

Introduction to Modern Algebra

David Joyce
Clark University

Version 1.2.7, 5 Dec 2017 ¹

¹Copyright (C) 2008,2017.

I dedicate this book to my friend and colleague Arthur Chou. Arthur encouraged me to write this book. I'm sorry that he did not live to see it finished.

Arthur was born in 1954 in Taipei, Taiwan. He received his bachelors in mathematics in 1976 from Tunghai University and his PhD from Stony Brook in 1982. After a year at the Institute for Advanced Study at Princeton, he joined Clark University in 1983. He was promoted to Associate Professor six years later and promoted to full professor in 2008, the year he died. Besides mathematics, he had many other interests. Among other things, he was the general manager of the North America Elite Youth Orchestra which performed at Dallas, Beijing, and Taipei, and he was the deacon of the Chinese Gospel Church in Southborough, Massachusetts.

Contents

List of Figures	vi
List of Tables	viii
1 Introduction	1
1.1 Algebra	1
1.2 Structures in Modern Algebra	2
1.2.1 Operations on sets	2
1.2.2 Fields	4
1.2.3 Rings	5
1.2.4 Groups	6
1.2.5 Other algebraic structures besides fields, rings, and groups	10
1.3 Isomorphisms, homomorphisms, etc.	11
1.3.1 Isomorphisms	11
1.3.2 Homomorphisms	12
1.3.3 Monomorphisms and epimorphisms	13
1.3.4 Endomorphisms and automorphisms	14
1.4 A little number theory	15
1.4.1 Mathematical induction on the natural numbers \mathbf{N}	15
1.4.2 Divisibility	16
1.4.3 Prime numbers	17
1.4.4 The Euclidean algorithm	19
1.5 The fundamental theorem of arithmetic	22
1.6 Polynomials.	25
1.6.1 Division for polynomials	26
1.6.2 Roots of unity and cyclotomic polynomials	28
2 Fields	31
2.1 Introduction to fields	31
2.1.1 Definition of fields	31
2.1.2 Subtraction, division, multiples, and powers	32
2.1.3 Properties that follow from the axioms	33
2.1.4 Subfields	34
2.1.5 Fields of rational functions	35
2.1.6 Vector spaces over arbitrary fields	35
2.2 Cyclic rings and finite fields	35

2.2.1	The cyclic ring \mathbf{Z}_n	36
2.2.2	The cyclic prime fields \mathbf{Z}_p	39
2.2.3	Characteristics of fields, and prime fields	41
2.3	Field Extensions, algebraic fields, the complex numbers	41
2.3.1	Algebraic fields	42
2.3.2	The field of complex numbers \mathbf{C}	43
2.3.3	General quadratic extensions	44
2.4	Real numbers and ordered fields	45
2.4.1	Ordered fields	45
2.4.2	Archimedean orders	47
2.4.3	Complete ordered fields	49
2.5	Skew fields (division rings) and the quaternions	50
2.5.1	Skew fields (division rings)	50
2.5.2	The quaternions \mathbf{H}	51
3	Rings	55
3.1	Introduction to rings	55
3.1.1	Definition and properties of rings	55
3.1.2	Products of rings	57
3.1.3	Integral domains	57
3.1.4	The Gaussian integers, $\mathbf{Z}[i]$	59
3.1.5	Finite fields again	59
3.2	Factoring \mathbf{Z}_n by the Chinese remainder theorem	60
3.2.1	The Chinese remainder theorem	60
3.2.2	Brahmagupta's solution	62
3.2.3	Qin Jiushao's solution	62
3.3	Boolean rings	63
3.3.1	Introduction to Boolean rings	64
3.3.2	Factoring Boolean rings	65
3.3.3	A partial order on a Boolean ring	66
3.4	The field of rational numbers, fields of fractions	67
3.5	Categories and the category of rings	69
3.5.1	The formal definition of categories	70
3.5.2	The category \mathcal{R} of rings	71
3.5.3	Monomorphisms and epimorphisms in a category	73
3.6	Kernels, ideals, and quotient rings	74
3.6.1	Kernels of ring homomorphisms	74
3.6.2	Ideals of a ring	74
3.6.3	Quotient rings, R/I	76
3.6.4	Prime and maximal ideals	78
3.7	Krull's theorem	79
3.8	UFDs, PIDs, and EDs	80
3.8.1	Divisibility in an integral domain	80
3.8.2	Unique factorization domains	81
3.8.3	Principal ideal domains	82
3.8.4	Euclidean domains	84

3.9	Real and complex polynomial rings $\mathbf{R}[x]$ and $\mathbf{C}[x]$	87
3.9.1	$\mathbf{C}[x]$ and the Fundamental Theorem of Algebra	87
3.9.2	The polynomial ring $\mathbf{R}[x]$	89
3.10	Rational and integer polynomial rings	90
3.10.1	Roots of polynomials	90
3.10.2	Gauss's lemma and Eisenstein's criterion	92
3.10.3	Prime cyclotomic polynomials	95
3.10.4	Polynomial rings with coefficients in a UFD, and polynomial rings in several variables.	95
3.11	Number fields and their rings of integers	97
4	Groups	99
4.1	Groups and subgroups	99
4.1.1	Definition and basic properties of groups	99
4.1.2	Subgroups	100
4.1.3	Cyclic groups and subgroups	101
4.1.4	Products of groups	102
4.1.5	Cosets and Lagrange's theorem	103
4.2	Symmetric Groups S_n	104
4.2.1	Permutations and the symmetric group	104
4.2.2	Even and odd permutations	106
4.2.3	Alternating and dihedral groups	107
4.3	Cayley's theorem and Cayley graphs	110
4.3.1	Cayley's theorem	110
4.3.2	Some small finite groups	112
4.4	The category of groups \mathcal{G}	115
4.5	Conjugacy classes and quandles	115
4.5.1	Conjugacy classes	116
4.5.2	Quandles and the operation of conjugation	117
4.6	Kernels, normal subgroups, and quotient groups	120
4.6.1	Kernels of group homomorphisms and normal subgroups	120
4.6.2	Quotient groups, and projections $\gamma : G \rightarrow G/N$	121
4.6.3	Isomorphism theorems	122
4.6.4	Internal direct products	123
4.7	Matrix rings and linear groups	124
4.7.1	Linear transformations	124
4.7.2	The general linear groups $GL_n(R)$	125
4.7.3	Other linear groups	126
4.7.4	Projective space and the projective linear group $PGL_n(F)$	127
4.8	Structure of finite groups	130
4.8.1	Simple groups	131
4.8.2	The Jordan-Hölder theorem	131
4.9	Abelian groups	134
4.9.1	The category \mathcal{A} of Abelian groups	136
4.9.2	Finite Abelian groups	137

Appendices	141
A Background mathematics	143
A.1 Logic and proofs	143
A.2 Sets	144
A.2.1 Basic set theory	144
A.2.2 Functions and relations	149
A.2.3 Equivalence relations	150
A.2.4 Axioms of set theory	151
A.3 Ordered structures	153
A.3.1 Partial orders and posets.	153
A.3.2 Lattices	155
A.3.3 Boolean algebras.	157
A.4 Axiom of choice	158
A.4.1 Zorn's lemma	158
A.4.2 Well-ordering principle	159
Index	161

List of Figures

1.1	Equilateral triangle with lines of symmetry	8
1.2	Unit circle S^1	10
1.3	Divisors of 432	17
1.4	Divisibility up through 12	17
2.1	Cyclic rings $\mathbf{Z}_6, \mathbf{Z}_{19}, \mathbf{Z}$	37
3.1	Lattice of Gaussian integers $\mathbf{Z}[i]$	59
3.2	Free Boolean ring on two elements	66
3.3	Lattice of Eisenstein integers	85
3.4	Primitive 7 th roots of unity	95
4.1	Subgroups of S_3	107
4.2	Symmetries of a pentagon	108
4.3	Symmetries of a cube and tetrahedron	109
4.4	Cayley graph for D_5	111
4.5	Cayley graph for A_4	113
4.6	Distributivity in a involutory quandle	118
4.7	A conjugacy class in the quaternion group	119
4.8	The conjugacy class of transpositions in S_4	120
4.9	The Fano plane $\mathbf{Z}_2 P^2$	128
4.10	The projective plane $\mathbf{Z}_3 P^2$	129
4.11	Cayley graph of the Frobenius group $F_{21} = C_7 \rtimes C_3$	134
4.12	Heptahedron on a torus	135
A.1	Lattice of the Powerset of 4 elements	157

List of Tables

1.1	Composition table for six particular rational functions.	9
3.1	Notations in Boolean algebras, set theory, and Boolean rings.	65
4.1	List of small groups	114
A.1	Standard logical symbols	144

Chapter 1

Introduction

1.1 Algebra

The word “algebra” means many things. The word dates back about 1200 years ago to part of the title of al-Khwārizmī’s book on the subject, but the subject itself goes back 4000 years ago to ancient Babylonia and Egypt. It was about solving numerical problems that we would now identify as linear and quadratic equations. Versions of the quadratic formula were used to find solutions to those quadratic equations. Al-Khwārizmī (ca. 780–ca. 850) codified the algorithms (“algorithm” is a word derived from his name) for solving these equations. He wrote all his equations out in words since symbolic algebra had yet to be invented.

Other places in the world also had algebra and developed various aspects of it. The ancient Chinese solved systems of simultaneous linear equations and later developed algorithms to find roots of polynomials of high degree. Various aspects of number theory were studied in China, in India, and by Greek mathematicians.

Symbolic algebra was developed in the 1500s. Symbolic algebra has symbols for the arithmetic operations of addition, subtraction, multiplication, division, powers, and roots as well as symbols for grouping expressions (such as parentheses), and most importantly, used letters for variables.

Once symbolic algebra was developed in the 1500s, mathematics flourished in the 1600s. Coordinates, analytic geometry, and calculus with derivatives, integrals, and series were developed in that century.

Algebra became more general and more abstract in the 1800s as more algebraic structures were invented. Hamilton (1805–1865) invented quaternions (see section 2.5.2) and Grassmann (1809–1977) developed exterior algebras in the 1840s, both of which led to vector spaces. (See section 2.1.6 for vector spaces.)

Groups were developed over the 1800s, first as particular groups of substitutions or permutations, then in the 1850’s Cayley (1821–1895) gave the general definition for a group. (See chapter 2 for groups.)

Several fields were studied in mathematics for some time including the field of real numbers the field of rational number, and the field of complex numbers, but there was no general definition for a field until the late 1800s. (See chapter 2 for fields.)

Rings also were studied in the 1800s. Noether (1882–1935) gave general concept of commutative ring in 1921 which was later generalized to include noncommutative rings. (See

chapter 3 for rings.)

We'll introduce the concepts of field, ring, and group in the Introduction, then study each in turn in the following chapters.

1.2 Structures in Modern Algebra

Fields, rings, and groups. We'll be looking at several kinds of algebraic structures this semester, the three major kinds being fields in chapter 2, rings in chapter 3, and groups in chapter 4, but also minor variants of these structures.

We'll start by examining the definitions and looking at some examples. For the time being, we won't prove anything; that will come in later chapters when we look at those structures in depth.

A note on notation. We'll use the standard notation for various kinds of numbers. The set of natural numbers, $\{0, 1, 2, \dots\}$ is denoted **N**. The set of integers $\{\dots, -2, -1, 0, 1, 2, \dots\}$ is denoted **Z** (for Zahlen, German for whole number). The set of rational numbers, that is, numbers of the form $\frac{m}{n}$ where m is an integer and n is a nonzero integer, is denoted **Q** (for “quotient”). The set of all real numbers, including all positive numbers, all negative numbers, and 0, is denoted **R**. And the set of complex numbers, that is, numbers of the form $x + iy$ where x and y are real numbers and $i^2 = -1$, is denoted **C**.

1.2.1 Operations on sets

For background on sets, see the section A.2 in the appendix.

We're familiar with many operations on the real numbers **R**—addition, subtraction, multiplication, division, negation, reciprocation, powers, roots, etc.

Addition, subtraction, and multiplication are examples of binary operations, that is, functions $\mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ which take two real numbers as their arguments and return another real number. Division is almost a binary operation, but since division by 0 is not defined, it's only a partially defined binary operation. Most of our operations will be defined everywhere, but some, like division, won't be.

Negation is a unary operation, that is, a function $\mathbf{R} \rightarrow \mathbf{R}$ which takes one real number as an argument and returns a real number. Reciprocation is a partial unary operation since the reciprocal of zero is not defined.

The operations we'll consider are all binary or unary. Ternary operations can certainly be defined, but useful ternary operations are rare.

Some of these operations satisfy familiar identities. For example, addition and multiplication are both commutative; they satisfy the identities

$$x + y = y + x \quad \text{and} \quad xy = yx.$$

A binary operation is said to be *commutative* when the order that the two arguments are applied doesn't matter, that is, interchanging them, or commuting one across the other, doesn't change the result. Subtraction and division, however, are not commutative.

Addition and multiplication are also associative binary operations

$$(x + y) + z = x + (y + z) \quad \text{and} \quad (xy)z = x(yz).$$

A binary operation is said to be *associative* when the parentheses can be associated with either the first pair or the second pair when the operation is applied to three arguments and the result is the same. Neither subtraction nor division are associative.

Both addition and multiplication also have identity elements

$$0 + x = x = x + 0 \quad \text{and} \quad 1x = x = x1.$$

An *identity element*, also called a *neutral element*, for a binary operation is an element in the set that doesn't change the value of other elements when combined with them under the operation. So, 0 is the identity element for addition, and 1 is the identity element for multiplication. Subtraction and division don't have identity elements. (Well, they do on the right, since $x - 0 = x$ and $\frac{x}{1} = x$, but not on the left, since usually $0 - x \neq x$ and $\frac{1}{x} \neq x$.)

Also, there are additive inverses and multiplicative inverses (for nonzero) elements. That is to say, given any x there is another element, namely $-x$, such that $x + (-x) = 0$, and given any nonzero x there is another element, namely $\frac{1}{x}$ such that $x \frac{1}{x} = 1$. Thus, a binary operation that has an identity element is said to have *inverses* if for each element there is an inverse element such that when combined by the operation they yield the identity element for the operation. Addition has inverses, and multiplication has inverses of nonzero elements.

Finally, there is a particular relation between the operations of addition and multiplication, that of distributivity:

$$x(y + z) = xy + xz \quad \text{and} \quad (y + z)x = yx + zx.$$

Multiplication distributes over addition, that is, when multiplying a sum by x we can distribute the x over the terms of the sum.

Exercise 1. On properties of operations.

- (a). Is the binary operation $x * y = \frac{xy}{x + y}$ for positive x and y a commutative operation?

That is, is it true that $x * y = y * x$ for all positive x and y ? Is it associative? Explain your answer.

(b). Is it true that $(w - x) - (y - z) = (w - y) - (x - z)$ is an identity for real numbers? Can you say why or why not? (The word "identity" is used for an equation which holds whenever both sides of the equation are defined and are equal.)

(c). Although multiplication in \mathbf{R} distributes over addition, addition doesn't distribute over multiplication. Give an example where it doesn't.

Algebraic structures. We'll define fields, rings, and groups as three kinds of algebraic structures. An algebraic structure will have an underlying set, binary operations, unary operations, and constants, that have some of the properties mentioned above like commutativity, associativity, identity elements, inverse elements, and distributivity. Different kinds of structures will have different operations and properties.

The algebraic structures are abstractions of familiar ones like those on the real numbers \mathbf{R} , but for each kind of structure there will be more than one example, as we'll see.

1.2.2 Fields

Informally, a field is a set equipped with four operations—addition, subtraction, multiplication, and division that have the usual properties. (They don't have to have the other operations that \mathbf{R} has, like powers, roots, logs, and the myriad other functions like $\sin x$.)

Definition 1.1 (Field). A *field* is a set equipped with two binary operations, one called *addition* and the other called *multiplication*, denoted in the usual manner, which are both commutative and associative, both have identity elements (the additive identity denoted 0 and the multiplicative identity denoted 1), addition has inverse elements (the inverse of x denoted $-x$), multiplication has inverses of nonzero elements (the inverse of x denoted $\frac{1}{x}$ or x^{-1}), multiplication distributes over addition, and $0 \neq 1$.

This definition will be spelled out in detail in chapter 2.

Of course, one example of a field is the field of real numbers \mathbf{R} . What are some others?

Example 1.2 (The field of rational numbers, \mathbf{Q}). Another example is the field of rational numbers. A rational number is the quotient of two integers a/b where the denominator is not 0. The set of all rational numbers is denoted \mathbf{Q} . We're familiar with the fact that the sum, difference, product, and quotient (when the denominator is not zero) of rational numbers is another rational number, so \mathbf{Q} has all the operations it needs to be a field, and since it's part of the field of the real numbers \mathbf{R} , its operations have the properties necessary to be a field. We say that \mathbf{Q} is a *subfield* of \mathbf{R} and that \mathbf{R} is an *extension* of \mathbf{Q} . But \mathbf{Q} is not all of \mathbf{R} since there are irrational numbers like $\sqrt{2}$.

Example 1.3 (The field of complex numbers, \mathbf{C}). Yet another example is the field of complex numbers \mathbf{C} . A complex number is a number of the form $a+bi$ where a and b are real numbers and $i^2 = -1$. The field of real numbers \mathbf{R} is a subfield of \mathbf{C} . We'll review complex numbers before we use them. See *Dave's Short Course on Complex Numbers* at <http://www.clarku.edu/~djoyce/complex>

In chapter 2, we'll study fields in detail, and we'll look at many other fields. Some will only have a finite number of elements. (They won't be subfields of \mathbf{Q} .) Some will have \mathbf{Q} as a subfield but be subfields themselves of \mathbf{R} or \mathbf{C} . Some will be even larger.

Exercise 2. On fields. None of the following are fields. In each case, the operations of addition and multiplication are the usual ones.

- (a). The integers \mathbf{Z} do not form a field. Why not?
- (b). The positive real numbers $\{x \in \mathbf{R} \mid x > 0\}$ do not form a field. Why not?
- (c). The set of real numbers between -10 and 10 , that is,

$$(-10, 10) = \{x \in \mathbf{R} \mid -10 < x < 10\}$$

is not a field. Why not?

1.2.3 Rings

Rings will have the three operations of addition, subtraction, and multiplication, but don't necessarily have division. Most of our rings will have commutative multiplication, but some won't, so we won't require that multiplication be commutative in our definition. All the rings we'll look at have a multiplicative identity, 1, so we'll include that in the definition.

Definition 1.4 (Ring). A *ring* is a set equipped with two binary operations, one called *addition* and the other called *multiplication*, denoted in the usual manner, which are both associative, addition is commutative, both have identity elements (the additive identity denoted 0 and the multiplicative identity denoted 1), addition has inverse elements (the inverse of x denoted $-x$), and multiplication distributes over addition. If multiplication is also commutative, then the ring is called a *commutative ring*.

Of course, all fields are automatically rings, in fact commutative rings, but what are some other rings?

Example 1.5 (The ring of integers, \mathbf{Z}). The ring of integers \mathbf{Z} includes all integers (whole numbers)—positive, negative, or 0. Addition, subtraction, and multiplication satisfy the requirements for a ring, indeed, a commutative ring. But there are no multiplicative inverses for any elements except 1 and -1 . For instance, $1/2$ is not an integer. We'll find that although the ring of integers looks like it has less structure than a field, this very lack of structure allows us to discover more about integers. We'll be able to talk about prime numbers, for example.

Example 1.6 (Polynomial rings). A whole family of examples are the rings of polynomials. Let R be any commutative ring (perhaps a field), and let $R[x]$ include all polynomials with coefficients in R . We know how to add, subtract, and multiply polynomials, and these operations have the properties required to make $R[x]$ a commutative ring. We have, for instance, the ring of polynomials with real coefficients $\mathbf{R}[x]$, the ring with integral coefficients $\mathbf{Z}[x]$, etc.

Example 1.7 (Matrix rings). How about an example ring that's not commutative? The ring of $n \times n$ matrices with entries in a commutative ring R gives such an example, this ring being denoted $M_n(R)$. This ring, $M_n(R)$, won't be commutative when $n \geq 2$. An example of a matrix ring is the ring of 2×2 matrices with real entries, $M_2(\mathbf{R})$. Addition and subtraction are computed coordinatewise. The additive identity, 0, of this matrix ring is the matrix with all 0 entries, $0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Matrix multiplication is not coordinatewise, but it is associative, and multiplication does distribute over addition. The multiplicative identity for this matrix ring is what's usually called the identity matrix, denoted I . It has 1's down the main diagonal and 0's elsewhere, $1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Sylvester (1814–1897), in 1850, called rectangular arrangements of numbers *matrices*, and Cayley wrote much about them in his papers of 1855–1858.

Example 1.8 (Integers modulo n). An important family of rings is the ring of integers modulo n . We'll study this in more detail later in section 2.2, but here's an incomplete overview. Fix a positive integer n . Think of two integers a and b as being the same modulo n if n divides $b - a$. In that case, we'll say that a and b are *congruent modulo n* , and we'll

use the notation Gauss (1777–1855) developed, $a \equiv b \pmod{n}$, to denote that congruence. Congruence is commonly used in the study of number theory. This meaning of the Latin word “modulo” was introduced into mathematics by Gauss in 1801.

Note that there are only n distinct integers modulo n , namely 0 through $n - 1$, since those are the only remainders you can get when you divide an integer by n . These remainders are also called “residues”. We can represent integers modulo n by these remainders from 0 through $n - 1$. Thus, we’ll say, for instance, that 5 plus 3 equals 1 modulo 7, by which we mean $5 + 3 \equiv 1 \pmod{7}$. Thus, we can turn congruence modulo n , which is an equivalence relation on \mathbf{Z} into equality on an n -element set. That n -element set is denoted $\mathbf{Z}/n\mathbf{Z}$, read \mathbf{Z} modulo $n\mathbf{Z}$, or more simply as \mathbf{Z}_n , read \mathbf{Z} mod n . So, we can take the elements of \mathbf{Z}_n to be the integers from 0 through $n - 1$, where we understand that addition, subtraction, and multiplication are done modulo n . And it turns out that this is a ring, as we’ll see when we study \mathbf{Z}_n in detail.

Incidentally, when n is a prime number p , then \mathbf{Z}_p is not just a ring, but a field, as will be discussed in section 2.2.

Exercise 3. On rings. None of the following are rings. In each case, the operations of addition and multiplication are the usual ones.

- (a). The set of nonzero integers, $\{x \in \mathbf{Z} \mid x \neq 0\}$ is not a ring. Why not?
- (b). The set of even integers $\{2x \mid x \in \mathbf{Z}\}$ is not a ring. Why not?
- (c). The set of odd degree polynomials with real coefficients

$$\{f(x) \in \mathbf{R}[x] \mid \text{the degree of } f(x) \text{ is odd}\}$$

is not a ring. Why not? (How about the set of even degree polynomials?)

Exercise 4. On noncommutative rings. Are the following rings? (The operations are the usual matrix operations.) Explain in a sentence or two, but a proof is not necessary.

- (a). The set of all matrices with real coefficients (all sizes).
- (b). The set of all 2×2 matrices with real entries of the form

$$\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}.$$

- (c). The set of all 2×2 matrices with real entries of the form

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

In chapter 3 we’ll analyze rings in more detail.

1.2.4 Groups

Unlike fields and rings which have two primary binary operations, groups only have one binary operation.

Definition 1.9 (Group). A *group* is a set equipped with a binary operation that is associative, has an identity element, and has inverse elements. If, furthermore, multiplication is

commutative, then the group is called a *commutative group* or an *Abelian group*. Abelian groups can be denoted either additively or multiplicatively, but nonabelian groups are usually denoted multiplicatively. We'll use the term *order of the group* to indicate how many elements a group G has and denote this order by $|G|$.

Example 1.10 (The underlying additive group of a ring). Of course, if you have a field or ring, and just consider addition (and forget about multiplication) you've got an Abelian group. Sometimes this is called the *underlying additive group* of the field or ring. We'll use the same notation for the underlying additive group as we do for the ring. Thus, \mathbf{Z} could mean either the ring of integers or the Abelian group of integers under addition, depending on the context.

Example 1.11 (Finite cyclic groups). The underlying group of the ring \mathbf{Z}_n is called a cyclic group. Its elements are, of course, $0, 1, 2, \dots, n - 1$ where n is congruent to 0. Cyclic groups are also written multiplicatively, and then the elements are $1, a, a^2, \dots, a^{n-1}$ where $a^n = 1$. A common notation for this cyclic group is C_n .

Definition 1.12 (Units in a ring). In order to use the multiplication for a group operation, we'll have to only include the units, also called invertible elements. A *unit* or *invertible element* of a ring R is an element $x \in R$ such that there exists another element $y \in R$ so that $xy = yx = 1$. The subset of units is denoted

$$R^* = \{x \in R \mid \exists y \in R, xy = 1\}.$$

You can easily show that the units form a group under multiplication, called the *multiplicative group of units* of R . When R is a field, then R^* is all of R except 0, but for rings there will be other elements than 0 that aren't invertible. The group R^* will be Abelian when the ring R is commutative, but usually it will be nonabelian when R is not commutative.

Examples 1.13. The units in the ring \mathbf{Z} are just 1 and -1 . The group of units \mathbf{Z}^* is a cyclic group of order 2.

We'll see later that the group of units \mathbf{Z}_p^* when p is prime is a cyclic group of order $p - 1$. It is usually the case that \mathbf{Z}_n^* when n is composite is not a cyclic group.

Example 1.14 (A general linear group, $GL_2(\mathbf{R})$). As a particular example of a multiplicative group of units, take the invertible elements of the matrix ring $M_2(\mathbf{R})$. The invertible 2×2 matrices are those matrices

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

whose determinants $ad - bc$ are nonzero. The group of invertible $n \times n$ matrices, $M_n(R)^*$, is the *general linear group* with coefficients in the ring R , denoted $GL_n(R)$. Note that $GL_n(R)$ is a nonabelian group for $n \geq 2$. The real general linear group $GL_2(\mathbf{R})$ can be interpreted as the group of invertible linear transformations of the plane \mathbf{R}^2 that leave the origin fixed.

We'll study $GL_2(\mathbf{R})$ and $GL_n(\mathbf{R})$ in more detail in section 4.7.2.

Exercise 5. Find two matrices in $GL_2(\mathbf{Z})$ that don't commute thereby proving $GL_2(\mathbf{Z})$ is a nonabelian group.

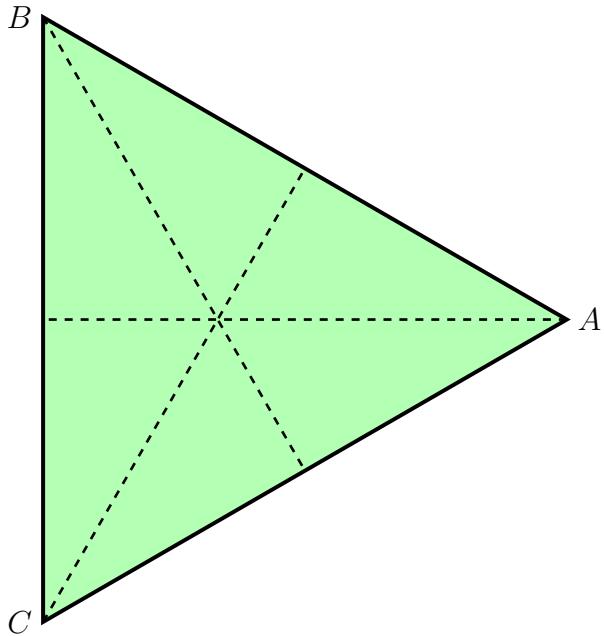


Figure 1.1: Equilateral triangle with lines of symmetry

There are many examples of finite nonabelian groups found in geometry. We'll look at the group of symmetries of an equilateral triangle.

Example 1.15 (The dihedral group D_3). Consider an equilateral triangle. Place a coordinate system on the plane of the triangle so that its center is at $(0, 0)$, one vertex, A , at $(1, 0)$, and the other two, B and C , at $(-\frac{1}{2}, \pm\frac{1}{2}\sqrt{3})$. This triangle has six symmetries. A symmetry is a transformation of the plane that preserves distance (that is, an *isometry*) that maps the triangle back to itself. Three of these symmetries are rotations by 0° , 120° , and 240° .

$$1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \rho = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix} \quad \rho^2 = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix}$$

The identity transformation, 1 , fixes A , B , and C ; the rotation ρ by 120° maps A to B , B to C , and C to A ; and the rotation ρ^2 by 240° maps A to C , B to A , and C to B . There are also three reflections.

$$\varphi = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \rho\varphi = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix} \quad \rho^2\varphi = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix}$$

The reflection φ fixes A , and interchanges B and C ; the reflection $\rho\varphi$ fixes C and interchanges A and B ; and the reflection $\rho^2\varphi$ fixes B and interchanges A and C . This is a particular nonabelian group that has 6 elements. It is a subgroup of $GL_2(\mathbf{R})$ mentioned above.

Example 1.16 (A group of functions). Many applications of group theory are to groups of invertible functions. Such a group includes invertible functions on some set such that the composition of any two of the functions is another one.

Let $f(x) = 1/x$ and $g(x) = 1 - x$. Both of those are invertible considered as rational functions, and, in fact, each is its own inverse: $(f \circ f)(x) = f(1/x) = 1$, and $(g \circ g)(x) =$

$g(1 - x) = 1 - (1 - x) = x$. Let's see what other functions we can derive from f and g by composing them.

First, consider $(f \circ g)(x) = f(g(x)) = f(1 - x) = \frac{1}{1 - x}$; call that composition h so that $h(x) = \frac{1}{1 - x}$. Next, consider $(g \circ f)(x) = g(f(x)) = g\left(\frac{1}{x}\right) = 1 - \frac{1}{x} = \frac{x - 1}{x}$; call that composition k so that $k(x) = \frac{x - 1}{x}$.

We can get more functions if we continue to compose these. Note that $(f \circ k)(x) = f\left(\frac{x - 1}{x}\right) = \frac{x}{x - 1}$; call that ℓ so that $\ell(x) = \frac{x}{x - 1}$. Also, $(g \circ h)(x) = g\left(\frac{1}{1 - x}\right) = 1 - \frac{1}{1 - x} = \frac{x}{x - 1}$. That function has already been called ℓ , so $g \circ h = \ell$.

A couple more computations show that $h \circ h = k$ and $k \circ k = h$.

Since f and g are each their own inverses, $f \circ f = i$ and $g \circ g = i$, where i is the identity function, $i(x) = x$. Also $h \circ k = k \circ h = i$, and $\ell \circ \ell = i$. Also, i composed with any function (on either side) is equal to that same function.

It turns out that these six functions are closed under composition. Table 1.1 gives all of their compositions.

	i	f	g	h	k	ℓ
i	i	f	g	h	k	ℓ
f	f	i	h	g	ℓ	k
g	g	k	i	ℓ	f	h
h	h	ℓ	f	k	i	g
k	k	g	ℓ	i	h	f
ℓ	ℓ	h	k	f	g	i

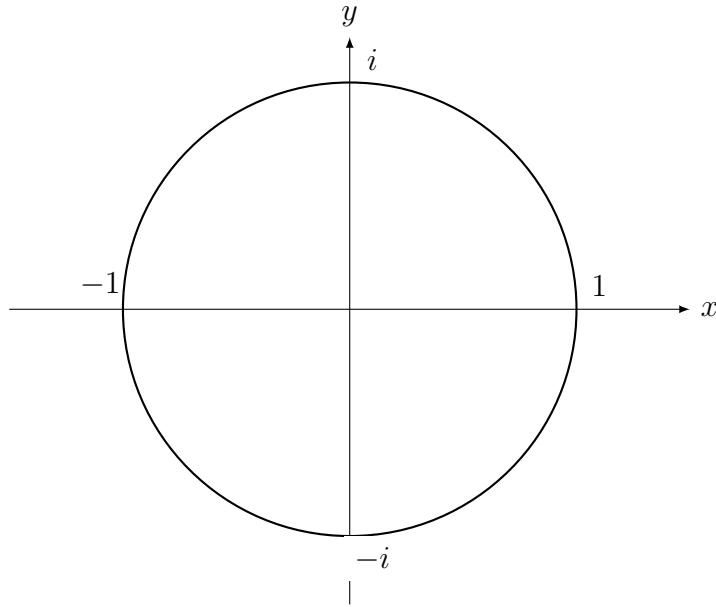
Table 1.1: Composition table for six particular rational functions.

Note that in each row and each column of the table, each one of the functions appears exactly once. That makes the entries of the table a Latin square. A *Latin square* is a square $n \times n$ array filled with n different symbols, each occurring exactly once in each row and exactly once in each column.

Example 1.17 (Euler's circle group). The unit circle, $S^1 = \{x + yi \in \mathbf{C} \mid x^2 + y^2 = 1\}$, is a group under multiplication. This is sometimes called Euler's circle group since Euler (1707–1783) introduced the unit circle in the complex plane for studying angles and trigonometric functions.

The product of two complex numbers on this unit circle is another number on the unit circle. You can directly verify that or you can show it by trigonometry. If $x + yi$ is on the unit circle, then we can identify x with $\cos \theta$ and y with $\sin \theta$ where θ is, as usual, the angle between the positive x -axis and the ray from 0 to $x + yi$. Then the product of two complex numbers on the unit circle corresponds to adding their angles together. The addition formulas for cosines and sines give this correspondence.

Exercise 6. Compute the product of $\cos \theta + i \sin \theta$ times $\cos \varphi + i \sin \varphi$. If $x + iy = (\cos \theta + i \sin \theta)(\cos \varphi + i \sin \varphi)$, then what is x , the real part of the product, in terms of θ and φ ? What is y , the imaginary part?

Figure 1.2: Unit circle S^1

Comment 1.18. Although the sphere

$$S^2 = \{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 = 1\}$$

has no group structure, the 3-sphere in 4-space does. The 3-sphere is

$$S^3 = \{(x, y, z, w) \in \mathbf{R}^4 \mid x^2 + y^2 + z^2 + w^2 = 1\}.$$

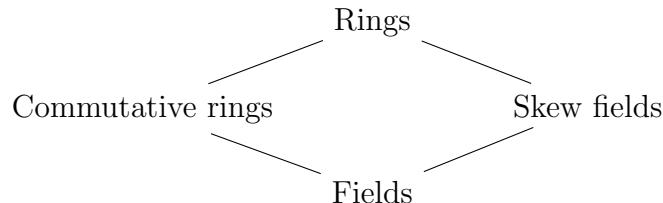
We don't have time or space to discuss that group structure here. (The 2-sphere S^2 , in fact, spheres in all dimensions, does have quandle structures, whatever a quandle might be. See section 4.5.2.)

In chapter 4 we'll study groups in detail.

1.2.5 Other algebraic structures besides fields, rings, and groups

There are an unlimited number of other algebraic structures. Some are similar to those listed above.

For instance, there are division rings (also called skew fields) that have all the properties of fields except multiplication doesn't have to be commutative. The primary example is the quaternions \mathbf{H} . We'll discuss quaternions later in section 2.5.2.



There are a number of structures that are just commutative rings that have nice properties, and we'll look at some of them including integral domains, unique factorization domains, principal ideal domains, and Euclidean domains.

Sometimes rings that don't have a multiplicative identity are studied, but for us, we'll always have 1.

You've already studied vector spaces over the real numbers. Most of the things that you've studied about vector spaces over \mathbf{R} also hold for vector spaces over other fields.

The analogous structure for vector spaces when a field is replaced by a ring is called a *module* over the ring. We won't study modules over a ring, but when we look at ideals in a ring, they are, in fact, examples of modules over the ring. Also, Abelian groups are modules over the ring \mathbf{Z} .

We'll discuss another algebraic structure, quandles, in section 4.5.2 when we discuss groups.

1.3 Isomorphisms, homomorphisms, etc.

Frequently, we look at two algebraic structures A and B of the same kind, for instance, two groups or two rings or two fields, and we'll want to compare them. For instance, we might think they're really the same thing, but they have different names for their elements. That leads to the concept of isomorphism $f : A \cong B$, and we'll talk about that first. Other times we'll know they're not the same thing, but there is a relation between them, and that will lead to the next concept, homomorphism, $f : A \rightarrow B$. We'll then look at some special homomorphisms such as monomorphisms. When we have a homomorphism $f : A \rightarrow A$, we'll call it an endomorphism, and when an isomorphism $f : A \cong A$, we'll call it an automorphism. We'll take each of these variants in turn.

The concepts of injection (one-to-one function), surjection (onto function), and bijection are described section A.2.2 in the appendix on functions.

We'll use the following theorem about finite sets when we consider homomorphisms between finite algebraic structures.

Theorem 1.19. Suppose that $f : A \rightarrow B$ is a function between two finite sets of the same cardinality. Then the following three conditions are equivalent: (1) f is a bijection, (2) f is an injection, and (3) f is a surjection.

Exercise 7. Prove that if $f : A \rightarrow B$ is a function between two finite sets of the same cardinality, then f is injective if and only if f is surjective.

1.3.1 Isomorphisms

We'll say two algebraic structures A and B are isomorphic if they have exactly the same structure, but their elements may be different. For instance, let A be the ring $\mathbf{R}[x]$ of polynomials in the variable x with real coefficients while B is the ring $\mathbf{R}[y]$ of polynomials in y . They're both just polynomials in one variable, it's just that the choice of variable is different in the two rings. We need to make this concept more precise.

Definition 1.20 (Ring isomorphism). Two rings A and B are *isomorphic* if there is a bijection $f : A \rightarrow B$ which preserves addition and multiplication, that is, for all x and y in A ,

$$f(x + y) = f(x) + f(y), \text{ and } f(xy) = f(x)f(y).$$

The correspondence f is called a *ring isomorphism*.

After we introduce homomorphisms, we'll have another way to describe isomorphisms.

You can prove various properties of ring isomorphism from this definition.

Exercise 8. Since the structure of rings is defined in terms of addition and multiplication, if f is a ring isomorphism, it will preserve structure defined in terms of them. Verify that f preserves 0, 1, negation, and subtraction.

Exercise 9. Prove that if f is a ring isomorphism, then so is its inverse function $f^{-1} : B \rightarrow A$.

Exercise 10. Prove that if $f : A \rightarrow B$ and $g : B \rightarrow C$ are both ring isomorphisms, then so is their composition $(g \circ f) : A \rightarrow C$.

Since a field is a special kind of ring, and its structure is defined in terms of addition and multiplication, we don't need a special definition for a field isomorphism. A field isomorphism is just a ring isomorphism between fields.

Exercise 11. Prove that if a ring is isomorphic to a field, then that ring is a field.

We do need a different definition for a group isomorphism since a group is defined in terms of just one binary operation instead of two.

Definition 1.21 (Group isomorphism). Two groups A and B are isomorphic if there is a bijection $f : A \rightarrow B$ which preserves the binary operation. If both are written additively, that means for all x and y in A , $f(x + y) = f(x) + f(y)$; if multiplicative notation is used in both, then $f(xy) = f(x)f(y)$; if additive in A but multiplicative in B , then $f(x + y) = f(x)f(y)$; and if multiplicative in A and additive in B , then $f(xy) = f(x) + f(y)$. The correspondence f is called a *group isomorphism*.

Usually A and B will use the same notation, both additive or both multiplicative, but not always.

Exercise 12. Suppose that both A and B are written multiplicatively and that $f : A \rightarrow B$ is a group isomorphism. Prove that $f(1) = 1$ and $f(x^{-1}) = f(x)^{-1}$ for all $x \in A$.

Example 1.22. Let $A = \mathbf{Z}$ be the group of integers under addition. Let B be the integral powers of 2, so $B = \{\dots, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, \dots\}$ with multiplication as the operation in B . Prove that an isomorphism $f : A \rightarrow B$ is defined by $f(n) = 2^n$. You'll need to show that $f(m+n) = f(m)f(n)$.

There's actually another isomorphism $g : A \rightarrow B$, too, defined by $g(n) = 2^{-n}$.

1.3.2 Homomorphisms

Whereas isomorphisms are bijections that preserve the algebraic structure, homomorphisms are simply functions that preserve the algebraic structure. Since the word homomorphism is so long, alternate words are often used like morphism and map, especially in spoken mathematics.

Definition 1.23 (Ring homomorphism). A *ring homomorphism* $f : A \rightarrow B$ between rings is a function that preserves addition, multiplication, and 1.

A *group homomorphism* $f : A \rightarrow B$ between groups preserves the binary operation (addition or multiplication depending on the notation used for the group).

Comment 1.24. It's a peculiarity of rings that preserving addition and multiplication doesn't imply that 1 is also preserved, so that condition has to be required as well. We'll see plenty of examples of homomorphisms in the course, and there are more examples in the next section on monomorphisms. Of course, isomorphisms are special cases of homomorphisms.

Example 1.25 (A ring homomorphism). Let $\mathbf{Z}[x]$ be the ring of polynomials with integral coefficients. Evaluating a polynomial $f(x)$ at a particular number, like 3, to give $f(3)$, is a ring homomorphism $\varphi : \mathbf{Z}[x] \rightarrow \mathbf{Z}$. It preserves addition since $\varphi(f(x) + g(x)) = f(3) + g(3) = \varphi(f(x)) + \varphi(g(x))$, and you can check that it preserves multiplication and 1.

Example 1.26 (A group homomorphism). Let A be the integers under addition, and let $B = \{1, -1\}$ with multiplication as the binary operation. Then $f : A \rightarrow B$ defined by $f(n) = (-1)^n$ is a group homomorphism.

You can prove several properties of homomorphisms from the definition, but for the time being I'll just mention two because they'll lead to the concept of *category* which will be introduced in section 3.5.

1. The composition of two homomorphisms (of the same kind) is another homomorphism.
2. The identity function $1_A : A \rightarrow A$, which maps every element to itself, is a homomorphism, indeed, it's an isomorphism.

When we have a homomorphism $f : A \rightarrow B$, we'll call A the *domain* of f and we'll call B the *codomain* of f . (Sometimes the word "range" is used for codomain, but some people prefer to use "range" to mean image, which is a different thing. To avoid ambiguity, we'll use "codomain".)

A more natural way to characterize isomorphism is in terms of homomorphisms. Two rings A and B are isomorphic if and only if there are two ring homomorphisms $f : A \rightarrow B$ and $g : B \rightarrow A$ such that $g \circ f$ is the identity on A and $f \circ g$ is the identity on B .

1.3.3 Monomorphisms and epimorphisms

Two common kinds of homomorphisms are monomorphisms and epimorphisms, often called monos and epis for short. When a homomorphism $f : A \rightarrow B$ is an injective function, it's called a *monomorphism*; and when it's a surjective function, it's an *epimorphism* (but, in the category of rings, we'll see there are more epimorphisms than just the surjective ring homomorphisms). You might wonder why we need these words when we've got more than enough words already to describe injective (one-to-one) and surjective (onto) as well as others not mentioned here. The main reason is that they're special kinds of injections or surjections—they preserve the algebraic structure. Another is that, although for group homomorphisms monos and epis have these particular correspondences to injective and surjective, there are other categories in which they don't.

Note that every isomorphism is simultaneously a monomorphism and an epimorphism. The converse holds for groups, but, surprisingly, not for rings.

Example 1.27 (Inclusion). Inclusions are monomorphisms. When one ring (or group) A is a subring (or subgroup) of another B , then the inclusion function $\iota : A \rightarrow B$, which maps an element to itself, is a monomorphism. That's an important example of a monomorphism, but there are others.

Example 1.28. For example, let A and B both be the additive group of integers \mathbf{Z} , and let $f(n) = 2n$. This f is a monomorphism, but it's not an inclusion (which in this case would be the identity map since A and B are the same).

Comment 1.29. Note that if $f : A \rightarrow B$ is a ring homomorphism where A is a field and $0 \neq 1$ in B , then f is always an injection, and so it's a monomorphism. You can prove this statement in two stages. First, show that if $f(x) = 0$ then $x = 0$. Second, show that if $f(x) = f(y)$, then $x = y$.

Thus, every field homomorphism is a monomorphism.

Example 1.30 (A group epimorphism). We'll see plenty of epimorphisms when we talk more about the integers modulo n , but for the time being, consider example 1.26 of a group epimorphism. The group A is the additive group of integers \mathbf{Z} , and the group B is the two element group $\{1, -1\}$ under multiplication. Then $f : A \rightarrow B$ defined by $f(n) = (-1)^n$ is a group epimorphism. Even numbers are sent to 1 and odd numbers to -1 .

1.3.4 Endomorphisms and automorphisms

An endomorphism is just a homomorphism $f : A \rightarrow A$ where the domain and codomain are the same, and an automorphism is just an isomorphism $f : A \rightarrow A$. These are important because we always have the identity automorphism $1_A : A \rightarrow A$ to compare f to, so we have more information when the domain and codomain are the same.

Example 1.31 (A field automorphism). Let \mathbf{C} be the complex field. Let $\phi : \mathbf{C} \rightarrow \mathbf{C}$ be *complex conjugation*, usually denoted by putting a bar above the complex number

$$\varphi(x + yi) = \overline{x + yi} = x - yi.$$

This is clearly a bijection since it is its own inverse, $\overline{\overline{x + yi}} = x + yi$. Also, it preserves addition, multiplication, and 1, so it's a ring isomorphism.

$$\begin{aligned}\overline{(x_1 + y_1i) + (x_2 + y_2i)} &= \overline{x_1 + y_1i} + \overline{x_2 + y_2i} \\ \overline{(x_1 + y_1i)(x_2 + y_2i)} &= \overline{x_1 + y_1i} \overline{x_2 + y_2i} \\ \overline{1} &= 1\end{aligned}$$

In fact, it's a field automorphism of \mathbf{C} .

The existence of this automorphism says that we can't distinguish between i and $-i$ in the sense that any true statement about the complex numbers remains true when all occurrences of i are replaced by $-i$.

Example 1.32 (Group endomorphisms and automorphisms). There are many group endomorphisms $f : \mathbf{Z} \rightarrow \mathbf{Z}$ from the additive group of integers to itself. Fix any integer n and let $f(x) = nx$. This is a group homomorphism since $f(x+y) = n(x+y) = nx+ny = f(x)+f(y)$.

For $n \neq 0$ it is also a monomorphism. For $n = -1$ this is negation, and it's a bijection, so it's a group automorphism. That says if we only consider addition, we can't distinguish between positive and negative numbers.

But negation is not a ring automorphism on the ring of integers because $-(xy)$ does not equal $(-x)(-y)$. Thus, with the use of multiplication, we can distinguish between positive and negative numbers.

1.4 A little number theory

In science nothing capable of proof ought to be accepted without proof. Though this demand seems so reasonable, yet I cannot regard it as having been met even in the most recent methods of laying the foundations for the simplest science; viz., that part of logic which deals with the theory of numbers.

Dedekind, 1888

This course is not meant to be a course in number theory, but we will need a little bit of it. We'll quickly review mathematical induction on the natural numbers \mathbf{N} , divisibility, prime numbers, greatest common divisors, and the Euclidean algorithm.

1.4.1 Mathematical induction on the natural numbers \mathbf{N}

Richard Dedekind (1831–1916) published in 1888 a paper entitled *Was sind und was sollen die Zahlen?* variously translated as *What are numbers and what should they be?* or *The Nature of Meaning of Numbers*. In that work he developed basic set theory and characterized the natural numbers as a simply infinite set.

Definition 1.33. (Dedekind) A set \mathbf{N} is said to be *simply infinite* when there exists a one-to-one function $\mathbf{N} \rightarrow \mathbf{N}$ called the *successor function*, such that there is an element, called the *initial element* and denoted 1, that is not the successor of any element, and if a subset S of \mathbf{N} contains 1 and is closed under the successor function, then $S = \mathbf{N}$.

Such a simply infinite set \mathbf{N} may be called the *natural numbers*. It is characterized by an element 1 and a transformation $\mathbf{N} \rightarrow \mathbf{N}$ satisfying the following conditions:

1. Injectivity: $\forall n, m, n \neq m$ implies $n' \neq m'$.
2. Initial element: $\forall n, 1 \neq n'$.
3. Induction: If $S \subseteq \mathbf{N}$, $1 \in S$, and $(\forall n, n \in S \text{ implies } n' \in S)$, then $S = \mathbf{N}$.

The Dedekind axioms, also called the Peano axioms, are this last characterization involving 1, the successor function, and the three conditions. Among other things, Peano (1858–1932) developed much of the notation in common use in set theory.

The last axiom is called mathematical induction. If you want to show a subset S of \mathbf{N} is all of N , first show that $1 \in S$. Then show for each natural number n that $n \in S$ implies $n + 1$ in S . Finally conclude that $S = \mathbf{N}$.

A principle that is logically equivalent to mathematical induction is the well-ordering principle, also called the minimization principle. It says that each nonempty subset of \mathbf{N} has a least element. To use it to prove a subset S of \mathbf{N} is all of \mathbf{N} , assume that it isn't, take the least element n in $\mathbf{N} - S$, and derive a contradiction, usually by showing there's a smaller element than n not in S .

Another principle logically equivalent to mathematical induction is Euclid's principle of infinite descent which says that there is no infinite decreasing sequence of positive integers. This principle was also used by Fermat (1607–1665).

1.4.2 Divisibility

We'll restrict our discussion now to \mathbf{N} , the natural numbers, that is, the set of positive integers.

Recall that an integer m *divides* an integer n , written $m|n$, if there exists an integer k such that $mk = n$. A few basic properties of divisibility follow directly from this definition. Euclid (fl. ca. 300 B.C.E.) uses some of these in Book VII of his *Elements*. You can find Joyce's translation of Euclid's *Elements* on the web at <http://aleph0.clarku.edu/~djoyce/java/elements/elements.html>

1. 1 divides every number. $1|n$.
2. Each number divides itself. $n|n$.
3. If one number m divides another number n , then m divides any multiple of n , that is, $m|n$ implies $m|kn$.
4. Divisibility is a transitive relation, that is, $m|n$ and $n|k$ imply $m|k$.
5. If one number divides two other numbers, then it divides both their sum and difference. $m|n$ and $m|k$ imply $m|(n + k)$ and $m|(n - k)$.
6. Cancellation law. One number divides another if and only if any multiple of that one number divides the same nonzero multiple of the other number. $m|n \iff kn|kn$. ($k \neq 0$)

Example 1.34. The divisors of a number can be displayed graphically in what is called a Hasse diagram of the lattice of divisors. As an example, consider the number 432. Its prime factorization is 2^43^3 , so its divisors are of the form 2^m3^n where $0 \leq m \leq 4$ and $0 \leq n \leq 3$. There are $5 \cdot 4 = 20$ of these divisors. They are

$$\begin{array}{ccccccccccccc} 1 & 2 & 3 & 4 & 6 & 8 & 9 & 12 & 16 & 18 \\ 24 & 27 & 36 & 48 & 54 & 72 & 108 & 144 & 216 & 432 \end{array}$$

We can display these numbers and emphasize which ones divide which other ones if we put the large numbers at the top of the diagram, and connect the smaller divisors to the larger ones with lines. That results in the Hasse diagram in figure 1.3.

Since divisibility is transitive, we don't have to include all possible connections. So long as there is a path of connections from a lower number to an upper one, then we can conclude the lower divides the upper. The resulting diagram is called a *Hasse diagram* in honor of Hasse (1898–1979).

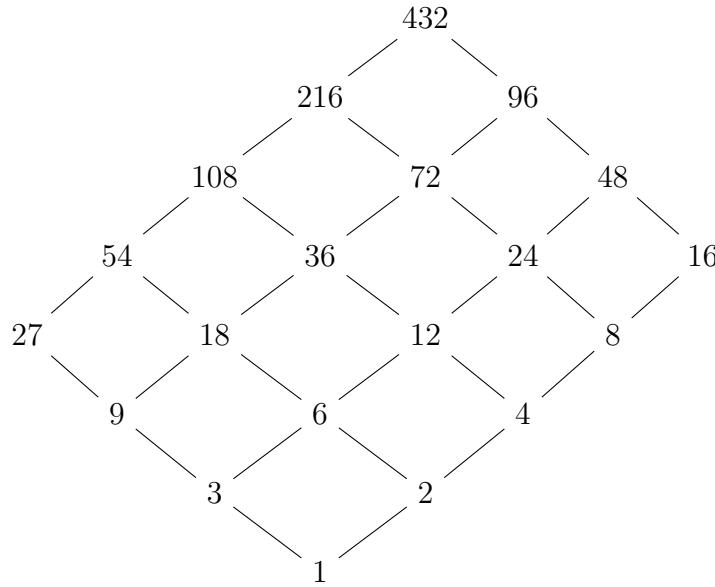


Figure 1.3: Divisors of 432

Exercise 13. Draw Hasse diagrams for the divisors of 30, 32, and 60.

The Hasse diagram for all positive integers under divisibility is, of course, infinite. Figure 1.4 shows the part of it up through 12.

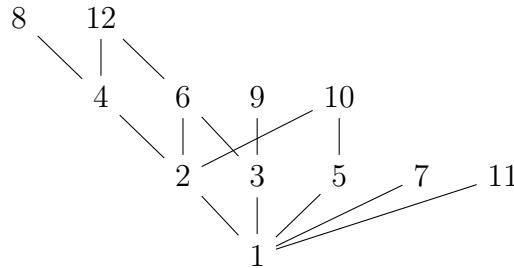


Figure 1.4: Divisibility up through 12

1.4.3 Prime numbers

Definition 1.35. A natural number greater than 1 is said to be a *prime number*, or more simply a *prime*, if its only divisors are 1 and itself, but if it has more divisors, it's called a *composite number*.

Two positive integers are said to be *relatively prime*, or *coprime* if the only positive integer that divides them both is 1.

Prime numbers were mentioned by the Pythagoreans Philolaus (470–385 B.C.E.) and Thymaridas (400–350 B.C.E.), and by Aristotle (384–322 B.C.E.) after them. The first recorded proofs about prime numbers occur in Euclid's *Elements*.

We know intuitively that there are infinitely many primes, and that every number is a product of primes. Now let's prove those statements. We'll start by proving something that will help us prove these two statements. If a theorem is not particularly interesting, but is useful in proving an interesting statement, then it's often called a lemma. This one is found in Euclid's *Elements*.

Lemma 1.36 (Euclid, VII.31). Every number greater than 1 has at least one prime divisor.

Proof. Let n be an integer greater than 1. We'll find a prime divisor of n . Let m be the smallest divisor of n greater than 1. (Note that we're using the minimization principle, also called the well-ordering principle, to conclude that such an m exists.) We'll show that m is prime thereby proving the lemma. We'll do that with a proof by contradiction, and that means that first we'll suppose that m is not prime, then derive a contradiction, and that will imply that m must be prime.

Suppose m is not prime, but composite. Then m is the product of two integers, j and k , each greater than 1. Now, $k|m$ and $m|n$, so $k|n$. But $k < m$. That gives us a divisor of n which is even smaller than m but still greater than 1. That contradicts the fact that m is the smallest divisor of n greater than 1. Thus, m is prime, and it's a divisor of n . Q.E.D.

Now we can prove one of the two statements.

Theorem 1.37. Every number greater than 1 is either a prime or the product of primes.

Proof. This will be another proof by contradiction that uses the well-ordering principle.

Suppose that the theorem is false. Then there is some composite number greater than 1 that is not the product of primes. Let n be the smallest such. By our lemma, this n has some prime divisor, call it p . Then $m = n/p$ is a number smaller than n but larger than 1, so, by the minimality of n , m is either prime or the product of primes. In the first case, when m is prime, then $n = pm$ is the product of two primes. In the second case when m is a product of primes, then $n = pm$ is also a product of primes. In any case, n is the product of primes, a contradiction. Thus, the theorem is true. Q.E.D.

This last theorem will form part of the so-called fundamental theorem of arithmetic that says every number greater than 1 can be uniquely factored as a product of primes. So far, we only have that every number is a product of primes, but we haven't seen the uniqueness. We'll prove that pretty soon.

Next, let's prove the other statement, that there are infinitely many primes. This is Euclid's proof.

Theorem 1.38 (Euclid IX.20). There are infinitely many primes.

Proof. Actually, Euclid proves something a little stronger. Given any finite list of primes, he finds a prime not on that list.

Suppose that p_1, p_2, \dots, p_k is a finite list of primes. Let n be the product of these primes,

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

By our lemma $n + 1$ has a prime factor, call it p . This prime p cannot equal any p_i , for then p would divide both n and $n + 1$, and so would divide the difference 1. But a prime p can't divide 1 since $p > 1$. This p is a prime not on the list.

It follows that there are infinitely many primes. Q.E.D.

The number of relatively prime integers. An important combinatorial count for number theory and algebra is the number $\varphi(n)$ of positive integers less than a given integer n . For example, we'll show later in corollary 2.11 that the number of units in the ring \mathbf{Z}_n is $\varphi(n)$. We'll also use it in our discussion of cyclotomic polynomials in section 1.6.2.

It's easy enough to compute $\varphi(n)$ when n is small. For example, $\varphi(12) = 4$, since there are four positive integers less than 12 which are relatively prime to 12, namely, 1, 5, 7, and 11.

Definition 1.39 (Euler's totient function). For a given positive integer n , the number of positive integers less than n that are relatively prime to n is denoted $\varphi(n)$. The function φ is called Euler's totient function.

The first few values of the totient function are listed in this table.

n	1	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\varphi(n)$	1	1	2	2	4	2	6	4	6	4	10	4	12	6	8	8	16	6

One obvious property of this function is that if p is prime, then $\varphi(p) = p - 1$.

A property that's not so obvious is that if m and n are relatively prime, then $\varphi(mn) = \varphi(m)\varphi(n)$. That property is summarized by saying that φ is a *multiplicative* function. It follows from the Chinese remainder theorem discussed in section 3.2.1.

That reduces the computation of φ to computing it on powers p^k of prime numbers. That can be found directly. The only positive integers less than or equal to p^k that aren't relatively prime to p^k are the multiples of p , which are $p, 2p, \dots, p^k$, and there are p^{k-1} of them. Therefore, $\varphi(p^k) = p^k - p^{k-1} = p^k \left(1 - \frac{1}{p}\right)$.

Theorem 1.40 (Euler's product formula).

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Proof. Write $n = p_1^{k_1} \cdots p_r^{k_r}$ as a product of powers of distinct primes. Then by the multiplicativity of φ ,

$$\varphi(n) = \varphi(p_1^{k_1} \cdots \varphi(p_r^{k_r})) = p_1^{k_1} \left(1 - \frac{1}{p_1}\right) \cdots p_r^{k_r} \left(1 - \frac{1}{p_r}\right) = \varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Q.E.D.

1.4.4 The Euclidean algorithm

The Euclidean algorithm is an algorithm to compute the greatest common divisor of two natural numbers m and n . Euclid described in Book VII of his *Elements*.

Euclid defined the *greatest common divisor* of two natural numbers m and n , often denoted $\text{GCD}(m, n)$ or more simply just (m, n) , as the largest number d which is at the same time a divisor of m and a divisor of n .

Among other things, greatest common divisors are used to reduce common fractions to lowest terms. For example, if you wanted to reduce the fraction $\frac{1417}{1853}$ to lowest terms, you would look for the greatest common divisor of the two numbers 1417 and 1853, which is 109. Then you could divide both the numerator and the denominator by that greatest common divisor to reduce $\frac{1417}{1853}$ to its lowest terms, namely, $\frac{13}{17}$.

There are two forms of the Euclidean algorithm. The first form, as Euclid stated it, repeatedly subtracts the smaller number from the larger replacing the larger by the difference, until the two numbers are reduced to the same number, and that's the greatest common divisor. (Note that the process has to stop by the well-ordering principle since at each step the larger number is reduced.)

The other form speeds up the process. Repeatedly divide the smaller number into the larger replacing the larger by the remainder. (This speeds up the process because if the smaller number is much smaller than the larger, you don't have to subtract it from the larger many times, just divide once and take the remainder which is the same as what you'd get if repeatedly subtracted it.)

Example 1.41. Let's find $\text{GCD}(6731, 5777)$. Since $6731 - 5777 = 954$, replace 6731 by 954. We've reduced the problem to finding $\text{GCD}(5777, 954)$.

Now repeatedly subtract 954 from 5777 until you get a number smaller than 954 and replace 5777 by that number. Alternatively, you could divide 954 into 5777 and replace 5777 by the remainder. You'll get the same thing, namely 53.

Next to find $\text{GCD}(954, 53)$. If you keep subtracting 53 from 954, eventually you'll get 0. Or if you're using division, when you divide 53 into 954, you'll get a remainder of 0. Either way, you can conclude 53 divides 954, so their GCD is 53 itself. Thus, $\text{GCD}(6731, 5777) = 53$

This Euclidean algorithm works to produce the GCD, and the argument only depended on two properties of divisibility mentioned above, namely that if one number divides two other numbers, then it divides both their sum and difference.

Sometimes the GCD of two numbers turns out to be 1, and in that case we say the two numbers are *relatively prime* or that they're *coprime*.

Theorem 1.42 (Euclidean algorithm). Let d be the result of applying the Euclidean algorithm to m and n . Then d is the greatest common divisor $\text{GCD}(m, n)$. Furthermore, the common divisors k of m and n are the divisors of $\text{GCD}(m, n)$.

Proof. One step of the Euclidean algorithm replaces the pair (m, n) by $(m - n, n)$. It was mentioned above in the properties of divisibility that if one number divides two other numbers, then it divides both their sum and difference. Therefore, a number k divides both m and n if and only if k divides $m - n$ and n . Since the pair (m, n) have the same set of divisors as the pair $(m - n, n)$, therefore $\text{GCD}(m, n) = \text{GCD}(m - n, n)$. Thus, at each step of the Euclidean algorithm the GCD remains invariant. Eventually, the two numbers are the same, but when that last step is reached, that number is the GCD. So, the end result of the Euclidean algorithm is $d = \text{GCD}(m, n)$.

The remarks above show that every divisor k of m and n also divides the result d of applying the Euclidean algorithm to m and n . Finally, if $k|d$, since $d|m$ and $d|n$, therefore $k|m$ and $k|n$. Q.E.D.

Extended Euclidean algorithm. There's still more that we can get out of the algorithm if we include the equations implicit in the computations. That will lead to the extended Euclidean algorithm.

Example 1.43. When we found $\text{GCD}(6731, 5777)$, if we kept track of the quotients as well as the remainders, then each step yields an equation.

$$\begin{aligned} 6731 - 1 \cdot 5777 &= 954 \\ 5777 - 6 \cdot 954 &= 53 \\ 954 - 18 \cdot 53 &= 0 \end{aligned}$$

Turning these equations around, we can find 53 as a linear combination of 6731 and 5777 as follows, starting with the next to the last equation.

$$\begin{aligned} 53 &= 5777 - 6 \cdot 954 \\ &= 5777 - 6 \cdot (6731 - 1 \cdot 5777) = 7 \cdot 5777 - 6 \cdot 6731 \end{aligned}$$

Thus, the GCD of 6731 and 5777 is a linear combination of them.

Here's the general situation to find $\text{GCD}(m, n)$ as a linear combination of m and n . Let's suppose that $m > n$ to begin with. We divide n into m and get a quotient of q_1 and remainder of r_1 , that is

$$m = q_1 n + r_1,$$

with r_1 between 1 and n . Then we work with n and r_1 instead of m and n . Divide r_1 into n to get a quotient of q_2 and a remainder of r_2 , that is,

$$n = q_2 r_1 + r_2.$$

And we keep going until eventually we get a remainder of 0.

$$\begin{aligned} r_1 &= q_3 r_2 + r_3 \\ r_2 &= q_4 r_3 + r_4 \\ &\vdots \\ r_{s-3} &= q_{s-1} r_{s-2} + r_{s-1} \\ r_{s-2} &= q_s r_{s-1} + 0 \end{aligned}$$

We have

$$m > n > r_1 > r_2 > \cdots > r_{s-1}$$

and r_{s-1} is d , the GCD we're looking for.

Each equation finds a remainder as a linear combination of the previous two remainders. Starting with the next to the last equation, we can find $d = r_{s-1}$ as a linear combination of r_{s-2} and r_{s-3} . The equation before that gives r_{s-2} in terms of r_{s-3} and r_{s-4} , so we can also get d in terms of r_{s-3} and r_{s-4} . Working our way back up, we can eventually get d as a linear combination of m and n .

Thus, we've shown the following theorem.

Theorem 1.44 (Extended Euclidean algorithm). The greatest common divisor $d = \text{GCD}(m, n)$ of m and n is a linear combination of m and n . That is, there exist integers a and b such that

$$d = am + bn.$$

Now that we have the major theorems on GCDs, there are a few more fairly elementary proprieties of GCDs that are straightforward to prove, such as these.

Theorem 1.45.

$$\text{GCD}(a, b + ka) = \text{GCD}(a, b).$$

$$\text{GCD}(ak, bk) = k\text{GCD}(a, b).$$

$$\text{If } d = \text{GCD}(a, b) \text{ then } \text{GCD}(a/d, b/d) = 1.$$

Exercise 14. Prove the statements in the theorem.

Greatest common divisors of more than two numbers The GCD of more than two numbers is defined the same way as for two numbers: the GCD of a set of numbers the largest number that divides them all. For example, $\text{GCD}(14, 49, 91) = 7$. To find a GCD of three numbers, a , b , and c , first find $d = \text{GCD}(a, b)$, then find $e = \text{GCD}(d, c)$. Thus,

$$\text{GCD}(a, b, c) = \text{GCD}(\text{GCD}(a, b), c),$$

a statement that is easy to show.

Pairwise relatively prime numbers A set of numbers is said to be *pairwise relatively prime* or *pairwise coprime* if any two of them are relatively prime. For instance, 15, 22, and 49 are three pairwise relatively prime numbers. Thus, a , b , and c are pairwise relatively prime when

$$\text{GCD}(a, b) = \text{GCD}(a, c) = \text{GCD}(b, c) = 1.$$

Note that $\text{GCD}(a, b, c)$ can be 1 without a , b , and c being pairwise relatively prime. For instance, $\text{GCD}(6, 10, 15) = 1$, but $\text{GCD}(6, 10) = 2$, $\text{GCD}(6, 15) = 3$, and $\text{GCD}(10, 15) = 5$.

Least common multiples The *least common multiple* of a set of positive integers is the smallest positive integer that they all divide. It is easy to show that the greatest common divisor of two integers times their least common multiple equals their product.

$$\text{GCD}(a, b) \text{ LCM}(a, b) = ab.$$

Least common multiples can be used to sum common fractions. For example, to add $\frac{5}{6} + \frac{4}{15}$, note that the least common multiple of 6 and 15 is 30, so each fraction can be expressed with the least common denominator 30 as $\frac{25}{30} + \frac{8}{30} = \frac{25+8}{30} = \frac{33}{30}$. Even using least common denominators, it may be that the sum can be simplified as it can in this case to $\frac{11}{10}$.

1.5 The fundamental theorem of arithmetic

We proved above that every natural number could be factored as a product of primes. But we want more than existence, we want uniqueness. We need to prove that there is only one way that it can be factored as a product of primes.

The unique factorization theorem, a.k.a., the fundamental theorem of arithmetic. Now, in order to make this general statement valid we have to extend a little bit what we mean by a product. For example, how do you write a prime number like 7 as a product of primes? It has to be written as the product 7 of only one prime. So we will have to accept a single number as being a product of one factor.

Even worse, what about 1? There are no primes that divide 1. One solution is to accept a product of no factors as being equal to 1. It's actually a reasonable solution to define the empty product to be 1, but until we find another need for an empty product, let's wait on that and restrict this unique factorization theorem to numbers greater than 1. So, here's the statement of the theorem we want to prove.

Theorem 1.46 (Unique factorization theorem). Each integer n greater than 1 can be uniquely factored as a product of primes. That is, if n equals the product $p_1 p_2 \cdots p_r$ of r primes, and it also equals the product $q_1 q_2 \cdots q_s$ of s primes, then the number of factors in the two products is the same, that is $r = s$, and the two lists of primes p_1, p_2, \dots, p_r and q_1, q_2, \dots, q_s are the same apart from the order the listings.

We'll prove this by using the strong form of mathematical induction. The form that we'll use is this:

In order to prove a statement $S(n)$ is true for all numbers, prove that $S(n)$ follows from the assumption that $S(k)$ is true for all $k < n$.

This principle of induction appears to be stronger than the one we've used before, but, in fact, it is equivalent to it. It's really the same as the minimization principle (i.e. well-ordering principle) applied to the negation of the statement. The advantage in using it is that a proof by contradiction is not needed making the proof more understandable.

Proof. We'll prove the unique factorization theorem in two cases. Case 1 will be where n is a prime number itself. Case 2 will be where n is composite.

Case 1: Suppose that n is a prime number. The only way that a prime number can be written as a product of primes is as itself; otherwise it would not be prime, but composite.

Case 2: Suppose that n is a composite number equal to both products of primes $p_1 p_2 \cdots p_r$ and $q_1 q_2 \cdots q_s$. Note that since n is composite, both r and s are at least 2; otherwise it would not be composite, but prime.

Now look at one of the primes, say p_1 . It divides n , so it divides the product of the other primes $q_1 q_2 \cdots q_s$. We suspect that that implies it has to be one of those other primes. Let's put that off for a bit; that is, logically before we prove this theorem, we need to prove another theorem, listed next, that if a prime divides a product of primes, then it is one of those primes; but we'll actually do that next. Assuming we've done that, then we can conclude that p_1 is one of the q_i 's. We can reorder the product $q_1 q_2 \cdots q_s$ to make it so that q_1 equals p_1 . Now, since $p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$ and the first first factors of the two products are equal, therefore $p_2 \cdots p_r = q_2 \cdots q_s$. Now, by our new induction principle, these are two prime factorizations of a number smaller than n , and hence are the same, except for their order. Therefore, they have the same number of factors, that is, $r = s$, and all the factors are the same except for their order. And the number n is that product times p_1 , which equals q_1 , therefore the original two products, $p_1 p_2 \cdots p_r$ and $q_1 q_2 \cdots q_s$, are the same except for order. Q.E.D.

Well, that finished the proof except we have to prove another theorem first, namely, the following one.

Theorem 1.47. If a prime divides a product of primes $q_1 q_2 \dots q_s$, then it equals one of the primes q_1, q_2, \dots, q_s .

We could do that, but we'll prove a slightly stronger theorem, namely, the following one.

Theorem 1.48. If a prime divides a product of numbers $b_1 b_2 \dots b_s$, then it divides one of the numbers b_1, b_2, \dots, b_s .

Now the reason this theorem implies the previous theorem is because if a prime p divides a product of primes $q_1 q_2 \dots q_s$, then it divides one of the primes q_1, q_2, \dots, q_s , but the only way that one prime can divide another is if it equals the other.

Proof. A product of s numbers $b_1 b_2 \dots b_s$ is actually a series of binary products. It's b_1 times $b_2 \dots b_s$, and $b_2 \dots b_s$ is b_2 times $b_3 \dots b_s$, etc, where the last product is $b_{s-1} b_s$ is the product of b_{s-1} times b_s . That means that if we knew the following theorem, then, using ordinary induction, we could conclude this one. Q.E.D.

Theorem 1.49. If a prime divides a product of two numbers, then it divides one of the numbers.

Now, we could prove this theorem directly, but it turns out that there is a slightly stronger version that we can use in other places, so let's prove it, the one listed next, instead, and show this theorem follows from it.

Theorem 1.50. If n and a are relatively prime, and $n|ab$, then $n|b$.

Proof that this theorem implies the previous one. Suppose that a prime p divides ab . If p doesn't divide a , then it's relatively prime to a , so by this theorem, it divides b . Therefore, either $p|a$ or $p|b$. Q.E.D.

Proof of this theorem. Suppose that $\text{GCD}(n, a) = 1$. Then, by the extended Euclidean algorithm, 1 is a linear combination of n and a , that is, there exist integers t and u such that

$$1 = tn + ua.$$

Multiply that equation by b to get

$$b = tnb + uab.$$

Now, if $n|ab$, then n divides the right hand side of the equation, but that equals the left hand side, so $n|b$. Q.E.D.

Comment 1.51. Typically in a mathematics book those theorems that come first logically are presented first. Here we started with our goal and discovered the theorems that were needed to prove the goal. (Actually, I made the list longer than it needed to be by strengthening a couple of them because the stronger versions are more useful, something you can only tell with hindsight.)

The advantage to presenting theorems in their logical order is that it is easier to follow the logic. The disadvantage is that the motivation for the preliminary theorems is not apparent until the final theorem, the interesting one, is reached.

Usually when we write the prime factorization of a number, we'll use exponents on those primes that are repeated. For instance, the number 40 had the prime factorization $2 \cdot 2 \cdot 2 \cdot 5$. An abbreviated form for this factorization is $2^3 \cdot 5$. We say that the prime 2 occurs with multiplicity 3, while the prime 5 occurs with multiplicity 1. The multiplicities are the exponents. So, in general, a number n has the prime factorization

$$n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$$

where the primes p_1, p_2, \dots, p_k are all distinct, and their multiplicities are the exponents e_1, e_2, \dots, e_k , respectively.

These exponents are called the orders of the primes in n . The *order* of p in n be the exponent of p in the prime factorization of n , denoted $\text{ord}_p n$.

Immediate corollaries to the unique factorization theorem. A corollary is a theorem that logically follows very simply from a theorem. Sometimes it follows from part of the proof of a theorem rather than from the statement of the theorem. In any case, it should be easy to see why it's true. We can draw a couple of corollaries from the unique factorization theorem.

Corollary 1.52. The only primes that can divide a number n are the ones that appear in its prime factorization $p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$.

Corollary 1.53. If the prime factorizations of m and n are $m = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ and $n = p_1^{f_1} p_2^{f_2} \cdots p_k^{f_k}$ (where here some of the e_i 's and f_i 's may equal 0 so we can use the same list of primes for both numbers), then their greatest common divisor $d = \text{GCD}(m, n)$ has the prime factorization $d = p_1^{g_1} p_2^{g_2} \cdots p_k^{g_k}$ where each exponent g_i is the minimum of the corresponding exponents e_i and f_i .

As an example of the last corollary, if $m = 1260 = 2^2 3^2 5^1 7^1$ and $n = 600 = 2^3 3^1 5^2$, then their GCD is $d = 2^2 3^1 5^1 = 60$.

1.6 Polynomials.

We'll frequently use polynomials in our study of fields and rings. We'll only consider polynomials with coefficients in fields and commutative rings, not with coefficients in noncommutative rings.

We won't formally define polynomials. For now, we'll only look at polynomials in one variable x , but later in section 3.10.4 we'll look at polynomials in two or more variables.

Informally a *polynomial* $f(x)$ with coefficients in a commutative ring R is an expression

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where each coefficient $a_i \in R$. We'll assume that the leading coefficient a_n is not zero so that $\deg f$, the degree of the polynomial, is n . When a_n is zero, the polynomial is called a *monic* polynomial.

It's convenient to denote a polynomial either by f or by $f(x)$. If the variable x is referred to somewhere nearby, then I'll use $f(x)$, otherwise I'll just use f . For instance, if I want to multiply two polynomials f and g together, I'll write fg , but if I want two multiply f by $x^2 - 3x + 2$, I'll write $f(x)(x^2 - 3x + 2)$ or $f(x) \cdot (x^2 - 3x + 2)$.

A *root* of a polynomial is an element a of R such that $f(a) = 0$, that is, it's a solution of the polynomial equation $f(x) = 0$.

The set of all polynomials with coefficients in a commutative ring R is denoted $R[x]$. It has addition, subtraction, and multiplication, and satisfies the requirements of a ring, that is, it has addition, subtraction, and multiplication with the usual properties. $R[x]$ is called the *ring of polynomials with coefficients in R* . Note that $R[x]$ doesn't have reciprocals even when R is a field, since x has no inverse in $R[x]$. Therefore, $R[x]$ is not a field. Nonetheless, the ring R is a subring of the ring $R[x]$ since we can identify the constant polynomials as the elements of R .

1.6.1 Division for polynomials

Although $R[x]$ doesn't have reciprocals, it does have a division algorithm, at least when the divisor is a monic polynomial.

Theorem 1.54 (The division algorithm for polynomials over a ring). Let R be a commutative ring and $R[x]$ its polynomial ring in one variable. Let f be a polynomial (the dividend) and g a monic polynomial (the divisor). Then there exist unique polynomials q (the quotient) and r (the remainder) such that $f = qg + r$ where either $r = 0$ or $\deg r < \deg g$.

Proof of existence. One case is when $f = 0$ or $\deg f < \deg g$. Since the dividend already has a lower degree, the quotient $q = 0$ and the remainder $r = f$.

That leaves the case when $\deg f \geq \deg g$. We'll prove it by induction on $n = \deg f$ where the base case is $n = 0$. That's the case where f and g are both constants in the ring R , but g is monic, so $g = 1$. Then $q = f$ and $r = 0$.

Now for the inductive step. We'll assume the inductive hypothesis that the theorem is correct for all polynomials f of degree less than n and show it's true for those of degree n . Let

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n \text{ and } g(x) = b_0 + b_1 x + \cdots + b_{m-1} x^{m-1} + x^m \text{ where } n \geq m.$$

The polynomial $f_1(x) = f(x) - a_n x^{n-m} g(x)$ has a 0 coefficient for x^n , so its degree is less than n . By inductive hypothesis, there are polynomials q_1 and r_1 such that $f_1 = q_1 g + r_1$ where $r_1 = 0$ or $\deg r_1 < \deg g$. Equating the right sides of the two equations involving f_1 , we may conclude that

$$f(x) = (a_1(x) + a_n x^{n-m})g(x) + f_1(x).$$

That gives us the desired representation $f(x) = q(x)g(x) + r(x)$, finishing the inductive proof for the existence half of the proof. Q.E.D.

Proof of uniqueness. Suppose there are also polynomials q' and r' such that $f = q'g + r'$ where either $r' = 0$ or $\deg r' < \deg g$. We'll show $r = r'$ and $q = q'$.

Since $f = qg + r$ and $f = q'g + r'$, therefore $qg + r = q'g + r'$ so $r - r' = g(q' - q)$. Suppose that $r \neq r'$. Then $q' - q \neq 0$, and since g is a monic polynomial, therefore $\deg g(q' - q) \geq \deg g$. Therefore $\deg(r - r') \geq \deg g$. But $\deg(r - r') < \deg g$ since both r and r' have degree less than $\deg g$, a contradiction. Therefore, $r = r'$.

Now we have $0 = g(q' - q)$, but g is monic, so $q' - q = 0$, and $q = q'$.

Q.E.D.

If R happens to be a field, there is a stronger version of the theorem that doesn't require g to be a monic polynomial.

