering that if $\{\alpha_i \mid i = 1, \dots, n\}$ is a basis for E over F and $\{\beta_j \mid j = 1, \dots, m\}$ is a basis for K over E , for fields $F \leq E \leq K$, then the set $\{\alpha_i\beta_j\}$ of mn products is a basis for K over F . Figure 40.5 gives a diagram for this situation. We shall illustrate this further in a moment.

40.6 Corollary If F_i is a field for $i = 1, \dots, r$ and F_{i+1} is a finite extension of F_i , then F_r is a finite extension of F_1 , and

$$[F_r : F_1] = [F_r : F_{r-1}][F_{r-1} : F_{r-2}] \cdots [F_2 : F_1].$$

Proof The proof is a straightforward extension of Theorem 40.4 by induction. \blacklozenge

40.7 Corollary If E is an extension field of F , $\alpha \in E$ is algebraic over F , and $\beta \in F(\alpha)$, then $\deg(\beta, F)$ divides $\deg(\alpha, F)$.

Proof By Corollary 39.23, $\deg(\alpha, F) = [F(\alpha) : F]$ and $\deg(\beta, F) = [F(\beta) : F]$. We have $F \leq F(\beta) \leq F(\alpha)$, so by Corollary 40.6 $[F(\beta) : F]$ divides $[F(\alpha) : F]$. \blacklozenge

The following example illustrates a type of argument one often makes using Theorem 40.4 or its corollaries.

40.8 Example By Corollary 40.7, there is no element of $\mathbb{Q}(\sqrt{2})$ that is a zero of $x^3 - 2$. Note that $\deg(\sqrt{2}, \mathbb{Q}) = 2$, while a zero of $x^3 - 2$ is of degree 3 over \mathbb{Q} , but 3 does not divide 2. \blacktriangle

Let E be an extension field of a field F , and let α_1, α_2 be elements of E , not necessarily algebraic over F . By definition, $F(\alpha_1)$ is the smallest extension field of F in E that contains α_1 . Similarly, $(F(\alpha_1))(\alpha_2)$ can be characterized as the smallest extension field of F in E containing both α_1 and α_2 . We could equally have started with α_2 , so $(F(\alpha_1))(\alpha_2) = (F(\alpha_2))(\alpha_1)$. We denote this field by $F(\alpha_1, \alpha_2)$. Similarly, for $\alpha_i \in E$, $F(\alpha_1, \dots, \alpha_n)$ is the smallest extension field of F in E containing all the α_i for $i = 1, \dots, n$. We obtain the field $F(\alpha_1, \dots, \alpha_n)$ from the field F by **joining to F the elements α_i in E** . Exercise 51 of Section 22 shows that, analogous to an intersection of subgroups of a group, an intersection of subfields of a field E is again a subfield of E . Thus $F(\alpha_1, \dots, \alpha_n)$ can be characterized as the intersection of all subfields of E containing F and all the α_i for $i = 1, \dots, n$.

40.9 Example Consider $\mathbb{Q}(\sqrt{2})$. Corollary 39.23 shows that $\{1, \sqrt{2}\}$ is a basis for $\mathbb{Q}(\sqrt{2})$ over \mathbb{Q} . Using the technique demonstrated in Example 39.10, we can easily discover that $\sqrt{2} + \sqrt{3}$ is a zero of $x^4 - 10x^2 + 1$. By the method demonstrated in Example 28.15, we can show that this polynomial is irreducible in $\mathbb{Q}[x]$. Thus $\text{irr}(\sqrt{2} + \sqrt{3}, \mathbb{Q}) = x^4 - 10x^2 + 1$, so $[\mathbb{Q}(\sqrt{2} + \sqrt{3}) : \mathbb{Q}] = 4$. Thus $(\sqrt{2} + \sqrt{3}) \notin \mathbb{Q}(\sqrt{2})$, so $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. Consequently, $\{1, \sqrt{3}\}$ is a basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = (\mathbb{Q}(\sqrt{2}))(\sqrt{3})$ over $\mathbb{Q}(\sqrt{2})$. The proof of Theorem 40.4 (see the comment following the theorem) then shows that $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ is a basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over \mathbb{Q} . \blacktriangle

40.10 Example Let $2^{1/3}$ be the real cube root of 2 and $2^{1/2}$ be the positive square root of 2. Then $2^{1/2} \notin \mathbb{Q}(2^{1/3})$ because $\deg(2^{1/2}, \mathbb{Q}) = 2$ and 2 is not a divisor of 3 = $\deg(2^{1/3}, \mathbb{Q})$. Thus $[\mathbb{Q}(2^{1/3}, 2^{1/2}) : \mathbb{Q}(2^{1/3})] = 2$. Hence $\{1, 2^{1/3}, 2^{1/2}\}$ is a basis for $\mathbb{Q}(2^{1/3})$ over \mathbb{Q} and $\{1, 2^{1/2}\}$ is a basis for $\mathbb{Q}(2^{1/3}, 2^{1/2})$ over $\mathbb{Q}(2^{1/3})$. Furthermore, by Theorem 40.4 (see the comment following the theorem),

$$\{1, 2^{1/2}, 2^{1/3}, 2^{5/6}, 2^{2/3}, 2^{7/6}\}$$

is a basis for $\mathbb{Q}(2^{1/2}, 2^{1/3})$ over \mathbb{Q} . Because $2^{7/6} = 2(2^{1/6})$, we have $2^{1/6} \in \mathbb{Q}(2^{1/2}, 2^{1/3})$. Now $2^{1/6}$ is a zero of $x^6 - 2$, which is irreducible over \mathbb{Q} , by Eisenstein's criterion, with $p = 2$. Thus

$$\mathbb{Q} \leq \mathbb{Q}(2^{1/6}) \leq \mathbb{Q}(2^{1/2}, 2^{1/3})$$

and by Theorem 40.4

$$\begin{aligned} 6 &= [\mathbb{Q}(2^{1/2}, 2^{1/3}) : \mathbb{Q}] = [\mathbb{Q}(2^{1/2}, 2^{1/3}) : \mathbb{Q}(2^{1/6})][\mathbb{Q}(2^{1/6}) : \mathbb{Q}] \\ &= [\mathbb{Q}(2^{1/2}, 2^{1/3}) : \mathbb{Q}(2^{1/6})](6). \end{aligned}$$

Therefore, we must have

$$[\mathbb{Q}(2^{1/2}, 2^{1/3}) : \mathbb{Q}(2^{1/6})] = 1,$$

so $\mathbb{Q}(2^{1/2}, 2^{1/3}) = \mathbb{Q}(2^{1/6})$, by the comment preceding Theorem 40.3. \blacktriangle

Example 40.10 shows that it is possible for an extension $F(\alpha_1, \dots, \alpha_n)$ of a field F to be actually a simple extension, even though $n > 1$.

Let us characterize extensions of F of the form $F(\alpha_1, \dots, \alpha_n)$ in the case that all the α_i are algebraic over F .

40.11 Theorem Let E be an algebraic extension of a field F . Then there exist a finite number of elements $\alpha_1, \dots, \alpha_n$ in E such that $E = F(\alpha_1, \dots, \alpha_n)$ if and only if E is a finite-dimensional vector space over F , that is, if and only if E is a finite extension of F .

Proof Suppose that $E = F(\alpha_1, \dots, \alpha_n)$. Since E is an algebraic extension of F , each α_i is algebraic over F , so each α_i is algebraic over every extension field of F in E . Thus $F(\alpha_1)$ is algebraic over F , and in general, $F(\alpha_1, \dots, \alpha_j)$ is algebraic over $F(\alpha_1, \dots, \alpha_{j-1})$ for $j = 2, \dots, n$. Corollary 40.6 applied to the sequence of finite extensions

$$F, F(\alpha_1), F(\alpha_1, \alpha_2), \dots, F(\alpha_1, \dots, \alpha_n) = E$$

then shows that E is a finite extension of F .

Conversely, suppose that E is a finite algebraic extension of F . If $[E : F] = 1$, then $E = F(1) = F$, and we are done. If $E \neq F$, let $\alpha_1 \in E$, where $\alpha_1 \notin F$. Then $[F(\alpha_1) : F] > 1$. If $F(\alpha_1) = E$, we are done; if not, let $\alpha_2 \in E$, where $\alpha_2 \notin F(\alpha_1)$. Continuing this process, we see from Theorem 40.4 that since $[E : F]$ is finite, we must arrive at α_n such that

$$F(\alpha_1, \dots, \alpha_n) = E.$$



Algebraically Closed Fields and Algebraic Closures

We have not yet observed that if E is an extension of a field F and $\alpha, \beta \in E$ are algebraic over F , then so are $\alpha + \beta, \alpha\beta, \alpha - \beta$, and α/β , if $\beta \neq 0$. This follows from Theorem 40.3 and is also included in the following theorem.

40.12 Theorem Let E be an extension field of F . Then

$$\bar{F}_E = \{\alpha \in E \mid \alpha \text{ is algebraic over } F\}$$

is a subfield of E , the **algebraic closure of F in E** .

Proof Let $\alpha, \beta \in \bar{F}_E$. Then Theorem 40.11 shows that $F(\alpha, \beta)$ is a finite extension of F , and by Theorem 40.3 every element of $F(\alpha, \beta)$ is algebraic over F , that is, $F(\alpha, \beta) \subseteq \bar{F}_E$. Thus \bar{F}_E contains $\alpha + \beta, \alpha\beta, \alpha - \beta$, and also contains α/β for $\beta \neq 0$, so \bar{F}_E is a subfield of E . \blacklozenge

40.13 Corollary The set of all algebraic numbers forms a field.

Proof Proof of this corollary is immediate from Theorem 40.12, because the set of all algebraic numbers is the algebraic closure of \mathbb{Q} in \mathbb{C} . \blacklozenge

It is well known that the complex numbers have the property that every nonconstant polynomial in $\mathbb{C}[x]$ has a zero in \mathbb{C} . This is known as the *Fundamental Theorem of Algebra*. An analytic proof of this theorem is given in Theorem 40.18. We now give a definition generalizing this important concept to other fields.

40.14 Definition A field F is **algebraically closed** if every nonconstant polynomial in $F[x]$ has a zero in F . \blacksquare

Note that a field F can be the algebraic closure of F in an extension field E without F being algebraically closed. For example, \mathbb{Q} is the algebraic closure of \mathbb{Q} in $\mathbb{Q}(x)$, but \mathbb{Q} is not algebraically closed because $x^2 + 1$ has no zero in \mathbb{Q} .

The next theorem shows that the concept of a field being algebraically closed can also be defined in terms of factorization of polynomials over the field.

40.15 Theorem A field F is algebraically closed if and only if every nonconstant polynomial in $F[x]$ factors in $F[x]$ into linear factors.

Proof Let F be algebraically closed, and let $f(x)$ be a nonconstant polynomial in $F[x]$. Then $f(x)$ has a zero $a \in F$. By Corollary 28.4, $x - a$ is a factor of $f(x)$, so $f(x) = (x - a)g(x)$. Then if $g(x)$ is nonconstant, it has a zero $b \in F$, and we have $f(x) = (x - a)(x - b)h(x)$. Continuing, we get a factorization of $f(x)$ in $F[x]$ into linear factors.

Conversely, suppose that every nonconstant polynomial of $F[x]$ has a factorization into linear factors. If $ax - b$ is a linear factor of $f(x)$, then b/a is a zero of $f(x)$. Thus F is algebraically closed. \blacklozenge

40.16 Corollary An algebraically closed field F has no proper algebraic extensions, that is, no algebraic extensions E with $F < E$.

Proof Let E be an algebraic extension of F , so $F \leq E$. Then if $\alpha \in E$, we have $\text{irr}(\alpha, F) = x - \alpha$, by Theorem 40.15, since F is algebraically closed. Thus $\alpha \in F$, and we must have $F = E$. \blacklozenge

In a moment we shall show that just as there exists an algebraically closed extension \mathbb{C} of the real numbers \mathbb{R} , for any field F there exists similarly an algebraic extension \bar{F} of F , with the property that \bar{F} is algebraically closed. Naively, to find \bar{F} we proceed

as follows. If a polynomial $f(x)$ in $F[x]$ has no zero in F , then adjoin a zero α of such an $f(x)$ to F , thus obtaining the field $F(\alpha)$. *Theorem 39.3, Kronecker's theorem, is strongly used here, of course.* If $F(\alpha)$ is still not algebraically closed, then continue the process further. The trouble is that, contrary to the situation for the algebraic closure \mathbb{C} of \mathbb{R} , we may have to do this a (possibly large) infinite number of times. It can be shown (see Exercises 33 and 36) that $\overline{\mathbb{Q}}$ is isomorphic to the field of all algebraic numbers, and that we cannot obtain $\overline{\mathbb{Q}}$ from \mathbb{Q} by adjoining a finite number of algebraic numbers. We shall have to first discuss some set-theoretic machinery, *Zorn's lemma*, in order to be able to handle such a situation. This machinery is a bit complex, so we are putting the proof under a separate heading. The existence theorem for \overline{F} is very important, and we state it here so that we will know this fact, even if we do not study the proof.

40.17 Theorem Every field F has an **algebraic closure**, that is, an algebraic extension \overline{F} that is algebraically closed.

The Fundamental Theorem of Algebra stating that \mathbb{C} is an algebraically closed field is well known. Interestingly, the simplest and shortest proofs of the Fundamental Theorem of Algebra are not algebraic proofs. There are much shorter and easier-to-follow proofs using either analysis or topology. However, behind both the analytical and the topological proofs lies a significant amount of machinery, so perhaps comparing the proofs is not completely fair. At any rate, we include a short proof for students who have studied functions of a complex variable and are familiar with Liouville's Theorem.

40.18 Theorem (Fundamental Theorem of Algebra) The field \mathbb{C} of complex numbers is an algebraically closed field.

Proof Let the polynomial $f(z) \in \mathbb{C}[z]$ have no zero in \mathbb{C} . Then $1/f(z)$ gives an entire function; that is, $1/f(z)$ is analytic everywhere. Also if $f(z) \notin \mathbb{C}$, $\lim_{|z| \rightarrow \infty} |f(z)| = \infty$, so $\lim_{|z| \rightarrow \infty} |1/f(z)| = 0$. Thus $1/f(z)$ must be bounded in the plane. Hence by Liouville's theorem of complex function theory, $1/f(z)$ is constant, and thus $f(z)$ is constant. Therefore, a nonconstant polynomial in $\mathbb{C}[z]$ must have a zero in \mathbb{C} , so \mathbb{C} is algebraically closed. ◆

Proof of the Existence of an Algebraic Closure

We shall prove that every field has an algebraic extension that is algebraically closed. Mathematics students should have the opportunity to see some proof involving the *Axiom of Choice* by the time they finish college. This is a natural place for such a proof. We shall use an equivalent form, *Zorn's lemma*, of the Axiom of Choice. To state Zorn's lemma, we have to give a set-theoretic definition.

40.19 Definition A **partial ordering of a set S** is given by a relation \leq defined for certain ordered pairs of elements of S such that the following conditions are satisfied:

1. $a \leq a$ for all $a \in S$ (**reflexive law**).
2. If $a \leq b$ and $b \leq a$, then $a = b$ (**antisymmetric law**).
3. If $a \leq b$ and $b \leq c$, then $a \leq c$ (**transitive law**). ■

In a *partially ordered set*, not every two elements need be **comparable**; that is, for $a, b \in S$, we need not have either $a \leq b$ or $b \leq a$. As usual, $a < b$ denotes $a \leq b$ but $a \neq b$.

A subset T of a partially ordered set S is a **chain** if every two elements a and b in T are comparable, that is, either $a \leq b$ or $b \leq a$ (or both). An element $u \in S$ is an

upper bound for a subset A of partially ordered set S if $a \leq u$ for all $a \in A$. Finally, an element m of a partially ordered set S is **maximal** if there is no $s \in S$ such that $m < s$.

40.20 Example The collection of all subsets of a set forms a partially ordered set under the relation \leq given by \subseteq . For example, if the whole set is \mathbb{R} , we have $\mathbb{Z} \subseteq \mathbb{Q}$. Note, however, that for \mathbb{Z} and \mathbb{Q}^+ , neither $\mathbb{Z} \subseteq \mathbb{Q}^+$ nor $\mathbb{Q}^+ \subseteq \mathbb{Z}$. \blacktriangleleft

40.21 Zorn's Lemma If S is a partially ordered set such that every chain in S has an upper bound in S , then S has at least one maximal element.

We do not prove Zorn's lemma. Instead we point out that it can be shown that Zorn's lemma is equivalent to the Axiom of Choice. Thus we are really taking Zorn's lemma here as an *axiom* for our set theory. Refer to the literature for a statement of the Axiom of Choice and a proof of its equivalence to Zorn's lemma. (See Edgerton [47].)

Zorn's lemma is often useful when we want to show the existence of a largest or maximal structure of some kind. If a field F has an algebraic extension \bar{F} that is algebraically closed, then \bar{F} will certainly be a maximal algebraic extension of F , for since \bar{F} is algebraically closed, it can have no proper algebraic extensions.

The idea of our proof of Theorem 40.17 is very simple. Given a field F , we shall first describe a class of algebraic extensions of F that is so large that it must contain (up to isomorphism) any conceivable algebraic extension of F . We then define a partial ordering, the ordinary subfield ordering, on this class, and show that the hypotheses of Zorn's lemma are satisfied. By Zorn's lemma, there will exist a maximal algebraic extension \bar{F} of F in this class. We shall then argue that, as a maximal element, this extension \bar{F} can have no proper algebraic extensions, so it must be algebraically closed.

Our proof differs a bit from the one found in many texts. We like it because it uses no algebra other than that derived from Theorems 39.3 and 40.4. Thus it throws into sharp relief the tremendous strength of both Kronecker's theorem and Zorn's lemma. The proof looks long, but only because we are writing out every little step. To the professional mathematician, the construction of the proof from the information in the preceding paragraph is a routine matter. This proof was suggested to the author during his graduate student days by a fellow graduate student, Norman Shapiro, who also had a strong preference for it.

We are now ready to carry out our proof of Theorem 40.17, which we restate here.

40.22 Restated Theorem 40.17 Every field F has an algebraic closure \bar{F} .

Proof It can be shown in set theory that given any set, there exists a set with *strictly more* elements. Suppose we form a set

$$A = \{\omega_f \mid f \in F[x]; i = 0, \dots, (\text{degree } f)\}$$

that has an element for every possible zero of any $f(x) \in F[x]$. Let Ω be a set with strictly more elements than A . Replacing Ω by $\Omega \cup F$ if necessary, we can assume $F \subset \Omega$. Consider all possible fields that are algebraic extensions of F and that, as sets, consist of elements of Ω . One such algebraic extension is F itself. If E is any extension field of F , and if $\gamma \in E$ is a zero $f(\gamma) \in F[x]$ for $\gamma \notin F$ and $\deg(f, F) = n$, then renaming γ by ω for $\omega \in \Omega$ and $\omega \notin F$, and renaming elements $a_0 + a_1\gamma + \dots + a_{n-1}\gamma^{n-1}$ of $F(\gamma)$ by distinct elements of Ω as the a_i range over F , we can consider our renamed $F(\gamma)$ to be an algebraic extension field $F(\omega)$ of F , with $F(\omega) \subset \Omega$ and $f(\omega) = 0$. The set Ω has enough elements to form $F(\omega)$, since Ω has more than enough elements to provide n different zeros for each element of each degree n in any subset of $F[x]$.

All algebraic extension fields E_j of F , with $E_j \subseteq \Omega$, form a set

$$S = \{E_j \mid j \in J\}$$

■ HISTORICAL NOTE

The Axiom of Choice, although used implicitly in the 1870s and 1880s, was first stated explicitly by Ernst Zermelo in 1904 in connection with his proof of the well-ordering theorem, the result that for any set A , there exists an order-relation $<$ such that every nonempty subset B of A contains a least element with respect to $<$. Zermelo's Axiom of Choice asserted that, given any set M and the set S of all subsets of M , there always exists a "choice" function, a function $f : S \rightarrow M$ such that $f(M') \in M'$ for every $M' \in S$. Zermelo noted, in fact, that "this logical principle cannot... be reduced to a still simpler one, but it is applied without hesitation everywhere in mathematical deduction." A few years later he included this axiom in his collection of axioms for set theory, a collection

that was slightly modified in 1930 into what is now called Zermelo–Fraenkel set theory, the axiom system generally used today as a basis of that theory.

Zorn's lemma was introduced by Max Zorn (1906–1993) in 1935. Although he realized that it was equivalent to the well-ordering theorem (itself equivalent to the Axiom of Choice), he claimed that his lemma was more natural to use in algebra because the well-ordering theorem was somehow a "transcendental" principle. Other mathematicians soon agreed with his reasoning. The lemma appeared in 1939 in the first volume of Nicolas Bourbaki's *Éléments de Mathématique: Les Structures Fondamentales de l'Analyse*. It was used consistently in that work and quickly became an essential part of the mathematician's toolbox.

that is partially ordered under our usual subfield inclusion \leq . One element of S is F itself. The preceding paragraphs shows that if F is far away from being algebraically closed, there will be many fields E_j in S .

Let $T = \{E_{j_k}\}$ be a chain in S , and let $W = \cup_k E_{j_k}$. We now make W into a field. Let $\alpha, \beta \in W$. Then there exist $E_{j_1}, E_{j_2} \in S$, with $\alpha \in E_{j_1}$ and $\beta \in E_{j_2}$. Since T is a chain, one of the fields E_{j_1} and E_{j_2} is a subfield of the other, say $E_{j_1} \leq E_{j_2}$. Then $\alpha, \beta \in E_{j_2}$, and we use the field operations of E_{j_2} to define the sum of α and β in W as $(\alpha + \beta) \in E_{j_2}$ and, likewise, the product as $(\alpha\beta) \in E_{j_2}$. These operations are well defined in W ; they are independent of our choice of E_{j_2} , since if $\alpha, \beta \in E_{j_3}$ also, for E_{j_3} in T , then one of the fields E_{j_2} and E_{j_3} is a subfield of the other, since T is a chain. Thus we have operations of addition and multiplication defined on W .

All the field axioms for W under these operations now follow from the fact that these operations were defined in terms of addition and multiplication in fields. Thus, for example, $1 \in F$ serves as multiplicative identity in W , since for $\alpha \in W$, if $1, \alpha \in E_{j_1}$, then we have $1\alpha = \alpha$ in E_{j_1} , so $1\alpha = \alpha$ in W , by definition of multiplication in W . As further illustration, to check the distributive laws, let $\alpha, \beta, \gamma \in W$. Since T is a chain, we can find one field in T containing all three elements α, β , and γ , and in this field the distributive laws for α, β , and γ hold. Thus they hold in W . Therefore, we can view W as a field, and by construction, $E_{j_k} \leq W$ for every $E_{j_k} \in T$.

If we can show that W is algebraic over F , then $W \in S$ will be an upper bound for T . But if $\alpha \in W$, then $\alpha \in E_{j_1}$ for some E_{j_1} in T , so α is algebraic over F . Hence W is an algebraic extension of F and is an upper bound for T .

The hypotheses of Zorn's lemma are thus fulfilled, so there is a maximal element \bar{F} of S . We claim that \bar{F} is algebraically closed. Let $f(x) \in \bar{F}[x]$, where $f(x) \notin \bar{F}$. Suppose that $f(x)$ has no zero in \bar{F} . Since Ω has many more elements than \bar{F} has, we can take $\omega \in \Omega$, where $\omega \notin \bar{F}$, and form a field $\bar{F}(\omega) \subseteq \Omega$, with ω a zero of $f(x)$, as we saw in the first paragraph of this proof. Let β be in $\bar{F}(\omega)$. Then by Theorem 39.19, β is a zero of a polynomial

$$g(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$$

in $\bar{F}[x]$, with $\alpha_i \in \bar{F}$, and hence each α_i is algebraic over F . Then by Theorem 40.11 the field $F(\alpha_0, \dots, \alpha_n)$ is a finite extension of F , and since β is algebraic over $F(\alpha_0, \dots, \alpha_n)$, we also see that $F(\alpha_0, \dots, \alpha_n, \beta)$ is a finite extension over $F(\alpha_0, \dots, \alpha_n)$. Theorem 40.4 then shows that $F(\alpha_0, \dots, \alpha_n, \beta)$ is a finite extension of F , so by Theorem 40.3, β is algebraic over F . Hence $\bar{F}(\omega) \in S$ and $\bar{F} < \bar{F}(\omega)$, which contradicts the choice of \bar{F} as maximal in S . Thus $f(x)$ must have had a zero in \bar{F} , so \bar{F} is algebraically closed. \blacklozenge

The mechanics of the preceding proof are routine to the professional mathematician. Since it may be the first proof that we have ever seen using Zorn's lemma, we wrote the proof out in detail.

■ EXERCISES 40

Computations

In Exercises 1 through 13, find the degree and a basis for the given field extension. Be prepared to justify your answers.

1. $\mathbb{Q}(\sqrt{2})$ over \mathbb{Q}
2. $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over \mathbb{Q}
3. $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{18})$ over \mathbb{Q}
4. $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3})$ over \mathbb{Q}
5. $\mathbb{Q}(\sqrt{2}, \sqrt[3]{2})$ over \mathbb{Q}
6. $\mathbb{Q}(\sqrt{2} + \sqrt{3})$ over \mathbb{Q}
7. $\mathbb{Q}(\sqrt{2}\sqrt{3})$ over \mathbb{Q}
8. $\mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$ over \mathbb{Q}
9. $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{6}, \sqrt[3]{24})$ over \mathbb{Q}
10. $\mathbb{Q}(\sqrt{2}, \sqrt{6})$ over $\mathbb{Q}(\sqrt{3})$
11. $\mathbb{Q}(\sqrt{2} + \sqrt{3})$ over $\mathbb{Q}(\sqrt{3})$
12. $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}(\sqrt{2} + \sqrt{3})$
13. $\mathbb{Q}(\sqrt{2}, \sqrt{6} + \sqrt{10})$ over $\mathbb{Q}(\sqrt{3} + \sqrt{5})$

Concepts

In Exercises 14 through 17, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

14. An *algebraic extension* of a field F is a field $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ where each α_i is a zero of some polynomial in $F[x]$.
15. A *finite extension field* of a field F is one that can be obtained by adjoining a finite number of elements to F .
16. The *algebraic closure* \bar{F}_E of a field F in an extension field E of F is the field consisting of all elements of E that are algebraic over F .
17. A field F is *algebraically closed* if and only if every polynomial has a zero in F .
18. Show by an example that for a proper extension field E of a field F , the algebraic closure of F in E need not be algebraically closed.
19. Determine whether each of the following is true or false.
 - a. If a field E is a finite extension of a field F , then E is a finite field.
 - b. Every finite extension of a field is an algebraic extension.
 - c. Every algebraic extension of a field is a finite extension.
 - d. The top field of a finite tower of finite extensions of fields is a finite extension of the bottom field.
 - e. \mathbb{Q} is its own algebraic closure in \mathbb{R} , that is, \mathbb{Q} is **algebraically closed** in \mathbb{R} .
 - f. \mathbb{C} is algebraically closed in $\mathbb{C}(x)$, where x is an indeterminate.
 - g. $\mathbb{C}(x)$ is algebraically closed, where x is an indeterminate.
 - h. The field $\mathbb{C}(x)$ has no algebraic closure, since \mathbb{C} already contains all algebraic numbers.
 - i. An algebraically closed field must be of characteristic 0.
 - j. If E is an algebraically closed extension field of F , then E is an algebraic extension of F .

Proof Synopsis

20. Give a one-sentence synopsis of the proof of Theorem 40.3.
 21. Give a one- or two-sentence synopsis of the proof of Theorem 40.4.

Theory

22. Let $(a + bi) \in \mathbb{C}$ where $a, b \in \mathbb{R}$ and $b \neq 0$. Show that $\mathbb{C} = \mathbb{R}(a + bi)$.
23. Show that if E is a finite extension of a field F and $[E : F]$ is a prime number, then E is a simple extension of F and, indeed, $E = F(\alpha)$ for every $\alpha \in E$ not in F .
24. Prove that $x^2 - 3$ is irreducible over $\mathbb{Q}(\sqrt[3]{2})$.
25. What degree field extensions can we obtain by successively adjoining to a field F a square root of an element of F not a square in F , then a square root of some nonsquare in this new field, and so on? Argue from this that a zero of $x^{14} - 3x^2 + 12$ over \mathbb{Q} can never be expressed as a rational function of square roots of rational functions of square roots, and so on, of elements of \mathbb{Q} .
26. Let E be a finite extension field of F . Let D be an integral domain such that $F \subseteq D \subseteq E$. Show that D is a field.
27. Prove in detail that $\mathbb{Q}(\sqrt{3} + \sqrt{7}) = \mathbb{Q}(\sqrt{3}, \sqrt{7})$.
28. Generalizing Exercise 27, show that if $\sqrt{a} + \sqrt{b} \neq 0$, then $\mathbb{Q}(\sqrt{a} + \sqrt{b}) = \mathbb{Q}(\sqrt{a}, \sqrt{b})$ for all a and b in \mathbb{Q} . [Hint: Compute $(a - b)/(\sqrt{a} + \sqrt{b})$.]
29. Let E be a finite extension of a field F , and let $p(x) \in F[x]$ be irreducible over F and have degree that is not a divisor of $[E : F]$. Show that $p(x)$ has no zeros in E .
30. Let E be an extension field of F . Let $\alpha \in E$ be algebraic of odd degree over F . Show that α^2 is algebraic of odd degree over F , and $F(\alpha) = F(\alpha^2)$.
31. Show that if F , E , and K are fields with $F \leq E \leq K$, then K is algebraic over F if and only if E is algebraic over F , and K is algebraic over E . (You must *not* assume the extensions are finite.)
32. Let E be an extension field of a field F . Prove that every $\alpha \in E$ that is not in the algebraic closure \bar{F}_E of F in E is transcendental over \bar{F}_E .
33. Let E be an algebraically closed extension field of a field F . Show that the algebraic closure \bar{F}_E of F in E is algebraically closed. (Applying this exercise to \mathbb{C} and \mathbb{Q} , we see that the field of all algebraic numbers is an algebraically closed field.)
34. Show that if E is an algebraic extension of a field F and contains all zeros in \bar{F} of every $f(x) \in F[x]$, then E is an algebraically closed field.
35. Show that no finite field of odd characteristic is algebraically closed. (Actually, no finite field of characteristic 2 is algebraically closed either.) [Hint: By counting, show that for such a finite field F , some polynomial $x^2 - a$, for some $a \in F$, has no zero in F . See Exercise 32, Section 39.]
36. Prove that, as asserted in the text, the algebraic closure of \mathbb{Q} in \mathbb{C} is not a finite extension of \mathbb{Q} .
37. Argue that every finite extension field of \mathbb{R} is either \mathbb{R} itself or is isomorphic to \mathbb{C} .
38. Use Zorn's lemma to show that every proper ideal of a ring R with unity is contained in some maximal ideal.

SECTION 41**[†]GEOMETRIC CONSTRUCTIONS**

In this section we digress briefly to give an application demonstrating the power of Theorem 40.4. For a more detailed study of geometric constructions, you are referred to Courant and Robbins [44, Chapter III].

In Euclid's *Elements*, geometry is approached from an axiomatic point of view. The first three axioms state that a line segment can be drawn between any two given points,

[†] This section is used only briefly in Section 48.

a line segment can be extended indefinitely to form a line, and a circle can be drawn centered at any given point with any given radius. A straightedge and a compass are objects that can accomplish the tasks set out in these axioms. A natural question asked by the ancient Greek geometers was exactly which points and configurations can be constructed using only a compass and a straightedge. They found many constructions that are well known today including bisecting angles, bisecting line segments, and trisecting line segments. Trisecting an arbitrary angle eluded mathematicians for well over 2000 years before field theory was developed and it was proven that there are some angles that are impossible to trisect using only a straightedge and compass. We shall discuss the impossibility of trisecting an angle and other classical questions.

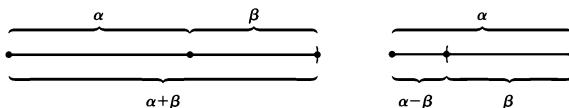
Constructible Numbers

Let us imagine that we are given only a single line segment that we shall define to be *one unit* in length. A real number α is **constructible** if we can construct a line segment of length $|\alpha|$ in a finite number of steps from this given segment of unit length by using a straightedge and a compass.

The rules of the game are pretty strict. We suppose that we are given just two points at the moment, the endpoints of our unit line segment, let us suppose that they correspond to the points $(0, 0)$ and $(1, 0)$ in the Euclidean plane. We are allowed to draw a line only with our straightedge through two points that we have already located. Thus we can start by using the straightedge and drawing the line through $(0, 0)$ and $(1, 0)$. We are allowed to open our compass only to a distance between points we have already found. Let us open our compass to the distance between $(0, 0)$ and $(1, 0)$. We can then place the point of the compass at $(1, 0)$ and draw a circle of radius 1, which passes through the point $(2, 0)$. Thus we now have located a third point, $(2, 0)$. Continuing in this way, we can locate points $(3, 0), (4, 0), (-1, 0), (-2, 0)$, and so on. Now open the compass the distance from $(0, 0)$ to $(0, 2)$, put the point at $(1, 0)$, and draw a circle of radius 2. Do the same with the point at $(-1, 0)$. We have now found two new points, where these circles intersect, and we can put our straightedge on them to draw what we think of as the *y-axis*. Then opening our compass to the distance from $(0, 0)$ to $(1, 0)$, we draw a circle with center at $(0, 0)$ and locate the point $(0, 1)$ where the circle intersects the *y-axis*. Continuing in this fashion, we can locate all points (x, y) with integer coordinates in any rectangle containing the point $(0, 0)$. Without going into more detail, it can be shown that it is possible, among other things, to erect a perpendicular to a given line at a known point on the line, and find a line passing through a known point and parallel to a given line. Our first result is the following theorem.

41.1 Theorem If α and β are constructible real numbers, then so are $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$, and α/β , if $\beta \neq 0$.

Proof We are given that α and β are constructible, so there are line segments of lengths $|\alpha|$ and $|\beta|$ available to us. For $\alpha, \beta > 0$, extend a line segment of length α with the straightedge. Start at one end of the original segment of length α , and lay off on the extension the length β with the compass. This constructs a line segment of length $\alpha + \beta$; $\alpha - \beta$ is similarly constructible (see Fig. 41.2). If α and β are not both positive, an obvious



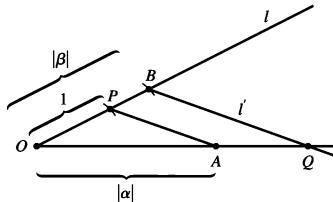
41.2 Figure

breakdown into cases according to their signs shows that $\alpha + \beta$ and $\alpha - \beta$ are still constructible.

The construction of $\alpha\beta$ is indicated in Fig. 41.3. We shall let \overline{OA} be the line segment from the point O to the point A , and shall let $|\overline{OA}|$ be the length of this line segment. If \overline{OA} is of length $|\alpha|$, construct a line l through O not containing \overline{OA} . (Perhaps, if O is at $(0, 0)$ and A is at $(a, 0)$, you use the line through $(0, 0)$ and $(4, 2)$.) Then find the points P and B on l such that \overline{OP} is of length 1 and \overline{OB} is of length $|\beta|$. Draw \overline{PA} and construct l' through B , parallel to \overline{PA} and intersecting \overline{OA} extended at Q . By similar triangles, we have

$$\frac{1}{|\alpha|} = \frac{|\beta|}{|\overline{OQ}|},$$

so \overline{OQ} is of length $|\alpha\beta|$.

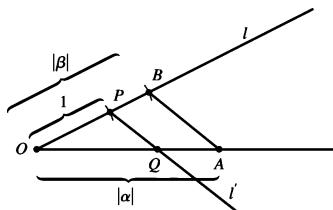


41.3 Figure

Finally, Fig. 41.4 shows that α/β is constructible if $\beta \neq 0$. Let \overline{OA} be of length $|\alpha|$, and construct l through O not containing OA . Then find B and P on l such that \overline{OB} is of length $|\beta|$ and \overline{OP} is of length 1. Draw \overline{BA} and construct l' through P , parallel to \overline{BA} , and intersecting \overline{OA} at Q . Again by similar triangles, we have

$$\frac{|\overline{OQ}|}{1} = \frac{|\alpha|}{|\beta|},$$

so \overline{OQ} is of length $|\alpha/\beta|$. ◆



41.4 Figure

41.5 Corollary The set of all constructible real numbers forms a subfield F of the field of real numbers.

Proof Proof of this corollary is immediate from Theorem 41.1. ◆

Thus the field F of all constructible real numbers contains \mathbb{Q} , the field of rational numbers, since \mathbb{Q} is the smallest subfield of \mathbb{R} .

From now on, we proceed analytically. We can construct any rational number. Regarding our given segment

$$0 \rule{1cm}{0.4pt} 1$$

of length 1 as the basic unit on an x -axis, we can locate any point (q_1, q_2) in the plane with both coordinates rational. Any further point in the plane that we can locate by using a compass and a straightedge can be found in one of the following three ways:

1. as an intersection of two lines, each of which passes through two known points having rational coordinates,
2. as an intersection of a line that passes through two points having rational coordinates and a circle whose center has rational coordinates and whose radius is rational.
3. as an intersection of two circles whose centers have rational coordinates and whose radii are rational.

Equations of lines and circles of the type discussed in 1, 2, and 3 are of the form

$$ax + by + c = 0$$

and

$$x^2 + y^2 + dx + ey + f = 0,$$

where a, b, c, d, e , and f are all in \mathbb{Q} . Since in Case 3 the intersection of two circles with equations

$$x^2 + y^2 + d_1x + e_1y + f_1 = 0$$

and

$$x^2 + y^2 + d_2x + e_2y + f_2 = 0$$

is the same as the intersection of the first circle having equation

$$x^2 + y^2 + d_1x + e_1y + f_1 = 0,$$

and the line (the common chord) having equation

$$(d_1 - d_2)x + (e_1 - e_2)y + f_1 - f_2 = 0,$$

we see that Case 3 can be reduced to Case 2. For Case 1, a simultaneous solution of two linear equations with rational coefficients can only lead to rational values of x and y , giving us no new points. However, finding a simultaneous solution of a linear equation with rational coefficients and a quadratic equation with rational coefficients, as in Case 2, leads, upon substitution, to a quadratic equation. Such an equation, when solved by the quadratic formula, may have solutions involving square roots of numbers that are not squares in \mathbb{Q} .

In the preceding argument, nothing was really used involving \mathbb{Q} except field axioms. If H is the smallest field containing those real numbers constructed so far, the argument shows that the “next new number” constructed lies in a field $H(\sqrt{\alpha})$ for some $\alpha \in H$, where $\alpha > 0$. We have proved half of our next theorem.

41.6 Theorem The field F of constructible real numbers consists precisely of all real numbers that we can obtain from \mathbb{Q} by taking square roots of positive numbers a finite number of times and applying a finite number of field operations.

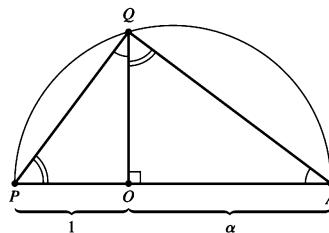
Proof We have shown that F can contain no numbers except those we obtain from \mathbb{Q} by taking a finite number of square roots of positive numbers and applying a finite number

of field operations. However, if $\alpha > 0$ is constructible, then Fig. 41.7 shows that $\sqrt{\alpha}$ is constructible. Let \overline{OA} have length α , and find P on \overline{OA} extended so that \overline{OP} has length 1. Find the midpoint of \overline{PA} and draw a semicircle with \overline{PA} as diameter. Erect a perpendicular to \overline{PA} at O , intersecting the semicircle at Q . Then the triangles OPQ and OQA are similar, so

$$\frac{|\overline{OQ}|}{|\overline{OA}|} = \frac{|\overline{OP}|}{|\overline{OQ}|},$$

and $|\overline{OQ}|^2 = 1\alpha = \alpha$. Thus \overline{OQ} is of length $\sqrt{\alpha}$. Therefore square roots of constructible numbers are constructible.

Theorem 41.1 showed that field operations are possible by construction. \diamond



41.7 Figure

41.8 Corollary If γ is constructible and $\gamma \notin \mathbb{Q}$, then there is a finite sequence of real numbers $\alpha_1, \dots, \alpha_n = \gamma$ such that $\mathbb{Q}(\alpha_1, \dots, \alpha_i)$ is an extension of $\mathbb{Q}(\alpha_1, \dots, \alpha_{i-1})$ of degree 2. In particular, $[\mathbb{Q}(\gamma) : \mathbb{Q}] = 2^r$ for some integer $r \geq 0$.

Proof The existence of the α_i is immediate from Theorem 41.6. Then

$$\begin{aligned} 2^n &= [\mathbb{Q}(\alpha_1, \dots, \alpha_n) : \mathbb{Q}] \\ &= [\mathbb{Q}(\alpha_1, \dots, \alpha_n) : \mathbb{Q}(\gamma)][\mathbb{Q}(\gamma) : \mathbb{Q}], \end{aligned}$$

by Theorem 40.4, which completes the proof. \diamond

The Impossibility of Certain Constructions

We can now show the impossibility of certain geometric constructions.

41.9 Theorem *Doubling the cube is impossible*, that is, given a side of a cube, it is not always possible to construct with a straightedge and a compass the side of a cube that has double the volume of the original cube.

Proof Let the given cube have a side of length 1, and hence a volume of 1. The cube being sought would have to have a volume of 2, and hence a side of length $\sqrt[3]{2}$. But $\sqrt[3]{2}$ is a zero of irreducible $x^3 - 2$ over \mathbb{Q} , so

$$[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3.$$

Corollary 41.8 shows that to double this cube of volume 1, we would need to have $3 = 2^r$ for some integer r , but no such r exists. \diamond

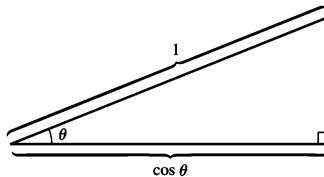
41.10 Theorem *Squaring the circle is impossible;* that is, given a circle, it is not always possible to construct with a straightedge and a compass a square having area equal to the area of the given circle.

Proof Let the given circle have a radius of 1, and hence an area of π . We would need to construct a square of side $\sqrt{\pi}$. But π is transcendental over \mathbb{Q} , so $\sqrt{\pi}$ is transcendental over \mathbb{Q} also. ◆

41.11 Theorem *Trisecting the angle is impossible;* that is, there exists an angle that cannot be trisected with a straightedge and a compass.

Proof Figure 41.12 indicates that the angle θ can be constructed if and only if a segment of length $|\cos \theta|$ can be constructed. Now 60° is a constructible angle, and we shall show that it cannot be trisected. Note that

$$\begin{aligned}\cos 3\theta &= \cos(2\theta + \theta) \\ &= \cos 2\theta \cos \theta - \sin 2\theta \sin \theta \\ &= (2\cos^2 \theta - 1)\cos \theta - 2\sin \theta \cos \theta \sin \theta \\ &= (2\cos^2 \theta - 1)\cos \theta - 2\cos \theta(1 - \cos^2 \theta) \\ &= 4\cos^3 \theta - 3\cos \theta.\end{aligned}$$



41.12 Figure

[We realize that some students have not seen the trigonometric identities we just used. Exercise 1 repeats Exercise 42 of Section 3 and asks you to prove the identity $\cos 3\theta = 4\cos^3 \theta - 3\cos \theta$ from Euler's formula.]

Let $\theta = 20^\circ$, so that $\cos 3\theta = \frac{1}{2}$, and let $\alpha = \cos 20^\circ$. From the identity $4\cos^3 \theta - 3\cos \theta = \cos 3\theta$, we see that

$$4\alpha^3 - 3\alpha = \frac{1}{2}.$$

Thus α is a zero of $8x^3 - 6x - 1$. This polynomial is irreducible in $\mathbb{Q}[x]$, since, by Theorem 28.12, it is enough to show that it does not factor in $\mathbb{Z}[x]$. But a factorization in $\mathbb{Z}[x]$ would entail a linear factor of the form $(8x \pm 1)$, $(4x \pm 1)$, $(2x \pm 1)$, or $(x \pm 1)$. We can quickly check that none of the numbers $\pm\frac{1}{8}, \pm\frac{1}{4}, \pm\frac{1}{2}$, and ± 1 is a zero of $8x^3 - 6x - 1$. Thus

$$[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3,$$

so by Corollary 41.8, α is not constructible. Hence 60° cannot be trisected. ◆

Note that the regular n -gon is constructible for $n \geq 3$ if and only if the angle $2\pi/n$ is constructible, which is the case if and only if a line segment of length $\cos(2\pi/n)$ is constructible.

■ HISTORICAL NOTE

Greek mathematicians as far back as the fourth century B.C. had tried without success to find geometric constructions using straightedge and compass to trisect the angle, double the cube, and square the circle. Although they were never able to prove that such constructions were impossible, they did manage to construct the solutions to these problems using other tools, including the conic sections.

