Abstract Algebra for Teachers

Abstract Algebra for Teachers

Thomas W. Judson
Stephen F. Austin State University

Oscar Levin
University of Northern Colorado

April 16, 2022

Edition: Preliminary Edition

Website: https://github.com/twjudson/AA-Teachers

For ...

Acknowledgements

Robert Beezer encouraged me to make *Abstract Algebra: Theory and Applications* available as an open source textbook, a decision that I have never regretted. With his assistance, the book has been rewritten in PreTeXt (pretextbook.org[1]), making it possible to quickly output print, web, PDF versions and more from the same source.

Abstract Algebra for Teachers grew out of *Abstract Algebra: Theory and Applications* in an attempt to provide a textbook for a one-semester course with the option of either placing an emphasis on rings or groups. The open source version of this book has received support from the National Science Foundation (Award #DUE–1821329).

[1] https://pretextbook.org

Preface

This textbook is intended for a one-semester undergraduate course in abstract algebra. The textbook is divided into three parts: Part I (An Introduction to Groups and Rings), Part II (Topics in Group Theory), and Part III (Topics in Ring Theory). Part I covers basic set theory and facts about the integers. Basic definitions, examples, and theorems for both groups and rings are introduced in Part I. It is expected that every course will cover Part I of the textbook. Part II (Topics in Group Theory) and Part III (Topics in Ring Theory) may be covered in any order, creating the flexibility for those who prefer to cover rings and fields early in a course.

For the Student. Though there are no specific prerequisites for a course in abstract algebra, students who have had other higher-level courses in mathematics will generally be more prepared than those who have not, because they will possess a bit more mathematical sophistication. Occasionally, we shall assume some basic linear algebra; that is, we shall take for granted an elementary knowledge of matrices and determinants. This should present no great problem, since most students taking a course in abstract algebra have been introduced to matrices and determinants elsewhere in their career, if they have not already taken a sophomore or junior-level course in linear algebra.

For the Teacher. Exercise sections are the heart of any mathematics text. An exercise set appears at the end of each section. The nature of the exercises ranges over several categories; computational, conceptual, and theoretical problems are included. A section presenting hints and solutions to many of the exercises appears at the end of the text. Often in the solutions a proof is only sketched, and it is up to the student to provide the details. The exercises range in difficulty from very easy to very challenging. Many of the more substantial problems require careful thought, so the student should not be discouraged if the solution is not forthcoming after a few minutes of work.

We have also included activites in each section. We encourage instructors to assign these activites are exercises to individual students or small groups of students during class time. By doing so, instructors will be able to receive immediate feedback on student understanding.

Thomas W. Judson
Nacogdoches, Texas

Oscar Levin
Greeley, Colorado
2022

Contents

Appendices

Back Matter

Part I

An Introduction to Groups and Rings

Chapter 1

Sets, Functions, and Equivalence Relations

Objectives

- To understand and be able to use sets and operations on sets.
- To understand and be able to apply the definition of Cartesian products.
- To understand the definitions of relations and mappings.
- To understand and be able to apply the definitions of one-to-one and onto functions.
- to understand the definitions of equivalence relations and partitions and the connection between the two.

A Short Note on Proofs. A certain amount of mathematical maturity is necessary to find and study applications of abstract algebra. A basic knowledge of set theory, mathematical induction, equivalence relations, and matrices is a must. Even more important is the ability to read and understand mathematical proofs. In this chapter we will outline the background needed for a course in abstract algebra.

Abstract mathematics is different from other sciences. In laboratory sciences such as chemistry and physics, scientists perform experiments to discover new principles and verify theories. Although mathematics is often motivated by physical experimentation or by computer simulations, it is made rigorous through the use of logical arguments. In studying abstract mathematics, we take what is called an axiomatic approach; that is, we take a collection of objects $\mathcal{S}$ and assume some rules about their structure. These rules are called **axioms**. Using the axioms for $\mathcal{S}$, we wish to derive other information about $\mathcal{S}$ by using logical arguments. We require that our axioms be consistent; that is, they should not contradict one another. We also demand that there not be too many axioms. If a system of axioms is too restrictive, there will be few examples of the mathematical structure.

A **statement** in logic or mathematics is an assertion that is either true or false. Consider the following examples:

- $3 + 56 - 13 + 8/2$.
- All cats are black.
- $2 + 3 = 5$.

- $2x = 6$ exactly when $x = 4$.
- If $ax^2 + bx + c = 0$ and $a \neq 0$, then
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$
- $x^3 - 4x^2 + 5x - 6$.

All but the first and last examples are statements, and must be either true or false.

A **mathematical proof** is nothing more than a convincing argument about the accuracy of a statement. Such an argument should contain enough detail to convince the audience; for instance, we can see that the statement "$2x = 6$ exactly when $x = 4$" is false by evaluating $2 \cdot 4$ and noting that $6 \neq 8$, an argument that would satisfy anyone. Of course, audiences may vary widely: proofs can be addressed to another student, to a professor, or to the reader of a text. If more detail than needed is presented in the proof, then the explanation will be either long-winded or poorly written. If too much detail is omitted, then the proof may not be convincing. Again it is important to keep the audience in mind. High school students require much more detail than do graduate students. A good rule of thumb for an argument in an introductory abstract algebra course is that it should be written to convince one's peers, whether those peers be other students or other readers of the text.

Let us examine different types of statements. A statement could be as simple as "$10/5 = 2$;" however, mathematicians are usually interested in more complex statements such as "If p, then q," where p and q are both statements. If certain statements are known or assumed to be true, we wish to know what we can say about other statements. Here p is called the **hypothesis** and q is known as the **conclusion**. Consider the following statement: If $ax^2 + bx + c = 0$ and $a \neq 0$, then
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$
The hypothesis is $ax^2 + bx + c = 0$ and $a \neq 0$; the conclusion is
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Notice that the statement says nothing about whether or not the hypothesis is true. However, if this entire statement is true and we can show that $ax^2 + bx + c = 0$ with $a \neq 0$ is true, then the conclusion *must* be true. A proof of this statement might simply be a series of equations:

$$\begin{aligned}
ax^2 + bx + c &= 0 \\
x^2 + \frac{b}{a}x &= -\frac{c}{a} \\
x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 &= \left(\frac{b}{2a}\right)^2 - \frac{c}{a} \\
\left(x + \frac{b}{2a}\right)^2 &= \frac{b^2 - 4ac}{4a^2} \\
x + \frac{b}{2a} &= \frac{\pm\sqrt{b^2 - 4ac}}{2a} \\
x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.
\end{aligned}$$

If we can prove a statement true, then that statement is called a **proposition**. A proposition of major importance is called a **theorem**. Sometimes instead of proving a theorem or proposition all at once, we break the proof down into modules; that is, we prove several supporting propositions, which are called **lemmas**, and use the results of these propositions to prove the main result. If we can prove a proposition or a theorem, we will often, with very little effort, be able to derive other related propositions called **corollaries**.

There are several different strategies for proving propositions. In addition to using different methods of proof, students often make some common mistakes when they are first learning how to prove theorems. To aid students who are studying abstract mathematics for the first time, we list here some of the difficulties that they may encounter and some of the strategies of proof available to them. It is a good idea to keep referring back to this list as a reminder. (Other techniques of proof will become apparent throughout this chapter and the remainder of the text.)

- A theorem cannot be proved by example; however, the standard way to show that a statement is not a theorem is to provide a counterexample.
- Quantifiers are important. Words and phrases such as *only*, *for all*, *for every*, and *for some* possess different meanings.
- Never assume any hypothesis that is not explicitly stated in the theorem. *You cannot take things for granted.*
- Suppose you wish to show that an object *exists* and is *unique*. First show that there actually is such an object. To show that it is unique, assume that there are two such objects, say r and s, and then show that $r = s$.
- Sometimes it is easier to prove the contrapositive of a statement. Proving the statement "If p, then q" is exactly the same as proving the statement "If not q, then not p."
- Although it is usually better to find a direct proof of a theorem, this task can sometimes be difficult. It may be easier to assume that the theorem that you are trying to prove is false, and to hope that in the course of your argument you are forced to make some statement that cannot possibly be true.

Remember that one of the main objectives of higher mathematics is proving theorems. Theorems are tools that make new and productive applications of mathematics possible. We use examples to give insight into existing theorems and to foster intuitions as to what new theorems might be true. Applications, examples, and proofs are tightly interconnected—much more so than they may seem at first appearance.

Activity 1.1 Consider each of the following arguments. Is the the argument valid? If not, why?

(a)
  i. All equilateral triangles are equiangular.
  ii. All equiangular triangles are isosceles.
  iii. Therefore, all equilateral triangles are isosceles.

(b)
  i. All equilateral triangles are equiangular.
  ii. All equiangular triangles are isosceles.
  iii. Therefore all isosceles triangles are equilateral.

(c)
  i. If Peyton studies every day, then Peyton will be successful.

ii. Peyton does not study every day.

iii. Therefore, Peyton will not be successful.

(d) i. If Peyton study every day, then Peyton will be successful.

ii. Peyton was not successful.

iii. Therefore, Peyton did not study every day.

(e) i. If the alarm goes off, then Ralph will call the police.

ii. Ralph called the police.

iii. So the alarm went off.

(f) i. If the alarm goes off, then Ralph will call the police.

ii. If Ralph calls the police, then Ralph will file a report.

iii. The alarm went off, so Ralph will file a report.

(g) i. If today is Friday, then tomorrow is Saturday.

ii. Tomorrow is Monday, so today is not Friday.

(h) i. All teachers are smart.

ii. Some teachers are funny.

iii. Therefore, some smart people are funny.

(i) i. If a student is a freshman, then the student takes English.

ii. Susan is a junior.

iii. Therefore, Susan does not take English.

1.1 Set Theory

A **set** is a well-defined collection of objects; that is, it is defined in such a manner that we can determine for any given object x whether or not x belongs to the set. The objects that belong to a set are called its **elements** or **members**. We will denote sets by capital letters, such as A or X; if a is an element of the set A, we write $a \in A$.

A set is usually specified either by listing all of its elements inside a pair of braces or by stating the property that determines whether or not an object x belongs to the set. We might write

$$X = \{x_1, x_2, \ldots, x_n\}$$

for a set containing elements $x_1, x_2, \ldots, x_n$ or

$$X = \{x : x \text{ satisfies } \mathcal{P}\}$$

if each x in X satisfies a certain property $\mathcal{P}$. For example, if E is the set of even positive integers, we can describe E by writing either

$$E = \{2, 4, 6, \ldots\} \quad \text{or} \quad E = \{x : x \text{ is an even integer and } x > 0\}.$$

We write $2 \in E$ when we want to say that 2 is in the set E, and $-3 \notin E$ to say that -3 is not in the set E.

Some of the more important sets that we will consider are the following:

$$\mathbb{N} = \{n : n \text{ is a natural number}\} = \{1, 2, 3, \ldots\};$$

$$\begin{aligned}\mathbb{Z} &= \{n : n \text{ is an integer}\} = \{\ldots, -1, 0, 1, 2, \ldots\};\\ \mathbb{Q} &= \{r : r \text{ is a rational number}\} = \{p/q : p, q \in \mathbb{Z} \text{ where } q \neq 0\};\\ \mathbb{R} &= \{x : x \text{ is a real number}\};\\ \mathbb{C} &= \{z : z \text{ is a complex number}\}.\end{aligned}$$

We can find various relations between sets as well as perform operations on sets. A set A is a **subset** of B, written $A \subset B$ or $B \supset A$, if every element of A is also an element of B. For example,

$$\{4, 5, 8\} \subset \{2, 3, 4, 5, 6, 7, 8, 9\}$$

and

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Trivially, every set is a subset of itself. A set B is a **proper subset** of a set A if $B \subset A$ but $B \neq A$. If A is not a subset of B, we write $A \not\subset B$; for example, $\{4, 7, 9\} \not\subset \{2, 4, 5, 8, 9\}$. Two sets are **equal**, written $A = B$, if we can show that $A \subset B$ and $B \subset A$.

It is convenient to have a set with no elements in it. This set is called the **empty set** and is denoted by $\emptyset$. Note that the empty set is a subset of every set.

To construct new sets out of old sets, we can perform certain operations: the **union** $A \cup B$ of two sets A and B is defined as

$$A \cup B = \{x : x \in A \text{ or } x \in B\};$$

the **intersection** of A and B is defined by

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

If $A = \{1, 3, 5\}$ and $B = \{1, 2, 3, 9\}$, then

$$A \cup B = \{1, 2, 3, 5, 9\} \quad \text{and} \quad A \cap B = \{1, 3\}.$$

We can consider the union and the intersection of more than two sets. In this case we write

$$\bigcup_{i=1}^{n} A_i = A_1 \cup \ldots \cup A_n$$

and

$$\bigcap_{i=1}^{n} A_i = A_1 \cap \ldots \cap A_n$$

for the union and intersection, respectively, of the sets $A_1, \ldots, A_n$.

Activity 1.2 Suppose that

$$\begin{aligned}A &= \{x : x \in \mathbb{N} \text{ and } x \text{ is even}\},\\ B &= \{x : x \in \mathbb{N} \text{ and } x \text{ is prime}\},\\ C &= \{x : x \in \mathbb{N} \text{ and } x \text{ is a multiple of } 5\}.\end{aligned}$$

Describe each of the following sets.

(a) $A \cap B$

(b) $B \cap C$

(c) $A \cup B$

(d) $A \cap (B \cup C)$

When two sets have no elements in common, they are said to be **disjoint**; for example, if E is the set of even integers and O is the set of odd integers, then E and O are disjoint. Two sets A and B are disjoint exactly when $A \cap B = \emptyset$.

Sometimes we will work within one fixed set U, called the **universal set**. For any set $A \subset U$, we define the **complement** of A, denoted by A', to be the set

$$A' = \{x : x \in U \text{ and } x \notin A\}.$$

We define the **difference** of two sets A and B to be

$$A \setminus B = A \cap B' = \{x : x \in A \text{ and } x \notin B\}.$$

Example 1.1 Let $\mathbb{R}$ be the universal set and suppose that

$$A = \{x \in \mathbb{R} : 0 < x \leq 3\} \quad \text{and} \quad B = \{x \in \mathbb{R} : 2 \leq x < 4\}.$$

Then

$$\begin{aligned} A \cap B &= \{x \in \mathbb{R} : 2 \leq x \leq 3\} \\ A \cup B &= \{x \in \mathbb{R} : 0 < x < 4\} \\ A \setminus B &= \{x \in \mathbb{R} : 0 < x < 2\} \\ A' &= \{x \in \mathbb{R} : x \leq 0 \text{ or } x > 3\}. \end{aligned}$$

□

Proposition 1.2 *Let A, B, and C be sets. Then*

1. $A \cup A = A$, $A \cap A = A$,*and* $A \setminus A = \emptyset$;
2. $A \cup \emptyset = A$ *and* $A \cap \emptyset = \emptyset$;
3. $A \cup (B \cup C) = (A \cup B) \cup C$ *and* $A \cap (B \cap C) = (A \cap B) \cap C$;
4. $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$;
5. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
6. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof. We will prove (1) and (3) and leave the remaining results to be proven in the activities and exercises.

(1) Observe that

$$\begin{aligned} A \cup A &= \{x : x \in A \text{ or } x \in A\} \\ &= \{x : x \in A\} \\ &= A \end{aligned}$$

and

$$\begin{aligned} A \cap A &= \{x : x \in A \text{ and } x \in A\} \\ &= \{x : x \in A\} \\ &= A. \end{aligned}$$

Also, $A \setminus A = A \cap A' = \emptyset$.

(3) For sets A, B, and C,

$$\begin{aligned} A \cup (B \cup C) &= A \cup \{x : x \in B \text{ or } x \in C\} \\ &= \{x : x \in A \text{ or } x \in B, \text{ or } x \in C\} \\ &= \{x : x \in A \text{ or } x \in B\} \cup C \end{aligned}$$

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

A similar argument proves that $A \cap (B \cap C) = (A \cap B) \cap C$. ■

Theorem 1.3 De Morgan's Laws. *Let A and B be sets.Then*

1. $(A \cup B)' = A' \cap B'$;

2. $(A \cap B)' = A' \cup B'$.

Proof. (1) If $A \cup B = \emptyset$, then the theorem follows immediately since both A and B are the empty set. Otherwise, we must show that $(A \cup B)' \subset A' \cap B'$ and $(A \cup B)' \supset A' \cap B'$. Let $x \in (A \cup B)'$. Then $x \notin A \cup B$. So x is neither in A nor in B, by the definition of the union of sets. By the definition of the complement, $x \in A'$ and $x \in B'$. Therefore, $x \in A' \cap B'$ and we have $(A \cup B)' \subset A' \cap B'$.

To show the reverse inclusion, suppose that $x \in A' \cap B'$. Then $x \in A'$ and $x \in B'$, and so $x \notin A$ and $x \notin B$. Thus $x \notin A \cup B$ and so $x \in (A \cup B)'$. Hence, $(A \cup B)' \supset A' \cap B'$ and so $(A \cup B)' = A' \cap B'$.

The proof of (2) is left as an exercise (Activity 1.3). ■

Example 1.4 Other relations between sets often hold true. For example,

$$(A \setminus B) \cap (B \setminus A) = \emptyset.$$

To see that this is true, observe that

$$\begin{aligned}(A \setminus B) \cap (B \setminus A) &= (A \cap B') \cap (B \cap A')\\ &= A \cap A' \cap B \cap B'\\ &= \emptyset.\end{aligned}$$

□

Activity 1.3 Prove the following statements.

(a) $A \cup \emptyset = A$ and $A \cap \emptyset = \emptyset$

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

(c) $A \subset B$ if and only if $A \cap B = A$

(d) $(A \cap B)' = A' \cup B'$

Reading Questions

1. Consider the sets $A = \{x : x - 3 \in \mathbb{N}\}$ and $B = \{x : \frac{x}{2} \in \mathbb{N}\}$. Find (describe) the sets $A \cap B$ and $A \cup B$.

2. What does the statement $A \cap B \subset A \cup B$ mean? Is this true? Explain why or why not in your own words.

3. What does it mean for two sets to be **disjoint**? Describe in words and in symbols, and give an example and non-example.

4. What do we mean by the word "proper" in the term "proper subset"? Explain what this means in your own words. Illustrate by giving an example of a proper subset of $A = \{1, 2, 3\}$ and a subset of A that is not a proper subset of A.

Exercises

1. Prove $A \cup B = B \cup A$ and $A \cap B = B \cap A$.

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

3. Prove $A \cup B = (A \cap B) \cup (A \setminus B) \cup (B \setminus A)$.

Hint. $(A \cap B) \cup (A \setminus B) \cup (B \setminus A) = (A \cap B) \cup (A \cap B') \cup (B \cap A') = [A \cap (B \cup B')] \cup (B \cap A') = A \cup (B \cap A') = (A \cup B) \cap (A \cup A') = A \cup B$.

4. Prove $(A \cup B) \times C = (A \times C) \cup (B \times C)$.

5. Prove $(A \cap B) \setminus B = \emptyset$.

6. Prove $(A \cup B) \setminus B = A \setminus B$.

7. Prove $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$.

Hint. $A \setminus (B \cup C) = A \cap (B \cup C)' = (A \cap A) \cap (B' \cap C') = (A \cap B') \cap (A \cap C') = (A \setminus B) \cap (A \setminus C)$.

8. Prove $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$.

9. Prove $(A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.

1.2 Cartesian Products and Mappings

Given sets A and B, we can define a new set $A \times B$,called the **Cartesian product** of A and B, as a set of ordered pairs. That is,

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

Example 1.5 If $A = \{x, y\}$, $B = \{1, 2, 3\}$, and $C = \emptyset$, then $A \times B$ is the set

$$\{(x, 1), (x, 2), (x, 3), (y, 1), (y, 2), (y, 3)\}$$

and

$$A \times C = \emptyset.$$

□

We define the **Cartesian product of n sets** to be

$$A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) : a_i \in A_i \text{ for } i = 1, \ldots, n\}.$$

If $A = A_1 = A_2 = \cdots = A_n$, we often write A^n for $A \times \cdots \times A$ (where A would be written n times). For example, the set $\mathbb{R}^3$ consists of all of 3-tuples of real numbers.

Subsets of $A \times B$ are called **relations**. We will define a **mapping** or **function** $f \subset A \times B$ from a set A to a set B to be the special type of relation where $(a, b) \in f$ if for every element $a \in A$ there exists a unique element $b \in B$. Another way of saying this is that for every element in A, f assigns a unique element in B. We usually write $f : A \to B$ or $A \xrightarrow{f} B$. Instead of writing down ordered pairs $(a, b) \in A \times B$, we write $f(a) = b$ or $f : a \mapsto b$. The set A is called the **domain** of f and

$$f(A) = \{f(a) : a \in A\} \subset B$$

is called the **range** or **image** of f. We can think of the elements in the function's domain as input values and the elements in the function's range as output values.

Example 1.6 Suppose $A = \{1, 2, 3\}$ and $B = \{a, b, c\}$. In Figure 1.7 we define relations f and g from A to B. The relation f is a mapping, but g is not because $1 \in A$ is not assigned to a unique element in B; that is, $g(1) = a$ and $g(1) = b$.

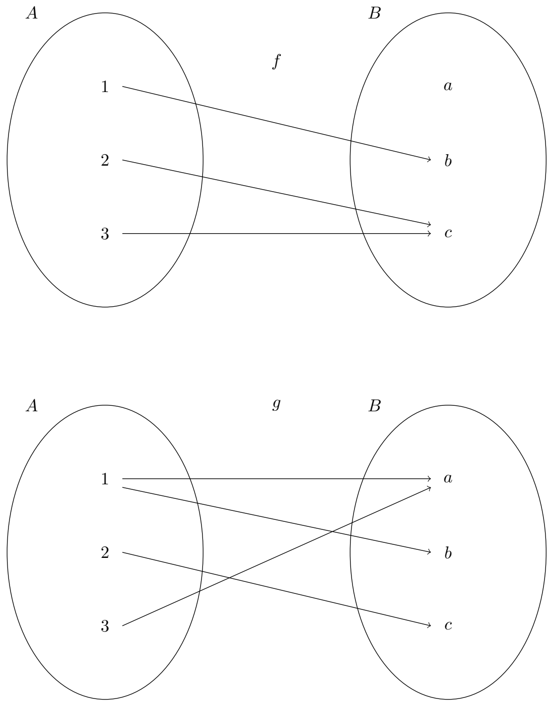


Figure 1.7 Mappings and relations

□

Given a function $f : A \to B$, it is often possible to write a list describing what the function does to each specific element in the domain. However, not all functions can be described in this manner. For example, the function $f : \mathbb{R} \to \mathbb{R}$ that sends each real number to its cube is a mapping that must be described by writing $f(x) = x^3$ or $f : x \mapsto x^3$.

Consider the relation $f : \mathbb{Q} \to \mathbb{Z}$ given by $f(p/q) = p$. We know that $1/2 = 3/6$, but is $f(1/2) = 1$ or 3? This relation cannot be a mapping because it is not well-defined. A relation is **well-defined** if each element in the domain is assigned to a *unique* element in the range.

If $f : A \to B$ is a map and the image of f is B, i.e., $f(A) = B$, then f is said to be **onto** or **surjective**. In other words, if there exists an $a \in A$ for each $b \in B$ such that $f(a) = b$, then f is onto. A map is **one-to-one** or **injective** if $a_1 \neq a_2$ implies $f(a_1) \neq f(a_2)$. Equivalently, a function is one-to-one if $f(a_1) = f(a_2)$ implies $a_1 = a_2$. A map that is both one-to-one and onto is called **bijective**.

Example 1.8 Let $f : \mathbb{Z} \to \mathbb{Q}$ be defined by $f(n) = n/1$. Then f is one-to-one but not onto. Define $g : \mathbb{Q} \to \mathbb{Z}$ by $g(p/q) = p$ where p/q is a rational number expressed in its lowest terms with a positive denominator. The function g is onto but not one-to-one. □

Activity 1.4 Which of the following relations $f : \mathbb{Q} \to \mathbb{Q}$ define a mapping? In each case, supply a reason why f is or is not a mapping.

(a) $f(p/q) = \dfrac{p+1}{p-2}$

(b) $f(p/q) = \dfrac{3p}{3q}$

(c) $f(p/q) = \dfrac{p+q}{q^2}$

(d) $f(p/q) = \dfrac{3p^2}{7q^2} - \dfrac{p}{q}$

Given two functions, we can construct a new function by using the range of the first function as the domain of the second function. Let $f : A \to B$ and $g : B \to C$ be mappings. Define a new map, the **composition** of f and g from A to C, by $(g \circ f)(x) = g(f(x))$.

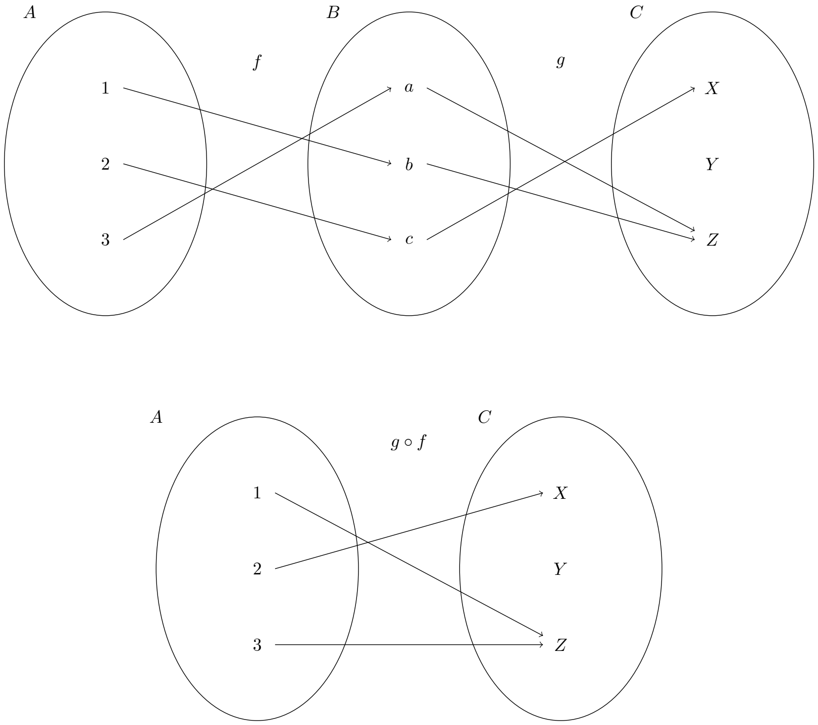


Figure 1.9 Composition of maps

Example 1.10 Consider the functions $f : A \to B$ and $g : B \to C$ that are defined in Figure 1.9 (top). The composition of these functions, $g \circ f : A \to C$, is defined in Figure 1.9 (bottom). □

Example 1.11 Let $f(x) = x^2$ and $g(x) = 2x + 5$. Then

$$(f \circ g)(x) = f(g(x)) = (2x + 5)^2 = 4x^2 + 20x + 25$$

and

$$(g \circ f)(x) = g(f(x)) = 2x^2 + 5.$$

In general, order makes a difference; that is, in most cases $f \circ g \neq g \circ f$. □

Example 1.12 Sometimes it is the case that $f \circ g = g \circ f$. Let $f(x) = x^3$ and $g(x) = \sqrt[3]{x}$. Then

$$(f \circ g)(x) = f(g(x)) = f(\sqrt[3]{x}\,) = (\sqrt[3]{x}\,)^3 = x$$

and

$$(g \circ f)(x) = g(f(x)) = g(x^3) = \sqrt[3]{x^3} = x.$$

□

Example 1.13 Given a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

we can define a map $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T_A(x, y) = (ax + by, cx + dy)$$

for (x, y) in $\mathbb{R}^2$. This is actually matrix multiplication; that is,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

Maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by matrices are called **linear maps** or **linear transformations**. □

Example 1.14 Suppose that $S = \{1, 2, 3\}$. Define a map $\pi : S \to S$ by

$$\pi(1) = 2, \qquad \pi(2) = 1, \qquad \pi(3) = 3.$$

This is a bijective map. An alternative way to write π is

$$\begin{pmatrix} 1 & 2 & 3 \\ \pi(1) & \pi(2) & \pi(3) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

Note that this is NOT a matrix, but simply a way to write a function using so-called **two-line notation**: the top row is the domain, the bottom row gives the image of each element above it.

For any set S, a one-to-one and onto mapping $\pi : S \to S$ is called a **permutation** of S. □

Theorem 1.15 *Let $f : A \to B$, $g : B \to C$, and $h : C \to D$. Then*

1. *The composition of mappings is associative; that is, $(h \circ g) \circ f = h \circ (g \circ f)$;*
2. *If f and g are both one-to-one, then the mapping $g \circ f$ is one-to-one;*
3. *If f and g are both onto, then the mapping $g \circ f$ is onto;*
4. *If f and g are bijective, then so is $g \circ f$.*

Proof. We will prove (1) and (3). Part (2) is left as an exercise. Part (4) follows directly from (2) and (3).

(1) We must show that

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

For $a \in A$ we have

$$\begin{aligned} (h \circ (g \circ f))(a) &= h((g \circ f)(a)) \\ &= h(g(f(a))) \\ &= (h \circ g)(f(a)) \\ &= ((h \circ g) \circ f)(a). \end{aligned}$$

(3) Assume that f and g are both onto functions. Given $c \in C$, we must show that there exists an $a \in A$ such that $(g \circ f)(a) = g(f(a)) = c$. However, since g is onto, there is an element $b \in B$ such that $g(b) = c$. Similarly, there is an $a \in A$ such that $f(a) = b$. Accordingly,

$$(g \circ f)(a) = g(f(a)) = g(b) = c.$$

■

Activity 1.5 Let $f : A \to B$ and $g : B \to C$ be maps.

(a) If f and g are both one-to-one functions, show that $g \circ f$ is one-to-one.

(b) If $g \circ f$ is onto, show that g is onto.

If S is any set, we will use id_S or id to denote the **identity mapping** from S to itself. Define this map by $id(s) = s$ for all $s \in S$. A map $g : B \to A$ is an **inverse mapping** of $f : A \to B$ if $g \circ f = id_A$ and $f \circ g = id_B$; in other words, the inverse function of a function simply "undoes" the function. A map is said to be **invertible** if it has an inverse. We usually write f^{-1} for the inverse of f.

Example 1.16 The function $f(x) = x^3$ has inverse $f^{-1}(x) = \sqrt[3]{x}$ by Example 1.12. □

Example 1.17 The natural logarithm and the exponential functions, $f(x) = \ln x$ and $f^{-1}(x) = e^x$, are inverses of each other provided that we are careful about choosing domains. Observe that

$$f(f^{-1}(x)) = f(e^x) = \ln e^x = x$$

and

$$f^{-1}(f(x)) = f^{-1}(\ln x) = e^{\ln x} = x$$

whenever composition makes sense. □

Example 1.18 Suppose that

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

Then A defines a map from $\mathbb{R}^2$ to $\mathbb{R}^2$ by

$$T_A(x, y) = (3x + y, 5x + 2y).$$

We can find an inverse map of T_A by simply inverting the matrix A; that is, $T_A^{-1} = T_{A^{-1}}$. In this example,

$$A^{-1} = \begin{pmatrix} 2 & -1 \\ -5 & 3 \end{pmatrix};$$

hence, the inverse map is given by

$$T_A^{-1}(x, y) = (2x - y, -5x + 3y).$$

It is easy to check that

$$T_A^{-1} \circ T_A(x, y) = T_A \circ T_A^{-1}(x, y) = (x, y).$$

Not every map has an inverse. If we consider the map

$$T_B(x, y) = (3x, 0)$$

given by the matrix

$$B = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix},$$

then an inverse map would have to be of the form

$$T_B^{-1}(x, y) = (ax + by, cx + dy)$$

and

$$(x, y) = T_B \circ T_B^{-1}(x, y) = (3ax + 3by, 0)$$

for all x and y. Clearly this is impossible because y might not be 0. □

Example 1.19 Given the permutation

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

on $S = \{1, 2, 3\}$, it is easy to see that the permutation defined by

$$\pi^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

is the inverse of π. In fact, any bijective mapping possesses an inverse, as we will see in the next theorem. □

Theorem 1.20 *A mapping is invertible if and only if it is both one-to-one and onto.*

Proof. Suppose first that $f : A \to B$ is invertible with inverse $g : B \to A$. Then $g \circ f = id_A$ is the identity map; that is, $g(f(a)) = a$. If $a_1, a_2 \in A$ with $f(a_1) = f(a_2)$, then $a_1 = g(f(a_1)) = g(f(a_2)) = a_2$. Consequently, f is one-to-one. Now suppose that $b \in B$. To show that f is onto, it is necessary to find an $a \in A$ such that $f(a) = b$, but $f(g(b)) = b$ with $g(b) \in A$. Let $a = g(b)$.

Conversely, let f be bijective and let $b \in B$. Since f is onto, there exists an $a \in A$ such that $f(a) = b$. Because f is one-to-one, a must be unique. Define g by letting $g(b) = a$. We have now constructed the inverse of f. ■

Reading Questions

1. If the set A contains 4 elements, and the set B contains 6 elements, how many elements with the set $A \times B$ contain? Explain why your answer makes sense.

2. Let $A = \{1, 2\}$ and $B = \{1, 2, 3, 4\}$. One function $f : A \to B$ might be defined by $f(x) = 2x$. Write this function as a **relation**, that is, as a subset of $A \times B$.

3. Your friend makes the bold claim that an alternative definition for a function to be one-to-one is that it is *not* onto. Is your friend correct? Explain why or why not.

Exercises

1. If $A = \{a, b, c\}$, $B = \{1, 2, 3\}$, $C = \{x\}$, and $D = \emptyset$, list all of the elements in each of the following sets.

(a) $A \times B$

(b) $B \times A$

(c) $A \times B \times C$

(d) $A \times D$

Hint. (a) $A \times B = \{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3), (c, 1), (c, 2), (c, 3)\}$; (d) $A \times D = \emptyset$.

2. Find an example of two nonempty sets A and B for which $A \times B = B \times A$ is true.

3. Determine which of the following functions are one-to-one and which are onto. If the function is not onto, determine its range.

(a) $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = e^x$

(b) $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = n^2 + 3$

(c) $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sin x$

(d) $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = x^2$

Hint. (a) f is one-to-one but not onto. $f(\mathbb{R}) = \{x \in \mathbb{R} : x > 0\}$. (c) f is neither one-to-one nor onto. $f(\mathbb{R}) = \{x : -1 \leq x \leq 1\}$.

4. Let $f : A \to B$ and $g : B \to C$ be invertible mappings; that is, mappings such that f^{-1} and g^{-1} exist. Show that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

5.

(a) Define a function $f : \mathbb{N} \to \mathbb{N}$ that is one-to-one but not onto.

(b) Define a function $f : \mathbb{N} \to \mathbb{N}$ that is onto but not one-to-one.

Hint. (a) $f(n) = n + 1$.

6. Let $f : A \to B$ and $g : B \to C$ be maps.

(a) If $g \circ f$ is one-to-one, show that f is one-to-one.

(b) If $g \circ f$ is one-to-one and f is onto, show that g is one-to-one.

(c) If $g \circ f$ is onto and g is one-to-one, show that f is onto.

Hint. (a) Let $x, y \in A$. Then $g(f(x)) = (g \circ f)(x) = (g \circ f)(y) = g(f(y))$. Thus, $f(x) = f(y)$ and $x = y$, so $g \circ f$ is one-to-one. (b) Let $c \in C$, then $c = (g \circ f)(x) = g(f(x))$ for some $x \in A$. Since $f(x) \in B$, g is onto.

7. Define a function on the real numbers by

$$f(x) = \frac{x+1}{x-1}.$$

What are the domain and range of f? What is the inverse of f? Compute $f \circ f^{-1}$ and $f^{-1} \circ f$.

Hint. $f^{-1}(x) = (x + 1)/(x - 1)$.

1.3 Equivalence Relations and Partitions

A fundamental notion in mathematics is that of equality. We can generalize equality with equivalence relations and equivalence classes. An **equivalence relation** on a set X is a relation $R \subset X \times X$ such that

- $(x, x) \in R$ for all $x \in X$ (**reflexive property**);
- $(x, y) \in R$ implies $(y, x) \in R$ (**symmetric property**);
- (x, y) and $(y, z) \in R$ imply $(x, z) \in R$ (**transitive property**).

Given an equivalence relation R on a set X, we usually write $x \sim y$ instead of $(x, y) \in R$. If the equivalence relation already has an associated notation such as $=$, $\equiv$, or $\cong$, we will use that notation.

Example 1.21 Let p, q, r, and s be integers, where q and s are nonzero. Define $p/q \sim r/s$ if $ps = qr$. Clearly $\sim$ is reflexive and symmetric. To show that it is also transitive, suppose that $p/q \sim r/s$ and $r/s \sim t/u$, with q, s, and u all nonzero. Then $ps = qr$ and $ru = st$. Therefore,

$$psu = qru = qst.$$

Since $s \neq 0$, $pu = qt$. Consequently, $p/q \sim t/u$. □

Example 1.22 Suppose that f and g are differentiable functions on $\mathbb{R}$. We can define an equivalence relation on such functions by letting $f(x) \sim g(x)$ if $f'(x) = g'(x)$. It is clear that $\sim$ is both reflexive and symmetric. To demonstrate transitivity, suppose that $f(x) \sim g(x)$ and $g(x) \sim h(x)$. From calculus we know that $f(x) - g(x) = c_1$ and $g(x) - h(x) = c_2$, where c_1 and c_2 are both constants. Hence,

$$f(x) - h(x) = (f(x) - g(x)) + (g(x) - h(x)) = c_1 + c_2$$

and $f'(x) - h'(x) = 0$. Therefore, $f(x) \sim h(x)$. □

Example 1.23 For (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$, define $(x_1, y_1) \sim (x_2, y_2)$ if $x_1^2 + y_1^2 = x_2^2 + y_2^2$. Then $\sim$ is an equivalence relation on $\mathbb{R}^2$. □

Example 1.24 Let A and B be 2×2 matrices with entries in the real numbers. We can define an equivalence relation on the set of 2×2 matrices, by saying $A \sim B$ if there exists an invertible matrix P such that $PAP^{-1} = B$. For example, if

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -18 & 33 \\ -11 & 20 \end{pmatrix},$$

then $A \sim B$ since $PAP^{-1} = B$ for

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

Let I be the 2×2 identity matrix; that is,

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

Then $IAI^{-1} = IAI = A$; therefore, the relation is reflexive. To show symmetry, suppose that $A \sim B$. Then there exists an invertible matrix P such that $PAP^{-1} = B$. So

$$A = P^{-1}BP = P^{-1}B(P^{-1})^{-1}.$$

Finally, suppose that $A \sim B$ and $B \sim C$. Then there exist invertible matrices P and Q such that $PAP^{-1} = B$ and $QBQ^{-1} = C$. Since

$$C = QBQ^{-1} = QPAP^{-1}Q^{-1} = (QP)A(QP)^{-1},$$

the relation is transitive. Two matrices that are equivalent in this manner are said to be **similar**. □

A **partition** $\mathcal{P}$ of a set X is a collection of nonempty sets $X_1, X_2, \ldots$ such that $X_i \cap X_j = \emptyset$ for $i \neq j$ and $\bigcup_k X_k = X$. Let $\sim$ be an equivalence relation on a set X and let $x \in X$. Then $[x] = \{y \in X : y \sim x\}$ is called the **equivalence class** of x. We will see that an equivalence relation gives rise to a partition via equivalence classes. Also, whenever a partition of a set exists, there is some natural underlying equivalence relation, as the following theorem demonstrates.

Theorem 1.25 *Given an equivalence relation $\sim$ on a set X, the equivalence classes of X form a partition of X. Conversely, if $\mathcal{P} = \{X_i\}$ is a partition of a set X, then there is an equivalence relation on X with equivalence classes X_i.*

Proof. Suppose there exists an equivalence relation $\sim$ on the set X. For any $x \in X$, the reflexive property shows that $x \in [x]$ and so $[x]$ is nonempty. Clearly $X = \bigcup_{x \in X} [x]$. Now let $x, y \in X$. We need to show that either $[x] = [y]$ or $[x] \cap [y] = \emptyset$. Suppose that the intersection of $[x]$ and $[y]$ is not empty and that $z \in [x] \cap [y]$. Then $z \sim x$ and $z \sim y$. By symmetry and transitivity $x \sim y$; hence, $[x] \subset [y]$. Similarly, $[y] \subset [x]$ and so $[x] = [y]$. Therefore, any two equivalence classes are either disjoint or exactly the same.

Conversely, suppose that $\mathcal{P} = \{X_i\}$ is a partition of a set X. Let two elements be equivalent if they are in the same partition. Clearly, the relation is reflexive. If x is in the same partition as y, then y is in the same partition as x, so $x \sim y$ implies $y \sim x$. Finally, if x is in the same partition as y and y is in the same partition as z, then x must be in the same partition as z, and transitivity holds. ■

Corollary 1.26 *Two equivalence classes of an equivalence relation are either disjoint or equal.*

Let us examine some of the partitions given by the equivalence classes in the last set of examples.

Example 1.27 In the equivalence relation in Example 1.21, two pairs of integers, (p, q) and (r, s), are in the same equivalence class when they reduce to the same fraction in its lowest terms. □

Example 1.28 In the equivalence relation in Example 1.22, two functions $f(x)$ and $g(x)$ are in the same partition when they differ by a constant. □

Example 1.29 We defined an equivalence class on $\mathbb{R}^2$ by $(x_1, y_1) \sim (x_2, y_2)$ if $x_1^2 + y_1^2 = x_2^2 + y_2^2$. Two pairs of real numbers are in the same partition when they lie on the same circle about the origin. □

Example 1.30 Let r and s be two integers and suppose that $n \in \mathbb{N}$. We say that r is **congruent** to s **modulo** n, or r is congruent to s mod n, if $r - s$ is evenly divisible by n; that is, $r - s = nk$ for some $k \in \mathbb{Z}$. In this case we write $r \equiv s \pmod{n}$. For example, $41 \equiv 17 \pmod{8}$ since $41 - 17 = 24$ is divisible by 8. We claim that congruence modulo n forms an equivalence relation of $\mathbb{Z}$. Certainly any integer r is equivalent to itself since $r - r = 0$ is divisible by n. We will now show that the relation is symmetric. If $r \equiv s \pmod{n}$, then $r - s = -(s - r)$ is divisible by n. So $s - r$ is divisible by n and $s \equiv r \pmod{n}$. Now suppose that $r \equiv s \pmod{n}$ and $s \equiv t \pmod{n}$. Then there exist integers k and l such that $r - s = kn$ and $s - t = ln$. To show transitivity, it is necessary to prove that $r - t$ is divisible by n. However,

$$r - t = r - s + s - t = kn + ln = (k + l)n,$$

and so $r - t$ is divisible by n.

If we consider the equivalence relation established by the integers modulo 3, then

$$\begin{aligned}
[0] &= \{\ldots, -3, 0, 3, 6, \ldots\},\\
[1] &= \{\ldots, -2, 1, 4, 7, \ldots\},\\
[2] &= \{\ldots, -1, 2, 5, 8, \ldots\}.
\end{aligned}$$

Notice that $[0] \cup [1] \cup [2] = \mathbb{Z}$ and also that the sets are disjoint. The sets $[0]$, $[1]$, and $[2]$ form a partition of the integers.

The integers modulo n are a very important example in the study of abstract algebra and will become quite useful in our investigation of various algebraic structures such as groups and rings. In our discussion of the integers modulo n we have actually assumed a result known as the division algorithm, which will be stated and proved in Chapter 2. □

Activity 1.6 Determine whether or not the following relations are equivalence relations on the given set. If the relation is an equivalence relation, describe the partition given by it. If the relation is not an equivalence relation, state why it fails to be one.

(a) $x \sim y$ in $\mathbb{R}$ if $x \geq y$

(b) $m \sim n$ in $\mathbb{Z}$ if $mn > 0$

(c) $x \sim y$ in $\mathbb{R}$ if $|x - y| \leq 4$

(d) $m \sim n$ in $\mathbb{Z}$ if $m \equiv n \pmod{6}$

Reading Questions

1. Consider the relation $R \subset \mathbb{Z} \times \mathbb{Z}$ that holds of (x, y) precisely when $x - y = 3$. Is R an equivalence relation? Not not, what specific properties of an equivalence relation does it not satisfy?

2. Consider a set X and a partition $\mathcal{P}$ of X. Is it possible for an element $a \in X$ to belong to two distinct sets of $\mathcal{P}$? Briefly explain.

3. Consider the equivalence relation established by the integers modulo 5. What does the notation $[4]$ mean? Be as specific as you can.

Exercises

1. Prove the relation defined on $\mathbb{R}^2$ by $(x_1, y_1) \sim (x_2, y_2)$ if $x_1^2 + y_1^2 = x_2^2 + y_2^2$ is an equivalence relation.

2. Define a relation $\sim$ on $\mathbb{R}^2$ by stating that $(a, b) \sim (c, d)$ if and only if $a^2 + b^2 \leq c^2 + d^2$. Show that $\sim$ is reflexive and transitive but not symmetric.

3. **Projective Real Line.** Define a relation on $\mathbb{R}^2 \setminus \{(0, 0)\}$ by letting $(x_1, y_1) \sim (x_2, y_2)$ if there exists a nonzero real number λ such that $(x_1, y_1) = (\lambda x_2, \lambda y_2)$. Prove that $\sim$ defines an equivalence relation on $\mathbb{R}^2 \setminus (0, 0)$. What are the corresponding equivalence classes? This equivalence relation defines the projective line, denoted by $\mathbb{P}(\mathbb{R})$, which is very important in geometry.

1.4 Summary and Additional Exercises

1.4.1 The Important Ideas

- A **set** is a well-defined collection of objects. The objects that belong to a set are called its **elements** or **members**. We will denote sets by capital letters, such as A or X; if a is an element of the set A, we write $a \in A$.

- A set A is a **subset** of B, written $A \subset B$ or $B \supset A$, if every element of A is also an element of B. Trivially, every set is a subset of itself. A set B is a **proper subset** of a set A if $B \subset A$ but $B \neq A$. If A is not a subset of B, we write $A \not\subset B$. Two sets are **equal**, written $A = B$, if we can show that $A \subset B$ and $B \subset A$. The **empty set** contains no elements of the set and is denoted by $\emptyset$. The empty set is a subset of every set.

- To construct new sets out of old sets, we can perform certain operations: the **union** $A \cup B$ of two sets A and B is defined as

$$A \cup B = \{x : x \in A \text{ or } x \in B\};$$

the **intersection** of A and B is defined by

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

The set U, called the **universal set**, is the set of all elements under discussion. For any set $A \subset U$, we define the **complement** of A, denoted by A', to be the set

$$A' = \{x : x \in U \text{ and } x \notin A\}.$$

- Let A, B, and C be sets. Then
  1. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
  2. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
  3. $(A \cup B)' = A' \cap B'$;
  4. $(A \cap B)' = A' \cup B'$.
- Subsets of $A \times B$ are called **relations**. We will define a **mapping** or **function** $f \subset A \times B$ from a set A to a set B to be the special type of relation where $(a, b) \in f$ if for every element $a \in A$ there exists a unique element $b \in B$. Another way of saying this is that for every element in A, f assigns a unique element in B. We usually write $f : A \to B$ or $A \xrightarrow{f} B$. Instead of writing down ordered pairs $(a, b) \in A \times B$, we write $f(a) = b$ or $f : a \mapsto b$. The set A is called the **domain** of f and

  $$f(A) = \{f(a) : a \in A\} \subset B$$

  is called the **range** or **image** of f.
- If $f : A \to B$ is a map and $f(A) = B$, then f is said to be **onto** or **surjective**. In other words, if there exists an $a \in A$ for each $b \in B$ such that $f(a) = b$, then f is onto. A map is **one-to-one** or **injective** if $f(a_1) = f(a_2)$ implies $a_1 = a_2$. A map that is both one-to-one and onto is called **bijective**. A mapping is invertible if and only if it is both one-to-one and onto.
- An **equivalence relation** on a set X is a relation $R \subset X \times X$ such that
  - ∘ $(x, x) \in R$ for all $x \in X$ (**reflexive property**);
  - ∘ $(x, y) \in R$ implies $(y, x) \in R$ (**symmetric property**);
  - ∘ (x, y) and $(y, z) \in R$ imply $(x, z) \in R$ (**transitive property**).
- A **partition** $\mathcal{P}$ of a set X is a collection of nonempty sets $X_1, X_2, \ldots$ such that $X_i \cap X_j = \emptyset$ for $i \neq j$ and $\bigcup_k X_k = X$. Let $\sim$ be an equivalence relation on a set X and let $x \in X$. Then $[x] = \{y \in X : y \sim x\}$ is called the **equivalence class** of x.
- An equivalence relation gives rise to a partition via equivalence classes, and whenever a partition of a set exists, there is a natural underlying equivalence relation.

1.4.2 Additional Exercises

1. Let $f : X \to Y$ be a map with $A_1, A_2 \subset X$ and $B_1, B_2 \subset Y$.

(a) Prove $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.

(b) Prove $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$. Give an example in which equality fails.

(c) Prove $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$, where

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

(d) Prove $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

(e) Prove $f^{-1}(Y \setminus B_1) = X \setminus f^{-1}(B_1)$.

Hint. (a) Let $y \in f(A_1 \cup A_2)$. Then there exists an $x \in A_1 \cup A_2$ such that $f(x) = y$. Hence, $y \in f(A_1)$ or $f(A_2)$. Therefore, $y \in f(A_1) \cup f(A_2)$. Consequently, $f(A_1 \cup A_2) \subset f(A_1) \cup f(A_2)$. Conversely, if $y \in f(A_1) \cup f(A_2)$, then $y \in f(A_1)$ or $f(A_2)$. Hence, there exists an x in A_1 or A_2 such that $f(x) = y$. Thus, there exists an $x \in A_1 \cup A_2$ such that $f(x) = y$. Therefore, $f(A_1) \cup f(A_2) \subset f(A_1 \cup A_2)$, and $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.

2. Find the error in the following argument by providing a counterexample. "The reflexive property is redundant in the axioms for an equivalence relation. If $x \sim y$, then $y \sim x$ by the symmetric property. Using the transitive property, we can deduce that $x \sim x$."

Hint. Let $X = \mathbb{N} \cup \{\sqrt{2}\}$ and define $x \sim y$ if $x + y \in \mathbb{N}$.

1.5 Connections to the Secondary Classroom—The Vertical Line Test

You will often see the "vertical line test" in many high school textbooks:

> If any vertical line intersects a graph more than once, then the graph is not a function.

We may also encounter the "horizontal line test":

> A function f has an inverse function if any possible horizontal line can intersect the graph of f at most once

or

> A function f is one-to-one if any possible horizontal line can intersect the graph of f at most once.

As we usually think of the graph of a function as a picture, both the vertical and horizontal line tests might be a bit problematic for functions or relations such as the ones in Example 1.6, Example 1.18, or Example 1.14. Or how do we describe the graph of a function or a relation to a blind student? Clearly, we need a better definition of the graph of a function.

Recall that we defined a relation between set A and B to be a subset of $A \times B$. A function from A to B can then be defined as a subset f of $A \times B$ such that

$$f = \{(a, b) : a \in A \text{ and for each } a \text{ there exists a unique } b\}.$$

So the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ can be written as $\{(x, x^2) : x \in \mathbb{R}\}$. The unit circle $x^2 + y^2 = 1$ can be written as $\{(x, y) : x^2 + y^2 = 1\}$.

We define the graph of a relation $R \subset A \times B$ to be the set of ordered pairs $(a, b) \in R$. Thus, the graph of a function $f : A \to B$ is the set of ordered pairs $(x, f(x))$.

Exercises

1. Rigorously explain the vertical line test in terms of the formal definition of the graph of a relation.

2. Rigorously explain the horizontal line test in terms of the formal definition of the graph of a relation.

1.6 References and Suggested Readings

[1] Artin, M. *Abstract Algebra.* 2nd ed. Pearson, Upper Saddle River, NJ, 2011.

[2] Childs, L. *A Concrete Introduction to Higher Algebra.* 2nd ed. Springer-Verlag, New York, 1995.

[3] Dummit, D. and Foote, R. *Abstract Algebra.* 3rd ed. Wiley, New York, 2003.

[4] Ehrlich, G. *Fundamental Concepts of Algebra.* PWS-KENT, Boston, 1991.

[5] Fraleigh, J. B. *A First Course in Abstract Algebra.* 7th ed. Pearson, Upper Saddle River, NJ, 2003.

[6] Gallian, J. A. *Contemporary Abstract Algebra.* 7th ed. Brooks/Cole, Belmont, CA, 2009.

[7] Halmos, P. *Naive Set Theory.* Springer, New York, 1991. One of the best references for set theory.

[8] Herstein, I. N. *Abstract Algebra.* 3rd ed. Wiley, New York, 1996.

[9] Hungerford, T. W. *Algebra.* Springer, New York, 1974. One of the standard graduate algebra texts.

[10] Lang, S. *Algebra.* 3rd ed. Springer, New York, 2002. Another standard graduate text.

[11] Lidl, R. and Pilz, G. *Applied Abstract Algebra.* 2nd ed. Springer, New York, 1998.

[12] Mackiw, G. *Applications of Abstract Algebra.* Wiley, New York, 1985.

[13] Nickelson, W. K. *Introduction to Abstract Algebra.* 3rd ed. Wiley, New York, 2006.

[14] Solow, D. *How to Read and Do Proofs.* 5th ed. Wiley, New York, 2009.

[15] van der Waerden, B. L. *A History of Algebra.* Springer-Verlag, New York, 1985. An account of the historical development of algebra.

[16] Wasserman, Nicholas, ed. *Connecting Abstract Algebra to Secondary Mathematics, for Secondary Mathematics Teachers.* Springer-Verlag, New York, 2018.

Chapter 2

The Integers

Objectives

- To understand and be able to apply the Principle of Mathematical Induction.
- To understand and be able to apply the Principle of Well Ordering and to understand how the Principle of Well Ordering and the Principle of Mathematical Induction are related.
- To understand and be able to apply the Division Algorithm.
- To understand and be able to apply the Euclidean Algorithm.
- To understand prime numbers and to understand and be able to apply the Fundamental Theorem of Arithmetic.

The integers are the building blocks of mathematics. In this chapter we will investigate the fundamental properties of the integers, including mathematical induction, the division algorithm, and the Fundamental Theorem of Arithmetic.

Suppose we wish to show that

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

for any natural number n. This formula is easily verified for small numbers such as $n = 1$, 2, 3, or 4, but it is impossible to verify for all natural numbers on a case-by-case basis. To prove the formula true in general, a more generic method is required.

Let us say that we have verified the equation for the first n cases. We will attempt to show that we can generate the formula for the $(n+1)$th case from this knowledge. The formula is true for $n = 1$ since

$$1 = \frac{1(1+1)}{2}.$$

If we have verified the first n cases, then

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= \frac{n(n+1)}{2} + n + 1 \\ &= \frac{n^2 + 3n + 2}{2} \\ &= \frac{(n+1)[(n+1)+1]}{2}. \end{aligned}$$

This is exactly the formula for the $(n+1)$th case.

Activity 2.1 Consider the formula for the sum of the first n squares,

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2.1}$$

for $n \in \mathbb{N}$.

(a) Verify (2.1) for $n = 1$.

(b) Assume (2.1) and show that

$$1^2 + 2^2 + \cdots + n^2 + (n+1)^2 = \frac{(n+1)[(n+1)+1][2(n+1)+1]}{6}.$$

2.1 Mathematical Induction

The method of proof outlined in the introduction is known as **mathematical induction**. Instead of attempting to verify a statement about some subset S of the positive integers $\mathbb{N}$ on a case-by-case basis, an impossible task if S is an infinite set, we give a specific proof for the smallest integer being considered, followed by a generic argument showing that *if* the statement holds for a given case, *then* it must also hold for the next case in the sequence. We summarize mathematical induction in the following axiom.

Principle 2.1 First Principle of Mathematical Induction. *Let $S(n)$ be a statement about integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If for all integers k with $k \geq n_0$, $S(k)$ implies that $S(k+1)$ is true, then $S(n)$ is true for all integers n greater than or equal to n_0.*

In practice, this principle is applied by first proving a **base case** and then proving the **inductive case**. To prove the base case means to prove $S(n_0)$. A proof of the inductive case is a proof of the implication $S(k) \to S(k+1)$, that is, for an arbitrary $k \geq n_0$, you assume $S(k)$ and prove $S(k+1)$. The First Principle of Mathematical Induction is the principle that proving the base case and the inductive case constitutes a proof of the statement $S(n)$ for all $n \geq n_0$.

Example 2.2 For all integers $n \geq 3$, $2^n > n + 4$. Since

$$8 = 2^3 > 3 + 4 = 7,$$

the statement is true for $n_0 = 3$. Assume that $2^k > k + 4$ for $k \geq 3$. Then $2^{k+1} = 2 \cdot 2^k > 2(k+4)$. But

$$2(k+4) = 2k + 8 > k + 5 = (k+1) + 4$$

since k is positive. Hence, by induction, the statement holds for all integers $n \geq 3$. □

Example 2.3 Every integer $10^{n+1} + 3 \cdot 10^n + 5$ is divisible by 9 for $n \in \mathbb{N}$. For $n = 1$,

$$10^{1+1} + 3 \cdot 10 + 5 = 135 = 9 \cdot 15$$

is divisible by 9. Suppose that $10^{k+1} + 3 \cdot 10^k + 5$ is divisible by 9 for $k \geq 1$. Then

$$\begin{aligned} 10^{(k+1)+1} + 3 \cdot 10^{k+1} + 5 &= 10^{k+2} + 3 \cdot 10^{k+1} + 50 - 45 \\ &= 10(10^{k+1} + 3 \cdot 10^k + 5) - 45 \end{aligned}$$

is divisible by 9. □

Example 2.4 We will prove the binomial theorem using mathematical induction; that is,

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

where a and b are real numbers, $n \in \mathbb{N}$, and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the binomial coefficient. We first show that

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

This result follows from

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\
&= \frac{(n+1)!}{k!(n+1-k)!} \\
&= \binom{n+1}{k}.
\end{aligned}$$

If $n = 1$, the binomial theorem is easy to verify. Now assume that the result is true for n greater than or equal to 1. Then

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)(a+b)^n \\
&= (a+b)\left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n+1-k} \\
&= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k-1} a^k b^{n+1-k} + \sum_{k=1}^{n} \binom{n}{k} a^k b^{n+1-k} + b^{n+1} \\
&= a^{n+1} + \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] a^k b^{n+1-k} + b^{n+1} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k}.
\end{aligned}$$

□

We have an equivalent statement of the Principle of Mathematical Induction that is often very useful.

Principle 2.5 Second Principle of Mathematical Induction. *Let $S(n)$ be a statement about integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If $S(n_0), S(n_0+1), \ldots, S(k)$ imply that $S(k+1)$ for $k \geq n_0$, then the statement $S(n)$ is true for all integers $n \geq n_0$.*

The main difference between this second principle and the first principle is how you prove the inductive case. In both versions of induction, you must prove $S(k+1)$ (for an arbitrary $k \geq n_0$). If you are applying the first principle, you get to assume $S(k)$ is true. But if you are applying the second principle, you get to assume that $S(k)$ is true, as well as $S(k-1)$ and $S(k-2)$... that $S(j)$ is true for all j starting at n_0 up through k.

Both principles of mathematical induction are equivalent to a statement about subsets of natural numbers called the Principle of Well Ordering.

A nonempty subset S of $\mathbb{Z}$ is **well-ordered** if S contains a least element. Notice that the set $\mathbb{Z}$ itself is not well-ordered since it does not contain a smallest element. However, the natural numbers are well-ordered.

Principle 2.6 Principle of Well-Ordering. *Every nonempty subset of the natural numbers is well-ordered.*

We will show that Principle of Well-Ordering follows from the Principle of Mathematical Induction. The converse is also true, although we will not prove it here.

Lemma 2.7 *The Principle of Mathematical Induction implies that* 1 *is the least positive natural number.*

Proof. Let $S = \{n \in \mathbb{N} : n \geq 1\}$. Then $1 \in S$. Assume that $n \in S$. Since $0 < 1$, it must be the case that $n = n + 0 < n + 1$. Therefore, $1 \leq n < n + 1$. Consequently, if $n \in S$, then $n + 1$ must also be in S, and by the Principle of Mathematical Induction, and $S = \mathbb{N}$. ■

Theorem 2.8 *The Principle of Mathematical Induction implies the Principle of Well-Ordering. That is, every nonempty subset of* $\mathbb{N}$ *contains a least element.*

Proof. We must show that if S is a nonempty subset of the natural numbers, then S contains a least element. If S contains 1, then the theorem is true by Lemma 2.7. Assume that if S contains an integer k such that $1 \leq k \leq n$, then S contains a least element. We will show that if a set S contains an integer less than or equal to $n + 1$, then S has a least element. If S does not contain an integer less than $n + 1$, then $n + 1$ is the smallest integer in S. Otherwise, since S is nonempty, S must contain an integer less than or equal to n. In this case, by induction, S contains a least element. ■

Induction can also be very useful in formulating definitions. For instance, there are two ways to define $n!$, the factorial of a positive integer n.

- The *explicit* definition: $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$.
- The *inductive* or *recursive* definition: $1! = 1$ and $n! = n(n-1)!$ for $n > 1$.

Every good mathematician or computer scientist knows that looking at problems recursively, as opposed to explicitly, often results in better understanding of complex issues.

Activity 2.2 Using the Principle of Mathematical Induction, prove each of the following statements.

(a) Prove that

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

for $n \in \mathbb{N}$.

(b) Prove that $n! > 2^n$ for $n \geq 4$.

(c) Prove that $10^{n+1} + 10^n + 1$ is divisible by 3 for $n \in \mathbb{N}$.

(d) If x is a nonnegative real number, then show that $(1+x)^n - 1 \geq nx$ for $n = 0, 1, 2, \ldots$.

Reading Questions

1. How does Example 2.2 use the First Principle of Mathematical Induction? That is, what in the example is $S(n)$ and what is n_0?

2. The Second Principle of Mathematical Induction is sometimes called **Strong Mathematical Induction**. Why do you think this might be a good name? In what ways is it *stronger* than the First Principle of Mathematical Induction?

3. Give an example to demonstrate the Principle of Well-Ordering does not hold true if we replace the natural numbers with the integers.

Exercises

1. Prove that

$$1^3 + 2^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$$

for $n \in \mathbb{N}$.

2. Prove that

$$x + 4x + 7x + \cdots + (3n-2)x = \frac{n(3n-1)x}{2}$$

for $n \in \mathbb{N}$.

3. Prove that $4 \cdot 10^{2n} + 9 \cdot 10^{2n-1} + 5$ is divisible by 99 for $n \in \mathbb{N}$.

4. Prove the Leibniz rule for $f^{(n)}(x)$, where $f^{(n)}$ is the nth derivative of f; that is, show that

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

5. Use induction to prove that $1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1} - 1$ for $n \in \mathbb{N}$.

6. Prove that

$$\frac{1}{2} + \frac{1}{6} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}$$

for $n \in \mathbb{N}$.

7. **Power Sets.** Let X be a set. Define the **power set** of X, denoted $\mathcal{P}(X)$, to be the set of all subsets of X. For example,

$$\mathcal{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$

For every positive integer n, show that a set with exactly n elements has a power set with exactly 2^n elements.

2.2 The Division and Euclidean Algorithms

The integers $\mathbb{Z}$, the set of positive and negative “whole” numbers, form a foundation for much of the mathematics we do. You can add, subtract, and multiply integers (and the result is always an integer). But can you divide? Sometimes you can, such as $12 \div 4 = 3$. But what about $23 \div 4$? We have two options. We can write $23 \div 4 = 5.75$, but then the result is not an integer. Or we can say that 4 goes into 23 a total of 5 times with a **remainder** of 3.

Does this always work? That is, can we always divide one integer by some other (positive) integer *with remainder*? Well, just keep subtracting the divisor until it is no longer possible to subtract, and what is left, is the remainder. Division is like splitting up marbles between children. We can give a marble to each child again and again until we do not have enough marbles left to fairly give them out, and we see how many marbles *remain.* If we begin with 23 marbles to be divided equally among 4 children, there will be 3 marbles leftover. Notice that $23 \div 4 = 5.75$ does not make sense in this example. How would you give a child 0.75 marbles?

We make this more precise by stating the following very useful theorem, called the **Division Algorithm**. Its proof is a nice application of the Principle of Well-Ordering.

Theorem 2.9 Division Algorithm. *Let a and b be integers, with $b > 0$. Then there exist unique integers q and r such that*

$$a = bq + r$$

where $0 \leq r < b$.

Proof. This is a perfect example of the existence-and-uniqueness type of proof. We must first prove that the numbers q and r actually exist. Then we must show that if q' and r' are two other such numbers, then $q = q'$ and $r = r'$.

Existence of q and r. Let

$$S = \{a - bk : k \in \mathbb{Z} \text{ and } a - bk \geq 0\}.$$

If $0 \in S$, then b divides a, and we can let $q = a/b$ and $r = 0$. If $0 \notin S$, we can use the Well-Ordering Principle. We must first show that S is nonempty. If $a > 0$, then $a - b \cdot 0 \in S$. If $a < 0$, then $a - b(2a) = a(1 - 2b) \in S$. In either case $S \neq \emptyset$. By the Well-Ordering Principle, S must have a smallest member, say $r = a - bq$. Therefore, $a = bq + r$, $r \geq 0$. We now show that $r < b$. Suppose that $r > b$. Then

$$a - b(q + 1) = a - bq - b = r - b > 0.$$

In this case we would have $a-b(q+1)$ in the set S. But then $a-b(q+1) < a-bq$, which would contradict the fact that $r = a - bq$ is the smallest member of S. So $r \leq b$. Since $0 \notin S$, $r \neq b$ and so $r < b$.

Uniqueness of q and r. Suppose there exist integers r, r', q, and q' such that

$$a = bq + r, 0 \leq r < b \quad \text{and} \quad a = bq' + r', 0 \leq r' < b.$$

Then $bq + r = bq' + r'$. Assume that $r' \geq r$. From the last equation we have $b(q - q') = r' - r$; therefore, b must divide $r' - r$ and $0 \leq r' - r \leq r' < b$. This is possible only if $r' - r = 0$. Hence, $r = r'$ and $q = q'$. ■

The division algorithm is, on the one hand, almost too obvious to write down. But it is also quite useful for proving less obvious facts about the integers. The next activity asks you to apply it in one such instance.

Activity 2.3 Using the division algorithm, show that every perfect square is of the form $4k$ or $4k + 1$ for some nonnegative integer k.

Of special interest will be the situation when the division algorithm results in a remainder of zero. Of course, this means that a is a multiple of b, and we could note that $a = kb$ for some integer k. We will also then say that b **divides** a and write $b \mid a$.

An integer d is called a **common divisor** of a and b if $d \mid a$ and $d \mid b$. The **greatest common divisor** of integers a and b is a positive integer d such that d is a common divisor of a and b and if d' is any other common divisor of a

and b,then $d' \mid d$. We write $d = \gcd(a, b)$; for example, $\gcd(24, 36) = 12$ and $\gcd(120, 102) = 6$. We say that two integers a and b are **relatively prime** if $\gcd(a, b) = 1$.

Theorem 2.10 *Let a and b be nonzero integers. Then there exist integers r and s such that*

$$\gcd(a, b) = ar + bs.$$

Furthermore, the greatest common divisor of a and b is unique.
Proof. Let

$$S = \{am + bn : m, n \in \mathbb{Z} \text{ and } am + bn > 0\}.$$

Clearly, the set S is nonempty; hence, by the Well-Ordering Principle S must have a smallest member, say $d = ar + bs$. We claim that $d = \gcd(a, b)$. Write $a = dq + r'$ where $0 \leq r' < d$. If $r' > 0$, then

$$\begin{aligned} r' &= a - dq \\ &= a - (ar + bs)q \\ &= a - arq - bsq \\ &= a(1 - rq) + b(-sq), \end{aligned}$$

which is in S. But this would contradict the fact that d is the smallest member of S. Hence, $r' = 0$ and d divides a. A similar argument shows that d divides b. Therefore, d is a common divisor of a and b.

Suppose that d' is another common divisor of a and b, and we want to show that $d' \mid d$. If we let $a = d'h$ and $b = d'k$, then

$$d = ar + bs = d'hr + d'ks = d'(hr + ks).$$

So d' must divide d. Hence, d must be the unique greatest common divisor of a and b. ■

Corollary 2.11 *Let a and b be two integers that are relatively prime. Then there exist integers r and s such that $ar + bs = 1$.*

Among other things, Theorem 2.10 allows us to compute the greatest common divisor of two integers.

Example 2.12 Let us compute the greatest common divisor of 945 and 2415. First observe that

$$\begin{aligned} 2415 &= 945 \cdot 2 + 525 \\ 945 &= 525 \cdot 1 + 420 \\ 525 &= 420 \cdot 1 + 105 \\ 420 &= 105 \cdot 4 + 0. \end{aligned}$$

Reversing our steps, 105 divides 420, 105 divides 525, 105 divides 945, and 105 divides 2415. Hence, 105 divides both 945 and 2415. If d were another common divisor of 945 and 2415, then d would also have to divide 105. Therefore, $\gcd(945, 2415) = 105$.

If we work backward through the above sequence of equations, we can also obtain numbers r and s such that $945r + 2415s = 105$. Observe that

$$\begin{aligned} 105 &= 525 + (-1) \cdot 420 \\ &= 525 + (-1) \cdot [945 + (-1) \cdot 525] \\ &= 2 \cdot 525 + (-1) \cdot 945 \\ &= 2 \cdot [2415 + (-2) \cdot 945] + (-1) \cdot 945 \end{aligned}$$

$$= 2 \cdot 2415 + (-5) \cdot 945.$$

So $r = -5$ and $s = 2$. Notice that r and s are not unique, since $r = 41$ and $s = -16$ would also work. □

To compute $\gcd(a, b) = d$, we are using repeated divisions to obtain a decreasing sequence of positive integers $r_1 > r_2 > \cdots > r_n = d$; that is,

$$\begin{aligned} b &= aq_1 + r_1 \\ a &= r_1 q_2 + r_2 \\ r_1 &= r_2 q_3 + r_3 \\ &\vdots \\ r_{n-2} &= r_{n-1} q_n + r_n \\ r_{n-1} &= r_n q_{n+1}. \end{aligned}$$

To find r and s such that $ar + bs = d$, we begin with this last equation and substitute results obtained from the previous equations:

$$\begin{aligned} d &= r_n \\ &= r_{n-2} - r_{n-1} q_n \\ &= r_{n-2} - q_n (r_{n-3} - q_{n-1} r_{n-2}) \\ &= -q_n r_{n-3} + (1 + q_n q_{n-1}) r_{n-2} \\ &\vdots \\ &= ra + sb. \end{aligned}$$

The algorithm that we have just used to find the greatest common divisor d of two integers a and b and to write d as the linear combination of a and b is known as the **Euclidean algorithm**.

Activity 2.4 For each of the following pairs of numbers a and b, calculate $\gcd(a, b)$ and find integers r and s such that $\gcd(a, b) = ra + sb$.

(a) 14 and 39

(b) 234 and 165

Reading Questions

1. What is the *Division Algorithm*? Illustrate your answer with two examples: first, with $a = 4$ and $b = 15$, and second with $a = -3$ and $b = 5$.

2. What does the Division Algorithm have to do with dividing numbers? Explain in your own words.

3. Use the definition of *greatest common divisor* given above to explain why $\gcd(28, 42) = 14$. Note, you should not use the phrase "14 is the largest number that" in your explanation.

Exercises

1. For each of the following pairs of numbers a and b, calculate $\gcd(a, b)$ and find integers r and s such that $\gcd(a, b) = ra + sb$.

(a) 1836 and 2940

(b) 471 and 562

(c) 2340 and 840

(d) 435 and 375

2. Let a and b be nonzero integers. If there exist integers r and s such that $ar + bs = 1$, show that a and b are relatively prime.

3. Let a and b be integers such that $\gcd(a, b) = 1$. Let r and s be integers such that $ar + bs = 1$. Prove that

$$\gcd(a, s) = \gcd(r, b) = \gcd(r, s) = 1.$$

4. Suppose that a, b, r, s are pairwise relatively prime and that

$$\begin{aligned} a^2 + b^2 &= r^2 \\ a^2 - b^2 &= s^2. \end{aligned}$$

Prove that a, r, and s are odd and b is even.