Theorem 1.55 (The division algorithm for polynomials over a field). Let F be a field and $F[x]$ its polynomial ring in one variable. Let f be a polynomial (the dividend) and g a nonzero polynomial (the divisor). Then there exist unique polynomials q (the quotient) and r (the remainder) such that $f = qg + r$ where either $r = 0$ or $\deg r < \deg g$.

Exercise 15. Prove the above theorem. Hint: divide g by its leading coefficient and use the division algorithm for polynomials over a ring. There will still be two parts, one for existence and one for uniqueness.

The remainder theorem and factor theorem. The remainder theorem is something that's frequently covered in high school algebra classes. It says when you divide a polynomial f by $x - a$, the remainder is $f(a)$. It works in general for polynomials with coefficients in an arbitrary ring.

Theorem 1.56 (Remainder theorem). Let R be a commutative ring and $R[x]$ its polynomial ring. For $f \in R[x]$ and $a \in R$, there is a polynomial q such that $f(x) = (x - a)q(x) + f(a)$.

Proof. Apply the division algorithm for $g(x) = x - a$. Then $f(x) = (x - a)q(x) + r$ where r is a constant. Setting x to a , we conclude $f(a) = r$.
Q.E.D.

The factor theorem is a corollary of the remainder theorem. —

Theorem 1.57 (Factor theorem). For $f \in R[x]$ and $a \in R$, a is a root of f if and only if $(x - a)$ divides $f(x)$.

Further properties of polynomials. There are a couple more properties of polynomials that apply only when the ring is a field or an integral domain. As described later in section 3.1.3, an integral domain is a commutative ring in which $0 \neq 1$ that satisfies one of the two equivalent conditions: it has no zero-divisors, or it satisfies the cancellation law. Thus, fields are special cases of integral domains.

One property is that a polynomial of degree n has at most n roots.

Theorem 1.58. The number of roots of a nonzero polynomial with coefficients in an integral domain is at most the degree of the polynomial.

Proof. We'll prove this by induction on n , the degree of the polynomial f .

If $n = 0$, then f is a constant, but it's not the zero constant, so it has no roots.

Assume the inductive hypothesis, namely, the theorem holds for all functions of degree n . We'll show it holds for each function f of degree $n + 1$. If f has no roots, then the theorem is true, so let r be a root of f . By the factor theorem, $f(x) = (x - r)q(x)$, where the degree of the quotient q equals n .

We'll show every other root $r' \neq r$ of f is also a root of q . Since r' is a root, therefore $0 = f(r') = (r' - r)q(r)$. Now $r' - r$ is not 0, and the ring is an integral domain which has no zero-divisors, therefore $0 = q(r)$. Thus all other roots of f are roots of q .

Since $\deg g = n$, by the inductive hypothesis, g has at most n roots, therefore f has at most $n + 1$ roots.

That completes the proof by induction.

Q.E.D.

Exercise 16. An example of a ring that is not an integral domain is \mathbf{Z}_8 . Show that the quadratic polynomial $f(x) = x^2 - 1$ in $\mathbf{Z}_8[x]$ has more than two roots in \mathbf{Z}_8 .

A couple of corollaries for polynomials with coefficients in an integral domain follow from the previous theorems.

Corollary 1.59. If $\deg f = n$, and a_1, a_2, \dots, a_n are n distinct roots of f , then

$$f(x) = a(x - a_1)(x - a_2) \cdots (x - a_n)$$

where a is the leading coefficient of f .

Corollary 1.60. If two monic polynomials f and g both of degree n have the same value at n places, then they are equal.

1.6.2 Roots of unity and cyclotomic polynomials

Definition 1.61 (Root of unity). A *root of unity*, also called a root of 1 is a complex number such that when raised to some positive integer power yields 1. If $z^n = 1$, then z is called an n^{th} root of unity. If n is the smallest positive integer power such that $z^n = 1$, then n is called a n^{th} primitive root of unity.

Among the real numbers, the only roots of unity are 1 and -1 . 1 is the only first primitive root of unity and -1 is the only primitive second root of unity.

The n^{th} roots of unity are equally spaced around the unit circle separated by angles of $2\pi/n$. See figure 3.4 for the primitive seventh roots of unity on the unit circle.

An n root of unity z is a root of the polynomial $z^n - 1$, but not all roots of such a polynomial are primitive. For example, roots of the polynomial $z^2 - 1$ are second roots of unity, but 1, being one of those two roots, is not a primitive second root of unity.

Example 1.62 (Sixth roots of unity). The sixth roots of unity are roots of the polynomial $z^6 - 1$. This polynomial factors as $(z^3 + 1)(z^3 - 1) = (z^2 - z + 1)(z + 1)(z^2 + z + 1)(z - 1)$. Of course, two of the roots of this polynomial are 1 and -1 which account for the factors $x - 1$ and $x + 1$. The roots of the factor $z^2 - z - 1$ are also roots of $z^3 - 1$, so are cube roots of unity, in fact, they're the two primitive third roots of unity. Those roots are $z = \frac{1}{2}(-1 + i\sqrt{3})$. If you call one of them $\omega = \frac{1}{2}(-1 + i\sqrt{3})$, then the other one is $\omega^2 = \frac{1}{2}(-1 - i\sqrt{3})$. You can see

them displayed in the complex plane in figure 3.3 which illustrates the lattice of Eisenstein integers.

The roots of the other factor $z^2 + z + 1$ are $z = \frac{1}{2}(1 + i\sqrt{3})$. They are the two primitive sixth roots of unity. Notice that they are $\omega + 1$ and $\omega^2 + 1$.

So, altogether, there are six sixth roots of unity. Two are primitive sixth roots, two are primitive third roots, one is a primitive second root, and one is a primitive first root.

Among the five fifth roots of unity, one of them, $z = 1$, is not primitive, the other four are. They are roots of the polynomial $\Phi_5(z) = \frac{z^5 - 1}{z - 1} = z^4 + z^3 + z^2 + z + 1$.

If z is a primitive n^{th} root of unity, then the entire list of n^{th} roots is $1, z, z^2, \dots, z^{n-1}$. The root z^k won't be primitive if there is a common divisor of n and k . That leaves only $\varphi(n)$ of the roots to be primitive, where $\varphi(n)$ is the number of positive integers less than n that are relatively prime to n . See definition 1.39 for a definition of Euler's totient function φ .

Definition 1.63 (Cyclotomic polynomial). The polynomial $\Phi_n(z) = \prod_{k=1}^{\varphi(n)} (z - z_k)$, where $z_1, z_2, \dots, z_{\varphi(n)}$ are the primitive n^{th} roots of unity, is called the n^{th} cyclotomic polynomial.

There are two primitive third roots of unity as mentioned in the example above, so $\Phi_3(z) = z^2 - z - 1$. There are also two primitive sixth roots, and $\Phi_6(z) = z^2 + z - 1$.

When p is a prime number, then $\Phi(p)$ has degree $\varphi(p) = p - 1$. Its value is $\Phi(p) = \frac{z^p - 1}{z - 1} = z^{p-1} + \dots + z + 1$.

Here's a short table of the first few cyclotomic polynomials.

n	$\Phi(n)$	n	$\Phi(n)$
1	$z - 1$	9	$z^6 + z^3 + 1$
2	$z + 1$	10	$z^4 - z^3 + z^2 - z + 1$
3	$z^2 + z + 1$	11	$z^{10} + z^9 + \dots + z + 1$
4	$z^2 + 1$	12	$z^4 - z^2 + 1$
5	$z^4 + z^3 + z^2 + z + 1$	13	$z^{12} + z^9 + \dots + z + 1$
6	$z^2 - z + 1$	14	$z^6 - z^5 + z^4 - z^3 + z^2 - z + 1$
7	$z^6 + z^5 + z^4 + z^3 + z^2 + z + 1$	15	$z^8 - z^7 + z^5 - z^4 + z^3 - z + 1$
8	$z^4 + 1$	16	$z^8 + 1$

It's interesting that the only coefficients that appear in the first one hundred cyclotomic polynomials are 0, 1, and -1 .

We'll use cyclotomic polynomials in section 3.10.3.

Chapter 2

Fields

Informally, a field is a set equipped with four operations—addition, subtraction, multiplication, and division that have the usual properties.

We'll study rings in chapter 3, which are like fields but need not have division.

2.1 Introduction to fields

A *field* is a set equipped with two binary operations, one called *addition* and the other called *multiplication*, denoted in the usual manner, which are both commutative and associative, both have identity elements (the additive identity denoted 0 and the multiplicative identity denoted 1), addition has inverse elements (the inverse of x being denoted $-x$), multiplication has inverses of nonzero elements (the inverse of x being denoted $\frac{1}{x}$), multiplication distributes over addition, and $0 \neq 1$.

Three fields that you already know are the field of real numbers \mathbf{R} , the field of rational numbers \mathbf{Q} , and the field of complex numbers \mathbf{C} .

We'll see that there are many other fields. When we have a generic field, will use a capital F to denote it.

2.1.1 Definition of fields

Here's a more complete definition.

Definition 2.1 (field). A *field* F consists of

1. a set, also denoted F and called the *underlying set* of the field;
2. a binary operation $+ : F \times F \rightarrow F$ called *addition*, which maps an ordered pair $(x, y) \in F \times F$ to its *sum* denoted $x + y$;
3. another binary operation $\cdot : F \times F \rightarrow F$ called *multiplication*, which maps an ordered pair $(x, y) \in F \times F$ to its *product* denoted $x \cdot y$, or more simply just xy ;
such that
4. addition is commutative, that is, for all elements x and y , $x + y = y + x$;

5. multiplication is commutative, that is, for all elements x and y , $xy = yx$;
6. addition is associative, that is, for all elements x , y , and z , $(x + y) + z = x + (y + z)$;
7. multiplication is associative, that is, for all elements x , y , and z , $(xy)z = x(yz)$;
8. there is an additive identity, an element of F denoted 0 , such that for all elements x , $0 + x = x$;
9. there is a multiplicative identity, an element of F denoted 1 , such that for all elements x , $1x = x$;
10. there are additive inverses, that is, for each element x , there exists an element y such that $x + y = 0$; such a y is called the *negation* of x ;
11. there are multiplicative inverses of nonzero elements, that is, for each nonzero element x , there exists an element y such that $xy = 1$; such a y is called a *reciprocal* of x ;
12. multiplication distributes over addition, that is, for all elements x , y , and z , $x(y + z) = xy + xz$; and
13. $0 \neq 1$.

The conditions for a field are often call the *field axioms*.

Caveat: We're using the terminology and notation of arithmetic that we use for numbers, but the elements of our fields need not be numbers; often they will be, but sometimes they won't.

Note that we'll use the standard notational conventions on precedence for all fields so we don't have to fully parenthesize every expression. Multiplication and division have a higher precedence than addition and subtraction, so that, for example, $x - y/z$ means $x - (y/z)$, not $(x - y)/z$. Also, operations are executed from left to right, so that $x - y - z$ means $(x - y) - z$, not $x - (y - z)$. (Usually operations are executed from left to right, but an exception is that exponentiation is executed from right to left, so that x^{m^n} means $x^{(m^n)}$, not $(x^m)^n$.)

Commutativity and associativity of addition imply that terms can be added in any order, so of course we won't put parentheses when we're adding more than two terms together. Likewise for multiplication.

Although in parts 10 and 11 of the definition only the existence of an additive and multiplicative inverses is required, you can easily show uniqueness follows from the definition. Once that is done we can note that the additive inverse of x is called the *negation* of x and denoted $-x$, and the multiplicative inverse of x , when x is not 0 , is called the *reciprocal* of x and denoted $1/x$, $\frac{1}{x}$, or x^{-1} .

2.1.2 Subtraction, division, multiples, and powers

With the help of negation, we can define subtraction as follows. The *difference* of two elements x and y is defined as $x - y = x + (-y)$.

Likewise, with the help of reciprocation, we can define division. The *quotient* of an element x and a nonzero element y is xy^{-1} , denoted x/y or $\frac{x}{y}$. The expected properties of subtraction

and division all follow from the definition of fields. For instance, multiplication distributes over subtraction, and division by z distributes over addition and subtraction.

Likewise, we can define integral multiples of elements in a field. First, we'll define nonnegative multiples inductively. For the base case, define $0x$ as 0. Then define $(n+1)x$ as $x + nx$ when n is a nonnegative integer. Thus nx is the sum of n x 's. For instance, $3x = x + x + x$. Then if $-n$ is a negative integer, we can define $-nx$ as $-(nx)$. The usual properties of multiples, like $(m+n)x = mx + nx$ will, of course, hold.

Furthermore, we can define integral powers of x . Define x^1 as x for a base case, and inductively for nonnegative n , define x^{n+1} as xx^n . Thus nx is the product of n x 's. For instance, $x^3 = xxx$. Next, define x^0 as 1, so long as $x \neq 0$. (0^0 should remain undefined, but for some purposes, especially in algebra, it's useful to define 0^0 to be 1.) Finally, if $-n$ is positive and $x \neq 0$, define x^{-n} as $(x^n)^{-1}$. The usual properties of integral powers hold, like $x^{m+n} = x^m x^n$ and $(xy)^n = x^n y^n$.

2.1.3 Properties that follow from the axioms

There are numerous useful properties that are logical consequences of the axioms. Generally speaking, the list of axioms should be short, if not minimal, and any properties that can be proved should be proved. Here's a list of several things that can be proved from the axioms. We'll prove a few in class, you'll prove some as homework, and we'll leave the rest. (They make good questions for quizzes and tests.)

In the following statements, unquantified statements are meant to be universal with the exception that whenever a variable appears in a denominator, that variable is not to be 0.

Exercise 17. Prove that 0 is unique. That is, there is only one element x of a field that has the property that for all y , $x + y = y$. (The proof that 1 is unique is similar.)

Exercise 18. Prove that each number has only one negation. That is, for each x there is only one y such that $x + y = 0$. (The proof that reciprocals of nonzero elements are unique is similar.)

Exercise 19. Prove that the inverses of the identity elements are themselves, that is, $-0 = 0$, and $1^{-1} = 1$.

Exercise 20. Prove that multiplication distributes over subtraction: $x(y - z) = xy - xz$.

Exercise 21. Prove that 0 times any element in a field is 0: $0x = 0$.

Exercise 22. Prove the following properties concerning multiplication by negatives: $(-1)x = -x$, $-(-x) = x$, $(-x)y = -(xy) = x(-y)$, and $(-x)(-y) = xy$.

Exercise 23. Prove the following properties concerning reciprocals: $(x^{-1})^{-1} = x$, and $(xy)^{-1} = x^{-1}y^{-1}$.

Exercise 24. Prove that when y and z are both nonzero that $\frac{x}{y} = \frac{w}{z}$ if and only if $xz = yw$.

Exercise 25. Prove the following properties concerning division:

$$\begin{aligned}\frac{x}{y} \pm \frac{w}{z} &= \frac{xz \pm yw}{yz}, \\ \frac{x}{y} \cdot \frac{w}{z} &= \frac{xw}{yz}, \text{ and} \\ \frac{x}{y} / \frac{w}{z} &= \frac{xz}{yw}.\end{aligned}$$

Assume that any time a term appears in a denominator that it does not equal 0.

Exercise 26. Prove that if $xy = 0$, then either $x = 0$ or $y = 0$.

2.1.4 Subfields

Frequently we'll find one field contained in another field. For instance, the field of rational numbers \mathbf{Q} is part of the field of real numbers \mathbf{R} , and \mathbf{R} is part of the field of complex numbers \mathbf{C} . They're not just subsets, $\mathbf{Q} \subset \mathbf{R} \subset \mathbf{C}$, but they have the same operations. Here's the precise definition of subfield.

Definition 2.2 (subfield). A field E is a *subfield* of a field F if

1. the underlying set of E is a subset of the underlying set of F ;
2. the addition operation $+_E$ on E is the restriction of the addition operation $+_F$ on F , that is, for all x and y in E , $x +_E y = x +_F y$; and
3. the multiplication operation \cdot_E on E is the restriction of the multiplication operation \cdot_F on F , that is, for all x and y in E , $x \cdot_E y = x \cdot_F y$.

When E is a subfield of F , we'll also say that F is an *extension* of E .

When you know one field is a subfield of another, there's no need to subscript the operations since they are the same.

There is an alternate characterization of subfield. The proof of the following theorem is straightforward, but there are many steps.

Theorem 2.3. If a subset E of a field F has 0, 1, and is closed under addition, multiplication, negation, and reciprocation of nonzero elements, then E is a subfield of F .

The field of rational numbers \mathbf{Q} . When we're trying to find the smallest example of a field, it looks like it will have to be \mathbf{Q} . Later in section 2.2 we'll see that it's not the smallest! But here's an argument (which must have a flaw in it) which says we need all the rational numbers to be in any field F .

To begin with, 0 and 1 have to be in F . But we also have to have $1 + 1$ in F and we'll denote that 2, of course. And we'll need $1 + 1 + 1 = 2 + 1$ which we'll denote 3. And so forth, so we've got 0 and all the positive integers in F . We also need negations of them, so all the negative integers are in F , too. But a rational number m/n is just an integer m divided by a positive integer n , so we'll have to have all rational numbers in F . That shows that \mathbf{Q} is a subfield of F .

Thus, it looks like every field F includes the smallest field \mathbf{Q} , the field of rational numbers.

There's one minor flaw in the argument above, but let's not pick it apart right now. Pretty soon we'll look at fields that don't contain \mathbf{Q} .

2.1.5 Fields of rational functions

A *rational function* with coefficients in F is a quotient of two polynomials $\frac{f(x)}{g(x)}$. Rational functions do form a field, the field $F(x)$ of rational functions with coefficients in F . Notice that the ring of polynomials $F[x]$ is denoted with square brackets while the field of rational functions $F(x)$ is denoted with round parentheses.

For example, one rational function in $\mathbf{Q}(x)$ is $\frac{5x^2 - 3x + 1/2}{x^3 + 27}$.

Note that the field F is a subfield of the $F(x)$. Again, we can identify the constant rational functions as the elements of F . For example, \mathbf{Q} is a subfield of $\mathbf{Q}(x)$, and both \mathbf{R} and $\mathbf{Q}(x)$ are subfields of $\mathbf{R}(x)$.

Also, the the ring of polynomials with coefficients is a subring of the field of rational functions. That is $F \subseteq F[x] \subseteq F(x)$.

2.1.6 Vector spaces over arbitrary fields

When you studied vector spaces, you may have studied only vector spaces over the real numbers, although vector spaces over other fields might have been mentioned. In fact, vector spaces over an arbitrary field F have the same basic properties as vector spaces over \mathbf{R} .

The n -dimensional standard vector space F^n is defined the same way as \mathbf{R}^n except the n -tuples have coordinates in F . Addition and scalar multiplication are defined the same way for F^n as they are for \mathbf{R}^n .

Furthermore, matrices in $M_{m \times n}(F)$ are defined the same way as matrices in $M_{m \times n}(\mathbf{R})$ except the entries are in F instead of \mathbf{R} . The matrix operations are the same. You can use the same methods of elimination to solve a system of linear equations with coefficients in F or find the inverse of a matrix in $M_{n \times n}(F)$ if its determinant is nonzero. Determinants have the same properties. You can use methods of linear algebra to study geometry in F^n just as you can for \mathbf{R}^n (although it may not be possible to visualize what F^n is supposed to look like, and things like areas of triangles have values in F).

The abstract theory of finite dimensional vector spaces over F is the same, too. Linear independence, span, basis, dimension are all the same. Rank and nullity of a matrix are the same. Change of basis is the same.

Eigenvalues, eigenvectors, and eigenspaces may have problems over some fields. In fact, when you studied transformations $\mathbf{R}^n \rightarrow \mathbf{R}^n$, sometimes you had complex eigenvalues, and their only eigenvectors were in \mathbf{C}^n . Likewise when looking at transformations $F^n \rightarrow F^n$ and the eigenvalues aren't in F , you'll may have to go to some field extension F' of F to find them and to F'^n to find the eigenvectors.

Likewise, canonical forms for matrices will depend on F .

2.2 Cyclic rings and finite fields

In this section we'll look at fields that are finite, and we'll discover that \mathbf{Q} actually isn't the smallest field. Although they're smaller fields—they're finite—they won't be subfields of \mathbf{Q} .

First we'll look a bit at the concept of congruence modulo n , where n is a positive integer. Then we'll look at the ring of integers modulo n , denoted $\mathbf{Z}/n\mathbf{Z}$ or more simply \mathbf{Z}_n . We'll

see why they're called cyclic rings. Finally, we'll look at the case where n is prime, and we'll denote it p then, where \mathbf{Z}_p turns out to be a field, and we'll examine some of the cyclic fields.

Definition 2.4 (Congruence modulo n). Fix n , a positive integer. We say that two integers x and y are *congruent modulo n* if n evenly divides the difference $x - y$. We'll use the standard notation from number theory

$$x \equiv y \pmod{n}$$

to indicate that x is congruent to y modulo n , and the notation $n|m$ to indicate that the integer n divides the integer m (with no remainder). Then

$$x \equiv y \pmod{n} \quad \text{iff} \quad n|(x - y).$$

When n doesn't divide the difference $x - y$, we say a is not congruent to b , denoted $x \not\equiv y \pmod{n}$.

You're familiar with congruence modulo 12; it's what 12-hour clocks use.

The general theory of equivalence relations in section [A.2.3](#).

Theorem 2.5. Congruence modulo n is an equivalence relation.

Proof. For reflexivity, $x \equiv x \pmod{n}$ holds since $n|(x - x)$.

For symmetry, we need to show that $x \equiv y \pmod{n}$ implies $y \equiv x \pmod{n}$. But if $n|(x - y)$, then $n|(y - x)$.

For transitivity, suppose that $x \equiv y \pmod{n}$ and $y \equiv z \pmod{n}$. Then $n|(x - y)$ and $n|(y - z)$, so there exist k and m such that $nk = x - y$ and $nm = y - z$. Therefore $n(k + m) = x - z$, showing that $n|(x - z)$. Hence $x \equiv z \pmod{n}$. Q.E.D.

2.2.1 The cyclic ring \mathbf{Z}_n

Definition 2.6 (Integers modulo n). The integers modulo n , \mathbf{Z}_n is the set of equivalence classes of integers under the equivalence relation which is congruence modulo n .

We'll denote these equivalence classes with square brackets subscripted by n . Thus, for instance, the element 0 in \mathbf{Z}_6 is really $[0]_6$, which we'll denote $[0]$ when modulo 6 is understood. This equivalence class is the set of all x such that $x \equiv 0 \pmod{6}$. This $[0]_6 = \{\dots, -18, -12, -6, 0, 6, 12, 18, \dots\}$. Likewise the element 1 in \mathbf{Z}_6 is really the equivalence class of 1, which is the set

$$[1]_6 = \{x \in \mathbf{Z} \mid x \equiv 1 \pmod{6}\} = \{\dots, -17, -11, -5, 1, 7, 13, 19, \dots\}.$$

Note that $[1]_6 = [7]_6 = [13]_6$ all name the same equivalence class.

An equation in equivalence classes, such as $[x]_6 + [3]_6 = [5]_6$, is the same thing as an congruence, $x + 3 \equiv 5 \pmod{6}$. The congruence notation is usually more convenient. When the modulus n is known by context, we'll dispense with the subscript n , and abusing notation, we'll frequently drop the square brackets.

There are two ways you can think about integers modulo n . One is to think of them as regular integers from 0 through $n - 1$, do the arithmetic modulo n , and adjust your answer so it's in the same range. For example, we can take $\mathbf{Z}_6 = \{0, 1, 2, 3, 4, 5\}$. Then, to do some

computation, say $5(2 - 4) \pmod{6}$, first compute $5(2 - 4)$ as integers to get -10 , and then, since $-10 \equiv 2 \pmod{6}$, say the answer is 2 . That works very well for computation, but it's pretty messy when you're trying to do anything with variables or trying to prove anything in general.

A better way is to say that an element of \mathbf{Z}_n is named by an integer, but two integers name the same element of \mathbf{Z}_n if they're congruent modulo n . Thus, x and y name the same element of \mathbf{Z}_n if $x \equiv y \pmod{n}$. This will work because congruence modulo n is an equivalence relation as we saw earlier.

In any case, it helps conceptually to think of the elements of \mathbf{Z}_n as being arranged on a circle like we imagine the elements of \mathbf{Z} being arranged on a line. See figure 2.1 of a couple of cyclic rings \mathbf{Z}_n to see where the word “ring” came from.

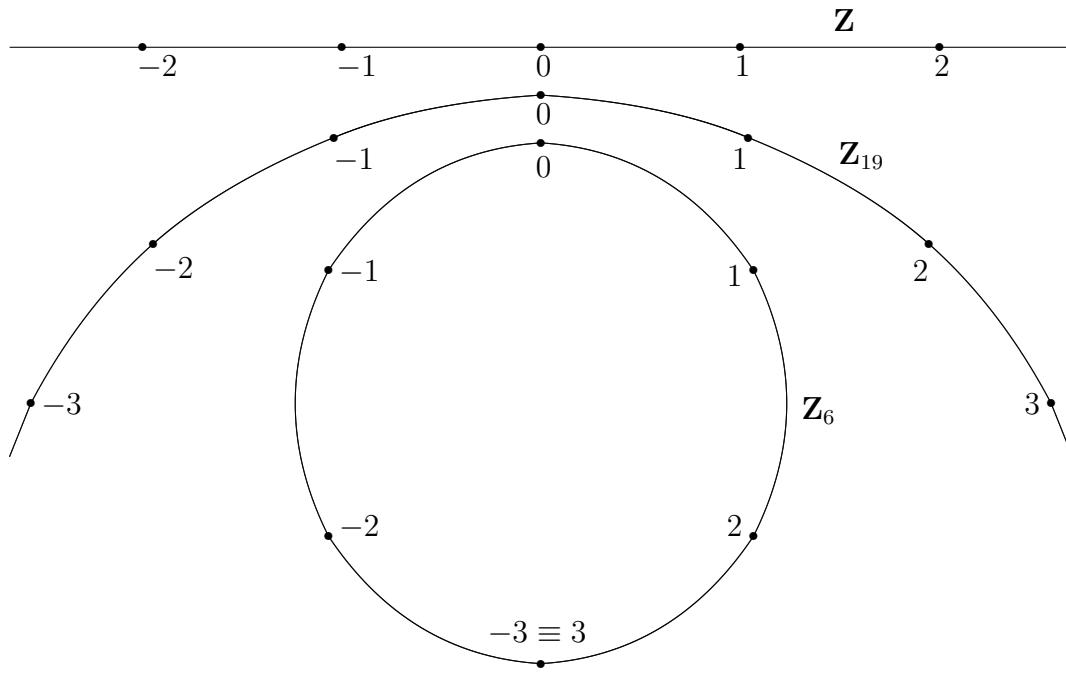


Figure 2.1: Cyclic rings $\mathbf{Z}_6, \mathbf{Z}_{19}, \mathbf{Z}$

The operations on \mathbf{Z}_n . Our equivalence relation is congruence modulo n , so our equivalence classes are also called congruence classes.

Congruence modulo n is more than just an equivalence relation; it works well with addition, subtraction, and multiplication, as you can easily show.

Theorem 2.7. If $x \equiv y \pmod{n}$, and $u \equiv v \pmod{n}$, then $x + u \equiv y + v \pmod{n}$, $x - u \equiv y - v \pmod{n}$, and $xu \equiv yv \pmod{n}$.

These properties will allow us to define a ring structure on \mathbf{Z}_n , as done below.

But congruence modulo n doesn't work so well with division. Although $6 \equiv 0 \pmod{6}$, it is not the case that $6/2 \equiv 0/2 \pmod{6}$. Thus, we can't expect that \mathbf{Z}_n will be a field, at least when $n = 6$.

Our job is to define addition, subtraction, and multiplication on \mathbf{Z}_n . Whenever a set is defined as a quotient set, that is, on equivalence classes, as \mathbf{Z}_n is, an extra step is required when defining an operation on it, as we'll see.

We would like to define addition on \mathbf{Z}_n by saying $[x] + [u] = [x + u]$, that is, the sum of the equivalence class of x and the equivalence class of u should be the equivalence class of $x + u$. But what if we named the equivalence class x by some other integer, say y , and the equivalence of u by some other integer v ? How do we know we that $[y + v]$ is the same equivalence class as $[x + u]$? We can state this question in a couple of other ways. How do we know

$$[x] = [y] \text{ and } [u] = [v] \text{ implies } [x + u] = [y + v]?$$

That asks the question: how do we know

$$x \equiv y \pmod{n} \text{ and } u \equiv v \pmod{n} \text{ implies } x + u \equiv y + v \pmod{n}?$$

That's one of the properties of congruence mentioned above. That property says addition on \mathbf{Z}_n is "well-defined".

Likewise, since multiplication works well with congruence,

$$x \equiv y \pmod{n} \text{ and } u \equiv v \pmod{n} \text{ imply } xu \equiv yv \pmod{n},$$

we can define multiplication on \mathbf{Z}_n by $[x] \cdot [u] = [xu]$.

Furthermore, all the ring axioms will be satisfied in \mathbf{Z}_n since they're satisfied in \mathbf{Z} . Thus, \mathbf{Z}_n is a ring, and it's called a *cyclic ring*.

The projection $\gamma : \mathbf{Z} \rightarrow \mathbf{Z}_n$. The function $\gamma : \mathbf{Z} \rightarrow \mathbf{Z}_n$ defined by $\gamma(k) = [k]$ maps an integer to its equivalence class modulo n . We defined addition and multiplication in \mathbf{Z}_n

$$[x + u] = [x] + [u] \quad \text{and} \quad [xu] = [x][u]$$

so γ preserves addition and multiplication. Furthermore, since $\gamma(1) = [1]$, it preserves 1. Therefore γ is a ring homomorphism. It is, of course, onto, so it is a ring epimorphism. It's called a *projection* or a *canonical homomorphism* to the quotient ring.

In section 3.6, we'll generalize this construction to rings besides \mathbf{Z} and their quotients, and we'll have projections for the generalizations, too.

The characteristic of a ring. What's weird about these cyclic rings is that if you start with 1 and add 1 over and over, you'll reach zero. For instance, in \mathbf{Z}_5 , we have $1+1+1+1+1 = 5 \equiv 0 \pmod{5}$. This corresponds to the geometric interpretation of these cyclic rings being shaped like rings.

Definition 2.8. If some multiple of 1 equals 0 in a ring, then the *characteristic* of the ring is the smallest such multiple. If no multiple of 1 equals 0, then the characteristic is said to be 0.

We're primarily interested in characteristics when we're talking about fields, and we'll see soon that the characteristic of a field is either 0 or a prime number.

Example 2.9. The characteristic of \mathbf{Z}_5 is 5, and, in general, the characteristic of a finite cyclic ring \mathbf{Z}_n is n .

2.2.2 The cyclic prime fields \mathbf{Z}_p

Since division doesn't work well with congruence, we can't expect \mathbf{Z}_n to always have reciprocals, so we don't expect it to be a field. Let's first see when an element in \mathbf{Z}_n is a unit. The term *unit* in a ring refers to an element x of the ring that does have a reciprocal. 1 is always a unit in a ring, and every nonzero element in a field is a unit.

Theorem 2.10. An element k in \mathbf{Z}_n is a unit if and only if k is relatively prime to n .

Proof. First, suppose that k is a unit in \mathbf{Z}_n . That means there exists l such that $kl \equiv 1 \pmod{n}$. Then $n|(kl - 1)$, and hence n is relatively prime to k .

Second, suppose that k is relatively prime to n . Then, by the extended Euclidean algorithm, their greatest common divisor, 1, is a linear combination of k and n . Thus, there are integers x and y so that $1 = xk + yn$. Then $1 \equiv xk \pmod{n}$, and k does have a reciprocal, namely x , in \mathbf{Z}_n . Thus k is a unit in \mathbf{Z}_n . Q.E.D.

Recall from definition 1.39 that the totient function $\varphi(n)$ denotes the number of positive integers less than n that are relatively prime to n .

Corollary 2.11 (Units in \mathbf{Z}_n). The number of units in \mathbf{Z}_n is $\varphi(n)$.

Theorem 2.12. The cyclic ring \mathbf{Z}_n is a field if and only if n is prime.

Proof. Part of this theorem is a direct corollary of the previous one. Suppose n is prime. Then every nonzero element of \mathbf{Z}_n is relatively prime to n . Therefore, \mathbf{Z}_n is a field.

Next we'll show that if n is composite, the ring is not a field. Let n be the product of two integers m and k , both greater than 1. Then neither m nor k can have a reciprocal in \mathbf{Z}_n . Why not? Suppose that m^{-1} did exist in \mathbf{Z}_n . Then

$$\begin{aligned} (m^{-1}m)k &\equiv 1k \equiv k \pmod{n} \\ m^{-1}(mk) &\equiv m^{-1}n \equiv 0 \pmod{n} \end{aligned}$$

But $k \not\equiv 0 \pmod{n}$, a contradiction. So m^{-1} doesn't exist. Therefore, \mathbf{Z}_n is not a field.

Q.E.D.

Corollary 2.13. The characteristic of a field is 0 or a prime number.

Proof. We'll show that if the characteristic n is finite, it must be a prime number. Suppose $n = st$. Then $0 = n \cdot 1 = (st) \cdot 1 = (s \cdot 1)(t \cdot 1)$. Therefore, either $s \cdot 1 = 0$ or $t \cdot 1 = 0$. But n is the smallest positive integer such that $n \cdot 1 = 0$, so either $s = n$ or $t = n$. Therefore n is prime. Q.E.D.

This proof works as well in integral domains introduced in section 3.1.3. This theorem will be mentioned again at that time.

Example 2.14. \mathbf{Z}_2 . Note that there is only one nonzero element, namely 1, and it is its own inverse. The addition and multiplication tables for \mathbf{Z}_2 are particularly simple.

+	0	1	·	0	1
0	0	1	0	0	0
1	1	0	1	0	1

Note that subtraction is the same as addition in \mathbf{Z}_2 since $x - y \equiv x + y \pmod{2}$.

Example 2.15. \mathbf{Z}_3 . Here, there are two nonzero elements, namely 1 and 2, but, for symmetry's sake, we'll call the two nonzero elements 1 and -1 . Note that each of these two are their own inverses. The addition and multiplication tables are still pretty simple.

+	-1	0	1	.	-1	0	1
-1	1	-1	0	-1	1	0	-1
0	-1	0	1	0	0	0	0
1	0	1	-1	1	-1	0	1

Example 2.16. \mathbf{Z}_{13} . What are the reciprocals of the 12 nonzero elements? We can name the nonzero elements as $\pm 1, \pm 2, \dots, \pm 6$. You can verify that this table gives their inverses.

x	± 1	± 2	± 3	± 4	± 5	± 6
x^{-1}	± 1	∓ 6	∓ 4	∓ 3	∓ 5	∓ 2

For instance, the reciprocal of 2 is -6 since $2(-6) \equiv -12 \equiv 1 \pmod{13}$.

These fields, \mathbf{Z}_p where p is prime, are the finite prime fields. But there are other finite fields.

Example 2.17. A field of order 9. We'll make an extension of \mathbf{Z}_3 to get a field of order 9. Note that -1 is not a square modulo 3. We can append $\sqrt{-1}$ to \mathbf{Z}_3 to get a field algebraic over it in exactly the same way we got \mathbf{C} from \mathbf{R} . Let's use i as an abbreviation for $\sqrt{-1}$, as usual. Then

$$\mathbf{Z}_3(i) = \{x + yi \mid x, y \in \mathbf{Z}_3\}$$

Addition, subtraction, and multiplication give us no problems. We just have to check that nonzero elements have inverses. That's exactly as before.

$$\frac{1}{x+yi} = \frac{x-yi}{(x+yi)(x-yi)} = \frac{x-yi}{x^2+y^2} = \frac{x}{x^2+y^2} + \frac{-y}{x^2+y^2}i$$

Thus, if $x+yi$ is not 0 in $\mathbf{Z}_3(i)$, that is, not both of x and y are congruent to 0 modulo 3, then $x^2+y^2 \not\equiv 0 \pmod{3}$, and the expression on the right gives $(x+yi)^{-1}$. Note that the characteristic of this field is 3 since $1+1+1$ is 0 in this field.

Exercise 27. You can construct a field of order 25 from \mathbf{Z}_5 , but it has to be done somewhat differently because $\sqrt{-1}$ already exists in \mathbf{Z}_5 since $(\pm 2)^2 = 4 \equiv -1$ in \mathbf{Z}_5 . The squares of the nonzero elements in \mathbf{Z}_5 include $1 = (\pm 1)^2$ and $4 = (\pm 2)^2$, but 2 is not among the squares. Show that the ring $\mathbf{Z}_5[\sqrt{2}]$ is a field by finding an inverse of a nonzero element $x+y\sqrt{2}$ where x and y are elements of \mathbf{Z}_5 but not both are 0. Hint: $(x-y\sqrt{2})(x+y\sqrt{2}) = x^2 - 2y^2$ cannot be 0.

In fact, there are finite fields of order p^n for each power of a prime p . These are called the Galois fields $GF(p^n)$. Note that cyclic prime field are the simplest Galois fields; \mathbf{Z}_p is $GF(p)$. The example constructed $GF(3^2)$ and the exercise $GF(5^2)$.

The proof that a finite field of characteristic p has to have p^n elements follows from the theorems in section 4.9.2 on Abelian groups. It's only dependent on addition in the finite field, not on multiplication.

2.2.3 Characteristics of fields, and prime fields

The characteristic of a ring was defined above, so we already have the definition for the characteristic of a field.

Those fields that have characteristic 0 all have \mathbf{Q} as a subfield. The flawed proof we saw earlier included the mistaken assumption that all the elements $0, 1, 2, \dots$ were distinct, which, as we've seen with these finite fields, isn't always the case. But we can correct the flawed proof to validate the following theorem. First, a definition.

Definition 2.18. A *prime field* is a field that contains no proper subfield. Equivalently, every element in it is a multiple of 1.

Theorem 2.19. Each field F has exactly one of the prime fields as a subfield. It will have \mathbf{Z}_p when it has characteristic p , but it will have \mathbf{Q} if it has characteristic 0.

The Frobenius endomorphism. Exponentiation to the power p has an interesting property when a commutative ring R has prime characteristic p :

$$(x + y)^p = x^p + y^p$$

There are various ways to prove this. For instance, you can show that the binomial coefficient $\binom{p}{k}$ is divisible by p when $1 < k < p$. Since $\binom{p}{k} = \frac{p!}{k!(n-k)!}$ and p divides the numerator but not the denominator, therefore it divides $\binom{p}{k}$.

This function $\varphi : R \rightarrow R$ defined by $\varphi(x) = x^p$ also preserves 1 and multiplication: $1^p = 1$ and $(xy)^p = x^p y^p$. Therefore, it is a ring endomorphism, called the *Frobenius endomorphism*.

We're most interested in the endomorphism when the ring is a field F of characteristic p . It's not particularly interesting when F is the prime field \mathbf{Z}_p because it's just the identity function then. For other finite fields of characteristic p it will be an automorphism—it's a bijection since it's an injection on a finite set—and it's not the identity function for those fields.

Example 2.20. In the example above of the Galois field $GF(3^2) = \mathbf{Z}_3(i)$, the characteristic of the field is 3, so $\varphi(x + yi) = (x + yi)^3 = x^3 + (yi)^3 = x^3 - y^3i = x - yi$. On the subfield \mathbf{Z}_3 , φ is the identity, but not on all of $GF(3^2) = \mathbf{Z}_3(i)$, since $\varphi(i) = -i$.

Exercise 28. Determine the value of $\varphi(\sqrt{2})$ in $GF(5^2)$.

2.3 Field Extensions, algebraic fields, the complex numbers

A lot of fields are found by extending known fields. For instance, the field of complex numbers \mathbf{C} is extended from the field of real numbers \mathbf{R} and $GF(3^2)$ is extended from $\mathbf{Z}_3 = GF(3)$. We'll look at the general case of extending fields by adding square roots to known fields, the smallest kind of extension, called a *quadratic extension*.

2.3.1 Algebraic fields

We've looked at some quadratic extensions of fields. Now we'll look at algebraic extensions in more detail.

Definition 2.21 (Algebraic and transcendental numbers). An *algebraic* number is a number that is a root of a polynomial with rational coefficients. If the polynomial is monic, then the algebraic number is an *algebraic integer*. A real number or a complex number that is not algebraic is called a *transcendental* number.

For instance, $x = \sqrt{2}$ is an algebraic number since it is the root of the polynomial $x^2 - 2$; in fact, it's an algebraic integer. On the other hand, $x = \sqrt{1/2}$ is a root of the polynomial $2x^2 - 1$, so it's an algebraic number, but not an algebraic integer.

There are many real numbers used in analysis that are transcendental. In 1873 Charles Hermite (1882–1901) proved that the number e is transcendental. It follows that many related numbers are transcendental such as e^2 and \sqrt{e} .

Definition 2.22 (Algebraic and transcendental field extensions). More generally, if x satisfies a polynomial equation $f(x) = 0$ where the polynomial f has coefficients in a field F , then we say x is *algebraic* over F . A field extension F' of F , all of whose elements are algebraic over F is said to be an *algebraic* extension of F . Field extensions that are not algebraic are called *transcendental* extensions. An algebraic extension of \mathbf{Q} is also called an *algebraic number field*, or more simply a number field.

In 1882 Lindemann extended Hermite's result to show that e^a is transcendental for all nonzero algebraic numbers a . Thus $e^{\sqrt{2}}$ is transcendental. More importantly, Lindemann's theorem shows that $\pi = e^i$ is transcendental, for if it were algebraic, then $e^{\pi i} = -1$ would be transcendental, which it isn't.

Weierstrass proved an even more general theorem in 1885. If a_1, \dots, a_n are distinct nonzero algebraic numbers, then the numbers e^{a_1}, \dots, e^{a_n} are algebraically independent meaning each e^{a_i} is transcendental over the field $\mathbf{Q}(e^{a_1}, \dots, \hat{e^{a_i}}, e^{a_n})$. The hat over e^{a_i} means that is omitted from the list.

Example 2.23. We know that the square root of 2, $\sqrt{2}$ is not a rational number. The field $\mathbf{Q}(\sqrt{2})$ is the smallest field that contains $\sqrt{2}$. In fact, its elements are all of the form

$$x + y\sqrt{2} \quad \text{where } x \in \mathbf{Q} \text{ and } y \in \mathbf{Q}.$$

It's pretty obvious that most of the field axioms hold. The only one that's not obvious is the existence of reciprocals of nonzero elements, that is to say, the statement " $(x + y\sqrt{2})^{-1}$ is of the form $x' + y'\sqrt{2}$ where x' and y' are rational and not both 0" is not so obvious. But the trick of "rationalizing the denominator" shows us how.

$$\frac{1}{x + y\sqrt{2}} = \frac{x - y\sqrt{2}}{(x + y\sqrt{2})(x - y\sqrt{2})} = \frac{x - y\sqrt{2}}{x^2 - 2y^2} = \frac{x}{x^2 - 2y^2} + \frac{-2y}{x^2 - 2y^2}\sqrt{2}$$

Note that $x^2 - 2y^2$ cannot be 0 when x and y are rational and not both 0. For if $x^2 - 2y^2 = 0$, then $2 = (x/y)^2$, and then $\sqrt{2}$ would be a rational number, which it isn't. Thus, $\mathbf{Q}(\sqrt{2})$ is a field.

The trick was to multiply and divide by the *conjugate*. Let's give a notation to this conjugate: $x + y\sqrt{2} = x - y\sqrt{2}$. Conjugation has some nice properties. It preserves all the elements of the base field \mathbf{Q} , that is, if $x \in \mathbf{Q}$, then $\bar{x} = x$. It preserves addition and multiplication, that is, if α and β are elements of $\mathbf{Q}(\sqrt{2})$, then $\bar{\alpha + \beta} = \bar{\alpha} + \bar{\beta}$ and $\bar{\alpha\beta} = \bar{\alpha}\bar{\beta}$. Finally, the operation of conjugation, $\bar{} : \mathbf{Q}(\sqrt{2}) \rightarrow \mathbf{Q}(\sqrt{2})$, is its own inverse, $\bar{\bar{x}} = x$. Thus, conjugation is a field automorphism. Furthermore, the elements α it fixes, $\bar{\alpha} = \alpha$, are just the elements of the base field \mathbf{Q} .

2.3.2 The field of complex numbers \mathbf{C}

In the same way we just adjoined $\sqrt{2}$ to \mathbf{Q} to get $\mathbf{Q}(\sqrt{2})$, we can adjoin $\sqrt{-1}$ to \mathbf{R} to get $\mathbf{R}(\sqrt{-1})$, which is \mathbf{C} . Algebraically, the process is identical, but conceptually it's a little different because we thought that $\sqrt{2}$, being a real number, existed before we appended it to \mathbf{Q} , while it may not be so clear that $\sqrt{-1}$ exists before we append it to \mathbf{R} . But $\sqrt{-1}$, usually denoted i , has the property $i^2 = -1$, so it is an algebraic number since it's the root of the polynomial $x^2 + 1$. In fact, $\mathbf{R}(i)$ consists of elements of the form

$$x + yi \quad \text{with} \quad x, y \in \mathbf{R}$$

as described by Euler. Addition and subtraction are “coordinatewise”

$$(x_1 + y_1 i) \pm (x_2 + y_2 i) = (x_1 \pm x_2) + (y_1 \pm y_2)i$$

while multiplication is only slightly more complicated

$$\begin{aligned} (x_1 + y_1 i)(x_2 + y_2 i) &= x_1 x_2 + x_1 y_2 i + x_2 y_1 i + y_1 y_2 i^2 \\ &= (x_1 x_2 - y_1 y_2) + (x_1 y_2 + x_2 y_1)i \end{aligned}$$

We can find reciprocals by rationalizing the denominator as we did above.

$$\frac{1}{x + yi} = \frac{x - yi}{(x + yi)(x - yi)} = \frac{x - yi}{x^2 + y^2} = \frac{x}{x^2 + y^2} + \frac{-y}{x^2 + y^2}i$$

We can define complex *conjugation* by $\bar{x + yi} = x - yi$. It's a field automorphism of \mathbf{C} , and its fixed subfield is \mathbf{R} .

We can also define a *norm* on \mathbf{C} once we have conjugation. For $z = x + yi \in \mathbf{Q}$, let

$$|z|^2 = z\bar{z} = (x + yi)(x - yi) = x^2 + y^2.$$

Since $|z|^2$ is a nonnegative real number, it has a square root $|z|$.

A matrix representation of \mathbf{C} . Consider the subset C of the matrix ring $M_2(\mathbf{R})$ consisting of matrices of the form

$$\begin{bmatrix} x & y \\ -y & x \end{bmatrix} \quad \text{where} \quad x, y \in \mathbf{R}.$$

You can easily show that this is a subring of $M_2(\mathbf{R})$ since the 0 matrix and the identity matrix are of this form, the sum and difference of matrices of this form are of this form, and so is the product as you can see here

$$\begin{bmatrix} x & y \\ -y & x \end{bmatrix} \begin{bmatrix} u & v \\ -v & u \end{bmatrix} = \begin{bmatrix} xu - yv & xv + yu \\ -yu - vx & -yv + xu \end{bmatrix}.$$

Thus, C is a subring of $M_2(\mathbf{R})$. Furthermore, it's a commutative subring even though $M_2(\mathbf{R})$ is not a commutative ring since the same product results when the two factors are interchanged:

$$\begin{bmatrix} u & v \\ -v & u \end{bmatrix} \begin{bmatrix} x & y \\ -y & x \end{bmatrix} = \begin{bmatrix} ux - vy & uy + vx \\ -vx - uy & -vy + ux \end{bmatrix}.$$

Furthermore C is a field because nonzero matrices in it have inverses. For suppose not both x and y are 0. Then

$$\begin{bmatrix} x & y \\ -y & x \end{bmatrix} \begin{bmatrix} \frac{x}{x^2+y^2} & \frac{-y}{x^2+y^2} \\ \frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In fact, C is isomorphic to the complex field \mathbf{C} as described above. The isomorphism is described by the one-to-one correspondence

$$\begin{bmatrix} x & y \\ -y & x \end{bmatrix} \leftrightarrow x + yi.$$

Note that a real number x corresponds to the matrix $\begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix}$ while a purely imaginary number yi corresponds to the matrix $\begin{bmatrix} 0 & y \\ -y & 0 \end{bmatrix}$.

Note that complex conjugation in this representation is just matrix transposition.

This alternate representation of the complex numbers as matrices directly explains how a complex number acts as a linear transformation on the real plane \mathbf{R}^2 . The complex number $x + yi$ maps a point (a, b) of \mathbf{R}^2 to the point $(ax + by, -ay + bx)$ since

$$\begin{bmatrix} x & y \\ -y & x \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} ax + by \\ -ay + bx \end{bmatrix}.$$

Matrix representations of various fields, rings, and groups are useful for two reasons. One is that they give us geometric interpretations for the elements as illustrated above. The other is that all the tools of linear algebra are available to us once we have the matrix representation.

2.3.3 General quadratic extensions

Now that we've seen a couple of quadratic extensions, let's see how it works in general.

Let F be a field and e an element of F that is not a square. In other words, the polynomial $x^2 - e$ has no roots in F . We'll consider ordered pairs $(a_1, a_2) \in F \times F$, but we'll write them as $a_1 + a_2\sqrt{e}$. We'll define addition coordinatewise

$$(a_1 + a_2\sqrt{e}) + (b_1 + b_2\sqrt{e}) = (a_1 + b_1) + (a_2 + b_2)\sqrt{e}$$

and define multiplication by

$$(a_1 + a_2\sqrt{e})(b_1 + b_2\sqrt{e}) = (a_1b_1 + ea_2b_2) + (a_1b_2 + a_2b_1)\sqrt{e}.$$

You can check that these definitions give us a ring. But, does it give us a field? As we did before, we'll find a reciprocal of a nonzero element $a_1 + a_2\sqrt{e}$

$$\frac{1}{a_1 + a_2\sqrt{e}} = \frac{a_1 - a_2\sqrt{e}}{(a_1 + a_2\sqrt{e})(a_1 - a_2\sqrt{e})} = \frac{a_1 - a_2\sqrt{e}}{a_1^2 - ea_2^2}$$

In order for this to be the reciprocal, all we have to do is show the denominator $a_1^2 - ea_2^2$ is not 0. In the case that $a_2 = 0$ we know $a_1 \neq 0$ since not both are 0, so in that case $a_1^2 - ea_2^2$ is not 0. That leaves us the case that $a_2 \neq 0$. Suppose that $a_1^2 - ea_2^2 = 0$. Then $ea_2^2 = a_1^2$, and dividing by a_2^2 , we conclude $e = (a_1/a_2)^2$. But e is not a square in F . Thus $a_1^2 - ea_2^2$ is not 0 in this case, too. Therefore, we've found the reciprocal.

Thus, we have a field, $F(\sqrt{e})$.

When we look at more general field extensions, we'll have a lot more theory, and we won't have details to check as we did here. That theory will involve the concept of "ideals" in a ring as discussed in section 3.6.

2.4 Real numbers and ordered fields

We'll look now at **R**, the field of real numbers. What's so special about the real number field? For one thing, it's got an order on it; we can compare two real numbers x and y and say which is smaller or if they're equal. That's an extra structure on a field. We'll start by looking at this concept of ordered field.

Before we get too far, you should know that that isn't enough to distinguish **R** from other fields. There are plenty of other ordered fields, such as **Q** and all the fields between **Q** and **R**.

2.4.1 Ordered fields

The easiest way to define an ordered field is by saying it's partitioned into positive elements, negative elements, and 0, and requiring a couple properties on these parts.

Definition 2.24 (Ordered field). An *ordered field* consists of a field F along with a subset P whose elements are called *positive* such that

1. F is partitioned into three parts: P , $\{0\}$, and N where

$$N = \{x \in F \mid -x \in P\}$$

the elements of N are called *negative*;

2. the sum of two positive elements is positive; and
3. the product of two positive elements is positive.

Properties of ordered fields. You can show from this definition that

1. the sum of negative elements is negative
2. the product of a negative element and a positive element is negative
3. the product of two negative elements is positive
4. 1 is positive, and -1 is negative

Exercise 29. Prove the four properties above.

Examples. \mathbf{R} , \mathbf{Q} , and all fields between them are ordered fields where the usual positive numbers in the field form P .

Although \mathbf{Q} and \mathbf{R} are ordered fields, finite fields and \mathbf{C} have no ordering.

Exercise 30. Show that \mathbf{C} is not an ordered field. Hint: show why i can't be positive, zero, or negative.

The binary order relations. From P we can define the binary order relations $<$, \leq , $>$, and \geq . For instance, $x \leq y$ means $y - x$ is zero or positive, while $x < y$ means $y - x$ is positive. That can be stated formally as follows:

$$x \leq y \quad \text{iff} \quad y - x \in P \cup \{0\}$$

$$x < y \quad \text{iff} \quad y - x \in P.$$

All the expected properties of these order relations follow. Here are a few.

1. Trichotomy: For each pair x, y , exactly one of the three relations $x < y$, $x = y$, or $x > y$ holds.
2. Transitivity: $x < y$ and $y < z$ imply $x < z$.
3. If x is positive and $y < z$, then $xy < xz$.
4. If x is negative and $y < z$, then $xy > xz$.
5. If x is positive, then so is $1/x$.
6. For positive x and y , if $x < y$, then $1/y < 1/x$.

Exercise 31. Prove the six properties above.

Theorem 2.25. The characteristic of an ordered field is 0.

Proof. Suppose F is an ordered field of characteristic $p \neq 0$. Since 1 is positive, then any sum of 1s will be positive. Then p is positive. But p equals 0 which is not positive. A contradiction. Therefore an ordered field cannot have nonzero characteristic. Q.E.D.

It follows that \mathbf{Q} is a subfield of every ordered field.

Example 2.26. An ordered extension of the real numbers with infinite elements and infinitesimal elements.

We can give the field of rational functions $\mathbf{R}(x)$ an order as follows. First, we'll define when a polynomial $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ is positive, and that will be when its leading coefficient a_n is a positive real number. Next, we'll define when a rational function $f(x)/g(x)$ is positive, and that will be when f and g are both positive polynomials or both negative polynomials. It follows that $f(x)/g(x)$ is negative one of f and g is positive and the other is negative. Only $0/g(x)$, which equals 0, won't be positive or negative. You can easily show that the sum and product of positive rational functions is positive.

The real numbers \mathbf{R} is an ordered subfield of $\mathbf{R}(x)$, meaning that it's a subfield and its elements have the same order whether the order on \mathbf{R} is used or the order on $\mathbf{R}(x)$ is used.

With this order, there are elements that are larger than any real number a , for example, $x > a$ since $x - a$ is positive. In other words, x is an infinite element. Likewise, there are positive elements that are smaller than any positive real number, $1/x$, for example, so $1/x$ is an infinitesimal number.

2.4.2 Archimedean orders

The last example is an example of an ordered field with infinite elements and infinitesimals. Every ordered field F is an extension of \mathbf{Q} , so we can define an infinite element of F to be an element $x \in F$ greater than every rational number, and we can define a positive infinitesimal element as a positive $x \in F$ smaller than every positive rational number. Note that the reciprocal of an infinite element is an infinitesimal, and vice versa.

Definition 2.27. An *Archimedean ordered field* or, more simply, an *Archimedean field*, is simply an ordered field F without infinite elements or infinitesimals.

Before Archimedes, Euclid used this property in his *Elements* in Book V and following books. The content of Book V is due to Eudoxus, so a better name for the Archimedean property would have been Eudoxus' property.

There are equivalent characteristics that could be used for the definition. Here are two. Each element of F is less than some integer. Each positive element of F is greater than the reciprocal of some positive integer.

Of course, the preceding example is a non-Archimedean field. Another interesting non-Archimedean field is that of surreal numbers created by John Conway. Surreal numbers include all real numbers, all ordinal numbers and more. Since ordinal numbers form a proper class, so do surreal numbers. For a nice introduction on surreal numbers, see Donald Knuth's book *Surreal Numbers*.

Still, there are loads of Archimedean fields, namely \mathbf{Q} , \mathbf{R} , and all the intermediate fields. We still haven't answered the question about what makes \mathbf{R} special. Before we go on, however, let's see how elements in an Archimedean field are determined by how they compare to rational numbers.

For an Archimedean field F , since F is ordered, it has characteristic 0, so it has as a subfield, indeed, an ordered subfield, the field of rational numbers \mathbf{Q} .

Theorem 2.28 (Density). Between any two distinct elements of an Archimedean field, there lies a rational number.

Proof. Let $x < y$ in an Archimedean field. We're looking for a rational number $\frac{m}{n}$ between x and y . If x is negative while y is positive, then the rational number 0 lies between them. We can reduce the case where they're both negative to the case where they're both positive by noting that if $\frac{m}{n}$ lies between $-x$ and $-y$, then $-\frac{m}{n}$ lies between x and y .

So we may assume that both x and y are positive. If we can find some multiple n of them so that $ny - nx > 1$, then some integer m lies between ny and nx , but $nx < m < ny$ gives $x < \frac{m}{n} < y$. And we can find such a multiple since $y - x$ is greater than the reciprocal $\frac{1}{n}$ of some positive integer since the field is Archimedean. Q.E.D.

An element a of F partitions \mathbf{Q} into two parts (L_a, R_a)

$$L_a = \{x \in \mathbf{Q} \mid x < a\} \text{ and } R_a = \{x \in \mathbf{Q} \mid x \geq a\}.$$

These two parts have a special property.

Definition 2.29. A *Dedekind cut* of the rational numbers is a partition of \mathbf{Q} into two nonempty parts (L, R) —a left part L and a right part R —such that every element of L is less than every element of R . Furthermore, the left part does not have a greatest element.

Theorem 2.30. An element a of an Archimedean field F is determined by its Dedekind cut (L_a, R_a) . That is, if $(L_a, R_a) = (L_b, R_b)$, then $a = b$.

Proof. If $a \neq b$, then there is a rational number between them, so that rational number will be in one left part but the other right part. Q.E.D.

In an Archimedean field F not every Dedekind cut has to determine an element. For example, in \mathbf{Q} , the cut (L, R) where $L = \{x \mid x < 0 \text{ or } x^2 \leq 2\}$ and $R = \{x \mid x > 0 \text{ and } x^2 > 2\}$ is not the cut of any rational number. But that same cut with $\sqrt{2}$ included in \mathbf{R} is the cut of $\sqrt{2}$. The real numbers are special in that every cut is the cut of some real number.

Although there might not be an element of F for every cut, the cuts are enough to determine, along with the order on F and the field structure of \mathbf{Q} , the field structure of F .

It helps in proofs to cut in half the information of a Dedekind cut from (L, R) to just L . It is sufficient to define a Dedekind cut just in terms of the left part. You can prove the following lemma to simplify the statement and the proof of the following theorem.

Lemma 2.31. If (L, R) is a Dedekind cut, then L has the following three properties

- i. L is a nonempty, proper subset of \mathbf{Q} ;
- ii. if $y \in L$ and $x \in \mathbf{Q}$ such that $x < y$, then $x \in L$; and
- iii. for each $x \in C$, there exists $y \in C$ such that $x < y$

Conversely, if L has these three properties, then (L, R) is a cut where R is the complement of L .

Theorem 2.32. In an Archimedean field F , addition and multiplication are determined by Dedekind cuts in the sense that If a and b are two elements of F , then the left part of their sum $a + b$ is determined by their left parts

$$L_{a+b} = \{x + y \mid x \in L_a \text{ and } y \in L_b\}.$$

If a and b are two positive elements of F , then the left part of their product is determined by their left parts

$$L_{ab} = \{xy \mid x \in L_a, x > 0, y \in L_b \text{ and } y > 0\} \cup \{x \mid x \leq 0\}.$$

2.4.3 Complete ordered fields

There are various definitions given for complete ordered fields, all logically equivalent. Here's one.

Definition 2.33. A *complete ordered field* is an Archimedean field that cannot be extended to a larger Archimedean field. Equivalently, every Dedekind cut determines an element of the field.

Completeness is the final property that characterizes \mathbf{R} . Actually, right now we haven't proved that there is *at least* one complete ordered field, and we haven't proved that there is *at most* one complete ordered field. Once we do, we can finally properly define \mathbf{R} .

Existence of a complete ordered field We'll start by stating the theorem which gives the components for one way of constructing a complete ordered field F . To make it complete, we just have to make sure that every Dedekind cut determines an element of the field. The way to do that, of course, to define the field to be the cuts, and the definition of the operations of addition and multiplication are determined by the cuts as seen in the last theorem.

Theorem 2.34. There is a complete ordered field F . Its elements are Dedekind cuts of \mathbf{Q} . If L_1 and L_2 are left parts of two cuts, then the left part of the sum is determined by the left part

$$L_+ = \{x + y \mid x \in L_1 \text{ and } y \in L_2\}.$$

If L is the left part a positive cut (one that contains at least one positive rational number), then its negation is determined by the left part

$$L_- = \{-x \mid x \notin L\}$$

except, if this L_- has a largest element, that largest element is removed. If L_1 and L_2 are left parts of two positive cuts, then the left part of the product is determined by the left part

$$L_\times = \{xy \mid x \in L_1, x > 0, y \in L_2 \text{ and } y > 0\} \cup \{x \mid x \leq 0\}.$$

There are many details to show to verify that R is a complete ordered field. First, that the sets L_+ , L_- , and L_\times are left parts. then the field axioms need to be verified, then the order axioms, then that's it's an Archimedean field. The last step, that it's complete is almost obvious from the construction. No one of these steps is difficult, but there are many details to check.

There are alternate ways to construct complete ordered fields. One is by means of Cauchy sequences. The spirit is different, but the result is the same, since, as we're about to see, there is only one complete ordered field.

Uniqueness of the complete ordered field We have to somehow exclude the possibility that there are two different Archimedean fields that can't be extended to larger Archimedean fields.

We don't want to count two isomorphic fields as being different, since, in essence, they're the same field but the names of the elements are just different. So, what we want is the following theorem.

Theorem 2.35. Any two complete ordered fields are isomorphic as ordered fields. Furthermore, there is only one isomorphism between them.

Proof. We may treat the field \mathbf{Q} as a subfield of the two complete ordered fields F_1 and F_2 . Then as a Dedekind cut determines an element $a_1 \in F_1$ and an element a_2 in F_2 , we have a bijection $F_1 \rightarrow F_2$. You only need to verify that preserves addition and multiplication, which it does, since in an Archimedean ring, addition and multiplication are determined by Dedekind cuts. Q.E.D.

R is the complete ordered field We now know that there is only one complete ordered field up to isomorphism. Any such complete ordered field may be taken as the real numbers.

2.5 Skew fields (division rings) and the quaternions

Sir William Rowan Hamilton, who early found that his road [to success with vectors] was obstructed—he knew not by what obstacle—so that many points which seemed within his reach were really inaccessible. He had done a considerable amount of good work, obstructed as he was, when, about the year 1843, he perceived clearly the obstruction to his progress in the shape of an old law which, prior to that time, had appeared like a law of common sense. The law in question is known as the *commutative* law of multiplication.

Kelland and Tait, 1873

2.5.1 Skew fields (division rings)

Skew fields, also called division rings, have all the properties of fields except that multiplication need not be commutative. When multiplication is not assumed to be commutative, a couple of the field axioms have to be stated in two forms, a left form and a right form. In particular, we require

1. there is a multiplicative identity, an element of F denoted 1, such that $\forall x, 1x = x = x1$;
2. there are multiplicative inverses of nonzero elements, that is, $\forall x \neq 0, \exists y, xy = 1 = yx$; and
3. multiplication distributes over addition, that is, $\forall x, \forall y, \forall z, x(y + z) = xy + xz$ and $\forall x, \forall y, \forall z, (y + z)x = yx + zx$.

All the other axioms remain the same, except we no longer require commutative multiplication.

The various properties of fields that follow from the field axioms also follow from the skew field axioms, although some have to be stated in two forms.

The most important skew field is the quaternions, mentioned next. Waring showed that there were no finite skew fields that weren't fields (a difficult proof).

2.5.2 The quaternions \mathbf{H}

We're not going to study skew fields, but one is of particular importance, the quaternions, denoted \mathbf{H} . The letter \mathbf{H} is in honor of Hamilton, their inventor.

We can define a quaternion a as an expression

$$a = a_0 + a_1i + a_2j + a_3k$$

where a_0, a_1, a_2 , and a_3 are real numbers and i, j , and k are formal symbols satisfying the properties

$$i^2 = j^2 = k^2 = -1$$

and

$$ij = k, jk = i, ki = j.$$

The i, j , and k are all square roots of -1 , but they don't commute as you can show from the definition that

$$ji = -k, kj = -i, ik = -j.$$

This doesn't lead to a commutative multiplication, but note that if a is real (i.e., its pure quaternion parts a_1, a_2 , and a_3 are all 0), then a will commute with any quaternion b .

Addition and subtraction are coordinatewise just like in \mathbf{C} . Here's multiplication.

$$\begin{aligned} & (a_0 + a_1i + a_2j + a_3k)(b_0 + b_1i + b_2j + b_3k) \\ &= (a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3) \\ &+ (a_0b_1 + a_1b_0 + a_2b_3 - a_3b_2)i \\ &+ (a_0b_2 - a_1b_3 + a_2b_0 + a_3b_1)j \\ &+ (a_0b_3 + a_1b_2 - a_2b_1 - a_3b_0)k \end{aligned}$$

It's easy to check that all the axioms for a noncommutative ring are satisfied. The only thing left to do in order to show that \mathbf{H} is a skew field is that reciprocals exist. We can use a variant of rationalizing the denominator to find the reciprocal of a quaternion.

$$\begin{aligned} \frac{1}{a_0 + a_1i + a_2j + a_3k} &= \frac{a_0 - a_1i - a_2j - a_3k}{(a_0 - a_1i - a_2j - a_3k)(a_0 + a_1i + a_2j + a_3k)} \\ &= \frac{a_0 - a_1i - a_2j - a_3k}{a_0^2 + a_1^2 + a_2^2 + a_3^2} \end{aligned}$$

Thus, a nonzero quaternion $a_0 + a_1i + a_2j + a_3k$, that is, one where not all of the real numbers a_0, a_1, a_2 , and a_3 are 0, has an inverse, since the denominator $a_0^2 + a_1^2 + a_2^2 + a_3^2$ is a nonzero real number.

The expression $a_0 - a_1i - a_2j - a_3k$ used to rationalize the denominator is the *conjugate* of the original quaternion $a_0 + a_1i + a_2j + a_3k$. It's worthwhile to have a notation for it.

$$\overline{a_0 + a_1i + a_2j + a_3k} = a_0 - a_1i - a_2j - a_3k,$$

as we do for \mathbf{C} . We'll also define the *norm* of a quaternion a by $|a|^2 = a\bar{a}$. It's a nonnegative real number, so it has a square root $|a|$. Note that $|a|^2 = a_0^2 + a_1^2 + a_2^2 + a_3^2$.

Thus, if a is a nonzero quaternion, then its inverse is $\frac{1}{a} = \frac{\bar{a}}{|a|^2}$.

For \mathbf{C} , the field of complex numbers, conjugation was a field automorphism, but for \mathbf{H} , it's not quite an automorphism. It has all of the properties of an automorphism except one. It preserves 0, 1, addition and subtraction $\overline{a \pm b} = \bar{a} \pm \bar{b}$, and reciprocation $\overline{1/a} = 1/\bar{a}$, but it reverses the order of multiplication $\overline{ab} = \bar{b} \bar{a}$. We'll call such a thing an *antiautomorphism*.

Note that \mathbf{H} extends \mathbf{C} in many ways. The assignment $x + iy \in \mathbf{C}$ to $x + iy \in \mathbf{H}$ is one, but $x + iy \in \mathbf{C}$ to $x + jy \in \mathbf{H}$ is another. There are, in fact, infinitely many ways that the skew field \mathbf{H} extends the field \mathbf{C} .

Theorem 2.36. The norm of a product is the product of the norms.

Proof. $|ab|^2 = ab\bar{ab} = ab\bar{b}\bar{a} = a|b|^2\bar{a} = a\bar{a}|b|^2 = |a|^2|b|^2$.

Q.E.D.

If we unpack the equation $|a|^2|b|^2 = |ab|^2$, we'll get as a corollary Lagrange's identity on real numbers which shows how to express the product of two sums of four squares as the sum of four squares.

Corollary 2.37 (Lagrange). The product of the sum of four squares of integers is a sum of four squares of integers

$$\begin{aligned} & (a_0^2 + a_1^2 + a_2^2 + a_3^2)(b_0^2 + b_1^2 + b_2^2 + b_3^2) \\ = & (a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3)^2 \\ + & (a_0b_1 + a_1b_0 + a_2b_3 - a_3b_2)^2 \\ + & (a_1b_2 + a_2b_1 + a_3b_1 - a_1b_3)^2 \\ + & (a_2b_3 + a_3b_2 + a_1b_2 - a_2b_1)^2 \end{aligned}$$

Note that this equation not only works for real numbers, but also for integers, indeed when the coefficients lie in any commutative ring. Lagrange used this identity to show that every nonnegative integer n is the sum of four squares. The identity above is used to reduce the general case to the case when n is prime. Lagrange still had work to do to take care of the prime case.

Frobenius's theorem and the octonions. The quaternions are very special in the sense that they're the only finite-dimensional division algebra over \mathbf{R} other than \mathbf{R} itself and \mathbf{C} . This theorem was proved by Frobenius in 1877.

A division algebra over the real numbers R is a division ring (skew field) that has the reals as a subfield. Its dimension is the dimension it has as a vector space over \mathbf{R} .

There is also an eight-dimensional non-associative algebra over the real numbers called the *octonions*, \mathbf{O} . Octonions were discovered by John T. Graves in 1843. Although \mathbf{O} is not associative, it does satisfy weaker associativity identities when two of the three variables are the same: $x(xy) = (xx)y$, $x(yy) = (xy)y$, and $(xy)x = x(yx)$. It also satisfies the Moufang identities: $z(x(zy)) = ((zx)z)y$, $x(z(yz)) = ((xz)y)z$, $(zx)(yz) = (z(xy))z$, and $(zx)(yz) = z((xy)z)$. Furthermore \mathbf{O} has a norm.

Octonions over \mathbf{R} are a special case of a Cayley algebra over a field.

A matrix representation for \mathbf{H} . There are various matrix representations for \mathbf{H} . This one will make \mathbf{H} a subring of the real matrix ring $M_4(\mathbf{R})$. We'll represent 1 by the identity matrix, and i , j , and k by three other matrices which, you can verify, satisfy $i^2 = j^2 = k^2 = -1$ and $ij = k, jk = i, ki = j$.

$$\begin{aligned} 1 &\leftrightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & i &\leftrightarrow \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\ j &\leftrightarrow \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} & k &\leftrightarrow \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Then a generic quaternion $a + bi + cj + dk$ corresponds to the matrix

$$\begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}$$

Quaternions and geometry. Each quaternion a is the sum of a real part a_0 and a pure quaternion part $a_1i + a_2j + a_3k$. Hamilton called the real part a *scalar* and pure quaternion part a *vector*. We can interpret $a_1i + a_2j + a_3k$ as a vector $\mathbf{a} = (a_1, a_2, a_3)$ in \mathbf{R}^3 . Addition and subtraction of pure quaternions then are just ordinary vector addition and subtraction.

Hamilton recognized that the product of two vectors (pure quaternions) had both a vector component and a scalar component (the real part). The vector component of the product \mathbf{ab} of two pure quaternions Hamilton called the *vector product*, now often denoted $\mathbf{a} \times \mathbf{b}$ or $\mathbf{a} \vee \mathbf{b}$, and called the *cross product* or the *outer product*. The negation of the scalar component Hamilton called the *scalar product*, now often denoted $\mathbf{a} \cdot \mathbf{b}$, (\mathbf{a}, \mathbf{b}) , $\langle \mathbf{a}, \mathbf{b} \rangle$, or $\langle \mathbf{a} | \mathbf{b} \rangle$ and called the *dot product* or the *inner product*. Thus

$$\mathbf{ab} = \mathbf{a} \times \mathbf{b} - \mathbf{a} \cdot \mathbf{b}.$$

Hamilton's quaternions were very successful in the 19th century in the study of three-dimensional geometry.

Here's a typical problem from Kelland and Tait's 1873 *Introduction to Quaternions*. If three mutually perpendicular vectors be drawn from a point to a plane, the sum of the reciprocals of the squares of their lengths is independent of their directions.

Matrices were invented later in the 19th century. (But determinants were invented earlier!) Matrix algebra supplanted quaternion algebra in the early 20th century because (1) they described linear transformations, and (2) they weren't restricted to three dimensions.

Exercise 32. Show that \mathbf{H} can be represented as a subring of the complex matrix ring $M_2(\mathbf{C})$ where

$$\begin{aligned} 1 &\leftrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & i &\leftrightarrow \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \\ j &\leftrightarrow \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} & k &\leftrightarrow \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \end{aligned}$$

so that a generic quaternion $a + bi + cj + dk$ corresponds to the matrix

$$\begin{bmatrix} a + bi & c + di \\ -c + di & a - bi \end{bmatrix}$$

Unit quaternions and S^3 as a group. The quaternions $a = a_0 + a_1i + a_2j + a_3k$ with norm 1 are called *unit quaternions*. Examples of unit quaternions are $\pm 1, \pm i, \pm j, \pm k$, but there are many more.

Unit quaternions are the quaternions for which $a_0^2 + a_1^2 + a_2^2 + a_3^2 = 1$. That equation is precisely the equation that defines the unit three sphere S^3 in 4-space \mathbf{R}^4 , although S^3 is usually described with different variables:

$$S^3 = \{(w, x, y, z) \in \mathbf{R}^4 \mid w^2 + x^2 + y^2 + z^2 = 1\}.$$

As we saw above, the product of the norms of two quaternions is the norm of the product, therefore multiplication is closed on this 3-sphere. Furthermore, 1 is a unit quaternion, and the reciprocal of a unit quaternion is another one, and, multiplication is associative, so multiplication of quaternions makes the 3-sphere S^3 into a group.

Chapter 3

Rings

Rings are things like \mathbf{Z} that have the three operations of addition, subtraction, and multiplication, but they don't need division. The lack of a division operation makes them more complicated and more interesting. The concept of *prime*, for example, is uninteresting for fields, but very interesting for \mathbf{Z} and other rings.

Most of our rings will have commutative multiplication, but some won't, so we won't require that multiplication be commutative in our definition. We will require that every ring have 1. The formal definition for rings is very similar to that for fields, but we leave out a couple of the requirements.

In this chapter we'll concentrate mainly on commutative rings and their properties. We'll consider commutative rings with various nice properties. Those rings with nice properties we'll give special names in increasing niceness such as *integral domain*, *unique factorization domain*, *principal ideal domain*, and *Euclidean domain*.

3.1 Introduction to rings

A *ring* is a set equipped with two binary operations, one called *addition* and the other called *multiplication*, denoted in the usual manner, which are both associative, addition is commutative, both have identity elements (the additive identity denoted 0 and the multiplicative identity denoted 1), addition has inverse elements (the inverse of x denoted $-x$), and multiplication distributes over addition. If, furthermore, multiplication is commutative, then the ring is called a *commutative ring*.

3.1.1 Definition and properties of rings

Here is a more complete definition.

Definition 3.1. A *ring* R consists of

1. a set, also denoted R and called the *underlying set* of the ring;
2. a binary operation $+$: $R \times R \rightarrow R$ called *addition*, which maps an ordered pair $(x, y) \in R \times R$ to its *sum* denoted $x + y$;

3. another binary operation $\cdot : R \times R \rightarrow R$ called *multiplication*, which maps an ordered pair $(x, y) \in R \times R$ to its *product* denoted $x \cdot y$, or more simply just xy ;
such that
4. addition is commutative, that is, $\forall x, \forall y, x + y = y + x$;
5. addition is associative, that is, $\forall x, \forall y, (x + y) + z = x + (y + z)$;
6. multiplication is associative, that is, $\forall x, \forall y, (xy)z = x(yz)$;
7. there is an additive identity, an element of F denoted 0 , such that $\forall x, 0 + x = x$;
8. there is a multiplicative identity, an element of F denoted 1 , such that $\forall x, 1x = x$;
9. there are additive inverses, that is, $\forall x, \exists y, x + y = 0$; and
10. multiplication distributes over addition, that is, $\forall x, \forall y, \forall z, x(y + z) = xy + xz$.

When multiplication is also commutative, that is, $\forall x, \forall y, xy = yx$, the ring is called a *commutative ring*. The conditions for a ring are often call the *ring axioms*.

Subtraction, multiples, and powers. As we did with fields, we can define subtraction, integral multiples, and nonnegative integral powers. We won't have division or negative integral powers since we don't have reciprocals.

As before, we define subtraction in terms of negation. The *difference* of two elements x and y is $x - y = x + (-y)$. The expected properties of subtraction all follow from the ring axioms. For instance, multiplication distributes over subtraction.

Likewise, we can define integral multiples of elements in a ring. Define $0x$ as 0 , then inductively define $(n + 1)x = x + nx$ when $n \geq 0$. Then if $-n$ is a negative integer, define $-nx$ as $-(nx)$. The usual properties of multiples, like $(m + n)x = mx + nx$ still hold.

Furthermore, we can define positive integral powers of x . Define x^1 as x for a base case, and inductively, $x^{n+1} = xx^n$. Thus nx is the product of n x 's. For instance, $x^3 = xxx$. Since rings needn't have reciprocals, we can't define negative integral powers of x .

Examples 3.2 (rings). Of course, all fields are automatically rings, but what are some other rings? We've talked about some others already, including

1. the ring of integers \mathbf{Z} which includes all integers (whole numbers)—positive, negative, or 0 .
2. the ring of polynomials $R[x]$ with coefficients in a commutative ring R .
3. the matrix ring $M_n(R)$ of $n \times n$ matrices with entries in a commutative ring R . This example is a noncommutative ring when $n \geq 2$.
4. the ring of upper triangular matrices is a subring of $M_n(R)$.
5. the cyclic ring \mathbf{Z}_n , the ring of integers modulo n , where n is a particular integer.
6. the powerset $\mathcal{P}(S)$ consisting of subsets of a set S becomes a ring, called a Boolean ring, where $A + B$ is the symmetric difference and AB is the intersection of two subsets A and B .

Properties that follow from the ring axioms. There are numerous useful properties that from the axioms, but not so many as follow from the field axioms. Here's a list of several of them.