It was Carl Gauss in the early nineteenth century who made a detailed study of constructibility in connection with his solution of cyclotomic equations, the equations of the form $x^p - 1 = 0$ with p prime whose roots form the vertices of a regular p -gon. He showed that although all such

equations are solvable using radicals, if $p - 1$ is not a power of 2, then the solutions must involve roots higher than the second. In fact, Gauss asserted that anyone who attempted to find a geometric construction for a p -gon where $p - 1$ is not a power of 2 would “spend his time uselessly.” Interestingly, Gauss did not prove the assertion that such constructions were impossible. That was accomplished in 1837 by Pierre Wantzel (1814–1848), who in fact proved Corollary 41.8 and also demonstrated Theorems 41.9 and 41.11. The proof of Theorem 41.10, on the other hand, requires a proof that π is transcendental, a result finally achieved in 1882 by Ferdinand Lindemann (1852–1939).

■ EXERCISES 41

Computations

- Prove the trigonometric identity $\cos 3\theta = 4\cos^3 \theta - 3\cos \theta$ from the Euler formula, $e^{i\theta} = \cos \theta + i\sin \theta$.

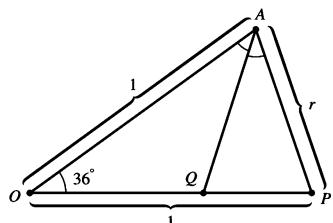
Concepts

- Determine whether each of the following is true or false.
 - It is impossible to double any cube of constructible edge by compass and straightedge constructions.
 - It is impossible to double every cube of constructible edge by compass and straightedge constructions.
 - It is impossible to square any circle of constructible radius by straightedge and compass constructions.
 - No constructible angle can be trisected by straightedge and compass constructions.
 - Every constructible number is of degree 2^r over \mathbb{Q} for some integer $r \geq 0$.
 - We have shown that every real number of degree 2^r over \mathbb{Q} for some integer $r \geq 0$ is constructible.
 - The fact that factorization of a positive integer into a product of primes is unique (up to order) was used strongly at the conclusion of Theorems 41.9 and 41.11.
 - Counting arguments are exceedingly powerful mathematical tools.
 - We can find any constructible number in a finite number of steps by starting with a given segment of unit length and using a straightedge and a compass.
 - We can find the totality of all constructible numbers in a finite number of steps by starting with a given segment of unit length and using a straightedge and a compass.

Theory

- Using the proof of Theorem 41.11, show that the regular 9-gon is not constructible.
- Show *algebraically* that it is possible to construct an angle of 30° .

5. Referring to Fig. 41.13, where \overline{AQ} bisects angle OAP , show that the regular 10-gon is constructible (and therefore that the regular pentagon is also). [Hint: Triangle OAP is similar to triangle APQ . Show algebraically that r is constructible.]



41.13 Figure

In Exercises 6 through 9 use the results of Exercise 5 where needed to show that the statement is true.

6. The regular 20-gon is constructible.
7. The regular 30-gon is constructible.
8. The angle 72° can be trisected.
9. The regular 15-gon can be constructed.
10. Suppose you wanted to explain roughly in just three or four sentences, for a high school plane geometry teacher who never had a course in abstract algebra, how it can be shown that it is impossible to trisect an angle of 60° . Write down what you would say.
11. Let $S = \{n \in \mathbb{Z} \mid \text{an } n \text{ degree angle is constructible with a compass and straightedge}\}$. We are assuming that if an angle of n degrees is constructed, then $n + k(360)$ is also constructed for any integer k . Prove that S is the principal ideal $(3) \subseteq \mathbb{Z}$. It may be helpful to use Exercise 5.
12. The proof of Theorem 41.11 can be simplified by making a different choice of the angle θ . Find a constructible real number α so that the angle $\theta = \frac{\arccos(\alpha)}{3}$ can be used in the formula $\cos 3\theta = 4\cos^3 \theta - 3\cos \theta$ to arrive at a polynomial that meets Eisenstein's criteria. Then finish the proof of Theorem 4.11.
13. Prove that for at least one constructible angle 5θ , θ is not constructible.
14. Continuing Exercise 13,
 - a. Use Euler's formula to show that for any integer $n \geq 2$, $\cos(n\theta) = 2\cos(\theta)\cos((n-1)\theta) - \cos((n-2)\theta)$.
 - b. Use part a to rewrite $\cos(7\theta)$ and use this formula to prove that there is a constructible angle 7θ such that θ is not constructible.
 - c. For each integer $1 \leq n \leq 10$, is there a constructible angle $n\theta$ such that θ is not constructible?

(The polynomials used in this exercise are called Chebyshev polynomials.)

SECTION 42

FINITE FIELDS

The purpose of this section is to determine the structure of all finite fields. We shall show that for every prime p and positive integer n , there is exactly one finite field (up to isomorphism) of order p^n . This field $\text{GF}(p^n)$ is usually referred to as the **Galois field of order p^n** . We shall be using quite a bit of our material on cyclic groups. The proofs are simple and elegant.

The Structure of a Finite Field

We now show that all finite fields must have prime-power order.

42.1 Theorem Let E be a finite extension of degree n over a finite field F . If F has q elements, then E has q^n elements.

Proof Let $\{\alpha_1, \dots, \alpha_n\}$ be a basis for E as a vector space over F . By Exercise 21 of Section 33, every $\beta \in E$ can be *uniquely* written in the form

$$\beta = b_1\alpha_1 + \dots + b_n\alpha_n$$

for $b_i \in F$. Since each b_i may be any of the q elements of F , the total number of such distinct linear combinations of the α_i is q^n . \diamond

42.2 Corollary If E is a finite field of characteristic p , then E contains exactly p^n elements for some positive integer n .

Proof Every finite field E is a finite extension of a prime field isomorphic to the field \mathbb{Z}_p , where p is the characteristic of E . The corollary follows at once from Theorem 42.1. \diamond

We now turn to the study of the multiplicative structure of a finite field. The following theorem will show us how any finite field can be formed from the prime subfield.

42.3 Theorem Let E be a field of p^n elements contained in an algebraic closure $\overline{\mathbb{Z}}_p$ of \mathbb{Z}_p . The elements of E are precisely the zeros in $\overline{\mathbb{Z}}_p$ of the polynomial $x^{p^n} - x$ in $\mathbb{Z}_p[x]$.

Proof The set E^* of nonzero elements of E forms a multiplicative group of order $p^n - 1$ under the field multiplication. For $\alpha \in E^*$, the order of α in this group divides the order $p^n - 1$ of the group. Thus for $\alpha \in E^*$, we have $\alpha^{p^n-1} = 1$, so $\alpha^{p^n} = \alpha$. Therefore, every element in E is a zero of $x^{p^n} - x$. Since $x^{p^n} - x$ can have at most p^n zeros, we see that E contains precisely the zeros of $x^{p^n} - x$ in $\overline{\mathbb{Z}}_p$. \diamond

42.4 Definition An element α of a field is an **n th root of unity** if $\alpha^n = 1$. It is a **primitive n th root of unity** if $\alpha^n = 1$ and $\alpha^m \neq 1$ for $0 < m < n$. \blacksquare

Thus the nonzero elements of a finite field of p^n elements are all $(p^n - 1)$ th roots of unity.

Recall that in Corollary 28.7, we showed that the multiplicative group of nonzero elements of a finite field is cyclic. This is a very important fact about finite fields; it has actually been applied to coding theory and combinatorics. For the sake of completeness in this section, we now state it here as a theorem, give a corollary, and illustrate with an example.

42.5 Theorem The multiplicative group $\langle E^*, \cdot \rangle$ of nonzero elements of a finite field E is cyclic.

Proof See Corollary 28.7. \diamond

42.6 Corollary A finite extension E of a finite field F is a simple extension of F .

Proof Let α be a generator for the cyclic group E^* of nonzero elements of E . Then $E = F(\alpha)$. \diamond

42.7 Example Consider the finite field \mathbb{Z}_{11} . By Theorem 42.5 $\langle \mathbb{Z}_{11}^*, \cdot \rangle$ is cyclic. Let us try to find a generator of \mathbb{Z}_{11}^* by brute force and ignorance. We start by trying 2. Since $|\mathbb{Z}_{11}^*| = 10$, 2 must be an element of \mathbb{Z}_{11}^* of order dividing 10, that is, either 2, 5, or 10. Now

$$2^2 = 4, \quad 2^4 = 4^2 = 5, \quad \text{and} \quad 2^5 = (2)(5) = 10 = -1.$$

Thus neither 2^2 nor 2^5 is 1, but, of course, $2^{10} = 1$, so 2 is a generator of \mathbb{Z}_{11}^* , that is, 2 is a primitive 10th root of unity in \mathbb{Z}_{11} . We were lucky.

By the theory of cyclic groups, all the generators of \mathbb{Z}_{11}^* , that is, all the primitive 10th roots of unity in \mathbb{Z}_{11} , are of the form 2^n , where n is relatively prime to 10. These elements are

$$2^1 = 2, \quad 2^3 = 8, \quad 2^7 = 7, \quad 2^9 = 6.$$

■ HISTORICAL NOTE

Although Carl F. Gauss had shown that the set of residues modulo a prime p satisfied the field properties, it was Evariste Galois (1811–1832) who first dealt with what he called “incommensurable solutions” to the congruence $F(x) \equiv 0 \pmod{p}$, where $F(x)$ is an n th degree irreducible polynomial modulo p . He noted in a paper written in 1830 that one should consider the roots of this congruence as “a variety of imaginary symbols” that one can use in calculations just as one uses $\sqrt{-1}$. Galois then showed that if α is any solution of $F(x) \equiv 0 \pmod{p}$, the expression $a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1}$ takes on precisely p^n different values. Finally, he proved results equivalent to Theorems 42.3 and 42.5 of the text.

Galois’ life was brief and tragic. He showed brilliance in mathematics early on, publishing

several papers before he was 20 and essentially established the basic ideas of Galois theory. He was, however, active in French revolutionary politics following the July revolution of 1830. In May 1831, he was arrested for threatening the life of King Louis-Philippe. Though he was acquitted, he was rearrested for participating, heavily armed, in a republican demonstration on Bastille Day of that year. Two months after his release from prison the following March, he was killed in a duel, “the victim of an infamous coquette and her two dupes”; the previous night he had written a letter to a friend clarifying some of his work in the theory of equations and requesting that it be studied by other mathematicians. Not until 1846, however, were his major papers published; it is from that date that his work became influential.

The primitive 5th roots of unity in \mathbb{Z}_{11} are of the form 2^m , where the gcd of m and 10 is 2, that is,

$$2^2 = 4, \quad 2^4 = 5, \quad 2^6 = 9, \quad 2^8 = 3.$$

The primitive square root of unity in \mathbb{Z}_{11} is $2^5 = 10 = -1$. ▲

The Existence of $\text{GF}(p^n)$

We turn now to the question of the existence of a finite field of order p^r for every prime power p^r , $r > 0$. We need the following lemma.

42.8 Lemma If F is a field of prime characteristic p with algebraic closure \bar{F} , then $x^{p^n} - x$ has p^n distinct zeros in \bar{F} .

Proof Because \bar{F} is algebraically closed, $x^{p^n} - x$ factors over that field into a product of linear factors $x - \alpha$, so it suffices to show that none of these factors occurs more than once in the factorization.

Exercise 15 uses derivatives to complete this proof. Although this is an elegant method, it requires some effort to develop derivatives for polynomials over an arbitrary field, so we proceed using long division. Observe that 0 is a zero of $x^{p^n} - x$ of multiplicity 1. Suppose $\alpha \neq 0$ is a zero of $x^{p^n} - x$, and hence is a zero of $f(x) = x^{p^n-1} - 1$. Then $x - \alpha$ is a factor of $f(x)$ in $\bar{F}[x]$, and by long division, we find that

$$\begin{aligned} \frac{f(x)}{(x - \alpha)} &= g(x) \\ &= x^{p^n-2} + \alpha x^{p^n-3} + \alpha^2 x^{p^n-4} + \cdots + \alpha^{p^n-3} x + \alpha^{p^n-2}. \end{aligned}$$

Now $g(x)$ has $p^n - 1$ summands, and in $g(\alpha)$, each summand is

$$\alpha^{p^n-2} = \frac{\alpha^{p^n-1}}{\alpha} = \frac{1}{\alpha}.$$

Thus

$$g(\alpha) = [(p^n - 1) \cdot 1] \frac{1}{\alpha} = -\frac{1}{\alpha}.$$

since we are in a field of characteristic p . Therefore, $g(\alpha) \neq 0$, so α is a zero of $f(x)$ of multiplicity 1. \blacklozenge

42.9 Lemma If F is a field of prime characteristic p , then $(\alpha + \beta)^{p^n} = \alpha^{p^n} + \beta^{p^n}$ for all $\alpha, \beta \in F$ and all positive integers n .

Proof Let $\alpha, \beta \in F$. Applying the binomial theorem to $(\alpha + \beta)^p$, we have

$$\begin{aligned} (\alpha + \beta)^p &= \alpha^p + (p \cdot 1)\alpha^{p-1}\beta + \left(\frac{p(p-1)}{2} \cdot 1\right)\alpha^{p-2}\beta^2 \\ &\quad + \cdots + (p-1)\alpha\beta^{p-1} + \beta^p \\ &= \alpha^p + 0\alpha^{p-1}\beta + 0\alpha^{p-2}\beta^2 + \cdots + 0\alpha\beta^{p-1} + \beta^p \\ &= \alpha^p + \beta^p. \end{aligned}$$

Proceeding by induction on n , suppose that we have $(\alpha + \beta)^{p^{n-1}} = \alpha^{p^{n-1}} + \beta^{p^{n-1}}$. Then $(\alpha + \beta)^{p^n} = [(\alpha + \beta)^{p^{n-1}}]^p = (\alpha^{p^{n-1}} + \beta^{p^{n-1}})^p = \alpha^{p^n} + \beta^{p^n}$. \blacklozenge

42.10 Theorem A finite field $\text{GF}(p^n)$ of p^n elements exists for every prime power p^n .

Proof Let $\overline{\mathbb{Z}}_p$ be an algebraic closure of \mathbb{Z}_p , and let K be the subset of $\overline{\mathbb{Z}}_p$ consisting of all zeros of $x^{p^n} - x$ in $\overline{\mathbb{Z}}_p$. Let $\alpha, \beta \in K$. Lemma 42.9 shows that $(\alpha + \beta) \in K$, and the equation $(\alpha\beta)^{p^n} = \alpha^{p^n}\beta^{p^n} = \alpha\beta$ shows that $\alpha\beta \in K$. From $\alpha^{p^n} = \alpha$ we obtain $(-\alpha)^{p^n} = (-1)^{p^n}\alpha^{p^n} = (-1)^{p^n}\alpha$. If p is an odd prime, then $(-1)^{p^n} = -1$ and if $p = 2$ then $-1 = 1$. Thus $(-\alpha)^{p^n} = -\alpha$, so $-\alpha \in K$. Now 0 and 1 are zeros of $x^{p^n} - x$. For $\alpha \neq 0, \alpha^{p^n} = \alpha$ implies that $(1/\alpha)^{p^n} = 1/\alpha$. Thus K is a subfield of $\overline{\mathbb{Z}}_p$ containing \mathbb{Z}_p . Therefore, K is the desired field of p^n elements, since Lemma 42.8 showed that $x^{p^n} - x$ has p^n distinct zeros in $\overline{\mathbb{Z}}_p$. \blacklozenge

42.11 Corollary If F is any finite field, then for every positive integer n , there is an irreducible polynomial in $F[x]$ of degree n .

Proof Let F have $q = p^r$ elements, where p is the characteristic of F . By Theorem 42.10, there is a field $K \leq \bar{F}$ containing \mathbb{Z}_p (up to isomorphism) and consisting precisely of the zeros of $x^{p^n} - x$. We want to show $F \leq K$. Every element of F is a zero of $x^{p^n} - x$, by Theorem 42.3. Now $p^{rn} = p^r p^{r(n-1)}$. Applying this equation repeatedly to the exponents and using the fact that for $\alpha \in F$ we have $\alpha^{p^r} = \alpha$, we see that for $\alpha \in F$,

$$\alpha^{p^m} = \alpha^{p^{r(n-1)}} = \alpha^{p^{r(n-2)}} = \cdots = \alpha^{p^r} = \alpha.$$

Thus $F \leq K$. Then Theorem 42.1 shows that we must have $[K : F] = n$. We have seen that K is simple over F in Corollary 42.6, so $K = F(\beta)$ for some $\beta \in K$. Therefore, $\text{irr}(\beta, F)$ must be of degree n . \blacklozenge

42.12 Theorem Let p be a prime and let $n \in \mathbb{Z}^+$. If E and E' are fields of order p^n , then $E \cong E'$.

Proof Both E and E' have \mathbb{Z}_p as prime field, up to isomorphism. By Corollary 42.6, E is a simple extension of \mathbb{Z}_p of degree n , so there exists an irreducible polynomial $f(x)$ of degree n in $\mathbb{Z}_p[x]$ such that $E \cong \mathbb{Z}_p[x]/(f(x))$. Because the elements of E are zeros of $x^{p^n} - x$, we see that $f(x)$ is a factor of $x^{p^n} - x$ in $\mathbb{Z}_p[x]$. Because E' also consists of

zeros of $x^{p^n} - x$, we see that E' also contains zeros of irreducible $f(x)$ in $\mathbb{Z}_p[x]$. Thus, because E' also contains exactly p^n elements, E' is also isomorphic to $\mathbb{Z}_p[x]/\langle f(x) \rangle$. ◆

In section 29 we saw that the field \mathbb{Z}_2 can be used to construct polynomial codes. Other finite fields have been used to construct algebraic codes with interesting properties. For example, in the *American Mathematical Monthly* 77 (1970): 249–258, Normal Levinson constructed an algebraic code that corrects three transmission errors. In this construction, the field of order 16 was used. Finite fields are also used in many other areas of mathematics including combinatorial designs, finite geometries, and algebraic topology.

■ EXERCISES 42

Computations

In Exercises 1 through 3, determine whether there exists a finite field having the given number of elements. (A calculator may be useful.)

1. 4096

2. 3127

3. 68,921

- 4.** Find the number of primitive 8th roots of unity in GF(9).
- 5.** Find the number of primitive 18th roots of unity in GF(19).
- 6.** Find the number of primitive 15th roots of unity in GF(31).
- 7.** Find the number of primitive 10th roots of unity in GF(23).

Concepts

- 8.** Determine whether each of the following is true or false.
 - a.** The nonzero elements of every finite field form a cyclic group under multiplication.
 - b.** The elements of every finite field form a cyclic group under addition.
 - c.** The zeros in \mathbb{C} of $(x^{28} - 1) \in \mathbb{Q}[x]$ form a cyclic group under multiplication.
 - d.** There exists a finite field of 60 elements.
 - e.** There exists a finite field of 125 elements.
 - f.** There exists a finite field of 36 elements.
 - g.** The complex number i is a primitive 4th root of unity.
 - h.** There exists an irreducible polynomial of degree 58 in $\mathbb{Z}_2[x]$.
 - i.** The nonzero elements of \mathbb{Q} form a cyclic group \mathbb{Q}^* under field multiplication.
 - j.** If F is a finite field, then every isomorphism mapping F onto a subfield of an algebraic closure \bar{F} of F is an automorphism of F .

Theory

- 9.** Let $\bar{\mathbb{Z}}_2$ be an algebraic closure of \mathbb{Z}_2 , and let $\alpha, \beta \in \bar{\mathbb{Z}}_2$ be zeros of $x^3 + x^2 + 1$ and of $x^3 + x + 1$, respectively. Using the results of this section, show that $\mathbb{Z}_2(\alpha) = \mathbb{Z}_2(\beta)$.
- 10.** Show that every irreducible polynomial in $\mathbb{Z}_p[x]$ is a divisor of $x^{p^n} - x$ for some n .
- 11.** Let F be a finite field of p^n elements containing the prime subfield \mathbb{Z}_p . Show that if $\alpha \in F$ is a generator of the cyclic group $\langle F^*, \cdot \rangle$ of nonzero elements of F , then $\deg(\alpha, \mathbb{Z}_p) = n$.
- 12.** Show that a finite field of p^n elements has exactly one subfield of p^m elements for each divisor m of n .
- 13.** Show that $x^{p^n} - x$ is the product of all monic irreducible polynomials in $\mathbb{Z}_p[x]$ of a degree d dividing n .
- 14.** Let p be an odd prime.

- a. Show that for $a \in \mathbb{Z}$, where $a \not\equiv 0 \pmod{p}$, the congruence $x^2 \equiv a \pmod{p}$ has a solution in \mathbb{Z} if and only if $a^{(p-1)/2} \equiv 1 \pmod{p}$. [Hint: Formulate an equivalent statement in the finite field \mathbb{Z}_p , and use the theory of cyclic groups.]
- b. Using part (a), determine whether or not the polynomial $x^2 - 6$ is irreducible in $\mathbb{Z}_{17}[x]$.
15. Let F be an arbitrary field. We define the **derivative** of $p(x) = \sum_{k=0}^n a_k x^k \in F[x]$ to be $D(p(x)) = \sum_{k=1}^n (k \cdot (a_k x^{k-1}))$. Let $p(x), q(x) \in F[x]$, and let n and m be nonnegative integers. Prove the following statements.
- $D(p(x) + q(x)) = D(p(x)) + D(q(x))$.
 - $D(ap(x)) = aD(p(x))$ for any $a \in F$.
 - $D(x^n x^m) = x^n D(x^m) + D(x^n)x^m$.
 - $D(p(x)x^m) = p(x)D(x^m) + D(p(x))x^m$.
 - $D(p(x)q(x)) = p(x)D(q(x)) + D(p(x))q(x)$.
 - $(x - a)^2$ divides $p(x)$ if and only if a is a zero of both $p(x)$ and $D(p(x))$.
 - Give a proof of Lemma 42.8 using part f.

Galois Theory

-
- Section 43** Introduction to Galois Theory
 - Section 44** Splitting Fields
 - Section 45** Separable Extensions
 - Section 46** Galois Theory
 - Section 47** Illustrations of Galois Theory
 - Section 48** Cyclotomic Extensions
 - Section 49** Insolvability of the Quintic

SECTION 43

INTRODUCTION TO GALOIS THEORY

An Example

We learned in high school that the quadratic formula provides zeros for polynomials of degree two. There are similar formulas for solutions to polynomials of degree three and four. These solutions all involve addition, subtraction, multiplication, division, and taking radicals. **Galois theory**, which provides an interesting connection between group theory and field theory, can be used to show that it is futile to seek a similar formula for polynomials of degree five or greater. Our main goal for the remainder of the book is to provide a proof of this fact.

43.1 Definition Let E be a field. An **automorphism** of E is a isomorphism of E onto itself. ■

43.2 Theorem The set of automorphisms of a field E is a group under function composition.

Proof The proof is Exercise 29. ◆

43.3 Example A field isomorphism $\phi : E \rightarrow K$ maps $1 \in E$ to $1 \in K$ and, it maps $0 \in E$ to $0 \in K$. Therefore, for any automorphism ϕ of \mathbb{Q} , $\phi(1) = 1$ and $\phi(0) = 0$. It follows by induction that $\phi(n) = n$ for any natural number n . Since $\phi(-x) = -\phi(x)$, for any integer n , $\phi(n) = n$. Every rational number is a ratio of integers, so $\phi(r) = r$ for every rational number r . Therefore the only automorphism of \mathbb{Q} is the identity map. ▲

43.4 Example Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. From Example 40.9, the degree of the extension K over \mathbb{Q} is 4 and a basis for the vector space K over the field \mathbb{Q} is $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$. Thus $K = \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \mid a, b, c, d \in \mathbb{Q}\}$. We determine the automorphisms of K .

Let ϕ be any automorphism of K . Since $\phi(1) = 1$, ϕ maps every rational number to itself as shown in the previous example. Since ϕ is a field automorphism, $\phi(\sqrt{2}^2 - 2) = \phi(0) = 0$. But $\phi(\sqrt{2}^2 - 2) = \phi(\sqrt{2})^2 - 2$. Thus $\phi(\sqrt{2})$ is a zero of the polynomial $x^2 - 2$, which means that $\phi(\sqrt{2})$ is either $\sqrt{2}$ or $-\sqrt{2}$. Similarly, $\phi(\sqrt{3}) = \pm\sqrt{3}$. Given any $a, b, c, d \in \mathbb{Q}$,

$$\alpha = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \in K,$$

then

$$\phi(\alpha) = a + b\phi(\sqrt{2}) + c\phi(\sqrt{3}) + d\phi(\sqrt{2})\phi(\sqrt{3}).$$

Once $\phi(\sqrt{2})$ and $\phi(\sqrt{3})$ are specified, there is at most one automorphism meeting these specifications. We have at most four automorphisms of K given by Table 43.5.

43.5 Table

	$\phi(\sqrt{2})$	$\phi(\sqrt{3})$	$\phi(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6})$
ι	$\sqrt{2}$	$\sqrt{3}$	$a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$
σ	$-\sqrt{2}$	$\sqrt{3}$	$a - b\sqrt{2} + c\sqrt{3} - d\sqrt{6}$
τ	$\sqrt{2}$	$-\sqrt{3}$	$a + b\sqrt{2} - c\sqrt{3} - d\sqrt{6}$
γ	$-\sqrt{2}$	$-\sqrt{3}$	$a - b\sqrt{2} - c\sqrt{3} + d\sqrt{6}$

It is tedious, but not difficult, to check that each of these maps is a field automorphism. By Theorem 43.2, $G = \{\iota, \sigma, \tau, \gamma\}$ forms a group under function composition. Every group with exactly four elements is isomorphic with the Klein 4-group or the cyclic group of order four. Each element of G has order one or two, so G is isomorphic with the Klein 4-group. ▲

43.6 Definition

If E and K are both field extensions of a field F and $\sigma : E \rightarrow K$ is a field isomorphism, then an element $\alpha \in E$ is **fixed by** σ if $\sigma(\alpha) = \alpha$. An element $\alpha \in E$ is **fixed by** a collection of isomorphisms if α is fixed by every isomorphism in the collection. A subset L of E is **fixed by** a collection of isomorphisms if every $\alpha \in L$ is fixed by the collection. Often we write **remains fixed** instead of simply fixed. ■

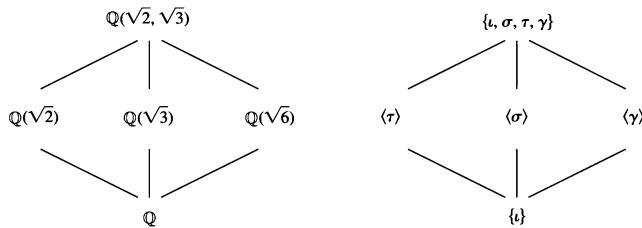
The discussion in Example 43.3 shows that for any field extensions E and K of \mathbb{Q} and isomorphism $\sigma : E \rightarrow K$, \mathbb{Q} remains fixed by σ .

43.7 Example

We find the elements fixed by each of the four automorphisms of G in Example 43.4. From Table 43.5 we see that each element fixes a subfield of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, and from this we can determine the field fixed by each subgroup of G .

- Each element K is fixed by ι . In other words, K remains fixed by the trivial subgroup $\{\iota\}$.
- Each element of $\{a + c\sqrt{3} \mid a, c \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{3})$ is fixed by σ . This implies that $\mathbb{Q}(\sqrt{3})$ remains fixed by the subgroup $\langle \sigma \rangle \leq G$.
- Each element of $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{2})$ is fixed by τ . So $\mathbb{Q}(\sqrt{2})$ remains fixed by the subgroup $\langle \tau \rangle$.
- Each element of $\{a + d\sqrt{6} \mid a, d \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{6})$ is fixed by γ . Therefore, $\mathbb{Q}(\sqrt{6})$ remains fixed by the subgroup $\langle \gamma \rangle$.
- The only elements that remain fixed by G are the elements of \mathbb{Q} .

Exercise 30 shows the five fields \mathbb{Q} , $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(\sqrt{3})$, $\mathbb{Q}(\sqrt{6})$, and $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, are the only subfields of K . Furthermore, from group theory we know that G , $\langle \sigma \rangle$, $\langle \tau \rangle$, $\langle \gamma \rangle$, and $\{\iota\}$ are the only subgroups of G . We have established a one-to-one correspondence between the subfields of K containing \mathbb{Q} , and the subgroups of the automorphism group of K that fix elements of \mathbb{Q} . Figure 43.8 shows the subfield diagram for K and the subgroup diagram for G . Notice that relabeling the fields by their corresponding subgroups gives the subgroup diagram, except that it is inverted. The reason that the diagrams are flipped is that if $H_1 \leq H_2$ are both subgroups of the automorphism group of K , then every element of K fixed by all the automorphisms in H_2 is also fixed by all the elements of H_1 . So the set that remains fixed by H_2 is a subset of the set that remains fixed by H_1 . ▲



43.8 Figure

The fact that the subfield diagram and the subgroup diagram correspond by associating a subfield of K with a subgroup of G is no accident. In fact, this is the heart of Galois theory. Before we can give precise statements of the Galois theorems, we need a few definitions and some background lemmas and theorems. It is advisable to have Examples 43.4 and 43.7 well in mind when reading the next few sections.

Subfields and Subgroups

We now investigate the Galois correspondence between subfields and subgroups of the automorphism group of a field. In Example 43.7, K was an extension field of \mathbb{Q} . In general we will investigate the automorphisms of a field that fix elements of a subfield that is not necessarily the rational numbers.

43.9 Theorem Let σ be an automorphism of the field E . Then the set E_σ of all the elements $a \in E$ that remain fixed by σ forms a subfield of E .

Proof Suppose that $a, b \in E$ remain fixed by σ , that is, $\sigma(a) = a$ and $\sigma(b) = b$. Since σ is a field automorphism, we have

$$\begin{aligned}\sigma(a \pm b) &= \sigma(a) \pm \sigma(b) = a \pm b, \\ \sigma(ab) &= \sigma(a)\sigma(b) = ab, \\ \sigma(a/b) &= \sigma(a)/\sigma(b) = a/b \quad \text{if } b \neq 0, \\ \sigma(0) &= 0, \text{ and} \\ \sigma(1) &= 1.\end{aligned}$$

Thus $a \pm b, ab, 0, 1 \in E_\sigma$ and if $b \neq 0$, $a/b \in E_\sigma$, which imply that E_σ is a subfield of E . \blacklozenge

43.10 Corollary Let $\{\sigma_i \mid i \in I\}$ be a collection of automorphisms of a field E . Then the set $E_{\{\sigma_i\}}$, of all $a \in E$ that remain fixed by every σ_i , for $i \in I$, is a subfield of E .

Proof The set $E_{\{\sigma_i\}} = \cap_{i \in I} E_{\{\sigma_i\}}$ is an intersection of subfields of E , so by Exercise 51 in Section 22, $E_{\{\sigma_i\}}$ is a subfield of E . \blacklozenge

We will continue to use the notation E_σ to denote the subfield of E that remains fixed by the automorphism σ and $E_{\{\sigma_i\}}$ to denote the subfield of E that remains fixed by σ_i for every $i \in I$.

43.11 Example Continuing Example 43.7, $\mathbb{Q}(\sqrt{2}, \sqrt{3})_{(\sigma)} = \mathbb{Q}(\sqrt{3})$ and $\mathbb{Q}(\sqrt{2}, \sqrt{3})_{(\gamma)} = \mathbb{Q}(\sqrt{6})$. \blacktriangle

In the above discussion we started with a set of automorphisms of a field K and saw that the elements fixed by the automorphisms form a subfield of K . This provides a way

to assign subfields of K to subgroups of the group of automorphisms. We now turn our attention to subfields $F \leq K$ and ask if there is a subgroup of the automorphism group of K that has exactly the set F as a fixed set.

43.12 Definition Let $F \leq K$ be a field extension. The set $G(K/F)$ is the set of all automorphisms of the field K that fix every element of the field F . ■

43.13 Example We see from Examples 43.4 and 43.7 that

$$\begin{aligned} G(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}) &= \{\iota, \sigma, \tau, \gamma\} \\ G(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}(\sqrt{3})) &= \{\iota, \sigma\} = \langle \sigma \rangle \\ G(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}(\sqrt{2})) &= \{\iota, \tau\} = \langle \tau \rangle \\ G(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}(\sqrt{6})) &= \{\iota, \gamma\} = \langle \gamma \rangle \\ G(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}(\sqrt{2}, \sqrt{3})) &= \{\iota\}. \end{aligned}$$
▲

If K is an extension field of F and $F \leq E \leq K$, then we will refer to E as an **intermediate field** of the extension. In Examples 43.11 and 43.13, we saw that for every intermediate field E of the extension $\mathbb{Q} \leq \mathbb{Q}(\sqrt{2}, \sqrt{3})$, there was a subgroup H of the automorphism group of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ that fixes every element of E and furthermore, $\mathbb{Q}(\sqrt{2}, \sqrt{3})_H = E$. This one-to-one correspondence between subgroups of the automorphism group and intermediate fields is the essence of Galois theory. There are field extensions where the correspondence fails, as shown by the next example. In order to have this one-to-one correspondence, a few technical conditions on the field extension need to be satisfied.

43.14 Example Let $K = \mathbb{Q}(\sqrt[3]{2}) = \{a + b\sqrt[3]{2} + c\sqrt[3]{4} \mid a, b \in \mathbb{Q}\}$. Then $G(K/\mathbb{Q})$ consists of all automorphisms of K that fix all the rational numbers. In $\mathbb{Q}(\sqrt[3]{2})$ there is only one zero of the polynomial $x^3 - 2$ since the other two zeros, $\sqrt[3]{2}(-1 \pm \sqrt{3}i)/2$, are complex numbers and $\mathbb{Q}(\sqrt[3]{2})$ is a subfield of the real numbers. For any automorphism $\sigma \in G(K/\mathbb{Q})$,

$$\begin{aligned} 0 &= \sigma(0) = \sigma(\sqrt[3]{2}^3 - 2) \\ &= (\sigma(\sqrt[3]{2}))^3 - 2. \end{aligned}$$

Thus $\sigma(\sqrt[3]{2})$ is a zero of $x^3 - 2$, which implies that $\sigma(\sqrt[3]{2}) = \sqrt[3]{2}$. Therefore any field automorphism of K fixes all elements of \mathbb{Q} and $\sqrt[3]{2}$, which implies that the only automorphism of K is the identity automorphism ι . Thus $G(K/\mathbb{Q}) = \{\iota\}$ and $K_{G(K/\mathbb{Q})} = K_{\{\iota\}} = K$. In this case, there is no subgroup $H \leq G(K/\mathbb{Q})$ with $K_H = \mathbb{Q}$. ▲

In the next two sections, we will investigate conditions on field extensions $F \leq K$ where there is a one-to-one correspondence between the intermediate fields and the subgroups of $G(K/F)$.

43.15 Theorem Let E be a field and let F be a subfield of E . Then the set $G(E/F)$ of all automorphisms that fix all the elements of F is a subgroup of the automorphism group of E . Furthermore, F is a subfield of $E_{G(E/F)}$.

Proof For $\sigma, \tau \in G(E/F)$ and $a \in F$,

$$(\sigma\tau)(a) = \sigma(\tau(a)) = \sigma(a) = a,$$

so $\sigma\tau \in G(E/F)$. Furthermore, $G(E/F)$ contains the identity map and $\sigma^{-1}(a) = a$. Thus $G(E/F)$ is a subgroup of the automorphisms of E .

Finally, since every automorphism in $G(E/F)$ fixes all the elements of F , F is a subset and, therefore, a subfield of $E_{G(E/F)}$. ◆

Since $G(E/F)$ is a subgroup of the automorphisms of E , we naturally call $G(E/F)$ the **group of automorphisms of E that fix F** . More briefly, we say that $G(E/F)$ is the **group of E over F** .

Beware! The symbol “ $/$ ” in $G(E/F)$ does not refer to a fraction or a quotient space. The symbol “ $/$ ” is used since we read $G(E/F)$ as the group of E over F and in the subfield diagram, E is written above F .

Since $G(E/F)$ is a subgroup of the automorphisms of E , we naturally call $G(E/F)$ the **group of automorphisms of E that fix F** . More briefly, we say that $G(E/F)$ is the **group of E over F** .

Beware! The symbol “ $/$ ” in $G(E/F)$ does not refer to a fraction or a quotient space. The symbol “ $/$ ” is used since we read $G(E/F)$ as the group of E over F and in the subfield diagram, E is written above F .

Conjugation Isomorphisms

In our ongoing example of $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$, we noticed that for any automorphism σ of K , $\sigma(\sqrt{2}) = \pm\sqrt{2}$ since the image of a zero of $x^2 - 2$ must also be a zero of $x^2 - 2$. This observation can be used for any polynomial and it is the basis for the Conjugation Theorem.

43.16 Definition Let E be an algebraic extension of the field F . Two elements α and β in E are **conjugates over F** , if both have the same minimal polynomial over F . That is, $\text{irr}(\alpha, F) = \text{irr}(\beta, F)$. ■

43.17 Example The definition of conjugates over F is consistent with our familiar use of the term complex conjugates in the setting of complex numbers. For $a, b \in \mathbb{R}$, the numbers $a + bi$ and $a - bi$ are complex conjugates. Both are zeros of the polynomial $f(x) = x^2 - 2ax + a^2 + b^2$ and, as long as $b \neq 0$, $f(x)$ is irreducible over \mathbb{R} . Thus for $b \neq 0$, $a + bi$ and $a - bi$ have the same minimal polynomial and they are conjugates over \mathbb{R} in the sense of Definition 43.16. ▲

43.18 Theorem (The Conjugation Isomorphism) Let F be a field, K an extension field of F , and $\alpha, \beta \in K$ algebraic over F with $\deg(\alpha, F) = n$. The map $\psi_{\alpha, \beta} : F(\alpha) \rightarrow F(\beta)$ defined by

$$\psi_{\alpha, \beta}(c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{n-1}\alpha^{n-1}) = c_0 + c_1\beta + c_2\beta^2 + \cdots + c_{n-1}\beta^{n-1},$$

for $c_i \in F$, is an isomorphism of $F(\alpha)$ onto $F(\beta)$ if and only if α and β are conjugate over F .

Proof We first assume that $\psi_{\alpha, \beta}$ is an isomorphism. Since $\deg(\alpha, F) = n$, $\text{irr}(\alpha, F) = x^n + g(x)$, for some polynomial $g(x) \in F[x]$, where the degree of $g(x)$ is less than n . Therefore

$$\alpha^n = -g(\alpha).$$

Since $\psi_{\alpha, \beta}$ is an isomorphism,

$$\psi_{\alpha, \beta}(\alpha^n) = (\psi_{\alpha, \beta}(\alpha))^n = \beta^n.$$

On the other hand,

$$\psi_{\alpha, \beta}(g(\alpha)) = g(\beta).$$

Thus

$$\beta^n = \psi_{\alpha, \beta}(\alpha^n) = \psi_{\alpha, \beta}(-g(\alpha)) = -g(\beta).$$

So β is a zero of the polynomial $x^n + g(x) = \text{irr}(\alpha, F)$, which is irreducible over F . By the definition of the minimal polynomial, $\text{irr}(\beta, F) = x^n + g(x) = \text{irr}(\alpha, F)$.

We next assume that α and β are conjugates over F with minimal polynomial $p(x) = \text{irr}(\alpha, F) = \text{irr}(\beta, F)$. By Corollary 39.14, the ideal $\langle p(x) \rangle \subseteq F[x]$ is the kernel of the evaluation homomorphism $\phi_\alpha : F[x] \rightarrow F(\alpha)$, which is onto. By the Fundamental Homomorphism Theorem 30.17, there is an isomorphism $\psi_\alpha : F[x]/\langle p(x) \rangle \rightarrow F(\alpha)$ with $\psi_\alpha(a + \langle p(x) \rangle) = a$ for all $a \in F$ and $\psi_\alpha(x + \langle p(x) \rangle) = \alpha$. Similarly, there is an isomorphism $\psi_\beta : F[x]/\langle p(x) \rangle \rightarrow F(\beta)$ with $\psi_\beta(a + \langle p(x) \rangle) = a$ for all $a \in F$, and $\psi_\beta(x + \langle p(x) \rangle) = \beta$. Since ψ_α and ψ_β are both isomorphisms, $\psi_\alpha^{-1} : F(\alpha) \rightarrow F[x]/\langle p(x) \rangle$ is an isomorphism and $\psi_{\alpha,\beta} = \psi_\beta \circ \psi_\alpha^{-1} : F(\alpha) \rightarrow F(\beta)$ is also an isomorphism. We have

$$\begin{aligned}\psi_{\alpha,\beta}(\alpha) &= \psi_\beta(\psi_\alpha^{-1}(\alpha)) \\ &= \psi_\beta(x + \langle p(x) \rangle) \\ &= \beta.\end{aligned}$$

We also note that for $c \in F$, $\psi_{\alpha,\beta}(c) = c$. Let $c_0, c_1, \dots, c_{n-1} \in F$ since $\psi_{\alpha,\beta}$ is an isomorphism,

$$\begin{aligned}\psi_{\alpha,\beta}(c_0 + c_1\alpha + \dots + c_{n-1}\alpha^{n-1}) &= \psi_{\alpha,\beta}(c_0) + \psi_{\alpha,\beta}(c_1)\psi_{\alpha,\beta}(\alpha) + \dots + \psi_{\alpha,\beta}(c_{n-1})\psi_{\alpha,\beta}(\alpha^{n-1}) \\ &= c_0 + c_1\psi_{\alpha,\beta}(\alpha) + \dots + c_{n-1}\psi_{\alpha,\beta}(\alpha)^{n-1} \\ &= c_0 + c_1\beta + \dots + c_{n-1}\beta^{n-1}.\end{aligned}\quad \diamond$$

43.19 Corollary Let K be a field extension of F with $\alpha \in K$ algebraic over F . Suppose that ψ is an isomorphism of $F(\alpha)$ onto a subfield of K , with the property that every element of F is fixed by ψ . Then ψ maps α to a conjugate over F of α . Conversely, if $\beta \in K$ is conjugate over F with α , then there is a unique isomorphism $\psi_{\alpha,\beta}$ mapping $F(\alpha)$ onto a subfield of K with the properties that each $a \in F$ is fixed by σ and $\sigma(\alpha) = \beta$.

Proof Let ψ be an isomorphism from $F(\alpha)$ onto a subfield of K with the property that every element of F is fixed by ψ . Let $\text{irr}(\alpha, F) = a_0 + a_1x + \dots + a_{n-1}x^{n-1}$. Then

$$a_0 + a_1\alpha + \dots + a_{n-1}\alpha^{n-1} = 0$$

and

$$0 = \psi(a_0 + a_1\alpha + \dots + a_{n-1}\alpha^{n-1}) = a_0 + a_1\psi(\alpha) + \dots + a_{n-1}\psi(\alpha)^{n-1}.$$

Thus $\beta = \psi(\alpha)$ has minimal polynomial $a_0 + a_1x + \dots + a_{n-1}x^{n-1}$ over F , which says that α and β are conjugates over F .

Now we let $\beta \in K$ be a conjugate of α over F . Theorem 43.18 provides an isomorphism $\psi_{\alpha,\beta} : F(\alpha) \rightarrow F(\beta)$ with the desired properties. Uniqueness follows since any isomorphism from $F(\alpha)$ to any field is completely determined by its values on elements of F and its value on α . \diamond

Theorem 43.9 and Corollary 43.19 formalize ideas used in Examples 43.7 and 43.14. Any automorphism σ of $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3})$ when restricted to the subfield $\mathbb{Q}(\sqrt[3]{2})$ is an isomorphism onto a subfield of $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3})$. Corollary 43.19 states that the automorphism maps $\sqrt[3]{2}$ to $\pm\sqrt[3]{2}$. Similarly $\sqrt[3]{3}$ maps to $\pm\sqrt[3]{3}$, making a total of at most four automorphisms of $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3})$.

Looking back at Example 43.14, any automorphism of $\mathbb{Q}(\sqrt[3]{2})$ maps $\sqrt[3]{2}$ to a conjugate over \mathbb{Q} and fixes elements of \mathbb{Q} . But $\sqrt[3]{2}$ has no conjugates in $\mathbb{Q}(\sqrt[3]{2})$ other than itself, so the only automorphism of $\mathbb{Q}(\sqrt[3]{2})$ is the identity map. Corollary 43.19 is essential for the rest of our study of Galois theory.

We now give a familiar corollary of Theorem 43.18 concerning complex number zeros of polynomials with real coefficients. Corollary 43.20 states that complex zeros of polynomials with real coefficients occur in conjugate pairs.