5. Let $n \in \mathbb{N}$. Use the division algorithm to prove that every integer is congruent mod n to precisely one of the integers $0, 1, \ldots, n-1$. Conclude that if r is an integer, then there is exactly one s in $\mathbb{Z}$ such that $0 \leq s < n$ and $[r] = [s]$. Hence, the integers are indeed partitioned by congruence mod n.

2.3 Prime Numbers

Let p be an integer such that $p > 1$. We say that p is a **prime number**, or simply p is **prime**, if the only positive numbers that divide p are 1 and p itself. An integer $n > 1$ that is not prime is said to be **composite**.

Lemma 2.13 Euclid. *Let a and b be integers and p be a prime number. If $p \mid ab$, then either $p \mid a$ or $p \mid b$.*

Proof. Suppose that p does not divide a. We must show that $p \mid b$. Since $\gcd(a, p) = 1$, there exist integers r and s such that $ar + ps = 1$. So

$$b = b(ar + ps) = (ab)r + p(bs).$$

Since p divides both ab and itself, p must divide $b = (ab)r + p(bs)$. ■

Theorem 2.14 Euclid. *There exist an infinite number of primes.*

Proof. We will prove this theorem by contradiction. Suppose that there are only a finite number of primes, say $p_1, p_2, \ldots, p_n$. Let $P = p_1 p_2 \cdots p_n + 1$. Then P must be divisible by some p_i for $1 \leq i \leq n$. In this case, p_i must divide $P - p_1 p_2 \cdots p_n = 1$, which is a contradiction. Hence, either P is prime or there exists an additional prime number $p \neq p_i$ that divides P. ■

Theorem 2.15 Fundamental Theorem of Arithmetic. *Let n be an integer such that $n > 1$. Then*

$$n = p_1 p_2 \cdots p_k,$$

where $p_1, \ldots, p_k$ are primes (not necessarily distinct). Furthermore, this factorization is unique; that is, if

$$n = q_1 q_2 \cdots q_l,$$

then $k = l$ and the q_i's are just the p_i's rearranged.

Proof. Uniqueness. To show uniqueness we will use induction on n. The theorem is certainly true for $n = 2$ since in this case n is prime. Now assume that the result holds for all integers m such that $1 \leq m < n$, and

$$n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l,$$

where $p_1 \leq p_2 \leq \cdots \leq p_k$ and $q_1 \leq q_2 \leq \cdots \leq q_l$. By Lemma 2.13, $p_1 \mid q_i$ for some $i = 1, \ldots, l$ and $q_1 \mid p_j$ for some $j = 1, \ldots, k$. Since all of the p_i's and q_i's are prime, $p_1 = q_i$ and $q_1 = p_j$. Hence, $p_1 = q_1$ since $p_1 \leq p_j = q_1 \leq q_i = p_1$. By the induction hypothesis,

$$n' = p_2 \cdots p_k = q_2 \cdots q_l$$

has a unique factorization. Hence, $k = l$ and $q_i = p_i$ for $i = 1, \ldots, k$.

Existence. To show existence, suppose that there is some integer that cannot be written as the product of primes. Let S be the set of all such numbers. By the Principle of Well-Ordering, S has a smallest number, say a. If the only positive factors of a are a and 1, then a is prime, which is a contradiction. Hence, $a = a_1 a_2$ where $1 < a_1 < a$ and $1 < a_2 < a$. Neither $a_1 \in S$ nor $a_2 \in S$, since a is the smallest element in S. So

$$\begin{aligned} a_1 &= p_1 \cdots p_r \\ a_2 &= q_1 \cdots q_s. \end{aligned}$$

Therefore,

$$a = a_1 a_2 = p_1 \cdots p_r q_1 \cdots q_s.$$

So $a \notin S$, which is a contradiction. ■

Activity 2.5 Let $x, y \in \mathbb{N}$ be relatively prime. If xy is a perfect square, prove that x and y must both be perfect squares.

2.3.1 Historical Note

Prime numbers were first studied by the ancient Greeks. Two important results from antiquity are Euclid's proof that an infinite number of primes exist and the Sieve of Eratosthenes, a method of computing all of the prime numbers less than a fixed positive integer n. One problem in number theory is to find a function f such that $f(n)$ is prime for each integer n. Pierre Fermat (1601?–1665) conjectured that $2^{2^n} + 1$ was prime for all n, but later it was shown by Leonhard Euler (1707–1783) that

$$2^{2^5} + 1 = 4{,}294{,}967{,}297$$

is a composite number. One of the many unproven conjectures about prime numbers is Goldbach's Conjecture. In a letter to Euler in 1742, Christian Goldbach stated the conjecture that every even integer with the exception of 2 seemed to be the sum of two primes: $4 = 2 + 2$, $6 = 3 + 3$, $8 = 3 + 5$, ... Although the conjecture has been verified for the numbers up through 4×10^{18}, it has yet to be proven in general. Since prime numbers play an important role in public key cryptography, there is currently a great deal of interest in determining whether or not a large number is prime.

2.3.2 Reading Questions

1. Is 1 a prime number? Answer this question based on the definition given above.

2. What is going on in Lemma 2.13? Illustrate what goes wrong with the lemma if p is not prime. For example, would the lemma still be true if $p = 6$?

3. Your sister shows you how to factor 24. First, we see that $24 = 6 \cdot 4$, and then we notice that $6 = 2 \cdot 3$ and $4 = 2 \cdot 2$. Then your brother says no, first you see that $24 = 8 \cdot 3$ and then you notice that $8 = 2 \cdot 2 \cdot 2$. What is this an example of and why?

2.3.3 Exercises

1. Let $p \geq 2$. Prove that if $2^p - 1$ is prime, then p must also be prime.

2. Prove that there are an infinite number of primes of the form $6n + 5$.

Hint. Every prime must be of the form 2, 3, $6n + 1$, or $6n + 5$. Suppose there are only finitely many primes of the form $6k + 5$.

3. Prove that there are an infinite number of primes of the form $4n - 1$.

4. Using the fact that 2 is prime, show that there do not exist integers p and q such that $p^2 = 2q^2$. Demonstrate that therefore $\sqrt{2}$ cannot be a rational number.

2.4 Summary and Additional Exercises

2.4.1 The Important Ideas

- The First Principle of Mathematical Induction (Principle 2.1): Let $S(n)$ be a statement about integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If for all integers k with $k \geq n_0$, $S(k)$ implies that $S(k+1)$ is true, then $S(n)$ is true for all integers n greater than or equal to n_0.

- Second Principle of Mathematical Induction (Principle 2.5): Let $S(n)$ be a statement about integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If $S(n_0), S(n_0+1), \ldots, S(k)$ imply that $S(k+1)$ for $k \geq n_0$, then the statement $S(n)$ is true for all integers $n \geq n_0$.

- Principle of Well-Ordering (Principle 2.6): Every nonempty subset of the natural numbers is well-ordered. The Principle of Well-Ordering is equivalent to the Principle of Mathematical Induction.

- The Division Algorithm (Theorem 2.9): Let a and b be integers, with $b > 0$. Then there exist unique integers q and r such that
$$a = bq + r$$
where $0 \leq r < b$.

- Fundamental Theorem of Arithmetic (Theorem 2.15): Let n be an integer such that $n > 1$. Then
$$n = p_1 p_2 \cdots p_k,$$
where $p_1, \ldots, p_k$ are primes (not necessarily distinct). Furthermore, this factorization is unique; that is, if
$$n = q_1 q_2 \cdots q_l,$$
then $k = l$ and the q_i's are just the p_i's rearranged.

2.4.2 Additional Exercises

1. Show that

$$\sqrt[n]{a_1 a_2 \cdots a_n} \leq \frac{1}{n} \sum_{k=1}^{n} a_k.$$

2. Prove that the two principles of mathematical induction stated in Section 2.1 are equivalent.

3. Show that the Principle of Well-Ordering for the natural numbers implies that 1 is the smallest natural number. Use this result to show that the Principle of Well-Ordering implies the Principle of Mathematical Induction; that is, show that if $S \subset \mathbb{N}$ such that $1 \in S$ and $n+1 \in S$ whenever $n \in S$, then $S = \mathbb{N}$.

4. **Fibonacci Numbers.** The Fibonacci numbers are

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots.$$

We can define them inductively by $f_1 = 1$, $f_2 = 1$, and $f_{n+2} = f_{n+1} + f_n$ for $n \in \mathbb{N}$.

(a) Prove that $f_n < 2^n$.

(b) Prove that $f_{n+1} f_{n-1} = f_n^2 + (-1)^n$, $n \geq 2$.

(c) Prove that $f_n = [(1+\sqrt{5}\,)^n - (1-\sqrt{5}\,)^n]/2^n \sqrt{5}$.

(d) Show that $\lim_{n \to \infty} f_n / f_{n+1} = (\sqrt{5} - 1)/2$.

(e) Prove that f_n and f_{n+1} are relatively prime.

5. Define the **least common multiple** of two nonzero integers a and b, denoted by $\operatorname{lcm}(a, b)$, to be the nonnegative integer m such that both a and b divide m, and if a and b divide any other integer n, then m also divides n. Prove there exists a unique least common multiple for any two integers a and b.

Hint. Use the Principle of Well-Ordering and the division algorithm.

6. If $d = \gcd(a, b)$ and $m = \operatorname{lcm}(a, b)$, prove that $dm = |ab|$.

7. Show that $\operatorname{lcm}(a, b) = ab$ if and only if $\gcd(a, b) = 1$.

8. Prove that $\gcd(a, c) = \gcd(b, c) = 1$ if and only if $\gcd(ab, c) = 1$ for integers a, b, and c.

9. Let $a, b, c \in \mathbb{Z}$. Prove that if $\gcd(a, b) = 1$ and $a \mid bc$, then $a \mid c$.

Hint. Since $\gcd(a, b) = 1$, there exist integers r and s such that $ar + bs = 1$. Thus, $acr + bcs = c$.

2.5 Connections to the Secondary Classroom—The Euclidean Algorithm

Recall how we used the Euclidean Algorithm to find the greatest common divisor of 945 and 2415 (Example 2.12). We calculated

$$\begin{aligned} 2415 &= 945 \cdot 2 + 525 \\ 945 &= 525 \cdot 1 + 420 \\ 525 &= 420 \cdot 1 + 105 \\ 420 &= 105 \cdot 4 + 0, \end{aligned}$$

and then reversed our steps, noting that 105 divides 420, 105 divides 525, 105 divides 945, and 105 divides 2415., and so 105 divides both 945 and 2415. Teaching the Euclidean Algorithm to high school students might prove to be a difficult task; however, we can associate a rectangle with the problem of finding the greatest common divisor of 945 and 2415 (Figure 2.16). Subdividing the 945 by 2415 rectangle in the perscribed way is equivalent to using the Euclidean Algorithm and finding the greatest common divisor.

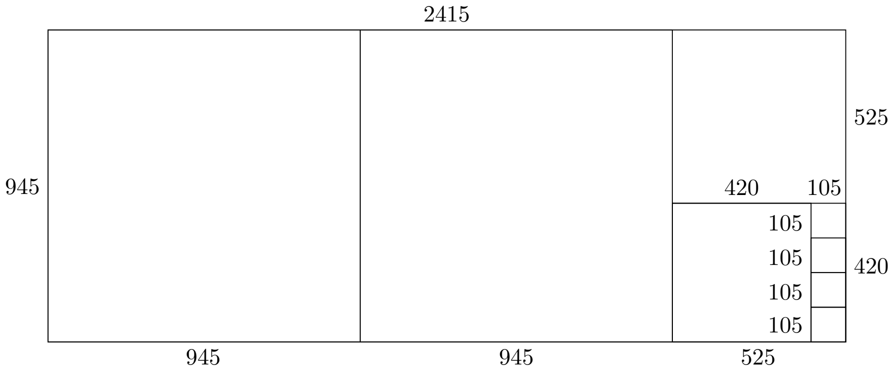


Figure 2.16 The greatest common divisor of 2415 and 945

Exercises

1. For each of the following pairs of numbers a and b, calculate $\gcd(a, b)$ by subdividing the appropriate rectangle.

(a) 14 and 39

(b) 234 and 165

(c) 1836 and 2940

(d) 471 and 562

(e) 2340 and 840

(f) 435 and 375

2. Explain the connection between the Euclidean Algorithm and the rectangle in Figure 2.16.

2.6 References and Suggested Readings

[1] Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers.* 6th ed. Oxford University Press, New York, 2008.

[2] Niven, I. and Zuckerman, H. S. *An Introduction to the Theory of Numbers.* 5th ed. Wiley, New York, 1991.

[3] Vanden Eynden, C. *Elementary Number Theory.* 2nd ed. Waveland Press, Long Grove IL, 2001.

Chapter 3

Groups

Objectives

- To understand and be able to apply the definition of a group.
- To understand and be able to apply the definition of a subgroup.
- To understand and be able to use examples of groups.
- To understand when two groups are isomorphic; that is, to understand when two groups are the "same."

We begin our study of algebraic structures by investigating sets associated with a single operation that satisfy certain reasonable axioms. In a way this is very natural as a child learns addition before multiplication. We want to define an operation on a set in a way that will generalize such familiar structures as the integers $\mathbb{Z}$ together with the single operation of addition, the integers modulo n with the operation of addition modulo n, or invertible 2×2 matrices together with the single operation of matrix multiplication. The integers, the integers modulo n, and the 2×2 matrices, together with their respective single operations, are examples of algebraic structures known as groups.

The theory of groups occupies a central position in mathematics. Modern group theory arose from an attempt to find the roots of a polynomial in terms of its coefficients. Groups now play a central role in such areas as coding theory, counting, and the study of symmetries. In addition, many areas of biology, chemistry, and physics have benefited from group theory.

Activity 3.1 Find all $x \in \mathbb{Z}$ satisfying each of the following equations.

(a) $3x \equiv 2 \pmod 7$

(b) $5x + 1 \equiv 13 \pmod{23}$

(c) $5x + 1 \equiv 13 \pmod{26}$

(d) $9x \equiv 3 \pmod 5$

(e) $5x \equiv 1 \pmod 6$

(f) $3x \equiv 1 \pmod 6$

3.1 Integer Equivalence Classes and Symmetries

Let us now investigate some mathematical structures that can be viewed as sets with single operations.

3.1.1 The Integers mod n

The integers mod n have become indispensable in the theory and applications of algebra. In mathematics they are used in cryptography, coding theory, and the detection of errors in identification codes.

We have already seen that two integers a and b are equivalent mod n if n divides $a - b$. The integers mod n also partition $\mathbb{Z}$ into n different equivalence classes; we will denote the set of these equivalence classes by $\mathbb{Z}_n$. Consider the integers modulo 12 and the corresponding partition of the integers:

$$\begin{aligned} [0] &= \{\ldots, -12, 0, 12, 24, \ldots\}, \\ [1] &= \{\ldots, -11, 1, 13, 25, \ldots\}, \\ &\;\vdots \\ [11] &= \{\ldots, -1, 11, 23, 35, \ldots\}. \end{aligned}$$

When no confusion can arise, we will use $0, 1, \ldots, 11$ to indicate the equivalence classes $[0], [1], \ldots, [11]$ respectively. We can do arithmetic on $\mathbb{Z}_n$. For two integers a and b, define addition modulo n to be $(a + b) \pmod{n}$; that is, the remainder when $a + b$ is divided by n. Similarly, multiplication modulo n is defined as $(ab) \pmod{n}$, the remainder when ab is divided by n.

Example 3.1 The following examples illustrate integer arithmetic modulo n:

$$\begin{array}{llll} 7 + 4 \equiv 1 & \pmod{5} & \qquad 7 \cdot 3 \equiv 1 & \pmod{5} \\ 3 + 5 \equiv 0 & \pmod{8} & \qquad 3 \cdot 5 \equiv 7 & \pmod{8} \\ 3 + 4 \equiv 7 & \pmod{12} & \qquad 3 \cdot 4 \equiv 0 & \pmod{12}. \end{array}$$

In particular, notice that it is possible that the product of two nonzero numbers modulo n can be equivalent to 0 modulo n. □

Example 3.2 Most, but not all, of the usual laws of arithmetic hold for addition and multiplication in $\mathbb{Z}_n$. For instance, it is not necessarily true that there is a multiplicative inverse. Consider the multiplication table for $\mathbb{Z}_8$ in Table 3.3. Notice that 2, 4, and 6 do not have multiplicative inverses; that is, for $n = 2$, 4, or 6, there is no integer k such that $kn \equiv 1 \pmod{8}$.

Table 3.3 Multiplication table for $\mathbb{Z}_8$

·	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	1	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	1	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	1

□

Proposition 3.4 *Let $\mathbb{Z}_n$ be the set of equivalence classes of the integers mod n and $a, b, c \in \mathbb{Z}_n$.*

1. *Addition and multiplication are commutative:*

$$\begin{aligned} a + b &\equiv b + a \pmod{n} \\ ab &\equiv ba \pmod{n}. \end{aligned}$$

2. *Addition and multiplication are associative:*

$$\begin{aligned} (a + b) + c &\equiv a + (b + c) \pmod{n} \\ (ab)c &\equiv a(bc) \pmod{n}. \end{aligned}$$

3. *There are both additive and multiplicative identities:*

$$\begin{aligned} a + 0 &\equiv a \pmod{n} \\ a \cdot 1 &\equiv a \pmod{n}. \end{aligned}$$

4. *Multiplication distributes over addition:*

$$a(b + c) \equiv ab + ac \pmod{n}.$$

5. *For every integer a there is an additive inverse $-a$:*

$$a + (-a) \equiv 0 \pmod{n}.$$

6. *Let a be a nonzero integer. Then* $\gcd(a, n) = 1$ *if and only if there exists a multiplicative inverse b for a* $\pmod{n}$*; that is, a nonzero integer b such that*

$$ab \equiv 1 \pmod{n}.$$

Proof. We will prove (1) and (6) and leave the remaining properties to be proven in the exercises.

(1) Addition and multiplication are commutative modulo n since the remainder of $a + b$ divided by n is the same as the remainder of $b + a$ divided by n.

(6) Suppose that $\gcd(a, n) = 1$. Then there exist integers r and s such that $ar + ns = 1$. Since $ns = 1 - ar$, it must be the case that $ar \equiv 1 \pmod{n}$. Letting b be the equivalence class of r, $ab \equiv 1 \pmod{n}$.

Conversely, suppose that there exists an integer b such that $ab \equiv 1 \pmod{n}$. Then n divides $ab - 1$, so there is an integer k such that $ab - nk = 1$. Let $d = \gcd(a, n)$. Since d divides $ab - nk$, d must also divide 1; hence, $d = 1$. ■

3.1.2 Symmetries

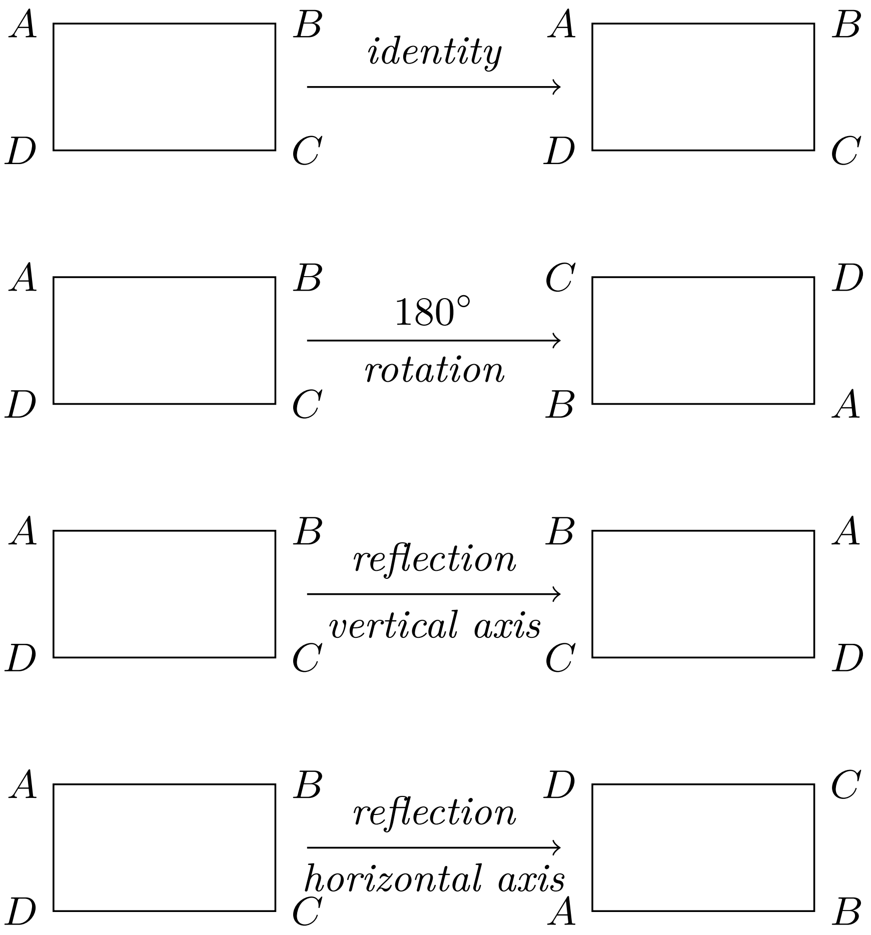
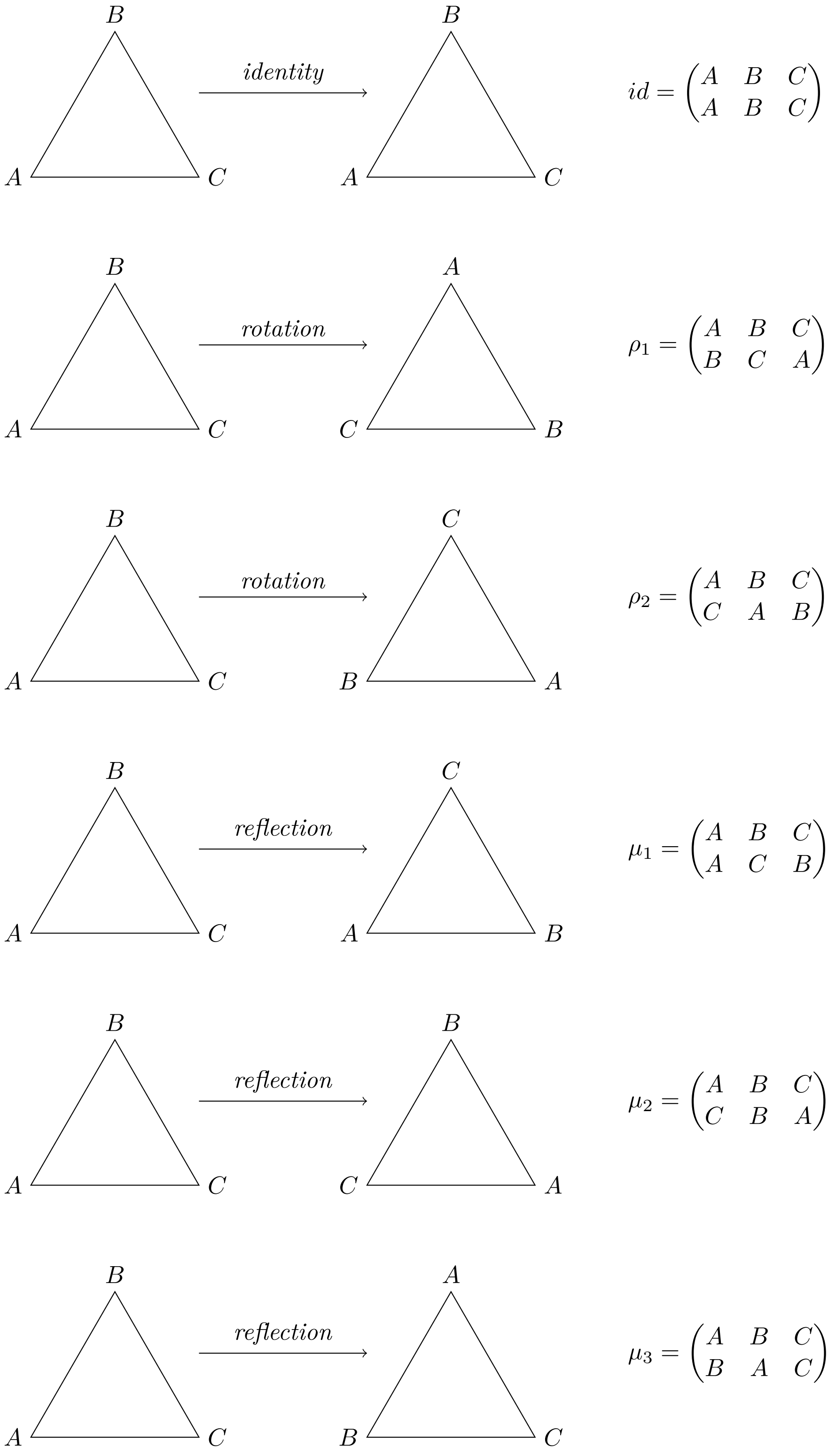


Figure 3.5 Rigid motions of a rectangle

A **symmetry** of a geometric figure is a rearrangement of the figure preserving the arrangement of its sides and vertices as well as its distances and angles. A map from the plane to itself preserving the symmetry of an object is called a **rigid motion**. For example, if we look at the rectangle in Figure 3.5, it is easy to see that a rotation of 180° or 360° returns a rectangle in the plane with the same orientation as the original rectangle and the same relationship among the vertices. A reflection of the rectangle across either the vertical axis or the horizontal axis can also be seen to be a symmetry. However, a 90° rotation in either direction cannot be a symmetry unless the rectangle is a square.



$$id = \begin{pmatrix} A & B & C \\ A & B & C \end{pmatrix}$$

$$\rho_1 = \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}$$

$$\rho_2 = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}$$

$$\mu_1 = \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix}$$

$$\mu_3 = \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix}$$

Figure 3.6 Symmetries of a triangle

Let us find the symmetries of the equilateral triangle $\triangle ABC$. To find a symmetry of $\triangle ABC$, we must first examine the permutations of the vertices A, B, and C and then ask if a permutation extends to a symmetry of the triangle. Recall that a **permutation** of a set S is a one-to-one and onto map $\pi : S \to S$. The three vertices have $3! = 6$ permutations, so the triangle has at most six symmetries. To see that there are six permutations, observe there are three different possibilities for the first vertex, and two for the second, and the remaining vertex is determined by the placement of the first two. So we have $3 \cdot 2 \cdot 1 = 3! = 6$ different arrangements. To denote the permutation of the vertices of an equilateral triangle that sends A to B, B to C, and C to A, we write the array

$$\begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}.$$

Notice that this particular permutation corresponds to the rigid motion of rotating the triangle by 120° in a clockwise direction. In fact, every permutation gives rise to a symmetry of the triangle. All of these symmetries are shown in Figure 3.6.

A natural question to ask is what happens if one motion of the triangle $\triangle ABC$ is followed by another. Which symmetry is $\mu_1\rho_1$; that is, what happens when we do the permutation ρ_1 and then the permutation μ_1? *Remember that we are composing functions here. Although we usually multiply left to right, we compose functions right to left.* We have

$$\begin{aligned} (\mu_1\rho_1)(A) &= \mu_1(\rho_1(A)) = \mu_1(B) = C \\ (\mu_1\rho_1)(B) &= \mu_1(\rho_1(B)) = \mu_1(C) = B \\ (\mu_1\rho_1)(C) &= \mu_1(\rho_1(C)) = \mu_1(A) = A. \end{aligned}$$

This is the same symmetry as μ_2. Suppose we do these motions in the opposite order, μ_1 then ρ_1. It is easy to determine that this is the same as the symmetry μ_3; hence, $\rho_1\mu_1 \neq \mu_1\rho_1$. A multiplication table for the symmetries of an equilateral triangle $\triangle ABC$ is given in Table 3.7.

Notice that in the multiplication table for the symmetries of an equilateral triangle, for every rigid motion of the triangle α there is another rigid motion β such that $\alpha\beta = \text{id}$; that is, for every rigid motion there is another rigid motion that takes the triangle back to its original orientation.

Table 3.7 Symmetries of an equilateral triangle

$\circ$	id	ρ_1	ρ_2	μ_1	μ_2	μ_3
id	id	ρ_1	ρ_2	μ_1	μ_2	μ_3
ρ_1	ρ_1	ρ_2	id	μ_3	μ_1	μ_2
ρ_2	ρ_2	id	ρ_1	μ_2	μ_3	μ_1
μ_1	μ_1	μ_2	μ_3	id	ρ_1	ρ_2
μ_2	μ_2	μ_3	μ_1	ρ_2	id	ρ_1
μ_3	μ_3	μ_1	μ_2	ρ_1	ρ_2	id

Activity 3.2 Consider a square with vertices A, B, C, and D.

(a) What are the symmetries of the square?

(b) Will it always be true that $\mu\rho = \rho\mu$ for two symmetries μ and ρ of the square?

3.1.3 Reading Questions

1. What does it mean to say that a number a in $\mathbb{Z}_8$ has a *multiplicative inverse*? Describe this in general, and illustrate with an example of an element that has this and one that doesn't.

2. Why are there six *symmetries* of an equilateral triangle? Does this mean there should be $4! = 24$ symmetries of a square? Explain.

3. Which of the operations discussed in this section are *commutative* and which are not? What does this mean? Illustrate with examples.

3.1.4 Exercises

1. Write addition and multiplication tables for $\mathbb{Z}_6$.

2. Write a multiplcation table for the symmetries of a square.

3. Show that there are $n!$ permutations of a set containing n items.

Hint. Let

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

be in S_n. All of the a_is must be distinct. There are n ways to choose a_1, $n-1$ ways to choose $a_2, \ldots$, 2 ways to choose a_{n-1}, and only one way to choose a_n. Therefore, we can form σ in $n(n-1)\cdots 2 \cdot 1 = n!$ ways.

4. Show that

$$0 + a \equiv a + 0 \equiv a \pmod{n}$$

for all $a \in \mathbb{Z}_n$.

5. Prove that there is a multiplicative identity for the integers modulo n:

$$a \cdot 1 \equiv a \pmod{n}.$$

6. For each $a \in \mathbb{Z}_n$ find an element $b \in \mathbb{Z}_n$ such that

$$a + b \equiv b + a \equiv 0 \pmod{n}.$$

7. Show that addition and multiplication mod n are well defined operations. That is, show that the operations do not depend on the choice of the representative from the equivalence classes mod n.

8. Show that addition and multiplication mod n are associative operations.

9. Show that multiplication distributes over addition modulo n:

$$a(b + c) \equiv ab + ac \pmod{n}.$$

3.2 Definitions and Examples

The integers mod n and the symmetries of a triangle or a rectangle are examples of groups. A **binary operation** or **law of composition** on a set G is a function $G \times G \to G$ that assigns to each pair $(a, b) \in G \times G$ a unique element $a \circ b$, or ab in G, called the composition of a and b. A **group** $(G, \circ)$ is a set G together with a law of composition $(a, b) \mapsto a \circ b$ that satisfies the following axioms.

- The law of composition is **associative**. That is,

$$(a \circ b) \circ c = a \circ (b \circ c)$$

for $a, b, c \in G$.

- There exists an element $e \in G$, called the **identity element**, such that for any element $a \in G$

$$e \circ a = a \circ e = a.$$

- For each element $a \in G$, there exists an **inverse element** in G, denoted by a^{-1}, such that
$$a \circ a^{-1} = a^{-1} \circ a = e.$$

A group G with the property that $a \circ b = b \circ a$ for all $a, b \in G$ is called **abelian** or **commutative**. Groups not satisfying this property are said to be **nonabelian** or **noncommutative**.

Example 3.8 The integers $\mathbb{Z} = \{\ldots, -1, 0, 1, 2, \ldots\}$ form a group under the operation of addition. The binary operation on two integers $m, n \in \mathbb{Z}$ is just their sum. Since the integers under addition already have a well-established notation, we will use the operator $+$ instead of $\circ$; that is, we shall write $m+n$ instead of $m \circ n$. The identity is 0, and the inverse of $n \in \mathbb{Z}$ is written as $-n$ instead of n^{-1}. Notice that the set of integers under addition have the additional property that $m+n = n+m$ and therefore form an abelian group. □

Most of the time we will write ab instead of $a \circ b$; however, if the group already has a natural operation such as addition in the integers, we will use that operation. That is, if we are adding two integers, we still write $m+n$, $-n$ for the inverse, and 0 for the identity as usual. We also write $m - n$ instead of $m + (-n)$.

It is often convenient to describe a group in terms of an addition or multiplication table. Such a table is called a **Cayley table**.

Example 3.9 The integers mod n form a group under addition modulo n. Consider $\mathbb{Z}_5$, consisting of the equivalence classes of the integers 0, 1, 2, 3, and 4. We define the group operation on $\mathbb{Z}_5$ by modular addition. We write the binary operation on the group additively; that is, we write $m+n$. The element 0 is the identity of the group and each element in $\mathbb{Z}_5$ has an inverse. For instance, $2+3 = 3+2 = 0$. Table 3.10 is a Cayley table for $\mathbb{Z}_5$. By Proposition 3.4, $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ is a group under the binary operation of addition mod n.

Table 3.10 Cayley table for $(\mathbb{Z}_5, +)$

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

□

Example 3.11 Not every set with a binary operation is a group. For example, if we let modular multiplication be the binary operation on $\mathbb{Z}_n$, then $\mathbb{Z}_n$ fails to be a group. The element 1 acts as a group identity since $1 \cdot k = k \cdot 1 = k$ for any $k \in \mathbb{Z}_n$; however, a multiplicative inverse for 0 does not exist since $0 \cdot k = k \cdot 0 = 0$ for every k in $\mathbb{Z}_n$. Even if we consider the set $\mathbb{Z}_n \setminus \{0\}$, we still may not have a group. For instance, let $2 \in \mathbb{Z}_6$. Then 2 has no multiplicative inverse since

$$\begin{aligned} 0 \cdot 2 &= 0 & 1 \cdot 2 &= 2 \\ 2 \cdot 2 &= 4 & 3 \cdot 2 &= 0 \\ 4 \cdot 2 &= 2 & 5 \cdot 2 &= 4. \end{aligned}$$

By Proposition 3.4, every nonzero k does have an inverse in $\mathbb{Z}_n$ if k is relatively prime to n. Denote the set of all such nonzero elements in $\mathbb{Z}_n$ by $U(n)$. Then $U(n)$ is a group called the **group of units** of $\mathbb{Z}_n$. Table 3.12 is a Cayley table for the group $U(8)$.

Table 3.12 Multiplication table for $U(8)$

$\cdot$	1	3	5	7
1	1	3	5	7
3	3	1	7	5
5	5	7	1	3
7	7	5	3	1

□

Example 3.13 The symmetries of an equilateral triangle described in Section 3.1 form a nonabelian group. As we observed, it is not necessarily true that $\alpha\beta = \beta\alpha$ for two symmetries α and β. Using Table 3.7, which is a Cayley table for this group, we can easily check that the symmetries of an equilateral triangle are indeed a group. We will denote this group by either S_3 or D_3, for reasons that will be explained later. □

Activity 3.3 Which of the following multiplication tables defined on the set $G = \{a, b, c, d\}$ form a group? Support your answer in each case.

(a)

$\circ$	a	b	c	d
a	a	c	d	a
b	b	b	c	d
c	c	d	a	b
d	d	a	b	c

(b)

$\circ$	a	b	c	d
a	a	b	c	d
b	b	a	d	c
c	c	d	a	b
d	d	c	b	a

(c)

$\circ$	a	b	c	d
a	a	b	c	d
b	b	c	d	a
c	c	d	a	b
d	d	a	b	c

(d)

$\circ$	a	b	c	d
a	a	b	c	d
b	b	a	c	d
c	c	b	a	d
d	d	d	b	c

Example 3.14 We use $\mathbb{M}_2(\mathbb{R})$ to denote the set of all 2×2 matrices. Let $GL_2(\mathbb{R})$ be the subset of $\mathbb{M}_2(\mathbb{R})$ consisting of invertible matrices; that is, a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in $GL_2(\mathbb{R})$ if there exists a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$, where

I is the 2×2 identity matrix. For A to have an inverse is equivalent to requiring that the determinant of A be nonzero; that is, $\det A = ad - bc \neq 0$. The set of invertible matrices forms a group called the **general linear group**. The identity of the group is the identity matrix

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

The inverse of $A \in GL_2(\mathbb{R})$ is

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The product of two invertible matrices is again invertible. Matrix multiplication is associative, satisfying the other group axiom. It is generally not true that $AB = BA$ for two matrices A and B; hence, $GL_2(\mathbb{R})$ is another example of a nonabelian group. □

Example 3.15 Let

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad I = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$J = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \qquad K = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

where $i^2 = -1$. One can easily verify the relations $I^2 = J^2 = K^2 = -1$, $IJ = K$, $JK = I$, $KI = J$, $JI = -K$, $KJ = -I$, and $IK = -J$. The set $Q_8 = \{\pm 1, \pm I, \pm J, \pm K\}$ is a group called the **quaternion group**. Notice that Q_8 is noncommutative. □

Example 3.16 Let $\mathbb{C}^*$ be the set of nonzero complex numbers. Then $\mathbb{C}^*$ forms a group under the operation of multiplication. The identity is 1. If $z = a + bi$ is a nonzero complex number, then

$$z^{-1} = \frac{a - bi}{a^2 + b^2}$$

is the inverse of z. It is easy to see that the remaining group axioms hold. □

A group is **finite**, or has **finite order**, if it contains a finite number of elements; otherwise, the group is said to be **infinite** or to have **infinite order**. The **order** of a finite group is the number of elements that it contains. If G is a group containing n elements, we write $|G| = n$. The group $\mathbb{Z}_5$ is a finite group of order 5; the integers $\mathbb{Z}$ form an infinite group under addition, and we sometimes write $|\mathbb{Z}| = \infty$.

3.2.1 Basic Properties of Groups

Proposition 3.17 *The identity element in a group G is unique; that is, there exists only one element $e \in G$ such that $eg = ge = g$ for all $g \in G$.*

Proof. Suppose that e and e' are both identities in G. Then $eg = ge = g$ and $e'g = ge' = g$ for all $g \in G$. We need to show that $e = e'$. If we think of e as the identity, then $ee' = e'$; but if e' is the identity, then $ee' = e$. Combining these two equations, we have $e = ee' = e'$. ■

Inverses in a group are also unique. If g' and g'' are both inverses of an element g in a group G, then $gg' = g'g = e$ and $gg'' = g''g = e$. We want to show that $g' = g''$, but $g' = g'e = g'(gg'') = (g'g)g'' = eg'' = g''$. We summarize this fact in the following proposition.

Proposition 3.18 *If g is any element in a group G, then the inverse of g, denoted by g^{-1}, is unique.*

Proposition 3.19 *Let G be a group. If $a, b \in G$, then $(ab)^{-1} = b^{-1}a^{-1}$.*

Proof. Let $a, b \in G$. Then $abb^{-1}a^{-1} = aea^{-1} = aa^{-1} = e$. Similarly, $b^{-1}a^{-1}ab = e$. But by the previous proposition, inverses are unique; hence, $(ab)^{-1} = b^{-1}a^{-1}$. ■

It may seem strange that the inverse of ab is $b^{-1}a^{-1}$ and not $a^{-1}b^{-1}$; however, if you think of a as the operation of putting on your sock and b as the operation of putting on your shoe, then putting on your shoe and sock must be done in the order ab. To take off your shoe and sock, $(ab)^{-1}$, you must remove your shoe first and then remove your sock, or $(ab)^{-1} = b^{-1}a^{-1}$.

Proposition 3.20 *Let G be a group. For any $a \in G$, $(a^{-1})^{-1} = a$.*

Proof. Observe that $a^{-1}(a^{-1})^{-1} = e$. Consequently, multiplying both sides of this equation by a, we have

$$(a^{-1})^{-1} = e(a^{-1})^{-1} = aa^{-1}(a^{-1})^{-1} = ae = a.$$

■

It makes sense to write equations with group elements and group operations. If a and b are two elements in a group G, does there exist an element $x \in G$ such that $ax = b$? If such an x does exist, is it unique? The following proposition answers both of these questions positively.

Proposition 3.21 *Let G be a group and a and b be any two elements in G. Then the equations $ax = b$ and $xa = b$ have unique solutions in G.*

Proof. Suppose that $ax = b$. We must show that such an x exists. We can multiply both sides of $ax = b$ by a^{-1} to find $x = ex = a^{-1}ax = a^{-1}b$.

To show uniqueness, suppose that x_1 and x_2 are both solutions of $ax = b$; then $ax_1 = b = ax_2$. So $x_1 = a^{-1}ax_1 = a^{-1}ax_2 = x_2$. The proof for the existence and uniqueness of the solution of $xa = b$ is similar. ■

Proposition 3.22 *If G is a group and $a, b, c \in G$, then $ba = ca$ implies $b = c$ and $ab = ac$ implies $b = c$.*

This proposition tells us that the **right and left cancellation laws** are true in groups. We leave the proof as an exercise.

We can use exponential notation for groups just as we do in ordinary algebra. If G is a group and $g \in G$, then we define $g^0 = e$. For $n \in \mathbb{N}$, we define

$$g^n = \underbrace{g \cdot g \cdots g}_{n \text{ times}}$$

and

$$g^{-n} = \underbrace{g^{-1} \cdot g^{-1} \cdots g^{-1}}_{n \text{ times}}.$$

Theorem 3.23 *In a group, the usual laws of exponents hold; that is, for all $g, h \in G$,*

1. *$g^m g^n = g^{m+n}$ for all $m, n \in \mathbb{Z}$;*
2. *$(g^m)^n = g^{mn}$ for all $m, n \in \mathbb{Z}$;*
3. *$(gh)^n = (h^{-1}g^{-1})^{-n}$ for all $n \in \mathbb{Z}$. Furthermore, if G is abelian, then $(gh)^n = g^n h^n$.*

We will leave the proof of this theorem as an exercise. Notice that $(gh)^n \neq g^n h^n$ in general, since the group may not be abelian. If the group is $\mathbb{Z}$ or $\mathbb{Z}_n$, we write the group operation additively and the exponential operation multiplicatively; that is, we write ng instead of g^n. The laws of exponents now become

1. $mg + ng = (m+n)g$ for all $m, n \in \mathbb{Z}$;
2. $m(ng) = (mn)g$ for all $m, n \in \mathbb{Z}$;
3. $m(g+h) = mg + mh$ for all $n \in \mathbb{Z}$.

It is important to realize that the last statement can be made only because $\mathbb{Z}$ and $\mathbb{Z}_n$ are commutative groups.

Activity 3.4 Let $S = \mathbb{R} \setminus \{-1\}$ and define a binary operation on S by $a * b = a + b + ab$. Prove that $(S, *)$ is an abelian group.

3.2.2 Historical Note

Although the first clear axiomatic definition of a group was not given until the late 1800s, group-theoretic methods had been employed before this time in the development of many areas of mathematics, including geometry and the theory of algebraic equations.

Joseph-Louis Lagrange used group-theoretic methods in a 1770–1771 memoir to study methods of solving polynomial equations. Later, Évariste Galois (1811–1832) succeeded in developing the mathematics necessary to determine exactly which polynomial equations could be solved in terms of the coefficients of the polynomial. Galois' primary tool was group theory.

The study of geometry was revolutionized in 1872 when Felix Klein proposed that geometric spaces should be studied by examining those properties that are invariant under a transformation of the space. Sophus Lie, a contemporary of Klein, used group theory to study solutions of partial differential equations. One of the first modern treatments of group theory appeared in William Burnside's *The Theory of Groups of Finite Order* [1], first published in 1897.

3.2.3 Reading Questions

1. What does it mean for a group to be *abelian*? Be as specific as possible.

2. What is the inverse of the element 5 in (a) the group $\mathbb{Z}_8$ and (b) the group $U(8)$? Briefly explain your answers.

3. What is the *order* of the group $U(8)$? Explain.

4. For any group G with elements a and b, is $(ab)^{-1} = a^{-1}b^{-1}$? Explain why or give a counterexample.

5. In any group G with elements a and b, is there a unique solution in the group to the equation $ax = b$? Explain why or give a counterexample.

6. What are the *right and left cancellation laws*? Do these hold for all groups, some groups, or no groups? Explain.

3.2.4 Exercises

1. Write out Cayley tables for groups formed by the symmetries of a rectangle and for $(\mathbb{Z}_4, +)$. How many elements are in each group? Are the groups the same? Why or why not?

2. Describe the symmetries of a rhombus and prove that the set of symmetries forms a group. Give Cayley tables for both the symmetries of a rectangle and the symmetries of a rhombus. Are the symmetries of a rectangle and those of a rhombus the same?

3. Describe the symmetries of a square and prove that the set of symmetries is a group. Give a Cayley table for the symmetries. How many ways can the vertices of a square be permuted? Is each permutation necessarily a symmetry of the square? The symmetry group of the square is denoted by D_4.

4. Give a multiplication table for the group $U(12)$.

Hint.

$\cdot$	1	5	7	11
1	1	5	7	11
5	5	1	11	7
7	7	11	1	5
11	11	7	5	1

5. Give an example of two elements A and B in $GL_2(\mathbb{R})$ with $AB \neq BA$.

Hint. Pick two matrices. Almost any pair will work.

6. Prove that the set of matrices of the form

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}$$

is a group under matrix multiplication. This group, known as the **Heisenberg group**, is important in quantum physics. Matrix multiplication in the Heisenberg group is defined by

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & x' & y' \\ 0 & 1 & z' \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x+x' & y+y'+xz' \\ 0 & 1 & z+z' \\ 0 & 0 & 1 \end{pmatrix}.$$

7. Prove that $\det(AB) = \det(A)\det(B)$ in $GL_2(\mathbb{R})$. Use this result to show that the binary operation in the group $GL_2(\mathbb{R})$ is closed; that is, if A and B are in $GL_2(\mathbb{R})$, then $AB \in GL_2(\mathbb{R})$.

8. Let $\mathbb{Z}_2^n = \{(a_1, a_2, \ldots, a_n) : a_i \in \mathbb{Z}_2\}$. Define a binary operation on $\mathbb{Z}_2^n$ by

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

Prove that $\mathbb{Z}_2^n$ is a group under this operation. This group is important in algebraic coding theory.

9. Given the groups $\mathbb{R}^*$ and $\mathbb{Z}$, let $G = \mathbb{R}^* \times \mathbb{Z}$. Define a binary operation $\circ$ on G by $(a, m) \circ (b, n) = (ab, m + n)$. Show that G is a group under this operation.

10. Prove or disprove that every group containing six elements is abelian.

Hint. There is a nonabelian group containing six elements.

11. Give a specific example of some group G and elements $g, h \in G$ where $(gh)^n \neq g^n h^n$.

Hint. Look at the symmetry group of an equilateral triangle or a square.

12. Give an example of three different groups with eight elements. Why are the groups different?

Hint. The are five different groups of order 8.

13. Let a and b be elements in a group G. Prove that $ab^na^{-1} = (aba^{-1})^n$ for $n \in \mathbb{Z}$.

Hint.

$$\begin{aligned}(aba^{-1})^n &= (aba^{-1})(aba^{-1}) \cdots (aba^{-1})\\ &= ab(aa^{-1})b(aa^{-1})b \cdots b(aa^{-1})ba^{-1}\\ &= ab^na^{-1}.\end{aligned}$$

14. Prove the right and left cancellation laws for a group G; that is, show that in the group G, $ba = ca$ implies $b = c$ and $ab = ac$ implies $b = c$ for elements $a, b, c \in G$.

15. Show that if $a^2 = e$ for all elements a in a group G, then G must be abelian.

Hint. Since $abab = (ab)^2 = e = a^2b^2 = aabb$, we know that $ba = ab$.

16. Let a and b be elements of a group G. If $a^4b = ba$ and $a^3 = e$, prove that $ab = ba$.

Hint. $ba = a^4b = a^3ab = ab$

17. If $xy = x^{-1}y^{-1}$ for all x and y in G, prove that G must be abelian.

18. Prove that the inverse of $g_1g_2 \cdots g_n$ is $g_n^{-1}g_{n-1}^{-1} \cdots g_1^{-1}$.

19. Prove the remainder of Proposition 3.21: if G is a group and $a, b \in G$, then the equation $xa = b$ has a unique solution in G.

3.3 Subgroups

3.3.1 Definitions and Examples

Sometimes we wish to investigate smaller groups sitting inside a larger group. The set of even integers $2\mathbb{Z} = \{\ldots, -2, 0, 2, 4, \ldots\}$ is a group under the operation of addition. This smaller group sits naturally inside of the group of integers under addition. We define a **subgroup** H of a group G to be a subset H of G such that when the group operation of G is restricted to H, H is a group in its own right. Observe that every group G with at least two elements will always have at least two subgroups, the subgroup consisting of the identity element alone and the entire group itself. The subgroup $H = \{e\}$ of a group G is called the **trivial subgroup**. A subgroup that is a proper subset of G is called a **proper subgroup**. In many of the examples that we have investigated up to this point, there exist other subgroups besides the trivial and improper subgroups.

Example 3.24 Consider the set of nonzero real numbers, $\mathbb{R}^*$, with the group operation of multiplication. The identity of this group is 1 and the inverse of any element $a \in \mathbb{R}^*$ is just $1/a$. We will show that

$$\mathbb{Q}^* = \{p/q : p \,\text{and}\, q \,\text{are nonzero integers}\}$$

is a subgroup of $\mathbb{R}^*$. The identity of $\mathbb{R}^*$ is 1; however, $1 = 1/1$ is the quotient of two nonzero integers. Hence, the identity of $\mathbb{R}^*$ is in $\mathbb{Q}^*$. Given two elements in $\mathbb{Q}^*$, say p/q and r/s, their product pr/qs is also in $\mathbb{Q}^*$. The inverse of any element $p/q \in \mathbb{Q}^*$ is again in $\mathbb{Q}^*$ since $(p/q)^{-1} = q/p$. Since multiplication in $\mathbb{R}^*$ is associative, multiplication in $\mathbb{Q}^*$ is associative. □

Example 3.25 Recall that $\mathbb{C}^*$ is the multiplicative group of nonzero complex numbers. Let $H = \{1, -1, i, -i\}$. Then H is a subgroup of $\mathbb{C}^*$. It is quite easy to verify that H is a group under multiplication and that $H \subset \mathbb{C}^*$. □

Example 3.26 Let $SL_2(\mathbb{R})$ be the subset of $GL_2(\mathbb{R})$ consisting of matrices of determinant one; that is, a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in $SL_2(\mathbb{R})$ exactly when $ad - bc = 1$. To show that $SL_2(\mathbb{R})$ is a subgroup of the general linear group, we must show that it is a group under matrix multiplication. The 2×2 identity matrix is in $SL_2(\mathbb{R})$, as is the inverse of the matrix A:

$$A^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

It remains to show that multiplication is closed; that is, that the product of two matrices of determinant one also has determinant one. We will leave this task as an exercise. The group $SL_2(\mathbb{R})$ is called the **special linear group**. □

Example 3.27 It is important to realize that a subset H of a group G can be a group without being a subgroup of G. For H to be a subgroup of G, it must inherit the binary operation of G. The set of all 2×2 matrices, $\mathbb{M}_2(\mathbb{R})$, forms a group under the operation of addition. The 2×2 general linear group is a subset of $\mathbb{M}_2(\mathbb{R})$ and is a group under matrix multiplication, but it is not a subgroup of $\mathbb{M}_2(\mathbb{R})$. If we add two invertible matrices, we do not necessarily obtain another invertible matrix. Observe that

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

but the zero matrix is not in $GL_2(\mathbb{R})$. □

Activity 3.5 Let G consist of the 2×2 matrices of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

where $\theta \in \mathbb{R}$. Prove that G is a subgroup of $SL_2(\mathbb{R})$.

Example 3.28 One way of telling whether or not two groups are the same is by examining their subgroups. Other than the trivial subgroup and the group itself, the group $\mathbb{Z}_4$ has a single subgroup consisting of the elements 0 and 2. From the group $\mathbb{Z}_2$, we can form another group of four elements as follows. As a set this group is $\mathbb{Z}_2 \times \mathbb{Z}_2$. We perform the group operation coordinatewise; that is, $(a, b) + (c, d) = (a + c, b + d)$. Table 3.29 is an addition table for $\mathbb{Z}_2 \times \mathbb{Z}_2$. Since there are three nontrivial proper subgroups of $\mathbb{Z}_2 \times \mathbb{Z}_2$, $H_1 = \{(0, 0), (0, 1)\}$, $H_2 = \{(0, 0), (1, 0)\}$, and $H_3 = \{(0, 0), (1, 1)\}$, $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ must be different groups.

Table 3.29 Addition table for $\mathbb{Z}_2 \times \mathbb{Z}_2$

$+$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,1)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$
$(1,0)$	$(1,0)$	$(1,1)$	$(0,0)$	$(0,1)$
$(1,1)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$

□

3.3.2 Some Subgroup Theorems

Let us examine some criteria for determining exactly when a subset of a group is a subgroup.

Proposition 3.30 *A subset H of G is a subgroup if and only if it satisfies the following conditions.*

1. *The identity e of G is in H.*
2. *If $h_1, h_2 \in H$, then $h_1 h_2 \in H$.*
3. *If $h \in H$, then $h^{-1} \in H$.*

Proof. First suppose that H is a subgroup of G. We must show that the three conditions hold. Since H is a group, it must have an identity e_H. We must show that $e_H = e$, where e is the identity of G. We know that $e_H e_H = e_H$ and that $e e_H = e_H e = e_H$; hence, $e e_H = e_H e_H$. By right-hand cancellation, $e = e_H$. The second condition holds since a subgroup H is a group. To prove the third condition, let $h \in H$. Since H is a group, there is an element $h' \in H$ such that $hh' = h'h = e$. By the uniqueness of the inverse in G, $h' = h^{-1}$.

Conversely, if the three conditions hold, we must show that H is a group under the same operation as G; however, these conditions plus the associativity of the binary operation are exactly the axioms stated in the definition of a group. ■

Proposition 3.31 *Let H be a subset of a group G. Then H is a subgroup of G if and only if $H \neq \emptyset$, and whenever $g, h \in H$ then gh^{-1} is in H.*

Proof. First assume that H is a subgroup of G. We wish to show that $gh^{-1} \in H$ whenever g and h are in H. Since h is in H, its inverse h^{-1} must also be in H. Because of the closure of the group operation, $gh^{-1} \in H$.

Conversely, suppose that $H \subset G$ such that $H \neq \emptyset$ and $gh^{-1} \in H$ whenever $g, h \in H$. If $g \in H$, then $gg^{-1} = e$ is in H. If $g \in H$, then $eg^{-1} = g^{-1}$ is also in H. Now let $h_1, h_2 \in H$. We must show that their product is also in H. However, $h_1(h_2^{-1})^{-1} = h_1 h_2 \in H$. Hence, H is a subgroup of G. ■

Activity 3.6 Prove that the intersection of two subgroups of a group G is also a subgroup of G.

3.3.3 Reading Questions

1. Is $\mathbb{Z}_4$ a subgroup of $\mathbb{Z}_8$? Explain.

2. What makes a subgroup *proper*? What makes a subgroup *non-trivial*?

3. The subgroup criterion in Proposition 3.31 says that for H to be a subgroup it is enough for $H \neq \emptyset$ and whenever $g, h \in H$ we also have $gh^{-1} \in H$. Does every subgroup need to contain the identity? If so, why does this criterion not say so?

3.3.4 Exercises

1. Prove that the product of two matrices in $SL_2(\mathbb{R})$ has determinant one.

2. Let G be the group of 2×2 matrices under addition and

$$H = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a + d = 0 \right\}.$$

Prove that H is a subgroup of G.

3. Prove or disprove: $SL_2(\mathbb{Z})$, the set of 2×2 matrices with integer entries and determinant one, is a subgroup of $SL_2(\mathbb{R})$.

4. List the subgroups of the quaternion group, Q_8.

5. Prove or disprove: If H and K are subgroups of a group G, then $H \cup K$ is a subgroup of G.

Hint. Look at S_3.

6. Give an example of an infinite group in which every nontrivial subgroup is infinite.

7. Prove or disprove: Every proper subgroup of a nonabelian group is nonabelian.

8. Let $\mathbb{T} = \{z \in \mathbb{C}^* : |z| = 1\}$. Prove that $\mathbb{T}$ is a subgroup of $\mathbb{C}^*$.

9. Prove that

$$G = \{a + b\sqrt{2} : a, b \in \mathbb{Q} \text{ and } a \text{ and } b \text{ are not both zero}\}$$

is a subgroup of $\mathbb{R}^*$ under the group operation of multiplication.

Hint. The identity of G is $1 = 1 + 0\sqrt{2}$. Since $(a + b\sqrt{2}\,)(c + d\sqrt{2}\,) = (ac + 2bd) + (ad + bc)\sqrt{2}$, G is closed under multiplication. Finally, $(a + b\sqrt{2}\,)^{-1} = a/(a^2 - 2b^2) - b\sqrt{2}/(a^2 - 2b^2)$.

10. Find all the subgroups of $\mathbb{Z}_3 \times \mathbb{Z}_3$. Use this information to show that $\mathbb{Z}_3 \times \mathbb{Z}_3$ is not the same group as $\mathbb{Z}_9$. (See Example 3.28 for a short description of the product of groups.)

11. Find all the subgroups of the symmetry group of an equilateral triangle.

Hint. $H_1 = \{\text{id}\}$, $H_2 = \{\text{id}, \rho_1, \rho_2\}$, $H_3 = \{\text{id}, \mu_1\}$, $H_4 = \{\text{id}, \mu_2\}$, $H_5 = \{\text{id}, \mu_3\}$, S_3.

12. Compute the subgroups of the symmetry group of a square.

13. Let $H = \{2^k : k \in \mathbb{Z}\}$. Show that H is a subgroup of $\mathbb{Q}^*$.

3.4 Isomorphisms

Many groups may appear to be different at first glance, but can be shown to be the same by a simple renaming of the group elements. For example, $\mathbb{Z}_4$ and the subgroup of the circle group $\mathbb{T}$ generated by i can be shown to be the same by demonstrating a one-to-one correspondence between the elements of the two groups and between the group operations. In such a case we say that the groups are isomorphic.

Two groups $(G, \cdot)$ and $(H, \circ)$ are **isomorphic** if there exists a one-to-one and onto map $\phi : G \to H$ such that the group operation is preserved; that is,

$$\phi(a \cdot b) = \phi(a) \circ \phi(b)$$

for all a and b in G. If G is isomorphic to H, we write $G \cong H$. The map ϕ is called an **isomorphism**.

Example 3.32 To show that $\mathbb{Z}_4 \cong \langle i \rangle$, define a map $\phi : \mathbb{Z}_4 \to \langle i \rangle$ by $\phi(n) = i^n$. We must show that ϕ is bijective and preserves the group operation. The map ϕ is one-to-one and onto because

$$\begin{aligned} \phi(0) &= 1 \\ \phi(1) &= i \\ \phi(2) &= -1 \\ \phi(3) &= -i. \end{aligned}$$

Since

$$\phi(m+n) = i^{m+n} = i^m i^n = \phi(m)\phi(n),$$

the group operation is preserved. □

Example 3.33 We can define an isomorphism ϕ from the additive group of real numbers$(\mathbb{R}, +)$ to the multiplicative group of positive real numbers $(\mathbb{R}^+, \cdot)$ with the exponential map; that is,

$$\phi(x+y) = e^{x+y} = e^x e^y = \phi(x)\phi(y).$$

Of course, we must still show that ϕ is one-to-one and onto, but this can be determined using calculus. □

Example 3.34 The integers are isomorphic to the subgroup of $\mathbb{Q}^*$ consisting of elements of the form 2^n. Define a map $\phi : \mathbb{Z} \to \mathbb{Q}^*$ by $\phi(n) = 2^n$. Then

$$\phi(m+n) = 2^{m+n} = 2^m 2^n = \phi(m)\phi(n).$$

By definition the map ϕ is onto the subset $\{2^n : n \in \mathbb{Z}\}$ of $\mathbb{Q}^*$. To show that the map is injective, assume that $m \neq n$. If we can show that $\phi(m) \neq \phi(n)$, then we are done. Suppose that $m > n$ and assume that $\phi(m) = \phi(n)$. Then $2^m = 2^n$ or $2^{m-n} = 1$, which is impossible since $m - n > 0$. □

Example 3.35 The groups $\mathbb{Z}_8$ and $\mathbb{Z}_{12}$ cannot be isomorphic since they have different orders; however, it is true that $U(8) \cong U(12)$. We know that

$$\begin{aligned} U(8) &= \{1, 3, 5, 7\} \\ U(12) &= \{1, 5, 7, 11\}. \end{aligned}$$

An isomorphism $\phi : U(8) \to U(12)$ is then given by

$$\begin{aligned} 1 &\mapsto 1 \\ 3 &\mapsto 5 \\ 5 &\mapsto 7 \\ 7 &\mapsto 11. \end{aligned}$$

The map ϕ is not the only possible isomorphism between these two groups. We could define another isomorphism ψ by $\psi(1) = 1$, $\psi(3) = 11$, $\psi(5) = 5$, $\psi(7) = 7$. In fact, both of these groups are isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$ (see Example 3.28). □

Example 3.36 Even though S_3 and $\mathbb{Z}_6$ possess the same number of elements, we would suspect that they are not isomorphic, since $\mathbb{Z}_6$ is abelian and S_3 is nonabelian. To demonstrate that this is indeed the case, suppose that $\phi : \mathbb{Z}_6 \to S_3$ is an isomorphism. Let $a, b \in S_3$ be two elements such that $ab \neq ba$. Since ϕ is an isomorphism, there exist elements m and n in $\mathbb{Z}_6$ such that

$$\phi(m) = a \quad \text{and} \quad \phi(n) = b.$$

However,

$$ab = \phi(m)\phi(n) = \phi(m+n) = \phi(n+m) = \phi(n)\phi(m) = ba,$$

which contradicts the fact that a and b do not commute. □

Theorem 3.37 *Let $\phi : G \to H$ be an isomorphism of two groups. Then the following statements are true.*

1. *$\phi^{-1} : H \to G$ is an isomorphism.*

2. $|G| = |H|$.

3. If G is abelian, then H is abelian.

4. If G has a subgroup of order n, then H has a subgroup of order n.

Proof. Assertions (1) and (2) follow from the fact that ϕ is a bijection. We will prove (3) here and leave the remainder of the theorem to be proved in the exercises.

(3) Suppose that h_1 and h_2 are elements of H. Since ϕ is onto, there exist elements $g_1, g_2 \in G$ such that $\phi(g_1) = h_1$ and $\phi(g_2) = h_2$. Therefore,

$$h_1 h_2 = \phi(g_1)\phi(g_2) = \phi(g_1 g_2) = \phi(g_2 g_1) = \phi(g_2)\phi(g_1) = h_2 h_1.$$

■

Activity 3.7 Find five non-isomorphic groups of order 8. Say why these groups cannot be isomorphic.

Theorem 3.38 *The isomorphism of groups determines an equivalence relation on the class of all groups.*

Hence, we can modify our goal of classifying all groups to classifying all groups **up to isomorphism**; that is, we will consider two groups to be the same if they are isomorphic.

One of the basic ideas of algebra is the concept of a homomorphism, a natural generalization of an isomorphism. If we relax the requirement that an isomorphism of groups be bijective, we have a homomorphism.

A **homomorphism** between groups $(G, \cdot)$ and $(H, \circ)$ is a map $\phi : G \to H$ such that

$$\phi(g_1 \cdot g_2) = \phi(g_1) \circ \phi(g_2)$$

for $g_1, g_2 \in G$. The range of ϕ in H is called the **homomorphic image** of ϕ.

Example 3.39 Let G be a group and $g \in G$. Define a map $\phi : \mathbb{Z} \to G$ by $\phi(n) = g^n$. Then ϕ is a group homomorphism, since

$$\phi(m + n) = g^{m+n} = g^m g^n = \phi(m)\phi(n).$$

□

Example 3.40 Let $G = GL_2(\mathbb{R})$. If

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in G, then the determinant is nonzero; that is, $\det(A) = ad - bc \neq 0$. Also, for any two elements A and B in G, $\det(AB) = \det(A)\det(B)$. Using the determinant, we can define a homomorphism $\phi : GL_2(\mathbb{R}) \to \mathbb{R}^*$ by $A \mapsto \det(A)$.

□

Reading Questions

1. List three properties of a group that are preserved by an isomorphism.

2. True or false: for two groups to be isomorphic, they must have the same operation. Briefly explain.

3. What is another name for a *bijective homomorphism*? Briefly explain.

Exercises

1. Prove that $\mathbb{C}^*$ is isomorphic to the subgroup of $GL_2(\mathbb{R})$ consisting of matrices of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Hint. Define $\phi : \mathbb{C}^* \to GL_2(\mathbb{R})$ by

$$\phi(a+bi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

2. Prove or disprove: $U(8) \cong \mathbb{Z}_4$.

Hint. False.

3. Prove that $U(8)$ is isomorphic to the group of matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

4. Show that $U(5)$ is isomorphic to $U(10)$, but $U(12)$ is not.

5. Let $G = \mathbb{R} \setminus \{-1\}$ and define a binary operation on G by

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

Prove that G is a group under this operation. Show that $(G, *)$ is isomorphic to the multiplicative group of nonzero real numbers.

6. Show that the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

form a group. Find an isomorphism of G with a more familiar group of order 6.

7. Let $\phi : G \to H$ be a group isomorphism. Show that $\phi(x) = e_H$ if and only if $x = e_G$, where e_G and e_H are the identities of G and H, respectively.

8. Let $\phi : G_1 \to G_2$ and $\psi : G_2 \to G_3$ be isomorphisms. Show that ϕ^{-1} and $\psi \circ \phi$ are both isomorphisms. Using these results, show that the isomorphism of groups determines an equivalence relation on the class of all groups.

9. Let G and H be isomorphic groups. If G has a subgroup of order n, prove that H must also have a subgroup of order n.

3.5 Summary and Additional Exercises

3.5.1 The Important Ideas

- A **group** $(G, \circ)$ is a set G together with a law of composition $(a, b) \mapsto a \circ b$ that is associative, has an identity, and each element of the group has an inverse. If $a \circ b = b \circ a$, then the group is **commutative** or

abelian. Groups not satisfying this property are said to be **nonabelian** or **noncommutative**.

- A **subgroup** H of a group G is a subset H of G such that when the group operation of G is restricted to H, H is also group. The subgroup $H = \{e\}$ of a group G is called the **trivial subgroup**. A subgroup that is a proper subset of G is called a **proper subgroup**.**Cayley tables** are useful for representing small groups.
- Groups have important properties, including
  - The uniqueness of the identity and inverse.
  - The inverse of the product ab is $(ab)^{-1} = b^{-1}a^{-1}$.
  - Right and left cancellation hold, $ba = ca$ implies $b = c$ and $ab = ac$ implies $b = c$ for elements $a, b, c \in G$.
- Important examples of groups include:
  - The integers ($\mathbb{Z}$), the real numbers ($\mathbb{R}$), and the complex numbers ($\mathbb{C}$) under addition.
  - The group of units ($U(n)$).
  - The nonzero rational numbers ($\mathbb{Q}^*$), the nonzero real numbers ($\mathbb{R}^*$), and the nonzero complex numbers ($\mathbb{C}^*$) under multiplication.
  - The integers mod n ($\mathbb{Z}_n$).
  - Symmetry groups such as the symmetries of an equilateral triangle (S_3), the symmetries of a rectangle, and the symmetries of a square (D_4).
  - The quaternion group, Q_8.
  - Groups of matrices such as $\mathbb{M}_2(\mathbb{R})$, $GL_2(\mathbb{R})$, and $SL_2(\mathbb{R})$.
- There are several tools that tell us about group properties (Proposition 3.17, Proposition 3.18, Proposition 3.19, Proposition 3.20, Proposition 3.21, Proposition 3.22, Theorem 3.23) or can be used to prove a subset is a subgroup (Proposition 3.30, Proposition 3.31).
- Two groups $(G, \cdot)$ and $(H, \circ)$ are **isomorphic** if there exists a one-to-one and onto map $\phi : G \to H$ such that

  $$\phi(a \cdot b) = \phi(a) \circ \phi(b)$$

  for all a and b in G. If G is isomorphic to H, we write $G \cong H$.
- If $\phi : G \to H$ is an isomorphism of two groups, then the following statements are true.
  1. $\phi^{-1} : H \to G$ is an isomorphism.
  2. $|G| = |H|$.
  3. If G is abelian, then H is abelian.
  4. If G has a subgroup of order n, then H has a subgroup of order n.
- A **homomorphism** between two groups $(G, \cdot)$ and $(H, \circ)$ is a map $\phi : G \to H$ that preserves the group operation. That is,

  $$\phi(a \cdot b) = \phi(a) \circ \phi(b)$$

  for all a and b in G.

3.5.2 Additional Exercises

1. Let $U(n)$ be the group of units in $\mathbb{Z}_n$. If $n > 2$, prove that there is an element $k \in U(n)$ such that $k^2 = 1$ and $k \neq 1$.

2. Prove Theorem 3.23.

3. Show that if G is a finite group of even order, then there is an $a \in G$ such that a is not the identity and $a^2 = e$.

4. Let G be a group and suppose that $(ab)^2 = a^2b^2$ for all a and b in G. Prove that G is an abelian group.

5. Let $n = 0, 1, 2, \ldots$ and $n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$. Prove that $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. Show that these subgroups are the only subgroups of $\mathbb{Z}$.

6. Prove or disprove: If H and K are subgroups of a group G, then $HK = \{hk : h \in H \text{ and } k \in K\}$ is a subgroup of G. What if G is abelian?

7. Let G be a group and $g \in G$. Show that

$$Z(G) = \{x \in G : gx = xg \text{ for all } g \in G\}$$

is a subgroup of G. This subgroup is called the **center** of G.

8. Let H be a subgroup of G and

$$C(H) = \{g \in G : gh = hg \text{ for all } h \in H\}.$$

Prove $C(H)$ is a subgroup of G. This subgroup is called the **centralizer** of H in G.

9. Let H be a subgroup of G. If $g \in G$, show that $gHg^{-1} = \{ghg^{-1} : h \in H\}$ is also a subgroup of G.

10. **ISBN Codes.** Every book has an International Standard Book Number (ISBN) code. This is a 10-digit code indicating the book's publisher and title. The tenth digit is a check digit satisfying

$$(d_1, d_2, \ldots, d_{10}) \cdot (10, 9, \ldots, 1) \equiv 0 \pmod{11}.$$

One problem is that d_{10} might have to be a 10 to make the inner product zero; in this case, 11 digits would be needed to make this scheme work. Therefore, the character X is used for the eleventh digit. So ISBN 3-540-96035-X is a valid ISBN code.

(a) Is ISBN 0-534-91500-0 a valid ISBN code? What about ISBN 0-534-91700-0 and ISBN 0-534-19500-0?

(b) Does this method detect all single-digit errors? What about all transposition errors?

(c) How many different ISBN codes are there?

(d) A publisher has houses in Germany and the United States. Its German prefix is 3-540. If its United States prefix will be 0-abc, find abc such that the rest of the ISBN code will be the same for a book printed in Germany and in the United States. Under the ISBN coding method the first digit identifies the language; German is 3 and English is 0. The next group of numbers identifies the publisher, and the last group identifies the specific book.

3.6 Connections to the Secondary Classroom—Symmetry

The Common Core State Standards describes the importance of symmetry in high school geometry.[1]

> The concepts of congruence, similarity, and symmetry can be understood from the perspective of geometric transformation. Fundamental are the rigid motions: translations, rotations, reflections, and combinations of these, all of which are here assumed to preserve distance and angles (and therefore shapes generally). Reflections and rotations each explain a particular type of symmetry, and the symmetries of an object offer insight into its attributes—as when the reflective symmetry of an isosceles triangle assures that its base angles are congruent.

A **rigid motion** occurs when a point or object is moved, but the size and shape remain the same. A rigid motion differs from non-rigid motion, such as a dilation, where the size of the object can increase or decrease. In other words, a rigid motion is any way of moving all the points in the plane such that the relative distance between points stays the same and the relative position of the points stays the same. The four types of rigid motions in the plan that are usually consider in the high school classroom are translations, rotations, reflections, and glide reflections.