1. 0 is unique. That is, there is only one element x of a ring that has the property that $\forall y, x + y = y$. Likewise, 1 is unique.
2. Multiplication distributes over subtraction. $x(y - z) = xy - xz$ and $(y - z)x = yx - zx$.
3. $-0 = 0$.
4. $0x = 0$.
5. $(-1)x = -x$, $(-x)y = -(xy) = x(-y)$, and $(-x)(-y) = xy$.

There are some expected properties that are not included here. I'll show why not using examples from \mathbf{Z}_6 .

1. If the product of two elements is 0, $xy = 0$, it does not follow that either $x = 0$ or $y = 0$. For example, in \mathbf{Z}_6 the product of 2 and 3 is 0.
2. Cancellation does not always work. That is, if $xy = xz$ and $x \neq 0$, it doesn't follow that $y = z$. For example, in \mathbf{Z}_6 , $3 \cdot 2 = 3 \cdot 4$, but $2 \neq 4$.

3.1.2 Products of rings

If R_1 and R_2 are two rings, you can construct their product ring R . The underlying set of R is the product $R_1 \times R_2$ of the underlying sets of the two rings, and addition, subtraction, and multiplication are coordinatewise. Thus,

$$(x_1, x_2) \pm (y_1, y_2) = (x_1 \pm y_1, x_2 \pm y_2) \quad \text{and} \quad (x_1, x_2)(y_1, y_2) = (x_1 y_1, x_2 y_2).$$

The additive identity in $R_1 \times R_2$ is $0 = (0, 0)$, and the multiplicative identity is $1 = (1, 1)$. Since all the operations are performed coordinatewise, the ring axioms are satisfied in $R_1 \times R_2$, so it's a ring.

The projection functions $\pi_1 : R_1 \times R_2 \rightarrow R_1$ and $\pi_2 : R_1 \times R_2 \rightarrow R_2$ defined by $\pi_1(x_1, x_2) = x_1$ and $\pi_2(x_1, x_2) = x_2$ are both ring homomorphisms. They preserve addition, multiplication, and 1.

Products of more than 2 rings can be defined analogously, even products of infinitely many rings.

We didn't discuss products of fields in the chapter on field because the product of two fields is not another field. It is at least a ring, however.

3.1.3 Integral domains

Much of the time we will want the cancellation property that was mentioned above to hold, so we'll give a special name to commutative rings that satisfy them. It will help if we make a couple of definitions.

Definition 3.3. A nonzero element x in a commutative ring is a *zero-divisor* if there exists a nonzero y such that $xy = 0$. Of course, 0 is always a zero-divisor. We'll say a commutative ring *has no zero divisors* if 0 is the only zero-divisor.

Definition 3.4. We'll say a commutative ring satisfies the *cancellation law* if

$$\forall x \neq 0, \forall y, \forall z, xy = xz \text{ implies } y = z.$$

We found in the example above that 2 and 3 are zero-divisors in \mathbf{Z}_6 , and that \mathbf{Z}_6 did not satisfy the cancellation law. You can examine \mathbf{Z}_n to determine which nonzero elements are zero-divisors and which have reciprocals.

There's a connection between zero-divisors and the cancellation law.

Theorem 3.5. A commutative ring satisfies the cancellation law if and only if it has no zero-divisors.

Proof. Suppose the ring satisfies the cancellation law. Let x be a nonzero element in the ring. If $xy = 0$, then $xy = x0$, so by that cancellation law, $y = 0$. Then x can't be a zero-divisor. Thus the ring has no zero-divisors.

Next suppose that the ring has no zero-divisors. We'll show it satisfies the cancellation law. If $x \neq 0$ and $xy = xz$, then $x(y - z) = 0$, and since x is not a zero divisor, therefore $y - z = 0$, so $y = z$. Thus the ring satisfies the cancellation law. Q.E.D.

Definition 3.6 (integral domain). An *integral domain* is a commutative ring D in which $0 \neq 1$ that satisfies one of the two equivalent conditions: it has no zero-divisors, or it satisfies the cancellation law.

All the fields and most of the examples of commutative rings we've looked at are integral domains, but \mathbf{Z}_n is not an integral domain if n is not a prime number.

Note that any subring of a field or an integral domain will be an integral domain since the subring still won't have any zero-divisors.

Note that products of (nontrivial) rings are never integral domains since they always have the zero divisors $(1, 0)$ and $(0, 1)$ whose product is 0.

Corollary 2.13 stated that the characteristic of a field is either 0 or a prime number. The proof there works as well for integral domains. The characteristic of an integral domain is either 0 or a prime number.

Group rings You can form a ring $\mathbf{Z}G$ out of a group G as follows. Assume that G is written multiplicatively. The finite formal sums of elements of G are the elements of $\mathbf{Z}G$. Thus, if n is a nonnegative integer and $a_1, \dots, a_n \in G$, then the formal sum $x_1a_1 + \dots + x_na_n$ names an element of the group ring $\mathbf{Z}G$. Addition is coordinatewise. Multiplication uses the group operation.

This definition can be generalized so that group rings have their coordinates in any commutative ring R , not just \mathbf{Z} . This results in a group ring RG .

Exercise 33. Let G be the two element cyclic group $G = \{1, a\}$ where $a^2 = 1$. A typical element of $\mathbf{Z}G$ is $x + ya$ where $x, y \in \mathbf{Z}$. Multiplication is defined by $(x_1 + y_1a)(x_2 + y_2a) = (x_1x_2 + y_1y_2) + (x_1y_2 + x_2y_1)a$. Show that the square of any nonzero element in $\mathbf{Z}G$ is not zero, but show that $\mathbf{Z}G$ does have zero-divisors by finding a pair of nonzero elements whose product is 0.

3.1.4 The Gaussian integers, $\mathbf{Z}[i]$

One important example of an integral domain is that of the Gaussian integers $\mathbf{Z}[i]$. Its elements are of the form $x + yi$ where $x, y \in \mathbf{Z}$, so they can be viewed as a lattice of points in the complex plane as in figure 3.1. You can check that $\mathbf{Z}[i]$ is closed under addition, subtraction, multiplication, and includes 1, so it is a subring of the field \mathbf{C} . Therefore, it's an integral domain. We'll see later that $\mathbf{Z}[i]$ is a particularly nice integral domain called a Euclidean domain.

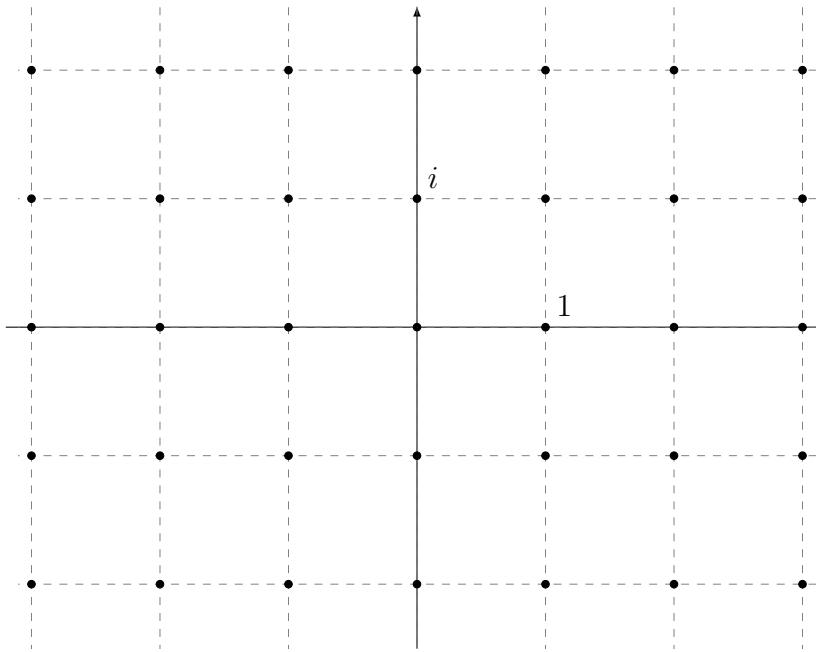


Figure 3.1: Lattice of Gaussian integers $\mathbf{Z}[i]$

There are four units (elements having reciprocals) in the Gaussian integers. Besides 1 and -1 , i and $-i$ are also units. Note that $(1+i)(1-i) = 2$, so 2 is not prime in $\mathbf{Z}[i]$ even though it is prime in \mathbf{Z} .

We'll come back to $\mathbf{Z}[i]$ when we study Euclidean domains in section 3.8.4. Also $\mathbf{Z}[i]$ is an example of a “ring of integers” to be defined in section 3.11.

Eisenstein integers. The Eisenstein integers are similar to the Gaussian integers, but instead of consisting of a square lattices of complex numbers, they consist of a triangular lattice of complex numbers. They include complex numbers of the form $z = x + y\omega$ where ω is the cube root of 1, $\omega = \frac{1}{2}(-1 + i\sqrt{3}) = e^{2\pi i/3}$. See figure 3.3 for the lattice of Eisenstein integers.

3.1.5 Finite fields again

We won't find any examples of finite integral domains that aren't fields because there aren't any.

Theorem 3.7 (Wedderburn). If R is a finite integral domain, then R is a field.

Proof. Let x be a nonzero element of R . Consider the positive powers of x :

$$x, x^2, x^3, \dots, x^n \dots$$

Since there are infinitely many powers, but only finitely many elements in R , therefore at least two distinct powers are equal. Let, then, $x^m = x^n$ with $m < n$. Cancel x^m from each side of the equation (which is possible because R is an integral domain) to conclude $x^{n-m} = 1$. Therefore, the reciprocal of x is x^{n-m-1} . Therefore, every nonzero element has an inverse. Q.E.D.

This theorem can be used to give a short proof that \mathbf{Z}_p is a field when p is a prime, since it's easy to show that \mathbf{Z}_p is an integral domain. We'll show it has no zero-divisors. Suppose that $xy \equiv 0 \pmod{p}$. Then $p|xy$. But if a prime divides a product, it divides one of the factors, so either $p|x$ or $p|y$, in other words, either $x \equiv 0 \pmod{p}$ or $y \equiv 0 \pmod{p}$. Thus, \mathbf{Z}_p is an integral domain, and hence, by the above theorem, it's a field.

Our earlier, more complicated proof used the extended Euclidean algorithm to find an inverse for x . That's actually a much more efficient way to find the inverse than to look through the powers of x .

3.2 Factoring \mathbf{Z}_n by the Chinese remainder theorem

We'll look at the structure of the cyclic ring \mathbf{Z}_n when n is composite in more detail. In particular, when n is not a power of a prime number, then \mathbf{Z}_n is a product of smaller cyclic rings.

3.2.1 The Chinese remainder theorem

This theorem says that if m and k are relatively prime and $n = mk$, then $\mathbf{Z}_n \cong \mathbf{Z}_m \times \mathbf{Z}_k$. Let's illustrate that with $m = 7$ and $k = 12$ to show how $\mathbf{Z}_{84} \cong \mathbf{Z}_7 \times \mathbf{Z}_{12}$. Starting with a number x modulo 84, we'll get a pair of numbers, one being x modulo 7, the other x modulo 12. We can display this in a 7×12 table where each row is a number modulo 7, each column a number modulo 12, and the entry at row i and column j is that number which is i modulo 7 and j modulo 12.

It's easy to construct the table. Start filling the diagonal. After you reach the last row, go next to the top row, and after you reach the right column, go next to the left column.

	0	1	2	3	4	5	6	7	8	9	10	11
0	0	49	14	63	28	77	42	7	56	21	70	35
1	36	1	50	15	64	29	78	43	8	57	22	71
2	72	37	2	51	16	65	30	79	44	9	58	23
3	24	73	38	3	52	17	66	31	80	45	10	59
4	60	25	74	39	4	53	18	67	32	81	46	11
5	12	61	26	75	40	5	54	19	68	33	82	47
6	48	13	62	27	76	41	6	55	20	69	34	83

All the numbers in the first row are congruent to 0 modulo 7, so they're divisible by 7, but looking at them, they seem to be rather randomly arranged. Likewise, all the numbers in the first column are divisible by 12.

The pair of linear congruences $x \equiv i \pmod{7}$ and $x \equiv j \pmod{12}$ can be easily solved for x in by looking in row i and column j .

For example, take this Chinese remainder problem. Find a number such that when you divide it by 7 you get a remainder of 3, but when you divide it by 12 you get a remainder of 8. The answer, 80, is right in the table.

Theorem 3.8 (Chinese remainder theorem). Suppose that $n = km$ where k and m are relatively prime. Then

$$\mathbf{Z}_n \cong \mathbf{Z}_k \times \mathbf{Z}_m.$$

More generally, if n is the product $k_1 \cdots k_r$ where the factors are pairwise relatively prime, then

$$\mathbf{Z}_n \cong \mathbf{Z}_{k_1} \times \cdots \times \mathbf{Z}_{k_r} = \prod_{i=1}^r \mathbf{Z}_{k_i}.$$

In particular, if the prime factorization of n is $n = p_1^{e_1} \cdots p_r^{e_r}$. Then the cyclic ring \mathbf{Z}_n factors as the product of the cyclic rings $\mathbf{Z}_{p_i^{e_i}}$, that is,

$$\mathbf{Z}_n \cong \prod_{i=1}^r \mathbf{Z}_{p_i^{e_i}}.$$

Proof. The third statement is a special case of the second.

The second follows from the first by induction on r .

That leaves us with the first statement. In one direction, $\mathbf{Z}_n \rightarrow \mathbf{Z}_k \times \mathbf{Z}_m$, the function giving the isomorphism is fairly obvious; it's built of the two functions $\mathbf{Z}_n \rightarrow \mathbf{Z}_k$ and $\mathbf{Z}_n \rightarrow \mathbf{Z}_m$ that are easy to describe.

There is an obvious candidate for a ring function $\mathbf{Z}_n \rightarrow \mathbf{Z}_k$, namely $[x]_n \mapsto [x]_k$ by which is meant the equivalence class of x modulo n is sent to the equivalence class of x modulo k .

First, we have to check that this function is well defined. Suppose $[x]_n = [y]_n$. Then $x \equiv y \pmod{n}$, so $n|(x - y)$. But $k|n$, therefore $k|(x - y)$. Hence, $x \equiv y \pmod{k}$, and $[x]_k = [y]_k$. So the function is well-defined.

You can check the rest, that this function preserves the ring operation so that it's a ring homomorphism.

Putting together the two ring homomorphisms $\mathbf{Z}_n \rightarrow \mathbf{Z}_k$ and $\mathbf{Z}_n \rightarrow \mathbf{Z}_m$ we have a ring homomorphism

$$\begin{aligned} \mathbf{Z}_n &\rightarrow \mathbf{Z}_k \times \mathbf{Z}_m \\ [x]_n &\mapsto ([x]_k, [x]_m) \end{aligned}$$

In order to show that this is an isomorphism, all we need to do is to show that it's a bijection, and for that, all we need to do is to show that it's an injection since the sets \mathbf{Z}_n and $\mathbf{Z}_k \times \mathbf{Z}_m$ have the same cardinality.

Suppose that $[x]_n$ and $[y]_n$ are sent to the same element in $\mathbf{Z}_k \times \mathbf{Z}_m$. Then $[x]_k = [y]_k$ and $[x]_m = [y]_m$, that is, $k|(x - y)$ and $m|(x - y)$. Since they both divide $x - y$, so does their least common multiple. But they're relatively prime, so their LCM is their product, n . Thus $n|(x - y)$, so $[x]_n = [y]_n$. Therefore, this is a one-to-one function, hence a one-to-one correspondence. Thus, the ring homomorphism is an isomorphism. Q.E.D.

The inverse. Well, since it's a bijection, it shouldn't be too hard to find its inverse $\mathbf{Z}_k \times \mathbf{Z}_m \rightarrow \mathbf{Z}_n$. In other words, solve for $x \pmod{n}$ the pair of simultaneous congruences

$$\begin{aligned} x &\equiv a \pmod{k} \\ x &\equiv b \pmod{m} \end{aligned}$$

It's too much work to construct the entire $k \times m$ table as was done for the 7×12 . There's a better way.

We can find a solution with the extended Euclidean algorithm. Since $\text{GCD}(m, k) = 1$, therefore 1 is a linear combination of m and k , that is, there are integers s and t so that $sm + tk = 1$. Multiply by $b - a$ to conclude $s(b - a)m + t(b - a)k = b - a$. Therefore, $t(b - a)k + a = b - s(b - a)m$. Let that be x . Then $x \equiv a \pmod{k}$ and $x \equiv b \pmod{m}$ as required.

Problems like this in indeterminate analysis were solved in ancient China and in ancient India. The earliest appeared in *Sunzi suanjing* (*Master Sun's Mathematical Manual*) in the about the fourth century C.E. in China. In 1247 Qin Jiushao gave a general method for solving linear congruences in his *Shushu jiuzhang* (*Mathematical Treatise in Nine Sections*).

3.2.2 Brahmagupta's solution

In India in the seventh century C.E., Brahmagupta also gave a general algorithm for solving these linear congruences in his *Brāhma-sphuṭasiddhānta* (*Correct Astronomical System of Brahma*). If more than two congruences were given, he first reduced the problem to solving pairs of congruences as we did above. His solution is the one described above.

As an example, find $x \pmod{210}$ if

$$\begin{aligned} x &\equiv 11 \pmod{45} \\ x &\equiv 4 \pmod{56} \end{aligned}$$

Here's how he did it in modern notation, explained with the numerical example above.

We're looking for a value of x so that $x = 45s + 11 = 56t + 4$ for some integers s and t . So we need s and t so that $45s + 7 = 56t$. That reduces to $45(s - t) + 7 = 11t$. Let $s' = s - t$. To solve $45s' + 7 = 11t$, since $45 = 4 \cdot 11 + 1$, reduce it to $s' + 7 = 11(t - 4s')$. Let $t' = t - 4s'$. We can solve $s' + 7 = 11t'$ by setting $s' = 4$ and $t' = 1$. Substituting these in the defining equations, we find $t = t' + 4s' = 17$, and $s = s' + t = 21$. Therefore, $x = 45s + 11 = 956$, the answer.

Of course, Brahmagupta did not use variables. His solution was described as a fairly simple algorithm that just used the four arithmetic operations.

3.2.3 Qin Jiushao's solution

The algorithm that Qin Jiushao described was fairly different and applied directly to many linear congruences so long as the moduli were pairwise relatively prime. Let's illustrate it with the system of three congruences

$$\begin{aligned} x &\equiv 45 \pmod{121} \\ x &\equiv 31 \pmod{63} \\ x &\equiv 30 \pmod{100} \end{aligned}$$

Since the moduli are pairwise relatively prime, we can find a unique solution to this system modulo 762300, the product of the moduli.

Step 1. For each modulus, find a reciprocal of the product of the remaining moduli modulo the given modulus. For the first modulus, 121, that means we need the reciprocal of 6300 modulo 121, that is, we need to solve

$$6300y \equiv 1 \pmod{121}.$$

That's the same as $8y \equiv 1 \pmod{121}$. The extended Euclidean algorithm gives us $1 = (-15) \cdot 8 + 1 \cdot 121$, so $y = -15$ is a solution.

For the second modulus, 63, we need the reciprocal of 12100 modulo 63. That's the same as the reciprocal of 4 modulo 63, which is 16.

For the third modulus, 100, we need the reciprocal of 7623 modulo 100. That's the same as the reciprocal of 23 modulo 100. The Chinese mathematicians called finding a reciprocal modulo n by the term “finding one”. By the extended Euclidean algorithm, $(-13) \cdot 23 + 3 \cdot 8 = 1$, so -13 is the reciprocal of 23 modulo 100.

Step 2. To get x sum three products abc , one for each congruence, where a is the constant in the congruence, b is the product of the other moduli, and c is the reciprocal found in the previous step. That gives us

$$\begin{aligned} & 45 \cdot 6300 \cdot (-15) \\ & + 31 \cdot 12100 \cdot 16 \\ & + 30 \cdot 7623 \cdot (-13) \\ & = -283515 + 6001600 - 2972970 = 2745115 \end{aligned}$$

and then reduce this number modulo the product 762300 of all three moduli. That gives a final answer of $x \equiv 458215 \pmod{762300}$.

Exercise 34. Solve the following system of simultaneous linear congruences. You can use either Brahmagupta's algorithm, Qin Jiushao's algorithm, or something of your own devising.

$$x \equiv 4 \pmod{33}$$

$$x \equiv 22 \pmod{35}$$

$$x \equiv 41 \pmod{53}$$

Be sure to show how you derived the solution.

3.3 Boolean rings

Representing by x the class “men,” and by y “Asiatics,” let z represent the adjective “white” to the collection of men expressed by the phrase “Men except Asiatics,” is the same as to say “White men except white Asiatics.” Hence we have

$$z(x - y) + zx - zy.$$

This is also in accordance with the laws of ordinary algebra.

George Boole, 1854. *An Investigation of the Laws of Thought on which are founded the mathematical theories of logic and probabilities.*

George Boole (1815-1864). Boole wanted to bring logic into the realm of mathematics, which he did by algebrizing it.

We'll incorporate his investigations in our study of ring theory, but change his notation slightly. Boole did not allow a sum of two things unless they were disjoint, so $x + x$ had no meaning for him. We'll just take $+$ to be an exclusive or (symmetric difference), so $x + x$ will be 0 for us.

3.3.1 Introduction to Boolean rings

We saw before that powerset $\mathcal{P}(S)$ of a set S becomes a ring when we define $A + B$ to be the symmetric difference and AB to be the intersection of two subsets A and B . The 0 element of the ring is the emptyset \emptyset , while the 1 element is S . The complement of a subset A is $1 - A$ (which equals $1 + A$).

We'll define what a Boolean ring is in terms of idempotents.

Definition 3.9. An element e of a ring is said to be *idempotent* when $e^2 = e$.

Notice that 0 and 1 are always idempotent in any ring.

Other examples of idempotent elements in rings are projections. The transformation $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ which projects a point in the plane to the x -axis, defined by $f(x, y) = (x, 0)$, is idempotent as is any projection from a space to a subspace of itself.

Definition 3.10. A *Boolean ring* is a ring in which every element is idempotent.

The ring $\mathcal{P}(S)$ is evidently an example of a Boolean ring.

Two properties that follow from the definition are (1) that the a Boolean ring has characteristic 2, (2) Boolean rings are commutative.

Theorem 3.11. A nontrivial Boolean ring has characteristic 2.

Proof. Since $1 + 1$ is idempotent, $(1 + 1)^2 = 1 + 1$. Therefore, $1 + 1 + 1 + 1 = 1 + 1$, and so $1 + 1 = 0$. Q.E.D.

As in any ring of characteristic 2, negation does nothing, $-x = x$, and subtraction is the same as addition, $x - y = x + y$.

Theorem 3.12. Boolean rings are commutative.

Proof. Let x and y be two elements of a Boolean ring. Since $x + y$ is idempotent, $(x + y)^2 = x + y$. Expanding that equation using the fact that multiplication distributes over addition in every ring, commutative or not, it follows that $x^2 + xy + yx + y^2 = x + y$. But $x^2 = x$ and $y^2 = y$, so that last equation simplifies to $xy + yx = 0$. Therefore, $xy = -yx$, and $-yx = yx$, so $xy = yx$. Q.E.D.

Boolean rings are the same thing as something called Boolean algebras, but the approaches are different. A Boolean ring is thought of as a special kind of ring, while a Boolean algebra is a special kind of partially ordered set whose elements are truth values. Boolean algebras are reviewed in the appendix section [A.3.3](#).

Table 3.1 compares common notations in Boolean algebras, set theory, and Boolean rings. Here, P and Q are propositions or predicates, A and B are subsets of a set Ω , and x and y

Boolean algebras	Set theory	Boolean rings
T (true)	Ω	1
F (false)	\emptyset	0
$P \wedge Q$ (and)	$A \cap B$	xy
$P \vee Q$ (inclusive or)	$A \cup B$	$x + y + xy$
$P \oplus Q$ (exclusive or)	$A \oplus B$	$x + y$
$\neg P$ (not)	A^c	$1 + x$
$P \iff Q$	$A = B$	$x = y$
$P \Rightarrow Q$	$A \subseteq B$	$xy = x$
$T \vee Q \iff T$	$\Omega \cup B = \Omega$	$1 + y + 1y = 1$
$F \vee Q \iff Q$	$\emptyset \cup B = B$	$0 + y + 0y = y$
$T \wedge Q \iff Q$	$\Omega \cap B = B$	$1y = y$
$F \wedge Q \iff F$	$\emptyset \cap B = \emptyset$	$0y = 0$
$P \wedge Q \iff Q \wedge P$	$P \cap Q = Q \cap P$	$xy = yx$
$P \vee Q \iff Q \vee P$	$P \cup Q = Q \cup P$	$x + y + xy = y + x + yx$
$\neg(P \wedge Q) \iff \neg P \vee \neg Q$	$(A \cap B)^c = A^c \cup B^c$	$1 + xy = (1+x) + (1+y) + (1+x)(1+y)$
$\neg(P \vee Q) \iff \neg P \wedge \neg Q$	$(A \cup B)^c = A^c \cap B^c$	$1 + (x + y + xy) = (1 + x)(1 + y)$

Table 3.1: Notations in Boolean algebras, set theory, and Boolean rings.

are elements of a Boolean ring. These are just a few correspondences. You can add many more.

For more on Boolean algebras, see section [A.3.3](#) in the appendix.

Free Boolean rings Some Boolean rings, called free Boolean rings, have special properties. Given a set S whose elements are calle *generators*, the *free Boolean ring* on S is the Boolean ring $B(S)$ which comes equipped with a function $\iota : S \rightarrow B(S)$ that satisfies the following universal property: for each Boolean ring R and function $f : S \rightarrow R$, there exists a unique ring homomorphism $\hat{f} : B(S) \rightarrow R$ such that $f \circ \iota = \hat{f}$.

Examples 3.13 (Free Boolean rings). If $S = \emptyset$ is the empty set, then $B(\emptyset)$ consists of only two elements, 0 and 1, which when identified with truth values are \perp and \top , respectively.

If $S = \{p\}$ is a singleton set, then $B(\{p\})$ has four elements, 0, p , $1 - p$, and 1, which when identified with truth values are \perp , p , $\neg p$, and \top .

If $S = \{p, q\}$ has two elements, then $B(\{p, q\})$ has 16 elements. They are displayed in figure [3.2](#) as truth values. It is isomorphic to the Boolean ring which is the powerset of a set of four elements in figure [A.1](#).

3.3.2 Factoring Boolean rings

Suppose that a set S is partitioned into subsets S_1, S_2, \dots, S_n . That means S is the union of all these subsets, and they are pairwise disjoint. Then the ring $\wp(S)$ is isomorphic to a product of the rings $\wp(S_i)$. The function

$$\begin{aligned}\wp(S) &\cong \wp(S_1) \times \wp(S_2) \times \cdots \times \wp(S_n) \\ A &\mapsto (A \cap S_1, A \cap S_2, \dots, A \cap S_n)\end{aligned}$$

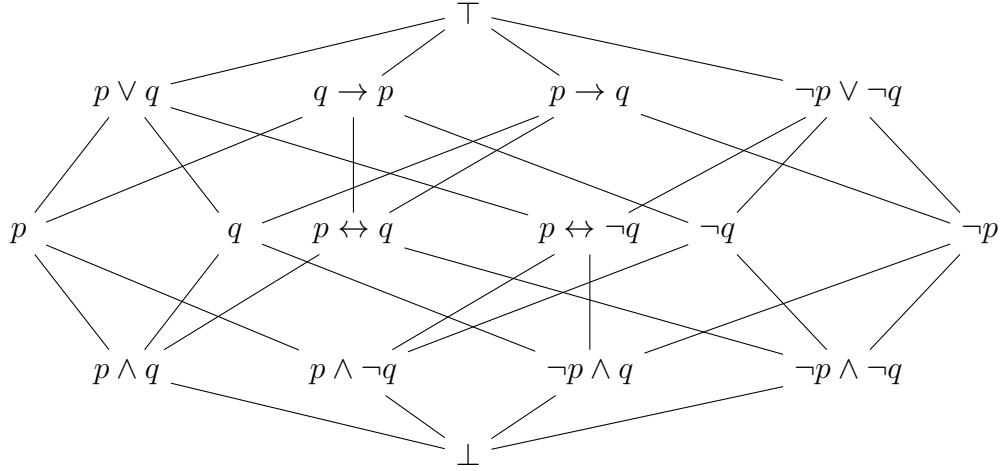


Figure 3.2: Free Boolean ring on two elements

gives the ring homomorphism in one direction, and it's a bijection since A is the disjoint union of the terms on the right.

In fact, this works even when S is partitioned into arbitrarily many subsets. Since S is the disjoint union of its singletons $S = \bigcup_{x \in S} \{x\}$, therefore $\wp = \prod_{x \in S} \wp(\{x\})$. In other words, \wp is a power of the 2-element ring.

Factoring works in a general Boolean ring as well as those of the form $\wp(S)$. Let R be a Boolean ring, and e any idempotent in it other than 0 or 1. Let $\bar{e} = 1 - e$, so that $1 = e + \bar{e}$ from which it follows that $x = xe + x\bar{e}$ for all $x \in R$. Let $R_e = \{xe \mid x \in R\}$, and let $R_{\bar{e}} = \{x\bar{e} \mid x \in R\}$. You can check that both R_e and $R_{\bar{e}}$ are Boolean rings, where the multiplicative identities are e and \bar{e} , respectively. Furthermore,

$$\begin{aligned} R &\cong R_e \times R_{\bar{e}} \\ x &\mapsto (xe, x\bar{e}) \end{aligned}$$

3.3.3 A partial order on a Boolean ring

If we define $x \preceq y$ to mean $xy = y$, then our Boolean ring will have a partial ordering.

Recall that a partial ordering \preceq on a set is a reflexive, antisymmetric, and transitive relation.

1. Reflexive: $x \preceq x$, since $x^2 = x$.
2. Antisymmetric: $x \preceq y$ and $y \preceq x$ imply $x = y$, since $xy = x$ and $yx = y$ imply $x = y$.
3. Transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$, since $xy = x$ and $yz = y$ imply $xz = x$.
(*Proof:* $xz = (xy)z = x(yz) = xy = x$.)

In this partial order, the product xy is the *meet* $x \wedge y$ of x and y , that is, it's the largest element z such that $z \preceq x$ and $z \preceq y$. Likewise, $x + y + xy$ is the *join* $x \vee y$ of x and y , that is, it's the smallest element z such that $x \preceq z$ and $y \preceq z$. A partial order that has meets and

joins of pairs of elements is called a *lattice*. Not all lattices have the distributive properties where meet and join distribute over each other

$$(x \vee y) \wedge z = (x \wedge z) \vee (y \wedge z) \text{ and } (x \wedge y) \vee z = (x \vee z) \wedge (y \vee z)$$

but Boolean rings do, so Boolean rings are examples of *distributive lattices*

A *minimal* element of a Boolean ring is a nonzero element such that there is no smaller nonzero element. Every element of a finite Boolean ring is a sum of the minimal elements less than or equal to it. Since there are no elements less than 0, 0 has to be treated as the empty sum.

Theorem 3.14. If R is a finite Boolean ring, then $R \cong \mathcal{O}(S)$ where

$$S = \{x \in R \mid x \text{ is minimal}\}.$$

Exercise 35. Prove the preceding theorem. Hint: see the section above 3.3.2 on factoring Boolean rings. Induction may help.

3.4 The field of rational numbers, fields of fractions

Suppose that we already have constructed the integral domain of integers \mathbf{Z} , but for some reason do not have the field of rational numbers \mathbf{Q} . Then we could construct \mathbf{Q} from \mathbf{Z} since each rational number can be named by a pair of integers. We'll do that. The steps we use only depend on \mathbf{Z} being an integral domain. That means that the construction we use can also be used to create a *field of fractions* F from any integral domain R . In the following, think of the integral domain R as being \mathbf{Z} and the field F as being \mathbf{Q} .

An equivalence relation on pairs of integers. First of all, a rational number $\frac{m}{n}$ can be named by a pair of integers (m, n) where the second integer n does not equal 0. But different pairs (m, n) and (k, l) can name the same integer $\frac{m}{n} = \frac{k}{l}$ if $ml = nk$. That suggests if we want to create rational numbers from integers, we'll need an equivalence relation on pairs of elements of the integral domain R .

We'll start with the set $R \times R_{\neq 0}$ of ordered pairs (m, n) of elements of an integral domain R with $n \neq 0$. Define a relation \equiv on this set by

$$(m, n) \equiv (k, l) \quad \text{iff} \quad ml = nk.$$

You can easily verify that this relation is an equivalence relation.

Reflexivity: $(m, n) \equiv (m, n)$. That's valid since $mn = mn$.

Symmetry: $(m, n) \equiv (k, l)$ implies $(k, l) \equiv (m, n)$. That's valid since $ml = nk$ implies $kn = lm$.

Transitivity: $(m, n) \equiv (k, l)$ and $(k, l) \equiv (s, t)$ imply $(m, n) \equiv (s, t)$. We need to show that $ml = nk$ and $kt = ls$ imply $mt = ns$. Multiply the first equation by t and the second by n . Then $mlt = nkt$ and $nkt = nls$, so $mlt = nls$. But R is an integral domain, so cancellation is valid when $l \neq 0$, so $mt = ns$.

Thus, \equiv is an equivalence relation on $R \times R_{\neq 0}$. Let F be the quotient set F_{\equiv} , and denote an element $[(m, n)]$ of F by $\frac{m}{n}$.

So far, we've got the underlying set for our proposed field F , but we don't have the operations for a field. Before we define them (and show they're well-defined), let's verify that the function $R \rightarrow R \times R_{\neq 0} \rightarrow F$ which sends an element m of R first to $(m, 1)$ then to $\frac{m}{1}$ is a one-to-one function. Suppose that $\frac{m}{n} = \frac{n}{l}$. That means $m1 = 1n$, so $m = n$. Thus we may interpret $R \rightarrow F$ as making R a subset of F by identifying m with $\frac{m}{1}$.

Addition on F . We'd like to define the sum

$$\frac{m}{n} + \frac{k}{l} \quad \text{as} \quad \frac{ml + nk}{nl},$$

but as our fractions are really equivalence classes, we need to show that's well defined. In detail, we need to show that

$$\frac{m}{n} = \frac{m'}{n'} \quad \text{and} \quad \frac{k}{l} = \frac{k'}{l'} \quad \text{imply} \quad \frac{ml + nk}{nl} = \frac{m'l' + n'k'}{n'l'}.$$

That reduces to showing that

$$mn' = nm' \quad \text{and} \quad kl' = lk' \quad \text{imply} \quad (ml + nk)n'l' = nl(m'l' + n'k').$$

But that can be shown by multiplying the first equation by ll' , the second by nn' and adding the two resulting equations. Thus, this addition on F is well-defined.

Multiplication on F . We'd like to define the product

$$\frac{m}{n} \frac{k}{l} \quad \text{as} \quad \frac{mk}{nl},$$

We need to show that's well defined. You'll find that the proof is easier than the one above for addition.

Next, we need to verify that with these definitions F satisfies the field axioms. A proof is needed for each field axiom.

Commutativity of addition. $\frac{m}{n} + \frac{k}{l} = \frac{k}{l} + \frac{m}{n}$. That's easily verified since $\frac{ml + nk}{nl} = \frac{kn + lm}{ln}$. (That depends on commutativity of addition and multiplication in R .)

Commutativity of multiplication. $\frac{m}{n} \frac{k}{l} = \frac{k}{l} \frac{m}{n}$. That's easily verified since $\frac{mk}{nl} = \frac{km}{ln}$.

Associativity of addition. You can easily show it, but it's a big mess.

Associativity of multiplication. Pretty easy.

Additive identity. $\frac{0}{1} + \frac{k}{l} = \frac{k}{l}$. Easy.

Multiplicative identity $\frac{1}{1} \frac{k}{l} = \frac{k}{l}$. Easy.

Negation. $\frac{m}{n} + \frac{-m}{n} = \frac{0}{1}$. Pretty easy.

Reciprocation. For $\frac{m}{n} \neq \frac{0}{1}$, $\frac{m}{n} \cdot \frac{n}{m} = \frac{1}{1}$. Pretty easy.

Multiplication distributes over addition. Easy but messy.

$0 \neq 1$. We need to show that $\frac{0}{1} \neq \frac{1}{1}$ in F . But that's the same as $0 \cdot 1 \neq 1 \cdot 1$ in the integral domain R , and part of the definition of integral domain requires $0 \neq 1$.

Thus, F is a field.

Exercise 36. Select four of the axioms above, and prove them. As always, your proofs should include justifications.

We'll summarize this result as a theorem.

Theorem 3.15. An integral domain R is a subring of a field F , called the *field of fractions*, where each element of F can be represented as $\frac{m}{n}$ where m and n are elements of R and $n \neq 0$.

This gives us another proof that the characteristic of an integral domain is either 0 or a prime number since it has the same characteristic of its field of fractions.

Examples 3.16. The primary example of this is the construction of \mathbf{Q} from \mathbf{Z} .

For another example, take the Gaussian integers $\mathbf{Z}[i]$ for the integral domain R . Then the field of fractions F is the field $\mathbf{Q}(i)$. The elements of $\mathbf{Q}(i)$ are of the form $x + yi$ where x and y are rational numbers.

Yet for another example, take the polynomial ring $F[x]$ with coefficients in a field F . It's an integral domain, and its field of fractions is the rational function field $F(x)$ with coefficients in F .

Stopping short of inverting all elements. Sometimes you may want to create reciprocals for some elements of an integral domain, but not for all elements. This can be done by a minor modification of the above process. Suppose, for instance, that you want to extend \mathbf{Z} to include the reciprocal of 2 but not of any other prime number. That would lead to the *domain of dyadic rationals* $\mathbf{Z}[\frac{1}{2}]$ where the denominators are powers of 2.

On the other hand, if you want to extend \mathbf{Z} to include the reciprocals of all the primes except 2, just include odd denominators. This is called localizing \mathbf{Z} at 2.

These other constructions are useful in mathematics, but we won't use them ourselves.

3.5 Categories and the category of rings

Categories are higher-order algebraic structures. We'll look at the category of rings in which the objects of the category are all the rings. The purpose of a category is to study the interrelations of its objects, and to do that the category includes morphisms between the objects. In the case of the category of rings, the morphisms are the ring homomorphisms.

We'll start with the formal definition of categories. We'll use the category of rings both to illustrate categorical concepts and to study rings. Category theory was developed by Eilenberg and Mac Lane in the 1940s.

3.5.1 The formal definition of categories

Unlike fields, rings, and groups, we won't require that categories build on sets. In a category the collection of all its objects won't be a set because the collection is larger than any set. That's not a problem since theories don't have to be built on set theory. Indeed, set theory itself is not built on set theory.

Definition 3.17. A *category* \mathcal{C} consists of

1. *objects* often denoted with uppercase letters, and
2. *morphisms* (also called *maps* or *arrows*) often denoted with lowercase letters.
3. Each morphism f has a *domain* which is an object and a codomain which is also an object. If the domain of f is A and the codomain is B , then we write $f : A \rightarrow B$ or $A \xrightarrow{f} B$. The collection of all morphisms from A to B is denoted $\text{Hom}(A, B)$.
4. For each object A there is a morphism $1_A : A \rightarrow A$ called the *identity morphism* on A . (When A can be determined by context, its denoted simply 1 .)
5. Given two morphisms $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ where the codomain of one is the same as the domain of the other there is another morphism $A \xrightarrow{g \circ f} C$ called the *composition* of the two morphisms. This composition is illustrated by the commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ & \searrow g \circ f & \downarrow g \\ & & C \end{array}$$

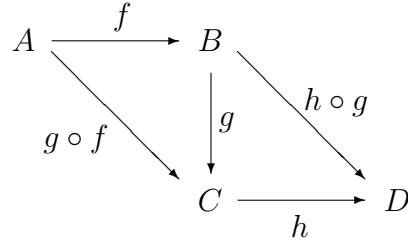
(Sometimes $g \circ f$ is denoted fg .)

A diagram of objects and morphisms in a category is said to *commute*, or be a *commutative diagram* if any two paths of morphisms (in the direction of the arrows) between any two objects yield equal compositions.

6. For all $A \xrightarrow{f} B$, $f \circ 1_A = f$ and $1_B \circ f = f$. These compositions are illustrated by the two commutative diagrams

$$\begin{array}{ccc} & & \\ A & \xrightarrow{f \circ 1_A} & B \\ \downarrow 1_A & \searrow & \\ A & \xrightarrow{f} & B \end{array} \quad \begin{array}{ccc} & & \\ A & \xrightarrow{f} & B \\ \downarrow 1_B & \searrow & \\ A & \xrightarrow{1_B \circ f} & B \end{array}$$

7. For all $A \xrightarrow{f} B$, $B \xrightarrow{g} C$, and $C \xrightarrow{h} D$, $(h \circ g) \circ f = h \circ (g \circ f)$. In the diagram below, if the two triangles in the diagram each commute, then the parallelogram commutes.



Isomorphisms in a category \mathcal{C} . Although only morphisms are defined in a category, it's easy to determine which ones are isomorphisms. A morphism $f : A \rightarrow B$ is an *isomorphism* if there exists another morphism $g : B \rightarrow A$, called its *inverse*, such that $f \circ g = 1_A$ and $g \circ f = 1_B$. Indeed, the main reason identity morphisms are included in the definition of categories is to be able to define isomorphisms.

Examples 3.18 (The categories of sets, groups, rings, and fields). Although we're more interested in the category of rings right now, the category \mathcal{S} of sets is also relevant. An object in \mathcal{S} is a set, and a morphism in \mathcal{S} is a function. The domain and codomain of a morphism are just the domain and codomain of the function, and composition is composition. Isomorphisms are bijections.

The objects of the category \mathcal{G} of groups are groups, and the morphisms of \mathcal{G} are group homomorphisms.

The objects of the category \mathcal{R} of rings are rings, and the morphisms of \mathcal{R} are ring homomorphisms.

The objects of the category of fields are fields, and its morphisms are field homomorphisms, which are just ring homomorphisms. The category of fields is a subcategory of the category of rings.

In each of these other three categories—groups, rings, fields—isomorphisms in the category are what we have called isomorphisms.

3.5.2 The category \mathcal{R} of rings

Recall that a ring homomorphism $f : A \rightarrow B$ between rings is a function that preserves addition, multiplication, and 1. The category of rings has as its objects all rings and as its morphisms all ring homomorphisms. The identity morphism 1_A on a ring is the identity homomorphism, and composition is the usual composition of homomorphisms. Thus, we have a category \mathcal{R} of rings.

If this were all there was to category theory, there wouldn't be much point to it. But by emphasizing the morphisms and deemphasizing elements in rings we can identify what's important about certain rings and certain ring constructions. We'll look at products of rings first to see what characterizes them. We'll also look at a couple of special rings, namely \mathbf{Z} and $\mathbf{Z}[x]$, for characterizing properties of them. We'll also see how to characterize monomorphisms.

The universal property of products. Recall that the product $R_1 \times R_2$ of two rings is consists of ordered pairs (x_1, x_2) with $x_1 \in R_1$ and $x_2 \in R_2$, and the ring operations for $R_1 \times R_2$ are performed coordinatewise. Furthermore, we have the projection ring homomorphisms $R_1 \times R_2 \xrightarrow{\pi_1} R_1$ and $R_1 \times R_2 \xrightarrow{\pi_2} R_2$ which pick out the two coordinates.

This product has the universal property that for each ring S and ring homomorphisms $S \xrightarrow{f_1} R_1$ and $S \xrightarrow{f_2} R_2$, there exists a unique ring homomorphism $S \rightarrow R_1 \times R_2$, which we will denote (f_1, f_2) , such that $f_1 = \pi_1 \circ (f_1, f_2)$ and $f_2 = \pi_2 \circ (f_1, f_2)$, as illustrated by the diagram below.

$$\begin{array}{ccccc}
 & & R_1 & & \\
 & f_1 \swarrow & & \searrow \pi_1 & \\
 S & \xrightarrow{(f_1, f_2)} & R_1 \times R_2 & \xrightarrow{\pi_2} & R_2 \\
 & \searrow & & \swarrow f_2 & \\
 & & R_2 & &
 \end{array}$$

In fact, the product is characterized by this universal property in the sense that if another ring R has this universal property, then there is a ring isomorphism $R \rightarrow R_1 \times R_2$. In more detail, if $R \xrightarrow{p_1} R_1$ and $R \xrightarrow{p_2} R_2$ have this product property (namely, that for each ring S and ring homomorphisms $S \xrightarrow{f_1} R_1$ and $S \xrightarrow{f_2} R_2$, there exists a unique ring homomorphism $S \xrightarrow{f} R$ such that $f_1 = p_1 \circ f$ and $f_2 = p_2 \circ f$), then there exists a unique ring isomorphism $R \xrightarrow{h} R_1 \times R_2$ such that $\pi_1 \circ h = p_1$ and $\pi_2 \circ h = p_2$.

Although this characterization of products was described for the category of rings, it is the definition for the product of two objects in any category. A product $R_1 \times R_2$ is characterized by the property that a morphism to the product correspond to a pair of morphisms to the factors. The product of two sets in the category \mathcal{S} of sets has this same universal property as does the product of two groups in the category \mathcal{G} of groups. There are, however, no products in the category of fields.

Z is the initial object in the category of rings. We can also use category theory to pin down what's so special about the ring **Z**. It has the property that given any ring R , there is a unique ring homomorphism $\mathbf{Z} \xrightarrow{f} R$, and it's defined by $f(n) = n$. An object in a category with that property is called the *initial object* in the category. Any two initial objects in a category are isomorphic.

The trivial ring is the final object in the category of rings. Dual to the initial object is a final object, which in the category of rings is the *trivial* or *degenerate* ring 0. This ring has only one element in which $0 = 1$. In fact, it's the only ring in which $0 = 1$ (since $0 = 0x = 1x = x$).

The *final object* in a category has the property that there's a unique morphism to it from each object in the category. The trivial ring has that property in the category of rings.

Exercise 37. Determine the initial object and the final object in the category \mathcal{S} of sets.

The universal property of the polynomial ring $\mathbf{Z}[x]$. Given any ring R and any element $a \in R$, there is a unique ring homomorphism $\mathbf{Z}[x] \rightarrow R$ that maps x to a . This homomorphism is just evaluation at a , and a polynomial $f(x)$ is mapped to the element $f(a)$ in R .

3.5.3 Monomorphisms and epimorphisms in a category

Although we defined a monomorphism $f : A \rightarrow B$ as a one-to-one homomorphism, we can characterize monomorphisms entirely in terms of category theory.

Definition 3.19. A morphism $f : A \rightarrow B$ is *monic*, or a *monomorphism*, when if g and h are any two morphisms from any another object C to A such that $f \circ g = f \circ h$, then $g = h$.

$$\begin{array}{ccc} C & \xrightarrow{\quad g \quad} & A & \xrightarrow{\quad f \quad} & B \\ & \xrightarrow{\quad h \quad} & & & \end{array}$$

A monomorphism in the category \mathcal{S} of sets is an injection.

This definition agrees with our previous definition for ring monomorphism in terms of elements, and one way to see the correspondence is to let C be $\mathbf{Z}[x]$. Likewise, a monomorphism in the category \mathcal{G} of groups agrees with our previous definition of group monomorphism.

Epimorphisms. The concept of epimorphism is dual to that of monomorphism. If we change the direction of all the arrows in the definition of monomorphism, we'll get the definition of epimorphism.

Definition 3.20. A morphism $f : A \rightarrow B$ is *epic*, or an *epimorphism*, when if g and h are any two morphisms from B to any another object C such that $g \circ f = h \circ f$, then $g = h$.

$$\begin{array}{ccc} A & \xrightarrow{\quad f \quad} & B & \xrightarrow{\quad g \quad} & C \\ & & \xrightarrow{\quad h \quad} & & \end{array}$$

In the category \mathcal{S} of sets, an epimorphism is a surjection. Likewise, it turns out that in the category \mathcal{G} of groups, an epimorphism is a surjection.

In the category \mathcal{R} of rings, it's easy enough to show that if f is a surjective ring homomorphism, then f is an epimorphism, but there are other epimorphisms that aren't surjections.

Example 3.21. Consider the inclusion function $\iota : \mathbf{Z} \rightarrow \mathbf{Q}$. We'll show that it's an epimorphism.

Let g and h be any two morphisms from \mathbf{Q} to any another ring C such that $g \circ \iota = h \circ \iota$. Then $g(n) = h(n)$ for any integer n . Let $\frac{m}{n}$ be a rational number with $n \neq 0$. Then $g(m) = h(m)$ and $g(n) = h(n)$. So,

$$g\left(\frac{m}{n}\right)g(n) = g\left(\frac{m}{n}n\right) = g(m) = h(m) = h\left(\frac{m}{n}n\right) = h\left(\frac{m}{n}\right)h(n) = h\left(\frac{m}{n}\right)g(n).$$

Since $n \neq 0$, therefore $g(n) \neq 0$ as well. Cancel the $g(n)$ at the ends of the continued equation to conclude $g\left(\frac{m}{n}\right) = h\left(\frac{m}{n}\right)$. Thus, $g = h$.

Therefore, the ring homomorphism $\iota : \mathbf{Z} \rightarrow \mathbf{Q}$ is an epimorphism in \mathcal{R} , the category of rings. It is also a monomorphism. But it is not an isomorphism.

In many categories, if a morphism is both monic and epic, then it's also an isomorphism. That's true in the category \mathcal{S} of sets and in the category \mathcal{G} of groups, but not in the category \mathcal{R} of rings. This example shows that \mathcal{R} is a somewhat unusual category.

3.6 Kernels, ideals, and quotient rings

These three concepts are closely related. For a ring homomorphism $f : R \rightarrow S$, the inverse image of 0 is a subset of R called the kernel of f and denoted $\text{Ker } f$. It can't be just any subset, as we'll see, since it's closed under addition and multiplication by elements of R . A subset with those properties we'll call an ideal of R . Every ideal I of R is the kernel of some ring homomorphism $f : R \rightarrow S$. We'll use an ideal I of a ring R to define a quotient ring R/I and a projection $\gamma : R \rightarrow R/I$. These projections will be generalizations of the projections $\mathbf{Z} \rightarrow \mathbf{Z}_n$ that we studied earlier.

3.6.1 Kernels of ring homomorphisms

Definition 3.22. Let $f : R \rightarrow S$ be a ring homomorphism. Those elements of R that are sent to 0 in S form the *kernel* of f .

$$\text{Ker } f = f^{-1}(0) = \{x \in R \mid f(x) = 0\}.$$

We'll look at properties of this kernel and see what it tells us about the function f .

Example 3.23. It's a good idea to have in mind an example or two whenever a new concept is defined. The definition of the kernel of a ring homomorphism is given above, and a good example for it is the ring homomorphism $f : \mathbf{Z} \rightarrow \mathbf{Z}_n$ where n is a fixed integer. That's an especially good example we can use it throughout this discussion of rings, ideals, and quotient rings.

For that $f : \mathbf{Z} \rightarrow \mathbf{Z}_n$, an element $x \in \mathbf{Z}$ is in $\text{Ker } f$ if it is sent to $[0]_n$, the 0 element in the ring \mathbf{Z}_n , that is, if $[x]_n = [0]_n$, or, more simply, if $n|x$. Therefore, the kernel of f consists of the multiples of n . A standard notation for the multiples of an integer n is $n\mathbf{Z}$. Thus, for this function f , $\text{Ker } f = n\mathbf{Z}$.

Kernels aren't just any subsets of R ; they have some special properties. We have, of course, $0 \in \text{Ker } f$, since $f(0) = 0$. Also, if x and y are both in $\text{Ker } f$, then $f(x + y) = f(x) + f(y) = 0 + 0 = 0$, so their sum $x + y$ is also in $\text{Ker } f$. Furthermore, if $x \in \text{Ker } f$ and y is any element of R , then $f(xy) = f(x)f(y) = 0f(y) = 0$, so $xy \in \text{Ker } f$, and likewise $yx \in \text{Ker } f$.

Besides telling us what elements are sent to 0 by f , the kernel of f also tells us when two elements are sent to the same element. Since $f(x) = f(y)$ if and only if $f(x - y) = 0$, therefore, f will send x and y to the same element of S if and only if $x - y \in \text{Ker } f$.

3.6.2 Ideals of a ring

The properties of kernels of homomorphisms that we just found we'll use to define ideals of rings. Historically, ideals had a different purpose, but we'll get to that purpose later. The word "ideal" is short for ideal number or ideal element.

Definition 3.24. An *ideal* I of a ring R is a subset that (1) includes 0, (2) is closed under addition, and (3) is closed under multiplication by elements of R . We can summarize these requirements symbolically by $0 \in I$, $I + I \subseteq I$, $RI \subseteq I$, and $IR \subseteq I$.

Both of the last two requirements, $RI \subseteq I$ and $IR \subseteq I$, are needed when R is a non-commutative ring. Most of the time we'll be dealing with commutative rings so one will do.

Note that $\{0\}$ is always an ideal in a ring R . It's called the *trivial* ideal. We'll usually just denote it 0. Also, the entire ring R is an ideal, but not a proper ideal. A *proper ideal* is any ideal $I \neq R$.

Theorem 3.25. The intersection of ideals is an ideal.

Proof. Here's the proof for two ideals I_1 and I_2 of a ring R . This proof can be generalized to any number, including an infinite number, of ideals.

We need to show that $I_1 \cap I_2$ (1) includes 0, (2) is closed under addition, and (3) is closed under multiplication by elements of R .

First, since $0 \in I_1$ and $0 \in I_2$, therefore $0 \in I_1 \cap I_2$.

Second, given two elements $x, y \in I_1 \cap I_2$, to show $x + y \in I_1 \cap I_2$. Since $x, y \in I_1 \cap I_2$, therefore $x, y \in I_1$ and $x, y \in I_2$. Therefore $x + y \in I_1$ and $x + y \in I_2$, and so $x + y \in I_1 \cap I_2$.

Third, given $x \in I_1 \cap I_2$ and $y \in R$, to show $xy \in I_1 \cap I_2$. Since $x \in I_1 \cap I_2$, therefore $x \in I_1$ and $x \in I_2$. Therefore, $xy \in I_1$ and $xy \in I_2$, and so $xy \in I_1 \cap I_2$. Q.E.D.

Principal ideals and ideals generated by a set. The simplest examples of ideals are what are called principal ideals. Let a be an element of a commutative ring R . The set of all multiples of a ,

$$(a) = \{xa \mid x \in R\},$$

is an ideal of R , as you can easily check. These ideals are called *principal ideals* because they are generated by one element. An alternate notation for the principal ideal generated by the element a is Ra or aR .

Note that (0), the ideal generated by 0, is just the 0 ideal, while (1), the ideal generated by 1, is all of R .

Sometimes it takes more than one element to generate an ideal. Let A be a subset of a commutative ring R . The smallest ideal that contains A is called the *ideal generated by A* . It must contain all linear combinations of elements of A since an ideal is closed under addition and closed under multiplication by elements of R , but that's enough. Usually, we're only interested in generating an ideal from a finite number of elements $A = \{a_1, a_2, \dots, a_k\}$. Then the ideal generated by A is

$$(a_1, a_2, \dots, a_k) = \{x_1a_1 + \dots + x_ka_k \mid \text{each } x_i \in R\}.$$

An example of an ideal generated by two elements but not principal (not by one element) is $(5, x^2)$ in $\mathbf{Z}[k]$, the polynomial ring with integral coefficients.

Exercise 38. As you know, if $n \in \mathbf{Z}$, then $n\mathbf{Z}$, also written (n) , is an ideal of the ring \mathbf{Z} . Consider the two ideals $I = 6\mathbf{Z}$ and $J = 10\mathbf{Z}$ of the \mathbf{Z} .

- (a). Determine their intersection $I \cap J$ as a principal ideal of \mathbf{Z} .
- (b). Prove that the union $I \cup J$ is not an ideal of \mathbf{Z} .

3.6.3 Quotient rings, R/I

As mentioned above the kernel of a ring homomorphism f tells us when two elements are sent to the same element: $f(x) = f(y)$ if and only if $x - y \in \text{Ker } f$. We can use $\text{Ker } f$ to construct a “quotient ring” $R/\text{Ker } f$ by identifying two elements x and y in R if their difference lies in $\text{Ker } f$. In fact, we can do this not just for kernels of homomorphisms, but for any ideal I . That is, we can use an ideal I of R to determine when two elements x and y are to be identified, $x \equiv y$, and we’ll end up with a ring R/I . The identification is called a congruence. This concept of congruence generalizes congruence modulo n on \mathbf{Z} .

Definition 3.26. A *congruence* \equiv on a ring R is an equivalence relation such that for all $x, x', y, y' \in R$,

$$x \equiv x' \text{ and } y \equiv y' \text{ imply } x + y \equiv x' + y' \text{ and } xy \equiv x'y'.$$

Since we’re dealing with rings with 1, we’ll usually insist that $0 \not\equiv 1$. The equivalence classes for a congruence are called *congruence classes*.

Theorem 3.27. If \equiv is a congruence on a ring R , then the quotient set R/\equiv , that is, the set of congruence classes, is a ring where addition is defined by $[x] + [y] = [x+y]$ and multiplication by $[x][y] = [xy]$.

Proof. First we need to show that the proposed definitions are actually well defined. That is, if a different representative x' is chosen from the congruence class $[x]$ and y' from $[y]$, then the same classes $[x'+y']$ and $[x'y']$ result. That is

$$[x] = [x'] \text{ and } [y] = [y'] \text{ imply } [x+y] = [x'+y'] \text{ and } [xy] = [x'y'].$$

That’s the same as the requirements met in the definition of congruence (which explains why they are in the definition).

Also, each of the axioms for a ring need to be verified, but they’re all automatic. Here’s commutativity of addition, for example.

$$[x] + [y] = [x+y] = [y+x] = [y] + [x].$$

We could say that the quotient ring inherits the properties from the ring.

Q.E.D.

In the next theorem we’ll see that an ideal I determines a congruence. We’ll write the congruence $x \equiv y \pmod{I}$ rather than just $x \equiv y$ when we want to emphasize the role of I . The congruence classes may be written $[x]$ or $[x]_I$, or $x+I$. The last notation is a good one since $[x] = \{x+y \mid y \in I\}$.

Theorem 3.28 (Congruence modulo an ideal). Let I be an ideal of a ring R . A congruence, called *congruence modulo I* , is defined by

$$x \equiv y \pmod{I} \text{ if and only if } x-y \in I.$$

The quotient ring, R/\equiv , is denoted R/I .

Proof. First, we need to show that it's an equivalence relation.

Reflexivity. $x \equiv x \pmod{I}$. That's okay since $x - x = 0 \in I$.

Symmetry. $x \equiv y \pmod{I}$ implies $y \equiv x \pmod{I}$. That's okay because if $x - y \in I$, then $y - x = -(x - y) \in I$.

Transitivity. $x \equiv y \pmod{I}$ and $y \equiv z \pmod{I}$ imply $x \equiv z \pmod{I}$. That's okay, too. If $x - y \in I$ and $y - z \in I$, then so is their sum $x - z \in I$.

Thus, it's an equivalence relation. Next to show that

$$x \equiv x' \pmod{I} \text{ and } y \equiv y' \pmod{I} \text{ imply } x + y \equiv x' + y' \pmod{I} \text{ and } xy \equiv x'y' \pmod{I}.$$

That requirement reduces to the statement

$$x - x' \in I \text{ and } y - y' \in I \text{ imply } (x + y) - (x' + y') \in I \text{ and } (xy - x'y') \in I,$$

which, you can check, follow from the definition of ideal.

Q.E.D.

Exercise 39. Prove the last statement above: if $x - x' \in I$ and $y - y' \in I$, then $(x + y) - (x' + y') \in I$ and $(xy - x'y') \in I$.

Example 3.29 (Cyclic rings). As we saw above, $I = n\mathbf{Z}$ is an ideal of \mathbf{Z} . The congruence defined here is the same one we had before. Thus, $x \equiv y \pmod{I}$ means $x \equiv y \pmod{n}$. The quotient ring is $\mathbf{Z}/n\mathbf{Z}$, which we have studied before and denoted \mathbf{Z}_n for short.

Comment 3.30. The ring structure on the quotient R/I was defined from the ring structure on R , so the projection $\gamma : R \rightarrow R/I$ is a ring homomorphism. This ring R/I is called a *quotient ring* of R . (It is also sometimes called a factor ring, but that term should be restricted to the case when R factors as a product of rings, one of which is R/I . An example of that is the Chinese remainder theorem.)

Examples 3.31 (Quadratic field extensions.). We've looked at $\mathbf{Q}(\sqrt{2})$, $\mathbf{C} = \mathbf{R}(i)$, and other quadratic field extensions. We can interpret them as quotient rings.

Let's take $\mathbf{Q}(\sqrt{2})$ first. Consider the ring $R = \mathbf{Q}[x]$ of polynomials with rational coefficients. An ideal in R is the principal ideal $I = (x^2 - 2)$ generated by the polynomial $x^2 - 2$. In the quotient ring $R/I = \mathbf{Q}[x]/(x^2 - 2)$, we have $x^2 - 2 \equiv 0 \pmod{I}$, that is, $x^2 \equiv 2 \pmod{I}$, so in R/I , we find that 2 does have a square root, namely x . Since in R/I every polynomial $a_nx^n + \dots + a_1x + a_0$ is congruent to a polynomial of degree 1 (because $x^2 \equiv 2$), but no two linear polynomials are congruent mod I (because $a_1x + a_0 \equiv b_1x + b_0 \pmod{I}$ implies $(a_1 - b_1)x + (a_0 - b_0) \in I$ so $a_1 = b_1$ and $a_0 = b_0$), therefore every element in R/I is uniquely represented as a linear polynomial $a_1x + a_0$. If we denote x by the symbol $\sqrt{2}$, then we find $\mathbf{Q}[x]/(x^2 - 2)$ is the same field as $\mathbf{Q}(\sqrt{2})$ that we described before.

Likewise, $\mathbf{R}[x]/(x^2 + 1)$ is \mathbf{C} .

We'll find this construction of new rings as quotient rings is very useful, especially when we take quotients rings of polynomial rings like we did here.

The image of a ring homomorphism is isomorphic to the ring modulo its kernel.
 Let $f : R \rightarrow S$ be a ring homomorphism. The image of f , denoted $f(R)$, is the set

$$f(R) = \{f(x) \in S \mid x \in R\}.$$

It is a subring of S , as you can easily verify. You can also show the following isomorphism theorem, called the first isomorphism for rings.

Theorem 3.32. If $f : R \rightarrow S$ is a ring homomorphism then the quotient ring $R/\text{Ker } f$ is isomorphic to the image ring $f(R)$, the isomorphism being given by

$$\begin{aligned} R/\text{Ker } f &\rightarrow f(R) \\ x + \text{Ker } f &\mapsto f(x) \end{aligned}$$

Exercise 40. Prove the preceding theorem.

(a). First show that the assignment $x + \text{Ker } f$ to $f(x)$ is well defined. That means that if $x + \text{Ker } f = x' + \text{Ker } f$, then $f(x) = f(x')$. Call that function $\phi(x)$.

(b). Show that assignment is a ring homomorphism. Show (1) $\phi(1) = 1$, (2) $\phi(x + y) = \phi(x) + \phi(y)$, and (3) $\phi(xy) = \phi(x)\phi(y)$.

This gives us two ways to look at the image, either as a quotient ring of the domain R or as a subring of the codomain S .

Furthermore, we can now treat a ring homomorphism $f : R \rightarrow S$ as a composition of three ring homomorphisms.

$$R \rightarrow R/\text{Ker } f \cong f(R) \rightarrow S$$

The first is the projection from R onto its quotient ring $R/\text{Ker } f$, the second is the isomorphism $R/\text{Ker } f \cong f(R)$, and the third is the inclusion of the image $f(R)$ as a subring of S .

3.6.4 Prime and maximal ideals

Sometimes it occurs that R/I is not just a ring, but either an integral domain or even a field. Those results occur when the ideal I is a prime ideal or a maximal ideal, respectively, as we'll define now.

Definition 3.33. An ideal I in a commutative ring R is said to be a *prime ideal* if R/I is an integral domain. Equivalently, I is a prime ideal if (1) $I \neq R$, and (2) $\forall x, y \in R$, if $xy \in I$, then either $x \in I$ or $y \in I$. An ideal I is said to be *maximal* if it's a proper ideal, but it is not contained in any larger proper ideal.

Exercise 41. Prove that R/I is an integral domain if and only if R/I satisfies both conditions (1) $I \neq R$, and (2) $\forall x, y \in R$, if $xy \in I$, then either $x \in I$ or $y \in I$.

Example 3.34. The ideals of \mathbf{Z} that are prime are those of the form $p\mathbf{Z}$ where p is a prime number, and the 0 ideal. In fact, $p\mathbf{Z}$ are maximal ideals, but 0 is not maximal.

In a field F there is only one proper ideal, namely 0.

In an integral domain, the 0 ideal is a prime ideal, and conversely, if 0 is an ideal in a commutative ring, then the ring is an integral domain.

Theorem 3.35. Every maximal ideal is prime.

Proof. Let I be a maximal ideal of a commutative ring R , and let $xy \in I$. Suppose $x \notin I$. Then $xR + I = \{xu + v \mid u \in R, v \in I\}$ is an ideal containing I . Since I is a maximal ideal, therefore $xR + I$ is not a proper ideal but all of R . Therefore $1 = xu + v$ for some $u \in R$, $v \in I$. Hence $y = yxu + yv \in Iu + I = I$. Thus, I satisfies the conditions to be a prime ideal. Q.E.D.

We won't show it right now, but we'll prove later Krull's theorem which says that every ideal is contained in a maximal ideal. We'll need to discuss the axiom of choice and Zorn's lemma before we can prove it.

Theorem 3.36 (Maximal ideal theorem). Let I be an ideal of a commutative ring R . Then I is a maximal ideal if and only if R/I is a field.

Proof. We'll use the notation $[x]$ for $x + I$ to stress that we're thinking of it as an element of R/I .

Suppose that I is a maximal ideal, and let $[x]$ be any nonzero element of R/I , that is $x \notin I$. As in the last proof, $xR + I = R$. Therefore $1 = xu + v$ for some $u \in R$, $v \in I$. Then, in R/I we have $[1] = [x][u] + [v] = [x][u] + [0] = [x][u]$. Therefore $[x]$ has a reciprocal, and R/I is a field.

Now suppose that R/I is a field. Let $x \notin I$. We'll show that $xR + I = R$ which will show that I is a maximal ideal. In R/I , $[x] \neq [0]$, so $[x]$ has an inverse $[y]$, $[x][y] = [1]$, so $1 - xy \in I$, so $1 \in xR + I$, hence $R = xR + I$. Q.E.D.

3.7 Krull's theorem

We'd like to prove Krull's theorem that every ideal in a commutative ring is contained in a maximal ideal, but in order to do that in general we'll need something called Zorn's lemma. It's a statement that's logically equivalent to the better known axiom of choice.

See section A.4 in the appendix for a review of the axiom of choice and Zorn's lemma.

Theorem 3.37 (Krull). Let I be a proper ideal of a commutative ring R . Then there is a maximal ideal J such that $I \subseteq J$.

Proof. Consider the collection \mathcal{M} of proper ideals of R that contain I . Note that \mathcal{M} is nonempty since $I \in \mathcal{M}$.

We'll show that every chain \mathcal{C} in \mathcal{M} has an upper bound in \mathcal{M} . Let $B = \bigcup_{A \in \mathcal{C}} A$. Certainly B is an upper bound for \mathcal{C} since B is just the union of elements of \mathcal{C} .

We still have to show B is an ideal, which requires $RB \subseteq B$ and $B + B \subseteq B$. For the first, $RB = R \left(\bigcup_{A \in \mathcal{C}} A \right) = \bigcup_{A \in \mathcal{C}} RA = \bigcup_{A \in \mathcal{C}} A = B$. Now let $x, y \in B$. Then $x \in A_1$ for some $A_1 \in \mathcal{C}$ and $y \in A_2$ for some $A_2 \in \mathcal{C}$. But \mathcal{C} is a chain, so either $A_1 \subseteq A_2$ or $A_2 \subseteq A_1$. In the first case, $x, y \in A_2$, so $x + y \in A_2 \subseteq B$, and in the second $x, y \in A_1$, so $x + y \in A_1 \subseteq B$. Thus, $B + B \subseteq B$.

Now we can apply Zorn's lemma. It implies \mathcal{M} has a maximal element J . Clearly, $I \subseteq J$, and J is a proper ideal of R , but there are no larger proper ideals of R that contain J , so J is a maximal ideal.

Q.E.D.

Note how we have not actually found J . There may be many different maximal ideals that contain I , and one was selected by a choice function, but we don't even know what the choice function is so we can't even determine J in principle.

It's actually the case that Krull's theorem is logically equivalent to the Axiom of Choice. That is, if Krull's theorem is taken as an axiom, then the Axiom of Choice can be proved from it.

There are many other applications of Zorn's lemma. For instance, you can prove that every vector space has a basis, even when the vector space is infinite dimensional.

3.8 Unique factorization domains, principal ideal domains, and Euclidean domains

Not every integral domain is as nice as the ring of integers. The ring of integers has three nice properties. One is unique factorization—every integer is uniquely a product of prime numbers. A second is that every ideal is a principal ideal. A third is that there is a division algorithm that is the basis of the Euclidean algorithm.

There aren't many rings that have all these properties, and some rings have none of them. We'll investigate these properties and their interrelations.

We'll use these three properties to define three special kinds of integral domains: unique factorization domains (UFDs), principal ideal domains (PIDs), and Euclidean domains (EDs). When we do we'll find every Euclidean domain is a principal ideal domain, every principal ideal domain is a unique factorization domain, every unique factorization domain is an integral domain; and every integral domain is a ring.

$$\text{EDs} \subset \text{PIDs} \subset \text{UFDs} \subset \text{Integral domains} \subset \text{Commutative rings}$$

3.8.1 Divisibility in an integral domain

We'll borrow the concepts of divisibility and greatest common divisor from \mathbf{Z} and apply them to integral domains. We'll separate the concept of prime number in \mathbf{Z} into two concepts since in some of the integral domains we'll look at they're actually different.

Definition 3.38. The following definitions apply to elements of an integral domain.

- Let a and b be nonzero elements. We'll say a divides b , written $a|b$, if there exists c such that $ac = b$.
- We'll say that d is a *greatest common divisor* of a and b , if d divides both a and b , and whenever another element e divides both a and b , then e divides d .
- An element x that is not zero and not a unit is *irreducible* if whenever $x = yz$, either y or z is a unit, otherwise it is *reducible*

- An element x that is not zero and not a unit is *prime* if whenever $x|yz$, then $x|y$ or $x|z$.

Note that we won't use the notation $d = \text{GCD}(a, b)$ when d is a greatest common divisor since there will be other greatest common divisors, that is, the greatest common divisor is only unique up to a unit. Later, when we look at principal ideal domains, we can use the notation $(c) = (a, b)$ for greatest common divisors which says the principal ideal (c) is the same as the ideal generated by a and b .

Exercise 42. Several properties of divisibility follow directly from the definition just like they do with the integral domain is \mathbf{Z} . Prove the following properties from the above definitions.

- 1 divides every element.
- Each element divides itself.
- If $a|b$ then $a|bc$.
- Divisibility is transitive.
- If one element divides two other elements, then it divides both their sum and difference.
- Cancellation: When $c \neq 0$, $a|b$ if and only if $ac|bc$.

Theorem 3.39. If an element in an integral domain is prime, then it irreducible.

Proof. Let x be prime. Suppose that $x = yz$. Then $x|yz$, so either $x|y$ or $x|z$. In the first case, $xw = y$ for some w . Therefore $xwz = yz = x$. Cancel the x to conclude $wz = 1$. Then z is a unit. Likewise, in the second case y is a unit. Therefore x is irreducible. Q.E.D.

The converse of this theorem does not hold. That is, there are integral domains where not all irreducible elements are prime. We'll see that in this next example. But then a little later, we'll see that in principal ideal domains (about to be defined), irreducible elements are prime.

Example 3.40 (a nonUFD). We'll find a number of other UFDs, but, it's important to know that not every integral domain has unique factorization. Consider the integral domain $R = \mathbf{Z}[\sqrt{10}]$. An element of it is of the form $x + y\sqrt{10}$ where x and y are integers. In this integral domain 9 can be factored in two ways.

$$9 = 3^2 = (\sqrt{10} + 1)(\sqrt{10} - 1),$$

but 3 , $\sqrt{10} + 1$, and $\sqrt{10} - 1$ are all irreducible. This integral domain, and many others, are not UFDs. Although the three elements 3 , $\sqrt{10} + 1$, and $\sqrt{10} - 1$ are irreducible, none divides any other, so none of them is prime, as you can see by the equation involving 9, above.

3.8.2 Unique factorization domains

Unique factorization is a property that we might expect, but it turns out it doesn't hold in every integral domain. Given any element x in a ring D , we expect that we can factor it into 'atoms,' things that can't be cut further, and that there's only one way to do that. Of course, with our experience with the integers, we know that there's a bit of difficulty in stating the uniqueness part of the claim. For one thing, the order of the factors is variable, and, for another, there are units, like 1 and -1 that can be inserted to change the formal listing of the factors. Still, these are small things that we can deal with.

Definition 3.41. An integral domain is a *unique factorization domain* (UFD) if every element in it is a product of irreducible elements and it is a product of irreducible elements in only one way apart from the order of the product and factors of units.

The ring \mathbf{Z} of integers is, of course, a unique factorization domain. An integer, such as 6 can be written in more than one way as a product of irreducible elements (primes, in the case of integers) $6 = 2 \cdot 3 = (-3) \cdot (-2)$, but the only difference is the order of the primes and the insertions of units in the factorization.

Recall that an ideal I in a commutative ring R is a prime ideal when R/I is an integral domain. Equivalently, I is a prime ideal if (1) $I \neq R$, and (2) for all $x, y \in R$, if $xy \in I$, then either $x \in I$ or $y \in I$.

Theorem 3.42. An nonzero element x is an integral domain D is prime if and only if the principal ideal (x) is a prime ideal.

Exercise 43. Prove the preceding theorem. Note that there are two things to prove in an if-and-only-if statement.

3.8.3 Principal ideal domains

A second nice property that the ring of integers has is that every ideal in \mathbf{Z} is generated by a single element. If I is an ideal in \mathbf{Z} , then the GCD of all its nonzero elements is an element of I and all other elements are multiples of this GCD. This will be our definition of a principal ideal domain (PID), and we'll show that every PID is a UFD. There are UFDs that aren't PIDs, for instance, $\mathbf{Z}[x]$, the ring of polynomials with integer coefficients is one; one nonprincipal ideal is generated by 2 and x .

Definition 3.43. An integral domain is a *principal ideal domain* (PID) if every ideal in the domain is principal, that is, generated by one element.

Besides \mathbf{Z} , other prominent PIDs are $F[x]$ where F is a field. We'll prove this in section 3.8.4 on Euclidean domains which are special kinds of PIDs.

We'll show in a couple of steps that every PID is a UFD. The first one makes a connection between greatest common divisors and ideals.

Theorem 3.44. Let D be a principal ideal domain with nonzero elements a and b . The ideal (a, b) is principal, so it is equal to (c) for some element c . Then c is a greatest common divisor of a and b .

Proof. Since $a \in c$, therefore $c|a$. Likewise, $c|b$. We also know that $c \in (a, b)$, so $c = xa + yb$ for some elements x and y .

To show that c is a greatest common divisor, suppose d is some other common divisor of a and b . Then $a = ud$ and $b = vd$ for some elements u and v . Now,

$$c = xa + yb = xud + yvd = (xu + yv)d.$$

Therefore, $d|c$. Thus c is a greatest common divisor of a and b . Q.E.D.

Theorem 3.45. In a principal ideal domain, irreducible elements are prime.

Proof. Suppose that p is irreducible and $p|ab$. We'll show either $p|a$ or $p|b$. We'll do that by showing that if p doesn't divide a , then it does divide b .

Suppose p does not divide a . Then the ideal (p, a) is (1) since p is irreducible. Since $1 \in (p, a)$, $1 = xp + ya$ for some elements x and y . Therefore, $b = bxp + aby$. Since $p|ab$, therefore $p|bxp + aby$, so $p|b$.

Thus, the irreducible element p is also prime. Q.E.D.

Next, we'll use the following lemma to show that elements have factorizations in PIDs. We'll still have to show they're unique after that. The condition in the lemma is called the *ascending chain condition* (ACC) on ideals, and rings that satisfy it are called *Noetherian rings* in honor of Noether who studied such rings.

Lemma 3.46. In a principal ideal domain, there are no infinitely ascending chains of ideals. That is,

$$(a_1) \subsetneq (a_2) \subsetneq (a_3) \subsetneq \cdots$$

does not exist.

Proof. Suppose there were such an infinitely ascending chain of ideals. Then the union $I = \bigcup_{i=1}^{\infty} (a_i)$ is an ideal, as you can easily check. It must be principal, so $I = (a)$ for some element a . But a is in the union, so it's in one of the ideals (a_i) . Then

$$(a) \subseteq (a_i) \subsetneq (a_{i+1}) \subseteq (a),$$

a contradiction. Q.E.D.

There are rings, in fact UFDs, that are not Noetherian. An example is a polynomial ring with infinitely many variables such as $\mathbf{Q}[x_1, x_2, x_3, \dots]$. An infinitely ascending chain of ideals in that ring is $(x_1) \subsetneq (x_1, x_2) \subsetneq (x_1, x_2, x_3) \subsetneq \cdots$.

Theorem 3.47. In a principal ideal domain, every element that is not zero and not a unit has a factorization into irreducible elements.

Proof. Suppose that a nonzero element a_1 has no factorization into irreducible elements. We'll derive a contradiction, but we'll need an element with no factorization with an extra property. We'll get that element, denoted a_n below, as follows.

Starting with the ideal (a_1) , form any ascending chain of ideals generated by other elements with no factorizations, and extend the chain as far as possible. By the lemma, it stops somewhere, say at (a_n) .

$$(a_1) \subsetneq (a_2) \subsetneq \cdots \subsetneq (a_n).$$

We now have an element a_n which has no factorization into irreducible elements with an extra property, namely, any ideal strictly containing (a_n) is generated by an element that does have such a factorization. Now, a_n is not irreducible itself, for that would be a factorization, so $a_n = bc$ where neither b nor c is a unit. Since $b|a_n$, therefore $(a_n) \subseteq (b)$. But $(a_n) \neq (b)$, for otherwise $b = a_nd$ for some d , and then $a_ndc = bc = a_n$, so $dc = 1$ making c a unit, which it is not.