43.20 Corollary Let $f(x) \in \mathbb{R}[x]$. If $a, b \in \mathbb{R}$ and $f(a + bi) = 0$, then $f(a - bi) = 0$.

Proof As a field $\mathbb{C} = \mathbb{R}(i)$. Both i and $-i$ have minimal polynomial $x^2 + 1$ over \mathbb{R} , so they are conjugate over \mathbb{R} . Theorem 43.18 assures us that there is an automorphism $\psi_{i,-i} : \mathbb{R}(i) = \mathbb{C} \rightarrow \mathbb{C}$ given by $\psi_{i,-i}(a + ib) = a - bi$. Let

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

and assume that $a + bi$ is a zero of $f(x)$. Then

$$0 = f(a + bi) = a_0 + a_1(a + bi) + a_2(a + bi)^2 + \cdots + a_n(a + bi)^n,$$

so

$$0 = \psi_{i,-i}(f(a + bi)) = a_0 + a_1(a - bi) + a_2(a - bi)^2 + \cdots + a_n(a - bi)^n = f(a - bi).$$

Thus $f(a - bi) = 0$. \diamond

In order to pursue the Galois connection between subgroups of $G(K/F)$ and intermediate fields $F \leq E \leq K$, we need a technical condition to be sure that K has enough conjugate elements so Theorem 43.18 and its Corollary 43.19 can be evoked as often as needed. When the technical condition on E is satisfied, then E is called a **splitting field**. We investigate splitting fields in Section 44.

There is one other technical condition that we will need in order to establish the Galois correspondence. That condition says that for each $\alpha \in K$, $\text{irr}(\alpha, F)$ has $\deg(\alpha, F)$ distinct zeros in the algebraic closure of K . If the field extension E over F satisfies this condition, then the extension is said to be **separable**. As we will see, essentially all of our applications and examples, as well as most field extensions with which we are familiar, are separable. But we are getting ahead of ourselves, as properties of separable extensions are the focus of Section 44.

■ EXERCISES 43

Computations

In Exercises 1 through 8, find all conjugates in \mathbb{C} of the given number over the given field.

- | | |
|--|--|
| 1. $\sqrt{2}$ over \mathbb{Q} | 2. $\sqrt{2}$ over \mathbb{R} |
| 3. $3 + \sqrt{2}$ over \mathbb{Q} | 4. $\sqrt{2} - \sqrt{3}$ over \mathbb{Q} |
| 5. $\sqrt{2} + i$ over \mathbb{Q} | 6. $\sqrt{2} + i$ over \mathbb{R} |
| 7. $\sqrt{1 + \sqrt{2}}$ over \mathbb{Q} | 8. $\sqrt{1 + \sqrt{2}}$ over $\mathbb{Q}(\sqrt{2})$ |

In Exercises 9 through 15, we consider the field $E = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$. It can be shown that $[E : \mathbb{Q}] = 8$. In the notation of Theorem 43.18, we have the following conjugation isomorphisms (which are here automorphisms of E):

$$\begin{aligned}\psi_{\sqrt{2},-\sqrt{2}} &: (\mathbb{Q}(\sqrt{3}, \sqrt{5}))(\sqrt{2}) \rightarrow (\mathbb{Q}(\sqrt{3}, \sqrt{5}))(-\sqrt{2}), \\ \psi_{\sqrt{3},-\sqrt{3}} &: (\mathbb{Q}(\sqrt{2}, \sqrt{5}))(\sqrt{3}) \rightarrow (\mathbb{Q}(\sqrt{2}, \sqrt{5}))(-\sqrt{3}), \\ \psi_{\sqrt{5},-\sqrt{5}} &: (\mathbb{Q}(\sqrt{2}, \sqrt{3}))(\sqrt{5}) \rightarrow (\mathbb{Q}(\sqrt{2}, \sqrt{3}))(-\sqrt{5}).\end{aligned}$$

For shorter notation, let $\tau_2 = \psi_{\sqrt{2},-\sqrt{2}}$, $\tau_3 = \psi_{\sqrt{3},-\sqrt{3}}$, and $\tau_5 = \psi_{\sqrt{5},-\sqrt{5}}$. Compute the indicated element of E .

- | | |
|--|--|
| 9. $\tau_2(\sqrt{3})$ | 10. $\tau_2(\sqrt{2} + \sqrt{5})$ |
| 11. $(\tau_3\tau_2)(\sqrt{2} + 3\sqrt{5})$ | 12. $(\tau_5\tau_3)\left(\frac{\sqrt{2} - 3\sqrt{5}}{2\sqrt{3} - \sqrt{2}}\right)$ |
| 13. $(\tau_5^2\tau_3\tau_2)(\sqrt{2} + \sqrt{45})$ | 14. $\tau_3[\tau_5(\sqrt{2} - \sqrt{3} + (\tau_2\tau_5)(\sqrt{30}))]$ |

15. $\tau_3\tau_5\tau_3^{-1}(\sqrt{3} - \sqrt{5})$

In Exercises 16 through 21, refer to the directions for Exercises 9 through 15 and find the fixed field of the automorphism or set of automorphisms of E .

16. τ_3

17. τ_3^2

18. $\{\tau_2, \tau_3\}$

19. $\tau_5\tau_2$

20. $\tau_5\tau_3\tau_2$

21. $\{\tau_2, \tau_3, \tau_5\}$

22. Refer to the directions for Exercises 9 through 15 for this exercise.

- Show that each of the automorphisms τ_2, τ_3 , and τ_5 is of order 2 in $G(E/\mathbb{Q})$. (Remember what is meant by the *order* of an element of a group.)
- Find the subgroup H of $G(E/\mathbb{Q})$ generated by the elements τ_2, τ_3 , and τ_5 , and give the group table. [Hint: There are eight elements.]
- Just as was done in Example 43.4, argue that the group H of part (b) is the full group $G(E/\mathbb{Q})$.

Concepts

In Exercises 23 and 24, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

- Two elements, α and β , of an algebraic extension E of a field F are *conjugate over F* if and only if they are both zeros of the same polynomial $f(x)$ in $F[x]$.
- Two elements, α and β , of an algebraic extension E of a field F are *conjugate over F* if and only if the evaluation homomorphisms $\phi_\alpha : F[x] \rightarrow E$ and $\phi_\beta : F[x] \rightarrow E$ have the same kernel.
- The fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(3 + \sqrt{2})$ are the same, of course. Let $\alpha = 3 + \sqrt{2}$.
 - Find a conjugate $\beta \neq \alpha$ of α over \mathbb{Q} .
 - Referring to part (a), compare the conjugation automorphism $\psi_{\sqrt{2}, -\sqrt{2}}$ of $\mathbb{Q}(\sqrt{2})$ with the conjugation automorphism $\psi_{\alpha, \beta}$.
- Determine whether each of the following is true or false.
 - For all $\alpha, \beta \in E$, there is always an automorphism of E mapping α onto β .
 - For α, β algebraic over a field F , there is always an isomorphism of $F(\alpha)$ onto $F(\beta)$.
 - For α, β algebraic and conjugate over a field F , there is always an isomorphism of $F(\alpha)$ onto $F(\beta)$.
 - Every automorphism of every field E fixes every element of the prime subfield of E .
 - Every automorphism of every field E fixes an infinite number of elements of E .
 - Every automorphism of every field E fixes at least two elements of E .
 - Every automorphism of every field E of characteristic 0 fixes an infinite number of elements of E .
 - All automorphisms of a field E form a group under function composition.
 - The set of all elements of a field E fixed by a single automorphism of E forms a subfield of E .
 - For fields $F \leq E \leq K$, $G(K/E) \leq G(K/F)$.

Proof Synopsis

- Give a one-sentence synopsis of the “if” part of Theorem 43.18.
- Give a one-sentence synopsis of the “only if” part of Theorem 43.18.

Theory

- Prove Theorem 43.2.
- Show that the only subfields of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ are \mathbb{Q} , $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(\sqrt{3})$, $\mathbb{Q}(\sqrt{6})$, and $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. [Hint: Show that a subfield of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ that is a degree 2 extension of \mathbb{Q} must be of the form $\mathbb{Q}(\sqrt{s})$ for some rational number s .]
- Let α be algebraic of degree n over F . Show that there are at most n different isomorphisms of $F(\alpha)$ onto a subfield of \bar{F} and leaving F fixed.
- Let $F(\alpha_1, \dots, \alpha_n)$ be an extension field of F . Show that any automorphism σ of $F(\alpha_1, \dots, \alpha_n)$ leaving F fixed is completely determined by the n values $\sigma(\alpha_i)$.

33. Let E be an algebraic extension of a field F , and let σ be an automorphism of E leaving F fixed. Let $\alpha \in E$. Show that σ induces a permutation of the set of all zeros of $\text{irr}(\alpha, F)$ that are in E .
34. Let E be an algebraic extension of a field F . Let $S = \{\sigma_i \mid i \in I\}$ be a collection of automorphisms of E such that every σ_i leaves each element of F fixed. Show that if S generates the subgroup H of $G(E/F)$, then $E_S = E_H$.
35. Let F be a finite field with characteristic p . Prove that the map $\phi : F \rightarrow F$ defined by $\phi(\alpha) = \alpha^p$ is a field automorphism. This automorphism is called the **Frobenius automorphism**.
36. Referring to Exercise 35, let F be a finite field of characteristic p , and let ϕ be the Frobenius automorphism on F . Prove that the fixed field $F_{\{\phi\}}$ is isomorphic with \mathbb{Z}_p .
37. Referring to Exercise 35, show that the finite assumption is necessary by finding an example of a field F with characteristic p , such that the map $\phi : F \rightarrow F$ given by $\phi(\alpha) = \alpha^p$ is not an automorphism.
38. We saw in Corollary 28.18 that the cyclotomic polynomial

$$\Phi_p(x) = \frac{x^{p-1}}{x-1} = x^{p-1} + x^{p-2} + \cdots + x + 1$$

is irreducible over \mathbb{Q} for every prime p . Let ζ be a zero of $\Phi_p(x)$, and consider the field $\mathbb{Q}(\zeta)$.

- a. Show that $\zeta, \zeta^2, \dots, \zeta^{p-1}$ are distinct zeros of $\Phi_p(x)$, and conclude that they are all the zeros of $\Phi_p(x)$.
- b. Deduce from Corollary 43.19 and part (a) of this exercise that $G(\mathbb{Q}(\zeta)/\mathbb{Q})$ is abelian of order $p - 1$.
- c. Show that the fixed field of $G(\mathbb{Q}(\zeta)/\mathbb{Q})$ is \mathbb{Q} . [Hint: Show that

$$\{\zeta, \zeta^2, \dots, \zeta^{p-1}\}$$

is a basis for $\mathbb{Q}(\zeta)$ over \mathbb{Q} , and consider which linear combinations of $\zeta, \zeta^2, \dots, \zeta^{p-1}$ are fixed by all elements of $G(\mathbb{Q}(\zeta)/\mathbb{Q})$.

39. Theorem 43.18 described conjugation isomorphisms for the case where α and β were conjugate algebraic elements over F . Is there a similar isomorphism of $F(\alpha)$ with $F(\beta)$ in the case that α and β are both transcendental over F ?
40. Let F be a field, and let x be an indeterminate over F . Determine all automorphisms of $F(x)$ leaving F fixed, by describing their values on x .
41. Prove the following sequence of theorems.
- a. An automorphism of a field E carries elements that are squares of elements in E onto elements that are squares of elements of E .
 - b. An automorphism of the field \mathbb{R} of real numbers carries positive numbers onto positive numbers.
 - c. If σ is an automorphism of \mathbb{R} and $a < b$, where $a, b \in \mathbb{R}$, then $\sigma(a) < \sigma(b)$.
 - d. The only automorphism of \mathbb{R} is the identity automorphism.

SECTION 44 SPLITTING FIELDS

In Example 43.4, $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ included all the zeros of the minimal polynomials for $\sqrt{2}$ and $\sqrt{3}$ over \mathbb{Q} . This is a key requirement in order to have the Galois correspondence between subgroups of the automorphism group of a field extension and intermediate fields. We saw in Example 43.14 that the correspondence failed due to the fact that the polynomial $x^3 - 2$ had only one zero in $\mathbb{Q}(\sqrt[3]{2})$. Definition 44.1 formalizes this idea.

- 44.1 Definition** Let F be a field and $P = \{f_1(x), f_2(x), \dots, f_s(x)\}$ be a finite set of polynomials in $F[x]$. An extension field K of F is a **splitting field of P over F** if every polynomial $f_k(x) \in P$ factors into linear factors in $K[x]$ and for any intermediate field E , $F \leq E < K$, at least one polynomial $f_j(x) \in P$ does not factor into linear factors in $E[x]$. A field K is a **splitting field for F** if E is a splitting field for some finite set of polynomials. ■

44.2 Example The field $E = \mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}(-1 + \sqrt{3}i)/2)$ is a splitting field of $\{x^3 - 2\}$ over \mathbb{Q} . The zeros of $x^3 - 2$ are

$$\sqrt[3]{2}, \quad \frac{\sqrt[3]{2}(-1 + \sqrt{3}i)}{2}, \quad \text{and} \quad \frac{\sqrt[3]{2}(-1 - \sqrt{3}i)}{2};$$

and each is an element of E . Also, any proper subfield of E would either not contain $\sqrt[3]{2}$ or not contain $\sqrt[3]{2}(-1 + \sqrt{3}i)/2$. \blacktriangle

Before using splitting fields, we need to verify that they actually exist! Here we give a proof based on the existence of an algebraic closure. Exercise 31 gives an alternative way of proving the existence without relying on an algebraic closure.

44.3 Theorem Let F be a field and $P = \{f_1, f_2, \dots, f_s\}$ a finite set of polynomials in $F[x]$. Then there is a splitting field K of P over F . Furthermore K is a finite extension of F .

Proof Let \bar{F} be an algebraic closure of F and let $\alpha_1, \alpha_2, \dots, \alpha_n \in \bar{F}$ be a list of all the zeros of all the polynomials in P . Let $K = F(\alpha_1, \alpha_2, \dots, \alpha_n)$. Since \bar{F} is algebraically closed, each polynomial f_k factors into linear factors in $\bar{F}[x]$ and, therefore, in $K[x]$. Furthermore, for any proper subfield E of K that contains F , E does not contain at least one α_j , which says that at least one f_k does not factor into linear factors in $E[x]$. Thus K is a splitting field of P over F . The fact that K is a finite extension of F follows from Corollary 40.6, the fact that each α_k is algebraic, and there are only a finite number of α_k . \blacklozenge

We only defined splitting fields using a finite collection of polynomials. It is also possible to use infinite sets of polynomials, but for our purposes finite sets of polynomials will do. We restrict our attention to splitting fields that are finite algebraic extensions.

In the proof of Theorem 44.3 we attached all the roots of all the polynomials in order to construct a splitting field. We will use the fact that a splitting field K of P over F has the property that $K = F(\alpha_1, \dots, \alpha_n)$, where $\alpha_1, \dots, \alpha_n$ are the zeros of the polynomials in P in an algebraic closure of F . Explicitly attaching some of the roots may be unnecessary. We saw in Example 44.2 that the splitting field of $x^3 - 2$ over \mathbb{Q} is $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}(-1 + \sqrt{3}i)/2)$. It is unnecessary to attach $\sqrt[3]{2}(-1 - \sqrt{3}i)/2$, the third zero of $x^3 - 2$, since it is already an element of $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}(-1 + \sqrt{3}i)/2)$, that is,

$$\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}(-1 + \sqrt{3}i)/2) = \mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}(-1 + \sqrt{3}i)/2, \sqrt[3]{2}(-1 - \sqrt{3}i)/2).$$

The Isomorphism Extension Theorem

Now that we know splitting fields exist, it is natural to ask if they are unique up to isomorphism. To answer the question we need the Isomorphism Extension Theorem. We first give a definition that makes the theorem easier to state.

44.4 Definition Let $\sigma : F \rightarrow F'$ be a field isomorphism; then $\sigma_x : F[x] \rightarrow F'[x]$, defined by

$$\sigma_x(a_0 + a_1x + \dots + a_nx^n) = \sigma(a_0) + \sigma(a_1)x + \dots + \sigma(a_n)x^n,$$

is the **polynomial extension of σ** . \blacksquare

The image $\sigma_x(f(x))$ is simply the polynomial in $F'[x]$ that corresponds to the polynomial $f(x)$ in $F[x]$ obtained by relabeling the coefficients via the isomorphism σ . It is intuitively clear that if $\sigma : F \rightarrow F'$ is an isomorphism, then $\sigma_x : F[x] \rightarrow F'[x]$ is also an isomorphism. You are asked to verify the details of the proof in Exercise 32.

44.5 Lemma Let $K = F(\alpha)$, where α is algebraic over F , and let $\sigma : F \rightarrow F'$ be a field isomorphism. If K' is an extension field of F' and $\beta \in K'$ is a zero of $\sigma_x(\text{irr}(\alpha, F))$, then there is a unique isomorphism $\phi : F(\alpha) \rightarrow F'(\beta)$ with $\sigma(a) = \phi(a)$ for all $a \in F$ and $\phi(\alpha) = \beta$.

Proof Let $p(x) = \text{irr}(\alpha, F)$ be the minimal polynomial for α over F . Then $p'(x) = \sigma_x(p(x))$ is an irreducible polynomial over F' since any factorization of $p'(x)$ in $F'[x]$ would give a factorization of $p(x)$ in $F[x]$. Since $p'(x)$ is irreducible over F' and β is a zero of $p'(x)$, $p'(\beta) = \text{irr}(\beta, F')$.

As in the proof of Kronecker's Theorem 39.3, we have an isomorphism $\psi_\alpha : F[x]/\langle p(x) \rangle \rightarrow F(\alpha)$, which is defined by the formula:

$$\psi_\alpha(f(x) + \langle p(x) \rangle) = f(\alpha)$$

for any $f(x) \in F[x]$.

We also have an isomorphism $\psi_\beta : F'[x]/\langle p'(x) \rangle \rightarrow F'(\beta)$ defined by

$$\psi_\beta(g(x) + \langle p'(x) \rangle) = g(\beta)$$

for any $g(x) \in F'[x]$.

A third isomorphism is $\theta : F[x]/\langle p(x) \rangle \rightarrow F'[x]/\langle p'(x) \rangle$ defined by

$$\theta(f(x) + \langle p(x) \rangle) = \sigma_x(f(x)) + \langle p'(x) \rangle$$

for any $f(x) \in F[x]$. The fact that θ is a homomorphism is Exercise 28 in Section 30. Also, using the homomorphism $\sigma_x^{-1} : F'[x] \rightarrow F[x]$, Exercise 28 in Section 30 shows that θ^{-1} is well defined. Therefore, θ is a one-to-one homomorphism mapping onto $F'[x]/\langle p'(x) \rangle$, or an isomorphism. The fact that θ is an isomorphism is intuitively clear since σ is an isomorphism between F and F' and the polynomial $p(x) \in F[x]$ corresponds to the polynomial $p'(x)$ by way of the isomorphism σ_x .

We now consider the isomorphism $\tau = \psi_\beta \circ \theta \circ \psi_\alpha^{-1}$. Let $a \in F$. We need to verify that $\tau(a) = \sigma(a)$. In the following calculation, we are thinking of a as a constant polynomial in $F[x]$, so $\sigma_x(a) = \sigma(a)$ and $\psi_\beta(\sigma_x(a) + \langle p'(x) \rangle) = \sigma(a)$. We have

$$\begin{aligned}\tau(a) &= \psi_\beta \circ \theta \circ \psi_\alpha^{-1}(a) \\ &= \psi_\beta(\theta(a + \langle p(x) \rangle)) \\ &= \psi_\beta(\sigma_x(a) + \langle p'(x) \rangle) \\ &= \psi_\beta(\sigma(a) + \langle p'(x) \rangle) \\ &= \sigma(a).\end{aligned}$$

Also,

$$\begin{aligned}\tau(\alpha) &= \psi_\beta \circ \theta \circ \psi_\alpha^{-1}(\alpha) \\ &= \psi_\beta(\theta(\alpha + \langle p(x) \rangle)) \\ &= \psi_\beta(\alpha + \langle p'(x) \rangle) \\ &= \beta.\end{aligned}$$

Uniqueness follows since an isomorphism $\rho : F(\alpha) \rightarrow F'(\beta)$ is completely determined by the values of $\rho(a)$ for $a \in F$ and $\rho(\alpha)$. \blacklozenge

44.6 Theorem (Isomorphism Extension Theorem) Let $K = F(\alpha_1, \alpha_2, \dots, \alpha_n)$ be a finite extension field of F , and let $\sigma : F \rightarrow F'$ be a field isomorphism. If K' contains a splitting field of $P = \{\sigma_x(\text{irr}(\alpha_k, F)) \mid 1 \leq k \leq n\}$ over F' , then σ can be extended to an isomorphism τ mapping K onto a subfield of K' .

Proof We first show that σ can be extended to an isomorphism mapping $F(\alpha_1)$ onto a subfield of K' . Since K' is the splitting field of P over F , there is a $\beta \in K'$ that is a zero of $\sigma_x(\text{irr}(\alpha_1, F))$. By Lemma 44.5, there is an isomorphism τ with the required properties.

We proceed by induction. Suppose τ is an isomorphism from $F(\alpha_1, \dots, \alpha_k)$ to some subfield of K' such that $\tau(\alpha) = \sigma(\alpha)$ for all $\alpha \in F$. We need to extend τ to an isomorphism from $F(\alpha_1, \dots, \alpha_k, \alpha_{k+1})$ to K' . There is a $\beta \in K'$ that is a zero of $g(x) = \tau_x(\text{irr}(\alpha_{k+1}, F(\alpha_1, \dots, \alpha_k)))$ since $g(x)$ is a factor of

$$\tau_x(\text{irr}(\alpha_{k+1}, F)) = \sigma_x(\text{irr}(\alpha_{k+1}, F)),$$

which factors into linear factors in K' . By Lemma 44.5, there is an isomorphism τ' mapping $F(\alpha_1, \dots, \alpha_{k+1})$ to a subfield of K' that extends τ . So by induction, σ can be extended to an isomorphism that maps K onto a subfield of K' . \blacklozenge

44.7 Example We consider the extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}(\sqrt{2})$. The map

$$\sigma : \mathbb{Q}(\sqrt{2}) \rightarrow \mathbb{Q}(\sqrt{2})$$

defined by

$$\sigma(a + b\sqrt{2}) = a - b\sqrt{2}$$

is an isomorphism. Since $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is a splitting field of $x^2 - 3$ over $\mathbb{Q}(\sqrt{2})$, σ can be extended to an isomorphism $\tau : \mathbb{Q}(\sqrt{2}, \sqrt{3}) \rightarrow \mathbb{Q}(\sqrt{2}, \sqrt{3})$ by Theorem 44.6. The proof of the theorem actually tells us more. We can choose τ to map to either zero of $x^2 - 3$, so we have two different choices for τ . In the notation of Example 43.4, these are the automorphisms σ and γ .

We could have used the identity map from $\mathbb{Q}(\sqrt{2})$ to $\mathbb{Q}(\sqrt{2})$ instead of σ for the isomorphism. Extending the identity isomorphism to $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ gives the automorphisms labeled ι and τ in Example 43.4. The Isomorphism Extension Theorem provides us with a simple way to show the existence of automorphisms that would otherwise be tedious to verify. \blacktriangle

Properties of Splitting Fields

We now show that given a finite set of polynomials, $P \subseteq F[x]$, a splitting field of P over F is unique up to isomorphism.

44.8 Theorem Let F be a field, $P = \{f_1, f_2, \dots, f_s\} \subseteq F[x]$ a finite set of polynomials, and both K and K' splitting fields of P over F . Then there is an isomorphism $\sigma : K \rightarrow K'$, which is the identity map on F .

Proof Let K and K' both be splitting fields of P over F . Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be the zeros of the polynomials of P in K and $\beta_1, \beta_2, \dots, \beta_m$ the zeros of the polynomials of P in K' . Then

$$K = F(\alpha_1, \dots, \alpha_n) \quad \text{and} \quad K' = F(\beta_1, \dots, \beta_m).$$

The extensions K and K' are finite over F since the α_i and β_j are algebraic. The extension K' over F is a splitting field of P , so it is also a splitting field of P' , the set of all irreducible factors over F of the polynomials in P . By Theorem 44.6, there is an isomorphism τ mapping K onto a subfield of K' that fixes the field F . Furthermore, since τ preserves the degree of the extension over F , the degree of the extension K' over F is greater than or equal to the degree of the extension K over F . Similarly, there is an isomorphism mapping K' onto a subfield of K , and the degree of the extension K over F is greater than or equal to the degree of the extension K' over F . Thus K and K' have the same degree as extensions of F . Since $\tau(K) \leq K'$, and each has the same degree as an extension over F , τ is an isomorphism mapping K onto K' . \blacklozenge

Theorem 44.8 says that it does not matter how you construct the splitting field for a fixed set of polynomials, you will always get the same field up to isomorphism fixing F . Because of this we will often speak of *the* splitting field of a set of polynomials instead of *a* splitting field.

44.9 Definition Let E be an extension field of F . A polynomial $f(x) \in F[x]$ **splits in E** if it factors into linear factors in $E[x]$. ■

44.10 Example The polynomial $x^4 - 5x^2 + 6 \in \mathbb{Q}[x]$ splits in the field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ since

$$x^4 - 5x^2 + 6 = (x^2 - 2)(x^2 - 3) = (x + \sqrt{2})(x - \sqrt{2})(x + \sqrt{3})(x - \sqrt{3}).$$



44.11 Theorem Let E be a finite extension of the field F . Then E is the splitting field of some finite set of polynomials in $F[x]$ if and only if for every field extension K over E and for every isomorphism σ that fixes all the elements of F and maps E onto a subfield of K , σ is an automorphism of E .

Proof We first assume that E is the splitting field for some set of polynomials

$$P = \{f_1(x), f_2(x), \dots, f_s(x)\}.$$

Let $\alpha_1, \dots, \alpha_n$ be the zeros in E of the polynomials in P . Then $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$. Let K be a field extension of E . Since all the polynomials $f_i(x)$ split in $E[x]$, all the zeros of $f_i(x)$ in K are actually in E . Let σ be an isomorphism from E to a subfield of K that fixes elements of F . Since σ maps each α_k to a zero of $f_i(x)$, for some i , $\sigma(\alpha_k) \in E$. Thus σ maps E into E . Since σ is an isomorphism, isomorphisms preserve the degree of the extension, and the degree of E over F is finite, σ is an isomorphism mapping E onto E . Thus σ is an automorphism of E .

We next assume that for any field extension K over E and any isomorphism σ that fixes all the elements of F and maps E to a subfield of K , σ is an automorphism of E . Since E is a finite extension of F , $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$ for some elements $\alpha_k \in E$ that are algebraic over F . Let $f_k(x) = \text{irr}(\alpha_k, F)$ be the minimal polynomial for α_k over F and $P = \{f_k(x) \mid 1 \leq k \leq n\}$. We show that E is the splitting field of P over F . Suppose by way of contradiction that some $f_k(x)$ does not split in E . By reordering the α_k we can assume that $k = 1$. Let \bar{E} be the algebraic closure of E . So $f_1(x)$ factors into linear factors in \bar{E} , which says that there is an element $\beta \in \bar{E}$, $\beta \notin E$, and β is a zero of $f_1(x) = \text{irr}(\alpha_1, F)$. Thus, α_1 and β are conjugates over F . By Theorem 43.18, there is an isomorphism

$$\psi_{\alpha_1, \beta} : F(\alpha_1) \rightarrow F(\beta)$$

that fixes all the elements of F and maps α_1 to β . Since \bar{E} contains the splitting field of $\{\psi_{\alpha_1, \beta}(f_k(\alpha_k)) \mid 1 \leq k \leq n\}$, by the Isomorphism Extension Theorem 44.6, $\psi_{\alpha_1, \beta}$ extends to an isomorphism σ mapping E onto a subfield of \bar{E} . But

$$\sigma(\alpha_1) = \psi_{\alpha_1, \beta}(\alpha_1) = \beta \notin E.$$

This gives a contradiction, which implies that each $f_k(x)$ splits in $E[x]$. Since each α_k is a zero of $f_k(x)$ and $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$, E is the smallest subfield of \bar{E} where each $f_k(x)$ splits. Thus E is a splitting field of P over F . ◆

The following corollary highlights one of the very strong properties of splitting fields.

44.12 Corollary If K is a finite splitting field over F and K contains one zero of an irreducible polynomial $f(x) \in F[x]$, then $f(x)$ splits in $K[x]$.

Proof Suppose by way of contradiction that $f(x)$ is irreducible over F , $f(x)$ has a zero α in $K = F(\alpha_1, \alpha_2, \dots, \alpha_n)$, and $f(x)$ does not split in K . Let \bar{K} be the algebraic closure of K . By our assumption, there is a $\beta \in \bar{K}$ that is a zero of $f(x)$ and $\beta \notin K$. Theorem 43.18, the Conjugation Isomorphism Theorem, says there is an isomorphism

$$\psi_{\alpha, \beta} : F(\alpha) \rightarrow F(\beta).$$

Since \bar{K} is algebraically closed, it contains the splitting field of

$$\{(\psi_{\alpha, \beta})_x(\text{irr}(\alpha_k, F(\alpha))) \mid 1 \leq k \leq n\}$$

over $F(\beta)$. The Isomorphism Extension Theorem allows us to extend $\psi_{\alpha, \beta}$ to an isomorphism σ mapping K onto a subfield of \bar{K} with $\sigma(\alpha) = \beta \notin K$, which contradicts Theorem 44.11. Thus $f(x)$ splits in $K[x]$. \blacklozenge

Corollary 44.12 tells us that if K is a splitting field of P over F and the irreducible polynomial $f(x) \in F[x]$ has a zero in K , then K contains the splitting field of $f(x)$ over F . It is surprising at first glance that a multiple of $f(x)$ need not be in the set P .

44.13 Example As we have seen, $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is the splitting field of $\{x^2 - 2, x^2 - 3\}$ over \mathbb{Q} . We have $\alpha = \sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$ and, as can easily be checked, α is a zero of

$$(x^2 - 5)^2 - 24 = x^4 - 10x^2 + 1.$$

With some effort, it can also be checked that $x^4 - 10x^2 + 1$ is irreducible over \mathbb{Q} . Thus $\text{irr}(\alpha, \mathbb{Q}) = x^4 - 10x^2 + 1$. By Corollary 44.12, $x^4 - 10x^2 + 1$ splits in $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ and $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ contains a splitting field K of $x^4 - 10x^2 + 1$ over \mathbb{Q} . Since

$$\mathbb{Q}(\sqrt{2} + \sqrt{3}) \leq K \leq \mathbb{Q}(\sqrt{2}, \sqrt{3})$$

and the two end fields have the same degree, 4, over \mathbb{Q} ,

$$\mathbb{Q}(\sqrt{2} + \sqrt{3}) = K = \mathbb{Q}(\sqrt{2}, \sqrt{3}).$$

We have two interesting results. First, the splitting field of $x^4 - 10x^2 + 1$ over \mathbb{Q} is $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, and second, although $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ does not appear to be a simple extension of \mathbb{Q} , it is. In the next section we will find that under mild conditions, every finite extension is a simple extension.

A challenging high school exercise is to use the quadratic formula to find all the zeros of $x^4 - 10x^2 + 1$ and rewrite them to see that they are all in both $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ and $\mathbb{Q}(\sqrt{2} + \sqrt{3})$. \blacktriangle

Theorem 44.11 gives a condition on a finite field extension $F \leq E$ that is equivalent to E being a splitting field. The condition involves looking at all possible extensions of E . Corollary 44.14 simplifies the condition significantly. Instead of looking at all extensions of E , Corollary 44.14 only requires looking at any one splitting field over F that contains E .

44.14 Corollary Let $F \leq E \leq K$ be fields with K a finite splitting field over F . Then E is a splitting field over F if and only if every isomorphism σ that fixes F and maps E to a subfield of K is an automorphism of E .

Proof Theorem 44.11 says that if E is a splitting field over F , then every isomorphism σ mapping E to a subfield of K that fixes F is an automorphism of E . This proves the only if direction.

We next assume that every isomorphism σ mapping E to a subfield of K that fixes F is an automorphism of E . Let $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$. Let $f_k(x) = \text{irr}(\alpha_k, F)$ and $P = \{f_k(x) \mid 1 \leq k \leq n\}$. We first show that in the algebraic closure, \bar{K} , of K , every conjugate

over F of every α_k is actually in E . By Theorems 43.18 and 44.6, for any conjugate $\beta \in \bar{K}$ of α_k over F , there is an isomorphism σ that fixes F , maps E onto a subfield of the algebraic closure \bar{K} , and maps α_k to β . Now $\sigma(\alpha_j)$ is a conjugate of α_j over F for each $1 \leq j \leq n$. That is, both α_j and $\sigma(\alpha_j)$ are zeros of $f_j(x)$. By Corollary 44.12, $f_j(x)$ splits in K , so $\sigma(\alpha_j) \in K$ for each j . Thus

$$\sigma(E) = \sigma(F(\alpha_1, \alpha_2, \dots, \alpha_n)) \subseteq K.$$

By our assumption, σ is an automorphism of E , so in particular, $\beta \in E$. We have shown that E contains all the conjugates of $\alpha_1, \alpha_2, \dots, \alpha_n \in \bar{K}$ over F . Since each $f_k(x)$ splits in the algebraically closed field \bar{K} , each $f_k(x)$ also splits in $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$. Since the splitting field of P over F contains E , E is the splitting field of P over F . ◆

44.15 Example We let $E = \mathbb{Q}(\sqrt[3]{2})$ and let K be the splitting field of the irreducible polynomial $x^3 - 2$ over \mathbb{Q} . The field K contains $\sqrt[3]{2}$, one zero of $\text{irr}(\sqrt[3]{2}, \mathbb{Q}) = x^3 - 2$, but it does not contain the other two zeros, $\sqrt[3]{2}(-1 \pm \sqrt{3}i)/2$. We can see that E is not the splitting field of any set of polynomials over \mathbb{Q} from Corollary 44.12. Alternatively, we can use the conjugation isomorphism theorem to show there is an isomorphism mapping $\mathbb{Q}(\sqrt[3]{2})$ to $\mathbb{Q}(\sqrt[3]{2}(-1 + \sqrt{3}i)/2) \leq K$. By Corollary 44.14, again we see that E is not a splitting field over \mathbb{Q} . ▲

■ EXERCISES 44

Computations

In Exercises 1 through 6, find the degree over \mathbb{Q} of the splitting field over \mathbb{Q} of the given polynomial in $\mathbb{Q}[x]$.

- | | | |
|--------------|--------------|-------------------------|
| 1. $x^2 + 3$ | 2. $x^4 - 1$ | 3. $(x^2 - 2)(x^2 - 3)$ |
| 4. $x^3 - 3$ | 5. $x^3 - 1$ | 6. $(x^2 - 2)(x^3 - 2)$ |

Refer to Example 44.2 for Exercises 7 through 9.

7. What is the order of $G(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$?
8. What is the order of $G(\mathbb{Q}(\sqrt[3]{2}, i\sqrt{3})/\mathbb{Q})$?
9. What is the order of $G(\mathbb{Q}(\sqrt[3]{2}, i\sqrt{3})/\mathbb{Q}(\sqrt[3]{2}))$?
10. Let α be a zero of $x^3 + x^2 + 1$ over \mathbb{Z}_2 . Show that $x^3 + x^2 + 1$ splits in $\mathbb{Z}_2(\alpha)$. [Hint: There are eight elements in $\mathbb{Z}_2(\alpha)$. Exhibit two more zeros of $x^3 + x^2 + 1$, in addition to α , among these eight elements. Alternatively, use the results of Section 42.]

Let $E = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$. It can be shown that $[E : \mathbb{Q}] = 8$. In Exercises 11 through 13, for the given isomorphic mappings of a subfield of E , give all extensions of the mapping to an isomorphic mapping of E onto a subfield of \mathbb{C} . Describe the extensions by giving values on the generating set $\{\sqrt{2}, \sqrt{3}, \sqrt{5}\}$ for E over \mathbb{Q} .

11. $\iota : \mathbb{Q}(\sqrt{2}, \sqrt{15}) \rightarrow \mathbb{Q}(\sqrt{2}, \sqrt{15})$, where ι is the identity map.
12. $\sigma : \mathbb{Q}(\sqrt{2}, \sqrt{15}) \rightarrow \mathbb{Q}(\sqrt{2}, \sqrt{15})$, where $\sigma(\sqrt{2}) = \sqrt{2}$ and $\sigma(\sqrt{15}) = -\sqrt{15}$.
13. $\Psi_{\sqrt{30}, -\sqrt{30}} : \mathbb{Q}(\sqrt{30}) \rightarrow \mathbb{Q}(\sqrt{30})$

In Exercises 14 through 16, let

$$\alpha_1 = \sqrt[3]{2}, \quad \alpha_2 = \frac{\sqrt[3]{2} - 1 + \sqrt{3}i}{2}, \quad \text{and} \quad \alpha_3 = \frac{\sqrt[3]{2} - 1 - \sqrt{3}i}{2},$$

where $\sqrt[3]{2}$ is the real number whose cube is 2. The zeros of $x^3 - 2$ are α_1, α_2 , and α_3 .

14. Describe all extensions of the identity map on \mathbb{Q} to an isomorphism mapping $\mathbb{Q}(\sqrt[3]{2})$ onto a subfield of \mathbb{C} .
15. Describe all extensions of the identity map on \mathbb{Q} to an isomorphism mapping $\mathbb{Q}(\sqrt{3}i, \sqrt[3]{2})$ onto a subfield of \mathbb{C} .

16. Describe all extensions of the automorphism $\Psi_{\sqrt{3}i, -\sqrt{3}i}$ on $\mathbb{Q}(\sqrt{3}i)$ to an isomorphism mapping $\mathbb{Q}(\sqrt{3}i, \sqrt[3]{2})$ onto a subfield of \mathbb{C} .
17. Let σ be an automorphism of $\mathbb{Q}(\pi)$ that maps π onto $-\pi$.
- Describe the fixed field of σ .
 - Describe all extensions of σ to an isomorphism mapping the field $\mathbb{Q}(\sqrt{\pi})$ onto a subfield of the splitting field of $x^2 + \pi$ over $\mathbb{Q}(\pi)$.

Concepts

In Exercise 18, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

18. A polynomial $f(x)$ in $F[x]$ splits in an extension field E of F if and only if it factors in $E[x]$ into a product of polynomials of lower degree.
19. Let $f(x)$ be a polynomial in $F[x]$ of degree n . Let E be a splitting field of $f(x)$ over F . What bounds can be put on $[E : F]$?
20. Determine whether each of the following is true or false.
- Let $\alpha, \beta \in E$, where E is a splitting field over F . Then there exists an automorphism of E leaving F fixed and mapping α onto β if and only if $\text{irr}(\alpha, F) = \text{irr}(\beta, F)$.
 - If $f(x) \neq g(x)$ are polynomials in $\mathbb{Q}[x]$, F is the splitting field of $f(x)$ over \mathbb{Q} , and K is the splitting field of $g(x)$ over \mathbb{Q} , then $F \neq K$.
 - \mathbb{R} is a splitting field over \mathbb{R} .
 - \mathbb{C} is a splitting field over \mathbb{R} .
 - $\mathbb{Q}(i)$ is a splitting field over \mathbb{Q} .
 - $\mathbb{Q}(\pi)$ is a splitting field over $\mathbb{Q}(\pi^2)$.
 - For every splitting field E over F , every isomorphic mapping of E is an automorphism of E .
 - For every splitting field E over F , where $E \leq K$, every isomorphism mapping E onto a subfield of K is an automorphism of E .
 - For every splitting field E over F , where $E \leq K$, every isomorphism mapping E onto a subfield of K and leaving F fixed is an automorphism of E .
 - If E is a splitting field over F and $\alpha \in E$, then $\deg(\alpha, F)$ divides $[E : F]$.

21. Show by an example that Corollary 44.12 is no longer true if the word *irreducible* is deleted.

22. Is $|G(E/F)|$ multiplicative for finite towers of finite extensions, that is, is

$$|G(K/F)| = |G(K/E)||G(E/F)| \quad \text{for} \quad F \leq E \leq K?$$

Why or why not? [Hint: Use Exercises 7 through 9.]

Theory

23. Show that if a finite extension E of a field F is a splitting field over F , then E is a splitting field of one polynomial in $F[x]$.
24. Show that if $[E : F] = 2$, then E is a splitting field over F .
25. Show that for $F \leq E \leq \bar{F}$, E is a splitting field over F if and only if E contains all conjugates over F in \bar{F} for each of its elements.
26. Show that the splitting field K of $\{x^2 - 2, x^2 - 5\}$ over \mathbb{Q} is $\mathbb{Q}(\sqrt{2} + \sqrt{5})$.
27. Show that

$$G(\mathbb{Q}(\sqrt[3]{2}, i\sqrt{3})/\mathbb{Q}(i\sqrt{3})) \simeq (\mathbb{Z}_3, +).$$

28. a. Show that an automorphism leaving F fixed of a splitting field E over F of a polynomial $f(x) \in F[x]$ permutes the zeros of $f(x)$ in E .
- b. Show that an automorphism leaving F fixed of a splitting field E over F of a polynomial $f(x) \in F[x]$ is completely determined by the permutation of the zeros of $f(x)$ in E given in part (a).

- c. Show that if E is a splitting field over F of a polynomial $f(x) \in F[x]$, then $G(E/F)$ can be viewed in a natural way as a certain group of permutations.
29. Let K be the splitting field of $x^3 - 2$ over \mathbb{Q} . Use Exercise 28 to show that $G(K/\mathbb{Q})$ is isomorphic with S_3 , the symmetric group on three letters. [Hint: Use Theorem 43.18 and complex conjugation to find enough elements that the Theorem of Lagrange and Exercise 28 imply the result.]
30. Show that for a prime p , the splitting field over \mathbb{Q} of $x^p - 1$ is of degree $p - 1$ over \mathbb{Q} . [Hint: Refer to Corollary 28.18.]
31. Let P be a finite set of polynomials with coefficients in the field F . Prove that a splitting field of P over F exists without using the algebraic closure of F .
32. Let $\sigma : F \rightarrow F'$ be a field isomorphism. Show that $\sigma_x : F[x] \rightarrow F'[x]$, as defined in Definition 44.4, is a ring isomorphism.

SECTION 45 SEPARABLE EXTENSIONS

Counting Zeros of Irreducible Polynomials

There is a technical issue in Galois theory that we have not yet discussed. Is it possible for an irreducible polynomial $f(x)$ over a field F to have fewer than $\deg(f(x))$ zeros in a splitting field over F ? We will see, for essentially all the fields we will consider, that the answer is no. We start with a calculus-based proof of this fact for subfields of the complex numbers.

45.1 Theorem Let $f(x)$ be an irreducible polynomial of degree n with coefficients in the field $F \subseteq \mathbb{C}$. Then the splitting field for $f(x)$ over F contains n distinct zeros of $f(x)$.

Proof We can assume that the coefficient of x^n is 1. We can also assume that $n \geq 2$ since otherwise the theorem is trivially true. Then

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0.$$

Since \mathbb{C} is algebraically closed, we can write $f(x) = (x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_n)$ where $\alpha_1, \alpha_2, \dots, \alpha_n$ are the zeros of $f(x)$ in \mathbb{C} . Since $f(x)$ is monic and irreducible, $f(x)$ is the minimal polynomial over F for α_k . We show that no two of the α_k are equal. We use proof by contradiction and assume that two of the α_k are equal. Then in $\mathbb{C}[x]$, $f(x) = (x - \alpha_k)^2 q(x)$ for some polynomial $q(x)$. Since we are considering polynomials over \mathbb{C} , we can use the derivative of a polynomial, which we denote in the standard way as $f'(x)$. The product rule gives

$$f'(x) = 2(x - \alpha_k)q(x) + (x - \alpha_k)^2 q'(x) = (x - \alpha_k)(2q(x) + (x - \alpha_k)q'(x)).$$

Therefore α_k is a zero of $f'(x)$. By the usual formula for the derivative of a polynomial

$$f'(x) = nx^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1$$

and $f'(x) \in F[x]$. Since $f'(x)$ has degree $n-1 \geq 1$ and α_k is a zero of $f'(x)$, $f(x)$ is not the minimal polynomial for α_k over F . This gives a contradiction and proves that the α_k are distinct.

We can construct a splitting field of $f(x)$ over F as a subfield of \mathbb{C} . So a splitting field for $f(x)$ over F contains n distinct zeros for $f(x)$. Since splitting fields are unique up to isomorphism, in any splitting field for $f(x)$ over F , $f(x)$ has n distinct zeros. ◆

Although the above proof applies to a special (but very important) case, essentially the same proof can be used for any field of characteristic zero. We can simply define the derivative of any polynomial $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \in F[x]$ to be

$$D(f)(x) = f'(x) = a_1 + (2 \cdot a_2)x + (3 \cdot a_3)x^2 + \cdots + (n \cdot a_n)x^{n-1},$$

where an integer times an element in a field has the usual meaning. In calculus this formula is derived using the limit definition of derivatives. For our purposes, we simply use this formula as the definition of the derivative, $D(f)(x) = f'(x)$. Exercise 15 in Section 42 uses this definition to prove the product rule and other essential rules that are required for the proof of Theorem 45.1. Exercise 13 in this Section asks for the details of the proof of Theorem 45.2.

45.2 Theorem Let $f(x)$ be an irreducible polynomial of degree n with coefficients in a field F of characteristic zero. Then $f(x)$ contains n distinct zeros in the splitting field for $f(x)$ over F .