In a **translation**, everything is moved by the same amount and in the same direction. Every translation has a direction and a distance. A **rotation** fixes one point and everything rotates by the same amount around that point. A **reflection** fixes a **line of reflection** in the plane and exchanges points from one side of the line with points on the other side of the line at the same distance from the line. A **glide reflection** is a reflection about a line followed by a translation parallel to the line of reflection.

For the time being, we will concentrate on rotations and reflections. Consider the pentagon in Figure 3.41. What can we say about rotational and reflective symmetry?

[1] Comon Core State Standards Initiative: High School Geometry. Retrieved from http://www.corestandards.org/Math/Content/HSG/introduction/, June 28, 2020.

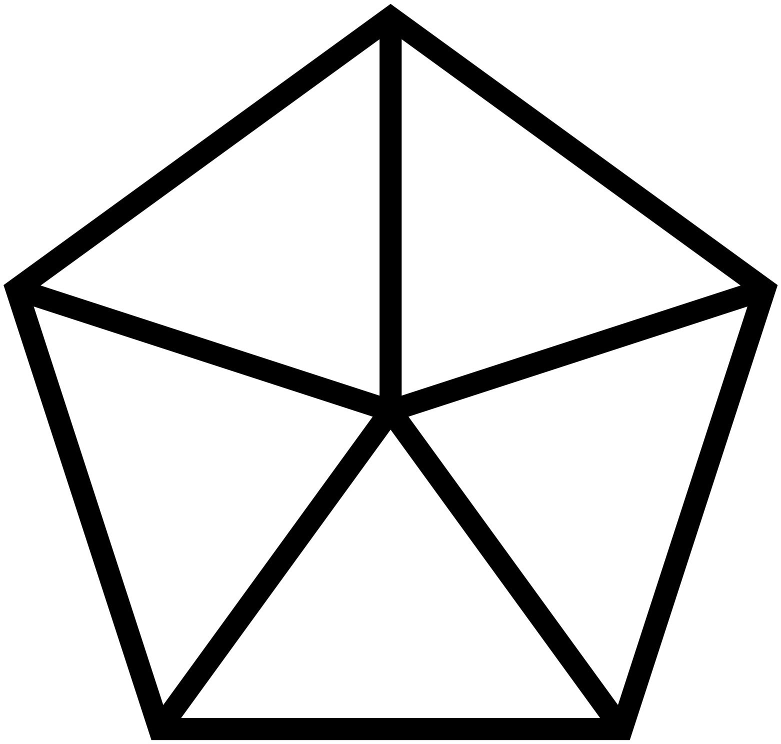


Figure 3.41 Symmetries of a Pentagon

Exercises

1. What are the reflective symmetries of the pentagon? Do the rotational symmetries form a group? Explain your answer.

2. Do the combined rotational and reflective symmetries form a group? Explain your answer.

3. Make a Cayley Table for the combined rotational and reflective symmetries.

4. Suppose that r is a rotation that is not the identity and s is some reflection. Is it true that $rs = sr$? If not, is there some other relation between r and s?

5. Explain why rotation and reflective symmetry are important in the secondary classroom.

6. An **automorphism** of a group G is an isomorphism with itself. Prove that complex conjugation is an automorphism of the additive group of complex numbers; that is, show that the map $\phi(a + bi) = a - bi$ is an isomorphism from $\mathbb{C}$ to $\mathbb{C}$.

7. Prove that $a + ib \mapsto a - ib$ is an automorphism of $\mathbb{C}^*$.

8. Prove that $A \mapsto B^{-1}AB$ is an automorphism of $SL_2(\mathbb{R})$ for all B in $GL_2(\mathbb{R})$.

9. We will denote the set of all automorphisms of G by $\operatorname{Aut}(G)$. Prove that $\operatorname{Aut}(G)$ is a subgroup of S_G, the group of permutations of G.

10. Find $\operatorname{Aut}(\mathbb{Z}_6)$.

Hint. Any automorphism of $\mathbb{Z}_6$ must send 1 to another generator of $\mathbb{Z}_6$.

11. Find $\operatorname{Aut}(\mathbb{Z})$.

12. Find two nonisomorphic groups G and H such that $\operatorname{Aut}(G) \cong \operatorname{Aut}(H)$.

13. Let G be a group and $g \in G$. Define a map $i_g : G \to G$ by $i_g(x) = gxg^{-1}$. Prove that i_g defines an automorphism of G. Such an automorphism is called an **inner automorphism**. The set of all inner automorphisms is denoted by $\operatorname{Inn}(G)$.

14. Prove that $\operatorname{Inn}(G)$ is a subgroup of $\operatorname{Aut}(G)$.

15. What are the inner automorphisms of the quaternion group Q_8? Is $\operatorname{Inn}(G) = \operatorname{Aut}(G)$ in this case?

3.7 References and Suggested Readings

[1] Burnside, W. *Theory of Groups of Finite Order.* 2nd ed. Cambridge University Press, Cambridge, 1911; Dover, New York, 1953. A classic. Also available at books.google.com.

[2] Gallian, J. A. and Winters, S. “Modular Arithmetic in the Marketplace,” *The American Mathematical Monthly* **95** (1988): 548–51.

[3] Gallian, J. A. *Contemporary Abstract Algebra.* 7th ed. Brooks/Cole, Belmont, CA, 2009.

[4] Hall, M. *Theory of Groups.* 2nd ed. American Mathematical Society, Providence, 1959.

[5] Kurosh, A. E. *The Theory of Groups*, vols. I and II. American Mathematical Society, Providence, 1979.

[6] Rotman, J. J. *An Introduction to the Theory of Groups.* 4th ed. Springer, New York, 1995.

Chapter 4

Rings

Objectives

- To understand and be able to apply the definition of a ring.
- To understand and be able to apply the definition of an integral domain and a field.
- To understand and be able to use examples of rings.

In Chapter 3 we studied sets with a single binary operation satisfying certain axioms, but we are often more interested in working with sets that have two binary operations. For example, one of the most natural algebraic structures to study is the integers, $\mathbb{Z}$, with the operations of addition and multiplication. These operations are related to one another by the distributive property. If we consider a set with two such related binary operations satisfying certain axioms, we have an algebraic structure called a ring. In a ring we add and multiply elements such as real numbers, complex numbers, matrices, and functions.

Activity 4.1 List three examples of rings. Find an example of a set with two operations that is not a ring.

4.1 Rings

A nonempty set R is a **ring** if it has two closed binary operations, addition and multiplication, satisfying the following conditions.

1. $a + b = b + a$ for $a, b \in R$.
2. $(a + b) + c = a + (b + c)$ for $a, b, c \in R$.
3. There is an element 0 in R such that $a + 0 = a$ for all $a \in R$.
4. For every element $a \in R$, there exists an element $-a$ in R such that $a + (-a) = 0$.
5. $(ab)c = a(bc)$ for $a, b, c \in R$.
6. For $a, b, c \in R$,

$$\begin{aligned} a(b + c) &= ab + ac \\ (a + b)c &= ac + bc. \end{aligned}$$

This last condition, the distributive axiom, relates the binary operations of addition and multiplication. Notice that the first four axioms simply require that a ring be an abelian group under addition (see Section 3.2), so we could also have defined a ring to be an abelian group $(R, +)$ together with a second binary operation satisfying Item 5 and Item 6 above.

If there is an element $1 \in R$ such that $1 \neq 0$ and $1a = a1 = a$ for each element $a \in R$, we say that R is a ring with **unity** or **identity**. A ring R for which $ab = ba$ for all a, b in R is called a **commutative ring**. A commutative ring R with identity is called an **integral domain** if, for every $a, b \in R$ such that $ab = 0$, either $a = 0$ or $b = 0$. A **division ring** is a ring R, with an identity, in which every nonzero element in R is a **unit**; that is, for each $a \in R$ with $a \neq 0$, there exists a unique element a^{-1} such that $a^{-1}a = aa^{-1} = 1$. A commutative division ring is called a **field**. The relationship among rings, integral domains, division rings, and fields is shown in Figure 4.1.

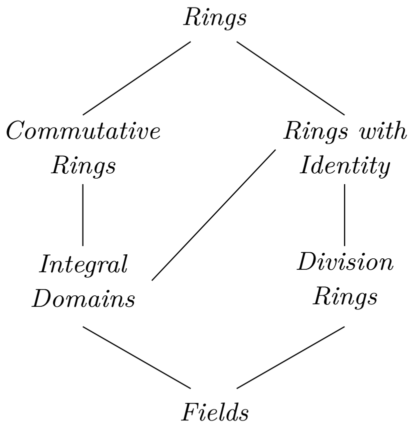


Figure 4.1 Types of rings

Example 4.2 As we have mentioned previously, the integers form a ring. In fact, $\mathbb{Z}$ is an integral domain. Certainly if $ab = 0$ for two integers a and b, either $a = 0$ or $b = 0$. However, $\mathbb{Z}$ is not a field. There is no integer that is the multiplicative inverse of 2, since $1/2$ is not an integer. The only integers with multiplicative inverses are 1 and -1. □

Example 4.3 Under the ordinary operations of addition and multiplication, all of the familiar number systems are rings: the rationals, $\mathbb{Q}$; the real numbers, $\mathbb{R}$; and the complex numbers, $\mathbb{C}$. Each of these rings is a field. □

Example 4.4 We can define the product of two elements a and b in $\mathbb{Z}_n$ by $ab \pmod{n}$. For instance, in $\mathbb{Z}_{12}$, $5 \cdot 7 \equiv 11 \pmod{12}$. This product makes the abelian group $\mathbb{Z}_n$ into a ring. Certainly $\mathbb{Z}_n$ is a commutative ring; however, it may fail to be an integral domain. If we consider $3 \cdot 4 \equiv 0 \pmod{12}$ in $\mathbb{Z}_{12}$, it is easy to see that a product of two nonzero elements in the ring can be equal to zero. □

A nonzero element a in a ring R is called a **zero divisor** if there is a nonzero element b in R such that $ab = 0$. In the previous example, 3 and 4 are zero divisors in $\mathbb{Z}_{12}$.

Example 4.5 In calculus the continuous real-valued functions on an interval $[a, b]$ form a commutative ring. We add or multiply two functions by adding or multiplying the values of the functions. If $f(x) = x^2$ and $g(x) = \cos x$, then $(f + g)(x) = f(x) + g(x) = x^2 + \cos x$ and $(fg)(x) = f(x)g(x) = x^2 \cos x$. □

Example 4.6 The 2×2 matrices with entries in $\mathbb{R}$ form a ring under the usual operations of matrix addition and multiplication. This ring is noncommutative, since it is usually the case that $AB \neq BA$. Also, it is possible that $AB = 0$ when neither A nor B is zero. □

Example 4.7 For an example of a noncommutative division ring, let

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

where $i^2 = -1$. These elements satisfy the following relations:

$$\begin{aligned} \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 &= -1 \\ \mathbf{ij} &= \mathbf{k} \\ \mathbf{jk} &= \mathbf{i} \\ \mathbf{ki} &= \mathbf{j} \\ \mathbf{ji} &= -\mathbf{k} \\ \mathbf{kj} &= -\mathbf{i} \\ \mathbf{ik} &= -\mathbf{j}. \end{aligned}$$

Let $\mathbb{H}$ consist of elements of the form $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, where a, b, c, d are real numbers. Equivalently, $\mathbb{H}$ can be considered to be the set of all 2×2 matrices of the form

$$\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix},$$

where $\alpha = a + di$ and $\beta = b + ci$ are complex numbers. We can define addition and multiplication on $\mathbb{H}$ either by the usual matrix operations or in terms of the generators 1, $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$:

$$\begin{aligned} &(a_1 + b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k}) + (a_2 + b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}) \\ &= (a_1 + a_2) + (b_1 + b_2)\mathbf{i} + (c_1 + c_2)\mathbf{j} + (d_1 + d_2)\mathbf{k} \end{aligned}$$

and

$$(a_1 + b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k})(a_2 + b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}) = \alpha + \beta\mathbf{i} + \gamma\mathbf{j} + \delta\mathbf{k},$$

where

$$\begin{aligned} \alpha &= a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2 \\ \beta &= a_1b_2 + a_2b_1 + c_1d_2 - d_1c_2 \\ \gamma &= a_1c_2 - b_1d_2 + c_1a_2 + d_1b_2 \\ \delta &= a_1d_2 + b_1c_2 - c_1b_2 + d_1a_2. \end{aligned}$$

Though multiplication looks complicated, it is actually a straightforward computation if we remember that we just add and multiply elements in $\mathbb{H}$ like polynomials and keep in mind the relationships between the generators $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$. The ring $\mathbb{H}$ is called the ring of **quaternions**.

To show that the quaternions are a division ring, we must be able to find an inverse for each nonzero element. Notice that

$$(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k})(a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}) = a^2 + b^2 + c^2 + d^2.$$

This element can be zero only if a, b, c, and d are all zero. So if $a+b\mathbf{i}+c\mathbf{j}+d\mathbf{k} \neq 0$,

$$(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) \left(\frac{a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}}{a^2 + b^2 + c^2 + d^2} \right) = 1.$$

□

Activity 4.2 Which of the following sets are rings with respect to the usual operations of addition and multiplication? If the set is a ring, is it also a field?

(a) $7\mathbb{Z}$

(b) $\mathbb{Z}_{18}$

(c) $\mathbb{Q}(\sqrt{2}\,) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$

(d) $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,) = \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} : a, b, c, d \in \mathbb{Q}\}$

(e) $\mathbb{Z}[\sqrt{3}\,] = \{a + b\sqrt{3} : a, b \in \mathbb{Z}\}$

(f) $R = \{a + b\sqrt[3]{3} : a, b \in \mathbb{Q}\}$

(g) $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z} \text{ and } i^2 = -1\}$

(h) $\mathbb{Q}(\sqrt[3]{3}\,) = \{a + b\sqrt[3]{3} + c\sqrt[3]{9} : a, b, c \in \mathbb{Q}\}$

Proposition 4.8 *Let R be a ring with $a, b \in R$. Then*

1. $a0 = 0a = 0$*;*

2. $a(-b) = (-a)b = -ab$*;*

3. $(-a)(-b) = ab$*.*

Activity 4.3 Prove Proposition 4.8.

Just as we have subgroups of groups, we have an analogous class of substructures for rings. A **subring** S of a ring R is a subset S of R such that S is also a ring under the inherited operations from R.

Example 4.9 The ring $n\mathbb{Z}$ is a subring of $\mathbb{Z}$. Notice that even though the original ring may have an identity, we do not require that its subring have an identity. We have the following chain of subrings:

$$\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

□

The following proposition gives us some easy criteria for determining whether or not a subset of a ring is indeed a subring. (We will leave the proof of this proposition as an exercise.)

Proposition 4.10 *Let R be a ring and S a subset of R. Then S is a subring of R if and only if the following conditions are satisfied.*

1. $S \neq \emptyset$*.*

2. $rs \in S$ *for all* $r, s \in S$*.*

3. $r - s \in S$ *for all* $r, s \in S$*.*

Example 4.11 Let $R = \mathbb{M}_2(\mathbb{R})$ be the ring of 2×2 matrices with entries in $\mathbb{R}$. If T is the set of upper triangular matrices in R; i.e.,

$$T = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in \mathbb{R} \right\},$$

then T is a subring of R. If

$$A = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} a' & b' \\ 0 & c' \end{pmatrix}$$

are in T, then clearly $A - B$ is also in T. Also,

$$AB = \begin{pmatrix} aa' & ab' + bc' \\ 0 & cc' \end{pmatrix}$$

is in T. □

In the study of groups, an isomorphism is a one-to-one and onto map that preserves the operation of the group. Similarly, a isomorphism of rings is a bijective map that preserves the operations of addition and multiplication in the ring. More specifically, if R and S are rings, then a **ring isomorphism** is a one-to-one and onto map $\phi : R \to S$ satisfying

$$\begin{aligned} \phi(a+b) &= \phi(a) + \phi(b) \\ \phi(ab) &= \phi(a)\phi(b) \end{aligned}$$

for all $a, b \in R$.

Example 4.12 If we define a map $\phi : \mathbb{C} \to \mathbb{M}_2(\mathbb{R})$ by

$$\phi(a + bi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

then ϕ is an isomorphism of $\mathbb{C}$ with its image in $\mathbb{M}_2(\mathbb{R})$. □

Reading Questions

1. How are groups and rings related to each other? How are they different?

2. What is a field and how do fields relate to rings?

3. What is an integral domain? How do integral domains relate to rings and fields?

Exercises

1. Let R be a ring and S a subset of R. Show that S is a subring of R if and only if each of the following conditions is satisfied.

(a) $S \neq \emptyset$.

(b) $rs \in S$ for all $r, s \in S$.

(c) $r - s \in S$ for all $r, s \in S$.

2. Let a be any element in a ring R with identity. Show that $(-1)a = -a$.

3. Let R and S be arbitrary rings. Show that their Cartesian product is a ring if we define addition and multiplication in $R \times S$ by

(a) $(r, s) + (r', s') = (r + r', s + s')$

(b) $(r, s)(r', s') = (rr', ss')$

4. List or characterize all of the units in each of the following rings.

(a) $\mathbb{Z}_{10}$

(b) $\mathbb{Z}_{12}$

(c) $\mathbb{Z}_7$

(d) $\mathbb{M}_2(\mathbb{Z})$,the 2×2 matrices with entries in $\mathbb{Z}$

(e) $\mathbb{M}_2(\mathbb{Z}_2)$, the 2×2 matrices with entries in $\mathbb{Z}_2$

Hint. (a) $\{1, 3, 7, 9\}$; (c) $\{1, 2, 3, 4, 5, 6\}$; (e)

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \right\}.$$

5. Let R be a ring with a collection of subrings $\{R_\alpha\}$. Prove that $\bigcap R_\alpha$ is a subring of R. Give an example to show that the union of two subrings is not necessarily a subring.

6. Let R be the ring of 2×2 matrices of the form

$$\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix},$$

where $a, b \in \mathbb{R}$. Show that although R is a ring that has no identity, we can find a subring S of R with an identity.

7. Let R be a ring with identity 1_R and S a subring of R with identity 1_S. Prove or disprove that $1_R = 1_S$.

4.2 Integral Domains and Fields

Let us briefly recall some definitions. If R is a ring and r is a nonzero element in R, then r is said to be a **zero divisor** if there is some nonzero element $s \in R$ such that $rs = 0$. A commutative ring with identity is said to be an **integral domain** if it has no zero divisors. If an element a in a ring R with identity has a multiplicative inverse, we say that a is a **unit**. If every nonzero element in a ring R is a unit, then R is called a **division ring**. A commutative division ring is called a **field**.

Example 4.13 If $i^2 = -1$, then the set $\mathbb{Z}[i] = \{m + ni : m, n \in \mathbb{Z}\}$ forms a ring known as the **Gaussian integers**. It is easily seen that the Gaussian integers are a subring of the complex numbers since they are closed under addition and multiplication. Let $\alpha = a + bi$ be a unit in $\mathbb{Z}[i]$. Then $\overline{\alpha} = a - bi$ is also a unit since if $\alpha\beta = 1$, then $\overline{\alpha}\overline{\beta} = 1$. If $\beta = c + di$, then

$$1 = \alpha\beta\overline{\alpha}\overline{\beta} = (a^2 + b^2)(c^2 + d^2).$$

Therefore, $a^2 + b^2$ must either be 1 or -1; or, equivalently, $a + bi = \pm 1$ or $a + bi = \pm i$. Therefore, units of this ring are ± 1 and $\pm i$; hence, the Gaussian integers are not a field. We will leave it as an activity to prove that the Gaussian integers are an integral domain. □

Example 4.14 The set of matrices

$$F = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

with entries in $\mathbb{Z}_2$ forms a field. □

Example 4.15 The set $\mathbb{Q}(\sqrt{2}\,) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ is a field. The inverse of an element $a + b\sqrt{2}$ in $\mathbb{Q}(\sqrt{2}\,)$ is

$$\frac{a}{a^2 - 2b^2} + \frac{-b}{a^2 - 2b^2}\sqrt{2}.$$

□

We have the following alternative characterization of integral domains.

Proposition 4.16 Cancellation Law. *Let D be a commutative ring with identity. Then D is an integral domain if and only if for all nonzero elements $a \in D$ with $ab = ac$, we have $b = c$.*

Proof. Let D be an integral domain. Then D has no zero divisors. Let $ab = ac$ with $a \neq 0$. Then $a(b - c) = 0$. Hence, $b - c = 0$ and $b = c$.

Conversely, let us suppose that cancellation is possible in D. That is, suppose that $ab = ac$ implies $b = c$. Let $ab = 0$. If $a \neq 0$, then $ab = a0$ or $b = 0$. Therefore, a cannot be a zero divisor. ■

The following surprising theorem is due to Wedderburn.

Theorem 4.17 *Every finite integral domain is a field.*

Proof. Let D be a finite integral domain and D^* be the set of nonzero elements of D. We must show that every element in D^* has an inverse. For each $a \in D^*$ we can define a map $\lambda_a : D^* \to D^*$ by $\lambda_a(d) = ad$. This map makes sense, because if $a \neq 0$ and $d \neq 0$, then $ad \neq 0$. The map λ_a is one-to-one, since for $d_1, d_2 \in D^*$,

$$ad_1 = \lambda_a(d_1) = \lambda_a(d_2) = ad_2$$

implies $d_1 = d_2$ by left cancellation. Since D^* is a finite set, the map λ_a must also be onto; hence, for some $d \in D^*$, $\lambda_a(d) = ad = 1$. Therefore, a has a left inverse. Since D is commutative, d must also be a right inverse for a. Consequently, D is a field. ■

For any nonnegative integer n and any element r in a ring R we write $r + \cdots + r$ (n times) as nr. We define the **characteristic** of a ring R to be the least positive integer n such that $nr = 0$ for all $r \in R$. If no such integer exists, then the characteristic of R is defined to be 0. We will denote the characteristic of R by $\operatorname{char} R$.

Example 4.18 For every prime p, $\mathbb{Z}_p$ is a field of characteristic p. By Proposition 3.4, every nonzero element in $\mathbb{Z}_p$ has an inverse; hence, $\mathbb{Z}_p$ is a field. If a is any nonzero element in the field, then $pa = 0$, since the order of any nonzero element in the abelian group $\mathbb{Z}_p$ is p. □

Lemma 4.19 *Let R be a ring with identity. If 1 has order n, then the characteristic of R is n.*

Proof. If 1 has order n, then n is the least positive integer such that $n1 = 0$. Thus, for all $r \in R$,

$$nr = n(1r) = (n1)r = 0r = 0.$$

On the other hand, if no positive n exists such that $n1 = 0$, then the characteristic of R is zero. ■

Theorem 4.20 *The characteristic of an integral domain is either prime or zero.*

Proof. Let D be an integral domain and suppose that the characteristic of D is n with $n \neq 0$. If n is not prime, then $n = ab$, where $1 < a < n$ and $1 < b < n$. By Lemma 4.19, we need only consider the case $n1 = 0$. Since $0 = n1 = (ab)1 = (a1)(b1)$ and there are no zero divisors in D, either $a1 = 0$ or $b1 = 0$. Hence, the characteristic of D must be less than n, which is a contradiction. Therefore, n must be prime. ■

Activity 4.4 Prove that the Gaussian integers, $\mathbb{Z}[i]$, are an integral domain.

4.2.1 Historical Note

Amalie Emmy Noether, one of the outstanding mathematicians of the twentieth century, was born in Erlangen, Germany in 1882. She was the daughter of Max Noether (1844–1921), a distinguished mathematician at the University of

Erlangen. Together with Paul Gordon (1837–1912), Emmy Noether's father strongly influenced her early education. She entered the University of Erlangen at the age of 18. Although women had been admitted to universities in England, France, and Italy for decades, there was great resistance to their presence at universities in Germany. Noether was one of only two women among the university's 986 students. After completing her doctorate under Gordon in 1907, she continued to do research at Erlangen, occasionally lecturing when her father was ill.

Noether went to Göttingen to study in 1916. David Hilbert and Felix Klein tried unsuccessfully to secure her an appointment at Göttingen. Some of the faculty objected to women lecturers, saying, "What will our soldiers think when they return to the university and are expected to learn at the feet of a woman?" Hilbert, annoyed at the question, responded, "Meine Herren, I do not see that the sex of a candidate is an argument against her admission as a Privatdozent. After all, the Senate is not a bathhouse." After Noether passed her habilitation examination in 1919, she was given a title and was paid a small sum for her lectures.

Over the next 11 years she used axiomatic methods to develop an abstract theory of rings and ideals. Though she was not good at lecturing, Noether was an inspiring teacher. One of her many students was B. L. van der Waerden, author of the first text treating abstract algebra from a modern point of view. Some of the other mathematicians Noether influenced or closely worked with were Alexandroff, Artin, Brauer, Courant, Hasse, Hopf, Pontryagin, von Neumann, and Weyl. One of the high points of her career was an invitation to address the International Congress of Mathematicians in Zurich in 1932. In spite of all the recognition she received from her colleagues, Noether's abilities were never recognized as they should have been during her lifetime. She was never promoted to full professor by the Prussian academic bureaucracy.

In 1933, Noether, who was Jewish, was banned from participation in all academic activities in Germany. She emigrated to the United States, took a position at Bryn Mawr College, and became a member of the Institute for Advanced Study at Princeton. Noether died suddenly on April 14, 1935. After her death she was eulogized by such notable scientists as Albert Einstein.

4.2.2 Reading Questions

1. What do we mean by a **zero divisor**? Explain and illustrate with an example.

2. What do we mean by a **unit**? Which elements of the field $\mathbb{Q}$ are units?

3. What is the *Cancellation Law* and how does this relate to zero divisors?

4.2.3 Exercises

1. Prove that $\mathbb{Z}[\sqrt{3}\,i] = \{a + b\sqrt{3}\,i : a, b \in \mathbb{Z}\}$ is an integral domain.

2. Let R be the ring of 2×2 matrices of the form

$$\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix},$$

where $a, b \in \mathbb{R}$. Show that although R is a ring that has no identity, we can find a subring S of R with an identity.

3. What is the characteristic of the field formed by the set of matrices

$$F = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

with entries in $\mathbb{Z}_2$?

4. Let R be a ring, where $a^3 = a$ for all $a \in R$. Prove that R must be a commutative ring.

5. Let R be a ring with identity 1_R and S a subring of R with identity 1_S. Prove or disprove that $1_R = 1_S$.

6. If we do not require the identity of a ring to be distinct from 0, we will not have a very interesting mathematical structure. Let R be a ring such that $1 = 0$. Prove that $R = \{0\}$.

7. Let S be a nonempty subset of a ring R. Prove that there is a subring R' of R that contains S.

4.3 Summary and Additional Exercises

4.3.1 The Important Ideas

- A **ring** is a nonempty set with two closed binary operations, addition and multiplication, satisfying the following conditions.

  1. $a + b = b + a$ for $a, b \in R$.
  2. $(a + b) + c = a + (b + c)$ for $a, b, c \in R$.
  3. There is an element 0 in R such that $a + 0 = a$ for all $a \in R$.
  4. For every element $a \in R$, there exists an element $-a$ in R such that $a + (-a) = 0$.
  5. $(ab)c = a(bc)$ for $a, b, c \in R$.
  6. For $a, b, c \in R$,

  $$\begin{aligned} a(b + c) &= ab + ac \\ (a + b)c &= ac + bc. \end{aligned}$$

- If there is an element $1 \in R$ such that $1 \neq 0$ and $1a = a1 = a$ for each element $a \in R$, we say that R is a ring with **unity** or **identity**. A ring R for which $ab = ba$ for all a, b in R is called a **commutative ring**. A commutative ring R with identity is called an **integral domain** if, for every $a, b \in R$ such that $ab = 0$, either $a = 0$ or $b = 0$. A **division ring** is a ring R, with an identity, in which every nonzero element in R is a **unit**; that is, for each $a \in R$ with $a \neq 0$, there exists a unique element a^{-1} such that $a^{-1}a = aa^{-1} = 1$. A commutative division ring is called a **field**.

- A **subring** S of a ring R is a subset S of R such that S is also a ring under the inherited operations from R.

- A nonzero element in a ring R is a **zero divisor** if there is some nonzero element $s \in R$ such that $rs = 0$.

- Let D be a commutative ring with identity. Then D is an integral domain if and only if for all nonzero elements $a \in D$ with $ab = ac$, we have $b = c$.

4.3.2 Additional Exercises

1. Let R be the ring of 2×2 matrices of the form

$$\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix},$$

where $a, b \in \mathbb{R}$. Show that although R is a ring that has no identity, we can find a subring S of R with an identity.

2. What is the characteristic of the field formed by the set of matrices

$$F = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

with entries in $\mathbb{Z}_2$?

3. A ring R is a **Boolean ring** if for every $a \in R$, $a^2 = a$. Show that every Boolean ring is a commutative ring.

Hint. Compute $(a+b)^2$ and $(-ab)^2$.

4. Let R be a ring, where $a^3 = a$ for all $a \in R$. Prove that R must be a commutative ring.

5. Let R be a ring with identity 1_R and S a subring of R with identity 1_S. Prove or disprove that $1_R = 1_S$.

6. If we do not require the identity of a ring to be distinct from 0, we will not have a very interesting mathematical structure. Let R be a ring such that $1 = 0$. Prove that $R = \{0\}$.

7. Let S be a nonempty subset of a ring R. Prove that there is a subring R' of R that contains S.

8. Let R be a ring. Define the **center** of R to be

$$Z(R) = \{a \in R : ar = ra \text{ for all } r \in R\}.$$

Prove that $Z(R)$ is a commutative subring of R.

9. Let p be prime. Prove that

$$\mathbb{Z}_{(p)} = \{a/b : a, b \in \mathbb{Z} \text{ and } \gcd(b, p) = 1\}$$

is a ring. The ring $\mathbb{Z}_{(p)}$ is called the **ring of integers localized at** p.

Hint. Let $a/b, c/d \in \mathbb{Z}_{(p)}$. Then $a/b + c/d = (ad + bc)/bd$ and $(a/b) \cdot (c/d) = (ac)/(bd)$ are both in $\mathbb{Z}_{(p)}$, since $\gcd(bd, p) = 1$.

10. An element x in a ring is called an **idempotent** if $x^2 = x$. Prove that the only idempotents in an integral domain are 0 and 1. Find a ring with a idempotent x not equal to 0 or 1.

Hint. Suppose that $x^2 = x$ and $x \neq 0$. Since R is an integral domain, $x = 1$. To find a nontrivial idempotent, look in $\mathbb{M}_2(\mathbb{R})$.

11. Let $\gcd(a, n) = d$ and $\gcd(b, d) \neq 1$. Prove that $ax \equiv b \pmod{n}$ does not have a solution.

4.4 Connections to the Secondary Classroom—Zero Divisors

Recall that a zero divisor is a nonzero element a in a ring R such that there is a nonzero element b in R where $ab = 0$. For example, 3 and 4 are zero divisors in $\mathbb{Z}_{12}$, since $3 \cdot 4 \equiv 0 \pmod{12}$. An integral domain is a commutative ring with identity that has no zero divisors. The integers, $\mathbb{Z}$, and the polynomials, $\mathbb{Z}[x]$, with coefficients in $\mathbb{Z}$ are examples of integral domains.

Checkpoint 4.21 Find all of the zero divisors in $\mathbb{Z}_{12} = \{0, 1, 2, \ldots, 11\}$.

Suppose that we wish to finds the roots of $f(x) = x^2 - 2x + 15 = (x-5)(x+3)$, where $f(x)$ is a polynomial with coeficients in F. To find the roots of $f(x)$, we could use the quadratic formula, but it is quicker to use the factored form of the polynomial. If $p(x) = (x-5)(x+3) = 0$, then either $x - 5 = 0$ or $x + 3 = 0$, since $\mathbb{Z}[x]$ is an integral domain. If $x - 5 = 0$, then $x = 5$. If $x + 3 = 0$, then $x = -3$. Since the polynomial has degree 2, it has exactly two roots.

Checkpoint 4.22 Find the zeros of $p(x) = x^2 - 7x + 12$.

On the other hand, suppose that $g(x) = x^2 + 6x + 8$ has coefficients in $\mathbb{Z}_{12}$. The polynomial factors as $g(x) = (x+2)(x+4)$ and must have roots $x = 10$ and $x = 8$. However, $x = 2$ and $x = 4$ are also roots.

Checkpoint 4.23 Explain why it possible for $g(x) = x^2 + 6x + 8$ to have more than two roots in $\mathbb{Z}_{12}$. Are there any additional roots? What are the possible factorizations of $g(x)$?

Exercises

1. Find all of the roots of each of the following polynomials. Factor each polynomial if possible.

(a) $p(x) = x^2 + 2x + 1$ with coefficients in $\mathbb{Z}$.

(b) $p(x) = x^2 + 2x + 1$ with coefficients in $\mathbb{Z}_5$.

(c) $p(x) = x^2 + 2x + 1$ with coefficients in $\mathbb{Z}_8$.

(d) $p(x) = x^2 + 2x + 1$ with coefficients in $\mathbb{Z}_3$.

(e) $p(x) = x^2 + 2x + 1$ with coefficients in $\mathbb{Z}_{10}$.

2. Explain the relationship between

4.5 References and Suggested Readings

[1] Anderson, F. W. and Fuller, K. R. *Rings and Categories of Modules.* 2nd ed. Springer, New York, 1992.

[2] Atiyah, M. F. and MacDonald, I. G. *Introduction to Commutative Algebra.* Westview Press, Boulder, CO, 1994.

[3] Herstein, I. N. *Noncommutative Rings.* Mathematical Association of America, Washington, DC, 1994.

[4] Kaplansky, I. *Commutative Rings.* Revised edition. University of Chicago Press, Chicago, 1974.

[5] Lidl, R. and Pilz, G. *Applied Abstract Algebra.* 2nd ed. Springer, New York, 1998. A good source for applications.

[6] Mackiw, G. *Applications of Abstract Algebra.* Wiley, New York, 1985.

[7] McCoy, N. H. *Rings and Ideals.* Carus Monograph Series, No. 8. Mathematical Association of America, Washington, DC, 1968.

[8] McCoy, N. H. *The Theory of Rings.* Chelsea, New York, 1972.

[9] Zariski, O. and Samuel, P. *Commutative Algebra*, vols. I and II. Springer, New York, 1975, 1960.

Part II

Topics in Group Theory

Chapter 5

Cyclic Groups

Objectives

- To understand and be able to apply the definition of a cyclic group.
- To understand and be able to use $\mathbb{Z}$ and $\mathbb{Z}_n$ as examples of groups.
- To understand and be able to apply the complex roots of unity.

The groups $\mathbb{Z}$ and $\mathbb{Z}_n$, which are among the most familiar and easily understood groups, are both examples of what are called cyclic groups. In this chapter we will study the properties of cyclic groups and cyclic subgroups, which play a fundamental part in the classification of all abelian groups.

5.1 Cyclic Subgroups

Often a subgroup will depend entirely on a single element of the group; that is, knowing that particular element will allow us to compute any other element in the subgroup.

Example 5.1 Suppose that we consider $3 \in \mathbb{Z}$ and look at all multiples (both positive and negative) of 3. As a set, this is

$$3\mathbb{Z} = \{\ldots, -3, 0, 3, 6, \ldots\}.$$

It is easy to see that $3\mathbb{Z}$ is a subgroup of the integers. This subgroup is completely determined by the element 3 since we can obtain all of the other elements of the group by taking multiples of 3. Every element in the subgroup is "generated" by 3. □

Example 5.2 If $H = \{2^n : n \in \mathbb{Z}\}$, then H is a subgroup of the multiplicative group of nonzero rational numbers, $\mathbb{Q}^*$. If $a = 2^m$ and $b = 2^n$ are in H, then $ab^{-1} = 2^m 2^{-n} = 2^{m-n}$ is also in H. By Proposition 3.31, H is a subgroup of $\mathbb{Q}^*$ determined by the element 2. □

Theorem 5.3 *Let G be a group and a be any element in G. Then the set*

$$\langle a \rangle = \{a^k : k \in \mathbb{Z}\}$$

is a subgroup of G. Furthermore, $\langle a \rangle$ is the smallest subgroup of G that contains a.

Proof. The identity is in $\langle a \rangle$ since $a^0 = e$. If g and h are any two elements in $\langle a \rangle$, then by the definition of $\langle a \rangle$ we can write $g = a^m$ and $h = a^n$ for some integers m and n. So $gh = a^m a^n = a^{m+n}$ is again in $\langle a \rangle$. Finally, if $g = a^n$ in $\langle a \rangle$, then the inverse $g^{-1} = a^{-n}$ is also in $\langle a \rangle$. Clearly, any subgroup H of G containing a must contain all the powers of a by closure; hence, H contains $\langle a \rangle$. Therefore, $\langle a \rangle$ is the smallest subgroup of G containing a. ■

Remark 5.4 If we are using the "+" notation, as in the case of the integers under addition, we write $\langle a \rangle = \{na : n \in \mathbb{Z}\}$.

For $a \in G$, we call $\langle a \rangle$ the **cyclic subgroup** generated by a. If G contains some element a such that $G = \langle a \rangle$, then G is a **cyclic group**. In this case a is a **generator** of G. If a is an element of a group G, we define the **order** of a to be the smallest positive integer n such that $a^n = e$, and we write $|a| = n$. If there is no such integer n, we say that the order of a is infinite and write $|a| = \infty$ to denote the order of a.

Example 5.5 Notice that a cyclic group can have more than a single generator. Both 1 and 5 generate $\mathbb{Z}_6$; hence, $\mathbb{Z}_6$ is a cyclic group. Not every element in a cyclic group is necessarily a generator of the group. The order of $2 \in \mathbb{Z}_6$ is 3. The cyclic subgroup generated by 2 is $\langle 2 \rangle = \{0, 2, 4\}$. □

The groups $\mathbb{Z}$ and $\mathbb{Z}_n$ are cyclic groups. The elements 1 and -1 are generators for $\mathbb{Z}$. We can certainly generate $\mathbb{Z}_n$ with 1 although there may be other generators of $\mathbb{Z}_n$, as in the case of $\mathbb{Z}_6$.

Example 5.6 The group of units, $U(9)$, in $\mathbb{Z}_9$ is a cyclic group. As a set, $U(9)$ is $\{1, 2, 4, 5, 7, 8\}$. The element 2 is a generator for $U(9)$ since

$$\begin{array}{ll} 2^1 = 2 & 2^2 = 4 \\ 2^3 = 8 & 2^4 = 7 \\ 2^5 = 5 & 2^6 = 1. \end{array}$$

□

Example 5.7 Not every group is a cyclic group. Consider the symmetry group of an equilateral triangle S_3. The multiplication table for this group is Table 3.7. The subgroups of S_3 are shown in Figure 5.8. Notice that every subgroup is cyclic; however, no single element generates the entire group.

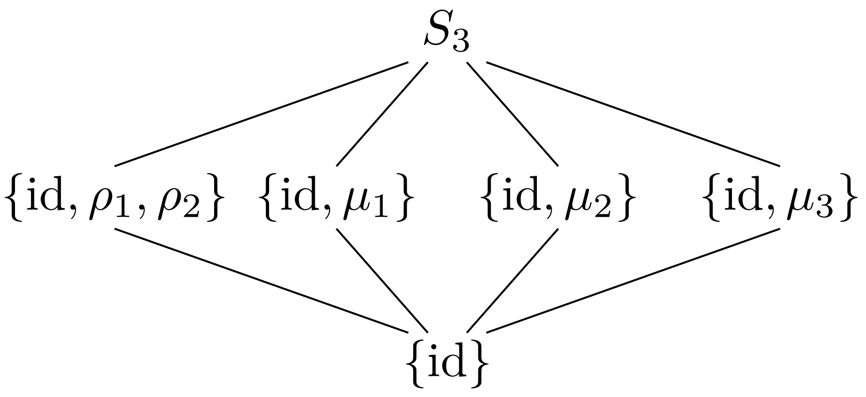


Figure 5.8 Subgroups of S_3

□

Theorem 5.9 *Every cyclic group is abelian.*

Proof. Let G be a cyclic group and $a \in G$ be a generator for G. If g and h are in G, then they can be written as powers of a, say $g = a^r$ and $h = a^s$. Since

$$gh = a^r a^s = a^{r+s} = a^{s+r} = a^s a^r = hg,$$

G is abelian. ■

Activity 5.1 Prove or disprove each of the following statements.

1. All of the generators of $\mathbb{Z}_{60}$ are prime.
2. $U(8)$ is cyclic.
3. $\mathbb{Q}$ is cyclic.
4. If every proper subgroup of a group G is cyclic, then G is a cyclic group.
5. A group with a finite number of subgroups is finite.

5.1.1 Subgroups of Cyclic Groups

We can ask some interesting questions about cyclic subgroups of a group and subgroups of a cyclic group. If G is a group, which subgroups of G are cyclic? If G is a cyclic group, what type of subgroups does G possess?

Theorem 5.10 *Every subgroup of a cyclic group is cyclic.*

Proof. The main tools used in this proof are the division algorithm and the Principle of Well-Ordering. Let G be a cyclic group generated by a and suppose that H is a subgroup of G. If $H = \{e\}$, then trivially H is cyclic. Suppose that H contains some other element g distinct from the identity. Then g can be written as a^n for some integer n. Since H is a subgroup, $g^{-1} = a^{-n}$ must also be in H. Since either n or $-n$ is positive, we can assume that H contains positive powers of a and $n > 0$. Let m be the smallest natural number such that $a^m \in H$. Such an m exists by the Principle of Well-Ordering.

We claim that $h = a^m$ is a generator for H. We must show that every $h' \in H$ can be written as a power of h. Since $h' \in H$ and H is a subgroup of G, $h' = a^k$ for some integer k. Using the division algorithm, we can find numbers q and r such that $k = mq + r$ where $0 \leq r < m$; hence,

$$a^k = a^{mq+r} = (a^m)^q a^r = h^q a^r.$$

So $a^r = a^k h^{-q}$. Since a^k and h^{-q} are in H, a^r must also be in H. However, m was the smallest positive number such that a^m was in H; consequently, $r = 0$ and so $k = mq$. Therefore,

$$h' = a^k = a^{mq} = h^q$$

and H is generated by h. ■

Corollary 5.11 *The subgroups of $\mathbb{Z}$ are exactly $n\mathbb{Z}$ for $n = 0, 1, 2, \ldots$.*

Proposition 5.12 *Let G be a cyclic group of order n and suppose that a is a generator for G. Then $a^k = e$ if and only if n divides k.*

Proof. First suppose that $a^k = e$. By the division algorithm, $k = nq + r$ where $0 \leq r < n$; hence,

$$e = a^k = a^{nq+r} = a^{nq} a^r = e a^r = a^r.$$

Since the smallest positive integer m such that $a^m = e$ is n, $r = 0$.

Conversely, if n divides k, then $k = ns$ for some integer s. Consequently,

$$a^k = a^{ns} = (a^n)^s = e^s = e.$$

■

Theorem 5.13 *Let G be a cyclic group of order n and suppose that $a \in G$ is a generator of the group. If $b = a^k$, then the order of b is n/d, where $d = \gcd(k, n)$.*

Proof. We wish to find the smallest integer m such that $e = b^m = a^{km}$. By Proposition 5.12, this is the smallest integer m such that n divides km or, equivalently, n/d divides $m(k/d)$. Since d is the greatest common divisor of n and k, n/d and k/d are relatively prime. Hence, for n/d to divide $m(k/d)$ it must divide m. The smallest such m is n/d. ■

Corollary 5.14 *The generators of $\mathbb{Z}_n$ are the integers r such that $1 \leq r < n$ and $\gcd(r, n) = 1$.*

Example 5.15 Let us examine the group $\mathbb{Z}_{16}$. The numbers 1, 3, 5, 7, 9, 11, 13, and 15 are the elements of $\mathbb{Z}_{16}$ that are relatively prime to 16. Each of these elements generates $\mathbb{Z}_{16}$. For example,

$$\begin{array}{ccc}
1 \cdot 9 = 9 & 2 \cdot 9 = 2 & 3 \cdot 9 = 11 \\
4 \cdot 9 = 4 & 5 \cdot 9 = 13 & 6 \cdot 9 = 6 \\
7 \cdot 9 = 15 & 8 \cdot 9 = 8 & 9 \cdot 9 = 1 \\
10 \cdot 9 = 10 & 11 \cdot 9 = 3 & 12 \cdot 9 = 12 \\
13 \cdot 9 = 5 & 14 \cdot 9 = 14 & 15 \cdot 9 = 7.
\end{array}$$

□

Activity 5.2

(a) List every generator $\mathbb{Z}_{32}$.

(b) List every generator for each subgroup of order 16 in $\mathbb{Z}_{32}$.

(c) List every generator for each subgroup of order 8 in $\mathbb{Z}_{32}$.

(d) List every generator for each subgroup of order 4 in $\mathbb{Z}_{32}$.

(e) List every generator for each subgroup of order 2 in $\mathbb{Z}_{32}$.

We are now in a position to characterize all cyclic groups.

Theorem 5.16 *All cyclic groups of infinite order are isomorphic to $\mathbb{Z}$.*

Proof. Let G be a cyclic group with infinite order and suppose that a is a generator of G. Define a map $\phi : \mathbb{Z} \to G$ by $\phi : n \mapsto a^n$. Then

$$\phi(m + n) = a^{m+n} = a^m a^n = \phi(m)\phi(n).$$

To show that ϕ is injective, suppose that m and n are two elements in $\mathbb{Z}$, where $m \neq n$. We can assume that $m > n$. We must show that $a^m \neq a^n$. Let us suppose the contrary; that is, $a^m = a^n$. In this case $a^{m-n} = e$, where $m - n > 0$, which contradicts the fact that a has infinite order. Our map is onto since any element in G can be written as a^n for some integer n and $\phi(n) = a^n$. ■

Theorem 5.17 *If G is a cyclic group of order n, then G is isomorphic to $\mathbb{Z}_n$.*

Proof. Let G be a cyclic group of order n generated by a and define a map $\phi : \mathbb{Z}_n \to G$ by $\phi : k \mapsto a^k$, where $0 \leq k < n$. The proof that ϕ is an isomorphism is one of the end-of-chapter exercises. ■

5.1.2 Reading Questions

1. Explain, in your own words, what it means for a group to be cyclic. Your explanation should use the word *generator* at least once.

2. Find two different generators for the group $U(9) = \{1, 2, 4, 5, 7, 8\}$. What does this say about whether $U(9)$ is cyclic? (Compare to the previous question.)

3. Suppose H is a subgroup of a cyclic group G. Must H be abelian? Explain.

5.1.3 Exercises

1. Find the order of each of the following elements.

(a) $5 \in \mathbb{Z}_{12}$

(b) $\sqrt{3} \in \mathbb{R}$

(c) $\sqrt{3} \in \mathbb{R}^*$

(d) $-i \in \mathbb{C}^*$

(e) $72 \in \mathbb{Z}_{240}$

(f) $312 \in \mathbb{Z}_{471}$

Hint. (a) 12; (c) infinite; (e) 10.

2. List all of the elements in each of the following subgroups.

(a) The subgroup of $\mathbb{Z}$ generated by 7

(b) The subgroup of $\mathbb{Z}_{24}$ generated by 15

(c) All subgroups of $\mathbb{Z}_{12}$

(d) All subgroups of $\mathbb{Z}_{60}$

(e) All subgroups of $\mathbb{Z}_{13}$

(f) All subgroups of $\mathbb{Z}_{48}$

(g) The subgroup generated by 3 in $U(20)$

(h) The subgroup generated by 5 in $U(18)$

(i) The subgroup of $\mathbb{R}^*$ generated by 7

(j) The subgroup of $\mathbb{C}^*$ generated by i where $i^2 = -1$

(k) The subgroup of $\mathbb{C}^*$ generated by $2i$

(l) The subgroup of $\mathbb{C}^*$ generated by $(1 + i)/\sqrt{2}$

(m) The subgroup of $\mathbb{C}^*$ generated by $(1 + \sqrt{3}\, i)/2$

Hint. (a) $7\mathbb{Z} = \{\ldots, -7, 0, 7, 14, \ldots\}$; (b) $\{0, 3, 6, 9, 12, 15, 18, 21\}$; (c) $\{0\}$, $\{0, 6\}$, $\{0, 4, 8\}$, $\{0, 3, 6, 9\}$, $\{0, 2, 4, 6, 8, 10\}$; (g) $\{1, 3, 7, 9\}$; (j) $\{1, -1, i, -i\}$.

3. Find the subgroups of $GL_2(\mathbb{R})$ generated by each of the following matrices.

(a) $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

(b) $\begin{pmatrix} 0 & 1/3 \\ 3 & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$

(e) $\begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$

(f) $\begin{pmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{pmatrix}$

Hint. (a)

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}, \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}.$$

(c)

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & -1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}-1 & 1\\ -1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ -1 & 1\end{pmatrix}, \begin{pmatrix}0 & -1\\ 1 & -1\end{pmatrix}, \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}.$$

4. Find the order of every element in $\mathbb{Z}_{18}$.

5. Find the order of every element in the symmetry group of the square, D_4.

6. What are all of the cyclic subgroups of the quaternion group, Q_8?

7. List all of the cyclic subgroups of $U(30)$.

8. Find all elements of finite order in each of the following groups. Here the "$*$" indicates the set with zero removed.

(a) $\mathbb{Z}$ (b) $\mathbb{Q}^*$ (c) $\mathbb{R}^*$

Hint. (a) 0; (b) $1, -1$.

9. If $a^{24} = e$ in a group G, what are the possible orders of a?

Hint. $1, 2, 3, 4, 6, 8, 12, 24$.

10. Find a cyclic group with exactly one generator. Can you find cyclic groups with exactly two generators? Four generators? How about n generators?

11. For $n \leq 20$, which groups $U(n)$ are cyclic? Make a conjecture as to what is true in general. Can you prove your conjecture?

12. Let

$$A = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix}0 & -1\\ 1 & -1\end{pmatrix}$$

be elements in $GL_2(\mathbb{R})$. Show that A and B have finite orders but AB does not.

13. Prove that $\mathbb{Z}_p$ has no nontrivial subgroups if p is prime.

14. If g and h have orders 15 and 16 respectively in a group G, what is the order of $\langle g \rangle \cap \langle h \rangle$?

Hint. $|\langle g \rangle \cap \langle h \rangle| = 1$.

15. Let $\phi : G \to H$ be an isomorphism of groups. If G is cyclic, prove that H must also be cyclic.

5.2 Multiplicative Group of Complex Numbers

The **complex numbers** are defined as

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

where $i^2 = -1$. If $z = a + bi$, then a is the **real part** of z and b is the **imaginary part** of z.

To add two complex numbers $z = a + bi$ and $w = c + di$, we just add the corresponding real and imaginary parts:

$$z + w = (a + bi) + (c + di) = (a + c) + (b + d)i.$$

Remembering that $i^2 = -1$, we multiply complex numbers just like polynomials. The product of z and w is

$$(a + bi)(c + di) = ac + bdi^2 + adi + bci = (ac - bd) + (ad + bc)i.$$

Every nonzero complex number $z = a + bi$ has a multiplicative inverse; that is, there exists a $z^{-1} \in \mathbb{C}^*$ such that $zz^{-1} = z^{-1}z = 1$. If $z = a + bi$, then

$$z^{-1} = \frac{a - bi}{a^2 + b^2}..$$

The **complex conjugate** of a complex number $z = a + bi$ is defined to be $\overline{z} = a - bi$. The **absolute value** or **modulus** of $z = a + bi$ is $|z| = \sqrt{a^2 + b^2}$.

Example 5.18 Let $z = 2 + 3i$ and $w = 1 - 2i$. Then

$$z + w = (2 + 3i) + (1 - 2i) = 3 + i$$

and

$$zw = (2 + 3i)(1 - 2i) = 8 - i.$$

Also,

$$\begin{aligned} z^{-1} &= \frac{2}{13} - \frac{3}{13}i \\ |z| &= \sqrt{13} \\ \overline{z} &= 2 - 3i. \end{aligned}$$

□

Activity 5.3 Evaluate each of the following.

1. $(3 - 2i) + (5i - 6)$
2. $(4 - 5i) - \overline{(4i - 4)}$
3. $(5 - 4i)(7 + 2i)$
4. $(9 - i)\overline{(9 - i)}$
5. i^{45}
6. $(1 + i) + \overline{(1 + i)}$

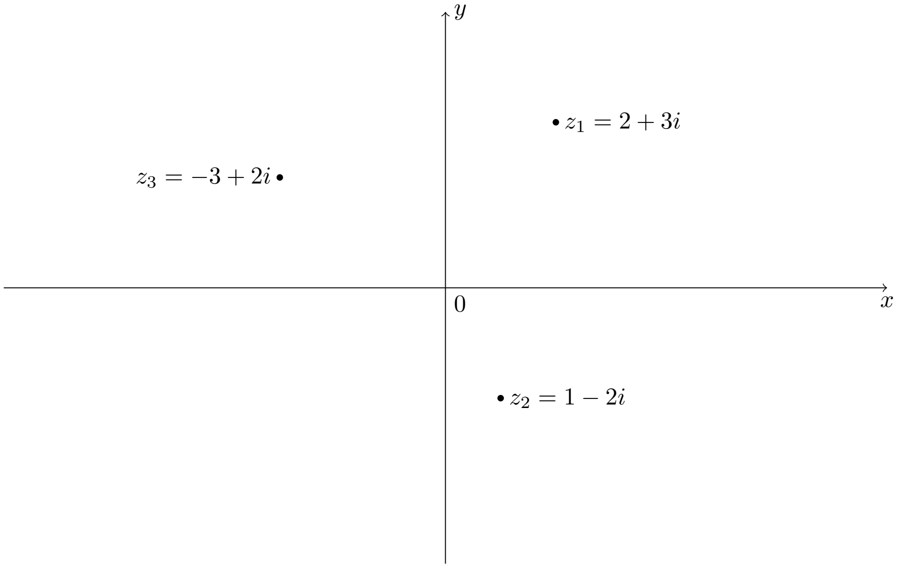


Figure 5.19 Rectangular coordinates of a complex number

There are several ways of graphically representing complex numbers. We can represent a complex number $z = a + bi$ as an ordered pair on the xy plane where a is the x (or real) coordinate and b is the y (or imaginary) coordinate. This is called the **rectangular** or **Cartesian** representation. The rectangular representations of $z_1 = 2 + 3i$, $z_2 = 1 - 2i$, and $z_3 = -3 + 2i$ are depicted in Figure 5.19.

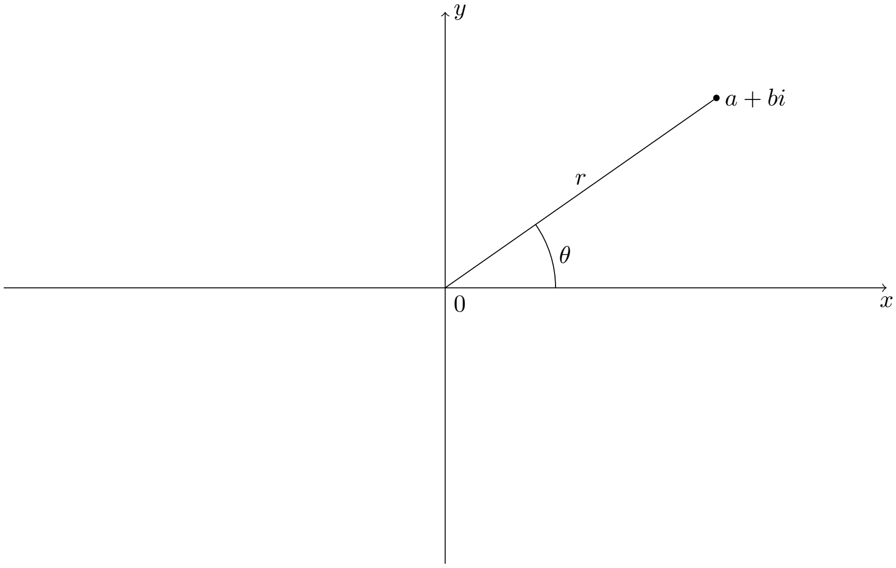


Figure 5.20 Polar coordinates of a complex number

Nonzero complex numbers can also be represented using **polar coordinates**. To specify any nonzero point on the plane, it suffices to give an angle θ from the positive x axis in the counterclockwise direction and a distance r from the origin, as in Figure 5.20. We can see that

$$z = a + bi = r(\cos\theta + i\sin\theta).$$

Hence,

$$r = |z| = \sqrt{a^2 + b^2}$$

and

$$\begin{aligned} a &= r\cos\theta \\ b &= r\sin\theta. \end{aligned}$$

We sometimes abbreviate $r(\cos\theta + i\sin\theta)$ as $r \operatorname{cis}\theta$. To assure that the representation of z is well-defined, we also require that $0° \leq \theta < 360°$. If the measurement is in radians, then $0 \leq \theta < 2\pi$.

Example 5.21 Suppose that $z = 2 \operatorname{cis} 60°$. Then

$$a = 2\cos 60° = 1$$

and

$$b = 2\sin 60° = \sqrt{3}.$$

Hence, the rectangular representation is $z = 1 + \sqrt{3}\, i$.

Conversely, if we are given a rectangular representation of a complex number, it is often useful to know the number's polar representation. If $z = 3\sqrt{2} - 3\sqrt{2}\, i$, then

$$r = \sqrt{a^2 + b^2} = \sqrt{36} = 6$$

and

$$\theta = \arctan\left(\frac{b}{a}\right) = \arctan(-1) = 315°,$$

so $3\sqrt{2} - 3\sqrt{2}\, i = 6 \operatorname{cis} 315°$. □

The polar representation of a complex number makes it easy to find products and powers of complex numbers. The proof of the following proposition is straightforward and is left as an exercise.

Proposition 5.22 *Let $z = r \operatorname{cis}\theta$ and $w = s \operatorname{cis}\phi$ be two nonzero complex numbers. Then*

$$zw = rs \operatorname{cis}(\theta + \phi).$$

Example 5.23 If $z = 3\operatorname{cis}(\pi/3)$ and $w = 2\operatorname{cis}(\pi/6)$, then $zw = 6\operatorname{cis}(\pi/2) = 6i$. □

Activity 5.4 Change the following complex numbers to polar representation.

1. $1 - i$
2. -5
3. $2 + 2i$
4. $\sqrt{3} + i$
5. $-3i$
6. $2i + 2\sqrt{3}$

Theorem 5.24 DeMoivre. *Let $z = r \operatorname{cis}\theta$ be a nonzero complex number. Then*

$$[r \operatorname{cis}\theta]^n = r^n \operatorname{cis}(n\theta)$$

for $n = 1, 2, \ldots$.

Proof. We will use induction on n. For $n = 1$ the theorem is trivial. Assume that the theorem is true for all k such that $1 \leq k \leq n$. Then

$$\begin{aligned} z^{n+1} &= z^n z \\ &= r^n(\cos n\theta + i \sin n\theta) r(\cos\theta + i\sin\theta) \\ &= r^{n+1}[(\cos n\theta\cos\theta - \sin n\theta\sin\theta) + i(\sin n\theta\cos\theta + \cos n\theta\sin\theta)] \\ &= r^{n+1}[\cos(n\theta + \theta) + i\sin(n\theta + \theta)] \\ &= r^{n+1}[\cos(n+1)\theta + i\sin(n+1)\theta]. \end{aligned}$$

■

Example 5.25 Suppose that $z = 1 + i$ and we wish to compute z^{10}. Rather than computing $(1+i)^{10}$ directly, it is much easier to switch to polar coordinates and calculate z^{10} using DeMoivre's Theorem:

$$\begin{aligned} z^{10} &= (1+i)^{10} \\ &= \left(\sqrt{2}\operatorname{cis}\left(\frac{\pi}{4}\right)\right)^{10} \\ &= (\sqrt{2}\,)^{10}\operatorname{cis}\left(\frac{5\pi}{2}\right) \\ &= 32\operatorname{cis}\left(\frac{\pi}{2}\right) \\ &= 32i. \end{aligned}$$

□

5.2.1 The Circle Group and the Roots of Unity

The multiplicative group of the complex numbers, $\mathbb{C}^*$, possesses some interesting subgroups. Whereas $\mathbb{Q}^*$ and $\mathbb{R}^*$ have no interesting subgroups of finite order,

$\mathbb{C}^*$ has many. We first consider the **circle group**,

$$\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}.$$

The following proposition is a direct result of Proposition 5.22.

Proposition 5.26 *The circle group is a subgroup of* $\mathbb{C}^*$*.*

Although the circle group has infinite order, it has many interesting finite subgroups. Suppose that $H = \{1, -1, i, -i\}$. Then H is a subgroup of the circle group. Also, 1, -1, i, and $-i$ are exactly those complex numbers that satisfy the equation $z^4 = 1$. The complex numbers satisfying the equation $z^n = 1$ are called the ***n*****th roots of unity**.

Theorem 5.27 *If* $z^n = 1$*, then the* n*th roots of unity are*

$$z = \operatorname{cis}\left(\frac{2k\pi}{n}\right),$$

where $k = 0, 1, \ldots, n-1$*. Furthermore, the* n*th roots of unity form a cyclic subgroup of* $\mathbb{T}$ *of order* n

Proof. By DeMoivre's Theorem,

$$z^n = \operatorname{cis}\left(n\frac{2k\pi}{n}\right) = \operatorname{cis}(2k\pi) = 1.$$

The z's are distinct since the numbers $2k\pi/n$ are all distinct and are greater than or equal to 0 but less than 2π. The fact that these are all of the roots of the equation $z^n = 1$ follows from from Corollary 10.10, which states that a polynomial of degree n can have at most n roots. We will leave the proof that the nth roots of unity form a cyclic subgroup of $\mathbb{T}$ as an exercise. ■

A generator for the group of the nth roots of unity is called a **primitive** ***n*****th root of unity**.

Example 5.28 The 8th roots of unity can be represented as eight equally spaced points on the unit circle (Figure 5.29). The primitive 8th roots of unity are

$$\begin{aligned}
\omega &= \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i \\
\omega^3 &= -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i \\
\omega^5 &= -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i \\
\omega^7 &= \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i.
\end{aligned}$$

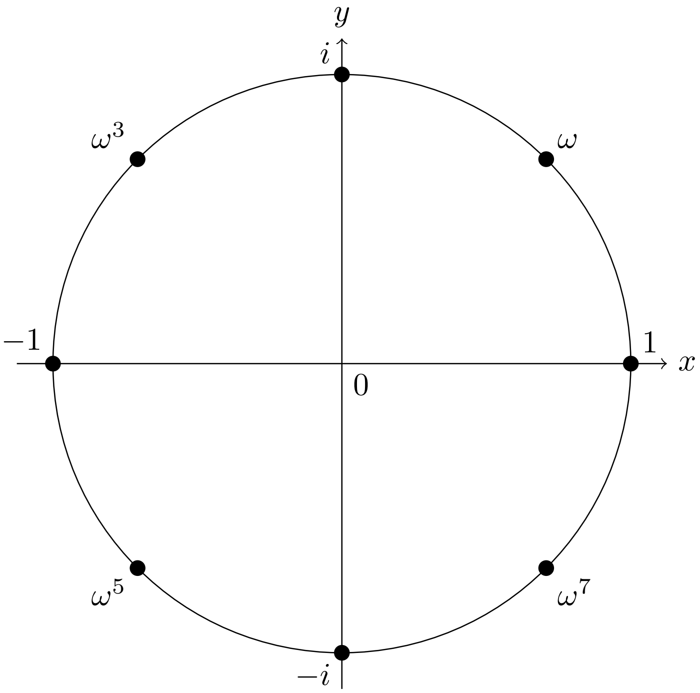


Figure 5.29 8th roots of unity

□

Activity 5.5 List and graph the 6th roots of unity. What are the generators of this group? What are the primitive 6th roots of unity?

5.2.2 Reading Questions

1. True or false: every real number is a complex number. Explain your answer.

2. Which of the following groups are cyclic: $\mathbb{C}$, $\mathbb{C}*$, and the 8th roots of unity. Explain why you are correct.

3. How many different ways can you write a single complex number using the *rectangular* representation? How many different ways can you write a single complex number using the *polar* representation?

4. What is the difference between a nth root of unity and a *primitive* nth root of unity?

5.2.3 Exercises

1. Convert the following complex numbers to the form $a + bi$.

(a) $2\operatorname{cis}(\pi/6)$

(b) $5\operatorname{cis}(9\pi/4)$

(c) $3\operatorname{cis}(\pi)$

(d) $\operatorname{cis}(7\pi/4)/2$

Hint. (a) $\sqrt{3}+i$; (c) -3.

2. Calculate each of the following expressions.

(a) $(1+i)^{-1}$

(b) $(1-i)^6$

(c) $(\sqrt{3}+i)^5$

(d) $(-i)^{10}$

(e) $((1-i)/2)^4$

(f) $(-\sqrt{2}-\sqrt{2}\,i)^{12}$

(g) $(-2+2i)^{-5}$

Hint. (a) $(1-i)/2$; (c) $16(i-\sqrt{3}\,)$; (e) $-1/4$.

3. Prove each of the following statements.

(a) $|z| = |\overline{z}|$

(b) $z\overline{z} = |z|^2$

(c) $z^{-1} = \overline{z}/|z|^2$

(d) $|z+w| \leq |z| + |w|$

(e) $|z-w| \geq ||z| - |w||$

(f) $|zw| = |z||w|$

4. List and graph the 5th roots of unity. What are the generators of this group? What are the primitive 5th roots of unity?

5. If $z = r(\cos\theta + i\sin\theta)$ and $w = s(\cos\phi + i\sin\phi)$ are two nonzero complex numbers, show that

$$zw = rs[\cos(\theta+\phi) + i\sin(\theta+\phi)].$$

6. Prove that the circle group is a subgroup of $\mathbb{C}^*$.

7. Prove that the nth roots of unity form a cyclic subgroup of $\mathbb{T}$ of order n. It follows from Theorem 5.17 that the nth roots of unity are isomorphic to $\mathbb{Z}_n$.

5.3 Direct Products

Given two groups G and H, it is possible to construct a new group from the Cartesian product of G and H, $G \times H$. Conversely, given a large group, it is sometimes possible to decompose the group; that is, a group is sometimes isomorphic to the direct product of two smaller groups. Rather than studying a large group G, it is often easier to study the component groups of G.

5.3.1 External Direct Products

If $(G, \cdot)$ and $(H, \circ)$ are groups, then we can make the Cartesian product of G and H into a new group. As a set, our group is just the ordered pairs $(g,h) \in G \times H$ where $g \in G$ and $h \in H$. We can define a binary operation on $G \times H$ by

$$(g_1, h_1)(g_2, h_2) = (g_1 \cdot g_2, h_1 \circ h_2);$$

that is, we just multiply elements in the first coordinate as we do in G and elements in the second coordinate as we do in H. We have specified the particular operations $\cdot$ and $\circ$ in each group here for the sake of clarity; we usually just write $(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2)$.

Proposition 5.30 *Let G and H be groups. The set $G \times H$ is a group under the operation $(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2)$ where $g_1, g_2 \in G$ and $h_1, h_2 \in H$.*

Proof. Clearly the binary operation defined above is closed. If e_G and e_H are the identities of the groups G and H respectively, then (e_G, e_H) is the identity of $G \times H$. The inverse of $(g,h) \in G \times H$ is (g^{-1}, h^{-1}). The fact that the operation is associative follows directly from the associativity of G and H. ■

Example 5.31 Let $\mathbb{R}$ be the group of real numbers under addition. The Cartesian product of $\mathbb{R}$ with itself, $\mathbb{R}\times\mathbb{R} = \mathbb{R}^2$, is also a group, in which the group operation is just addition in each coordinate; that is, $(a,b)+(c,d) = (a+c, b+d)$. The identity is $(0,0)$ and the inverse of (a,b) is $(-a,-b)$. □

Example 5.32 Consider

$$\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(0,0),(0,1),(1,0),(1,1)\}.$$

Although $\mathbb{Z}_2 \times \mathbb{Z}_2$ and $\mathbb{Z}_4$ both contain four elements, they are not isomorphic. Every element (a,b) in $\mathbb{Z}_2\times\mathbb{Z}_2$ has order 2, since $(a,b)+(a,b) = (0,0)$; however, $\mathbb{Z}_4$ is cyclic. □

The group $G \times H$ is called the **external direct product** of G and H. Notice that there is nothing special about the fact that we have used only two groups to build a new group. The direct product

$$\prod_{i=1}^{n} G_i = G_1 \times G_2 \times \cdots \times G_n$$

of the groups $G_1, G_2, \ldots, G_n$ is defined in exactly the same manner. If $G = G_1 = G_2 = \cdots = G_n$, we often write G^n instead of $G_1 \times G_2 \times \cdots \times G_n$.

Example 5.33 The group $\mathbb{Z}_2^n$, considered as a set, is just the set of all binary n-tuples. The group operation is the "exclusive or" of two binary n-tuples. For example,

$$(01011101) + (01001011) = (00010110).$$

This group is important in coding theory, in cryptography, and in many areas of computer science. □

Theorem 5.34 *Let $(g,h) \in G \times H$. If g and h have finite orders r and s respectively, then the order of (g,h) in $G \times H$ is the least common multiple of r and s.*

Proof. Suppose that m is the least common multiple of r and s and let $n = |(g,h)|$. Then

$$(g,h)^m = (g^m, h^m) = (e_G, e_H)$$
$$(g^n, h^n) = (g,h)^n = (e_G, e_H).$$

Hence, n must divide m, and $n \leq m$. However, by the second equation, both r and s must divide n; therefore, n is a common multiple of r and s. Since m is the *least common multiple* of r and s, $m \leq n$. Consequently, m must be equal to n. ■

Corollary 5.35 *Let $(g_1, \ldots, g_n) \in \prod G_i$. If g_i has finite order r_i in G_i, then the order of $(g_1, \ldots, g_n)$ in $\prod G_i$ is the least common multiple of $r_1, \ldots, r_n$.*

Example 5.36 Let $(8,56) \in \mathbb{Z}_{12} \times \mathbb{Z}_{60}$. Since $\gcd(8,12) = 4$, the order of 8 is $12/4 = 3$ in $\mathbb{Z}_{12}$. Similarly, the order of 56 in $\mathbb{Z}_{60}$ is 15. The least common multiple of 3 and 15 is 15; hence, $(8,56)$ has order 15 in $\mathbb{Z}_{12} \times \mathbb{Z}_{60}$. □

Example 5.37 The group $\mathbb{Z}_2 \times \mathbb{Z}_3$ consists of the pairs

$$(0,0), \qquad (0,1), \qquad (0,2), \qquad (1,0), \qquad (1,1), \qquad (1,2).$$

In this case, unlike that of $\mathbb{Z}_2 \times \mathbb{Z}_2$ and $\mathbb{Z}_4$, it is true that $\mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6$. We need only show that $\mathbb{Z}_2 \times \mathbb{Z}_3$ is cyclic. It is easy to see that $(1,1)$ is a generator for $\mathbb{Z}_2 \times \mathbb{Z}_3$. □

The next theorem tells us exactly when the direct product of two cyclic groups is cyclic.

Theorem 5.38 *The group $\mathbb{Z}_m \times \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_{mn}$ if and only if* $\gcd(m, n) = 1$.

Proof. We will first show that if $\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$, then $\gcd(m, n) = 1$. We will prove the contrapositive; that is, we will show that if $\gcd(m, n) = d > 1$, then $\mathbb{Z}_m \times \mathbb{Z}_n$ cannot be cyclic. Notice that mn/d is divisible by both m and n; hence, for any element $(a, b) \in \mathbb{Z}_m \times \mathbb{Z}_n$,

$$\underbrace{(a, b) + (a, b) + \cdots + (a, b)}_{mn/d \text{ times}} = (0, 0).$$

Therefore, no (a, b) can generate all of $\mathbb{Z}_m \times \mathbb{Z}_n$.

The converse follows directly from Theorem 5.34 sincelcm$(m, n) = mn$ if and only if $\gcd(m, n) = 1$. ■

Corollary 5.39 *Let $n_1, \ldots, n_k$ be positive integers. Then*

$$\prod_{i=1}^{k} \mathbb{Z}_{n_i} \cong \mathbb{Z}_{n_1 \cdots n_k}$$

if and only if $\gcd(n_i, n_j) = 1$ *for* $i \neq j$.