So $(a_n) \subsetneq (b)$ and likewise $(a_n) \subsetneq (c)$, therefore both b and c have factorizations, and the product of those factorizations gives a factorization for a_n , a contradiction. Q.E.D.

Theorem 3.48. Every principal ideal domain is a unique factorization domain.

Proof. The last theorem gave the existence of at least one factorization for an element a . We still have to show that there's at most one factorization.

Suppose that a has two factorizations as products of irreducible elements.

$$a = p_1 \cdots p_n = q_1 \cdots q_m$$

Since the irreducible element p_1 is prime (in a PID), p_1 divides one of the q_i 's, which we can renumber as q_1 . Then $p_1 = u_1 q_1$ where u_1 is a unit. Substitute $u_1 q_1$ for p_1 , and cancel q_1 to get the equation

$$u_1 p_2 \cdots p_n = q_2 \cdots q_m.$$

That completes the inductive step of mathematical induction on n . The base case, when $n = 1$, is left to the reader. Q.E.D.

3.8.4 Euclidean domains

The third nice property that \mathbf{Z} has is that there is a division algorithm that is the basis of the Euclidean algorithm.

Some example Euclidean domains besides \mathbf{Z} that we'll discuss in this section include the Gaussian integers $\mathbf{Z}[i]$, the Eisenstein integers $\mathbf{Z}[\omega]$ where ω is a primitive cube root of 1, and polynomial rings $F[x]$ over a field F .

For the integers, the division algorithm starts with an integer a (the dividend) and a nonzero integer b (the divisor) and delivers q (the quotient) and r (the remainder) such that

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

This property allowed us to construct the Euclidean algorithm for finding GCDs as well as the extended Euclidean algorithm to show that the greatest common divisor of two numbers is a linear combination of them.

There are a few other integral domains that have the same kind of division algorithm where the remainder is somehow “smaller” than the divisor, but the concept of smaller and how to find q and r differs from domain to domain.

Definition 3.49. A *Euclidean valuation* on an integral domain D is a function $v : D - 0 \rightarrow \mathbf{Z}_{\geq 0}$ that satisfies the conditions

1. for nonzero elements a and b , $v(a) \leq v(ab)$, and
2. for each element a (the dividend) and nonzero element b (the divisor), there are elements q (the quotient) and r (the remainder) such that

$$a = qb + r \quad \text{where either } r = 0 \text{ or } v(r) < v(b).$$

An integral domain that admits a Euclidean valuation is called *Euclidean domain*.

Of course, \mathbf{Z} is a Euclidean domain with the valuation being the absolute value $v(a) = |a|$.

Another class of Euclidean domains are the rings of polynomials (in one variable) with coefficients in a given field. The following theorem is essentially just long division for polynomials. We'll make it simple by making the divisor $g(x)$ a monic polynomial, that is, a polynomial whose leading coefficient is 1.

It directly follows from the division algorithm for polynomials over a field, theorem 1.55, that a field's polynomial ring is a Euclidean domain.

Corollary 3.50. The polynomial ring $F[x]$ with coefficients in a field F is a Euclidean domain where the valuation v assigns to a polynomial $f(x)$ the degree of f .

Soon we'll study polynomial rings in more detail.

There are other Euclidean domains including the Gaussian integers and the Eisenstein integers.

The Gaussian integers $\mathbf{Z}[i]$ is a Euclidean domain. The ring of Gaussian integers is $\mathbf{Z}[i] = \{a_1 + a_2i \mid a_1, a_2 \in \mathbf{Z}\}$. Its valuation function, also called the norm, is $v(a_1 + a_2i) = a_1^2 + a_2^2$, the square of the distance to the origin. In order to divide one Gaussian integer $b_1 + b_2i$ into another $a_1 + a_2i$ to get a quotient $q_1 + q_2i$ and remainder $r_1 + r_2i$, you can perform the complex division $\frac{a_1 + a_2i}{b_1 + b_2i}$ to get an exact quotient, and choose $q_1 + q_2i$ to be the closest Gaussian integer to that exact quotient. The remainder is then determined.

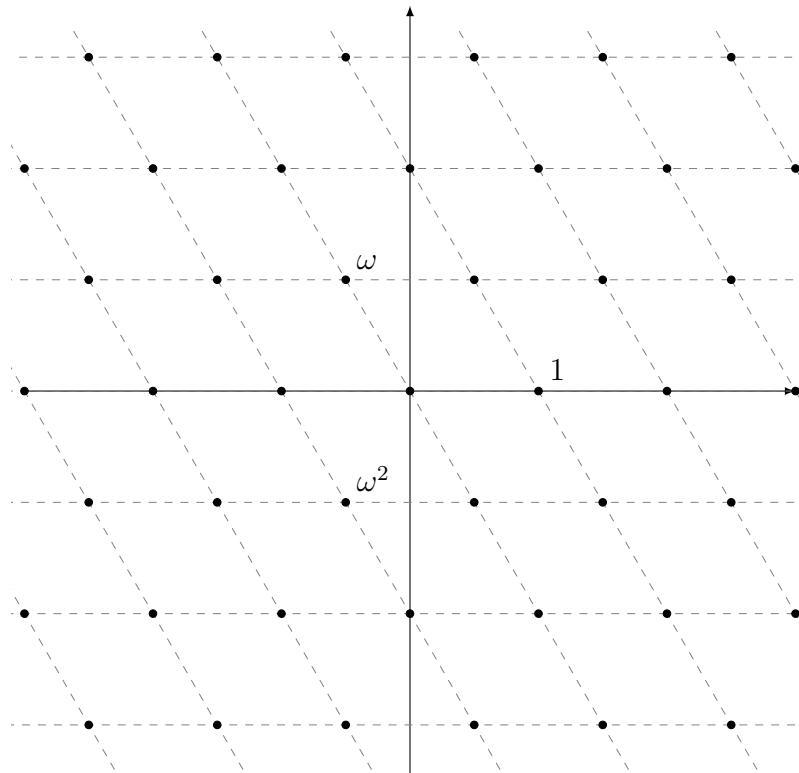


Figure 3.3: Lattice of Eisenstein integers

Eisenstein integers Whereas a basis for the Gaussian integers consists of 1 and i , a basis for the Eisenstein integers consists of 1 and ω where $\omega = \frac{1}{2}(-1+i\sqrt{3})$ is a primitive cube root of unity. A primitive cube root of unity satisfies the equation $\frac{x^3 - 1}{x - 1} = 0$ which simplifies to $x^2 + x + 1 = 0$. The lattice of Eisenstein integers is a triangular lattice since $1 + \omega + \omega^2 = 0$. The lattice is shown in figure 3.3. The dotted lines show coordinates relative to the basis consisting of 1 and ω .

There are six units in the Eisenstein integers. They are the six sixth roots of unity: 1 itself, a primitive sixth root $\omega - 1$, a primitive cube root ω , the primitive square root -1 , and a primitive sixth root $-\omega$. They are equally spaced at 60° around the unit circle.

Like the Gaussian integers, the Eisenstein integers also are a Euclidean domain. The valuation is $v(a + b\omega) = a^2 - ab + b^2$.

The Euclidean algorithm in Euclidean domains. First, we'll show that Euclidean domains are principal ideal domains, and since PIDs are also UFDs, therefore Euclidean domains are also unique factorization domains. Then we'll look at an example of the Euclidean algorithm in a Euclidean domain other than \mathbf{Z} .

Theorem 3.51. A Euclidean domain is a principal ideal domain.

Proof. Let I be an ideal in a Euclidean domain D with valuation v . We'll show I is a principal ideal. If I is the zero ideal (0) , then it's principal of course.

Assume now that I has a nonzero element, and let $S = \{v(x) \mid 0 \neq x \in I\}$. This is a nonempty subset of the nonnegative integers, so it has a least element, and let that be $v(a)$. Thus, a is a nonzero element of I , so $(a) \subseteq I$. Let x be any other nonzero element in I . Then $v(a) \leq v(x)$. Furthermore, there are elements q and r in D such that $x = aq + r$ and either $r = 0$ or $v(r) < v(a)$. But $r = x - aq \in I$, so if $r \neq 0$, then $v(r) > v(a)$ contradicts $v(a) \leq v(r)$. Therefore, $r = 0$, and hence $x = aq$, so $a|x$. Therefore, $I = (a)$. Thus, D is a PID. Q.E.D.

The Euclidean algorithm works in any Euclidean domain the same way it does for integers. It will compute the greatest common divisor (up to a unit), and the extended Euclidean algorithm will construct the greatest common divisor as a linear combination of the original two elements.

Example 3.52. Let's take an example from the polynomial ring $\mathbf{Q}[x]$. Let's find the greatest common divisor of $f_1(x) = x^4 + 2x^3 - x - 2$ and $f_2(x) = x^4 - x^3 - 4x^2 - 5x - 3$. They have the same degree, so we can take either one of them as the divisor; let's take $f_2(x)$. Divide f_2 into f_1 to get a quotient of 1 and remainder of $f_3(x) = 3x^3 + 4x^2 + 4x + 1$. Then divide f_3 into f_2 to get a quotient and a remainder f_4 , and continue until the remainder is 0 (which occurs on the next iteration).

$$\begin{array}{ll} f_1(x) = x^4 + 2x^3 - x - 2 & f_1(x) = 1 \cdot f_2(x) + f_3(x) \\ f_2(x) = x^4 - x^3 - 4x^2 - 5x - 3 & f_2(x) = (\frac{1}{3}x - \frac{7}{9})f_3(x) + f_4(x) \\ f_3(x) = 3x^3 + 4x^2 + 4x + 1 & f_3(x) = (\frac{27}{20}x - \frac{9}{20})f_4(x) \\ f_4(x) = -\frac{20}{9}x^2 - \frac{20}{9}x - \frac{20}{9} & \end{array}$$

Thus, a greatest common divisor is $f_4(x)$, which differs by a unit factor from the simpler greatest common divisor $x^2 + x + 1$. We can read the equations on the right in reverse to get f_4 as a linear combination of f_1 and f_2 .

$$\begin{aligned} f_4(x) &= f_2(x) - \left(\frac{1}{3}x - \frac{7}{9}\right)f_3(x) \\ &= f_2(x) - \left(\frac{1}{3}x - \frac{7}{9}\right)(f_1(x) - f_2(x)) \\ &= \left(\frac{1}{3}x + \frac{2}{9}\right)f_2(x) - \left(\frac{1}{3}x - \frac{7}{9}\right)f_1(x) \end{aligned}$$

3.9 Real and complex polynomial rings $\mathbf{R}[x]$ and $\mathbf{C}[x]$

We know a fair amount about $F[x]$, the ring of polynomials over a field F . It has a division algorithm, so it's a Euclidean domain where the Euclidean valuation is the degree of a polynomial, so it has division and Euclidean algorithms. Since it's Euclidean, it's also a principal ideal domain, and that means irreducible elements are prime. And since it's a PID, it's also a unique factorization domain, that is, every polynomial uniquely factors as a product of irreducible polynomials.

Rather than calling irreducible polynomials prime polynomials, we'll use the term "irreducible polynomial". That's the common practice.

The nonzero prime ideals of $F[x]$ are just the principal ideals (f) generated by irreducible polynomials $f \in F[x]$, and, furthermore, they're maximal ideals, so $F[x]/(f)$ is a field. We've seen examples of this, for instance, $\mathbf{R}[x]/(x^2 + 1) \cong \mathbf{R}[i] = \mathbf{C}$, $\mathbf{Q}[x]/(x^2 - 2) \cong \mathbf{Q}(\sqrt{2})$, and $\mathbf{Z}_3[x]/(x^2 + 1) \cong \mathbf{Z}_3(i)$.

So, irreducible polynomials in $F[x]$ give field extensions of F .

The main question for $F[x]$ is: what are the irreducible polynomials?

We'll start with $\mathbf{C}[x]$ and $\mathbf{R}[x]$ followed by $\mathbf{Q}[x]$ and $\mathbf{Z}[x]$.

3.9.1 $\mathbf{C}[x]$ and the Fundamental Theorem of Algebra

In the 16th century Cardano (1501–1576) and Tartaglia (1500–1557) and others found formulas for roots of cubic and quartic equations in terms of square roots and cube roots. At the time, only positive numbers were completely legitimate, negative numbers were still somewhat mysterious, and the first inkling of a complex number appeared. Incidentally, at this time symbolic algebra was still being developed, and they wrote their equations in words instead of symbols!

Here's an illustration of how complex numbers arose. One of Cardano's cubic formulas gives the solution to the equation $x^3 = cx + d$ as

$$x = \sqrt[3]{d/2 + \sqrt{e}} + \sqrt[3]{d/2 - \sqrt{e}}$$

where $e = (d/2)^2 - (c/3)^3$. Bombelli used this to solve the equation $x^3 = 15x + 4$, which was known to have 4 as a solution, to get the solution

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

Now, $\sqrt{-121}$ is not a real number; it's neither positive, negative, nor zero. Bombelli continued to work with this expression until he found equations that lead him to the solution 4. Assuming that the usual operations of arithmetic held for these “numbers,” he determined that

$$\sqrt[3]{2 + \sqrt{-121}} = 2 + \sqrt{-1} \quad \text{and} \quad \sqrt[3]{2 - \sqrt{-121}} = 2 - \sqrt{-1}$$

and, therefore, the solution $x = 4$.

Cardano had noted that the sum of the three solutions of a cubic equation $x^3+bx^2+cx+d=0$ is $-b$, the negation of the coefficient of x^2 . By the 17th century the theory of equations had developed so far as to allow Girard (1595–1632) to state a principle of algebra, what we call now “the fundamental theorem of algebra.”

His formulation, which he didn’t prove, also gives a general relation between the n solutions to an n^{th} degree equation and its n coefficients.

For a generic equation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

Girard recognized that there could be n solutions, if you allow all roots and count roots with multiplicity. So, for example, the equation $x^2+1=0$ has the two solutions $\sqrt{-1}$ and $-\sqrt{-1}$, and the equation $x^2-2x+1=0$ has the two solutions 1 and 1. Girard wasn’t particularly clear what form his solutions were to have, just that there were n of them: x_1, x_2, \dots, x_n .

Girard gave the relation between the n roots x_1, x_2, \dots, x_n and the n coefficients a_1, \dots, a_n that extended Cardano’s remark. First, the sum of the roots $x_1 + x_2 + \cdots + x_n$ is $-a_1$ (Cardano’s remark). Next, the sum of all products of pairs of solutions is a_2 . Next, the sum of all products of triples of solutions is $-a_3$. And so on until the product of all n solutions is either a_n (when n is even) or $-a_n$ (when n is odd). He figured this out by using a version of one of the properties of polynomials mentioned above, namely, if a_1, a_2, \dots, a_n are roots of a monic polynomial $f(x)$ of degree n , then

$$f(x) = (x - a_1)(x - a_2) \cdots (x - a_n).$$

If you expand the right side of the equation, you’ll derive his result.

Here’s an example. The 4th degree equation

$$x^4 - 6x^3 + 3x^2 + 26x - 24 = 0$$

has the four solutions $-2, 1, 3$, and 4 . The sum of the solutions equals 6, that is $-2+1+3+4=6$. The sum of all products of pairs (six of them) is

$$(-2)(1) + (-2)(3) + (-2)(4) + (1)(3) + (1)(4) + (3)(4)$$

which is 3. The sum of all products of triples (four of them) is

$$(-2)(1)(3) + (-2)(1)(4) + (-2)(3)(4) + (1)(3)(4)$$

which is -26 . And the product of all four solutions is -24 .

Over the remainder of the 17th century, negative numbers rose in status to be full-fledged numbers. But complex numbers remained suspect through much of the 18th century. They

weren't considered to be real numbers, but they were useful in the theory of equations and becoming more and more useful in analysis. It wasn't even clear what form the solutions to equations might take. Certainly "numbers" of the form $a + b\sqrt{-1}$ were sufficient to solve quadratic equations, even cubic and quartic equations.

Euler did a pretty good job of studying complex numbers. For instance, he studied the unit circle assigning the value $\cos \theta + i \sin \theta$ to the point on the unit circle at an angle θ clockwise from the positive real axis. He measured the angle by the length of the arc cut off by the angle. We call that measurement radians now, but the word "radian" wasn't coined until later.

In his study of this circle he developed what we call Euler's identity

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

This was an especially useful observation in the solution of differential equations. Because of this and other uses of i , it became quite acceptable for use in mathematics. By the end of the 18th century numbers of the form $x + iy$ were in fairly common use by research mathematicians, and it became common to represent them as points in the plane.

Yet maybe some other form of "number" was needed for higher-degree equations. The part of the Fundamental Theorem of Algebra which stated there actually are n solutions of an n^{th} degree equation was yet to be proved, pending, of course, some description of the possible forms that the solutions might take.

Still, at nearly the end of the 18th century, it wasn't yet certain what form all the solutions of a polynomial equation might take. Leibniz, for example, stated in 1702 that $x^4 - a^4$ didn't have roots of the form $x + y\sqrt{-1}$, but Euler showed it did in 1742. D'Alembert, Euler, de Foncenex, Lagrange, and Laplace developed partial proofs. Finally, in 1799, Gauss (1777–1855) published his first proof of the Fundamental Theorem of Algebra.

We won't look at his or any other proof of the theorem. That's usually proved in a course in complex analysis. We will, however, use the theorem.

Definition 3.53. A field F is *algebraically closed* if every polynomial $f(x) \in F[x]$ factors as a product of linear factors. Equivalently, a polynomial $f(x)$ of degree n has n roots in F counting multiplicities.

A weaker definition could be made, and that's that every polynomial of degree at least 1 has at least one root in F . By induction, the remaining roots can be shown to exist.

Thus, the Fundamental Theorem of Algebra is a statement that \mathbf{C} is an algebraically closed field. Therefore, the algebra of $\mathbf{C}[x]$ is particularly simple. The irreducible polynomials are the linear polynomials.

3.9.2 The polynomial ring $\mathbf{R}[x]$

Let's turn our attention now to polynomials with real coefficients. Much of what we can say about $\mathbf{R}[x]$ comes from the relation of \mathbf{R} as a subfield \mathbf{C} , and consequently= from the relation of $\mathbf{R}[x]$ as a subring of $\mathbf{C}[x]$. That is to say, we can interpret a polynomial $f(x)$ with real coefficients as a polynomial with complex coefficients.

Theorem 3.54. If a polynomial $f(x)$ with real coefficients has a complex root z , then its complex conjugate \bar{z} is also a root.

Proof. Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$ where each $a_i \in \mathbf{R}$. If z is a root of f , then $f(z) = a_n z^n + \cdots + a_1 z + a_0 = 0$. Take the complex conjugate of the equation, and note that $\bar{a}_i = a_i$. Then $f(\bar{z}) = a_n \bar{z}^n + \cdots + a_1 \bar{z} + a_0 = 0$. Thus, \bar{z} is also a root. Q.E.D.

This theorem tells us for a polynomial $f(x)$ with real coefficients, its roots either come in k pairs of a complex number or singly as real numbers. We can name the $2k$ complex roots as

$$z_1, \bar{z}_1, z_2, \bar{z}_2, \dots, z_k, \bar{z}_k.$$

Writing $z_1 = x_1 + y_1 i, \dots, z_k = x_i + y_k i$, the complex roots are

$$x_1 + y_1 i, x_1 - y_1 i, x_2 + y_2 i, x_2 - y_2 i, \dots, x_k + y_k i, x_k - y_k i$$

and the $n - 2k$ real roots as

$$r_{2k+1}, \dots, r_n.$$

Using the fact that \mathbf{C} is algebraically closed, we can write $f(x)$ as

$$\begin{aligned} f(x) &= a_n (x - z_1)(x - \bar{z}_1) \cdots (x - z_k)(x - \bar{z}_k)(x - r_{2k+1}) \cdots (x - r_n) \\ &= a_n (x^2 - 2x_1 x + x_1^2 + y_1^2) \cdots (x^2 - 2x_k x + x_k^2 + y_k^2)(x - r_{2k+1}) \cdots (x - r_n) \end{aligned}$$

This last expression has factored $f(x)$ as a product of irreducible quadratic and linear polynomials with real coefficients.

Theorem 3.55. The irreducible polynomials in $\mathbf{R}[x]$ are the linear polynomials and the quadratic polynomials with negative discriminant.

Proof. The remarks above show that only linear and quadratic polynomials can be irreducible. Linear polynomials are always irreducible. A quadratic polynomial will have no real roots when its discriminant is negative. Q.E.D.

3.10 Rational and integer polynomial rings

We've studied the irreducible polynomials in $\mathbf{C}[x]$ and $\mathbf{R}[x]$ with the help of the Fundamental Theorem of Algebra and found them to be easily classified. The irreducible polynomials in $\mathbf{C}[x]$ are the linear polynomials, and irreducible polynomials in $\mathbf{R}[x]$ are the linear polynomials and quadratic polynomials with negative discriminant.

Determining which polynomials in $\mathbf{Q}[x]$ are irreducible is much harder. Of course, all the linear ones are, and we'll be able to tell which quadratic and cubic ones are irreducible fairly easily. After that it becomes difficult.

3.10.1 Roots of polynomials

The quadratic case. Consider a quadratic polynomial $f(x) = ax^2 + bx + c$ with coefficients in \mathbf{Q} .

Its roots are given by the quadratic formula $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ which can be shown by the process known as completing the square. The *discriminant* of a quadratic polynomial is

$\Delta = b^2 - 4ac$. When Δ is positive, there are two real roots; when 0, there is one double root; and when negative, the roots are a pair of complex conjugate numbers.

When Δ is a perfect rational square, that is, the square of a rational number, then $f(x)$ factors, that is, it's reducible. Otherwise, it's irreducible.

Thus, $f(x)$ is irreducible if and only if the discriminant is not a perfect square.

The cubic case. It is more difficult to determine when a cubic polynomial $f(x) = ax^3 + bx^2 + cx + d$ with rational coefficients is irreducible, but not too difficult. Note that if $f(x)$ factors, then one of the factors has to be linear, so the question of reducibility reduces to the existence of a rational root of $f(x)$.

Various solutions of a cubic equation $ax^3 + bx^2 + cx + d = 0$ have been developed. Here's one. First, we may assume that f is monic by dividing by the leading coefficient. Our equation now has the form $x^3 + bx^2 + cx + d = 0$. Second, we can eliminate the quadratic term by replacing x by $y - \frac{1}{3}b$. The new polynomial in y will have different roots, but they're only translations by $\frac{1}{3}b$. We now have the cubic equation

$$y^3 + (c - \frac{1}{3}b^2)y + (\frac{2}{27}b^3 - \frac{1}{3}bc + d) = 0$$

which we'll write as

$$y^3 + py + q = 0.$$

By the way, this substitution which results in a polynomial whose term after the leading term is 0 has a name. It is called a Tschirnhaus substitution. The roots of the new polynomial will sum to 0.

We'll follow Viète's method and perform another substitution. Replace y by $z - \frac{p}{3z}$. After simplifying and clearing the denominators we'll have the equation

$$z^6 + qz^3 - \frac{p^3}{27z} = 0$$

which is a quadratic equation in z^3 . Its complex solutions are

$$z^3 = \frac{-q \pm \sqrt{q^2 + 4p^3/27}}{2} = -\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}.$$

Taking complex cube roots to get three values for z , then using $y = z - \frac{p}{3z}$ to determine y and $x = y - \frac{1}{3}b$ to determine x , we have the all three complex solutions to the original equation. At least one of these three complex solutions is real, and perhaps all three.

We have a way of determining whether a cubic polynomial is reducible. First $q^2 + 4p^3/27$ needs to be a perfect rational square r^2 , then one of $-q + r$ and $-q - r$ needs to be a perfect rational cube.

There is another way to determine if there is a rational root.

Rational roots of a polynomial. If we're looking for the roots of a polynomial with rational coefficients, we can simplify the job a little bit by clearing the denominators so that all the coefficients are integers. The following theorem helps in finding roots.

Theorem 3.56 (Rational root theorem). Let $f(x) = a_nx^n + \cdots + a_1x + a_0$ be a polynomial with integral coefficients. If r/s is a rational root of f with r/s in lowest terms, then r divides the constant a_0 and s divides the leading coefficient a_n .

Proof. Since r/s is a root, therefore

$$f(x) = a_n(r/s)^n + a_n(r/s)^{n-1} + \cdots + a_1(r/s) + a_0 = 0,$$

and so, clearing the denominators, we have

$$a_nr^n + a_nr^{n-1}s + \cdots + a_1rs^{n-1} + a_0s^n = 0.$$

We can rewrite this equation as

$$(a_nr^{n-1} + a_nr^{n-2}s + \cdots + a_1s^{n-1})r = -a_0s^n.$$

Now, since r divides $-a_0s^n$, and r is relatively prime to s , and hence to s^n , therefore r divides a_0 . In like manner, you can show s divides a_n . Q.E.D.

For example, to find the rational roots r/s of $f(x) = 27x^4 + 30x^3 + 26x^2 - x - 4$, r will have to divide 4, so the possibilities for r are $\pm 1, \pm 2, \pm 4$, and s will have to divide 27, so the possibilities for s are $1, 3, 9, 27$ (since we may assume s is positive). That gives 24 rational numbers to check, and among them will be found the two rational roots $\frac{1}{3}$ and $-\frac{4}{9}$. After one, $\frac{r}{s}$, is found $f(x)$ can be divided by $x - \frac{r}{s}$ to lower the degree of the polynomial to find the rest of the roots.

If a polynomial does have a rational root, then it's clearly reducible since that rational root determines a linear factor of the polynomial. That gives us another way to determine if a cubic polynomial is reducible.

For polynomials of degree 4 or higher, knowing that there are no rational roots is insufficient to conclude the polynomial is irreducible. It still may factor as quadratic and higher degree terms. For example, $x^4 + x^2 + 1$ has no rational roots, but it factors as $(x^2 + x + 1)(x^2 - x + 1)$, so it is reducible.

3.10.2 Gauss's lemma and Eisenstein's criterion

Further study of $\mathbf{Q}[x]$ will require looking at $\mathbf{Z}[x]$. In other words, in order to study polynomials with rational coefficients, we'll have to look at polynomials with integral coefficients. We can take a polynomial with rational coefficients and multiply it by the least common multiple of the denominators of its coefficients to get another polynomial with the same roots but with integral coefficients. We can also divide by the greatest common divisor of the resulting coefficients to get yet another polynomial with the same roots, with integral coefficients, and the greatest common divisor of all its coefficients is 1. Such a polynomial is called *primitive*.

After that, we'll be able to prove Gauss's lemma which says that a primitive polynomial $f(x) \in \mathbf{Z}[x]$ is reducible in $\mathbf{Q}[x]$ if and only if it's reducible in $\mathbf{Z}[x]$.

We can make more use of these results if, instead of considering just the case of the domain \mathbf{Z} and its field of fractions \mathbf{Q} , we generalize to any unique factorization domain D and its field of fractions F . So, for the following discussion, fix a UFD D , and let F denote its field

of fractions. Though, keep in mind the basic case when $D = \mathbf{Z}$, $F = \mathbf{Q}$, $D/(p) = \mathbf{Z}_p$, and $D/(p)[x] = \mathbf{Z}_p[x]$ to get a better idea of what's going on.

When we have a prime p in D , the projection $\gamma : D \rightarrow D/(p)$ induces a ring epimorphism $D[x] \rightarrow D/(p)[x]$ between polynomial rings where the coefficients of f are reduced modulo p giving a polynomial in $D/(p)[x]$. We'll denote the resulting polynomial in $D/(p)[x]$ by f_p .

Definition 3.57. The *content* of a polynomial in $D[x]$ is the greatest common divisor of all of its coefficients. If the content is 1, the polynomial is called *primitive*.

For example, if $f(x) = 3x^2 - 9x + 6$ then the content of f is 3. Also, the content of a monic polynomial is 1, so all monic polynomials are primitive.

The content of a polynomial is only defined up to a unit.

Evidently, every polynomial in $D[x]$ equals a constant times a primitive polynomial, the constant being its content.

Lemma 3.58 (Gauss). The product of two primitive polynomials in $D[x]$ is primitive, and the content of the product of any two polynomials in $D[x]$ is the product of their contents (up to a unit).

Proof. In order to show the first statement, we'll show if the product is not primitive, then one of the two polynomials is not primitive.

Let f and g be primitive polynomials and suppose that their product fg is not primitive. Then some prime p of D divides the content of fg , so p divides every coefficient of fg . Therefore, in $D/(p)[x]$, $(fg)_p = 0$, so $f_p g_p = 0$. But $D/(p)[x]$ is an integral domain (in fact, a UFD), so either $f_p = 0$ or $g_p = 0$. Therefore, p either divides all the coefficients of f or all the coefficients of g , hence one or the other is not primitive.

The second statement follows from the first just by using the fact that a polynomial equals its content times a primitive polynomial. Q.E.D.

Theorem 3.59 (Gauss's lemma). If a primitive polynomial in $D[x]$ can be factored as the product of two polynomials in $F[x]$, then it can be factored as the product of two polynomials in $D[x]$ of the same degrees.

Proof. Given $f \in D[x]$ as a product gh with $g, h \in F[x]$. We can write $gh = \frac{p}{q}uv$ where u and v are primitive polynomials in $D[x]$, and p and q are relatively prime integers. Then $qf = puv$. Since f is primitive, the content of qf equals the content of q . Since u and v are primitive, so is uv , and therefore the content of puv equals the content of p . Thus $p = q$, and they're both 1, and so $f = uv$. Note that the degrees of u and v are the same as the degrees of g and h , respectively. Q.E.D.

The following corollary is sometimes called Gauss's lemma. It follows directly from the above since monic polynomials are primitive.

Corollary 3.60. A monic polynomial in $D[x]$ is reducible over $F[x]$ if and only if it's reducible over $D[x]$.

There are irreducibility tests for polynomials with integer coefficients, so by this corollary, we'll be able to test irreducibility for polynomials with rational coefficients.

One test for irreducibility of polynomials with integer coefficients is to move to a quotient ring \mathbf{Z}_p . That also generalizes to any UFD D . If you can factor it in D , you can factor it in a quotient ring, at least if the leading term doesn't disappear in the quotient.

Theorem 3.61 (Modulo p irreducibility test.). Let p be a prime integer, and let f be a polynomial whose leading coefficient is not divisible by p . If f is reducible in $F[x]$, then f_p is reducible in $D/(p)[x]$. If f_p is irreducible in $D/(p)[x]$, then f is irreducible in $F[x]$.

Proof. Suppose f is reducible in $F[x]$. Then there exist $g, h \in D[x]$ such that $f = gh$ where the degrees of g and h are at least 1. Since $f = gh$, therefore, $f_p = g_p h_p$. Since p does not divide the leading coefficient of f , neither does it divide the leading coefficients of g or h . Therefore $\deg g_p = \deg g \geq 1$ and $\deg h_p = \deg h \geq 1$. Thus, f_p is reducible.

The last statement of the theorem is the contrapositive of the first statement. Q.E.D.

Example 3.62. Consider any cubic polynomial f in $\mathbf{Q}[x]$ with an odd leading coefficient, an odd constant, and one of the other two coefficients odd, for instance, $f(x) = 77x^3 + 15x^2 + 8x + 105$. By Gauss's lemma, it's reducible in $\mathbf{Q}[x]$ if and only if it's reducible in $\mathbf{Z}[x]$. To determine that, use the modulo 2 irreducibility test. For $f(x) = 77x^3 + 15x^2 + 8x + 105$, you'll get $f_2(x) = x^3 + x^2 + 1$. The resulting f_2 will have no roots in \mathbf{Z}_2 since it has three nonzero terms. A cubic polynomial with no roots is irreducible, so f_2 is irreducible in $\mathbf{Z}_2[x]$. Hence, f is irreducible in $\mathbf{Q}[x]$.

The converse of the mod p irreducibility test is not valid. A polynomial can be reducible mod p but irreducible in $\mathbf{Z}[x]$. Take $f(x) = 77x^3 + 15x^2 + 8x + 105$, for example, which we know is irreducible in $\mathbf{Z}[x]$. Modulo $p = 5$, however, it factors into linear factors: $f_5(x) \equiv 2x^3 - 2x = 2(x+1)x(x-1)$, so is reducible.

Exercise 44. Show the polynomial $f(x) = x^4 + x^3 + 2x^2 + 2x + 1$ is irreducible in $\mathbf{Q}[x]$. Hint: Consider it modulo 2. First check for roots, then see if it's divisible by a irreducible quadratic. There aren't many irreducible quadratics modulo 2; $x^2 + 1$ isn't since it factors as $(x+1)^2$ modulo 2. Neither are x^2 or $x^2 + x$ since they're both divisible by x .

Another useful irreducibility test is Eisenstein's criterion.

Theorem 3.63 (Eisenstein's criterion). Let $f \in D[x]$. If a prime p does not divide the leading coefficient of f , but it does divide all the other coefficients, and p^2 does not divide the constant of f , then f is irreducible in $F[x]$.

Proof. Suppose f is reducible. As in the previous theorem, there exist $g, h \in D[x]$ such that $f = gh$ where the degrees of g and h are at least 1. Reduce everything modulo p . Then $a_n x^n = f_p(x) = g_p(x)h_p(x)$ where a_n is the leading coefficient of f . Now $\mathbf{Z}_p[x]$ is a UFD, and since $f_p(x)$ is the unit a_n times the irreducible x raised to the n^{th} power, therefore x divides both $g_p(x)$ and $h_p(x)$. Therefore $g_p(0) = h_p(0) = 0$. That means that p divides the constant terms of both g and h , which implies p^2 divides the constant term of f , contrary to the assumption. Q.E.D.

Example 3.64. Consider the polynomial $f(x) = x^n - a$. As long as a has a prime factor that appears to the first power, then Eisenstein's criterion implies f is irreducible.

Exercise 45. Show that the polynomial $f(x) = x^n + 10x + 15$ is irreducible in $\mathbf{Q}[x]$

3.10.3 Prime cyclotomic polynomials

Cyclotomic polynomials were introduced in defintion 1.63.

For a prime p , the p^{th} cyclotomic polynomial is

$$\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + \cdots + x + 1.$$

We'll use Eisenstein's criterion to show Φ_p is irreducible, but not directly. First, we'll use a translation. Let

$$f(x) = \Phi_p(x + 1) = \frac{(x + 1)^p - 1}{x} = x^{p-1} + \binom{p}{p-1}x^{p-2} + \cdots + \binom{p}{2}x + \binom{p}{1}.$$

Then Eisenstein's criterion applies to f . Since f is irreducible, so is Φ_p .

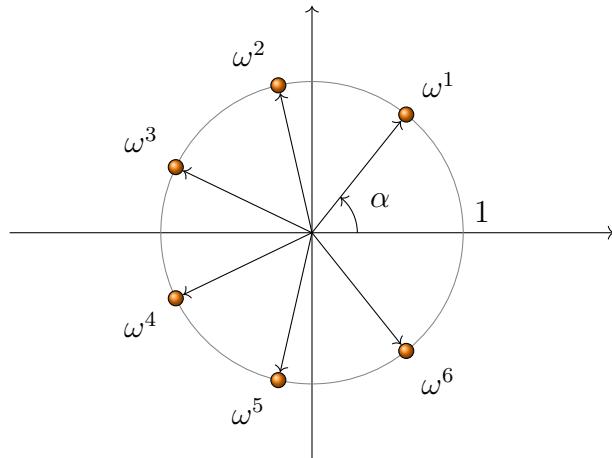


Figure 3.4: Primitive 7th roots of unity

The roots of Φ_p are the $p - 1$ primitive p^{th} roots of unity. In the case that $p = 7$, there are six primitive 7th roots of unity. One, labelled ω in figure 3.4, is located at an angle of $\alpha = 2\pi/7$, and the others are powers of it. There is one more 7th root of unity, namely 1. 1 is not primitive 7th root of unity but instead a primitive first root of unity since it's a root of the polynomial $x - 1$ of degree 1.

3.10.4 Polynomial rings with coefficients in a UFD, and polynomial rings in several variables.

Gauss's lemma has more uses than we've used it for. We can use it to show that if D is a UFD, then so is the polynomial ring $D[x]$. And we can apply that statement to conclude a polynomial ring $D[x, y]$ in two or $D[x_1, \dots, x_n]$ more variables is also a UFD. Although these rings are UFDs, they're not PIDs.

Theorem 3.65. Let D be a unique factorization domain and F its ring of fractions. Then $D[x]$ is also a UFD. The irreducible polynomials in $D[x]$ are either irreducible elements of D or have content 1 and are irreducible polynomials in $F[x]$.

Proof. Let f be a nonzero polynomial in $D[x]$. It is equal to its content times a primitive polynomial. Its content is an element of D , and, since D is a UFD, its content uniquely factors (up to a unit) as a product of irreducible elements of D .

We're reduced the proof to showing that that a primitive polynomial f in $D[x]$ of degree at least 1 uniquely factors as a product of irreducible polynomials.

Since f is a polynomial in $D[x]$, it's also a polynomial in $F[x]$, and we know $F[x]$ is a UFD being a polynomial ring with coefficients in a field F . Thus, f uniquely factors in F :

$$f(x) = f_1(x)f_2(x) \cdots f_k(x)$$

where each $f_i(x)$ is irreducible in $F[x]$. We only need to show that this factorization can be carried out in $D[x]$. Each polynomial $f_i(x)$ is a element a_i of F times a primitive polynomial $f'_i(x)$ in $D[x]$, so

$$f(x) = a_1 \cdots a_k f'_1(x) \cdots f'_k(x).$$

Since $f(x)$ is primitive and the product $f'_1(x) \cdots f'_k(x)$ is also primitive, therefore $a_1 \cdots a_k$ is a unit in D . Thus, $f(x)$ factors in $D[x]$. You can also show that it can factor in only one way in $D[x]$ since it only factors in one way in $F[x]$. Q.E.D.

Corollary 3.66. If D is a UFD, then a polynomial ring in several variables $D[x_1, x_2, \dots, x_r]$ with coefficients in D is also a UFD.

In general, these aren't PIDs. For example, $(2, x)$ is not a principal ideal in $\mathbf{Z}[x]$, and (x, y) is not a principal ideal in $\mathbf{Q}[x, y]$.

Irreducible polynomials and field extensions. We'll see irreducible polynomials in a field $F[x]$ correspond to maximal ideals. Quotient rings by ideals are fields if and only if the ideal is a maximal ideal, as shown in section 3.6.4. Therefore irreducible polynomials correspond to field extensions.

Theorem 3.67. The ideal generated by a polynomial with coefficients in a field is maximal if and only if the polynomial is irreducible over the field.

Proof. Let $f \in F[x]$. Suppose first that (f) is a maximal ideal of $F[x]$. We'll show that f can't be reducible. Suppose that f factors as gh in $F[x]$ where g and h have lower degrees than f , but neither g nor h is a unit. Then $(f) \subseteq (g) \subseteq F[x]$. Since (f) is maximal, therefore (g) is equal to either (f) or $F[x]$. Neither of these can occur, for if $(g) = (f)$, then they have the same degree; but if $(g) = F[x]$, then g is a unit. Therefore f is not a reducible polynomial.

Next suppose that f is an irreducible polynomial in $F[x]$. Let I be any ideal such that $(f) \subseteq I \subset F[x]$. Since $F[x]$ is a Euclidean domain (see section 3.8.4), it is also a principal ideal domain, so $I = (g)$ for some polynomial g . Therefore $f = gh$ for some $h \in F[x]$. But f is irreducible, so either g or h is a unit. But g is not a unit since $(g) = I \neq F[x]$, so h is a unit. Therefore $(f) = (g)$. So any proper ideal I that contains (f) is (f) itself. Thus (f) is a maximal ideal. Q.E.D.

Corollary 3.68. The quotient ring $F[x]/(f)$ of a polynomial ring over a field F by the ideal generated by an irreducible polynomial is a field extension of F .

This corollary follows directly the preceding theorem and the maximal ideal theorem, theorem 3.36, .

Example 3.69. We saw earlier that any cubic polynomial f in $\mathbf{Q}[x]$ with an odd leading coefficient, an odd constant, and one of the other two coefficients odd is irreducible. So, for example, $f(x) = x^3 - x - 1$ is irreducible over \mathbf{Q} . That means $K = \mathbf{Q}[x]/(f)$ is a field. This field K can also be denoted $K = \mathbf{Q}[x]/(x^3 - x - 1)$ since in the quotient, $x^3 - x - 1$ is 0. Rather than using the symbol x in the quotient, it would be better to have a different symbol so that x can still be used as our variable. Let's use w .

Then every element in it is of the form $aw^2 + bw + c$. Addition is done as usual for polynomials, and multiplication is as usual except whenever w^3 appears, it is replaced by $w+1$. For example, the product $(w^2 + 3w - 3)(2w^2 - 5) = 2w^3 + w^2 - 9w + 15 \equiv w^2 - 7w + 17$.

As this is a field, there are reciprocals of nonzero elements and division by nonzero elements. Finding reciprocals is not so easy. For example, to find the reciprocal of w , we need an element $aw^2 + bw + c$ such that its product with w equals 1. Now, $(aw^2 + bw + c)w = aw^3 + bw^2 + cw = bw^2 + (a+c)w + a$, so for that to equal 1, $a = 1$, $b = 0$, and $c = -1$. So the reciprocal is $w^{-1} = w^2 - 1$.

Exercise 46. Find an irreducible cubic polynomial in $\mathbf{Z}_2[x]$ to construct a field with eight elements. Write down a multiplication table for that field. You can leave out 0 and 1 from the table since it's obvious how they multiply, but have six rows and columns labeled $a = x$, $b = x + 1$, $c = x^2$, $d = x^2 + 1$, $e = x^2 + x$, and $f = x^2 + x + 1$.

3.11 Number fields and their rings of integers

In section 2.3 a *number field* K was defined a finite field extension of the rational number field \mathbf{Q} . Some examples were $\mathbf{Q}[\sqrt{2}]$ and $\mathbf{Q}[\sqrt{-1}] = \mathbf{Q}[i]$.

The Gaussian integers $\mathbf{Z}[i]$ is an example of what is called a ring of integers. We'll see in this section what a ring of integers is and study some of their properties.

A number field was defined to be an algebraic extension of \mathbf{Q} , and an algebraic integer was defined in section 2.3.1 to be a root of a monic polynomial with coefficients in the integers.

Our first goal is to show that the set of all algebraic integers in number field K , that set being denoted \mathcal{O}_K is a subring of K . That will take a few steps.

Define the minimal polynomial of an algebraic integer a is being that monic polynomial f in \mathbf{Q} of minimal degree such that $f(a) = 0$. Note that we're not requiring f to have coefficients in \mathbf{Z} , but we'll prove that below.

Lemma 3.70. The minimal polynomial of an algebraic integer divides every polynomial in $\mathbf{Q}[x]$ of which it is a root.

Proof. Let a be an algebraic integer with minimal polynomial f . By the division algorithm, there are polynomials q and r in $\mathbf{Q}[x]$ such that $g = qf+r$, where either $r = 0$ or $\deg r < \deg f$. Then $r(a) = g(a) - q(a)f(a) = 0$, so a is a root of r . Since f is the polynomial of least positive degree with root a , so $r = 0$. Q.E.D.

Lemma 3.71. The minimal polynomial of an algebraic integer has coefficients in \mathbf{Z} .

Proof. Let f be the minimal polynomial of a , and let g be a monic polynomial in $\mathbf{Z}[x]$ such that $g(a) = 0$. By the previous lemma, $g = fh$ for some $h \in \mathbf{Q}[x]$.

Suppose that $f \notin \mathbf{Z}[x]$, then some prime number p divides the denominator of some coefficient of f . Let p^i be the largest power of p dividing that denominator, so $i \geq 1$. Let p^j be the largest power of p that divides some denominator of a coefficient of h , with $j \geq 0$. Then $p^{i+j}g = (p^i f)(p^j h)$. Now take that equation modulo p . Modulo p , the left side is 0, but neither polynomial on the right side is 0, a contradiction since $\mathbf{Z}_p[x]$ is an integral domain. Q.E.D.

Theorem 3.72. The set of all algebraic integers, \mathcal{O}_K , in a number field K is a subring of that field.

This ring \mathcal{O}_K is called *the ring of integers* in the number field K .

Exercise 47. Prove that the Gaussian integers $\mathbf{Z}[i]$ is the ring of integers in the number field $\mathbf{Q}[i]$. Since i is the root of the monic polynomial $x^2 - 1$, all that's needed to prove is that there are no other integral elements in $\mathbf{Q}[i]$ other than those in $\mathbf{Z}[i]$.

Chapter 4

Groups

Recall that a group is a set equipped with one binary operation that is associative, has an identity element, and has inverse elements. If that binary operation is commutative, then the group is called an Abelian group.

4.1 Groups and subgroups

4.1.1 Definition and basic properties of groups

We'll look at basic properties of groups, and since we'll discuss groups in general, we'll use a multiplicative notation even though some of the example groups are Abelian.

Definition 4.1. The axioms for a group are very few. A group G has an underlying set, also denoted G , and a binary operation $G \times G \rightarrow G$ that satisfies three properties.

1. Associativity. $(xy)z = x(yz)$.
2. Identity. There is an element 1 such that $1x = x = x1$.
3. Inverses. For each element x there is an element x^{-1} such that $xx^{-1} = x^{-1}x = 1$.

Theorem 4.2. From these few axioms several properties of groups immediately follow.

1. Uniqueness of the identity. There is only one element e such that $ex = x = xe$, and it is $e = 1$.

Outline of proof. The definition says that there is at least one such element. To show that it's the only one, suppose e also has the property of an identity and prove $e = 1$.

2. Uniqueness of inverses. For each element x there is only one element y such that $xy = yx = 1$.

Outline of proof. The definition says that there is at least one such element. To show that it's the only one, suppose that y also has the property of an inverse of x and prove $y = x^{-1}$.

3. Inverse of an inverse. $(x^{-1})^{-1} = x$.

Outline of proof. Show that x has the property of an inverse of x^{-1} and use the previous result.

4. Inverse of a product. $(xy)^{-1} = y^{-1}x^{-1}$.

Outline of proof. Show that $y^{-1}x^{-1}$ has the property of an inverse of xy .

5. Cancellation. If $xy = xz$, then $y = z$, and if $xz = yz$, then $x = y$.
6. Solutions to equations. Given elements a and b there are unique solutions to each of the equations $ax = b$ and $ya = b$, namely, $x = a^{-1}b$ and $y = ba^{-1}$.
7. Generalized associativity. The value of a product $x_1x_2 \cdots x_n$ is not affected by the placement of parentheses.

Outline of proof. The associativity in the definition of groups is for $n = 3$. Induction is needed for $n > 3$.

8. Powers of an element. You can define x^n for nonnegative values of n inductively. For the base case, define $x^0 = 1$, and for the inductive step, define $x^{n+1} = xx^n$. For negative values of n , define $x^n = (x^{-n})^{-1}$.
9. Properties of powers. Using the definition above, you can prove using induction the following properties of powers where m and n are any integers: $x^m x^n = x^{m+n}$, $(x^m)^n = x^{mn}$.

Note that $(xy)^n$ does not equal $x^n y^n$ in general, although it does for Abelian groups.

4.1.2 Subgroups

A subgroup H of G is a group whose underlying set is a subset of the underlying set of G and has the same binary operation, that is, for $x, y \in H$, $x \cdot_H y = x \cdot_G y$ where \cdot_H denotes the multiplication in H while \cdot_G denotes the multiplication in G . Since they are the same, we won't have to subscript the multiplication operation.

An alternate description of a subgroup H is that it is a subset of G that is closed under multiplication, has 1, and is closed under inverses.

Of course, G is a subgroup of itself. All other subgroups of G , that is, those subgroups that don't have every element of G in them, are called *proper subgroups*.

Also, $\{1\}$ is a subgroup of G , usually simply denoted 1. It's called the *trivial subgroup* of G .

Example 4.3. Consider the cyclic group of six elements $G = \{1, a, a^2, a^3, a^4, a^5\}$ where $a^6 = 1$. Besides the trivial subgroup 1 and the entire subgroup G , there are two other subgroups of G . One is the 3-element subgroup $\{1, a^2, a^4\}$ and the other is the 2-element subgroup $\{1, a^3\}$.

The intersection $H \cap K$ of two subgroups H and K is also a subgroup, as you can easily show. Indeed, the intersection of any number of subgroups is a subgroup.

The union of two subgroups is never a subgroup unless one of the two subgroups is contained in the other.

Exercise 48. About intersections and unions of subgroups.

- (a). Show that the intersection of two subgroups is also a subgroup.
- (b). Give a counterexample where the union of two subgroups is not a subgroup.

Example 4.4 (Subgroups of \mathbf{Z}). Consider the group \mathbf{Z} under addition. A subgroup of \mathbf{Z} has to be closed under addition, include 0, and be closed under negation. Besides 0 and \mathbf{Z} itself, what are the subgroups of \mathbf{Z} ? If the subgroup is nontrivial, then it has a smallest positive element, n . But if n lies in a subgroup, then all multiples, both positive and negative, of n also must be in the subgroup. Thus, $n\mathbf{Z}$ is that subgroup of \mathbf{Z} .

Useful subgroups of a group. There are a number of other subgroups of a group that are important in studying nonabelian groups such as the center of a group and the centralizer of an element of a group.

Definition 4.5. Center and centralizer.

The *center* of a group G is $Z(G) = \{x \in G \mid ax = xa \text{ for all } a \in G\}$,

For $a \in G$, the *centralizer* of a is $Z_a(G) = \{x \in G \mid ax = xa\}$.

Exercise 49. Show the following properties about centers and centralizers.

- (a). Prove that $Z(G)$ is a subgroup of G .
- (b). Prove that the center of G is the intersection of all the centralizer subgroups of G .
- (c). Prove that $Z_a(G)$ is a subgroup of G .

Definition 4.6 (Commutator subgroup). The commutator of two elements x and y in a group G is the element $x^{-1}y^{-1}xy$. It is denoted $[x, y]$.

The subgroup of G generated by all the commutators of its elements is called the *commutator subgroup* of G , denoted G' .

Note that for an Abelian group, all the commutators are 1, and the the commutator subgroup is trivial.

If S is a subset of G , then there is a smallest subgroup $\langle S \rangle$ of G containing S . It can be described as the intersection of all subgroups H containing S ,

$$\langle S \rangle = \bigcap_{S \subseteq H} H.$$

Alternatively, it can be described as the subset of G of all products of powers of elements of S ,

$$\langle S \rangle = \{x_1^{e_1}x_2^{e_2} \cdots x_n^{e_n} \mid n \geq 0, \text{ each } x_i \in S, \text{ and each } e_i \in \mathbf{Z}\}.$$

4.1.3 Cyclic groups and subgroups

If a is an element of a group G , then the subset of G generated by a

$$\langle a \rangle = \{a^n \mid n \in \mathbf{Z}\}$$

is a subgroup of G . This subgroup generated by a is called a *cyclic subgroup* of G . If G itself is generated by some element a , then G is called a *cyclic group*.

Definition 4.7 (Order and involution). The *order* of a group G is the number of elements in it, that is, the cardinality of its underlying set. It's usually denoted $|G|$.

The *order* of an element a in a group is the smallest positive integer n such that $a^n = 1$. It's denoted $\text{ord } a$. If every positive power $a^n \neq 1$, then the order of n is defined to be ∞ . So, for example, the order of 1 is 1 since $1^1 = 1$.

An *involution* a is an element of a group which is its own inverse, $a^{-1} = a$. Clearly, the order of an involution a is 2 unless $a = 1$, in which case the order of a is 1.

Exercise 50. Prove that the order of a is also equal to the order of the cyclic group (a) generated by a . That is, $\text{ord } a = |\langle a \rangle|$.

An abstract cyclic group of order n is often denoted $C_n = \{1, a, a^2, \dots, a^{n-1}\}$ when the operation is written multiplicatively. It is isomorphic to the underlying additive group of the ring \mathbf{Z}_n where an isomorphism is $f : \mathbf{Z}_n \rightarrow C_n$ is defined by $f(k) = a^k$.

Exercise 51. Prove that any subgroup of a cyclic group is itself cyclic.

Exercise 52. Let G be a cyclic group of order n and a an element of G . Prove that a generates G , that is, $\langle a \rangle = G$, if and only if $\text{ord } a = n$.

Cyclic groups are all Abelian, since $a^n a^m = a^{m+n} = a^m a^n$. The integers \mathbf{Z} under addition is an infinite cyclic group, while \mathbf{Z}_n , the integers modulo n , is a finite cyclic group of order n .

Exercise 53. Prove that every cyclic group is isomorphic either to \mathbf{Z} or to \mathbf{Z}_n for some n .

Exercise 54. Prove that if k is relatively prime to n , then k generates \mathbf{Z}_n .

4.1.4 Products of groups

Just as products of rings are defined coordinatewise, so are products of groups. Using multiplicative notation, if G and H are two groups then $G \times H$ is a group where the product $(x_1, y_1)(x_2, y_2)$ is defined by $(x_1 x_2, y_1 y_2)$. The identity element in $G \times H$ is $(1, 1)$, and the inverse $(x, y)^{-1}$ is (x^{-1}, y^{-1}) . The projections $\pi_1 : G \times H \rightarrow G$ and $\pi_2 : G \times H \rightarrow H$ are group epimorphisms where $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$.

Also, $\iota_1 : G \rightarrow G \times H$ and $\iota_2 : H \rightarrow G \times H$ are group monomorphisms where $\iota_1(x) = (x, 1)$ and $\iota_2(y) = (1, y)$. Thus, we can interpret G and H as subgroups of $G \times H$.

Note that G and H are both Abelian groups if and only if $G \times H$ is an Abelian group. The product of two Abelian groups is also called their *direct sum*, denoted $G \oplus H$.

The underlying additive group of a ring is an Abelian group, and some of the results we have for rings give us theorems for Abelian groups. In particular, the Chinese remainder theorem for cyclic rings \mathbf{Z}_n gives us a theorem for cyclic groups C_n .

Theorem 4.8 (Chinese remainder theorem for groups). Suppose that $n = km$ where k and m are relatively prime. Then the cyclic group C_n is isomorphic to $C_k \times C_m$. More generally, if n is the product $k_1 \cdots k_r$ where the factors are pairwise relatively prime, then

$$C_n \cong C_{k_1} \times \cdots \times C_{k_r} = \prod_{i=1}^r C_{k_i}.$$

In particular, if the prime factorization of n is $n = p_1^{e_1} \cdots p_r^{e_r}$. Then the cyclic group C_n factors as the product of the cyclic groups $C_{p_i^{e_i}}$, that is,

$$C_n \cong \prod_{i=1}^r C_{p_i^{e_i}}.$$

4.1.5 Cosets and Lagrange's theorem

Cosets are useful in developing the combinatorics of finite groups, that is, for counting subgroups and other things related to a finite group. They come in both left and right forms as you'll see in the definition below, but we'll only use left cosets. Our first combinatorial theorem is called Lagrange's theorem which says that the order of a subgroup divides the order of a group. Since the subgroup $\langle a \rangle$ generated by a single element has an order that divides the order of the group, therefore the order of an element divides the order of the group, too. We'll have our first classification theorem as a corollary, and that is that a group whose order is a prime number is cyclic. Thus, up to isomorphism, there is only one group of that order.

Definition 4.9. Let H be a subgroup of G . A *left coset* of H is a set of the form

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

while a *right coset* is of the form $Ha = \{ha \mid h \in H\}$.

Theorem 4.10. Several properties of cosets follow from this definition.

1. The coset $1H$ is just the subgroup H itself. In fact, if $h \in H$ then $hH = H$.
2. More generally, $aH = bH$ if and only if $ab^{-1} \in H$. Thus, the same coset can be named in many different ways.
3. Cosets are disjoint. If $aH \neq bH$, then $aH \cap bH = \emptyset$.

Outline of proof. It's probably easier to show the contrapositive: if $aH \cap bH \neq \emptyset$ then $aH \neq bH$. Suppose an element is in the intersection. Then it can be written as ah or as bh' where both h and h' are elements of H . The rest relies on the previous statement.

4. Cosets of H all have the same cardinality.

Outline of proof. Check that the function $f(ah) = bh$ is a bijection $aH \rightarrow bH$.

5. Thus, the cosets of H partition G into subsets all having the same cardinality.
6. *Lagrange's theorem.* If G is a finite group, and H a subgroup of G , then $|H|$ divides $|G|$. Moreover, $|G|/|H|$ is the number of cosets of H .

Outline of proof. Follows from the preceding statement.

Definition 4.11. The *index* of a subgroup H of a group G is the number of cosets of H . The index is denoted $[G : H]$. By Lagrange's theorem, $[G : H] = |G|/|H|$ when G is a finite group.

Corollary 4.12. If the order of a group is a prime number, then the group is cyclic.

Proof. Let $|G| = p$, a prime. Since p has no divisors except 1 and p , therefore, by Lagrange's theorem, G only has itself and the trivial subgroup as its subgroups. Let $a \neq 1$ be an element of G . It generates a cyclic subgroup $\langle a \rangle$ which isn't trivial, so $\langle a \rangle = G$. Thus G is cyclic. Q.E.D.

Corollary 4.13. If a group is finite, then the order of every element divides the order of the group.

Proof. Let a be an element of a finite group G . Then the order of the subgroup $\langle a \rangle$ divides $|G|$. But $\text{ord } a$ is the order of $\langle a \rangle$. Therefore $\text{ord } a$ divides $|G|$. Q.E.D.

Products of subsets in a group. Occasionally we'll want to look at products HK of subsets H and K , especially when H and K are subgroups of a group G . This product is defined by

$$HK = \{xy \mid x \in H, y \in K\}.$$

Even when H and K are subgroups, it isn't necessary that HK is a subgroup, but there is a simple criterion to test if it is.

Abelian groups are often written additively. In that case, rather than using the notation HK , the notation $H + K$ is preferred: $H + K = \{x + y \mid x \in H, y \in K\}$.

Theorem 4.14. Let H and K be subgroups of G . Then HK is also a subgroup of G if and only if $HK = KH$.

Proof. \implies : Suppose that HK is a subgroup. We'll show that $KH \subseteq HK$. Let $xy \in KH$ with $x \in K$ and $y \in H$. Since $x = 1x \in HK$ and $y = y1 \in HK$, therefore their product xy is also in HK . Thus, $KH \subseteq HK$. Likewise $HK \subseteq KH$. Therefore $HK = KH$.

\impliedby : Suppose that $HK = KH$. To show it's a subgroup, first note $1 \in HK$ since $1 \in H$ and $1 \in K$.

Second, we'll show that HK is closed under multiplication. Let x_1y_1 and x_2y_2 be elements of HK with $x_1, x_2 \in H$ and $y_1, y_2 \in K$. Then $y_1x_2 \in KH = HK$, so $y_1x_2 = x_3y_3$ where $x_3 \in H$ and $y_3 \in K$. Therefore, $(x_1y_1)(x_2y_2) = (x_1x_3)(y_3y_2) \in HK$.

Third, we'll show that HK is closed under inverses. Let $xy \in HK$ with $x \in H$ and $y \in K$. Then $(xy)^{-1} = y^{-1}x^{-1} \in KH = HK$. Q.E.D.

Corollary 4.15. If H and K are subgroups of an Abelian group G , then $H + K$ is also a subgroup of G .

4.2 Symmetric Groups S_n

We've looked at several examples of groups already. It's time to examine some in more detail.

4.2.1 Permutations and the symmetric group

Definition 4.16. A *permutation* of a set X is just a bijection $\rho : X \rightarrow X$ on that set. The permutations on X form a group called the *symmetric group*. We're primarily interested in permutations on a finite set. We'll call the elements of the finite set letters, but we'll denote them with numbers. The symmetric group on n elements $1, 2, \dots, n$ is denoted S_n .

Note that the order of the symmetric group on n letters is $|S_n| = n!$.

Example 4.17. Consider the permutation ρ on set $X = \{1, 2, 3, 4, 5, 6\}$ that exchanges 2 with 4, sends 1 to 3, 3 to 5, and 5 to 1, and leaves 6 fixed. You can describe ρ in a table like this:

n	1	2	3	4	5	6
$\rho(n)$	3	4	5	2	1	6

That table has a lot of redundant information. The first row is just the names of the elements. To describe ρ on an ordered set like X , it's enough to list the elements in the second row: 3, 4, 5, 2, 1, 6. Unfortunately, that makes it harder to figure out where ρ sends an element. The cycle notation, mentioned next, is compact and makes it easier to see how ρ acts. For ρ , this notation will look like $(135)(24)$.

The three elements form a 3-cycle $1 \xrightarrow{\rho} 3 \xrightarrow{\rho} 5 \xrightarrow{\rho} 1$ of ρ denoted (135) . Also note $2 \xrightarrow{\rho} 4 \xrightarrow{\rho} 2$, so (24) is a 2-cycle of ρ . Another name for a 2-cycle is *transposition*. Since $\rho(6) = 6$, therefore (6) by itself is a 1-cycle, also called a *fixed point*, of ρ . The cycle notation for this permutation is $\rho = (135)(24)$. Note that fixed points are not denoted in this notation. Alternatively, this permutation could be denoted $(24)(135)$, or $(531)(42)$, or several other variants.

Since fixed points aren't denoted in cycle notation, we'll need a special notation for the identity permutation since it fixes all points. We'll use 1 to denote the identity since we're using 1 to denote the identity in a group written multiplicatively. In many textbooks the identity is denoted e .

There's a bit of experience needed to quickly multiply two permutations together when they're in cycle notation. Let $\rho = (146)(23)$ and $\sigma = (15)(2643)$. By $\rho\sigma$ mean first perform the permutation ρ then perform σ (in other words, the composition $\sigma \circ \rho$ if we think of these permutations as functions). Then we need simplify the cycle notation

$$\rho\sigma = (146)(23) (15)(2643).$$

Note that first ρ sends 1 to 4, then σ sends 4 to 3, therefore $\rho\sigma$ sends 1 to 3. Next $3 \xrightarrow{\rho} 2 \xrightarrow{\sigma} 6$, so $3 \xrightarrow{\rho\sigma} 6$, likewise $6 \xrightarrow{\rho} 1 \xrightarrow{\sigma} 5$, so $6 \xrightarrow{\rho\sigma} 5$, and $5 \xrightarrow{\rho} 4 \xrightarrow{\sigma} 1$, so $5 \xrightarrow{\rho\sigma} 1$. Thus, we have a cycle of $\rho\sigma$, namely, (1365) . You can check that (2) and (4) are fixed points of $\rho\sigma$. Thus, we found the product. $(146)(23) (15)(2643) = (1365)$.

Incidentally, finding the inverse of a permutation in cycle notation is very easy—just reverse all the cycles. The inverse of $\rho = (146)(23)$ is $\rho^{-1} = (641)(32)$.

Small symmetric groups When $n = 0$ or $n = 1$, there's nothing in the symmetric group except the identity.

The symmetric group on two letters, S_2 , has one nontrivial element, namely, the transposition (12) . This is the smallest nontrivial group, and it's isomorphic to any group of order 2. It is, of course, an Abelian group.

The symmetric group on three letters, S_3 , has order 6. We can name its elements using the cycle notation.

$$1, (12), (13), (23), (123), (132)$$

Besides the identity, there are three transpositions and two 3-cycles. This is not an Abelian group. For instance $(12)(13) = (123)$, but $(13)(12) = (132)$.

The symmetric group on four letters, S_4 , has order 24. Besides the identity, there are $\binom{4}{2} = 6$ transpositions, $\binom{4}{3} \cdot 2 = 8$ 3-cycles, 6 4-cycles, and 3 products of two 2-cycles, like $(12)(34)$.

Exercise 55. Complete the following table listing all 24 of the elements of S_4 .

the identity	1
transpositions	$(12), (13), (14), (23), (24), (34)$
3-cycles	
4-cycles	
products of 2 transpositions	

4.2.2 Even and odd permutations

First we'll note that every cycle, and therefore every permutation, can be expressed as a product of transpositions. We'll soon see after that that a permutation can either be expressed as a product of an even number of transpositions or as a product of an odd number of transpositions, but not both. That will justify the definition of even and odd permutations.

Theorem 4.18. Any cycle can be expressed as a product of transpositions.

Proof. The cycle $(a_1a_2a_3 \cdots a_k)$ is the product $(a_1a_2)(a_1a_3) \dots (a_1a_k)$. Q.E.D.

We'll look at an invariant that will help us distinguish even from odd permutations. It is P_n , the product of all differences of the form $i - j$ where $0 < i < j \leq n$.

$$\begin{aligned} P_n &= \prod_{0 < i < j \leq n} (i - j) \\ &= (1 - 2)(1 - 3) \cdots (1 - n) \\ &\quad (2 - 3) \cdots (2 - n) \\ &\quad \dots \\ &\quad ((n - 1) - n) \end{aligned}$$

Lemma 4.19. The effect of applying a transposition to the integers that make up P_n is to change the sign of P_n .

Proof. Let the transposition be (ab) where $0 < a < b \leq n$. The product P_n is made of three factors $P_n = P'P''P'''$ where $P' = (a - b)$, P'' is the product of factors that have either a or b but not both, and P''' is the product of factors that don't have either a or b . Now the transposition (ab) has no effect at all on P''' but negates P' . Its effect on P'' is more complicated. Suppose c is another letter.

Case 1. $c < a < b$. The factors $(c - a)$ and $(c - b)$ of P'' are interchanged by the transposition (ab) .

Case 2. $a < c < b$. The factors $(a - c)$ and $(c - b)$ are interchanged and both negated.

Case 3. $a < b < c$. Like case 1. Thus P'' does not change its value. Since only P' is negated, P_n is negated. Q.E.D.

Theorem 4.20. A permutation is either the product of an even number of transpositions or the product of an odd number of transpositions, but it can't be both.

Proof. Since each transposition negates P_n , the product of an even number of transpositions leaves P_n alone, but the product of an odd number of transpositions negates P_n . It can't be both since P_n is not 0. Q.E.D.

Definition 4.21. A permutation is *even* if it's the product of an even number of transpositions, it's *odd* if it's the product of an odd number of transpositions. The identity 1 is an even permutation.

Note that a cycle is an even permutation if it has an odd length, but it's an odd permutation if it has an even length.

Also, the product of two even permutations is even, the product of two odds is even, and the product of an even and an odd is odd.

Examples 4.22. The symmetric group S_3 has order 6. Its elements are 1, (12), (13), (23), (123), and (132). Three of them, namely 1, (123), and (132) are even while the other three (12), (13), and (23) are odd.

The symmetric group S_4 has 12 even permutations (the identity, eight 3-cycles, and three products of two 2-cycles) and 12 odd permutations (six transpositions and six 4-cycles).

4.2.3 Alternating and dihedral groups

Definition 4.23 (The alternating group A_n). Since the product of even permutations is even, and the inverse of an even permutation is even, therefore the set of even permutations in the symmetric group S_n is a subgroup of S_n . It is called the *alternating group* on n letters, denoted A_n .

For $n \geq 2$, the number of even permutations in S_n is the same as the number of odd permutations, since multiplying by the transposition (12) sets up the bijection. Therefore, the order of A_n is half the order of S_n . So $|A_n| = \frac{1}{n} n!$.

Example 4.24 (Subgroups of S_3). The symmetric group S_3 only has six elements, so it doesn't have many elements. There's the trivial subgroup 1 of order 1. There are three cyclic subgroups of order 2 each isomorphic to C_2 ; besides 1, the other element is one of the transpositions (12), (13) or (23). There's one subgroup of order three, namely, $D_3 = \{1, (123), (132)\}$. (Note that A_3 is the same group as D_3 . The Hasse diagram for the subgroups is fairly simple.

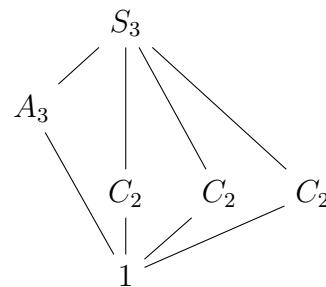


Figure 4.1: Subgroups of S_3

Example 4.25 (The dihedral group D_5). The dihedral groups D_n are the symmetry groups of regular n -gons. We already looked at the case $n = 3$ of an equilateral triangle. Consider a regular polygon with $n = 5$ vertices.

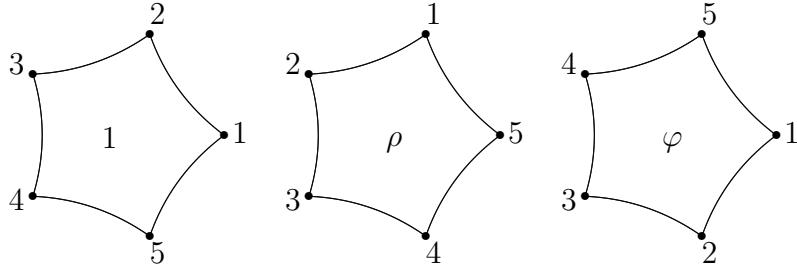


Figure 4.2: Symmetries of a pentagon

We can label the vertices in order from 1 to n . A symmetry of a plane figure is a transformation of the plane that maps the figure to itself. We're only interested in isometries, transformations that preserve distance, right now, but other transformations have their applications, too.

Figure 4.2 shows shows a pentagon. (The pentagon shown here is in the hyperbolic plane, but that doesn't matter.) One of its symmetries ρ is the one that rotates the pentagon 72° counterclockwise. It maps the vertex labelled 1 to 2, maps 2 to 3, and so forth. Knowing where the vertices are mapped is enough to determine the transformation, so we can identify ρ with the permutation it describes on the set of vertices. This ρ is the permutation (12345) .

Another of the symmetries of the pentagon is a reflection like φ shown above, a reflection across a horizontal axis. In cycle notation $\varphi = (25)(34)$.

In fact, there are 10 symmetries of the regular pentagon, so $|D_5| = 10$. In general $|D_n| = 2n$. In D_5 , besides the identity, there are four rotations and five reflections.

$$\begin{aligned} \text{identity} &= 1 & \rho &= (12345) & \rho^2 &= (13524) & \rho^3 &= (14253) & \rho^4 &= (15432) \\ \varphi &= (25)(34) & \varphi\rho &= (12)(35) & \varphi\rho^2 &= (13)(45) & \varphi\rho^3 &= (14)(23) & \varphi\rho^4 &= (15)(24) \end{aligned}$$

There are no more symmetries although we can write more expressions in terms of φ and ρ , for instance $\rho\varphi = (15)(24)$ which is $\varphi\rho^4$.

Thus, we can see now how to represent the dihedral group, D_5 , as a subgroup of the symmetric group S_5 . In fact, it's represented as a subgroup of the alternating group, A_5 as well, since all the permutations are even permutations.

Example 4.26 (Symmetries of a cube and tetrahedron). Consider a cube with vertices $1'2'3'4'1'23'4$ and the inscribed regular tetrahedron 1234 shown in figure 4.3. The four diagonals of the cube, $11'$, $22'$, $33'$, and $44'$, are drawn in green.

There are many symmetries of a tetrahedron. They permute the vertices 1234 . There are rotations of 120° and 240° about any of the four diagonals. Those rotations about the line $11'$ are the permutations (234) and (243) . The rotations about the other three diagonals are (123) , (132) , (124) , (142) , (134) , and (143) . Besides these rotations, there are three 180° rotations about the three lines joining the midpoints of the opposite edges of the tetrahedron. Along with the identity, that makes 12 permutations, all of which preserve orientation, that

is, they're rigid motions. The group of orientation preserving symmetries of the tetrahedron form the group S_4 .

Besides these, there are symmetries of the tetrahedron which are reflections across planes. They are orientation reversing symmetries. For example, the reflection across the plane passing through vertices 1 and 2 and the midpoint of edge 34 leaves vertices 1 and 2 fixed but it exchanges vertices 3 and 4; it's the transposition (34). The group of all the symmetries of the tetrahedron, including both the orientation preserving symmetries and the orientation reversing ones, form the group S_4 .

Each of the symmetries of the tetrahedron gives a symmetry of the enclosing cube. The symmetries of a cube permute its eight vertices 12341'2'3'4'. For example, the symmetry (123) of the tetrahedron gives the symmetry (123)(1'2'3') of the cube.

But there are other symmetries of a cube since the tetrahedron 1234 doesn't have to be preserved under a symmetry of the cube; it could be sent to the opposite tetrahedron 1'2'3'4'. Other orientation preserving symmetries that send the tetrahedron to the opposite tetrahedron include the six 90° and 270° rotations about the centers of the faces and the six 180° rotations about the line joining midpoints of opposite sides. That makes 24 orientation preserving symmetries for the cube. Each one permutes the four diagonals, and no two of them permute the four diagonals in the same way, so this symmetry group is S_4 .

Note that the symmetry (11')(22')(33')(44') that exchanges a vertex with its opposite vertex reverses orientation.

The entire group of symmetries of the cube includes the 24 orientation preserving symmetries and each of those times (11')(22')(33')(44'). That makes 48 symmetries of the cube.

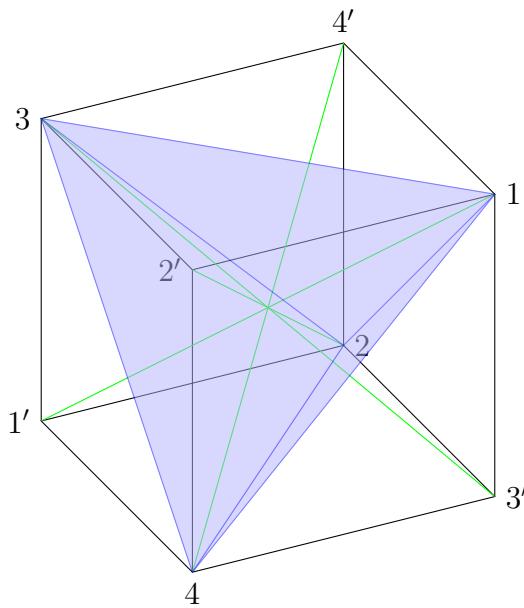


Figure 4.3: Symmetries of a cube and tetrahedron

Exercise 56. Verify the statements made in the example.

- (a). The orientation preserving symmetries of a tetrahedron form the group A_4 .
- (b). The group of all the symmetries of the tetrahedron form the group S_4 .

- (c). The orientation preserving symmetries of a tetrahedron form the group S_4 .
- (c). Explain why the permutation $(11')(22')(33')(44')$ of the cube reverses orientation.

Presentations by generators and relations. Although it's nice to have a group represented in a symmetric group, sometimes it's more convenient to describe it more algebraically in terms of generators and relations. For D_5 we can see that ρ and φ are sufficient to generate the whole group in the sense that every element in the group can be written as some expression involving ρ and φ . But there are certain relations, actually equations, that ρ and φ satisfy in this group, namely $\rho^5 = 1$, $\varphi^2 = 1$, and $\rho\varphi = \varphi\rho^{-1}$. Thus, we can present the group as

$$D_5 = \langle \rho, \varphi : \rho^5 = 1, \varphi^2 = 1, \rho\varphi = \varphi\rho^{-1} \rangle.$$

The difficulty with a presentation of this type is knowing when you have enough generators and relations. If you don't have enough generators, you won't generate the whole group. If you don't have enough relations, you'll generate a larger group, but not the one you want. A proof needs to be supplied to be assured that this is the right presentation. Frequently, a diagram of some sort fills the bill.

4.3 Cayley's theorem and Cayley graphs

One of the reasons symmetric groups are so important is that every group is isomorphic to a subgroup of a symmetric group, a result of Cayley. This gives us another way to look at groups, especially small finite ones.

We'll prove Cayley's theorem, then look at a few Cayley graphs which depend on Cayley's theorem.

4.3.1 Cayley's theorem

Recall that a permutation of a set X is just a bijection $\rho : X \rightarrow X$ on that set and permutations on X form a group called the *symmetric group* $S(X)$. When the set is finite, we can write it as $\{1, 2, \dots, n\}$, and S_n denotes its symmetric group.

Cayley's theorem can be stated for infinite groups as well as finite groups.

Theorem 4.27 (Cayley). Let G be a group, and let $S(G)$ be the symmetric group on G , that is, the group of permutations on the underlying set of G . The function $\varphi : G \rightarrow S(G)$ defined by $\varphi(a)(x) = ax$ is a group monomorphism. Therefore, G is isomorphic to a subgroup of $S(G)$.

Proof. $\varphi(a)$ is the permutation on G that maps x to ax . It's a bijection since its inverse sends x to $a^{-1}x$. To show that it's a group homomorphism, it is only necessary to show that $\varphi(ab) = \varphi(a)\varphi(b)$ for a and b in G . But $\varphi(ab)(x) = abx$, and $(\varphi(a)\varphi(b))(x) = \varphi(a)(\varphi(b)(x)) = \varphi(a)(bx) = abx$. Finally, $\varphi : G \rightarrow S(G)$ is a monomorphism since if $\varphi(a) = \varphi(b)$, then evaluating the two permutations at 1 gives $a1 = b1$, so $a = b$. Q.E.D.

Although this representation theorem does show that every group is a subgroup of a symmetric group (up to isomorphism), it's practically not all that useful since if the group G has order n , it's being represented in a group of order $n!$, which is much too large to deal with if n is at all large. Still, it's a useful representation for theoretical purposes.

Cayley graphs. With a Cayley graph we can represent a group G by a graph with vertices and labeled, directed edges. Each element of G is a vertex of the graph, and for each element a , we also have a directed edge labeled a from a vertex x to the vertex ax . In other words, the Cayley graph is a representation of G by the Cayley theorem to $S(G)$.

For a small example, let G be the cyclic group $G = \{1, a, b\}$ where $a^2 = b$ and $a^3 = 1$. The Cayley graph for G has three vertexes, labeled 1, a , and b . Each node has a loop on it labeled 1 since $1x = x$. There are three edges labelled a , $1 \xrightarrow{a} a \xrightarrow{a} b \xrightarrow{a} 1$, and three edges labelled b , $1 \xrightarrow{b} b \xrightarrow{b} a \xrightarrow{b} 1$. This is probably most conveniently drawn in a triangular figure.

There's a lot of redundancy in the graph in the sense that you don't need all the information to reconstruct the group. The loops labelled 1 might just as well be dropped since for any group $1x = x$. If we know the edges labelled a , then we can determine the edges labelled b since you just travel two a -edges to get a b -edge. That leaves just the triangle $1 \xrightarrow{a} a \xrightarrow{a} b \xrightarrow{a} 1$. More generally, if you know the edges for generators of a group, then all the other edges are determined.