Proof See Exercise 13. ◆

45.3 Definition Let $f(x) \in F[x]$, and let α be a zero of $f(x)$ in a splitting field E over F . If v is the largest positive integer such that $(x - \alpha)^v$ is a factor of $f(x)$ in $E[x]$, then α is a zero of $f(x)$ with multiplicity v . ■

Definition 45.3 does not specify which splitting field over F is to be used. It can be any splitting field that contains α . In Exercise 19, you are asked to show that the definition is independent of which splitting field is used. An equivalent statement of Theorem 45.2 is that if $f(x)$ is an irreducible polynomial with coefficients in a field F of characteristic 0, then every zero of $f(x)$ in any splitting field over F has multiplicity one.

Characteristic p

We have seen that for irreducible polynomials over a field F of characteristic zero, zeros of multiplicity two or greater cannot occur. We now turn our attention to the case where the characteristic is not zero.

45.4 Theorem Let F be a finite field of characteristic p . Any irreducible polynomial $f(x) \in F[x]$ has $k = \deg(f(x))$ distinct zeros in its splitting field.

Proof Let $E \leq \bar{F}$ be the splitting field of $f(x)$ over F and α a zero $f(x)$ in E . Then $|E| = p^n$, for some n . We can assume that the leading coefficient of $f(x)$ is 1. Therefore $f(x)$ is the minimal polynomial for α over F . By Theorem 42.3, α is a zero of the polynomial $x^{p^n} - x$, which implies that $f(x)$ divides $x^{p^n} - x$. The p^n zeros of $x^{p^n} - x$ are distinct in the algebraic closure \bar{E} by Lemma 42.8, so the linear factors of

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_k)$$

must also be distinct in E , and therefore, $f(x)$ has k distinct zeros in its splitting field. ◆

Theorems 45.2 and 45.4 say that any irreducible $f(x) \in F[x]$, where the field F is either finite or it has characteristic zero, has $\deg(f(x))$ distinct zeros of multiplicity one in its splitting field over F . The next example shows that this is not the case for all infinite fields of characteristic p .

45.5 Example Let p be a prime and $E = \mathbb{Z}_p(y)$, where y is an indeterminate. We let $F = \mathbb{Z}_p(y^p) \leq E$. For convenience, we let $t = y^p$, so $F = \mathbb{Z}_p(t)$. The extension E over F is algebraic since y is a zero of the polynomial $x^p - t$. By Corollary 39.14, $\text{irr}(y, F)$ divides $x^p - t$. We can factor $x^p - t$ in E as

$$x^p - t = x^p - y^p = (x - y)^p,$$

since we are in characteristic p . Furthermore, $y \notin F$, so the degree of $\text{irr}(y, F)$ is at least two. Therefore the number of distinct zeros of $\text{irr}(y, F)$ is one, but its degree is greater than one, showing that the finite assumption in Theorem 45.4 is necessary. \blacktriangle

Counting Automorphisms

Our goal is to associate intermediate fields in a field extension $F \leq K$ with subgroups of the group of automorphisms of K that fix F . In order to apply this correspondence, it is very helpful to know the number of automorphisms under consideration. Knowing that all irreducible polynomials in F have zeros of multiplicity one in a splitting field K gives us the information we need to count the automorphisms in $G(K/F)$.

- 45.6 Definition** An irreducible polynomial $f(x) \in F[x]$ of degree n is **separable** if in the splitting field K of $f(x)$ over F , $f(x)$ has n distinct zeros. An element α in an extension field of F is **separable** if $\text{irr}(\alpha, F)$ is a separable polynomial. A field extension $F \leq E$ is **separable** if every $\alpha \in E$ is separable over F . If every finite extension of a field F is separable, then F is **perfect**. \blacksquare

It is clear that an irreducible polynomial $f(x) \in F[x]$ is separable if and only if every zero of $f(x)$ in its splitting field over F has multiplicity one.

- 45.7 Theorem** Every field of characteristic 0 is perfect and every finite field is perfect.

Proof Let $F \leq E$ be a finite field extension, where either F has characteristic 0 or F is finite. Let $g(x)$ be the minimal polynomial for $\alpha \in E$ over F . By Theorems 45.2 and 45.4, $g(x)$ has $\deg(g(x))$ distinct zeros in the splitting field of $g(x)$ over F . \blacklozenge

- 45.8 Theorem** Let K be a separable extension of the field F and E an intermediate field. Then both the extensions K over E and E over F are separable.

Proof Let $\alpha \in K$. Then $\text{irr}(\alpha, F)$ has $\deg(\alpha, F)$ zeros, each with multiplicity one, in the splitting field of $\text{irr}(\alpha, F)$ over F , as K is separable over F . Since $\text{irr}(\alpha, E)$ divides $\text{irr}(\alpha, F)$, $\text{irr}(\alpha, E)$ has $\deg(\alpha, E)$ zeros, each with multiplicity one. Thus α is separable over E and K is a separable extension of E .

We now let $\beta \in E$. Therefore, $\beta \in K$, which implies that $\text{irr}(\beta, F)$ has $\deg(\beta, F)$ zeros in the splitting field of $\text{irr}(\beta, F)$ over F . Thus E is a separable extension of F . \blacklozenge

The following theorem is very useful for our purposes. It counts the number of isomorphisms mapping an intermediate field E , of a separable splitting field $F \leq K$, onto a subfield of K . In particular, using $E = K$, the theorem tells us the number of automorphisms in $G(K/F)$.

- 45.9 Theorem** Let K be a splitting field over F and $F \leq E \leq K$. If K is a separable extension over F , then the number of isomorphisms that map E onto a subfield of K that fix all the elements of F is $[E : F]$.

Proof Let $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$. We use induction on n to show that for each $1 \leq k \leq n$ the number of isomorphisms that fix elements of F and map $F(\alpha_1, \alpha_2, \dots, \alpha_k)$ onto a subfield of K is $[F(\alpha_1, \alpha_2, \dots, \alpha_k) : F]$. To save writing, we let $E_k = F(\alpha_1, \alpha_2, \dots, \alpha_k)$. For $n = 1$, $E_1 = F(\alpha_1)$ and $[E_1 : F] = \deg(\alpha_1, F)$. Furthermore, Corollary 43.19 says there is exactly one isomorphism that fixes elements of F and maps E_1 onto a subfield of K for each zero of $\text{irr}(\alpha_1, F)$ in K . Since K is a splitting field and α_1 is a zero of $\text{irr}(\alpha_1, F)$, Corollary 44.12 says that $\text{irr}(\alpha_1, F)$ splits into linear factors. But the extension E over F

is separable, so $\text{irr}(\alpha_1, F)$ has exactly $\deg(\alpha_1, F)$ zeros in K . Thus the number of isomorphisms mapping E_1 onto a subfield of K is

$$\deg(\alpha, F) = [E_1 : F].$$

We proceed with the induction step. We assume there are $[E_k : F]$ isomorphisms that fix F and map E_k onto a subfield of K . Let σ be one of these isomorphisms. We let $g(x) = \text{irr}(\alpha_{k+1}, E_k)$. Since $\sigma_x(g(x))$ is a factor of $\sigma_x(\text{irr}(\alpha_{k+1}, F)) = \text{irr}(\alpha_{k+1}, F)$, it follows that $\sigma_x(g(x))$ factors into $\deg(g(x))$ linear factors in $K[x]$. By Theorem 45.8, K is separable over E_k . Thus $\sigma_x(g(x))$ has exactly $\deg(g(x))$ distinct zeros in K . The map σ can be extended to exactly

$$\deg(\text{irr}(\alpha_{k+1}, E_k)) = [E_{k+1} : E_k]$$

isomorphisms from $E_{k+1} = E_k(\alpha_{k+1})$ onto a subfield of K by Lemma 44.5. By the induction hypothesis there are exactly $[E_k : F]$ isomorphisms σ that fix elements of F and map E_k onto a subfield of K , giving us a total of

$$[E_{k+1} : E_k] \cdot [E_k : F] = [E_{k+1} : F]$$

isomorphisms that fix elements of F and map E_{k+1} onto a subfield of K . This completes the induction step, and we conclude that there are exactly $[E_n : F] = [E : F]$ isomorphisms mapping E onto a subfield of K that fix elements of F . ◆

In the case where E is a separable splitting field over F , Theorems 44.11 and 45.9 imply that $G(E/F) = [E : F]$. However, if E is not a splitting field of F , then some of the isomorphisms fixing F and mapping E onto a subfield of K will map onto a subfield of K other than E .

45.10 Corollary Let E be a separable splitting field over F . Then $|G(E/F)| = [E : F]$.

Proof The Corollary follows immediately from Theorem 45.9. ◆

45.11 Corollary Let E be a splitting field over F where F is either a field of characteristic 0 or a finite field. Then $|G(E/F)| = [E : F]$.

Proof Since F is perfect by Theorem 45.7, the result follows from Corollary 45.10. ◆

45.12 Example In Example 43.4, we explicitly determined four automorphisms of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ that fix elements of \mathbb{Q} . Corollary 45.11 shows that there are exactly

$$[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{3})] \cdot [\mathbb{Q}(\sqrt{3}) : \mathbb{Q}] = 2 \cdot 2 = 4$$

automorphisms without explicitly writing out the formulas.

On the other hand, in Example 43.14, even though

$$[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3,$$

there is only one automorphism of $\mathbb{Q}(\sqrt[3]{2})$. This is due to the fact that that $\mathbb{Q}(\sqrt[3]{2})$ is not a splitting field over \mathbb{Q} and Theorem 45.9 does not apply. ▲

The Primitive Element Theorem

The primitive element theorem is a classic of field theory. It says that any finite separable extension of a field is a simple extension. We will find it useful in Section 46 where we prove the Galois correspondence.

45.13 Theorem **Primitive Element Theorem** Let E be a finite separable extension of a field F . Then there is an $\alpha \in E$ such that $E = F(\alpha)$. Any such element α is called a **primitive element**.

Proof We first consider the case where F is a finite field. In this case E^* , the units in E , is a cyclic group under multiplication by Theorem 28.7. Clearly $E = F(\alpha)$ where α is a generator of the cyclic group E^* .

We next assume that F is infinite and show that if $E = F(\beta, \gamma)$, then E has a primitive element. If $\gamma \in F$, then β is a primitive element. We therefore assume that γ is not in F . We seek a primitive element of the form $\alpha = \beta + a\gamma$ where $a \in F$. Let $f(x) = \text{irr}(\beta, F)$ and $g(x) = \text{irr}(\gamma, F)$. In the splitting field of $\{f(x), g(x)\}$ over E , let $\beta = \beta_1, \beta_2, \dots, \beta_n$ be the zeros of $f(x)$ and $\gamma = \gamma_1, \gamma_2, \dots, \gamma_m$ the zeros of $g(x)$. Since γ is not in F , $m \geq 2$. The only conditions we need on the element a are that $a \in F$ and that for each $1 \leq i \leq n$ and $2 \leq j \leq m$,

$$a \neq \frac{\beta_i - \beta}{\gamma - \gamma_j}.$$

There is certainly such an element $a \in F$ due to the fact that F is infinite and we only eliminate a finite number of possible values for a . Since $\beta_1 = \beta$ and $m \geq 2$, $a \neq 0$. We let

$$\alpha = \beta + a\gamma.$$

If $\alpha = \beta_i + a\gamma_j$ for some i and $j \neq 1$, then

$$\begin{aligned}\beta_i + a\gamma_j &= \beta + a\gamma \quad \text{and} \\ a &= \frac{\beta_i - \beta}{\gamma - \gamma_j},\end{aligned}$$

which is a contradiction. Thus for any i and $j \neq 1$, $\alpha \neq \beta_i + a\gamma_j$ or equivalently,

$$\alpha - a\gamma_j \neq \beta_i.$$

We now let

$$h(x) = f(\alpha - ax) \in F(\alpha)[x].$$

Since $h(x)$ is in $F(\alpha)[x]$ and $f(x) = \text{irr}(\beta, F) \in F[x] \leq F(\alpha)[x]$, the greatest common divisor of $h(x)$ and $f(x)$ is a polynomial in $F(\alpha)[x]$. For $j \neq 1$,

$$\begin{aligned}h(\gamma) &= f(\alpha - a\gamma) = f(\beta) = 0 \quad \text{and} \\ h(\gamma_j) &= f(\alpha - a\gamma_j) \neq 0,\end{aligned}$$

since the only zeros of $f(x)$ are the β_i and $\alpha - a\gamma_j \neq \beta_i$ for any i . Therefore the only common factor of $h(x)$ and $f(x)$ is $x - \beta$, so the greatest common divisor of $h(x)$ and $f(x)$ is $x - \beta$. Thus $x - \beta \in F(\alpha)[x]$ and $\beta \in F(\alpha)$. Since $\alpha, \beta \in F(\alpha)$,

$$\gamma = \frac{\alpha - \beta}{a} \in F(\alpha).$$

We conclude that $F(\beta, \gamma) \leq F(\alpha)$. Clearly $F(\alpha) \leq F(\beta, \gamma)$, so

$$F(\alpha) = F(\beta, \gamma).$$

By a straightforward induction argument, any finite separable extension of F has a primitive element. ◆

To illustrate the construction of a primitive element for a given extension we provide the following example.

45.14 Example In Example 44.13, we saw that $\sqrt{2} + \sqrt{3}$ is a primitive element for the extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. Let us follow the proof of Theorem 45.13 to find other primitive elements $\alpha \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Let $f(x) = \text{irr}(\sqrt{2}, \mathbb{Q}) = x^2 - 2$ and $g(x) = \text{irr}(\sqrt{3}, \mathbb{Q}) = x^2 - 3$. So

$$\beta = \beta_1 = \sqrt{2}, \quad \beta_2 = -\sqrt{2}, \quad \gamma = \gamma_1 = \sqrt{3}, \quad \text{and} \quad \gamma_2 = -\sqrt{3}.$$

Thus a can be any rational number other than

$$\frac{\sqrt{2} - \sqrt{2}}{\sqrt{3} - (-\sqrt{3})} = 0 \quad \text{and} \quad \frac{-\sqrt{2} - \sqrt{2}}{\sqrt{3} - (-\sqrt{3})} = -\frac{\sqrt{2}}{\sqrt{3}}.$$

Since $-\frac{\sqrt{2}}{\sqrt{3}}$ is not a rational number, we can take $a = 1, 2, 1/2, -17/42$, or any rational number other than 0. Using $a = 2$, we have that

$$\alpha = \beta + a\gamma = \sqrt{2} + 2\sqrt{3}.$$

Thus

$$\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + 2\sqrt{3})$$

and in general,

$$\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + a\sqrt{3})$$

for any rational number a other than 0. ▲

45.15 Corollary If F is either a finite field or a field of characteristic 0, then every finite extension of F is a simple extension.

Proof This is an immediate consequence of Theorems 45.7 and 45.13. ◆

Normal Extensions

We have now investigated the essential conditions on a field extension $F \leq E$ that are required in order to apply Galois theory. The requirements are that E is a separable splitting field over F .

45.16 Definition A finite extension E of F is a **normal extension of F** if E is a separable splitting field over F . If E is a normal extension of F , then $G(E/F)$ is the **Galois group of E over F** . The Galois group is sometimes denoted by $\text{Gal}(E/F)$. ■

Although one can define an infinite normal extension, for our purposes we will restrict our attention to finite extensions. In what follows, when we refer to a normal extension, it will be assumed that the extension is finite.

45.17 Theorem Let K be a normal extension of F and let E be an intermediate field of the extension, $F \leq E \leq K$. Then K is a normal extension of E and $|G(K/E)| = [K : E]$.

Proof Since K is a splitting field over F , there are polynomials $f_1(x), f_2(x), \dots, f_r(x) \in F[x]$ with zeros $\alpha_1, \alpha_2, \dots, \alpha_k$ such that $K = F(\alpha_1, \alpha_2, \dots, \alpha_k)$ and each f_i factors into linear factors in K . Then $K = E(\alpha_1, \alpha_2, \dots, \alpha_k)$, also. Thus K is the splitting field of $\{f_1(x), f_2(x), \dots, f_r(x)\}$ over E . Furthermore, Theorem 45.8 states that K is a separable extension of E , which implies that K is a normal extension of E .

Since K is a separable splitting field over E , Corollary 45.10 says that $|G(K/E)| = [K : E]$. ◆

45.18 Corollary If $F \leq E \leq K$ where K is a normal extension of F , then $G(K/E)$ is a subgroup of $G(K/F)$ with index $(G(K/F) : G(K/E)) = [E : F]$.

Proof Theorem 45.17 says that K is a normal extension of E . Each isomorphism $\sigma \in G(K/E)$ fixes all the elements of E and, therefore, σ fixes all the elements of F . Thus $\sigma \in G(K/F)$ and $G(K/E) \leq G(K/F)$.

We have

$$\begin{aligned}(G(K/F) : G(K/E)) &= \frac{|G(K/F)|}{|G(K/E)|} \\ &= \frac{[K : F]}{[K : E]} \\ &= [E : F].\end{aligned}$$

◆

45.19 Example In Example 44.2 we saw that the splitting field of $x^3 - 2$ over \mathbb{Q} is

$$K = \mathbb{Q}\left(\sqrt[3]{2}, \sqrt[3]{2}^{-1} + \frac{\sqrt{3}i}{2}\right) = \mathbb{Q}(\sqrt[3]{2}, \sqrt{3}i).$$

The degree of the extension of K over $\mathbb{Q}(\sqrt[3]{2})$ is

$$[K : \mathbb{Q}(\sqrt[3]{2})] = \deg(\sqrt{3}i, \mathbb{Q}(\sqrt[3]{2})) = 2.$$

Also, the degree of the extension $\mathbb{Q}(\sqrt[3]{2})$ over \mathbb{Q} is

$$[\mathbb{Q}(\sqrt[3]{2}), \mathbb{Q}] = \deg(\sqrt[3]{2}, \mathbb{Q}) = 3,$$

and

$$[K : \mathbb{Q}] = [K : \mathbb{Q}(\sqrt[3]{2})] \cdot [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 6.$$

Since \mathbb{Q} is a perfect field, K is a separable and, therefore, a normal extension of \mathbb{Q} . Thus Corollary 45.18 applies and we have

$$|G(K/\mathbb{Q})| = 6, \quad |G(K/\mathbb{Q}(\sqrt[3]{2}))| = 2, \quad \text{and} \quad (G(K/\mathbb{Q}) : G(K/\mathbb{Q}(\sqrt[3]{2}))) = 3.$$

Up to isomorphism, there are two groups of order 6, \mathbb{Z}_6 and S_3 . We will see in Section 46 that $G(K/\mathbb{Q})$ is isomorphic with S_3 . ▲

■ EXERCISES 45

Computations

In Exercises 1 through 4, find an α such that the given field is $\mathbb{Q}(\alpha)$. Show that your α is indeed in the given field. Verify by direct computation that the given generators for the extension of \mathbb{Q} can indeed be expressed as formal polynomials in your α with coefficients in \mathbb{Q} .

- | | |
|--|--|
| 1. $\mathbb{Q}(\sqrt{2}, \sqrt[3]{2})$ | 2. $\mathbb{Q}(\sqrt[3]{2}, \sqrt{2})$ |
| 3. $\mathbb{Q}(\sqrt{2}, \sqrt{5})$ | 4. $\mathbb{Q}(i, \sqrt[3]{2})$ |

Concepts

In Exercises 5 and 6, correct the definition of the italicized term without reference to the text, if correction is needed, so that it is in a form acceptable for publication.

5. Let E be a splitting field over F . The *multiplicity of a zero* $\alpha \in E$ of a polynomial $f(x) \in F[x]$ is $v \in \mathbb{Z}^+$ if and only if $(x - \alpha)^v$ is a factor of $f(x)$ in $F[x]$.
6. Let E be an extension of a field F . An element α in E is *separable over F* if and only if α is a zero of multiplicity 1 of $\text{irr}(\alpha, F)$.
7. Give an example of an $f(x) \in \mathbb{Q}[x]$ that has no zeros in \mathbb{Q} but whose zeros in \mathbb{C} are all of multiplicity 2. Explain how this is consistent with Theorem 45.7, which shows that \mathbb{Q} is perfect.
8. Determine whether each of the following is true or false.
 - a. Every finite extension of every field F is separable over F .
 - b. Every finite extension of every finite field F is separable over F .

- c. Every field of characteristic 0 is perfect.
- d. Every polynomial of degree n over every field F always has n distinct zeros in \bar{F} .
- e. Every polynomial of degree n over every perfect field F always has n distinct zeros in \bar{F} .
- f. Every irreducible polynomial of degree n over every perfect field F always has n distinct zeros in \bar{F} .
- g. Every algebraically closed field is perfect.
- h. Every field F has an algebraic extension E that is perfect.
- i. If E is a finite separable splitting field extension of F , then $|G(E/F)| = [E : F]$.
- j. If a field F is neither finite nor of characteristic 0, then F is not a perfect field.

Theory

9. Show that $\{1, y, \dots, y^{p-1}\}$ is a basis for $\mathbb{Z}_p(y)$ over $\mathbb{Z}_p(y^p)$, where y is an indeterminate. Referring to Example 45.5, conclude by a degree argument that $x^p - t$ is irreducible over $\mathbb{Z}_p(t)$, where $t = y^p$.

10. Prove that if E is an algebraic extension of a perfect field F , then E is perfect.

11. Let E be a finite field of order p^n .

- a. Show that the Frobenius automorphism σ_p , defined in Exercise 35 of Section 43, has order n .
- b. Deduce from part (a) that $G(E/\mathbb{Z}_p)$ is cyclic of order n with generator σ_p . [Hint: Remember that

$$|G(E/F)| = [E : F]$$

for a normal field extension E over F .]

12. Let $f(x) \in F[x]$, and let $\alpha \in \bar{F}$ be a zero of $f(x)$ of multiplicity v . Show that $v > 1$ if and only if α is also a zero of $f'(x)$, the derivative of $f(x)$. [Hint: Apply Exercise 15 in Section 42 to the factorization $f(x) = (x - \alpha)^v g(x)$ of $f(x)$ in the ring $\bar{F}[x]$.]

13. Show from Exercise 12 that every irreducible polynomial over a field F of characteristic 0 is separable.

14. Show from Exercise 12 that an irreducible polynomial $q(x)$ over a field F of characteristic $p \neq 0$ is not separable if and only if each exponent of each term of $q(x)$ is divisible by p .

15. Generalize Exercise 12, showing that $f(x) \in F[x]$ has no zero of multiplicity > 1 if and only if $f(x)$ and $f'(x)$ have no common factor in $\bar{F}[x]$ of degree > 0 .

16. Working a bit harder than in Exercise 15, show that $f(x) \in F[x]$ has no zero of multiplicity > 1 if and only if $f(x)$ and $f'(x)$ have no common nonconstant factor in $F[x]$. [Hint: Use Theorem 35.9 to show that if 1 is a gcd of $f(x)$ and $f'(x)$ in $F[x]$, it is a gcd of these polynomials in $E[x]$ for E any splitting field of F , also.]

17. Describe a feasible computational procedure for determining whether $f(x) \in F[x]$ has a zero of multiplicity > 1 , without actually finding the zeros of $f(x)$. [Hint: Use Exercise 16.]

18. Let $F \leq E \leq K$ be field extensions with K a normal extension of F . By Corollary 45.18, $G(K/E)$ is a subgroup of $G(K/F)$. For two automorphisms $\sigma, \tau \in G(K/F)$, show that they are in the same left cosets of $G(K/E) \leq G(K/F)$ if and only if $\sigma(\alpha) = \tau(\alpha)$ for all $\alpha \in E$.

19. Prove that Definition 45.3 does not depend on which splitting field over F is used.

SECTION 46

GALOIS THEORY

The Galois Theorems

In this section we present the main theorems of Galois theory. These theorems provide precise statements regarding the correspondence between intermediate fields of a normal field extension and subgroups of the Galois group. But first we state key definitions related to the correspondence.

46.1 Definition Let K be a normal extension of F , E an intermediate field of the extension, and H a subgroup of $G(K/F)$. The set of all $\alpha \in K$ such that each element of H fixes α is an

intermediate field of the extension K over F , and it is called the **fixed field for H** . We write K_H to denote the fixed field for H .

We let $\lambda(E)$ be the set of all $\sigma \in G(K/F)$ that fix all the elements of E , that is, $\lambda(E) = G(K/E)$. We call $\lambda(E)$ the **group of E** .

If K is the splitting field of $f(x) \in F[x]$, then we say that $G(K/F)$ is the **group of the polynomial $f(x)$** . ■

46.2 Example Let K be the splitting field of $f(x) = x^3 - 2$ over \mathbb{Q} . As we saw in Example 45.19, K is a normal extension of \mathbb{Q} and

$$K = \mathbb{Q}(\sqrt[3]{2}, \sqrt{3}i).$$

The group of $f(x)$ is $G(K/\mathbb{Q})$. Also,

$$\lambda(\mathbb{Q}(\sqrt[3]{2})) = \{\sigma \in G(K/\mathbb{Q}) \mid \sigma(\sqrt[3]{2}) = \sqrt[3]{2}\} = G(K/\mathbb{Q}(\sqrt[3]{2})).$$

Both the identity, ι , and complex conjugation, $\sigma(a + bi) = a - bi$, fix $\sqrt[3]{2}$. Therefore,

$$\langle \sigma \rangle = \{\iota, \sigma\} \leq \lambda(\mathbb{Q}(\sqrt[3]{2}))$$

and

$$K_\sigma = \mathbb{Q}(\sqrt[3]{2}).$$

At this point, we can only say $\langle \sigma \rangle$ is a subgroup of $\lambda(\mathbb{Q}(\sqrt[3]{2}))$ since it is conceivable that there could be other automorphisms of $G(K/\mathbb{Q})$ that fix $\mathbb{Q}(\sqrt[3]{2})$. As we will soon see, this cannot be the case, and the two subgroups are equal. ▲

We now present a series of related theorems that together make up the essence of Galois Theory.

46.3 Theorem Let K be a normal extension of a field F and E an intermediate field. The fixed field for the set of all automorphisms of K that fix E is exactly E . That is,

$$E = K_{\lambda(E)}.$$

Proof Clearly $E \subseteq K_{\lambda(E)}$. We show that $K_{\lambda(E)} \subseteq E$. Let α be an element of K that is not in E . The minimal polynomial for α over E has degree at least two and therefore α has a conjugate $\beta \in K$, with $\beta \neq \alpha$, by Corollary 44.12 and the fact that K is a separable extension of F . Theorem 43.18, the Conjugation Isomorphism Theorem, says there is an isomorphism

$$\psi_{\alpha,\beta} : E(\alpha) \rightarrow E(\beta)$$

that maps α to β and fixes all the elements of E . The map $\psi_{\alpha,\beta}$ can be extended to an automorphism $\sigma : K \rightarrow K$ by the Isomorphism Extension Theorem, Theorem 44.6. Thus $\sigma \in \lambda(E)$, and σ does not fix α . We have shown that if $\alpha \notin E$, then $\lambda(E)$ does not fix α ; or equivalently, $K_{\lambda(E)} \subseteq E$, which completes the proof. ◆

46.4 Theorem Let K be a normal extension of a field F and E an intermediate field. The degree of the extension K over E is the order of the group $\lambda(E)$:

$$[K : E] = |\lambda(E)| = |G(K/E)|.$$

Furthermore, the number of left cosets of $\lambda(E)$ in $G(K/F)$ is the degree of the extension of E over F . That is,

$$(G(K/F) : \lambda(E)) = [E : F].$$

Proof Since $\lambda(E) = G(K/E)$, this theorem is simply a restatement of Corollary 45.18. ◆

46.5 Example Continuing Example 46.2, $K = \mathbb{Q}(\sqrt[3]{2}, \sqrt{3}i)$ is the splitting field of $x^3 - 2$ over \mathbb{Q} . We saw that if σ is complex conjugation, then $\langle \sigma \rangle \leq \lambda(\mathbb{Q}(\sqrt[3]{2}))$. The degree of the extension K over $\mathbb{Q}(\sqrt[3]{2})$ is two since $\alpha = \sqrt[3]{2}i \notin K$, but α is a zero of the degree 2 polynomial $x^2 + 3 \in \mathbb{Q}(\sqrt[3]{2})[x]$. By Theorem 46.4,

$$2 = [K : \mathbb{Q}(\sqrt[3]{2})] = |\lambda(\mathbb{Q}(\sqrt[3]{2}))|.$$

Since $\langle \sigma \rangle \leq \lambda(\mathbb{Q}(\sqrt[3]{2}))$ and both finite groups have the same number of elements,

$$\langle \sigma \rangle = \lambda(\mathbb{Q}(\sqrt[3]{2})). \quad \blacktriangle$$

46.6 Theorem Let K be a normal extension of a field F and H a subgroup of the Galois group $G(K/F)$. The subgroup of $G(K/F)$ that fixes all the elements fixed by K_H is exactly H . That is,

$$\lambda(K_H) = H.$$

Proof It is clear that H is a subgroup of $\lambda(K_H)$. We will verify that the two groups are equal by checking that they have the same number of elements. Let $k = |H|$.

By Theorem 45.13, the field extension K over F has a primitive element, α , so $K = F(\alpha)$. Let $E = K_H$, the subfield of K that is fixed by every element in H . Then $K = E(\alpha)$ and $[K : E] = \deg(\alpha, E)$. We let $n = [K : E]$. We next let

$$f(x) = \prod_{\sigma \in H} (x - \sigma(\alpha)) \in K[x].$$

The degree of f is $k = |H|$. Let $\tau \in H$, so τ is an isomorphism from K onto K . Since H is a group, multiplying all the elements of H by τ on the left simply permutes the elements of H . That is,

$$H = \{\sigma_1, \sigma_2, \dots, \sigma_k\} = \{\tau\sigma_1, \tau\sigma_2, \dots, \tau\sigma_k\}.$$

By Exercise 32 in Section 44, the map $\tau_x : K[x] \rightarrow K[x]$ is an isomorphism and

$$\tau_x(f(x)) = \prod_{\sigma \in H} (x - \tau\sigma(\alpha)) = \prod_{\sigma \in H} (x - \sigma(\alpha)) = f(x).$$

Writing $f(x) = a_0 + a_1x + a_2x^2 + \dots + a_kx^k$,

$$\tau_x(f(x)) = \tau(a_0) + \tau(a_1)x + \tau(a_2)x^2 + \dots + \tau(a_k)x^k.$$

Equating coefficients in $\tau_x(f(x)) = f(x)$, we see that for any $\tau \in H$, and for any i , $a_i = \tau(a_i)$. But the only elements of K that are fixed by every element in H are the elements of $E = K_H$, which implies that each a_i is in E . Therefore, $f(x) \in E[x]$. Since the identity map is in H , α is a zero of $f(x)$. Thus $\text{irr}(\alpha, E)$ divides $f(x)$ and

$$k = \deg(f(x)) \geq \deg(\text{irr}(\alpha, E)) = \deg(\alpha, E) = n.$$

Since H is a subgroup of $\lambda(K_H)$,

$$k = |H| \leq |\lambda(K_H)| = [K : E] = n.$$

Thus we have $k = n$ and $\lambda(K_H) = H$. ◆

Theorems 46.3 and 46.6 together imply that for normal extensions, the map λ , which maps the intermediate fields of the extension K over F to subgroups of $G(K/F)$, is both one-to-one and onto. Furthermore the inverse map λ^{-1} is simply the map that sends a subgroup $H \leq G(K/E)$ to the intermediate field $K_H = G(K/E)$.

46.7 Example Let $K = \mathbb{Q}(\sqrt[3]{2}, \sqrt{3}i)$ over \mathbb{Q} . In Example 45.19, we determined that $|G(K/\mathbb{Q})| = 6$. In this example, we take an alternate route to arrive at the same conclusion. In

Example 46.5, we saw that for $\sigma \in G(K/\mathbb{Q})$ given by complex conjugation, $\lambda(\mathbb{Q}(\sqrt[3]{2})) = \langle \sigma \rangle$. Therefore,

$$K_{\langle \sigma \rangle} = \mathbb{Q}(\sqrt[3]{2}).$$

Theorem 46.4 says that

$$(G(K/\mathbb{Q}) : \langle \sigma \rangle) = (G(K/\mathbb{Q}) : \lambda(\mathbb{Q}(\sqrt[3]{2}))) = [\mathbb{Q}(\sqrt[3]{2} : \mathbb{Q}] = 3.$$

Since $|\langle \sigma \rangle| = 2$,

$$|G(K/\mathbb{Q})| = (G(K/\mathbb{Q}) : \langle \sigma \rangle) \cdot |\langle \sigma \rangle| = 6. \quad \blacktriangle$$

46.8 Theorem Let K be a normal extension of a field F and E an intermediate field of the extension. Then E is a normal extension of F if and only if $\lambda(E)$ is a normal subgroup of $G(K/F)$. Furthermore, if E is a normal extension of F , then $G(E/F)$ is isomorphic with $G(K/F)/G(K/E)$.

Proof We first assume that E is a normal extension of F . Let $\tau \in \lambda(E) = G(K/E)$ and $\sigma \in G(K/F)$. We verify that $\sigma \tau \sigma^{-1} \in G(K/E)$ to show that $\lambda(E)$ is a normal subgroup of $G(K/F)$. By Theorem 44.11, σ^{-1} maps E isomorphically onto E . Therefore, for any $\alpha \in E$, $\tau(\sigma^{-1}(\alpha)) = \sigma^{-1}(\alpha)$ and

$$\sigma \tau \sigma^{-1}(\alpha) = \sigma(\tau(\sigma^{-1}(\alpha))) = \sigma(\sigma^{-1}(\alpha)) = \alpha.$$

Thus, $\sigma \tau \sigma^{-1} \in G(K/E)$, and $\lambda(E)$ is a normal subgroup of $G(K/F)$.

We next assume that $\lambda(E)$ is a normal subgroup of $G(K/F)$ and show that E is a normal extension of F . We must show that E is a splitting field over F and that E is a separable extension of F . Theorem 45.8 says that the extension is separable. We use Corollary 44.14 to show that E is a splitting field over F . Let σ_1 be any isomorphism fixing F and mapping E onto a subfield of K . Let σ be an extension of σ_1 to an isomorphism from K onto K . Such an extension exists by the Isomorphism Extension Theorem 44.6. Let $\alpha \in E$ and $\tau \in \lambda(E)$. By assumption $\sigma^{-1} \tau \sigma(\alpha) = \alpha$. Thus

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

Since every $\tau \in \lambda(E)$ fixes $\sigma(\alpha)$, $\sigma(\alpha) \in K_{\lambda(E)}$. By Theorem 46.3, $\sigma(\alpha) \in E$. Thus the isomorphism σ_1 maps E isomorphically onto E . By Corollary 44.14, E is a splitting field over F , and therefore, E is a normal extension of F .

Assuming that E is a normal extension of F , we know that any isomorphism of K that fixes F maps E isomorphically onto itself. Using this fact, we can define $\phi : G(K/F) \rightarrow G(E/F)$ by simply letting $\phi(\sigma)$ be the map σ restricted to E . By the Isomorphism Extension Theorem 44.6, ϕ maps onto $G(E/F)$. Also

$$\text{Ker}(\phi) = \{\sigma \in G(K/F) \mid \sigma(\alpha) = \alpha \text{ for all } \alpha \in E\} = G(K/E).$$

Therefore $G(E/F)$ is isomorphic with

$$G(K/F)/\text{Ker}(\phi) = G(K/F)/G(K/E). \quad \blacklozenge$$

46.9 Example Recall from Example 43.14, $x^3 - 2$ has only one zero in the field extension $\mathbb{Q}(\sqrt[3]{2})$ over \mathbb{Q} . Thus $\mathbb{Q}(\sqrt[3]{2})$ over \mathbb{Q} is not a splitting field, and therefore, not a normal extension. As we saw, the only automorphism of $G(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ is the identity map ι . So the set of automorphisms of $\mathbb{Q}(\sqrt[3]{2})$ fixing $\mathbb{Q}(\sqrt[3]{2})$ is the same as the set of automorphisms of $\mathbb{Q}(\sqrt[3]{2})$ fixing \mathbb{Q} . In the case of a normal extension, this could not happen since λ gives a one-to-one correspondence.

In Example 46.7, we saw that

$$|G(\mathbb{Q}(\sqrt[3]{2}, \sqrt{3}i)/\mathbb{Q})| = 6.$$

Up to isomorphism, there are only two groups of order 6, \mathbb{Z}_6 and S_3 . Since $\mathbb{Q}(\sqrt[3]{2})$ over \mathbb{Q} is not a normal extension of \mathbb{Q} , $\lambda(\mathbb{Q}(\sqrt[3]{2}))$ is not a normal subgroup of $G(\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3}i)/\mathbb{Q})$. Every subgroup of \mathbb{Z}_6 is a normal subgroup, so $G(\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3}i)/\mathbb{Q})$ is isomorphic with S_3 . \blacktriangle

The next theorem completes the basic facts that are used to apply Galois theory to particular problems. The theorem is usually interpreted in terms of subgroup and subfield diagrams. It says that the diagram for subfields is the same as the diagram for the corresponding subgroups, except that the diagrams are inverted.

46.10 Theorem Let K be a normal extension of F , with E_1 and E_2 intermediate fields. Then

$$E_1 \text{ is a subfield of } E_2 \text{ if and only if } \lambda(E_2) \text{ is a subgroup of } \lambda(E_1).$$

Proof Suppose that $E_1 \leq E_2$. Then any automorphism of K that fixes every element of E_2 clearly fixes every element of E_1 , which means that $\lambda(E_2) \leq \lambda(E_1)$.

Now suppose that $\lambda(E_2) \leq \lambda(E_1)$. Let $\alpha \in K_{\lambda(E_1)}$. Then α is fixed by every element of $\lambda(E_1)$ and, therefore, by every element of $\lambda(E_2)$. So $\alpha \in K_{\lambda(E_2)}$. Thus

$$E_1 = K_{\lambda(E_1)} \leq K_{\lambda(E_2)} = E_2. \quad \blacklozenge$$

46.11 Example As before, we let K be the splitting field for $f(x) = x^3 - 2$ over \mathbb{Q} . Each $\sigma \in G(K/\mathbb{Q})$ is completely determined by how σ permutes the zeros of $f(x)$. Since $f(x)$ has three zeros, $G(K/E)$ is isomorphic with a subgroup of the symmetric group S_3 . As we saw in Example 46.9, $G(K/\mathbb{Q})$ is isomorphic with S_3 . By labeling the zeros of $f(x)$ by

$$\begin{aligned} r_1 &= \sqrt[3]{2} \\ r_2 &= \frac{\sqrt[3]{2}}{2} (1 + \sqrt{3}i) \\ r_3 &= \frac{\sqrt[3]{2}}{2} (1 - \sqrt{3}i) \end{aligned}$$

the isomorphism $\phi : G(K/\mathbb{Q}) \rightarrow S_3$ can be defined by letting $\phi(\sigma)$ be the permutation of the zeros of $f(x)$ by the automorphism σ , that is, $\phi(\sigma)(i) = j$ if $\sigma(r_i) = r_j$. We will abuse notation slightly and identify elements in $G(K/\mathbb{Q})$ with their corresponding elements in the symmetric group S_3 .

The subgroup diagram for S_3 is displayed in Figure 46.12 (a). By Theorem 46.10, the subfield diagram for K is given in Figure 46.12 (b).

The automorphism corresponding to the transposition $(2, 3)$ maps r_2 to r_3 , maps r_3 to r_2 , and fixes r_1 , which is complex conjugation. As we saw in Example 46.7, $K_{\langle(2,3)\rangle} = \mathbb{Q}(\sqrt[3]{2}) = \mathbb{Q}(r_1)$. The automorphism corresponding to $(1, 2)$ fixes r_3 , maps r_1 to r_2 , and maps r_2 to r_1 . Thus $K_{\langle(1,2)\rangle} = \mathbb{Q}(r_3)$. Similarly, $K_{\langle(1,3)\rangle} = \mathbb{Q}(r_2)$.

We can determine $K_{\langle(1,2,3)\rangle}$ by noticing that

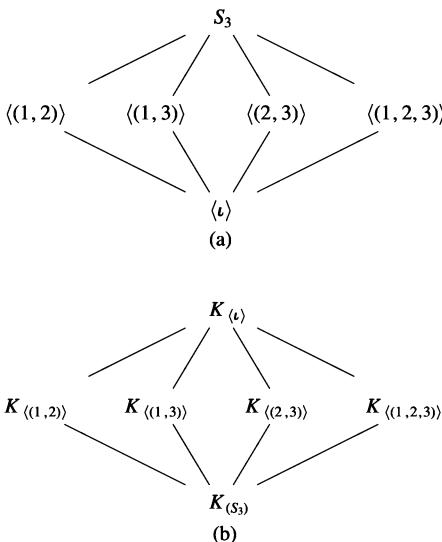
$$\begin{aligned} |\lambda(\mathbb{Q}(\sqrt[3]{3}i))| &= [K : \mathbb{Q}(\sqrt[3]{3}i)] \\ &= [\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3}i) : \mathbb{Q}(\sqrt[3]{3}i)] \\ &= [\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3}i) : \mathbb{Q}] / [\mathbb{Q}(\sqrt[3]{3}i) : \mathbb{Q}] \\ &= 3. \end{aligned}$$

The subfield diagram shows that there is only one intermediate field E with $[K : E] = 3$, so $K_{\langle(1,2,3)\rangle} = \mathbb{Q}(\sqrt[3]{3}i)$. Summarizing:

$$K_{\{1\}} = K$$

$$K_{S_3} = \mathbb{Q}$$

$$K_{\langle(2,3)\rangle} = \mathbb{Q}(r_1) = \mathbb{Q}(\sqrt[3]{2})$$



46.12 Figure

$$K_{\langle(1,2)\rangle} = \mathbb{Q}(r_3) = \mathbb{Q}\left(\frac{\sqrt[3]{2}}{2}(-1 - \sqrt{3}i)\right)$$

$$K_{\langle(1,3)\rangle} = \mathbb{Q}(r_2) = \mathbb{Q}\left(\frac{\sqrt[3]{2}}{2}(-1 + \sqrt{3}i)\right)$$

$$K_{\langle(1,2,3)\rangle} = \mathbb{Q}(\sqrt{3}i)$$

Normal subgroups of $G(K/\mathbb{Q})$ correspond to subfields of K that are normal extensions of \mathbb{Q} by Theorem 46.8. Thus the only intermediate fields of K that are normal extensions of \mathbb{Q} are K , \mathbb{Q} , and $\mathbb{Q}(\sqrt{3}i)$ corresponding to the normal subgroups of S_3 , namely $\{1\}$, S_3 , and $\langle(1,2,3)\rangle$, respectively. \blacktriangle

Not every subgroup diagram of a Galois group looks like its own inversion, as we will see in the next section.

■ EXERCISES 46

Computations

The field $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$ is a finite normal extension of \mathbb{Q} . It can be shown that $[K : \mathbb{Q}] = 8$. In Exercises 1 through 8, compute the indicated numerical quantity. The notation is that of Definition 46.1.

- | | |
|---|--|
| 1. $[K : \mathbb{Q}\sqrt{2}]$ | 2. $ G(K/\mathbb{Q}) $ |
| 3. $ \lambda(\mathbb{Q}) $ | 4. $ \lambda(\mathbb{Q}(\sqrt{2}, \sqrt{3})) $ |
| 5. $ \lambda(\mathbb{Q}(\sqrt{6})) $ | 6. $ \lambda(\mathbb{Q}(\sqrt{30})) $ |
| 7. $ \lambda(\mathbb{Q}(\sqrt{2} + \sqrt{6})) $ | 8. $ \lambda(K) $ |

9. Describe the group of the polynomial $(x^4 - 1) \in \mathbb{Q}[x]$ over \mathbb{Q} .
10. Let G be the group of the polynomial $x^3 + 2$ over \mathbb{Q} . Find the order of G and identify a well-known group that is isomorphic with G .
11. Let K be the splitting field of $x^3 - 5$ over \mathbb{Q} .
 - a. Show that $K = \mathbb{Q}(\sqrt[3]{5}, i\sqrt{3})$.
 - b. Describe the six elements of $G(K/\mathbb{Q})$ by giving their values on $\sqrt[3]{5}$ and $i\sqrt{3}$.
 - c. To what group have we seen before is $G(K/\mathbb{Q})$ isomorphic?
 - d. Using the notation given in the answer to part (b) in the back of the text, give the diagrams for the subfields of K and for the subgroups of $G(K/\mathbb{Q})$, indicating corresponding intermediate fields and subgroups, as we did in Example 46.11.
12. Describe the group of the polynomial $(x^4 - 5x^2 + 6) \in \mathbb{Q}[x]$ over \mathbb{Q} .
13. Describe the group of the polynomial $(x^3 - 1) \in \mathbb{Q}[x]$ over \mathbb{Q} .