Corollary 5.40 *If*

$$m = p_1^{e_1} \cdots p_k^{e_k},$$

where the p_is are distinct primes, then

$$\mathbb{Z}_m \cong \mathbb{Z}_{p_1^{e_1}} \times \cdots \times \mathbb{Z}_{p_k^{e_k}}.$$

Proof. Since the greatest common divisor of $p_i^{e_i}$ and $p_j^{e_j}$ is 1 for $i \neq j$, the proof follows from Corollary 5.39. ■

Activity 5.6 List all of the elements of $\mathbb{Z}_4 \times \mathbb{Z}_2$. What is the order of each element?

5.3.2 Internal Direct Products

The external direct product of two groups builds a large group out of two smaller groups. We would like to be able to reverse this process and conveniently break down a group into its direct product components; that is, we would like to be able to say when a group is isomorphic to the direct product of two of its subgroups.

Let G be a group with subgroups H and K satisfying the following conditions.

- $G = HK = \{hk : h \in H, k \in K\}$;
- $H \cap K = \{e\}$;
- $hk = kh$ for all $k \in K$ and $h \in H$.

Then G is the **internal direct product** of H and K.

Example 5.41 The group $U(8)$ is the internal direct product of

$$H = \{1, 3\} \quad \text{and} \quad K = \{1, 5\}.$$

□

Example 5.42 The dihedral group D_6 is an internal direct product of its two subgroups

$$H = \{\text{id}, r^3\} \quad \text{and} \quad K = \{\text{id}, r^2, r^4, s, r^2 s, r^4 s\}.$$

It can easily be shown that $K \cong S_3$; consequently, $D_6 \cong \mathbb{Z}_2 \times S_3$. □

Example 5.43 Not every group can be written as the internal direct product of two of its proper subgroups. If the group S_3 were an internal direct product of its proper subgroups H and K, then one of the subgroups, say H, would have to have order 3. In this case H is the subgroup $\{(1), (123), (132)\}$. The subgroup K must have order 2, but no matter which subgroup we choose for K, the condition that $hk = kh$ will never be satisfied for $h \in H$ and $k \in K$. □

Theorem 5.44 *Let G be the internal direct product of subgroups H and K. Then G is isomorphic to $H \times K$.*

Proof. Since G is an internal direct product, we can write any element $g \in G$ as $g = hk$ for some $h \in H$ and some $k \in K$. Define a map $\phi : G \to H \times K$ by $\phi(g) = (h, k)$.

The first problem that we must face is to show that ϕ is a well-defined map; that is, we must show that h and k are uniquely determined by g. Suppose that $g = hk = h'k'$. Then $h^{-1}h' = k(k')^{-1}$ is in both H and K, so it must be the identity. Therefore, $h = h'$ and $k = k'$, which proves that ϕ is, indeed, well-defined.

To show that ϕ preserves the group operation, let $g_1 = h_1 k_1$ and $g_2 = h_2 k_2$ and observe that

$$\begin{aligned} \phi(g_1 g_2) &= \phi(h_1 k_1 h_2 k_2) \\ &= \phi(h_1 h_2 k_1 k_2) \\ &= (h_1 h_2, k_1 k_2) \\ &= (h_1, k_1)(h_2, k_2) \\ &= \phi(g_1)\phi(g_2). \end{aligned}$$

We will leave the proof that ϕ is one-to-one and onto as an exercise. ■

Example 5.45 The group $\mathbb{Z}_6$ is an internal direct product isomorphic to $\{0, 2, 4\} \times \{0, 3\}$. □

We can extend the definition of an internal direct product of G to a collection of subgroups $H_1, H_2, \ldots, H_n$ of G, by requiring that

- $G = H_1 H_2 \cdots H_n = \{h_1 h_2 \cdots h_n : h_i \in H_i\}$;
- $H_i \cap \langle \cup_{j \neq i} H_j \rangle = \{e\}$;
- $h_i h_j = h_j h_i$ for all$h_i \in H_i$ and $h_j \in H_j$.

We will leave the proof of the following theorem as an exercise.

Theorem 5.46 *Let G be the internal direct product of subgroups H_i, where $i = 1, 2, \ldots, n$. Then G is isomorphic to $\prod_i H_i$.*

5.3.3 Reading Questions

1. If G is a group with order 4 and H is a group with order 6, what is the order of the group $G \times H$? Briefly explain.

2. True or false: if G and H are cyclic groups, then the group $G \times H$ is also cyclic. Briefly explain why or give a counterexample.

3. What is the difference between an external direct product and an internal direct product? Explain in your own words how these are different and how these are the same.

5.3.4 Exercises

1. Find the order of each of the following elements.

(a) $(3, 4)$ in $\mathbb{Z}_4 \times \mathbb{Z}_6$

(b) $(6, 15, 4)$ in $\mathbb{Z}_{30} \times \mathbb{Z}_{45} \times \mathbb{Z}_{24}$

(c) $(5, 10, 15)$ in $\mathbb{Z}_{25} \times \mathbb{Z}_{25} \times \mathbb{Z}_{25}$

(d) $(8, 8, 8)$ in $\mathbb{Z}_{10} \times \mathbb{Z}_{24} \times \mathbb{Z}_{80}$

Hint. (a) 12; (c) 5.

2. Prove that D_4 cannot be the internal direct product of two of its proper subgroups.

3. Prove that the subgroup of $\mathbb{Q}^*$ consisting of elements of the form $2^m 3^n$ for $m, n \in \mathbb{Z}$ is an internal direct product isomorphic to $\mathbb{Z} \times \mathbb{Z}$.

4. Prove or disprove the following assertion. Let G, H, and K be groups. If $G \times K \cong H \times K$, then $G \cong H$.

5. Prove or disprove: There is a noncyclic abelian group of order 51.

6. Prove or disprove: There is a noncyclic abelian group of order 52.

Hint. True.

7. Prove that $A \times B$ is abelian if and only if A and B are abelian.

8. If G is the internal direct product of $H_1, H_2, \ldots, H_n$, prove that G is isomorphic to $\prod_i H_i$.

9. Let H_1 and H_2 be subgroups of G_1 and G_2, respectively. Prove that $H_1 \times H_2$ is a subgroup of $G_1 \times G_2$.

10. Let G be a group of order 20. If G has subgroups H and K of orders 4 and 5 respectively such that $hk = kh$ for all $h \in H$ and $k \in K$, prove that G is the internal direct product of H and K.

11. If $G \cong \overline{G}$ and $H \cong \overline{H}$, show that $G \times H \cong \overline{G} \times \overline{H}$.

12. Prove that $G \times H$ is isomorphic to $H \times G$.

5.4 Summary and Additional Exercises

5.4.1 The Important Ideas

- For $a \in G$, we call $\langle a \rangle$ the **cyclic subgroup** generated by a. If G contains some element a such that $G = \langle a \rangle$, then G is a **cyclic group**. In this case a is a **generator** of G. If a is an element of a group G, we define the **order** of a to be the smallest positive integer n such that $a^n = e$, and we write $|a| = n$. If there is no such integer n, we say that the order of a is infinite and write $|a| = \infty$ to denote the order of a.

- Let G be a group and a be any element in G. Then the set

$$\langle a \rangle = \{a^k : k \in \mathbb{Z}\}$$

is a subgroup of G. Furthermore, $\langle a \rangle$ is the smallest subgroup of G that contains a.

- Every cyclic group is abelian.

- Every subgroup of a cyclic group is cyclic. The subgroups of $\mathbb{Z}$ are exactly $n\mathbb{Z}$ for $n = 0, 1, 2, \ldots$.

- Let G be a cyclic group of order n and suppose that $a \in G$ is a generator of the group. If $b = a^k$, then the order of b is n/d, where $d = \gcd(k, n)$.

- The generators of $\mathbb{Z}_n$ are the integers r such that $1 \leq r < n$ and $\gcd(r, n) = 1$.

- Nonzero complex numbers can also be represented using **polar coordinates**. To specify any nonzero point on the plane, it suffices to give an angle θ from the positive x axis in the counterclockwise direction and a distance r from the origin. In this case

  $$z = a + bi = r(\cos\theta + i\sin\theta).$$

  Hence,

  $$r = |z| = \sqrt{a^2 + b^2}$$

  and

  $$\begin{aligned} a &= r\cos\theta \\ b &= r\sin\theta. \end{aligned}$$

  We sometimes abbreviate $r(\cos\theta + i\sin\theta)$ as $r \operatorname{cis}\theta$. If $z = r\operatorname{cis}\theta$ and $w = s\operatorname{cis}\phi$, then

  $$zw = rs\operatorname{cis}(\theta + \phi).$$

- DeMoivre's Theorem: Let $z = r\operatorname{cis}\theta$ be a nonzero complex number. Then

  $$[r\operatorname{cis}\theta]^n = r^n\operatorname{cis}(n\theta)$$

  for $n = 1, 2, \ldots$.

- The **circle group**,

  $$\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$$

  is a subgroup of $\mathbb{C}^*$. The complex numbers satisfying the equation $z^n = 1$ are called the **nth roots of unity**, which form a cyclic subgroup of $\mathbb{T}$ of order n. A generator for the group of the nth roots of unity is called a **primitive nth root of unity**.

- If G and H be groups, the set $G \times H$ is a group under the operation $(g_1, h_1)(g_2, h_2) = (g_1g_2, h_1h_2)$ where $g_1, g_2 \in G$ and $h_1, h_2 \in H$. The group $G \times H$ is called the **external direct product** of G and H.

- The group $\mathbb{Z}_m \times \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_{mn}$ if and only if $\gcd(m, n) = 1$.

- Let G be a group with subgroups H and K satisfying the following conditions.

  - $G = HK = \{hk : h \in H, k \in K\}$;
  - $H \cap K = \{e\}$;
  - $hk = kh$ for all $k \in K$ and $h \in H$.

  Then G is the **internal direct product** of H and K. If G is the internal direct product of subgroups H and K, then G is isomorphic to $H \times K$.

5.4.2 Additional Exercises

1. Let $a, b \in G$. Prove the following statements.

(a) The order of a is the same as the order of a^{-1}.

(b) For all $g \in G$, $|a| = |g^{-1}ag|$.

(c) The order of ab is the same as the order of ba.

2. Let p and q be distinct primes. How many generators does $\mathbb{Z}_{pq}$ have?

3. Let p be prime and r be a positive integer. How many generators does $\mathbb{Z}_{p^r}$ have?

4. Prove or disprove: Every abelian group of order divisible by 3 contains a subgroup of order 3.

5. Let a be an element in a group G. What is a generator for the subgroup $\langle a^m \rangle \cap \langle a^n \rangle$?

6. Prove that $\mathbb{Z}_n$ has an even number of generators for $n > 2$.

7. Suppose that G is a group and let a, $b \in G$. Prove that if $|a| = m$ and $|b| = n$ with $\gcd(m, n) = 1$, then $\langle a \rangle \cap \langle b \rangle = \{e\}$.

8. 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.

Hint. The identity element in any group has finite order. Let $g, h \in G$ have orders m and n, respectively. Since $(g^{-1})^m = e$ and $(gh)^{mn} = e$, the elements of finite order in G form a subgroup of G.

9. Let G be a finite cyclic group of order n generated by x. Show that if $y = x^k$ where $\gcd(k, n) = 1$, then y must be a generator of G.

10. If G is an abelian group that contains a pair of cyclic subgroups of order 2, show that G must contain a subgroup of order 4. Does this subgroup have to be cyclic?

11. Let G be an abelian group of order pq where $\gcd(p, q) = 1$. If G contains elements a and b of order p and q respectively, then show that G is cyclic.

12. Prove that the subgroups of $\mathbb{Z}$ are exactly $n\mathbb{Z}$ for $n = 0, 1, 2, \ldots$.

13. Prove that the generators of $\mathbb{Z}_n$ are the integers r such that $1 \leq r < n$ and $\gcd(r, n) = 1$.

14. Prove that if G has no proper nontrivial subgroups, then G is a cyclic group.

Hint. If g is an element distinct from the identity in G, g must generate G; otherwise, $\langle g \rangle$ is a nontrivial proper subgroup of G.

15. Prove that the order of an element in a cyclic group G must divide the order of the group.

16. Prove that if G is a cyclic group of order m and $d \mid m$, then G must have a subgroup of order d.

17. For what integers n is -1 an nth root of unity?

18. Let $\alpha \in \mathbb{T}$. Prove that $\alpha^m = 1$ and $\alpha^n = 1$ if and only if $\alpha^d = 1$ for $d = \gcd(m, n)$.

19. Let $z \in \mathbb{C}^*$. If $|z| \neq 1$, prove that the order of z is infinite.

20. Let $z = \cos\theta + i \sin\theta$ be in $\mathbb{T}$ where $\theta \in \mathbb{Q}$. Prove that the order of z is infinite.

21. Prove $U(5) \cong \mathbb{Z}_4$. Can you generalize this result for $U(p)$, where p is prime?

22. Let G be the internal direct product of subgroups H and K. Show that the map $\phi : G \to H \times K$ defined by $\phi(g) = (h, k)$ for $g = hk$, where $h \in H$ and $k \in K$, is one-to-one and onto.

Hint. To show that ϕ is one-to-one, let $g_1 = h_1 k_1$ and $g_2 = h_2 k_2$ and consider $\phi(g_1) = \phi(g_2)$.

23. Let $n_1, \ldots, n_k$ be positive integers. Show that

$$\prod_{i=1}^{k} \mathbb{Z}_{n_i} \cong \mathbb{Z}_{n_1 \cdots n_k}$$

if and only if $\gcd(n_i, n_j) = 1$ for $i \neq j$.

24. Let $m, n \in \mathbb{Z}$. Prove that $\langle m, n \rangle = \langle d \rangle$ if and only if $d = \gcd(m, n)$.

25. Let $m, n \in \mathbb{Z}$. Prove that $\langle m \rangle \cap \langle n \rangle = \langle l \rangle$ if and only if $l = \operatorname{lcm}(m, n)$.

5.5 Connections to the Secondary Classroom—Modular Arithmetic

This appendix will connect cyclic groups and modular arithmetic to the high school classroom.

Chapter 6

Permutation Groups

Objectives

- To understand and be able to apply the definition of a permutation group.
- To understand and be proficient in cycle notation.
- To understand and be able to apply the definition of a dihedral group.

Permutation groups are central to the study of geometric symmetries and to Galois theory, the study of finding solutions of polynomial equations. They also provide abundant examples of nonabelian groups.

Let us recall for a moment the symmetries of the equilateral triangle $\triangle ABC$ from Chapter 3. The symmetries actually consist of permutations of the three vertices, where a **permutation** of the set $S = \{A, B, C\}$ is a one-to-one and onto map $\pi : S \to S$. The three vertices have the following six permutations,

$$\begin{pmatrix} A & B & C \\ A & B & C \end{pmatrix} \qquad \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix} \qquad \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}$$
$$\begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} \qquad \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix} \qquad \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix}.$$

We have used the array

$$\begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}$$

to denote the permutation that sends A to B, B to C, and C to A. That is,

$$\begin{aligned} A &\mapsto B \\ B &\mapsto C \\ C &\mapsto A. \end{aligned}$$

The symmetries of a triangle form a group. In this chapter we will study groups of this type.

6.1 Definitions and Notation

In general, the permutations of a set X form a group S_X. If X is a finite set, we can assume $X = \{1, 2, \ldots, n\}$. In this case we write S_n instead of S_X. The following theorem says that S_n is a group. We call this group the **symmetric group** on n letters.

Theorem 6.1 *The symmetric group on n letters, S_n, is a group with $n!$ elements, where the binary operation is the composition of maps.*

Proof. The identity of S_n is just the identity map that sends 1 to 1, 2 to 2, ..., n to n. If $f : S_n \to S_n$ is a permutation, then f^{-1} exists, since f is one-to-one and onto; hence, every permutation has an inverse. Composition of maps is associative, which makes the group operation associative. We leave the proof that $|S_n| = n!$ as an exercise. ■

A subgroup of S_n is called a **permutation group**.

Example 6.2 Consider the subgroup G of S_5 consisting of the identity permutation id and the permutations

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

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

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

The following table tells us how to multiply elements in the permutation group G.

$\circ$	id	σ	τ	μ
id	id	σ	τ	μ
σ	σ	id	μ	τ
τ	τ	μ	id	σ
μ	μ	τ	σ	id

□

Remark 6.3 Though it is natural to multiply elements in a group from left to right, functions are composed from right to left. Let σ and τ be permutations on a set X. To compose σ and τ as functions, we calculate $(\sigma \circ \tau)(x) = \sigma(\tau(x))$. That is, we do τ first, then σ. There are several ways to approach this inconsistency. *We will adopt the convention of multiplying permutations right to left. To compute $\sigma\tau$, do τ first and then σ.* That is, by $\sigma\tau(x)$ we mean $\sigma(\tau(x))$. (Another way of solving this problem would be to write functions on the right; that is, instead of writing $\sigma(x)$, we could write $(x)\sigma$. We could also multiply permutations left to right to agree with the usual way of multiplying elements in a group. Certainly all of these methods have been used.

Example 6.4 Permutation multiplication is not usually commutative. Let

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

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

Then

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

but

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

□

6.1.1 Cycle Notation

The notation that we have used to represent permutations up to this point is cumbersome, to say the least. To work effectively with permutation groups, we need a more streamlined method of writing down and manipulating permutations.

A permutation $\sigma \in S_X$ is a **cycle of length k** if there exist elements $a_1, a_2, \ldots, a_k \in X$ such that

$$\begin{aligned} \sigma(a_1) &= a_2 \\ \sigma(a_2) &= a_3 \\ &\vdots \\ \sigma(a_k) &= a_1 \end{aligned}$$

and $\sigma(x) = x$ for all other elements $x \in X$. We will write $(a_1, a_2, \ldots, a_k)$ to denote the cycle σ. Cycles are the building blocks of all permutations.

Example 6.5 The permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 3 & 5 & 1 & 4 & 2 & 7 \end{pmatrix} = (162354)$$

is a cycle of length 6, whereas

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 2 & 3 & 5 & 6 \end{pmatrix} = (243)$$

is a cycle of length 3.

Not every permutation is a cycle. Consider the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 1 & 3 & 6 & 5 \end{pmatrix} = (1243)(56).$$

This permutation actually contains a cycle of length 2 and a cycle of length 4. □

Example 6.6 It is very easy to compute products of cycles. Suppose that

$$\sigma = (1352) \quad \text{and} \quad \tau = (256).$$

If we think of σ as

$$1 \mapsto 3, \qquad 3 \mapsto 5, \qquad 5 \mapsto 2, \qquad 2 \mapsto 1,$$

and τ as

$$2 \mapsto 5, \qquad 5 \mapsto 6, \qquad 6 \mapsto 2,$$

then for $\sigma\tau$ remembering that we apply τ first and then σ, it must be the case that

$$1 \mapsto 3, \qquad 3 \mapsto 5, \qquad 5 \mapsto 6, \qquad 6 \mapsto 2 \mapsto 1,$$

or $\sigma\tau = (1356)$. If $\mu = (1634)$, then $\sigma\mu = (1652)(34)$. □

Two cycles in S_X, $\sigma = (a_1, a_2, \ldots, a_k)$ and $\tau = (b_1, b_2, \ldots, b_l)$, are **disjoint** if $a_i \neq b_j$ for all i and j.

Example 6.7 The cycles (135) and (27) are disjoint; however, the cycles (135) and (347) are not. Calculating their products, we find that

$$(135)(27) = (135)(27)$$

$$(135)(347) = (13475).$$

The product of two cycles that are not disjoint may reduce to something less complicated; the product of disjoint cycles cannot be simplified. □

Proposition 6.8 *Let σ and τ be two disjoint cycles in S_X. Then $\sigma\tau = \tau\sigma$.*

Proof. Let $\sigma = (a_1, a_2, \ldots, a_k)$ and $\tau = (b_1, b_2, \ldots, b_l)$. We must show that $\sigma\tau(x) = \tau\sigma(x)$ for all $x \in X$. If x is neither in $\{a_1, a_2, \ldots, a_k\}$ nor $\{b_1, b_2, \ldots, b_l\}$, then both σ and τ fix x. That is, $\sigma(x) = x$ and $\tau(x) = x$. Hence,

$$\sigma\tau(x) = \sigma(\tau(x)) = \sigma(x) = x = \tau(x) = \tau(\sigma(x)) = \tau\sigma(x).$$

Do not forget that we are multiplying permutations right to left, which is the opposite of the order in which we usually multiply group elements. Now suppose that $x \in \{a_1, a_2, \ldots, a_k\}$. Then $\sigma(a_i) = a_{(i \bmod k)+1}$; that is,

$$\begin{aligned} a_1 &\mapsto a_2 \\ a_2 &\mapsto a_3 \\ &\vdots \\ a_{k-1} &\mapsto a_k \\ a_k &\mapsto a_1. \end{aligned}$$

However, $\tau(a_i) = a_i$ since σ and τ are disjoint. Therefore,

$$\begin{aligned} \sigma\tau(a_i) &= \sigma(\tau(a_i)) \\ &= \sigma(a_i) \\ &= a_{(i \bmod k)+1} \\ &= \tau(a_{(i \bmod k)+1}) \\ &= \tau(\sigma(a_i)) \\ &= \tau\sigma(a_i). \end{aligned}$$

Similarly, if $x \in \{b_1, b_2, \ldots, b_l\}$, then σ and τ also commute. ■

Theorem 6.9 *Every permutation in S_n can be written as the product of disjoint cycles.*

Proof. We can assume that $X = \{1, 2, \ldots, n\}$. If $\sigma \in S_n$ and we define X_1 to be $\{\sigma(1), \sigma^2(1), \ldots\}$, then the set X_1 is finite since X is finite. Now let i be the first integer in X that is not in X_1 and define X_2 by $\{\sigma(i), \sigma^2(i), \ldots\}$. Again, X_2 is a finite set. Continuing in this manner, we can define finite disjoint sets $X_3, X_4, \ldots$. Since X is a finite set, we are guaranteed that this process will end and there will be only a finite number of these sets, say r. If σ_i is the cycle defined by

$$\sigma_i(x) = \begin{cases} \sigma(x) & x \in X_i \\ x & x \notin X_i, \end{cases}$$

then $\sigma = \sigma_1 \sigma_2 \cdots \sigma_r$. Since the sets $X_1, X_2, \ldots, X_r$ are disjoint, the cycles $\sigma_1, \sigma_2, \ldots, \sigma_r$ must also be disjoint. ■

Example 6.10 Let

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

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

Using cycle notation, we can write

$$\begin{aligned} \sigma &= (1624) \\ \tau &= (13)(456) \\ \sigma\tau &= (136)(245) \\ \tau\sigma &= (143)(256). \end{aligned}$$

□

Remark 6.11 From this point forward we will find it convenient to use cycle notation to represent permutations. When using cycle notation, we often denote the identity permutation by (1).

Activity 6.1 Compute each of the following.

1. (1345)(234)
2. (143)(23)(24)
3. (1254)(13)(25)
4. $(1254)^{-1}(123)(45)(1254)$
5. $(12)^{-1}$
6. $(12537)^{-1}$

6.1.2 Transpositions

The simplest permutation is a cycle of length 2. Such cycles are called **transpositions**. Since

$$(a_1, a_2, \ldots, a_n) = (a_1 a_n)(a_1 a_{n-1}) \cdots (a_1 a_3)(a_1 a_2),$$

any cycle can be written as the product of transpositions, leading to the following proposition.

Proposition 6.12 *Any permutation of a finite set containing at least two elements can be written as the product of transpositions.*

Example 6.13 Consider the permutation

$$(16)(253) = (16)(23)(25) = (16)(45)(23)(45)(25).$$

As we can see, there is no unique way to represent permutation as the product of transpositions. For instance, we can write the identity permutation as (12)(12), as (13)(24)(13)(24), and in many other ways. However, as it turns out, no permutation can be written as the product of both an even number of transpositions and an odd number of transpositions. For instance, we could represent the permutation (16) by

$$(23)(16)(23)$$

or by

$$(35)(16)(13)(16)(13)(35)(56),$$

but (16) will always be the product of an odd number of transpositions. □

Lemma 6.14 *If the identity is written as the product of r transpositions,*

$$\mathrm{id} = \tau_1 \tau_2 \cdots \tau_r,$$

then r is an even number.

Proof. We will employ induction on r. A transposition cannot be the identity; hence, $r > 1$. If $r = 2$, then we are done. Suppose that $r > 2$. In this case the product of the last two transpositions, $\tau_{r-1}\tau_r$, must be one of the following cases:

$$\begin{aligned}
(ab)(ab) &= \mathrm{id} \\
(bc)(ab) &= (ac)(bc) \\
(cd)(ab) &= (ab)(cd) \\
(ac)(ab) &= (ab)(bc),
\end{aligned}$$

where a, b, c, and d are distinct.

The first equation simply says that a transposition is its own inverse. If this case occurs, delete $\tau_{r-1}\tau_r$ from the product to obtain

$$\mathrm{id} = \tau_1\tau_2 \cdots \tau_{r-3}\tau_{r-2}.$$

By induction $r - 2$ is even; hence, r must be even.

In each of the other three cases, we can replace $\tau_{r-1}\tau_r$ with the right-hand side of the corresponding equation to obtain a new product of r transpositions for the identity. In this new product the last occurrence of a will be in the next-to-the-last transposition. We can continue this process with $\tau_{r-2}\tau_{r-1}$ to obtain either a product of $r - 2$ transpositions or a new product of r transpositions where the last occurrence of a is in τ_{r-2}. If the identity is the product of $r - 2$ transpositions, then again we are done, by our induction hypothesis; otherwise, we will repeat the procedure with $\tau_{r-3}\tau_{r-2}$.

At some point either we will have two adjacent, identical transpositions canceling each other out or a will be shuffled so that it will appear only in the first transposition. However, the latter case cannot occur, because the identity would not fix a in this instance. Therefore, the identity permutation must be the product of $r - 2$ transpositions and, again by our induction hypothesis, we are done. ■

Theorem 6.15 *If a permutation σ can be expressed as the product of an even number of transpositions, then any other product of transpositions equaling σ must also contain an even number of transpositions. Similarly, if σ can be expressed as the product of an odd number of transpositions, then any other product of transpositions equaling σ must also contain an odd number of transpositions.*

Proof. Suppose that

$$\sigma = \sigma_1\sigma_2 \cdots \sigma_m = \tau_1\tau_2 \cdots \tau_n,$$

where m is even. We must show that n is also an even number. The inverse of σ is $\sigma_m \cdots \sigma_1$. Since

$$\mathrm{id} = \sigma\sigma_m \cdots \sigma_1 = \tau_1 \cdots \tau_n\sigma_m \cdots \sigma_1,$$

n must be even by Lemma 6.14. The proof for the case in which σ can be expressed as an odd number of transpositions is left as an exercise. ■

In light of Theorem 6.15, we define a permutation to be **even** if it can be expressed as an even number of transpositions and **odd** if it can be expressed as an odd number of transpositions.

6.1.3 The Alternating Groups

One of the most important subgroups of S_n is the set of all even permutations, A_n. The group A_n is called the **alternating group on n letters**.

Theorem 6.16 *The set A_n is a subgroup of S_n.*
Proof. Since the product of two even permutations must also be an even permutation, A_n is closed. The identity is an even permutation and therefore is in A_n. If σ is an even permutation, then

$$\sigma = \sigma_1 \sigma_2 \cdots \sigma_r,$$

where σ_i is a transposition and r is even. Since the inverse of any transposition is itself,

$$\sigma^{-1} = \sigma_r \sigma_{r-1} \cdots \sigma_1$$

is also in A_n. ■

Proposition 6.17 *The number of even permutations in S_n, $n \geq 2$, is equal to the number of odd permutations; hence, the order of A_n is $n!/2$.*
Proof. Let A_n be the set of even permutations in S_n and B_n be the set of odd permutations. If we can show that there is a bijection between these sets, they must contain the same number of elements. Fix a transposition σ in S_n. Since $n \geq 2$, such a σ exists. Define

$$\lambda_\sigma : A_n \to B_n$$

by

$$\lambda_\sigma(\tau) = \sigma\tau.$$

Suppose that $\lambda_\sigma(\tau) = \lambda_\sigma(\mu)$. Then $\sigma\tau = \sigma\mu$ and so

$$\tau = \sigma^{-1}\sigma\tau = \sigma^{-1}\sigma\mu = \mu.$$

Therefore, λ_σ is one-to-one. We will leave the proof that λ_σ is surjective to the reader. ■

Example 6.18 The group A_4 is the subgroup of S_4 consisting of even permutations. There are twelve elements in A_4:

$$\begin{array}{llll} (1) & (12)(34) & (13)(24) & (14)(23) \\ (123) & (132) & (124) & (142) \\ (134) & (143) & (234) & (243). \end{array}$$

One of the end-of-chapter exercises will be to write down all the subgroups of A_4. You will find that there is no subgroup of order 6. Does this surprise you? □

6.1.4 Cayley's Theorem

Cayley proved that every group G is isomorphic to a group of permutations on some set; hence, every group is a permutation group. Cayley's Theorem is what we call a representation theorem. The aim of representation theory is to find an isomorphism of some group G that we wish to study into a group that we know a great deal about, such as a group of permutations or matrices.

Example 6.19 Consider the group $\mathbb{Z}_3$. The Cayley table for $\mathbb{Z}_3$ is as follows.

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

The addition table of $\mathbb{Z}_3$ suggests that it is the same as the permutation group $G = \{(0), (012), (021)\}$. The isomorphism here is

$$\begin{aligned} 0 &\mapsto \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix} = (0) \\ 1 &\mapsto \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \end{pmatrix} = (012) \\ 2 &\mapsto \begin{pmatrix} 0 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix} = (021). \end{aligned}$$

□

Theorem 6.20 Cayley. *Every group is isomorphic to a group of permutations.*

Proof. Let G be a group. We must find a group of permutations $\overline{G}$ that is isomorphic to G. For any $g \in G$, define a function $\lambda_g : G \to G$ by $\lambda_g(a) = ga$. We claim that λ_g is a permutation of G. To show that λ_g is one-to-one, suppose that $\lambda_g(a) = \lambda_g(b)$. Then

$$ga = \lambda_g(a) = \lambda_g(b) = gb.$$

Hence, $a = b$. To show that λ_g is onto, we must prove that for each $a \in G$, there is a b such that $\lambda_g(b) = a$. Let $b = g^{-1}a$.

Now we are ready to define our group $\overline{G}$. Let

$$\overline{G} = \{\lambda_g : g \in G\}.$$

We must show that $\overline{G}$ is a group under composition of functions and find an isomorphism between G and $\overline{G}$. We have closure under composition of functions since

$$(\lambda_g \circ \lambda_h)(a) = \lambda_g(ha) = gha = \lambda_{gh}(a).$$

Also,

$$\lambda_e(a) = ea = a$$

and

$$(\lambda_{g^{-1}} \circ \lambda_g)(a) = \lambda_{g^{-1}}(ga) = g^{-1}ga = a = \lambda_e(a).$$

We can define an isomorphism from G to $\overline{G}$ by $\phi : g \mapsto \lambda_g$. The group operation is preserved since

$$\phi(gh) = \lambda_{gh} = \lambda_g \lambda_h = \phi(g)\phi(h).$$

It is also one-to-one, because if $\phi(g)(a) = \phi(h)(a)$, then

$$ga = \lambda_g a = \lambda_h a = ha.$$

Hence, $g = h$. That ϕ is onto follows from the fact that $\phi(g) = \lambda_g$ for any $\lambda_g \in \overline{G}$. ■

The isomorphism $g \mapsto \lambda_g$ is known as the **left regular representation** of G.

Activity 6.2 Write out the permutations associated with each element of S_3 in the proof of Cayley's Theorem.

6.1.5 Historical Note

Arthur Cayley was born in England in 1821, though he spent much of the first part of his life in Russia, where his father was a merchant. Cayley was educated at Cambridge, where he took the first Smith's Prize in mathematics. A lawyer for much of his adult life, he wrote several papers in his early twenties before entering the legal profession at the age of 25. While practicing law he continued his mathematical research, writing more than 300 papers during this period of his life. These included some of his best work. In 1863 he left law to become a professor at Cambridge. Cayley wrote more than 900 papers in fields such as group theory, geometry, and linear algebra. His legal knowledge was very valuable to Cambridge; he participated in the writing of many of the university's statutes. Cayley was also one of the people responsible for the admission of women to Cambridge.

6.1.6 Reading Questions

1. What is the order of the group S_4? What is the order of the group A_4? Then, give an example of an element in S_4 that is not in A_4.

2. Write the permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 2 & 7 & 5 & 6 & 4 & 1 \end{pmatrix}$ in cycle notation.

3. Write $\sigma = (142)(243)$ as a single cycle or product of disjoint cycles. Then, find $\sigma(2)$.

6.1.7 Exercises

1. Write the following permutations in cycle notation.

(a)
$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 5 & 3 \end{pmatrix}$$

(b)
$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 5 & 1 & 3 \end{pmatrix}$$

(c)
$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 4 & 2 \end{pmatrix}$$

(d)
$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 2 & 5 \end{pmatrix}$$

Hint. (a) (12453); (c) (13)(25).

2. Compute each of the following.

(a) (12)(1253)

(b) (1423)(34)(56)(1324)

(c) $(1254)(13)(25)^2$

(d) $(1254)^2(123)(45)$

(e) $(123)(45)(1254)^{-2}$

(f) $(1254)^{100}$

(g) $|(1254)|$

(h) $|(1254)^2|$

(i) $[(12)(34)(12)(47)]^{-1}$

(j) $[(1235)(467)]^{-1}$

Hint. (a) (135)(24); (c) (14)(23); (e) (1324); (g) (134)(25); (n) (17352).

3. Express the following permutations as products of transpositions and identify them as even or odd.

(a) (14356)

(b) (156)(234)

(c) (1426)(142)

(d) (17254)(1423)(154632)

(e) (142637)

Hint. (a) (16)(15)(13)(14); (c) (16)(14)(12).

4. Find $(a_1, a_2, \ldots, a_n)^{-1}$.

Hint. $(a_1, a_2, \ldots, a_n)^{-1} = (a_1, a_n, a_{n-1}, \ldots, a_2)$

5. List all of the subgroups of S_4. Find each of the following sets:

(a) $\{\sigma \in S_4 : \sigma(1) = 3\}$

(b) $\{\sigma \in S_4 : \sigma(2) = 2\}$

(c) $\{\sigma \in S_4 : \sigma(1) = 3 \text{ and } \sigma(2) = 2\}$.

Are any of these sets subgroups of S_4?

Hint. (a) $\{(13), (13)(24), (132), (134), (1324), (1342)\}$ is not a subgroup.

6. Find all of the subgroups in A_4. What is the order of each subgroup?

7. Find all possible orders of elements in S_7 and A_7.

8. Show that A_{10} contains an element of order 15.

Hint. $(12345)(678)$.

9. Does A_8 contain an element of order 26?

10. Find an element of largest order in S_n for $n = 3, \ldots, 10$.

11. What are the possible cycle structures of elements of A_5? What about A_6?

Hint. Permutations of the form

$$(1), (a_1, a_2)(a_3, a_4), (a_1, a_2, a_3), (a_1, a_2, a_3, a_4, a_5)$$

are possible for A_5.

12. Let $\sigma \in S_n$ have order n. Show that for all integers i and j, $\sigma^i = \sigma^j$ if and only if $i \equiv j \pmod{n}$.

13. Prove that S_n is nonabelian for $n \geq 3$.

Hint. Calculate $(123)(12)$ and $(12)(123)$.

14. Show that A_n is nonabelian for $n \geq 4$.

6.2 Dihedral Groups

Another special type of permutation group is the dihedral group. Recall the symmetry group of an equilateral triangle in Chapter 3. Such groups consist of the rigid motions of a regular n-sided polygon or n-gon. For $n = 3, 4, \ldots$, we define the **nth dihedral group** to be the group of rigid motions of a regular n-gon. We will denote this group by D_n. We can number the vertices of a regular n-gon by $1, 2, \ldots, n$ (Figure 6.21). Notice that there are exactly n choices to replace the first vertex. If we replace the first vertex by k, then the second vertex must be replaced either by vertex $k+1$ or by vertex $k-1$; hence, there are $2n$ possible rigid motions of the n-gon. We summarize these results in the following theorem.

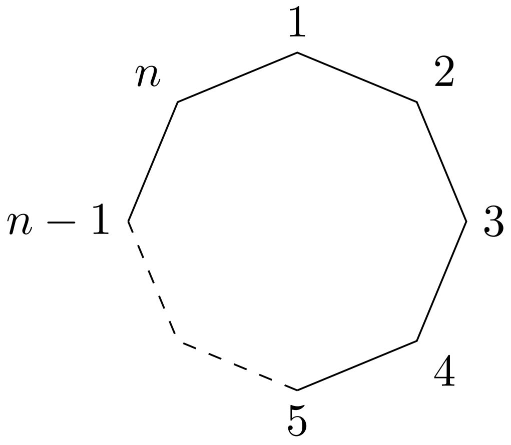


Figure 6.21 A regular n-gon

Theorem 6.22 *The dihedral group, D_n, is a subgroup of S_n of order $2n$.*

Theorem 6.23 *The group D_n, $n \geq 3$, consists of all products of the two elements r and s, satisfying the relations*

$$\begin{aligned} r^n &= 1 \\ s^2 &= 1 \\ srs &= r^{-1}. \end{aligned}$$

Proof. The possible motions of a regular n-gon are either reflections or rotations (Figure 6.24). There are exactly n possible rotations:

$$\mathrm{id}, \frac{360^\circ}{n}, 2 \cdot \frac{360^\circ}{n}, \ldots, (n-1) \cdot \frac{360^\circ}{n}.$$

We will denote the rotation $360^\circ/n$ by r. The rotation r generates all of the other rotations. That is,

$$r^k = k \cdot \frac{360^\circ}{n}.$$

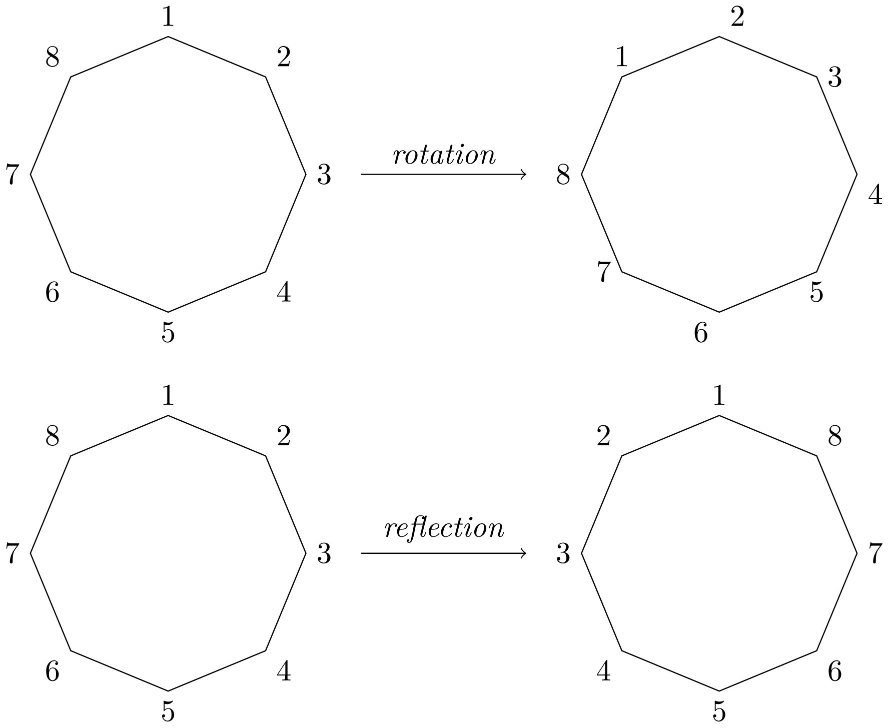


Figure 6.24 Rotations and reflections of a regular n-gon

Label the n reflections $s_1, s_2, \ldots, s_n$, where s_k is the reflection that leaves vertex k fixed. There are two cases of reflections, depending on whether n is even or odd. If there are an even number of vertices, then two vertices are left fixed by a reflection, and $s_1 = s_{n/2+1}, s_2 = s_{n/2+2}, \ldots, s_{n/2} = s_n$. If there are an odd number of vertices, then only a single vertex is left fixed by a reflection and $s_1, s_2, \ldots, s_n$ are distinct (Figure 6.25). In either case, the order of each s_k is two. Let $s = s_1$. Then $s^2 = 1$ and $r^n = 1$. Since any rigid motion t of the n-gon replaces the first vertex by the vertex k, the second vertex must be replaced by either $k+1$ or by $k-1$. If the second vertex is replaced by $k+1$, then $t = r^k$. If the second vertex is replaced by $k-1$, then $t = sr^k$. Hence, r and s generate D_n. That is, D_n consists of all finite products of r and s,

$$D_n = \{1, r, r^2, \ldots, r^{n-1}, s, sr, sr^2, \ldots, sr^{n-1}\}.$$

We will leave the proof that $srs = r^{-1}$ as an exercise.

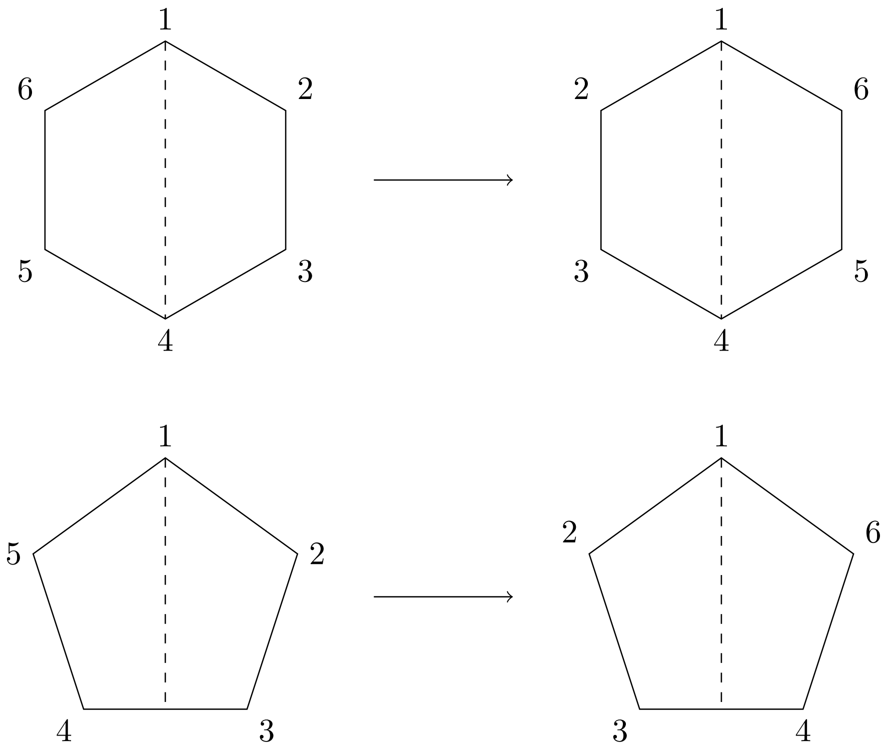


Figure 6.25 Types of reflections of a regular n-gon

■

Example 6.26 The group of rigid motions of a square, D_4, consists of eight elements. With the vertices numbered 1, 2, 3, 4 (Figure 6.27), the rotations are

$$\begin{aligned} r &= (1234) \\ r^2 &= (13)(24) \\ r^3 &= (1432) \\ r^4 &= (1) \end{aligned}$$

and the reflections are

$$\begin{aligned} s_1 &= (24) \\ s_2 &= (13). \end{aligned}$$

The order of D_4 is 8. The remaining two elements are

$$\begin{aligned} rs_1 &= (12)(34) \\ r^3 s_1 &= (14)(23). \end{aligned}$$

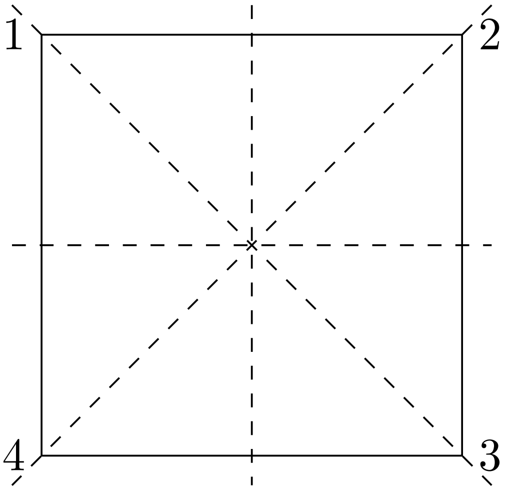


Figure 6.27 The group D_4

□

Activity 6.3 Using cycle notation, list the elements in D_5. What are r and s? Write every element as a product of r and s.

6.2.1 The Motion Group of a Cube

We can investigate the groups of rigid motions of geometric objects other than a regular n-sided polygon to obtain interesting examples of permutation groups. Let us consider the group of rigid motions of a cube. One of the first questions that we can ask about this group is "what is its order?" A cube has 6 sides. If a particular side is facing upward, then there are four possible rotations of the cube that will preserve the upward-facing side. Hence, the order of the group is $6 \cdot 4 = 24$. We have just proved the following proposition.

Proposition 6.28 *The group of rigid motions of a cube contains* 24 *elements.*

Theorem 6.29 *The group of rigid motions of a cube is* S_4.

Proof. From Proposition 6.28, we already know that the motion group of the cube has 24 elements, the same number of elements as there are in S_4. There are exactly four diagonals in the cube. If we label these diagonals 1, 2, 3, and 4, we must show that the motion group of the cube will give us any permutation of the diagonals (Figure 6.30). If we can obtain all of these permutations, then S_4 and the group of rigid motions of the cube must be the same. To obtain a transposition we can rotate the cube 180° about the axis joining the midpoints of opposite edges (Figure 6.31). There are six such axes, giving all transpositions in S_4. Since every element in S_4 is the product of a finite number of transpositions, the motion group of a cube must be S_4.

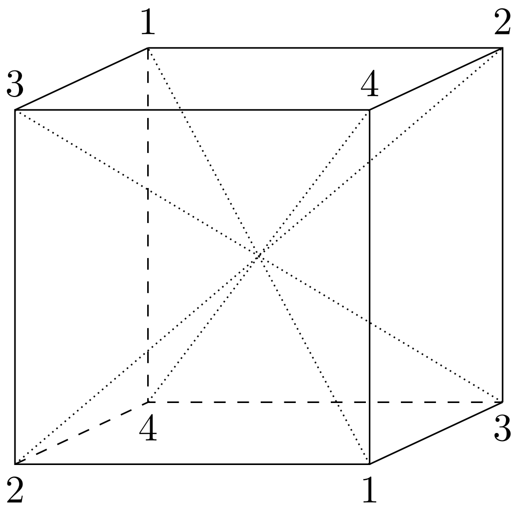


Figure 6.30 The motion group of a cube

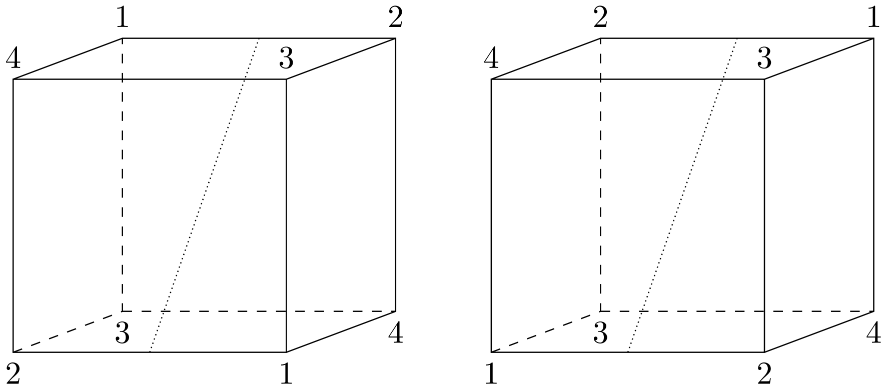


Figure 6.31 Transpositions in the motion group of a cube

■

6.2.2 Reading Questions

1. How many elements are in the group D_5? Give an example of one of the elements.

2. Give an example of an element in S_5 that is not in D_5, and explain why the element is not in D_5.

3. Is S_4 a dihedral group? Explain.

6.2.3 Exercises

1. Prove that D_n is nonabelian for $n \geq 3$.

2. If the diagonals of a cube are labeled as Figure 6.30, to which motion of the cube does the permutation $(12)(34)$ correspond? What about the other permutations of the diagonals?

3. Prove S_4 is not isomorphic to D_{12}.

4. Let $\omega = \operatorname{cis}(2\pi/n)$ be a primitive nth root of unity. Prove that the matrices

$$A = \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

generate a multiplicative group isomorphic to D_n.

5. Show that the set of all matrices of the form

$$\begin{pmatrix} \pm 1 & k \\ 0 & 1 \end{pmatrix},$$

is a group isomorphic to D_n, where all entries in the matrix are in $\mathbb{Z}_n$.

6. Find the group of rigid motions of a tetrahedron. Show that this is the same group as A_4.

7. Let r and s be the elements in D_n described in Theorem 6.23

(a) Show that $srs = r^{-1}$.

(b) Show that $r^k s = sr^{-k}$ in D_n.

(c) Prove that the order of $r^k \in D_n$ is $n/\gcd(k, n)$.

6.3 Summary and Additional Exercises

6.3.1 The Important Ideas

- A **permutation** of a set X is a one-to-one and onto map $\pi : X \to X$. The permutations of a set X form a group S_X. If X is a finite set, we can assume $X = \{1, 2, \ldots, n\}$. In this case we write S_n instead of S_X. We call this group the **symmetric group**. on n letters. A subgroup of S_n is called a **permutation group**. Since we compose functions right to left, we will multiply permutations from right to left.

- A permutation $\sigma \in S_X$ is a **cycle of length** k if there exist elements $a_1, a_2, \ldots, a_k \in X$ such that

$$\begin{aligned} \sigma(a_1) &= a_2 \\ \sigma(a_2) &= a_3 \\ &\vdots \\ \sigma(a_k) &= a_1 \end{aligned}$$

and $\sigma(x) = x$ for all other elements $x \in X$. We write $(a_1, a_2, \ldots, a_k)$ to denote the cycle σ. Every permutation in S_n can be written as the product of disjoint cycles.

- A permutation of cycle of length 2 is called **transpositions**. Any permutation of a finite set containing at least two elements can be written as the product of transpositions. If a permutation σ can be expressed as the product of an even number of transpositions, then any other product of transpositions equaling σ must also contain an even number of transpositions. Similarly, if σ can be expressed as the product of an odd number of transpositions, then any other product of transpositions equaling σ must also contain an odd number of transpositions.

- The set of all even permutations, A_n, is called the **alternating group on** n **letters** and is a subgroup of S_n of order $n!/2$.

- For $n = 3, 4, \ldots$, we define the **nth dihedral group**, D_n, to be the group of rigid motions of reegular n-sided polygon or n-gon. The dihedral group, D_n, is a subgroup of S_n of order $2n$ consisiting of all products of the two elements r and s, satisfying the relations

$$r^n = 1$$
$$s^2 = 1$$
$$srs = r^{-1}.$$

6.3.2 Exercises

1. Let $\sigma = \sigma_1 \cdots \sigma_m \in S_n$ be the product of disjoint cycles. Prove that the order of σ is the least common multiple of the lengths of the cycles $\sigma_1, \ldots, \sigma_m$.

2. Let $\sigma \in S_n$ be a cycle. Prove that σ can be written as the product of at most $n - 1$ transpositions.

3. Let $\sigma \in S_n$. If σ is not a cycle, prove that σ can be written as the product of at most $n - 2$ transpositions.

4. If σ can be expressed as an odd number of transpositions, show that any other product of transpositions equaling σ must also be odd.

5. If σ is a cycle of odd length, prove that σ^2 is also a cycle.

6. Show that a 3-cycle is an even permutation.

7. Prove that in A_n with $n \geq 3$, any permutation is a product of cycles of length 3.

Hint. Consider the cases $(ab)(bc)$ and $(ab)(cd)$.

8. Prove that any element in S_n can be written as a finite product of the following permutations.

(a) $(12), (13), \ldots, (1n)$

(b) $(12), (23), \ldots, (n - 1, n)$

(c) $(12), (12 \ldots n)$

9. Let G be a group and define a map $\lambda_g : G \to G$ by $\lambda_g(a) = ga$. Prove that λ_g is a permutation of G.

10. Prove that there exist $n!$ permutations of a set containing n elements.

11. Recall that the **center** of a group G is

$$Z(G) = \{g \in G : gx = xg \text{ for all } x \in G\}.$$

Find the center of D_8. What about the center of D_{10}? What is the center of D_n?

12. Let $\tau = (a_1, a_2, \ldots, a_k)$ be a cycle of length k.

(a) Prove that if σ is any permutation, then

$$\sigma\tau\sigma^{-1} = (\sigma(a_1), \sigma(a_2), \ldots, \sigma(a_k))$$

is a cycle of length k.

(b) Let μ be a cycle of length k. Prove that there is a permutation σ such that $\sigma\tau\sigma^{-1} = \mu$.

Hint. For (a), show that $\sigma\tau\sigma^{-1}(\sigma(a_i)) = \sigma(a_{i+1})$.

13. For α and β in S_n, define $\alpha \sim \beta$ if there exists an $\sigma \in S_n$ such that $\sigma\alpha\sigma^{-1} = \beta$. Show that $\sim$ is an equivalence relation on S_n.

14. Let $\sigma \in S_X$. If $\sigma^n(x) = y$, we will say that $x \sim y$.

(a) Show that $\sim$ is an equivalence relation on X.

(b) If $\sigma \in A_n$ and $\tau \in S_n$, show that $\tau^{-1}\sigma\tau \in A_n$.

(c) Define the **orbit** of $x \in X$ under $\sigma \in S_X$ to be the set

$$\mathcal{O}_{x,\sigma} = \{y : x \sim y\}.$$

Compute the orbits of each of the following elements in S_5:

$$\begin{aligned} \alpha &= (1254) \\ \beta &= (123)(45) \\ \gamma &= (13)(25). \end{aligned}$$

(d) If $\mathcal{O}_{x,\sigma} \cap \mathcal{O}_{y,\sigma} \neq \emptyset$, prove that $\mathcal{O}_{x,\sigma} = \mathcal{O}_{y,\sigma}$. The orbits under a permutation σ are the equivalence classes corresponding to the equivalence relation $\sim$.

(e) A subgroup H of S_X is **transitive** if for every $x, y \in X$, there exists a $\sigma \in H$ such that $\sigma(x) = y$. Prove that $\langle \sigma \rangle$ is transitive if and only if $\mathcal{O}_{x,\sigma} = X$ for some $x \in X$.

15. Let $\alpha \in S_n$ for $n \geq 3$. If $\alpha\beta = \beta\alpha$ for all $\beta \in S_n$, prove that α must be the identity permutation; hence, the center of S_n is the trivial subgroup.

16. If α is even, prove that α^{-1} is also even. Does a corresponding result hold if α is odd?

17. Show that $\alpha^{-1}\beta^{-1}\alpha\beta$ is even for $\alpha, \beta \in S_n$.

18. Prove that $S_3 \times \mathbb{Z}_2$ is isomorphic to D_6. Can you make a conjecture about D_{2n}? Prove your conjecture.

Hint. Draw the picture.

19. Show that S_n is isomorphic to a subgroup of A_{n+2}.

20. Prove that D_n is isomorphic to a subgroup of S_n.

21. Let G be a group and $g \in G$. Define maps $\lambda_g : G \to G$ and $\rho_g : G \to G$ by $\lambda_g(x) = gx$ and $\rho_g(x) = xg^{-1}$. Show that $i_g = \rho_g \circ \lambda_g$ is an automorphism of G. The isomorphism $g \mapsto \rho_g$ is called the **right regular representation** of G.

22. **Groups of order $2p$.** In this series of exercises we will classify all groups of order $2p$, where p is an odd prime.

(a) Assume G is a group of order $2p$, where p is an odd prime. If $a \in G$, show that a must have order 1, 2, p, or $2p$.

(b) Suppose that G has an element of order $2p$. Prove that G is isomorphic to $\mathbb{Z}_{2p}$. Hence, G is cyclic.

(c) Suppose that G does not contain an element of order $2p$. Show that G must contain an element of order p. *Hint*: Assume that G does not contain an element of order p.

(d) Suppose that G does not contain an element of order $2p$. Show that G must contain an element of order 2.

(e) Let P be a subgroup of G with order p and $y \in G$ have order 2. Show that $yP = Py$.

(f) Suppose that G does not contain an element of order $2p$ and $P = \langle z \rangle$ is a subgroup of order p generated by z. If y is an element of order 2, then $yz = z^k y$ for some $2 \leq k < p$.

(g) Suppose that G does not contain an element of order $2p$. Prove that G is not abelian.

(h) Suppose that G does not contain an element of order $2p$ and $P = \langle z \rangle$ is a subgroup of order p generated by z and y is an element of order 2. Show that we can list the elements of G as $\{z^i y^j \mid 0 \leq i < p, 0 \leq j < 2\}$.

(i) Suppose that G does not contain an element of order $2p$ and $P = \langle z \rangle$ is a subgroup of order p generated by z and y is an element of order 2. Prove that the product $z^i y^j)(z^r y^s)$ can be expressed as a uniquely as $z^m y^n$ for some non negative integers m, n. Thus, conclude that there is only one possibility for a non-abelian group of order $2p$, it must therefore be the one we have seen already, the dihedral group.

6.4 Connections to the Secondary Classroom—Permutation Groups

This appendix will connect permutation groups to the high school classroom, especially nonabelian groups.

Chapter 7

Cosets and Lagrange's Theorem

Objectives

- To understand that left cosets of a subgroup H in a group $> G$ partition G and that the **index** of H in G is the number of left cosets of H in G, which we will denote the index by $[G : H]$.
- To understand and be able to apply Lagrange's Theorem and that the order of a subgroup must divide the order of the group.
- To understand and be able to apply the **Euler ϕ-function**.

Lagrange's Theorem, one of the most important results in finite group theory, states that the order of a subgroup must divide the order of the group. This theorem provides a powerful tool for analyzing finite groups; it gives us an idea of exactly what type of subgroups we might expect a finite group to possess. Central to understanding Lagranges's Theorem is the notion of a coset.

7.1 Cosets

Let G be a group and H a subgroup of G. Define a **left coset** of H with **representative** $g \in G$ to be the set

$$gH = \{gh : h \in H\}.$$

Right cosets can be defined similarly by

$$Hg = \{hg : h \in H\}.$$

If left and right cosets coincide or if it is clear from the context to which type of coset that we are referring, we will use the word *coset* without specifying left or right.

Example 7.1 Let H be the subgroup of $\mathbb{Z}_6$ consisting of the elements 0 and 3. The cosets are

$$\begin{aligned} 0 + H &= 3 + H = \{0, 3\} \\ 1 + H &= 4 + H = \{1, 4\} \\ 2 + H &= 5 + H = \{2, 5\}. \end{aligned}$$

We will always write the cosets of subgroups of $\mathbb{Z}$ and $\mathbb{Z}_n$ with the additive notation we have used for cosets here. In a commutative group, left and right cosets are always identical. □

Example 7.2 Let H be the subgroup of S_3 defined by the permutations $\{(1), (123), (132)\}$. The left cosets of H are

$$\begin{gathered}(1)H = (123)H = (132)H = \{(1), (123), (132)\} \\ (12)H = (13)H = (23)H = \{(12), (13), (23)\}.\end{gathered}$$

The right cosets of H are exactly the same as the left cosets:

$$\begin{gathered}H(1) = H(123) = H(132) = \{(1), (123), (132)\} \\ H(12) = H(13) = H(23) = \{(12), (13), (23)\}.\end{gathered}$$

It is not always the case that a left coset is the same as a right coset. Let K be the subgroup of S_3 defined by the permutations $\{(1), (12)\}$. Then the left cosets of K are

$$\begin{gathered}(1)K = (12)K = \{(1), (12)\} \\ (13)K = (123)K = \{(13), (123)\} \\ (23)K = (132)K = \{(23), (132)\};\end{gathered}$$

however, the right cosets of K are

$$\begin{gathered}K(1) = K(12) = \{(1), (12)\} \\ K(13) = K(132) = \{(13), (132)\} \\ K(23) = K(123) = \{(23), (123)\}.\end{gathered}$$

□

Activity 7.1 List the left and right cosets of the subgroups in each of the following.

1. $\langle 8 \rangle$ in $\mathbb{Z}_{24}$
2. $3\mathbb{Z}$ in $\mathbb{Z}$
3. A_4 in S_4
4. $\mathbb{T}$ in $\mathbb{C}^*$

The following lemma is quite useful when dealing with cosets. (We leave its proof as an exercise.

Lemma 7.3 *Let H be a subgroup of a group G and suppose that $g_1, g_2 \in G$. The following conditions are equivalent.*

1. *$g_1 H = g_2 H$;*
2. *$H g_1^{-1} = H g_2^{-1}$;*
3. *$g_1 H \subset g_2 H$;*
4. *$g_2 \in g_1 H$;*
5. *$g_1^{-1} g_2 \in H$.*

In all of our examples the cosets of a subgroup H partition the larger group G. The following theorem proclaims that this will always be the case.

Theorem 7.4 *Let H be a subgroup of a group G. Then the left cosets of H in G partition G. That is, the group G is the disjoint union of the left cosets of H in G.*

Proof. Let g_1H and g_2H be two cosets of H in G. We must show that either $g_1H \cap g_2H = \emptyset$ or $g_1H = g_2H$. Suppose that $g_1H \cap g_2H \neq \emptyset$ and $a \in g_1H \cap g_2H$. Then by the definition of a left coset, $a = g_1h_1 = g_2h_2$ for some elements h_1 and h_2 in H. Hence, $g_1 = g_2h_2h_1^{-1}$ or $g_1 \in g_2H$. By Lemma 7.3, $g_1H = g_2H$. ■

Remark 7.5 There is nothing special in this theorem about left cosets. Right cosets also partition G; the proof of this fact is exactly the same as the proof for left cosets except that all group multiplications are done on the opposite side of H.

Let G be a group and H be a subgroup of G. Define the **index** of H in G to be the number of left cosets of H in G. We will denote the index by $[G : H]$.

Example 7.6 Let $G = \mathbb{Z}_6$ and $H = \{0, 3\}$. Then $[G : H] = 3$. □

Example 7.7 Suppose that $G = S_3$, $H = \{(1), (123), (132)\}$, and $K = \{(1), (12)\}$. Then $[G : H] = 2$ and $[G : K] = 3$. □

Theorem 7.8 *Let H be a subgroup of a group G. The number of left cosets of H in G is the same as the number of right cosets of H in G.*

Proof. Let $\mathcal{L}_H$ and $\mathcal{R}_H$ denote the set of left and right cosets of H in G, respectively. If we can define a bijective map $\phi : \mathcal{L}_H \to \mathcal{R}_H$, then the theorem will be proved. If $gH \in \mathcal{L}_H$, let $\phi(gH) = Hg^{-1}$. By Lemma 7.3, the map ϕ is well-defined; that is, if $g_1H = g_2H$, then $Hg_1^{-1} = Hg_2^{-1}$. To show that ϕ is one-to-one, suppose that

$$Hg_1^{-1} = \phi(g_1H) = \phi(g_2H) = Hg_2^{-1}.$$

Again by Lemma 7.3, $g_1H = g_2H$. The map ϕ is onto since $\phi(g^{-1}H) = Hg$. ■

Activity 7.2 What fails in the proof of Theorem 7.8 if $\phi : \mathcal{L}_H \to \mathcal{R}_H$ is defined by $\phi(gH) = Hg$?

Reading Questions

1. Why do we sometimes need to specify *left* or *right* cosets instead of just calling them cosets? When do we not need to worry about this?

2. Let $G = \mathbb{Z}_6$ and $H = \{0, 2, 4\}$. List all cosets of H in G. How many are there?

3. Let $G = S_4$ and $H = \{(1), (12)(34), (13)(24), (14)(23)\}$. What is the index of H in G? Briefly explain.

Exercises

1. List the left and right cosets of the subgroups in each of the following.

(a) $\langle 3 \rangle$ in $U(8)$

(b) A_n in S_n

(c) D_4 in S_4

(d) $H = \{(1), (123), (132)\}$ in S_4

Hint. (a) $\langle 8 \rangle$, $1 + \langle 8 \rangle$, $2 + \langle 8 \rangle$, $3 + \langle 8 \rangle$, $4 + \langle 8 \rangle$, $5 + \langle 8 \rangle$, $6 + \langle 8 \rangle$, and $7 + \langle 8 \rangle$; (c) $3\mathbb{Z}$, $1 + 3\mathbb{Z}$, and $2 + 3\mathbb{Z}$.

2. Describe the left cosets of $SL_2(\mathbb{R})$ in $GL_2(\mathbb{R})$. What is the index of $SL_2(\mathbb{R})$ in $GL_2(\mathbb{R})$?

3. Show that the integers have infinite index in the additive group of rational numbers.

4. Show that the additive group of real numbers has infinite index in the additive group of the complex numbers.

5. If $ghg^{-1} \in H$ for all $g \in G$ and $h \in H$, show that right cosets are identical to left cosets. That is, show that $gH = Hg$ for all $g \in G$.

Hint. Let $g_1 \in gH$. Show that $g_1 \in Hg$ and thus $gH \subset Hg$.

6. Let H be a subgroup of a group G and suppose that $g_1, g_2 \in G$. Prove that the following conditions are equivalent.

(a) $g_1 H = g_2 H$

(b) $Hg_1^{-1} = Hg_2^{-1}$

(c) $g_1 H \subset g_2 H$

(d) $g_2 \in g_1 H$

(e) $g_1^{-1} g_2 \in H$

7. Suppose that $[G : H] = 2$. If a and b are not in H, show that $ab \in H$.

8. If $[G : H] = 2$, prove that $gH = Hg$.

7.2 Lagrange's Theorem

Our goal is to uncover the connection between the size of a group and the possible order of elements inside the group. Since the order of an element a is the same as the order of the subgroup generated by a, i.e., $\operatorname{ord}(a) = |\langle a \rangle|$, it will be enough to find a connection between the order of a group and the orders of its subgroups.

The connection is Lagrange's theorem, stated below. It is a remarkable theorem both in terms of its content and the simplicity of its proof. The proof is a consequence of some facts about cosets, in particular, that cosets create a partition of the group into equal sized sets. We prove this fact first.

Proposition 7.9 *Let H be a subgroup of G with $g \in G$ and define a map $\phi : H \to gH$ by $\phi(h) = gh$. The map ϕ is bijective; hence, the number of elements in H is the same as the number of elements in gH.*

Proof. We first show that the map ϕ is one-to-one. Suppose that $\phi(h_1) = \phi(h_2)$ for elements $h_1, h_2 \in H$. We must show that $h_1 = h_2$, but $\phi(h_1) = gh_1$ and $\phi(h_2) = gh_2$. So $gh_1 = gh_2$, and by left cancellation $h_1 = h_2$. To show that ϕ is onto is easy. By definition every element of gH is of the form gh for some $h \in H$ and $\phi(h) = gh$. ■

We can now easily prove Lagrange's theorem.

Theorem 7.10 Lagrange. *Let G be a finite group and let H be a subgroup of G. Then $|G|/|H| = [G : H]$ is the number of distinct left cosets of H in G. In particular, the number of elements in H must divide the number of elements in G.*

Proof. The group G is partitioned into $[G : H]$ distinct left cosets. Each left coset has $|H|$ elements; therefore, $|G| = [G : H]|H|$. ■

Now using the relationship between order of elements and sizes of subgroups, we arrive at our desired results.

Corollary 7.11 *Suppose that G is a finite group and $g \in G$. Then the order of g must divide the number of elements in G.*

Corollary 7.12 *Let $|G| = p$ with p a prime number. Then G is cyclic and any $g \in G$ such that $g \neq e$ is a generator. Furthermore, G is isomorphic to $\mathbb{Z}_p$.*

Proof. Let g be in G such that $g \neq e$. Then by Corollary 7.11, the order of g must divide the order of the group. Since $|\langle g \rangle| > 1$, it must be p. Hence, g generates G. ■

Corollary 7.12 suggests that groups of prime order p must somehow look like $\mathbb{Z}_p$.

Corollary 7.13 *Let H and K be subgroups of a finite group G such that $G \supset H \supset K$. Then*

$$[G : K] = [G : H][H : K].$$

Proof. Observe that

$$[G : K] = \frac{|G|}{|K|} = \frac{|G|}{|H|} \cdot \frac{|H|}{|K|} = [G : H][H : K].$$

■

Activity 7.3 Suppose that G is a finite group with 60 elements. What are the orders of possible subgroups of G?

Remark 7.14 The converse of Lagrange's Theorem is false. The group A_4 has order 12; however, it can be shown that it does not possess a subgroup of order 6. According to Lagrange's Theorem, subgroups of a group of order 12 can have orders of either 1, 2, 3, 4, or 6. However, we are not guaranteed that subgroups of every possible order exist. To prove that A_4 has no subgroup of order 6, we will assume that it does have such a subgroup H and show that a contradiction must occur. Since A_4 contains eight 3-cycles, we know that H must contain a 3-cycle. We will show that if H contains one 3-cycle, then it must contain more than 6 elements.

Proposition 7.15 *The group A_4 has no subgroup of order* 6.

Proof. Since $[A_4 : H] = 2$, there are only two cosets of H in A_4. Inasmuch as one of the cosets is H itself, right and left cosets must coincide; therefore, $gH = Hg$ or $gHg^{-1} = H$ for every $g \in A_4$. Since there are eight 3-cycles in A_4, at least one 3-cycle must be in H. Without loss of generality, assume that (123) is in H. Then $(123)^{-1} = (132)$ must also be in H. Since $ghg^{-1} \in H$ for all $g \in A_4$ and all $h \in H$ and

$$\begin{aligned}(124)(123)(124)^{-1} &= (124)(123)(142) = (243)\\ (243)(123)(243)^{-1} &= (243)(123)(234) = (142)\end{aligned}$$

we can conclude that H must have at least seven elements

$$(1), (123), (132), (243), (243)^{-1} = (234), (142), (142)^{-1} = (124).$$

Therefore, A_4 has no subgroup of order 6. ■

In fact, we can say more about when two cycles have the same length.

Theorem 7.16 *Two cycles τ and μ in S_n have the same length if and only if there exists a $\sigma \in S_n$ such that $\mu = \sigma\tau\sigma^{-1}$.*

Proof. Suppose that

$$\begin{aligned}\tau &= (a_1, a_2, \ldots, a_k)\\ \mu &= (b_1, b_2, \ldots, b_k).\end{aligned}$$

Define σ to be the permutation

$$\begin{aligned}\sigma(a_1) &= b_1\\ \sigma(a_2) &= b_2\end{aligned}$$

$$\vdots$$

$$\sigma(a_k) = b_k.$$

Then $\mu = \sigma\tau\sigma^{-1}$.

Conversely, suppose that $\tau = (a_1, a_2, \ldots, a_k)$ is a k-cycle and $\sigma \in S_n$. If $\sigma(a_i) = b$ and $\sigma(a_{(i \bmod k)+1}) = b'$, then $\mu(b) = b'$. Hence,

$$\mu = (\sigma(a_1), \sigma(a_2), \ldots, \sigma(a_k)).$$

Since σ is one-to-one and onto, μ is a cycle of the same length as τ. ■

Reading Questions

1. Why must the index of a subgroup H of a finite group G be a whole number (i.e., a positive integer)? Briefly explain.

2. How is Proposition 7.9 used in the proof of Theorem 7.10? Why is it important?

3. True or false: if a group has order n and k is a divisor of n, then the group has an element of order k. Briefly explain.

Exercises

1. Suppose that G is a finite group with an element g of order 5 and an element h of order 7. Why must $|G| \geq 35$?

Hint. The order of g and the order h must both divide the order of G.

2. Prove or disprove: Every subgroup of the integers has finite index.

Hint. This is true for every proper nontrivial subgroup.

3. Prove or disprove: Every subgroup of the integers has finite order.

Hint. False.

7.3 Fermat's and Euler's Theorems

The **Euler ϕ-function** is the map $\phi : \mathbb{N} \to \mathbb{N}$ defined by $\phi(n) = 1$ for $n = 1$, and, for $n > 1$, $\phi(n)$ is the number of positive integers m with $1 \leq m < n$ and $\gcd(m, n) = 1$.