Example 4.28 (D_5). Recall that the dihedral group D_5 has 10 elements and the presentation

$$D_5 = \langle \rho, \varphi : \rho^5 = \varphi^2 = (\varphi\rho)^2 = 1 \rangle.$$

The first relation, $\rho^5 = 1$ gives us a five cycle

$$1 \xrightarrow{\rho} \rho \xrightarrow{\rho} \rho^2 \xrightarrow{\rho} \rho^3 \xrightarrow{\rho} \rho^4 \xrightarrow{\rho} 1$$

which we can draw as a pentagon, the center pentagon in the graph below. The second relation, $\varphi^2 = 1$, means we have the 2-cycle $1 \xrightarrow{\varphi} \varphi \xrightarrow{\varphi} 1$, and, more generally, for any element a , we have a 2-cycle $a \xrightarrow{\varphi} a\varphi \xrightarrow{\varphi} a$. We'll draw 2-cycles as undirected edges $a \xleftarrow{\varphi} a$. We get five of these edges, one at each vertex of the center pentagon. The third relation, $(\varphi\rho)^2 = 1$, describes a square

$$a \xrightarrow{\varphi} a\varphi \xrightarrow{\rho} a\varphi\rho \xrightarrow{\varphi} a\varphi\rho\varphi \xrightarrow{\rho} a.$$

Starting at each of the new outer vertices of the graph, follow three edges to reach another outer vertex, and draw a ρ -edge back to where you started. When you finish, you have the Cayley graph for D_5 in figure 4.4

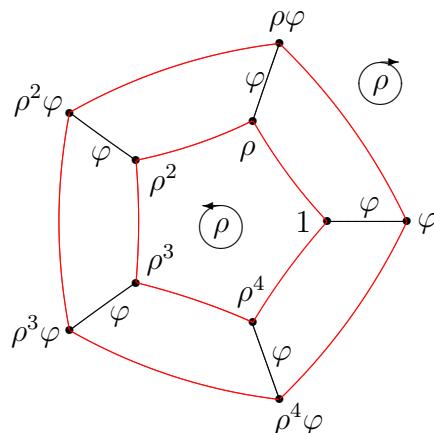


Figure 4.4: Cayley graph for D_5

Notice that the graph is completely symmetric. You could label any vertex 1 and fill in the names of the rest of the vertices by following the labelled arcs. For that reason, the vertices of a Cayley graph needn't be labelled.

There is another presentation for D_5 that gives a different looking Cayley graph. Let $\psi = \rho\phi$. Then

$$D_5 = \langle \varphi, \psi : \varphi = \psi^2 = (\varphi\psi)^5 \rangle.$$

The Cayley graph has the same ten vertices, but the edges are all undirected and they form a cycle of length 10 with labels alternating between φ and ψ .

Example 4.29 (A_4). Recall that the alternating group on $\{1, 2, 3, 4\}$ has 12 elements. It's not cyclic, so at least two generators are required to generate it. In fact, two will do. Consider the three elements

$$\begin{aligned} a &= (123) \\ b &= (124) \\ c &= ab = (14)(23) \end{aligned}$$

The two elements a and b are sufficient to generate A_4 as are the two elements a and c and many other pairs of elements (but not all pairs will do). In fact, A_4 can be represented in either of the following two ways:

$$\begin{aligned} \langle a, b : a^3 = b^3 = (ab)^2 = 1 \rangle \\ \langle a, c : a^3 = c^2 = (ac)^2 = 1 \rangle \end{aligned}$$

So, if we have the Cayley graph with only a - and b -edges, then we have enough information to determine A_4 , or if we have the graph with only a - and c -edges, then that's enough. Although these two graphs both have 12 vertices (since $|A_4| = 12$), they don't look very much alike. Let's look at the Cayley graph with all three kinds of edges, a -edges and b -edges and c -edges.

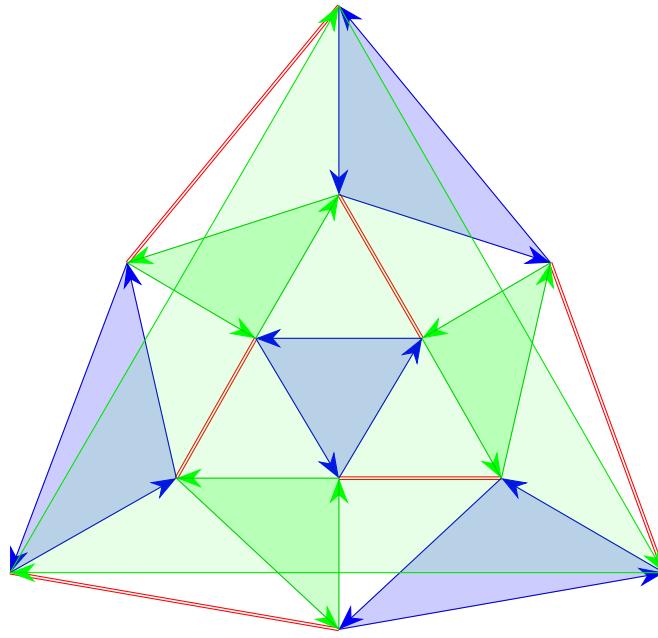
It's displayed in figure 4.5 as a planar graph, but more of the symmetry would be apparent if it were displayed in three dimensions where the vertices and edges were those of an icosahedron. Some of the triangles in the figure are blue. Their three sides of the triangle are a -edges. Likewise, some triangles are green with b -edges. Note that all the a - and b -triangles are oriented counterclockwise except the outer b -triangle. The remaining edges are the red c -edges, and to save space, since c is an involution, rather than putting in two edges, one pointing one way and the other pointing the other way, just a single thick undirected edge is included. Each vertex in the graph has an a -edge coming in and one coming out, a b -edge coming in and one coming out, and an undirected c -edge meaning that it goes both in and out.

Since it only takes two of these three elements to generate A_4 , this graph has superfluous information. All the edges labelled by one of the letters can be removed making the graph simpler.

Exercise 57. Find a Cayley graph for the symmetric group S_4 . There are various pairs or triples of generators you can use. One is the pair $a = (1234), b = (12)$.

4.3.2 Some small finite groups

We've seen a few families of finite groups including C_n the cyclic group of order n , D_n the dihedral group of order $2n$, S_n the symmetric group of order $n!$, and A_n the alternating group of order $n!/2$.

Figure 4.5: Cayley graph for A_4

The classification of finite groups (up to isomorphism, of course) is extremely difficult. Daniel Gorenstein (1923–1992) was a leader of mathematicians who eventually classified finite simple groups. He was faculty member at Clark University for 13 years.

We'll look at a few more small finite groups. Later, we'll look at the classification of finite Abelian groups, and find that they're all products of cyclic groups.

Table 4.1 lists the small groups up to isomorphism of order up through 24.

order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
number of groups	1	1	1	2	1	2	1	5	2	2	1	5	1	2	1

We won't prove that these are all of them, but we will look at them all. There are combinatorial theorems, the most important being the Sylow theorems, that help in classifying finite groups.

We know nearly all of these 27 groups. The cyclic groups C_n account for 15 of them. There are 12 others. Some of them are products of smaller ones, for instance, the other group of order 4 is $C_2 \oplus C_2$, sometimes called the *Klein 4-group*.

The second group of order 6 is D_3 , which is the same as S_3 .

Two of the groups of order 8 are products, namely, $C_4 \oplus C_2$ and $C_2 \oplus C_2 \oplus C_2$. Another is D_4 and the remaining one is called the quaternion group.

Example 4.30 (The quaternion group). This group consists of eight of the units of the division ring \mathbf{H} , the quaternions. Let $Q = \{\pm 1, \pm i, \pm j, \pm k\}$. Recall that the multiplication of quaternions has $i^2 = j^2 = k^2 = -1$, $ij = k$, $jk = i$, and $ki = j$, so this set of units is closed under multiplication and forms a group, called the *quaternion group*.

Exercise 58. Construct a Cayley graph for the quaternion group.

The second group of order 9 is the Abelian group $C_3 \oplus C_3$, and the second group of order 10 is D_5 . We already know the other groups of order 12: D_6 , $C_2 \oplus C_6$, $D_3 \times C_2$, and A_4 , and

Order	1	2	3	4	5	6	7	8
Abelian groups	C_1	C_2	C_3	C_4 $C_2 \oplus C_2$	C_5	C_6	C_7	C_8 $C_4 \oplus C_2$ $C_2 \oplus C_2 \oplus C_2$
Non-Abelian						D_3		D_4 Q
Order	9	10	11	12	13	14	15	16
Abelian groups	C_9	C_{10}	C_{11}	C_{12} $C_3 \oplus C_3$	C_{13}	C_{14}	C_{15}	C_{16} $C_2 \oplus C_8$ $C_2 \oplus C_2 \oplus C_4$ $C_2 \oplus C_2 \oplus C_2 \oplus C_2$ $C_4 \oplus C_4$
Non-Abelian		D_5		D_6 A_4 $C_3 \times C_4$		D_7		D_8 8 others
Order	17	18	19	20	21	22	23	24
Abelian groups	C_{17}	C_{18}	C_{19}	C_{20} $C_3 \oplus C_6$	C_{21}	C_{22}	C_{23}	C_{24} $C_2 \oplus C_{12}$ $C_2 \oplus C_2 \oplus C_6$
Non-Abelian		D_9 $S_3 \times C_3$ $C_3^2 \rtimes C_2$		D_{10} Dic_5 $C_5 \rtimes C_4$	$C_7 \rtimes C_3$	D_{11}		D_{12} 11 others

Table 4.1: List of small groups

the other group of order 14 is D_7 .

4.4 The category of groups \mathcal{G}

The category \mathcal{G} of groups was mentioned briefly in section 3.5 when the category of rings was introduced. The objects in this category are groups, and the morphisms are group homomorphisms.

Products in a category are defined by a universal property rather than by ordered pairs, but in \mathcal{G} , the product of two groups is what was called the the product of groups in section 4.1.4:

$$G \times H = \{(x, y) \mid x \in G, y \in H\}$$

Other categorical concepts include the initial and final object. In \mathcal{G} these are the same group, namely the trivial group that has only one element 1. There is a unique group homomorphism from each group to 1, and there's a unique group homomorphism from 1 to each group.

The universal property of an infinite cyclic group in the category of groups. The addition operation on \mathbf{Z} makes it an infinite cyclic group since each element in it is a multiple of 1. You can also write it multiplicatively as $C_\infty = \langle a \rangle = \{\dots, a^{-2}, a^{-1}, 1, a, a^2, \dots\}$.

This infinite cyclic group has the following universal property. Given any group G and any element $c \in G$, there is a unique group homomorphism $\langle a \rangle \rightarrow G$ that maps a to c . In general, it maps a^n to c^n . The image of this homomorphism is the subgroup of G generated by c .

Free groups. The infinite cyclic group $\langle a \rangle$ is a special case of a free group. It's a free group on one element. There are free groups on more than one element with analogous universal properties. We'll look at the free group on two elements.

Let a and b be two symbols. Form the group $\langle a, b \rangle$ as follows. An element in it is named by a string of a 's and b 's raised to various integral powers, such as $ba^{-3}b^2b^4a^2$. Different names are to be considered to name the same element if adjacent symbols are the same, in which case, they can be combined by the usual power rule. For example, $ba^{-3}b^2b^4a^2 = ba^{-3}b^6a^2$. Also, any symbol to the power 0 is to be treated as the identity element 1, and 1 times any string simplifies to that string.

A formal proof that $\langle a, b \rangle$ is, in fact, a group requires induction. We'll omit that proof.

This group $\langle a, b \rangle$ has the following universal property. Given any group G and any elements $c, d \in G$, there is a unique group homomorphism $\langle a, b \rangle \rightarrow G$ that maps a to c and b to d .

4.5 Conjugacy classes and quandles

We'll consider another way to examine the structure of groups. That depends on analyzing the operation of conjugation in a group.

Definition 4.31 (Conjugate element in a group). If x and y are elements of a group G , then yxy^{-1} is called *conjugates* of x . In that case, elements y and yxy^{-1} are said to be conjugates.

Exercise 59. Show that being an conjugate in a group is an equivalence relation by proving that (a) any element is conjugate to itself, (b) if one element is conjugate to second, then the second is conjugate to the first, and (c) if one element is conjugate to a second and the second conjugate to a third, then the first is conjugate to the third.

4.5.1 Conjugacy classes

Since being conjugates in a group is an equivalence relation, the corresponding equivalence classes can say a lot about the group.

Definition 4.32. Each of the equivalence classes of a group under conjugacy is called a *conjugacy class*, and the set of all conjugates of a particular element x is called the *conjugacy class* of x .

Exercise 60. If x is an element of order n in a group, show every conjugate of x also has order n .

Example 4.33 (Conjugacy classes in symmetric groups). Conjugation and conjugacy classes in symmetric groups are particularly easy to identify using cycle notation. Let $x = (13)(245)$ and $y = (142)$ be two elements in S_n . Then $y^{-1}xy = (124)(13)(245)(142) = (43)(125)$. Note how y conjugates the cycle (13) to the cycle (43) , and it conjugates the cycle (245) to (125) . The cycle structures for x and $y^{-1}xy$ are the same, but the elements in the cycles are permuted by y . This is generally the case for symmetric groups. It follows that a conjugacy class in S_n consists of all the elements in S_n with a given structure. Thus, for example, the conjugacy class of $(13)(235)$ consists of all elements of the form $(ab)(cde)$ where a, b, c, d , and e are 5 distinct integers between 1 and n . For S_5 the size of that conjugacy class is $\binom{5}{2} \cdot 2 = 20$.

Exercise 61. Determine all the conjugacy classes of S_5 and their sizes. (The sum of their sizes will equal 120, of course.)

Theorem 4.34. If H is a subgroup of G , and $x \in G$, then xHx^{-1} is also a subgroup of G , called a subgroup *conjugate* to H .

Proof. First, $1 \in xHx^{-1}$ since $x1x^{-1} = 1$. Next, if $xyx^{-1}, xzx^{-1} \in xHx^{-1}$ with $y, z \in H$, then their product $xyx^{-1}xzx^{-1} = x(yz)x^{-1} \in xHx^{-1}$. Finally, given $xyx^{-1} \in xHx^{-1}$ with $y \in H$, then the inverse $(xyx^{-1})^{-1} = xy^{-1}x^{-1} \in xHx^{-1}$. Therefore, xHx^{-1} is a subgroup of G . Q.E.D.

Similarly to the argument in the exercise above, being conjugate subgroups of a given group is an equivalence relation.

Theorem 4.35. If no other subgroup of G has the same order as H , then H is normal.

Proof. Since any conjugate subgroup xHx^{-1} is in one-to-one correspondence with H , it has the same number of elements, so must equal H . Q.E.D.

Exercise 62. If H is a subgroup of G and N is a normal subgroup of G , prove that $H \cap N$ is a normal subgroup of H .

Exercise 63. If H is a subgroup of G and N is a normal subgroup of G , prove that HN is a subgroup of G . (Hint: show $HN = NH$.)

Exercise 64. Prove that the intersection of two normal subgroups is also a normal subgroup.

Exercise 65. Prove that if H and N are normal subgroups of G , then their product is also a normal subgroup of G , in fact, it's the subgroup generated by $H \cup N$.

4.5.2 Quandles and the operation of conjugation

The operations of conjugation have certain properties. If we think of $y^{-1}xy$ as a binary operation $x \triangleright y$, and yxy^{-1} as another operation $x \triangleright^{-1} y$, then these two operations satisfy the properties stated in the next definition.

Definition 4.36. A *quandle* is a set equipped with two operations, \triangleright and \triangleright^{-1} satisfying the following three conditions for all elements x , y , and z .

- Q1.** $x \triangleright x = x$.
- Q2.** $(x \triangleright y) \triangleright^{-1} y = x = (x \triangleright^{-1} y) \triangleright y$.
- Q3.** $(x \triangleright y) \triangleright z = (x \triangleright z) \triangleright (y \triangleright z)$.

The symbol \triangleright is pronounced “through”, and \triangleright^{-1} “backthrough”.

Exercise 66. Prove that if Q is a conjugacy class in a group G then Q is a quandle where the operation $x \triangleright y$ is $y^{-1}xy$, and $x \triangleright^{-1} y$ is yxy^{-1} .

Involutory quandles. A quandle satisfying the identity $x \triangleright y \triangleright y = x$, equivalently $x \triangleright y = x \triangleright^{-1} y$, is called an *involutory quandle* or a *2-quandle*. The two operations of a quandle are the same in an involutory quandle.

There is an analogous definition for an n -quandle. First define $x \triangleright^n y$ as $x \triangleright y \triangleright \cdots \triangleright y$ where $\triangleright y$ occurs n times. An n -quandle is a quandle that satisfies the identity $x \triangleright^n y = x$.

A conjugacy class of involutions in a group is an involutory quandle, while the conjugacy class of an element of order n is an n -quandle.

Conjugacy classes of involutions are useful in the study of groups.

Besides conjugacy classes of groups, involutory quandles appear as cores of a group. The *core* of a group G has the same elements as G but with the operation $x \triangleright y = yx^{-1}y$.

Exercise 67. Prove that the core of a group is an involutory quandle.

Involutory quandles with geodesics. Involutory quandles have a nice geometric interpretation where the elements are points and the lines are determined by the operation.

Example 4.37 (The plane as a quandle). Consider the Euclidean plane \mathbf{R}^2 with the operation which sends a point p through a point q to yield the point $p \triangleright q$ on the line that passes through p and q and on the opposite side of q that p lies on but equally far away from q . If p happens to equal q , then define $p \triangleright q$ to be q . Algebraically, $p \triangleright q = 2q - p$.

This operation makes \mathbf{R}^2 an involutory quandle. The self distributivity axiom **Q3**, which says $(p \triangleright q) \triangleright r = (p \triangleright r) \triangleright (q \triangleright r)$, is illustrated in figure 4.6.

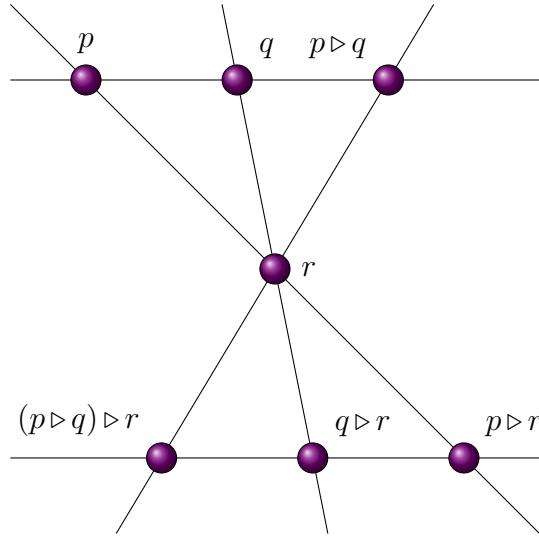


Figure 4.6: Distributivity in a involutory quandle

Symmetric spaces. A symmetric space is a particular kind of manifold. At each point in the space, there is an isometry (that is, a translation that preserves distance) for which that point is an isolated singularity. Ottmar Loos discovered in 1967 that the intrinsic algebraic structure of a symmetric is an involutory quandle. Thus, a symmetric space as a differentiable involutory quandle in which every point is an isolated fixed point of the symmetry through it.

Besides the plane \mathbf{R}^2 , every vector spaces V over any field is a symmetric spaces. The operation that makes it an involutory quandle is given by $\mathbf{v} \triangleright \mathbf{w} = 2\mathbf{w} - \mathbf{v}$. That's the same operation as described above for \mathbf{R}^2 .

There are lots of other symmetric spaces. The ordinary sphere S^2 as well as higher dimensional spheres S^n are all symmetric spaces. So are other geometric spaces including projective spaces, hyperbolic spaces, and inversive spaces. They can all be used to represent quandles geometrically as subspaces.

Geodesics. A *geodesic* in a manifold is a curve which for points close together is the curve of shortest length that joins them. In Euclidean space, a geodesic is a straight line. On the sphere S^2 , a geodesic is a great circle, that is, the intersection of a plane passing through the center of the sphere with the sphere. Geodesics on manifolds have metrics, that is, there's a distance between any two points on the geodesic.

Given two points p and q , the entire involutive quandle generated by them lies on one geodesic. That means that any other expression that can be made from p and q lie on a geodesic. In particular, the points $p \triangleright (q \triangleright p)$, $q \triangleright p$, p , q , $p \triangleright q$, and $q \triangleright (p \triangleright q)$ lie on a geodesic, and they're equally spaced on it.

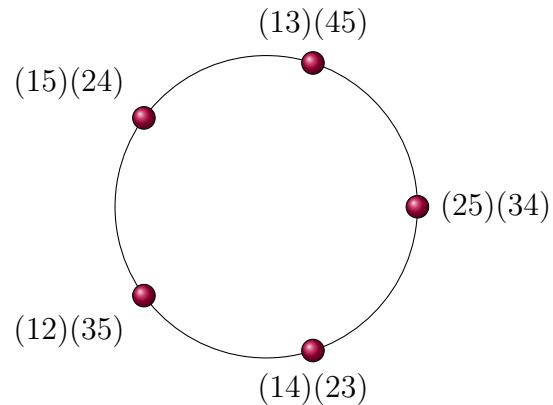


Since geodesics have all the information needed to describe the quandle structure, we can draw involutory quandles, at least the small ones, to see what they look like. Some will be

generated by two elements, so they'll either look like the line above, or be quotients of it.

Example 4.38 (Reflections of a pentagon). The group D_5 of symmetries of the pentagon was illustrated in figure 4.2.

The reflections of a pentagon are involutions, and they form a conjugacy class in D_5 . There are five reflections: $p = (25)(34)$, $q = (13)(25)$, $p \triangleright q = (15)(24)$, $q \triangleright p = (14)(23)$, and $p \triangleright (q \triangleright p) = q \triangleright (p \triangleright q) = (12)(35)$. Since $p \triangleright (q \triangleright p) = q \triangleright (p \triangleright q)$, the five of them lie on a circle as illustrated to the right.



Exercise 68. Let the seven vertices of a regular heptagon be denoted 1, 2, 3, 4, 5, 6, and 7. Describe the how the symmetry $(27)(36)(45)$ acts on the heptagon in words. Determine the conjugacy class of $(27)(36)(45)$ in D_7 and illustrate it as points on a circle.

Some conjugacy classes of involutions are cyclic like the ones in D_5 above, but most aren't. Here are two examples of 6-element conjugacy classes in small groups.

Example 4.39 (A conjugacy class in the quaternion group). The quaternion group was introduced in section 4.30. It has eight elements, namely 1, -1 , i , $-i$, j , $-j$, k , and $-k$. Six of them, all those except ± 1 , are involutions and they form a conjugacy class. It's illustrated in figure 4.7. Note that $i \triangleright j = jij = -i$, so i , j , $-i$, and $-j$ are equally spaced around a circle. Likewise for i , k , $-i$, and $-k$ and for k , j , $-k$, and $-j$. Although the spacing doesn't appear equal on the Euclidean plane as shown, it is when represented on a sphere.

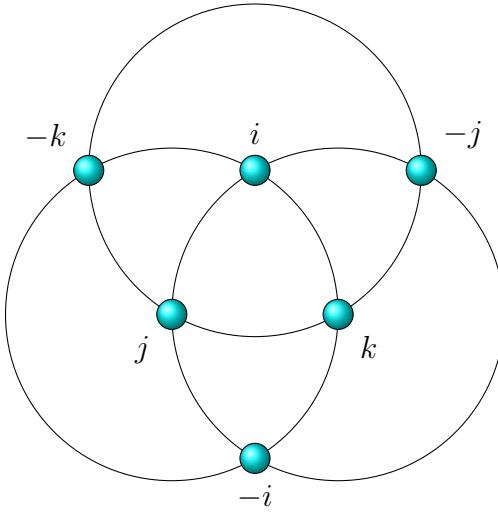


Figure 4.7: A conjugacy class in the quaternion group

Example 4.40 (The conjugacy class of transpositions in S_4). The transpositions in a symmetric group form a conjugacy class. In the symmetric group S_4 there are six transpositions,

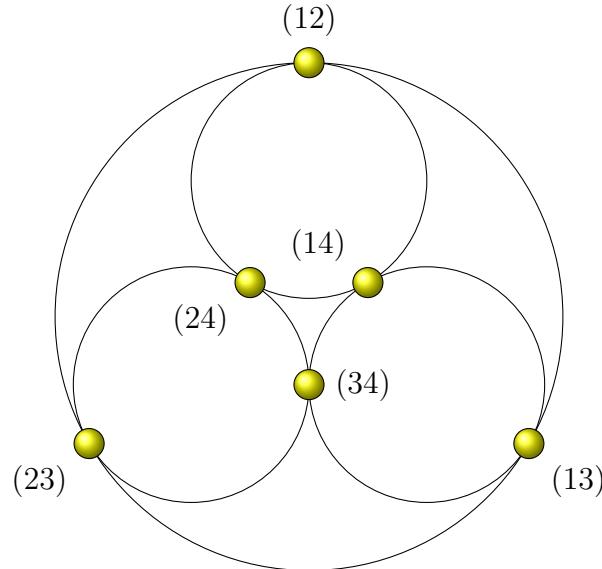


Figure 4.8: The conjugacy class of transpositions in S_4

namely, (12) , (13) , (14) , (23) , (24) , and (34) . The involutory quandle they form is shown in figure 4.8. There are four circles, each with three transpositions. One circle, for example, includes (12) , (13) , and (23) since $(12) \triangleright (13) = (23)$, $(13) \triangleright (23) = (12)$, and $(23) \triangleright (12) = (13)$. Also note that $(12) \triangleright (34) = (12)$, but no geodesic is shown having those two transpositions. It reduces the clutter in the diagram to suppress geodesics with only two elements.

4.6 Kernels, normal subgroups, and quotient groups

The kernel $\text{Ker } f$ of a group homomorphism $f : G \rightarrow H$ plays the same role as the kernel of a ring homomorphism. It's defined as the inverse image of the identity. It is a subgroup of the domain G , but a particular kind of subgroup called a normal subgroup. We'll see that every normal subgroup N of G is the kernel of some group homomorphism, in fact, of a projection $G \rightarrow G/N$ where G/N is a quotient group of G .

4.6.1 Kernels of group homomorphisms and normal subgroups

We'll use multiplicative notation.

Definition 4.41. Let $f : G \rightarrow H$ be a group homomorphism. Those elements of G that are sent to the identity 1 in H form the *kernel* of f .

$$\text{Ker } f = f^{-1}(1) = \{x \in G \mid f(x) = 1\}.$$

Example 4.42. Let G be the symmetric group S_n and $f : G \rightarrow \{1, -1\}$ map even permutations to 1 and odd permutations to -1 . Then f is a group homomorphism, and $\text{Ker } f = A_n$, the alternating subgroup of S_n .

Theorem 4.43. The kernel of a group homomorphism $f : G \rightarrow H$ is a subgroup $N = \text{Ker } f$ of G such that for each $x \in G$, $xN x^{-1} \subseteq N$.

Proof. To show that N is a subgroup of G , note that (1) it's closed under multiplication, (2) it includes 1, and (3) it's closed under inverses. For (1), if $x, y \in N$, then $f(x) = f(y) = 1$, so $f(xy) = f(x)f(y) = 1$, therefore $xy \in N$. (2) is obvious. For (3), if $x \in N$, then $f(x) = 1$, so $f(x^{-1}) = f(x)^{-1} = 1^{-1} = 1$, therefore $x^{-1} \in N$. Thus N is a subgroup of G .

Now to show that for $x \in G$, $xNx^{-1} \subseteq N$. Consider xyx^{-1} where $y \in N$. Then $f(y) = 1$, so $f(xyx^{-1}) = f(x)f(y)f(x)^{-1} = f(x)1f(x)^{-1} = f(x)f(x)^{-1} = 1$. Therefore, $xyx^{-1} \in N$. Thus, $xNx^{-1} \subseteq N$. Q.E.D.

Besides telling us what elements are sent to 1 by f , the kernel of f also tells us when two elements are sent to the same element. Since $f(x) = f(y)$ if and only if $f(xy^{-1}) = 1$, therefore, f will send x and y to the same element of S if and only if $xy^{-1} \in \text{Ker } f$.

The properties of kernels of group homomorphisms that we just found determine the following definition.

Definition 4.44. A subgroup N of a group G is said to be a *normal subgroup* if for each $x \in G$, $xNx^{-1} \subseteq N$.

Note that since a normal subgroup N of a group G is closed under conjugation, therefore a normal subgroup of G is the union of some of the conjugacy classes in G .

Exercise 69. Show that a subgroup N is normal in G if and only if for each $x \in G$, $xNx^{-1} = N$.

Exercise 70. Show that a subgroup N is normal in G if and only if for each $x \in G$, $xN = Nx$.

Both the trivial subgroup of G and G itself are always normal subgroups.

If G is an Abelian group, then every subgroup of G is a normal subgroup.

Theorem 4.45. Any subgroup of index 2 is a normal subgroup.

Proof. Let N be a subgroup of a group G of index 2. We'll show that $xN = Nx$ for each $x \in G$. In case $x \in N$, then $xN = N = Nx$. Now consider the case $x \notin N$. Then there are two left cosets of N , namely N itself and xN , and there are two right cosets, N and Nx . That gives us two partitions of G , but since N is a part of each partition, the other parts, namely xN and Nx must be equal. Q.E.D.

4.6.2 Quotient groups, and projections $\gamma : G \rightarrow G/N$

As mentioned above the kernel of a group homomorphism f tells us when two elements are sent to the same element: $f(x) = f(y)$ if and only if $xy^{-1} \in \text{Ker } f$. We can use $\text{Ker } f$ to construct a “quotient group” $G/\text{Ker } f$ by identifying two elements x and y in G if xy^{-1} lies in $\text{Ker } f$. In fact, we can do this not just for kernels of homomorphisms, but for any normal subgroup N . That is, we can use a normal subgroup N of G to determine when two elements x and y are to be identified, $x \equiv y$, and we'll end up with a group G/N .

Definition 4.46. A *congruence* \equiv on a group G is an equivalence relation such that for all $x, x', y, y' \in G$,

$$x \equiv x' \text{ and } y \equiv y' \text{ imply } xy \equiv x'y'.$$

The equivalence classes for a congruence are called *congruence classes*.

Theorem 4.47. If \equiv is a congruence on a group G , then the quotient set G/\equiv , that is, the set of congruence classes, is a group where the binary operation is defined by $[x][y] = [xy]$.

Proof. First we need to show that the proposed definitions are actually well defined. That is, if a different representative x' is chosen from the congruence class $[x]$ and y' from $[y]$, then the same class $[x'y']$ results. That is

$$[x] = [x'] \text{ and } [y] = [y'] \text{ imply } [xy] = [x'y'].$$

But that is the requirement in the definition of congruence.

Also, each of the axioms for a group need to be verified, but they're all automatic as they're inherited from the group G . Q.E.D.

Just as an ideal in a ring determines a congruence on the ring, a normal subgroup of a group determines a congruence on a group, and the proof is similar.

Theorem 4.48 (Congruence modulo a normal subgroup). Let N be a normal subgroup of a group G . A congruence, called *congruence modulo N* , is defined by

$$x \equiv y \pmod{N} \text{ if and only if } xy^{-1} \in N.$$

The quotient group, G/\equiv , is denoted G/N . The congruence classes are cosets of N , that is $[x] = xN$. The function $\gamma : G \rightarrow G/N$ defined by $\gamma(x) = [x] = xN$ is a group homomorphism, in fact, an epimorphism. It's called a *projection* or a *canonical homomorphism* to the quotient group. Its kernel is N .

Exercise 71. If \equiv is a congruence on a group G , show that the congruence class of the identity, $[1] = N$, is a normal subgroup of G , and the congruence determined by N is the original congruence.

4.6.3 Isomorphism theorems

The image of a group homomorphism is isomorphic to the group modulo its kernel. Let $f : G \rightarrow H$ be a ring homomorphism. The image of f , denoted $f(G)$, is the set

$$f(G) = \{f(x) \in H \mid x \in G\}.$$

Exercise 72. Verify that the image $f(G)$ is a subgroup of H .

Exercise 73. Prove the following theorem. You'll need to show that the proposed function is well-defined, that it is a group homomorphism, and then that it's an isomorphism.

Theorem 4.49 (First isomorphism theorem, Jordan, 1870). If $f : G \rightarrow H$ is a group homomorphism then the quotient group $G/\text{Ker } f$ is isomorphic to the image ring $f(G)$, the isomorphism being given by

$$\begin{aligned} G/\text{Ker } f &\rightarrow f(G) \\ x \text{Ker } f &\mapsto f(x) \end{aligned}$$

This gives us two ways to look at the image, either as a quotient group of the domain G or as a subgroup of the codomain H .

Furthermore, we can now treat a group homomorphism $f : G \rightarrow H$ as a composition of three group homomorphisms.

$$G \xrightarrow{\gamma} G/\text{Ker } f \cong f(G) \xrightarrow{\iota} H$$

The first is the projection from G onto its quotient ring $G/\text{Ker } f$, the second is the isomorphism $G/\text{Ker } f \cong f(G)$, and the third is the inclusion of the image $f(G)$ as a subgroup of H .

Theorem 4.50 (Second isomorphism theorem). If H is a subgroup of G and N is a normal subgroup of G , then

$$H/(H \cap N) \cong (HN)/N.$$

Proof. Let $f : H \rightarrow (HN)/N$ be defined by $f(x) = xN$. This f is a group homomorphism since $f(xy) = xyN = xNyN = f(x)f(y)$.

Next, we'll show that f is an epimorphism. Let $xN \in (HN)/N$ where $x \in HN$. Then $x = yz$ for some $y \in H$ and $z \in N$. So $xN = yzN = yN = f(y)$. Thus, f is an epimorphism, that is, $f(H) = (HN)/N$. by the first isomorphism theorem, we have

$$H/\text{Ker } f \cong (HN)/N.$$

Finally, we'll show that $\text{Ker } f = H \cap K$ which will imply $H/(H \cap N) \cong (HN)/N$. Let x be an element of H which lies in $\text{Ker } f$. Then xN is the identity element N in $(HN)/N$, so $x \in N$. But $x \in H$ also, so $x \in H \cap N$. Conversely, $x \in H \cap N$ implies $x \in \text{Ker } f$. Q.E.D.

Theorem 4.51 (Third isomorphism theorem). If H and K are both normal subgroups of G with $H \subseteq K$, then

$$(G/H)/(K/H) \cong G/K.$$

Exercise 74. Prove the third isomorphism theorem. Define $f : G/H \rightarrow G/K$ by $f(aH) = aK$. Check that this is a well-defined homomorphism. Show $\text{Ker } f = H$. Show the image of f is all of G/K . Apply the first isomorphism theorem to finish the proof.

Theorem 4.52 (Correspondence theorem). Let N be a normal subgroup of G . The subgroups of G containing N are in one-to-one correspondence with the subgroups of G/N . Thus, if H is a subgroup of G containing N , then H/N is a subgroup of G/N , and every subgroup of G/N so arises. Furthermore, H is normal in G if and only if H/N is normal in G/N .

Exercise 75. Prove the correspondence theorem. Show that for $H \supseteq N$ that H/N is, indeed, a subgroup of G/N . Show that if \bar{H} is any subgroup of G/N that the set $H = \{x \in G \mid x/N \in \bar{H}\}$ is a subgroup of G containing N . Verify that these two operations are inverse to each other. Finally, verify the last statement.

4.6.4 Internal direct products

We can recognize when a group G is isomorphic to a product of two or more groups. Recall that if $G = M \times N$, then we can interpret M and N as subgroups of G . As such they are normal subgroups of G and their intersection is trivial. Furthermore, $G = MN$.

Definition 4.53. A group G is said to be an *internal direct product* of two subgroups M and N if $M \cap N = 1$, $MN = G$, and both M and N are normal subgroups of G .

We'll show in a moment that if G is the internal direct product of M and N , then G is isomorphic to the product group $M \times N$. But first, a lemma.

Lemma 4.54. If M and N are two normal subgroups of G whose intersection is trivial, then elements of M commute with elements of N .

Proof. Let $m \in M$ and $n \in N$. In order to show that $mn = nm$, we'll show the equivalent $mnm^{-1}n^{-1} = 1$. Let $x = mnm^{-1}n^{-1}$. Since $x = (mnm^{-1})n^{-1}$, and both mnm^{-1} and n^{-1} are elements of the normal subgroup N , therefore $x \in N$. But since $x = m(nm^{-1}n^{-1})$, and both m and $nm^{-1}n^{-1}$ are elements of the normal subgroup M , therefore $x \in M$. Since $x \in M \cap N = 1$, therefore $x = 1$. Q.E.D.

Theorem 4.55. If G is the internal direct product of M and N , then $M \times N \cong G$ where the isomorphism is given by $(m, n) \mapsto mn$.

Proof. Outline. Use the lemma to verify that the proposed isomorphism is a homomorphism. It's evidently a surjection since $MN = G$. To show that it's an injection, show that the kernel is trivial. Suppose $(m, n) \mapsto mn = 1$. Then $m = n^{-1}$ lies in both M and N , so it's trivial, that is, $m = n = 1$. Q.E.D.

Exercise 76. Prove that G is an internal direct product of two normal subgroups M and N if and only if every element $x \in G$ can be uniquely represented as a product mn with $m \in M$ and $n \in N$.

Although we've only looked at internal direct products of two subgroups, the definition can be generalized to more than two subgroups. We'll say that G is the *internal direct product* of r normal subgroups N_1, N_2, \dots, N_r if (1) they jointly generate G , that is, $N_1N_2 \cdots N_r = G$, and (2) the intersection of any one N_i with the subgroup generated by the rest is trivial. It follows that $N_1 \times N_2 \times \cdots \times N_r \cong G$. Furthermore, an equivalent condition to being a internal direct product of the normal subgroups N_1, N_2, \dots, N_r is that every element $x \in G$ can be uniquely represented as a product $n_1n_2 \cdots n_r$ with each $n_i \in N_i$.

4.7 Matrix rings and linear groups

The representation of rings and groups as subrings or subgroups of matrix rings is very helpful for a couple of reasons. One is that matrices describe linear transformations. That means that the elements of the ring or group can be interpreted as geometric transformations. A second is that matrix notation is so very convenient. Usually the coefficients are taken to be elements of a familiar field like **C**, **R**, or **Q**, but for special purposes the coefficients may be taken in some other integral domain such as **Z**.

For example, the field complex numbers **C** can be represented as a certain subring of $M_2(\mathbf{R})$, the ring of 2×2 matrices with coefficients in **R**, and the division ring of quaternions **H** can be represented as a certain subring of $M_4(\mathbf{R})$.

Most of our examples have n equal to 2 or 3 and the coefficients are real.

4.7.1 Linear transformations

The ring of $n \times n$ matrices with real coefficients, $M_2(\mathbf{R})$, is a noncommutative ring when $n \geq 2$. We can interpret each matrix $A \in M_2(\mathbf{R})$ as a linear transformation $A : \mathbf{R}^n \rightarrow \mathbf{R}^n$

where a (column) n -vector $\mathbf{x} \in \mathbf{R}^n$ is mapped to another n -vector

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{bmatrix}$$

The identity matrix

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

corresponds to the identity transformation $I : \mathbf{R}^n \rightarrow \mathbf{R}^n$ where $I\mathbf{x} = \mathbf{x}$.

A linear transformation from a vector space to itself is also called a linear operator.

4.7.2 The general linear groups $GL_n(R)$

The invertible $n \times n$ matrices in $M_n(R)$, that is, the units in the ring $M_n(R)$, form the *general linear group* with coefficients in the commutative ring R , denoted $GL_n(R)$. They describe nonsingular transformations $R^n \rightarrow R^n$. Recall that a matrix A has an inverse if and only if its determinant $|A|$ is a unit in R .

Let's interpret some of these in the case when $n = 2$. The determinant of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $|A| = ad - bc$, and when that's a unit in R , the inverse of A is $A^{-1} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Note that the determinant is a group homomorphism $GL_n(R) \rightarrow R^*$ from the general linear group to the invertible elements of R . The determinant of the identity matrix is 1, the determinant of the product of two matrices is the product of their determinants, and the determinant of the inverse of a matrix is the reciprocal of the determinant of the matrix.

Let's let R be the field of real numbers \mathbf{R} . The real general linear group $GL_2(\mathbf{R})$ can be interpreted as the group of invertible linear transformations of the plane \mathbf{R}^2 that leave the origin fixed. Here are a few linear transformations of the plane.

Rotation by an angle θ about the origin is described by the matrix

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

since a point $\begin{bmatrix} x \\ y \end{bmatrix}$ in \mathbf{R}^2 is sent to the point $\begin{bmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{bmatrix}$. The determinant of a rotation matrix is 1.

Reflection across a line through the origin at an angle θ to the x -axis is described by the matrix

$$\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}.$$

The determinant is -1 .

Expansions and contractions are described by scalar matrices $\begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$ where r is the ratio. If $r > 1$, then it's an expansion (also called dilation), but if $0 < r < 1$, then it's a contraction.

There are numerous other kinds of transformations. Here's just one more example $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, an example of a shear parallel to the x -axis. Points above the x -axis are moved right, points below left, and points on the x -axis are fixed.

In three dimensions you can describe rotations, reflections, and so forth, as well.

4.7.3 Other linear groups

There are a number of interesting subgroups of $GL_n(R)$.

The special linear groups $SL_n(R)$. There are several subgroups of $GL_n(R)$, one of which is the special linear group $SL_n(R)$ which consists of matrices whose determinants equal 1, also called *unimodular* matrices. (There are other linear groups called “special” and in each case it means the determinant is 1.)

Among the examples in $GL_2(\mathbf{R})$ mentioned above, the rotations and shears are members of $SL_2(\mathbf{R})$, but reflections have determinant -1 and expansions and contractions have determinants greater or less than 1, so none of them belong to the special linear group.

Since the absolute value of the determinant is the Jacobian of the transformation $\mathbf{R}^n \rightarrow \mathbf{R}^n$, therefore transformations in $SL_2(\mathbf{R})$ preserve area. Since the determinant is positive, these transformations preserve orientation. Thus, transformations in $SL_2(\mathbf{R})$ are the linear transformations that preserve orientation and area. More generally those in $SL_n(\mathbf{R})$ preserve orientation and n -dimensional content. Rotations and shears, and their products, are always in $SL_n(\mathbf{R})$.

Exercise 77. Show that the matrix $\begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}$ lies in $SL_2(\mathbf{R})$. Describe in words how this transformation acts on the plane.

The orthogonal groups $\mathcal{O}(n)$. These are subgroups of $GL_n(\mathbf{R})$. An *orthogonal* transformation is one that preserves inner products (also called dot products or scalar products).

I'll use the notation

$$\langle \mathbf{a} | \mathbf{b} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

for the inner product of the vectors $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$. Other common notations are (\mathbf{a}, \mathbf{b}) or $\mathbf{a} \cdot \mathbf{b}$.

For the transformation described by the matrix A to preserve inner products means that $\langle A\mathbf{a} | A\mathbf{b} \rangle = \langle \mathbf{a} | \mathbf{b} \rangle$. Since the length of a vector $|\mathbf{a}|$ is determined by the inner product, $|\mathbf{a}|^2 = \langle \mathbf{a} | \mathbf{a} \rangle$, therefore an orthogonal transformation preserves distance, too: $|A\mathbf{a}| = |\mathbf{a}|$. Conversely, if A preserves distance, it preserves inner products.

Note that since distance is preserved, so is area in dimension 2 or n -dimensional content in dimension n .

It's a theorem from linear algebra that a matrix A describes an orthogonal transformation if and only if its inverse equals its transform: $A^{-1} = A^T$; equivalently, $AA^T = 1$. These ma-

trices, of course, are called *orthogonal* matrices. Note that the determinant of an orthogonal matrix is ± 1 .

The orthogonal group $\mathcal{O}(n)$ is the subgroup of $GL_n(\mathbf{R})$ of orthogonal matrices. It's not a subgroup of $SL_n(\mathbf{R})$ since half the orthogonal matrices have determinant -1 , meaning they reverse orientation. The special orthogonal group $S\mathcal{O}(n)$ is the subgroup of $\mathcal{O}(n)$ of matrices with determinant 1 .

In two dimensions $\mathcal{O}(2)$ consists of rotations and reflections while $S\mathcal{O}(n)$ consists of only the rotations. In three dimensions $\mathcal{O}(3)$ consists of rotations (by some angle around some line through 0) and reflections (across some plane through 0). Again, $S\mathcal{O}(3)$ only has the rotations.

The unitary groups $\mathcal{U}(n)$. For matrices with complex coefficients, the most useful analogous group corresponding to the orthogonal group for real coefficients is something called a unitary group.

The inner product, also called the *Hermitian*, for the complex vector space \mathbf{C}^n is defined as

$$\langle \mathbf{a} | \mathbf{b} \rangle = a_1 \bar{b}_1 + a_2 \bar{b}_2 + \cdots + a_n \bar{b}_n$$

for the complex vectors $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$ where the bar indicates complex conjugation. A matrix A , and the transformation $\mathbf{C}^n \rightarrow \mathbf{C}^n$ that it describes, are called *unitary* if it preserves the Hermitian. The collection of all unitary matrices in $GL_n(\mathbf{C})$ is called the unitary group $\mathcal{U}(n)$.

Another theorem from linear algebra is that a matrix A is unitary if and only if its inverse is the transform of its conjugate, $A^{-1} = \overline{A}^T$, equivalently, $A\overline{A}^T = I$.

There are many properties of complex unitary matrices that correspond to properties of real orthogonal matrices.

4.7.4 Projective space and the projective linear group $PGL_n(F)$

Projective planes and projective space. Projective geometry differs from Euclidean geometry in a couple of ways: all lines in a plane intersect, and distance and angles are not considered.

Let's start with Euclidean plane geometry, then drop distance and angles, then add points at infinity to get the projective plane.

When distance and angles are not considered in Euclidean geometry, what's left is called *affine geometry*. Points and lines still remain. The affine plane is still modelled by \mathbf{R}^2 , but affine transformations don't have to preserve distance or angles. So, for instance, the linear transformations known as expansions, contractions, and shear transformations are all affine transformations. In fact every element in $GL_2(\mathbf{R})$ describes an affine planar transformation. These are the affine transformations that fix the origin. Also, translations, which are not linear transformations, are affine transformations. Similarly, in dimension n , affine transformations are composed of translations and elements of $GL_n(\mathbf{R})$.

Affine spaces F^n can be similarly defined for other fields F besides the reals \mathbf{R} .

So far, we've dropped distance and angles, but parallel lines remain in affine geometry. The next step is to add enough points, called points at infinity, so that parallel lines meet at them.

Parallelism is an equivalence relation on lines. We'll assume that a line is parallel to itself, so parallelism is reflexive. It's also symmetric, and it's transitive: if one line is parallel to another, and the other to a third, then the first is parallel to the third.

For each parallelism equivalence class, add one point, a *point at infinity* to affine space and specify that every line in that equivalence class passes through that point. Add one more line, the *line at infinity*, and specify that every point at infinity passes through it. The resulting space is the *projective space* corresponding to the affine space.

Projective space and projective coordinates. Let F be a field, such as the field of real numbers. The projective linear group $PGL_n(F)$ is used to study projective space.

A more formal way to define projective n -space over a field F is by modelling points of the projective plane by lines in affine $n + 1$ -space, F^{n+1} , through the origin by means of an algebraic equivalence relation.

Two points $\mathbf{a} = (a_0, a_1, \dots, a_n)$ and $\mathbf{b} = (b_0, b_1, \dots, b_n)$ of F^{n+1} name the same point of FP^n if their coordinates are proportional, that is, if there exists a nonzero element $\lambda \in F$ such that $b_i/a_i = \lambda$ for $i = 0, 1, \dots, n$. We'll let $[a_0, a_1, \dots, a_n]$ denote the point in FP^n named by $(a_0, a_1, \dots, a_n) \in F^{n+1}$. Thus, $[a_0, a_1, \dots, a_n] = [\lambda a_0, \lambda a_1, \dots, \lambda a_n]$. The notation $[a_0, a_1, \dots, a_n]$ is called *projective coordinates*.

Geometrically, this construction adds points at infinity to the affine plane, one point for each set of parallel lines.

Lines can also be named with projective coordinates $\mathbf{b} = [b_0, b_1, \dots, b_n]$. If you do that, then a point $\mathbf{a} = [a_0, a_1, \dots, a_n]$ lies on the line \mathbf{b} if their inner product $\langle \mathbf{a} | \mathbf{b} \rangle$ is 0.

Example 4.56 (The Fano plane $\mathbf{Z}_2 P^2$). The projective plane $\mathbf{Z}_2 P^2$ has a name, the Fano plane, named after Gino Fano (1871–1952), a founder of finite geometries.

Figure 4.9 shows a representation of $\mathbf{Z}_2 P^2$. There are 7 points and 7 lines, each line with 3 points, and each point on 3 lines.

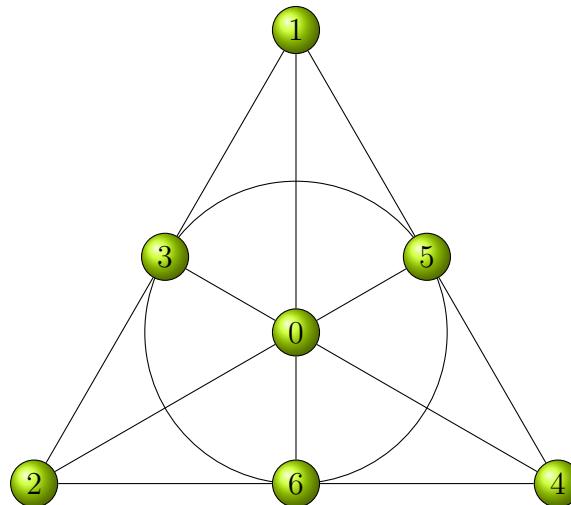


Figure 4.9: The Fano plane $\mathbf{Z}_2 P^2$

Example 4.57 (The projective plane $\mathbf{Z}_3 P^2$). Figure 4.10 shows a representation of the finite projective plane $\mathbf{Z}_3 P^2$. There are 13 points and 13 lines, each line with 4 points, and each point on 4 lines.

We can name the 9 points in the affine plane \mathbf{Z}_3^2 with third coordinate 1, and the 4 points at infinity with third coordinate 0. The four points at infinity lie on a line at infinity. Each of these points at infinity lie on all those lines with a particular slope. For instance, the point $[1, -1, 0]$ lies on the three lines with slope -1 (and it lies on the line at infinity, too).

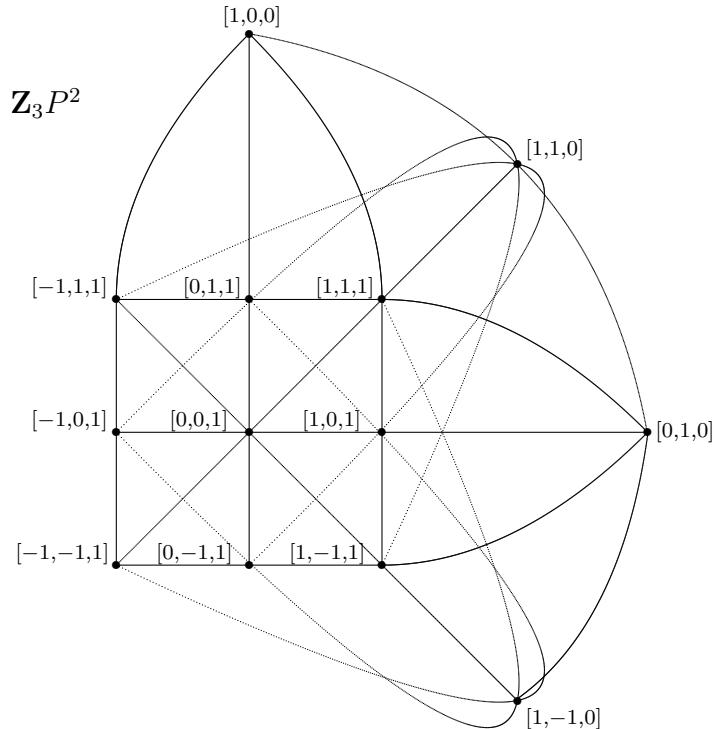


Figure 4.10: The projective plane $\mathbf{Z}_3 P^2$

Finite projective planes. There's a simpler definition of a projective plane that can be made axiomatically. It states that two points determine a line, and two lines determine a point. A nondegeneracy axiom is also required that there are at least three points which don't all lie on the same line (from which it follows that there are at least three lines which don't all meet at one point). It turns out that this axiomatic definition admits projective planes that don't derive from fields. We'll look at the ones that do.

Finite projective planes exist for each finite field. Let $GF(p^n)$ be a Galois field of $q = p^n$ elements. There will be q^2 points on the affine plane $GF(p^n)^2$ with third coordinate 1, and $q + 1$ points on the line at infinity with third coordinate 0. So the finite projective plane $GF(p^n)P^2$ has $q^2 + q + 1$ points altogether. It has the same number of lines.

These projective planes all have a couple of nice properties. They are all Desarguesian and Pappian, that is, Desargue's theorem and Pappas's theorem both hold for these projective planes. These two theorems state that certain configurations of points and lines hold for the

projective plane. Desargues developed projective geometry in the 1600s, and one of Pappus's theorems apply to projective geometry. There are other projective planes that aren't based on finite fields that aren't Desarguesian and Pappian.

Projective linear group $PGL_n(F)$. As we defined projective $n - 1$ -space over a field F as a quotient of nonzero elements of n -space, so too we can define a quotient of $GL_n(F)$ to get the projective linear group $PGL_n(F)$ acting on projective $n - 1$ -space. Two matrices A and B in $GL_n(F)$ name the same element of $PGL_n(F)$ if each is a multiple of the other, that is, there exists $\lambda \neq 0 \in F$ such that $B = \lambda A$. Then $PGL_n(F)$ acts on FP^{n-1} , since $A\mathbf{a}$ and $\lambda A\mathbf{a}$ name the same element of FP^{n-1} .

If F is a finite field with q elements, then the order of the group $PGL_n(F)$ is the order of $GL(n, F)$ divided by $q - 1$, so $|PGL_n(F)| = \frac{(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1})}{q - 1}$.

Projective special linear group $PSL_n(F)$. The *projective special linear group*, $PSL_n(F)$, is the subgroup of $PGL_n(F)$ named by unimodular matrices. It's $SL_n(F)$ modulo scalar matrices ωI where ω is an n th root of unity.

The order of $PSL_n(F)$ is equal to the order of $PGL_n(F)$ divided by $\text{GCD}(n, q - 1)$ where q is the number of elements of F .

Except for small values of n the projective special linear groups are all simple. Simplicity is defined in the next section.

The groups $PSL_3(\mathbf{Z}_3)$ is actually the same as $PGL_3(\mathbf{Z}_3)$ since 3 and 2 are relatively prime.

Example 4.58. The projective linear group $PGL_3(\mathbf{Z}_2) = PSL_3(\mathbf{Z}_2)$ which acts on the Fano plane $\mathbf{Z}_2 P^2$ has $\frac{7 \cdot 6 \cdot 4}{1} = 168$ elements.

It's small enough so that its conjugacy classes can be determined without resorting to advanced methods. There are six conjugacy classes of sizes 1, 21, 56, 42, 24, and 24. As always, the identity forms a conjugacy class of 1 element. Refer to figure 4.9 to name elements. The conjugacy class of the involution $(12)(56)$ has size 21. The conjugacy class of $(124)(365)$ has 56 elements. The conjugacy class of $(0124)(36)$ has 42 elements. The conjugacy class of (0125463) has 24 elements, and the conjugacy class of its inverse also has 24 elements.

Also, $PGL_3(\mathbf{Z}_3) = PSL_3(\mathbf{Z}_3)$, acting on the projective plane $\mathbf{Z}_3 P^2$, has order $\frac{26 \cdot 24 \cdot 18}{2} = 5616$.

4.8 Structure of finite groups

The classification of finite groups is extremely difficult, but there are tools we can use to see how that classification begins. In the next section we'll classify finite Abelian groups and see that they're isomorphic to products of cyclic groups, but the situation for general groups much more complicated.

4.8.1 Simple groups

The way we'll analyze groups is by their normal subgroups and quotients. In particular, if N is a maximal, proper normal subgroup of G , then G/N has no subgroups, for if it did, by the correspondence theorem, there would be a normal subgroup between N and G .

Definition 4.59. A nontrivial group is said to be *simple* if it has no proper, nontrivial, normal subgroups.

Exercise 78. Prove that the only Abelian simple groups are cyclic of prime order.

There are many nonabelian simple groups. There are several infinite families of them, and a few that aren't in infinite families, called *sporadic* simple groups. One infinite family of simple groups consists of alternating groups A_n with $n \geq 5$. Indeed, A_5 is the smallest nonabelian simple group. The projective special linear groups mentioned in the section above form another family of finite simple groups.

Exercise 79 (Nonsimplicity of A_4). Verify that there are five conjugacy classes in A_4 as shown in the following table.

Generator	Size	Order
1	1	1
(12)(34)	3	2
(123)	4	3
(132)	4	3

A normal subgroup of A_4 would be a union of some of these conjugacy classes including the identity conjugacy class of size 1, but its order would have to divide 12. Find all the proper nontrivial normal subgroups of A_4 .

Exercise 80 (Simplicity of A_5). Verify that there are five conjugacy classes in A_5 as shown in the following table.

Generator	Size	Order
1	1	1
(12)(34)	15	2
(123)	20	3
(12345)	12	5
(12354)	12	5

A normal subgroup of A_5 would be a union of some of these conjugacy classes including the identity conjugacy class of size 1, but its order would have to divide 60. Verify that no combination of the numbers 1, 15, 12, 12, and 20, where 1 is included in the the combination, yields a sum that divides 60 (those numbers being 2, 3, 4, 6, 10, 12, 15, 20, and 30) except just 1 itself and the sum of all five numbers. Thus, there is no proper nontrivial normal subgroup of A_5 .

4.8.2 The Jordan-Hölder theorem

Definition 4.60. A *composition series* for a group G is a finite chain of subgroups

$$1 = N_n \subseteq N_{n-1} \subseteq \cdots \subseteq N_1 \subseteq N_0 = G$$

such that each N_{i-1} is a maximal proper normal subgroup of N_i . The number n is called the *length* of the composition series, and the n quotient groups

$$N_{n-1}/1, \dots, N_1/N_2, G/N_1$$

which are all simple groups, are called *composition factors* determined by the composition series.

It is evident that any finite group G has at least one composition series. Just take N_1 to be a maximal proper normal subgroup of G , N_2 to be a maximal proper normal subgroup of N_1 , etc. Infinite groups may also have composition series, but not all infinite groups do.

Exercise 81. Find a composition series for the symmetric group S_4 .

Exercise 82. Prove that an infinite cyclic group has no (finite) composition series.

Although a finite group may have more than one composition series, the length of the series is determined by the group as are composition factors at least up to isomorphism as we'll see in a moment. Thus, these are invariants of the group. They do not, however, completely determine the group.

Exercise 83. Show that the dihedral group D_5 and the cyclic group C_{10} have composition series with the same length and same factors.

Theorem 4.61 (Jordan-Hölder). Any two composition series for a finite group have the same length and there is a one-to-one correspondence between the composition factors of the two composition series for which the corresponding composition factors are isomorphic.

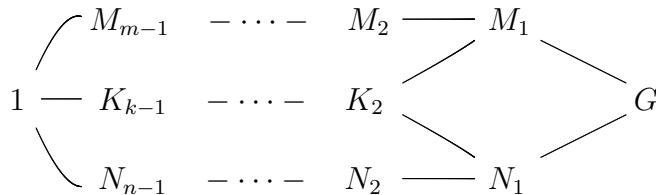
Proof. We'll prove this by induction on the order of the group under question. The base case is for the trivial group which has only the trivial composition series.

Assume now that a group G has two composition series

$$1 = N_m \subseteq M_{m-1} \subseteq \cdots \subseteq M_1 \subseteq M_0 = G, \text{ and } 1 = N_n \subseteq N_{n-1} \subseteq \cdots \subseteq N_1 \subseteq N_0 = G$$

If $M_1 = N_1$, then by induction we conclude that the lengths of the rest of the composition are equal and the composition factors the rest of the rest of the series are the same, and of course, the factors G/M_1 and G/N_1 are equal, so the case $M_1 = N_1$ is finished.

Consider now the case $M_1 \neq N_1$. Since both M_1 and N_1 are normal subgroups of G , so is their intersection $K_2 = M_1 \cap N_1$. Let $1 = K_k \subseteq K_{k-1} \subseteq \cdots \subseteq K_3 \subseteq K_2$ be a composition series for their intersection. These subgroups of G are illustrated in the following diagram.



By the second isomorphism theorem, we have $M_1/(M_1 \cap N_1) \cong G/N_1$. Therefore, K_2 is a maximal normal subgroup of M_1 . Thus, we have two composition series for M_1 , and by the inductive hypothesis, they have the same length, so $m = k$, and they have the same factors

up to isomorphism in some order. Likewise we have two composition series for N_1 , and they have the same length, so $k = n$, and the same factors up to isomorphism in some order. We now have four composition series for G , two including M_1 and two including N_1 . They all have the same length, and since $G/M_1 \cong N_1/K_2$ and $G/N_1 \cong M_1/K_2$, they all have the same factors up to isomorphism in some order.

Q.E.D.

There is a generalization of this theorem that applies to infinite groups that have composition series but its proof is considerably longer.

The list of composition factors is not enough to characterize the group. That is to say, there are non-isomorphic groups that have the same composition factors. The smallest pair of such groups are A_3 and C_6 of order 6.

A sporadic group. Most finite simple groups come in infinite parameterized families such as the cyclic groups C_p for prime p , and the alternating groups A_n for $n \geq 5$. There are several of these infinite families of simple groups. There are also a few simple groups that don't belong to any of these infinite families. We'll look at one of them, the Mathieu group M_{11} .

Mathieu discovered M_{11} in 1861. It's the smallest sporadic group, and it has order $7920 = 8 \cdot 9 \cdot 10 \cdot 11$. It can be described as a subgroup of the symmetric group S_{11} generated by the pair of permutations $(123456789te)$ and $(37e8)(4t56)$. (Here t is used for 10 and e for 11.)

M_{11} has elements of order 1, 2, 3, 4, 5, 6, 8, and 11. It has $165 = 3 \cdot 5 \cdot 11$ elements of order 2, that is, involutions. They are all conjugates of $(3t)(49)(56)(8e)$.

As S_{11} acts on a set of 11 elements, so does M_{11} . In fact, the action is sharply 4-transitive. *Transitive* means that for any pair x and y of elements in the set, there is a group element that maps the x to y . *Doubly transitive* means that for x_1, x_2 and y_1, y_2 , distinct pairs, there's a group element that sends x_1 to y_1 at the same time as sending x_2 to y_2 . More generally, and *n-transitive action* is one such that for all pairwise distinct n -tuples x_1, \dots, x_n and pairwise distinct y_1, \dots, y_n there is a group element that maps each x_i to y_i . When there is exactly one group element for pair of n -tuples, the group is said to act *sharply*.

Solvable groups One of the applications of group theory is Galois theory for algebraic fields. The groups of automorphisms of these fields are closely related to the solutions of algebraic equations. In particular, these groups can tell you if the equations have solutions that can be expressed in terms of radicals, that is square roots, cube roots, and higher roots. The condition for such solvability is none the factors in a composition series for a group are nonabelian simple groups, equivalently, that all the factors are cyclic groups of prime order.

Definition 4.62. A group is said to be *solvable* if it has a composition series all of whose factors are cyclic.

Exercise 84. Prove that if the order of a group is a power of a prime number, then that group is solvable.

Example 4.63 (The Frobenius group $F_{21} = C_7 \rtimes C_3$). This group will have 21 elements. It is what is called a semidirect product of the cyclic group $C_7 = \{1, a, a^2, \dots, a^6\}$ of 7 elements with the cyclic group $C_3 = \{1, b, b^2\}$ of 3 elements. Each element can be written in the form $b^n a^m$ with $0 \leq b \leq 2$ and $0 \leq a \leq 6$, but a and b don't commute. For this group, $bab^{-1} = a^2$.

The group is denoted $C_7 \rtimes C_3$. The group C_7 is a normal subgroup, but C_3 is not a normal subgroup.

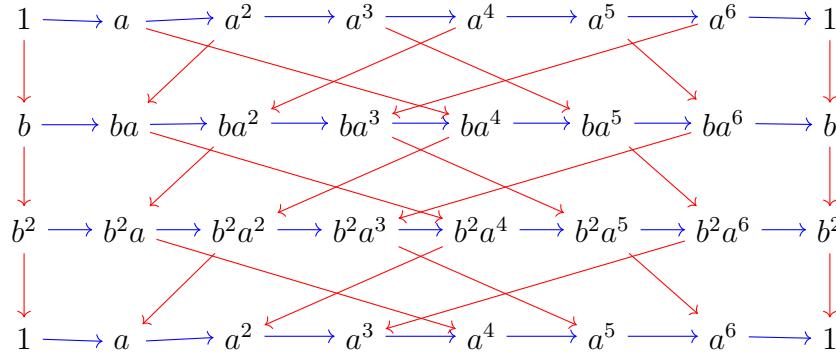


Figure 4.11: Cayley graph of the Frobenius group $F_{21} = C_7 \rtimes C_3$

This group can be presented as $\langle a, b : a^7 = 1, b^3 = 1, ba = a^2b \rangle$.

Its Cayley graph is shown in figure 4.11 with the understanding that the top line is identified with the bottom line, and the left line is identified with the right line. A blue arrow indicates multiplication by a while a red one is multiplication by b .

The group $C_7 \rtimes C_3$ is a group of symmetries of a heptahedron on a torus.

A heptahedron has 7 hexagonal faces which meet three at a time at a vertex, 14 vertices, and 27 edges. It is a tiling of the torus which is illustrated in figure 4.12. Each of the seven hexagons is labelled 1 through 7 and colored a different color. The outer edges are to be identified so that the edges $ABCD$ are identified on the upper left and lower right, the edges $DEFA$ are identified on the upper left and lower right, and the edges $AGHD$ are identified on the left and the right. The resulting topological space is a torus.

You can also interpret this as a coloring of the tiling of the plane by hexagons where the labels of some nearby hexagons are shown in the figure.

The group $C_7 \rtimes C_3$ is a subgroup of the group of symmetries of this heptahedron. The element a of order 7 describes the permutation of the faces (1234567) which moves the hexagons to the upper right. The element b of order 3 describes the permutation $(142)(356)$ which is a rotation about hexagon 7 by 120° . It's easily verified that ba and a^2b both describe the permutation $(157)(364)$ which is a rotation about hexagon 2.

Exercise 85. Verify that the rotation $c = (154623)$ about hexagon 7 by 60° is a symmetry of the heptahedron. Evidently $c^2 = b$.

- (a). Determine the relation between a and c of the form $ca = a^k c$, that is, find k .
- (b). This group is a semidirect product $C_7 \rtimes C_6$. Draw its Cayley graph.

Much more can be said about solvable groups than we have time for.

4.9 Abelian groups

Commutative groups are called Abelian groups in honor of Neils Henrik Abel (1802–1829) who worked with groups of substitutions in order to understand solutions of polynomial equations.

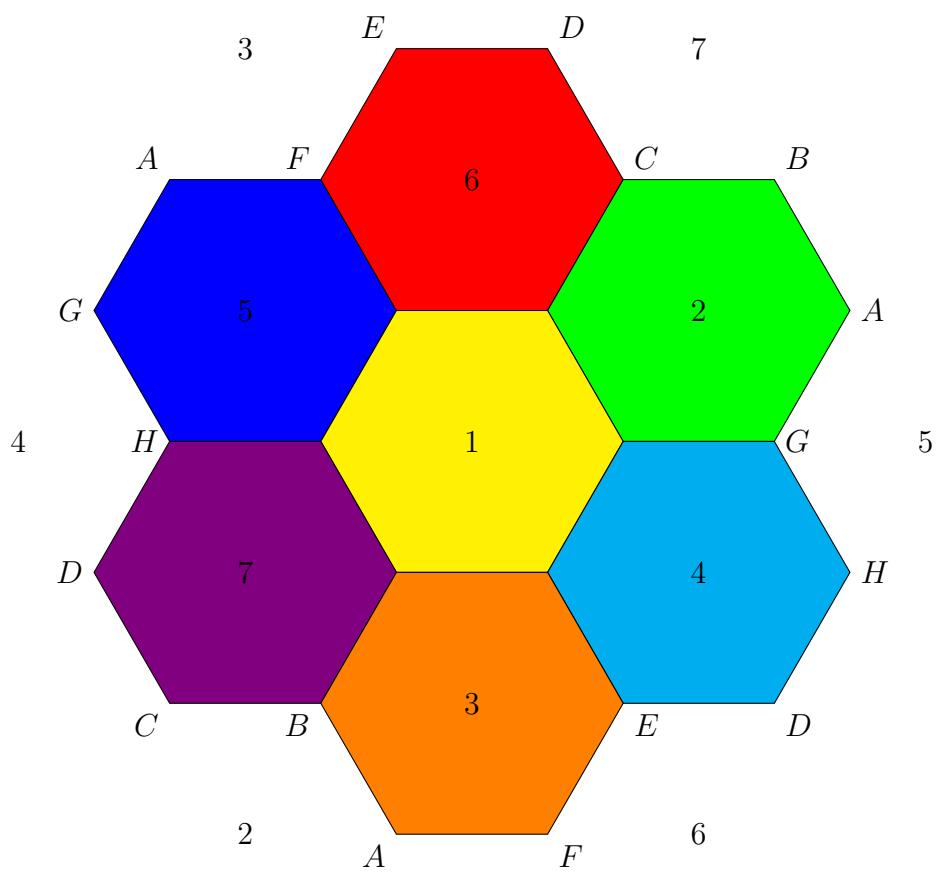


Figure 4.12: Heptahedron on a torus

We'll use additive notation throughout this section on Abelian groups. Also, we'll call the product of two Abelian groups A and B a direct sum and denote it $A \oplus B$ rather than $A \times B$.

Every subgroup of an Abelian group is normal, so we'll just refer to them as subgroups and leave off the adjective "normal."

We already know a fair amount about Abelian groups. We know about cyclic groups and the Chinese remainder theorem.

For example, we know $\mathbf{Z}_{12} \cong \mathbf{Z}_3 \oplus \mathbf{Z}_4$ where an element n modulo 12 corresponds to the pair n modulo 3 and n modulo 4. Likewise, $\mathbf{Z}_6 \cong \mathbf{Z}_2 \oplus \mathbf{Z}_3$. This gives us three ways to treat the group $\mathbf{Z}_2 \oplus \mathbf{Z}_3 \oplus \mathbf{Z}_4$ since it is isomorphic to both $\mathbf{Z}_2 \oplus \mathbf{Z}_{12}$ and $\mathbf{Z}_6 \oplus \mathbf{Z}_4$.

Our characterization of internal direct product looks a little different when the group is written additively. Here it is, rewritten for Abelian groups.

An Abelian group G is the *internal direct sum* of subgroups M and N if (1) they jointly generate G , that is, $M + N = G$, and (2) the intersection $M \cap N = 0$. If G is the internal direct sum of M and N , then $M \oplus N = G$. Furthermore, an equivalent condition to being a internal direct sum is that every element $x \in G$ can be uniquely represented as a sum $m + n$ with $m \in M$ and $n \in N$.

For the example $\mathbf{Z}_2 \oplus \mathbf{Z}_3 \oplus \mathbf{Z}_4$ above, it is the internal direct sum of \mathbf{Z}_2 and $0 \oplus \mathbf{Z}_3 \oplus \mathbf{Z}_4$ as well as the internal direct sum of $\mathbf{Z}_2 \oplus \mathbf{Z}_3 \oplus 0$ and \mathbf{Z}_4 .

4.9.1 The category \mathcal{A} of Abelian groups

The category of Abelian groups is a particularly nice category. Not only does it have products, but it also has coproducts, to be defined next, and the products are coproducts, and that's why we're calling them direct sums. It's not the only category with direct sums. The category of vector spaces over a fixed field has them too.

Coproducts in a category and their universal property When all the arrows in a diagram are reversed, a similar diagram, called the *dual* results. Recall that products in a category are characterized by a diagram.

The product $A \times B$ in a category along with the two projections $A \times B \xrightarrow{\pi_1} A$ and $A \times B \xrightarrow{\pi_2} B$ has the universal property that for each object X and morphisms $X \rightarrow A$ and $X \rightarrow B$, there is a unique morphism $X \rightarrow A \times B$, such that the diagram below commutes.

$$\begin{array}{ccccc}
 & & A & & \\
 & & \pi_1 & & \\
 X & \xrightarrow{\quad} & A \times B & \xrightarrow{\quad} & B \\
 & & \pi_2 & &
 \end{array}$$

If we turn around all the arrows, we'll get the characterizing property for coproducts. The coproduct $A \coprod B$ in a category along with the two injections $A \xrightarrow{\gamma_1} A \coprod B$ and $B \xrightarrow{\gamma_2} A \coprod B$ has the universal property that for each object X and morphisms $A \rightarrow X$ and $B \rightarrow X$, there is a unique morphism $A \coprod B \rightarrow X$, such that the diagram below commutes.

$$\begin{array}{ccc}
 A & & \\
 \searrow \gamma_1 & \nearrow & \searrow \\
 & A \coprod B & \longrightarrow X \\
 \swarrow \gamma_2 & & \swarrow \\
 B & &
 \end{array}$$

In the category \mathcal{S} of sets coproducts are disjoint unions. The disjoint union of two sets S and T has one element for each element of S and a different element for each element of T . So the cardinality of their disjoint union is $|S| + |T|$.

Exercise 86. In the category of Abelian groups, the coproduct object $A \coprod B$ is what we've called the direct sum $A \oplus B$, which is the same as the product $A \times B$. The injections $A \xrightarrow{\gamma_1} A \coprod B$ and $B \xrightarrow{\gamma_2} A \coprod B$ for Abelian groups are defined by $\gamma_1(x) = (x, 0)$ and $\gamma_2(y) = (0, y)$. Verify that the universal property holds.

4.9.2 Finite Abelian groups

The classification of finite groups is very difficult, but the classification of finite Abelian is not so difficult. It turns out, as we'll see, that a finite Abelian group is isomorphic to a product of cyclic groups, and there's a certain uniqueness to this representation. This classification is sometimes called the fundamental theorem of finite Abelian groups. The theorem above on internal direct sums is essential in this classification.

Theorem 4.64. Let G be a finite Abelian group of order mn where m and n are relatively prime, both greater than 1. Let $M = \{x \in G \mid mx = 0\}$ and $N = \{x \in G \mid nx = 0\}$. Then M and N are subgroups of G , and G is the internal direct sum of M and N . Furthermore, $|M| = m$ and $|N| = n$.

Proof. Outline. That M and N are subgroups is quickly verified. Since m and n are relatively prime, therefore 1 is a linear combination of them, that is, there are integers s and t such that $1 = sm + tn$. Their intersection $M \cap N$ is trivial since if $x \in M \cap N$, then $mx = nx = 0$, hence $x = 1x = (sm + tn)x = smx + tnx = 0$. Together M and N generate G , since for $x \in G$, $x = smx + tnx$, but $smx \in N$ since $nsmx = (nm)sx = 0$, likewise $tnx \in M$. Thus $M + N = G$. Therefore, G is the internal direct sum of M and N . Q.E.D.

p -primary groups. Let G be a Abelian group and p a prime number. The set

$$G(p) = \{x \mid p^k x = 0 \text{ for some } k \geq 0\}$$

is a subgroup of G . It is called the *p-primary component* of G .

As a corollary to the above theorem consider the case when $|G|$ is factored as a power of primes.

Corollary 4.65 (Primary decomposition theorem). Let G be a finite Abelian group whose order has prime factorization $p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r}$. Then G is a direct sum of the p_i -primary components

$$G \cong G(p_1) \oplus G(p_2) \oplus \cdots \oplus G(p_r)$$

and $|G(p_i)| = p_i^{e_i}$ for each i .

We've reduced the problem of classifying finite Abelian groups to classifying those whose orders are powers of a prime p . Such groups are called *p-primary groups* or simply *p-groups*. If the power is greater than 1, then there are different groups of that order. For example, there are three distinct Abelian groups of order 125, namely, \mathbf{Z}_{125} , $\mathbf{Z}_{25} \oplus \mathbf{Z}_5$ and $\mathbf{Z}_5 \oplus \mathbf{Z}_5 \oplus \mathbf{Z}_5$. The first has an element of order 125, but the other two don't, while the second has an element of order 25, but the third doesn't. Hence, they are not isomorphic.

We'll see soon that every p -group G is isomorphic to unique direct sum of cyclic p -groups

$$G \cong \mathbf{Z}_p^{e_1} \oplus \mathbf{Z}_{p^2}^{e_2} \oplus \cdots \oplus \mathbf{Z}_{p^r}^{e_r}$$

where the sum $e_1 + 2e_2 + \cdots + re_r$ is equal to n , where $|G| = p^n$.

Example 4.66. We'll find all the 2-groups of order 32 up to isomorphism. Since $32 = 2^5$, We'll need $e_1 + 2e_2 + \cdots + re_r = 5$. Each solution will give us a way of making a sum of positive integers equal to 5. A *partition* of n is a list of positive integers that sum to n . Here's a table which gives all the partitions of 5 and the associated 2-groups.

5	\mathbf{Z}_{32}
$1 + 4$	$\mathbf{Z}_2 \oplus \mathbf{Z}_{16}$
$2 + 3$	$\mathbf{Z}_4 \oplus \mathbf{Z}_8$
$1 + 1 + 3$	$\mathbf{Z}_2^2 \oplus \mathbf{Z}_4$
$1 + 2 + 2$	$\mathbf{Z}_2 \oplus \mathbf{Z}_4^2$
$1 + 1 + 1 + 2$	$\mathbf{Z}_2^3 \oplus \mathbf{Z}_4$
$1 + 1 + 1 + 1 + 1$	\mathbf{Z}_2^5

Exercise 87. Complete this table of the number of partitions of n up through $n = 10$. Work it out yourself.

n	0	1	2	3	4	5	6	7	8	9	10
	1	1	2	3	5	7					

Fundamental theorem of finite Abelian groups .

Our strategy for a p -primary group will be to pick off direct summands containing elements of maximal orders, one at a time. That will show that a p -primary group is a direct sum of cyclic groups whose orders are nonincreasing powers of p . We'll then show those powers of p are determined by the p -primary group.