Concepts

14. Give an example of two normal extensions K_1 and K_2 of the same field F such that K_1 and K_2 are not isomorphic fields but $G(K_1/F) \cong G(K_2/F)$.
15. Determine whether each of the following is true or false.
 - a. Two different subgroups of a Galois group may have the same fixed field.
 - b. If $F \leq E < L \leq K$ are field extensions and K is a normal extension of F , then $\lambda(E) < \lambda(L)$.
 - c. If K is a normal extension of F , then K is a normal extension of E , where $F \leq E \leq K$.
 - d. If two normal extensions E and L of a field F have isomorphic Galois groups, then $[E : F] = [L : F]$.
 - e. If E is a normal extension of F and H is a normal subgroup of $G(E/F)$, then E_H is a normal extension of F .
 - f. If E is any normal simple extension of a field F , then the Galois group $G(E/F)$ is a simple group.
 - g. No Galois group is simple.
 - h. If two intermediate fields E_1 and E_2 of a normal extension K over F have isomorphic groups, $\lambda(E_1)$ and $\lambda(E_2)$, then $K_{\lambda(E_1)}$ is isomorphic with $K_{\lambda(E_2)}$.
 - i. An extension E of degree 2 over a field F is always a normal extension of F .
 - j. An extension E of degree 2 over a field F is always a normal extension of F if the characteristic of F is zero.

Theory

16. A normal extension K of a field F is **abelian over F** if $G(K/F)$ is an abelian group. Show that if K is abelian over F and $F \leq E \leq K$, then K is abelian over E and E is abelian over F .
17. Let K be a normal extension of a field F . Prove that for every $\alpha \in K$, the **norm of α over F** , given by

$$N_{K/F}(\alpha) = \prod_{\sigma \in G(K/F)} \sigma(\alpha),$$

and the **trace of α over F** , given by

$$Tr_{K/F}(\alpha) = \sum_{\sigma \in G(K/F)} \sigma(\alpha),$$

are elements of F .

18. Consider $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Referring to Exercise 17, compute each of the following (see Table 43.5).
 - a. $N_{K/\mathbb{Q}}(\sqrt{2})$
 - b. $N_{K/\mathbb{Q}}(\sqrt{2} + \sqrt{3})$
 - c. $N_{K/\mathbb{Q}}(\sqrt{6})$
 - d. $N_{K/\mathbb{Q}}(2)$
 - e. $Tr_{K/\mathbb{Q}}(\sqrt{2})$
 - f. $Tr_{K/\mathbb{Q}}(\sqrt{2} + \sqrt{3})$
 - g. $Tr_{K/\mathbb{Q}}(\sqrt{6})$
 - h. $Tr_{K/\mathbb{Q}}(2)$

19. Let $K = F(\alpha)$ be a normal extension of F . Let

$$\text{irr}(\alpha, F) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0.$$

Referring to Exercise 17, show that

a. $N_{K/F}(\alpha) = (-1)^n a_0$,

b. $\text{Tr}_{K/F}(\alpha) = -a_{n-1}$.

20. Let $f(x) \in F[x]$ be a polynomial of degree n such that each irreducible factor is separable over F . Show that the order of the group of $f(x)$ over F divides $n!$.

21. Let $f(x) \in F[x]$ be a polynomial such that every irreducible factor of $f(x)$ is a separable polynomial over F . Show that the group of $f(x)$ over F can be viewed in a natural way as a group of permutations of the zeros of $f(x)$ in \bar{F} .

22. Let F be a field and let ζ be a primitive n th root of unity in \bar{F} , where the characteristic of F is 0.

- a. Show that $F(\zeta)$ is a normal extension of F .

- b. Show that $G(F(\zeta)/F)$ is abelian. [Hint: Every $\sigma \in G(F(\zeta)/F)$ maps ζ onto some ζ^r and is completely determined by this value r .]

23. A normal extension K of a field F is **cyclic over F** if $G(K/F)$ is a cyclic group.

- a. Show that if K is cyclic over F and $F \leq E \leq K$, then E is cyclic over F and K is cyclic over E .

- b. Show that if K is cyclic over F , then there exists exactly one field E , $F \leq E \leq K$, of degree d over F for each divisor d of $[K : F]$.

24. Let K be a normal extension of F .

- a. For $\alpha \in K$, show that

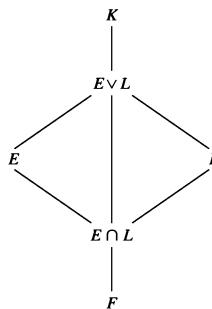
$$f(x) = \prod_{\sigma \in G(K/F)} (x - \sigma(\alpha))$$

is in $F[x]$.

- b. Referring to part (a), show that $f(x)$ is a power of $\text{irr}(\alpha, F)$, and $f(x) = \text{irr}(\alpha, F)$ if and only if $K = F(\alpha)$.

25. Let K be a normal extension of the field F . If E and L are both intermediate fields of the extension, the **join**, $E \vee L$, is the intersection of all intermediate fields of the extension that contain both E and L . See Figure 46.13. Describe $G(K/(E \vee L))$ in terms of $G(K/E)$ and $G(K/L)$.

26. With reference to the situation in Exercise 25, describe $G(K/(E \cap L))$ in terms of $G(K/E)$ and $G(K/L)$.



46.13 Figure

SECTION 47

ILLUSTRATIONS OF GALOIS THEORY

Symmetric Functions

Let F be a field of characteristic zero, and let y_1, \dots, y_n be indeterminates. There are some natural automorphisms of $F(y_1, \dots, y_n)$ leaving F fixed, namely, those defined by permutations of $\{y_1, \dots, y_n\}$. To be more explicit, let σ be a permutation of $\{1, \dots, n\}$, that is, $\sigma \in S_n$. Then σ gives rise to a natural map $\bar{\sigma} : F(y_1, \dots, y_n) \rightarrow F(y_1, \dots, y_n)$ given by

$$\bar{\sigma}\left(\frac{f(y_1, \dots, y_n)}{g(y_1, \dots, y_n)}\right) = \frac{f(y_{\sigma(1)}, \dots, y_{\sigma(n)})}{g(y_{\sigma(1)}, \dots, y_{\sigma(n)})}$$

for $f(y_1, \dots, y_n), g(y_1, \dots, y_n) \in F[y_1, \dots, y_n]$, with $g(y_1, \dots, y_n) \neq 0$. It is immediate that $\bar{\sigma}$ is an automorphism of $F(y_1, \dots, y_n)$ leaving F fixed. The elements of $F(y_1, \dots, y_n)$ left fixed by all $\bar{\sigma}$, for all $\sigma \in S_n$, are those rational functions that are *symmetric* in the indeterminates y_1, \dots, y_n .

47.1 Definition An element of the field $F(y_1, \dots, y_n)$ is a **symmetric function in y_1, \dots, y_n over F** , if it is fixed by all permutations of y_1, \dots, y_n , in the sense just explained. ■

Let $\overline{S_n}$ be the group of all the automorphisms $\bar{\sigma}$ for $\sigma \in S_n$. Observe that $\overline{S_n}$ is naturally isomorphic to S_n . Let K be the subfield of $F(y_1, \dots, y_n)$, which is the fixed field of $\overline{S_n}$. Consider the polynomial

$$f(x) = \prod_{i=1}^n (x - y_i);$$

this polynomial $f(x) \in (F(y_1, \dots, y_n))[x]$ is a **general polynomial of degree n** . Let $\bar{\sigma}_x$ be the polynomial extension of $\bar{\sigma}$, as defined in Definition 44.4, to $(F(y_1, \dots, y_n))[x]$, where $\bar{\sigma}_x(x) = x$. Now $f(x)$ is fixed by each map $\bar{\sigma}_x$ for $\sigma \in S_n$; that is,

$$\prod_{i=1}^n (x - y_i) = \prod_{i=1}^n (x - y_{\sigma(i)}).$$

Thus the coefficients of $f(x)$ are in K ; they are, except for sign, the *elementary symmetric functions* in y_1, \dots, y_n . As illustration, note that the constant term of $f(x)$ is

$$(-1)^n y_1 y_2 \cdots y_n,$$

the coefficient of x^{n-1} is $-(y_1 + y_2 + \cdots + y_n)$, and so on. These are symmetric functions in y_1, \dots, y_n .

The first elementary symmetric function in y_1, \dots, y_n is

$$s_1 = y_1 + y_2 + \cdots + y_n,$$

the second is $s_2 = y_1 y_2 + y_1 y_3 + \cdots + y_{n-1} y_n$, and so on, and the n th is $s_n = y_1 y_2 \cdots y_n$.

Consider the field $E = F(s_1, \dots, s_n)$. Of course, $E \leq K$, where K is the field of all symmetric functions in y_1, \dots, y_n over F . Since the characteristic of E is zero, the extension K over E is a separable extension. Thus $F(y_1, \dots, y_n)$ is a finite normal extension of E , namely, the splitting field of

$$f(x) = \prod_{i=1}^n (x - y_i)$$

over E . Since the degree of $f(x)$ is n , we have at once that

$$[F(y_1, \dots, y_n) : E] \leq n!$$

(see Exercise 19, Section 44). However, since K is the fixed field of $\overline{S_n}$ and

$$|\overline{S_n}| = |S_n| = n!,$$

we have also

$$n! = [F(y_1, \dots, y_n) : K].$$

Therefore,

$$n! = [F(y_1, \dots, y_n) : K] \leq [F(y_1, \dots, y_n) : E] \leq n!,$$

so

$$K = E.$$

The full Galois group of $F(y_1, \dots, y_n)$ over E is therefore $\overline{S_n}$. The fact that $K = E$ shows that every symmetric function can be expressed as a rational function of the elementary symmetric functions s_1, \dots, s_n . We summarize these results in a theorem.

47.2 Theorem Let F be a field with characteristic zero. Let s_1, \dots, s_n be the elementary symmetric functions in the indeterminates y_1, \dots, y_n . Then every symmetric function of y_1, \dots, y_n over F is a rational function of the elementary symmetric functions. Also, $F(y_1, \dots, y_n)$ is a finite normal extension of degree $n!$ of $F(s_1, \dots, s_n)$, and the Galois group of this extension is naturally isomorphic to S_n .

In view of Cayley's Theorem 8.11, it can be deduced from Theorem 47.2 that any finite group can occur as a Galois group (up to isomorphism). (See Exercise 11.)

The proof of Theorem 47.2 only uses the fact that the characteristic of F is zero to conclude that the extension $F(y_1, y_2, \dots, y_n)$ over E is a separable extension. With a bit more work, the proof can be modified to allow F to be an arbitrary field.

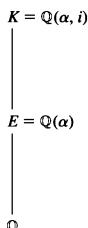
Examples

Let us give our promised example of a finite normal extension having a Galois group whose subgroup diagram does not look like its own inversion.

47.3 Example Consider the splitting field in \mathbb{C} of $x^4 - 2$ over \mathbb{Q} . Now $x^4 - 2$ is irreducible over \mathbb{Q} , by Eisenstein's criterion, with $p = 2$. Let $\alpha = \sqrt[4]{2}$ be the real positive zero of $x^4 - 2$. Then the four zeros of $x^4 - 2$ in \mathbb{C} are $\alpha, -\alpha, i\alpha$, and $-i\alpha$, where i is the usual zero of $x^2 + 1$ in \mathbb{C} . The splitting field K of $x^4 - 2$ over \mathbb{Q} thus contains $(i\alpha)/\alpha = i$. Since α is a real number, $\mathbb{Q}(\alpha) < \mathbb{R}$, so $\mathbb{Q}(\alpha) \neq K$. However, since $\mathbb{Q}(\alpha, i)$ contains all zeros of $x^4 - 2$, we see that $\mathbb{Q}(\alpha, i) = K$. Letting $E = \mathbb{Q}(\alpha)$, we have the diagram in Fig. 47.4.

Now $\{1, \alpha, \alpha^2, \alpha^3\}$ is a basis for E over \mathbb{Q} , and $\{1, i\}$ is a basis for K over E . Thus

$$\{1, \alpha, \alpha^2, \alpha^3, i, i\alpha, i\alpha^2, i\alpha^3\}$$



47.4 Figure $\rho \in G(K/\mathbb{Q})$ be the automorphism with $\rho(\alpha) = i\alpha$ and $\rho(i) = i$; and we let $\mu \in G(K/\mathbb{Q})$ be the automorphism with $\mu(\alpha) = \alpha$ and $\mu(i) = -i$. We have

$$\begin{aligned}\rho^2(\alpha) &= \rho(\rho(\alpha)) = \rho(i\alpha) = i(i\alpha) = -\alpha, \\ \rho^2(i) &= \rho(\rho(i)) = \rho(i) = i.\end{aligned}$$

47.5 Table

	ι	ρ	ρ^2	ρ^3	μ	$\mu\rho$	$\mu\rho^2$	$\mu\rho^3$
$\alpha \rightarrow$	α	$i\alpha$	$-\alpha$	$-i\alpha$	α	$-i\alpha$	$-\alpha$	$i\alpha$
$i \rightarrow$	i	i	i	i	$-i$	$-i$	$-i$	$-i$

Similarly, we have

$$\begin{aligned}\mu\rho(\alpha) &= \mu(\rho(\alpha)) = \mu(i\alpha) = -i\alpha, \\ \mu\rho(i) &= \mu(\rho(i)) = \mu(i) = -i.\end{aligned}$$

Table 47.5 shows the results of similar computations for $\iota, \rho, \rho^2, \rho^3, \mu, \mu\rho, \mu\rho^2$, and $\mu\rho^3$. These automorphisms account for all eight of the elements of $G(K/\mathbb{Q})$. The table looks remarkably like the dihedral group D_4 . To verify that $G(K/\mathbb{Q})$ is isomorphic with D_4 we check the relations $\mu^2 = \iota$, $\rho^4 = \iota$, and $\rho\mu = \mu\rho^3$ by evaluating each at α and i .

$$\begin{aligned}\mu^2(\alpha) &= \mu(\alpha) = \alpha, \\ \mu^2(i) &= \mu(-i) = i,\end{aligned}$$

$$\begin{aligned}\rho^4(\alpha) &= \rho(\rho^3(\alpha)) = \rho(-i\alpha) = -i^2\alpha = \alpha, \\ \rho^4(i) &= \rho(\rho^3(i)) = \rho(i) = i,\end{aligned}$$

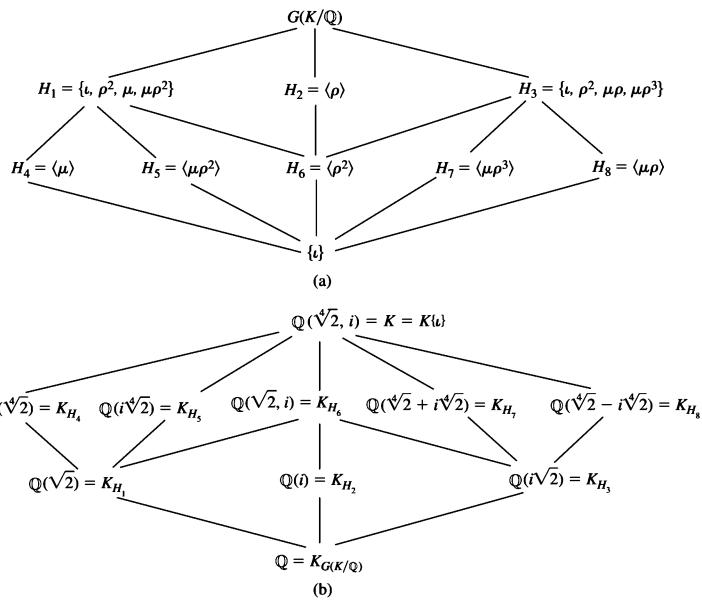
$$\begin{aligned}\rho\mu(\alpha) &= \rho(\alpha) = i\alpha = \mu\rho^3(\alpha), \\ \rho\mu(i) &= \rho(-i) = -i = \mu\rho^3(i).\end{aligned}$$

The subgroup diagram for the dihedral group is given in Figure 47.6 (a) with the corresponding subfield diagram in Figure 47.6 (b). This provides a good illustration of how one diagram is the inversion of the other.

The determination of the fixed fields K_{H_i} sometimes requires a bit of ingenuity. Let's illustrate. To find K_{H_2} , we merely have to find an extension of \mathbb{Q} of degree 2 fixed by $\{\iota, \rho, \rho^2, \rho^3\}$. Since all ρ^j leave i fixed, $\mathbb{Q}(i)$ is the field we are after. To find K_{H_4} , we have to find an extension of \mathbb{Q} of degree 4 fixed by ι and μ . Since μ leaves α fixed and α is a zero of $\text{irr}(\alpha, \mathbb{Q}) = x^4 - 2$, we see that $\mathbb{Q}(\alpha)$ is of degree 4 over \mathbb{Q} and is fixed by $\{\iota, \mu\}$. By *Galois theory, it is the only such field*. Here we are using strongly the one-to-one correspondence given by the Galois theory. If we find one field that fits the bill, it is the one we are after. Finding K_{H_7} requires more ingenuity. Since H_7 is a group, for any $\beta \in K$, $\iota(\beta) + \mu\rho^3(\beta)$ is fixed by ι and $\mu\rho^3$, the elements of H_7 . Letting $\beta = \alpha$ we see that $\iota(\alpha) + \mu\rho^3(\alpha) = \alpha + i\alpha$ is fixed by H_7 . By checking all eight automorphisms in Table 47.5, we see that only ι and $\mu\rho^3$ fix $\alpha + i\alpha$. Thus by the one-to-one correspondence, we must have

$$\mathbb{Q}(\alpha + i\alpha) = \mathbb{Q}(\sqrt[4]{2} + i\sqrt[4]{2}) = K_{H_7}.$$

Suppose we wish to find $\text{irr}(\alpha + i\alpha, \mathbb{Q})$. If $\gamma = \alpha + i\alpha$, then for every conjugate of γ over \mathbb{Q} , there exists an automorphism of K mapping γ into that conjugate. Thus we need only compute the various different values $\sigma(\gamma)$ for $\sigma \in G(K/\mathbb{Q})$ to find the other zeros of $\text{irr}(\gamma, \mathbb{Q})$. Every element in D_4 can be written in the form $\rho^i(\mu\rho^3)^j$ where $0 \leq i \leq 3$ and j is either 0 or 1. But $\mu\rho^3(\alpha + i\alpha) = \alpha + i\alpha$, so to compute the conjugates of $\alpha + i\alpha$ we only need to compute $\iota(\alpha + i\alpha)$, $\rho(\alpha + i\alpha)$, $\rho^2(\alpha + i\alpha)$, and $\rho^3(\alpha + i\alpha)$.



47.6 Figure (a) Group diagram. (b) Field diagram.

The conjugates of $\gamma = \alpha + i\alpha$ are thus $\alpha + i\alpha, i\alpha - \alpha, -\alpha - i\alpha$, and $-i\alpha + \alpha$. Hence

$$\begin{aligned}\text{irr}(\gamma, \mathbb{Q}) &= [(x - (\alpha + i\alpha))(x - (i\alpha - \alpha))] \\ &\quad \cdot [(x - (-\alpha - i\alpha))(x - (-i\alpha + \alpha))] \\ &= (x^2 - 2i\alpha x - 2\alpha^2)(x^2 + 2i\alpha x - 2\alpha^2) \\ &= x^4 + 4\alpha^4 = x^4 + 8.\end{aligned}$$

▲

We have seen examples in which the splitting field of a quartic (4th degree) polynomial over a field F is an extension of F of degree 8 (Example 47.3) and of degree 24 (Theorem 47.2, with $n = 4$). The degree of an extension of a field F that is a splitting field of a quartic over F must always divide $4! = 24$. The splitting field of $(x - 2)^4$ over \mathbb{Q} is \mathbb{Q} , an extension of degree 1, and the splitting field of $(x^2 - 2)^2$ over \mathbb{Q} is $\mathbb{Q}(\sqrt{2})$, an extension of degree 2. Our last example will give an extension of degree 4 for the splitting field of a quartic.

47.7 Example Consider the splitting field of $x^4 + 1$ over \mathbb{Q} . By Theorem 28.12, we can show that $x^4 + 1$ is irreducible over \mathbb{Q} , by arguing that it does not factor in $\mathbb{Z}[x]$. (See Exercise 1.) The work on complex numbers in Section 3 shows that the zeros of $x^4 + 1$ are $(1 \pm i)/\sqrt{2}$ and $(-1 \pm i)/\sqrt{2}$. A computation shows that if

$$\alpha = \frac{1+i}{\sqrt{2}},$$

then

$$\alpha^3 = \frac{-1+i}{\sqrt{2}}, \quad \alpha^5 = \frac{-1-i}{\sqrt{2}}, \quad \text{and} \quad \alpha^7 = \frac{1-i}{\sqrt{2}}.$$

Thus the splitting field K of $x^4 + 1$ over \mathbb{Q} is $\mathbb{Q}(\alpha)$, and $[K : \mathbb{Q}] = 4$. Let us compute $G(K/\mathbb{Q})$ and give the group and field diagrams. Since there exist automorphisms of K mapping α onto each conjugate of α , and since an automorphism σ of $\mathbb{Q}(\alpha)$ is completely determined by $\sigma(\alpha)$, we see that the four elements of $G(K/\mathbb{Q})$ are defined by Table 47.8.

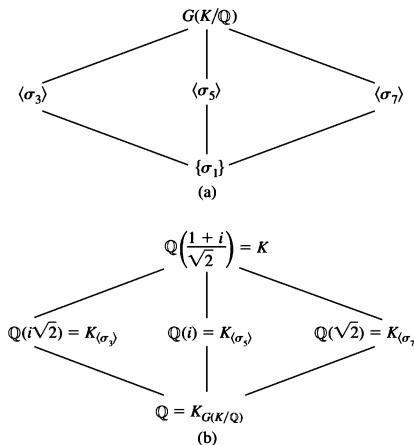
47.8 Table

	σ_1	σ_3	σ_5	σ_7
$\alpha \rightarrow$	α	α^3	α^5	α^7

Since

$$(\sigma_j \sigma_k)(\alpha) = \sigma_j(\alpha^k) = (\alpha^j)^k = \alpha^{jk}$$

and $\alpha^8 = 1$, we see that $G(K/\mathbb{Q})$ is isomorphic to the group $\{1, 3, 5, 7\}$ under multiplication modulo 8. There are only two groups of order 4 up to isomorphism, the cyclic group and the Klein 4-group. Since each element of $\{1, 3, 5, 7\}$ has order 2 or 1, $G(K/\mathbb{Q})$ is isomorphic with the Klein 4-group. The diagrams are given in Fig. 47.9.



47.9 Figure (a) Group diagram. (b) Field diagram.

To find $K_{\langle \sigma_3 \rangle}$, it is only necessary to find an element of K not in \mathbb{Q} fixed by $\{\sigma_1, \sigma_3\}$, since $[K_{\langle \sigma_3 \rangle} : \mathbb{Q}] = 2$. Clearly $\sigma_1(\alpha) + \sigma_3(\alpha)$ is fixed by both σ_1 and σ_3 , since $\{\sigma_1, \sigma_3\} = \langle \sigma_3 \rangle$ is a group. We have

$$\sigma_1(\alpha) + \sigma_3(\alpha) = \alpha + \alpha^3 = i\sqrt{2}.$$

Similarly,

$$\sigma_1(\alpha) + \sigma_7(\alpha) = \alpha + \alpha^7 = \sqrt{2}$$

is fixed by $\langle \sigma_7 \rangle = \{\sigma_1, \sigma_7\}$. This technique is of no use in finding $E_{\langle \sigma_5 \rangle}$, for

$$\sigma_1(\alpha) + \sigma_5(\alpha) = \alpha + \alpha^5 = 0,$$

and $0 \in \mathbb{Q}$. But by a similar argument, $\sigma_1(\alpha)\sigma_5(\alpha)$ is fixed by both σ_1 and σ_5 , and

$$\sigma_1(\alpha)\sigma_5(\alpha) = \alpha\alpha^5 = -i.$$

Thus $\mathbb{Q}(-i) = \mathbb{Q}(i)$ is the field we are after. ▲

■ EXERCISES 47

Computations (requiring more than the usual amount of theory)

1. Show that $x^4 + 1$ is irreducible in $\mathbb{Q}[x]$, as we asserted in Example 47.7.
2. Verify that the intermediate fields given in the field diagram in Fig. 47.6 are correct. (Some are verified in the text. Verify the rest.)
3. For each field in the field diagram in Fig. 47.6, find a primitive element generating the field over \mathbb{Q} (see Theorem 45.13) and give its irreducible polynomial over \mathbb{Q} .
4. Let ζ be a primitive 5th root of unity in \mathbb{C} .
 - a. Show that $\mathbb{Q}(\zeta)$ is the splitting field of $x^5 - 1$ over \mathbb{Q} .
 - b. Show that every automorphism of $K = \mathbb{Q}(\zeta)$ maps ζ onto some power ζ^r of ζ .
 - c. Using part (b), describe the elements of $G(K/\mathbb{Q})$.
 - d. Give the group and field diagrams for $\mathbb{Q}(\zeta)$ over \mathbb{Q} , computing the intermediate fields as we did in Examples 47.3 and 47.7.
5. Describe the group of the polynomial $(x^5 - 2) \in (\mathbb{Q}(\zeta))[x]$ over $\mathbb{Q}(\zeta)$, where ζ is a primitive 5th root of unity.
6. Recpat Exercise 4 for ζ a primitive 7th root of unity in \mathbb{C} .
7. In the easiest way possible, describe the group of the polynomial

$$(x^8 - 1) \in \mathbb{Q}[x]$$
 over \mathbb{Q} .
8. Find the splitting field K in \mathbb{C} of the polynomial $(x^4 - 4x^2 - 1) \in \mathbb{Q}[x]$. Compute the group of the polynomial over \mathbb{Q} and exhibit the correspondence between the subgroups of $G(K/\mathbb{Q})$ and the intermediate fields. In other words, do the complete job.
9. Express each of the following symmetric functions in y_1, y_2, y_3 over \mathbb{Q} as a rational function of the elementary symmetric functions s_1, s_2, s_3 .
 - a. $y_1^2 + y_2^2 + y_3^2$
 - b. $\frac{y_1}{y_2} + \frac{y_2}{y_1} + \frac{y_1}{y_3} + \frac{y_3}{y_1} + \frac{y_2}{y_3} + \frac{y_3}{y_2}$
10. Let $\alpha_1, \alpha_2, \alpha_3$ be the zeros in \mathbb{C} of the polynomial

$$(x^3 - 4x^2 + 6x - 2) \in \mathbb{Q}[x].$$

Find the polynomial having as zeros precisely the following:

- a. $\alpha_1 + \alpha_2 + \alpha_3$
- b. $\alpha_1^2, \alpha_2^2, \alpha_3^2$

Theory

11. Show that every finite group is isomorphic to some Galois group $G(K/F)$ for some finite normal extension K of some field F .

12. Let $f(x) \in F[x]$ be a monic polynomial of degree n having all its irreducible factors separable over F . Let K be the splitting field of $f(x)$ over F , and suppose that $f(x)$ factors in $K[x]$ into

$$\prod_{i=1}^n (x - \alpha_i).$$

Let

$$\Delta(f) = \prod_{i < j} (\alpha_i - \alpha_j);$$

the product $(\Delta(f))^2$ is the **discriminant** of $f(x)$.

- a. Show that $\Delta(f) = 0$ if and only if $f(x)$ has as a factor the square of some irreducible polynomial in $F[x]$.
 - b. Show that $(\Delta(f))^2 \in F$.
 - c. $G(K/F)$ may be viewed as a subgroup of $\overline{S_n}$, where $\overline{S_n}$ is the group of all permutations of $\{\alpha_i \mid i = 1, \dots, n\}$. Show that $G(K/F)$, when viewed in this fashion, is a subgroup of $\overline{A_n}$, the group formed by all even permutations of $\{\alpha_i \mid i = 1, \dots, n\}$, if and only if $\Delta(f) \in F$.
13. An element of \mathbb{C} is an **algebraic integer** if it is a zero of some *monic* polynomial in $\mathbb{Z}[x]$. Show that the set of all algebraic integers forms a subring of \mathbb{C} .

SECTION 48 CYCLOTOMIC EXTENSIONS

The Galois Group of a Cyclotomic Extension

In this section we consider subfields of the complex numbers obtained by adjoining roots of unity to the rational numbers, \mathbb{Q} . We apply Galois theory to these extensions to determine which regular n -gons are constructible.

- 48.1 Definition** The splitting field of $x^n - 1$ over a field F is the **n th cyclotomic extension of F** . ■

Since fields of characteristic zero are perfect, the splitting field, K , of $f(x) = x^n - 1$ over \mathbb{Q} is separable and thus a normal extension of \mathbb{Q} . The distinct n zeros of $f(x)$ form a cyclic group U_n . (See Section 3.) We saw in Corollary 6.17 that the number of generators of a cyclic group of order n is the number of positive integers less than n that are relatively prime to n , which we defined to be the Euler phi-function $\varphi(n)$. These $\varphi(n)$ generators are exactly the primitive n th roots of unity.

- 48.2 Definition** The polynomial

$$\Phi_n(x) = \prod_{i=1}^{\varphi(n)} (x - \alpha_i)$$

where the α_i are the primitive n th roots of unity in \mathbb{Q} , is the **n th cyclotomic polynomial over \mathbb{Q}** . ■

Let $\sigma \in G(K/\mathbb{Q})$. Since an automorphism of the Galois group $G(K/\mathbb{Q})$ must permute the primitive n th roots of unity, we see that $\sigma_x(\Phi_n(x)) = \Phi_n(x)$, where $\sigma_x : K[x] \rightarrow K[x]$ is the polynomial extension of σ . Thus the coefficients of Φ_n are fixed by every $\sigma \in G(K/\mathbb{Q})$, and therefore, all the coefficients of $\Phi_n(x)$ are rational numbers. That is, $\Phi_n(x) \in \mathbb{Q}[x]$. The n th cyclotomic polynomial $\Phi_n(x)$ must divide $x^n - 1$, so $\Phi_n(x) \in \mathbb{Z}[x]$ by Theorem 28.12. For p a prime number, Corollary 28.18 says that $\Phi_p(x)$ is irreducible over \mathbb{Q} . Although we do not prove it here, it can be shown that $\Phi_n(x)$ is irreducible even when $n \geq 2$ is not a prime number.

■ HISTORICAL NOTE

Carl Gauss considered cyclotomic polynomials in the final chapter of his *Disquisitiones Arithmeticae* of 1801. In that chapter, he gave a constructive procedure for actually determining the roots of $\Phi_p(x)$ in the case where p is prime. Gauss's method, which became an important example for Galois in the development of the general theory, was to solve a series of auxiliary equations, each of degree a prime factor of $p - 1$, with the coefficients of each in turn being determined by the roots of the previous equation. Gauss, of course, knew that the roots of $\Phi_p(x)$ were all powers of one of them, say ζ . He determined the auxiliary equations by taking certain sets of sums of the roots ζ^j , which were the desired roots of these equations. For example, in the case where $p = 19$ (and $p - 1 = 18 = 3 \times 3 \times 2$), Gauss needed to find two equations of degree 3 and one of degree 2 as his auxiliaries.

It turned out that the first one had the three roots, $\alpha_1 = \zeta + \zeta^8 + \zeta^7 + \zeta^{18} + \zeta^{11} + \zeta^{12}$, $\alpha_2 = \zeta^2 + \zeta^{16} + \zeta^{14} + \zeta^{17} + \zeta^3 + \zeta^5$, and $\alpha_3 = \zeta^4 + \zeta^{13} + \zeta^9 + \zeta^{15} + \zeta^6 + \zeta^{10}$. In fact, these three values are the roots of the cubic equation $x^3 + x^2 - 6x - 7$. Gauss then found a second cubic equation, with coefficients involving the α 's, whose roots were sums of two of the powers of ζ , and finally a quadratic equation, whose coefficients involved the roots of the previous equation, which had ζ as one of its roots. Gauss then asserted (without a complete proof) that each auxiliary equation can in turn be reduced to an equation of the form $x^m - A$, which clearly can be solved by radicals. That is, he showed that the solvability of the Galois group in this case, the cyclic group of order $p - 1$, implied that the cyclotomic equation was solvable in terms of radicals. (See Section 49.)

Let i be the usual complex zero of $x^2 + 1$. Our work with complex numbers in Section 3 shows that

$$\left(\cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n} \right)^n = \cos 2\pi + i \sin 2\pi = 1,$$

so $\cos(2\pi/n) + i \sin(2\pi/n)$ is an n th root of unity. The least integer m such that $(\cos(2\pi/n) + i \sin(2\pi/n))^m = 1$ is n . Thus $\cos(2\pi/n) + i \sin(2\pi/n)$ is a primitive n th root of unity, a zero of

$$\Phi_n(x) \in \mathbb{Q}[x].$$

48.3 Example A primitive 8th root of unity in \mathbb{C} is

$$\begin{aligned} \zeta &= \cos \frac{2\pi}{8} + i \sin \frac{2\pi}{8} \\ &= \cos \frac{\pi}{4} + i \sin \frac{\pi}{4} \\ &= \frac{1}{\sqrt{2}} + i \frac{1}{\sqrt{2}} = \frac{1+i}{\sqrt{2}}. \end{aligned}$$

By the theory of cyclic groups, in particular by Corollary 6.17, all the primitive 8th roots of unity in \mathbb{Q} are ζ, ζ^3, ζ^5 , and ζ^7 , so

$$\Phi_8(x) = (x - \zeta)(x - \zeta^3)(x - \zeta^5)(x - \zeta^7).$$

We can compute, directly from this expression, $\Phi_8(x) = x^4 + 1$ (see Exercise 1). Compare this with Example 47.7. ▲

Let us assume, without proof, that $\Phi_n(x)$ is irreducible over \mathbb{Q} . Let

$$\zeta = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n},$$

so that ζ is a primitive n th root of unity. Note that ζ is a generator of the cyclic multiplicative group of order n consisting of *all* n th roots of unity. All the primitive n th roots of unity, that is, all the generators of this group, are of the form ζ^m for $1 \leq m < n$ and m relatively prime to n . The field $\mathbb{Q}(\zeta)$ is the whole splitting field of $x^n - 1$ over \mathbb{Q} . Let $K = \mathbb{Q}(\zeta)$. If ζ^m is another primitive n th root of unity, then since ζ and ζ^m are conjugate over \mathbb{Q} , there is an automorphism τ_m in $G(K/\mathbb{Q})$ mapping ζ onto ζ^m . Let τ_r be the similar automorphism in $G(K/\mathbb{Q})$ corresponding to a primitive n th root of unity ζ^r . Then

$$(\tau_m \tau_r)(\zeta) = \tau_m(\zeta^r) = (\tau_m(\zeta))^r = (\zeta^m)^r = \zeta^{mr}.$$

This shows that composing automorphisms τ_m and τ_r corresponds to multiplying m and r modulo n . In other words, the Galois group $G(K/\mathbb{Q})$ is isomorphic with the group of units in \mathbb{Z}_n under multiplication, with the isomorphism mapping τ_m to m . The units in \mathbb{Z}_n are the numbers less than n that are relatively prime to n . Thus $G(K/\mathbb{Q})$ is an abelian group with $\varphi(n)$ elements.

Special cases of this material have appeared several times in the text and exercises. For example, α of Example 47.7 is a primitive 8th root of unity, and we made arguments in that example identical to those given here. We summarize these results in a theorem.

48.4 Theorem The Galois group of the n th cyclotomic extension of \mathbb{Q} has $\varphi(n)$ elements and is isomorphic to the group consisting of the positive integers less than n and relatively prime to n under multiplication modulo n .

48.5 Example Example 47.7 illustrates this theorem, for it is easy to see that the splitting field of $x^4 + 1$ is the same as the splitting field of $x^8 - 1$ over \mathbb{Q} . This follows from the fact that $\Phi_8(x) = x^4 + 1$ (see Example 48.3 and Exercise 1). \blacktriangle

48.6 Corollary The Galois group of the p th cyclotomic extension of \mathbb{Q} for a prime p is cyclic of order $p - 1$.

Proof By Theorem 48.4, the Galois group of the p th cyclotomic extension of \mathbb{Q} has $\varphi(p) = p - 1$ elements, and is isomorphic to the group of positive integers less than p and relatively prime to p under multiplication modulo p . This is exactly the multiplicative group (\mathbb{Z}_p^*, \cdot) of nonzero elements of the field \mathbb{Z}_p under field multiplication. By Corollary 28.7, this group is cyclic. \blacklozenge

Constructible Polygons

We conclude with an application determining which regular n -gons are constructible with a compass and a straightedge. We saw in Section 41 that the regular n -gon is constructible if and only if $\cos(2\pi/n)$ is a constructible real number. Now let

$$\zeta = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}.$$

Then

$$\frac{1}{\zeta} = \cos \frac{2\pi}{n} - i \sin \frac{2\pi}{n},$$

for

$$\left(\cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n} \right) \left(\cos \frac{2\pi}{n} - i \sin \frac{2\pi}{n} \right) = \cos^2 \frac{2\pi}{n} + \sin^2 \frac{2\pi}{n} = 1.$$

But then

$$\zeta + \frac{1}{\zeta} = 2 \cos \frac{2\pi}{n}.$$

Thus Corollary 41.8 shows that the regular n -gon is constructible only if $\zeta + 1/\zeta$ generates an extension of \mathbb{Q} of degree a power of 2.

If K is the splitting field of $x^n - 1$ over \mathbb{Q} , then $[K : \mathbb{Q}] = \varphi(n)$, by Theorem 48.4. If $\sigma \in G(K/\mathbb{Q})$ and $\sigma(\zeta) = \zeta^r$, then

$$\begin{aligned}\sigma\left(\zeta + \frac{1}{\zeta}\right) &= \zeta^r + \frac{1}{\zeta^r} \\ &= \left(\cos \frac{2\pi r}{n} + i \sin \frac{2\pi r}{n}\right) + \left(\cos \frac{2\pi r}{n} - i \sin \frac{2\pi r}{n}\right) \\ &= 2 \cos \frac{2\pi r}{n}.\end{aligned}$$

But for $1 < r < n$, we have $2 \cos(2\pi r/n) = 2 \cos(2\pi/n)$ only in the case that $r = n - 1$. Thus the only elements of $G(K/\mathbb{Q})$ carrying $\zeta + 1/\zeta$ onto itself are the identity automorphism and the automorphism τ , with $\tau(\zeta) = \zeta^{n-1} = 1/\zeta$. This shows that the subgroup of $G(K/\mathbb{Q})$ leaving $\mathbb{Q}(\zeta + 1/\zeta)$ fixed is of order 2, so by Galois theory,

$$\left[\mathbb{Q}\left(\zeta + \frac{1}{\zeta}\right) : \mathbb{Q}\right] = \frac{\varphi(n)}{2}.$$

Hence the regular n -gon is constructible only if $\varphi(n)/2$, and therefore also $\varphi(n)$, is a power of 2.

It can be shown by elementary arguments in number theory that if

$$n = 2^v p_1^{s_1} \cdots p_t^{s_t},$$

where the p_i are the distinct odd primes dividing n , then

$$\varphi(n) = 2^{v-1} p_1^{s_1-1} \cdots p_t^{s_t-1} (p_1 - 1) \cdots (p_t - 1). \quad (1)$$

If $\varphi(n)$ is to be a power of 2, then every odd prime dividing n must appear only to the first power and must be one more than a power of 2. Thus we must have each

$$p_i = 2^m + 1$$

for some m . Since -1 is a zero of $x^q + 1$ for q an odd prime, $x + 1$ divides $x^q + 1$ for q an odd prime. Thus, if $m = qu$, where q is an odd prime, then $2^m + 1 = (2^u)^q + 1$ is divisible by $2^u + 1$. Therefore, for $p_i = 2^m + 1$ to be prime, it must be that m is divisible by 2 only, so p_i has to have the form

$$p_i = 2^{(2^k)} + 1,$$

a Fermat prime. Fermat conjectured that these numbers $2^{(2^k)} + 1$ were prime for all nonnegative integers k . Euler showed that while $k = 0, 1, 2, 3$, and 4 give the primes 3, 5, 17, 257, and 65537, for $k = 5$, the integer $2^{(2^5)} + 1$ is divisible by 641. It has been shown that for $5 \leq k \leq 19$, all the numbers $2^{(2^k)} + 1$ are composite. The case $k = 20$ is still unsolved as far as we know. For at least 60 values of k greater than 20, including $k = 9448$, it has been shown that $2^{(2^k)} + 1$ is composite. It is unknown whether the number of Fermat primes is finite or infinite.

We have thus shown that the only regular n -gons that might be constructible are those where the odd primes dividing n are Fermat primes whose squares do not divide n . In particular, the only regular p -gons that might be constructible for p a prime greater than 2 are those where p is a Fermat prime.

- 48.7 Example** The regular 7-gon is not constructible, since 7 is not a Fermat prime. Similarly, the regular 18-gon is not constructible, for while 3 is a Fermat prime, its square divides 18. ▲

We now demonstrate that all these regular n -gons that are candidates for being constructible are indeed actually constructible. Let ζ again be the primitive n th root of unity $\cos(2\pi/n) + i\sin(2\pi/n)$. We saw above that

$$2 \cos \frac{2\pi}{n} = \zeta + \frac{1}{\zeta},$$

and that

$$\left[\mathbb{Q}\left(\zeta + \frac{1}{\zeta}\right) : \mathbb{Q} \right] = \frac{\varphi(n)}{2}.$$

Suppose now that $\varphi(n)$ is a power 2^s of 2. We saw above that $\mathbb{Q}(\zeta + 1/\zeta)$ is the subfield of $K = \mathbb{Q}(\zeta)$ fixed by $H_1 = \{\iota, \tau\}$, where ι is the identity element of $G(K/\mathbb{Q})$ and $\tau(\zeta) = 1/\zeta$. By Sylow theory, there exist additional subgroups H_j of order 2^j of $G(\mathbb{Q}(\zeta)/\mathbb{Q})$ for $j = 0, 2, 3, \dots, s$ such that

$$\{\iota\} = H_0 < H_1 < \dots < H_s = G(\mathbb{Q}(\zeta)/\mathbb{Q}).$$

By Galois theory,

$$\mathbb{Q} = K_{H_s} < K_{H_{s-1}} < \dots < K_{H_1} = \mathbb{Q}\left(\zeta + \frac{1}{\zeta}\right),$$

and $[K_{H_{j-1}} : K_H] = 2$. Note that $(\zeta + 1/\zeta) \in \mathbb{R}$, so $\mathbb{Q}(\zeta + 1/\zeta) \subset \mathbb{R}$. If $K_{H_{j-1}} = K_H(\alpha_j)$, then α_j is a zero of some $(a_jx^2 + b_jx + c_j) \in K_H[x]$. By the familiar “quadratic formula,” we have

$$K_{H_{j-1}} = K_H\left(\sqrt{b_j^2 - 4a_jc_j}\right).$$

Since we saw in Section 41 that construction of square roots of positive constructible numbers can be achieved by a straightedge and a compass, we see that every element in $\mathbb{Q}(\zeta + 1/\zeta)$, in particular $\cos(2\pi/n)$, is constructible. Hence the regular n -gons where $\varphi(n)$ is a power of 2 are constructible.

We summarize our work under this heading in a theorem.

- 48.8 Theorem** The regular n -gon is constructible with a compass and a straightedge if and only if all the odd primes dividing n are Fermat primes whose squares do not divide n .

- 48.9 Example** The regular 60-gon is constructible, since $60 = (2^2)(3)(5)$ and 3 and 5 are both Fermat primes. ▲

■ EXERCISES 48

Computations

- Referring to Example 48.3, complete the indicated computation, showing that $\Phi_8(x) = x^4 + 1$. [Suggestion: Compute the product in terms of ζ , and then use the fact that $\zeta^8 = 1$ and $\zeta^4 = -1$ to simplify the coefficients.]
- Classify the group of the polynomial $(x^{20} - 1) \in \mathbb{Q}[x]$ over \mathbb{Q} according to the Fundamental Theorem of Finitely Generated Abelian Groups. Theorem 9.12. [Hint: Use Theorem 48.4.]
- Using the formula for $\varphi(n)$ in terms of the factorization of n , as given in Eq. (1), compute the indicated value:

a. $\varphi(60)$	b. $\varphi(1000)$	c. $\varphi(8100)$
-------------------------	---------------------------	---------------------------
- Give the first 30 values of $n \geq 3$ for which the regular n -gon is constructible with a straightedge and a compass.

5. Find the smallest angle of integral degree, that is, $1^\circ, 2^\circ, 3^\circ$, and so on, constructible with a straightedge and a compass. [Hint: Constructing a 1° angle amounts to constructing the regular 360-gon, and so on.]
6. Let K be the splitting field of $x^{12} - 1$ over \mathbb{Q} .
- Find $[K : \mathbb{Q}]$.
 - Show that for $\sigma \in G(K/\mathbb{Q})$, σ^2 is the identity automorphism. Classify $G(K/\mathbb{Q})$ according to the Fundamental Theorem 9.12 of finitely generated abelian groups.