From Proposition 3.4, we know that the order of $U(n)$, the group of units in $\mathbb{Z}_n$, is $\phi(n)$. For example, $|U(12)| = \phi(12) = 4$ since the numbers that are relatively prime to 12 are 1, 5, 7, and 11. For any prime p, $\phi(p) = p - 1$. We state these results in the following theorem.

Theorem 7.17 *Let $U(n)$ be the group of units in $\mathbb{Z}_n$. Then $|U(n)| = \phi(n)$.*

The following theorem is an important result in number theory, due to Leonhard Euler.

Theorem 7.18 Euler's Theorem. *Let a and n be integers such that $n > 0$ and $\gcd(a, n) = 1$. Then $a^{\phi(n)} \equiv 1 \pmod{n}$.*

Proof. By Theorem 7.17 the order of $U(n)$ is $\phi(n)$. Consequently, $a^{\phi(n)} = 1$ for all $a \in U(n)$; or $a^{\phi(n)} - 1$ is divisible by n. Therefore, $a^{\phi(n)} \equiv 1 \pmod{n}$. ■

If we consider the special case of Euler's Theorem in which $n = p$ is prime and recall that $\phi(p) = p - 1$, we obtain the following result, due to Pierre de Fermat.

Theorem 7.19 Fermat's Little Theorem. *Let p be any prime number and suppose that $p \nmid a$ (p does not divide a). Then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Furthermore, for any integer b, $b^p \equiv b \pmod{p}$.

Activity 7.4 Verify Euler's Theorem for $n = 15$ and $a = 4$.

7.3.1 Historical Note

Joseph-Louis Lagrange (1736–1813), born in Turin, Italy, was of French and Italian descent. His talent for mathematics became apparent at an early age. Leonhard Euler recognized Lagrange's abilities when Lagrange, who was only 19, That year he was also named a professor at the Royal Artillery School in Turin. At the age of 23 he joined the Berlin Academy. Frederick the Great had written to Lagrange proclaiming that the "greatest king in Europe" should have the "greatest mathematician in Europe" at his court. For 20 years Lagrange held the position vacated by his mentor, Euler. His works include contributions to number theory, group theory, physics and mechanics, the calculus of variations, the theory of equations, and differential equations. Along with Laplace and Lavoisier, Lagrange was one of the people responsible for designing the metric system. During his life Lagrange profoundly influenced the development of mathematics,leaving much to the next generation of mathematicians in the form of examples and new problems to be solved.

7.3.2 Reading Questions

1. The group $U(35)$ contains 24 elements. What is the value of $\phi(35)$?
2. True or false: $a^{34} \equiv 1 \pmod{35}$ for any a that is not a multiple of 35. Briefly explain.

7.3.3 Exercises

1. Use Fermat's Little Theorem to show that if $p = 4n + 3$ is prime, there is no solution to the equation $x^2 \equiv -1 \pmod{p}$.
2. Let G be a cyclic group of order n. Show that there are exactly $\phi(n)$ generators for G.

7.4 Summary and Additional Exercises

7.4.1 The Important Ideas

- Let G be a group. Define a **left coset** of a subgroup H with **representative** $g \in G$ to be the set

  $$gH = \{gh : h \in H\}.$$

  Right cosets are defined similarly by

  $$Hg = \{hg : h \in H\}.$$

  If left and right cosets coincide or if it is clear from the context to which type of coset that we are referring, we will use the word **coset** without specifying left or right.

- If H is a subgroup of a group G, then the left cosets of H in G partition G. The **index** of H in G is the number of left cosets of H in G and is denoted by $[G : H]$. The number of left cosets is the same as the number of right cosets.

- Let H be a subgroup of G with $g \in G$ and define a map $\phi : H \to gH$ by $\phi(h) = gh$. The map ϕ is bijective; hence, the number of elements in H is the same as the number of elements in gH.

- If G is a finite group and H is a subgroup of G, then **Lagrange's Theorem** states that $|G|/|H| = [G : H]$ is the number of distinct left cosets of H in G. In particular, the number of elements in H must divide the number of elements in G. In particular, if H and K are subgroups of a finite group G such that $G \supset H \supset K$, then

$$[G : K] = [G : H][H : K].$$

- The converse of Lagrange's Theorem is false. The group A_4 has order 12 but does not possess a subgroup of order 6.

- The **Euler ϕ-function** is the map $\phi : \mathbb{N} \to \mathbb{N}$ defined by $\phi(n) = 1$ for $n = 1$, and, for $n > 1$, $\phi(n)$ is the number of positive integers m with $1 \leq m < n$ and $\gcd(m, n) = 1$.

- Let a and n be integers such that $n > 0$ and $\gcd(a, n) = 1$. Then Euler's Theorem states that $a^{\phi(n)} \equiv 1 \pmod{n}$.

- Let p be any prime number and suppose that $p \nmid a$ (p does not divide a). Then Fermat's Little Theorem states that

$$a^{p-1} \equiv 1 \pmod{p}.$$

Furthermore, for any integer b, $b^p \equiv b \pmod{p}$.

7.4.2 Exercises

1. Suppose that $g^n = e$. Show that the order of g divides n.

2. The **cycle structure** of a permutation σ is defined as the unordered list of the sizes of the cycles in the cycle decomposition σ. For example, the permutation $\sigma = (12)(345)(78)(9)$ has cycle structure $(2, 3, 2, 1)$ which can also be written as $(1, 2, 2, 3)$.

Show that any two permutations $\alpha, \beta \in S_n$ have the same cycle structure if and only if there exists a permutation γ such that $\beta = \gamma\alpha\gamma^{-1}$. If $\beta = \gamma\alpha\gamma^{-1}$ for some $\gamma \in S_n$, then α and β are **conjugate**.

3. If $|G| = 2n$, prove that the number of elements of order 2 is odd. Use this result to show that G must contain a subgroup of order 2.

4. Let H and K be subgroups of a group G. Prove that $gH \cap gK$ is a coset of $H \cap K$ in G.

Hint. Show that $g(H \cap K) = gH \cap gK$.

5. Let H and K be subgroups of a group G. Define a relation $\sim$ on G by $a \sim b$ if there exists an $h \in H$ and a $k \in K$ such that $hak = b$. Show that this relation is an equivalence relation. The corresponding equivalence classes are called **double cosets**. Compute the double cosets of $H = \{(1), (123), (132)\}$ in A_4.

6. Let $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, where $p_1, p_2, \ldots, p_k$ are distinct primes. Prove that

$$\phi(n) = n \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_k}\right).$$

Hint. If $\gcd(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$ (Exercise 2.4.2.8 in Chapter 2).

7. Show that

$$n = \sum_{d|n} \phi(d)$$

for all positive integers n.

7.5 Connections to the Secondary Classroom—The Euler ϕ-function

This appendix will connect the Euler ϕ-function to the high school classroom.

Chapter 8

Normal Subgroups and Homomorphisms

Objectives

- To understand and be able to apply the definition of a normal subgroup.
- Given a group G and a normal subgoup N, to understand and be able the definition of a quotient group G/N.
- To understand and be able to apply the definition of a group homomorphism.
- To understand the relationship between a group homomorphism and normal subgroups.

If H is a subgroup of a group G, then right cosets are not always the same as left cosets; that is, it is not always the case that $gH = Hg$ for all $g \in G$. The subgroups for which this property holds play a critical role in group theory—they allow for the construction of a new class of groups, called factor or quotient groups. Factor groups may be studied directly or by using homomorphisms, a generalization of isomorphisms (Section 3.4.).

8.1 Factor Groups and Normal Subgroups

8.1.1 Normal Subgroups

A subgroup H of a group G is **normal** in G if $gH = Hg$ for all $g \in G$. That is, a normal subgroup of a group G is one in which the right and left cosets are precisely the same.

Example 8.1 Let G be an abelian group. Every subgroup H of G is a normal subgroup. Since $gh = hg$ for all $g \in G$ and $h \in H$, it will always be the case that $gH = Hg$. □

Example 8.2 Let H be the subgroup of S_3 consisting of elements (1) and (12). Since

$$(123)H = \{(123), (13)\} \quad \text{and} \quad H(123) = \{(123), (23)\},$$

H cannot be a normal subgroup of S_3. However, the subgroup N, consisting of the permutations (1), (123), and (132), is normal since the cosets of N are

$$\begin{aligned} N &= \{(1), (123), (132)\} \\ (12)N = N(12) &= \{(12), (13), (23)\}. \end{aligned}$$

□

The following theorem is fundamental to our understanding of normal subgroups.

Theorem 8.3 *Let G be a group and N be a subgroup of G. Then the following statements are equivalent.*

1. *The subgroup N is normal in G.*
2. *For all $g \in G$, $gNg^{-1} \subset N$.*
3. *For all $g \in G$, $gNg^{-1} = N$.*

Proof. (1) ⇒ (2). Since N is normal in G, $gN = Ng$ for all $g \in G$. Hence, for a given $g \in G$ and $n \in N$, there exists an n' in N such that $gn = n'g$. Therefore, $gng^{-1} = n' \in N$ or $gNg^{-1} \subset N$.

(2) ⇒ (3). Let $g \in G$. Since $gNg^{-1} \subset N$, we need only show $N \subset gNg^{-1}$. For $n \in N$, $g^{-1}ng = g^{-1}n(g^{-1})^{-1} \in N$. Hence, $g^{-1}ng = n'$ for some $n' \in N$. Therefore, $n = gn'g^{-1}$ is in gNg^{-1}.

(3) ⇒ (1). Suppose that $gNg^{-1} = N$ for all $g \in G$. Then for any $n \in N$ there exists an $n' \in N$ such that $gng^{-1} = n'$. Consequently, $gn = n'g$ or $gN \subset Ng$. Similarly, $Ng \subset gN$. ■

8.1.2 Factor Groups

If N is a normal subgroup of a group G, then the cosets of N in G form a group G/N under the operation $(aN)(bN) = abN$. This group is called the **factor** or **quotient group** of G and N. Our first task is to prove that G/N is indeed a group.

Theorem 8.4 *Let N be a normal subgroup of a group G. The cosets of N in G form a group G/N of order $[G : N]$.*

Proof. The group operation on G/N is $(aN)(bN) = abN$. This operation must be shown to be well-defined; that is, group multiplication must be independent of the choice of coset representative. Let $aN = bN$ and $cN = dN$. We must show that

$$(aN)(cN) = acN = bdN = (bN)(dN).$$

Then $a = bn_1$ and $c = dn_2$ for some n_1 and n_2 in N. Hence,

$$\begin{aligned} acN &= bn_1dn_2N \\ &= bn_1dN \\ &= bn_1Nd \\ &= bNd \\ &= bdN. \end{aligned}$$

The remainder of the theorem is easy: $eN = N$ is the identity and $g^{-1}N$ is the inverse of gN. The order of G/N is, of course, the number of cosets of N in G. ■

It is very important to remember that the elements in a factor group are *sets of elements* in the original group.

Example 8.5 Consider the normal subgroup of S_3, $N = \{(1), (123), (132)\}$. The cosets of N in S_3 are N and $(12)N$. The factor group S_3/N has the following multiplication table.

	N	$(12)N$
N	N	$(12)N$
$(12)N$	$(12)N$	N

This group is isomorphic to $\mathbb{Z}_2$. At first, multiplying cosets seems both complicated and strange; however, notice that S_3/N is a smaller group. The factor group displays a certain amount of information about S_3. Actually, $N = A_3$, the group of even permutations, and $(12)N = \{(12), (13), (23)\}$ is the set of odd permutations. The information captured in G/N is parity; that is, multiplying two even or two odd permutations results in an even permutation, whereas multiplying an odd permutation by an even permutation yields an odd permutation. □

Example 8.6 Consider the normal subgroup $3\mathbb{Z}$ of $\mathbb{Z}$. The cosets of $3\mathbb{Z}$ in $\mathbb{Z}$ are

$$\begin{aligned} 0 + 3\mathbb{Z} &= \{\ldots, -3, 0, 3, 6, \ldots\} \\ 1 + 3\mathbb{Z} &= \{\ldots, -2, 1, 4, 7, \ldots\} \\ 2 + 3\mathbb{Z} &= \{\ldots, -1, 2, 5, 8, \ldots\}. \end{aligned}$$

The group $\mathbb{Z}/3\mathbb{Z}$ is given by the Cayley table below.

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

In general, the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$ is normal. The cosets of $\mathbb{Z}/n\mathbb{Z}$ are

$$\begin{gathered} n\mathbb{Z} \\ 1 + n\mathbb{Z} \\ 2 + n\mathbb{Z} \\ \vdots \\ (n-1) + n\mathbb{Z}. \end{gathered}$$

The sum of the cosets $k + n\mathbb{Z}$ and $l + n\mathbb{Z}$ is $k + l + n\mathbb{Z}$. Notice that we have written our cosets additively, because the group operation is integer addition. □

Example 8.7 Consider the dihedral group D_n, generated by the two elements r and s, satisfying the relations

$$\begin{aligned} r^n &= \mathrm{id} \\ s^2 &= \mathrm{id} \\ srs &= r^{-1}. \end{aligned}$$

The element r actually generates the cyclic subgroup of rotations, R_n, of D_n. Since $srs^{-1} = srs = r^{-1} \in R_n$, the group of rotations is a normal subgroup of D_n; therefore, D_n/R_n is a group. Since there are exactly two elements in this group, it must be isomorphic to $\mathbb{Z}_2$. □

Activity 8.1 For each of the following groups G, determine whether H is a normal subgroup of G. If H is a normal subgroup, write out a Cayley table for the factor group G/H.

1. $G = S_4$ and $H = A_4$
2. $G = A_5$ and $H = \{(1), (123), (132)\}$
3. $G = S_4$ and $H = D_4$
4. $G = Q_8$ and $H = \{1, -1, I, -I\}$
5. $G = \mathbb{Z}$ and $H = 5\mathbb{Z}$

8.1.3 Reading Questions

1. Give two ways we can can be sure that a subgroup H of a group G is normal.

2. What does the notation G/H represent? What does the notation $[G : H]$ represent? How are these related?

3. The subgroup $8\mathbb{Z}$ is normal in $\mathbb{Z}$. In the factor group $\mathbb{Z}/8\mathbb{Z}$, perform the computation $(3 + 8\mathbb{Z}) + (7 + 8\mathbb{Z})$.

8.1.4 Exercises

1. Find all the subgroups of D_4. Which subgroups are normal? What are all the factor groups of D_4 up to isomorphism?

2. Find all the subgroups of D_4. Which subgroups are normal? What are all the factor groups of D_4 up to isomorphism?

3. Find all the subgroups of the quaternion group, Q_8. Which subgroups are normal? What are all the factor groups of Q_8 up to isomorphism?

4. Let T be the group of nonsingular upper triangular 2×2 matrices with entries in $\mathbb{R}$; that is, matrices of the form

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix},$$

where a, b, $c \in \mathbb{R}$ and $ac \neq 0$. Let U consist of matrices of the form

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix},$$

where $x \in \mathbb{R}$.

(a) Show that U is a subgroup of T.

(b) Prove that U is abelian.

(c) Prove that U is normal in T.

(d) Show that T/U is abelian.

(e) Is T normal in $GL_2(\mathbb{R})$?

5. Show that the intersection of two normal subgroups is a normal subgroup.

6. If G is abelian, prove that G/H must also be abelian.

7. Prove or disprove: If H is a normal subgroup of G such that H and G/H are abelian, then G is abelian.

8. If G is cyclic, prove that G/H must also be cyclic.

Hint. If $a \in G$ is a generator for G, then aH is a generator for G/H.

9. Prove or disprove: If H and G/H are cyclic, then G is cyclic.

10. Let H be a subgroup of index 2 of a group G. Prove that H must be a normal subgroup of G. Conclude that S_n is not simple for $n \geq 3$.

8.2 Group Homomorphisms

A **homomorphism** between groups $(G, \cdot)$ and $(H, \circ)$ is a map $\phi : G \to H$ such that

$$\phi(g_1 \cdot g_2) = \phi(g_1) \circ \phi(g_2)$$

for $g_1, g_2 \in G$. The range of ϕ in H is called the **homomorphic image** of ϕ.

Two groups are related in the strongest possible way if they are isomorphic; however, a weaker relationship may exist between two groups. For example, the symmetric group S_n and the group $\mathbb{Z}_2$ are related by the fact that S_n can be divided into even and odd permutations that exhibit a group structure like that $\mathbb{Z}_2$, as shown in the following multiplication table.

	even	odd
even	even	odd
odd	odd	even

We use homomorphisms to study relationships such as the one we have just described.

Example 8.8 Let G be a group and $g \in G$. Define a map $\phi : \mathbb{Z} \to G$ by $\phi(n) = g^n$. Then ϕ is a group homomorphism, since

$$\phi(m + n) = g^{m+n} = g^m g^n = \phi(m)\phi(n).$$

This homomorphism maps $\mathbb{Z}$ onto the cyclic subgroup of G generated by g. □

Example 8.9 Let $G = GL_2(\mathbb{R})$. If

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in G, then the determinant is nonzero; that is, $\det(A) = ad - bc \neq 0$. Also, for any two elements A and B in G, $\det(AB) = \det(A)\det(B)$. Using the determinant, we can define a homomorphism $\phi : GL_2(\mathbb{R}) \to \mathbb{R}^*$ by $A \mapsto \det(A)$. □

Example 8.10 Recall that the circle group $\mathbb{T}$ consists of all complex numbers z such that $|z| = 1$. We can define a homomorphism ϕ from the additive group of real numbers $\mathbb{R}$ to $\mathbb{T}$ by $\phi : \theta \mapsto \cos\theta + i \sin\theta$. Indeed,

$$\begin{aligned} \phi(\alpha + \beta) &= \cos(\alpha + \beta) + i \sin(\alpha + \beta) \\ &= (\cos\alpha \cos\beta - \sin\alpha \sin\beta) + i(\sin\alpha \cos\beta + \cos\alpha \sin\beta) \\ &= (\cos\alpha + i \sin\alpha)(\cos\beta + i \sin\beta) \\ &= \phi(\alpha)\phi(\beta). \end{aligned}$$

Geometrically, we are simply wrapping the real line around the circle in a group-theoretic fashion. □

The following proposition lists some basic properties of group homomorphisms.

Proposition 8.11 *Let $\phi : G_1 \to G_2$ be a homomorphism of groups. Then*

1. *If e is the identity of G_1, then $\phi(e)$ is the identity of G_2;*
2. *For any element $g \in G_1$, $\phi(g^{-1}) = [\phi(g)]^{-1}$;*
3. *If H_1 is a subgroup of G_1, then $\phi(H_1)$ is a subgroup of G_2;*
4. *If H_2 is a subgroup of G_2, then $\phi^{-1}(H_2) = \{g \in G_1 : \phi(g) \in H_2\}$ is a subgroup of G_1. Furthermore, if H_2 is normal in G_2, then $\phi^{-1}(H_2)$ is normal in G_1.*

Proof. (1) Suppose that e and e' are the identities of G_1 and G_2, respectively; then

$$e'\phi(e) = \phi(e) = \phi(ee) = \phi(e)\phi(e).$$

By cancellation, $\phi(e) = e'$.

(2) This statement follows from the fact that

$$\phi(g^{-1})\phi(g) = \phi(g^{-1}g) = \phi(e) = e'.$$

(3) The set $\phi(H_1)$ is nonempty since the identity of G_2 is in $\phi(H_1)$. Suppose that H_1 is a subgroup of G_1 and let x and y be in $\phi(H_1)$. There exist elements $a, b \in H_1$ such that $\phi(a) = x$ and $\phi(b) = y$. Since

$$xy^{-1} = \phi(a)[\phi(b)]^{-1} = \phi(ab^{-1}) \in \phi(H_1),$$

$\phi(H_1)$ is a subgroup of G_2 by Proposition 3.31.

(4) Let H_2 be a subgroup of G_2 and define H_1 to be $\phi^{-1}(H_2)$; that is, H_1 is the set of all $g \in G_1$ such that $\phi(g) \in H_2$. The identity is in H_1 since $\phi(e) = e'$. If a and b are in H_1, then $\phi(ab^{-1}) = \phi(a)[\phi(b)]^{-1}$ is in H_2 since H_2 is a subgroup of G_2. Therefore, $ab^{-1} \in H_1$ and H_1 is a subgroup of G_1. If H_2 is normal in G_2, we must show that $g^{-1}hg \in H_1$ for $h \in H_1$ and $g \in G_1$. But

$$\phi(g^{-1}hg) = [\phi(g)]^{-1}\phi(h)\phi(g) \in H_2,$$

since H_2 is a normal subgroup of G_2.Therefore, $g^{-1}hg \in H_1$. ■

Let $\phi : G \to H$ be a group homomorphism and suppose that e is the identity of H. By Proposition 8.11, $\phi^{-1}(\{e\})$ is a subgroup of G. This subgroup is called the **kernel** of ϕ and will be denoted by $\ker \phi$. In fact, this subgroup is a normal subgroup of G since the trivial subgroup is normal in H. We state this result in the following theorem, which says that with every homomorphism of groups we can naturally associate a normal subgroup.

Theorem 8.12 *Let $\phi : G \to H$ be a group homomorphism. Then the kernel of ϕ is a normal subgroup of G.*

Example 8.13 Let us examine the homomorphism $\phi : GL_2(\mathbb{R}) \to \mathbb{R}^*$ defined by $A \mapsto \det(A)$. Since 1 is the identity of $\mathbb{R}^*$, the kernel of this homomorphism is all 2×2 matrices having determinant one. That is, $\ker \phi = SL_2(\mathbb{R})$. □

Example 8.14 The kernel of the group homomorphism $\phi : \mathbb{R} \to \mathbb{C}^*$ defined by $\phi(\theta) = \cos\theta + i \sin\theta$ is $\{2\pi n : n \in \mathbb{Z}\}$. Notice that $\ker \phi \cong \mathbb{Z}$. □

Example 8.15 Suppose that we wish to determine all possible homomorphisms ϕ from $\mathbb{Z}_7$ to $\mathbb{Z}_{12}$. Since the kernel of ϕ must be a subgroup of $\mathbb{Z}_7$, there are only two possible kernels, $\{0\}$ and all of $\mathbb{Z}_7$. The image of a subgroup of $\mathbb{Z}_7$ must be a subgroup of $\mathbb{Z}_{12}$. Hence, there is no injective homomorphism; otherwise, $\mathbb{Z}_{12}$ would have a subgroup of order 7, which is impossible. Consequently, the only possible homomorphism from $\mathbb{Z}_7$ to $\mathbb{Z}_{12}$ is the one mapping all elements to zero. □

Example 8.16 Let G be a group. Suppose that $g \in G$ and ϕ is the homomorphism from $\mathbb{Z}$ to G given by $\phi(n) = g^n$. If the order of g is infinite, then the kernel of this homomorphism is $\{0\}$ since ϕ maps $\mathbb{Z}$ onto the cyclic subgroup of G generated by g. However, if the order of g is finite, say n, then the kernel of ϕ is $n\mathbb{Z}$. □

Activity 8.2 Which of the following maps are homomorphisms? If the map is a homomorphism, what is the kernel?

1. $\phi : \mathbb{R}^* \to GL_2(\mathbb{R})$ defined by

$$\phi(a) = \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$$

2. $\phi : \mathbb{R} \to GL_2(\mathbb{R})$ defined by

$$\phi(a) = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}$$

3. $\phi : GL_2(\mathbb{R}) \to \mathbb{R}$ defined by

$$\phi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = a + d$$

4. $\phi : GL_2(\mathbb{R}) \to \mathbb{R}^*$ defined by

$$\phi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = ad - bc$$

5. $\phi : \mathbb{M}_2(\mathbb{R}) \to \mathbb{R}$ defined by

$$\phi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = b,$$

   where $\mathbb{M}_2(\mathbb{R})$ is the additive group of 2×2 matrices with entries in $\mathbb{R}$.

Reading Questions

1. What is the difference between a *homomorphism* and an *isomorphism*?

2. What do we mean by the **kernel** of a homomorphism? What does this say about the kernel of any *isomorphism*?

3. What can you say about the image of a subgroup under a homomorphism? That is, what can you say about $\phi(H)$ for a subgroup H of G and homomorphism $\phi : G \to G'$?

Exercises

1. Prove that $\det(AB) = \det(A)\det(B)$ for $A, B \in GL_2(\mathbb{R})$. This shows that the determinant is a homomorphism from $GL_2(\mathbb{R})$ to $\mathbb{R}^*$.

2. Let A be an $m \times n$ matrix. Show that matrix multiplication, $x \mapsto Ax$, defines a homomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^m$.

3. Let $\phi : \mathbb{Z} \to \mathbb{Z}$ be given by $\phi(n) = 7n$. Prove that ϕ is a group homomorphism. Find the kernel and the image of ϕ.

Hint. Since $\phi(m + n) = 7(m + n) = 7m + 7n = \phi(m) + \phi(n)$, ϕ is a homomorphism.

4. Describe all of the homomorphisms from $\mathbb{Z}$ to $\mathbb{Z}_{12}$.

5. If G is an abelian group and $n \in \mathbb{N}$, show that $\phi : G \to G$ defined by $g \mapsto g^n$ is a group homomorphism.

6. If $\phi : G \to H$ is a group homomorphism and G is abelian, prove that $\phi(G)$ is also abelian.

Hint. Let $a, b \in G$. Then $\phi(a)\phi(b) = \phi(ab) = \phi(ba) = \phi(b)\phi(a)$.

7. If $\phi : G \to H$ is a group homomorphism and G is cyclic, prove that $\phi(G)$ is also cyclic.

8. Show that a homomorphism defined on a cyclic group is completely determined by its action on the generator of the group.

9. If a group G has exactly one subgroup H of order k, prove that H is normal in G.

10. Let $\phi : G \to H$ be a group homomorphism. Show that ϕ is one-to-one if and only if $\phi^{-1}(e) = \{e\}$.

11. Given a homomorphism $\phi : G \to H$ define a relation $\sim$ on G by $a \sim b$ if $\phi(a) = \phi(b)$ for $a, b \in G$. Show this relation is an equivalence relation and describe the equivalence classes.

8.3 The Isomorphism Theorems

Although it is not evident at first, factor groups correspond exactly to homomorphic images, and we can use factor groups to study homomorphisms. We already know that with every group homomorphism $\phi : G \to H$ we can associate a normal subgroup of G, $\ker \phi$. The converse is also true; that is, every normal subgroup of a group G gives rise to homomorphism of groups.

Let H be a normal subgroup of G. Define the **natural** or **canonical homomorphism**

$$\phi : G \to G/H$$

by

$$\phi(g) = gH.$$

This is indeed a homomorphism, since

$$\phi(g_1 g_2) = g_1 g_2 H = g_1 H g_2 H = \phi(g_1)\phi(g_2).$$

The kernel of this homomorphism is H. The following theorems describe the relationships between group homomorphisms, normal subgroups, and factor groups.

Theorem 8.17 First Isomorphism Theorem. *If $\psi : G \to H$ is a group homomorphism with $K = \ker \psi$, then K is normal in G. Let $\phi : G \to G/K$ be the canonical homomorphism. Then there exists a unique isomorphism $\eta : G/K \to \psi(G)$ such that $\psi = \eta\phi$.*

Proof. We already know that K is normal in G. Define $\eta : G/K \to \psi(G)$ by $\eta(gK) = \psi(g)$. We first show that η is a well-defined map. If $g_1 K = g_2 K$, then for some $k \in K$, $g_1 k = g_2$; consequently,

$$\eta(g_1 K) = \psi(g_1) = \psi(g_1)\psi(k) = \psi(g_1 k) = \psi(g_2) = \eta(g_2 K).$$

Thus, η does not depend on the choice of coset representatives and the map

$\eta : G/K \to \psi(G)$ is uniquely defined since $\psi = \eta\phi$. We must also show that η is a homomorphism. Indeed,

$$\begin{aligned}\eta(g_1 K g_2 K) &= \eta(g_1 g_2 K) \\ &= \psi(g_1 g_2) \\ &= \psi(g_1)\psi(g_2) \\ &= \eta(g_1 K)\eta(g_2 K).\end{aligned}$$

Clearly, η is onto $\psi(G)$.To show that η is one-to-one, suppose that $\eta(g_1 K) = \eta(g_2 K)$. Then $\psi(g_1) = \psi(g_2)$. This implies that $\psi(g_1^{-1} g_2) = e$, or $g_1^{-1} g_2$ is in the kernel of ψ; hence, $g_1^{-1} g_2 K = K$; that is, $g_1 K = g_2 K$. ■

Mathematicians often use diagrams called **commutative diagrams** to describe such theorems. The following diagram "commutes" since $\psi = \eta\phi$.

$$\begin{tikzcd} G \arrow[rr, "\psi"] \arrow[dr, "\phi"'] & & H \\ & G/K \arrow[ur, "\eta"'] & \end{tikzcd}$$

Example 8.18 Let G be a cyclic group with generator g. Define a map $\phi : \mathbb{Z} \to G$ by $n \mapsto g^n$. This map is a surjective homomorphism since

$$\phi(m+n) = g^{m+n} = g^m g^n = \phi(m)\phi(n).$$

Clearly ϕ is onto. If $|g| = m$, then $g^m = e$. Hence, $\ker\phi = m\mathbb{Z}$ and $\mathbb{Z}/\ker\phi = \mathbb{Z}/m\mathbb{Z} \cong G$. On the other hand, if the order of g is infinite, then $\ker\phi = 0$ and ϕ is an isomorphism of G and $\mathbb{Z}$. Hence, two cyclic groups are isomorphic exactly when they have the same order. Up to isomorphism, the only cyclic groups are $\mathbb{Z}$ and $\mathbb{Z}_n$. □

Theorem 8.19 Second Isomorphism Theorem. *Let H be a subgroup of a group G (not necessarily normal in G) and N a normal subgroup of G. Then HN is a subgroup of G, $H \cap N$ is a normal subgroup of H, and*

$$H/H \cap N \cong HN/N.$$

Proof. We will first show that $HN = \{hn : h \in H, n \in N\}$ is a subgroup of G. Suppose that $h_1 n_1, h_2 n_2 \in HN$. Since N is normal, $(h_2)^{-1} n_1 h_2 \in N$. So

$$(h_1 n_1)(h_2 n_2) = h_1 h_2 ((h_2)^{-1} n_1 h_2) n_2$$

is in HN. The inverse of $hn \in HN$ is in HN since

$$(hn)^{-1} = n^{-1} h^{-1} = h^{-1}(h n^{-1} h^{-1}).$$

Next, we prove that $H \cap N$ is normal in H. Let $h \in H$ and $n \in H \cap N$. Then $h^{-1} n h \in H$ since each element is in H. Also, $h^{-1} n h \in N$ since N is normal in G; therefore, $h^{-1} n h \in H \cap N$.

Now define a map ϕ from H to HN/N by $h \mapsto hN$. The map ϕ is onto, since any coset $hnN = hN$ is the image of h in H. We also know that ϕ is a homomorphism because

$$\phi(hh') = hh'N = hNh'N = \phi(h)\phi(h').$$

By the First Isomorphism Theorem, the image of ϕ is isomorphic to $H/\ker\phi$; that is,

$$HN/N = \phi(H) \cong H/\ker\phi.$$

Since

$$\ker\phi = \{h \in H : h \in N\} = H \cap N,$$

$HN/N = \phi(H) \cong H/H \cap N$. ■

Theorem 8.20 Correspondence Theorem. *Let N be a normal subgroup of a group G. Then $H \mapsto H/N$ is a one-to-one correspondence between the set of subgroups H containing N and the set of subgroups of G/N. Furthermore, the normal subgroups of G containing N correspond to normal subgroups of G/N.*

Proof. Let H be a subgroup of G containing N. Since N is normal in H, H/N makes is a factor group. Let aN and bN be elements of H/N. Then $(aN)(b^{-1}N) = ab^{-1}N \in H/N$; hence, H/N is a subgroup of G/N.

Let S be a subgroup of G/N. This subgroup is a set of cosets of N. If $H = \{g \in G : gN \in S\}$, then for $h_1, h_2 \in H$, we have that $(h_1N)(h_2N) = h_1h_2N \in S$ and $h_1^{-1}N \in S$. Therefore, H must be a subgroup of G. Clearly, H contains N. Therefore, $S = H/N$. Consequently, the map $H \mapsto H/N$ is onto.

Suppose that H_1 and H_2 are subgroups of G containing N such that $H_1/N = H_2/N$. If $h_1 \in H_1$, then $h_1N \in H_1/N$. Hence, $h_1N = h_2N \subset H_2$ for some h_2 in H_2. However, since N is contained in H_2, we know that $h_1 \in H_2$ or $H_1 \subset H_2$. Similarly, $H_2 \subset H_1$. Since $H_1 = H_2$, the map $H \mapsto H/N$ is one-to-one.

Suppose that H is normal in G and N is a subgroup of H. Then it is easy to verify that the map $G/N \to G/H$ defined by $gN \mapsto gH$ is a homomorphism. The kernel of this homomorphism is H/N, which proves that H/N is normal in G/N.

Conversely, suppose that H/N is normal in G/N. The homomorphism given by

$$G \to G/N \to \frac{G/N}{H/N}$$

has kernel H. Hence, H must be normal in G. ■

Notice that in the course of the proof of Theorem 8.20, we have also proved the following theorem.

Theorem 8.21 Third Isomorphism Theorem. *Let G be a group and N and H be normal subgroups of G with $N \subset H$. Then*

$$G/H \cong \frac{G/N}{H/N}.$$

Example 8.22 By the Third Isomorphism Theorem,

$$\mathbb{Z}/m\mathbb{Z} \cong (\mathbb{Z}/mn\mathbb{Z})/(m\mathbb{Z}/mn\mathbb{Z}).$$

Since $|\mathbb{Z}/mn\mathbb{Z}| = mn$ and $|\mathbb{Z}/m\mathbb{Z}| = m$, we have $|m\mathbb{Z}/mn\mathbb{Z}| = n$. □

8.3.1 Historical Note

One of the foremost problems of group theory has been to classify all simple finite groups. This problem is over a century old and has been solved only in the last few decades of the twentieth century. In a sense, finite simple groups are the building blocks of all finite groups. The first nonabelian simple groups to be discovered were the alternating groups. Galois was the first to prove that A_5 was simple. Later, mathematicians such as C. Jordan and L. E. Dickson found several infinite families of matrix groups that were simple. Other families of simple groups were discovered in the 1950s. At the turn of the century, William Burnside conjectured that all nonabelian simple groups must have even order. In 1963, W. Feit and J. Thompson proved Burnside's conjecture and published their results in the paper "Solvability of Groups of Odd Order," which appeared in the *Pacific Journal of Mathematics*. Their proof, running over 250 pages, gave impetus to a program in the 1960s and 1970s to classify all finite simple groups. Daniel Gorenstein was the organizer of this remarkable effort. One of the last simple groups was the "Monster," discovered by R. Greiss. The Monster, a 196,833 × 196,833 matrix group, is one of the 26 sporadic, or special, simple groups. These sporadic simple groups are groups that fit into no infinite family of simple groups. Some of the sporadic groups play an important role in physics.

8.3.2 Reading Questions

1. If $\phi : G \to H$ is a group homomorphism, what can you say about $\ker \phi$?

2. If N is a normal subgroup of G, must there be a homomorphism that has N as its kernel? If so, what is it?

3. What is the point of showing that $\eta(g_1 K g_2) = \eta(g_1 K)\eta(g_2 K)$ in the proof of Theorem 8.17? Also, why is the proof not done directly after showing this?

8.3.3 Exercises

1. In the group $\mathbb{Z}_{24}$, let $H = \langle 4 \rangle$ and $N = \langle 6 \rangle$.

(a) List the elements in HN (we usually write $H + N$ for these additive groups)and $H \cap N$.

(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 HN/N and $H/(H \cap N)$ described in the proof of the Second Isomorphism Theorem.

2. Prove or disprove: $\mathbb{Q}/\mathbb{Z} \cong \mathbb{Q}$.

3. Let G be a finite group and N a normal subgroup of G. If H is a subgroup of G/N, prove that $\phi^{-1}(H)$ is a subgroup in G of order $|H| \cdot |N|$, where $\phi : G \to G/N$ is the canonical homomorphism.

8.4 Summary and Additional Exercises

8.4.1 The Important Ideas

- A subgroup N of a group G is **normal** in G if $gN = Ng$ for all $g \in G$.
- If N is a normal subgroup of a group G, then the cosets of N in G form a group G/N, under the operation $(aN)(bN) = abN$. The group G/N is called the **factor group** or **quotient group** of G and N.
- A **homomorphism** between groups $(G, \cdot)$ and $(H, \circ)$ is a map $\phi : G \to H$ such that
$$\phi(g_1 \cdot g_2) = \phi(g_1) \circ \phi(g_2)$$
for $g_1, g_2 \in G$.

- If $\phi : G_1 \to G_2$ is a homomorphism of groups, then
  1. If e is the identity of G_1, then $\phi(e)$ is the identity of G_2;
  2. For any element $g \in G_1$, $\phi(g^{-1}) = [\phi(g)]^{-1}$;
  3. If H_1 is a subgroup of G_1, then $\phi(H_1)$ is a subgroup of G_2;
  4. If H_2 is a subgroup of G_2, then $\phi^{-1}(H_2) = \{g \in G_1 : \phi(g) \in H_2\}$ is a subgroup of G_1. Furthermore, if H_2 is normal in G_2, then $\phi^{-1}(H_2)$ is normal in G_1.

- If $\phi : G \to H$ be a group homomorphism and e is the identity of H, then $\ker \phi = \phi^{-1}(\{e\})$ is a normal subgroup of G called the **kernel** of ϕ.

- Let H be a normal subgroup of G. The **natural** or **canonical homomorphism** $\phi : G \to G/H$ is $\phi(g) = gH$.

- **First Isomorphism Theorem** says that if $\psi : G \to H$ is a group homomorphism with $K = \ker \psi$, then K is normal in G. Let $\phi : G \to G/K$ be the canonical homomorphism. Then there exists a unique isomorphism $\eta : G/K \to \psi(G)$ such that $\psi = \eta\phi$.

- **Second Isomorphism Theorem**: Let H be a subgroup of a group G (not necessarily normal in G) and N a normal subgroup of G. Then HN is a subgroup of G, $H \cap N$ is a normal subgroup of H, and

$$H/H \cap N \cong HN/N.$$

- **Third Isomorphism Theorem**: Let G be a group and N and H be normal subgroups of G with $N \subset H$. Then

$$G/H \cong \frac{G/N}{H/N}.$$

- **Correspondence Theorem**: Let N be a normal subgroup of a group G. Then $H \mapsto H/N$ is a one-to-one correspondence between the set of subgroups H containing N and the set of subgroups of G/N. Furthermore, the normal subgroups of G containing N correspond to normal subgroups of G/N.

8.4.2 Exercises

1. Recall that the **center** of a group G is the set

$$Z(G) = \{x \in G : xg = gx \text{ for all } g \in G\}.$$

(a) Calculate the center of S_3.

(b) Calculate the center of $GL_2(\mathbb{R})$.

(c) Show that the center of any group G is a normal subgroup of G.

(d) If $G/Z(G)$ is cyclic, show that G is abelian.

2. Let G be a group and let $G' = \langle aba^{-1}b^{-1} \rangle$; that is, G' is the subgroup of all finite products of elements in G of the form $aba^{-1}b^{-1}$. The subgroup G' is called the **commutator subgroup** of G.

(a) Show that G' is a normal subgroup of G.

(b) Let N be a normal subgroup of G. Prove that G/N is abelian if and only if N contains the commutator subgroup of G.

Hint. (a) Let $g \in G$ and $h \in G'$. If $h = aba^{-1}b^{-1}$, then

$$\begin{aligned} ghg^{-1} &= gaba^{-1}b^{-1}g^{-1} \\ &= (gag^{-1})(gbg^{-1})(ga^{-1}g^{-1})(gb^{-1}g^{-1}) \\ &= (gag^{-1})(gbg^{-1})(gag^{-1})^{-1}(gbg^{-1})^{-1}. \end{aligned}$$

We also need to show that if $h = h_1 \cdots h_n$ with $h_i = a_i b_i a_i^{-1} b_i^{-1}$, then ghg^{-1} is a product of elements of the same type. However, $ghg^{-1} = gh_1 \cdots h_n g^{-1} = (gh_1g^{-1})(gh_2g^{-1}) \cdots (gh_ng^{-1})$.

3. Describe all of the homomorphisms from $\mathbb{Z}_{24}$ to $\mathbb{Z}_{18}$.

Hint. For any homomorphism $\phi : \mathbb{Z}_{24} \to \mathbb{Z}_{18}$, the kernel of ϕ must be a subgroup of $\mathbb{Z}_{24}$ and the image of ϕ must be a subgroup of $\mathbb{Z}_{18}$. Now use the fact that a generator must map to a generator.

4. Let G_1 and G_2 be groups, and let H_1 and H_2 be normal subgroups of G_1 and G_2 respectively. Let $\phi : G_1 \to G_2$ be a homomorphism. Show that ϕ induces a homomorphism $\overline{\phi} : (G_1/H_1) \to (G_2/H_2)$ if $\phi(H_1) \subset H_2$.

5. If H and K are normal subgroups of G and $H \cap K = \{e\}$, prove that G is isomorphic to a subgroup of $G/H \times G/K$.

6. Let $\phi : G_1 \to G_2$ be a surjective group homomorphism. Let H_1 be a normal subgroup of G_1 and suppose that $\phi(H_1) = H_2$. Prove or disprove that $G_1/H_1 \cong G_2/H_2$.

Hint. Find a counterexample.

7. Let $\text{Aut}(G)$ be the set of all automorphisms of G; that is, isomorphisms from G to itself. Prove this set forms a group and is a subgroup of the group of permutations of G; that is, $\text{Aut}(G) \leq S_G$.

8. An **inner automorphism** of G,

$$i_g : G \to G,$$

is defined by the map

$$i_g(x) = gxg^{-1},$$

for $g \in G$. Show that $i_g \in \text{Aut}(G)$.

9. The set of all inner automorphisms is denoted by $\text{Inn}(G)$. Show that $\text{Inn}(G)$ is a subgroup of $\text{Aut}(G)$.

10. Find an automorphism of a group G that is not an inner automorphism.

11. Let G be a group and i_g be an inner automorphism of G, and define a map

$$G \to \text{Aut}(G)$$

by

$$g \mapsto i_g.$$

Prove that this map is a homomorphism with image $\text{Inn}(G)$ and kernel $Z(G)$. Use this result to conclude that

$$G/Z(G) \cong \text{Inn}(G).$$

12. Compute $\text{Aut}(S_3)$ and $\text{Inn}(S_3)$. Do the same thing for D_4.

13. Find all of the homomorphisms $\phi : \mathbb{Z} \to \mathbb{Z}$. What is $\text{Aut}(\mathbb{Z})$?

14. Find all of the automorphisms of $\mathbb{Z}_8$. Prove that $\operatorname{Aut}(\mathbb{Z}_8) \cong U(8)$.

15. For $k \in \mathbb{Z}_n$, define a map $\phi_k : \mathbb{Z}_n \to \mathbb{Z}_n$ by $a \mapsto ka$. Prove that ϕ_k is a homomorphism.

16. Prove that ϕ_k is an isomorphism if and only if k is a generator of $\mathbb{Z}_n$.

17. Show that every automorphism of $\mathbb{Z}_n$ is of the form ϕ_k, where k is a generator of $\mathbb{Z}_n$.

18. Prove that $\psi : U(n) \to \operatorname{Aut}(\mathbb{Z}_n)$ is an isomorphism, where $\psi : k \mapsto \phi_k$.

The Simplicity of A_5. Of special interest are groups with no nontrivial normal subgroups. Such groups are called **simple groups**. Of course, we already have a whole class of examples of simple groups, $\mathbb{Z}_p$, where p is prime. These groups are trivially simple since they have no proper subgroups other than the subgroup consisting solely of the identity. Other examples of simple groups are not so easily found. We can, however, show that the alternating group, A_n, is simple for $n \geq 5$. The following exercises (Exercise Group 8.4.2.19–22) will guide us through a proof that the alternating groups, A_n, are simple for $n \geq 5$.

19. Show that the alternating group A_n is generated by 3-cycles for $n \geq 3$.

Solution. To show that the 3-cycles generate A_n, we need only show that any pair of transpositions can be written as the product of 3-cycles. Since $(ab) = (ba)$, every pair of transpositions must be one of the following:

$$\begin{aligned}(ab)(ab) &= \mathrm{id}\\ (ab)(cd) &= (acb)(acd)\\ (ab)(ac) &= (acb).\end{aligned}$$

20. Let N be a normal subgroup of A_n, where $n \geq 3$. If N contains a 3-cycle, show that $N = A_n$.

Solution. We will first show that A_n is generated by 3-cycles of the specific form (ijk), where i and j are fixed in $\{1, 2, \ldots, n\}$ and we let k vary. Every 3-cycle is the product of 3-cycles of this form, since

$$\begin{aligned}(iaj) &= (ija)^2\\ (iab) &= (ijb)(ija)^2\\ (jab) &= (ijb)^2(ija)\\ (abc) &= (ija)^2(ijc)(ijb)^2(ija).\end{aligned}$$

Now suppose that N is a nontrivial normal subgroup of A_n for $n \geq 3$ such that N contains a 3-cycle of the form (ija). Using the normality of N, we see that

$$[(ij)(ak)](ija)^2[(ij)(ak)]^{-1} = (ijk)$$

is in N. Hence, N must contain all of the 3-cycles (ijk) for $1 \leq k \leq n$. These 3-cycles generate A_n; hence, $N = A_n$.

21. For $n \geq 5$, show that every nontrivial normal subgroup N of A_n contains a 3-cycle. Let σ be an arbitrary element in a normal subgroup N and consider each of the following possible cycle structures for σ.

- σ is a 3-cycle.
- σ is the product of disjoint cycles, $\sigma = \tau(a_1 a_2 \cdots a_r) \in N$, where $r > 3$.

- σ is the product of disjoint cycles, $\sigma = \tau(a_1a_2a_3)(a_4a_5a_6)$.
- $\sigma = \tau(a_1a_2a_3)$, where τ is the product of disjoint 2-cycles.
- $\sigma = \tau(a_1a_2)(a_3a_4)$, where τ is the product of an even number of disjoint 2-cycles.

Solution. Let σ be an arbitrary element in a normal subgroup N. There are several possible cycle structures for σ.

- σ is a 3-cycle.
- σ is the product of disjoint cycles, $\sigma = \tau(a_1a_2\cdots a_r) \in N$, where $r > 3$.
- σ is the product of disjoint cycles, $\sigma = \tau(a_1a_2a_3)(a_4a_5a_6)$.
- $\sigma = \tau(a_1a_2a_3)$, where τ is the product of disjoint 2-cycles.
- $\sigma = \tau(a_1a_2)(a_3a_4)$, where τ is the product of an even number of disjoint 2-cycles.

If σ is a 3-cycle, then we are done. If N contains a product of disjoint cycles, σ, and at least one of these cycles has length greater than 3, say $\sigma = \tau(a_1a_2\cdots a_r)$, then

$$(a_1a_2a_3)\sigma(a_1a_2a_3)^{-1}$$

is in N since N is normal; hence,

$$\sigma^{-1}(a_1a_2a_3)\sigma(a_1a_2a_3)^{-1}$$

is also in N. Since

$$\begin{aligned}\sigma^{-1}(a_1a_2a_3)\sigma(a_1a_2a_3)^{-1} &= \sigma^{-1}(a_1a_2a_3)\sigma(a_1a_3a_2)\\ &= (a_1a_2\cdots a_r)^{-1}\tau^{-1}(a_1a_2a_3)\tau(a_1a_2\cdots a_r)(a_1a_3a_2)\\ &= (a_1a_ra_{r-1}\cdots a_2)(a_1a_2a_3)(a_1a_2\cdots a_r)(a_1a_3a_2)\\ &= (a_1a_3a_r),\end{aligned}$$

N must contain a 3-cycle; hence, $N = A_n$.

Now suppose that N contains a disjoint product of the form

$$\sigma = \tau(a_1a_2a_3)(a_4a_5a_6).$$

Then

$$\sigma^{-1}(a_1a_2a_4)\sigma(a_1a_2a_4)^{-1} \in N$$

since

$$(a_1a_2a_4)\sigma(a_1a_2a_4)^{-1} \in N.$$

So

$$\begin{aligned}\sigma^{-1}(a_1a_2a_4)\sigma(a_1a_2a_4)^{-1} &= [\tau(a_1a_2a_3)(a_4a_5a_6)]^{-1}(a_1a_2a_4)\tau(a_1a_2a_3)(a_4a_5a_6)(a_1a_2a_4)^{-1}\\ &= (a_4a_6a_5)(a_1a_3a_2)\tau^{-1}(a_1a_2a_4)\tau(a_1a_2a_3)(a_4a_5a_6)(a_1a_4a_2)\\ &= (a_4a_6a_5)(a_1a_3a_2)(a_1a_2a_4)(a_1a_2a_3)(a_4a_5a_6)(a_1a_4a_2)\\ &= (a_1a_4a_2a_6a_3).\end{aligned}$$

So N contains a disjoint cycle of length greater than 3, and we can apply the previous case.

Suppose N contains a disjoint product of the form $\sigma = \tau(a_1a_2a_3)$, where τ is the product of disjoint 2-cycles. Since $\sigma \in N$, $\sigma^2 \in N$, and

$$\begin{aligned}\sigma^2 &= \tau(a_1a_2a_3)\tau(a_1a_2a_3)\\ &= (a_1a_3a_2).\end{aligned}$$

So N contains a 3-cycle.

The only remaining possible case is a disjoint product of the form

$$\sigma = \tau(a_1a_2)(a_3a_4),$$

where τ is the product of an even number of disjoint 2-cycles. But

$$\sigma^{-1}(a_1a_2a_3)\sigma(a_1a_2a_3)^{-1}$$

is in N since $(a_1a_2a_3)\sigma(a_1a_2a_3)^{-1}$ is in N; and so

$$\begin{aligned}\sigma^{-1}(a_1a_2a_3)\sigma(a_1a_2a_3)^{-1} &= \tau^{-1}(a_1a_2)(a_3a_4)(a_1a_2a_3)\tau(a_1a_2)(a_3a_4)(a_1a_2a_3)^{-1}\\ &= (a_1a_3)(a_2a_4).\end{aligned}$$

Since $n \geq 5$, we can find $b \in \{1, 2, \ldots, n\}$ such that $b \neq a_1, a_2, a_3, a_4$. Let $\mu = (a_1a_3b)$. Then

$$\mu^{-1}(a_1a_3)(a_2a_4)\mu(a_1a_3)(a_2a_4) \in N$$

appendix

$$\begin{aligned}\mu^{-1}(a_1a_3)(a_2a_4)\mu(a_1a_3)(a_2a_4) &= (a_1ba_3)(a_1a_3)(a_2a_4)(a_1a_3b)(a_1a_3)(a_2a_4)\\ &= (a_1a_3b).\end{aligned}$$

Therefore, N contains a 3-cycle. This completes the proof of the lemma.

22. The alternating group, A_n, is simple for $n \geq 5$.

8.5 Connections to the Secondary Classroom—Quotient Groups

This appendix will examine why quotient groups are important in the secondary classroom.

Part III

Topics in Ring and Field Theory

Chapter 9

Ideals

9.1 Ring Homomorphisms and Ideals

As we saw in Section 3.4, one way to study relationships between groups is to consider the "nice" functions between them: two groups are basically the same if there exists an *isomorphism* between them. Relaxing the requirement that the isomorphism be bijective, we get a *homomorphism*: a function that preserves the operation of the group. We now extend this idea to rings, and consider functions that preserve both ring operations.

A homomorphism between rings preserves the operations of addition and multiplication in the ring. More specifically, if R and S are rings, then a **ring homomorphism** is a map $\phi : R \to S$ satisfying

$$\begin{aligned} \phi(a+b) &= \phi(a) + \phi(b) \\ \phi(ab) &= \phi(a)\phi(b) \end{aligned}$$

for all $a, b \in R$. If $\phi : R \to S$ is a one-to-one and onto homomorphism, then ϕ is called an **isomorphism** of rings.

The set of elements that a ring homomorphism maps to 0 plays a fundamental role in the theory of rings. For any ring homomorphism $\phi : R \to S$, we define the **kernel** of a ring homomorphism to be the set

$$\ker \phi = \{r \in R : \phi(r) = 0\}.$$

Example 9.1 For any integer n we can define a ring homomorphism $\phi : \mathbb{Z} \to \mathbb{Z}_n$ by $a \mapsto a \pmod{n}$. This is indeed a ring homomorphism, since

$$\begin{aligned} \phi(a+b) &= (a+b) \pmod{n} \\ &= a \pmod{n} + b \pmod{n} \\ &= \phi(a) + \phi(b) \end{aligned}$$

and

$$\begin{aligned} \phi(ab) &= ab \pmod{n} \\ &= a \pmod{n} \cdot b \pmod{n} \\ &= \phi(a)\phi(b). \end{aligned}$$

The kernel of the homomorphism ϕ is $n\mathbb{Z}$. □

Example 9.2 Let $C[a, b]$ be the ring of continuous real-valued functions on an interval $[a, b]$ as in Example 4.5. For a fixed $\alpha \in [a, b]$, we can define a ring homomorphism $\phi_\alpha : C[a,b] \to \mathbb{R}$ by $\phi_\alpha(f) = f(\alpha)$. This is a ring homomorphism since

$$\phi_\alpha(f+g) = (f+g)(\alpha) = f(\alpha) + g(\alpha) = \phi_\alpha(f) + \phi_\alpha(g)$$
$$\phi_\alpha(fg) = (fg)(\alpha) = f(\alpha)g(\alpha) = \phi_\alpha(f)\phi_\alpha(g).$$

Ring homomorphisms of the type ϕ_α are called **evaluation homomorphisms**. □

In the next proposition we will examine some fundamental properties of ring homomorphisms. The proof of the proposition is left as an exercise.

Proposition 9.3 *Let $\phi : R \to S$ be a ring homomorphism.*

1. *If R is a commutative ring, then $\phi(R)$ is a commutative ring.*
2. *$\phi(0) = 0$.*
3. *Let 1_R and 1_S be the identities for R and S, respectively. If ϕ is onto, then $\phi(1_R) = 1_S$.*
4. *If R is a field and $\phi(R) \neq \{0\}$, then $\phi(R)$ is a field.*

Part 2 of the proposition above says that the additive identity of a ring is always mapped to the additive identity by a homomorphism. In other words, the additive identity is an element of the kernel of a homomorphism. The homomorphism in Example 9.1 had the set of multiples of n as its kernel. So what sort of set must the kernel of a homomorphism be?

Suppose a and b in R are two elements of the kernel of ϕ. That means that $\phi(a) = 0$ and $\phi(b) = 0$. But since ϕ is a homomorphism, we have

$$\phi(a+b) = \phi(a) + \phi(b) = 0 + 0 = 0$$

and

$$\phi(ab) = \phi(a)\phi(b) = 0 \cdot 0 = 0.$$

So the kernel is closed under addition and multiplication. Further, since $\phi(-a) = -\phi(a)$ (that is, the additive inverse of a must be mapped to the additive inverse of $\phi(a)$), we see that $-a$ is also in the kernel.

All of this is to say that the kernel of a ring homomorphism is necessarily a subring of the domain. But we can say more. Suppose a is in the kernel of ϕ and $r \in R$, but not necessarily in the kernel. We still have that ra and ar are in the kernel:

$$\phi(ra) = \phi(r) \cdot 0 = 0 \text{ and } \phi(ar) = 0 \cdot \phi(r) = 0.$$

So the kernel of a ring homomorphism is a subring that has the additional property that it is closed under multiplication by ring elements outside of itself.

Subrings with this property get a special name. An **ideal** in a ring R is a subring I of R such that if a is in I and r is in R, then both ar and ra are in I; that is, $rI \subset I$ and $Ir \subset I$ for all $r \in R$.

We can summarize what we have defined and shown above as follows.

Proposition 9.4 *The kernel of any ring homomorphism $\phi : R \to S$ is an ideal in R.*

Ideals are interesting to study in their own right.

Example 9.5 Every ring R has at least two ideals, $\{0\}$ and R. These ideals are called the **trivial ideals**. □

Let R be a ring with identity and suppose that I is an ideal in R such that 1 is in I. Since for any $r \in R$, $r1 = r \in I$ by the definition of an ideal, $I = R$.

Example 9.6 If a is any element in a commutative ring R with identity, then the set

$$\langle a \rangle = \{ar : r \in R\}$$

is an ideal in R. Certainly, $\langle a \rangle$ is nonempty since both $0 = a0$ and $a = a1$ are in $\langle a \rangle$. The sum of two elements in $\langle a \rangle$ is again in $\langle a \rangle$ since $ar + ar' = a(r + r')$. The inverse of ar is $-ar = a(-r) \in \langle a \rangle$. Finally, if we multiply an element $ar \in \langle a \rangle$ by an arbitrary element $s \in R$, we have $s(ar) = a(sr)$. Therefore, $\langle a \rangle$ satisfies the definition of an ideal. □

If R is a commutative ring with identity, then an ideal of the form $\langle a \rangle = \{ar : r \in R\}$ is called a **principal ideal**.

Theorem 9.7 *Every ideal in the ring of integers $\mathbb{Z}$ is a principal ideal.*
Proof. The zero ideal $\{0\}$ is a principal ideal since $\langle 0 \rangle = \{0\}$. If I is any nonzero ideal in $\mathbb{Z}$, then I must contain some positive integer m. There exists a least positive integer n in I by the Principle of Well-Ordering. Now let a be any element in I. Using the division algorithm, we know that there exist integers q and r such that

$$a = nq + r$$

where $0 \leq r < n$.This equation tells us that $r = a - nq \in I$, but r must be 0 since n is the least positive element in I. Therefore, $a = nq$ and $I = \langle n \rangle$. ■

Example 9.8 The set $n\mathbb{Z}$ is an ideal in the ring of integers. If na is in $n\mathbb{Z}$ and b is in $\mathbb{Z}$, then nab is in $n\mathbb{Z}$ as required. In fact, by Theorem 9.7, these are the only ideals of $\mathbb{Z}$. □

It should not come as a surprise that all the multiples of n form an ideal. After all, this is really just saying that any multiple of a multiple of n is still a multiple of n. We also saw in Example 9.1 that $n\mathbb{Z}$ is the kernel of a homomorphism, which we know is an ideal. This raises the interesting question of whether *every* ideal is the kernel of some homomorphism. To answer this question, we will need to consider a new ring created by an ideal called a *factor ring*, which we will consider in the next section.

Remark 9.9 In our definition of an ideal we have required that $rI \subset I$ and $Ir \subset I$ for all $r \in R$. Such ideals are sometimes referred to as **two-sided ideals**. We can also consider **one-sided ideals**; that is, we may require only that either $rI \subset I$ or $Ir \subset I$ for $r \in R$ hold but not both. Such ideals are called **left ideals** and **right ideals**, respectively. Of course, in a commutative ring any ideal must be two-sided. In this text we will concentrate on two-sided ideals.

9.2 Factor Rings

We will now consider how to construct a new ring from an ideal. First, an example.

Example 9.10 Start with the ring of integers $\mathbb{Z}$. Consider the (principle) ideal $\langle 3 \rangle$ of all multiples of 5. Now for each element $a \in \mathbb{Z}$, we can create the set

$$a + \langle 3 \rangle = \{a + j : j \in \langle 3 \rangle\}.$$

For example, $2 + \langle 3 \rangle = \{\ldots, -4, -1, 2, 5, 8, \ldots\}$. Note that this is exactly the same set as $8 + \langle 3 \rangle$. In fact, for any $a \in 2 + \langle 3 \rangle$, we have $a + \langle 3 \rangle = 2 + \langle 3 \rangle$.

Of course, there are lots of elements of $\mathbb{Z}$ that are not in $2 + \langle 3 \rangle$. Upon further experimenting, we find that there are exactly 3 different sets we get as $a + \langle 3 \rangle$:

$$0 + \langle 3 \rangle, \quad 1 + \langle 3 \rangle, \text{ and } 2 + \langle 3 \rangle.$$

These sets form a *partition* of the ring (every element of $\mathbb{Z}$ is in exactly one of these sets). It is exactly the same partition that we saw in Example 1.30 by considering the equivalence classes $[0], [1], [2]$ established by the integers modulo 3.

Of course, the integers modulo 3 also form a ring (with operations addition and multiplication mod 3). We can also think of these operations as being performed on the set of three *cosets* $\{0 + \langle 3\rangle, 1 + \langle 3\rangle, 2 + \langle 3\rangle\}$. We define

$$(a + \langle 3\rangle) + (b + \langle 3\rangle) = a + b + \langle 3\rangle$$

and

$$(a + \langle 3\rangle)(b + \langle 3\rangle) = ab + \langle 3\rangle.$$

□

The example above illustrates exactly what happens in general. Given any ring R and any subring J, we can always form a set of **cosets**, each a set of the form

$$a + J = \{a + j : j \in J\}.$$

The cosets always form a partition of R. We write R/J for the *set* of all those cosets (and say R *mod* J). We call R/J a **factor group** because we can define a group operation by

$$(a + J) + (b + J) = a + b + J.$$

(All of these facts are proved in Section 7.1 and Section 8.1.)

If we insist that J is an *ideal*, then we can also define a multiplication operation by

$$(a + J)(b + J) = ab + J$$

which makes R/J into a ring, called a **factor ring** or **quotient ring**.

Theorem 9.11 *Let I be an ideal of R. The factor group R/I is a ring with multiplication defined by*

$$(r + I)(s + I) = rs + I.$$

Proof. We prove that R/I is an abelian group under addition in Theorem 8.4. To prove that R/I is a ring, let $r + I$ and $s + I$ be in R/I. We must show that the product $(r + I)(s + I) = rs + I$ is independent of the choice of coset; that is, if $r' \in r + I$ and $s' \in s + I$, then $r's'$ must be in $rs + I$. Since $r' \in r + I$, there exists an element a in I such that $r' = r + a$. Similarly, there exists a $b \in I$ such that $s' = s + b$. Notice that

$$r's' = (r + a)(s + b) = rs + as + rb + ab$$

and $as + rb + ab \in I$ since I is an ideal; consequently, $r's' \in rs + I$. We will leave as an exercise the verification of the associative law for multiplication and the distributive laws. ■

In the previous section we saw that a natural way that ideals appear is as the kernel of a homomorphism. We asked whether *every* ideal was a kernel. If this is going to be true, the homomorphism would have domain R. But what should the codomain be? It turns out that the factor ring R/I is the right choice.

Theorem 9.12 *Let I be an ideal of R. The map $\phi : R \to R/I$ defined by $\phi(r) = r + I$ is a ring homomorphism of R onto R/I with kernel I.*

Proof. To show that ϕ is a ring homomorphism, let r and s be in R. Then

$$\phi(r) + \phi(s) = (r + I) + (s + I) = r + s + I = \phi(r + s)$$

and

$$\phi(r)\phi(s) = (r+I)(s+I) = rs + I = \phi(rs).$$

To see that ϕ is surjective, note that any coset $r+I$ in R/I is the image of r under ϕ. ■

The map $\phi : R \to R/I$ is often called the **natural** or **canonical homomorphism**.

So given any ideal I of a ring R, we can factor ring R/I, which is the homomorphic image of R under the canonical homomorphism ϕ which has kernel I. Alternatively, given any homomorphism $\phi : R \to S$ there is some kernel I that must be an ideal, and since I is an ideal, we can build a factor ring R/I. How are these two processes related? The next theorem says that they are basically the same, meaning that R/I is isomorphic to S.

Theorem 9.13 First Isomorphism Theorem. *Let $\psi : R \to S$ be a ring homomorphism. Then $\ker\psi$ is an ideal of R. If $\phi : R \to R/\ker\psi$ is the canonical homomorphism, then there exists a unique isomorphism $\eta : R/\ker\psi \to \psi(R)$ such that $\psi = \eta\phi$.*

Proof. Let $K = \ker\psi$. By `[cross-reference to target(s) "homomorph-theorem-first-isomorphism" missing or not unique]` there exists a well-defined group homomorphism $\eta : R/K \to \psi(R)$ defined by $\eta(r+K) = \psi(r)$ for the additive abelian groups R and R/K. To show that this is a ring homomorphism, we need only show that $\eta((r+K)(s+K)) = \eta(r+K)\eta(s+K)$; but

$$\begin{aligned}\eta((r+K)(s+K)) &= \eta(rs+K)\\ &= \psi(rs)\\ &= \psi(r)\psi(s)\\ &= \eta(r+K)\eta(s+K).\end{aligned}$$

■

When studying groups, there is more that can be said about the relationship between group homomorphisms and normal subgroups, and these extend to the context of ring homomorphisms and ideals. We state the analogous theorems for rings here and leave their proofs as exercises.

Theorem 9.14 Second Isomorphism Theorem. *Let I be a subring of a ring R and J an ideal of R. Then $I \cap J$ is an ideal of I and*

$$I/I \cap J \cong (I+J)/J.$$

Theorem 9.15 Third Isomorphism Theorem. *Let R be a ring and I and J be ideals of R where $J \subset I$. Then*

$$R/I \cong \frac{R/J}{I/J}.$$

Theorem 9.16 Correspondence Theorem. *Let I be an ideal of a ring R. Then $S \mapsto S/I$ is a one-to-one correspondence between the set of subrings S containing I and the set of subrings of R/I. Furthermore, the ideals of R containing I correspond to ideals of R/I.*

9.3 Maximal and Prime Ideals

In this particular section we are especially interested in certain ideals of commutative rings. These ideals give us special types of factor rings. More specifically, we would like to characterize those ideals I of a commutative ring R such that R/I is an integral domain or a field.

A proper ideal M of a ring R is a **maximal ideal** of R if the ideal M is not a proper subset of any ideal of R except R itself.That is, M is a maximal ideal if for any ideal I properly containing M, $I = R$. The following theorem completely characterizes maximal ideals for commutative rings with identity in terms of their corresponding factor rings.

Theorem 9.17 *Let R be a commutative ring with identity and M an ideal in R. Then M is a maximal ideal of R if and only if R/M is a field.*

Proof. Let M be a maximal ideal in R. If R is a commutative ring, then R/M must also be a commutative ring. Clearly, $1 + M$ acts as an identity for R/M. We must also show that every nonzero element in R/M has an inverse. If $a + M$ is a nonzero element in R/M, then $a \notin M$. Define I to be the set $\{ra + m : r \in R \text{ and } m \in M\}$. We will show that I is an ideal in R. The set I is nonempty since $0a + 0 = 0$ is in I. If $r_1a + m_1$ and $r_2a + m_2$ are two elements in I, then

$$(r_1a + m_1) - (r_2a + m_2) = (r_1 - r_2)a + (m_1 - m_2)$$

is in I. Also, for any $r \in R$ it is true that $rI \subset I$; hence, I is closed under multiplication and satisfies the necessary conditions to be an ideal. Therefore, by Proposition 4.10 and the definition of an ideal, I is an ideal properly containing M. Since M is a maximal ideal, $I = R$; consequently, by the definition of I there must be an m in M and an element b in R such that $1 = ab + m$. Therefore,

$$1 + M = ab + M = ba + M = (a + M)(b + M).$$

Conversely, suppose that M is an ideal and R/M is a field. Since R/M is a field, it must contain at least two elements: $0 + M = M$ and $1 + M$. Hence, M is a proper ideal of R. Let I be any ideal properly containing M. We need to show that $I = R$. Choose a in I but not in M. Since $a + M$ is a nonzero element in a field, there exists an element $b + M$ in R/M such that $(a + M)(b + M) = ab + M = 1 + M$. Consequently, there exists an element $m \in M$ such that $ab + m = 1$ and 1 is in I. Therefore, $r1 = r \in I$ for all $r \in R$. Consequently, $I = R$. ■

Example 9.18 Let $p\mathbb{Z}$ be an ideal in $\mathbb{Z}$, where p is prime. Then $p\mathbb{Z}$ is a maximal ideal since $\mathbb{Z}/p\mathbb{Z} \cong \mathbb{Z}_p$ is a field. □

A proper ideal P in a commutative ring R is called a **prime ideal** if whenever $ab \in P$, then either $a \in P$ or $b \in P$.[1]

Example 9.19 It is easy to check that the set $P = \{0, 2, 4, 6, 8, 10\}$ is an ideal in $\mathbb{Z}_{12}$. This ideal is prime. In fact, it is a maximal ideal. □

Proposition 9.20 *Let R be a commutative ring with identity 1, where $1 \neq 0$. Then P is a prime ideal in R if and only if R/P is an integral domain.*

Proof. First let us assume that P is an ideal in R and R/P is an integral domain. Suppose that $ab \in P$. If $a + P$ and $b + P$ are two elements of R/P such that $(a + P)(b + P) = 0 + P = P$, then either $a + P = P$ or $b + P = P$. This means that either a is in P or b is in P, which shows that P must be prime.

Conversely, suppose that P is prime and

$$(a + P)(b + P) = ab + P = 0 + P = P.$$

Then $ab \in P$. If $a \notin P$, then b must be in P by the definition of a prime ideal;

[1] It is possible to define prime ideals in a noncommutative ring. See [1] or [3].

hence, $b + P = 0 + P$ and R/P is an integral domain. ■

Example 9.21 Every ideal in $\mathbb{Z}$ is of the form $n\mathbb{Z}$. The factor ring $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$ is an integral domain only when n is prime. It is actually a field. Hence, the nonzero prime ideals in $\mathbb{Z}$ are the ideals $p\mathbb{Z}$, where p is prime. This example really justifies the use of the word "prime" in our definition of prime ideals. □

Since every field is an integral domain, we have the following corollary.

Corollary 9.22 *Every maximal ideal in a commutative ring with identity is also a prime ideal.*

9.3.1 Historical Note

Amalie Emmy Noether, one of the outstanding mathematicians of the twentieth century, was born in Erlangen, Germany in 1882. She was the daughter of Max Noether (1844–1921), a distinguished mathematician at the University of Erlangen. Together with Paul Gordon (1837–1912), Emmy Noether's father strongly influenced her early education. She entered the University of Erlangen at the age of 18. Although women had been admitted to universities in England, France, and Italy for decades, there was great resistance to their presence at universities in Germany. Noether was one of only two women among the university's 986 students. After completing her doctorate under Gordon in 1907, she continued to do research at Erlangen, occasionally lecturing when her father was ill.

Noether went to Göttingen to study in 1916. David Hilbert and Felix Klein tried unsuccessfully to secure her an appointment at Göttingen. Some of the faculty objected to women lecturers, saying, "What will our soldiers think when they return to the university and are expected to learn at the feet of a woman?" Hilbert, annoyed at the question, responded, "Meine Herren, I do not see that the sex of a candidate is an argument against her admission as a Privatdozent. After all, the Senate is not a bathhouse." At the end of World War I, attitudes changed and conditions greatly improved for women. After Noether passed her habilitation examination in 1919, she was given a title and was paid a small sum for her lectures.

In 1922, Noether became a Privatdozent at Göttingen. Over the next 11 years she used axiomatic methods to develop an abstract theory of rings and ideals. Though she was not good at lecturing, Noether was an inspiring teacher. One of her many students was B. L. van der Waerden, author of the first text treating abstract algebra from a modern point of view. Some of the other mathematicians Noether influenced or closely worked with were Alexandroff, Artin, Brauer, Courant, Hasse, Hopf, Pontryagin, von Neumann, and Weyl. One of the high points of her career was an invitation to address the International Congress of Mathematicians in Zurich in 1932. In spite of all the recognition she received from her colleagues, Noether's abilities were never recognized as they should have been during her lifetime. She was never promoted to full professor by the Prussian academic bureaucracy.

In 1933, Noether, who was Jewish, was banned from participation in all academic activities in Germany. She emigrated to the United States, took a position at Bryn Mawr College, and became a member of the Institute for Advanced Study at Princeton. After her death she was eulogized by such notable scientists as Albert Einstein.

9.4 Exercises

1. Which of the following sets are rings with respect to the usual operations of addition and multiplication? If the set is a ring, is it also a field?