A difficulty in the proof is that there are many choices to be made resulting in different direct sums, but we'll see that the orders of the cyclic subgroups turns out to be the same no matter how we make the choices.

The proof of the theorem is particularly technical, so we'll separate parts of the proof as lemmas.

Lemma 4.67. Let G be a noncyclic p -primary group and a an element of G of maximal order. Then there is an element b in the complement of $\langle a \rangle$ of order p .

Proof. Let c be an element in the complement of $\langle a \rangle$ of smallest order. Since the order of pc is $1/p$ times the order of c , which is a smaller order than the order of c , therefore pc lies in $\langle a \rangle$. So $pc = ka$ for some integer k . Let p^m denote the ord a , the largest order of any element in G . Then $\text{ord}(ka) \leq p^{m-1}$ since $p^{m-1}(ka) = p^{m-1}pc = p^m c = 0$. Therefore, ka is not a generator of the cyclic group $\langle a \rangle$ since that group has p^m elements. Hence, $\text{GCD}(p^m, k) \neq 1$, and so p divides k . Let $k = pj$. Then $pb = ka = pji$. Let $b = c - ja$. Then $pb = 0$, but $b \notin \langle a \rangle$ as $c = b + ka \notin \langle a \rangle$. Q.E.D.

Proof. Let $|G| = p^n$ and $\text{ord } a = p^m$ with $m < n$.

We'll prove the lemma by induction. Assume it is valid for all groups of order less than p^n . Let b be an element in the complement of $\langle a \rangle$ of order p shown to exist in the previous lemma. Since $\text{ord } b = p$ and $b \notin \langle a \rangle$, therefore $\langle a \rangle \cap \langle b \rangle = 0$.

We'll reduce modulo $\langle b \rangle$ to a smaller p -primary group $G/\langle b \rangle$ where we can use the inductive hypothesis, then bring the results back up to G .

First, we'll show that $a + \langle b \rangle$, which is the image of a in $G/\langle b \rangle$, has the same order that a does in G , namely p^m , which implies that $a + \langle b \rangle$ is an element of maximal order in the group $G/\langle b \rangle$. Suppose $\text{ord}(a + \langle b \rangle) < p^m$. Then $p^{m-1}(a + \langle b \rangle)$ is the 0 element of $G/\langle b \rangle$, in other words, $p^{m-1}a \in \langle b \rangle$. But $p^{m-1}a \in \langle a \rangle$, and the intersection of $\langle a \rangle$ and $\langle b \rangle$ is trivial. Therefore, $p^{m-1}a = 0$ which contradicts $\text{ord } a = p^m$.

We now know $a + \langle b \rangle$ is an element of maximal order in the group $G/\langle b \rangle$, so we can apply the inductive hypothesis to conclude that $G/\langle b \rangle$ is the direct sum of the cyclic subgroup generated by $a + \langle b \rangle$ and another subgroup $K/\langle b \rangle$. Note that by the correspondence theorem, every subgroup of a quotient group $G/\langle b \rangle$ is the image of a group in G , so we may take K to be a subgroup of G .

We'll show that $G = \langle a \rangle \oplus K$ by showing that (1) $\langle a \rangle \cap K = 0$, and (2) $\langle a \rangle K = G$.

(1). If $x \in \langle a \rangle \cap K$, then its image $x + \langle b \rangle$ in the quotient group $G/\langle b \rangle$ lies in both the cyclic subgroup generated by $a + \langle b \rangle$ and $K/\langle b \rangle$. But their intersection is the 0 element in $G/\langle b \rangle$, therefore $x \in \langle b \rangle$. Since $x \in \langle a \rangle$ also, and $x \in \langle a \rangle \cap \langle b \rangle$ is trivial, therefore $x = 0$.

(2). We can show $\langle a \rangle K$ is all of G by a counting argument. We know that the order of $G/\langle b \rangle$ is the product of the order of the cyclic subgroup generated by $a + \langle b \rangle$ and the order of $K/\langle b \rangle$, the order of G is p times the order of $G/\langle b \rangle$, the order of $\langle a \rangle$ is the same as the order of the cyclic subgroup generated by $a + \langle b \rangle$, and the order of K is p times the order of

$K\langle b \rangle$. Therefore, the order of G equals the product of the order of $\langle a \rangle$ and the order of K . Thus $\langle a \rangle K = G$. Q.E.D.

You can prove the first statement of following theorem by induction using the lemma we just proved, then apply the primary decomposition theorem for the second statement. This is the existence half of the theorem we want. We'll still need some kind of uniqueness of the terms in the direct sum.

Theorem 4.68. A p -primary group is a direct sum of cyclic groups whose orders are powers of p . A finite Abelian group is the direct sum of cyclic groups.

There are a couple of ways to describe the uniqueness of the terms. Since we've been using cyclic groups whose orders are prime powers, let's stick to that.

There's a concept we'll need in the following lemma. If G is an Abelian group and p an integer, then the subset $G^p = \{x \mid px = 0\}$ is a subgroup of G . In fact, it's just the kernel of the group homomorphism $G \rightarrow G$ that maps x to px .

Exercise 88. Show that it is, indeed, a group homomorphism.

Lemma 4.69. Suppose that G is a p -primary group that can be written as a direct sum of nontrivial cyclic subgroups in two ways

$$G = H_1 \oplus H_2 \oplus \cdots \oplus H_m = K_1 \oplus K_2 \oplus \cdots \oplus K_n$$

where $|H_1| \geq |H_2| \geq \cdots \geq |H_m|$ and $|K_1| \geq |K_2| \geq \cdots \geq |K_n|$. Then $m = n$ and for each i , $|H_i| = |K_i|$.

Proof. Outline. By induction on the order of G . First verify that

$$G^p = H_1^p \oplus H_2^p \oplus \cdots \oplus H_m^p = K_1^p \oplus K_2^p \oplus \cdots \oplus K_n^p.$$

If any of the groups H_i^p or K_j^p are trivial, then drop them to get

$$G^p = H_1^p \oplus H_2^p \oplus \cdots \oplus H_{m'}^p = K_1^p \oplus K_2^p \oplus \cdots \oplus K_{n'}^p$$

to get two direct sums of nontrivial cyclic subgroups. By induction, $m' = n'$ and for each $i \leq m'$, $|H_i^p| = |K_i^p|$. Since $|H_i| = p|H_i^p|$ and $|K_i| = p|K_i^p|$, therefore $|H_i| = |K_i|$ for each $i \leq m'$. Finish with a counting argument to show that the number of trivial groups that were dropped is the same for the H 's as for the K 's. They're the subgroups H_i and K_i of order n . Q.E.D.

Putting the last theorem and lemma together, we have the following theorem.

Theorem 4.70 (Fundamental theorem of finite Abelian groups). A finite Abelian group is the direct sum of cyclic groups whose orders are prime powers. The number of terms in the direct sum and the orders of the cyclic groups are determined by the group.

Appendices

Appendix A

Background mathematics

A.1 Logic and proofs

Theorems. Logic and proofs are at the heart of mathematics. A statement will not be accepted by a mathematician if there's no proof of it. A theorem is a statement that has an accompanying proof.

If a statement is suspected to be true, but there's no proof yet, then it will be called a conjecture. Sometimes someone will supply the conjecture with a proof, then it becomes a theorem; sometimes a counterexample to the conjecture is discovered so it fails to be a theorem.

A typical theorem begins with the word “Theorem” followed by the statement of the theorem. That statement usually doesn't have much mathematical symbolism or variables, but it's written as much as possible in English sentences. After the proof is complete, it's ended with Q.E.D. (“Quod Erat Demonstrandum”, Latin for “that which was to be shown”) or some special symbol like a box \square .

Corollaries and lemmas are also theorems. A corollary is a theorem that follows quite easily from the preceding theorem. Sometimes the proof of a corollary is omitted and left to the reader to provide.

A lemma is a theorem that precedes another theorem. Lemmas are often technical and of little interest in themselves, but are necessary for the theorems which follows them. Sometimes a complicated proof will be split up and parts declared as lemmas. That makes it easier to understand the logical flow of the proof.

Some standard symbols seen in proofs. There are a whole lot of symbols and abbreviations that are used in proofs. Some are listed in table A.1. Although these are fairly standard, sometimes other symbols are used instead. They are used a lot when writing mathematics on a blackboard to save time. They're not as common in textbooks.

Besides these symbols, the symbol \therefore stands for “since”, and the symbol \therefore stands for “therefore”. They rarely appear in textbooks, but often on blackboards.

An example of universal quantification is the expression $\forall x, (x > 2 \Rightarrow x^2 > 4)$ which means for all x , if x is greater than 2, then x^2 is greater than 4. Typically conditions like $x > 2$ after universal quantifiers are included in the quantifier so that the implication doesn't have to be expressed separately. That last expression can be abbreviated as $\forall x > 2, x^2 > 4$.

Operation, symbol	Read As	Explanation
Conjunction, \wedge	and	The statement $A \wedge B$ is true if A and B are both true; else it is false.
Disjunction, \vee	(inclusive) or	The statement $A \vee B$ is true if A or B (or both) are true; if both are false, the statement is false.
Negation, \neg	not	The statement $\neg A$ is true just when A is false
Implication, \Rightarrow	implies; if... then	$A \Rightarrow B$ means if A is true, then B is also true; if A is false, then nothing is said about B .
Bi-implication, \iff	“iff”, if and only if	$A \iff B$ means <i>both</i> $A \Rightarrow B$ and $B \Rightarrow A$.
Universal quantification, \forall	for all; for any; for each	when it's true universally
Existential quantification, \exists	there exists; there is an	when there's at least one
Unique existential quantification, $\exists!$	there exists a unique	when there is exactly one

Table A.1: Standard logical symbols

An example of existential quantification is the expression $\exists x, (x > 1 \wedge x^2 = 4)$ which means there is an x such that x is greater than 1 and $x^2 = 4$. Typically conditions like $x > 1$ after existential quantifiers are included in the quantifier so that the conjunction doesn't have to be expressed separately. That last expression can be abbreviated as $\exists x > 1, x^2 = 4$.

A.2 Sets

Just a little bit about sets. We'll use the language of sets throughout the course, but we're not using much of set theory. This note just collects the background that you need to know about sets in one place

A.2.1 Basic set theory

A set itself is just supposed to be something that has elements. It doesn't have to have any structure but just have elements. The elements can be anything, but usually they'll be things of the same kind.

If you've only got one set, however, there's no need to even mention sets. It's when several sets are under consideration that the language of sets becomes useful.

There are ways to construct new sets, too, and these constructions are important. The most important of these is a way to select some of the elements in a set to form another set, a subset of the first.

Examples. Let's start with sets of numbers. There are ways of constructing these sets, but let's not deal with that now. Let's assume that we already have these sets.

The natural numbers. These are the counting numbers, that is, whole nonnegative numbers. That means we'll include 0 as a natural number. (Sometimes 0 isn't included.) There is a structure on \mathbf{N} , namely there are operations of addition, subtraction, etc., but as a set, it's just the numbers. You'll often see \mathbf{N} defined as

$$\mathbf{N} = \{0, 1, 2, 3, \dots\}$$

which is read as “ \mathbf{N} is the set whose elements are 0, 1, 2, 3, and so forth.” That's just an informal way of describing what \mathbf{N} is. A complete description couldn't get away with “and so forth.” If you want to see all of what “and so forth” entails, you can read Dedekind's 1888 paper *Was sind und was sollen die Zahlen?* and Joyce's comments on it. In that article Dedekind starts off developing set theory and ends up with the natural numbers.

The real numbers. These include all positive numbers, negative numbers, and 0. Besides the natural numbers, their negations and 0 are included, fractions like $\frac{22}{7}$, algebraic numbers like $\sqrt{5}$, and transcendental numbers like π and e . If a number can be named decimaly with infinitely many digits, then it's a real number. We'll use \mathbf{R} to denote the set of all real numbers. Like \mathbf{N} , \mathbf{R} has lots of operations and functions associated with it, but treated as a set, all it has is its elements, the real numbers.

Note that \mathbf{N} is a subset of \mathbf{R} since every natural number is a real number.

Elements and membership. The standard notation to say an element x is a member of a set S is $x \in S$. The \in symbol varies a bit. Sometimes it appears as an epsilon ϵ or ε or \mathcal{E} . Read $x \in S$ as “ x is an element of S ,” or as “ x belongs to S , or more simply “ x is in S .”

Its negation is the symbol \notin . So, for example $\sqrt{5} \in \mathbf{R}$, but $\sqrt{5} \notin \mathbf{N}$.

As mentioned above, sets are completely determined by their elements, so two sets are equal if they have exactly the same elements.

$$S = T \text{ if and only if (1) for all } x \in S, x \in T, \text{ and (2) for all } x \in T, x \in S.$$

The two halves of the condition on the right lead to the concept of subset.

Subsets. If you have a set and a language to talk about elements in that set, then you can form subsets of that set by properties of elements in that language.

For instance, we have arithmetic on \mathbf{R} , so solutions to equations are subsets of \mathbf{R} . The solutions to the equation $x^3 = x$ are 0, 1, and -1 . We can describe its solution set using the notation

$$S = \{x \in \mathbf{R} \mid x^3 = x\}$$

which is read as “ S is the set of x in \mathbf{R} such that $x^3 = x$.” We could also describe that set by listing its elements, $S = \{0, 1, -1\}$. When you name a set by listing its elements, the order that you name them doesn't matter. We could have also written $S = \{-1, 0, 1\}$ for the same set. This set S is a subset of \mathbf{R} .

A set S is a subset of a set T if every element of S is also an element of T , that is

$$S \subseteq T \text{ if and only if for all } x \in S, x \in T.$$

Read $S \subseteq T$ as “ S is a subset of T .”

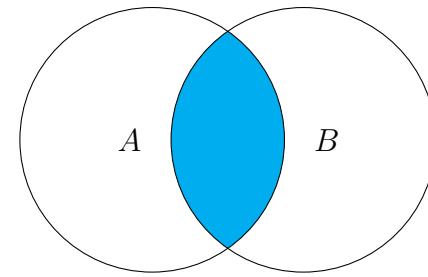
Note that $S = T$ if and only if $S \subseteq T$ and $T \subseteq S$.

There are a couple of notations for subsets. We'll use the notation $A \subseteq S$ to say that A is a subset of S . We allow $S \subseteq S$, that is, we consider a set S to be a subset of itself. If a subset A doesn't include all the elements of S , then A is called a *proper* subset of S . The only subset of S that's not a proper subset is S itself. We'll use the notation $A \subset S$ to indicate that A is a proper subset of S .

(Warning. There's an alternate notational convention for subsets. In that notation $A \subset S$ means A is any subset of S , while $A \subsetneq S$ means A is a proper subset of S . I prefer the the notation we're using because it's analogous to the notations \leq for less than or equal, and $<$ for less than.)

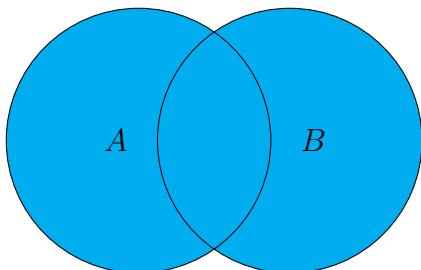
Operations on subsets. Frequently you deal with several subsets of a set, and there are operations of intersection, union, and difference that describe new subsets in terms of previously known subsets.

The intersection $A \cap B$ of two subsets A and B of a given set S is the subset of S that includes all the elements that are in both A and B , as shown in the Venn diagram below. (It's interesting that Venn called them Euler circles as Euler had used them earlier, but Leibniz had also used them, and Ramon Llull (Raymond Lully) in the 13th century.) Read $A \cap B$ as “the intersection of A and B ” or as “ A intersect B .” Note that the operation of intersection is associative and commutative.



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

Two sets A and B are said to be *disjoint* if their union is empty, $A \cap B = \emptyset$. Several sets are said to be *pairwise disjoint* if each pair of those sets are disjoint.

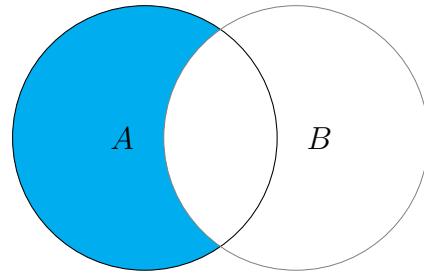


$$A \cup B = \{x \in S \mid x \in A \text{ or } x \in B\}.$$

The union $A \cup B$ of two subsets A and B of a given set S is the subset of S that includes all the elements that are in A or in B or in both. Read $A \cup B$ as “the union of A and B ” or as “ A union B .” Like intersection, the operation of union is also associative and commutative. It is usual in mathematics to take the word “or” to mean an inclusive or. It implicitly includes “or both.”

Intersection and union each distribute over the other:

$$\begin{aligned}(A \cap B) \cup C &= (A \cup C) \cap (B \cup C) \\ (A \cup B) \cap C &= (A \cap C) \cup (B \cap C)\end{aligned}$$



The difference $A - B$ of two subsets A and B of a given set S is the subset of S that includes all the elements that are in A but not in B .

$$A - B = \{x \in S \mid x \in A \text{ and } x \notin B\}$$

There's also the complement of a subset A of a set S . The complement is just $S - A$, all the elements of S that aren't in A . When the set S is understood, the complement of A often is denoted more simply as either A^c , \bar{A} , or A' rather than $S - A$. I prefer the notation A^c .

These operations satisfy lots of identities. I'll just name a couple of important ones.

De Morgan's laws describe a duality between intersection and union. They can be written as

$$(A \cap B)^c = A^c \cup B^c \quad \text{and} \quad (A \cup B)^c = A^c \cap B^c$$

Unions and intersections sometimes are taken of many subsets, even infinitely many. Suppose that A_1, A_2, \dots, A_n are subsets of S . The intersection of all of them can be written in an indexed notation as

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \cdots \cap A_n$$

and their union as

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \cdots \cup A_n.$$

And when there are infinitely many, $A_1, A_2, \dots, A_n, \dots$, as

$$\bigcap_{i=1}^{\infty} A_i = \{x \in S \mid x \in A_i \text{ for all } i\}$$

and their union as

$$\bigcup_{i=1}^{\infty} A_i = \{x \in S \mid x \in A_i \text{ for at least one } i\}.$$

DeMorgan's laws and the distributivity laws also apply to indexed intersections and unions.

$$\begin{aligned} \left(\bigcap_{i=1}^n A_i \right)^c &= \bigcup_{i=1}^n A_i^c \\ \left(\bigcup_{i=1}^n A_i \right)^c &= \bigcap_{i=1}^n A_i^c \\ \left(\bigcap_{i=1}^n A_i \right) \cup C &= \bigcap_{i=1}^n (A_i \cup C) \\ \left(\bigcup_{i=1}^n A_i \right) \cap C &= \bigcup_{i=1}^n (A_i \cap C) \end{aligned}$$

Partitions. A set S is said to be *partitioned* into subsets A_1, A_2, \dots, A_n when each element of S belongs to exactly one of the subsets A_1, A_2, \dots, A_n . That's logically equivalent to saying that S is the disjoint union of the A_1, A_2, \dots, A_n .

When you have a partition A_1, A_2, \dots, A_n of a set S like that, it induces a partition $E \cap A_1, E \cap A_2, \dots, E \cap A_n$ on each subset E of S . Each element of E belongs to exactly one of its subsets $E \cap A_1, E \cap A_2, \dots, E \cap A_n$.

Products of sets. So far we've looked at creating sets within set. There are some operations on sets that create bigger sets, the most important being creating products of sets. These depend on the concept of ordered pairs of elements. The notation for ordered pair (a, b) of two elements extends the usual notation we use for coordinates in the xy -plane. The important property of ordered pairs is that two ordered pairs are equal if and only if they have the same first and second coordinates:

$$(a, b) = (c, d) \text{ iff } a = c \text{ and } b = d.$$

The product of two sets S and T consists of all the ordered pairs where the first element comes from S and the second element comes from T :

$$S \times T = \{(a, b) \mid a \in S \text{ and } b \in T\}.$$

Thus, the usual xy -plane is $\mathbf{R} \times \mathbf{R}$, usually denoted \mathbf{R}^2 .

Besides binary products $S \times T$, you can analogously define ternary products $S \times T \times U$ in terms of triples (a, b, c) where $a \in S$, $b \in T$, and $c \in U$, and higher products, too.

Sets of subsets; power sets. Another way to create bigger sets is to form sets of subsets. If you collect all the subsets of a given set S into a set, then the set of all those subsets is called the *power set* of S , denoted $\mathcal{P}(S)$ or sometimes 2^S .

For example, let S be a set with 3 elements, $S = \{a, b, c\}$. Then S has eight subsets. There are three singleton subsets, that is, subsets having exactly one element, namely $\{a\}$, $\{b\}$, and $\{c\}$. There are three subsets having exactly two elements, namely $\{a, b\}$, $\{a, c\}$, and $\{b, c\}$. There's one subset having all three elements, namely S itself. And there's one subset that has no elements. You could denote it $\{\}$, but it's always denoted \emptyset and called the *empty set* or *null set*. Thus, the power set of S has eight elements

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

Cardinality, countable versus uncountable sets. The *cardinality* of a set S is the number of elements in it, denoted $|S|$. So, for example, if $S = \{a, b, c\}$, then $|S| = 3$, and $|\mathcal{P}(S)| = 2^3 = 8$.

Some sets are infinite, so their cardinality is not a finite number. A more careful definition is needed. Two sets S and T are said to have the *same cardinality* if there is a one-to-one correspondence of their elements. That means that there is some function $f : S \rightarrow T$ which is injective (also called one-to-one) and surjective (also called onto). A function which is both injective and surjective is called a *bijection*. For a bijection $f : S \rightarrow T$, the inverse function $f^{-1} : T \rightarrow S$ is also a bijection. The notation $|S| = |T|$ indicates S and T have the same cardinality.

If there is an injection $S \rightarrow T$, then the cardinality of S is less than or equal to that of T , written $|S| \leq |T|$. It is evident that \leq is a transitive relation on cardinalities. The Schröder-Bernstein theorem states that if there are injections both ways between S and T , then they have the same cardinality. Thus, \leq is a partial order on cardinalities.

The notation $|S| < |T|$ means $|S| \leq |T|$ but not $|S| = |T|$.

As Georg Cantor (1845–1918) discovered, not all infinite sets have the same cardinality. Some infinite sets are bigger than others. Using his famous diagonal proof, he proved that for any set, even if it's infinite, $|S| < |\mathcal{P}(S)|$.

The smallest size an infinite set can be is that of the natural numbers \mathbf{N} . A set that has the same cardinality as \mathbf{N} is called a *countably infinite* set. An infinite set that doesn't have the same cardinality as \mathbf{N} is called an *uncountable* set. The set of real numbers \mathbf{R} is uncountable.

Finite sets are also said to be countable. Thus, a set is *countable* if it's either finite or countably infinite.

A.2.2 Functions and relations

A function f is associated to a pair of sets, a *domain* S and a *codomain* T . The usual notations for that are $f : S \rightarrow T$ and $S \xrightarrow{f} T$. In order to be a function, each element $x \in S$ must be associated to a particular element of T , denoted $f(x)$.

The *graph* of a function f is a subset of the product $S \times T$, namely, the set $F = \{(x, y) \in S \times T \mid y = f(x)\}$.

Two functions are said to be the same if they have the same graph, so the graph characterizes the function. Frequently, textbooks define a function $f : S \rightarrow T$ as its graph, that is, a subset F of $S \times T$ such that for all $x \in X$, there is a unique $y \in T$ such that $(x, y) \in F$.

When $f : S \rightarrow T$, it is said that f maps S to T , and that f maps x to $f(x)$. This element $f(x)$ is called the *image* of x under f . The mapping of x to $f(x)$ is denoted $x \mapsto f(x)$.

The concept of image is extended to subsets of the domain. If $A \subseteq S$, then f maps A to the set $f(A) = \{f(x) \mid x \in A\}$, called the *image* of A under f .

Another related concept is that of preimage, also called inverse image. If B is a subset of the codomain T , then the preimage of B under f is the set $f^{-1}(B) = \{x \in A \mid f(x) \in B\}$.

Composition. If $f : S \rightarrow T$ and $g : T \rightarrow U$, then the composition $g \circ f : S \rightarrow U$ is defined by $(g \circ f)(x) = g(f(x))$.

Composition is associative. $(h \circ g) \circ f = h \circ (g \circ f)$. Since composition is associative, parentheses are not necessary when composing three or more functions.

For each set S there is an identity function $1_S : S \rightarrow S$ which maps every element in S to itself, $1_S(x) = x$. The identity functions act as units for composition. If $f : S \rightarrow T$, then $1_T \circ f = f$ and $f = f \circ 1_S$.

Injections, surjections, and bijections. These are words that describe certain functions $f : S \rightarrow T$ from one set to another. An *injection*, also called a *one-to-one function* is a function that maps distinct elements to distinct elements, that is, if $x \neq y$, then $f(x) \neq f(y)$. Equivalently, if $f(x) = f(y)$ then $x = y$. If S is a subset of T , then there is a natural injection $\iota : S \rightarrow T$, called the *inclusion function*, defined by $\iota(x) = x$.

A *surjection*, also called an *onto function* is one that includes all of T in its image, that is, if $y \in T$, then there is an $x \in S$ such that $f(x) = y$.

A *bijection*, also called a *one-to-one correspondence*, is a function that is simultaneously injective and bijective. Another way to describe a bijection is to say that there is an inverse function $g : T \rightarrow S$ so that the composition $g \circ f : S \rightarrow S$ is the identity function on S while $f \circ g : T \rightarrow T$ is the identity function on T . The usual notation for the function inverse to f is f^{-1} . In this situation f and g are inverse to each other, that is, if g is f^{-1} , then f is g^{-1} . Thus, $(f^{-1})^{-1} = f$.

Relations. Relations include functions, but are more general. A binary *relation* $R : S \rightarrow T$ doesn't have to associate each element of S to exactly one element of T . It can associate an element of S to any number of elements in T including the possibility of no elements in T at all. In other words, a relation $R : S \rightarrow T$ is determined by an arbitrary subset of $S \times T$.

The most useful relations are those that have special properties. The next section discusses equivalence relations. A typical equivalence relation is congruence modulo n . Order relations are discussed in section A.3. A typical order relation is \leq on numbers.

A.2.3 Equivalence relations

There are various symbols used for equivalence relations, such as \cong , \equiv , \approx , \asymp , \simeq , \sim , and so forth. We'll use \equiv for a generic equivalence relation.

Definition A.1 (Equivalence relation). An *equivalence relation* \equiv on a set S is a relation that is reflexive, symmetric, and transitive.

A relation on a set S may be identified with a subset of the product $S \times S$. For an equivalence relation \equiv , this means $x \equiv y$ corresponds to the statement that the ordered pair (x, y) is an element of that subset.

Reflexivity: For all x , $x \equiv x$.

Symmetry: For all x and y , $x \equiv y$ implies $y \equiv x$.

Transitivity: For all x , y , and z , $x \equiv y$ and $y \equiv z$ implies $x \equiv z$.

Equivalence classes and partitions of sets. An equivalence relation on a set determines a partition on that set, and conversely, as we'll see presently.

Definition A.2 (Equivalence class). Given an equivalence relation on a set, an *equivalence class* of an element x , denoted $[x]$, is the set of all elements equivalent to x ,

$$[x] = \{y \mid y \equiv x\}.$$

You can easily show the several properties of equivalence classes.

Theorem A.3. If \equiv is an equivalence relation on a set S , then the following four statements are equivalent

1. $x \equiv y$.
2. $[x] = [y]$.

3. $x \in [y]$.
4. $[x] \cap [y] \neq \emptyset$.

Furthermore, for each $x \in S$, there is exactly one equivalence class containing x , namely, $[x]$.

Definition A.4 (Partition of a set). A *partition* of a set S is a collection of nonempty subsets, called *parts*, of S which are pairwise disjoint and whose union is all of S . Thus, each element of S belongs to exactly one of the parts.

The above theorem shows that the equivalence classes form a partition. The converse is also true as you can easily show.

Theorem A.5. For each equivalence class on a set, the equivalence classes partition the set. Conversely, a partition of a set determines an equivalence relation where two elements are equivalent if they're in the same part.

The set of equivalence classes is sometimes denoted S/\equiv , and it's sometimes called a quotient set. Using equivalence classes to construct new sets of things is a common practice in mathematics and especially in algebra.

Keep in mind that you can always name an element of S/\equiv by naming an element of S , but two elements x and y of S will name the same element of S/\equiv , that is, $[x] = [y]$, if $x \equiv y$.

The function $\gamma : S \rightarrow S/\equiv$ defined by $\gamma(x) = [x]$ is called a *projection*, or the *canonical function*, from the set to its quotient set.

A.2.4 Axioms of set theory

Although the axioms of set theory don't play an important role in an introductory course in modern algebra, occasionally they may be useful. Here is a summary of axioms of Zermelo-Fraenkel set theory, abbreviated ZF set theory.

Axiom of extensionality . This is the axiom that says two sets are the same if they have the same elements.

$$\forall A, \forall B, (\forall x, (x \in A \Leftrightarrow x \in B) \Leftrightarrow A = B).$$

Axiom of separation . This axiom is also called the axiom of specification. It says if you have a predicate φ on sets and a given set A , then there is a subset B of A on which that predicate holds.

$$\forall A, \exists B, \forall x, (x \in B \Leftrightarrow x \in A \wedge \varphi(x)).$$

It's an axiom schema rather than a single axiom because a different axiom is needed for each predicate φ .

This axiom allows the creation of smaller sets from a given set. For example, if $A = \mathbf{R}$ is the set of real numbers, by the axiom of separation there is a set B such that the elements of B are the real numbers that satisfy the equation $x^3 - 3x = 1$. Here, the predicate φ at x , written above as $\varphi(x)$ is that equation. The axiom of separation is the justification for the “set building” notation $B = \{x \in \mathbf{R} \mid x^3 - 3x = 1\}$.

Axiom of pairing. The axiom of pairing allows the creation of a set containing two elements (or one if they're the same element).

$$\forall x, \forall y, \exists A, (z \in A \iff z = x \vee z = y).$$

The set A is usually denoted $\{x, y\}$.

If it happens that $x = y$, then A only has one element instead of two since $\{x, x\} = \{x\}$. An set with only one element is called a *singleton set*, or just a *singleton*.

Axiom of union. Given a set A of sets, this says the union C of the sets in A is a set.

$$\forall A, \exists C, \forall x, (x \in C \iff \exists B, (x \in B \wedge B \in A)).$$

The usual notation for C is $\bigcup A$, or $\bigcup_{B \in A} B$.

When A is the pair $\{D, E\}$ then $\bigcup_{B \in A} B$ is the pairwise union $D \cup E$.

There doesn't need to be an axiom for intersections or for relative compliments because intersections and relative complements can be proved from the axiom of separation.

Axiom of powersets. It says given a set A , there is a set which contains all the subsets of A .

$$\forall A, \exists B, \forall C, (C \in B \iff C \subseteq A).$$

One common notation for the powerset B of A is $\wp(A)$.

Axiom of infinity. So far, there are no axioms that say there are any sets at all. This axiom says that there is an infinite set which contains the emptyset \emptyset , so among other things, it says the emptyset exists. When studying the theory of finite sets, the axiom of infinity is not included, but an explicit axiom is needed to say the emptyset exists.

Define $S(A)$ to denote $A \cup \{A\}$ where A is a set. $S(A)$ is called the *successor* of A . The axiom of pairing says that if A is a set, then so is $\{A\}$, and the axiom of union then implies that $A \cup \{A\}$ is a set.

The axiom of infinity says that there is a set B that has \emptyset as an element and is closed under S .

$$\exists B, (\emptyset \in B \wedge \forall y \in B, S(y) \in B).$$

Along with the axiom of regularity, this axiom implies that there is at least one infinite set. With the other axioms, it can be shown that there is a smallest such set. That smallest set is a model for the set of natural numbers \mathbf{N} . In that model, the emptyset \emptyset acts as 0, its successor $S(\emptyset)$ acts as 1, $S(S(\emptyset))$ acts as 2, and so forth.

Axiom of regularity. This axioms is also called the axiom of foundation. It is a technical axiom that says that given a nonempty set A , there is an element of A which is disjoint from it.

$$\forall A \neq \emptyset, \exists x \in A, \forall y \in x, y \notin A.$$

The axiom of regularity implies that no set is an element of itself, nor is there a finite cycle of memberships where $A_1 \in A_2 \in \cdots \in A_n \in A_1$. Furthermore, there is no infinite descending memberships $\cdots \in A_n \in \cdots A_2 \in A_1$. One of the main reasons for the axiom of regularity is to develop the theory of ordinals.

Axiom of replacement. Like the axiom of separation, this is another axiom schema. This technical axiom creates images of functions described by predicates. A predicate φ describes a function if for all x , there exists a unique y such that $\varphi(x, y)$. In that case, a function symbol like f is used so that $f(x) = y$ expresses $\varphi(x, y)$. (For this axiom, the predicate can have other arguments that won't be mentioned explicitly.)

$$\forall A, ((\forall x \in A, \exists!y, \varphi(x, y)) \Rightarrow \exists B, \forall x \in A, \exists y \in B, \varphi(x, y)).$$

The B in the axiom is usually denoted $f(A)$, the image of A under f .

Axiom of choice. The axiom of choice is not part of ZF set theory, but when it's included, the set theory is denoted ZFC set theory, Zermelo-Fraenkel set theory with the axiom of choice. This axiom is discussed in more detail in the section [A.4](#).

Von Neumann–Bernays–Gödel set theory (NBG). This is an extension of ZF that includes proper classes. Whereas sets can be elements of other sets and proper classes, proper classes cannot be elements. NBG is a conservative extension of ZF in that sense that any theorem not mentioning classes and provable in one theory can be proved in the other. NBG makes it possible to talk about things like the class of all sets, or the class of all groups, etc.

A.3 Ordered structures

Several mathematical structures are defined in terms of an order relation. These order relations have something in common with the order relation \leq “less than or equal” on the real numbers. Many of them are not total orders like \leq , but only partial orders. Having fewer nice properties than \leq , however, can make them more interesting.

In particular, we'll look at partial orders, lattices, and Boolean algebras.

A.3.1 Partial orders and posets.

You're familiar with the order \leq on real numbers. It's a binary relation with the following four properties.

1. Reflexivity: for all x , $x \leq x$.
2. Anti-symmetry: for all x and y , if $x \leq y$ and $y \leq x$, then $x = y$.
3. Transitivity: for all x , y , and z , if $x \leq y$ and $y \leq z$, then $x \leq z$.
4. Totality: for all x and y , either $x \leq y$ or $y \leq x$ (or both in which case $x = y$).

There are other useful binary relations in mathematics with either those four properties or at least the first three. Although sometimes such binary relations are denoted with the same \leq sign, frequently a similar but visually distinct sign such as \preceq is used. Both are read “less than or equal to”. Of course, there's also a greater than or equal to, written \succeq and defined by $x \succeq y$ if and only if $y \preceq x$.

Definition A.6 (Total order). A *total order*, also called a *linear order*, on a set is a binary relation having the four properties: reflexivity, anti-symmetry, transitivity, and totality.

A set with a specified total order is called a *totally ordered set* or a *chain*.

The strict form \prec of a total order (or a partial order defined below) \preceq is defined by

$$x \prec y \text{ if and only if } x \preceq y \text{ and } x \neq y.$$

A useful weakening of total orders is what is called a partial order.

Definition A.7 (Partial order). A *partial order* on a set is a binary relation having the first three of those properties: reflexivity, anti-symmetry, and transitivity.

A set with a specified partial order is called a *partially ordered set* or *poset* for short.

Two elements x and y in a partially ordered set are said to be *comparable* if either $x \preceq y$ or $y \preceq x$. Otherwise they're *incomparable*.

Example A.8. The positive integers are partially ordered by divisibility. Divisibility is reflexive since $x|x$; it's anti-symmetric since if $x|y$ and $y|x$, then $x = y$, and it's transitive since if $x|y$ and $y|z$, then $x|z$.

Divisibility is not a partial order on all integers since $2|-2$ and $-2|2$ but $2 \neq -2$. It is, however, a pre-order. A *pre-order* is reflexive and transitive but need not be anti-symmetric.

Example A.9. Any collection \mathcal{T} of subsets of a set S is a partially ordered set where the binary relation is \subseteq . In particular, the power set $\wp(S)$ consisting of all the subsets of S is partially ordered.

Hasse diagrams. A partially ordered set can be described with a kind of a graph called a Hasse diagram. The elements of the set are the vertices of the graph, and the edges indicate which elements are less than or equal to which other elements. If $a \prec b$, then an edge is drawn from a to b with the larger element above the smaller one. Transitivity of the order relation is assumed so that if $a \prec b \prec c$, than an edge doesn't have to be drawn between a and c .

Definition A.10. An *upper bound* of a subset S in a poset is any element in the poset which is greater than or equal to all elements in S . That element needn't be an element of the subset S . Likewise, a *lower bound* of S is an element that is less than or equal to all the elements in S .

A *least upper bound*, also called *supremum* of S is an upper bound of S which is less than or equal to all other upper bounds of S . It is denoted $\text{lub } S$ or $\sup S$. Likewise, a *greatest lower bound*, also called *infimum* of S is an lower bound of S which is greater than or equal to all other lower bounds of S . It is denoted $\text{glb } S$ or $\inf S$.

Least upper bounds and greatest lower bounds of subsets need not always exist.

Example A.11. With the usual ordering on the real numbers \mathbf{R} , both the open interval $(2, 3)$ and the closed interval $[2, 3]$ have the same least upper bound 3 and the same greatest lower bound 2.

With the usual ordering on the rational numbers \mathbf{Q} , the subset $S = \{x \mid x^2 = 2\}$ has neither a least upper bound nor a greatest lower bound since $\sqrt{2}$ and $-\sqrt{2}$ are not rational numbers.

Definition A.12 (Maximal and minimal elements). A *maximal element* in a partially ordered set is an element which is not less than or equal to any other element. A *minimal element* in a partially ordered set is an element which is not greater than or equal to any other element.

Maximal and minimal elements don't have to be unique. A partially ordered set can have more than one of each no none at all.

Definition A.13 (Meet and join). The *meet* of two elements a and b is the greatest lower bound of the set $\{a, b\}$. That is, it is an element x less than or equal to both a and b and greater than or equal to all other elements greater than or equal to both a and b . If that meet exists, it is denoted $a \wedge b$.

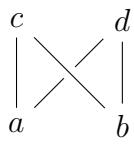
The *join* of two elements a and b in a partially ordered set is the least upper bound of the set $\{a, b\}$. That is, it is an element x greater than or equal to both a and b and less than or equal to all other elements greater than or equal to both a and b . If that join exists, it is denoted $a \vee b$.

Meets and joins aren't particularly interesting in totally ordered sets. In a totally ordered set, the meet of two elements is the minimum of the two while the join of two elements is the maximum of the two.

Example A.14. Consider the positive integers partially ordered by divisibility. The meet of two integers m and n is that number d which divides them both for which any other divisor of both divides d . In other words, a meet in this partially ordered set is the greatest common divisor.

Likewise, a join is the least common multiple.

Example A.15. Sometimes meets and joins don't exist in a partially ordered set. Consider the poset with four elements, $\{a, b, c, d\}$ where both a and b are less than both c and d .



The join $c \vee d$ doesn't exist since there is no upper bound for c and d . The join $a \vee b$ doesn't exist because there are two upper bounds for a and b , but no least upper bound. Likewise, the two meets $c \wedge d$ and $a \wedge b$ don't exist.

A.3.2 Lattices

Lattices are partially ordered sets that have meets and joins

Definition A.16 (Lattice). A *lattice* is a partially ordered set in which all meets and joins of two elements exist, has a smallest element (often denoted 0 or \perp) and a largest element (often denoted 1 or \top), in which the following identities hold.

Idempotency: $x = x \wedge x = x \vee x$.

Commutativity: $x \wedge y = y \wedge x$ and $x \vee y = y \vee x$.

Associativity: $(x \wedge y) \wedge z = x \wedge (y \wedge z)$ and $(x \vee y) \vee z = x \vee (y \vee z)$.

Absorption: $a \wedge (a \vee b) = a$ and $a \vee (a \wedge b) = a$.

Identity: $a \wedge 1 = a$ and $a \vee 0 = a$.

Lattices can be defined without reference to a partial order as the relation $a \leq b$ can be characterized in terms of meets and joins as in the following theorem.

Theorem A.17. The following three conditions are equivalent: $a \preceq b$, $a \wedge b = a$ and $a \vee b = b$.

Proof. First, suppose $a \preceq b$, then by definition, the meet of a and b is a while the join of a and b is b . Thus, the first condition in the statement of the theorem implies the other two.

Now suppose $a \wedge b = a$, since $a \wedge b \preceq b$, therefore $a \preceq b$. Thus the second condition implies the first. Similarly, the third condition implies the first. Q.E.D.

Since \preceq can be characterized in terms of \wedge and \vee , there is an alternate definition of lattice.

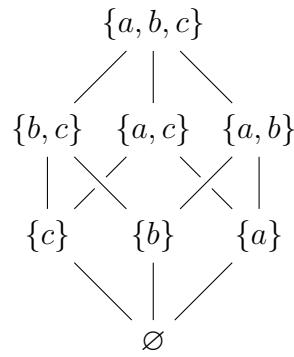
Definition A.18 (Alternate definition of lattice). A *lattice* is a set equipped with two binary operations \wedge and \vee and two constants 0 and 1 which satisfy the identities in the previous definition.

It follows from that definition that $a = a \wedge b$ if and only if $b = b \vee a$. (Proof: $a = a \wedge b$ implies $b = b \vee (b \wedge a) = (a \wedge b) \vee b = a \vee b$ by commutativity and absorption.)

The partial order can then be recovered by defining $a \preceq b$ if and only if $a \wedge b = a$ and $a \vee b = b$.

There are a couple other identities that follow from the definition, namely, $0 \wedge a = 0$ and $1 \vee a = 1$.

Example A.19. The powerset $\mathcal{P}(S)$ of a set S is a lattice. (It's actually a Boolean ring, discussed later.) Here's the Hasse diagram for $\mathcal{P}(\{a, b, c\})$.



The powerset of a set with four elements has 16 elements. It's a little harder to draw as a Hasse diagram which is displayed in figure A.1. The names of the subsets are abbreviated so that, for example, the subset $\{a, b, c\}$ is displayed as abc .

Modular and distributive lattices. Note that distributivity is not listed among the identities above. That's because it doesn't hold in all lattices. Another identity that doesn't hold in all lattices is modularity.

Definition A.20. A lattice is said to be *modular* if for all elements a , b , and c for which $a \leq c$, it is the case that $a \vee (b \wedge c) = (a \vee b) \wedge c$.

A lattice is said to be *distributive* if for all elements a , b , and c , it is the case that $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ and $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$.

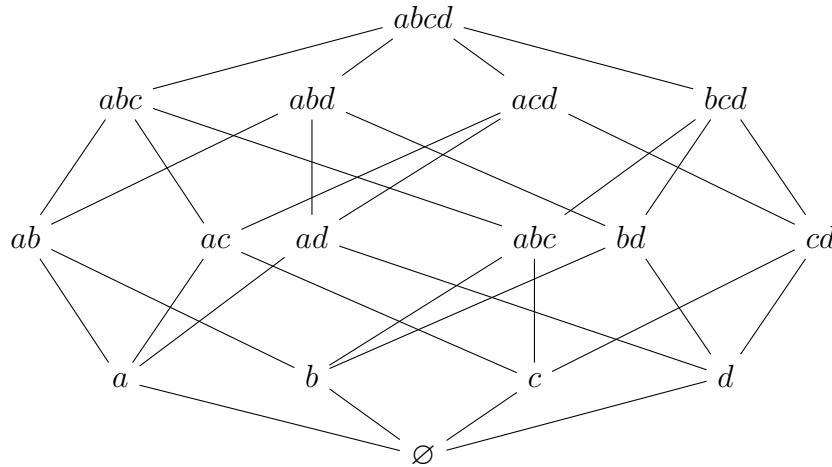


Figure A.1: Lattice of the Powerset of 4 elements

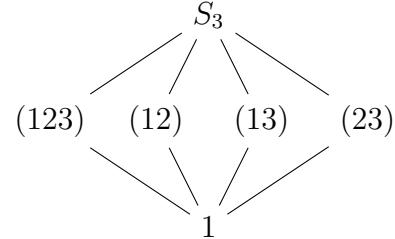
It can be shown that every distributive lattice is also modular, but there are modular lattices that are not distributive. It can also be shown that either one of the distributive identities imply the other.

The powerset $\wp(S)$ lattice is a distributive lattice.

The subgroups of a group with inclusion as a partial order always form a modular lattice, but not always a distributive lattice.

Example A.21.

The symmetric group S_3 has four subgroups besides itself and the trivial subgroup. The subgroup generated by the permutation (123) has order 3 while the three subgroups generated by the three transpositions (12) , (13) , and (23) each have order 2. The lattice of subgroups is modular, but it's not distributive.



A.3.3 Boolean algebras.

A Boolean algebra is a distributive lattice with one more operation.

Definition A.22 (Boolean algebra). A *Boolean algebra* is a distributive lattice with a unary operation, called *complementation* or *negation*, denoted \neg satisfying the identities $a \vee \neg a = 1$ and $a \wedge \neg a = 0$.

Actually, not all the identities from boolean lattices are necessary for the definition since absorption can be shown from the rest. Other identities that follow from the definition include $\neg 0 = 1$, $\neg 1 = 0$, and $\neg \neg a = a$.

As described in section 3.3, Boolean algebras are the same thing as Boolean rings. The only difference is notational.

Truth values. The two-element Boolean algebra that consists only of 0 and 1 is used in logic. 0, or \perp is the truth value “false” while 1, or \top is the truth value “true”.

A.4 Axiom of choice

Given a collection of nonempty sets, the axiom of choice says that there is a function that chooses one element from each set.

This is an axiom of set theory. There are many axioms of set theory, most of which are fairly obvious and uncontroversial.

More precisely, the axiom of choice says that given any set S , there exists a “choice function” $\gamma : \wp(S) - \emptyset \rightarrow S$ which chooses from any nonempty set $T \subseteq S$ an element $\gamma(T) \in T$.

In some sense, any theorem that relies on the axiom of choice is flawed since the axiom of choice is not constructive. So, for instance, after proving an ideal is a subideal of a maximal ideal, we won’t have any way to identify that maximal ideal.

Here’s a simple theorem that relies on the axiom of choice.

Theorem A.23. Let $f : A \rightarrow B$ be a surjective function between sets A and B . Then there exists $g : B \rightarrow A$ such that $f \circ g$ is the identity function on B .

Proof. Let γ be a choice function for A . Then g is the function

$$g(y) = \gamma(f^{-1}(y)) = \gamma(\{x \mid f(x) = y\}).$$

Since f is surjective, $f^{-1}(y)$ is not the empty set, so the choice function γ will choose some element x out of $f^{-1}(y)$ with $f(x) = y$. Q.E.D.

That theorem is actually logically equivalent to the axiom of choice, that is, the axiom of choice follows from it.

Independence of the axiom of choice. The axiom of choice is independent of the rest of the axioms of set theory. Gödel proved in 1938 that set theory with the axioms of choice added is as consistent as set theory, while Cohen in 1963 proved that set theory with the negation of the axiom of choice added is as consistent as set theory. In other words, the axiom of choice is independent of the rest of the axioms.

A.4.1 Zorn’s lemma

Although the axiom of choice is easy to state, it’s not usually easy to use. Zorn’s lemma, which is logically equivalent to the axiom of choice is hard to state, but easy to use. Another is the well-ordering principle.

This lemma is applied to a nonempty collection \mathcal{M} of subsets of a set S .

Section A.3.1 on partially-ordered sets defined a chain, upper bound, and maximal element. A chain in \mathcal{M} is a collection \mathcal{C} of elements of \mathcal{M} linearly ordered by subset inclusion. In other words, if A and B are elements of \mathcal{C} , either $A \subseteq B$ or $B \subseteq A$. An upper bound of \mathcal{C} is a subset B of S which contains all elements of \mathcal{C} . A maximal element B of \mathcal{M} is one not contained in any larger element of \mathcal{M} .

Zorn’s lemma. If every chain in \mathcal{M} has an upper bound in \mathcal{M} , then \mathcal{M} has a maximal element.

We won’t prove that the Axiom of Choice is equivalent to Zorn’s lemma because it would take too long.

A.4.2 Well-ordering principle

The most common form of the axioms of choice used in algebra is Zorn's lemma. Another that's sometimes used is the well-ordering principle.

Definition A.24 (Well-ordering). A partially ordered set is well ordered if every nonempty subset of it has a least element.

It follows from the definition that every well-ordering is totally ordered. Given two elements x and y , the subset $\{x, y\}$ has a smallest element, either x in which case $x \preceq y$, or y in which case $y \preceq x$.

Example A.25. The natural numbers \mathbf{N} is well-ordered by its usual ordering. The integers \mathbf{Z} is not well ordered by its usual ordering because the entire set doesn't have a smallest element. For the same reason \mathbf{R} is not well-ordered. The non-negative real numbers aren't well-ordered by its usual ordering because any open interval (a, b) doesn't have a least element.

Any subset of a well-ordered set is well-ordered by the same ordering.

Lexicographic ordering. The product $\mathbf{N} \times \mathbf{N}$ has a well-ordering called the lexicographic ordering. The ordered pair (a, b) is defined to be less than or equal to the ordered pair (c, d) if either $a = c$ and $b \leq d$ or $a \leq c$. Thus, the elements of $\mathbf{N} \times \mathbf{N}$ listed in increasing order are

$$(0, 0), (0, 1), (0, 2), \dots; (1, 0), (1, 1), (1, 2), \dots; (2, 0), (2, 1), (2, 2), \dots; \dots$$

More generally, if A and B are both well ordered, then the lexicographic order on $A \times B$ is a well-ordering.

Furthermore, finite products $A_0 \times A_1 \times \dots \times A_n$ of well-ordered sets are well ordered by a lexicographic ordering.

The well-ordering principle. This principle states that every set has a well-ordering, that is, for each set, there is some well-ordering of that set.

The axiom of choice, Zorn's lemma, and the well-ordering principle can each be proved from the each other. Here's a proof that the well-ordering principle follows from Zorn's lemma.

Theorem A.26. The well-ordering principle follows from Zorn's lemma.

Proof. Let S be a set. Let \mathcal{W} be the set of well-orderings of subsets of S . Partially order \mathcal{W} so given subsets A and B both with well-orderings, define $A \preceq B$ if $A \subseteq B$ and the two orderings agree on A . In other words, the ordering on A extends to that on B .

To use Zorn's lemma, we need to show that every chain \mathcal{C} in \mathcal{W} has an upper bound. A chain \mathcal{C} consists of subsets A and B where if $A \preceq B$, the ordering of A is extended to B . The union of all these subsets is a set C which, when given the extended ordering, so lies in \mathcal{W} , is itself a well-ordered set that contains every subset $A \in \mathcal{C}$. Thus, every chain in \mathcal{W} has an upper bound.

By Zorn's lemma, \mathcal{W} has a maximal element M . This M is a well-ordered subset of S which cannot be extended (since it's maximal). If there were an element of $S - M$, the

ordering on M could be extended to a well-ordering by making that element less than every element in M . Therefore, there are no elements in $S - M$. Thus, $S = M$, and so S has a well-ordering.

Q.E.D.

This principle implies that there is some well-ordering of the real numbers \mathbf{R} . It's not the usual order, of course, since the usual order does not well order \mathbf{R} . In fact, no particular well-ordering of \mathbf{R} can ever be described.

Index

- A_4 , 108, 112
 A_n , *see* Alternating group
 D_3 , 8
 D_5 , 108, 111, 119
 D_n , *see* Dihedral group
 $F_{21} = C_7 \rtimes C_3$, 133
 $GF(p^n)$, *see* Galois field
 $GL_2(\mathbf{R})$, 7
 $GL_n(R)$, *see* General linear group
 $PGL_n(F)$, *see* Projective linear group
 S^1 , *see* Unit circle
 S^2 , *see* Sphere
 S^3 , *see* 3-sphere, *see* 3-sphere
 S_4 , 119
 S_5 , 108
 S_n , *see* Symmetric group
 \mathbf{C} , *see* Complex numbers
 $\mathbf{C}[x]$, *see* Complex polynomials
 \mathbf{N} , *see* Natural numbers
 Φ_p , *see* Prime cyclotomic polynomials
 \mathbf{Q} , *see* Rational numbers
 \mathbf{R} , *see* Real numbers
 $\mathbf{R}[x]$, *see* Real polynomials
 \Rightarrow , *see* Implication
 \mathbf{Z} , *see* Integers
 $\mathbf{Z}[i]$, *see* Gaussian integers
 \mathbf{Z}_3P^2 , 128
 \mathbf{Z}_n , *see* Integers modulo n
 \triangleright^{-1} back through, 117
 \therefore , since, 143
 \cong , *see* Isomorphism
 $|$, *see* Divisibility
 \equiv , *see* Equivalence relation
 $\equiv (\text{mod } n)$, *see* Congruence modulo n
 \exists , *see* Existential quantification
 \forall , *see* Universal quantification
 \iff , *see* Bi-implication
 \wedge , *see* Conjunction
 \vee , *see* Disjunction
 \mathbf{H} , *see* Quaternions
 \mathcal{C} , *see* Category
- \mathcal{G} , *see* Category of groups
 \mathcal{R} , *see* Category of rings
 \mathcal{S} , *see* Category of sets
 \neg , *see* Negation (logical)
 \oplus , *see* Direct sum
 \wp , *see* Powerset
 \preceq , *see* Partial order
 \therefore , therefore, 143
 \triangleright through, 117
 p -group, 138
 p -primary component, 138
 p -primary group, 138
3-Sphere, 54
- Abel, Niels Henrik (1802–1829), 134
Abelian group, 7, 99
finite, 137–140
ACC (ascending chain condition), 83
Affine geometry, 127
al-Khwārizmī (ca. 780–ca. 850), 1
Algebra, 1
Boolean, 157
Cayley, 52
division, 52
Algebraic field extension, 42
Algebraic fields, 42
Algebraic integer, 42
Algebraic number, 42
Algebraic structure, 2–11
Algebraically closed field, 89
Algorithm
Brahmagupta's, 62
division, 84
Euclidean, 19–20, 86
extended Euclidean, 22
Qin Jiushao's, 62
Alternating group A_n , 107, 112
Antiautomorphism, 52
Antisymmetry, 66
Archimedean ordered field, 47–48
Archimedes of Syracuse (ca. 287–212 B.C.E.), 47

- Aristotle (384–322 B.C.E.), 17
- Arrow, *see* morphism
- Ascending chain condition, 83
- Associativity, 2
- Automorphism, 14
 - field, 43
- Axiom
 - of choice, 153, 158–160
 - of extensionality, 151
 - of infinity, 152
 - of pairing, 151
 - of powersets, 152
 - of regularity, 152
 - of replacement, 152
 - of separation, 151
 - of union, 152
- Axioms
 - Dedekind/Peano, 15
 - field, 31
 - group, 99
 - of set theory, 151–153
 - ring, 55
- Back through \triangleright^{-1} , 117
- Bernays, Paul Isaak (1888–1977), 153
- Bernstein, Felix (1878–1956), 148
- Bi-implication \iff , 144
- Bijection, 150
- Binary operation, 2
- Binary order relation, 46
- Bombelli, Rafael (1526–1572), 87
- Boole, George (1815–1864), 56, 63
- Boolean algebra, 157
- Boolean ring, 56, 63–67
- Bound
 - greatest lower, 154
 - least upper, 154
 - lower, 154
 - upper, 154
- Brahmagupta (598–670), 62
- Brahmagupta’s algorithm, 62
- Cancellation, 58
- Canonical function, 151
- Canonical homomorphism, 38
- Cardano, Gerolamo (1501–1576), 87
- Cardinality, 101, 148
- Category, 69–74
 - coproduct, 136
 - final object, 72, 115
 - generic, \mathcal{C} , 13
 - initial object, 72, 115
 - of Abelian groups \mathcal{A} , 136–137
 - of fields, 71
 - of groups \mathcal{G} , 71, 115
 - of rings \mathcal{R} , 71–74
 - of sets \mathcal{S} , 71
- Cauchy sequence, 49
- Cauchy, Augustin Louis (1789–1857), 49
- Cayley algebra, 52
- Cayley’s theorem, 110–112
- Cayley, Arthur (1821–1895), 1, 5, 52, 110
- Center of a group, 101
- Centralizer, 101
- Chain, 158
- Characteristic
 - of a field, 41
 - of a ring, 38
 - of an integral domain, 58, 69
- Charles Hermite (1882–1901), 42
- Chinese remainder theorem, 60–63, 77, 102
- Circle
 - unit, 9, 89
- Codomain, 70, 149
- Cohen, Paul (1934–2007), 158
- Commutative diagram, 70
- Commutative group, *see* Abelian group
- Commutative ring, 55
- Commutativity, 2
- Commutator
 - of two group elements, 101
 - subgroup, 101
- Comparable elements, 154
- Complete ordered field, 49–50
- Complex conjugation, 14, 43, 89
- Complex numbers \mathbf{C} , 2, 4, 43–44, 87–89
- Complex polynomials $\mathbf{C}[x]$, 87
- Composite number, 17
- Composition
 - of functions, 149
 - of homomorphisms, 13
 - of morphisms, 70
- Composition factor, 131
- Composition series, 131
- Congruence
 - group, 121
 - ring, 76
- Congruence class, 37, 76, 121

- Congruence modulo n , 5, 36
 Conjecture, 143
 Conjugacy class, 116
 Conjugate
 element in a group, 115
 subgroup, 116
 Conjugation
 complex, 14, 43, 89
 for a quadratic extension field, 43
 quaternion, 51
 Conjunction \wedge , 144
 Content of a polynomial, 93
 Contraction, 125
 Conway, John H., 47
 Coprime, 17
 Coproduct
 in a category, 136
 Core of a group, 117
 Corollary, 143
 Correspondence theorem for groups, 123
 Coset, 103–104
 Countable set, 149
 Cross product, 53
 CRT, *see* Chinese remainder theorem
 Cubic equation, 87
 Cubic polynomial, 91
 Cycle notation, 105
 Cyclic field, 39–40
 Cyclic group, 101–102
 Cyclotomic polynomial, 29
- d'Alembert, Jean le Rond (1717–1783), 89
 Dave's Short Course on Complex Numbers, 4
 de Foncenex, François Daviet (1734–1799), 89
 De Morgan's laws, 147
 De Morgan, Augustus (1806–1871), 147
 Dedekind cut, 48
 Dedekind, Richard (1831–1916), 15
 Dedekind/Peano axioms, 15
 Desargues
 Girard (1591–1661), 129
 Determinants
 as group homomorphisms, 125
 Diagram
 commutative, 70
 Dihedral group D_n , 108, 111
 Dilation, 125
 Direct sum \oplus , 102
 Disjoint
 pairwise, 146
 sets, 146
 union, 137
 Disjunction \vee , 144
 Distributivity, 3, 33, 67
 Divisibility \cong , 80–81
 Divisibility $|$, 16–17
 Division algorithm, 84
 for polynomials, 26
 Division ring, 10, 50–54
 Domain, 70, 149
 Euclidean, 84–87
 integral, 57–60, 78, 80
 principal ideal, 82–84, 86
 unique factorization, 81–82, 84
 Dot product, 53, *see* inner product
 Dyadic rational, 69
- ED, *see* Euclidean domain
 Eilenberg, Samuel (1913–1998), 69
 Eisenstein integers, 59, 86
 Eisenstein's criterion, 92–95
 Eisenstein, Gotthold (1823–1852), 59, 86, 92
 Element
 identity, 3
 initial, 15
 inverse, 3
 irreducible, 80–82
 maximal, 158
 order of, 101
 positive and negative, 45
 prime, 81, 82
 Elements, 145
 Elements of Euclid, 16–18
 Endomorphism, 14
 Epimorphism, 13, 73
 Equivalence class, 150–151
 Equivalence relation \equiv , 67, 77, 150–151
 Euclid of Alexandria (fl. ca. 300 B.C.E.), 16–19, 47
 Euclidean algorithm, 19–20, 86
 Euclidean domain, 84–87
 Euclidean geometry, 117, 127
 Euclidean valuation, 84
 Eudoxus (fl. 350 B.C.E.), 47
 Euler's circle group, 9
 Euler's identity, 89
 Euler, Leonhard (1707–1783), 9, 19, 43, 89, 146
 Even permutation, 106–107

- Existential quantification \exists , 144
 Expansion, 125
 Extended Euclidean algorithm, 22
 Extension field, 41–43
- Factor theorem, 27
 Fano plane, 128
 Fano, Gino (1871–1952), 128
 Fermat, Pierre de (1607–1665), 16
 Field, 2, 4, 31–54, 79
 algebraic, 42
 algebraic number, 42
 algebraically closed, 89
 Archimedean, 47–48
 axioms, 31
 category, 71
 complete ordered, 49–50
 definition, 4, 31
 extension, 41–43
 homomorphism, 14
 isomorphism, 12
 number, 42, 97–98
 of complex numbers, 43–44
 of rational functions, 35, 69
 of rational numbers, 34, 67–69
 ordered, 45–50
 prime, 39–41
 skew, 10, 50–54
- Field extension
 algebraic, 42
 quadratic, 41–45, 77
 transcendental, 42
- Final object, 72, 115
 Finding one, 63
 Finite Abelian group, 137–140
 Finite group, 103, 104, 112–115
 First isomorphism theorem for groups, 122
 First isomorphism theorem for rings, 78
 Fixed point, 104, 105
 Four squares identity, 52
 Fraenkel, Abraham (1891–1965), 151
 Free Boolean ring, 65
 Free group, 115
 Frobenius endomorphism, 41
 Frobenius, Ferdinand Georg (1849–1917), 41, 52, 133
 FTA, *see* Fundamental theorem of algebra
 Function, 149
 canonical, 151
- choice, 158
 codomain, 149
 composition, 149
 domain, 149
 graph, 149
 identity, 13, 149
 image, 149
 inclusion, 149
 injective, 13, 73, 149
 inverse, 150
 preimage, 149
 projection, 151
 rational, 35, 69
 successor, 15
 surjective, 13, 73, 149
- Fundamental theorem
 of algebra, 88–89
 of arithmetic, 22–24
 of finite Abelian groups, 140
- Gödel, Kurt (1906–1978), 153, 158
 Galois field $GF(p^n)$, 40, 129
 Galois, Evariste (1811–1832), 40, 133
 Gauss's lemma, 92–95
 Gauss, Carl Friedrich (1777–1855), 5, 59, 69, 85, 89, 93
 Gaussian integers, 59
 Gaussian integers $\mathbf{Z}[i]$, 69, 85
 GCD, *see* greatest common divisor
 General linear group $GL_n(R)$ $GL_n(R)$, 7, 125–126
 Geodesic, 118
 Geodesics, 117
 Geometry
 affine, 127
 Euclidean, 117, 127
- Georg Cantor (1845–1918), 149
 Girard, Albert (1595–1632), 88
 Gorenstein, Daniel (1923–1992), 112
 Grassmann, Hermann (1809–1977), 1
 Graves, John T. (1806–1870), 52
 Greatest common divisor, 19, 22, 80
 Greatest lower bound, 154
 Group, 2, 6–10, 99–140
 Abelian, 7, 99, 134–140
 alternating, 107, 112
 axioms, 99
 category, 71, 115
 center, 101

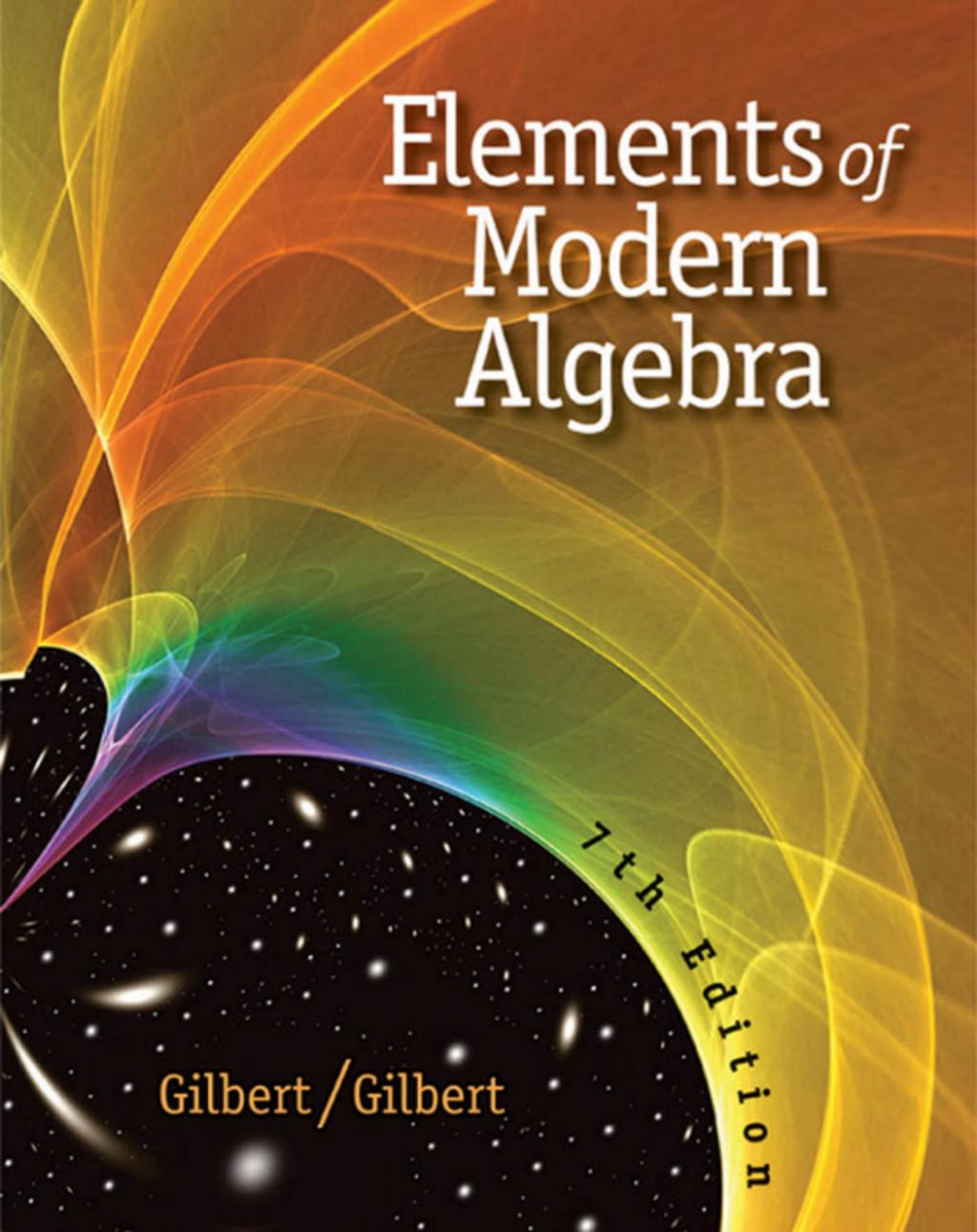
- circle, 9
- core, 117
- cyclic, 7, 101–102, 111
- definition, 6
- dihedral, 108, 111
- finite, 103, 104, 112–115
- finite Abelian, 137–140
- free, 115
- Frobenius, 133
- general linear, 7, 125–126
- homomorphism, 13
- isomorphism, 12
- Klein 4-group, 113
- linear, 124–130
 - of units in a ring, 7
 - order, 7, 101, 103
 - orthogonal, 126
 - presentation, 110
 - primary, 138
 - projective linear, 127, 130
 - projective special linear, 130
 - quaternion, 113, 119
 - quotient, 121–123
 - simple, 131–134
 - solvable, 133
 - special linear, 126
 - sporadic, 133
 - symmetric, 104–107, 110, 112, 116, 119
 - unitary, 127
- Group action
 - transitive, 133
- Group ring, 58
- Hölder, Otto (1859–1937), 131
- Hamilton, William Rowan (1805–1865), 1, 50
- Hasse diagram, 16, 154
- Hasse, Helmut (1898–1979), 16, 154
- Heptahedron, 134
- Hermite, Charles (1822–1901), 127
- Hermitian, 127
- Hom set, 70
- Homomorphism, 12–13
 - field, 14
 - group, 13
 - ring, 13, 71
- Hyperbolic space, 118
- Ideal, 74–79
 - generated by a set, 75
- maximal, 78–79
- prime, 78–79, 82
- principal, 75
- proper, 75
- trivial, 75
- Idempotent element, 64
- Identity element, 3
- Identity morphism, 70
- Image, 149
- Implication \Rightarrow , 144
- Inclusion, 14
- Inclusion function, 149
- Index of a subgroup, 103
- Infimum, *see* Greatest lower bound
- Initial element, 15
- Initial object, 72, 115
- Injection, 13, 73, 149
- Inner product, 53, 126
- Integer
 - algebraic, 42
- Integers
 - Eisenstein, 59, 86
 - Gaussian, 59, 69, 85
- Integers \mathbf{Z} , 2, 4, 5, 69, 72
- Integers modulo n , \mathbf{Z}_n , 5, 7, 35–39, 56, 60–63, 74, 77, 102
 - definition, 36
- Integral domain, 57–60, 78, 80
- Internal direct product, 123–124
- Internal direct sum, 136
- Intersection, 146, 147
 - of subgroups, 100
- Inverse element, 3
- Inverse function, 150
- Inversive space, 118
- Invertible element, *see* unit
- Involution, 101
- Involutory quandle, 117
- Irreducibility test
 - Eisenstein’s criterion, 94
 - modulo p , 93
- Irreducible element, 80–82
- Irreducible polynomial, 87, 89–97
- Isomorphism \cong , 11–12, 71, 74
 - field, 12
 - group, 12
 - ring, 11
- Isomorphism theorem
 - first for groups, 122

- first for rings, 78
- second for groups, 123
- third for groups, 123
- Join, 66
- Jordan, Camille (1838–1922), 122, 131
- Jordan-Hölder theorem, 131–133
- Joyce, David, 4, 16, 145
- Kelland, Philip (1808–1879), 50
- Kernel
 - of a group homomorphism, 120–124
 - of a ring homomorphism, 74
- Klein, Felix (1849–1925), 113
- Knuth, Donald, 47
- Krull's theorem, 79
- Krull, Wolfgang (1899–1971), 79
- Lagrange's theorem, 103–104
- Lagrange, Joseph-Louis (1736–1813), 52, 89, 103
- Laplace, Pierre-Simon (1749–1827), 89
- Latin square, 9
- Lattice, 67, 155
 - distributive, 67, 156
 - modular, 156
- Least common denominator, 22
- Least common multiple, 22
- Least upper bound, 154
- Leibniz, Gottfried Wilhelm (1646–1716), 89, 146
- Lemma, 143
- Lexicographic ordering, 159
- Lindemann, Ferdinand von (1852–1939), 42
- Linear group, 124–130
- Linear order, 153
- Linear transformation, 7, 44, 53, 124–127
- Llull, Ramon (ca. 1232–ca. 1315), 146
- Localization, 69
- Logical symbols, 143
- Loos, Ottmar, 117
- Lower bound, 154
- Mac Lane, Saunders (1909–2005), 69
- Map, 12, *see* morphism, 149
- Mathematical induction, 15
 - strong form, 23
- Mathieu, Émile Léonard (1835–1890), 133
- Matrix
 - unimodular, 126
 - unitary, 127
- Matrix representation
 - of \mathbf{C} , 43
 - of \mathbf{H} , 53
- Matrix ring, 5, 35, 43, 124–130
- Maximal ideal, 78–79
- Meet, 66
- Membership, 145
- Minimization principle, 15
- Module, 11
- Modulo p irreducibility test, 93
- Monomorphism, 13, 73
- Morphism, 12, 70
- Moufang, Ruth (1905–1977), 52
- Multiplicative function, 19
- Multiplicative group of units, 7
- Natural numbers \mathbf{N} , 2, 15
- Negation (logical) \neg , 144
- Neutral element, *see* identity element
- Noether, Emmy (1882–1935), 1
- Noether, Emmy Amalie (1882–1935), 83
- Noetherian ring, 83
- Norm
 - of a complex number, 43
 - of a quaternion, 51
- Normal subgroup, 120–124
- Number
 - algebraic, 42
 - complex, 2, 4, 43–44, 87–89
 - composite, 17
 - greatest common divisor, 22
 - integer, 2, 4, 5, 69, 72
 - natural, 2, 15, 144
 - prime, 17–19, 22–39
 - rational, 2, 4, 34, 67–69
 - real, 2, 45–50, 89–90, 145
 - relatively prime, 17, 19–20, 24, 39, 61, 102, 137
 - surreal, 47
 - transcendental, 42
 - whole, *see* integers
- Number field, 42, 97–98
- Number theory, 15–25
- Object, 70
- Octonions, 52
- Odd permutation, 106–107
- One-to-one correspondence, *see* bijection
- One-to-one function, *see* injection
- Onto function, *see* surjection

- Operation, 2–3
 - associative, 2
 - binary, 2
 - commutative, 2
 - unary, 2
- Order
 - lexicographic, 159
 - linear, 153
 - of a group, 7, 101, 103
 - of a prime in a number, 25
 - of an element in a group, 101
 - partial, 66, 153–155
 - total, 153
- Ordered field, 45–50
 - Archimedean, 47–48
 - complete, 49–50
- Orthogonal group, 126
- Orthogonal transformation, 126
- Outer product, 53
- Pairwise relatively prime numbers, 22
- Pappus of Alexandria (ca. 290–ca. 350), 129
- Partial order \preceq , 66, 153–155
- Partition, 151
 - Partition of a number, 138
 - Partition of a set, 148, 150–151
- Peano, Giuseppe (1858–1932), 15
- Permutation, 104
 - even and odd, 106–107
- Philolaus (470–385 B.C.E.), 17
- PID, *see* principal ideal domain
- Polynomial, 25–29
 - complex, 87
 - content, 93
 - cubic, 91
 - cyclotomic, 29
 - irreducible, 87, 89–97
 - monic, 25
 - prime cyclotomic, 95
 - primitive, 92
 - quadratic, 90
 - rational root theorem, 91
 - real, 89
 - root, 26
- Polynomial evaluation, 13, 73
- Polynomial ring, 5, 26, 73, 85–97
- Poset, 153–155
- Powerset \wp , 56, 64, 148
- Pre-order, 154
- Preimage, 149
- Presentation by generators and relations, 110
- Primary component, 138
- Primary decomposition theorem, 138
- Primary group, 138
- Prime cyclotomic polynomials Φ_p , 95
- Prime element, 81, 82
- Prime field, 39–41
- Prime ideal, 78–79, 82
- Prime number, 17–19, 22–39
 - infinitely many, 18
- Primitive polynomial, 92
- Primitive root of unity, 28
- Primitive roots of unity, 95
- Principal ideal, 75
- Principal ideal domain, 82–84, 86
- Principle of infinite descent, 16
- Product
 - in a category, 71
 - internal direct, 123–124
 - of groups, 102
 - of rings, 57, 71
 - of sets, 148
 - semidirect, 133
- Products of subsets in a group, 104
- Projection, 38, 151
- Projective linear group $PGL_n(F)$, 127, 130
- Projective plane
 - Desarguesian, 129
 - finite, 129
 - Pappian, 129
- Projective space, 118, 127
- Projective special linear group $PSL_n(F)$, 130
- Q.E.D., 143
- Qin Jiushao (1202–1261), 62
- Qin Jiushao's algorithm, 62
- Quadratic field extension, 41–45, 77
- Quadratic polynomial, 90
- Quandle, 10, 11, 117
 - involutory, 117
 - with geodesics, 117
- Quaternion group, 113, 119
- Quaternions \mathbf{H} , 10, 50–54
 - unit, 54
- Quotient group, 121–123
- Quotient ring, 76–79
- Quotient set, 37, 68, 76, 151

- Radian, 89
 Rational function, 35, 69
 Rational numbers, 2, 4, 34, 67–69
 Rational root theorem, 91
 Real numbers, 45–50, 89–90
 Real numbers \mathbf{R} , 2
 Real polynomials $\mathbf{R}[x]$, 89
 Reducible, 80
 Reflection, 125
 Reflexivity, 66, 150
 Relation, 150
 - antisymmetric, 66
 - binary order, 46
 - equivalence, 67, 77, 150–151
 - partial order, *see* Partial order
 - reflexive, 66, 150
 - symmetric, 150
 - transitive, 16, 66, 150
 Relatively prime, 17, 19–20, 24, 39, 61, 102, 137
 - pairwise, 22
 Remainder theorem, 27
 Residue, 6
 Ring, 2, 5–6, 55–98
 - algebraic integers, 97–98
 - axioms, 55
 - Boolean, 56, 63–67
 - category, 71–74
 - commutative, 55
 - cyclic, 35–38, 56, 77
 - definition, 5
 - division, 10, 50–54
 - free Boolean, 65
 - homomorphism, 13, 71
 - isomorphism, 11
 - matrix, 5, 35, 124–130
 - Noetherian, 83
 - of integers, *see* integers
 - of polynomials, 5, 26, 73, 85–97
 - quotient, 76–79
 - trivial, 72
 Root of unity, 28–29, 95
 - primitive, 28, 95
 Rotation, 125
 Scalar, 53
 Scalar product, 53
 Schröder, Ernst (1841–1902), 148
 Second isomorphism theorem for groups, 123
 Semidirect product, 133
 Set, 15, 144–149
 - category 71
 - complement, 147
 - countable, 149
 - difference, 147
 - element, 145
 - finite, 11
 - infinite, 15
 - intersection, 146, 147
 - membership, 145
 - operation on, 2–3
 - partially ordered, 153–155
 - partition, 148, 151
 - permutation, 104
 - power, 56, 64, 148
 - product of, 148
 - quotient, 37, 68, 76, 151
 - singleton, 152
 - subset, 145
 - uncountable, 149
 - underlying, 3, 31, 55, 99
 - union, 146, 147
 Set theory
 - axioms, 151–153
 Shear, 126
 Simple group, 131–134
 Simply infinite, 15
 Singleton set, 152
 Skew field, 10, 50–54
 Solvable group, 133
 Space
 - hyperbolic, 118
 - inversive, 118
 - projective, 118, 127
 Special linear group, 126
 Sphere, 118
 Sphere S^2 , 10
 Structure
 - algebraic, 2–11
 Subfield, 34
 - definition, 34
 Subgroup, 14, 100–104
 - commutator, 101
 - conjugate, 116
 - generated by a set, 101
 - generated by an element, 101
 - index, 103
 - normal, 120–124
 - of \mathbf{Z} , 101

- of S_3 , 107
- proper, 100
- trivial, 100
- Subring, 14, 58
- Subset, 145
- Substitution
 - Tschirnhaus, 91
- Successor function, 15
- Sun Zi (fl. 400), 62
- Supremum, *see* Least upper bound
- Surjection, 13, 73, 149
- Surreal numbers, 47
- Sylvester, James Joseph (1814–1897), 5
- Symmetric group S_n , 104–107, 110, 112, 116, 119
- Symmetric space, 117
- Symmetries
 - of a cube, 108
 - of a pentagon, 108, 111, 119
 - of a tetrahedron, 108
 - of a triangle, 8
- Symmetry, 150
- Tait, Peter Guthrie (1831–1901), 50
- Tartaglia, Nicolo Fontana (1500–1557), 87
- Theorem, 143
- Third isomorphism theorem for groups, 123
- Through \triangleright , 117
- Thymaridas (400–350 B.C.E.), 17
- Torus, 134
- Total order, 153
- Totient function, 19, 29, 39
- Transcendence
 - of π , 42
 - of e , 42
- Transcendental field extensions, 42
- Transcendental number, 42
- Transformation
 - linear, 7, 44, 53, 124–127
- Transitive group action, 133
- Transitivity, 16, 46, 66, 150
- Transposition, 105, 106
- Trichotomy, 46
- Trivial ring, 72
- Tschirnhaus substitution, 91
- Tschirnhaus, Ehrenfried Walther von (1651–1708), 91
- UFD, *see* unique factorization domain
- Unary operation, 2
- Uncountable set, 149
- Underlying set, 3
- Unimodular matrix, 126
- Union, 146, 147
 - disjoint, 137
- Unique factorization domain, 81–82, 84, 95
- Unique factorization theorem, 22–24
- Unit
 - circle, 9
 - in \mathbf{Z}_n , 19
 - in a ring, 7
- Unit circle S^1 , 89
- Unitary group, 127
- Unitary matrix, 127
- Unitary transformation, 127
- Unity
 - root of, 28–29, 95
- Universal property
 - of an infinite cyclic group, 115
 - of coproducts, 136
 - of final objects, 72, 115
 - of free groups, 115
 - of initial objects, 72, 115
 - of products, 71
 - of the ring \mathbf{Z} , 72, 73
- Universal quantification \forall , 144
- Upper bound, 154
- Valuation
 - Euclidean, 84
- Vector, 53
- Vector product, 53
- Vector space, 35
- Venn diagram, 146
- Venn, John (1834–1923), 146
- Viète, François (1540–1603), 91
- von Neumann, John (1903–1957), 153
- Waring, Edward (1736–1798), 50
- Wedderburn, Joseph (1882–1948), 59
- Weierstrass, Karl (1815–1897), 42
- Well-ordering principle, 15, 159
- Zermelo, Ernst (1871–1953), 151
- Zero-divisor, 58
- Zorn's lemma, 79, 158–160
- Zorn, Max August (1906–1993), 158



Elements *of* Modern Algebra

7th Edition

Gilbert/Gilbert

Elements of Modern Algebra

This page intentionally left blank

S E V E N T H E D I T I O N

Elements of Modern Algebra

Linda Gilbert

University of South Carolina Upstate

Jimmie Gilbert

Late of University of South Carolina Upstate



BROOKS/COLE
CENGAGE Learning™

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

**Elements of Modern Algebra,
Seventh Edition**

Linda Gilbert, Jimmie Gilbert
Editor: Molly Taylor
Development Editor: Stacy Green
Assistant Editor: Dan Seibert
Editorial Assistant: Cynthia Ashton
Marketing Specialist: Ashley Pickering
Marketing Assistant: Angela Kim
Marketing Communications Manager:
Mary Anne Payumbo
Project Manager, Editorial Production:
Cheryll Linthicum
Creative Director: Rob Hugel
Art Director: Vernon Boes
Print Buyer: Paula Vang
Permissions Editor: Tim Sisler
Production Service: ICC Macmillan Inc.
Text Designer: Roy Neuhaus
Photo Researcher: Martha Hall
Copy Editor: AmyLyn Reynolds
Illustrator: ICC Macmillan Inc.
Cover Designer: Matt Gilbert
Cover Image: Marsha Cohen
Compositor: ICC Macmillan Inc.