Concepts

7. Determine whether each of the following is true or false.
- $\Phi_n(x)$ is irreducible over every subfield of \mathbb{C} .
 - Every zero in \mathbb{C} of $\Phi_n(x)$ is a primitive n th root of unity.
 - The group of $\Phi_n(x) \in \mathbb{Q}[x]$ over \mathbb{Q} has order n .
 - The group of $\Phi_n(x) \in \mathbb{Q}[x]$ over \mathbb{Q} is abelian.
 - The Galois group of the splitting field of $\Phi_n(x)$ over \mathbb{Q} has order $\varphi(n)$.
 - The regular 25-gon is constructible with a straightedge and a compass.
 - The regular 17-gon is constructible with a straightedge and a compass.
 - For a prime p , the regular p -gon is constructible if and only if p is a Fermat prime.
 - All integers of the form $2^{(2^k)} + 1$ for nonnegative integers k are Fermat primes.
 - All Fermat primes are numbers of the form $2^{(2^k)} + 1$ for nonnegative integers k .

Theory

8. Show that

$$x^n - 1 = \prod_{d|n} \Phi_d(x)$$

- in $\mathbb{Q}[x]$, where the product is over all divisors d of n .
9. Find the cyclotomic polynomial $\Phi_n(x)$ over \mathbb{Q} for $n = 1, 2, 3, 4, 5$, and 6. [Hint: Use Exercise 8.]
10. Find $\Phi_{12}(x)$ in $\mathbb{Q}[x]$. [Hint: Use Exercises 8 and 9.]
11. Show that in $\mathbb{Q}[x]$, $\Phi_{2n}(x) = \Phi_n(-x)$ for odd integers $n > 1$. [Hint: If ζ is a primitive n th root of unity for n odd, what is the order of $-\zeta$?]
12. Let $n, m \in \mathbb{Z}^+$ be relatively prime. Show that the splitting field in \mathbb{C} of $x^{nm} - 1$ over \mathbb{Q} is the same as the splitting field in \mathbb{C} of $(x^n - 1)(x^m - 1)$ over \mathbb{Q} .
13. Let $n, m \in \mathbb{Z}^+$ be relatively prime. Show that the group of $(x^{nm} - 1) \in \mathbb{Q}[x]$ over \mathbb{Q} is isomorphic to the direct product of the groups of $(x^n - 1) \in \mathbb{Q}[x]$ and of $(x^m - 1) \in \mathbb{Q}[x]$ over \mathbb{Q} . [Hint: Using Galois theory, show that the groups of $x^m - 1$ and $x^n - 1$ can both be regarded as subgroups of the group of $x^{nm} - 1$. Then use Exercises 50 and 51 of Section 9.]

SECTION 49

INSOLVABILITY OF THE QUINTIC

The Problem

We are familiar with the fact that a quadratic polynomial $f(x) = ax^2 + bx + c, a \neq 0$, with real coefficients has $(-b \pm \sqrt{b^2 - 4ac})/2a$ as zeros in \mathbb{C} . Actually, this is true for $f(x) \in F[x]$, where F is any field of characteristic $\neq 2$ and the zeros are in \bar{F} . Exercise 4 asks us to show this. Thus, for example, $(x^2 + 2x + 3) \in \mathbb{Q}[x]$ has its zeros in $\mathbb{Q}(\sqrt{-2})$. You may wonder whether the zeros of a cubic polynomial over \mathbb{Q} can also always be expressed in terms of radicals. The answer is yes, and indeed, even the zeros of a polynomial of degree 4 over \mathbb{Q} can be expressed in terms of radicals. After mathematicians had tried for years to find the “radical formula” for zeros of a 5th degree polynomial, it was a triumph when Abel proved that a quintic need not be solvable by radicals. Our first job will be to describe precisely what this means. A large amount of the algebra we have developed is used in the forthcoming discussion.

Extensions by Radicals

49.1 Definition

An extension K of a field F is an **extension of F by radicals** if there are elements $\alpha_1, \dots, \alpha_r \in K$ and positive integers n_1, \dots, n_r such that $K = F(\alpha_1, \dots, \alpha_r)$, $\alpha_1^{n_1} \in F$ and $\alpha_i^{n_i} \in F(\alpha_1, \dots, \alpha_{i-1})$ for $1 < i \leq r$. A polynomial $f(x) \in F[x]$ is **solvable by radicals over F** if the splitting field E of $f(x)$ over F is contained in an extension of F by radicals. ■

A polynomial $f(x) \in F[x]$ is thus solvable by radicals over F if we can obtain every zero of $f(x)$ by using a finite sequence of the operations of addition, subtraction, multiplication, division, and taking n th roots, starting with elements of F . Now to say that the quintic is not solvable in the classic case, that is, characteristic 0, is not to say that no quintic is solvable, as the following example shows.

49.2 Example

The polynomial $x^5 - 1$ is solvable by radicals over \mathbb{Q} . The splitting field K of $x^5 - 1$ is generated over \mathbb{Q} by a primitive 5th root ζ of unity. Then $\zeta^5 = 1$, and $K = \mathbb{Q}(\zeta)$. Similarly, $x^5 - 2$ is solvable by radicals over \mathbb{Q} , for its splitting field over \mathbb{Q} is generated by $\sqrt[5]{2}$ and ζ , where $\sqrt[5]{2}$ is the real zero of $x^5 - 2$. ▲

To say that the quintic is insolvable in the classic case means that there exists *some* polynomial of degree 5 with real coefficients that is not solvable by radicals. We shall show this. *We assume throughout this section that all fields mentioned have characteristic 0.*

The outline of the argument is as follows, and it is worthwhile to try to remember it.

1. *We shall show that a polynomial $f(x) \in F[x]$ is solvable by radicals over F (if and) only if its splitting field E over F has a solvable Galois group.* Recall that a solvable group is one having a composition series with abelian quotients. While this theorem goes both ways, we shall not prove the “if” part.
2. *We shall show that there is a polynomial $f(x) \in \Phi[x]$ of degree 5 with a splitting field E over \mathbb{Q} such that $G(E/\mathbb{Q}) \cong S_5$, the symmetric group on 5 letters.* Recall that a composition series for S_5 is $\{\mathbb{Q}\} < A_5 < S_5$. Since A_5 is not abelian, we will be done.

The following lemma does most of our work for Step 1.

49.3 Lemma

Let F be a field of characteristic 0, and let $a \in F$. If K is the splitting field of $x^n - a$ over F , then $G(K/F)$ is a solvable group.

■ HISTORICAL NOTE

The first publication of a formula for solving cubic equations in terms of radicals was in 1545 in the *Ars Magna* of Girolamo Cardano, although the initial discovery of the method is in part also due to Scipione del Ferro and Niccolo Tartaglia. Cardano's student, Lodovico Ferrari, discovered a method for solving quartic equations by radicals, which also appeared in Cardano's work.

After many mathematicians had attempted to solve quintics by similar methods, it was Joseph-Louis Lagrange who in 1770 first attempted a detailed analysis of the general principles underlying the solutions for polynomials of degree 3 and 4, and showed why these methods fail for those of higher degree. His basic insight was that in the former cases there were rational functions of the roots that took on two and three values, respectively, under all

possible permutations of the roots, hence these rational functions could be written as roots of equations of degree less than that of the original. No such functions were evident in equations of higher degree.

The first mathematician to claim to have a proof of the insolvability of the quintic equation was Paolo Ruffini (1765–1822) in his algebra text of 1799. His proof was along the lines suggested by Lagrange, in that he in effect determined all of the subgroups of S_5 and showed how these subgroups acted on rational functions of the roots of the equation. Unfortunately, there were several gaps in his various published versions of the proof. It was Niels Henrik Abel who, in 1824 and 1826, published a complete proof, closing all of Ruffini's gaps and finally settling this centuries-old question.

Proof Suppose first that F contains all the n th roots of unity. By Corollary 28.7 the n th roots of unity form a cyclic subgroup of $\langle F^*, \cdot \rangle$. Let ζ be a generator of the subgroup. (Actually, the generators are exactly the *primitive* n th roots of unity.) Then the n th roots of unity are

$$1, \zeta, \zeta^2, \dots, \zeta^{n-1}.$$

If $\beta \in \bar{F}$ is a zero of $(x^n - a) \in F[x]$, then all zeros of $x^n - a$ are

$$\beta, \zeta\beta, \zeta^2\beta, \dots, \zeta^{n-1}\beta.$$

Since $K = F(\beta)$, an automorphism σ in $G(K/F)$ is determined by the value $\sigma(\beta)$ of the automorphism σ on β . Now if $\sigma(\beta) = \zeta^i\beta$ and $\tau(\beta) = \zeta^j\beta$, where $\tau \in G(K/F)$, then

$$(\tau\sigma)(\beta) = \tau(\sigma(\beta)) = \tau(\zeta^i\beta) = \zeta^i\tau(\beta) = \zeta^i\zeta^j\beta,$$

since $\zeta^i \in F$. Similarly,

$$(\sigma\tau)(\beta) = \zeta^j\zeta^i\beta.$$

Thus $\sigma\tau = \tau\sigma$, and $G(K/F)$ is abelian and therefore solvable.

Now suppose that F does not contain a primitive n th root of unity. Let ζ be a generator of the cyclic group of n th roots of unity under multiplication in \bar{F} . Let β again be a zero of $x^n - a$. Since β and $\zeta\beta$ are both in the splitting field K of $x^n - a$, $\zeta = (\zeta\beta)/\beta$ is in K . Let $F' = F(\zeta)$, so we have $F < F' \leq K$. Now F' is a normal extension of F , since F' is the splitting field of $x^n - 1$. Since $F' = F(\zeta)$, an automorphism η in $G(F'/F)$ is determined by $\eta(\zeta)$, and we must have $\eta(\zeta) = \zeta^i$ for some i , since all zeros of $x^n - 1$ are powers of ζ . If $\mu(\zeta) = \zeta^j$ for $\mu \in G(F'/F)$, then

$$(\mu\eta)(\zeta) = \mu(\eta(\zeta)) = \mu(\zeta^i) = \mu(\zeta)^i = (\zeta^j)^i = \zeta^{ij},$$

and, similarly,

$$(\eta\mu)(\zeta) = \zeta^{ij}.$$

Thus $G(F'/F)$ is abelian. By Theorem 46.10,

$$\{\iota\} \leq G(K/F') \leq G(K/F)$$

is a normal series and hence a subnormal series of groups. The first part of the proof shows that $G(K/F')$ is abelian, and Galois theory tells us that $G(K/F)/G(K/F')$ is isomorphic to $G(F'/F)$, which is abelian. Exercise 6 shows that if a group has a subnormal series of subgroups with abelian quotient groups, then any refinement of this series also has abelian quotient groups. Thus a composition series of $G(K/F)$ must have abelian quotient groups, so $G(K/F)$ is solvable. \diamond

The following theorem will complete Step 1 of our program.

49.4 Theorem Let F be a field of characteristic zero, and let $F \leq E \leq K \leq \bar{F}$, where E is a normal extension of F and K is an extension of F by radicals. Then $G(E/F)$ is a solvable group.

Proof We first show that K is contained in a finite normal extension L of F by radicals and that the group $G(L/F)$ is solvable. Since K is an extension of F by radicals, $K = F(\alpha_1, \dots, \alpha_r)$ where $\alpha_i^{n_i} \in F(\alpha_1, \dots, \alpha_{i-1})$ for $1 < i \leq r$ and $\alpha_1^{n_1} \in F$. To form L , we first form the splitting field L_1 of $f_1(x) = x^{n_1} - \alpha_1^{n_1}$ over F . Then L_1 is a normal extension of F , and Lemma 49.3 shows that $G(L_1/F)$ is a solvable group. Now $\alpha_2^{n_2} \in L_1$ and we form the polynomial

$$f_2(x) = \prod_{\sigma \in G(L_1/F)} [(x^{n_2} - \sigma(\alpha_2)^{n_2})].$$

Since this polynomial is invariant under action by any σ in $G(L_1/F)$, we see that $f_2(x) \in F[x]$. We let L_2 be the splitting field of $f_2(x)$ over L_1 . Then L_2 is a splitting field over F also and is a normal extension of F by radicals. We can form L_2 from L_1 via repeated steps as in Lemma 49.3, passing to a splitting field of $x^{n_2} - \sigma(\alpha_2)^{n_2}$ at each step. By Lemma 49.3 and Exercise 7, we see that the Galois group over F of each new extension thus formed continues to be solvable. We continue this process of forming splitting fields over F in this manner: At stage i , we form the splitting field of the polynomial

$$f_i(x) = \prod_{\sigma \in G(L_{i-1}/F)} [(x^{n_i} - \sigma(\alpha_i)^{n_i})]$$

over L_{i-1} . We finally obtain a field $L = L_r$ that is a normal extension of F by radicals, and we see that $G(L/F)$ is a solvable group. We see from construction that $K \leq L$.

To conclude, we need only note that by Theorem 46.8, we have $G(E/F) \cong G(L/F)/G(L/E)$. Thus $G(E/F)$ is a factor group, and hence a homomorphic image, of $G(L/F)$. Since $G(L/F)$ is solvable, Exercise 31 of Section 18 shows that $G(E/F)$ is solvable. \diamond

The Insolvability of the Quintic

It remains to find a polynomial $f(x) \in \mathbb{Q}[x]$, whose splitting field has Galois group S_5 . The polynomial $f(x) = 2x^5 - 5x^4 + 5$ does the trick as we now show. To begin, we prove the following theorem that gives a simple condition for a subgroup of S_5 to actually be all of S_5 . We will use this theorem to show that the Galois group of the splitting field of $f(x)$ over \mathbb{Q} is isomorphic with S_5 .

49.5 Theorem Let H be a subgroup of S_5 . If H has a transposition and a 5-cycle, then $H = S_5$.

Proof We can assume, by relabeling the points being permuted, that the 5-cycle $(0, 1, 2, 3, 4)$ is in H and the transposition $(0, j)$, $j \neq 0$, is also in H . We think of S_5 as permuting the

elements in \mathbb{Z}_5 . By conjugating with the 5-cycle $(0, 1, 2, 3, 4)^r$, we have that for each $0 \leq r \leq 4$,

$$(0, 1, 2, 3, 4)^r(0, j)(0, 1, 2, 3, 4)^{-r} = (r, r + j),$$

where addition is in \mathbb{Z}_5 . Thus $(0, j)$ and $(j, 2j)$ are both in H ; and therefore

$$(0, j)(j, 2j)(0, j) = (0, 2j) \in H.$$

As before, by conjugating $(0, 2j)$ with $(0, 1, 2, 3, 4)^r$ we have $(r, r + 2j) \in H$ for $0 \leq r \leq 4$. Now,

$$(0, 2j)(2j, 3j)(0, 2j) = (0, 3j) \in H.$$

Again, conjugating with powers of $(0, 1, 2, 3, 4)$ shows that $(r, r + 3j) \in H$. Furthermore,

$$(0, 3j)(3j, 4j)(0, 3j) = (0, 4j) \in H$$

and $(r, r + 4j) \in H$. Summarizing, we have that

$$\{(r, r + sj) \mid r \in \mathbb{Z}_5 \text{ and } s = \mathbb{Z}_5^*\} \in H.$$

But

$$\{sj \mid s \in \mathbb{Z}_5^*\} = \mathbb{Z}_5^*$$

since $j \neq 0$ is a unit in the field \mathbb{Z}_5 . Therefore, H contains all the transpositions in the set

$$\{(r, r + sj) \mid r \in \mathbb{Z}_5 \text{ and } s \in \mathbb{Z}_5^*\} = \{(r, r + t) \mid r \in \mathbb{Z}_5 \text{ and } t \in \mathbb{Z}_5^*\}.$$

This is the set of all transpositions in S_5 . By Theorem 8.15, $H = S_5$. \blacklozenge

We now turn our attention back to the polynomial $f(x) = 2x^5 - 5x^4 + 5$. We first observe that $f(x)$ is irreducible over \mathbb{Q} by Eisenstein's criterion, Theorem 28.16, using $p = 5$. In order to better understand the zeros of $f(x)$, we compute the following.

$$f(-1) = -2 < 0$$

$$f(0) = 5 > 0$$

$$f(2) = -11 < 0$$

$$f(3) = 86 > 0$$

The intermediate value theorem from calculus says that if $f(x)$ is a continuous function and $f(a)$ and $f(b)$ have opposite signs, then $f(x)$ has a zero between a and b . Since the polynomial $f(x)$ is continuous, we have at least three real number zeros of $f(x)$; one between -1 and 0 , one between 0 and 2 , and the third between 2 and 3 . The derivative of $f(x)$ is

$$f'(x) = 10x^4 - 20x^3 = 10x^3(x - 2).$$

The only zeros of $f'(x)$ are 0 and 2 . By the mean value theorem from calculus, between any two real number zeros of $f'(x)$, there is a zero of $f'(x)$. Therefore, there cannot be more than three zeros of $f'(x)$ that are real numbers. Let K be the splitting field of $f(x)$ over \mathbb{Q} . Since $f(x)$ is irreducible over \mathbb{Q} and \mathbb{Q} is a perfect field, K contains five distinct zeros of $f(x)$. Thus $f(x)$ has two complex zeros, and they are complex conjugates.

We have that $G(K/\mathbb{Q})$ is isomorphic with a subgroup of the permutation group of the zeros of $f(x)$. The isomorphism maps each σ to the permutation given by the map σ restricted to the zeros of $f(x)$. For any zero $\alpha \in K$ of $f(x)$,

$$[\mathbb{Q}(\alpha) : \mathbb{Q}] = \deg(\text{irr}(\alpha, \mathbb{Q})) = \deg(f(x)) = 5.$$

Thus

$$|G(K/\mathbb{Q})| = [K : \mathbb{Q}] = [K : \mathbb{Q}(\alpha)] \cdot [\mathbb{Q}(\alpha) : \mathbb{Q}] = [K : \mathbb{Q}(\alpha)] \cdot 5.$$

Cauchy's Theorem 14.20 says that $G(K/\mathbb{Q})$ has a subgroup of order 5 and therefore an element of order 5. So $G(K/\mathbb{Q})$ has an element that permutes the zeros of $f(x)$ in a 5-cycle, since the only elements in S_5 that have order 5 are the 5-cycles. Also, complex conjugation is an element in $G(K/\mathbb{Q})$ that fixes each real zero of $f(x)$ and switches the two complex zeros. Thus an element of $G(K/\mathbb{Q})$ permutes the zeros of $f(x)$ by a 2-cycle. By Theorem 49.5, $G(K/\mathbb{Q})$ is isomorphic with S_5 .

49.6 Theorem There is a polynomial in $\mathbb{Q}[x]$ that is not solvable by radicals.

Proof As we just discovered, $f(x) = 2x^5 - 5x^4 + 5$ is one such polynomial since its splitting field over \mathbb{Q} has Galois group isomorphic with S_5 , which is not a solvable group. ◆

There are many other polynomials that are not solvable by radicals. In the proof given above, we need a few key properties for the polynomial $f(x)$ to have Galois group S_5 . First we require that $f(x)$ be an irreducible polynomial over \mathbb{Q} with degree 5. We then exploit the fact that $f(x)$ has three real zeros and two complex zeros. As long as a polynomial meets these conditions, the polynomial is not solvable by radicals. The set of polynomials with these properties is infinite as shown in Exercise 8.

Exercise 9 gives a different approach to constructing polynomials in $F[x]$ whose Galois group over F is S_5 , where F is a subfield of \mathbb{R} . This approach also shows that there are an infinite number of polynomials that are not solvable by radicals, although the polynomials do not have rational coefficients. While Exercises 8 and 9 both produce an infinite number of polynomials that are not solvable by radicals, in Exercise 8 the number is countable while in Exercise 9 the number is uncountable.

■ EXERCISES 49

Concepts

1. Can the splitting field K of $x^2 + x + 1$ over \mathbb{Z}_2 be obtained by adjoining a square root to \mathbb{Z}_2 of an element in \mathbb{Z}_2 ? Is K an extension of \mathbb{Z}_2 by radicals?
2. Is every polynomial in $F[x]$ of the form $ax^8 + bx^6 + cx^4 + dx^2 + e$, where $a \neq 0$, solvable by radicals over F , if $F \leq \mathbb{R}$? Why or why not?
3. Determine whether each of the following is true or false.
 - a. Let F be a field of characteristic 0. A polynomial in $F[x]$ is solvable by radicals if and only if its splitting field in \bar{F} is contained in an extension of F by radicals.
 - b. Let F be a field of characteristic 0. A polynomial in $F[x]$ is solvable by radicals if and only if its splitting field in \bar{F} has a solvable Galois group over F .
 - c. The splitting field of $x^{17} - 5$ over \mathbb{Q} has a solvable Galois group.
 - d. If $f(x) \in \mathbb{Q}[x]$ is any polynomial of degree five having three real zeros and two complex zeros, then $f(x)$ is not solvable by radicals.
 - e. The Galois group of a finite extension of a finite field is solvable.
 - f. No quintic polynomial is solvable by radicals over any field.
 - g. Every 4th degree polynomial over a field $F \leq \mathbb{R}$ is solvable by radicals.
 - h. The zeros of a cubic polynomial over a field $F \leq \mathbb{R}$ can always be attained by means of a finite sequence of operations of addition, subtraction, multiplication, division, and taking square roots starting with elements in F .

- i. The zeros of a cubic polynomial over a field F of characteristic 0 can never be attained by means of a finite sequence of operations of addition, subtraction, multiplication, division, and taking square roots, starting with elements in F .
- j. The theory of subnormal series of groups plays an important role in applications of Galois theory.

Theory

4. Let F be a field, and let $f(x) = ax^2 + bx + c$ be in $F[x]$, where $a \neq 0$. Show that if the characteristic of F is not 2, the splitting field of $f(x)$ over F is $F(\sqrt{b^2 - 4ac})$. [Hint: Complete the square, just as in your high school work, to derive the “quadratic formula.”]

5. Show that if F is a field of characteristic different from 2 and

$$f(x) = ax^4 + bx^2 + c,$$

where $a \neq 0$, then $f(x)$ is solvable by radicals over F .

6. Show that for a finite group, every refinement of a subnormal series with abelian quotients also has abelian quotients, thus completing the proof of Lemma 49.3. [Hint: Use Theorem 16.8.]

7. Show that for a finite group, a subnormal series with solvable quotient groups can be refined to a composition series with abelian quotients, thus completing the proof of Theorem 49.4. [Hint: Use Theorem 16.8.]

8. Let p be a prime number and $f(x) = x^5 - p^2x + p \in \mathbb{Q}[x]$. Prove that $f(x)$ is not solvable by radicals. [Hint: Mimic the proof that $2x^5 - 5x^4 + 5$ is not solvable by radicals.]

9. This is an alternate method of finding a polynomial with coefficients in a field F , a subfield of \mathbb{R} , whose Galois group over F is S_5 .

- a. Suppose that $F \leq \mathbb{R}$ is a countable field and x is an indeterminate. Show that $F[x]$, $F(x)$, and the set of real numbers that are algebraic over F are all countable sets.
- b. Show that there is a sequence of real numbers $y_1, y_2, \dots \in \mathbb{R}$, with y_1 transcendental over \mathbb{Q} , and for each $i > 1$, y_i is transcendental over $\mathbb{Q}(y_1, y_2, \dots, y_{i-1})$.
- c. Using the notation of part b), let $K = \mathbb{Q}(y_1, y_2, y_3, y_4, y_5)$ and

$$f(x) = \prod_{i=1}^5 (x - y_i).$$

Let s_1, s_2, s_3, s_4, s_5 be the value of the elementary symmetric functions (as defined in Section 47) evaluated at y_1, y_2, y_3, y_4, y_5 . Show that $f(x) \in F = \mathbb{Q}(s_1, s_2, s_3, s_4, s_5)$.

- d. Show that K is the splitting field of $f(x)$ over F and $G(K/F)$ is isomorphic with S_5 .

This page is intentionally left blank

Appendix: Matrix Algebra

We give a brief summary of matrix algebra here. Matrices appear in examples in some chapters of the text and also are involved in several exercises.

A **matrix** is a rectangular array of numbers. For example, the array

$$\begin{bmatrix} 2 & -1 & 4 \\ 3 & 1 & 2 \end{bmatrix} \quad (1)$$

is a matrix having two rows and three columns. A matrix having m rows and n columns is an $m \times n$ matrix, so Matrix (1) is a 2×3 matrix. If $m = n$, the matrix is **square**. Entries in a matrix may be any type of number—integer, rational, real, or complex. We let $M_{m \times n}(\mathbb{R})$ be the set of all $m \times n$ matrices with real number entries. If $m = n$, the notation is abbreviated to $M_n(\mathbb{R})$. We can similarly consider $M_n(\mathbb{Z})$, $M_{2 \times 3}(\mathbb{C})$, etc.

Two matrices having the same number m of rows and the same number n of columns can be added in the obvious way: we add entries in corresponding positions.

A1 Example In $M_{2 \times 3}(\mathbb{Z})$, we have

$$\begin{bmatrix} 2 & -1 & 4 \\ 3 & 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & -3 \\ 2 & -7 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 1 \\ 5 & -6 & 3 \end{bmatrix}. \quad \blacktriangle$$

We will use uppercase letters to denote matrices. If A , B , and C are $m \times n$ matrices, it is easily seen that $A + B = B + A$ and that $A + (B + C) = (A + B) + C$.

Matrix multiplication, AB , is defined only if the number of columns of A is equal to the number of rows of B . That is, if A is an $m \times n$ matrix, then B must be an $n \times s$ matrix for some integer s . We start by defining as follows the product AB where A is a $1 \times n$ matrix and B is an $n \times 1$ matrix:

$$AB = [a_1 \quad a_2 \quad \cdots \quad a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n. \quad (2)$$

Note that the result is a number. (We shall not distinguish between a number and the 1×1 matrix having that number as its sole entry.) You may recognize this product as the *dot product* of vectors. Matrices having only one *row* or only one *column* are **row vectors** or **column vectors**, respectively.

A2 Example We find that

$$\begin{bmatrix} 3 & -7 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} = (3)(1) + (-7)(4) + (2)(5) = -15. \quad \blacktriangle$$

Let A be an $m \times n$ matrix and let B be an $n \times s$ matrix. Note that the number n of entries in each row of A is the same as the number n of entries in each column of B . The product $C = AB$ is an $m \times s$ matrix. The entry in the i th row and j th column of AB is the product of the i th row of A times the j th column of B as defined by Eq. (2) and illustrated in Example A2.

A3 Example Compute

$$AB = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 4 & 6 \end{bmatrix} \begin{bmatrix} 3 & 1 & 2 & 1 \\ 1 & 4 & 1 & -1 \\ -1 & 0 & 2 & 1 \end{bmatrix}.$$

Solution Note that A is 2×3 and B is 3×4 . Thus AB will be 2×4 . The entry in its second row and third column is

$$(2\text{nd row } A)(3\text{rd column } B) = [1 \quad 4 \quad 6] \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = 2 + 4 + 12 = 18.$$

Computing all eight entries of AB in this fashion, we obtain

$$AB = \begin{bmatrix} 2 & -2 & 9 & 6 \\ 1 & 17 & 18 & 3 \end{bmatrix}. \quad \blacktriangle$$

A4 Example The product

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

is not defined, since the number of entries in a row of the first matrix is not equal to the number of entries in a column of the second matrix. \blacktriangle

For square matrices of the same size, both addition and multiplication are always defined. Exercise 10 asks us to illustrate the fact that matrix multiplication is not commutative.

That is, AB need not equal BA even when both products are defined, as for $A, B \in M_2(\mathbb{Z})$. It can be shown that $A(BC) = (AB)C$ and $A(B + C) = AB + AC$ whenever all these expressions are defined.

We let I_n be the $n \times n$ matrix with entries 1 along the diagonal from the upper-left corner to the lower-right corner, and entries 0 elsewhere. For example,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is easy to see that if A is any $n \times s$ matrix and B is any $r \times n$ matrix, then $I_n A = A$ and $B I_n = B$. That is, the matrix I_n acts much as the number 1 does for multiplication when multiplication by I_n is defined.

Let A be an $n \times n$ matrix and consider a matrix equation of the form $AX = B$, where A and B are known but X is unknown. If we can find an $n \times n$ matrix A^{-1} such that $A^{-1}A = AA^{-1} = I_n$, then we can conclude that

$$A^{-1}(AX) = A^{-1}B, \quad (A^{-1}A)X = A^{-1}B, \quad I_nX = A^{-1}B, \quad X = A^{-1}B,$$

and we have found the desired matrix X . Such a matrix A^{-1} acts like the reciprocal of a number: $A^{-1}A = I_n$ and $(1/r)r = 1$. This is the reason for the notation A^{-1} .

If A^{-1} exists, the square matrix A is **invertible** and A^{-1} is the **inverse** of A . If A^{-1} does not exist, then A is said to be **singular**. It can be shown that if there exists an $n \times n$ matrix A^{-1} such that $A^{-1}A = I_n$, then $AA^{-1} = I_n$ also, and furthermore, there is only one matrix A^{-1} having this property.

A5 Example Let

$$A = \begin{bmatrix} 2 & 9 \\ 1 & 4 \end{bmatrix}.$$

We can check that

$$\begin{bmatrix} -4 & 9 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 9 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 9 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -4 & 9 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus,

$$A^{-1} = \begin{bmatrix} -4 & 9 \\ 1 & -2 \end{bmatrix}. \quad \blacktriangle$$

We leave the problems of determining the existence of A^{-1} and its computation to a course in linear algebra.

Associated with each square $n \times n$ matrix A is a number called the *determinant* of A and denoted by $\det(A)$. This number can be computed as sums and differences of certain products of the numbers that appear in the matrix A . For example, the determinant of the 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $ad - bc$. Note that an $n \times 1$ matrix with real number entries can be viewed as giving coordinates of a point in n -dimensional Euclidean space \mathbb{R}^n . Multiplication of such a single column matrix on the left by a real $n \times n$ matrix A produces another such single column matrix corresponding to another point in \mathbb{R}^n . This multiplication on the left by A thus gives a map of \mathbb{R}^n into itself. It can be shown that a piece of \mathbb{R}^n of volume V is mapped by this multiplication by A into a piece of volume $|\det(A)| \cdot V$. This is one of the reasons that determinants are important.

The following properties of determinants for $n \times n$ matrices A and B are of interest in this text:

1. $\det(I_n) = 1$
2. $\det(AB) = \det(A)\det(B)$
3. $\det(A) \neq 0$ if and only if A is an invertible matrix
4. If B is obtained from A by interchanging two rows (or two columns) of A , then $\det(B) = -\det(A)$
5. If every entry of A is zero above the *main diagonal* from the upper left corner to the lower right corner, then $\det(A)$ is the product of the entries on this diagonal. The same is true if all entries below the main diagonal are zero.

■ EXERCISES A

In Exercises 1 through 9, compute the given arithmetic matrix expression, if it is defined.

1. $\begin{bmatrix} -2 & 4 \\ 1 & 5 \end{bmatrix} + \begin{bmatrix} 4 & -3 \\ 1 & 2 \end{bmatrix}$

2. $\begin{bmatrix} 1+i & -2 & 3-i \\ 4 & i & 2-i \end{bmatrix} + \begin{bmatrix} 3 & i-1 & -2+i \\ 3-i & 1+i & 0 \end{bmatrix}$

3. $\begin{bmatrix} i & -1 \\ 4 & 1 \\ 3 & -2i \end{bmatrix} - \begin{bmatrix} 3-i & 4i \\ 2 & 1+i \\ 3 & -i \end{bmatrix}$

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

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

6. $\begin{bmatrix} 4 & -1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 7 \\ 3 & 1 \end{bmatrix}$

7. $\begin{bmatrix} i & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 3i & 1 \\ 4 & -2i \end{bmatrix}$

8. $\begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}^4$

9. $\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}^4$

10. Give an example in $M_2(\mathbb{Z})$ showing that matrix multiplication is not commutative.

11. Find $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}^{-1}$, by experimentation if necessary.

12. Find $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -1 \end{bmatrix}^{-1}$, by experimentation if necessary.

13. If $A = \begin{bmatrix} 3 & 0 & 0 \\ 10 & -2 & 0 \\ 4 & 17 & 8 \end{bmatrix}$, find $\det(A)$.

14. Prove that if $A, B \in M_n(\mathbb{C})$ are invertible, then AB and BA are invertible also.

Bibliography

Classic Works

1. N. Bourbaki, *Éléments de Mathématique*, Book II of Part I, *Algèbre*. Paris: Hermann, 1942–58.
2. N. Jacobson, *Lectures in Abstract Algebra*. Princeton, NJ: Van Nostrand, vols. I, 1951, II, 1953, and III, 1964.
3. O. Schreier and E. Sperner, *Introduction to Modern Algebra and Matrix Theory* (English translation), 2nd Ed. New York: Chelsea, 1959.
4. B. L. van der Waerden, *Modern Algebra* (English translation). New York: Ungar, vols. I, 1949, and II, 1950.

General Algebra Texts

5. M. Artin, *Algebra*. (Classic Version), 2nd Edition, London: Pearson, 2018.
6. A. A. Albert, *Fundamental Concepts of Higher Algebra*. Chicago: University of Chicago Press, 1956.
7. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, 3rd Ed. New York: Macmillan, 1965.
8. J. A. Gallian, *Contemporary Abstract Algebra*, 8th Ed. Boston, MA: Brook/Cole, 2013.
9. I. N. Herstein, *Topics in Algebra*. New York: Blaisdell, 1964.
10. T. W. Hungerford, *Algebra*. New York: Springer, 1974.
11. S. Lang, *Algebra*. Reading, MA: Addison-Wesley, 1965.
12. S. MacLane and G. Birkhoff, *Algebra*. New York: Macmillan, 1967.
13. N. H. McCoy and G. J. Janusz, *Introduction to Modern Algebra*. Cambridge, MA: Academic Press, 2001.
14. G. D. Mostow, J. H. Sampson, and J. Meyer, *Fundamental Structures of Algebra*. New York: McGraw-Hill, 1963.
15. W. W. Sawyer, *A Concrete Approach to Abstract Algebra*. Mineola, NY: Dover, 1978.

Group Theory

16. W. Burnside, *Theory of Groups of Finite Order*, 2nd Ed. Cambridge, UK: Cambridge University Press, 2012.
17. H. S. M. Coxeter and W. O. Moser, *Generators and Relations for Discrete Groups*, 2nd Ed. Berlin: Springer, 1965.
18. M. Hall, Jr., *The Theory of Groups*. Mineola, NY: Dover, 2018.
19. A. G. Kurosh, *The Theory of Groups* (English translation). New York: Chelsea, vols. I, 1955, and II, 1956.
20. W. Ledermann, *Introduction to the Theory of Finite Groups*, 4th rev. Ed. New York: Interscience, 1961.

Bibliography

21. J. G. Thompson and W. Feit, "Solvability of Groups of Odd Order." *Pac. J. Math.*, **13** (1963), 775–1029.
22. M. A. Rabin, "Recursive Unsolvability of Group Theoretic Problems." *Ann. Math.*, **67** (1958), 172–194.

Ring Theory

23. W. W. Adams and P. Loustaunau, *An Introduction to Gröbner Bases* (Graduate Studies in Mathematics, vol. 3). Providence, RI: American Mathematical Society, 1994.
24. E. Artin, C. J. Nesbitt, and R. M. Thrall, *Rings with Minimum Condition*. Ann Arbor: University of Michigan Press, 1964.
25. N. H. McCoy, *Rings and Ideals* (Carus Monograph No. 8), 5th Ed. Buffalo: The Mathematical Association of America, 1971.
26. N. H. McCoy, *The Theory of Rings*. New York: Macmillan, 1964.

Field Theory

27. E. Artin, *Galois Theory* (Notre Dame Mathematical Lecture No. 2), 2nd Ed. Notre Dame, IN: University of Notre Dame Press, 1944.
28. O. Zariski and P. Samuel, *Commutative Algebra*. Princeton, NJ: Van Nostrand, vol. I, 1958.

Number Theory

29. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th Ed. Oxford: Oxford University Press, 2008.
30. S. Lang, *Algebraic Numbers*. Reading, MA: Addison-Wesley, 1964.
31. W. J. LeVeque, *Elementary Theory of Numbers*. Mineola, NY: Dover, 1990.
32. W. J. LeVeque, *Topics in Number Theory*. Mineola, NY: Dover, 2002.
33. T. Nagell, *Introduction to Number Theory*, 2nd Ed. Providence, RI: American Mathematical Society, 2001.
34. I. Niven and H. S. Zuckerman, *An Introduction to the Theory of Numbers*, 5th Ed. New York: Wiley, 1991.
35. H. Pollard, *The Theory of Algebraic Numbers* (Carus Monograph No. 9). Buffalo: The Mathematical Association of America; New York: Wiley, 1950.
36. D. Shanks, *Solved and Unsolved Problems in Number Theory*. Washington, DC: Spartan Books, vol. I, 1962.
37. B. M. Stewart, *Theory of Numbers*, 2nd Ed. New York: Macmillan, 1964.
38. J. V. Uspensky and M. H. Heaslet, *Elementary Number Theory*. New York: McGraw-Hill, 1939.
39. E. Weiss, *Algebraic Number Theory*. Mineola, NY: Dover, 1998.

Homological Algebra

40. J. P. Jans, *Rings and Homology*. New York: Holt, 1964.
41. S. MacLane, *Homology*. Berlin: Springer, 1963.

Other References

42. A. A. Albert (ed.), *Studies in Modern Algebra* (MAA Studies in Mathematics, vol. 2). Buffalo: The Mathematical Association of America; Englewood Cliffs, NJ: Prentice-Hall, 1963.
43. E. Artin, *Geometric Algebra*. New York: Interscience, 1957.
44. R. Courant and H. Robbins, *What Is Mathematics?* Oxford University Press, 1941.
45. H. S. M. Coxeter, *Introduction to Geometry*, 2nd Ed. New York: Wiley, 1969.
46. R. H. Crowell and R. H. Fox, *Introduction to Knot Theory*. New York: Ginn, 1963.
47. H. B. Edgerton, *Elements of Set Theory*. San Diego: Academic Press, 1977.
48. C. Schumacher, *Chapter Zero*. Reading, MA: Addison-Wesley, 1996.

Notations

$\epsilon, a \in S$	membership, 1
\emptyset	empty set, 1
$\notin, a \notin S$	nonmembership, 1
$\{x \mid P(x)\}$	set of all x such that $P(x)$, 1
$B \subseteq A$	set inclusion, 2
$B \subset A$	subset $B \neq A$, 2
$A \times B$	Cartesian product of sets, 2
\mathbb{Z}	integers, 2
\mathbb{Q}	rational numbers, 2
\mathbb{R}	real numbers, 3
\mathbb{C}	complex numbers, 3
$\mathbb{Z}^+, \mathbb{Q}^+, \mathbb{R}^+$	positive elements of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, 3
$\mathbb{Z}^*, \mathbb{Q}^*, \mathbb{R}^*, \mathbb{C}^*$	nonzero elements of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, 3
\mathcal{R}	relation, 3
$ A $	number of elements in A , 3; as order of group, 41
$\phi : A \rightarrow B$	mapping of A into B by ϕ , 3
$\phi(a)$	image of element a under ϕ , 4
$\phi[A]$	image of set A under ϕ , 4
\leftrightarrow	one-to-one correspondence, 4
ϕ^{-1}	the inverse function of ϕ , 4
\aleph_0	cardinality of \mathbb{Z}^+ , 4
\bar{x}	cell containing $x \in S$ in a partition of S , 6
$\mathbb{Z}/n\mathbb{Z}$	residue classes modulo n , 6
$\equiv_n, a \equiv b \pmod{n}$	congruence modulo n , 6
$\mathcal{P}(A)$	power set of A , 8
U	set of all $z \in \mathbb{C}$ such that $ z = 1$, 36
\mathbb{R}_c	set of all $x \in \mathbb{R}$ such that $0 \leq x < c$, 33
$+_c$	addition modulo c , 33
U_n	group of n th roots of unity, 37

\mathbb{Z}_n	$\{0, 1, 2, \dots, n - 1\}$, 32 cyclic group $\{0, 1, \dots, n - 1\}$ under addition modulo n , 38 ring $\{0, 1, \dots, n - 1\}$ under addition and multiplication modulo n , 189
$*_*, a *_b$	binary operation, 11
$\circ, f \circ g, \sigma \tau$	function composition, 14, 41
e	identity element, 20
$M_{m \times n}(S)$	$m \times n$ matrices with entries from S , 22, 393
$M_n(S)$	$n \times n$ matrices with entries from S , 22, 393
$GL(n, \mathbb{R})$	general linear group of degree n , 22, 23
$\det(A)$	determinant of square matrix A , 28, 395
$a^{-1}, -a$	inverse of a , 15
$H \leq G; K \leq L$	subgroup inclusion, 52; substructure inclusion, 192
$H < G; K < L$	subgroup $H \neq G$, 52; substructure $K \neq L$, 192
$\langle a \rangle$	cyclic subgroup generated by a , 56
$n\mathbb{Z}$	principal ideal generated by n , 256
A^T	subgroup of \mathbb{Z} generated by n , 245
\gcd	transpose of A , 57
$\cap_{i \in I} S_i,$	greatest common divisor, 63, 283, 304
$S_1 \cap S_2 \cap \dots \cap S_n$	intersection of sets, 71
S_A	group of permutations of A , 43
ι	identity map, 43
S_n	symmetric group on n letters, 43
$n!$	n factorial, 43
D_n	n th dihedral group, 47
A_n	alternating group on n letters, 84
$aH, a + H$	left coset of H containing a , 98
$Ha, H + a$	right coset of H containing a , 100
$(G : H)$	index of H in G , 137
$\prod_{i=1}^n B_i,$	Euler phi-function, 105, 204
$B_1 \times B_2 \times \dots \times B_n$	Cartesian product of sets, 88
$\prod_{i=1}^n G_i$	direct product of groups, 88, 89
$\oplus_{i=1}^n G_i$	direct sum of groups, 89
lcm	least common multiple, 90
\bar{G}_i	natural subgroup of $\prod_{i=1}^n G_i$, 91
$H \leq G$	H normal subgroup of G , 116
$\text{SL}(n, \mathbb{R})$	special linear group, 117
ϕ_a	evaluation homomorphism, 191
π_i	projection onto i th component, 248
$\phi^{-1}[B]$	inverse image of the set B under ϕ , 78
$\text{Ker}(\phi)$	kernel of homomorphism ϕ , 78
$G/N; R/N$	factor group, 117; factor ring, 247
γ	canonical residue class map, 118, 119
i_g	inner automorphism, 120
$Z(G)$	center of the group G , 130
C	commutator subgroup, 130
X_g	subset of elements of X fixed by g , 136
G_x	isotropy subgroup of elements of G leaving x fixed, 137
Gx	orbit of x under G , 143
$R[x]$	polynomial ring with coefficients in R , 220
$R[x_1, x_2, \dots, x_n]$	polynomials in n indeterminates, 299
$F(x)$	field of quotients of $F[x]$, 222
$F(x_1, \dots, x_n)$	field of rational functions in n indeterminates, 223

$\Phi_p(x)$	cyclotomic polynomial of degree $p - 1$, 236
$\text{End}(A)$	endomorphisms of A , 260
RG	group ring, 262
FG	group algebra over the field F , 263
\mathbb{H}	quaternions, 264, 265
ACC	ascending chain condition, 280
F^n	Cartesian product, 299
$F[x]$	ring of polynomials in x_1, \dots, x_n over F , 299
$V(S)$	algebraic variety of polynomials in S , 300
$\langle b_1, \dots, b_r \rangle$	ideal generated by elements b_1, \dots, b_r , 300
$\text{lt}(f)$	leading term of the polynomial f , 306
$\text{lp}(f)$	power product of $\text{lt}(f)$, 306
$\text{irr}(\alpha, F)$	irreducible polynomial for α over F , 317
$\deg(\alpha, F)$	degree of α over F , 317
$F(\alpha)$	field obtained by adjoining α to field F , 317
$[E : F]$	degree of E over F , 321
$F(\alpha_1, \dots, \alpha_n)$	field obtained by adjoining $\alpha_1, \dots, \alpha_n$ to F , 323
$\overline{F_E}$	algebraic closure of F in E , 325
\overline{F}	an algebraic closure of F , 325, 326
$\text{GF}(p^n)$	Galois field of order p^n , 337
HN	product set, 148
$H \vee N$	subgroup join, 148
$N[H]$	normalizer of H , 152
$F[A]$	free group on A , 175
$(x_j : r_i)$	group presentation, 181
$a b$	a divides (is a factor of) b , 278
UFD	unique factorization domain, 278
PID	principal ideal domain, 278
$\cup_{i \in I} S_i$	union of sets, 280
$S_1 \cup S_2 \cup \dots \cup S_n$	
v	Euclidean norm, 288
$N(\alpha)$	norm of α , 294, 296
$\psi_{\alpha, \beta}$	conjugation isomorphism of $F(\alpha)$ with $F(\beta)$, 347
$E_{\{\sigma_i\}}, E_H$	subfield of E fixed by all σ_i or all $\sigma \in H$, 345
$G(E/F)$	automorphism group of E over F , 346
$\lambda(E)$	automorphisms that fix E , 367

This page is intentionally left blank

Answers to Odd-Numbered Exercises Not Asking for Definitions or Proofs

SECTION 0

1. $\{-\sqrt{3}, \sqrt{3}\}$
3. $\{1, -1, 2, -2, 3, -3, 4, -4, 5, -5, 6, -6, 10, -10, 12, -12, 15, -15, 20, -20, 30, -30, 60, -60\}$
5. Not a set (not well defined). A case can also be made for the empty set \emptyset .
7. The set \emptyset
9. It is not a well-defined set.
11. $(a, 1), (a, 2), (a, c), (b, 1), (b, 2), (b, c), (c, 1), (c, 2), (c, c)$
13. Draw the line through P and x , and let y be the point where it intersects the line segment CD .
17. Conjecture: $n(\mathcal{P}(A)) = 2^s$. (Proofs are usually omitted from answers.)
21. $10^2, 10^5, 10^{80} = 12^{80} = 2^{80} = |\mathbb{R}|$. (The numbers x where $0 \leq x \leq 1$ can be written to base 12 and to base 2 as well as to base 10.)
23. 1 25. 5 27. 52
29. Not an equivalence relation
31. Not an equivalence relation.
33. An equivalence relation;
 $\overline{1} = \{1, 2, \dots, 9\}$,
 $\overline{10} = \{10, 11, \dots, 99\}$,
 $\overline{100} = \{100, 101, \dots, 999\}$, and in general
 $\overline{10^n} = \{10^n, 10^n + 1, \dots, 10^{n+1} - 1\}$
35. a. $\{\dots, -3, 0, 3, \dots\}, \{\dots, -2, 1, 4, \dots\}, \{\dots, -5, -2, 1, \dots\}$
b. $\{\dots, -4, 0, 4, \dots\}, \{\dots, -3, 1, 4, \dots\}, \{\dots, -6, -2, 2, \dots\}, \{\dots, -5, -1, 3, \dots\}$
c. $\{\dots, -5, 0, 5, \dots\}, \{\dots, -4, 1, 6, \dots\}, \{\dots, -3, 2, 7, \dots\}, \{\dots, -2, 3, 8, \dots\}, \{\dots, -1, 4, 9, \dots\}$
37. $\overline{1} = \{x \in \mathbb{Z} \mid x \div n \text{ has remainder } 1\}$ depends on the value of n .
41. The name *two-to-two function* suggests that such a function f should carry every pair of distinct points into two distinct points. Such a function is one to one in the conventional sense. (If the domain has only one element, a function cannot fail to be two to two, since the only way it can fail to be two to two is to carry two points into one point, and the set does not have two points.) Conversely, every function that is one to one in the conventional sense carries any pair of points into two distinct points. Thus the functions conventionally called one to one are precisely those that carry two points into two points, which is a much more intuitive unidirectional way of regarding them. Also, the standard way of trying to show a function is one to one is precisely to show that it does not carry two points into just one point. Thus, proving a function is one to one becomes more natural in the two-to-two terminology.