(a) $7\mathbb{Z}$

(b) $\mathbb{Z}_{18}$

(c) $\mathbb{Q}(\sqrt{2}\,) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$

(d) $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,) = \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} : a, b, c, d \in \mathbb{Q}\}$

(e) $\mathbb{Z}[\sqrt{3}\,] = \{a + b\sqrt{3} : a, b \in \mathbb{Z}\}$

(f) $R = \{a + b\sqrt[3]{3} : a, b \in \mathbb{Q}\}$

(g) $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z} \text{ and } i^2 = -1\}$

(h) $\mathbb{Q}(\sqrt[3]{3}\,) = \{a + b\sqrt[3]{3} + c\sqrt[3]{9} : a, b, c \in \mathbb{Q}\}$

Hint. (a) $7\mathbb{Z}$ is a ring but not a field; (c) $\mathbb{Q}(\sqrt{2}\,)$ is a field; (f) R is not a ring.

2. Let R be the ring of 2×2 matrices of the form

$$\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix},$$

where $a, b \in \mathbb{R}$. Show that although R is a ring that has no identity, we can find a subring S of R with an identity.

3. List or characterize all of the units in each of the following rings.

(a) $\mathbb{Z}_{10}$

(b) $\mathbb{Z}_{12}$

(c) $\mathbb{Z}_7$

(d) $\mathbb{M}_2(\mathbb{Z})$, the 2×2 matrices with entries in $\mathbb{Z}$

(e) $\mathbb{M}_2(\mathbb{Z}_2)$, the 2×2 matrices with entries in $\mathbb{Z}_2$

Hint. (a) $\{1, 3, 7, 9\}$; (c) $\{1, 2, 3, 4, 5, 6\}$; (e)

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \right\}.$$

4. Find all of the ideals in each of the following rings. Which of these ideals are maximal and which are prime?

(a) $\mathbb{Z}_{18}$

(b) $\mathbb{Z}_{25}$

(c) $\mathbb{M}_2(\mathbb{R})$, the 2×2 matrices with entries in $\mathbb{R}$

(d) $\mathbb{M}_2(\mathbb{Z})$, the 2×2 matrices with entries in $\mathbb{Z}$

(e) $\mathbb{Q}$

Hint. (a) $\{0\}$, $\{0, 9\}$, $\{0, 6, 12\}$, $\{0, 3, 6, 9, 12, 15\}$, $\{0, 2, 4, 6, 8, 10, 12, 14, 16\}$; (c) there are no nontrivial ideals.

5. For each of the following rings R with ideal I, give an addition table and a multiplication table for R/I.

(a) $R = \mathbb{Z}$ and $I = 6\mathbb{Z}$

(b) $R = \mathbb{Z}_{12}$ and $I = \{0, 3, 6, 9\}$

6. Find all homomorphisms $\phi : \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/15\mathbb{Z}$.

7. Prove that $\mathbb{R}$ is not isomorphic to $\mathbb{C}$.

Hint. Assume there is an isomorphism $\phi : \mathbb{C} \to \mathbb{R}$ with $\phi(i) = a$.

8. Prove or disprove: The ring $\mathbb{Q}(\sqrt{2}\,) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ is isomorphic to the ring $\mathbb{Q}(\sqrt{3}\,) = \{a + b\sqrt{3} : a, b \in \mathbb{Q}\}$.

Hint. False. Assume there is an isomorphism $\phi : \mathbb{Q}(\sqrt{2}\,) \to \mathbb{Q}(\sqrt{3}\,)$ such that $\phi(\sqrt{2}\,) = a$.

9. What is the characteristic of the field formed by the set of matrices

$$F = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

with entries in $\mathbb{Z}_2$?

10. Prove that the Gaussian integers, $\mathbb{Z}[i]$, are an integral domain.

11. Prove that $\mathbb{Z}[\sqrt{3}\,i] = \{a + b\sqrt{3}\,i : a, b \in \mathbb{Z}\}$ is an integral domain.

12. If R is a field, show that the only two ideals of R are $\{0\}$ and R itself.

Hint. If $I \neq \{0\}$, show that $1 \in I$.

13. Let a be any element in a ring R with identity. Show that $(-1)a = -a$.

14. Let $\phi : R \to S$ be a ring homomorphism. Prove each of the following statements.

(a) If R is a commutative ring, then $\phi(R)$ is a commutative ring.

(b) $\phi(0) = 0$.

(c) Let 1_R and 1_S be the identities for R and S, respectively. If ϕ is onto, then $\phi(1_R) = 1_S$.

(d) If R is a field and $\phi(R) \neq 0$, then $\phi(R)$ is a field.

Hint. (a) $\phi(a)\phi(b) = \phi(ab) = \phi(ba) = \phi(b)\phi(a)$.

15. Prove that the associative law for multiplication and the distributive laws hold in R/I.

16. Prove the Second Isomorphism Theorem for rings: Let I be a subring of a ring R and J an ideal in R. Then $I \cap J$ is an ideal in I and

$$I/I \cap J \cong I + J/J.$$

17. Prove the Third Isomorphism Theorem for rings: Let R be a ring and I and J be ideals of R, where $J \subset I$. Then

$$R/I \cong \frac{R/J}{I/J}.$$

18. Prove the Correspondence Theorem: Let I be an ideal of a ring R. Then $S \to S/I$ is a one-to-one correspondence between the set of subrings S containing I and the set of subrings of R/I. Furthermore, the ideals of R correspond to ideals of R/I.

19. Let R be a ring and S a subset of R. Show that S is a subring of R if and only if each of the following conditions is satisfied.

(a) $S \neq \emptyset$.

(b) $rs \in S$ for all $r, s \in S$.

(c) $r - s \in S$ for all $r, s \in S$.

20. Let R be a ring with a collection of subrings $\{R_\alpha\}$. Prove that $\bigcap R_\alpha$ is a subring of R. Give an example to show that the union of two subrings is not necessarily a subring.

21. Let $\{I_\alpha\}_{\alpha \in A}$ be a collection of ideals in a ring R. Prove that $\bigcap_{\alpha \in A} I_\alpha$ is also an ideal in R. Give an example to show that if I_1 and I_2 are ideals in R, then $I_1 \cup I_2$ may not be an ideal.

22. Let R be an integral domain. Show that if the only ideals in R are $\{0\}$ and R itself, R must be a field.

Hint. Let $a \in R$ with $a \neq 0$. Then the principal ideal generated by a is R. Thus, there exists a $b \in R$ such that $ab = 1$.

23. Let R be a commutative ring. An element a in R is **nilpotent** if $a^n = 0$ for some positive integer n. Show that the set of all nilpotent elements forms an ideal in R.

24. A ring R is a **Boolean ring** if for every $a \in R$, $a^2 = a$. Show that every Boolean ring is a commutative ring.

Hint. Compute $(a + b)^2$ and $(-ab)^2$.

25. Let R be a ring, where $a^3 = a$ for all $a \in R$. Prove that R must be a commutative ring.

26. Let R be a ring with identity 1_R and S a subring of R with identity 1_S. Prove or disprove that $1_R = 1_S$.

27. If we do not require the identity of a ring to be distinct from 0, we will not have a very interesting mathematical structure. Let R be a ring such that $1 = 0$. Prove that $R = \{0\}$.

28. Let S be a nonempty subset of a ring R. Prove that there is a subring R' of R that contains S.

29. Let R be a ring. Define the **center** of R to be

$$Z(R) = \{a \in R : ar = ra \text{ for all } r \in R\}.$$

Prove that $Z(R)$ is a commutative subring of R.

30. Let p be prime. Prove that

$$\mathbb{Z}_{(p)} = \{a/b : a, b \in \mathbb{Z} \text{ and } \gcd(b, p) = 1\}$$

is a ring. The ring $\mathbb{Z}_{(p)}$ is called the **ring of integers localized at** p.

Hint. Let $a/b, c/d \in \mathbb{Z}_{(p)}$. Then $a/b + c/d = (ad + bc)/bd$ and $(a/b) \cdot (c/d) = (ac)/(bd)$ are both in $\mathbb{Z}_{(p)}$, since $\gcd(bd, p) = 1$.

31. Prove or disprove: Every finite integral domain is isomorphic to $\mathbb{Z}_p$.

32. Let R be a ring with identity.

(a) Let u be a unit in R. Define a map $i_u : R \to R$ by $r \mapsto uru^{-1}$. Prove that i_u is an automorphism of R. Such an automorphism of R is called an inner automorphism of R. Denote the set of all inner automorphisms of R by $\operatorname{Inn}(R)$.

(b) Denote the set of all automorphisms of R by $\text{Aut}(R)$. Prove that $\text{Inn}(R)$ is a normal subgroup of $\text{Aut}(R)$.

(c) Let $U(R)$ be the group of units in R. Prove that the map

$$\phi : U(R) \to \text{Inn}(R)$$

defined by $u \mapsto i_u$ is a homomorphism. Determine the kernel of ϕ.

(d) Compute $\text{Aut}(\mathbb{Z})$,$\text{Inn}(\mathbb{Z})$, and $U(\mathbb{Z})$.

33. Let R and S be arbitrary rings. Show that their Cartesian product is a ring if we define addition and multiplication in $R \times S$ by

(a) $(r, s) + (r', s') = (r + r', s + s')$

(b) $(r, s)(r', s') = (rr', ss')$

34. An element x in a ring is called an **idempotent** if $x^2 = x$. Prove that the only idempotents in an integral domain are 0 and 1. Find a ring with a idempotent x not equal to 0 or 1.

Hint. Suppose that $x^2 = x$ and $x \neq 0$. Since R is an integral domain, $x = 1$. To find a nontrivial idempotent, look in $\mathbb{M}_2(\mathbb{R})$.

35. Let $\gcd(a, n) = d$ and $\gcd(b, d) \neq 1$. Prove that $ax \equiv b \pmod{n}$ does not have a solution.

9.5 References and Suggested Readings

[1] Anderson, F. W. and Fuller, K. R. *Rings and Categories of Modules.* 2nd ed. Springer, New York, 1992.

[2] Atiyah, M. F. and MacDonald, I. G. *Introduction to Commutative Algebra.* Westview Press, Boulder, CO, 1994.

[3] Herstein, I. N. *Noncommutative Rings.* Mathematical Association of America, Washington, DC, 1994.

[4] Kaplansky, I. *Commutative Rings.* Revised edition. University of Chicago Press, Chicago, 1974.

[5] Knuth, D. E. *The Art of Computer Programming: Semi-Numerical Algorithms*, vol. 2. 3rd ed. Addison-Wesley Professional, Boston, 1997.

[6] Lidl, R. and Pilz, G. *Applied Abstract Algebra.* 2nd ed. Springer, New York, 1998. A good source for applications.

[7] Mackiw, G. *Applications of Abstract Algebra.* Wiley, New York, 1985.

[8] McCoy, N. H. *Rings and Ideals.* Carus Monograph Series, No. 8. Mathematical Association of America, Washington, DC, 1968.

[9] McCoy, N. H. *The Theory of Rings.* Chelsea, New York, 1972.

[10] Zariski, O. and Samuel, P. *Commutative Algebra*, vols. I and II. Springer, New York, 1975, 1960.

9.6 Ideals in the Secondary Classroom

This appendix will relate ideals to the secondary classroom.

Chapter 10

Polynomials

Most people are fairly familiar with polynomials by the time they begin to study abstract algebra. When we examine polynomial expressions such as

$$\begin{aligned} p(x) &= x^3 - 3x + 2 \\ q(x) &= 3x^2 - 6x + 5, \end{aligned}$$

we have a pretty good idea of what $p(x) + q(x)$ and $p(x)q(x)$ mean. We just add and multiply polynomials as functions; that is,

$$\begin{aligned} (p + q)(x) &= p(x) + q(x) \\ &= (x^3 - 3x + 2) + (3x^2 - 6x + 5) \\ &= x^3 + 3x^2 - 9x + 7 \end{aligned}$$

and

$$\begin{aligned} (pq)(x) &= p(x)q(x) \\ &= (x^3 - 3x + 2)(3x^2 - 6x + 5) \\ &= 3x^5 - 6x^4 - 4x^3 + 24x^2 - 27x + 10. \end{aligned}$$

It is probably no surprise that polynomials form a ring. In this chapter we shall emphasize the algebraic structure of polynomials by studying polynomial rings. We can prove many results for polynomial rings that are similar to the theorems we proved for the integers. Analogs of prime numbers, the division algorithm, and the Euclidean algorithm all exist for polynomials.

10.1 Polynomial Rings

When you first encounter polynomials, you might see things like

$$x^3 + 7x^2 + 1$$

or maybe something like

$$x^3 + \frac{7}{2}x^2 + \frac{1}{2}.$$

These two examples seem fundamentally different, and the reason is that they have different sorts of *coefficients.* When we study polynomials, we must be very clear what sets of coefficients are allowed. We will see that the coefficinets need to at least come from a communtative ring with identity.

Throughout this chapter we shall assume R is a commutative ring with identity. Any expression of the form

$$f(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where $a_i \in R$ and $a_n \neq 0$, is called a **polynomial over** R with **indeterminate** x. The elements $a_0, a_1, \ldots, a_n$ are called the **coefficients** of f. The coefficient a_n is called the **leading coefficient**. A polynomial is called **monic** if the leading coefficient is 1. If n is the largest nonnegative number for which $a_n \neq 0$, we say that the **degree** of f is n and write $\deg f(x) = n$. If no such n exists—that is, if $f(x) = 0$ is the zero polynomial—then the degree of f is defined to be $-\infty$. We will denote the set of all polynomials with coefficients in a ring R by $R[x]$. Two polynomials are equal exactly when their corresponding coefficients are equal; that is, if we let

$$\begin{aligned} p(x) &= a_0 + a_1 x + \cdots + a_n x^n \\ q(x) &= b_0 + b_1 x + \cdots + b_m x^m, \end{aligned}$$

then $p(x) = q(x)$ if and only if $a_i = b_i$ for all $i \geq 0$.

To show that the set of all polynomials forms a ring, we must first define addition and multiplication. We define the sum of two polynomials as follows. Let

$$\begin{aligned} p(x) &= a_0 + a_1 x + \cdots + a_n x^n \\ q(x) &= b_0 + b_1 x + \cdots + b_m x^m. \end{aligned}$$

Then the sum of $p(x)$ and $q(x)$ is

$$p(x) + q(x) = c_0 + c_1 x + \cdots + c_k x^k,$$

where $c_i = a_i + b_i$ for each i. We define the product of $p(x)$ and $q(x)$ to be

$$p(x)q(x) = c_0 + c_1 x + \cdots + c_{m+n} x^{m+n},$$

where

$$c_i = \sum_{k=0}^{i} a_k b_{i-k} = a_0 b_i + a_1 b_{i-1} + \cdots + a_{i-1} b_1 + a_i b_0$$

for each i. Notice that in each case some of the coefficients may be zero.

Example 10.1 Suppose that

$$p(x) = 3 + 0x + 0x^2 + 2x^3 + 0x^4$$

and

$$q(x) = 2 + 0x - x^2 + 0x^3 + 4x^4$$

are polynomials in $\mathbb{Z}[x]$. If the coefficient of some term in a polynomial is zero, then we usually just omit that term. In this case we would write $p(x) = 3 + 2x^3$ and $q(x) = 2 - x^2 + 4x^4$. The sum of these two polynomials is

$$p(x) + q(x) = 5 - x^2 + 2x^3 + 4x^4.$$

The product,

$$p(x)q(x) = (3 + 2x^3)(2 - x^2 + 4x^4) = 6 - 3x^2 + 4x^3 + 12x^4 - 2x^5 + 8x^7,$$

can be calculated either by determining the c_is in the definition or by simply multiplying polynomials in the same way as we have always done. □

Example 10.2 Let

$$p(x) = 3 + 3x^3 \qquad \text{and} \qquad q(x) = 4 + 4x^2 + 4x^4$$

be polynomials in $\mathbb{Z}_{12}[x]$. The sum of $p(x)$ and $q(x)$ is $7 + 4x^2 + 3x^3 + 4x^4$. The product of the two polynomials is the zero polynomial. This example tells us that we can not expect $R[x]$ to be an integral domain if R is not an integral domain. □

Theorem 10.3 *Let R be a commutative ring with identity. Then $R[x]$ is a commutative ring with identity.*

Proof. Our first task is to show that $R[x]$ is an abelian group under polynomial addition. The zero polynomial, $f(x) = 0$, is the additive identity. Given a polynomial $p(x) = \sum_{i=0}^{n} a_i x^i$, the inverse of $p(x)$ is easily verified to be $-p(x) = \sum_{i=0}^{n}(-a_i)x^i = -\sum_{i=0}^{n} a_i x^i$. Commutativity and associativity follow immediately from the definition of polynomial addition and from the fact that addition in R is both commutative and associative.

To show that polynomial multiplication is associative, let

$$\begin{aligned} p(x) &= \sum_{i=0}^{m} a_i x^i, \\ q(x) &= \sum_{i=0}^{n} b_i x^i, \\ r(x) &= \sum_{i=0}^{p} c_i x^i. \end{aligned}$$

Then

$$\begin{aligned} [p(x)q(x)]r(x) &= \left[\left(\sum_{i=0}^{m} a_i x^i\right)\left(\sum_{i=0}^{n} b_i x^i\right)\right]\left(\sum_{i=0}^{p} c_i x^i\right) \\ &= \left[\sum_{i=0}^{m+n}\left(\sum_{j=0}^{i} a_j b_{i-j}\right) x^i\right]\left(\sum_{i=0}^{p} c_i x^i\right) \\ &= \sum_{i=0}^{m+n+p}\left[\sum_{j=0}^{i}\left(\sum_{k=0}^{j} a_k b_{j-k}\right) c_{i-j}\right] x^i \\ &= \sum_{i=0}^{m+n+p}\left(\sum_{j+k+l=i} a_j b_k c_l\right) x^i \\ &= \sum_{i=0}^{m+n+p}\left[\sum_{j=0}^{i} a_j \left(\sum_{k=0}^{i-j} b_k c_{i-j-k}\right)\right] x^i \\ &= \left(\sum_{i=0}^{m} a_i x^i\right)\left[\sum_{i=0}^{n+p}\left(\sum_{j=0}^{i} b_j c_{i-j}\right) x^i\right] \\ &= \left(\sum_{i=0}^{m} a_i x^i\right)\left[\left(\sum_{i=0}^{n} b_i x^i\right)\left(\sum_{i=0}^{p} c_i x^i\right)\right] \\ &= p(x)[q(x)r(x)]. \end{aligned}$$

The commutativity and distribution properties of polynomial multiplication are proved in a similar manner. We shall leave the proofs of these properties as an exercise. ■

Proposition 10.4 *Let $p(x)$ and $q(x)$ be polynomials in $R[x]$, where R is an integral domain. Then $\deg p(x) + \deg q(x) = \deg(p(x)q(x))$. Furthermore, $R[x]$ is an integral domain.*

Proof. Suppose that we have two nonzero polynomials

$$p(x) = a_m x^m + \cdots + a_1 x + a_0$$

and

$$q(x) = b_n x^n + \cdots + b_1 x + b_0$$

with $a_m \neq 0$ and $b_n \neq 0$. The degrees of $p(x)$ and $q(x)$ are m and n, respectively. The leading term of $p(x)q(x)$ is $a_m b_n x^{m+n}$, which cannot be zero since R is an integral domain; hence, the degree of $p(x)q(x)$ is $m + n$, and $p(x)q(x) \neq 0$. Since $p(x) \neq 0$ and $q(x) \neq 0$ imply that $p(x)q(x) \neq 0$, we know that $R[x]$ must also be an integral domain. ■

We also want to consider polynomials in two or more variables, such as $x^2 - 3xy + 2y^3$. Let R be a ring and suppose that we are given two indeterminates x and y. Certainly we can form the ring $(R[x])[y]$. It is straightforward but perhaps tedious to show that $(R[x])[y] \cong R([y])[x]$. We shall identify these two rings by this isomorphism and simply write $R[x, y]$. The ring $R[x, y]$ is called the **ring of polynomials in two indeterminates x and y with coefficients in R.** We can define the **ring of polynomials in n indeterminates with coefficients in R** similarly. We shall denote this ring by $R[x_1, x_2, \ldots, x_n]$.

Theorem 10.5 *Let R be a commutative ring with identity and $\alpha \in R$. Then we have a ring homomorphism $\phi_\alpha : R[x] \to R$ defined by*

$$\phi_\alpha(p(x)) = p(\alpha) = a_n \alpha^n + \cdots + a_1 \alpha + a_0,$$

where $p(x) = a_n x^n + \cdots + a_1 x + a_0$.

Proof. Let $p(x) = \sum_{i=0}^{n} a_i x^i$ and $q(x) = \sum_{i=0}^{m} b_i x^i$. It is easy to show that $\phi_\alpha(p(x) + q(x)) = \phi_\alpha(p(x)) + \phi_\alpha(q(x))$. To show that multiplication is preserved under the map ϕ_α, observe that

$$\begin{aligned} \phi_\alpha(p(x))\phi_\alpha(q(x)) &= p(\alpha)q(\alpha) \\ &= \left(\sum_{i=0}^{n} a_i \alpha^i\right)\left(\sum_{i=0}^{m} b_i \alpha^i\right) \\ &= \sum_{i=0}^{m+n} \left(\sum_{k=0}^{i} a_k b_{i-k}\right) \alpha^i \\ &= \phi_\alpha(p(x)q(x)). \end{aligned}$$

■

The map $\phi_\alpha : R[x] \to R$ is called the **evaluation homomorphism** at α.

Exercises

1. List out all polynomials of degree 2 in the ring $\mathbb{Z}_2[x]$.

Solution. The coefficients must be either 0 or 1. The coefficient of x^2 must be 1, otherwise the degree would be less than 2. For each of the other terms, we could have coefficient 0 or 1, so there are 4 polynomials:x^2+x+1, $x^2 + x$, $x^2 + 1$, and x^2.

2. How many polynomials of degree 2 are there in $\mathbb{Z}_3[x]$? Explain.

Solution. The coefficient of x^2 must be a 1 or 2. The coefficient of x

and the constant can be 0, 1, or 2. Thus there are $2 \cdot 3 \cdot 3 = 18$ polynomials of degree 2 in this ring.

3. Are there polynomials of degree 4 in $\mathbb{Z}_3[x]$? Explain.

Solution. Yes! The *coefficients* must be from $\mathbb{Z}_3$. The powers of x are any integer. In particular, there are infinitely many polynomials in $\mathbb{Z}_3[x]$.

4. Consider the two polynomials $p(x) = 5x^2 + 3x + 6$ and $q(x) = 4x^2 - x + 9$.

(a) Compute $p(x) + q(x)$ and $p(x)q(x)$ in the ring $\mathbb{Z}[x]$.

(b) Compute $p(x) + q(x)$ and $p(x)q(x)$ in the ring $\mathbb{R}[x]$.

(c) Compute $p(x) + q(x)$ and $p(x)q(x)$ in the ring $\mathbb{Z}_{12}[x]$.

Solution.

(a) $p(x) + q(x) = 9x^2 + 2x + 15$ and

$$\begin{aligned} p(x)q(x) &= 20x^4 - 5x^3 + 45x^2 + 12x^3 - 3x^2 + 18x + 24x^2 - 6x + 54 \\ &= 20x^4 + 7x^3 + 66x^2 + 12x + 54. \end{aligned}$$

(b) We will get the same thing as in the previous part.

(c) We can take our same calculations from (a) and reduce the results mod 12. Or we could have reduced as we went (thanks to the homomorphism property). So we get $p(x) + q(x) = 9x^2 + 2x + 3$ and $p(x)q(x) = 8x^4 + 7x^3 + 6x^2 + 6$.

Checkpoint 10.6 Suppose $p(x)$ and $q(x)$ are polynomials in $R[x]$ where R is a commutative ring with unity. We want to understand how the degree of $p(x) + q(x)$ and $p(x)q(x)$ relate to the degrees of $p(x)$ and $q(x)$.

1. Give an example of polynomials $p(x)$ and $q(x)$ in $\mathbb{Z}[x]$ so that $\deg(p(x) + q(x)) < \max(\deg p(x), \deg q(x))$.

2. Give an example of polynomials $p(x)$ and $q(x)$ in $\mathbb{Z}_{12}[x]$ so that $\deg(p(x)q(x)) < \deg p(x) + \deg q(x)$.

3. Can you repeat the previous part if $p(x)$ and $q(x)$ were polynomials in $\mathbb{Z}_{11}[x]$? Explain.

4. What can you say in general? In particular, in what situations can you claim that $\deg(p(x) + q(x)) = \max(\deg p(x), \deg q(x))$ and in what situations can you claim that $\deg(p(x)q(x)) = \deg p(x) + \deg q(x)$?

10.2 The Division Algorithm

Recall that the division algorithm for integers (Theorem 2.9) says that if a and b are integers with $b > 0$, then there exist unique integers q and r such that $a = bq + r$, where $0 \leq r < b$. The algorithm by which q and r are found is just long division. A similar theorem exists for polynomials. The division algorithm for polynomials has several important consequences. Since its proof is very similar to the corresponding proof for integers, it is worthwhile to review Theorem 2.9 at this point.

Theorem 10.7 Division Algorithm. *Let $f(x)$ and $g(x)$ be polynomials in $F[x]$, where F is a field and $g(x)$ is a nonzero polynomial. Then there exist* *unique polynomials $q(x), r(x) \in F[x]$ such that*

$$f(x) = g(x)q(x) + r(x),$$

where either $\deg r(x) < \deg g(x)$ *or $r(x)$ is the zero polynomial.*

Proof. We will first consider the existence of $q(x)$ and $r(x)$. If $f(x)$ is the zero polynomial, then

$$0 = 0 \cdot g(x) + 0;$$

hence, both q and r must also be the zero polynomial. Now suppose that $f(x)$ is not the zero polynomial and that $\deg f(x) = n$ and $\deg g(x) = m$. If $m > n$, then we can let $q(x) = 0$ and $r(x) = f(x)$. Hence, we may assume that $m \leq n$ and proceed by induction on n. If

$$\begin{aligned} f(x) &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \\ g(x) &= b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0 \end{aligned}$$

the polynomial

$$f'(x) = f(x) - \frac{a_n}{b_m} x^{n-m} g(x)$$

has degree less than n or is the zero polynomial. By induction, there exist polynomials $q'(x)$ and $r(x)$ such that

$$f'(x) = q'(x)g(x) + r(x),$$

where $r(x) = 0$ or the degree of $r(x)$ is less than the degree of $g(x)$. Now let

$$q(x) = q'(x) + \frac{a_n}{b_m} x^{n-m}.$$

Then

$$f(x) = g(x)q(x) + r(x),$$

with $r(x)$ the zero polynomial or $\deg r(x) < \deg g(x)$.

To show that $q(x)$ and $r(x)$ are unique, suppose that there exist two other polynomials $q_1(x)$ and $r_1(x)$ such that $f(x) = g(x)q_1(x) + r_1(x)$ with $\deg r_1(x) < \deg g(x)$ or $r_1(x) = 0$, so that

$$f(x) = g(x)q(x) + r(x) = g(x)q_1(x) + r_1(x),$$

and

$$g(x)[q(x) - q_1(x)] = r_1(x) - r(x).$$

If $q(x) - q_1(x)$ is not the zero polynomial, then

$$\deg(g(x)[q(x) - q_1(x)]) = \deg(r_1(x) - r(x)) \geq \deg g(x).$$

However, the degrees of both $r(x)$ and $r_1(x)$ are strictly less than the degree of $g(x)$; therefore, $r(x) = r_1(x)$ and $q(x) = q_1(x)$. ■

Example 10.8 The division algorithm merely formalizes long division of polynomials, a task we have been familiar with since high school. For example, suppose that we divide $x^3 - x^2 + 2x - 3$ by $x - 2$.

$$\begin{array}{rrrrrrrrrr}
 & & & x^2 & + & x & + & 4 & & \\ \cline{4-10}
x & - & 2 & \multicolumn{1}{|r}{x^3} & - & x^2 & + & 2x & - & 3 \\
 & & & x^3 & - & 2x^2 & & & & \\ \cline{4-10}
 & & & & & x^2 & + & 2x & - & 3 \\
 & & & & & x^2 & - & 2x & & \\ \cline{6-10}
 & & & & & & & 4x & - & 3 \\
 & & & & & & & 4x & - & 8 \\ \cline{8-10}
 & & & & & & & & & 5
\end{array}$$

Hence, $x^3 - x^2 + 2x - 3 = (x - 2)(x^2 + x + 4) + 5$. □

Let $p(x)$ be a polynomial in $F[x]$ and $\alpha \in F$. We say that α is a **zero** or **root** of $p(x)$ if $p(x)$ is in the kernel of the evaluation homomorphism ϕ_α. All we are really saying here is that α is a zero of $p(x)$ if $p(\alpha) = 0$.

Corollary 10.9 *Let F be a field. An element $\alpha \in F$ is a zero of $p(x) \in F[x]$ if and only if $x - \alpha$ is a factor of $p(x)$ in $F[x]$.*

Proof. Suppose that $\alpha \in F$ and $p(\alpha) = 0$. By the division algorithm, there exist polynomials $q(x)$ and $r(x)$ such that

$$p(x) = (x - \alpha)q(x) + r(x)$$

and the degree of $r(x)$ must be less than the degree of $x - \alpha$. Since the degree of $r(x)$ is less than 1, $r(x) = a$ for $a \in F$; therefore,

$$p(x) = (x - \alpha)q(x) + a.$$

But

$$0 = p(\alpha) = 0 \cdot q(\alpha) + a = a;$$

consequently, $p(x) = (x - \alpha)q(x)$, and $x - \alpha$ is a factor of $p(x)$.

Conversely, suppose that $x - \alpha$ is a factor of $p(x)$; say $p(x) = (x - \alpha)q(x)$. Then $p(\alpha) = 0 \cdot q(\alpha) = 0$. ■

Corollary 10.10 *Let F be a field. A nonzero polynomial $p(x)$ of degree n in $F[x]$ can have at most n distinct zeros in F.*

Proof. We will use induction on the degree of $p(x)$. If $\deg p(x) = 0$, then $p(x)$ is a constant polynomial and has no zeros. Let $\deg p(x) = 1$. Then $p(x) = ax + b$ for some a and b in F. If α_1 and α_2 are zeros of $p(x)$, then $a\alpha_1 + b = a\alpha_2 + b$ or $\alpha_1 = \alpha_2$.

Now assume that $\deg p(x) > 1$. If $p(x)$ does not have a zero in F, then we are done. On the other hand, if α is a zero of $p(x)$, then $p(x) = (x - \alpha)q(x)$ for some $q(x) \in F[x]$ by Corollary 10.9. The degree of $q(x)$ is $n - 1$ by Proposition 10.4. Let β be some other zero of $p(x)$ that is distinct from α. Then $p(\beta) = (\beta - \alpha)q(\beta) = 0$. Since $\alpha \neq \beta$ and F is a field, $q(\beta) = 0$. By our induction hypothesis, $q(x)$ can have at most $n - 1$ zeros in F that are distinct from α. Therefore, $p(x)$ has at most n distinct zeros in F. ■

Let F be a field. A monic polynomial $d(x)$ is a **greatest common divisor** of polynomials $p(x), q(x) \in F[x]$ if $d(x)$ evenly divides both $p(x)$ and $q(x)$; and, if for any other polynomial $d'(x)$ dividing both $p(x)$ and $q(x)$, $d'(x) \mid d(x)$. We write $d(x) = \gcd(p(x), q(x))$. Two polynomials $p(x)$ and $q(x)$ are **relatively prime** if $\gcd(p(x), q(x)) = 1$.

Proposition 10.11 *Let F be a field and suppose that $d(x)$ is a greatest common divisor of two polynomials $p(x)$ and $q(x)$ in $F[x]$. Then there exist polynomials $r(x)$ and $s(x)$ such that*

$$d(x) = r(x)p(x) + s(x)q(x).$$

Furthermore, the greatest common divisor of two polynomials is unique.

Proof. Let $d(x)$ be the monic polynomial of smallest degree in the set

$$S = \{f(x)p(x) + g(x)q(x) : f(x), g(x) \in F[x]\}.$$

We can write $d(x) = r(x)p(x) + s(x)q(x)$ for two polynomials $r(x)$ and $s(x)$ in $F[x]$. We need to show that $d(x)$ divides both $p(x)$ and $q(x)$. We shall first show that $d(x)$ divides $p(x)$. By the division algorithm, there exist polynomials

$a(x)$ and $b(x)$ such that $p(x) = a(x)d(x) + b(x)$, where $b(x)$ is either the zero polynomial or $\deg b(x) < \deg d(x)$. Therefore,

$$\begin{aligned} b(x) &= p(x) - a(x)d(x) \\ &= p(x) - a(x)(r(x)p(x) + s(x)q(x)) \\ &= p(x) - a(x)r(x)p(x) - a(x)s(x)q(x) \\ &= p(x)(1 - a(x)r(x)) + q(x)(-a(x)s(x)) \end{aligned}$$

is a linear combination of $p(x)$ and $q(x)$ and therefore must be in S. However, $b(x)$ must be the zero polynomial since $d(x)$ was chosen to be of smallest degree; consequently, $d(x)$ divides $p(x)$. A symmetric argument shows that $d(x)$ must also divide $q(x)$; hence, $d(x)$ is a common divisor of $p(x)$ and $q(x)$.

To show that $d(x)$ is a greatest common divisor of $p(x)$ and $q(x)$, suppose that $d'(x)$ is another common divisor of $p(x)$ and $q(x)$. We will show that $d'(x) \mid d(x)$. Since $d'(x)$ is a common divisor of $p(x)$ and $q(x)$, there exist polynomials $u(x)$ and $v(x)$ such that $p(x) = u(x)d'(x)$ and $q(x) = v(x)d'(x)$. Therefore,

$$\begin{aligned} d(x) &= r(x)p(x) + s(x)q(x) \\ &= r(x)u(x)d'(x) + s(x)v(x)d'(x) \\ &= d'(x)[r(x)u(x) + s(x)v(x)]. \end{aligned}$$

Since $d'(x) \mid d(x)$, $d(x)$ is a greatest common divisor of $p(x)$ and $q(x)$.

Finally, we must show that the greatest common divisor of $p(x)$ and $q(x)$ is unique. Suppose that $d'(x)$ is another greatest common divisor of $p(x)$ and $q(x)$. We have just shown that there exist polynomials $u(x)$ and $v(x)$ in $F[x]$ such that $d(x) = d'(x)[r(x)u(x) + s(x)v(x)]$. Since

$$\deg d(x) = \deg d'(x) + \deg[r(x)u(x) + s(x)v(x)]$$

and $d(x)$ and $d'(x)$ are both greatest common divisors,$\deg d(x) = \deg d'(x)$. Since $d(x)$ and $d'(x)$ are both monic polynomials of the same degree, it must be the case that $d(x) = d'(x)$. ■

Notice the similarity between the proof of Proposition 10.11 and the proof of Theorem 2.10.

Exercises

1. In $\mathbb{Z}_7[x]$, find the quotient and remainder as in the division algorithm for $a(x) = 5x^3 + 6x^2 - 3x + 4$ and $b(x) = x - 2$. Then verify that $r(x)$ is $a(2)$.

Solution. $5x^3 + 6x^2 - 3x + 4 = (5x^2 + 2x + 1)(x - 2) + 6$.

2. In $\mathbb{Z}_5[x]$, find the quotient and remainder as in the division algorithm for $a(x) = 4x^5 - x^3 + x^2 + 4$ and $b(x) = x^3 - 2$.

Solution. $4x^5 - x^3 + x^2 + 4 = (4x^2 + 4)(x^3 + 3) + 4x^2 + 2$.

3. Suppose the ideal J of $\mathbb{Z}$ is the smallest ideal that contains both 231 and 429. Find a single element $p \in \mathbb{Z}$ such that $J = \langle p \rangle$.

Solution. We have $J = \langle 33 \rangle$. Note that $\langle 3 \rangle$ and $\langle 11 \rangle$ and $\langle 1 \rangle$ are also ideals that contains 231 and 429, but these are larger ideals.

4. Suppose the ideal J of $\mathbb{Q}[x]$ is the smallest ideal that contains both $x^2 - 2x - 3$ and $x^3 - 5x^2 + 6x$. Find a single element $p \in \mathbb{Q}[x]$ such that $J = \langle p \rangle$.

Hint. Factor the polynomials

Solution. The polynomials factor as $(x-1)(x-3)$ and $x(x-2)(x-3)$. Thus the greatest common divisor for the two polynomials is $x-3$ so both are in the ideal $\langle x-3\rangle$.

5. Suppose the ideal J of $\mathbb{Z}_5[x]$ is the smallest ideal that contains both $2x^7+4$ and x^7+1. Find a single element $p \in \mathbb{Z}_5[x]$ such that $J = \langle p\rangle$.

Solution. Note that if the ideal contains these polynomials then it also contains $2x^7+2$ and thus also $2x^7+4-(2x^7+2) = 2$, and then also 1 (since $2\cdot 3 \in J$ too). Thus $J = \langle 1\rangle = \mathbb{Z}_5[x]$.

6. Let J be an ideal that contains polynomials $a(x)$ and $b(x)$ in $F[x]$ for some field F. If $r(x)$ is the remainder when $a(x)$ is divided by $b(x)$, prove that $r(x)$ is also in the ideal J.

Solution. By the division algorithm, $a(x) = q(x)b(x) + r(x)$, which in turn means that $a(x) - q(x)b(x) = r(x)$. But $-q(x)b(x)\ inJ$ (since J absorbs products) and thus so is $a(x) - q(x)b(x) = r(x)$.

7. In the previous problem you proved that if $a(x), b(x) \in J$ and $a(x) = q(x)b(x) + r(x)$, then $r(x) \in J$. If you then apply the division algorithm again, to $b(x)$ and $r(x)$, to find $b(x) = q'(x)r(x) + r'(x)$, you would have that $r'(x) \in J$ as well.

What would happen if you kept doing this? Would it go on forever? If it stops, how would it? And what does this have to do with the GCD?

Solution. We know that the process cannot go on forever, since each time the new remainder is smaller than the previous remainder (the new divisor). Thus eventually this must stop (by having a remainder of zero).

The last non-zero remainder will in fact be the greatest common divisor of $a(x)$ and $b(x)$, since it will be the smallest degree polynomial in the ideal.

8. Does $3x^3 - 4x^2 - x + 4$, as a polynomial in $\mathbb{Z}_5[x]$ have any roots? What does this tell you about whether the polynomial can factor in $\mathbb{Z}_5[x]$?

Solution. There are no roots. Since any factorization would need to include a linear term, this says that the polynomial is irreducible.

9. Find two different factorizations of $x^2 + x + 8$ in $\mathbb{Z}_{10}[x]$

Hint. One factorization is $(x+2)(x+9)$. Are there other ways to multiply to 8 in $\mathbb{Z}_{10}$?

10. Let F be a field and $p(x)$ be a polynomial in $F[x]$. Let $a \in F$ be some element in F. What will the remainder be when you divide $p(x)$ by $x-a$? We saw in class that if a is a root of $p(x)$, then the remainder will be zero. But what if a is not a root?

(a) Try a few examples and conjecture a formula for the remainder. Then use the division algorithm to prove your conjecture.

(b) Restate the result you proved in the language of ideals. In other words, which coset of the ideal $\langle x-a\rangle$ will the remainder belong to?

11. Let F be a field, $a \in F$ and $p(x) \in F[x]$.

(a) Prove that when $p(x)$ is divided by $x-a$, the remainder is the constant $p(a)$.

(b) Prove that for any polynomial $b(x) \in F[x]$, the polynomial $p(x)$ and the remainder when $p(x)$ is divided by $b(x)$ are in the same coset of $F[x]/\langle b(x)\rangle$.

10.3 Irreducible Polynomials

A nonconstant polynomial $f(x) \in F[x]$ is **irreducible** over a field F if $f(x)$ cannot be expressed as a product of two polynomials $g(x)$ and $h(x)$ in $F[x]$, where the degrees of $g(x)$ and $h(x)$ are both smaller than the degree of $f(x)$. Irreducible polynomials function as the "prime numbers" of polynomial rings.

Example 10.12 The polynomial $x^2 - 2 \in \mathbb{Q}[x]$ is irreducible since it cannot be factored any further over the rational numbers. Similarly, $x^2 + 1$ is irreducible over the real numbers. □

Example 10.13 The polynomial $p(x) = x^3 + x^2 + 2$ is irreducible over $\mathbb{Z}_3[x]$. Suppose that this polynomial was reducible over $\mathbb{Z}_3[x]$. By the division algorithm there would have to be a factor of the form $x - a$, where a is some element in $\mathbb{Z}_3[x]$. Hence, it would have to be true that $p(a) = 0$. However,

$$\begin{aligned} p(0) &= 2 \\ p(1) &= 1 \\ p(2) &= 2. \end{aligned}$$

Therefore, $p(x)$ has no zeros in $\mathbb{Z}_3$ and must be irreducible. □

Lemma 10.14 *Let $p(x) \in \mathbb{Q}[x]$. Then*

$$p(x) = \frac{r}{s}(a_0 + a_1 x + \cdots + a_n x^n),$$

where $r, s, a_0, \ldots, a_n$ are integers, the a_i's are relatively prime, and r and s are relatively prime.

Proof. Suppose that

$$p(x) = \frac{b_0}{c_0} + \frac{b_1}{c_1}x + \cdots + \frac{b_n}{c_n}x^n,$$

where the b_i's and the c_i's are integers. We can rewrite $p(x)$ as

$$p(x) = \frac{1}{c_0 \cdots c_n}(d_0 + d_1 x + \cdots + d_n x^n),$$

where $d_0, \ldots, d_n$ are integers. Let d be the greatest common divisor of $d_0, \ldots, d_n$. Then

$$p(x) = \frac{d}{c_0 \cdots c_n}(a_0 + a_1 x + \cdots + a_n x^n),$$

where $d_i = da_i$ and the a_i's are relatively prime. Reducing $d/(c_0 \cdots c_n)$ to its lowest terms, we can write

$$p(x) = \frac{r}{s}(a_0 + a_1 x + \cdots + a_n x^n),$$

where $\gcd(r, s) = 1$. ■

Theorem 10.15 **Gauss's Lemma.** *Let $p(x) \in \mathbb{Z}[x]$ be a monic polynomial such that $p(x)$ factors into a product of two polynomials $\alpha(x)$ and $\beta(x)$ in $\mathbb{Q}[x]$, where the degrees of both $\alpha(x)$ and $\beta(x)$ are less than the degree of $p(x)$. Then $p(x) = a(x)b(x)$, where $a(x)$ and $b(x)$ are monic polynomials in $\mathbb{Z}[x]$ with $\deg \alpha(x) = \deg a(x)$ and $\deg \beta(x) = \deg b(x)$.*

Proof. By Lemma 10.14, we can assume that

$$\begin{aligned}\alpha(x) &= \frac{c_1}{d_1}(a_0 + a_1 x + \cdots + a_m x^m) = \frac{c_1}{d_1}\alpha_1(x)\\ \beta(x) &= \frac{c_2}{d_2}(b_0 + b_1 x + \cdots + b_n x^n) = \frac{c_2}{d_2}\beta_1(x),\end{aligned}$$

where the a_i's are relatively prime and the b_i's are relatively prime. Consequently,

$$p(x) = \alpha(x)\beta(x) = \frac{c_1 c_2}{d_1 d_2}\alpha_1(x)\beta_1(x) = \frac{c}{d}\alpha_1(x)\beta_1(x),$$

where c/d is the product of c_1/d_1 and c_2/d_2 expressed in lowest terms. Hence, $dp(x) = c\alpha_1(x)\beta_1(x)$.

If $d = 1$, then $ca_m b_n = 1$ since $p(x)$ is a monic polynomial. Hence, either $c = 1$ or $c = -1$. If $c = 1$, then either $a_m = b_n = 1$ or $a_m = b_n = -1$. In the first case $p(x) = \alpha_1(x)\beta_1(x)$, where $\alpha_1(x)$ and $\beta_1(x)$ are monic polynomials with $\deg \alpha(x) = \deg \alpha_1(x)$ and $\deg \beta(x) = \deg \beta_1(x)$. In the second case $a(x) = -\alpha_1(x)$ and $b(x) = -\beta_1(x)$ are the correct monic polynomials since $p(x) = (-\alpha_1(x))(-\beta_1(x)) = a(x)b(x)$. The case in which $c = -1$ can be handled similarly.

Now suppose that $d \neq 1$. Since $\gcd(c, d) = 1$, there exists a prime p such that $p \mid d$ and $p \nmid c$. Also, since the coefficients of $\alpha_1(x)$ are relatively prime, there exists a coefficient a_i such that $p \nmid a_i$. Similarly, there exists a coefficient b_j of $\beta_1(x)$ such that $p \nmid b_j$. Let $\alpha_1'(x)$ and $\beta_1'(x)$ be the polynomials in $\mathbb{Z}_p[x]$ obtained by reducing the coefficients of $\alpha_1(x)$ and $\beta_1(x)$ modulo p. Since $p \mid d$, $\alpha_1'(x)\beta_1'(x) = 0$ in $\mathbb{Z}_p[x]$. However, this is impossible since neither $\alpha_1'(x)$ nor $\beta_1'(x)$ is the zero polynomial and $\mathbb{Z}_p[x]$ is an integral domain. Therefore, $d = 1$ and the theorem is proven. ■

Corollary 10.16 *Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be a polynomial with coefficients in $\mathbb{Z}$ and $a_0 \neq 0$. If $p(x)$ has a zero in $\mathbb{Q}$, then $p(x)$ also has a zero α in $\mathbb{Z}$. Furthermore, α divides a_0.*

Proof. Let $p(x)$ have a zero $a \in \mathbb{Q}$. Then $p(x)$ must have a linear factor $x - a$. By Gauss's Lemma, $p(x)$ has a factorization with a linear factor in $\mathbb{Z}[x]$. Hence, for some $\alpha \in \mathbb{Z}$

$$p(x) = (x - \alpha)(x^{n-1} + \cdots - a_0/\alpha).$$

Thus $a_0/\alpha \in \mathbb{Z}$ and so $\alpha \mid a_0$. ■

Example 10.17 Let $p(x) = x^4 - 2x^3 + x + 1$. We shall show that $p(x)$ is irreducible over $\mathbb{Q}[x]$. Assume that $p(x)$ is reducible. Then either $p(x)$ has a linear factor, say $p(x) = (x - \alpha)q(x)$, where $q(x)$ is a polynomial of degree three, or $p(x)$ has two quadratic factors.

If $p(x)$ has a linear factor in $\mathbb{Q}[x]$, then it has a zero in $\mathbb{Z}$. By Corollary 10.16, any zero must divide 1 and therefore must be ± 1; however, $p(1) = 1$ and $p(-1) = 3$. Consequently, we have eliminated the possibility that $p(x)$ has any linear factors.

Therefore, if $p(x)$ is reducible it must factor into two quadratic polynomials, say

$$\begin{aligned}p(x) &= (x^2 + ax + b)(x^2 + cx + d)\\ &= x^4 + (a + c)x^3 + (ac + b + d)x^2 + (ad + bc)x + bd,\end{aligned}$$

where each factor is in $\mathbb{Z}[x]$ by Gauss's Lemma. Hence,

$$\begin{aligned}a + c &= -2\\ ac + b + d &= 0\end{aligned}$$

$$
\begin{aligned}
ad + bc &= 1 \\
bd &= 1.
\end{aligned}
$$

Since $bd = 1$, either $b = d = 1$ or $b = d = -1$. In either case $b = d$ and so

$$ad + bc = b(a + c) = 1.$$

Since $a + c = -2$, we know that $-2b = 1$. This is impossible since b is an integer. Therefore, $p(x)$ must be irreducible over $\mathbb{Q}$. □

Theorem 10.18 Eisenstein's Criterion. *Let p be a prime and suppose that*

$$f(x) = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x].$$

If $p \mid a_i$ for $i = 0, 1, \ldots, n-1$, but $p \nmid a_n$ and $p^2 \nmid a_0$, then $f(x)$ is irreducible over $\mathbb{Q}$.

Proof. By Gauss's Lemma, we need only show that $f(x)$ does not factor into polynomials of lower degree in $\mathbb{Z}[x]$. Let

$$f(x) = (b_r x^r + \cdots + b_0)(c_s x^s + \cdots + c_0)$$

be a factorization in $\mathbb{Z}[x]$, with b_r and c_s not equal to zero and $r, s < n$. Since p^2 does not divide $a_0 = b_0 c_0$, either b_0 or c_0 is not divisible by p. Suppose that $p \nmid b_0$ and $p \mid c_0$. Since $p \nmid a_n$ and $a_n = b_r c_s$, neither b_r nor c_s is divisible by p. Let m be the smallest value of k such that $p \nmid c_k$. Then

$$a_m = b_0 c_m + b_1 c_{m-1} + \cdots + b_m c_0$$

is not divisible by p, since each term on the right-hand side of the equation is divisible by p except for $b_0 c_m$. Therefore, $m = n$ since a_i is divisible by p for $m < n$. Hence, $f(x)$ cannot be factored into polynomials of lower degree and therefore must be irreducible. ■

Example 10.19 The polynomial

$$f(x) = 16x^5 - 9x^4 + 3x^2 + 6x - 21$$

is easily seen to be irreducible over $\mathbb{Q}$ by Eisenstein's Criterion if we let $p = 3$. □

Eisenstein's Criterion is more useful in constructing irreducible polynomials of a certain degree over $\mathbb{Q}$ than in determining the irreducibility of an arbitrary polynomial in $\mathbb{Q}[x]$: given an arbitrary polynomial, it is not very likely that we can apply Eisenstein's Criterion. The real value of Theorem 10.18 is that we now have an easy method of generating irreducible polynomials of any degree.

10.3.1 Ideals in $F[x]$

Let F be a field. Recall that a principal ideal in $F[x]$ is an ideal $\langle p(x) \rangle$ generated by some polynomial $p(x)$; that is,

$$\langle p(x) \rangle = \{p(x)q(x) : q(x) \in F[x]\}.$$

Example 10.20 The polynomial x^2 in $F[x]$ generates the ideal $\langle x^2 \rangle$ consisting of all polynomials with no constant term or term of degree 1. □

Theorem 10.21 *If F is a field, then every ideal in $F[x]$ is a principal ideal.*

Proof. Let I be an ideal of $F[x]$. If I is the zero ideal, the theorem is easily true. Suppose that I is a nontrivial ideal in $F[x]$, and let $p(x) \in I$ be a nonzero element of minimal degree. If $\deg p(x) = 0$, then $p(x)$ is a nonzero constant and 1 must be in I. Since 1 generates all of $F[x]$, $\langle 1 \rangle = I = F[x]$ and I is again a principal ideal.

Now assume that $\deg p(x) \geq 1$ and let $f(x)$ be any element in I. By the division algorithm there exist $q(x)$ and $r(x)$ in $F[x]$ such that $f(x) = p(x)q(x) + r(x)$ and $\deg r(x) < \deg p(x)$. Since $f(x), p(x) \in I$ and I is an ideal, $r(x) = f(x) - p(x)q(x)$ is also in I. However, since we chose $p(x)$ to be of minimal degree, $r(x)$ must be the zero polynomial. Since we can write any element $f(x)$ in I as $p(x)q(x)$ for some $q(x) \in F[x]$, it must be the case that $I = \langle p(x) \rangle$. ■

Example 10.22 It is not the case that every ideal in the ring $F[x, y]$ is a principal ideal. Consider the ideal of $F[x, y]$ generated by the polynomials x and y. This is the ideal of $F[x, y]$ consisting of all polynomials with no constant term. Since both x and y are in the ideal, no single polynomial can generate the entire ideal. □

Theorem 10.23 *Let F be a field and suppose that $p(x) \in F[x]$. Then the ideal generated by $p(x)$ is maximal if and only if $p(x)$ is irreducible.*

Proof. Suppose that $p(x)$ generates a maximal ideal of $F[x]$. Then $\langle p(x) \rangle$ is also a prime ideal of $F[x]$. Since a maximal ideal must be properly contained inside $F[x]$, $p(x)$ cannot be a constant polynomial. Let us assume that $p(x)$ factors into two polynomials of lesser degree, say $p(x) = f(x)g(x)$. Since $\langle p(x) \rangle$ is a prime ideal one of these factors, say $f(x)$, is in $\langle p(x) \rangle$ and therefore be a multiple of $p(x)$. But this would imply that $\langle p(x) \rangle \subset \langle f(x) \rangle$, which is impossible since $\langle p(x) \rangle$ is maximal.

Conversely, suppose that $p(x)$ is irreducible over $F[x]$. Let I be an ideal in $F[x]$ containing $\langle p(x) \rangle$. By Theorem 10.21, I is a principal ideal; hence, $I = \langle f(x) \rangle$ for some $f(x) \in F[x]$. Since $p(x) \in I$, it must be the case that $p(x) = f(x)g(x)$ for some $g(x) \in F[x]$. However, $p(x)$ is irreducible; hence, either $f(x)$ or $g(x)$ is a constant polynomial. If $f(x)$ is constant, then $I = F[x]$ and we are done. If $g(x)$ is constant, then $f(x)$ is a constant multiple of I and $I = \langle p(x) \rangle$. Thus, there are no proper ideals of $F[x]$ that properly contain $\langle p(x) \rangle$. ■

10.3.2 Historical Note

Throughout history, the solution of polynomial equations has been a challenging problem. The Babylonians knew how to solve the equation $ax^2 + bx + c = 0$. Omar Khayyam (1048–1131) devised methods of solving cubic equations through the use of geometric constructions and conic sections. The algebraic solution of the general cubic equation $ax^3 + bx^2 + cx + d = 0$ was not discovered until the sixteenth century. An Italian mathematician, Luca Pacioli (ca. 1445–1509), wrote in *Summa de Arithmetica* that the solution of the cubic was impossible. This was taken as a challenge by the rest of the mathematical community.

Scipione del Ferro (1465–1526), of the University of Bologna, solved the "depressed cubic,"

$$ax^3 + cx + d = 0.$$

He kept his solution an absolute secret. This may seem surprising today, when mathematicians are usually very eager to publish their results, but in the days of the Italian Renaissance secrecy was customary. Academic appointments were not easy to secure and depended on the ability to prevail in public contests. Such challenges could be issued at any time. Consequently, any major new discovery was a valuable weapon in such a contest. If an opponent presented a list of problems to be solved, del Ferro could in turn present a list of depressed cubics. He kept the secret of his discovery throughout his life, passing it on only on his deathbed to his student Antonio Fior (ca. 1506–?).

Although Fior was not the equal of his teacher, he immediately issued a challenge to Niccolo Fontana (1499–1557). Fontana was known as Tartaglia (the Stammerer). As a youth he had suffered a blow from the sword of a French soldier during an attack on his village. He survived the savage wound, but his speech was permanently impaired. Tartaglia sent Fior a list of 30 various mathematical problems; Fior countered by sending Tartaglia a list of 30 depressed cubics. Tartaglia would either solve all 30 of the problems or absolutely fail. After much effort Tartaglia finally succeeded in solving the depressed cubic and defeated Fior, who faded into obscurity.

At this point another mathematician, Gerolamo Cardano (1501–1576), entered the story. Cardano wrote to Tartaglia, begging him for the solution to the depressed cubic. Tartaglia refused several of his requests, then finally revealed the solution to Cardano after the latter swore an oath not to publish the secret or to pass it on to anyone else. Using the knowledge that he had obtained from Tartaglia, Cardano eventually solved the general cubic

$$ax^3 + bx^2 + cx + d = 0.$$

Cardano shared the secret with his student, Ludovico Ferrari (1522–1565), who solved the general quartic equation,

$$ax^4 + bx^3 + cx^2 + dx + e = 0.$$

In 1543, Cardano and Ferrari examined del Ferro's papers and discovered that he had also solved the depressed cubic. Cardano felt that this relieved him of his obligation to Tartaglia, so he proceeded to publish the solutions in *Ars Magna* (1545), in which he gave credit to del Ferro for solving the special case of the cubic. This resulted in a bitter dispute between Cardano and Tartaglia, who published the story of the oath a year later.

10.3.3 Exercises

1. True or false: $3x + 9$ is irreducible in $\mathbb{Q}[x]$.

Solution. True. While $3x + 9 = 3(x + 3)$, we have not written the polynomial as the product of two polynomials of smaller degree. If fact, all degree 1 polynomials in $F[x]$ were F is a field are irreducible because you can't have two degree 0 polynomials multiply to be a degree 1 polynomial.

2. Factor $x + 2$ in $\mathbb{Z}_6[x]$ into two degree 1 polynomials. Does this mean that $x + 2$ is not irreducible?

Solution. We have $x + 2 = (2x + 1)(3x + 2)$. This is strange. Under our definition, the polynomial is still irreducible, but it does now factor into polynomials of positive degree (which would be an equivalent definition if the coefficients came from a field).

3. Factor the polynomial $x^5 - 6x^3 - 6x^2 - 7x - 6$ over $\mathbb{Q}$ into irreducible factors (and explain how you know the irreducible factors are irreducible).

Solution. Using either the corollary to Gauss's lemma or the rational roots theorem, we know the only possible roots are ± 1, ± 2, ± 3, and ± 6. Checking each of these shows that $x = -2$, $x = -1$ and $x = 3$ are roots. After doing long division, we find the polynomial factors as

$$(x - 3)(x + 1)(x + 2)(x^2 + 1).$$

This is completely factored over $\mathbb{Q}$ now. The last quadratic factor is irreducible because if it were to factor, it would factor as linear terms, but it has no roots in $\mathbb{Q}$.

4. Factor the polynomial $x^5 + x^4 + 3x^3 + 3x^2 + 2x + 2$ completely over $\mathbb{Z}_5$.

Solution. We start by checking for roots (there are only 5 possibilities in $\mathbb{Z}_5$). We find that 2, 3, and 4 are roots. Dividing by the corresponding factors gives us

$$(x - 2)(x - 3)(x - 4)(x^2 + 2) = (x + 3)(x + 2)(x + 1)(x^2 + 2).$$

5. Factor $4x^7 + 10x^4 - 15x^2 + 5$ completely over $\mathbb{Q}$ (and explain how you know the irreducible factors are irreducible).

Solution. The polynomial is already factored. It is irreducible over $\mathbb{Q}$ by Eisenstein's criterion using $p = 5$.

6. Give an example of a polynomial with integer coefficients which has no roots in $\mathbb{Q}$ but is *not* irreducible in $\mathbb{Q}[x]$. Explain how you know your example works.

Solution. Here is such a polynomial: $x^4 + 2x^2 + 1 = (x^2 + 1)(x^2 + 1)$. Clearly it is not irreducible, and the only possible roots are $pm1$ neither of which are roots. The point is that because the degree of the polynomial is greater than 3, it could factor as non-linear terms; roots only correspond to linear factors.

7. Give an example of a polynomial with integer coefficients that is irreducible but that you can't use Eisenstein's criterion to prove is irreducible.

Solution. For example, $x^2 + 4x + 8$ must be irreducible, because it is degree 2 and has not roots. However, the only prime that we could use for Eisenstein's criterion would be $p = 2$ and $2^2 = 4$ divides the constant, so Eisenstein's criterion does not apply.

8. Explain how you know each of the following polynomials are irreducible in the given ring. Reference the appropriate theorem when appropriate.

(a) $f(x) = x^3 + 3x^2 + 4x + 5$ in $\mathbb{Q}[x]$.

(b) $g(x) = x^2 + 3x + 5$ in $\mathbb{Z}_7[x]$.

(c) $h(x) = x^5 + 6x^4 + 9x^2 + 3$ in $\mathbb{Q}[x]$.

9. You cannot apply Eisenstein's criterion to the polynomial $p(x) = x^3 - 3x + 1$. However, we can apply it to $p(x + c)$ for a carefully picked value of c.

(a) Prove that for any polynomial $p(x)$ and any constant c, if $p(x + c)$ is irreducible, then so is $p(x)$

(b) For $p(x) = x^3 - 3x + 1$, prove that $p(x + 2)$ is irreducible.

Hint. For part (a), prove the contrapositive.

10.4 Factoring over $\mathbb{C}$ and $\mathbb{R}$

We will now consider how to factor over the larger fields $\mathbb{R}$ and $\mathbb{C}$. It will turn out that even if we only care about factoring polynomials over the real numbers, working in the complex numbers is helpful. Thus we begin by reviewing a bit about the complex numbers and describing some of their group structure.

10.4.1 Multiplicative Group of Complex Numbers

A more complete discussion of the complex numbers and their connection to cyclic groups is discussed in Subsection 5.1.1. Here though, is a short summary of the parts relevant to factoring some polynomials over $\mathbb{C}$.

Each complex number can be written in the form $a + bi$ where a and b are real numbers and i is the imaginary number satisfying $i^2 = -1$. We can add and multiply complex numbers as you would expect (treating them akin to polynomials).

Each complex number $z = a + bi$ has a **complex conjugate** defined to be $\overline{z} = a - bi$. Note that when you multiply a complex number and its conjugate you get a real number:

$$z \cdot \overline{z} = (a + bi)(a - bi) = a^2 + b^2.$$

We can view the complex numbers graphically by plotting them on a Cartisian plane, thinking of the real part (the a in $a + bi$) as the horizontal direction and the imaginary part (the b) as the vertical direction. In other words, we plot the complex number $a + bi$ at the point (a, b).

The reason this is so helpful is that it allows us to represent each complex number in **polar form**: each point in the Cartesian plane has a radius r from the origin and an angle θ measured counterclockwise from the positive horizontal axis. Consider Figure 10.24.

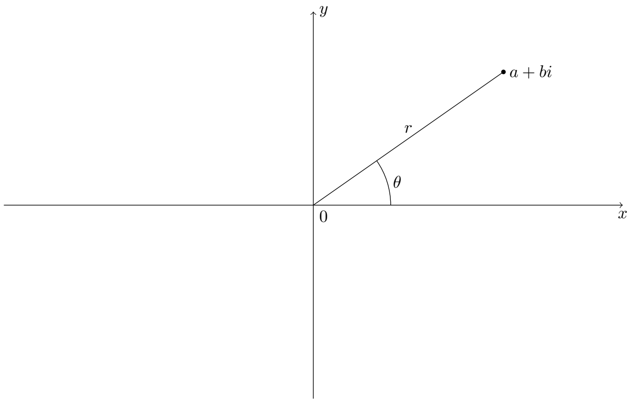


Figure 10.24 Polar coordinates of a complex number

We can see that

$$z = a + bi = r(\cos\theta + i\sin\theta).$$

Hence,

$$r = |z| = \sqrt{a^2 + b^2}$$

and

$$\begin{aligned} a &= r\cos\theta \\ b &= r\sin\theta. \end{aligned}$$

We sometimes abbreviate $r(\cos\theta + i\sin\theta)$ as $r \operatorname{cis}\theta$. To assure that the representation of z is well-defined, we also require that $0° \leq \theta < 360°$. If the measurement is in radians, then $0 \leq \theta < 2\pi$.

The advantage of using the polar form of a complex number makes it becomes very easy to find products and powers of complex numbers. The following propsition says how to multiply two complex numbers:

Proposition 10.25 *Let $z = r\operatorname{cis}\theta$ and $w = s\operatorname{cis}\phi$ be two nonzero complex numbers. Then*

$$zw = rs\operatorname{cis}(\theta + \phi).$$

Example 10.26 If $z = 3\operatorname{cis}(\pi/3)$ and $w = 2\operatorname{cis}(\pi/6)$, then $zw = 6\operatorname{cis}(\pi/2) = 6i$. □

Even more important for the purposes of factoring is DeMoivre's Theorem, which tells us how to compute powers of complex numbers.

Theorem 10.27 DeMoivre. *Let $z = r\operatorname{cis}\theta$ be a nonzero complex number. Then*

$$[r\operatorname{cis}\theta]^n = r^n\operatorname{cis}(n\theta)$$

for $n = 1, 2, \ldots$.

Proof. We will use induction on n. For $n = 1$ the theorem is trivial. Assume that the theorem is true for all k such that $1 \leq k \leq n$. Then

$$\begin{aligned} z^{n+1} &= z^n z \\ &= r^n(\cos n\theta + i\sin n\theta)r(\cos\theta + i\sin\theta) \\ &= r^{n+1}[(\cos n\theta\cos\theta - \sin n\theta\sin\theta) + i(\sin n\theta\cos\theta + \cos n\theta\sin\theta)] \\ &= r^{n+1}[\cos(n\theta + \theta) + i\sin(n\theta + \theta)] \\ &= r^{n+1}[\cos(n+1)\theta + i\sin(n+1)\theta]. \end{aligned}$$

■

Example 10.28 Suppose that $z = 1 + i$ and we wish to compute z^{10}. Rather than computing $(1+i)^{10}$ directly, it is much easier to switch to polar coordinates and calculate z^{10} using DeMoivre's Theorem:

$$\begin{aligned} z^{10} &= (1+i)^{10} \\ &= \left(\sqrt{2}\operatorname{cis}\left(\frac{\pi}{4}\right)\right)^{10} \\ &= (\sqrt{2}\,)^{10}\operatorname{cis}\left(\frac{5\pi}{2}\right) \\ &= 32\operatorname{cis}\left(\frac{\pi}{2}\right) \\ &= 32i. \end{aligned}$$

□

Roots of Unity and Other Numbers. Using DeMoivre's theorem "backward", we can now find nth roots of real numbers. Consider first the *unity* 1. What is $\sqrt[5]{1}$, for example? In the real numbers, we are simply asking for those numbers x such that $x^5 = 1$, and we quickly realize that the only real numbers that satisfies this equation is 1 itself. But there are four other complex numbers that work here!

Consider $z = \operatorname{cis}(2\pi/5)$. By DeMoivre's theorem, we check that

$$z^5 = \operatorname{cis}(5 \cdot 2\pi/5) = \operatorname{cis}(2\pi).$$

But of course, $\operatorname{cis}(2pi) = 1$, so we have identified a complex root of unity. In fact, since $\operatorname{cis}(2k\pi) = 1$ for any integer k, we similarly see that $\operatorname{cis}(4\pi/5)$, $\operatorname{cis}(6\pi/5)$, and $\operatorname{cis}(8\pi/5)$ are also roots of unity. Further, it is easy to see if you plot these points on the complex plane that they are all distinct complex numbers: they fall equally spaced around a unit circle.

These five fifth roots of unity have a nice group structure: they form a cyclic group under multiplication. There is nothing special about 5 here: there are always n distinct nth roots of unity, specifically

$$\{\text{cis}\left(\frac{2k\pi}{n}\right) : n = 0, 1, \ldots, n-1\}.$$

This is a cyclic group of order n, generated by $\text{cis}(2\pi/n)$. That generator (or any other generator of the group) is called a **primitive nth root of unity**.

In the next section we will see how to use these roots of unity to factor polynomials of the form $x^n - k$. The only new piece there will be to decide how to find the nth roots of non-unities. This does not complicate things much at all though, as the next example shows.

Example 10.29 There are five distinct fifth roots of 3. To find those complex numbers z satisfying $z^5 = 3$, we multiply $\sqrt[5]{3}$ by each of the fifth roots of unity:

$$z = \sqrt[5]{3}\,\text{cis}\left(\frac{2\pi}{5}\right), \quad z = \sqrt[5]{3}\,\text{cis}\left(\frac{4\pi}{5}\right), \ldots.$$

□

10.4.2 Factoring With and Without Complex Numbers

Every odd degree polynomial has a root in $\mathbb{R}$ (how do we know?), so no odd degree polynomial (of degree at least 3) can be irreducible over $\mathbb{R}$. What about even degree polynomials? For example, what about $x^2 + 3$?

Well, that polynomial has roots, but they are complex roots, in particular, non-real complex roots.

Let's see how to factor in the complex numbers. It turns out (although hard to prove) that over $\mathbb{C}$, every polynomial factors into linear terms. The *Fundamental Theorem of Algebra* says: Every non-constant polynomial in $\mathbb{C}[x]$ has a complex root.

What does this tell us about the irreducible polynomials in $\mathbb{C}[x]$? They are exactly the degree 1 polynomials. If $a(x)$ is a polynomial for degree greater than 1 in $\mathbb{C}[x]$, then it must have a complex root c, so $x - c$ is a factor. Applying this repeatedly, we find

$$a(x) = k(x - c_1)(x - c_2)\cdots(x - c_n).$$

There will be exactly n roots (although note that the roots might not be distinct).

What about simple polynomials like $x^5 - 1$. Okay, that is not irreducible, since 1 is a root. Does it have any other roots? We use the observations made in the previous section.

Example 10.30 Factor $x^5 - 1$.

Solution. We write $x^5 = 1$. There are lots of ways to write 1 as a complex number though: $1 = \text{cis}(0) = \text{cis}(2\pi) = \text{cis}(4\pi) = \text{cis}(6\pi) = \text{cis}(8\pi)$ and so on. However, we can stop there since we are taking 5th roots.

Now if $x^5 = \text{cis}(2\pi)$, then $x = (x^5)^{\frac{1}{5}} = \text{cis}(2\pi/5)$.

On the other hand, $x^5 = \text{cis}(4\pi)$, so this says that $x = \text{cis}(4\pi/5)$. This is a different complex number than $\text{cis}(2\pi/5)$, since it has a different angle.

If we do this for each of the five representations of 1, we get the following five fifth-roots of 1:

$$1 = \text{cis}(0\pi/5), \quad \text{cis}(2\pi/5), \quad \text{cis}(4\pi/5), \quad \text{cis}(6\pi/5), \quad \text{cis}(8\pi/5).$$

Note that while $\operatorname{cis}(10\pi/5)$ is also a solution to the equation, this is already listed: $\operatorname{cis}(10\pi/5) = \operatorname{cis}(2\pi) = \operatorname{cis}(0) = 1$. Similarly, $\operatorname{cis}(12\pi/5) = \operatorname{cis}(2\pi/5)$, and so on.

If we plot these five complex numbers we will get five equally spaced points on the unit circle. Notice that if we raise each to a power we jump around the circle, landing on the point at $(1, 0)$ on the fifth power, as expected. □

The example above is perhaps a little simplistic. In particular, $r = 1$ here, so we didn't need to do anything with that. In general, if we had $r \operatorname{cis}(\theta)$ and took the nth root, we would get $\sqrt[n]{r} \operatorname{cis}(\theta/n)$.

Example 10.31 Factor $x^3 + 5$ over the complex numbers.

Solution. This is different from the previous example for two reasons: we will have an $r \neq 1$, and we will be finding roots of a negative number.

Write $x^3 = -5 = 5 \operatorname{cis}(\pi) = 5 \operatorname{cis}(3\pi) = 5 \operatorname{cis}(5\pi)$. Then take third roots: $x = \sqrt[3]{5} \operatorname{cis}(\pi/3)$ and $x = \sqrt[3]{5} \operatorname{cis}(3\pi/3)$ and $x = \sqrt[3]{5} \operatorname{cis}(5\pi/3)$.

Plot these three points on the complex plane. One will be on the negative real axis (x-axis), the other two will be vertically aligned in the first and fourth quadrants.

This says that $x^3 + 5 = (x + \sqrt[3]{5})(x - \sqrt[3]{5} \operatorname{cis}(\pi/3))(x - \sqrt[3]{5} \operatorname{cis}(5\pi/3))$. As expected, we can factor the polynomial into linear factors over $\mathbb{C}$. □

Could we factor the polynomial $x^3 + 5$ over $\mathbb{R}$? We could simply use long division to factor out $x + \sqrt[3]{5}$, although that might be messy. Notice though that if we did, then the quotient will be a degree 2 polynomial with real coefficients.

There is a better way. We should be able to get the same polynomial by multiplying the two complex factors from the example above.

Example 10.32 Factor $x^3 + 5$ over the real numbers.

Solution. Pick up where we left off when factoring of $\mathbb{C}$. Then multiply the two factors corresponding to the complex conjugate roots. We find

$$(x - \sqrt[3]{5} \operatorname{cis}(\pi/3))(x - \sqrt[3]{5} \operatorname{cis}(5\pi/3)) - x^2 - \sqrt[3]{5} \operatorname{cis}(5\pi/3)x - \sqrt[3]{5} \operatorname{cis}(\pi/3)x + {\sqrt[3]{5}}^2 \operatorname{cis}(6\pi/3).$$

The last term is just ${\sqrt[3]{5}}^2$. This makes sense if you think about adding the angles.

What should be do with the middle terms? If we write these using trig functions we can see how to add them. Or use parallelograms. □

Will something like this always work? If $a + bi$ is a root of $a(x)$, then $a - bi$ is also a root of $a(x)$ (in $\mathbb{C}$). That is, if r is a root, so is its complex conjugate $\overline{r}$ is as well.

How do we know? Well the function $f(r) = \overline{r}$ is a ring homomorphism from $\mathbb{C} \to \mathbb{C}$. Check this. But then if $a(r) = 0$, apply f to both sides to get $a(\overline{r}) = 0$.