© 2009, 2005 Brooks/Cole, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at
Cengage Learning Customer & Sales Support, 1-800-354-9706.

For permission to use material from this text or product,
submit all requests online at www.cengage.com/permissions.
Further permissions questions can be e-mailed to
permissionrequest@cengage.com.

Library of Congress Control Number: 2008928823

ISBN-13: 978-0-495-56136-1

ISBN-10: 0-495-56136-3

Brooks/Cole
10 Davis Drive
Belmont, CA 94002-3098
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at international.cengage.com/region.

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your course and learning solutions, visit academic.cengage.com.

Purchase any of our products at your local college store or at our preferred online store www.ichapters.com.

To: Jimmie
~~Linda

This page intentionally left blank

Contents

Preface xi

1 Fundamentals 1

- 1.1 Sets 1
- 1.2 Mappings 12
- 1.3 Properties of Composite Mappings (Optional) 25
- 1.4 Binary Operations 30
- 1.5 Permutations and Inverses 37
- 1.6 Matrices 42
- 1.7 Relations 55
 - Key Words and Phrases* 62
 - A Pioneer in Mathematics: Arthur Cayley* 62

2 The Integers 65

- 2.1 Postulates for the Integers (Optional) 65
- 2.2 Mathematical Induction 71
- 2.3 Divisibility 81
- 2.4 Prime Factors and Greatest Common Divisor 86
- 2.5 Congruence of Integers 95
- 2.6 Congruence Classes 107
- 2.7 Introduction to Coding Theory (Optional) 114
- 2.8 Introduction to Cryptography (Optional) 123
 - Key Words and Phrases* 134
 - A Pioneer in Mathematics: Blaise Pascal* 135

3 Groups 137

- 3.1 Definition of a Group 137
- 3.2 Properties of Group Elements 145

3.3	Subgroups	152
3.4	Cyclic Groups	163
3.5	Isomorphisms	174
3.6	Homomorphisms	183
<i>Key Words and Phrases</i>		188
<i>A Pioneer in Mathematics: Niels Henrik Abel</i>		189

4 More on Groups 191

4.1	Finite Permutation Groups	191
4.2	Cayley's Theorem	205
4.3	Permutation Groups in Science and Art (Optional)	208
4.4	Cosets of a Subgroup	215
4.5	Normal Subgroups	223
4.6	Quotient Groups	230
4.7	Direct Sums (Optional)	239
4.8	Some Results on Finite Abelian Groups (Optional)	246
<i>Key Words and Phrases</i>		255
<i>A Pioneer in Mathematics: Augustin Louis Cauchy</i>		256

5 Rings, Integral Domains, and Fields 257

5.1	Definition of a Ring	257
5.2	Integral Domains and Fields	270
5.3	The Field of Quotients of an Integral Domain	276
5.4	Ordered Integral Domains	284
<i>Key Words and Phrases</i>		291
<i>A Pioneer in Mathematics: Richard Dedekind</i>		292

6 More on Rings 293

6.1	Ideals and Quotient Rings	293
6.2	Ring Homomorphisms	303
6.3	The Characteristic of a Ring	313
6.4	Maximal Ideals (Optional)	319
<i>Key Words and Phrases</i>		324
<i>A Pioneer in Mathematics: Amalie Emmy Noether</i>		324

7 Real and Complex Numbers 325

- 7.1 The Field of Real Numbers 325
- 7.2 Complex Numbers and Quaternions 333
- 7.3 De Moivre's Theorem and Roots of Complex Numbers 343
- Key Words and Phrases* 352
- A Pioneer in Mathematics: William Rowan Hamilton* 353

8 Polynomials 355

- 8.1 Polynomials over a Ring 355
- 8.2 Divisibility and Greatest Common Divisor 367
- 8.3 Factorization in $F[x]$ 375
- 8.4 Zeros of a Polynomial 384
- 8.5 Solution of Cubic and Quartic Equations by Formulas (Optional) 397
- 8.6 Algebraic Extensions of a Field 409
- Key Words and Phrases* 421
- A Pioneer in Mathematics: Carl Friedrich Gauss* 422

APPENDIX: The Basics of Logic 423

Answers to True/False and Selected Computational Exercises 435

Bibliography 499

Index 503

This page intentionally left blank

Preface

As the earlier editions were, this book is intended as a text for an introductory course in algebraic structures (groups, rings, fields, and so forth). Such a course is often used to bridge the gap from manipulative to theoretical mathematics and to help prepare secondary mathematics teachers for their careers.

A minimal amount of mathematical maturity is assumed in the text; a major goal is to develop mathematical maturity. The material is presented in a theorem-proof format, with definitions and major results easily located thanks to a user-friendly format. The treatment is rigorous and self-contained, in keeping with the objectives of training the student in the techniques of algebra and providing a bridge to higher-level mathematical courses.

Groups appear in the text before rings. The standard topics in elementary group theory are included, and the last two sections in Chapter 4 provide an optional sample of more advanced work in finite abelian groups.

The treatment of the set \mathbf{Z}_n of congruence classes modulo n is a unique and popular feature of this text, in that it threads throughout most of the book. The first contact with \mathbf{Z}_n is early in Chapter 2, where it appears as a set of equivalence classes. Binary operations of addition and multiplication are defined in \mathbf{Z}_n at a later point in that chapter. Both the additive and multiplicative structures are drawn upon for examples in Chapters 3 and 4. The development of \mathbf{Z}_n continues in Chapter 5, where it appears in its familiar context as a ring. This development culminates in Chapter 6 with the final description of \mathbf{Z}_n as a quotient ring of the integers by the principal ideal (n) .

Some flexibility is provided by including more material than would normally be taught in one course, and a dependency diagram of the chapters/sections (Figure P.1) is included at the end of this preface. Several sections are marked “optional” and may be skipped by instructors who prefer to spend more time on later topics.

Several users of the text have inquired as to what material the authors themselves teach in their courses. Our basic goal in a single course has always been to reach the end of Section 5.3 “The Field of Quotients of an Integral Domain,” omitting the last two sections of Chapter 4 along the way. Other optional sections could also be omitted if class meetings are in short supply. The sections on applications naturally lend themselves well to outside student projects involving additional writing and library research.

For the most part, the problems in an exercise set are arranged in order of difficulty, with easier problems first, but exceptions to this arrangement occur if it violates logical order. If one problem is needed or useful in another problem, the more basic problem appears first. When teaching from this text, we use a ground rule that any previous result, including prior exercises, may be used in constructing a proof. Whether to adopt this ground rule is, of course, completely optional.

Some users have indicated that they omit Chapter 7 (Real and Complex Numbers) because their students are already familiar with it. Others cover Chapter 8 (Polynomials) before Chapter 7. These and other options are diagrammed in Figure P.1 at the end of this preface.

The following *user-friendly* features are retained from the sixth edition:

- **Descriptive labels and titles** are placed on definitions and theorems to indicate their content and relevance.
- **Strategy boxes** that give guidance and explanation about techniques of proof are included. This feature forms a component of the bridge that enables students to become more proficient in constructing proofs.
- **Symbolic marginal notes** such as “ $(p \wedge q) \Rightarrow r$ ” and “ $\sim p \Leftarrow (\sim q \wedge \sim r)$ ” are used to help students analyze the logic in the proofs of theorems without interrupting the natural flow of the proof.
- A **reference system** provides guideposts to continuations and interconnections of exercises throughout the text. For example, consider Exercise 8 in Section 4.4. The marginal notation “Sec. 3.3, #11 \Rightarrow ” indicates that Exercise 8 of Section 4.4 is *connected* to Exercise 11 in the *earlier* Section 3.3. The marginal notation “Sec. 4.8, #7 \Leftarrow ” indicates that Exercise 8 of Section 4.4 has a *continuation* in Exercise 7 of Section 4.8. Instructors, as well as students, have found this system useful in anticipating which exercises are needed or helpful in later sections/chapters.
- An **appendix** on the basics of logic and methods of proof is included.
- A **biographical sketch** of a great mathematician whose contributions are relevant to that material concludes each chapter.
- A **gradual introduction and development** of concepts is used, proceeding from the simplest structures to the more complex.
- An **abundance of examples** that are designed to develop the student’s intuition are included.
- Enough **exercises** to allow instructors to make different assignments of approximately the same difficulty are included.
- **Exercise sets** are designed to develop the student’s maturity and ability to construct proofs. They contain many problems that are elementary or of a computational nature.
- A **summary of key words and phrases** is included at the end of each chapter.
- A **list of special notations** used in the book appears on the front endpapers.
- **Group tables** for the most common examples are on the back endpapers.
- An **updated bibliography** is included.

Between this edition and the previous one, my coauthor and beloved husband, Jimmie Gilbert, passed away. As I worked on this edition, Jimmie was sitting on my shoulder whispering do’s and don’ts to me, and for this reason, his profound influence is still being reflected in this edition. The most significant changes that “we” made include:

- enhancing the treatment of congruences to systems by introducing the Chinese Remainder Theorem (Section 2.5);
- splitting Section 3.1 so that the variety of groups can be appreciated before the group properties are emphasized;

- splitting Section 4.4 so that cosets can be completely understood before introducing normal subgroups;
- expanding the treatment of irreducibility of polynomials (Section 8.4);
- introducing the discriminant of a cubic polynomial to characterize the solutions of cubic equations (Section 8.5);
- fine-tuning the links between exercises from one section/chapter to another;
- including around 300 True/False statements that encourage the students to thoroughly understand the statements of definitions and results of theorems;
- adding nearly 400 new exercises, a majority of which are theoretical and the remainder computational; and, of course,
- minor rewriting throughout the text.

Acknowledgments

We are grateful to the following people for their helpful comments, suggestions for improvements, and corrections for this and earlier editions:

Lateef A. Adelani, <i>Harris-Stowe College</i>	Sharon Emerson-Stonnell, <i>Longwood University</i>
Philip C. Almes, <i>Wayland Baptist University</i>	Paul J. Fairbanks, <i>Bridgewater State College</i>
Edwin F. Baumgartner, <i>Le Moyne College</i>	Howard Frisinger, <i>Colorado State University</i>
Brian Beasley, <i>Presbyterian College</i>	Marcus Greferath, <i>San Diego State University</i>
Bruce M. Bemis, <i>Westminster College</i>	Jacqueline Hall, <i>Longwood University</i>
Steve Benson, <i>St. Olaf College</i>	Nickolas Heerema, <i>Florida State University</i>
Louise M. Berard, <i>Wilkes College</i>	Edward K. Hinson, <i>University of New Hampshire</i>
Thomas D. Bishop, <i>Arkansas State University</i>	J. Taylor Hollist, <i>State University of New York at Oneonta</i>
David M. Bloom, <i>Brooklyn College of the City University of New York</i>	David L. Johnson, <i>Lehigh University</i>
James C. Bradford, <i>Abilene Christian University</i>	Kenneth Kalmanson, <i>Montclair State University</i>
Shirley Branan, <i>Birmingham Southern College</i>	William J. Keane, <i>Boston College</i>
Joel Brawley, <i>Clemson University</i>	William F. Keigher, <i>Rutgers University</i>
Gordon Brown, <i>University of Colorado, Boulder</i>	Robert E. Kennedy, <i>Central Missouri State University</i>
Harmon C. Brown, <i>Harding University</i>	Andre E. Kezdy, <i>University of Louisville</i>
Marshall Cates, <i>California State University, Los Angeles</i>	Stanley M. Lukawecski, <i>Clemson University</i>
Patrick Costello, <i>Eastern Kentucky University</i>	Joan S. Morrison, <i>Goucher College</i>
Richard Cowan, <i>Shorter College</i>	Richard J. Painter, <i>Colorado State University</i>
Elwyn H. Davis, <i>Pittsburg State University</i>	
David J. DeVries, <i>Georgia College</i>	
John D. Elwin, <i>San Diego State University</i>	

Carl R. Spitznagel, <i>John Carroll University</i>	Carroll G. Wells, <i>Western Kentucky University</i>
Ralph C. Steinlage, <i>University of Dayton</i>	
James J. Tattersall, <i>Providence College</i>	Burdette C. Wheaton, <i>Mankato State University</i>
Mark L. Teply, <i>University of Wisconsin-Milwaukee</i>	John Woods, <i>Southwestern Oklahoma State University</i>
Krishnanand Verma, <i>University of Minnesota, Duluth</i>	Henry Wyzinski, <i>Indiana University Northwest</i>
Robert P. Webber, <i>Longwood College</i>	
Diana Y. Wei, <i>Norfolk State University</i>	

I wish to express my most sincere gratitude to Molly Taylor, Stacy Green, Dan Seibert, and Cynthia Ashton for their outstanding editorial guidance, to Cheryll Linthicum, Lynn Lustberg, AmyLyn Reynolds, and Vernon Boes for their excellent supervision of the production, and to Ian Crewe for his accuracy checking of answers.

Finally, my sincere thanks to Matt who showered me with his font flower bouquets, and to Beckie who gently lifted me from the darkness back to writing.

Linda Gilbert

Chapters/Sections Dependency Diagram

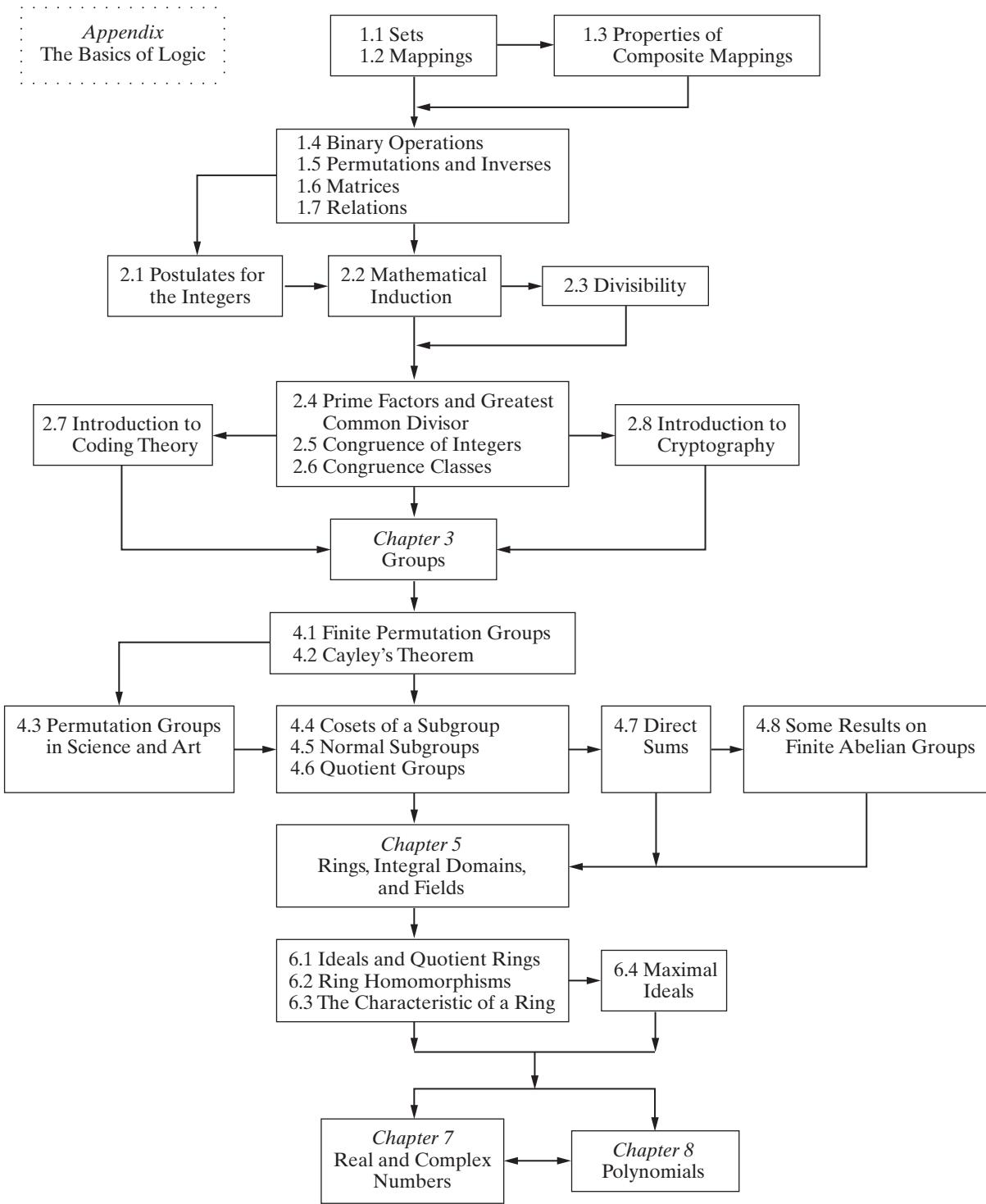


Figure P.1

This page intentionally left blank

Fundamentals

■ Introduction

This chapter presents the fundamental concepts of set, mapping, binary operation, and relation. It also contains a section on matrices, which will serve as a basis for examples and exercises from time to time in the remainder of the text. Much of the material in this chapter may be familiar from earlier courses. If that is the case, appropriate omissions can be made to expedite the study of later topics.

1.1 Sets

Abstract algebra had its beginnings in attempts to address mathematical problems such as the solution of polynomial equations by radicals and geometric constructions with straight-edge and compass. From the solutions of specific problems, general techniques evolved that could be used to solve problems of the same type, and treatments were generalized to deal with whole classes of problems rather than individual ones.

In our study of abstract algebra, we shall make use of our knowledge of the various number systems. At the same time, in many cases we wish to examine how certain properties are consequences of other, known properties. This sort of examination deepens our understanding of the system. As we proceed, we shall be careful to distinguish between the properties we have assumed and made available for use and those that must be deduced from these properties. We must accept without definition some terms that are basic objects in our mathematical systems. Initial assumptions about each system are formulated using these undefined terms.

One such undefined term is **set**. We think of a set as a collection of objects about which it is possible to determine whether or not a particular object is a member of the set. Sets are usually denoted by capital letters and are sometimes described by a list of their elements, as illustrated in the following examples.

Example 1 We write

$$A = \{0, 1, 2, 3\}$$

to indicate that the set A contains the elements 0, 1, 2, 3, and no other elements. The notation $\{0, 1, 2, 3\}$ is read as “the set with elements 0, 1, 2, and 3.” ■

Example 2 The set B , consisting of all the nonnegative integers, is written

$$B = \{0, 1, 2, 3, \dots\}.$$

The three dots \dots , called an *ellipsis*, mean that the pattern established before the dots continues indefinitely. The notation $\{0, 1, 2, 3, \dots\}$ is read as “the set with elements 0, 1, 2, 3, and so on.” ■

As in Examples 1 and 2, it is customary to avoid repetition when listing the elements of a set. Another way of describing sets is called *set-builder notation*. Set-builder notation uses braces to enclose a property that is the qualification for membership in the set.

Example 3 The set B in Example 2 can be described using set-builder notation as

$$B = \{x | x \text{ is a nonnegative integer}\}.$$

The vertical slash is shorthand for “such that,” and we read “ B is the set of all x such that x is a nonnegative integer.” ■

There is also a shorthand notation for “is an element of.” We write “ $x \in A$ ” to mean “ x is an element of the set A .” We write “ $x \notin A$ ” to mean “ x is not an element of the set A .” For the set A in Example 1, we can write

$$2 \in A \quad \text{and} \quad 7 \notin A.$$

Definition 1.1 ■ Subset

Let A and B be sets. Then A is called a **subset** of B if and only if every element of A is an element of B . Either the notation $A \subseteq B$ or the notation $B \supseteq A$ indicates that A is a subset of B .

The notation $A \subseteq B$ is read “ A is a subset of B ” or “ A is contained in B .” Also, $B \supseteq A$ is read as “ B contains A .” The symbol \in is reserved for elements, whereas the symbol \subseteq is reserved for subsets.

Example 4 We write

$$a \in \{a, b, c, d\} \quad \text{or} \quad \{a\} \subseteq \{a, b, c, d\}.$$

However,

$$a \subseteq \{a, b, c, d\} \quad \text{and} \quad \{a\} \in \{a, b, c, d\}$$

are both *incorrect* uses of set notation. ■

Definition 1.2 ■ Equality of Sets

Two sets are **equal** if and only if they contain exactly the same elements.

The sets A and B are equal, and we write $A = B$, if each member of A is also a member of B and if each member of B is also a member of A . Typically, a proof that two sets are

equal is presented in two parts. The first shows that $A \subseteq B$, the second that $B \subseteq A$. We then conclude that $A = B$. We shall have an example of this type of proof shortly.

Strategy ■ One method that can be used to prove that $A \neq B$ is to exhibit an element that is in either set A or set B but is not in both.

Example 5 Suppose $A = \{1, 1\}$, $B = \{-1, 1\}$, and $C = \{1\}$. Now $A \neq B$ since $-1 \in B$ but $-1 \notin A$, whereas $A = C$ since $A \subseteq C$ and $A \supseteq C$. ■

Definition 1.3 ■ Proper Subset

If A and B are sets, then A is a **proper subset** of B if and only if $A \subseteq B$ and $A \neq B$.

We sometimes write $A \subset B$ to denote that A is a proper subset of B .

Example 6 The following statements illustrate the notation for proper subsets and equality of sets.

$$\{1, 2, 4\} \subset \{1, 2, 3, 4, 5\} \quad \{a, c\} = \{c, a\}$$

There are two basic operations, *union* and *intersection*, that are used to combine sets. These operations are defined as follows.

Definition 1.4 ■ Union, Intersection

If A and B are sets, the **union** of A and B is the set $A \cup B$ (read “ A union B ”), given by

$$A \cup B = \{x | x \in A \text{ or } x \in B\}.$$

The **intersection** of A and B is the set $A \cap B$ (read “ A intersection B ”), given by

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

The union of two sets A and B is the set whose elements are either in A or in B or are in both A and B . The intersection of sets A and B is the set of those elements common to both A and B .

Example 7 Suppose $A = \{2, 4, 6\}$ and $B = \{4, 5, 6, 7\}$. Then

$$A \cup B = \{2, 4, 5, 6, 7\}$$

and

$$A \cap B = \{4, 6\}.$$

The operations of union and intersection of two sets have some properties that are analogous to properties of addition and multiplication of numbers.

Example 8 It is easy to see that for any sets A and B , $A \cup B = B \cup A$:

$$\begin{aligned} A \cup B &= \{x | x \in A \text{ or } x \in B\} \\ &= \{x | x \in B \text{ or } x \in A\} \\ &= B \cup A. \end{aligned}$$

Because of the fact that $A \cup B = B \cup A$, we say that the operation union has the **commutative property**. It is just as easy to show that $A \cap B = B \cap A$, and we say also that the operation intersection has the commutative property. ■

It is easy to find sets that have no elements at all in common. For example, the sets

$$A = \{1, -1\} \quad \text{and} \quad B = \{0, 2, 3\}$$

have no elements in common. Hence, there are no elements in their intersection, $A \cap B$, and we say that the intersection is *empty*. Thus it is logical to introduce the *empty set*.

Definition 1.5 ■ Empty Set, Disjoint Sets

The **empty set** is the set that has no elements, and the empty set is denoted by \emptyset or $\{\}$. Two sets A and B are called **disjoint** if and only if $A \cap B = \emptyset$.

The sets $\{1, -1\}$ and $\{0, 2, 3\}$ are disjoint, since

$$\{1, -1\} \cap \{0, 2, 3\} = \emptyset.$$

There is only one empty set \emptyset , and \emptyset is a subset of every set. For a set A with n elements (n a nonnegative integer), we can write out all the subsets of A . For example, if

$$A = \{a, b, c\},$$

then the subsets of A are

$$\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, A.$$

Definition 1.6 ■ Power Set

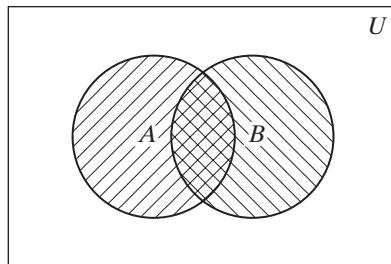
For any set A , the **power set** of A , denoted by $\mathcal{P}(A)$, is the set of all subsets of A and is written

$$\mathcal{P}(A) = \{X | X \subseteq A\}.$$

Example 9 For $A = \{a, b, c\}$, the power set of A is

$$\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, A\}. \quad \blacksquare$$

It is often helpful to draw a picture or diagram of the sets under discussion. When we do this, we assume that all the sets we are dealing with, along with all possible unions and intersections of those sets, are subsets of some **universal set**, denoted by U . In Figure 1.1, we let two overlapping circles represent the two sets A and B . The sets A and B are subsets of the universal set U , represented by the rectangle. Hence the circles are contained in the rectangle. The intersection of A and B , $A \cap B$, is the crosshatched region where the two circles overlap. This type of pictorial representation is called a **Venn diagram**.



$\text{---} \text{---}$: A

$\text{---} \text{---}$: B

$\text{---} \text{---} \text{---}$: $A \cap B$

■ **Figure 1.1**

Another special subset is defined next.

Definition 1.7 ■ Complement

For arbitrary subsets A and B of the universal set U , the **complement** of B in A is

$$A - B = \{x \in U \mid x \in A \text{ and } x \notin B\}.$$

The special notation A' is reserved for a particular complement, $U - A$:

$$A' = U - A = \{x \in U \mid x \notin A\}.$$

We read A' simply as “the complement of A ” rather than as “the complement of A in U .”

Example 10 Let

$$U = \{x \mid x \text{ is an integer}\}$$

$$A = \{x \mid x \text{ is an even integer}\}$$

$$B = \{x \mid x \text{ is a positive integer}\}.$$

Then

$$\begin{aligned} B - A &= \{x \mid x \text{ is a positive odd integer}\} \\ &= \{1, 3, 5, 7, \dots\} \end{aligned}$$

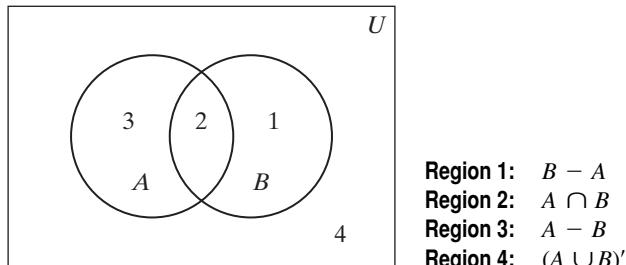
$$\begin{aligned} A - B &= \{x \mid x \text{ is a nonpositive even integer}\} \\ &= \{0, -2, -4, -6, \dots\} \end{aligned}$$

$$\begin{aligned} A' &= \{x \mid x \text{ is an odd integer}\} \\ &= \{\dots, -3, -1, 1, 3, \dots\} \end{aligned}$$

$$\begin{aligned} B' &= \{x \mid x \text{ is a nonpositive integer}\} \\ &= \{0, -1, -2, -3, \dots\}. \end{aligned}$$

■

Example 11 The overlapping circles representing the sets A and B separate the interior of the rectangle representing U into four regions, labeled 1, 2, 3, and 4, in the Venn diagram in Figure 1.2. Each region represents a particular subset of U .



Many of the examples and exercises in this book involve familiar systems of numbers, and we adopt the following standard notations for some of these systems:

- Z** denotes the set of all **integers**.
- Z⁺** denotes the set of all **positive integers**.
- Q** denotes the set of all **rational numbers**.
- R** denotes the set of all **real numbers**.
- R⁺** denotes the set of all **positive real numbers**.
- C** denotes the set of all **complex numbers**.

We recall that a **complex number** is defined as a number of the form $a + bi$, where a and b are real numbers and $i = \sqrt{-1}$. Also, a real number x is **rational** if and only if x can be written as a quotient of integers that has a nonzero denominator. That is,

$$\mathbf{Q} = \left\{ \frac{m}{n} \mid m \in \mathbf{Z}, n \in \mathbf{Z}, \text{ and } n \neq 0 \right\}.$$

The relationships that some of the number systems have to each other are indicated by the Venn diagram in Figure 1.3.

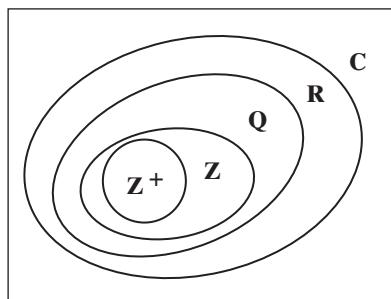


Figure 1.3 $\mathbf{Z}^+ \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R} \subset \mathbf{C}$

Our work in this book usually assumes a knowledge of the various number systems that would be familiar from a precalculus or college algebra course. Some exceptions occur when we wish to examine how certain properties are consequences of other properties in a particular system. Exceptions of this kind occur with the integers in Chapter 2 and the complex numbers in Chapter 7, and these exceptions are clearly indicated when they occur.

The operations of union and intersection can be applied repeatedly. For instance, we might form the intersection of A and B , obtaining $A \cap B$, and then form the intersection of this set with a third set C : $(A \cap B) \cap C$.

Example 12 The sets $(A \cap B) \cap C$ and $A \cap (B \cap C)$ are equal, since

$$\begin{aligned}(A \cap B) \cap C &= \{x \mid x \in A \text{ and } x \in B\} \cap C \\ &= \{x \mid x \in A \text{ and } x \in B \text{ and } x \in C\} \\ &= A \cap \{x \mid x \in B \text{ and } x \in C\} \\ &= A \cap (B \cap C).\end{aligned}$$

In analogy with the associative property

$$(x + y) + z = x + (y + z)$$

for addition of numbers, we say that the operation of intersection is **associative**. When we work with numbers, we drop the parentheses for convenience and write

$$x + y + z = x + (y + z) = (x + y) + z.$$

Similarly, for sets A , B , and C , we write

$$A \cap B \cap C = A \cap (B \cap C) = (A \cap B) \cap C. \quad \blacksquare$$

Just as simply, we can show (see Exercise 18 in this section) that the union of sets is an associative operation. We write

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

Example 13 A separation of a nonempty set A into mutually disjoint nonempty subsets is called a **partition** of the set A . If

$$A = \{a, b, c, d, e, f\},$$

then one partition of A is

$$X_1 = \{a, d\}, \quad X_2 = \{b, c, f\}, \quad X_3 = \{e\},$$

since

$$A = X_1 \cup X_2 \cup X_3$$

with $X_1 \neq \emptyset, X_2 \neq \emptyset, X_3 \neq \emptyset$, and

$$X_1 \cap X_2 = \emptyset, \quad X_1 \cap X_3 = \emptyset, \quad X_2 \cap X_3 = \emptyset.$$

The concept of a partition is fundamental to many of the topics encountered later in this book. ■

The operations of intersection, union, and forming complements can be combined in all sorts of ways, and several nice equalities can be obtained that relate some of these results. For example, it can be shown that

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

and that

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

Because of the resemblance between these equations and the familiar distributive property $x(y + z) = xy + xz$ for numbers, we call these equations **distributive properties**.

We shall prove the first of these distributive properties in the next example and leave the last one as an exercise. To prove the first, we shall show that $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ and that $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. This illustrates the point made earlier in the discussion of equality of sets, immediately after Definition 1.2.

The symbol \Rightarrow is shorthand for “implies,” and \Leftarrow is shorthand for “is implied by.” We use them in the next example.

Example 14

To prove

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

we first let $x \in A \cap (B \cup C)$. Now

$$\begin{aligned} x \in A \cap (B \cup C) &\Rightarrow x \in A \quad \text{and} \quad x \in (B \cup C) \\ &\Rightarrow x \in A, \quad \text{and} \quad x \in B \quad \text{or} \quad x \in C \\ &\Rightarrow x \in A \quad \text{and} \quad x \in B, \quad \text{or} \quad x \in A \quad \text{and} \quad x \in C \\ &\Rightarrow x \in A \cap B, \quad \text{or} \quad x \in A \cap C \\ &\Rightarrow x \in (A \cap B) \cup (A \cap C). \end{aligned}$$

Thus $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

Conversely, suppose $x \in (A \cap B) \cup (A \cap C)$. Then

$$\begin{aligned} x \in (A \cap B) \cup (A \cap C) &\Rightarrow x \in A \cap B, \quad \text{or} \quad x \in A \cap C \\ &\Rightarrow x \in A \quad \text{and} \quad x \in B, \quad \text{or} \quad x \in A \quad \text{and} \quad x \in C \\ &\Rightarrow x \in A, \quad \text{and} \quad x \in B \quad \text{or} \quad x \in C \\ &\Rightarrow x \in A \quad \text{and} \quad x \in (B \cup C) \\ &\Rightarrow x \in A \cap (B \cup C). \end{aligned}$$

Therefore, $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$, and we have shown that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

It should be evident that the second part of the proof can be obtained from the first simply by reversing the steps. That is, when each \Rightarrow is replaced by \Leftarrow , a valid implication results. In fact, then, we could obtain a proof of both parts by replacing \Rightarrow with \Leftrightarrow , where \Leftrightarrow is short for “if and only if.” Thus

$$\begin{aligned}x \in A \cap (B \cup C) &\Leftrightarrow x \in A \quad \text{and} \quad x \in (B \cup C) \\&\Leftrightarrow x \in A, \quad \text{and} \quad x \in B \quad \text{or} \quad x \in C \\&\Leftrightarrow x \in A \quad \text{and} \quad x \in B, \quad \text{or} \quad x \in A \quad \text{and} \quad x \in C \\&\Leftrightarrow x \in A \cap B, \quad \text{or} \quad x \in A \cap C \\&\Leftrightarrow x \in (A \cap B) \cup (A \cap C).\end{aligned}$$
■

Strategy ■ In proving an equality of sets S and T , we can often use the technique of showing that $S \subseteq T$ and then check to see whether the steps are reversible. In many cases, the steps are indeed reversible, and we obtain the other part of the proof easily. However, this method should not obscure the fact that there are still two parts to the argument: $S \subseteq T$ and $T \subseteq S$.

There are some interesting relations between complements and unions or intersections. For example, it is true that

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

This statement is one of two that are known as **De Morgan's[†] Laws**. De Morgan's other law is the statement that

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

Stated somewhat loosely in words, the first law says that the complement of an intersection is the union of the individual complements. The second similarly says that the complement of a union is the intersection of the individual complements.

Exercises 1.1

True or False

Label each of the following statements as either true or false.

1. Two sets are equal if and only if they contain exactly the same elements.
2. If A is a subset of B and B is a subset of A , then A and B are equal.
3. The empty set is a subset of every set except itself.
4. $A - A = \emptyset$ for all sets A .
5. $A \cup A = A \cap A$ for all sets A .

[†]Augustus De Morgan (1806–1871) coined the term mathematical induction and is responsible for rigorously defining the concept. Not only does he have laws of logic bearing his name but also the headquarters of the London Mathematical Society and a crater on the moon.

6. $A \subset A$ for all sets A .
 7. $\{a, b\} = \{b, a\}$
 8. $\{a, b\} = \{b, a, b\}$
 9. $A - B = C - B$ implies $A = C$, for all sets A, B , and C .
 10. $A - B = A - C$ implies $B = C$, for all sets A, B , and C .
-

Exercises

1. For each set A , describe A by indicating a property that is a qualification for membership in A .
 - a. $A = \{0, 2, 4, 6, 8, 10\}$
 - b. $A = \{1, -1\}$
 - c. $A = \{-1, -2, -3, \dots\}$
 - d. $A = \{1, 4, 9, 16, 25, \dots\}$
2. Decide whether or not each statement is true for $A = \{2, 7, 11\}$ and $B = \{1, 2, 9, 10, 11\}$.
 - a. $2 \subseteq A$
 - b. $\{11, 2, 7\} \subseteq A$
 - c. $2 = A \cap B$
 - d. $\{7, 11\} \in A$
 - e. $A \subseteq B$
 - f. $\{7, 11, 2\} = A$
3. Decide whether or not each statement is true, where A and B are arbitrary sets.
 - a. $B \cup A \subseteq A$
 - b. $B \cap A \subseteq A \cup B$
 - c. $\emptyset \subseteq A$
 - d. $0 \in \emptyset$
 - e. $\emptyset \in \{\emptyset\}$
 - f. $\emptyset \subseteq \{\emptyset\}$
 - g. $\{\emptyset\} \subseteq \emptyset$
 - h. $\{\emptyset\} = \emptyset$
 - i. $\emptyset \in \emptyset$
 - j. $\emptyset \subseteq \emptyset$
4. Decide whether or not each of the following is true for all sets A, B , and C .
 - a. $A \cap A' = \emptyset$
 - b. $A \cap \emptyset = A \cup \emptyset$
 - c. $A \cap (B \cup C) = A \cup (B \cap C)$
 - d. $A \cup (B' \cap C') = A \cup (B \cup C)'$
 - e. $A \cup (B \cap C) = (A \cup B) \cap C$
 - f. $(A \cap B) \cup C = A \cap (B \cup C)$
 - g. $A \cup (B \cap C) = (A \cap C) \cup (B \cap C)$
 - h. $A \cap (B \cup C) = (A \cup B) \cap (A \cup C)$
5. Evaluate each of the following sets, where

$$U = \{0, 1, 2, 3, \dots, 10\}$$

$$A = \{0, 1, 2, 3, 4, 5\}$$

$$B = \{0, 2, 4, 6, 8, 10\}$$

$$C = \{2, 3, 5, 7\}.$$

- a. $A \cup B$
- b. $A \cap C$
- c. $A' \cup B$
- d. $A \cap B \cap C$
- e. $A' \cap B \cap C$
- f. $A \cup (B \cap C)$
- g. $A \cap (B \cup C)$
- h. $(A \cup B)'$
- i. $A - B$
- j. $B - A$
- k. $A - (B - C)$
- l. $C - (B - A)$
- m. $(A - B) \cap (C - B)$
- n. $(A - B) \cap (A - C)$

6. Write each of the following as either A , A' , U , or \emptyset , where A is an arbitrary subset of the universal set U .

- | | |
|-----------------------|-----------------------|
| a. $A \cap A$ | b. $A \cup A$ |
| c. $A \cap A'$ | d. $A \cup A'$ |
| e. $A \cup \emptyset$ | f. $A \cap \emptyset$ |
| g. $A \cap U$ | h. $A \cup U$ |
| i. $U \cup A'$ | j. $A - \emptyset$ |
| k. \emptyset' | l. U' |
| m. $(A')'$ | n. $\emptyset - A$ |

7. Write out the power set, $\mathcal{P}(A)$, for each set A .

- | | |
|------------------------|---------------------------------------|
| a. $A = \{a\}$ | b. $A = \{0, 1\}$ |
| c. $A = \{a, b, c\}$ | d. $A = \{1, 2, 3, 4\}$ |
| e. $A = \{1, \{1\}\}$ | f. $A = \{\{1\}\}$ |
| g. $A = \{\emptyset\}$ | h. $A = \{\emptyset, \{\emptyset\}\}$ |

8. Describe two partitions of each of the following sets.

- | | |
|---|---|
| a. $\{x \mid x \text{ is an integer}\}$ | b. $\{a, b, c, d\}$ |
| c. $\{1, 5, 9, 11, 15\}$ | d. $\{x \mid x \text{ is a complex number}\}$ |

9. Write out all the different partitions of the given set A .

- | | |
|----------------------|-------------------------|
| a. $A = \{1, 2, 3\}$ | b. $A = \{1, 2, 3, 4\}$ |
|----------------------|-------------------------|

10. Suppose the set A has n elements where $n \in \mathbf{Z}^+$.

- a. How many elements does the power set $\mathcal{P}(A)$ have?

Sec. 2.2, #33–36 ⇐
b. If $0 \leq k \leq n$, how many elements of the power set $\mathcal{P}(A)$ contain exactly k elements?

11. State the most general conditions on the subsets A and B of U under which the given equality holds.

- | | |
|---------------------------|-----------------------------|
| a. $A \cap B = A$ | b. $A \cup B' = A$ |
| c. $A \cup B = A$ | d. $A \cap B' = A$ |
| e. $A \cap B = U$ | f. $A' \cap B' = \emptyset$ |
| g. $A \cup \emptyset = U$ | h. $A' \cap U = \emptyset$ |

12. Let \mathbf{Z} denote the set of all integers, and let

$$\begin{aligned} A &= \{x \mid x = 3p - 2 \text{ for some } p \in \mathbf{Z}\} \\ B &= \{x \mid x = 3q + 1 \text{ for some } q \in \mathbf{Z}\}. \end{aligned}$$

Prove that $A = B$.

13. Let \mathbf{Z} denote the set of all integers, and let

$$\begin{aligned} C &= \{x \mid x = 3r - 1 \text{ for some } r \in \mathbf{Z}\} \\ D &= \{x \mid x = 3s + 2 \text{ for some } s \in \mathbf{Z}\}. \end{aligned}$$

Prove that $C = D$.

In Exercises 14–33, prove each statement.

14. $A \cap B \subseteq A \cup B$
15. $(A')' = A$
16. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.
17. $A \subseteq B$ if and only if $B' \subseteq A'$.
18. $A \cup (B \cup C) = (A \cup B) \cup C$
19. $(A \cup B)' = A' \cap B'$
20. $(A \cap B)' = A' \cup B'$
21. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
22. $A \cap (A' \cup B) = A \cap B$
23. $A \cup (A' \cap B) = A \cup B$
24. $A \cup (A \cap B) = A \cap (A \cup B)$
25. If $A \subseteq B$, then $A \cup C \subseteq B \cup C$.
26. If $A \subseteq B$, then $A \cap C \subseteq B \cap C$.
27. $B - A = B \cap A'$
28. $A \cap (B - A) = \emptyset$
29. $A \cup (B - A) = A \cup B$
30. $(A \cup B) - C = (A - C) \cup (B - C)$
31. $(A - B) \cup (A \cap B) = A$
32. $A \subseteq B$ if and only if $A \cup B = B$.
33. $A \subseteq B$ if and only if $A \cap B = A$.
34. Prove or disprove that $A \cup B = A \cup C$ implies $B = C$.
35. Prove or disprove that $A \cap B = A \cap C$ implies $B = C$.
36. Prove or disprove that $\mathcal{P}(A \cup B) = \mathcal{P}(A) \cup \mathcal{P}(B)$.
37. Prove or disprove that $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cap \mathcal{P}(B)$.
38. Prove or disprove that $\mathcal{P}(A - B) = \mathcal{P}(A) - \mathcal{P}(B)$.
39. Express $(A \cup B) - (A \cap B)$ in terms of unions and intersections that involve A , A' , B , and B' .
40. Let the operation of addition be defined on subsets A and B of U by $A + B = (A \cup B) - (A \cap B)$. Use a Venn diagram with labeled regions to illustrate each of the following statements.
 - a. $A + B = (A - B) \cup (B - A)$
 - b. $A + (B + C) = (A + B) + C$
 - c. $A \cap (B + C) = (A \cap B) + (A \cap C)$.
41. Let the operation of addition be as defined in Exercise 40. Prove each of the following statements.
 - a. $A + A = \emptyset$
 - b. $A + \emptyset = A$

1.2

Mappings

The concept of a function is fundamental to nearly all areas of mathematics. The term *function* is the one most widely used for the concept that we have in mind, but it has become traditional to use the terms *mapping* and *transformation* in algebra. It is likely that these words are used because they express an intuitive feel for the association between the elements involved. The basic idea is that correspondences of a certain type exist between

the elements of two sets. There is to be a rule of association between the elements of a first set and those of a second set. The association is to be such that for each element in the first set, there is one and only one associated element in the second set. This rule of association leads to a natural pairing of the elements that are to correspond, and then to the formal statement in Definition 1.9.

By an **ordered pair** of elements we mean a pairing (a, b) , where there is to be a distinction between the pair (a, b) and the pair (b, a) , if a and b are different. That is, there is to be a first position and a second position such that $(a, b) = (c, d)$ if and only if both $a = c$ and $b = d$. This ordering is altogether different from listing the elements of a set, for there the order of listing is of no consequence at all. The sets $\{1, 2\}$ and $\{2, 1\}$ have exactly the same elements, and $\{1, 2\} \neq \{2, 1\}$. When we speak of ordered pairs, however, we do not consider $(1, 2)$ and $(2, 1)$ equal. With these ideas in mind, we make the following definition.

Definition 1.8 ■ Cartesian[†] Product

For two nonempty sets A and B , the **Cartesian product** $A \times B$ is the set of all ordered pairs (a, b) of elements $a \in A$ and $b \in B$. That is,

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

Example 1 If $A = \{1, 2\}$ and $B = \{3, 4, 5\}$, then

$$A \times B = \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5)\}.$$

We observe that the order in which the sets appear is important. In this example,

$$B \times A = \{(3, 1), (3, 2), (4, 1), (4, 2), (5, 1), (5, 2)\},$$

so $A \times B$ and $B \times A$ are quite distinct from each other. ■

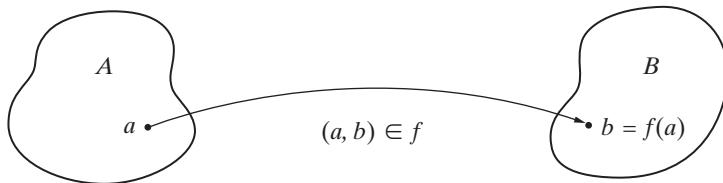
We now make our formal definition of a mapping.

Definition 1.9 ■ Mapping, Image

Let A and B be nonempty sets. A subset f of $A \times B$ is a **mapping** from A to B if and only if for each $a \in A$ there is a unique (one and only one) element $b \in B$ such that $(a, b) \in f$. If f is a mapping from A to B and the pair (a, b) belongs to f , we write $b = f(a)$ and call b the **image** of a under f .

Figure 1.4 illustrates the pairing between a and $f(a)$. A mapping f from A to B is the same as a function from A to B , and the image of $a \in A$ under f is the same as the value of the function f at a . Two mappings f from A to B and g from A to B are **equal** if and only if $f(x) = g(x)$ for all $x \in A$.

[†]The Cartesian product is named for René Descartes (1596–1650), who has been called the “Father of Modern Philosophy” and the “Father of Modern Mathematics.”

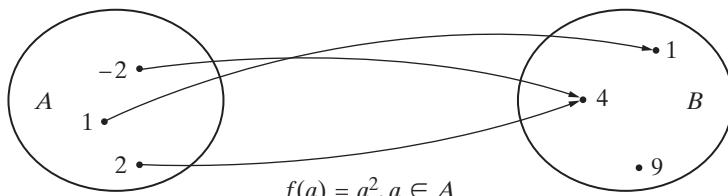


■ Figure 1.4

Example 2 Let $A = \{-2, 1, 2\}$, and let $B = \{1, 4, 9\}$. The set f given by

$$f = \{(-2, 4), (1, 1), (2, 4)\}$$

is a mapping from A to B , since for each $a \in A$ there is a unique element $b \in B$ such that $(a, b) \in f$. As is frequently the case, this mapping can be efficiently described by giving the rule for the image under f . In this case, $f(a) = a^2$, $a \in A$. This mapping is illustrated in Figure 1.5.



■ Figure 1.5

■

When it is possible to describe a mapping by giving a simple rule for the image of an element, it is certainly desirable to do so. We must keep in mind, however, that the set A , the set B , and the rule must all be known before the mapping is determined. If f is a mapping from A to B , we write $f: A \rightarrow B$ or $A \xrightarrow{f} B$ to indicate this.

Definition 1.10 ■ Domain, Codomain, Range

Let f be a mapping from A to B . The set A is called the **domain** of f , and B is called the **codomain** of f . The **range** of f is the set

$$C = \{y | y \in B \text{ and } y = f(x) \text{ for some } x \in A\}.$$

The range of f is denoted by $f(A)$.

Example 3 Let $A = \{-2, 1, 2\}$ and $B = \{1, 4, 9\}$, and let f be the mapping described in the previous example:

$$f = \{(a, b) | f(a) = a^2, a \in A\}.$$

The domain of f is A , the codomain of f is B , and the range of f is $\{1, 4\} \subset B$.

■

If $f: A \rightarrow B$, the notation used in Definition 1.10 can be extended as follows to arbitrary subsets $S \subseteq A$.

Definition 1.11 ■ Image, Inverse Image

If $f: A \rightarrow B$ and $S \subseteq A$, then

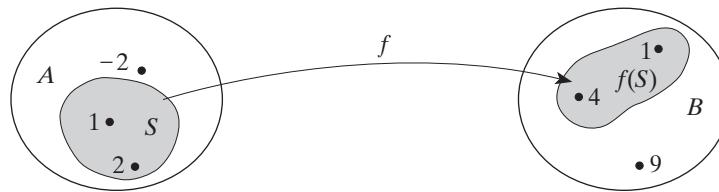
$$f(S) = \{y | y \in B \text{ and } y = f(x) \text{ for some } x \in S\}.$$

The set $f(S)$ is called the **image** of S under f . For any subset T of B , the **inverse image** of T is denoted by $f^{-1}(T)$ and is defined by

$$f^{-1}(T) = \{x | x \in A \text{ and } f(x) \in T\}.$$

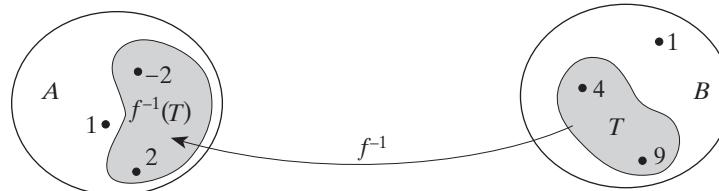
We note that the image $f(A)$ is the same as the range of f . Also, both notations $f(S)$ and $f^{-1}(T)$ in Definition 1.11 denote *sets*, not values of a mapping. We illustrate these notations in the next example.

Example 4 Let $f: A \rightarrow B$ as in Example 3. If $S = \{1, 2\}$, then $f(S) = \{1, 4\}$ as shown in Figure 1.6.



■ Figure 1.6

With $T = \{4, 9\}$, $f^{-1}(T)$ is given by $f^{-1}(T) = \{-2, 2\}$ as shown in Figure 1.7.



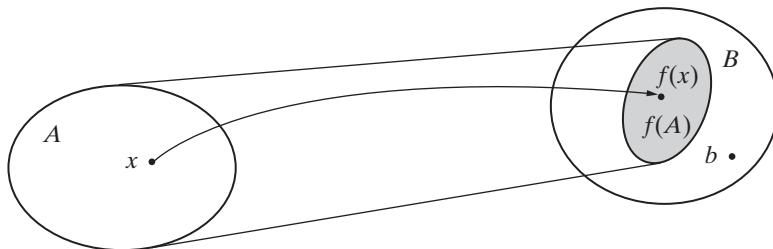
■ Figure 1.7

Among the various mappings from a nonempty set A to a nonempty set B , there are some that have properties worthy of special designation. We make the following definition.

Definition 1.12 ■ Onto, Surjective

Let $f: A \rightarrow B$. Then f is called **onto**, or **surjective**, if and only if $B = f(A)$. Alternatively, an onto mapping f is called a mapping from A **onto** B .

We begin our discussion of *onto* mappings by describing what is meant by a mapping that does not satisfy the requirement in Definition 1.12. To show that a given mapping $f: A \rightarrow B$ is **not onto**, we need only find a single element b in B for which no $x \in A$ exists such that $f(x) = b$. Such an element b and the sets A , B , and $f(A)$ are diagrammed in Figure 1.8.



■ **Figure 1.8**

Example 5 Suppose we have $f: A \rightarrow B$, where $A = \{-1, 0, 1\}$, $B = \{4, -4\}$, and $f = \{(-1, 4), (0, 4), (1, 4)\}$. The mapping f is not onto, since there is no $a \in A$ such that $f(a) = -4 \in B$. ■

Strategy

According to our definition, a mapping f from A to B is onto if and only if every element of B is the image of at least one element in A . A standard way to demonstrate that $f: A \rightarrow B$ is onto is to take an arbitrary element b in B and show (usually by some kind of formula) that there exists an element $a \in A$ such that $b = f(a)$.

Example 6 Let $f: \mathbf{Z} \rightarrow \mathbf{Z}$, where \mathbf{Z} is the set of integers. If f is defined by

$$f = \{(a, 2 - a) | a \in \mathbf{Z}\},$$

then we write $f(a) = 2 - a$, $a \in \mathbf{Z}$.

To show that f is onto (surjective), we choose an arbitrary element $b \in \mathbf{Z}$. Then there exists $2 - b \in \mathbf{Z}$ such that

$$(2 - b, b) \in f$$

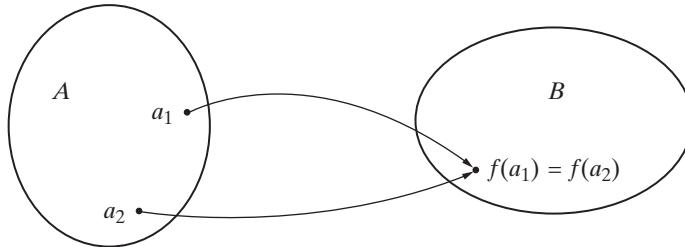
since $f(2 - b) = 2 - (2 - b) = b$, and hence f is onto. ■

Definition 1.13 ■ **One-to-One, Injective**

Let $f: A \rightarrow B$. Then f is called **one-to-one**, or **injective**, if and only if different elements of A always have different images under f .

In an approach analogous to our treatment of the onto property, we first examine the situation when a mapping fails to have the one-to-one property. To show that f is **not one-to-one**,

we need only find two elements $a_1 \in A$ and $a_2 \in A$ such that $a_1 \neq a_2$ and $f(a_1) = f(a_2)$. A pair of elements with this property is shown in Figure 1.9.



■ **Figure 1.9**

Strategy ■ The preceding discussion illustrates how *only one exception is needed to show that a given statement is false*. An example that provides such an exception is referred to as a **counterexample**.

Example 7 Suppose we reconsider the mapping $f: A \rightarrow B$ from Example 5 where $A = \{-1, 0, 1\}$, $B = \{4, -4\}$, and $f = \{(-1, 4), (0, 4), (1, 4)\}$. We see that f is not one-to-one, since

$$f(-1) = f(0) = 4 \quad \text{but} \quad -1 \neq 0. \quad \blacksquare$$

A mapping $f: A \rightarrow B$ is one-to-one if and only if it has the property that $a_1 \neq a_2$ in A always implies that $f(a_1) \neq f(a_2)$ in B . This is just a precise statement of the fact that different elements always have different images. The trouble with this statement is that it is formulated in terms of unequal quantities, whereas most of the manipulations in mathematics deal with equalities. For this reason, we take the logically equivalent *contrapositive* statement “ $f(a_1) = f(a_2)$ always implies $a_1 = a_2$ ” as our working form of the definition.

Strategy ■ We usually show that f is one-to-one by assuming that $f(a_1) = f(a_2)$ and proving that this implies that $a_1 = a_2$.

This strategy is used to show that the mapping in Example 6 is one-to-one.

Example 8 Suppose $f: \mathbf{Z} \rightarrow \mathbf{Z}$ is defined by

$$f = \{(a, 2 - a) | a \in \mathbf{Z}\}.$$

To show that f is one-to-one (injective), we assume that for $a_1 \in \mathbf{Z}$ and $a_2 \in \mathbf{Z}$,

$$f(a_1) = f(a_2).$$

Then we have

$$2 - a_1 = 2 - a_2,$$

and this implies that $a_1 = a_2$. Thus f is one-to-one. ■

Definition 1.14 ■ One-to-One Correspondence, Bijection

Let $f: A \rightarrow B$. The mapping f is called **bijective** if and only if f is both surjective and injective. A bijective mapping from A to B is called a **one-to-one correspondence** from A to B , or a **bijection** from A to B .

Example 9 The mapping $f: \mathbf{Z} \rightarrow \mathbf{Z}$ defined in Example 6 by

$$f = \{(a, 2 - a) | a \in \mathbf{Z}\}$$

is both onto and one-to-one. Thus f is a one-to-one correspondence. ■

Just after Example 11 in Section 1.1, the symbols \mathbf{Z} , \mathbf{Z}^+ , \mathbf{Q} , \mathbf{R} , \mathbf{R}^+ , and \mathbf{C} were introduced as standard notations for some of the number systems. Another set of numbers that we use often enough to justify a special notation is the set of all even integers. The set \mathbf{E} of all even integers includes 0 and all negative even integers, $-2, -4, -6, \dots$, as well as the positive even integers, $2, 4, 6, \dots$. Thus

$$\mathbf{E} = \{\dots, -6, -4, -2, 0, 2, 4, 6, \dots\},$$

and we define n to be an **even integer** if and only if $n = 2k$ for some integer k . An integer n is defined to be an **odd integer** if and only if $n = 2k + 1$ for some integer k , and the set of all odd integers is the complement of \mathbf{E} in \mathbf{Z} :

$$\mathbf{Z} - \mathbf{E} = \{\dots, -5, -3, -1, 1, 3, 5, \dots\}.$$

Note that we could also define an odd integer by using the expression $n = 2j - 1$ for some integer j .

The next two examples show that a mapping may be onto but not one-to-one, or it may be one-to-one but not onto.

Example 10 In this example, we encounter a mapping that is onto but not one-to-one. Let $h: \mathbf{Z} \rightarrow \mathbf{Z}$ be defined by

$$h(x) = \begin{cases} \frac{x-2}{2} & \text{if } x \text{ is even} \\ \frac{x-3}{2} & \text{if } x \text{ is odd.} \end{cases}$$

To attempt a proof that h is onto, let b be an arbitrary element in \mathbf{Z} and consider the equation $h(x) = b$. There are two possible values for $h(x)$, depending on whether x is even or odd. Considering both of these values, we have

$$\frac{x-2}{2} = b \quad \text{for } x \text{ even,} \quad \text{or} \quad \frac{x-3}{2} = b \quad \text{for } x \text{ odd.}$$

Solving each of these equations separately for x yields

$$x = 2b + 2 \quad \text{for } x \text{ even,} \quad \text{or} \quad x = 2b + 3 \quad \text{for } x \text{ odd.}$$

We note that $2b + 2 = 2(b + 1)$ is an even integer for every choice of b in \mathbf{Z} and that $2b + 3 = 2(b + 1) + 1$ is an odd integer for every choice of b in \mathbf{Z} . Thus there are two values, $2b + 2$ and $2b + 3$, for x in \mathbf{Z} such that

$$h(2b + 2) = b \quad \text{and} \quad h(2b + 3) = b.$$

This proves that h is onto. Since $2b + 2 \neq 2b + 3$ and $h(2b + 2) = h(2b + 3)$, we have also proved that h is not one-to-one. ■

Example 11 Consider now the mapping $f: \mathbf{Z} \rightarrow \mathbf{Z}$ defined by

$$f(x) = 2x + 1.$$

To attempt a proof that f is onto, consider an arbitrary element b in \mathbf{Z} . We have

$$\begin{aligned} f(x) = b &\Leftrightarrow 2x + 1 = b \\ &\Leftrightarrow 2x = b - 1, \end{aligned}$$

and the equation $2x = b - 1$ has a solution x in \mathbf{Z} if and only if $b - 1$ is an even integer—that is, if and only if b is an odd integer. Thus only odd integers are in the range of f , and therefore f is not onto.

The proof that f is one-to-one is straightforward:

$$\begin{aligned} f(m) = f(n) &\Rightarrow 2m + 1 = 2n + 1 \\ &\Rightarrow 2m = 2n \\ &\Rightarrow m = n. \end{aligned}$$

Thus f is one-to-one even though it is not onto. ■

In Section 3.1 and other places in our work, we need to be able to apply two mappings in succession, one after the other. In order for this successive application to be possible, the mappings involved must be compatible, as required in the next definition.

Definition 1.15 ■ Composite Mapping

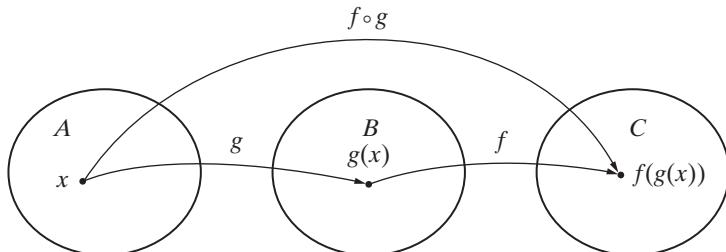
Let $g: A \rightarrow B$ and $f: B \rightarrow C$. The **composite mapping** $f \circ g$ is the mapping from A to C defined by

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

for all $x \in A$.

The process of forming the composite mapping is called **composition of mappings**, and the result $f \circ g$ is sometimes called the **composition** of g and f . Readers familiar with calculus will recognize this as the setting for the *chain rule* of derivatives.

The composite mapping $f \circ g$ is diagrammed in Figure 1.10. Note that the domain of f must contain the range of g before the composition $f \circ g$ is defined.



■ **Figure 1.10**

Example 12 Let \mathbf{Z} be the set of integers, A the set of nonnegative integers, and B the set of nonpositive integers. Suppose the mappings g and f are defined as

$$\begin{aligned} g: \mathbf{Z} &\rightarrow A, & g(x) = x^2 \\ f: A &\rightarrow B, & f(x) = -x. \end{aligned}$$

Then the composition $f \circ g$ is a mapping from \mathbf{Z} to B with

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

Note that $f \circ g$ is not onto, since $-3 \in B$, but there is no integer x such that

$$(f \circ g)(x) = -x^2 = -3.$$

Also, $f \circ g$ is not one-to-one, since

$$(f \circ g)(-2) = -(-2)^2 = -4 = (f \circ g)(2)$$

and

$$-2 \neq 2. \quad \blacksquare$$

In connection with the composition of mappings, a word of caution about notation is in order. Some mathematicians use the notation xf to indicate the image of x under f . That is, both notations xf and $f(x)$ represent the value of f at x . When the xf notation is used, mappings are applied from left to right, and the composite mapping $f \circ g$ is defined by the equation $x(f \circ g) = (xf)g$. We consistently use the $f(x)$ notation in this book, but the xf notation is found in some other texts on algebra.

When the composite mapping can be formed, we have an operation defined that is **associative**. If $h: A \rightarrow B$, $g: B \rightarrow C$, and $f: C \rightarrow D$, then

$$\begin{aligned} ((f \circ g) \circ h)(x) &= (f \circ g)(h(x)) \\ &= f[g(h(x))] \\ &= f((g \circ h)(x)) \\ &= (f \circ (g \circ h))(x) \end{aligned}$$

for all $x \in A$. Thus the compositions $(f \circ g) \circ h$ and $f \circ (g \circ h)$ are the same mapping from A to D .

Exercises 1.2

True or False

Label each of the following statements as either true or false.

1. $A \times A = A$, for every set A .
2. $A \times \emptyset = \emptyset$, for every set A .
3. Let $f: A \rightarrow B$ where A and B are nonempty. Then $f^{-1}(f(S)) = S$ for every subset S of A .
4. Let $f: A \rightarrow B$ where A and B are nonempty. Then $f(f^{-1}(T)) = T$ for every subset T of B .
5. Let $f: A \rightarrow B$. Then $f(A) = B$ for all nonempty sets A and B .
6. Every bijection is both one-to-one and onto.
7. A mapping is onto if and only if its codomain and range are equal.
8. Let $g: A \rightarrow A$ and $f: A \rightarrow A$. Then $(f \circ g)(a) = (g \circ f)(a)$ for every a in A .
9. Composition of mappings is an associative operation.

Exercises

1. For the given sets, form the indicated Cartesian product.
 - a. $A \times B; A = \{a, b\}, B = \{0, 1\}$
 - b. $B \times A; A = \{a, b\}, B = \{0, 1\}$
 - c. $A \times B; A = \{2, 4, 6, 8\}, B = \{2\}$
 - d. $B \times A; A = \{1, 5, 9\}, B = \{-1, 1\}$
 - e. $B \times A; A = B = \{1, 2, 3\}$
2. For each of the following mappings, state the domain, the codomain, and the range, where $f: E \rightarrow Z$.

a. $f(x) = x/2, x \in E$ c. $f(x) = x , x \in E$	b. $f(x) = x, x \in E$ d. $f(x) = x + 1, x \in E$
--	--
3. For each of the following mappings, write out $f(S)$ and $f^{-1}(T)$ for the given S and T , where $f: Z \rightarrow Z$.

a. $f(x) = x ; S = Z - E, T = \{1, 3, 4\}$ b. $f(x) = \begin{cases} x + 1 & \text{if } x \text{ is even} \\ x & \text{if } x \text{ is odd;} \end{cases} S = \{0, 1, 5, 9\}, T = Z - E$	c. $f(x) = x^2; S = \{-2, -1, 0, 1, 2\}, T = \{2, 7, 11\}$ d. $f(x) = x - x; S = T = \{-7, -1, 0, 2, 4\}$
---	--
4. For each of the following mappings $f: Z \rightarrow Z$, determine whether the mapping is onto and whether it is one-to-one. Justify all negative answers.

a. $f(x) = 2x$ c. $f(x) = x + 3$ e. $f(x) = x $	b. $f(x) = 3x$ d. $f(x) = x^3$ f. $f(x) = x - x $
--	--

g. $f(x) = \begin{cases} x & \text{if } x \text{ is even} \\ 2x - 1 & \text{if } x \text{ is odd} \end{cases}$

i. $f(x) = \begin{cases} x & \text{if } x \text{ is even} \\ \frac{x-1}{2} & \text{if } x \text{ is odd} \end{cases}$

h. $f(x) = \begin{cases} x & \text{if } x \text{ is even} \\ x - 1 & \text{if } x \text{ is odd} \end{cases}$

j. $f(x) = \begin{cases} x - 1 & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd} \end{cases}$

5. For each of the following mappings $f: \mathbf{R} \rightarrow \mathbf{R}$, determine whether the mapping is onto and whether it is one-to-one. Justify all negative answers. (Compare these results with the corresponding parts of Exercise 4.)

a. $f(x) = 2x$

c. $f(x) = x + 3$

e. $f(x) = |x|$

b. $f(x) = 3x$

d. $f(x) = x^3$

f. $f(x) = x - |x|$

6. For the given subsets A and B of \mathbf{Z} , let $f(x) = 2x$ and determine whether $f: A \rightarrow B$ is onto and whether it is one-to-one. Justify all negative answers.

a. $A = \mathbf{Z}, B = \mathbf{E}$

b. $A = \mathbf{E}, B = \mathbf{E}$

7. For the given subsets A and B of \mathbf{Z} , let $f(x) = |x|$ and determine whether $f: A \rightarrow B$ is onto and whether it is one-to-one. Justify all negative answers.

a. $A = \mathbf{Z}, B = \mathbf{Z}^+ \cup \{0\}$

c. $A = \mathbf{Z}^+, B = \mathbf{Z}^+$

b. $A = \mathbf{Z}^+, B = \mathbf{Z}$

d. $A = \mathbf{Z} - \{0\}, B = \mathbf{Z}^+$

8. For the given subsets A and B of \mathbf{Z} , let $f(x) = |x + 4|$ and determine whether $f: A \rightarrow B$ is onto and whether it is one-to-one. Justify all negative answers.

a. $A = \mathbf{Z}, B = \mathbf{Z}$

b. $A = \mathbf{Z}^+, B = \mathbf{Z}^+$

9. For the given subsets A and B of \mathbf{Z} , let $f(x) = 2^x$ and determine whether $f: A \rightarrow B$ is onto and whether it is one-to-one. Justify all negative answers.

a. $A = \mathbf{Z}^+, B = \mathbf{Z}$

b. $A = \mathbf{Z}^+, B = \mathbf{Z}^+ \cap \mathbf{E}$

10. For each of the following parts, give an example of a mapping from \mathbf{E} to \mathbf{E} that satisfies the given conditions.

a. one-to-one and onto

c. onto and not one-to-one

b. one-to-one and not onto

d. not one-to-one and not onto

11. For the given $f: \mathbf{Z} \rightarrow \mathbf{Z}$, determine whether f is onto and whether it is one-to-one. Prove that your conclusions are correct.

a. $f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ 0 & \text{if } x \text{ is odd} \end{cases}$

c. $f(x) = \begin{cases} 2x + 1 & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd} \end{cases}$

e. $f(x) = \begin{cases} 3x & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd} \end{cases}$

b. $f(x) = \begin{cases} 0 & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd} \end{cases}$

d. $f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{x-3}{2} & \text{if } x \text{ is odd} \end{cases}$

f. $f(x) = \begin{cases} 2x - 1 & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd} \end{cases}$

12. Let $A = \mathbf{R} - \{0\}$ and $B = \mathbf{R}$. For the given $f: A \rightarrow B$, determine whether f is onto and whether it is one-to-one. Prove that your decisions are correct.

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

c. $f(x) = \frac{x}{x^2 + 1}$

b. $f(x) = \frac{2x-1}{x}$

d. $f(x) = \frac{2x-1}{x^2 + 1}$

13. For the given $f: A \rightarrow B$, determine whether f is onto and whether it is one-to-one. Prove that your conclusions are correct.

a. $A = \mathbf{Z} \times \mathbf{Z}, B = \mathbf{Z} \times \mathbf{Z}, f(x, y) = (y, x)$

b. $A = \mathbf{Z} \times \mathbf{Z}, B = \mathbf{Z}, f(x, y) = x + y$

c. $A = \mathbf{Z} \times \mathbf{Z}, B = \mathbf{Z}, f(x, y) = x$

d. $A = \mathbf{Z}, B = \mathbf{Z} \times \mathbf{Z}, f(x) = (x, 1)$

e. $A = \mathbf{Z}^+ \times \mathbf{Z}^+, B = \mathbf{Q}, f(x, y) = x/y$

f. $A = \mathbf{R} \times \mathbf{R}, B = \mathbf{R}, f(x, y) = 2^{x+y}$

14. Let $f: \mathbf{Z} \rightarrow \{-1, 1\}$ be given by $f(x) = \begin{cases} 1 & \text{if } x \text{ is even} \\ -1 & \text{if } x \text{ is odd.} \end{cases}$

a. Prove or disprove that f is onto.

b. Prove or disprove that f is one-to-one.

c. Prove or disprove that $f(x_1 + x_2) = f(x_1)f(x_2)$.

d. Prove or disprove that $f(x_1x_2) = f(x_1)f(x_2)$.

15. a. Show that the mapping f given in Example 2 is neither onto nor one-to-one.

b. For this mapping f , show that if $S = \{1, 2\}$, then $f^{-1}(f(S)) \neq S$.

c. For this same f and $T = \{4, 9\}$, show that $f(f^{-1}(T)) \neq T$.

16. Let $g: \mathbf{Z} \rightarrow \mathbf{Z}$ be given by $g(x) = \begin{cases} x & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$

a. For $S = \{3, 4\}$, find $g(S)$ and $g^{-1}(g(S))$.

b. For $T = \{5, 6\}$, find $g^{-1}(T)$ and $g(g^{-1}(T))$.

17. Let $f: \mathbf{Z} \rightarrow \mathbf{Z}$ be given by $f(x) = \begin{cases} 2x-1 & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd.} \end{cases}$

a. For $S = \{0, 1, 2\}$, find $f(S)$ and $f^{-1}(f(S))$.

b. For $T = \{-1, 1, 4\}$, find $f^{-1}(T)$ and $f(f^{-1}(T))$.

18. Let $f: \mathbf{Z} \rightarrow \mathbf{Z}$ and $g: \mathbf{Z} \rightarrow \mathbf{Z}$ be defined as follows. In each case, compute $(f \circ g)(x)$ for arbitrary $x \in \mathbf{Z}$.

a. $f(x) = 2x, g(x) = \begin{cases} x & \text{if } x \text{ is even} \\ 2x-1 & \text{if } x \text{ is odd} \end{cases}$

b. $f(x) = 2x, g(x) = x^3$

c. $f(x) = x + |x|$, $g(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ -x & \text{if } x \text{ is odd} \end{cases}$

d. $f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ x + 1 & \text{if } x \text{ is odd} \end{cases}$ $g(x) = \begin{cases} x - 1 & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd} \end{cases}$

e. $f(x) = x^2$, $g(x) = x - |x|$

19. Let f and g be defined in the various parts of Exercise 18. In each part, compute $(g \circ f)(x)$ for arbitrary $x \in \mathbf{Z}$.

In Exercises 20–22, suppose m and n are positive integers, A is a set with exactly m elements, and B is a set with exactly n elements.

20. How many mappings are there from A to B ?
21. If $m = n$, how many one-to-one correspondences are there from A to B ?
22. If $m \leq n$, how many one-to-one mappings are there from A to B ?
23. Let a and b be constant integers with $a \neq 0$, and let the mapping $f: \mathbf{Z} \rightarrow \mathbf{Z}$ be defined by $f(x) = ax + b$.
- Prove that f is one-to-one.
 - Prove that f is onto if and only if $a = 1$ or $a = -1$.
24. Let $f: A \rightarrow B$, where A and B are nonempty.
- Prove that $f(S_1 \cup S_2) = f(S_1) \cup f(S_2)$ for all subsets S_1 and S_2 of A .
 - Prove that $f(S_1 \cap S_2) \subseteq f(S_1) \cap f(S_2)$ for all subsets S_1 and S_2 of A .
 - Give an example where there are subsets S_1 and S_2 of A such that
- $$f(S_1 \cap S_2) \neq f(S_1) \cap f(S_2).$$
- d. Prove that $f(S_1) - f(S_2) \subseteq f(S_1 - S_2)$ for all subsets S_1 and S_2 of A .
- e. Give an example where there are subsets S_1 and S_2 of A such that
- $$f(S_1) - f(S_2) \neq f(S_1 - S_2).$$
25. Let $f: A \rightarrow B$, where A and B are nonempty, and let T_1 and T_2 be subsets of B .
- Prove that $f^{-1}(T_1 \cup T_2) = f^{-1}(T_1) \cup f^{-1}(T_2)$.
 - Prove that $f^{-1}(T_1 \cap T_2) = f^{-1}(T_1) \cap f^{-1}(T_2)$.
 - Prove that $f^{-1}(T_1) - f^{-1}(T_2) = f^{-1}(T_1 - T_2)$.
 - Prove that if $T_1 \subseteq T_2$, then $f^{-1}(T_1) \subseteq f^{-1}(T_2)$.
26. Let $g: A \rightarrow B$ and $f: B \rightarrow C$. Prove that $(f \circ g)^{-1}(T) = g^{-1}(f^{-1}(T))$ for any subset T of C .