SECTION 1

1. e, b, a
3. a, c . * is not associative.
5. Top row: d ; second row: a ; fourth row: c, b .
7. Not commutative, not associative
9. Commutative, associative, has identity.
11. Not commutative, not associative
13. $8, 729, n^{[n(n+1)/2]}$
15. $n^{(n-1)^2}$
19. An identity in the set S with operation * is an element $e \in S$ such that for all $a \in S$, $a * e = e * a = a$.
21. Yes 23. No. Condition 2 is violated.
25. No. Condition 1 is violated.
27. a. Yes b. Yes
29. Let $S = \{?, \Delta\}$. Define * and *' on S by $a * b = ?$ and $a *' b = \Delta$ for all $a, b \in S$. (Other answers are possible.)
31. True 33. True
35. False. Let $f(x) = x^2, g(x) = x$, and $h(x) = 2x + 1$. Then
 $(f(x) - g(x)) - h(x) = x^2 - 3x - 1$ but
 $f(x) - (g(x) - h(x)) = x^2 - (-x - 1) = x^2 + x + 1$.
37. True
39. True 41. False. Let * be + and let *' be on \mathbb{Z} .

SECTION 2

1. No. \mathcal{G}_3 fails.
3. No. \mathcal{G}_1 fails.
5. No. \mathcal{G}_1 fails.
7. \mathcal{G}_3
9. \mathcal{G}_1
11. Yes 13. Yes
15. No. The matrix with all entries 0 is upper triangular, but has no inverse.
17. Yes.
19. (Proofs are omitted.) c. $-1/3$
21. 2, 3. (It gets harder for 4 elements, where the answer is *not* 4.)
25. a. F c. T e. F g. T i. F
35. b^2a^{12}

SECTION 3

1. $-i$
3. -1
5. $20 - 9i$
7. $17 - 15i$
9. $-4 + 4i$
11. $\sqrt{\pi^2 + e^2}$
13. $\sqrt{2} \left(-\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i \right)$
15. $\sqrt{34} \left(\frac{-3}{\sqrt{34}} + \frac{5}{\sqrt{34}}i \right)$
17. $\frac{1}{\sqrt{2}} \pm \frac{1}{\sqrt{2}}i, -\frac{1}{\sqrt{2}} \pm \frac{1}{\sqrt{2}}i$
19. $3i, \pm \frac{3\sqrt{3}}{2} - \frac{3}{2}i$
21. $\sqrt{3} \pm i, \pm 2i, -\sqrt{3} \pm i$
23. 7
25. $\frac{3}{8}$
27. $\sqrt{2}$
29. $x = 6$
31. 5
33. 1, 7
37. $\zeta^0 \leftrightarrow 0, \zeta^3 \leftrightarrow 7, \zeta^4 \leftrightarrow 4, \zeta^5 \leftrightarrow 1, \zeta^6 \leftrightarrow 6, \zeta^7 \leftrightarrow 3$
39. With $\zeta \leftrightarrow 4$, we must have $\zeta^2 \leftrightarrow 2, \zeta^3 \leftrightarrow 0$, and $\zeta^4 \leftrightarrow 4$ again, which is impossible for a one-to-one correspondence.
41. Multiplying, we obtain

$$z_1 z_2 = |z_1| |z_2| [(\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + (\cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2)i]$$

- and the desired result follows at once from Exercise 40 and the equation $|z_1| |z_2| = |z_1 z_2|$.
45. Let $f : \mathbb{R}_b \rightarrow \mathbb{R}_c$ be given by $f(x) = \frac{c}{b}x$.

SECTION 4

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

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

7. ι 9. ι

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

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

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

13. a. ρ^6 b. ρ c. $\mu\rho^{10}$ d. $\mu\rho^{10}$

15. {1, 2, 3, 4, 5, 6} 17. {1, 5}

19. a. These are “elementary permutation matrices,” resulting from permuting the rows of the identity matrix. When another matrix A is multiplied on the left by one of these matrices P , the rows of A are permuted in the same fashion that the rows of the 3×3 identity matrix were permuted to obtain P . Because all 6 possible permutations of the three rows are present, we see they will act just like the elements of S_3 in permuting the entries 1, 2, 3 of the given column vector. Thus they form a group because S_3 is a group.

b. The symmetric group S_3 .

21. Need to add that ϕ is one-to-one and onto.

23. This is a good definition.

25. Not a permutation 27. Not a permutation

29. a. T c. T e. F g. F

SECTION 5

1. Yes 3. Yes 5. Yes 7. \mathbb{Q}^+ and $\{\pi^n \mid n \in \mathbb{Z}\}$ 9. Yes

11. No. Not closed under multiplication.

13. Yes

15. a. Yes b. No. It is not even a subset of $\tilde{\mathbb{F}}$.

17. a. No. Not closed under addition. b. Yes

19. a. Yes b. No. The zero constant function is not in $\tilde{\mathbb{F}}$.

21. a. $-50, -25, 0, 25, 50$ b. $4, 2, 1, 1/2, 1/4$ c. $1, \pi, \pi^2, 1/\pi, 1/\pi^2$ d. $\iota, \rho^3, \rho^6, \rho^9, \rho^{12}, \rho^{15}$
e. $\iota, (1, 2, 3)(5, 6), (1, 3, 2), (5, 6), (1, 2, 3), (1, 3, 2)(5, 6)$

23. All matrices $\begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$ for $n \in \mathbb{Z}$

25. All matrices of the form $\begin{bmatrix} 4^n & 0 \\ 0 & 4^n \end{bmatrix}$ or $\begin{bmatrix} 0 & -2^{2n+1} \\ -2^{2n+1} & 0 \end{bmatrix}$ for $n \in \mathbb{Z}$

27. 4 29. 3 31. 4 33. 2 35. 3

39. a. T c. T e. F g. F i. T

41. This is a subgroup: ι is in the set, the set is closed under function composition, and if $\sigma(b) = b$, then $\sigma^{-1}(b) = b$.

43. This is not a subgroup. Let $A = \mathbb{Z}$, $B = \mathbb{Z}^+$, and $b = 1$. The permutation $\sigma(n) = n + 1$ mapping \mathbb{Z} to \mathbb{Z} is in the set, but σ^{-1} is not in the set.

SECTION 6

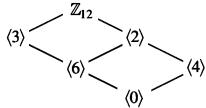
1. $q = 4, r = 6$ 3. $q = -5, r = 3$ 5. 8 7. 60

9. 4 11. 24 13. 2 15. 2 17. 6 19. 4

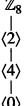
21. An infinite cyclic group

23. 75 25. 6 27. 12 29. 30

31.



33.



41.

a. T c. F e. T g. F i. T

43. $<Q, +>$ 45. There is none.47. $i, -i$ 49. $\frac{\sqrt{2}}{2}(\pm 1 \pm i)$ 63. $(p-1)p^{r-1}$

SECTION 7

1. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

3. \mathbb{Z}_{25} 5. $\dots, -24, -18, -12, -6, 0, 6, 12, 18, 24, \dots$ 7. $\{\iota, \rho^2, \rho^4, \rho^6, \mu, \mu\rho^2, \mu\rho^4, \mu\rho^6\}$

9. a. c b. e c. d

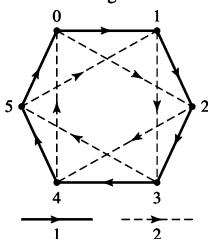
11.

	e	a	b	c	d	f
e	e	a	b	c	d	f
a	a	e	c	b	f	d
b	b	d	e	f	a	c
c	c	f	a	d	e	b
d	d	b	f	e	c	a
f	f	c	d	a	b	e

13. Choose a pair of generating directed arcs, call them $arc1$ and $arc2$, start at any vertex of the digraph, and see if the sequences $arc1, arc2$ and $arc2, arc1$ lead to the same vertex. (This corresponds to asking if the two corresponding group generators commute.) The group is commutative if and only if these two sequences lead to the same vertex for every pair of generating directed arcs.

15. It is not obvious, since a digraph of a cyclic group might be formed using a generating set of two or more elements, no one of which generates the group.

17.



19. a. Starting from any vertex a , every path through the graph that terminates at that same vertex a represents a product of generators or their inverses that is equal to the identity and thus gives a relation.

b. $a^4 = e, b^2 = e, (ab)^2 = e$

SECTION 8

1. Homomorphism
3. Homomorphism
5. Not a homomorphism
7. Homomorphism
9. Not a homomorphism
11. $\text{Ker}(\phi) = \langle 4 \rangle$.
13. $\text{Ker}(\phi) = \langle (5, 3) \rangle$.
15. $\{0, 0\}$
17. $\{1, 2, 5\}, \{3\}, \{4, 6\}$
19. $\{1, 2, 3, 4, 5\}, \{6\}, \{7, 8\}$
21. $\{2n \mid n \in \mathbb{Z}\}, \{2n + 1 \mid n \in \mathbb{Z}\}$
23. $(1, 8)(3, 6, 4)(5, 7) = (1, 8)(3, 4)(3, 6)(5, 7)$
25. $(1, 5, 4, 8)(2, 3)(6, 7) = (1, 8)(1, 4)(1, 5)(2, 3)(6, 7)$
31. a. F c. F e. F g. T i. T k. T

SECTION 9

1. Element	Order	Element	Order
$(0, 0)$	1	$(0, 2)$	2
$(1, 0)$	2	$(1, 2)$	2
$(0, 1)$	4	$(0, 3)$	4
$(1, 1)$	4	$(1, 3)$	4

The group is not cyclic

3. 2 5. 9 7. 60
9. $\{(0, 0), (0, 1)\}, \{(0, 0), (1, 0)\}, \{(0, 0), (1, 1)\}$
11. $\{(0, 0), (0, 1), (0, 2), (0, 3)\}$
 $\{(0, 0), (0, 2), (1, 0), (1, 2)\}$
 $\{(0, 0), (1, 1), (0, 2), (1, 3)\}$
13. $\mathbb{Z}_{20} \times \mathbb{Z}_3, \mathbb{Z}_{15} \times \mathbb{Z}_4, \mathbb{Z}_{12} \times \mathbb{Z}_5, \mathbb{Z}_5 \times \mathbb{Z}_3 \times \mathbb{Z}_4$
15. 12
17. 168
19. 180
21. $\mathbb{Z}_8, \mathbb{Z}_2 \times \mathbb{Z}_4, \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$
23. $\mathbb{Z}_{32}, \mathbb{Z}_2 \times \mathbb{Z}_{16}, \mathbb{Z}_4 \times \mathbb{Z}_8, \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_8, \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_4,$
 $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4, \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$
25. $\mathbb{Z}_9 \times \mathbb{Z}_{121}, \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_{121}, \mathbb{Z}_9 \times \mathbb{Z}_{11} \times \mathbb{Z}_{11}, \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_{11} \times \mathbb{Z}_{11}$

29. a.	n	2	3	4	5	6	7	8
	number of groups	2	3	5	7	11	15	22

- b. i) 225 ii) 225 iii) 110
31. a. It is abelian when the arrows on both n -gons have the same (clockwise or counterclockwise) direction.
b. $\mathbb{Z}_2 \times \mathbb{Z}_n$
c. When n is odd.
d. The dihedral group D_n .
33. \mathbb{Z}_2 is an example.
35. S_3 is an example.
37. The numbers are the same. 41. $\{-1, 1\}$

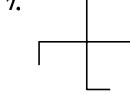
SECTION 10

1. $4\mathbb{Z} = \{\dots, -8, -4, 0, 4, 8, \dots\}$
 $1 + 4\mathbb{Z} = \{\dots, -7, -3, 1, 5, 9, \dots\}$
 $2 + 4\mathbb{Z} = \{\dots, -6, -2, 2, 6, 10, \dots\}$
 $3 + 4\mathbb{Z} = \{\dots, -5, -1, 3, 7, 11, \dots\}$
3. $\{0, 3, 6, 9, 12, 15\}, \{1, 4, 7, 10, 13, 16\}, \{2, 5, 8, 11, 14, 17\}$
5. $\langle 18 \rangle = \{0, 18\}, 1 + \langle 18 \rangle = \{1, 19\}, 2 + \langle 18 \rangle = \{2, 20\}, \dots, 17 + \langle 18 \rangle = \{17, 35\}$
7. $\{\iota, \mu\rho\}, \{\rho, \mu\rho^2\}, \{\rho^2, \mu\rho^3\}, \{\rho^3, \mu\}$ They are different.
9. $\{\iota, \rho^2\}, \{\rho, \rho^3\}, \{\mu, \mu\rho^2\}, \{\mu\rho, \mu\rho^3\}$
11. 4
13. 12
15. 24
21. $2\mathbb{Z} \leq \mathbb{Z}$ has only 2 cosets.
23. $G = \mathbb{Z}_2$, subgroup $H = \mathbb{Z}_2$.
25. Impossible. The number of cells must divide the order of the group, and 12 does not divide 6.

SECTION 11

1. a. The only isometries of \mathbb{R} leaving a number c fixed are the reflection through c that carries $c+x$ to $c-x$ for all $x \in \mathbb{R}$, and the identity map.
b. The isometries of \mathbb{R}^2 that leave a point P fixed are the rotations about P through any angle θ where $0 \leq \theta < 360^\circ$ and the reflections across any axis that passes through P .
c. The only isometries of \mathbb{R} that carry a line segment into itself are the reflection through the midpoint of the line segment (see the answer to part (a)) and the identity map.
d. The isometries of \mathbb{R}^2 that carry a line segment into itself are a rotation of 180° about the midpoint of the line segment, a reflection in the axis containing the line segment, a reflection in the axis perpendicular to the line segment at its midpoint, and the identity map.
e. The isometries of \mathbb{R}^3 that carry a line segment into itself include rotations through any angle about an axis that contains the line segment, reflections across any plane that contains the line segment, and reflection across the plane perpendicular to the line segment at its midpoint.
3.

	τ	ρ	μ	γ
τ	τ	ρ	$\mu\gamma$	$\mu\gamma$
ρ	ρ	$\rho\tau$	$\mu\gamma$	$\mu\gamma$
μ	$\mu\gamma$	$\mu\gamma$	$\tau\rho$	$\tau\rho$
γ	$\mu\gamma$	$\mu\gamma$	$\tau\rho$	$\tau\rho$

5. 
7. 

9. *Translation:* order ∞
Rotation: order any $n \geq 2$ or ∞
Reflection: order 2
Glide reflection: order ∞
11. Rotations 13. Only the identity and reflections.
17. Yes. The product of two translations is a translation and the inverse of a translation is a translation.
19. Yes. There is only one reflection μ across one particular line L , and μ^2 is the identity, so we have a group isomorphic to \mathbb{Z}_2 .

21. Only reflections and rotations (and the identity) because translations and glide reflections do not have finite order in the group of all plane isometries.
25. a. No b. No c. Yes d. No e. D_∞
 27. a. Yes b. No c. No d. No e. D_∞
 29. a. No b. No c. No d. Yes e. \mathbb{Z}
 31. a. Yes. $90^\circ, 180^\circ$ b. Yes c. No
 33. a. No b. No c. No
 35. a. Yes. 180° b. Yes c. No
 37. a. Yes. 120° b. Yes c. No
 39. a. Yes. 120° b. No c. No d. $(1, 0), (1, \sqrt{3})$

SECTION 12

1. 3 3. 4 5. 3 7. 1
 9. 4 11. 3 13. 5 15. 1
21. a. When working with a factor group G/H , you would let a and b be elements of G , not elements of G/H . The student probably does not understand what elements of G/H look like and can write nothing sensible concerning them.
 b. We must show that G/H is abelian. Let aH and bH be two elements of G/H .
23. a. T c. T e. T g. T i. T
35. Example: Let $G = N = S_3$, and let $H = \{p_0, \mu_1\}$. Then N is normal in G , but $H \cap N = H$ is not normal in G .

SECTION 13

1. \mathbb{Z}_2 3. \mathbb{Z}_4 5. $\mathbb{Z}_4 \times \mathbb{Z}_8$ 7. $\mathbb{Z} \times \mathbb{Z}_2$ 9. $\mathbb{Z}_3 \times \mathbb{Z} \times \mathbb{Z}_4$
 11. $\mathbb{Z}_2 \times \mathbb{Z}$ 13. $\mathbb{Z} \times \mathbb{Z}_2$
 15. $Z(D_4) = C = \{\iota, \rho^2\}$
 17. $Z(S_3 \times D_4) = \{(\iota, \iota), (\iota, \rho)\}, C = A_3 \times \{\iota, \rho\}$.
 21. a. T c. F e. F g. F i. T
 23. $\{f \in F^* | f(0) = 1\}$
 25. Yes. Let $f(x) = 1$ for $x \geq 0$ and $f(x) = -1$ for $x < 0$. Then $f(x) \cdot f(x) = 1$ for all x , so $f^2 \in K^*$ but f is not in K^* . Thus $f K^*$ has order 2 in F^*/K^* .
 27. U
 29. The multiplicative group U of complex numbers of absolute value 1
 31. Let $G = \mathbb{Z}_2 \times \mathbb{Z}_4$. Then $H = \langle(1, 0)\rangle$ is isomorphic to $K = \langle(0, 2)\rangle$, but G/H is isomorphic to \mathbb{Z}_4 while G/K is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$.
 33. a. $\{e\}$ b. The whole group

SECTION 14

1. $X_\iota = X, X_\rho = \{C\}, X_{\rho^2} = \{m_1, m_2, d_1, d_2, C\}, X_{\rho^3} = \{C\},$
 $X_{\mu_\rho} = \{s_1, s_3, m_1, m_2, C, P_1, P_3\}, X_{\mu_\rho^2} = \{s_2, s_0, m_1, m_2, C, P_2, P_0\},$
 $X_\mu = \{2, 0, d_1, d_2, C\}, X_{\mu_\rho^3} = \{1, 3, d_1, d_2, C\}.$
 3. $\{(1, 2, 3, 0), \{s_1, s_2, s_3, s_0\}, \{m_1, m_2\}, \{d_1, d_2\}, \{C\}, \{P_1, P_2, P_3, P_0\}$
 5. $\{\iota\}$
 7. There are three orbits: $\{(\mu), (\mu\rho^2)\}, \{(\mu\rho), (\mu\rho^3)\}, \{(\rho^2)\}$
 9. $3, 3; 3, 1, 1; 1, 1, 1, 1, 1$
 11. $8, 2; 8, 1, 1; 4, 4, 2; 4, 4, 1, 1; 4, 2, 2, 2; 4, 2, 2, 1, 1; 4, 2, 1, 1, 1, 1; 4, 1, 1, 1, 1, 1, 1; 2, 2, 2, 2, 2; 2, 2, 2, 2, 1, 1;$
 $2, 2, 2, 1, 1, 1; 2, 2, 1, 1, 1, 1, 1, 1; 2, 1, 1, 1, 1, 1, 1, 1, 1; 1, 1, 1, 1, 1, 1, 1, 1, 1$
 15. A transitive G -set has just one orbit.
 17. a. $\{s_1, s_2, s_3, s_0\}$ and $\{P_1, P_2, P_3, P_0\}$
 b. The set of points on the circle with center at the origin and passing through P
 c. The cyclic subgroup (2π) of $G = \mathbb{R}$

25. a. $K = g_0 H g_0^{-1}$.

b. *Conjecture:* H and K should be conjugate subgroups of G .

27.

	X	Y	Z	
	a	a	b	a
	a	a	b	b
0	a	a	b	a
1	a	b	a	b
2	a	a	b	c
3	a	b	a	a
4	a	a	b	b
5	a	b	a	c

 There are four of them: X, Y, Z , and \mathbb{Z}_6 .

	X	Y	Z	
	a	a	b	a
	a	a	b	b
0	a	a	b	a
1	a	b	a	b
2	a	a	b	c
3	a	b	a	a
4	a	a	b	b
5	a	b	a	c

SECTION 15

1. 5 3. 2 5. 11,712
 7. a. 45 b. 231
 9. a. 90 b. 6,426

SECTION 16

1. a. $K = \{0, 3, 6, 9\}$.
 b. $0 + K = \{0, 3, 6, 9\}, 1 + K = \{1, 4, 7, 10\}, 2 + K = \{2, 5, 8, 11\}$.
 c. $\mu(0 + K) = 0, \mu(1 + K) = 2, \mu(2 + K) = 1$.
 3. a. $HN = \{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22\}, H \cap N = \{0, 12\}$.
 b. $0 + N = \{0, 6, 12, 18\}, 2 + N = \{2, 8, 14, 20\}, 4 + N = \{4, 10, 16, 22\}$.
 c. $0 + (H \cap N) = \{0, 12\}, 4 + (H \cap N) = \{4, 16\}, 8 + (H \cap N) = \{8, 20\}$.
 d. $\mu(0 + (H \cap N)) = 0 + N = \{0, 6, 12, 18\}, \mu(4 + (H \cap N)) = 4 + N = \{4, 10, 16, 20\}$,
 $\mu(8 + (H \cap N)) = 8 + N = \{2, 8, 13, 20\}$.
 5. a. $0 + H = \{0, 4, 8, 12, 16, 20\}, 1 + H = \{1, 5, 9, 13, 17, 21\}$,
 $2 + H = \{2, 6, 10, 14, 18, 22\}, 3 + H = \{3, 7, 11, 15, 19, 23\}$.
 b. $0 + K = \{0, 8, 16\}, 1 + K = \{1, 9, 17\}, 2 + K = \{2, 10, 18\}$,
 $3 + K = \{3, 11, 19\}$,
 $4 + K = \{4, 12, 20\}, 5 + K = \{5, 13, 21\}, 6 + K = \{6, 14, 22\}$,
 $7 + K = \{7, 15, 23\}$.
 c. $0 + K = \{0, 8, 16\}, 4 + K = \{4, 12, 20\}$.
 d. $(0 + K) + (H/K) = H/K = \{0 + K, 4 + K\} = \{\{0, 8, 16\}, 4, 12, 20\}$
 $(1 + K) + (H/K) = \{1 + K, 5 + K\} = \{\{1, 9, 17\}, \{5, 13, 21\}\}$
 $(2 + K) + (H/K) = \{2 + K, 6 + K\} = \{\{2, 10, 18\}, \{6, 14, 22\}\}$
 $(3 + K) + (H/K) = \{3 + K, 7 + K\} = \{\{3, 11, 19\}, \{7, 15, 23\}\}$.
 e. $\phi(\{0, 8, 16\}, \{4, 12, 20\}) = \{0, 4, 8, 12, 16, 20\}, \phi(\{1, 9, 17\}, \{5, 13, 21\})$
 $= \{1, 5, 9, 13, 17, 21\}, \phi(\{2, 10, 18\}, \{6, 14, 22\}) = \{2, 6, 10, 14, 18, 22\}$,
 $\phi(\{3, 11, 19\}, \{7, 15, 23\}) = \{3, 7, 11, 15, 19, 23\}$

SECTION 17

1. 3 3. 1, 3
 5. The Sylow 3-subgroups are $\langle (1, 2, 3) \rangle, \langle (1, 2, 4) \rangle, \langle (1, 3, 4) \rangle$, and $\langle (2, 3, 4) \rangle$. Also $(3, 4)\langle (1, 2, 3) \rangle(3, 4) = \langle (1, 2, 4) \rangle$, etc.
 7. 1, 2, 3, 4, 5, 7, 9, 11, 13, 15, 17, 19.
 13. a. T c. F e. T g. T i. F

SECTION 18

1. The refinements $\{0\} < 250\mathbb{Z} < 10\mathbb{Z} < \mathbb{Z}$ of $\{0\} < 10\mathbb{Z} < \mathbb{Z}$ and $\{0\} < 250\mathbb{Z} < 25\mathbb{Z} < \mathbb{Z}$ of $0 < 25\mathbb{Z} < \mathbb{Z}$ are isomorphic.
3. $\{0\} < \langle 27 \rangle < \langle 9 \rangle < \mathbb{Z}_{54}$ and $\{0\} < \langle 18 \rangle < \langle 2 \rangle < \mathbb{Z}_{54}$
5. The refinements $\{(0, 0)\} < (4800\mathbb{Z}) \times \mathbb{Z} < (240\mathbb{Z}) \times \mathbb{Z} < (60\mathbb{Z}) \times \mathbb{Z} < (10\mathbb{Z}) \times \mathbb{Z} < \mathbb{Z} \times \mathbb{Z}$ of the first series and $\{(0, 0)\} < \mathbb{Z} \times (4800\mathbb{Z}) < \mathbb{Z} \times (480\mathbb{Z}) < \mathbb{Z} \times (80\mathbb{Z}) < \mathbb{Z} \times (20\mathbb{Z}) < \mathbb{Z} \times \mathbb{Z}$ of the second series are isomorphic refinements.
7. $\{0\} < \langle 16 \rangle < \langle 8 \rangle < \langle 4 \rangle < \langle 2 \rangle < \mathbb{Z}_{48}$
 $\{0\} < \langle 24 \rangle < \langle 8 \rangle < \langle 4 \rangle < \langle 2 \rangle < \mathbb{Z}_{48}$
 $\{0\} < \langle 24 \rangle < \langle 12 \rangle < \langle 4 \rangle < \langle 2 \rangle < \mathbb{Z}_{48}$
 $\{0\} < \langle 24 \rangle < \langle 12 \rangle < \langle 6 \rangle < \langle 2 \rangle < \mathbb{Z}_{48}$
 $\{0\} < \langle 24 \rangle < \langle 12 \rangle < \langle 6 \rangle < \langle 3 \rangle < \mathbb{Z}_{48}$
9. $\{(t, 0)\} < A_3 \times \{0\} < S_3 \times \{0\} < S_3 \times \mathbb{Z}_2$
 $\{(t, 0)\} < \{t\} \times \mathbb{Z}_2 < A_3 \times \mathbb{Z}_2 < S_3 \times \mathbb{Z}_2$
 $\{(t, 0)\} < A_3 \times \{0\} < A_3 \times \mathbb{Z}_2 < S_3 \times \mathbb{Z}_2$
11. $\{t\} \times \mathbb{Z}_4$ 13. $\{t\} \times \mathbb{Z}_4 \leq \{t\} \times \mathbb{Z}_4 \leq \{t\} \times \mathbb{Z}_4 \leq \cdots$
17. a. T c. T e. F g. F i. T
- i. The Jordan-Hölder theorem applied to the groups \mathbb{Z}_n implies the Fundamental Theorem of Arithmetic.
19. Yes. $\{t\} < \langle t, \rho \rangle < \langle t, \rho, \rho^2, \rho^3 \rangle < D_4$ is a composition (actually a principal) series and all factor groups are isomorphic to \mathbb{Z}_2 and are thus abelian.
21. Chain (3) Chain (4)

$$\begin{array}{ll} \{0\} \leq \langle 12 \rangle \leq \langle 12 \rangle \leq \langle 12 \rangle & \{0\} \leq \langle 12 \rangle < \langle 12 \rangle \leq \langle 6 \rangle \\ \leq \langle 12 \rangle \leq \langle 12 \rangle \leq \langle 4 \rangle & \leq \langle 6 \rangle \leq \langle 6 \rangle \leq \langle 3 \rangle \\ \leq \langle 2 \rangle \leq \mathbb{Z}_{24} \leq \mathbb{Z}_{24} & \leq \langle 3 \rangle \leq \mathbb{Z}_{24} \leq \mathbb{Z}_{24} \end{array}$$

Isomorphisms

$$\begin{aligned} \langle 12 \rangle / \{0\} &\simeq \langle 12 \rangle / \{0\} \simeq \mathbb{Z}_2, & \langle 12 \rangle / \langle 12 \rangle &\simeq \langle 6 \rangle / \langle 6 \rangle \simeq \{0\}, \\ \langle 12 \rangle / \langle 12 \rangle &\simeq \langle 3 \rangle / \langle 3 \rangle \simeq \{0\}, & \langle 12 \rangle / \langle 12 \rangle &\simeq \langle 12 \rangle / \langle 12 \rangle \simeq \{0\}, \\ \langle 12 \rangle / \langle 12 \rangle &\simeq \langle 6 \rangle / \langle 6 \rangle \simeq \{0\}, & \langle 4 \rangle / \langle 12 \rangle &\simeq \mathbb{Z}_{24} / \langle 3 \rangle \simeq \mathbb{Z}_3 \\ \langle 2 \rangle / \langle 4 \rangle &\simeq \langle 6 \rangle / \langle 12 \rangle \simeq \mathbb{Z}_2, & \mathbb{Z}_{24} / \langle 2 \rangle &\simeq \langle 3 \rangle / \langle 6 \rangle \simeq \mathbb{Z}_2 \\ \mathbb{Z}_{24} / \mathbb{Z}_{24} &\simeq \mathbb{Z}_{24} / \mathbb{Z}_{24} \simeq \{0\} & & \end{aligned}$$

SECTION 19

1. $\{(1, 1, 1), (1, 2, 1), (1, 1, 2)\}$
3. No. $n(2, 1) + m(4, 1)$ can never yield an odd number for first coordinate.
7. $2\mathbb{Z} < \mathbb{Z}$, rank $r = 1$

SECTION 20

1. a. $a^2b^2a^3c^3b^{-2}, b^2c^{-3}a^{-3}b^{-2}a^{-2}$ b. $a^{-1}b^3a^4c^6a^{-1}, ac^{-6}a^{-4}b^{-3}a$
3. a. 16 b. 36 c. 36
5. a. 16 b. 36 c. 18
11. a. Partial answer: $\{1\}$ is a basis for \mathbb{Z}_4 . c. Yes
13. c. A blob group on S is isomorphic to the free group $F[S]$ on S .

SECTION 21

1. $(a : a^4 = 1); (a, b : a^4 = 1, b = a^2); (a, b, c : a = 1, b^4 = 1, c = 1)$. (Other answers are possible.)
 3. Octic group:

	1	a	a^2	a^3	b	ab	a^2b	a^3b
1	1	a	a^2	a^3	b	ab	a^2b	a^3b
a	a	a^2	a^3	1	ab	a^2b	a^3b	b
a^2	a^2	a^3	1	a	a^2b	a^3b	b	ab
a^3	a^3	1	a	a^2	a^3b	b	ab	a^2b
b	b	a^3b	a^2b	ab	1	a^3	a^2	a
ab	ab	b	a^3b	a^2b	a	1	a^3	a^2
a^2b	a^2b	ab	b	a^3b	a^2	a	1	a^3
a^3b	a^3b	a^2b	ab	b	a^3	a^2	a	1

Quaternion group: The same as the table for the octic group except that the 16 entries in the lower right corner are

a^2	a	1	a^3
a^3	a^2	a	1
1	a^3	a^2	a
a	1	a^3	a^2

5. \mathbb{Z}_{21} . $(a, b : a^7 = 1, b^3 = 1, ba = a^2b)$

SECTION 22

1. 0 3. 1 5. (1, 6)
 7. Commutative ring, no unity, not a field
 9. Commutative ring with unity, not a field
 11. Commutative ring with unity, not a field
 13. No. $\{ri \mid r \in \mathbb{R}\}$ is not closed under multiplication.
 15. $(1, 1), (1, -1), (-1, 1), (-1, -1)$
 17. All nonzero $q \in \mathbb{Q}$ 19. 1, 3
 21. Let $\mathbb{R} = \mathbb{Z}$ with unity 1 and $\mathbb{R}' = \mathbb{Z} \times \mathbb{Z}$ with unity $1' = (1, 1)$. Let $\phi : R \rightarrow R'$ be defined by $\phi(n) = (n, 0)$. Then $\phi(1) = (1, 0) \neq 1'$.
 23. $\phi_1 : \mathbb{Z} \rightarrow \mathbb{Z}$ where $\phi_1(n) = 0$, $\phi_2 : \mathbb{Z} \rightarrow \mathbb{Z}$ where $\phi_2(n) = n$
 25. $\phi_1 : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ where $\phi_1(n, m) = 0$, $\phi_2 : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ where $\phi_2(n, m) = n$,
 $\phi_3 : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ where $\phi_3(n, m) = m$
 27. The reasoning is not correct since a product $(X - I_3)(X + I_3)$ of two matrices may be the zero matrix 0 without having either matrix be 0. Counterexample:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}^2 = I_3.$$

29. 2, 10. In the ring \mathbb{Z}_{14} there are nonzero elements 2 and 7 which when multiplied give 0. This is not the case in \mathbb{Z}_{13} .

33. $a = 2, b = 3$ in \mathbb{Z}_6
 35. a. T c. F e. T g. T i. T

SECTION 23

1. 0, 3, 5, 8, 9, 11 3. No solutions 5. 0 7. 0 9. 12
 11. 1, 5 are units; 2, 3, 4 are 0 divisors.
 13. 1, 2, 4, 7, 8, 11, 13, 14 are units; 3, 5, 6, 9, 10, 12 are 0 divisors.
 15. (1, 1), (1, 2), (2, 1), (2, 2) are units; (0, 1), (0, 2), (1, 0), (2, 0) are 0 divisors.
 17. $a^4 + 2a^2b^2 + b^4$ 19. $a^6 + 2a^3b^3 + b^6$
 23. a. F c. F e. T g. F i. F
 25. 1. $\text{Det}(A) = 0$. 2. The column vectors of A are dependent.
 3. The row vectors of A are dependent. 4. Zero is an eigenvalue of A .
 5. A is not invertible.

SECTION 24

1. 3 or 5 3. Any of 3, 5, 6, 7, 10, 11, 12, or 14. 5. 2
 7. $\varphi(1) = 1$ $\varphi(7) = 6$ $\varphi(13) = 12$ $\varphi(19) = 18$ $\varphi(25) = 20$
 $\varphi(2) = 1$ $\varphi(8) = 4$ $\varphi(14) = 6$ $\varphi(20) = 8$ $\varphi(26) = 12$
 $\varphi(3) = 2$ $\varphi(9) = 6$ $\varphi(15) = 8$ $\varphi(21) = 12$ $\varphi(27) = 18$
 $\varphi(4) = 2$ $\varphi(10) = 4$ $\varphi(16) = 8$ $\varphi(22) = 10$ $\varphi(28) = 12$
 $\varphi(5) = 4$ $\varphi(11) = 10$ $\varphi(17) = 16$ $\varphi(23) = 22$ $\varphi(29) = 28$
 $\varphi(6) = 2$ $\varphi(12) = 4$ $\varphi(18) = 6$ $\varphi(24) = 8$ $\varphi(30) = 8$
 9. $(p - 1)(q - 1)$ 11. $1 + 4\mathbb{Z}, 3 + 4\mathbb{Z}$ 13. No solutions
 15. No solutions
 17. $3 + 65\mathbb{Z}, 16 + 65\mathbb{Z}, 29 + 65\mathbb{Z}, 42 + 65\mathbb{Z}, 55 + 65\mathbb{Z}$
 19. 1 21. 9
 23. a. F c. T e. T g. F i. F

SECTION 25

1. $n = pq = 15$, $(p - 1)(q - 1) = 8$, so the pairs are (3, 3), (5, 5)
 3. $n = pq = 33$, $(p - 1)(q - 1) = 20$, so the pairs are (3, 7), (7, 3), (9, 9), (11, 11), (13, 17), (17, 13)
 5. $s = 77$
 7. a. $y = 64$ b. $r = 13$ c. $64^{13} \equiv 25 \pmod{143}$
 9. Private key is $p = 257$, $q = 359$, $n = 92263$, $r = 1493$. Public key is $n = 92263$ and $s = 9085$.

SECTION 26

1. $\{q_1 + q_2i \mid q_1, q_2 \in \mathbb{Q}\}$
 15. It is isomorphic to the ring D of all rational numbers that can be expressed as a quotient of integers with denominator some power of 2.
 17. It runs into trouble when we try to prove the transitive property in the proof of Lemma 5.4.2, for multiplicative cancellation may not hold. For $R = \mathbb{Z}_6$ and $T = \{1, 2, 4\}$ we have $(1, 2) \sim (2, 4)$ since $(1)(4) = (2)(2) = 4$ and $(2, 4) \sim (2, 1)$ since $(2)(1) = (4)(2)$ in \mathbb{Z}_6 . However, $(1, 2)$ is not equivalent to $(2, 1)$ because $(1)(1) \neq (2)(2)$ in \mathbb{Z}_6 .

SECTION 27

1. $f(x) + g(x) = 2x^2 + 5$, $f(x)g(x) = 6x^2 + 4x + 6$
 3. $f(x) + g(x) = 5x^2 + 5x + 1$, $f(x)g(x) = x^3 + 5x$
 5. 16 7. 7 9. 2 11. 0 13. 2, 3 15. 0, 2, 4
 17. 0, 1, 2, 3
 21. $0, x - 5, 2x - 10, x^2 - 25, x^2 - 5x, x^4 - 5x^3$. (Other answers are possible.)
 23. a. T c. T e. F g. T i. T
 25. a. They are the units of D . b. 1, -1 c. 1, 2, 3, 4, 5, 6
 27. b. F c. $F[x]$ 31. a. 4, 27 b. $\mathbb{Z}_2 \times \mathbb{Z}_2, \mathbb{Z}_3 \times \mathbb{Z}_3$

SECTION 28

1. $q(x) = x^4 + x^3 + x^2 + x - 2, r(x) = 4x + 3$
3. $q(x) = 6x^4 + 7x^3 + 2x^2 - x + 2, r(x) = 4$
5. 2, 3 7. 3, 10, 5, 11, 14, 7, 12, 6
9. $(x - 1)(x + 1)(x - 2)(x + 2)$
11. $(x - 3)(x + 3)(2x + 3)$
13. Yes. It is of degree 3 with no zeros in \mathbb{Z}_5 .
 $2x^3 + x^2 + 2x + 2$
15. Partial answer: $g(x)$ is irreducible over \mathbb{R} , but it is not irreducible over \mathbb{C} .
19. Yes. $p = 3$ 21. Yes. $p = 5$
25. a. T c. T e. T g. T i. T
27. $x^2 + x + 1$
29. $x^2 + 1, x^2 + x + 2, x^2 + 2x + 2, 2x^2 + 2, 2x^2 + x + 1, 2x^2 + 2x + 1$
31. $p(p - 1)^2/2$

SECTION 29

1. 32
3. \mathbb{Z}_2^5
5. a. $\{(0, 0)\}, \{(0, 0), (1, 1)\}, \mathbb{Z}_2^2$ b. $\{(0, 0, 0), \{(0, 0, 0), (1, 1, 1)\}, \mathbb{Z}_2^3$
c. $\{(0, 0, 0, 0)\}, \{(0, 0, 0, 0), (1, 1, 1, 1)\}, \{(0, 0, 0, 0), (0, 1, 0, 1), (1, 0, 1, 0), (1, 1, 1, 1)\}, \mathbb{Z}_2^4$
7. a. $x^7 + 1 = (x^3 + x + 1)(x^4 + x^2 + x + 1)$ b. C consists of the cyclic shifts of $x^3 + x + 1, x^4 + x^3 + x^2 + 1$ together with 0 and $x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$. c. A single bit error can be detected and corrected. d. A two-bit error can be detected, but not corrected.
9. a. Use long division to verify that $(x^3 + 1)(x^6 + x^3 + 1) = x^9 + 1$.
b. $C = \{x^6 + x^3 + 1, x^7 + x^4 + x, x^8 + x^5 + x^2, x^7 + x^6 + x^4 + x^3 + x + 1, x^8 + x^7 + x^5 + x^4 + x^2 + x, x^8 + x^6 + x^5 + x^3 + x^2 + 1, x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x^1 + 1, 0\}$
c. The minimal weight among the nonzero words is 3, so the minimum distance between two different code words is 3. So C detects and corrects a one bit error.
d. A two-bit error would be detected, but it could not be corrected.
11. $x^9 + 1 = (x + 1)(x^2 + x + 1)(x^6 + x^3 + 1)$, so the polynomials $x + 1, x^2 + x + 1, x^6 + x^3 + 1, (x + 1)(x^2 + x + 1)$, $(x + 1)(x^6 + x^3 + 1)$, and $(x^2 + x + 1)(x^6 + x^3 + 1)$ all generated cyclic codes with code word length 9.

SECTION 30

1. There are just nine possibilities:
 $\phi(1, 0) = (1, 0)$ while $\phi(0, 1) = (0, 0)$ or $(0, 1)$,
 $\phi(1, 0) = (0, 1)$ while $\phi(0, 1) = (0, 0)$ or $(1, 0)$,
 $\phi(1, 0) = (1, 1)$ while $\phi(0, 1) = (0, 0)$, and
 $\phi(1, 0) = (0, 0)$ while $\phi(0, 1) = (0, 0), (1, 0), (0, 1)$, or $(1, 1)$.
3. $\langle 0 \rangle = \{0\}, \mathbb{Z}_{12}/\langle 0 \rangle \cong \mathbb{Z}_{12}$
 $\langle 1 \rangle = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}, \mathbb{Z}_{12}/\langle 1 \rangle \cong \{0\}$
 $\langle 2 \rangle = \{0, 2, 4, 6, 8, 10\}, \mathbb{Z}_{12}/\langle 2 \rangle \cong \mathbb{Z}_2$
 $\langle 3 \rangle = \{0, 3, 6, 9\}, \mathbb{Z}_{12}/\langle 3 \rangle \cong \mathbb{Z}_3$
 $\langle 4 \rangle = \{0, 4, 8\}, \mathbb{Z}_{12}/\langle 4 \rangle \cong \mathbb{Z}_4$
 $\langle 6 \rangle = \{0, 6\}, \mathbb{Z}_{12}/\langle 6 \rangle \cong \mathbb{Z}_6$
9. Let $\phi : \mathbb{Z} \rightarrow \mathbb{Z} \times \mathbb{Z}$ be given by $\phi(n) = (n, 0)$ for $n \in \mathbb{Z}$.
11. R/R and $R/\{0\}$ are not of real interest because R/R is the ring containing only the zero element, and $R/\{0\}$ is isomorphic to R .
13. \mathbb{Z} is an integral domain. $\mathbb{Z}/4\mathbb{Z}$ is isomorphic to \mathbb{Z}_4 , which has a divisor 2 of 0.
15. $\{(n, n) \mid n \in \mathbb{Z}\}$. (Other answers are possible.)
31. The nilradical of \mathbb{Z}_{12} is $\{0, 6\}$. The nilradical of \mathbb{Z} is $\{0\}$ and the nilradical of \mathbb{Z}_{32} is $\{0, 2, 4, 6, 8, \dots, 30\}$.

- 35.** **a.** Let $R = \mathbb{Z}$ and let $N = 4\mathbb{Z}$. Then $\sqrt{N} = 2\mathbb{Z} \neq 4\mathbb{Z}$
b. Let $R = \mathbb{Z}$ and let $N = 2\mathbb{Z}$. Then $\sqrt{N} = N$.