This is great: if $x - r$ is a factor of $a(x)$ in $\mathbb{C}[x]$, then so is $x - \overline{r}$. But notice:

$$(x - r)(x - \overline{r}) = x^2 - 2ax + (a^2 + b^2)$$

is a quadratic polynomial with *real* coefficients.

This shows that every polynomial in $\mathbb{R}[x]$ can be factored into polynomials of degree 1 or 2 in $\mathbb{R}[x]$.

Thus the irreducible polynomials in $\mathbb{R}[x]$ are exactly the linear polynomials and the quadratics with negative discriminant (i.e., $b^2 - 4ac < 0$).

Here is another example.

Example 10.33 Factor $p(x) = x^9 - 2x^8 - 3x^7 - 2x^2 + 4x + 6$ completely over $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.

Solution. Note that the rational roots theorem tells us that the only possible rational roots are ±1, ±2, ±3 and ±6. Plugging these in, we see quickly that -1 and 3 are roots.

After dividing by $(x+1)(x-3)$ we are left with (x^7-2), which is irreducible over $\mathbb{Q}$ by Eisenstein.

Now let's work over $\mathbb{C}$. We can find the roots of x^7-2 by writing $2 = 2\operatorname{cis}(0) = 2\operatorname{cis}(2\pi) = 2\operatorname{cis}(4\pi) = \cdots = 2\operatorname{cis}(12\pi)$.

Taking 7th roots:

$$x = \sqrt[7]{2}, x = \sqrt[7]{2}\operatorname{cis}(2\pi/7), x = \sqrt[7]{2}\operatorname{cis}(4\pi/7), \ldots, x = \sqrt[7]{2}\operatorname{cis}(12\pi/7).$$

This allows us to write down the 9 factors of degree 1 for $p(x)$ over $\mathbb{C}$.

$$x^7 - 2 = (x-\sqrt[7]{2})(x-\sqrt[7]{2}\operatorname{cis}(2\pi/7))(x-\sqrt[7]{2}\operatorname{cis}(4\pi/7))\cdots(x-\sqrt[7]{2}\operatorname{cis}(12\pi/7)).$$

Finally, over $\mathbb{R}$ we can multiply out each pair of complex conjugate factors. To find them, look at the unit circle. This gives:

$$p(x) = (x+1)(x-3)(x-\sqrt[7]{2})(x^2-2^{8/7}\cos(2\pi/7)x+2^{2/7})\cdot$$
$$\cdot(x^2-2^{8/7}\cos(4\pi/7)x+2^{2/7})(x^2-2^{8/7}\cos(6\pi/7)x+2^{2/7})$$

□

10.4.3 Exercises

1. Evaluate each of the following.

(a) $(3-2i)+(5i-6)$

(b) $(4-5i)-\overline{(4i-4)}$

(c) $(5-4i)(7+2i)$

(d) $(9-i)\overline{(9-i)}$

(e) i^{45}

(f) $(1+i)+\overline{(1+i)}$

Hint. (a) $-3+3i$; (c) $43-18i$; (e) i

2. Convert the following complex numbers to the form $a+bi$.

(a) $2\operatorname{cis}(\pi/6)$

(b) $5\operatorname{cis}(9\pi/4)$

(c) $3\operatorname{cis}(\pi)$

(d) $\frac{1}{2}\operatorname{cis}(7\pi/4)$

Hint. (a) $\sqrt{3}+i$; (c) -3.

3. Change the following complex numbers to polar representation.

(a) $1-i$

(b) -5

(c) $2+2i$

(d) $\sqrt{3}+i$

(e) $-3i$

(f) $2i+2\sqrt{3}$

Hint. (a) $\sqrt{2}\operatorname{cis}(7\pi/4)$; (c) $2\sqrt{2}\operatorname{cis}(\pi/4)$; (e) $3\operatorname{cis}(3\pi/2)$.

4. Calculate each of the following expressions.

(a) $(1+i)^{-1}$

(b) $(1-i)^6$

(c) $(\sqrt{3}+i)^5$

(d) $(-i)^{10}$

(e) $((1-i)/2)^4$

(f) $(-\sqrt{2}-\sqrt{2}\,i)^{12}$

(g) $(-2+2i)^{-5}$

Hint. (a) $(1-i)/2$; (c) $16(i-\sqrt{3})$; (e) $-1/4$.

5. Prove that the function $\phi : \mathbb{C} \to \mathbb{C}$ given by $\phi(z) = \bar{z}$ is a ring homomorphism. Here $\bar{z} = a - bi$ is the complex conjugate of $z = a + bi$

Solution. Consider the two complex numbers $a+bi$ and $c+di$. We have

$$\phi(a+bi)+\phi(c+di) = a-bi+c-di = a+c-(b+d)i = \phi(a+c+(b+d)i) = \phi((a+bi)+(c+di))$$

and

$$\phi(a+bi)\phi(c+di) = (a-bi)(c-di) = ac+bd-(cb+ad)i = \phi(ac+bd+(cb+ad)i) = \phi((a+bi)(c+di)).$$

6. Let z be a complex number. Prove that the sum $z + \bar{z}$ and product $z \cdot \bar{z}$ or the number with its conjugate are always real numbers.

Solution. For the sum, it is easiest to think of $z = a + bi$, since now $a + bi + a - bi = 2a$. For the product, write the number in polar form: $z = r\,\mathrm{cis}(\theta)$ so $\bar{z} = r\,\mathrm{cis}(-\theta)$. Thus $z \cdot \bar{z} = r\,\mathrm{cis}(\theta) r\,\mathrm{cis}(-\theta) = r^2\,\mathrm{cis}(0) = r^2$.

7. Is $x^7 - 1$ irreducible over $\mathbb{C}$? How many roots should it have? Find all of them. Hint: use the polar form of complex numbers, $r\,\mathrm{cis}(\vartheta)$.

Solution. The polynomial is not irreducible over $\mathbb{C}$ because the only irreducible polynomials with real coefficients over $\mathbb{C}$ are the linear polynomials. It should have exactly 7 roots. If we write $1 = \mathrm{cis}(0) = \mathrm{cis}(2\pi) = \cdots$ and raise each to the $1/7$ we get the roots. They are:

$$\mathrm{cis}(0) = 1, \quad \mathrm{cis}(\frac{2\pi}{7}), \quad \mathrm{cis}(\frac{4\pi}{7}), \quad \mathrm{cis}(\frac{6\pi}{7}), \quad \mathrm{cis}(\frac{8\pi}{7}), \quad \mathrm{cis}(\frac{10\pi}{7}), \quad \mathrm{cis}(\frac{12\pi}{7})$$

8. Factor $x^8 - 5x^7 - 14x^6 + x^2 - 5x - 14$ completely over $\mathbb{Q}$, $\mathbb{C}$ and $\mathbb{R}$

Solution. First, find rational roots. They are -2 and 7. Dividing we get

$$(x+2)(x-7)(x^6+1)$$

The last term has no roots but is still not irreducible. We can use sum of cubes:

$$(x+2)(x-7)(x^2+1)(x^4-x^2+1)$$

The last degree 4 polynomial is irreducible - use the quadratic formula for x^2 to see that there are no quadratic factors over $\mathbb{Q}$. Now let's factor (x^6+1) over $\mathbb{C}$. There are 6 roots, namely

$$\mathrm{cis}(\pi/6), \quad \mathrm{cis}(3\pi/6), \quad \mathrm{cis}(5\pi/6), \quad \mathrm{cis}(7\pi/6), \quad \mathrm{cis}(9\pi/6), \quad \mathrm{cis}(11\pi/6)$$

These are all "nice" angles, so it is easy to convert back to rectangular form. So over $\mathbb{C}$ we have the factorization

$$(x+2)(x-7)(x-\frac{\sqrt{3}}{2}-\frac{1}{2}i)(x-\frac{\sqrt{3}}{2}+\frac{1}{2}i)(x-i)(x+i)(x+\frac{\sqrt{3}}{2}-\frac{1}{2}i)(x+\frac{\sqrt{3}}{2}+\frac{1}{2}i)$$

Finally, to get the factorization over $\mathbb{R}$ we multiply conjugate pairs:

$$(x+2)(x-7)(x^2+1)(x^2-\sqrt{3}+1)(x^2+\sqrt{3}x+1)$$

9. Factor the polynomial $p(x) = x^7 + 2x^6 - 3x - 6$ completely (into irreducible factors) over $\mathbb{Q}$, then over $\mathbb{C}$, and then over $\mathbb{R}$.

Hint. Do the factoring in that order.

10. True or false: $x^4 + 20x^3 + 5x^2 + 10x + 15$ is irreducible in $\mathbb{R}[x]$. Briefly explain.

10.5 Exercises

1. List all of the polynomials of degree 3 or less in $\mathbb{Z}_2[x]$.

2. Compute each of the following.

(a) $(5x^2 + 3x - 4) + (4x^2 - x + 9)$ in $\mathbb{Z}_{12}$

(b) $(5x^2 + 3x - 4)(4x^2 - x + 9)$ in $\mathbb{Z}_{12}$

(c) $(7x^3 + 3x^2 - x) + (6x^2 - 8x + 4)$ in $\mathbb{Z}_9$

(d) $(3x^2 + 2x - 4) + (4x^2 + 2)$ in $\mathbb{Z}_5$

(e) $(3x^2 + 2x - 4)(4x^2 + 2)$ in $\mathbb{Z}_5$

(f) $(5x^2 + 3x - 2)^2$ in $\mathbb{Z}_{12}$

Hint. (a) $9x^2 + 2x + 5$; (b) $8x^4 + 7x^3 + 2x^2 + 7x$.

3. Use the division algorithm to find $q(x)$ and $r(x)$ such that $a(x) = q(x)b(x) + r(x)$ with $\deg r(x) < \deg b(x)$ for each of the following pairs of polynomials.

(a) $a(x) = 5x^3 + 6x^2 - 3x + 4$ and $b(x) = x - 2$ in $\mathbb{Z}_7[x]$

(b) $a(x) = 6x^4 - 2x^3 + x^2 - 3x + 1$ and $b(x) = x^2 + x - 2$ in $\mathbb{Z}_7[x]$

(c) $a(x) = 4x^5 - x^3 + x^2 + 4$ and $b(x) = x^3 - 2$ in $\mathbb{Z}_5[x]$

(d) $a(x) = x^5 + x^3 - x^2 - x$ and $b(x) = x^3 + x$ in $\mathbb{Z}_2[x]$

Hint. (a) $5x^3+6x^2-3x+4 = (5x^2+2x+1)(x-2)+6$; (c) $4x^5-x^3+x^2+4 = (4x^2+4)(x^3+3)+4x^2+2$.

4. Find the greatest common divisor of each of the following pairs $p(x)$ and $q(x)$ of polynomials. If $d(x) = \gcd(p(x), q(x))$, find two polynomials $a(x)$ and $b(x)$ such that $a(x)p(x) + b(x)q(x) = d(x)$.

(a) $p(x) = x^3 - 6x^2 + 14x - 15$ and $q(x) = x^3 - 8x^2 + 21x - 18$, where $p(x), q(x) \in \mathbb{Q}[x]$

(b) $p(x) = x^3 + x^2 - x + 1$ and $q(x) = x^3 + x - 1$, where $p(x), q(x) \in \mathbb{Z}_2[x]$

(c) $p(x) = x^3 + x^2 - 4x + 4$ and $q(x) = x^3 + 3x - 2$, where $p(x), q(x) \in \mathbb{Z}_5[x]$

(d) $p(x) = x^3 - 2x + 4$ and $q(x) = 4x^3 + x + 3$, where $p(x), q(x) \in \mathbb{Q}[x]$

5. Find all of the zeros for each of the following polynomials.

(a) $5x^3 + 4x^2 - x + 9$ in $\mathbb{Z}_{12}$

(b) $3x^3 - 4x^2 - x + 4$ in $\mathbb{Z}_5$

(c) $5x^4 + 2x^2 - 3$ in $\mathbb{Z}_7$

(d) $x^3 + x + 1$ in $\mathbb{Z}_2$

Hint. (a) No zeros in $\mathbb{Z}_{12}$; (c) 3, 4.

6. Find all of the units in $\mathbb{Z}[x]$.

7. Find a unit $p(x)$ in $\mathbb{Z}_4[x]$ such that $\deg p(x) > 1$.

Hint. Look at $(2x + 1)$.

8. Which of the following polynomials are irreducible over $\mathbb{Q}[x]$?

(a) $x^4 - 2x^3 + 2x^2 + x + 4$

(b) $x^4 - 5x^3 + 3x - 2$

(c) $3x^5 - 4x^3 - 6x^2 + 6$

(d) $5x^5 - 6x^4 - 3x^2 + 9x - 15$

Hint. (a) Reducible; (c) irreducible.

9. Find all of the irreducible polynomials of degrees 2 and 3 in $\mathbb{Z}_2[x]$.

10. Give two different factorizations of$x^2 + x + 8$ in $\mathbb{Z}_{10}[x]$.

Hint. One factorization is $x^2 + x + 8 = (x + 2)(x + 9)$.

11. Prove or disprove: There exists a polynomial $p(x)$ in $\mathbb{Z}_6[x]$ of degree n with more than n distinct zeros.

12. If F is a field, show that $F[x_1, \ldots, x_n]$ is an integral domain.

13. Show that the division algorithm does not hold for $\mathbb{Z}[x]$. Why does it fail?

Hint. The integers $\mathbb{Z}$ do not form a field.

14. Prove or disprove: $x^p + a$ is irreducible for any $a \in \mathbb{Z}_p$, where p is prime.

Hint. False.

15. Let $f(x)$ be irreducible in $F[x]$, where F is a field. If $f(x) \mid p(x)q(x)$, prove that either $f(x) \mid p(x)$ or $f(x) \mid q(x)$.

16. Suppose that R and S are isomorphic rings. Prove that $R[x] \cong S[x]$.

Hint. Let $\phi : R \to S$ be an isomorphism. Define $\overline{\phi} : R[x] \to S[x]$ by $\overline{\phi}(a_0 + a_1x + \cdots + a_nx^n) = \phi(a_0) + \phi a_1)x + \cdots + \phi(a_n)x^n$.

17. Let F be a field and $a \in F$. If $p(x) \in F[x]$, show that $p(a)$ is the remainder obtained when $p(x)$ is divided by $x - a$.

18. **The Rational Root Theorem.** Let

$$p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{Z}[x],$$

where $a_n \neq 0$. Prove that if $p(r/s) = 0$, where $\gcd(r, s) = 1$, then $r \mid a_0$ and $s \mid a_n$.

19. Let $\mathbb{Q}^*$ be the multiplicative group of positive rational numbers. Prove that $\mathbb{Q}^*$ is isomorphic to $(\mathbb{Z}[x], +)$.

20. **Cyclotomic Polynomials.** The polynomial

$$\Phi_n(x) = \frac{x^n - 1}{x - 1} = x^{n-1} + x^{n-2} + \cdots + x + 1$$

is called the **cyclotomic polynomial.**. Show that $\Phi_p(x)$ is irreducible over $\mathbb{Q}$ for any prime p.

Hint. The polynomial

$$\Phi_n(x) = \frac{x^n - 1}{x - 1} = x^{n-1} + x^{n-2} + \cdots + x + 1$$

is called the **cyclotomic polynomial.** Show that $\Phi_p(x)$ is irreducible over $\mathbb{Q}$ for any prime p.

21. If F is a field, show that there are infinitely many irreducible polynomials in $F[x]$.

22. Let R be a commutative ring with identity. Prove that multiplication is commutative in $R[x]$.

23. Let R be a commutative ring with identity. Prove that multiplication is distributive in $R[x]$.

24. Show that $x^p - x$ has p distinct zeros in $\mathbb{Z}_p$, for any prime p. Conclude that

$$x^p - x = x(x - 1)(x - 2) \cdots (x - (p - 1)).$$

25. Let F be a field and $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be in $F[x]$. Define $f'(x) = a_1 + 2a_2x + \cdots + na_nx^{n-1}$ to be the **derivative** of $f(x)$.

(a) Prove that

$$(f + g)'(x) = f'(x) + g'(x).$$

Conclude that we can define a homomorphism of abelian groups$D : F[x] \to F[x]$ by $D(f(x)) = f'(x)$.

(b) Calculate the kernel of D if $\operatorname{char} F = 0$.

(c) Calculate the kernel of D if $\operatorname{char} F = p$.

(d) Prove that

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x).$$

(e) Suppose that we can factor a polynomial $f(x) \in F[x]$ into linear factors, say

$$f(x) = a(x - a_1)(x - a_2) \cdots (x - a_n).$$

Prove that $f(x)$ has no repeated factors if and only if $f(x)$ and $f'(x)$ are relatively prime.

26. Let F be a field. Show that $F[x]$ is never a field.

Hint. Find a nontrivial proper ideal in $F[x]$.

27. Let R be an integral domain. Prove that $R[x_1, \ldots, x_n]$ is an integral domain.

28. Let R be a commutative ring with identity. Show that $R[x]$ has a subring R' isomorphic to R.

29. Let $p(x)$ and $q(x)$ be polynomials in $R[x]$, where R is a commutative ring with identity. Prove that $\deg(p(x) + q(x)) \leq \max(\deg p(x), \deg q(x))$.

10.6 Quadratics, Cubic, and Quartic Polynomials in the Secondary Classroom

A standard topic for the secondary classroom is solving quadratic equations and introducing the quadrtic formula. The general quadratic equation

$$ax^2 + bx + c = 0$$

can be solved to obtain

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{10.1}$$

Indeed, we can complete the square to obtain (10.1)

$$\begin{aligned} ax^2 + bx + c &= 0 \\ x^2 + \frac{b}{a}x &= -\frac{c}{a} \\ x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 &= \left(\frac{b}{2a}\right)^2 - \frac{c}{a} \\ \left(x + \frac{b}{2a}\right)^2 &= \frac{b^2 - 4ac}{4a^2} \\ x + \frac{b}{2a} &= \frac{\pm\sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The **discriminant** of the quadratic equation $\Delta = b^2 - 4ac$ gives us a great deal of information about the quadratic, since it determines the nature of the solutions of the equation. If $\Delta > 0$, the equation has two distinct real solutions. If $\Delta = 0$, the equation has a single repeated real root. If $\Delta < 0$, there are two distinct imaginary solutions. By determining the sign of the discriminant alone, we can determine if the graph of any quadratic function intersects the x-axis once, twice, or not at all. We do not need to compute solutions of a quadratic.

We might ask what generalizes to higher degree polynomials such as cubics and quartics. The following exercises will give us some sense of what is going on.

Exercises

1. Show that any cubic equation of the form

$$x^3 + bx^2 + cx + d = 0$$

can be reduced to the form $y^3 + py + q = 0$ by making the substitution $x = y - b/3$.

2. Prove that the cube roots of 1 are given by

$$\begin{aligned} \omega &= \frac{-1 + i\sqrt{3}}{2} \\ \omega^2 &= \frac{-1 - i\sqrt{3}}{2} \\ \omega^3 &= 1. \end{aligned}$$

3. Make the substitution

$$y = z - \frac{p}{3z}$$

for y in the equation $y^3 + py + q = 0$ and obtain two solutions A and B for z^3.

4. Show that the product of the solutions obtained in (4) is $-p^3/27$, deducing that $\sqrt[3]{AB} = -p/3$.

5. Prove that the possible solutions for z in (4) are given by

$$\sqrt[3]{A}, \quad \omega\sqrt[3]{A}, \quad \omega^2\sqrt[3]{A}, \quad \sqrt[3]{B}, \quad \omega\sqrt[3]{B}, \quad \omega^2\sqrt[3]{B}$$

and use this result to show that the three possible solutions for y are

$$\omega^i \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}} + \omega^{2i} \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}},$$

where $i = 0, 1, 2$.

6. The **discriminant** of the cubic equation is

$$\Delta = \frac{p^3}{27} + \frac{q^2}{4}.$$

Show that $y^3 + py + q = 0$

(a) has three real roots, at least two of which are equal, if $\Delta = 0$.

(b) has one real root and two conjugate imaginary roots if $\Delta > 0$.

(c) has three distinct real roots if $\Delta < 0$.

7. Solve the following cubic equations.

(a) $x^3 - 4x^2 + 11x + 30 = 0$

(b) $x^3 - 3x + 5 = 0$

(c) $x^3 - 3x + 2 = 0$

(d) $x^3 + x + 3 = 0$

8. Show that the general quartic equation

$$x^4 + ax^3 + bx^2 + cx + d = 0$$

can be reduced to

$$y^4 + py^2 + qy + r = 0$$

by using the substitution $x = y - a/4$.

9. Show that

$$\left(y^2 + \frac{1}{2}z\right)^2 = (z-p)y^2 - qy + \left(\frac{1}{4}z^2 - r\right).$$

10. Show that the right-hand side of Exercise 10.6.9 can be put in the form $(my+k)^2$ if and only if

$$q^2 - 4(z-p)\left(\frac{1}{4}z^2 - r\right) = 0.$$

11. From Exercise 10.6.10 obtain the **resolvent cubic equation**

$$z^3 - pz^2 - 4rz + (4pr - q^2) = 0.$$

Solving the resolvent cubic equation, put the equation found in Exercise 10.6.9 in the form

$$\left(y^2 + \frac{1}{2}z\right)^2 = (my+k)^2$$

to obtain the solution of the quartic equation.

12. Use this method to solve the following quartic equations.

(a) $x^4 - x^2 - 3x + 2 = 0$

(b) $x^4 + x^3 - 7x^2 - x + 6 = 0$

(c) $x^4 - 2x^2 + 4x - 3 = 0$

(d) $x^4 - 4x^3 + 3x^2 - 5x + 2 = 0$

Chapter 11

Integral Domains

One of the most important rings we study is the ring of integers. It was our first example of an algebraic structure: the first polynomial ring that we examined was $\mathbb{Z}[x]$. We also know that the integers sit naturally inside the field of rational numbers, $\mathbb{Q}$. The ring of integers is the model for all integral domains. In this chapter we will examine integral domains in general, answering questions about the ideal structure of integral domains, polynomial rings over integral domains, and whether or not an integral domain can be embedded in a field.

11.1 Fields of Fractions

Every field is also an integral domain; however, there are many integral domains that are not fields. For example, the integers $\mathbb{Z}$ form an integral domain but not a field. A question that naturally arises is how we might associate an integral domain with a field. There is a natural way to construct the rationals $\mathbb{Q}$ from the integers: the rationals can be represented as formal quotients of two integers. The rational numbers are certainly a field. In fact, it can be shown that the rationals are the smallest field that contains the integers. Given an integral domain D, our question now becomes how to construct a smallest field F containing D. We will do this in the same way as we constructed the rationals from the integers.

An element $p/q \in \mathbb{Q}$ is the quotient of two integers p and q; however, different pairs of integers can represent the same rational number. For instance, $1/2 = 2/4 = 3/6$. We know that

$$\frac{a}{b} = \frac{c}{d}$$

if and only if $ad = bc$. A more formal way of considering this problem is to examine fractions in terms of equivalence relations. We can think of elements in $\mathbb{Q}$ as ordered pairs in $\mathbb{Z} \times \mathbb{Z}$. A quotient p/q can be written as (p, q). For instance, $(3, 7)$ would represent the fraction $3/7$. However, there are problems if we consider all possible pairs in $\mathbb{Z} \times \mathbb{Z}$. There is no fraction $5/0$ corresponding to the pair $(5, 0)$. Also, the pairs $(3, 6)$ and $(2, 4)$ both represent the fraction $1/2$. The first problem is easily solved if we require the second coordinate to be nonzero. The second problem is solved by considering two pairs (a, b) and (c, d) to be equivalent if $ad = bc$.

If we use the approach of ordered pairs instead of fractions, then we can study integral domains in general. Let D be any integral domain and let

$$S = \{(a, b) : a, b \in D \text{ and } b \neq 0\}.$$

Define a relation on S by $(a, b) \sim (c, d)$ if $ad = bc$.

Lemma 11.1 *The relation $\sim$ between elements of S is an equivalence relation.*
Proof. Since D is commutative, $ab = ba$; hence, $\sim$ is reflexive on D. Now suppose that $(a, b) \sim (c, d)$. Then $ad = bc$ or $cb = da$. Therefore, $(c, d) \sim (a, b)$ and the relation is symmetric. Finally, to show that the relation is transitive, let $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. In this case $ad = bc$ and $cf = de$. Multiplying both sides of $ad = bc$ by f yields

$$afd = adf = bcf = bde = bed.$$

Since D is an integral domain, we can deduce that $af = be$ or $(a, b) \sim (e, f)$. ■

We will denote the set of equivalence classes on S by F_D. We now need to define the operations of addition and multiplication on F_D. Recall how fractions are added and multiplied in $\mathbb{Q}$:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd};$$
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

It seems reasonable to define the operations of addition and multiplication on F_D in a similar manner. If we denote the equivalence class of $(a, b) \in S$ by $[a, b]$, then we are led to define the operations of addition and multiplication on F_D by

$$[a, b] + [c, d] = [ad + bc, bd]$$

and

$$[a, b] \cdot [c, d] = [ac, bd],$$

respectively. The next lemma demonstrates that these operations are independent of the choice of representatives from each equivalence class.

Lemma 11.2 *The operations of addition and multiplication on F_D are well-defined.*
Proof. We will prove that the operation of addition is well-defined. The proof that multiplication is well-defined is left as an exercise. Let $[a_1, b_1] = [a_2, b_2]$ and $[c_1, d_1] = [c_2, d_2]$. We must show that

$$[a_1d_1 + b_1c_1, b_1d_1] = [a_2d_2 + b_2c_2, b_2d_2]$$

or, equivalently, that

$$(a_1d_1 + b_1c_1)(b_2d_2) = (b_1d_1)(a_2d_2 + b_2c_2).$$

Since $[a_1, b_1] = [a_2, b_2]$ and $[c_1, d_1] = [c_2, d_2]$, we know that $a_1b_2 = b_1a_2$ and $c_1d_2 = d_1c_2$. Therefore,

$$\begin{aligned}(a_1d_1 + b_1c_1)(b_2d_2) &= a_1d_1b_2d_2 + b_1c_1b_2d_2 \\ &= a_1b_2d_1d_2 + b_1b_2c_1d_2 \\ &= b_1a_2d_1d_2 + b_1b_2d_1c_2 \\ &= (b_1d_1)(a_2d_2 + b_2c_2).\end{aligned}$$

■

Lemma 11.3 *The set of equivalence classes of S, F_D, under the equivalence relation $\sim$, together with the operations of addition and multiplication defined* *by*

$$\begin{aligned} [a,b]+[c,d] &= [ad+bc, bd] \\ [a,b]\cdot[c,d] &= [ac, bd], \end{aligned}$$

is a field.

Proof. The additive and multiplicative identities are $[0,1]$ and $[1,1]$, respectively. To show that $[0,1]$ is the additive identity, observe that

$$[a,b]+[0,1] = [a1+b0, b1] = [a,b].$$

It is easy to show that $[1,1]$ is the multiplicative identity. Let $[a,b] \in F_D$ such that $a \neq 0$. Then $[b,a]$ is also in F_D and $[a,b]\cdot[b,a] = [1,1]$; hence, $[b,a]$ is the multiplicative inverse for $[a,b]$. Similarly, $[-a,b]$ is the additive inverse of $[a,b]$. We leave as exercises the verification of the associative and commutative properties of multiplication in F_D. We also leave it to the reader to show that F_D is an abelian group under addition.

It remains to show that the distributive property holds in F_D; however,

$$\begin{aligned} [a,b][e,f]+[c,d][e,f] &= [ae, bf]+[ce, df] \\ &= [aedf + bfce, bdf^2] \\ &= [aed+bce, bdf] \\ &= [ade+bce, bdf] \\ &= ([a,b]+[c,d])[e,f] \end{aligned}$$

and the lemma is proved. ■

The field F_D in Lemma 11.3 is called the **field of fractions** or **field of quotients** of the integral domain D.

Theorem 11.4 *Let D be an integral domain. Then D can be embedded in a field of fractions F_D, where any element in F_D can be expressed as the quotient of two elements in D. Furthermore, the field of fractions F_D is unique in the sense that if E is any field containing D, then there exists a map $\psi : F_D \to E$ giving an isomorphism with a subfield of E such that $\psi(a) = a$ for all elements $a \in D$, where we identify a with its image in F_D.*

Proof. We will first demonstrate that D can be embedded in the field F_D. Define a map $\phi : D \to F_D$ by $\phi(a) = [a,1]$. Then for a and b in D,

$$\phi(a+b) = [a+b,1] = [a,1]+[b,1] = \phi(a)+\phi(b)$$

and

$$\phi(ab) = [ab,1] = [a,1][b,1] = \phi(a)\phi(b);$$

hence, ϕ is a homomorphism. To show that ϕ is one-to-one, suppose that $\phi(a) = \phi(b)$. Then $[a,1] = [b,1]$, or $a = a1 = 1b = b$. Finally, any element of F_D can be expressed as the quotient of two elements in D, since

$$\phi(a)[\phi(b)]^{-1} = [a,1][b,1]^{-1} = [a,1]\cdot[1,b] = [a,b].$$

Now let E be a field containing D and define a map $\psi : F_D \to E$ by $\psi([a,b]) = ab^{-1}$. To show that ψ is well-defined, let $[a_1,b_1] = [a_2,b_2]$. Then $a_1b_2 = b_1a_2$. Therefore, $a_1b_1^{-1} = a_2b_2^{-1}$ and $\psi([a_1,b_1]) = \psi([a_2,b_2])$.

If $[a,b]$ and $[c,d]$ are in F_D, then

$$\begin{aligned} \psi([a,b]+[c,d]) &= \psi([ad+bc, bd]) \\ &= (ad+bc)(bd)^{-1} \end{aligned}$$

$$
\begin{aligned}
&= ab^{-1} + cd^{-1} \\
&= \psi([a,b]) + \psi([c,d])
\end{aligned}
$$

and

$$
\begin{aligned}
\psi([a,b] \cdot [c,d]) &= \psi([ac, bd]) \\
&= (ac)(bd)^{-1} \\
&= ab^{-1}cd^{-1} \\
&= \psi([a,b])\psi([c,d]).
\end{aligned}
$$

Therefore, ψ is a homomorphism.

To complete the proof of the theorem, we need to show that ψ is one-to-one. Suppose that $\psi([a,b]) = ab^{-1} = 0$. Then $a = 0b = 0$ and $[a,b] = [0,b]$. Therefore, the kernel of ψ is the zero element $[0,b]$ in F_D, and ψ is injective. ■

Example 11.5 Since $\mathbb{Q}$ is a field, $\mathbb{Q}[x]$ is an integral domain. The field of fractions of $\mathbb{Q}[x]$ is the set of all rational expressions $p(x)/q(x)$, where $p(x)$ and $q(x)$ are polynomials over the rationals and $q(x)$ is not the zero polynomial. We will denote this field by $\mathbb{Q}(x)$. □

We will leave the proofs of the following corollaries of Theorem 11.4 as exercises.

Corollary 11.6 *Let F be a field of characteristic zero. Then F contains a subfield isomorphic to $\mathbb{Q}$.*

Corollary 11.7 *Let F be a field of characteristic p. Then F contains a subfield isomorphic to $\mathbb{Z}_p$.*

11.1.1 Historical Note

Karl Friedrich Gauss, born in Brunswick, Germany on April 30, 1777, is considered to be one of the greatest mathematicians who ever lived. Gauss was truly a child prodigy. At the age of three he was able to detect errors in the books of his father's business. Gauss entered college at the age of 15. Before the age of 20, Gauss was able to construct a regular 17-sided polygon with a ruler and compass. This was the first new construction of a regular n-sided polygon since the time of the ancient Greeks. Gauss succeeded in showing that if $N = 2^{2^n} + 1$ was prime, then it was possible to construct a regular N-sided polygon.

Gauss obtained his Ph.D. in 1799 under the direction of Pfaff at the University of Helmstedt. In his dissertation he gave the first complete proof of the Fundamental Theorem of Algebra, which states that every polynomial with real coefficients can be factored into linear factors over the complex numbers. The acceptance of complex numbers was brought about by Gauss, who was the first person to use the notation of i for $\sqrt{-1}$.

Gauss then turned his attention toward number theory; in 1801, he published his famous book on number theory, *Disquisitiones Arithmeticae*. Throughout his life Gauss was intrigued with this branch of mathematics. He once wrote, "Mathematics is the queen of the sciences, and the theory of numbers is the queen of mathematics."

In 1807, Gauss was appointed director of the Observatory at the University of Göttingen, a position he held until his death. This position required him to study applications of mathematics to the sciences. He succeeded in making contributions to fields such as astronomy, mechanics, optics, geodesy, and magnetism. Along with Wilhelm Weber, he coinvented the first practical electric telegraph some years before a better version was invented by Samuel F. B. Morse.

Gauss was clearly the most prominent mathematician in the world in the early nineteenth century. His status naturally made his discoveries subject to intense scrutiny. Gauss's cold and distant personality many times led him to ignore the work of his contemporaries, making him many enemies. He did not enjoy teaching very much, and young mathematicians who sought him out for encouragement were often rebuffed. Nevertheless, he had many outstanding students, including Eisenstein, Riemann, Kummer, Dirichlet, and Dedekind. Gauss also offered a great deal of encouragement to Sophie Germain (1776–1831), who overcame the many obstacles facing women in her day to become a very prominent mathematician. Gauss died at the age of 78 in Göttingen on February 23, 1855.

11.2 Exercises

1. Let $z = a + b\sqrt{3}\,i$ be in $\mathbb{Z}[\sqrt{3}\,i]$. If $a^2 + 3b^2 = 1$, show that z must be a unit. Show that the only units of $\mathbb{Z}[\sqrt{3}\,i]$ are 1 and -1.

Hint. Note that $z^{-1} = 1/(a+b\sqrt{3}\,i) = (a-b\sqrt{3}\,i)/(a^2+3b^2)$ is in $\mathbb{Z}[\sqrt{3}\,i]$ if and only if $a^2 + 3b^2 = 1$. The only integer solutions to the equation are $a = \pm 1, b = 0$.

2. The Gaussian integers, $\mathbb{Z}[i]$, are a UFD. Factor each of the following elements in $\mathbb{Z}[i]$ into a product of irreducibles.

(a) 5

(b) $1 + 3i$

(c) $6 + 8i$

(d) 2

Hint. (a) $5 = -i(1 + 2i)(2 + i)$; (c) $6 + 8i = -i(1 + i)^2(2 + i)^2$.

3. Let D be an integral domain.

(a) Prove that F_D is an abelian group under the operation of addition.

(b) Show that the operation of multiplication is well-defined in the field of fractions, F_D.

(c) Verify the associative and commutative properties for multiplication in F_D.

4. Prove or disprove: Any subring of a field F containing 1 is an integral domain.

Hint. True.

5. Prove or disprove: If D is an integral domain, then every prime element in D is also irreducible in D.

6. Let F be a field of characteristic zero. Prove that F contains a subfield isomorphic to $\mathbb{Q}$.

7. Let F be a field.

(a) Prove that the field of fractions of $F[x]$, denoted by $F(x)$, is isomorphic to the set all rational expressions $p(x)/q(x)$, where $q(x)$ is not the zero polynomial.

(b) Let $p(x_1, \ldots, x_n)$ and $q(x_1, \ldots, x_n)$ be polynomials in $F[x_1, \ldots, x_n]$. Show that the set of all rational expressions $p(x_1, \ldots, x_n)/q(x_1, \ldots, x_n)$ is isomorphic to the field of fractions of $F[x_1, \ldots, x_n]$. We denote the field of fractions of $F[x_1, \ldots, x_n]$ by $F(x_1, \ldots, x_n)$.

8. Let p be prime and denote the field of fractions of $\mathbb{Z}_p[x]$ by $\mathbb{Z}_p(x)$. Prove that $\mathbb{Z}_p(x)$ is an infinite field of characteristic p.

9. Prove that the field of fractions of the Gaussian integers, $\mathbb{Z}[i]$, is

$$\mathbb{Q}(i) = \{p + qi : p, q \in \mathbb{Q}\}.$$

Hint. Let $z = a+bi$ and $w = c+di \neq 0$ be in $\mathbb{Z}[i]$. Prove that $z/w \in \mathbb{Q}(i)$.

10. A field F is called a **prime field** if it has no proper subfields. If E is a subfield of F and E is a prime field, then E is a **prime subfield** of F.

(a) Prove that every field contains a unique prime subfield.

(b) If F is a field of characteristic 0, prove that the prime subfield of F is isomorphic to the field of rational numbers, $\mathbb{Q}$.

(c) If F is a field of characteristic p, prove that the prime subfield of F is isomorphic to $\mathbb{Z}_p$.

11. Let $\mathbb{Z}[\sqrt{2}\,] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$.

(a) Prove that $\mathbb{Z}[\sqrt{2}\,]$ is an integral domain.

(b) Find all of the units in $\mathbb{Z}[\sqrt{2}\,]$.

(c) Determine the field of fractions of $\mathbb{Z}[\sqrt{2}\,]$.

(d) Prove that $\mathbb{Z}[\sqrt{2}i]$ is a Euclidean domain under the Euclidean valuation $\nu(a + b\sqrt{2}\,i) = a^2 + 2b^2$.

12. Let D be a UFD. An element $d \in D$ is a **greatest common divisor of** a **and** b **in** D if $d \mid a$ and $d \mid b$ and d is divisible by any other element dividing both a and b.

(a) If D is a PID and a and b are both nonzero elements of D, prove there exists a unique greatest common divisor of a and b up to associates. That is, if d and d' are both greatest common divisors of a and b, then d and d' are associates. We write $\gcd(a, b)$ for the greatest common divisor of a and b.

(b) Let D be a PID and a and b be nonzero elements of D. Prove that there exist elements s and t in D such that $\gcd(a, b) = as + bt$.

13. Let D be an integral domain. Define a relation on D by $a \sim b$ if a and b are associates in D. Prove that $\sim$ is an equivalence relation on D.

14. Let D be a Euclidean domain with Euclidean valuation ν. If u is a unit in D, show that $\nu(u) = \nu(1)$.

15. Let D be a Euclidean domain with Euclidean valuation ν. If a and b are associates in D, prove that $\nu(a) = \nu(b)$.

Hint. Let $a = ub$ with u a unit. Then $\nu(b) \leq \nu(ub) \leq \nu(a)$. Similarly, $\nu(a) \leq \nu(b)$.

16. Show that $\mathbb{Z}[\sqrt{5}\,i]$ is not a unique factorization domain.

Hint. Show that 21 can be factored in two different ways.

17. Prove or disprove: Every subdomain of a UFD is also a UFD.

18. An ideal of a commutative ring R is said to be **finitely generated** if there exist elements $a_1, \ldots, a_n$ in R such that every element $r \in R$ can be written as $a_1r_1 + \cdots + a_nr_n$ for some $r_1, \ldots, r_n$ in R. Prove that R satisfies the ascending chain condition if and only if every ideal of R is finitely generated.

19. Let D be an integral domain with a descending chain of ideals $I_1 \supset I_2 \supset I_3 \supset \cdots$. Suppose that there exists an N such that $I_k = I_N$ for all $k \geq N$. A ring satisfying this condition is said to satisfy the **descending chain condition**, or **DCC**. Rings satisfying the DCC are called **Artinian rings**, after Emil Artin. Show that if D satisfies the descending chain condition, it must satisfy the ascending chain condition.

20. Let R be a commutative ring with identity. We define a **multiplicative subset** of R to be a subset S such that $1 \in S$ and $ab \in S$ if $a, b \in S$.

(a) Define a relation $\sim$ on $R \times S$ by $(a, s) \sim (a', s')$ if there exists an $s^* \in S$ such that $s^*(s'a - sa') = 0$. Show that $\sim$ is an equivalence relation on $R \times S$.

(b) Let a/s denote the equivalence class of $(a, s) \in R \times S$ and let $S^{-1}R$ be the set of all equivalence classes with respect to $\sim$. Define the operations of addition and multiplication on $S^{-1}R$ by

$$\frac{a}{s} + \frac{b}{t} = \frac{at + bs}{st}$$
$$\frac{a}{s}\frac{b}{t} = \frac{ab}{st},$$

respectively. Prove that these operations are well-defined on $S^{-1}R$ and that $S^{-1}R$ is a ring with identity under these operations. The ring $S^{-1}R$ is called the **ring of quotients** of R with respect to S.

(c) Show that the map $\psi : R \to S^{-1}R$ defined by $\psi(a) = a/1$ is a ring homomorphism.

(d) If R has no zero divisors and $0 \notin S$, show that ψ is one-to-one.

(e) Prove that P is a prime ideal of R if and only if $S = R \setminus P$ is a multiplicative subset of R.

(f) If P is a prime ideal of R and $S = R \setminus P$, show that the ring of quotients $S^{-1}R$ has a unique maximal ideal. Any ring that has a unique maximal ideal is called a **local ring**.

11.3 Fields of Fractions in the Secondary Classroom

This appendix will relate how fields of fractions to the secondary classroom.

11.4 References and Suggested Readings

[1] Atiyah, M. F. and MacDonald, I. G. *Introduction to Commutative Algebra.* Westview Press, Boulder, CO, 1994.

[2] Zariski, O. and Samuel, P. *Commutative Algebra*, vols. I and II. Springer, New York, 1975, 1960.

Chapter 12

Vector Spaces

In a physical system a quantity can often be described with a single number. For example, we need to know only a single number to describe temperature, mass, or volume. However, for some quantities, such as location, we need several numbers. To give the location of a point in space, we need x, y, and z coordinates. Temperature distribution over a solid object requires four numbers: three to identify each point within the object and a fourth to describe the temperature at that point. Often n-tuples of numbers, or vectors, also have certain algebraic properties, such as addition or scalar multiplication.

In this chapter we will examine mathematical structures called vector spaces. As with groups and rings, it is desirable to give a simple list of axioms that must be satisfied to make a set of vectors a structure worth studying.

12.1 Definitions and Examples

A **vector space** V over a field F is an abelian group with a **scalar product** $\alpha \cdot v$ or αv defined for all $\alpha \in F$ and all $v \in V$ satisfying the following axioms.

- $\alpha(\beta v) = (\alpha\beta)v$;
- $(\alpha + \beta)v = \alpha v + \beta v$;
- $\alpha(u + v) = \alpha u + \alpha v$;
- $1v = v$;

where $\alpha, \beta \in F$ and $u, v \in V$.

The elements of V are called **vectors**; the elements of F are called **scalars**. It is important to notice that in most cases two vectors cannot be multiplied. In general, it is only possible to multiply a vector with a scalar. To differentiate between the scalar zero and the vector zero, we will write them as 0 and $\mathbf{0}$, respectively.

Let us examine several examples of vector spaces. Some of them will be quite familiar; others will seem less so.

Example 12.1 The n-tuples of real numbers, denoted by $\mathbb{R}^n$, form a vector space over $\mathbb{R}$. Given vectors $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ in $\mathbb{R}^n$ and α in $\mathbb{R}$, we can define vector addition by

$$u + v = (u_1, \ldots, u_n) + (v_1, \ldots, v_n) = (u_1 + v_1, \ldots, u_n + v_n)$$

and scalar multiplication by

$$\alpha u = \alpha(u_1, \ldots, u_n) = (\alpha u_1, \ldots, \alpha u_n).$$

□

Example 12.2 If F is a field, then $F[x]$ is a vector space over F. The vectors in $F[x]$ are simply polynomials, and vector addition is just polynomial addition. If $\alpha \in F$ and $p(x) \in F[x]$, then scalar multiplication is defined by $\alpha p(x)$. □

Example 12.3 The set of all continuous real-valued functions on a closed interval $[a, b]$ is a vector space over $\mathbb{R}$. If $f(x)$ and $g(x)$ are continuous on $[a, b]$, then $(f + g)(x)$ is defined to be $f(x) + g(x)$. Scalar multiplication is defined by $(\alpha f)(x) = \alpha f(x)$ for $\alpha \in \mathbb{R}$. For example, if $f(x) = \sin x$ and $g(x) = x^2$, then $(2f + 5g)(x) = 2 \sin x + 5x^2$. □

Example 12.4 Let $V = \mathbb{Q}(\sqrt{2}\,) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. Then V is a vector space over $\mathbb{Q}$. If $u = a + b\sqrt{2}$ and $v = c + d\sqrt{2}$, then $u + v = (a + c) + (b + d)\sqrt{2}$ is again in V. Also, for $\alpha \in \mathbb{Q}$, αv is in V. We will leave it as an exercise to verify that all of the vector space axioms hold for V. □

Proposition 12.5 *Let V be a vector space over F. Then each of the following statements is true.*

1. *$0v = \mathbf{0}$ for all $v \in V$.*

2. *$\alpha\mathbf{0} = \mathbf{0}$ for all $\alpha \in F$.*

3. *If $\alpha v = \mathbf{0}$, then either $\alpha = 0$ or $v = \mathbf{0}$.*

4. *$(-1)v = -v$ for all $v \in V$.*

5. *$-(\alpha v) = (-\alpha)v = \alpha(-v)$ for all $\alpha \in F$ and all $v \in V$.*

Proof. To prove (1), observe that

$$0v = (0 + 0)v = 0v + 0v;$$

consequently, $\mathbf{0} + 0v = 0v + 0v$. Since V is an abelian group, $\mathbf{0} = 0v$.

The proof of (2) is almost identical to the proof of (1). For (3), we are done if $\alpha = 0$. Suppose that $\alpha \neq 0$. Multiplying both sides of $\alpha v = \mathbf{0}$ by $1/\alpha$, we have $v = \mathbf{0}$.

To show (4), observe that

$$v + (-1)v = 1v + (-1)v = (1 - 1)v = 0v = \mathbf{0},$$

and so $-v = (-1)v$. We will leave the proof of (5) as an exercise. ■

12.2 Subspaces

Just as groups have subgroups and rings have subrings, vector spaces also have substructures. Let V be a vector space over a field F, and W a subset of V. Then W is a **subspace** of V if it is closed under vector addition and scalar multiplication; that is, if $u, v \in W$ and $\alpha \in F$, it will always be the case that $u + v$ and αv are also in W.

Example 12.6 Let W be the subspace of $\mathbb{R}^3$ defined by $W = \{(x_1, 2x_1 + x_2, x_1 - x_2) : x_1, x_2 \in \mathbb{R}\}$. We claim that W is a subspace of $\mathbb{R}^3$. Since

$$\alpha(x_1, 2x_1 + x_2, x_1 - x_2) = (\alpha x_1, \alpha(2x_1 + x_2), \alpha(x_1 - x_2))$$

$$= (\alpha x_1, 2(\alpha x_1) + \alpha x_2, \alpha x_1 - \alpha x_2),$$

W is closed under scalar multiplication. To show that W is closed under vector addition, let $u = (x_1, 2x_1 + x_2, x_1 - x_2)$ and $v = (y_1, 2y_1 + y_2, y_1 - y_2)$ be vectors in W. Then

$$u + v = (x_1 + y_1, 2(x_1 + y_1) + (x_2 + y_2), (x_1 + y_1) - (x_2 + y_2)).$$

□

Example 12.7 Let W be the subset of polynomials of $F[x]$ with no odd-power terms. If $p(x)$ and $q(x)$ have no odd-power terms, then neither will $p(x) + q(x)$. Also, $\alpha p(x) \in W$ for $\alpha \in F$ and $p(x) \in W$. □

Let V be any vector space over a field F and suppose that $v_1, v_2, \ldots, v_n$ are vectors in V and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are scalars in F. Any vector w in V of the form

$$w = \sum_{i=1}^{n} \alpha_i v_i = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

is called a **linear combination** of the vectors $v_1, v_2, \ldots, v_n$. The **spanning set** of vectors $v_1, v_2, \ldots, v_n$ is the set of vectors obtained from all possible linear combinations of $v_1, v_2, \ldots, v_n$. If W is the spanning set of $v_1, v_2, \ldots, v_n$, then we say that W is **spanned** by $v_1, v_2, \ldots, v_n$.

Proposition 12.8 *Let $S = \{v_1, v_2, \ldots, v_n\}$ be vectors in a vector space V. Then the span of S is a subspace of V.*

Proof. Let u and v be in S. We can write both of these vectors as linear combinations of the v_i's:

$$u = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$
$$v = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n.$$

Then

$$u + v = (\alpha_1 + \beta_1) v_1 + (\alpha_2 + \beta_2) v_2 + \cdots + (\alpha_n + \beta_n) v_n$$

is a linear combination of the v_i's. For $\alpha \in F$,

$$\alpha u = (\alpha \alpha_1) v_1 + (\alpha \alpha_2) v_2 + \cdots + (\alpha \alpha_n) v_n$$

is in the span of S. ■

12.3 Linear Independence

Let $S = \{v_1, v_2, \ldots, v_n\}$ be a set of vectors in a vector space V. If there exist scalars $\alpha_1, \alpha_2 \ldots \alpha_n \in F$ such that not all of the α_i's are zero and

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \mathbf{0},$$

then S is said to be **linearly dependent**. If the set S is not linearly dependent, then it is said to be **linearly independent**. More specifically, S is a linearly independent set if

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \mathbf{0}$$

implies that

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

for any set of scalars $\{\alpha_1, \alpha_2 \ldots \alpha_n\}$.

Proposition 12.9 *Let $\{v_1, v_2, \ldots, v_n\}$ be a set of linearly independent vectors in a vector space. Suppose that*

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n.$$

Then $\alpha_1 = \beta_1, \alpha_2 = \beta_2, \ldots, \alpha_n = \beta_n$.
Proof. If

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n,$$

then

$$(\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_n - \beta_n)v_n = \mathbf{0}.$$

Since $v_1, \ldots, v_n$ are linearly independent,$\alpha_i - \beta_i = 0$ for $i = 1, \ldots, n$. ■

The definition of linear dependence makes more sense if we consider the following proposition.

Proposition 12.10 *A set $\{v_1, v_2, \ldots, v_n\}$ of vectors in a vector space V is linearly dependent if and only if one of the v_i's is a linear combination of the rest.*
Proof. Suppose that $\{v_1, v_2, \ldots, v_n\}$ is a set of linearly dependent vectors. Then there exist scalars $\alpha_1, \ldots, \alpha_n$ such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \mathbf{0},$$

with at least one of the α_i's not equal to zero. Suppose that $\alpha_k \neq 0$. Then

$$v_k = -\frac{\alpha_1}{\alpha_k} v_1 - \cdots - \frac{\alpha_{k-1}}{\alpha_k} v_{k-1} - \frac{\alpha_{k+1}}{\alpha_k} v_{k+1} - \cdots - \frac{\alpha_n}{\alpha_k} v_n.$$

Conversely, suppose that

$$v_k = \beta_1 v_1 + \cdots + \beta_{k-1} v_{k-1} + \beta_{k+1} v_{k+1} + \cdots + \beta_n v_n.$$

Then

$$\beta_1 v_1 + \cdots + \beta_{k-1} v_{k-1} - v_k + \beta_{k+1} v_{k+1} + \cdots + \beta_n v_n = \mathbf{0}.$$

■

The following proposition is a consequence of the fact that any system of homogeneous linear equations with more unknowns than equations will have a nontrivial solution. We leave the details of the proof for the end-of-chapter exercises.

Proposition 12.11 *Suppose that a vector space V is spanned by n vectors. If $m > n$, then any set of m vectors in V must be linearly dependent.*

A set $\{e_1, e_2, \ldots, e_n\}$ of vectors in a vector space V is called a **basis** for V if $\{e_1, e_2, \ldots, e_n\}$ is a linearly independent set that spans V.

Example 12.12 The vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$ form a basis for $\mathbb{R}^3$. The set certainly spans $\mathbb{R}^3$, since any arbitrary vector (x_1, x_2, x_3) in $\mathbb{R}^3$ can be written as $x_1 e_1 + x_2 e_2 + x_3 e_3$. Also, none of the vectors e_1, e_2, e_3 can be written as a linear combination of the other two; hence, they are linearly independent. The vectors e_1, e_2, e_3 are not the only basis of $\mathbb{R}^3$: the set $\{(3, 2, 1), (3, 2, 0), (1, 1, 1)\}$ is also a basis for $\mathbb{R}^3$. □

Example 12.13 Let $\mathbb{Q}(\sqrt{2}\,) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. The sets $\{1, \sqrt{2}\,\}$ and $\{1 + \sqrt{2}, 1 - \sqrt{2}\,\}$ are both bases of $\mathbb{Q}(\sqrt{2}\,)$. □

From the last two examples it should be clear that a given vector space has several bases. In fact, there are an infinite number of bases for both of these examples. *In general, there is no unique basis for a vector space.* However, every basis of $\mathbb{R}^3$ consists of exactly three vectors, and every basis of $\mathbb{Q}(\sqrt{2}\,)$ consists of exactly two vectors. This is a consequence of the next proposition.

Proposition 12.14 *Let $\{e_1, e_2, \ldots, e_m\}$ and $\{f_1, f_2, \ldots, f_n\}$ be two bases for a vector space V. Then $m = n$.*

Proof. Since $\{e_1, e_2, \ldots, e_m\}$ is a basis, it is a linearly independent set. By Proposition 12.11, $n \leq m$. Similarly, $\{f_1, f_2, \ldots, f_n\}$ is a linearly independent set, and the last proposition implies that $m \leq n$. Consequently, $m = n$. ■

If $\{e_1, e_2, \ldots, e_n\}$ is a basis for a vector space V,then we say that the **dimension** of V is n and we write $\dim V = n$. We will leave the proof of the following theorem as an exercise.

Theorem 12.15 *Let V be a vector space of dimension n.*

1. *If $S = \{v_1, \ldots, v_n\}$ is a set of linearly independent vectors for V, then S is a basis for V.*

2. *If $S = \{v_1, \ldots, v_n\}$ spans V, then S is a basis for V.*

3. *If $S = \{v_1, \ldots, v_k\}$ is a set of linearly independent vectors for V with $k < n$, then there exist vectors $v_{k+1}, \ldots, v_n$ such that*

$$\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$$

is a basis for V.

12.4 Exercises

1. If F is a field, show that $F[x]$ is a vector space over F, where the vectors in $F[x]$ are polynomials. Vector addition is polynomial addition, and scalar multiplication is defined by $\alpha p(x)$ for $\alpha \in F$.

2. Prove that $\mathbb{Q}(\sqrt{2}\,)$ is a vector space.

3. Let $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,)$ be the field generated by elements of the form $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$, where a, b, c, d are in $\mathbb{Q}$. Prove that $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,)$ is a vector space of dimension 4 over $\mathbb{Q}$. Find a basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,)$.

Hint. $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,)$ has basis $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\,\}$ over $\mathbb{Q}$.

4. Prove that the complex numbers are a vector space of dimension 2 over $\mathbb{R}$.

5. Prove that the set P_n of all polynomials of degree less than n form a subspace of the vector space $F[x]$. Find a basis for P_n and compute the dimension of P_n.

Hint. The set $\{1, x, x^2, \ldots, x^{n-1}\}$ is a basis for P_n.

6. Let F be a field and denote the set of n-tuples of F by F^n.Given vectors $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ in F^n and α in F, define vector addition by

$$u + v = (u_1, \ldots, u_n) + (v_1, \ldots, v_n) = (u_1 + v_1, \ldots, u_n + v_n)$$

and scalar multiplication by

$$\alpha u = \alpha(u_1, \ldots, u_n) = (\alpha u_1, \ldots, \alpha u_n).$$

`Prove that F^n is a vector space of dimension n under these operations.

7. Which of the following sets are subspaces of $\mathbb{R}^3$? If the set is indeed a subspace, find a basis for the subspace and compute its dimension.

(a) $\{(x_1, x_2, x_3) : 3x_1 - 2x_2 + x_3 = 0\}$

(b) $\{(x_1, x_2, x_3) : 3x_1 + 4x_3 = 0, 2x_1 - x_2 + x_3 = 0\}$

(c) $\{(x_1, x_2, x_3) : x_1 - 2x_2 + 2x_3 = 2\}$

(d) $\{(x_1, x_2, x_3) : 3x_1 - 2x_2^2 = 0\}$

Hint. (a) Subspace of dimension 2 with basis $\{(1, 0, -3), (0, 1, 2)\}$; (d) not a subspace.

8. Show that the set of all possible solutions $(x, y, z) \in \mathbb{R}^3$ of the equations

$$\begin{aligned} Ax + By + Cz &= 0 \\ Dx + Ey + Cz &= 0 \end{aligned}$$

form a subspace of $\mathbb{R}^3$.

9. Let W be the subset of continuous functions on $[0, 1]$ such that $f(0) = 0$. Prove that W is a subspace of $C[0, 1]$.

10. Let V be a vector space over F. Prove that $-(\alpha v) = (-\alpha)v = \alpha(-v)$ for all $\alpha \in F$ and all $v \in V$.

Hint. Since $0 = \alpha 0 = \alpha(-v + v) = \alpha(-v) + \alpha v$, it follows that $-\alpha v = \alpha(-v)$.

11. Let V be a vector space of dimension n. Prove each of the following statements.

(a) If $S = \{v_1, \ldots, v_n\}$ is a set of linearly independent vectors for V, then S is a basis for V.

(b) If $S = \{v_1, \ldots, v_n\}$ spans V, then S is a basis for V.

(c) If $S = \{v_1, \ldots, v_k\}$ is a set of linearly independent vectors for V with $k < n$, then there exist vectors $v_{k+1}, \ldots, v_n$ such that

$$\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$$

is a basis for V.

12. Prove that any set of vectors containing $\mathbf{0}$ is linearly dependent.

Hint. Let $v_0 = 0, v_1, \ldots, v_n \in V$ and $\alpha_0 \neq 0, \alpha_1, \ldots, \alpha_n \in F$. Then $\alpha_0 v_0 + \cdots + \alpha_n v_n = 0$.

13. Let V be a vector space. Show that $\{\mathbf{0}\}$ is a subspace of V of dimension zero.

14. If a vector space V is spanned by n vectors, show that any set of m vectors in V must be linearly dependent for $m > n$.

15. **Linear Transformations.** Let V and W be vector spaces over a field F, of dimensions m and n, respectively. If $T : V \to W$ is a map satisfying

$$\begin{aligned} T(u + v) &= T(u) + T(v) \\ T(\alpha v) &= \alpha T(v) \end{aligned}$$

for all $\alpha \in F$ and all $u, v \in V$, then T is called a **linear transformation** from V into W.

(a) Prove that the **kernel** of T, $\ker(T) = \{v \in V : T(v) = \mathbf{0}\}$, is a subspace of V. The kernel of T is sometimes called the **null space** of T.

(b) Prove that the **range** or **range space** of T, $R(V) = \{w \in W : T(v) = w \text{ for some } v \in V\}$, is a subspace of W.

(c) Show that $T : V \to W$ is injective if and only if $\ker(T) = \{\mathbf{0}\}$.

(d) Let $\{v_1, \ldots, v_k\}$ be a basis for the null space of T. We can extend this basis to be a basis $\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_m\}$ of V. Why? Prove that $\{T(v_{k+1}), \ldots, T(v_m)\}$ is a basis for the range of T. Conclude that the range of T has dimension $m - k$.

(e) Let $\dim V = \dim W$. Show that a linear transformation $T : V \to W$ is injective if and only if it is surjective.

Hint. (a) Let $u, v \in \ker(T)$ and $\alpha \in F$. Then

$$T(u + v) = T(u) + T(v) = 0$$
$$T(\alpha v) = \alpha T(v) = \alpha 0 = 0.$$

Hence, $u + v, \alpha v \in \ker(T)$, and $\ker(T)$ is a subspace of V.

(c) The statement that $T(u) = T(v)$ is equivalent to $T(u - v) = T(u) - T(v) = 0$, which is true if and only if $u - v = 0$ or $u = v$.

16. Let V and W be finite dimensional vector spaces of dimension n over a field F. Suppose that $T : V \to W$ is a vector space isomorphism. If $\{v_1, \ldots, v_n\}$ is a basis of V, show that $\{T(v_1), \ldots, T(v_n)\}$ is a basis of W. Conclude that any vector space over a field F of dimension n is isomorphic to F^n.

17. Direct Sums. Let U and V be subspaces of a vector space W. The sum of U and V, denoted $U + V$, is defined to be the set of all vectors of the form $u + v$, where $u \in U$ and $v \in V$.

(a) Prove that $U + V$ and $U \cap V$ are subspaces of W.

(b) If $U + V = W$ and $U \cap V = \mathbf{0}$, then W is said to be the **direct sum.** In this case, we write $W = U \oplus V$. Show that every element $w \in W$ can be written uniquely as $w = u + v$, where $u \in U$ and $v \in V$.

(c) Let U be a subspace of dimension k of a vector space W of dimension n. Prove that there exists a subspace V of dimension $n - k$ such that $W = U \oplus V$. Is the subspace V unique?

(d) If U and V are arbitrary subspaces of a vector space W, show that

$$\dim(U + V) = \dim U + \dim V - \dim(U \cap V).$$

Hint. (a) Let $u, u' \in U$ and $v, v' \in V$. Then

$$(u + v) + (u' + v') = (u + u') + (v + v') \in U + V$$
$$\alpha(u + v) = \alpha u + \alpha v \in U + V.$$

18. Dual Spaces. Let V and W be finite dimensional vector spaces over a field F.

(a) Show that the set of all linear transformations from V into W, denoted by $\text{Hom}(V, W)$, is a vector space over F, where we define vector addition as follows:

$$(S+T)(v) = S(v) + T(v)$$
$$(\alpha S)(v) = \alpha S(v),$$

where $S, T \in \mathrm{Hom}(V, W)$, $\alpha \in F$, and $v \in V$.

(b) Let V be an F-vector space. Define the **dual space** of V to be $V^* = \mathrm{Hom}(V, F)$. Elements in the dual space of V are called **linear functionals.** Let $v_1, \ldots, v_n$ be an ordered basis for V. If $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ is any vector in V, define a linear functional $\phi_i : V \to F$ by $\phi_i(v) = \alpha_i$. Show that the ϕ_i's form a basis for V^*. This basis is called the **dual basis** of $v_1, \ldots, v_n$ (or simply the dual basis if the context makes the meaning clear).

(c) Consider the basis $\{(3,1), (2,-2)\}$ for $\mathbb{R}^2$. What is the dual basis for $(\mathbb{R}^2)^*$?

(d) Let V be a vector space of dimension n over a field F and let V^{**} be the dual space of V^*. Show that each element $v \in V$ gives rise to an element λ_v in V^{**} and that the map $v \mapsto \lambda_v$ is an isomorphism of V with V^{**}.

12.5 Vector Spaces in the Secondary Classroom

This appendix will relate vector spaces to the secondary classroom.

12.6 References and Suggested Readings

[1] Beezer, R. *A First Course in Linear Algebra* . Available online at http://linear.ups.edu/. 2004–2014.

[2] Bretscher, O. *Linear Algebra with Applications.* 4th ed. Pearson, Upper Saddle River, NJ, 2009.

[3] Curtis, C. W. *Linear Algebra: An Introductory Approach.* 4th ed. Springer, New York, 1984.

[4] Hoffman, K. and Kunze, R. *Linear Algebra.* 2nd ed. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[5] Johnson, L. W., Riess, R. D., and Arnold, J. T. *Introduction to Linear Algebra.* 6th ed. Pearson, Upper Saddle River, NJ, 2011.

[6] Leon, S. J. *Linear Algebra with Applications.* 8th ed. Pearson, Upper Saddle River, NJ, 2010.

Chapter 13

Fields

It is natural to ask whether or not some field F is contained in a larger field. We think of the rational numbers, which reside inside the real numbers, while in turn, the real numbers live inside the complex numbers. We can also study the fields between $\mathbb{Q}$ and $\mathbb{R}$ and inquire as to the nature of these fields.

More specifically if we are given a field F and a polynomial $p(x) \in F[x]$, we can ask whether or not we can find a field E containing F such that $p(x)$ factors into linear factors over $E[x]$. For example, if we consider the polynomial

$$p(x) = x^4 - 5x^2 + 6$$

in $\mathbb{Q}[x]$, then $p(x)$ factors as $(x^2 - 2)(x^2 - 3)$. However, both of these factors are irreducible in $\mathbb{Q}[x]$. If we wish to find a zero of $p(x)$, we must go to a larger field. Certainly the field of real numbers will work, since

$$p(x) = (x - \sqrt{2})(x + \sqrt{2})(x - \sqrt{3})(x + \sqrt{3}).$$

It is possible to find a smaller field in which $p(x)$ has a zero, namely

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

We wish to be able to compute and study such fields for arbitrary polynomials over a field F.

13.1 Extension Fields

A field E is an **extension field** of a field F if F is a subfield of E. The field F is called the **base field**. We write $F \subset E$.

Example 13.1 For example, let

$$F = \mathbb{Q}(\sqrt{2}\,) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$$

and let $E = \mathbb{Q}(\sqrt{2} + \sqrt{3}\,)$ be the smallest field containing both $\mathbb{Q}$ and $\sqrt{2} + \sqrt{3}$. Both E and F are extension fields of the rational numbers. We claim that E is an extension field of F. To see this, we need only show that $\sqrt{2}$ is in E. Since $\sqrt{2} + \sqrt{3}$ is in E, $1/(\sqrt{2} + \sqrt{3}\,) = \sqrt{3} - \sqrt{2}$ must also be in E. Taking linear combinations of $\sqrt{2} + \sqrt{3}$ and $\sqrt{3} - \sqrt{2}$, we find that $\sqrt{2}$ and $\sqrt{3}$ must both be in E. □

Example 13.2 Let $p(x) = x^2 + x + 1 \in \mathbb{Z}_2[x]$. Since neither 0 nor 1 is a root of this polynomial, we know that $p(x)$ is irreducible over $\mathbb{Z}_2$. We will construct a field extension of $\mathbb{Z}_2$ containing an element α such that $p(\alpha)=0$. By Theorem 10.23, the ideal $\langle p(x)\rangle$ generated by $p(x)$ is maximal; hence, $\mathbb{Z}_2[x]/\langle p(x)\rangle$ is a field. Let $f(x)+\langle p(x)\rangle$ be an arbitrary element of $\mathbb{Z}_2[x]/\langle p(x)\rangle$. By the division algorithm,

$$f(x)=(x^2+x+1)q(x)+r(x),$$

where the degree of $r(x)$ is less than the degree of x^2+x+1. Therefore,

$$f(x)+\langle x^2+x+1\rangle = r(x)+\langle x^2+x+1\rangle.$$

The only possibilities for $r(x)$ are then 0, 1, x, and $1+x$. Consequently, $E=\mathbb{Z}_2[x]/\langle x^2+x+1\rangle$ is a field with four elements and must be a field extension of $\mathbb{Z}_2$, containing a zero α of $p(x)$. The field $\mathbb{Z}_2(\alpha)$ consists of elements

$$\begin{aligned} 0+0\alpha &= 0\\ 1+0\alpha &= 1\\ 0+1\alpha &= \alpha\\ 1+1\alpha &= 1+\alpha. \end{aligned}$$

Notice that $\alpha^2+\alpha+1=0$; hence, if we compute $(1+\alpha)^2$,

$$(1+\alpha)(1+\alpha)=1+\alpha+\alpha+(\alpha)^2=\alpha.$$

Other calculations are accomplished in a similar manner. We summarize these computations in the following tables, which tell us how to add and multiply elements in E. □

Table 13.3 Addition Table for $\mathbb{Z}_2(\alpha)$

$+$	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

Table 13.4 Multiplication Table for $\mathbb{Z}_2(\alpha)$

$\cdot$	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

The following theorem, due to Kronecker, is so important and so basic to our understanding of fields that it is often known as the Fundamental Theorem of Field Theory.

Theorem 13.5 *Let F be a field and let $p(x)$ be a nonconstant polynomial in $F[x]$. Then there exists an extension field E of F and an element $\alpha\in E$ such that $p(\alpha)=0$.*

Proof. To prove this theorem, we will employ the method that we used to construct Example 13.2. Clearly, we can assume that $p(x)$ is an irreducible polynomial. We wish to find an extension field E of F containing an element α such that $p(\alpha)=0$. The ideal $\langle p(x)\rangle$ generated by $p(x)$ is a maximal ideal in $F[x]$ by Theorem 10.23; hence, $F[x]/\langle p(x)\rangle$ is a field. We claim that $E=F[x]/\langle p(x)\rangle$

is the desired field.

We first show that E is a field extension of F. We can define a homomorphism of commutative rings by the map $\psi : F \to F[x]/\langle p(x)\rangle$, where $\psi(a) = a + \langle p(x)\rangle$ for $a \in F$. It is easy to check that ψ is indeed a ring homomorphism. Observe that

$$\psi(a) + \psi(b) = (a + \langle p(x)\rangle) + (b + \langle p(x)\rangle) = (a + b) + \langle p(x)\rangle = \psi(a + b)$$

and

$$\psi(a)\psi(b) = (a + \langle p(x)\rangle)(b + \langle p(x)\rangle) = ab + \langle p(x)\rangle = \psi(ab).$$

To prove that ψ is one-to-one, assume that

$$a + \langle p(x)\rangle = \psi(a) = \psi(b) = b + \langle p(x)\rangle.$$

Then $a - b$ is a multiple of $p(x)$, since it lives in the ideal $\langle p(x)\rangle$. Since $p(x)$ is a nonconstant polynomial, the only possibility is that $a - b = 0$. Consequently, $a = b$ and ψ is injective. Since ψ is one-to-one, we can identify F with the subfield $\{a + \langle p(x)\rangle : a \in F\}$ of E and view E as an extension field of F.

It remains for us to prove that $p(x)$ has a zero $\alpha \in E$. Set $\alpha = x + \langle p(x)\rangle$. Then α is in E. If $p(x) = a_0 + a_1x + \cdots + a_nx^n$, then

$$\begin{aligned} p(\alpha) &= a_0 + a_1(x + \langle p(x)\rangle) + \cdots + a_n(x + \langle p(x)\rangle)^n \\ &= a_0 + (a_1x + \langle p(x)\rangle) + \cdots + (a_nx^n + \langle p(x)\rangle) \\ &= a_0 + a_1x + \cdots + a_nx^n + \langle p(x)\rangle \\ &= 0 + \langle p(x)\rangle. \end{aligned}$$

Therefore, we have found an element $\alpha \in E = F[x]/\langle p(x)\rangle$ such that α is a zero of $p(x)$. ■

Example 13.6 Let $p(x) = x^5 + x^4 + 1 \in \mathbb{Z}_2[x]$. Then $p(x)$ has irreducible factors $x^2 + x + 1$ and $x^3 + x + 1$. For a field extension E of $\mathbb{Z}_2$ such that $p(x)$ has a root in E, we can let E be either $\mathbb{Z}_2[x]/\langle x^2 + x + 1\rangle$ or $\mathbb{Z}_2[x]/\langle x^3 + x + 1\rangle$. We will leave it as an exercise to show that $\mathbb{Z}_2[x]/\langle x^3 + x + 1\rangle$ is a field with $2^3 = 8$ elements. □

13.1.1 Algebraic Elements

An element α in an extension field E over F is **algebraic** over F if $f(\alpha) = 0$ for some nonzero polynomial $f(x) \in F[x]$. An element in E that is not algebraic over F is **transcendental** over F. An extension field E of a field F is an **algebraic extension** of F if every element in E is algebraic over F. If E is a field extension of F and $\alpha_1, \ldots, \alpha_n$ are contained in E, we denote the smallest field containing F and $\alpha_1, \ldots, \alpha_n$ by $F(\alpha_1, \ldots, \alpha_n)$. If $E = F(\alpha)$ for some $\alpha \in E$,then E is a **simple extension** of F.

Example 13.7 Both $\sqrt{2}$ and i are algebraic over$\mathbb{Q}$ since they are zeros of the polynomials $x^2 - 2$ and $x^2 + 1$, respectively. Clearly π and e are algebraic over the real numbers; however, it is a nontrivial fact that they are transcendental over $\mathbb{Q}$. Numbers in $\mathbb{R}$ that are algebraic over $\mathbb{Q}$ are in fact quite rare. Almost all real numbers are transcendental over $\mathbb{Q}$.[1](In many cases we do not know whether or not a particular number is transcendental; for example, it is still not known whether $\pi + e$ is transcendental or algebraic.) □

[1] The probability that a real number chosen at random from the interval $[0, 1]$ will be transcendental over the rational numbers is one.

A complex number that is algebraic over $\mathbb{Q}$ is an **algebraic number**. A **transcendental number** is an element of $\mathbb{C}$ that is transcendental over $\mathbb{Q}$.

Example 13.8 We will show that $\sqrt{2+\sqrt{3}}$ is algebraic over $\mathbb{Q}$. If $\alpha = \sqrt{2+\sqrt{3}}$, then $\alpha^2 = 2+\sqrt{3}$. Hence, $\alpha^2 - 2 = \sqrt{3}$ and $(\alpha^2-2)^2 = 3$. Since $\alpha^4 - 4\alpha^2 + 1 = 0$, it must be true that α is a zero of the polynomial $x^4 - 4x^2 + 1 \in \mathbb{Q}[x]$. □

It is very easy to give an example of an extension field E over a field F, The following theorem characterizes transcendental extensions.

Theorem 13.9 *Let E be an extension field of F and $\alpha \in E$. Then α is transcendental over F if and only if $F(\alpha)$ is isomorphic to $F(x)$, the field of fractions of $F[x]$.*

Proof. Let $\phi_\alpha : F[x] \to E$ be the evaluation homomorphism for α. Then α is transcendental over F if and only if $\phi_\alpha(p(x)) = p(\alpha) \neq 0$ for all nonconstant polynomials $p(x) \in F[x]$. This is true if and only if $\ker \phi_\alpha = \{0\}$; that is, it is true exactly when ϕ_α is one-to-one. Hence, E must contain a copy of $F[x]$. The smallest field containing $F[x]$ is the field of fractions $F(x)$. By Theorem 11.4, E must contain a copy of this field. ■

We have a more interesting situation in the case of algebraic extensions.

Theorem 13.10 *Let E be an extension field of a field F and $\alpha \in E$ with α algebraic over F. Then there is a unique irreducible monic polynomial $p(x) \in F[x]$ of smallest degree such that $p(\alpha) = 0$. If $f(x)$ is another polynomial in $F[x]$ such that $f(\alpha) = 0$, then $p(x)$ divides $f(x)$.*

Proof. Let $\phi_\alpha : F[x] \to E$ be the evaluation homomorphism. The kernel of ϕ_α is a principal ideal generated by some $p(x) \in F[x]$ with $\deg p(x) \geq 1$. We know that such a polynomial exists, since $F[x]$ is a principal ideal domain and α is algebraic. The ideal $\langle p(x) \rangle$ consists exactly of those elements of $F[x]$ having α as a zero. If $f(\alpha) = 0$ and $f(x)$ is not the zero polynomial, then $f(x) \in \langle p(x) \rangle$ and $p(x)$ divides $f(x)$. So $p(x)$ is a polynomial of minimal degree having α as a zero. Any other polynomial of the same degree having α as a zero must have the form $\beta p(x)$ for some $\beta \in F$.

Suppose now that $p(x) = r(x)s(x)$ is a factorization of $p(x)$ into polynomials of lower degree. Since $p(\alpha) = 0$, $r(\alpha)s(\alpha) = 0$; consequently, either $r(\alpha) = 0$ or $s(\alpha) = 0$, which contradicts the fact that p is of minimal degree. Therefore, $p(x)$ must be irreducible. ■

Let E be an extension field of F and $\alpha \in E$ be algebraic over F. The unique monic polynomial $p(x)$ of the last theorem is called the **minimal polynomial** for α over F. The degree of $p(x)$ is the **degree of α over F**.

Example 13.11 Let $f(x) = x^2 - 2$ and $g(x) = x^4 - 4x^2 + 1$. These polynomials are the minimal polynomials of $\sqrt{2}$ and $\sqrt{2+\sqrt{3}}$, respectively. □

Proposition 13.12 *Let E be a field extension of F and $\alpha \in E$ be algebraic over F. Then $F(\alpha) \cong F[x]/\langle p(x) \rangle$, where $p(x)$ is the minimal polynomial of α over F.*

Proof. Let $\phi_\alpha : F[x] \to E$ be the evaluation homomorphism. The kernel of this map is $\langle p(x) \rangle$, where $p(x)$ is the minimal polynomial of α. By the First Isomorphism Theorem for rings, the image of ϕ_α in E is isomorphic to $F(\alpha)$ since it contains both F and α. ■

Theorem 13.13 *Let $E = F(\alpha)$ be a simple extension of F, where $\alpha \in E$ is algebraic over F. Suppose that the degree of α over F is n. Then every element $\beta \in E$ can be expressed uniquely in the form*

$$\beta = b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}$$

for $b_i \in F$.

Proof. Since $\phi_\alpha(F[x]) \cong F(\alpha)$, every element in $E = F(\alpha)$ must be of the form $\phi_\alpha(f(x)) = f(\alpha)$, where $f(\alpha)$ is a polynomial in α with coefficients in F. Let

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$$

be the minimal polynomial of α. Then $p(\alpha) = 0$; hence,

$$\alpha^n = -a_{n-1}\alpha^{n-1} - \cdots - a_0.$$

Similarly,

$$\begin{aligned} \alpha^{n+1} &= \alpha\alpha^n \\ &= -a_{n-1}\alpha^n - a_{n-2}\alpha^{n-1} - \cdots - a_0\alpha \\ &= -a_{n-1}(-a_{n-1}\alpha^{n-1} - \cdots - a_0) - a_{n-2}\alpha^{n-1} - \cdots - a_0\alpha. \end{aligned}$$

Continuing in this manner, we can express every monomial α^m, $m \geq n$, as a linear combination of powers of α that are less than n. Hence, any $\beta \in F(\alpha)$ can be written as

$$\beta = b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}.$$

To show uniqueness, suppose that

$$\beta = b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1} = c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}$$

for b_i and c_i in F. Then

$$g(x) = (b_0 - c_0) + (b_1 - c_1)x + \cdots + (b_{n-1} - c_{n-1})x^{n-1}$$

is in $F[x]$ and $g(\alpha) = 0$. Since the degree of $g(x)$ is less than the degree of $p(x)$, the irreducible polynomial of α, $g(x)$ must be the zero polynomial. Consequently,

$$b_0 - c_0 = b_1 - c_1 = \cdots = b_{n-1} - c_{n-1} = 0,$$

or $b_i = c_i$ for $i = 0, 1, \ldots, n-1$. Therefore, we have shown uniqueness. ■

Example 13.14 Since $x^2 + 1$ is irreducible over $\mathbb{R}$, $\langle x^2 + 1 \rangle$ is a maximal ideal in $\mathbb{R}[x]$. So $E = \mathbb{R}[x]/\langle x^2 + 1 \rangle$ is a field extension of $\mathbb{R}$ that contains a root of $x^2 + 1$. Let $\alpha = x + \langle x^2 + 1 \rangle$. We can identify E with the complex numbers. By Proposition 13.12, E is isomorphic to $\mathbb{R}(\alpha) = \{a + b\alpha : a, b \in \mathbb{R}\}$. We know that $\alpha^2 = -1$ in E, since

$$\begin{aligned} \alpha^2 + 1 &= (x + \langle x^2 + 1 \rangle)^2 + (1 + \langle x^2 + 1 \rangle) \\ &= (x^2 + 1) + \langle x^2 + 1 \rangle \\ &= 0. \end{aligned}$$

Hence, we have an isomorphism of $\mathbb{R}(\alpha)$ with $\mathbb{C}$ defined by the map that takes $a + b\alpha$ to $a + bi$. □

Let E be a field extension of a field F. If we regard E as a vector space over F, then we can bring the machinery of linear algebra to bear on the problems that we will encounter in our study of fields. The elements in the field E are vectors; the elements in the field F are scalars. We can think of addition in E as adding vectors. When we multiply an element in E by an element of F, we are multiplying a vector by a scalar. This view of field extensions is especially fruitful if a field extension E of F is a finite dimensional vector space over F, and Theorem 13.13 states that $E = F(\alpha)$ is finite dimensional vector space over F with basis $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$.

If an extension field E of a field F is a finite dimensional vector space over F of dimension n, then we say that E is a **finite extension of degree n over F**. We write

$$[E : F] = n.$$

to indicate the dimension of E over F.

Theorem 13.15 *Every finite extension field E of a field F is an algebraic extension.*
Proof. Let $\alpha \in E$. Since $[E : F] = n$, the elements

$$1, \alpha, \ldots, \alpha^n$$

cannot be linearly independent. Hence, there exist $a_i \in F$, not all zero, such that

$$a_n\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0.$$

Therefore,

$$p(x) = a_nx^n + \cdots + a_0 \in F[x]$$

is a nonzero polynomial with $p(\alpha) = 0$. ■

Remark 13.16 Theorem 13.15 says that every finite extension of a field F is an algebraic extension. The converse is false, however. We will leave it as an exercise to show that the set of all elements in $\mathbb{R}$ that are algebraic over $\mathbb{Q}$ forms an infinite field extension of $\mathbb{Q}$.

The next theorem is a counting theorem, similar to Lagrange's Theorem in group theory. Theorem 13.17 will prove to be an extremely useful tool in our investigation of finite field extensions.

Theorem 13.17 *If E is a finite extension of F and K is a finite extension of E, then K is a finite extension of F and*

$$[K : F] = [K : E][E : F].$$

Proof. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis for E as a vector space over F and $\{\beta_1, \ldots, \beta_m\}$ be a basis for K as a vector space over E. We claim that $\{\alpha_i\beta_j\}$ is a basis for K over F. We will first show that these vectors span K. Let $u \in K$. Then $u = \sum_{j=1}^m b_j\beta_j$ and $b_j = \sum_{i=1}^n a_{ij}\alpha_i$, where $b_j \in E$ and $a_{ij} \in F$. Then

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

So the mn vectors $\alpha_i\beta_j$ must span K over F.

We must show that $\{\alpha_i\beta_j\}$ are linearly independent. Recall that a set of vectors $v_1, v_2, \ldots, v_n$ in a vector space V are linearly independent if

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$$

implies that

$$c_1 = c_2 = \cdots = c_n = 0.$$

Let

$$u = \sum_{i,j} c_{ij}(\alpha_i\beta_j) = 0$$

for $c_{ij} \in F$. We need to prove that all of the c_{ij}'s are zero. We can rewrite u as

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

where $\sum_i c_{ij}\alpha_i \in E$. Since the β_j's are linearly independent over E, it must be the case that

$$\sum_{i=1}^{n} c_{ij}\alpha_i = 0$$

for all j. However, the α_j are also linearly independent over F. Therefore, $c_{ij} = 0$ for all i and j, which completes the proof. ■

The following corollary is easily proved using mathematical induction.

Corollary 13.18 *If F_i is a field for $i = 1, \ldots, k$, and F_{i+1} is a finite extension of F_i, then F_k is a finite extention of F_1 and*

$$[F_k : F_1] = [F_k : F_{k-1}] \cdots [F_2 : F_1].$$

Corollary 13.19 *Let E be an extension field of F. If $\alpha \in E$ is algebraic over F with minimal polynomial $p(x)$ and $\beta \in F(\alpha)$ with minimal polynomial $q(x)$, then $\deg q(x)$ divides $\deg p(x)$.*

Proof. We know that $\deg p(x) = [F(\alpha) : F]$ and $\deg q(x) = [F(\beta) : F]$. Since $F \subset F(\beta) \subset F(\alpha)$,

$$[F(\alpha) : F] = [F(\alpha) : F(\beta)][F(\beta) : F].$$

■

Example 13.20 Let us determine an extension field of $\mathbb{Q}$ containing $\sqrt{3} + \sqrt{5}$. It is easy to determine that the minimal polynomial of $\sqrt{3} + \sqrt{5}$ is $x^4 - 16x^2 + 4$. It follows that

$$[\mathbb{Q}(\sqrt{3} + \sqrt{5}\,) : \mathbb{Q}] = 4.$$

We know that $\{1, \sqrt{3}\,\}$ is a basis for $\mathbb{Q}(\sqrt{3}\,)$ over $\mathbb{Q}$. Hence, $\sqrt{3} + \sqrt{5}$ cannot be in $\mathbb{Q}(\sqrt{3}\,)$. It follows that $\sqrt{5}$ cannot be in $\mathbb{Q}(\sqrt{3}\,)$ either. Therefore, $\{1, \sqrt{5}\,\}$ is a basis for $\mathbb{Q}(\sqrt{3}, \sqrt{5}\,) = (\mathbb{Q}(\sqrt{3}\,))(\sqrt{5}\,)$ over $\mathbb{Q}(\sqrt{3}\,)$ and $\{1, \sqrt{3}, \sqrt{5}, \sqrt{3}\sqrt{5} = \sqrt{15}\,\}$ is a basis for $\mathbb{Q}(\sqrt{3}, \sqrt{5}\,) = \mathbb{Q}(\sqrt{3} + \sqrt{5}\,)$ over $\mathbb{Q}$. This example shows that it is possible that some extension $F(\alpha_1, \ldots, \alpha_n)$ is actually a simple extension of F even though $n > 1$. □

Example 13.21 Let us compute a basis for $\mathbb{Q}(\sqrt[3]{5}, \sqrt{5}\,i)$, $\sqrt[3]{5}$ is the real cube root of 5. We know that $\sqrt{5}\,i \notin \mathbb{Q}(\sqrt[3]{5}\,)$, so

$$[\mathbb{Q}(\sqrt[3]{5}, \sqrt{5}\,i) : \mathbb{Q}(\sqrt[3]{5}\,)] = 2.$$

It is easy to determine that $\{1, \sqrt{5}i\,\}$ is a basis for $\mathbb{Q}(\sqrt[3]{5}, \sqrt{5}\,i)$ over $\mathbb{Q}(\sqrt[3]{5}\,)$. We also know that $\{1, \sqrt[3]{5}, (\sqrt[3]{5}\,)^2\}$ is a basis for $\mathbb{Q}(\sqrt[3]{5}\,)$ over $\mathbb{Q}$. Hence, a basis for $\mathbb{Q}(\sqrt[3]{5}, \sqrt{5}\,i)$ over $\mathbb{Q}$ is

$$\{1, \sqrt{5}\,i, \sqrt[3]{5}, (\sqrt[3]{5}\,)^2, (\sqrt[6]{5}\,)^5 i, (\sqrt[6]{5}\,)^7 i = 5\sqrt[6]{5}\,i \text{ or } \sqrt[6]{5}\,i\}.$$

Notice that $\sqrt[6]{5}\,i$ is a zero of $x^6 + 5$. We can show that this polynomial is irreducible over $\mathbb{Q}$ using Eisenstein's Criterion, where we let $p = 5$. Consequently,

$$\mathbb{Q} \subset \mathbb{Q}(\sqrt[6]{5}\,i) \subset \mathbb{Q}(\sqrt[3]{5}, \sqrt{5}\,i).$$

But it must be the case that $\mathbb{Q}(\sqrt[6]{5}\,i) = \mathbb{Q}(\sqrt[3]{5}, \sqrt{5}\,i)$, since the degree of both of these extensions is 6. □

Theorem 13.22 *Let E be a field extension of F. Then the following statements are equivalent.*

1. *E is a finite extension of F.*

2. *There exists a finite number of algebraic elements $\alpha_1, \ldots, \alpha_n \in E$ such that $E = F(\alpha_1, \ldots, \alpha_n)$.*

3. *There exists a sequence of fields*

$$E = F(\alpha_1, \ldots, \alpha_n) \supset F(\alpha_1, \ldots, \alpha_{n-1}) \supset \cdots \supset F(\alpha_1) \supset F,$$

where each field $F(\alpha_1, \ldots, \alpha_i)$ is algebraic over $F(\alpha_1, \ldots, \alpha_{i-1})$.

Proof. (1) ⇒ (2). Let E be a finite algebraic extension of F. Then E is a finite dimensional vector space over F and there exists a basis consisting of elements $\alpha_1, \ldots, \alpha_n$ in E such that $E = F(\alpha_1, \ldots, \alpha_n)$. Each α_i is algebraic over F by Theorem 13.15.

(2) ⇒ (3). Suppose that $E = F(\alpha_1, \ldots, \alpha_n)$, where every α_i is algebraic over F. Then

$$E = F(\alpha_1, \ldots, \alpha_n) \supset F(\alpha_1, \ldots, \alpha_{n-1}) \supset \cdots \supset F(\alpha_1) \supset F,$$

where each field $F(\alpha_1, \ldots, \alpha_i)$ is algebraic over $F(\alpha_1, \ldots, \alpha_{i-1})$.

(3) ⇒ (1). Let

$$E = F(\alpha_1, \ldots, \alpha_n) \supset F(\alpha_1, \ldots, \alpha_{n-1}) \supset \cdots \supset F(\alpha_1) \supset F,$$

where each field $F(\alpha_1, \ldots, \alpha_i)$ is algebraic over $F(\alpha_1, \ldots, \alpha_{i-1})$. Since

$$F(\alpha_1, \ldots, \alpha_i) = F(\alpha_1, \ldots, \alpha_{i-1})(\alpha_i)$$

is simple extension and α_i is algebraic over $F(\alpha_1, \ldots, \alpha_{i-1})$, it follows that

$$[F(\alpha_1, \ldots, \alpha_i) : F(\alpha_1, \ldots, \alpha_{i-1})]$$

is finite for each i. Therefore, $[E : F]$ is finite. ■

13.1.2 Algebraic Closure

Given a field F, the question arises as to whether or not we can find a field E such that every polynomial $p(x)$ has a root in E. This leads us to the following theorem.

Theorem 13.23 *Let E be an extension field of F. The set of elements in E that are algebraic over F form a field.*

Proof. Let $\alpha, \beta \in E$ be algebraic over F. Then $F(\alpha, \beta)$ is a finite extension of F. Since every element of $F(\alpha, \beta)$ is algebraic over F, $\alpha \pm \beta$, $\alpha\beta$, and α/β ($\beta \neq 0$) are all algebraic over F. Consequently, the set of elements in E that are algebraic over F form a field. ■

Corollary 13.24 *The set of all algebraic numbers forms a field; that is, the set of all complex numbers that are algebraic over $\mathbb{Q}$ makes up a field.*

Let E be a field extension of a field F. We define the **algebraic closure** of a field F in E to be the field consisting of all elements in E that are algebraic over F. A field F is **algebraically closed** if every nonconstant polynomial in $F[x]$ has a root in F.

Theorem 13.25 *A field F is algebraically closed if and only if every nonconstant polynomial in $F[x]$ factors into linear factors over $F[x]$.*

Proof. Let F be an algebraically closed field. If $p(x) \in F[x]$ is a nonconstant polynomial, then $p(x)$ has a zero in F, say α. Therefore, $x - \alpha$ must be a factor of $p(x)$ and so $p(x) = (x - \alpha)q_1(x)$, where $\deg q_1(x) = \deg p(x) - 1$. Continue this process with $q_1(x)$ to find a factorization

$$p(x) = (x - \alpha)(x - \beta)q_2(x),$$

where $\deg q_2(x) = \deg p(x) - 2$. The process must eventually stop since the degree of $p(x)$ is finite.

Conversely, suppose that every nonconstant polynomial $p(x)$ in $F[x]$ factors into linear factors. Let $ax - b$ be such a factor. Then $p(b/a) = 0$. Consequently, F is algebraically closed. ■

Corollary 13.26 *An algebraically closed field F has no proper algebraic extension E.*

Proof. Let E be an algebraic extension of F; then $F \subset E$. For $\alpha \in E$, the minimal polynomial of α is $x - \alpha$. Therefore, $\alpha \in F$ and $F = E$. ■

Theorem 13.27 *Every field F has a unique algebraic closure.*

It is a nontrivial fact that every field has a unique algebraic closure. The proof is not extremely difficult, but requires some rather sophisticated set theory. We refer the reader to [3], [4], or [8] for a proof of this result.

We now state the Fundamental Theorem of Algebra, first proven by Gauss at the age of 22 in his doctoral thesis. This theorem states that every polynomial with coefficients in the complex numbers has a root in the complex numbers.

Theorem 13.28 **Fundamental Theorem of Algebra.** *The field of complex numbers is algebraically closed.*

13.1.3 Historical Note

Algebraic number theory uses the tools of algebra to solve problems in number theory. Modern algebraic number theory began with Pierre de Fermat (1601–1665). Certainly we can find many positive integers that satisfy the equation $x^2 + y^2 = z^2$; Fermat conjectured that the equation $x^n + y^n = z^n$ has no positive integer solutions for $n \geq 3$. He stated in the margin of his copy of the Latin translation of Diophantus' *Arithmetica* that he had found a marvelous proof of this theorem, but that the margin of the book was too narrow to contain it. Building on work of other mathematicians, it was Andrew Wiles who finally succeeded in proving Fermat's Last Theorem in the 1990s. Wiles's achievement was reported on the front page of the *New York Times*.

Attempts to prove Fermat's Last Theorem have led to important contributions to algebraic number theory by such notable mathematicians as Leonhard Euler (1707–1783). Significant advances in the understanding of Fermat's Last Theorem were made by Ernst Kummer (1810–1893). Kummer's student, Leopold Kronecker (1823–1891), became one of the leading algebraists of the nineteenth century. Kronecker's theory of ideals and his study of algebraic number theory added much to the understanding of fields.

David Hilbert (1862–1943) and Hermann Minkowski (1864–1909) were among the mathematicians who led the way in this subject at the beginning of the twentieth century. Hilbert and Minkowski were both mathematicians at Göttingen University in Germany. Göttingen was truly one the most important centers of mathematical research during the last two centuries. The large number of exceptional mathematicians who studied there included Gauss, Dirichlet, Riemann, Dedekind, Noether, and Weyl.

André Weil answered questions in number theory using algebraic geometry, a field of mathematics that studies geometry by studying commutative rings. From about 1955 to 1970, Alexander Grothendieck dominated the field of algebraic geometry. Pierre Deligne, a student of Grothendieck, solved several of Weil's number-theoretic conjectures. One of the most recent contributions to algebra and number theory is Gerd Falting's proof of the Mordell-Weil conjecture. This conjecture of Mordell and Weil essentially says that certain polynomials $p(x, y)$ in $\mathbb{Z}[x, y]$ have only a finite number of integral solutions.

13.2 Exercises

1. Show that each of the following numbers is algebraic over $\mathbb{Q}$ by finding the minimal polynomial of the number over $\mathbb{Q}$.

(a) $\sqrt{1/3 + \sqrt{7}}$

(b) $\sqrt{3} + \sqrt[3]{5}$

(c) $\sqrt{3} + \sqrt{2}\, i$

(d) $\cos\theta + i\sin\theta$ for $\theta = 2\pi/n$ with $n \in \mathbb{N}$

(e) $\sqrt{\sqrt[3]{2} - i}$

Hint. (a) $x^4 - (2/3)x^2 - 62/9$; (c) $x^4 - 2x^2 + 25$.

2. Find a basis for each of the following field extensions. What is the degree of each extension?

(a) $\mathbb{Q}(\sqrt{3}, \sqrt{6}\,)$ over $\mathbb{Q}$

(b) $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{3}\,)$ over $\mathbb{Q}$

(c) $\mathbb{Q}(\sqrt{2}, i)$ over $\mathbb{Q}$

(d) $\mathbb{Q}(\sqrt{3}, \sqrt{5}, \sqrt{7}\,)$ over $\mathbb{Q}$

(e) $\mathbb{Q}(\sqrt{2}, \sqrt[3]{2}\,)$ over $\mathbb{Q}$

(f) $\mathbb{Q}(\sqrt{8}\,)$ over $\mathbb{Q}(\sqrt{2}\,)$

(g) $\mathbb{Q}(i, \sqrt{2} + i, \sqrt{3} + i)$ over $\mathbb{Q}$

(h) $\mathbb{Q}(\sqrt{2} + \sqrt{5}\,)$ over $\mathbb{Q}(\sqrt{5}\,)$

(i) $\mathbb{Q}(\sqrt{2}, \sqrt{6} + \sqrt{10}\,)$ over $\mathbb{Q}(\sqrt{3} + \sqrt{5}\,)$

Hint. (a) $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\,\}$; (c) $\{1, i, \sqrt{2}, \sqrt{2}\, i\}$; (e) $\{1, 2^{1/6}, 2^{1/3}, 2^{1/2}, 2^{2/3}, 2^{5/6}\}$.

3. Find the splitting field for each of the following polynomials.

(a) $x^4 - 10x^2 + 21$ over $\mathbb{Q}$

(b) $x^4 + 1$ over $\mathbb{Q}$

(c) $x^3 + 2x + 2$ over $\mathbb{Z}_3$

(d) $x^3 - 3$ over $\mathbb{Q}$

Hint. (a) $\mathbb{Q}(\sqrt{3}, \sqrt{7}\,)$.

4. Consider the field extension $\mathbb{Q}(\sqrt[4]{3}, i)$ over $\mathbb{Q}$.

(a) Find a basis for the field extension $\mathbb{Q}(\sqrt[4]{3}, i)$ over $\mathbb{Q}$. Conclude that $[\mathbb{Q}(\sqrt[4]{3}, i) : \mathbb{Q}] = 8$.

(b) Find all subfields F of$\mathbb{Q}(\sqrt[4]{3}, i)$ such that $[F : \mathbb{Q}] = 2$.

(c) Find all subfields F of $\mathbb{Q}(\sqrt[4]{3}, i)$ such that $[F : \mathbb{Q}] = 4$.

5. Show that $\mathbb{Z}_2[x]/\langle x^3+x+1\rangle$ is a field with eight elements. Construct a multiplication table for the multiplicative group of the field.

Hint. Use the fact that the elements of $\mathbb{Z}_2[x]/\langle x^3+x+1\rangle$ are 0, 1, α, $1+\alpha$, α^2, $1+\alpha^2$, $\alpha+\alpha^2$, $1+\alpha+\alpha^2$ and the fact that $\alpha^3+\alpha+1=0$.

6. Prove that $\mathbb{Q}(\sqrt{3}, \sqrt[4]{3}, \sqrt[8]{3}, \ldots)$ is an algebraic extension of $\mathbb{Q}$ but not a finite extension.

7. Prove or disprove: π is algebraic over $\mathbb{Q}(\pi^3)$.

8. Let $p(x)$ be a nonconstant polynomial of degree n in $F[x]$. Prove that there exists a splitting field E for $p(x)$ such that $[E:F]\leq n!$.

9. Prove or disprove: $\mathbb{Q}(\sqrt{2}\,)\cong\mathbb{Q}(\sqrt{3}\,)$.

10. Prove that the fields $\mathbb{Q}(\sqrt[4]{3}\,)$ and $\mathbb{Q}(\sqrt[4]{3}\,i)$ are isomorphic but not equal.

11. Let K be an algebraic extension of E, and E an algebraic extension of F. Prove that K is algebraic over F. [*Caution*: Do not assume that the extensions are finite.]

Hint. Suppose that E is algebraic over F and K is algebraic over E. Let $\alpha\in K$. It suffices to show that α is algebraic over some finite extension of F. Since α is algebraic over E, it must be the zero of some polynomial $p(x)=\beta_0+\beta_1x+\cdots+\beta_nx^n$ in $E[x]$. Hence α is algebraic over $F(\beta_0,\ldots,\beta_n)$.

12. Prove or disprove: $\mathbb{Z}[x]/\langle x^3-2\rangle$ is a field.

13. Let F be a field of characteristic p. Prove that $p(x)=x^p-a$ either is irreducible over F or splits in F.

14. Let E be the algebraic closure of a field F. Prove that every polynomial $p(x)$ in $F[x]$ splits in E.

15. If every irreducible polynomial $p(x)$ in $F[x]$ is linear, show that F is an algebraically closed field.

16. Prove that if α and β are constructible numbers such that $\beta\neq 0$, then so is α/β.

17. Show that the set of all elements in $\mathbb{R}$ that are algebraic over $\mathbb{Q}$ form a field extension of $\mathbb{Q}$ that is not finite.

18. 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 the minimal polynomial of α that are in E.

19. Show that $\mathbb{Q}(\sqrt{3},\sqrt{7}\,)=\mathbb{Q}(\sqrt{3}+\sqrt{7}\,)$. Extend your proof to show that $\mathbb{Q}(\sqrt{a},\sqrt{b}\,)=\mathbb{Q}(\sqrt{a}+\sqrt{b}\,)$, where $\gcd(a,b)=1$.

Hint. Since $\{1,\sqrt{3},\sqrt{7},\sqrt{21}\,\}$ is a basis for $\mathbb{Q}(\sqrt{3},\sqrt{7}\,)$ over $\mathbb{Q}$, $\mathbb{Q}(\sqrt{3},\sqrt{7}\,)\supset\mathbb{Q}(\sqrt{3}+\sqrt{7}\,)$. Since $[\mathbb{Q}(\sqrt{3},\sqrt{7}\,):\mathbb{Q}]=4$, $[\mathbb{Q}(\sqrt{3}+\sqrt{7}\,):\mathbb{Q}]=2$ or 4. Since the degree of the minimal polynomial of $\sqrt{3}+\sqrt{7}$ is 4, $\mathbb{Q}(\sqrt{3},\sqrt{7}\,)=\mathbb{Q}(\sqrt{3}+\sqrt{7}\,)$.

20. Let E be a finite extension of a field F. If $[E:F]=2$, show that E is a splitting field of F for some polynomial $f(x)\in F[x]$.

21. Prove or disprove: Given a polynomial $p(x)$ in $\mathbb{Z}_6[x]$, it is possible to construct a ring R such that $p(x)$ has a root in R.

22. Let E be a field extension of F and $\alpha\in E$. Determine $[F(\alpha):F(\alpha^3)]$.

23. Let α,β be transcendental over $\mathbb{Q}$. Prove that either $\alpha\beta$ or $\alpha+\beta$ is also transcendental.

24. Let E be an extension field of F and $\alpha \in E$ be transcendental over F. Prove that every element in $F(\alpha)$ that is not in F is also transcendental over F.

Hint. Let $\beta \in F(\alpha)$ not in F. Then $\beta = p(\alpha)/q(\alpha)$, where p and q are polynomials in α with $q(\alpha) \neq 0$ and coefficients in F. If β is algebraic over F, then there exists a polynomial $f(x) \in F[x]$ such that $f(\beta) = 0$. Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. Then

$$0 = f(\beta) = f\left(\frac{p(\alpha)}{q(\alpha)}\right) = a_0 + a_1\left(\frac{p(\alpha)}{q(\alpha)}\right) + \cdots + a_n\left(\frac{p(\alpha)}{q(\alpha)}\right)^n .$$

Now multiply both sides by $q(\alpha)^n$ to show that there is a polynomial in $F[x]$ that has α as a zero.

25. Let α be a root of an irreducible monic polynomial $p(x) \in F[x]$, with $\deg p = n$. Prove that $[F(\alpha) : F] = n$.

Hint. See the comments following Theorem 13.13.

13.3 Fields in the Secondary Classroom

This appendix will relate field extensions to the secondary classroom.

13.4 References and Suggested Readings

[1] Dean, R. A. *Elements of Abstract Algebra* . Wiley, New York, 1966.

[2] Dudley, U. *A Budget of Trisections.* Springer-Verlag, New York, 1987. An interesting and entertaining account of how not to trisect an angle.

[3] Fraleigh, J. B. *A First Course in Abstract Algebra.* 7th ed. Pearson, Upper Saddle River, NJ, 2003.

[4] Kaplansky, I. *Fields and Rings*, 2nd ed. University of Chicago Press, Chicago, 1972.

[5] Klein, F. *Famous Problems of Elementary Geometry.* Chelsea, New York, 1955.

[6] Martin, G. *Geometric Constructions.* Springer, New York, 1998.

[7] H. Pollard and H. G. Diamond. *Theory of Algebraic Numbers*, Dover, Mineola, NY, 2010.

[8] Walker, E. A. *Introduction to Abstract Algebra.* Random House, New York, 1987. This work contains a proof showing that every field has an algebraic closure.

Chapter 14

Constructions

14.1 Geometric Constructions

In ancient Greece, three classic problems were posed. These problems are geometric in nature and involve straightedge-and-compass constructions from what is now high school geometry; that is, we are allowed to use only a straightedge and compass to solve them. The problems can be stated as follows.

1. Given an arbitrary angle, can one trisect the angle into three equal subangles using only a straightedge and compass?
2. Given an arbitrary circle, can one construct a square with the same area using only a straightedge and compass?
3. Given a cube, can one construct the edge of another cube having twice the volume of the original? Again, we are only allowed to use a straightedge and compass to do the construction.

After puzzling mathematicians for over two thousand years, each of these constructions was finally shown to be impossible. We will use the theory of fields to provide a proof that the solutions do not exist. It is quite remarkable that the long-sought solution to each of these three geometric problems came from abstract algebra.

First, let us determine more specifically what we mean by a straightedge and compass, and also examine the nature of these problems in a bit more depth. To begin with, *a straightedge is not a ruler*. We cannot measure arbitrary lengths with a straightedge. It is merely a tool for drawing a line through two points. The statement that the trisection of an arbitrary angle is impossible means that there is at least one angle that is impossible to trisect with a straightedge-and-compass construction. Certainly it is possible to trisect an angle in special cases. We can construct a 30° angle; hence, it is possible to trisect a 90° angle. However, we will show that it is impossible to construct a 20° angle. Therefore, we cannot trisect a 60° angle.

14.1.1 Constructible Numbers

A real number α is **constructible** if we can construct a line segment of length $|\alpha|$ in a finite number of steps from a segment of unit length by using a straightedge and compass.

Theorem 14.1 *The set of all constructible real numbers forms a subfield F of the field of real numbers.*

Proof. Let α and β be constructible numbers. We must show that $\alpha+\beta$, $\alpha-\beta$, $\alpha\beta$, and α/β ($\beta \neq 0$) are also constructible numbers. We can assume that both α and β are positive with $\alpha > \beta$. It is quite obvious how to construct $\alpha+\beta$ and $\alpha-\beta$. To find a line segment with length $\alpha\beta$, we assume that $\beta > 1$ and construct the triangle in Figure 14.2 such that triangles $\triangle ABC$ and $\triangle ADE$ are similar. Since $\alpha/1 = x/\beta$, the line segment x has length $\alpha\beta$. A similar construction can be made if $\beta < 1$. We will leave it as an exercise to show that the same triangle can be used to construct α/β for $\beta \neq 0$. ■

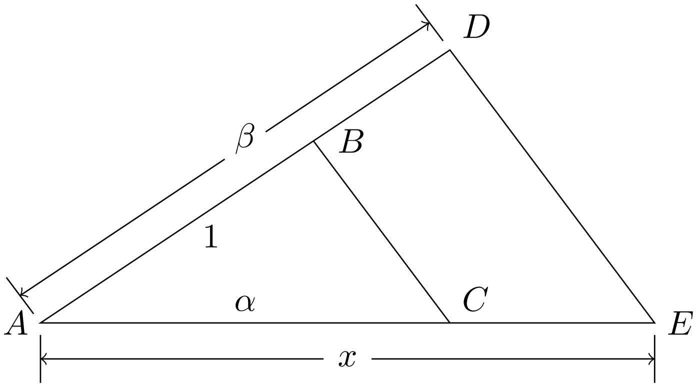


Figure 14.2 Construction of products

Lemma 14.3 *If α is a constructible number, then $\sqrt{\alpha}$ is a constructible number.*

Proof. In Figure 14.4 the triangles $\triangle ABD$, $\triangle BCD$, and $\triangle ABC$ are similar; hence, $1/x = x/\alpha$, or $x^2 = \alpha$. ■

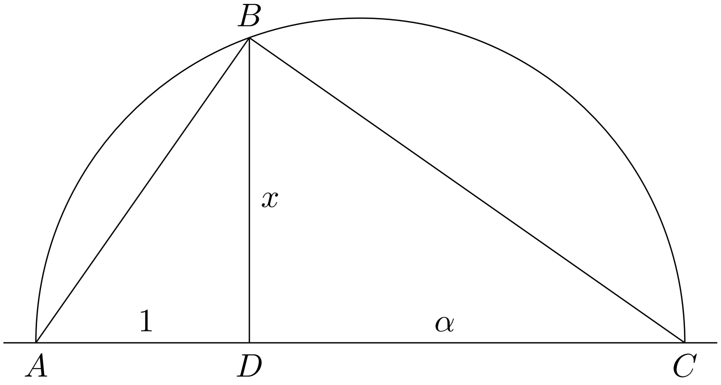


Figure 14.4 Construction of roots

By Theorem 14.1, we can locate in the plane any point $P = (p, q)$ that has rational coordinates p and q. We need to know what other points can be constructed with a compass and straightedge from points with rational coordinates.

Lemma 14.5 *Let F be a subfield of $\mathbb{R}$.*

1. *If a line contains two points in F, then it has the equation $ax+by+c=0$, where a, b, and c are in F.*

2. *If a circle has a center at a point with coordinates in F and a radius that is also in F, then it has the equation $x^2+y^2+dx+ey+f=0$, where d, e, and f are in F.*

Proof. Let (x_1, y_1) and (x_2, y_2) be points on a line whose coordinates are in F. If $x_1 = x_2$, then the equation of the line through the two points is $x - x_1 = 0$, which has the form $ax + by + c = 0$. If $x_1 \neq x_2$, then the equation of the line through the two points is given by

$$y - y_1 = \left(\frac{y_2 - y_1}{x_2 - x_1} \right) (x - x_1),$$

which can also be put into the proper form.

To prove the second part of the lemma, suppose that (x_1, y_1) is the center of a circle of radius r. Then the circle has the equation

$$(x - x_1)^2 + (y - y_1)^2 - r^2 = 0.$$

This equation can easily be put into the appropriate form. ■

Starting with a field of constructible numbers F, we have three possible ways of constructing additional points in $\mathbb{R}$ with a compass and straightedge.

1. To find possible new points in $\mathbb{R}$, we can take the intersection of two lines, each of which passes through two known points with coordinates in F.

2. The intersection of a line that passes through two points that have coordinates in F and a circle whose center has coordinates in F with radius of a length in F will give new points in $\mathbb{R}$.

3. We can obtain new points in $\mathbb{R}$ by intersecting two circles whose centers have coordinates in F and whose radii are of lengths in F.

The first case gives no new points in $\mathbb{R}$, since the solution of two equations of the form $ax + by + c = 0$ having coefficients in F will always be in F. The third case can be reduced to the second case. Let

$$\begin{aligned} x^2 + y^2 + d_1 x + e_1 y + f_1 &= 0 \\ x^2 + y^2 + d_2 x + e_2 y + f_2 &= 0 \end{aligned}$$

be the equations of two circles, where d_i, e_i, and f_i are in F for $i = 1, 2$. These circles have the same intersection as the circle

$$x^2 + y^2 + d_1 x + e_1 x + f_1 = 0$$

and the line

$$(d_1 - d_2)x + b(e_2 - e_1)y + (f_2 - f_1) = 0.$$

The last equation is that of the chord passing through the intersection points of the two circles. Hence, the intersection of two circles can be reduced to the case of an intersection of a line with a circle.

Considering the case of the intersection of a line and a circle, we must determine the nature of the solutions of the equations

$$\begin{aligned} ax + by + c &= 0 \\ x^2 + y^2 + dx + ey + f &= 0. \end{aligned}$$

If we eliminate y from these equations, we obtain an equation of the form $Ax^2 + Bx + C = 0$, where A, B, and C are in F. The x coordinate of the intersection points is given by

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

and is in $F(\sqrt{\alpha}\,)$, where $\alpha = B^2 - 4AC > 0$. We have proven the following lemma.

Lemma 14.6 *Let F be a field of constructible numbers. Then the points determined by the intersections of lines and circles in F lie in the field $F(\sqrt{\alpha}\,)$ for some α in F.*

Theorem 14.7 *A real number α is a constructible number if and only if there exists a sequence of fields*

$$\mathbb{Q} = F_0 \subset F_1 \subset \cdots \subset F_k$$

such that $F_i = F_{i-1}(\sqrt{\alpha_i}\,)$ with$\alpha_i \in F_i$ and $\alpha \in F_k$. In particular, there exists an integer $k > 0$ such that $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 2^k$.

Proof. The existence of the F_i's and the α_i's is a direct consequence of Lemma 14.6 and of the fact that

$$[F_k : \mathbb{Q}] = [F_k : F_{k-1}][F_{k-1} : F_{k-2}] \cdots [F_1 : \mathbb{Q}] = 2^k.$$

■

Corollary 14.8 *The field of all constructible numbers is an algebraic extension of $\mathbb{Q}$.*

As we can see by the field of constructible numbers, not every algebraic extension of a field is a finite extension.

14.1.2 Doubling the Cube and Squaring the Circle

We are now ready to investigate the classical problems of doubling the cube and squaring the circle. We can use the field of constructible numbers to show exactly when a particular geometric construction can be accomplished.

Doubling the cube is impossible. Given the edge of the cube, it is impossible to construct with a straightedge and compass the edge of the cube that has twice the volume of the original cube. Let the original cube have an edge of length 1 and, therefore, a volume of 1. If we could construct a cube having a volume of 2, then this new cube would have an edge of length $\sqrt[3]{2}$. However, $\sqrt[3]{2}$ is a zero of the irreducible polynomial $x^3 - 2$ over $\mathbb{Q}$; hence,

$$[\mathbb{Q}(\sqrt[3]{2}\,) : \mathbb{Q}] = 3.$$

This is impossible, since 3 is not a power of 2.

Squaring the circle. Suppose that we have a circle of radius 1. The area of the circle is π; therefore, we must be able to construct a square with side $\sqrt{\pi}$. This is impossible since π and consequently $\sqrt{\pi}$ are both transcendental. Therefore, using a straightedge and compass, it is not possible to construct a square with the same area as the circle.

14.1.3 Trisecting an Angle

Trisecting an arbitrary angle is impossible. We will show that it is impossible to construct a 20° angle. Consequently, a 60° angle cannot be trisected. We first need to calculate the triple-angle formula for the cosine:

$$\begin{aligned}
\cos 3\theta &= \cos(2\theta + \theta) \\
&= \cos 2\theta \cos\theta - \sin 2\theta \sin\theta \\
&= (2\cos^2\theta - 1)\cos\theta - 2\sin^2\theta\cos\theta \\
&= (2\cos^2\theta - 1)\cos\theta - 2(1 - \cos^2\theta)\cos\theta
\end{aligned}$$

$$= 4\cos^3\theta - 3\cos\theta.$$

The angle θ can be constructed if and only if $\alpha = \cos\theta$ is constructible. Let $\theta = 20^\circ$. Then $\cos 3\theta = \cos 60^\circ = 1/2$. By the triple-angle formula for the cosine,

$$4\alpha^3 - 3\alpha = \frac{1}{2}.$$

Therefore, α is a zero of $8x^3 - 6x - 1$. This polynomial has no factors in $\mathbb{Z}[x]$, and hence is irreducible over $\mathbb{Q}[x]$. Thus, $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$. Consequently, α cannot be a constructible number.

14.2 Exercises

1. Show that the regular 9-gon is not constructible with a straightedge and compass, but that the regular 20-gon is constructible.

2. Prove that the cosine of one degree ($\cos 1^\circ$) is algebraic over $\mathbb{Q}$ but not constructible.

3. Can a cube be constructed with three times the volume of a given cube?

Hint. False.

4. Prove that if α and β are constructible numbers such that $\beta \neq 0$, then so is α/β.

14.3 Geometric in the Secondary Classroom

This appendix will the field of constructible numbers to the secondary classroom.

Appendices

Appendix A

GNU Free Documentation License

Version 1.3, 3 November 2008

0. PREAMBLE. The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS. This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “publisher” means any person or entity that distributes copies of the Document to the public.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING. You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY. If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS. You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties — for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS. You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS. You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS. A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION. Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION. You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE. The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING. “Massive Multiauthor Collaboration Site” (or “MMC Site”) means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A “Massive Multiauthor Collaboration” (or “MMC”) contained in the site means any set of copyrightable works thus published on the MMC site.

“CC-BY-SA” means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

“Incorporate” means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is “eligible for relicensing” if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents. To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  YEAR  YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.3
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with… Texts.” line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Appendix B

Hints and Answers to Selected Exercises

I · An Introduction to Groups and Rings
1 · Sets, Functions, and Equivalence Relations
1.1 · Set Theory

· Exercises

1.1.3.

Hint. $(A \cap B) \cup (A \setminus B) \cup (B \setminus A) = (A \cap B) \cup (A \cap B') \cup (B \cap A') = [A \cap (B \cup B')] \cup (B \cap A') = A \cup (B \cap A') = (A \cup B) \cap (A \cup A') = A \cup B.$

1.1.7.

Hint. $A \setminus (B \cup C) = A \cap (B \cup C)' = (A \cap A) \cap (B' \cap C') = (A \cap B') \cap (A \cap C') = (A \setminus B) \cap (A \setminus C)$.

1.2 · Cartesian Products and Mappings

· Exercises

1.2.1.

Hint. (a) $A \times B = \{(a,1),(a,2),(a,3),(b,1),(b,2),(b,3),(c,1),(c,2),(c,3)\}$; (d) $A \times D = \emptyset$.

1.2.3.

Hint. (a) f is one-to-one but not onto. $f(\mathbb{R}) = \{x \in \mathbb{R} : x > 0\}$. (c) f is neither one-to-one nor onto. $f(\mathbb{R}) = \{x : -1 \leq x \leq 1\}$.

1.2.5.

Hint. (a) $f(n) = n + 1$.

1.2.6.

Hint. (a) Let $x, y \in A$. Then $g(f(x)) = (g \circ f)(x) = (g \circ f)(y) = g(f(y))$. Thus, $f(x) = f(y)$ and $x = y$, so $g \circ f$ is one-to-one. (b) Let $c \in C$, then $c = (g \circ f)(x) = g(f(x))$ for some $x \in A$. Since $f(x) \in B$, g is onto.

1.2.7.

Hint. $f^{-1}(x) = (x+1)/(x-1)$.

1.4 · Summary and Additional Exercises
1.4.2 · Additional Exercises

1.4.2.1.

Hint. (a) Let $y \in f(A_1 \cup A_2)$. Then there exists an $x \in A_1 \cup A_2$ such that $f(x) = y$. Hence, $y \in f(A_1)$ or $f(A_2)$. Therefore, $y \in f(A_1) \cup f(A_2)$. Consequently, $f(A_1 \cup A_2) \subset f(A_1) \cup f(A_2)$. Conversely, if $y \in f(A_1) \cup f(A_2)$, then $y \in f(A_1)$ or $f(A_2)$. Hence, there exists an x in A_1 or A_2 such that $f(x) = y$. Thus, there exists an $x \in A_1 \cup A_2$ such that $f(x) = y$. Therefore, $f(A_1) \cup f(A_2) \subset f(A_1 \cup A_2)$, and $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.

1.4.2.2.

Hint. Let $X = \mathbb{N} \cup \{\sqrt{2}\}$ and define $x \sim y$ if $x + y \in \mathbb{N}$.

2 · The Integers

2.3 · Prime Numbers

2.3.3 · Exercises

2.3.3.2.

Hint. Every prime must be of the form 2, 3, $6n + 1$, or $6n + 5$. Suppose there are only finitely many primes of the form $6k + 5$.

2.4 · Summary and Additional Exercises

2.4.2 · Additional Exercises

2.4.2.5.

Hint. Use the Principle of Well-Ordering and the division algorithm.

2.4.2.9.

Hint. Since $\gcd(a, b) = 1$, there exist integers r and s such that $ar + bs = 1$. Thus, $acr + bcs = c$.

3 · Groups

3.1 · Integer Equivalence Classes and Symmetries

3.1.4 · Exercises

3.1.4.3.

Hint. Let

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

be in S_n. All of the a_is must be distinct. There are n ways to choose a_1, $n - 1$ ways to choose $a_2, \ldots$, 2 ways to choose a_{n-1}, and only one way to choose a_n. Therefore, we can form σ in $n(n-1)\cdots 2 \cdot 1 = n!$ ways.

3.2 · Definitions and Examples

3.2.4 · Exercises

3.2.4.4.

Hint.

·	1	5	7	11
1	1	5	7	11
5	5	1	11	7
7	7	11	1	5
11	11	7	5	1

3.2.4.5.

Hint. Pick two matrices. Almost any pair will work.

3.2.4.10.

Hint. There is a nonabelian group containing six elements.

3.2.4.11.

Hint. Look at the symmetry group of an equilateral triangle or a square.

3.2.4.12.

Hint. The are five different groups of order 8.

3.2.4.13.

Hint.

$$\begin{aligned}(aba^{-1})^n &= (aba^{-1})(aba^{-1})\cdots(aba^{-1})\\ &= ab(aa^{-1})b(aa^{-1})b\cdots b(aa^{-1})ba^{-1}\\ &= ab^na^{-1}.\end{aligned}$$

3.2.4.15.

Hint. Since $abab = (ab)^2 = e = a^2b^2 = aabb$, we know that $ba = ab$.

3.2.4.16.

Hint. $ba = a^4b = a^3ab = ab$

3.3 · Subgroups

3.3.4 · Exercises

3.3.4.5.

Hint. Look at S_3.

3.3.4.9.

Hint. The identity of G is $1 = 1 + 0\sqrt{2}$. Since $(a + b\sqrt{2}\,)(c + d\sqrt{2}\,) = (ac+2bd)+(ad+bc)\sqrt{2}$, G is closed under multiplication. Finally, $(a+b\sqrt{2}\,)^{-1} = a/(a^2 - 2b^2) - b\sqrt{2}/(a^2 - 2b^2)$.

3.3.4.11.

Hint. $H_1 = \{\mathrm{id}\}$, $H_2 = \{\mathrm{id}, \rho_1, \rho_2\}$, $H_3 = \{\mathrm{id}, \mu_1\}$, $H_4 = \{\mathrm{id}, \mu_2\}$, $H_5 = \{\mathrm{id}, \mu_3\}$, S_3.

3.4 · Isomorphisms

· Exercises

3.4.1.

Hint. Define $\phi : \mathbb{C}^* \to GL_2(\mathbb{R})$ by

$$\phi(a + bi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

3.4.2.

Hint. False.

3.6 · Connections to the Secondary Classroom—Symmetry

· Exercises

3.6.10.

Hint. Any automorphism of $\mathbb{Z}_6$ must send 1 to another generator of $\mathbb{Z}_6$.

4 · Rings

4.1 · Rings

· Exercises

4.1.4.

Hint. (a) $\{1,3,7,9\}$; (c) $\{1,2,3,4,5,6\}$; (e)

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \right\}.$$

4.3 · Summary and Additional Exercises

4.3.2 · Additional Exercises

4.3.2.3.

Hint. Compute $(a+b)^2$ and $(-ab)^2$.

4.3.2.9.

Hint. Let $a/b, c/d \in \mathbb{Z}_{(p)}$. Then $a/b + c/d = (ad+bc)/bd$ and $(a/b)\cdot(c/d) = (ac)/(bd)$ are both in $\mathbb{Z}_{(p)}$, since $\gcd(bd,p)=1$.

4.3.2.10.

Hint. Suppose that $x^2 = x$ and $x \neq 0$. Since R is an integral domain, $x = 1$. To find a nontrivial idempotent, look in $\mathbb{M}_2(\mathbb{R})$.

II · Topics in Group Theory

5 · Cyclic Groups

5.1 · Cyclic Subgroups

5.1.3 · Exercises

5.1.3.1.

Hint. (a) 12; (c) infinite; (e) 10.

5.1.3.2.

Hint. (a) $7\mathbb{Z} = \{\ldots,-7,0,7,14,\ldots\}$; (b) $\{0,3,6,9,12,15,18,21\}$; (c) $\{0\}$, $\{0,6\}$, $\{0,4,8\}$, $\{0,3,6,9\}$, $\{0,2,4,6,8,10\}$; (g) $\{1,3,7,9\}$; (j) $\{1,-1,i,-i\}$.

5.1.3.3.

Hint. (a)

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

(c)

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

5.1.3.8.

Hint. (a) 0; (b) $1,-1$.

5.1.3.9.

Hint. $1,2,3,4,6,8,12,24$.

5.1.3.14.

Hint. $|\langle g\rangle \cap \langle h\rangle| = 1$.

5.2 · Multiplicative Group of Complex Numbers

5.2.3 · Exercises

5.2.3.1.

Hint. (a) $\sqrt{3}+i$; (c) -3.

5.2.3.2.

Hint. (a) $(1-i)/2$; (c) $16(i-\sqrt{3})$; (e) $-1/4$.

5.3 · Direct Products

5.3.4 · Exercises

5.3.4.1.

Hint. (a) 12; (c) 5.

5.3.4.6.

Hint. True.

5.4 · Summary and Additional Exercises

5.4.2 · Additional Exercises

5.4.2.8.

Hint. The identity element in any group has finite order. Let $g, h \in G$ have orders m and n, respectively. Since $(g^{-1})^m = e$ and $(gh)^{mn} = e$, the elements of finite order in G form a subgroup of G.

5.4.2.14.

Hint. If g is an element distinct from the identity in G, g must generate G; otherwise, $\langle g \rangle$ is a nontrivial proper subgroup of G.

5.4.2.22.

Hint. To show that ϕ is one-to-one, let $g_1 = h_1 k_1$ and $g_2 = h_2 k_2$ and consider $\phi(g_1) = \phi(g_2)$.

6 · Permutation Groups

6.1 · Definitions and Notation

6.1.7 · Exercises

6.1.7.1.

Hint. (a) (12453); (c) (13)(25).

6.1.7.2.

Hint. (a) (135)(24); (c) (14)(23); (e) (1324); (g) (134)(25); (n) (17352).

6.1.7.3.

Hint. (a) (16)(15)(13)(14); (c) (16)(14)(12).

6.1.7.4.

Hint. $(a_1, a_2, \ldots, a_n)^{-1} = (a_1, a_n, a_{n-1}, \ldots, a_2)$

6.1.7.5.

Hint. (a) $\{(13), (13)(24), (132), (134), (1324), (1342)\}$ is not a subgroup.

6.1.7.8.

Hint. (12345)(678).

6.1.7.11.

Hint. Permutations of the form

$$(1), (a_1, a_2)(a_3, a_4), (a_1, a_2, a_3), (a_1, a_2, a_3, a_4, a_5)$$

are possible for A_5.

6.1.7.13.

Hint. Calculate (123)(12) and (12)(123).

6.3 · Summary and Additional Exercises

6.3.2 · Exercises

6.3.2.7.

Hint. Consider the cases $(ab)(bc)$ and $(ab)(cd)$.

6.3.2.12.

Hint. For (a), show that $\sigma\tau\sigma^{-1}(\sigma(a_i)) = \sigma(a_{i+1})$.

6.3.2.18.

Hint. Draw the picture.

7 · Cosets and Lagrange's Theorem

7.1 · Cosets

· Exercises

7.1.1.

Hint. (a) $\langle 8 \rangle$, $1 + \langle 8 \rangle$, $2 + \langle 8 \rangle$, $3 + \langle 8 \rangle$, $4 + \langle 8 \rangle$, $5 + \langle 8 \rangle$, $6 + \langle 8 \rangle$, and $7 + \langle 8 \rangle$; (c) $3\mathbb{Z}$, $1 + 3\mathbb{Z}$, and $2 + 3\mathbb{Z}$.

7.1.5.

Hint. Let $g_1 \in gH$. Show that $g_1 \in Hg$ and thus $gH \subset Hg$.

7.2 · Lagrange's Theorem

· Exercises

7.2.1.

Hint. The order of g and the order h must both divide the order of G.

7.2.2.

Hint. This is true for every proper nontrivial subgroup.

7.2.3.

Hint. False.

7.4 · Summary and Additional Exercises

7.4.2 · Exercises

7.4.2.4.

Hint. Show that $g(H \cap K) = gH \cap gK$.

7.4.2.6.

Hint. If $\gcd(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$ (Exercise 2.4.2.8 in Chapter 2).

8 · Normal Subgroups and Homomorphisms

8.1 · Factor Groups and Normal Subgroups

8.1.4 · Exercises

8.1.4.8.

Hint. If $a \in G$ is a generator for G, then aH is a generator for G/H.

8.2 · Group Homomorphisms

· Exercises

8.2.3.

Hint. Since $\phi(m + n) = 7(m + n) = 7m + 7n = \phi(m) + \phi(n)$, ϕ is a homomorphism.

8.2.6.

Hint. Let $a, b \in G$. Then $\phi(a)\phi(b) = \phi(ab) = \phi(ba) = \phi(b)\phi(a)$.

8.4 · Summary and Additional Exercises

8.4.2 · Exercises

8.4.2.2.

Hint. (a) Let $g \in G$ and $h \in G'$. If $h = aba^{-1}b^{-1}$, then

$$\begin{aligned} ghg^{-1} &= gaba^{-1}b^{-1}g^{-1} \\ &= (gag^{-1})(gbg^{-1})(ga^{-1}g^{-1})(gb^{-1}g^{-1}) \\ &= (gag^{-1})(gbg^{-1})(gag^{-1})^{-1}(gbg^{-1})^{-1}. \end{aligned}$$

We also need to show that if $h = h_1 \cdots h_n$ with $h_i = a_i b_i a_i^{-1} b_i^{-1}$, then ghg^{-1} is a product of elements of the same type. However, $ghg^{-1} = gh_1 \cdots h_n g^{-1} = (gh_1g^{-1})(gh_2g^{-1}) \cdots (gh_ng^{-1})$.

8.4.2.3.

Hint. For any homomorphism $\phi : \mathbb{Z}_{24} \to \mathbb{Z}_{18}$, the kernel of ϕ must be a subgroup of $\mathbb{Z}_{24}$ and the image of ϕ must be a subgroup of $\mathbb{Z}_{18}$. Now use the fact that a generator must map to a generator.

8.4.2.6.

Hint. Find a counterexample.

III · Topics in Ring and Field Theory

9 · Ideals

9.4 · Exercises

9.4.1.

Hint. (a) $7\mathbb{Z}$ is a ring but not a field; (c) $\mathbb{Q}(\sqrt{2}\,)$ is a field; (f) R is not a ring.

9.4.3.

Hint. (a) $\{1, 3, 7, 9\}$; (c) $\{1, 2, 3, 4, 5, 6\}$; (e)

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \right\}.$$

9.4.4.

Hint. (a) $\{0\}$, $\{0, 9\}$, $\{0, 6, 12\}$, $\{0, 3, 6, 9, 12, 15\}$, $\{0, 2, 4, 6, 8, 10, 12, 14, 16\}$; (c) there are no nontrivial ideals.

9.4.7.

Hint. Assume there is an isomorphism $\phi : \mathbb{C} \to \mathbb{R}$ with $\phi(i) = a$.

9.4.8.

Hint. False. Assume there is an isomorphism $\phi : \mathbb{Q}(\sqrt{2}\,) \to \mathbb{Q}(\sqrt{3}\,)$ such that $\phi(\sqrt{2}\,) = a$.

9.4.12.

Hint. If $I \neq \{0\}$, show that $1 \in I$.

9.4.14.

Hint. (a) $\phi(a)\phi(b) = \phi(ab) = \phi(ba) = \phi(b)\phi(a)$.

9.4.22.

Hint. Let $a \in R$ with $a \neq 0$. Then the principal ideal generated by a is R. Thus, there exists a $b \in R$ such that $ab = 1$.

9.4.24.

Hint. Compute $(a+b)^2$ and $(-ab)^2$.

9.4.30.

Hint. Let $a/b, c/d \in \mathbb{Z}_{(p)}$. Then $a/b + c/d = (ad+bc)/bd$ and $(a/b)\cdot(c/d) = (ac)/(bd)$ are both in $\mathbb{Z}_{(p)}$, since $\gcd(bd, p) = 1$.

9.4.34.

Hint. Suppose that $x^2 = x$ and $x \neq 0$. Since R is an integral domain, $x = 1$. To find a nontrivial idempotent, look in $\mathbb{M}_2(\mathbb{R})$.

10 · Polynomials

10.2 · The Division Algorithm

· Exercises

10.2.4.

Hint. Factor the polynomials

10.2.9.

Hint. One factorization is $(x+2)(x+9)$. Are there other ways to multiply to 8 in $\mathbb{Z}_{10}$?

10.3 · Irreducible Polynomials

10.3.3 · Exercises

10.3.3.9.

Hint. For part (a), prove the contrapositive.

10.4 · Factoring over $\mathbb{C}$ and $\mathbb{R}$

10.4.3 · Exercises

10.4.3.1.

Hint. (a) $-3+3i$; (c) $43-18i$; (e) i

10.4.3.2.

Hint. (a) $\sqrt{3}+i$; (c) -3.

10.4.3.3.

Hint. (a) $\sqrt{2}\operatorname{cis}(7\pi/4)$; (c) $2\sqrt{2}\operatorname{cis}(\pi/4)$; (e) $3\operatorname{cis}(3\pi/2)$.

10.4.3.4.

Hint. (a) $(1-i)/2$; (c) $16(i-\sqrt{3})$; (e) $-1/4$.

10.4.3.9.

Hint. Do the factoring in that order.

10.5 · Exercises

10.5.2.

Hint. (a) $9x^2+2x+5$; (b) $8x^4+7x^3+2x^2+7x$.

10.5.3.

Hint. (a) $5x^3+6x^2-3x+4 = (5x^2+2x+1)(x-2)+6$; (c) $4x^5-x^3+x^2+4 = (4x^2+4)(x^3+3)+4x^2+2$.

10.5.5.

Hint. (a) No zeros in $\mathbb{Z}_{12}$; (c) 3, 4.

10.5.7.

Hint. Look at $(2x+1)$.

10.5.8.

Hint. (a) Reducible; (c) irreducible.

10.5.10.

Hint. One factorization is $x^2+x+8=(x+2)(x+9)$.

10.5.13.

Hint. The integers $\mathbb{Z}$ do not form a field.

10.5.14.

Hint. False.

10.5.16.

Hint. Let $\phi : R \to S$ be an isomorphism. Define $\overline{\phi} : R[x] \to S[x]$ by $\overline{\phi}(a_0+a_1x+\cdots+a_nx^n)=\phi(a_0)+\phi a_1)x+\cdots+\phi(a_n)x^n$.

10.5.20. Cyclotomic Polynomials.

Hint. The polynomial

$$\Phi_n(x)=\frac{x^n-1}{x-1}=x^{n-1}+x^{n-2}+\cdots+x+1$$

is called the **cyclotomic polynomial.** Show that $\Phi_p(x)$ is irreducible over $\mathbb{Q}$ for any prime p.

10.5.26.

Hint. Find a nontrivial proper ideal in $F[x]$.

11 · Integral Domains

11.2 · Exercises

11.2.1.

Hint. Note that $z^{-1}=1/(a+b\sqrt{3}\,i)=(a-b\sqrt{3}\,i)/(a^2+3b^2)$ is in $\mathbb{Z}[\sqrt{3}\,i]$ if and only if $a^2+3b^2=1$. The only integer solutions to the equation are $a=\pm1, b=0$.

11.2.2.

Hint. (a) $5=-i(1+2i)(2+i)$; (c) $6+8i=-i(1+i)^2(2+i)^2$.

11.2.4.

Hint. True.

11.2.9.

Hint. Let $z=a+bi$ and $w=c+di\neq0$ be in $\mathbb{Z}[i]$. Prove that $z/w\in\mathbb{Q}(i)$.

11.2.15.

Hint. Let $a=ub$ with u a unit. Then $\nu(b)\leq\nu(ub)\leq\nu(a)$. Similarly, $\nu(a)\leq\nu(b)$.

11.2.16.

Hint. Show that 21 can be factored in two different ways.

12 · Vector Spaces

12.4 · Exercises

12.4.3.

Hint. $\mathbb{Q}(\sqrt{2},\sqrt{3}\,)$ has basis $\{1,\sqrt{2},\sqrt{3},\sqrt{6}\,\}$ over $\mathbb{Q}$.

12.4.5.

Hint. The set $\{1, x, x^2, \ldots, x^{n-1}\}$ is a basis for P_n.

12.4.7.

Hint. (a) Subspace of dimension 2 with basis $\{(1, 0, -3), (0, 1, 2)\}$; (d) not a subspace.

12.4.10.

Hint. Since $0 = \alpha 0 = \alpha(-v + v) = \alpha(-v) + \alpha v$, it follows that $-\alpha v = \alpha(-v)$.

12.4.12.

Hint. Let $v_0 = 0, v_1, \ldots, v_n \in V$ and $\alpha_0 \neq 0, \alpha_1, \ldots, \alpha_n \in F$. Then $\alpha_0 v_0 + \cdots + \alpha_n v_n = 0$.

12.4.15. Linear Transformations.

Hint. (a) Let $u, v \in \ker(T)$ and $\alpha \in F$. Then

$$T(u + v) = T(u) + T(v) = 0$$
$$T(\alpha v) = \alpha T(v) = \alpha 0 = 0.$$

Hence, $u + v, \alpha v \in \ker(T)$, and $\ker(T)$ is a subspace of V.

(c) The statement that $T(u) = T(v)$ is equivalent to $T(u-v) = T(u)-T(v) = 0$, which is true if and only if $u - v = 0$ or $u = v$.

12.4.17. Direct Sums.

Hint. (a) Let $u, u' \in U$ and $v, v' \in V$. Then

$$(u + v) + (u' + v') = (u + u') + (v + v') \in U + V$$
$$\alpha(u + v) = \alpha u + \alpha v \in U + V.$$

13 · Fields

13.2 · Exercises

13.2.1.

Hint. (a) $x^4 - (2/3)x^2 - 62/9$; (c) $x^4 - 2x^2 + 25$.

13.2.2.

Hint. (a) $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$; (c) $\{1, i, \sqrt{2}, \sqrt{2}\, i\}$; (e) $\{1, 2^{1/6}, 2^{1/3}, 2^{1/2}, 2^{2/3}, 2^{5/6}\}$.

13.2.3.

Hint. (a) $\mathbb{Q}(\sqrt{3}, \sqrt{7})$.

13.2.5.

Hint. Use the fact that the elements of $\mathbb{Z}_2[x]/\langle x^3 + x + 1\rangle$ are 0, 1, α, $1 + \alpha$, α^2, $1 + \alpha^2$, $\alpha + \alpha^2$, $1 + \alpha + \alpha^2$ and the fact that $\alpha^3 + \alpha + 1 = 0$.

13.2.11.

Hint. Suppose that E is algebraic over F and K is algebraic over E. Let $\alpha \in K$. It suffices to show that α is algebraic over some finite extension of F. Since α is algebraic over E, it must be the zero of some polynomial $p(x) = \beta_0 + \beta_1 x + \cdots + \beta_n x^n$ in $E[x]$. Hence α is algebraic over $F(\beta_0, \ldots, \beta_n)$.

13.2.19.

Hint. Since $\{1, \sqrt{3}, \sqrt{7}, \sqrt{21}\}$ is a basis for $\mathbb{Q}(\sqrt{3}, \sqrt{7})$ over $\mathbb{Q}$, $\mathbb{Q}(\sqrt{3}, \sqrt{7}) \supset \mathbb{Q}(\sqrt{3} + \sqrt{7})$. Since $[\mathbb{Q}(\sqrt{3}, \sqrt{7}) : \mathbb{Q}] = 4$, $[\mathbb{Q}(\sqrt{3} + \sqrt{7}) : \mathbb{Q}] = 2$ or 4. Since the degree of the minimal polynomial of $\sqrt{3} + \sqrt{7}$ is 4, $\mathbb{Q}(\sqrt{3}, \sqrt{7}) = \mathbb{Q}(\sqrt{3} + \sqrt{7})$.

13.2.24.

Hint. Let $\beta \in F(\alpha)$ not in F. Then $\beta = p(\alpha)/q(\alpha)$, where p and q are polynomials in α with $q(\alpha) \neq 0$ and coefficients in F. If β is algebraic over F, then there exists a polynomial $f(x) \in F[x]$ such that $f(\beta) = 0$. Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. Then

$$0 = f(\beta) = f\left(\frac{p(\alpha)}{q(\alpha)}\right) = a_0 + a_1\left(\frac{p(\alpha)}{q(\alpha)}\right) + \cdots + a_n\left(\frac{p(\alpha)}{q(\alpha)}\right)^n.$$

Now multiply both sides by $q(\alpha)^n$ to show that there is a polynomial in $F[x]$ that has α as a zero.

13.2.25.

Hint. See the comments following Theorem 13.13.

14 · Constructions

14.2 · Exercises

14.2.3.

Hint. False.

Appendix C

Notation

The following table defines the notation used in this book. Page numbers or references refer to the first appearance of each symbol.

Symbol	Description	Page
$a \in A$	a is in the set A	5
$\mathbb{N}$	the natural numbers	6
$\mathbb{Z}$	the integers	6
$\mathbb{Q}$	the rational numbers	6
$\mathbb{R}$	the real numbers	6
$\mathbb{C}$	the complex numbers	6
$A \subset B$	A is a subset of B	6
$\emptyset$	the empty set	6
$A \cup B$	the union of sets A and B	6
$A \cap B$	the intersection of sets A and B	6
A'	complement of the set A	7
$A \setminus B$	difference between sets A and B	7
$A \times B$	Cartesian product of sets A and B	9
A^n	$A \times \cdots \times A$ (n times)	9
id	identity mapping	13
f^{-1}	inverse of the function f	13
$a \equiv b \pmod{n}$	a is congruent to b modulo n	17
$n!$	n factorial	24
$\binom{n}{k}$	binomial coefficient $n!/(k!(n-k)!)$	24
$\mathcal{P}(X)$	power set of X	26
$a \mid b$	a divides b	28
$\gcd(a,b)$	greatest common divisor of a and b	28
$\operatorname{lcm}(m,n)$	the least common multiple of m and n	33
$\mathbb{Z}_n$	the integers modulo n	36
$U(n)$	group of units in $\mathbb{Z}_n$	42
$\mathbb{M}_n(\mathbb{R})$	the $n \times n$ matrices with entries in $\mathbb{R}$	43
$\det A$	the determinant of A	44
$GL_n(\mathbb{R})$	the general linear group	44
Q_8	the group of quaternions	44
$\mathbb{C}^*$	the multiplicative group of complex numbers	44
$\lvert G \rvert$	the order of a group	44
$\mathbb{R}^*$	the multiplicative group of real numbers	48

(Continued on next page)

Symbol	**Description**	**Page**
$\mathbb{Q}^*$	the multiplicative group of rational numbers	48
$SL_n(\mathbb{R})$	the special linear group	49
$G \cong H$	G is isomorphic to a group H	51
$Z(G)$	the center of a group	56
$\operatorname{Aut}(G)$	automorphism group of a group G	58
i_g	$i_g(x) = gxg^{-1}$	59
$\operatorname{Inn}(G)$	inner automorphism group of a group G	59
$\mathbb{H}$	the ring of quaternions	62
$\mathbb{Z}[i]$	the Gaussian integers	65
$\operatorname{char} R$	characteristic of a ring R	66
$\mathbb{Z}_{(p)}$	ring of integers localized at p	69
$\langle a \rangle$	cyclic group generated by a	73
$\|a\|$	the order of an element a	74
$\operatorname{cis} \theta$	$\cos \theta + i \sin \theta$	80
$\mathbb{T}$	the circle group	82
S_n	the symmetric group on n letters	92
$(a_1, a_2, \ldots, a_k)$	cycle of length k	94
A_n	the alternating group on n letters	97
D_n	the dihedral group	101
ρ_g	right regular representation	109
$[G : H]$	index of a subgroup H in a group G	113
$\mathcal{L}_H$	the set of left cosets of a subgroup H in a group G	113
$\mathcal{R}_H$	the set of right cosets of a subgroup H in a group G	113
$a \nmid b$	a does not divide b	117
G/N	factor group of G mod N	121
$\ker \phi$	kernel of ϕ	125
G'	commutator subgroup of G	131
$\mathbb{Z}_{(p)}$	ring of integers localized at p	146
$\deg f(x)$	degree of a polynomial	149
$R[x]$	ring of polynomials over a ring R	149
$R[x_1, x_2, \ldots, x_n]$	ring of polynomials in n indeterminants	151
ϕ_α	evaluation homomorphism at α	151
$\operatorname{cis} \theta$	$\cos \theta + i \sin \theta$	163
$\mathbb{Q}(x)$	field of rational functions over $\mathbb{Q}$	177
$F(x)$	field of rational functions in x	178
$F(x_1, \ldots, x_n)$	field of rational functions in $x_1, \ldots, x_n$	178
$\dim V$	dimension of a vector space V	185
$U \oplus V$	direct sum of vector spaces U and V	187
$\operatorname{Hom}(V, W)$	set of all linear transformations from U into V	188
V^*	dual of a vector space V	188
$F(\alpha_1, \ldots, \alpha_n)$	smallest field containing F and $\alpha_1, \ldots, \alpha_n$	191
$[E : F]$	dimension of a field extension of E over F	194

Index

Colophon

This book was authored and produced with PreTeXt[1].

[1] https://pretextbook.org