SECTION 31

1. $\{0, 2, 4\}$ and $\{0, 3\}$ are both prime and maximal.
3. $\{(0, 0), (1, 0)\}$ and $\{(0, 0), (0, 1)\}$ are both prime and maximal.
5. 1 7. 2 9. 1, 4 15. $2\mathbb{Z} \times \mathbb{Z}$ 17. $4\mathbb{Z} \times \{0\}$
19. Yes. $x^2 - 6x + 6$ is irreducible over \mathbb{Q} by Eisenstein with $p = 2$.
27. Yes. $\mathbb{Z}_2 \times \mathbb{Z}_3$
29. No. Enlarging the domain to a field of quotients, you would have to have a field containing two different prime fields \mathbb{Z}_p and \mathbb{Z}_q , which is impossible.

SECTION 32

1. $1e + 0a + 3b$ 3. $2e + 2a + 2b$ 5. j 7. $(1/50)j - (3/50)k$
9. \mathbb{R}^* , that is, $\{a_1 + 0i + 0j + 0k \mid a_1 \in \mathbb{R}, a_1 \neq 0\}$
11. **a.** F **c.** F **e.** F **g.** T **i.** T
c. If $|A| = 1$, then $\text{End}(A) = \{0\}$. **e.** $0 \in \text{End}(A)$ is not in $\text{Iso}(A)$.
19. **a.** $K = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.
b. Denoting by B the matrix with coefficient b and by C the matrix with coefficient c and the 2×2 identity matrix by I , we must check that

$$B^2 = -I, C^2 = -I, K^2 = -I,$$

$$CK = B, KB = C, CB = -K, KC = -B, \text{ and } BK = -C.$$

- c.** We should check that ϕ is one-to-one.

SECTION 33

1. $\{(0, 1), (1, 0)\}, \{(1, 1), (-1, 1)\}, \{(2, 1), (1, 2)\}$. (Other answers are possible.)
3. No. $2(-1, 1, 2) - 4(2, -3, 1) + (10, -14, 0) = (0, 0, 0)$
5. $1, \sqrt{2}$ (answers can vary)
7. Infinite Dimensional
9. Infinite Dimensional
15. **a.** T **c.** T **e.** F **g.** F **i.** F
17. **a.** The subspace of V generated by S is the intersection of all subspaces of V containing S .
19. *Partial answer:* A basis for F^n is

$$\{(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)\}$$

where 1 is the multiplicative identity of F .

- 25.** **a.** A homomorphism
b. *Partial answer:* The kernel (or nullspace) of ϕ is $\{\alpha \in V \mid \phi(\alpha) = 0\}$.
c. ϕ is an isomorphism of V with V' if $\text{Ker}(\phi) = \{0\}$ and ϕ maps V onto V' .

SECTION 34

1. Yes 3. No 5. No. 7. Yes
9. In $\mathbb{Z}[x]$: only $2x - 7, -2x + 7$
In $\mathbb{Q}[x]$: $4x - 14, x - \frac{7}{2}, 6x - 21, -8x + 28$
In $\mathbb{Z}_{11}[x]$: $2x - 7, 10x - 2, 6x + 1, 3x - 5, 5x - 1$
11. 26, -26 13. 198, -198
15. It is already “primitive” because every nonzero element of \mathbb{Q} is a unit. Indeed $18ax^2 - 12ax + 48a$ is primitive for all $a \in \mathbb{Q}, a \neq 0$.

17. $2ax^2 - 3ax + 6a$ is primitive for all $a \neq 0$ in \mathbb{Z}_7 because every such element a is a unit in \mathbb{Z}_7 .
 21. a. T c. T e. T g. F i. F
 i. Either p or one of its associates must appear in every factorization *into irreducibles*.
 23. $2x + 4$ is irreducible in $\mathbb{Q}[x]$ but not in $\mathbb{Z}[x]$.
 31. Partial answer: $x^3 - y^3 = (x - y)(x^2 + xy + y^2)$

SECTION 35

1. Yes 3. No. (1) is violated. 5. Yes
 7. 61 9. $x^3 + 2x - 1$ 11. 66
 13. a. T c. T e. T g. T i. T
 23. Partial answer: The equation $ax = b$ has a solution in \mathbb{Z}_n for nonzero $a, b \in \mathbb{Z}_n$ if and only if the positive gcd of a and n divides b .

SECTION 36

1. $5 = (1 + 2i)(1 - 2i)$ 3. $4 + 3i = (1 + 2i)(2 - i)$
 5. $6 = (2)(3) = (-1 + \sqrt{-5})(-1 - \sqrt{-5})$ 7. $7 - i$
 15. c. i) order 9, characteristic 3 ii) order 2, characteristic 2
 iii) order 5, characteristic 5

SECTION 37

1. $\{x, y\}$
 3. $\{x + 4, y - 5\}$
 5. By multiplying the first polynomial by -2 and adding to the second polynomial, we have $I = \langle x + y + z, -y + z - 4 \rangle$.
 The algebraic variety is $\{4 - 2z, z - 4, z \mid z \in \mathbb{R}\}$ which is a line through $(4, -4, 0)$.
 7. After two careful long divisions, $I = \langle x^2 + x - 2 \rangle$. The algebraic variety is $\{1, -2\}$.
 9. F^2
 11. a. \emptyset , b. $\{i, -i\}$
 13. a. T c. T e. T g. T i. T

SECTION 38

1. $-3x^3 + 7x^2y^2z - 5x^2yz^3 + 2xy^3z^5$
 3. $2x^2yz^2 - 2xy^2z^2 - 7x + 3y + 10z^3$
 5. $2z^3y^3x - 5z^3yx^2 + 7zy^2x^2 - 3x^3$
 7. $10z^3 - 2z^2y^2 + 2z^2yx^2 + 3y - 7x$
 9. $1 < z < y < x < z^2 < yz < y^2 < xz < xy < x^2 < z^3 < yz^2 < y^2z < y^3 < xz^2 < xyz <$
 $xy^2 < x^2z < x^2y < x^3 < \dots$
 11. $3y^2z^5 - 8z^7 + 5y^3z^3 - 4x$ 13. $3yz^3 - 8xy - 4xz + 2yz + 38$
 15. $\langle y^5 + y^3, y^3 + z, x - y^4 \rangle$ 17. $\langle y^2z^3 + 3, -3y - 2z, y^2z^2 + 3 \rangle$
 19. $\{1\}$ 21. $\{x - 1\}$
 23. $\{2x + y - 5, y^2 - 9y + 18\}$
 The algebraic variety is $\{(1, 3), (-\frac{1}{2}, 6)\}$.
 25. $\{x + y, y^3 - y + 1\}$
 The algebraic variety consists of one point $(a, -a)$ where $a \approx 1.3247$.
 27. a. F c. T e. T g. T i. F
 29. Any order with d_1 and d_2 (in either order) the largest.

SECTION 39

1. $x^2 - 2x - 1$
3. $x^2 - 2x + 2$
5. $x^{12} + 3x^8 - 4x^6 + 3x^4 + 12x^2 + 5$
7. $\text{Irr}(\alpha, \mathbb{Q}) = x^4 - \frac{2}{3}x^2 - \frac{62}{9}; \deg(\alpha, \mathbb{Q}) = 4$
9. Algebraic, $\deg(\alpha, F) = 2$
11. Transcendental
13. Algebraic, $\deg(\alpha, F) = 2$
15. Algebraic, $\deg(\alpha, F) = 1$
17. $x^2 + x + 1 = (x - \alpha)(x + 1 + \alpha)$
23. a. T c. T e. F g. F i. F
25. b. $x^3 + x^2 + 1 = (x - \alpha)(x - \alpha^2)[x - (1 + \alpha + \alpha^2)]$
27. The polynomial $\text{irr}(\alpha, F)$ is a generator of the principal ideal of all polynomials in $F[x]$ that have α as a zero. Therefore, irr is the monic polynomial of **minimum degree** that has α as a zero. Also, $\text{irr}(\alpha, F)$ is the only **irreducible** monic polynomial that has α as a zero.

SECTION 40

1. $2, \{1, \sqrt{2}\}$
3. $4, \{1, \sqrt{3}, \sqrt{2}, \sqrt{6}\}$
5. $6, \{1, \sqrt{2}, \sqrt[3]{2}, \sqrt[2]{\sqrt[3]{2}}, (\sqrt[3]{2})^2, \sqrt{2}(\sqrt[3]{2})^2\}$
7. $2, \{1, \sqrt{6}\}$
9. $9, \{1, \sqrt[3]{2}, \sqrt[3]{4}, \sqrt[3]{3}, \sqrt[3]{6}, \sqrt[3]{12}, \sqrt[3]{9}, \sqrt[3]{18}, \sqrt[3]{36}\}$
11. $2, \{1, \sqrt{2}\}$
13. $2, \{1, \sqrt{2}\}$
19. a. F c. F e. F g. F i. F
23. *Partial answer:* Extensions of degree 2^n for $n \in \mathbb{Z}^+$ are obtained.

SECTION 41

All odd-numbered answers require proofs and are not listed here.

SECTION 42

1. Yes
3. Yes
5. 6
7. 0

SECTION 43

1. $\sqrt{2}, -\sqrt{2}$
3. $3 + \sqrt{2}, 3 - \sqrt{2}$
5. $\sqrt{2} + i, \sqrt{2} - i, -\sqrt{2} + i, -\sqrt{2} - i$
7. $\sqrt{1 + \sqrt{2}}, -\sqrt{1 + \sqrt{2}}, \sqrt{1 - \sqrt{2}}, -\sqrt{1 - \sqrt{2}}$
9. $\sqrt{3}$
11. $-\sqrt{2} + 3\sqrt{5}$
13. $-\sqrt{2} + \sqrt{45}$
15. $\frac{\sqrt{3} + \sqrt{5}}{2}$
17. $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$
19. $\mathbb{Q}(\sqrt{3}, \sqrt{10})$
21. \mathbb{Q}
25. a. $3 - \sqrt{2}$
- b. They are the same maps.
39. Yes

SECTION 44

1. 2
3. 4
5. 2
7. 1
9. 2
11. $\sqrt{2} \rightarrow \sqrt{2}, \sqrt{3} \rightarrow \sqrt{3}, \sqrt{5} \rightarrow \sqrt{5}$; and $\sqrt{2} \rightarrow \sqrt{2}, \sqrt{3} \rightarrow -\sqrt{3}, \sqrt{5} \rightarrow -\sqrt{5}$
13. $\sqrt{2} \rightarrow \sqrt{2}, \sqrt{3} \rightarrow \sqrt{3}, \sqrt{5} \rightarrow -\sqrt{5}$; $\sqrt{2} \rightarrow \sqrt{2}, \sqrt{3} \rightarrow -\sqrt{3}, \sqrt{5} \rightarrow \sqrt{5}$; $\sqrt{2} \rightarrow -\sqrt{2}, \sqrt{3} \rightarrow \sqrt{3}, \sqrt{5} \rightarrow \sqrt{5}$
15. There are six extensions. One for each of the combinations where $\sqrt{3}i$ maps to one of $\alpha_1, \alpha_2, \alpha_3$.

17. a. $\mathbb{Q}(\pi^2)$ b. $\sqrt{\pi}$ can map to either $\pm\sqrt{\pi}i$.

19. $1 \leq [E : F] \leq n!$

21. Let $F = \mathbb{Q}$ and $E = \mathbb{Q}(\sqrt{2})$. Then

$$f(x) = x^4 - 5x^2 + 6 = (x^2 - 2)(x^2 - 3)$$

has a zero in E , but does not split in E .

SECTION 45

1. $\alpha = \sqrt[6]{2} = 2/(\sqrt[3]{2}\sqrt{2}), \sqrt{2} = (\sqrt[6]{2})^3, \sqrt[3]{2} = (\sqrt[6]{2})^2$. (Other answers are possible.)
 3. $\alpha = \sqrt{2} + \sqrt{5}, \sqrt{2} = \frac{1}{6}\alpha^3 - \frac{11}{6}\alpha, \sqrt{5} = \frac{17}{6}\alpha - \frac{1}{6}\alpha^3$. (Other answers are possible.)
 7. $f(x) = x^4 - 4x^2 + 4 = (x^2 - 2)^2$. Here $f(x)$ is not an irreducible polynomial. Every irreducible factor of $f(x)$ has zeros of multiplicity 1 only.

SECTION 46

1. 4 3. 8 5. 4 7. 2

9. The group has two elements, the identity automorphism ι of $\mathbb{Q}(i)$ and σ such that $\sigma(i) = -i$.

11. b. Let $\alpha_1 = \sqrt[3]{5}$, $\alpha_2 = \sqrt[3]{5} \frac{-1+i\sqrt{3}}{2}$, and $\alpha_3 = \sqrt[3]{5} \frac{-1-i\sqrt{3}}{2}$.

The maps are

ι , where ι is the identity map;

ρ , where $\rho(\alpha_1) = \alpha_2$ and $\rho(i\sqrt{3}) = i\sqrt{3}$;

ρ^2 , where $\rho^2(\alpha_1) = \alpha_3$ and $\rho^2(i\sqrt{3}) = i\sqrt{3}$;

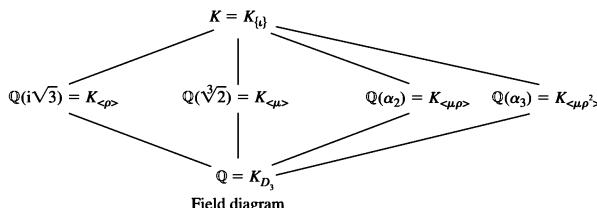
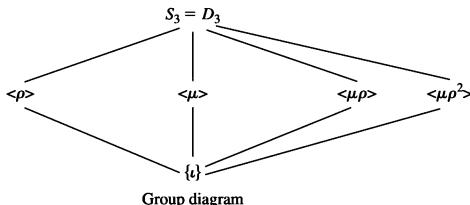
μ , where $\mu(\alpha_1) = \alpha_1$ and $\mu(i\sqrt{3}) = -i\sqrt{3}$;

$\mu\rho$, where $\mu\rho(\alpha_1) = \alpha_3$ and $\mu\rho(i\sqrt{3}) = -i\sqrt{3}$;

$\mu\rho^2$, where $\mu\rho^2(\alpha_1) = \alpha_2$ and $\mu\rho^2(i\sqrt{3}) = -i\sqrt{3}$.

- c. S_3 . The notation in (a) was chosen to coincide with the standard notation for $D_3 \cong S_3$.

d.



13. The splitting field of $(x^3 - 1) \in \mathbb{Q}[x]$ is $\mathbb{Q}(i\sqrt{3})$, and the group is cyclic of order 2 with elements: ι , where ι is the identity map of $\mathbb{Q}(i\sqrt{3})$, and σ , where $\sigma(i\sqrt{3}) = -i\sqrt{3}$.
15. a. F c. T e. T g. F i. F
 25. Partial answer: $G(K/(E \vee L)) = G(K/E) \cap G(K/L)$

SECTION 47

3. $\mathbb{Q}(\sqrt[4]{2}, i)$: $\sqrt[4]{2} + i, x^8 + 4x^6 + 2x^4 + 28x^2 + 1$;

$\mathbb{Q}(\sqrt[4]{2})$: $\sqrt[4]{2}, x^4 - 2$;

$\mathbb{Q}(i\sqrt[4]{2})$: $i(\sqrt[4]{2}), x^4 - 2$;

$\mathbb{Q}(\sqrt[4]{2}, i)$: $\sqrt[4]{2} + i, x^4 - 2x^2 + 9$;

$\mathbb{Q}(\sqrt[4]{2} + i\sqrt[4]{2})$: $\sqrt[4]{2} + i(\sqrt[4]{2}), x^4 + 8$;

$\mathbb{Q}(\sqrt[4]{2} - i\sqrt[4]{2})$: $\sqrt[4]{2} - i(\sqrt[4]{2}), x^4 + 8$;

$\mathbb{Q}(\sqrt[4]{2})$: $\sqrt[4]{2}, x^2 - 2$;

$\mathbb{Q}(i)$: $i, x^2 + 1$;

$\mathbb{Q}(i\sqrt[4]{2})$: $i\sqrt[4]{2}, x^2 + 2$;

\mathbb{Q} : $1, x - 1$

5. The group is cyclic of order 5, and its elements are

	ι	σ_1	σ_2	σ_3	σ_4
$\sqrt[5]{2} \rightarrow$	$\sqrt[5]{2}$	$\zeta(\sqrt[5]{2})$	$\zeta^2(\sqrt[5]{2})$	$\zeta^3(\sqrt[5]{2})$	$\zeta^4(\sqrt[5]{2})$

where $\sqrt[5]{2}$ is the real 5th root of 2.

7. The splitting field of $x^8 - 1$ over \mathbb{Q} is the same as the splitting field of $x^4 + 1$ over \mathbb{Q} , so a complete description is contained in Example 47.7. (This is the easiest way to answer the problem.)

9. a. $s_1^2 - 2s_2$ b. $\frac{s_1s_2 - 3s_3}{s_3}$

SECTION 48

3. a. 16 b. 400 c. 2160

5. 3^0

7. a. T c. F e. T g. T i. F

9. $\Phi_1(x) = x - 1$

$\Phi_2(x) = x + 1$

$\Phi_3(x) = x^2 + x + 1$

$\Phi_4(x) = x^2 + 1$

$\Phi_5(x) = x^4 + x^3 + x^2 + x + 1$

$\Phi_6(x) = x^2 - x + 1$

SECTION 49

1. No. Yes, K is an extension of \mathbb{Z}_2 by radicals.

3. a. T c. T e. T g. T i. F ($x^3 - 2x$ over \mathbb{Q} gives a counterexample.)

APPENDIX

1. $\begin{bmatrix} 2 & 1 \\ 2 & 7 \end{bmatrix}$

3. $\begin{bmatrix} -3+2i & -1-4i \\ 2 & -i \\ 0 & -i \end{bmatrix}$

5. $\begin{bmatrix} 5 & 16 & -3 \\ 0 & -18 & 24 \end{bmatrix}$

7. $\begin{bmatrix} 1 & -i \\ 4-6i & -2-2i \end{bmatrix}$

9. $\begin{bmatrix} 8 & -8i \\ 8i & 8 \end{bmatrix}$

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

13. -48

This page is intentionally left blank

Index

- Abel, Niels Henrik, 21, 150, 191, 385
Abelian extension, 370
Abelian group(s), 20
 finitely-generated, fundamental theorems of, 92, 93
 free, 167
 torsion free, 96, 119
Absolute value, 34
Action
 faithful, 133
 on a group, 132
 transitive, 133
Addition
 modulo 2π , 37
 modulo c , 33
 modulo n , 32, 65
Additive identity, 14
Adjoining elements to field, 321
Al-Tusi Sharaf al-Din, 224
Algebra
 fundamental theorem of, 257, 324
 group, 261
Algebraic closure, 324
Algebraic closure of F in E , 323
Algebraic element over F , 313
Algebraic extension, 319
Algebraic integer, 378
Algebraic number, 314
Algebraic number theory, 294
Algebraic variety, 298
Algebraically closed field, 323, 327
Alphabet, 172
Alternating group on n letters, 84
Antisymmetric law, 324
Arc
 of a diagraph, 72
 of a graph, 308
Arithmetic, fundamental theorem of, 280
Artin, Emil, 225
Ascending central series, 163
Ascending chain condition, 278, 286
Aschbacher, Michael, 127
Associates, 276
Associative operation, 13, 20
Automorphism, 341
 of a field, 345
 Frobenius, 349
 of a group, 68, 118
 group of, 345
 inner, 118
Axiom of choice, 324, 325
Axis of reflection, 105
Banach, Stefan, 268
Basis
 for a finitely-generated abelian group, 176
 for a free abelian group, 167
 Gröbner, 305
 for an ideal, 298
 for a vector space, 270
Bessy, Bernard Frenicle de, 201
Betti number, 92
Bijection, 4
Binary operation, 11
Bit, 237
Bit errors, 239
Blip group, 177
Bloom, David M., 82
Blop group, 177
Boolean ring, 194
Bourbaki, Nicholas, 4, 177, 326
Brahmagupta, 288
Burnside, William, 127, 152
Burnside's formula, 141
Cancellation laws, 23, 196
Cardano, Girolamo, 224, 385
Cardinality, 3, 4
Cartesian product, 2, 88
Cauchy Augustin-Louis, 42
Cauchy's theorem, 136
Cayley, Arthur, 72, 81, 178
Cayley digraph, 72
Cayley's theorem, 79
Cell, 5
Center of a group, 60, 128, 163
Chain(s), 324
Chain condition,
 ascending, 278, 286
 descending, 286
Characteristic of a ring, 198
Chief series, 161
Class
 equivalence, 7
 residue modulo n , 6
Class equation, 156
Closed interval, 8
Closed set under an operation, 12, 37
Closure, algebraic, 323
Code, 238
 polynomial, 241
Code word, 238
Codomain, 3
Coefficients
 of a polynomial, 219
 torsion, 93, 96
Coloring of graph, 308
Commensurable numbers, 223
Commutative operation, 13
Commutative ring, 189
Commutator, 120, 128
Commutator subgroup, 120, 128
Comparable elements, 324
Complex number, 3, 33
 absolute value of, 34
 multiplication of, 34
Composition, function, 12, 14
 associativity of, 14
Composition series, 161
Congruence modulo n , 6
Conjugate elements over F , 345
Conjugate subgroups, 119, 120
Conjugation, 118
Conjugation isomorphism, 345

- Consequence, 179
 Constant polynomial, 219
 Constructible number, 329
 Constructible polygon, 380
 Content of a polynomial, 237, 281
 Contraction, elementary, 173
 Correction of bit errors, 239
 Correspondence, one-to-one, 4
 Coset, 97
 - double, 104
 - left, 98
 - right, 100
 Coset group, 97
 Crelle, August, 21
 Cycle(s), 44
 - disjoint, 44, 45
 - k -, 44
 Cyclic extension, 371
 Cyclic group, 56, 61
 - elementary properties of, 61
 - finite, subgroups of, 66
 - structure of, 64
 Cyclic linear code, 238
 Cyclic subgroup, 55, 61
 Cyclotomic extension, 378
 Cyclotomic polynomial, 234, 378
- Decomposable group, 93
 Decryption, 205
 Dedekind, Richard, 191, 244
 Definitions, 1
 Degree
 - of α over F , 315
 - of an extension, 319
 - of a polynomial, 219
 Derivative, 340
 Descartes, René, 218
 Descending chain condition, 286
 Detection of bit error, 239
 Determinant of a square matrix, 28
 Diagonal matrix, 28
 Diagram, subgroup, 53
 Digraph, 72
 - arc of, 72
 - vertex of, 72
 Dihedral group, 46
 Dimension of a vector space over F , 272
 Direct product, 89
 - external, 91
 - internal, 91
 - of rings, 187
 Direct sum, 89
 - of vector spaces, 274
 Dirichlet, Peter Lejeune, 191
 Discrete frieze group, 107
 Discriminant of a polynomial, 378
 Disjoint cycles, 44, 45
 Disjoint sets, 5
 Disjoint union of G -sets, 140
 Distance between two strings, 238
 Distributive law, 185
 Division algorithm
 - for \mathbb{Z} , 62
 - for $F[x]$, 228, 286, 300
 Division ring, 190
 Divisor, 276, 300
 - greatest common, 63, 281
 - of a polynomial, 234
 - of zero, 194
 Domain
 - Euclidean, 286
 - of a function, 3
 - integral, 196
 - principal ideal, 276
 - unique factorization, 276
 Double coset, 104
 Doubling the cube, 332
- Edges of graph, 308
 Eisenstein criterion, 233
 Element(s), 1
 - adjoining to field, 321
 - algebraic over F , 313
 - comparable, 324
 - conjugate over F , 345
 - equivalent, 212
 - fixed, 342
 - idempotent, 19, 31, 193, 199
 - identity, 14, 20
 - inverse of, 20
 - irreducible, 276
 - maximal, 325
 - nilpotent, 193, 250
 - orbit of, 82, 135
 - order of, 61
 - prime, 280
 - primitive, 360
 - separable over F , 359
 - transcendent over F , 313
 Elementary contraction, 173
 Elementary symmetric function, 372
 Empty set, 1
 Empty word, 172
 Encryption, 205
 - RSA, 206
 Endomorphism, 258
 Equation, class, 156
 Equivalence class, 7
 Equivalence relation, 6
 Equivalent elements, 212
 Errors, bit, 239
 Escher M. C., 109
 Euclid, 201, 288
 Euclidean algorithm, 288
 Euclidean domain, 286
 Euclidean norm, 286
 Euler, Leonard, 21, 34, 202, 381
 Euler formula, 34
 Euler phi-function, 105, 202
 Euler's theorem, 202
 Evaluation homomorphism, 120, 189, 221
 Even permutation, 83
 Extension(s), 311
 - abelian, 370
 - algebraic, 319
 - cyclic, 371
 - cyclotomic, 378
 - degree of, 319
 finite, 319
 finite normal, 362
 join of, 371
 of a map, 351
 by radicals, 384
 separable, 347, 359
 simple, 315
 Extension field, 311
 External direct product, 91
- Factor, 276, 300
 - of a polynomial, 300
 Factor group, 113, 115
 Factor theorem, 229
 Faithful action, 133
 Feit, Walter, 127, 152
 Fermat, Pierre de, 201
 Fermat prime, 381
 Fermat's last theorem, 277
 Fermat's $p = a^2 + b^2$ theorem, 295
 Fermat's theorem, 200
 - Euler's generalization of, 202
 Ferrari, Lodovico, 385
 Ferro, Scipione del, 385
 Field, 190
 - adjoining elements to, 321
 - algebraic closure of, 323
 - algebraic closure in E , 323
 - algebraically closed, 323, 327
 - extension of, 311
 - fixed, 365
 - Galois, 335
 - intermediate, 344
 - perfect, 359
 - prime, 254
 - of quotients, 211, 215
 - of rational functions, 221
 - separable extension of, 359
 - simple extension of, 315
 - skew, 190
 - splitting, 347, 349, 353
 - strictly skew, 190
 - subfield of, 190
 - tower of, 311
 Field extension, 311
 - simple, 315
 Finite-basis condition, 286
 Finite extension, 319
 - degree of, 319
 Finite generating set, 286
 Finite group, 25
 Finite presentation, 179
 Finite-dimensional vector space, 269
 Finitely-generated group, 71
 Fixed elements, 342
 Fixed field, 365
 Fixed point, 110
 Free abelian group, 167
 - basis for, 167
 - rank of, 168
 Free generators, 174
 Free group, 173
 - rank of, 174
 Frey, Gerhard, 277
 Frieze group, 107

- Frobenius, Georg, 150
 Frobenius automorphism, 349
 Frobenius homomorphism, 249
 Function(s), 3
 codomain of, 3
 composite, 12, 14
 composition of, 12, 14
 domain of, 3
 elementary symmetric, 372
 Euler phi-, 105, 202
 image of A under, 78
 inverse of, 4
 one-to-one, 4
 onto, 4
 phi-, 105, 202
 polynomial on F , 227
 range of, 3
 rational, 221
 symmetric, 372
 trap door, 206
 two-to-two, 9
- Fundamental homomorphism theorem, 117, 247
 Fundamental theorem of algebra, 257, 324
 Fundamental theorem of arithmetic, 280
 Fundamental theorem of finitely-generated abelian groups, 170
 invariant factor version of, 93
 primary factor version of, 92
 G -set(s), 132
 applications to counting, 140
 applications to finite groups, 136
 disjoint union of, 140
 isomorphic, 139
 orbits of, 135
 sub-, 139
 transitive, 133
- Gallian, Joseph A., 108
 Galois, Evariste, 162, 191, 337, 379
 Galois field, 335
 Galois group, 362
 Galois theory, 341
 main theorems of, 364
 Gauss, Carl F., 21, 92, 292, 334, 337, 379
 Gauss's lemma, 282
 Gaussian integer, 216, 292
 Gaussian norm, 292
 General linear group, 22
 General polynomial of degree n , 372
 Generating set, 70, 71
 Generator(s), 56, 61, 70, 71
 for a presentation, 179
 free, 174
 of a group, 56, 61
 of a principal ideal, 254
 relation on, 75, 179
 for a vector space, 269
 Glide reflection, 106
 nontrivial, 108
- Graph, 308
 Grassmann, Hermann, 268
 Greatest common divisor, 63, 281
 Griess, Robert L. Jr., 127
 Gröbner basis, 305
 Group(s), 20
 abelian, 20
 alternating on n letters, 84
 ascending central series of, 163
 automorphism of, 68, 118
 of automorphisms, 345
 blip, 177
 blob, 177
 center of, 60, 128, 163
 commutator in a, 120
 of cosets, 97
 cyclic, 56, 61
 decomposable, 93
 dihedral, 46
 direct product of, 89
 direct sum of, 89
 discrete frieze, 107
 elementary properties of, 23
 endomorphism of, 258
 factor, 115
 finite, 25
 finitely-generated, 71
 free, 173
 free on A , 173
 free abelian, 167
 frieze, 107
 Galois, 362
 general linear, 22
 generator(s) of, 56, 61, 70, 71
 indecomposable, 93
 inner automorphism of, 118
 isomorphic, 25, 26
 Klein 4-, 53
 notation and terminology, 39
 octic, 182
 order of, 41
 p -, 137
 permutation, 41, 77, 79
 plane crystallographic, 108
 of a polynomial, 365
 presentation of, 178, 179
 quaternion, 183
 quotient, 115
 regular representation of, 80
 series of, 157
 simple, 126
 solvable, 163
 special linear, 115
 subgroup of, 52
 symmetric on n letters, 43
 of symmetries, 105
 torsion, 119
 torsion free, 119
 wallpaper, 108
- Group action, 132
 Group algebra, 261
 Group homomorphism, 77
 Group ring, 261
 Group table, 25
 properties of, 27
- Hamilton, Sir William Rowan, 262, 268
 Hamming distance, 238
 Hamming weight, 238
 Hilbert, David, 186
 Hilbert basis theorem, 299
 Hölder, Otto, 162, 178
 Homomorphism, 113, 188
 evaluation, 120, 189, 221
 Frobenius, 249
 fundamental theorem for, 117, 247
 group, 77
 kernel of, 78, 188, 247
 projection, 246
 of a ring, 188, 245
- Homomorphism property, 26
- Ideal(s), 244
 ascending chain condition for, 278, 286
 basis for, 298
 descending chain condition for, 286
 finite-basis condition for, 286
 improper, 251
 left, 258
 maximal, 251
 maximum condition for, 286
 minimum condition for, 286
 nilradical of, 250
 prime, 252
 principal, 254
 product of, 258
 proper nontrivial, 251
 quotient of, 258
 radical of, 250
 right, 258
 sum of, 257
 trivial, 251
- Idempotent element, 19, 31, 193, 199
 Identity element, 14, 20
 left, 25
- Image
 of A , 78
 inverse, 78
 under a map, 78
- Imaginary number, 34
 Improper ideal, 251
 Improper subgroup, 52, 59
 Improper subset, 2
- Indecomposable group, 93
 Indeterminate, 218
 Index of a subgroup, 99
 Induced operation, 12, 114
 Infinite order, 61
 Infinite set, 4
 Information rate, 238
 Injection, 4
 Injection map, 4, 214
 Inner automorphism, 118
 Integer(s), 2
 algebraic, 378
 Gaussian, 216, 292
 rational, 292
 relatively prime, 64

- Integral domain**, 196
 associates in, 276
 Euclidean norm on, 286
 prime element of, 280
 field of quotients of, 211, 215
 unit in, 276
Intermediate field, 344
Internal direct product, 91
Intersection, 60, 71
Interval, closed, 8
Invariant factors, 93
Invariant series, 157
Invariant subgroup, 118
Inverse
 of an element, 20
 left, 25
 multiplicative, 190
Inverse function, 4
Inverse map, 4
Irreducible element, 276
Irreducible polynomial, 231
 for α over F , 315, 319
 in $F[x]$, 231
Isometry, 22, 105
Isomorphic G -sets, 139
Isomorphic groups, 26
Isomorphic presentations, 180
Isomorphic rings, 189
Isomorphic series, 157
Isomorphism
 conjugation, 345
 of a G -set, 139
 of a group, 26
 of a ring, 189
 up to, 92
 of a vector space, 274
Isomorphism extension theorem, 351
Isomorphism theorems, 145–148
Isosceles triangle, 2
Isotropy subgroup, 135

Join
 of extension fields, 371
 of subgroups, 146
Jordan, Camille, 21, 162
Jordan-Hölder theorem, 161

 k -cycle, 44
Kernel, 78, 188, 247
 of a linear transformation, 275
Khayyam, Omar, 224
Klein 4-group, 32, 53
Kronecker, Leopold, 92, 191, 312
Kronecker's theorem, 312
Kummer, Ernst, 92, 244, 277

Lagrange, Joseph-Louis, 21, 42, 385
 theorem of, 99, 123
Lame, Gabriel, 277
Law
 antisymmetric, 324
 cancellation, 23, 196
 distributive, 185
 reflexive, 324
 transitive, 324

Least common multiple, 69, 90, 291
Left cancellation law, 23
Left coset, 98
Left distributive law, 185
Left ideal, 258
Left identity, 25
Left inverse, 25
Left R -module, 272
Left regular representation, 80
Length of a code word, 238
Letter, 172
Levi ben Gerson, 42
Levinson, Norman, 339
Lexicographical order, 304
Lindemann, Ferdinand, 334
Linear code, 238
Linear combination, 269
Linear group, special, 115
Linear transformation, 275
 kernel of, 275
Linearly dependent vectors over F , 269
Linearly independent vectors over F , 269
Liouville, Joseph, 277

Main diagonal of a matrix, 28
Main theorems of Galois theory, 364
Map, 3
 extension of, 351
 image under, 78
 injection, 214
 inverse of, 4
 range of, 3
Matrix
 determinant of, 28
 diagonal, 28
 main diagonal of, 28
 orthogonal, 57
 transpose of, 57
 upper-triangular, 28
Maximal element, 325
Maximal ideal, 251
Maximal normal subgroup, 127
Maximum condition, 286
Mersenne prime, 201
Minimal polynomial for α over F , 315, 319
Minimal subset, 55*n*
Minimum condition, 286
Monic polynomial, 314
Monoid, 25
Multiple, least common, 69, 90, 291
Multiplication
 by components, 88
 modulo n , 187
 permutation, 41
Multiplicative identity, 14
Multiplicative inverse, 190
Multiplicative norm, 294
Multiplicity of a zero, 358

Nilpotent element, 193, 250
Nilradical, 250
Noether, Emmy, 186

Noetherian ring, 278
Nontrivial ideal, proper, 251
Nontrivial subgroup, 52
Norm
 Euclidean, 286
 Gaussian, 292
 multiplicative, 294
 over F , 370
Normal extension, finite, 362
Normal series, 157
Normal subgroup, 114, 118, 248
 maximal, 127
Normalizer of a subgroup, 150
Nullstellensatz, 257
Number(s)
 algebraic, 314
 Betti, 92
 commensurable, 223
 complex, 3, 33
 constructible, 329
 imaginary, 34
 rational, 2
 real, 2
 transcendental, 314
Number theory, algebraic, 294

Octic group, 182
Odd permutation, 83
One-to-one correspondence, 4
One-to-one function, 4
Onto function, 4
Operation
 associative, 13, 20
 binary, 11
 commutative, 13
 induced, 12
 well-defined, 16
Orbit, 51, 82, 135
Order
 of a group, 41
 of an element, 61
 infinite, 61
 of ring, 192
 term, 304
Ordering
 lexicographical, 304
 partial, 324
 of power products, 303
Orientation, 106
Orthogonal matrix, 57

 p -group, 137
 p -subgroup, 137
Partial ordering, 324
Partition, 5
 cells of, 5
Pattern, periodic, 108
Peano, Giuseppe, 268
Perfect field, 359
Periodic pattern, 108
Permutation, 41
 even, 83
 groups of, 77, 79
 movement of elements in, 87
 multiplication, 41

- odd, 83
- orbits of, 82
- sign of, 84
- Phi-function, 105, 202
- Plane
 - isometry of, 22
 - translation of, 105
- Plane crystallographic group, 108
- Plane isometry, 105
- Point, fixed, 110
- Polygon, constructible, 380
- Polynomial(s), 219
 - coefficients of, 219
 - constant, 219
 - content of, 237, 281
 - cyclotomic, 234, 378
 - degree of, 219
 - discriminant of, 378
 - divisor of, 234, 300
 - Eisenstein, 233
 - factor of, 300
 - general of degree n , 372
 - group of, 365
 - irreducible for α over F , 315, 319
 - irreducible over F , 231
 - irreducible, 231
 - minimal for α over F , 319
 - monic, 314
 - primitive, 237, 281
 - reducible, 231
 - ring of, 220
 - separable over F , 359
 - solvable by radicals over F , 384
 - splitting field of, 347, 349, 353
 - term ordering of, 304
 - zero of, 223, 298
- Polynomial code, 241
- Polynomial extension, 350
- Polynomial function on F , 227
- Power product, 303
 - ordering of, 303
- Power set, 8
- Presentation, 178, 179
 - finite, 179
 - generators for, 179
 - isomorphic, 180
- Prime, 280
 - Fermat, 381
 - Mersenne, 201
 - relatively, 291
- Prime field, 254
- Prime ideal, 252
- Primitive element, 360
- Primitive element theorem, 360
- Primitive n th root of unity, 69, 336
- Primitive polynomial, 237, 281
- Principal ideal, 254
 - generator of, 254
- Principal ideal domain, 276
- Principal series, 161
- Private key, 206
- Product
 - Cartesian, 2, 88
 - direct, 89, 187
- of ideals, 258
- power, 303
- Projection homomorphism, 246
- Proper nontrivial ideal, 251
- Proper subgroup, 52
- Proper subset, 2
- Public key, 206, 207
- Pythagorean theorem, 224
- Qin Jushao, 288
- Quaternion group, 183
- Quaternions, 262
- Quotient
 - in the division algorithm, 62
 - of ideals, 258
- Quotient group, 115
- Quotient space, 275
- Rabin, Michael, 180
- Radical(s)
 - extension by, 384
 - of an ideal, 250
- Range of a map, 3
- Rank, 168, 174
- Rate of linear code, 238
- Rational function, 221
- Rational integer, 292
- Rational number, 2
- Real number, 3
- Reduced word, 173
- Reducible polynomial, 231
- Reduction modulo n , 116
- Refinement of a series, 157
- Reflection, 105
 - axis of, 105
 - glide, 106
- Reflexive law, 324
- Reflexive relation, 6, 7
- Regular representation, 80
 - left, 80
 - right, 81
- Relation(s), 3, 75, 179
 - consequence of, 179
 - equality, 3
 - equivalence, 6
 - reflexive, 6, 7
 - symmetric, 6, 7
 - transitive, 6, 7
- Relatively prime, 64, 291
- Relator, 179
- Remainder in the division algorithm, 62
- Representation
 - left regular, 80
 - right regular, 81
- Residue class modulo n , 6
- Ribet, Ken, 277
- Right cancellation law, 23
- Right coset, 100
- Right distributive law, 185
- Right ideal, 258
- Right R -module, 273
- Right regular representation, 81
- Ring(s), 185
 - additive group of, 186
 - Boolean, 194
 - characteristic of, 198
 - commutative, 189
 - division, 190
 - of endomorphisms, 258
 - factor, 245
 - group, 261
 - homomorphism, 188, 245
 - ideal of, 248
 - isomorphic, 189
 - isomorphism of, 189
 - maximal ideal of, 251
 - modules over, 272
 - nilradical of, 250
 - Noetherian, 278
 - order of, 192
 - of polynomials, 220
 - prime ideal of, 252
 - quotient, 245
 - radical of, 250
 - simple, 257
 - subring of, 190
 - unit in a, 190, 276
 - with unity, 189
 - zero, 189
- Roots of unity, 37
 - n th, 37
 - primitive n th, 69, 336
- Rotation, 105
- RSA encryption, 206
- Ruffini, Paolo, 385
- Scalar, 267
- Schreier theorem, 160
- Sefer Yetzirah, 42
- Semigroup, 25
- Separable element over F , 359
- Separable extension, 347, 359
- Separable polynomial over F , 359
- Series
 - ascending central, 163
 - chief, 161
 - composition, 161
 - invariant, 157
 - isomorphic, 157
 - normal, 157
 - principal, 161
 - refinement of, 157
 - subnormal, 157
- Set(s), 1
 - binary operation on, 11
 - cardinality of, 3
 - Cartesian product of, 2, 88
 - closed under an operation, 12
 - disjoint, 5
 - element of, 1
 - empty, 1
 - finite generating, 286
 - G -, 132
 - generating, 70, 71
 - infinite, 4
 - intersection of, 60, 71
 - partial ordering of, 324

- Set(s) (cont.)**
- partition of, 5
 - permutation of, 41
 - power, 8
 - subset of, 2
 - union of, 278
 - well-defined, 1
- Shimura, Goro, 277
- Sign of a permutation, 84
- Simple extension, 315
- Simple group, 126
- Simple ring, 257
- Skew field, 190
- Smallest subset, 55*n*
- Solvable group, 163
- Solvable polynomial over F , 384
- Span, 269
- Special linear group, 115
- Splitting field, 347, 349, 353
- Square matrix
- determinant of, 28
 - main diagonal of, 28
- Squaring the circle, 333
- Standard form of dihedral group element, 48
- Strictly skew field, 190
- Sub- G -set, 139
- Subfield, 190
- Subgroup(s), 52
- commutator, 120, 128
 - conjugate, 119, 120
 - cyclic, 55, 61
 - improper, 52, 59
 - index of, 99
 - invariant, 118
 - isotropy, 135
 - join of, 146
 - maximal normal, 127
 - nontrivial, 52
 - normal, 114, 118, 248
 - normalizer of, 150
 - p -, 137
 - proper, 52
 - torsion, 96
 - trivial, 52
- Subgroup diagram, 53
- Subnormal series, 157
- Subring, 190
- generated by a , 193
- Subset, 2
- improper, 2
 - minimal, 55*n*
 - proper, 2
 - smallest, 55*n*
- upper bound for, 325
- Subspace of a vector space, 274
- Sum
- direct, 89
 - of ideals, 257
 - modulo n , 65
- Surjection, 4
- Syllable, 172
- Sylow, Peter Ludvig Mejdell, 150
- Sylow p -subgroup, 151
- Sylow theorems, 150
- Symmetric function, 372
- elementary, 372
- Symmetric group on n letters, 43
- Symmetric relation, 6, 7
- Symmetries, group of, 105
- Table, group, 25
- properties of, 27
- Taniyama, Yutaka, 277
- Tartaglia, Niccolo, 385
- Taylor, Richard, 277
- Term ordering, 304
- Thompson, John G., 127, 152
- Torsion coefficient, 93, 96
- Torsion free, 96, 119
- Torsion group, 119
- Torsion subgroup, 96
- Tower of fields, 311
- Trace over F , 370
- Transcendental element over F , 313
- Transcendental number, 314
- Transitive action, 133
- Transitive G -set, 133
- Transitive law, 324
- Transitive relation, 6, 7
- Translation, 105
- Transpose of a matrix, 57
- Transposition, 45, 81
- Trap door functions, 206
- Triangle, isosceles, 2
- Trisection of an angle, 333
- Trivial ideal, 251
- Trivial subgroup, 52
- Two-to-two function, 9
- Union
- of sets, 278
 - of G -sets, 140
- Unique factorization domain, 276
- Unit, 190, 276
- Unit circle, 37
- Unity, 189
- n*th root of, 37, 336
- primitive *n*th root of, 69, 336
- Upper bound for a subset, 325
- Upper-triangular matrix, 28
- Variety, algebraic, 298
- Vector(s), 267
- linear combination of, 269
 - linearly dependent over F , 269
 - linearly independent over F , 269
- Vector space(s), 267
- basis for, 270
 - dimension over F , 272
 - direct sum of, 274
 - finite-dimensional, 269
 - isomorphism of, 274
 - linear transformation of, 275
 - subspace of, 274
- Vertex/vertices
- of a digraph, 72
 - of graph, 308
- Viete, Francois, 218
- Von Dyck, Walther, 21, 81
- Wallpaper group, 108
- Wantzel, Pierre, 334
- Weber, Heinrich, 21, 191
- Wedderburn, Joseph Henry Maclagan, 262
- Wedderburn theorem, 263
- Weierstrass, Karl, 312
- Weight of a string, 238
- Well-defined operation, 16
- Well-defined set, 1
- Weyl, Hermann, 268
- Weyl algebra, 260
- Wiles, Andrew, 277
- Wilson's theorem, 205
- Word(s), 172
- empty, 172
 - reduced, 173
- Word problem, 180
- Zassenhaus, Hans, 158
- Zassenhaus lemma, 158
- Zermelo, Ernst, 326
- Zero
- multiplicity of, 358
 - of a polynomial, 223, 298
- Zero divisors, 194
- Zero ring, 189
- Zorn, Max, 325
- Zorn's lemma, 324, 325