27. Let $f: A \rightarrow B$, where A and B are nonempty. Prove that f has the property that $f^{-1}(f(S)) = S$ for every subset S of A if and only if f is one-to-one. (Compare with Exercise 15b.)
28. Let $f: A \rightarrow B$, where A and B are nonempty. Prove that f has the property that $f(f^{-1}(T)) = T$ for every subset T of B if and only if f is onto. (Compare with Exercise 15c.)

1.3

Properties of Composite Mappings (Optional)

In many cases, we will be dealing with mappings of a set into itself; that is, the domain and codomain of the mappings are the same. In these cases, the mappings $f \circ g$ and $g \circ f$ are both defined, and the question of whether $f \circ g$ and $g \circ f$ are equal arises. That is, is mapping composition commutative when the domain and codomain are equal? The following example shows that the answer is no.

Example 1 Let \mathbf{Z} be the set of all integers, and let the mappings $f: \mathbf{Z} \rightarrow \mathbf{Z}$ and $g: \mathbf{Z} \rightarrow \mathbf{Z}$ be defined for each $n \in \mathbf{Z}$ by

$$\begin{aligned} f(n) &= 2n \\ g(n) &= \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ 4 & \text{if } n \text{ is odd.} \end{cases} \end{aligned}$$

In this case, the composition mappings $f \circ g$ and $g \circ f$ are both defined. We have, on the one hand,

$$\begin{aligned} (g \circ f)(n) &= g(f(n)) \\ &= g(2n) \\ &= n, \end{aligned}$$

so $(g \circ f)(n) = n$ for all $n \in \mathbf{Z}$. On the other hand,

$$\begin{aligned} (f \circ g)(n) &= f(g(n)) \\ &= \begin{cases} f\left(\frac{n}{2}\right) = n & \text{if } n \text{ is even} \\ f(4) = 8 & \text{if } n \text{ is odd,} \end{cases} \end{aligned}$$

so $f \circ g \neq g \circ f$. Thus mapping composition is not commutative. ■

In the next example we use the same functions f , g , $g \circ f$, and $f \circ g$ as in Example 1. For each of them, we determine whether the mapping is onto and whether it is one-to-one.

Example 2 Let f and g be the same as in Example 1. We see that f is one-to-one since

$$\begin{aligned} f(m) = f(n) &\Rightarrow 2m = 2n \\ &\Rightarrow m = n. \end{aligned}$$

To show that f is not onto, consider the equation $f(n) = 1$:

$$\begin{aligned} f(n) = 1 &\Rightarrow 2n = 1 \\ &\Rightarrow n = \frac{1}{2}, \end{aligned}$$

and $\frac{1}{2}$ is not an element of \mathbf{Z} . Thus 1 is not in the range of f .

We see that g is not one-to-one since

$$g(3) = 4 \quad \text{and} \quad g(5) = 4.$$

However, we can show that g is onto. For any $m \in \mathbf{Z}$, the integer $2m$ is in \mathbf{Z} and

$$\begin{aligned} g(2m) &= \frac{2m}{2} \text{ since } 2m \text{ is even} \\ &= m. \end{aligned}$$

Thus every $m \in \mathbf{Z}$ is in the range of g , and g is onto.

Using the computed values from Example 1, we have

$$(g \circ f)(n) = n$$

and

$$(f \circ g) = \begin{cases} n & \text{if } n \text{ is even} \\ 8 & \text{if } n \text{ is odd.} \end{cases}$$

The value $(g \circ f)(n) = n$ shows that $g \circ f$ is both onto and one-to-one. Since

$$(f \circ g)(1) = 8 \quad \text{and} \quad (f \circ g)(3) = 8,$$

$f \circ g$ is not one-to-one. Since $(f \circ g)(n)$ is always an even integer, there is no $n \in \mathbf{Z}$ such that

$$(f \circ g)(n) = 5,$$

and hence $f \circ g$ is not onto.

Summarizing our results, we have that

f is one-to-one and not onto.

g is onto and not one-to-one.

$g \circ f$ is both onto and one-to-one.

$f \circ g$ is neither onto nor one-to-one.

■

Considerations such as those in Example 2 raise the question of how the one-to-one and onto properties of the mappings f , g , and $f \circ g$ are related. General statements concerning these relationships are given in the next two theorems, and others can be found in the exercises.

Strategy

- To show that $f \circ g$ is onto in the proof of the next theorem, we use the standard procedure described on p. 16: We take an arbitrary $c \in C$ and prove that there exists an $a \in A$ such that $(f \circ g)(a) = c$.

Theorem 1.16 ■ Composition of Onto Mappings

Let $g: A \rightarrow B$ and $f: B \rightarrow C$. If f and g are both onto, then $f \circ g$ is onto.

$(p \wedge q) \Rightarrow r^{\dagger}$ **Proof** Suppose f and g satisfy the stated conditions. The composition $f \circ g$ maps A to C . Suppose $c \in C$. Since f is onto, there exists $b \in B$ such that

$$f(b) = c.$$

Since g is onto, every element in B is an image under g . In particular, for the specific b such that $f(b) = c$, there exists $a \in A$ such that

$$g(a) = b.$$

Hence, for $c \in C$, there exists $a \in A$ such that

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

and $f \circ g$ is onto.

Theorem 1.17 ■ Composition of One-to-One Mappings

Let $g: A \rightarrow B$ and $f: B \rightarrow C$. If f and g are both one-to-one, then $f \circ g$ is one-to-one.

$(p \wedge q) \Rightarrow r$ **Proof** Suppose f and g satisfy the stated conditions. Let a_1 and a_2 be elements in A such that

$$(f \circ g)(a_1) = (f \circ g)(a_2)$$

or

$$f(g(a_1)) = f(g(a_2)).$$

Since f is one-to-one, then

$$g(a_1) = g(a_2),$$

and since g is one-to-one, then

$$a_1 = a_2.$$

Thus $f \circ g$ is one-to-one.

The mappings in Example 3 provide a combination of properties that is different from the one in Example 2.

Example 3 Let $f: \mathbf{Z} \rightarrow \mathbf{Z}$ and $g: \mathbf{Z} \rightarrow \mathbf{Z}$ be defined as follows:

$$f(x) = \begin{cases} x & \text{if } x \text{ is even} \\ \frac{x-1}{2} & \text{if } x \text{ is odd,} \end{cases}$$

$$g(x) = 4x.$$

[†]The notation describing the logic of the proofs is defined in the Appendix.

We shall determine which of the mappings f , g , $f \circ g$, and $g \circ f$ are onto, and also which of these mappings are one-to-one.

For arbitrary $n \in \mathbf{Z}$, $2n + 1$ is odd in \mathbf{Z} , and $f(2n + 1) = n$. Thus f is onto. We have $f(2) = 2$ and also $f(5) = 2$, so f is not one-to-one.

Since $g(x)$ is always a multiple of 4, there is no $x \in \mathbf{Z}$ such that $g(x) = 3$. Hence g is not onto. However,

$$\begin{aligned} g(x) = g(z) &\Rightarrow 4x = 4z \\ &\Rightarrow x = z, \end{aligned}$$

so g is one-to-one.

Now

$$(f \circ g)(x) = f(g(x)) = f(4x) = 4x.$$

This means that $(f \circ g)(x) = g(x)$ for all $x \in \mathbf{Z}$. Therefore, $f \circ g = g$ is not onto and is one-to-one.

Computing $(g \circ f)(x)$, we obtain

$$\begin{aligned} (g \circ f)(x) &= g(f(x)) \\ &= \begin{cases} g(x) & \text{if } x \text{ is even} \\ g\left(\frac{x-1}{2}\right) & \text{if } x \text{ is odd} \end{cases} \\ &= \begin{cases} 4x & \text{if } x \text{ is even} \\ 2(x-1) & \text{if } x \text{ is odd.} \end{cases} \end{aligned}$$

Since $(g \circ f)(x)$ is never odd, there is no x such that $(g \circ f)(x) = 1$, and $g \circ f$ is not onto. Also, since $(g \circ f)(2) = 8$ and $(g \circ f)(5) = 8$, $g \circ f$ is not one-to-one.

We can summarize our results as follows:

- f is onto and not one-to-one.
- g is one-to-one and not onto.
- $f \circ g$ is one-to-one and not onto.
- $g \circ f$ is neither onto nor one-to-one.



Exercises 1.3

True or False

Label each of the following statements as either true or false.

1. Mapping composition is a commutative operation.
2. The composition of two bijections is also a bijection.
3. Let f , g , and h be mappings from A into A such that $f \circ g = h \circ g$. Then $f = h$.
4. Let f , g , and h be mappings from A into A such that $f \circ g = f \circ h$. Then $g = h$.

5. Let $g: A \rightarrow B$ and $f: B \rightarrow C$ such that $f \circ g$ is onto. Then both f and g are onto.
6. Let $g: A \rightarrow B$ and $f: B \rightarrow C$ such that $f \circ g$ is one-to-one. Then both f and g are one-to-one.
-

Exercises

1. For each of the following pairs $f: \mathbf{Z} \rightarrow \mathbf{Z}$ and $g: \mathbf{Z} \rightarrow \mathbf{Z}$, decide whether $f \circ g$ is onto or one-to-one and justify all negative answers.
 - a. $f(x) = 2x, \quad g(x) = \begin{cases} x & \text{if } x \text{ is even} \\ 2x - 1 & \text{if } x \text{ is odd} \end{cases}$
 - b. $f(x) = 2x, \quad g(x) = x^3$
 - c. $f(x) = x + |x|, \quad g(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ -x & \text{if } x \text{ is odd} \end{cases}$
 - d. $f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ x + 1 & \text{if } x \text{ is odd} \end{cases}, \quad g(x) = \begin{cases} x - 1 & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd} \end{cases}$
 - e. $f(x) = x^2, \quad g(x) = x - |x|$
2. For each pair f, g given in Exercise 1, decide whether $g \circ f$ is onto or one-to-one, and justify all negative answers.
3. Give an example of mappings f and g such that one of f or g is not onto but $f \circ g$ is onto.
4. Give an example of mappings f and g , different from those in Example 3, such that one of f or g is not one-to-one but $f \circ g$ is one-to-one.
5. a. Give an example of mappings f and g , different from those in Example 2, where f is one-to-one, g is onto, and $f \circ g$ is not one-to-one.
b. Give an example of mappings f and g , different from those in Example 2, where f is one-to-one, g is onto, and $f \circ g$ is not onto.
6. a. Give an example of mappings f and g , where f is onto, g is one-to-one, and $f \circ g$ is not one-to-one.
b. Give an example of mappings f and g , different from those in Example 3, where f is onto, g is one-to-one, and $f \circ g$ is not onto.
7. Suppose f, g , and h are all mappings of a set A into itself.
 - a. Prove that if g is onto and $f \circ g = h \circ g$, then $f = h$.
 - b. Prove that if f is one-to-one and $f \circ g = f \circ h$, then $g = h$.
8. a. Find mappings f, g , and h of a set A into itself such that $f \circ g = h \circ g$ and $f \neq h$.
b. Find mappings f, g , and h of a set A into itself such that $f \circ g = f \circ h$ and $g \neq h$.

9. Let $g: A \rightarrow B$ and $f: B \rightarrow C$. Prove that f is onto if $f \circ g$ is onto.
10. Let $g: A \rightarrow B$ and $f: B \rightarrow C$. Prove that g is one-to-one if $f \circ g$ is one-to-one.
11. Let $f: A \rightarrow B$ and $g: B \rightarrow A$. Prove that f is one-to-one and onto if $f \circ g$ is one-to-one and $g \circ f$ is onto.

1.4

Binary Operations

We are familiar with the operations of addition, subtraction, and multiplication on real numbers. These are examples of *binary operations*. When we speak of a binary operation on a set, we have in mind a process that combines two elements of the set to produce a third element of the set. This third element, the result of the operation on the first two, must be unique. That is, there must be one and only one result from the combination. Also, it must always be possible to combine the two elements, no matter which two are chosen. This discussion is admittedly a bit vague, in that the terms *process* and *combine* are somewhat indefinite. To eliminate this vagueness, we make the following formal definition.

Definition 1.18 ■ Binary Operation

A **binary operation** on a nonempty set A is a mapping f from $A \times A$ to A .

It is conventional in mathematics to assume that when a formal definition is made, it is automatically biconditional. That is, it is understood to be an “if and only if” statement, without this being written out explicitly. In Definition 1.18, for example, it is understood as part of the definition that f is a binary operation on the nonempty set A if and only if f is a mapping from $A \times A$ to A . Throughout the remainder of this book, we will adhere to this convention when we make definitions.

We now have a precise definition of the term *binary operation*, but some of the feel for the concept may have been lost. However, the definition gives us what we want. Suppose f is a mapping from $A \times A$ to A . Then $f(x, y)$ is defined for every ordered pair (x, y) of elements of A , and the image $f(x, y)$ is unique. In other words, we can combine any two elements x and y of A to obtain a unique third element of A by finding the value $f(x, y)$. The result of performing the binary operation on x and y is $f(x, y)$, and the only thing unfamiliar about this is the notation for the result. We are accustomed to indicating results of binary operations by symbols such as $x + y$ and $x - y$. We can use a similar notation and write $x * y$ in place of $f(x, y)$. Thus $x * y$ represents the result of an arbitrary binary operation $*$ on A , just as $f(x, y)$ represents the value of an arbitrary mapping from $A \times A$ to A .

Example 1 Two examples of binary operations on \mathbf{Z} are the mappings from $\mathbf{Z} \times \mathbf{Z}$ to \mathbf{Z} , defined as follows:

1. $x * y = x + y - 1$, for $(x, y) \in \mathbf{Z} \times \mathbf{Z}$.
2. $x * y = 1 + xy$, for $(x, y) \in \mathbf{Z} \times \mathbf{Z}$.



Example 2 The operation of forming the intersection $A \cap B$ of subsets A and B of a universal set U is a binary operation on the collection of all subsets of U . This is also true of the operation of forming the union. ■

Since we are dealing with ordered pairs in connection with a binary operation, the results $x * y$ and $y * x$ may well be different.

Definition 1.19 ■ Commutativity, Associativity

If $*$ is a binary operation on the nonempty set A , then $*$ is called **commutative** if $x * y = y * x$ for all x and y in A . If $x * (y * z) = (x * y) * z$ for all x, y, z in A , then we say that the binary operation is **associative**.

Example 3 The usual binary operations of addition and multiplication on the integers are both commutative and associative. However, the binary operation of subtraction on the integers does not have either of these properties. For example, $5 - 7 \neq 7 - 5$, and $9 - (8 - 3) \neq (9 - 8) - 3$. ■

Suppose we consider the two binary operations given in Example 1.

Example 4 The binary operation $*$ defined on \mathbf{Z} by

$$x * y = x + y - 1$$

is commutative, since

$$x * y = x + y - 1 = y + x - 1 = y * x.$$

Note that $*$ is also associative, since

$$\begin{aligned} x * (y * z) &= x * (y + z - 1) \\ &= x + (y + z - 1) - 1 \\ &= x + y + z - 2 \end{aligned}$$

and

$$\begin{aligned} (x * y) * z &= (x + y - 1) * z \\ &= (x + y - 1) + z - 1 \\ &= x + y + z - 2. \end{aligned}$$

Example 5 The binary operation $*$ defined on \mathbf{Z} by

$$x * y = 1 + xy$$

is commutative, since

$$x * y = 1 + xy = 1 + yx = y * x.$$

To check whether $*$ is associative, we compute

$$x * (y * z) = x * (1 + yz) = 1 + x(1 + yz) = 1 + x + xyz$$

and

$$(x * y) * z = (1 + xy) * z = 1 + (1 + xy)z = 1 + z + xyz.$$

Thus we can demonstrate that $*$ is not associative by choosing x , y , and z in \mathbf{Z} with $x \neq z$. Using $x = 1$, $y = 2$, $z = 3$, we get

$$1 * (2 * 3) = 1 * (1 + 6) = 1 * 7 = 1 + 7 = 8$$

and

$$(1 * 2) * 3 = (1 + 2) * 3 = 3 * 3 = 1 + 9 = 10.$$

Hence $*$ is not associative on \mathbf{Z} . ■

The commutative and associative properties are properties of the binary operation itself. In contrast, the property described in the next definition depends on the set under consideration as well as on the binary operation.

Definition 1.20 ■ Closure

Suppose that $*$ is a binary operation on a nonempty set A , and let $B \subseteq A$. If $x * y$ is an element of B for all $x \in B$ and $y \in B$, then B is **closed** with respect to $*$.

In the special case where $B = A$ in Definition 1.20, the property of being closed is automatic, since the result $x * y$ is required to be in A by the definition of a binary operation on A .

Example 6 Consider the binary operation $*$ defined on the set of integers \mathbf{Z} by

$$x * y = |x| + |y|, \quad (x, y) \in \mathbf{Z} \times \mathbf{Z}.$$

The set B of negative integers is not closed with respect to $*$, since $x = -1 \in B$ and $y = -2 \in B$, but

$$x * y = (-1) * (-2) = |-1| + |-2| = 3 \notin B. \quad \blacksquare$$

Example 7 The definition of an odd integer that was stated in Section 1.2 can be used to prove that the set S of all odd integers is closed under multiplication.

Let x and y be arbitrary odd integers. According to the definition of an odd integer, this means that $x = 2m + 1$ for some integer m and $y = 2n + 1$ for some integer n . Forming the product, we obtain

$$\begin{aligned} xy &= (2m + 1)(2n + 1) \\ &= 4mn + 2m + 2n + 1 \\ &= 2(2mn + m + n) + 1 \\ &= 2k + 1, \end{aligned}$$

where $k = 2mn + m + n \in \mathbf{Z}$, and therefore xy is an odd integer. Hence the set S of all odd integers is closed with respect to multiplication. ■

Definition 1.21 ■ Identity Element

Let $*$ be a binary operation on the nonempty set A . An element e in A is called an **identity element** with respect to the binary operation $*$ if e has the property that

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

for all $x \in A$.

The phrase “with respect to the binary operation” is critical in this definition because the particular binary operation being considered is all-important. This is pointed out in the next example.

Example 8 The integer 1 is an identity with respect to the operation of multiplication ($1 \cdot x = x \cdot 1 = x$), but not with respect to the operation of addition ($1 + x \neq x$). ■

Example 9 The element 1 is the identity element with respect to the binary operation $*$ given by

$$x * y = x + y - 1, \quad (x, y) \in \mathbf{Z} \times \mathbf{Z},$$

since

$$x * 1 = x + 1 - 1 = x \quad \text{and} \quad 1 * x = 1 + x - 1 = x. \quad \blacksquare$$

Example 10 There is no identity element with respect to the binary operation $*$ defined by

$$x * y = 1 + xy, \quad (x, y) \in \mathbf{Z} \times \mathbf{Z},$$

since there is no fixed integer z such that

$$x * z = z * x = 1 + xz = x, \quad \text{for all } x \in \mathbf{Z}. \quad \blacksquare$$

Whenever a set has an identity element with respect to a binary operation on the set, it is in order to raise the question of inverses.

Definition 1.22 ■ Right Inverse, Left Inverse, Inverse

Suppose that e is an identity element for the binary operation $*$ on the set A , and let $a \in A$. If there exists an element $b \in A$ such that $a * b = e$, then b is called a **right inverse** of a with respect to this operation. Similarly, if $b * a = e$, then b is called a **left inverse** of a . If both of $a * b = e$ and $b * a = e$ hold, then b is called an **inverse of a** , and a is called an **invertible** element of A .

Sometimes an inverse is referred to as a *two-sided inverse* to emphasize that both of the required equations hold.

Strategy ■ Exercise 13 of this section requests a proof that the inverse of an element with respect to an associative binary operation is unique. A standard way to prove the uniqueness of an entity is to assume that two such entities exist and then prove the two to be equal.

Example 11 Each element $x \in \mathbf{Z}$ has a two-sided inverse $(-x + 2) \in \mathbf{Z}$ with respect to the binary operation $*$ given by

$$x * y = x + y - 1, \quad (x, y) \in \mathbf{Z} \times \mathbf{Z},$$

since

$$x * (-x + 2) = (-x + 2) * x = -x + 2 + x - 1 = 1 = e. \blacksquare$$

Exercises 1.4

True or False

Label each of the following statements as either true or false.

1. If a binary operation on a nonempty set A is commutative, then an identity element will exist in A .
2. If $*$ is a binary operation on a nonempty set A , then A is closed with respect to $*$.
3. Let $A = \{a, b, c\}$. The power set $\mathcal{P}(A)$ is closed with respect to the binary operation \cap of forming intersections.
4. Let $A = \{a, b, c\}$. The empty set \emptyset is the identity element in $\mathcal{P}(A)$ with respect to the binary operation \cap .
5. Let $A = \{a, b, c\}$. The power set $\mathcal{P}(A)$ is closed with respect to the binary operation \cup of forming unions.
6. Let $A = \{a, b, c\}$. The empty set \emptyset is the identity element in $\mathcal{P}(A)$ with respect to the binary operation \cup .
7. Any binary operation defined on a set containing a single element is commutative and associative.
8. An identity and inverses exist in a set containing a single element upon which a binary operation is defined.
9. The set of all bijections from A to A is closed with respect to the binary operation of composition defined on the set of all mappings from A to A .

Exercises

1. Decide whether the given set B is closed with respect to the binary operation defined on the set of integers \mathbf{Z} . If B is not closed, exhibit elements $x \in B$ and $y \in B$, such that $x * y \notin B$.
 - a. $x * y = xy, \quad B = \{-1, -2, -3, \dots\}$
 - b. $x * y = x - y, \quad B = \mathbf{Z}^+$

c. $x * y = x^2 + y^2$, $B = \mathbf{Z}^+$

d. $x * y = \operatorname{sgn} x + \operatorname{sgn} y$, $B = \{-2, -1, 0, 1, 2\}$ where $\operatorname{sgn} x = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$

e. $x * y = |x| - |y|$, $B = \mathbf{Z}^+$

f. $x * y = x + xy$, $B = \mathbf{Z}^+$

g. $x * y = xy - x - y$, B is the set of all odd integers.

h. $x * y = x^y$, B is the set of positive odd integers.

2. In each part following, a rule is given that determines a binary operation $*$ on the set \mathbf{Z} of all integers. Determine in each case whether the operation is commutative or associative and whether there is an identity element. Also find the inverse of each invertible element.

a. $x * y = x + xy$

b. $x * y = x$

c. $x * y = x + 2y$

d. $x * y = 3(x + y)$

e. $x * y = 3xy$

f. $x * y = x - y$

g. $x * y = x + xy + y$

h. $x * y = x + y + 3$

i. $x * y = x - y + 1$

j. $x * y = x + xy + y - 2$

k. $x * y = |x| - |y|$

l. $x * y = |x - y|$

m. $x * y = x^y$ for $x, y \in \mathbf{Z}^+$

n. $x * y = 2^{xy}$ for $x, y \in \mathbf{Z}^+$

3. Let S be a set of three elements given by $S = \{A, B, C\}$. In the following table, all of the elements of S are listed in a row at the top and in a column at the left. The result $x * y$ is found in the row that starts with x at the left and in the column that has y at the top. For example, $B * C = C$ and $C * B = A$.

*	A	B	C
A	C	A	B
B	A	B	C
C	B	A	C

a. Is the binary operation $*$ commutative? Why?

b. Determine whether there is an identity element in S for $*$.

Sec. 1.6, #8 <

c. If there is an identity element, which elements have inverses?

4. Let S be the set of three elements given by $S = \{A, B, C\}$ with the following table.

*	A	B	C
A	A	B	C
B	B	C	A
C	C	A	B

- a. Is the binary operation $*$ commutative? Why?
- b. Determine whether there is an identity element in S for $*$.
- c. If there is an identity element, which elements have inverses?
5. Let S be a set of four elements given by $S = \{A, B, C, D\}$ with the following table.
- | $*$ | A | B | C | D |
|-----|-----|-----|-----|-----|
| A | B | C | A | B |
| B | C | D | B | A |
| C | A | B | C | D |
| D | A | B | D | D |
- a. Is the binary operation $*$ commutative? Why?
- b. Determine whether there is an identity element in S for $*$.
- c. If there is an identity element, which elements have inverses?
6. Let S be the set of four elements given by $S = \{A, B, C, D\}$ with the following table.
- | $*$ | A | B | C | D |
|-----|-----|-----|-----|-----|
| A | A | A | A | A |
| B | A | B | A | B |
| C | A | A | C | C |
| D | A | B | C | D |

- a. Is the binary operation $*$ commutative? Why?
- b. Determine whether there is an identity element in S for $*$.
- c. If there is an identity element, which elements have inverses?
7. Prove or disprove that the set of nonzero integers is closed with respect to division.
8. Prove or disprove that the set of all odd integers is closed with respect to addition.
9. The definition of an even integer was stated in Section 1.2. Prove or disprove that the set \mathbf{E} of all even integers is closed with respect to
- addition
 - multiplication.
10. Assume that $*$ is an associative binary operation on the nonempty set A . Prove that

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

for all a, b, c , and d in A .

11. Assume that $*$ is a binary operation on a nonempty set A , and suppose that $*$ is both commutative and associative. Use the definitions of the commutative and associative properties to show that

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

for all a, b, c , and d in A .

12. Let $*$ be a binary operation on the nonempty set A . Prove that if A contains an identity element with respect to $*$, the identity element is unique. (*Hint:* Assume that both e_1 and e_2 are identity elements for $*$, and then prove that $e_1 = e_2$.)
13. Assume that $*$ is an associative binary operation on A with an identity element. Prove that the inverse of an element is unique when it exists.

1.5

Permutations and Inverses

The set of all mappings of a set into itself is of special interest, and we consider such a set next.

Definition 1.23 ■ Permutation

A one-to-one correspondence from a set A to itself is called a **permutation** on A . For any nonempty set A , we adopt the notation $\mathcal{S}(A)$ as standard for the set of all permutations on A . The set of all mappings from A to A will be denoted by $\mathcal{M}(A)$.

From the discussion at the end of Section 1.2, we know that composition of mappings is an associative binary operation on $\mathcal{M}(A)$. The **identity mapping** I_A is defined by

$$I_A(x) = x \quad \text{for all } x \in A.$$

For any f in $\mathcal{M}(A)$,

$$(I_A \circ f)(x) = I_A(f(x)) = f(x)$$

and

$$(f \circ I_A)(x) = f(I_A(x)) = f(x),$$

so $I_A \circ f = f \circ I_A = f$. That is, I_A is an identity element for mapping composition. Once an identity element is established for a binary operation, the next natural question is whether inverses exist. Consider the mappings in the next example.

Example 1 In Example 1 of Section 1.3, we defined the mappings $f: \mathbf{Z} \rightarrow \mathbf{Z}$ and $g: \mathbf{Z} \rightarrow \mathbf{Z}$ by

$$f(n) = 2n$$

and

$$g(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ 4 & \text{if } n \text{ is odd.} \end{cases}$$

For these mappings, $(g \circ f)(n) = n$ for all $n \in \mathbb{Z}$, so $g \circ f = I_{\mathbb{Z}}$ and g is a left inverse for f . Note, however, that

$$(f \circ g)(n) = \begin{cases} n & \text{if } n \text{ is even} \\ 8 & \text{if } n \text{ is odd.} \end{cases}$$

Thus $f \circ g \neq I_{\mathbb{Z}}$, and g is not a right inverse for f . ■

Example 1 furnishes some insight into the next two lemmas.[†]

Strategy

- Each of these lemmas makes a statement of the form “ p if and only if q .” For this kind of statement, there are two things to be proved:
 1. ($p \Leftarrow q$) The “if” part, where we assume q is true and prove that p must then be true, and
 2. ($p \Rightarrow q$) The “only if” part, where we assume that p is true and prove that q must then be true.

Lemma 1.24

■ Left Inverses and the One-to-One Property

Let A be a nonempty set, and let $f: A \rightarrow A$. Then f is one-to-one if and only if f has a left inverse.

$p \Leftarrow q$ **Proof** Assume first that f has a left inverse g , and suppose that $f(a_1) = f(a_2)$. Since $g \circ f = I_A$, we have

$$\begin{aligned} a_1 &= I_A(a_1) = (g \circ f)(a_1) = g(f(a_1)) = g(f(a_2)) \\ &= (g \circ f)(a_2) = I_A(a_2) = a_2. \end{aligned}$$

Thus $f(a_1) = f(a_2)$ implies $a_1 = a_2$, and f is one-to-one.

$p \Rightarrow q$ Conversely, now assume that f is one-to-one. We shall define a left inverse g of f . Let a_0 represent an arbitrarily chosen but fixed element in A . For each x in A , $g(x)$ is defined by this rule:

1. If there is an element y in A such that $f(y) = x$, then $g(x) = y$.
2. If no such element y exists in A , then $g(x) = a_0$.

[†]A **lemma** is a proposition whose main purpose is to help prove another proposition.

When the first part of the rule applies, the element y is unique because f is one-to-one ($f(y_1) = x = f(y_2) \Rightarrow y_1 = y_2 = g(x)$). Thus $g(x)$ is unique in this case. When the second part of the rule applies, $g(x) = a_0$ is surely unique, and g is a mapping from A to A . For all a in A , we have

$$(g \circ f)(a) = g(f(a)) = a$$

because $x = f(a)$ requires $g(x) = a$. Thus g is a left inverse for f .

There is a connection between the onto property and right inverses that is similar to the one between the one-to-one property and left inverses. This connection is stated in Lemma 1.25, and its proof involves using the **Axiom of Choice**. In one of its simplest forms, this axiom states that it is possible to make a choice of an element from each of the sets in a nonempty collection of nonempty sets. We assume the Axiom of Choice in this text, and it should be noted that this is an *assumption*.

Lemma 1.25

■ Right Inverses and the Onto Property

Let A be a nonempty set, and $f: A \rightarrow A$. Then f is an onto mapping if and only if f has a right inverse.

$p \Leftarrow q$ **Proof** Assume that f has a right inverse g , and let a_0 be an arbitrarily chosen element of A . Now $g(a_0)$ is an element of A , and

$$\begin{aligned} f(g(a_0)) &= (f \circ g)(a_0) \\ &= I_A(a_0) \quad \text{since } g \text{ is a right inverse of } f \\ &= a_0. \end{aligned}$$

Thus a_0 is the image of $g(a_0)$ under f , and this proves that f is onto if f has a right inverse.

$p \Rightarrow q$ Let us assume now that f is onto, and we shall define a right inverse of f as follows: Let a_0 be an arbitrary element of A . Since f is onto, there exists at least one element x of A such that $f(x) = a_0$. Choose[†] one of these elements, say, x_0 , and define $g(a_0)$ by

$$g(a_0) = x_0.$$

For each a_0 in A , we have a unique value $g(a_0)$ such that

$$\begin{aligned} (f \circ g)(a_0) &= f(g(a_0)) \\ &= f(x_0) \\ &= a_0 \quad \text{by the choice of } x_0. \end{aligned}$$

Therefore, $f \circ g = I_A$, and g is a right inverse of f .

Lemmas 1.24 and 1.25 enable us to prove the following important theorem.

[†]The Axiom of Choice implies that this is possible.

Theorem 1.26 ■ Inverses and Permutations

Let $f: A \rightarrow A$. Then f is invertible if and only if f is a permutation on A .

$p \Rightarrow q$ **Proof** If f has an inverse g , then $g \circ f = I_A$ and $f \circ g = I_A$. Note that $g \circ f = I_A$ implies that f is one-to-one by Lemma 1.24, and $f \circ g = I_A$ implies that f is onto by Lemma 1.25. Thus f is a permutation on A .

$p \Leftarrow q$ Now suppose that f is a permutation on A . Then f has a left inverse g by Lemma 1.24 and a right inverse h by Lemma 1.25. We have $g \circ f = I_A$ and $f \circ h = I_A$, so

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

That is, $g = h$, and f has an inverse.

The last theorem shows that the members of the set $\mathcal{S}(A)$ are special in that each of them is invertible. From Exercise 13 of the last section, we know that the inverse of an element with respect to an associative binary operation is unique. Thus we denote the unique inverse of a permutation f by f^{-1} . It is left as an exercise to prove that f^{-1} is a permutation on A .

There is one other property of the set $\mathcal{S}(A)$ that is significant. Whenever f and g are in $\mathcal{S}(A)$, then $f \circ g$ is also in $\mathcal{S}(A)$. (See Exercise 8 of this section.) Thus $\mathcal{S}(A)$ is *closed* under mapping composition.

Some of the preceding results are illustrated in the following example.

Example 2 From Example 11 of Section 1.2, we know that the mapping $f: \mathbf{Z} \rightarrow \mathbf{Z}$ defined by

$$f(x) = 2x + 1$$

is one-to-one and not onto. According to Lemmas 1.24 and 1.25, f has a left inverse but fails to have a right inverse. The two-part rule for g in the proof of Lemma 1.24 can be used as a guide in defining a left inverse of the f under consideration here.

The first part of the rule reads as follows: If there is an element y in \mathbf{Z} such that $f(y) = x$, then $g(x) = y$. Since we have $f(x) = 2x + 1$ here, the equation $f(y) = x$ requires that x be odd and that $2y + 1 = x$. Solving this equation for y , we obtain

$$y = \frac{x - 1}{2}.$$

Thus the equation $g(x) = y$ becomes

$$g(x) = \frac{x - 1}{2} \quad \text{for } x \text{ odd}$$

in this instance.

According to the second part of the rule for g in the proof of Lemma 1.24, we may choose an arbitrary fixed a_0 in \mathbf{Z} and define $g(x) = a_0$ when x is not in the range of f . Choosing $a_0 = 4$ gives us a left inverse g of f defined as follows:

$$g(x) = \begin{cases} \frac{x-1}{2} & \text{if } x \text{ is odd} \\ 4 & \text{if } x \text{ is even.} \end{cases}$$

■

Exercises 1.5

True or False

Label each of the following statements as either true or false.

1. Every permutation has an inverse.
2. Let $A \neq \emptyset$ and $f: A \rightarrow A$. Then f is one-to-one if and only if f has a right inverse.
3. Let $A \neq \emptyset$ and $f: A \rightarrow A$. Then f is onto if and only if f has a left inverse.

Exercises

1. For each of the following mappings $f: \mathbf{Z} \rightarrow \mathbf{Z}$, exhibit a right inverse of f with respect to mapping composition whenever one exists.
 - a. $f(x) = 2x$
 - b. $f(x) = 3x$
 - c. $f(x) = x + 2$
 - d. $f(x) = 1 - x$
 - e. $f(x) = x^3$
 - f. $f(x) = x^2$
 - g. $f(x) = \begin{cases} x & \text{if } x \text{ is even} \\ 2x - 1 & \text{if } x \text{ is odd} \end{cases}$
 - h. $f(x) = \begin{cases} x & \text{if } x \text{ is even} \\ x - 1 & \text{if } x \text{ is odd} \end{cases}$
 - i. $f(x) = |x|$
 - j. $f(x) = x - |x|$
 - k. $f(x) = \begin{cases} x & \text{if } x \text{ is even} \\ \frac{x-1}{2} & \text{if } x \text{ is odd} \end{cases}$
 - l. $f(x) = \begin{cases} x - 1 & \text{if } x \text{ is even} \\ 2x & \text{if } x \text{ is odd} \end{cases}$
 - m. $f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ x + 2 & \text{if } x \text{ is odd} \end{cases}$
 - n. $f(x) = \begin{cases} x + 1 & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd} \end{cases}$
2. For each of the mappings f given in Exercise 1, determine whether f has a left inverse. Exhibit a left inverse whenever one exists.
3. If n is a positive integer and the set A has n elements, how many elements are in the set $S(A)$ of all permutations on A ?
4. Let $f: A \rightarrow A$, where A is nonempty. Prove that f has a left inverse if and only if $f^{-1}(f(S)) = S$ for every subset S of A .

5. Let $f: A \rightarrow A$, where A is nonempty. Prove that f has a right inverse if and only if $f(f^{-1}(T)) = T$ for every subset T of A .
6. Prove that if f is a permutation on A , then f^{-1} is a permutation on A .
7. Prove that if f is a permutation on A , then $(f^{-1})^{-1} = f$.
8.
 - a. Prove that the set of all onto mappings from A to A is closed under composition of mappings.
 - b. Prove that the set of all one-to-one mappings from A to A is closed under mapping composition.
9. Let f and g be permutations on A . Prove that $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.
10. Let f and g be mappings from A to A . Prove that if $f \circ g$ is invertible, then f is onto and g is one-to-one.

1.6 Matrices

The material in this section provides a rich source of examples for many of the concepts treated later in the text. The basic element under consideration here will be a **matrix** (plural **matrices**).

The word *matrix* is used in mathematics to denote a rectangular array of elements in rows and columns. The elements in the array are usually numbers, and brackets may be used to mark the beginning and the end of the array. Two illustrations of this type of matrix are

$$\begin{bmatrix} 5 & -1 & 0 & 3 \\ 2 & 1 & -2 & 7 \\ 4 & -6 & 4 & 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 9 & 1 \\ -1 & 0 \\ 6 & -3 \end{bmatrix}.$$

The formal notation for a matrix is introduced in the following definition. We shall soon see that this notation is extremely useful in proving certain facts about matrices.

Definition 1.27 ■ Matrix

An **m by n matrix** over a set S is a rectangular array of elements of S , arranged in m rows and n columns. It is customary to write an m by n matrix using notation such as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

where the uppercase letter A denotes the matrix and the lowercase a_{ij} denotes the element in row i and column j of the matrix A . The rows are numbered from the top down, and the columns are numbered from left to right. The matrix A is referred to as a matrix of **dimension $m \times n$** (read “ m by n ”).

The $m \times n$ matrix A in Definition 1.27 can be written compactly as $A = [a_{ij}]_{m \times n}$ or simply as $A = [a_{ij}]$ if the dimension is known from the context.

Example 1 In compact notation, $B = [b_{ij}]_{2 \times 4}$ is shorthand for the matrix

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \end{bmatrix}.$$

As a more concrete example, the matrix A defined by $A = [a_{ij}]_{3 \times 3}$ with $a_{ij} = (-1)^{i+j}$ would appear written out as

$$A = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

(This matrix describes the sign pattern in the cofactor expansion of third-order determinants that is used with Cramer's Rule for solving systems of linear equations in intermediate algebra.) ■

An $n \times n$ matrix is called a **square matrix of order n** , and a square matrix $A = [a_{ij}]_{n \times n}$ with $a_{ij} = 0$ whenever $i \neq j$ is known as a **diagonal matrix**. The matrices

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 8 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

are diagonal matrices.

Definition 1.28 ■ Matrix Equality

Two matrices $A = [a_{ij}]_{m \times n}$ and $B = [b_{ij}]_{p \times q}$ over a set S are **equal** if and only if $m = p$, $n = q$, and $a_{ij} = b_{ij}$ for all pairs i, j .

The set of all $m \times n$ matrices over S will be denoted in this book by $M_{m \times n}(S)$. When $m = n$, we simply write $M_n(S)$ instead of $M_{n \times n}(S)$. For the remainder of this section, we will restrict our attention to the sets $M_{m \times n}(\mathbf{R})$, where \mathbf{R} is the set of all real numbers. Our goal is to define binary operations of addition and multiplication on certain sets of matrices and to investigate the basic properties of these operations.

Definition 1.29 ■ Matrix Addition

Addition in $M_{m \times n}(\mathbf{R})$ is defined by

$$[a_{ij}]_{m \times n} + [b_{ij}]_{m \times n} = [c_{ij}]_{m \times n}$$

where $c_{ij} = a_{ij} + b_{ij}$.

To form the sum of two elements in $M_{m \times n}(\mathbf{R})$, we simply add the elements that are placed in corresponding positions.

Example 2 In $M_{2 \times 3}(\mathbf{R})$, an example of addition is

$$\begin{bmatrix} 3 & -1 & 1 \\ 2 & -7 & -4 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 1 \\ 3 & -4 & -5 \end{bmatrix}.$$

We note that a sum of two matrices with *different* dimensions is *not defined*. For instance, the sum

$$\begin{bmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

is undefined because the dimensions of the two matrices involved are not equal. ■

Definition 1.29 can be written in shorter form as

$$[a_{ij}]_{m \times n} + [b_{ij}]_{m \times n} = [a_{ij} + b_{ij}]_{m \times n},$$

and this shorter form is efficient to use in proving the basic properties of addition in $M_{m \times n}(\mathbf{R})$. These basic properties are stated in the next theorem.

Theorem 1.30 ■ Properties of Matrix Addition

Addition in $M_{m \times n}(\mathbf{R})$ has the following properties.

- a. Addition as defined in Definition 1.29 is a binary operation on $M_{m \times n}(\mathbf{R})$.
- b. Addition is associative in $M_{m \times n}(\mathbf{R})$.
- c. $M_{m \times n}(\mathbf{R})$ contains an identity element for addition.
- d. Each element of $M_{m \times n}(\mathbf{R})$ has an additive inverse in $M_{m \times n}(\mathbf{R})$.
- e. Addition is commutative in $M_{m \times n}(\mathbf{R})$.

Proof Let $A = [a_{ij}]_{m \times n}$, $B = [b_{ij}]_{m \times n}$, and $C = [c_{ij}]_{m \times n}$ be arbitrary elements of $M_{m \times n}(\mathbf{R})$.

- a. The addition defined in Definition 1.29 is a binary operation on $M_{m \times n}(\mathbf{R})$ because the rule

$$[a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}]$$

yields a result that is both unique and an element of $M_{m \times n}(\mathbf{R})$.

- b. The following equalities establish the associative property for addition.

$$\begin{aligned} A + (B + C) &= [a_{ij}] + [b_{ij} + c_{ij}] && \text{by Definition 1.29} \\ &= [a_{ij} + (b_{ij} + c_{ij})] && \text{by Definition 1.29} \\ &= [(a_{ij} + b_{ij}) + c_{ij}] && \text{since addition in } \mathbf{R} \text{ is associative} \\ &= [a_{ij} + b_{ij}] + [c_{ij}] && \text{by Definition 1.29} \\ &= (A + B) + C && \text{by Definition 1.29} \end{aligned}$$

- c. Let $O_{m \times n}$ denote the $m \times n$ matrix that has all elements zero. Then

$$\begin{aligned} A + O_{m \times n} &= [a_{ij}]_{m \times n} + [0]_{m \times n} \\ &= [a_{ij} + 0]_{m \times n} && \text{by Definition 1.29} \\ &= [a_{ij}]_{m \times n} && \text{since } 0 \text{ is the additive identity in } \mathbf{R} \\ &= A. \end{aligned}$$

A similar computation shows that $O_{m \times n} + A = A$, and therefore $O_{m \times n}$ is the additive identity for $M_{m \times n}(\mathbf{R})$, called the **zero matrix** of dimension $m \times n$.

- d. It is left as an exercise to verify that the matrix $-A$ defined by

$$-A = [-a_{ij}]_{m \times n}$$

is the additive inverse for A in $M_{m \times n}(\mathbf{R})$.

- e. The proof that addition in $M_{m \times n}(\mathbf{R})$ is commutative is also left as an exercise.

Part **d** of Theorem 1.30 leads to the definition of **subtraction** in $M_{m \times n}(\mathbf{R})$: For A and B in $M_{m \times n}(\mathbf{R})$,

$$A - B = A + (-B),$$

where $-B = [-b_{ij}]$ is the additive inverse of $B = [b_{ij}]$.

The definition of **multiplication** that we present is a standard definition universally used in linear algebra, operations research, and other branches of mathematics. Its widespread acceptance is due to its usefulness in a great variety of important applications, not to its simplicity, for the definition of multiplication is much more complicated and much less “intuitive” than the definition of addition. We first state the definition and then illustrate it with an example.

Definition 1.31 ■ Matrix Multiplication

The **product** of an $m \times n$ matrix A over \mathbf{R} and an $n \times p$ matrix B over \mathbf{R} is an $m \times p$ matrix $C = AB$, where the element c_{ij} in row i and column j of AB is found by using the elements in row i of A and the elements in column j of B in the following manner:

$$\begin{matrix} & \text{column } j \\ & \text{of } B \\ \text{row } i \\ \text{of } A & \left[\begin{array}{ccccc} \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & & \vdots \end{array} \right] \cdot \left[\begin{array}{ccc} \cdots & b_{1j} & \cdots \\ \cdots & b_{2j} & \cdots \\ \cdots & b_{3j} & \cdots \\ \vdots & & \vdots \\ \cdots & b_{nj} & \cdots \end{array} \right] = \left[\begin{array}{ccc} & \vdots & \\ & c_{ij} & \cdots \\ & \vdots & \vdots \end{array} \right] \text{row } i \\ \text{of } C & \end{matrix}$$

where

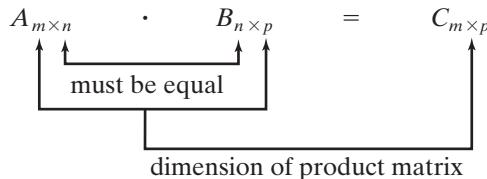
$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}$$

That is, the element

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}$$

in row i and column j of AB is found by adding the products formed from corresponding elements of row i in A and column j in B (first times first, second times second, and so on). Note that the elements of C are real numbers.

Note that the number of columns in A *must* equal the number of rows in B in order to form the product AB . If this is the case, then A and B are said to be **conformable for multiplication**. A simple diagram illustrates this fact.



Example 3 Consider the products that can be formed using the matrices

$$A = \begin{bmatrix} 3 & -2 \\ 0 & 4 \\ 1 & -3 \\ 5 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix}.$$

Since the number of columns in A is equal to the number of rows in B , the product AB is defined. Performing the multiplication, we obtain

$$\begin{aligned} AB &= \begin{bmatrix} 3 & -2 \\ 0 & 4 \\ 1 & -3 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix} \\ &= \begin{bmatrix} 3(2) + (-2)(4) & 3(1) + (-2)(-3) & 3(0) + (-2)(7) \\ 0(2) + 4(4) & 0(1) + 4(-3) & 0(0) + 4(7) \\ 1(2) + (-3)(4) & 1(1) + (-3)(-3) & 1(0) + (-3)(7) \\ 5(2) + 1(4) & 5(1) + 1(-3) & 5(0) + 1(7) \end{bmatrix}. \end{aligned}$$

Thus AB is the 4×3 matrix given by

$$AB = \begin{bmatrix} -2 & 9 & -14 \\ 16 & -12 & 28 \\ -10 & 10 & -21 \\ 14 & 2 & 7 \end{bmatrix}.$$

Since the number of columns in B is not equal to the number of rows in A , the product BA is not defined. Similarly, the products $A \cdot A$ and $B \cdot B$ are not defined. ■

The work in Example 3 shows that multiplication of matrices does not have the commutative property. Some of the computations in the exercises for this section illustrate cases where $AB \neq BA$, even when both products are defined and have the same dimension.

It should also be noted in connection with Example 3 that the product of matrices we are working with is not a true binary operation as defined in Section 1.4. With a binary operation on a set A , it must always be possible to combine any two elements of A and obtain a unique result of the operation. Multiplication of matrices does not have this feature, since the product of two matrices may not be defined. If consideration is restricted to the set $M_n(\mathbf{R})$ of all $n \times n$ matrices of a fixed order n , this difficulty disappears, and multiplication is a true binary operation on $M_n(\mathbf{R})$.

Although matrix multiplication is not commutative, it does have several properties that are analogous to corresponding properties in the set \mathbf{R} of all real numbers. The *sigma notation* is useful in writing out proofs of these properties.

In the **sigma notation**, the capital Greek letter Σ (sigma) is used to indicate a sum:

$$\sum_{i=1}^n a_i = a_1 + a_2 + \cdots + a_n.$$

The variable i is called the **index of summation**, and the notations below and above the sigma indicate the value of i at which the sum starts and the value of i at which it ends. For example,

$$\sum_{i=3}^5 b_i = b_3 + b_4 + b_5.$$

The index of summation is sometimes called a “dummy variable” because the value of the sum is unaffected if the index is changed to a different letter:

$$\sum_{i=0}^3 a_i = \sum_{j=0}^3 a_j = \sum_{k=0}^3 a_k = a_0 + a_1 + a_2 + a_3.$$

Using the distributive properties in \mathbf{R} , we can write

$$\begin{aligned} a \left(\sum_{k=1}^n b_k \right) &= a(b_1 + b_2 + \cdots + b_n) \\ &= ab_1 + ab_2 + \cdots + ab_n \\ &= \sum_{k=1}^n ab_k. \end{aligned}$$

Similarly,

$$\left(\sum_{k=1}^n b_k \right) a = \sum_{k=1}^n b_k a.$$

In the definition of the matrix product AB , the element

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

can be written compactly by use of the sigma notation as

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}.$$

If all necessary conformability is assumed, the following theorem asserts that matrix multiplication is associative.

Theorem 1.32 ■ Associative Property of Multiplication

Let $A = [a_{ij}]_{m \times n}$, $B = [b_{ij}]_{n \times p}$, and $C = [c_{ij}]_{p \times q}$ be matrices over \mathbf{R} . Then $A(BC) = (AB)C$.

Proof From Definition 1.31, $BC = [d_{ij}]_{n \times q}$ where $d_{ij} = \sum_{k=1}^p b_{ik}c_{kj}$, and $A(BC) = \left[\sum_{r=1}^n a_{ir}d_{rj} \right]_{m \times q}$ where

$$\begin{aligned} \sum_{r=1}^n a_{ir}d_{rj} &= \sum_{r=1}^n a_{ir} \left(\sum_{k=1}^p b_{rk}c_{kj} \right) \\ &= \sum_{r=1}^n \left(\sum_{k=1}^p a_{ir}(b_{rk}c_{kj}) \right). \end{aligned}$$

Also, $AB = [f_{ij}]_{m \times p}$ where $f_{ij} = \sum_{r=1}^n a_{ir}b_{rj}$, and $(AB)C = \left[\sum_{k=1}^p f_{ik}c_{kj} \right]_{m \times q}$ where

$$\begin{aligned} \sum_{k=1}^p f_{ik}c_{kj} &= \sum_{k=1}^p \left(\sum_{r=1}^n a_{ir}b_{rk} \right) c_{kj} \\ &= \sum_{k=1}^p \left(\sum_{r=1}^n (a_{ir}b_{rk})c_{kj} \right) \\ &= \sum_{k=1}^p \left(\sum_{r=1}^n a_{ir}(b_{rk}c_{kj}) \right). \end{aligned}$$

The last equality follows from the associative property

$$(a_{ir}b_{rk})c_{kj} = a_{ir}(b_{rk}c_{kj})$$

of multiplication of real numbers. Comparing the elements in row i , column j , of $A(BC)$ and $(AB)C$, we see that

$$\sum_{r=1}^n \left(\sum_{k=1}^p a_{ir}(b_{rk}c_{kj}) \right) = \sum_{k=1}^p \left(\sum_{r=1}^n a_{ir}(b_{rk}c_{kj}) \right),$$

since each of these double sums consists of all the np terms that can be made by using a product of the form $a_{ir}(b_{rk}c_{kj})$ with $1 \leq r \leq n$ and $1 \leq k \leq p$. Thus $A(BC) = (AB)C$.

Similar but simpler use of the sigma notation can be made to prove the distributive properties stated in the following theorem. Proofs are requested in the exercises.

Theorem 1.33 ■ Distributive Properties

Let A be an $m \times n$ matrix over \mathbf{R} , let B and C be $n \times p$ matrices over \mathbf{R} , and let D be a $p \times q$ matrix over \mathbf{R} . Then

- a. $A(B + C) = AB + AC$, and
- b. $(B + C)D = BD + CD$.

For each positive integer n , we define a special matrix I_n by

$$I_n = [\delta_{ij}]_{n \times n} \quad \text{where} \quad \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

(The symbol δ_{ij} used in defining I_n is called the **Kronecker delta**.) For $n = 2$ and $n = 3$, these special matrices are given by

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The matrices I_n have special properties in matrix multiplication, as stated in Theorem 1.34.

Theorem 1.34 ■ Special Properties of I_n

Let A be an arbitrary $m \times n$ matrix over \mathbf{R} . With I_n as defined in the preceding paragraph,

- a. $I_m A = A$, and
- b. $A I_n = A$.

Proof To prove part a, let $A = [a_{ij}]_{m \times n}$ and consider $I_m A$. By Definition 1.31,

$$I_m A = [c_{ij}]_{m \times n}$$

where

$$c_{ij} = \sum_{k=1}^m \delta_{ik} a_{kj}.$$

Since $\delta_{ik} = 0$ for $k \neq i$ and $\delta_{ii} = 1$, the expression for c_{ij} simplifies to

$$c_{ij} = \delta_{ii} a_{ij} = 1 \cdot a_{ij} = a_{ij}.$$

Thus $c_{ij} = a_{ij}$ for all pairs i, j and $I_m A = A$.

The proof that $A I_n = A$ is left as an exercise.

Because the equations $I_m A = A$ and $A I_n = A$ hold for any $m \times n$ matrix A , the matrix I_n is called the **identity matrix of order n** . In a more general context, the terms *left identity* and *right identity* are defined as follows.

Definition 1.35 ■ Left Identity, Right Identity

Let $*$ be a binary operation on the nonempty set A . If an element e in A has the property that

$$e * x = x \text{ for all } x \in A,$$

then e is called a **left identity element** with respect to $*$. Similarly, if

$$x * e = x \text{ for all } x \in A,$$

then e is a **right identity element** with respect to $*$.

If the same element e is both a left identity and a right identity with respect to $*$, then e is an *identity element* as defined in Definition 1.21. An identity element is sometimes called a **two-sided identity** to emphasize that both of the required equations hold.

Even though matrix multiplication is not a binary operation on $M_{m \times n}(\mathbf{R})$ when $m \neq n$, we call I_m a *left identity* and I_n a *right identity* for multiplication with elements of $M_{m \times n}(\mathbf{R})$. In the set $M_n(\mathbf{R})$ of all square matrices of order n over \mathbf{R} , I_n is a two-sided identity element with respect to multiplication.

The fact that I_n is a multiplicative identity for $M_n(\mathbf{R})$ leads immediately to the question: Does every nonzero element A of $M_n(\mathbf{R})$ have a multiplicative inverse? The answer is not what one might expect, because some nonzero square matrices do not have multiplicative inverses. This fact is illustrated in the next example.

Example 4 Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, and consider the problem of finding a matrix

$B = \begin{bmatrix} x & z \\ y & w \end{bmatrix}$ such that $AB = I_2$. Computation of AB leads at once to

$$\begin{bmatrix} x + 3y & z + 3w \\ 2x + 6y & 2z + 6w \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

or

$$\begin{bmatrix} x + 3y & z + 3w \\ 2(x + 3y) & 2(z + 3w) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This matrix equality is equivalent to the following system of four linear equations.

$$\begin{aligned} x + 3y &= 1 & z + 3w &= 0 \\ 2(x + 3y) &= 0 & 2(z + 3w) &= 1 \end{aligned}$$

Since $x + 3y = 1$ requires $2(x + 3y) = 2$, and this contradicts $2(x + 3y) = 0$, there is no solution to the system of equations and therefore no matrix B such that $AB = I_2$. That is, A does not have a multiplicative inverse. ■

When we work with matrices, the convention is to use the term *inverse* to mean “multiplicative inverse.” If the matrix A has an inverse, Exercise 13 of Section 1.4 assures us that the inverse is unique. In this case, A is **invertible**, and its inverse is denoted by A^{-1} . A few properties of inverses are included in the exercises for this section, but an in-depth investigation of inverses is more appropriate for a linear algebra course.

Exercises 1.6

True or False

Label each of the following statements as either true or false.

1. Matrix addition is a binary operation from $M_{m \times n}(\mathbf{R}) \times M_{m \times n}(\mathbf{R})$ to $M_{m \times n}(\mathbf{R})$.
2. Matrix multiplication is a binary operation from $M_{m \times n}(\mathbf{R}) \times M_{m \times n}(\mathbf{R})$ to $M_{m \times n}(\mathbf{R})$.
3. $AB = BA$ for all square matrices A and B of order n over \mathbf{R} .
4. $(AB)^n = A^nB^n$ for all square matrices A and B of order n over \mathbf{R} .
5. Let A be a nonzero element in $M_{m \times n}(\mathbf{R})$ and B and C elements in $M_{n \times p}(\mathbf{R})$. If $AB = AC$, then $B = C$.
6. Let A be in $M_{m \times n}(\mathbf{R})$ and B be in $M_{n \times p}(\mathbf{R})$. If $AB = O_{m \times p}$ then either $A = O_{m \times n}$ or $B = O_{n \times p}$.
7. The set of diagonal matrices of order n over \mathbf{R} is closed with respect to matrix addition.
8. $(A + B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$ for all square matrices A and B of order n over \mathbf{R} .
9. The products AB and BA are defined if and only if both A and B are square matrices of the same order.
10. Let A be in $M_{m \times n}(\mathbf{R})$ and B be in $M_{n \times p}(\mathbf{R})$. If the j th column of A contains all zeros, then the j th column of AB contains all zeros.
11. Let A be in $M_{m \times n}(\mathbf{R})$ and B be in $M_{n \times p}(\mathbf{R})$. If the i th row of A contains all zeros, then the i th row of AB contains all zeros.
12. Let A be a square matrix of order n over \mathbf{R} such that $A^2 - 3A + I_n = O_n$. Then $A^{-1} = 3I_n - A$.

Exercises

1. Write out the matrix that matches the given description.
 - a. $A = [a_{ij}]_{3 \times 2}$ with $a_{ij} = 2i - j$
 - b. $A = [a_{ij}]_{4 \times 2}$ with $a_{ij} = (-1)^j$
 - c. $B = [b_{ij}]_{2 \times 4}$ with $b_{ij} = (-1)^{i+j}$
 - d. $B = [b_{ij}]_{3 \times 4}$ with $b_{ij} = 1$ if $i < j$ and $b_{ij} = 0$ if $i \geq j$
 - e. $C = [c_{ij}]_{4 \times 3}$ with $c_{ij} = i + j$ if $i \geq j$ and $c_{ij} = 0$ if $i < j$
 - f. $C = [c_{ij}]_{4 \times 3}$ with $c_{ij} = 0$ if $i \neq j$ and $c_{ij} = 1$ if $i = j$

2. Perform the indicated operations, if possible.

$$\begin{array}{ll}
 \text{a. } \begin{bmatrix} -1 & 2 & 5 \\ 0 & -3 & 7 \end{bmatrix} + \begin{bmatrix} 4 & -2 & -9 \\ 8 & -5 & -1 \end{bmatrix} & \text{b. } \begin{bmatrix} 8 & 9 \\ 3 & 7 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 6 & 5 \end{bmatrix} \\
 \text{c. } \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \end{bmatrix} + \begin{bmatrix} 4 & 9 \\ -5 & -8 \\ 6 & 7 \end{bmatrix} & \text{d. } \begin{bmatrix} 3 & 0 \\ 8 & 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 4 \end{bmatrix}
 \end{array}$$

3. Perform the following multiplications, if possible.

a. $\begin{bmatrix} 2 & 0 & -3 \\ -4 & 1 & -1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 5 & 6 \\ 1 & -1 \end{bmatrix}$

b. $\begin{bmatrix} -1 & 2 \\ 5 & 6 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & -3 \\ -4 & 1 & -1 \end{bmatrix}$

c. $\begin{bmatrix} 2 & 0 \\ 0 & -3 \\ -1 & 5 \end{bmatrix} \begin{bmatrix} 3 & 2 & -1 \\ 6 & -2 & 0 \\ 1 & 0 & 4 \end{bmatrix}$

d. $\begin{bmatrix} 3 & 2 & -1 \\ 6 & -2 & 0 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ -1 & 5 \end{bmatrix}$

e. $\begin{bmatrix} -6 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$

f. $\begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -6 & 4 \\ 1 & 3 \end{bmatrix}$

g. $\begin{bmatrix} 5 \\ -3 \\ 2 \end{bmatrix} \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}$

h. $[-4 \ 6 \ 2] [-1 \ 0 \ 5]$

i. $[3 \ -2 \ 1] \begin{bmatrix} -4 \\ -5 \\ 6 \end{bmatrix}$

j. $\begin{bmatrix} -4 \\ -5 \\ 6 \end{bmatrix} [3 \ -2 \ 1]$

4. Let $A = [a_{ij}]_{2 \times 3}$ where $a_{ij} = i + j$, and let $B = [b_{ij}]_{3 \times 4}$ where $b_{ij} = 2i - j$. If $AB = [c_{ij}]_{2 \times 4}$, write a formula for c_{ij} in terms of i and j .

5. Show that the matrix equation

$$\begin{bmatrix} 1 & -2 & 7 \\ 5 & -1 & 6 \\ 3 & 4 & -8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ -4 \\ 2 \end{bmatrix}$$

is equivalent to a system of linear equations in x , y , and z .

6. Write a single matrix equation of the form $AX = B$ that is equivalent to the following system of equations.

$$\begin{aligned} w + 6x - 3y + 2z &= 9 \\ 4w - 7x + y + 5z &= 0 \end{aligned}$$

7. Let δ_{ij} denote the Kronecker delta: $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$. Find the value of the following expressions.

a. $\sum_{i=1}^n \left(\sum_{j=1}^n \delta_{ij} \right)$

b. $\sum_{i=1}^n \left(\sum_{j=1}^n (1 - \delta_{ij}) \right)$

c. $\sum_{i=1}^5 \left(\sum_{j=1}^4 (-1)^{\delta_{ij}} \right)$

d. $\sum_{j=1}^n \delta_{ij} \delta_{jk}$

Sec. 1.4, #3 ➤

- 8.** Let S be the set of four matrices $S = \{I, A, B, C\}$, where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Follow the procedure described in Exercise 3 of Section 1.4 to complete the following multiplication table for S . (In this case, the product $BC = A$ is entered as shown in the row with B at the left end and in the column with C at the top.) Is S closed under multiplication?

•	I	A	B	C
I				
A				
B	B	C	I	A
C				

- 9.** Find two square matrices A and B such that $AB \neq BA$.
- 10.** Find two nonzero matrices A and B such that $AB = BA$.
- 11.** Find two nonzero matrices A and B such that $AB = O_{n \times n}$.
- 12.** Let A , B , and C be elements of $M_2(\mathbf{R})$, where A is not a zero matrix. Prove or disprove that $AB = AC$ implies $B = C$.
- 13.** Positive integral powers of a square matrix are defined by $A^1 = A$ and $A^{n+1} = A^n \cdot A$ for every positive integer n . Evaluate $(A - B)(A + B)$ and $A^2 - B^2$ and compare the results for
- $$A = \begin{bmatrix} 1 & 2 \\ 4 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix}.$$
- 14.** For the matrices in Exercise 13, evaluate $(A + B)^2$ and $A^2 + 2AB + B^2$ and compare the results.
- 15.** Assume that A^{-1} exists and find a solution X to $AX = B$ where A and B are in $M_n(\mathbf{R})$.
- 16.** Assume that A , B , C , and X are in $M_n(\mathbf{R})$, and $AXC = B$ with A and C invertible. Solve for X .
- 17. a.** Prove part **d** of Theorem 1.30.
b. Prove part **e** of Theorem 1.30.
- 18. a.** Prove part **a** of Theorem 1.33.
b. Prove part **b** of Theorem 1.33.

- 19.** Prove part **b** of Theorem 1.34.
- 20.** Prove that if $A \in M_{m \times n}(\mathbf{R})$, then $A \cdot O_{n \times p} = O_{m \times p}$.
- 21.** Suppose that A is an invertible matrix over \mathbf{R} and O is a zero matrix. Prove that if $AX = O$, then $X = O$.
- 22.** Let G be the set of all elements of $M_2(\mathbf{R})$ that have one row that consists of zeros and one row of the form $[a \ a]$, with $a \neq 0$.
- Show that G is closed under multiplication.
 - Show that for each x in G , there is an element y in G such that $xy = yx = x$.
 - Show that G does not have an identity element with respect to multiplication.
- 23.** Prove that the set $S = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \mid a, b \in \mathbf{R} \right\}$ is closed with respect to matrix addition and multiplication.
- 24.** Prove or disprove that the set of diagonal matrices of order n over \mathbf{R} is closed with respect to matrix multiplication.
- 25.** Let A and B be square matrices of order n over \mathbf{R} . Prove or disprove that the product AB is a diagonal matrix of order n over \mathbf{R} if B is a diagonal matrix.
- 26.** Let A and B be square matrices of order n over \mathbf{R} . Prove or disprove that if AB is a diagonal matrix of order n over \mathbf{R} , then at least one of A or B is a diagonal matrix.
- 27.** A square matrix $A = [a_{ij}]_n$ with $a_{ij} = 0$ for all $i > j$ is called **upper triangular**. Prove or disprove each of the following statements.
- The set of all upper triangular matrices is closed with respect to matrix addition.
 - The set of all upper triangular matrices is closed with respect to matrix multiplication.
 - If A and B are square and the product AB is upper triangular then at least one of A or B is upper triangular.
- 28.** Let a, b, c , and d be real numbers. If $ad - bc \neq 0$, show that the multiplicative inverse of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by
- Sec. 3.3, #14g ≪ Sec. 3.6, #9 ≪
- $$\begin{bmatrix} \frac{d}{ad - bc} & \frac{-b}{ad - bc} \\ \frac{-c}{ad - bc} & \frac{a}{ad - bc} \end{bmatrix}.$$
- 29.** Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ over \mathbf{R} . Prove that if $ad - bc = 0$, then A does not have an inverse.
- 30.** Let A, B , and C be square matrices of order n over \mathbf{R} . Prove that if A is invertible and $AB = AC$, then $B = C$.
- 31.** Let A and B be $n \times n$ matrices over \mathbf{R} such that A^{-1} and B^{-1} exist. Prove that $(AB)^{-1}$ exists and that $(AB)^{-1} = B^{-1}A^{-1}$. (This result is known as the **reverse order law** for inverses.)
- Sec. 2.2, #27 ≪

1.7**Relations**

In the study of mathematics, we deal with many examples of relations between elements of various sets. In working with the integers, we encounter relations such as “ x is less than y ” and “ x is a factor of y . ” In calculus, one function may be the derivative of some other function, or perhaps an integral of another function. The property that these examples of relations have in common is that there is an association of some sort between two elements of a set, and the ordering of the elements is important. These relations can all be described by the following definition.

Definition 1.36 ■ Relation

A **relation** (or a **binary relation**) on a nonempty set A is a nonempty set R of ordered pairs (x, y) of elements x and y of A .

That is, a relation R is a subset of the Cartesian product $A \times A$. If the pair (a, b) is in R , we write aRb and say that a has the relation R to b . If $(a, b) \notin R$, we write $a \not R b$. This notation agrees with the customary notations for relations, such as $a = b$ and $a < b$.

Example 1 Let $A = \{-2, -5, 2, 5\}$ and $R = \{(5, -2), (5, 2), (-5, -2), (-5, 2)\}$. Then $5R2$, $-5R2$, $5R(-2)$, and $(-5)R(-2)$, but $2 \not R 5$, $5 \not R 5$, and so on. As is frequently the case, this relation can be described by a simple rule: xRy if and only if the absolute value of x is the same as $y^2 + 1$ —that is, if $|x| = y^2 + 1$. ■

Any mapping from A to A is an example of a relation, but not all relations are mappings, as Example 1 illustrates. We have $(5, 2) \in R$ and $(5, -2) \in R$, and for a mapping from A to A , the second element y in $(5, y)$ would have to be unique.

Our main concern is with relations that have additional special properties. More precisely, we are interested for the most part in *equivalence relations*.

Definition 1.37 ■ Equivalence Relation

A relation R on a nonempty set A is an **equivalence relation** if the following conditions are satisfied for arbitrary x, y, z in A :

- | | |
|--------------------------------------|----------------------------|
| 1. xRx for all $x \in A$. | Reflexive Property |
| 2. If xRy , then yRx . | Symmetric Property |
| 3. If xRy and yRz , then xRz . | Transitive Property |
-

Properties 1, 2, and 3 of Definition 1.37 are familiar basic properties of equality.

Example 2 The relation R defined on the set of integers \mathbf{Z} by

$$xRy \text{ if and only if } |x| = |y|$$

is reflexive, symmetric, and transitive. For arbitrary x, y , and z in \mathbf{Z} ,

1. xRx , since $|x| = |x|$.
2. $xRy \Rightarrow |x| = |y|$
 $\Rightarrow |y| = |x|$
 $\Rightarrow yRx$.
3. xRy and $yRz \Rightarrow |x| = |y|$ and $|y| = |z|$
 $\Rightarrow |x| = |z|$
 $\Rightarrow xRz.$

■

Example 3 The relation R defined on the set of integers \mathbf{Z} by

$$xRy \text{ if and only if } x > y$$

is not an equivalence relation, since it is neither reflexive nor symmetric.

1. $x \not> x$ for all $x \in \mathbf{Z}$.
2. $x > y \not\Rightarrow y > x$.

Note that R is transitive:

3. $x > y$ and $y > z \Rightarrow x > z.$

■

The following example is a special case of an equivalence relation on the integers that will be extremely important in later work.

Example 4 The relation “congruence modulo 4” is defined on the set \mathbf{Z} of all integers as follows: x is congruent to y modulo 4 if and only if $x - y$ is a multiple of 4. We write $x \equiv y \pmod{4}$ as shorthand for “ x is congruent to y modulo 4.” Thus $x \equiv y \pmod{4}$ if and only if $x - y = 4k$ for some integer k . We demonstrate that this is an equivalence relation. For arbitrary x, y, z in \mathbf{Z} ,

1. $x \equiv x \pmod{4}$, since $x - x = (4)(0)$.
2. $x \equiv y \pmod{4} \Rightarrow x - y = 4k$ for some $k \in \mathbf{Z}$
 $\Rightarrow y - x = 4(-k)$ and $-k \in \mathbf{Z}$
 $\Rightarrow y \equiv x \pmod{4}$.
3. $x \equiv y \pmod{4}$ and $y \equiv z \pmod{4}$
 $\Rightarrow x - y = 4k$ and $y - z = 4m$ for some $k, m \in \mathbf{Z}$
 $\Rightarrow x - z = x - y + y - z = 4(k + m)$, and $k + m \in \mathbf{Z}$
 $\Rightarrow x \equiv z \pmod{4}$.

Thus congruence modulo 4 has the reflexive, symmetric, and transitive properties and is an equivalence relation on \mathbf{Z} .

■

Definition 1.38 ■ Equivalence Class

Let R be an equivalence relation on the nonempty set A . For each $a \in A$, the set

$$[a] = \{x \in A \mid xRa\}$$

is called the **equivalence class** containing a .

Example 5 The relation R in Example 2 defined on \mathbf{Z} by $xRy \Leftrightarrow |x| = |y|$ is an equivalence relation. The equivalence class containing 0 is

$$[0] = \{0\}$$

since 0 is the only element $x \in \mathbf{Z}$ such that $|x| = 0$. Some other equivalence classes are given by

$$[1] = \{1, -1\} \quad \text{and} \quad [-3] = \{-3, 3\}.$$

For $a \neq 0$, the equivalence class $[a]$ is given by

$$[a] = \{-a, a\}$$

since a and $-a$ are the only elements in \mathbf{Z} with absolute value equal to $|a|$. ■

Example 6 The relation “congruence modulo 4” was shown in Example 4 to be an equivalence relation. Since $x \equiv y \pmod{4}$ if and only if $x - y$ is a multiple of 4, the equivalence class $[a]$ consists of all those integers that differ from a by a multiple of 4. Thus $[0]$ consists of all multiples of 4:

$$[0] = \{\dots, -8, -4, 0, 4, 8, \dots\}.$$

Similarly, the other equivalence classes are given by:

$$[1] = \{\dots, -7, -3, 1, 5, 9, \dots\}.$$

$$[2] = \{\dots, -6, -2, 2, 6, 10, \dots\}.$$

$$[3] = \{\dots, -5, -1, 3, 7, 11, \dots\}. ■$$

In both Examples 5 and 6, the equivalence classes separate the set \mathbf{Z} into mutually disjoint nonempty subsets. Recall from Section 1.1 that a separation of the elements of a nonempty set A into mutually disjoint nonempty subsets is called a *partition* of A . It is not difficult to show that if R is an equivalence relation on A , then the distinct equivalence classes of R form a partition of A . Conversely, if a partition of A is given, then we can find an equivalence relation R on A that has the given subsets as its equivalence classes. We simply define R by xRy if and only if x and y are in the same subset. The proofs of these statements are requested in the exercises for this section.

The discussion in the last paragraph illustrates a situation where we are dealing with a collection of sets about which very little is explicit. For example, the collection may be finite, or it may be infinite. In such situations, it is sometimes desirable to use the notational

convenience known as indexing. We assume that the sets in the collection are **labeled**, or **indexed**, by a set \mathcal{L} of symbols λ . That is, a typical set in the collection is denoted by a symbol such as A_λ , and the index λ takes on values from the set \mathcal{L} . For such a collection $\{A_\lambda\}$, we write $\bigcup_{\lambda \in \mathcal{L}} A_\lambda$ for the union of the collection of sets, and we write $\bigcap_{\lambda \in \mathcal{L}} A_\lambda$ for the intersection. That is,

$$\bigcup_{\lambda \in \mathcal{L}} A_\lambda = \{x | x \in A_\lambda \text{ for at least one } \lambda \in \mathcal{L}\}$$

and

$$\bigcap_{\lambda \in \mathcal{L}} A_\lambda = \{x | x \in A_\lambda \text{ for every } \lambda \in \mathcal{L}\}.$$

If the indexing set \mathcal{L} is given by $\mathcal{L} = \{1, 2, \dots, n\}$, then the union of the collection of sets $\{A_i\}$ might be written in any one of the following three ways.

$$A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i \in \mathcal{L}} A_i = \bigcup_{i=1}^n A_i$$

The index notation is useful in describing a partition of a set. An alternative definition can be made in the following manner.

Definition 1.39 ■ Partition

Let $\{A_\lambda\}$, $\lambda \in \mathcal{L}$, be a collection of subsets of the nonempty set A . Then $\{A_\lambda\}$ is a **partition** of A if all these conditions are satisfied:

1. Each A_λ is nonempty.
2. $A = \bigcup_{\lambda \in \mathcal{L}} A_\lambda$.
3. If $A_\alpha \cap A_\beta \neq \emptyset$, then $A_\alpha = A_\beta$.

Exercises 1.7

True or False

Label each of the following statements as either true or false.

1. Every mapping on a nonempty set A is a relation.
2. Every relation on a nonempty set A is a mapping.
3. If R is an equivalence relation on a nonempty set A , then the distinct equivalence classes of R form a partition of A .
4. If R is an equivalence relation on a nonempty set A , then any two equivalence classes of R contain the same number of elements.

5. Let R be an equivalence relation on a nonempty set A and let a and b be in A . If $b \in [a]$, then $[b] = [a]$.
 6. Let R be a relation on a nonempty set A that is symmetric and transitive. Since R is symmetric xRy implies yRx . Since R is transitive xRy and yRx implies xRx . Hence R is also reflexive and thus an equivalence relation on A .
-

Exercises

1. For $A = \{1, 3, 5\}$, determine which of the following relations on A are mappings from A to A , and justify your answer.
 - a. $\{(1, 3), (3, 5), (5, 1)\}$
 - b. $\{(1, 1), (3, 1), (5, 1)\}$
 - c. $\{(1, 1), (1, 3), (1, 5)\}$
 - d. $\{(1, 3), (3, 1), (5, 5)\}$
 - e. $\{(1, 5), (3, 3), (5, 3)\}$
 - f. $\{(5, 1), (5, 3), (5, 5)\}$
2. In each of the following parts, a relation R is defined on the set \mathbf{Z} of all integers. Determine in each case whether or not R is reflexive, symmetric, or transitive. Justify your answers.
 - a. xRy if and only if $x = 2y$.
 - b. xRy if and only if $x = -y$.
 - c. xRy if and only if $y = xk$ for some k in \mathbf{Z} .
 - d. xRy if and only if $x < y$.
 - e. xRy if and only if $x \geq y$.
 - f. xRy if and only if $x = |y|$.
 - g. xRy if and only if $|x| \leq |y| + 1$.
 - h. xRy if and only if $xy \geq 0$.
 - i. xRy if and only if $xy \leq 0$.
 - j. xRy if and only if $|x - y| = 1$.
 - k. xRy if and only if $|x - y| < 1$.
3. a. Let R be the equivalence relation defined on \mathbf{Z} in Example 2, and write out the elements of the equivalence class $[3]$.

b. Let R be the equivalence relation “congruence modulo 4” that is defined on \mathbf{Z} in Example 4. For this R , list five members of the equivalence class $[7]$.
4. Let R be the relation “congruence modulo 5” defined on \mathbf{Z} as follows: x is congruent to y modulo 5 if and only if $x - y$ is a multiple of 5, and we write $x \equiv y \pmod{5}$.
 - a. Prove that “congruence modulo 5” is an equivalence relation.
 - b. List five members of each of the equivalence classes $[0], [1], [2], [8]$, and $[-4]$.
5. Let R be the relation “congruence modulo 7” defined on \mathbf{Z} as follows: x is congruent to y modulo 7 if and only if $x - y$ is a multiple of 7, and we write $x \equiv y \pmod{7}$.
 - a. Prove that “congruence modulo 7” is an equivalence relation.
 - b. List five members of each of the equivalence classes $[0], [1], [3], [9]$, and $[-2]$.

In Exercises 6–10, a relation R is defined on the set \mathbf{Z} of all integers. In each case, prove that R is an equivalence relation. Find the distinct equivalence classes of R and list at least four members of each.

6. xRy if and only if $x^2 + y^2$ is a multiple of 2.
7. xRy if and only if $x^2 - y^2$ is a multiple of 5.
8. xRy if and only if $x + 3y$ is a multiple of 4.
9. xRy if and only if $3x - 10y$ is a multiple of 7.
10. xRy if and only if $(-1)^x = (-1)^y$.
11. Consider the set $\mathcal{P}(A) - \{\emptyset\}$ of all nonempty subsets of $A = \{1, 2, 3, 4, 5\}$. Determine whether the given relation R on $\mathcal{P}(A) - \{\emptyset\}$ is reflexive, symmetric, or transitive. Justify your answers.
 - a. xRy if and only if x is a subset of y .
 - b. xRy if and only if x is a proper subset of y .
 - c. xRy if and only if x and y have the same number of elements.
12. In each of the following parts, a relation is defined on the set of all human beings. Determine whether the relation is reflexive, symmetric, or transitive. Justify your answers.
 - a. xRy if and only if x lives within 400 miles of y .
 - b. xRy if and only if x is the father of y .
 - c. xRy if and only if x is a first cousin of y .
 - d. xRy if and only if x and y were born in the same year.
 - e. xRy if and only if x and y have the same mother.
 - f. xRy if and only if x and y have the same hair color.
13. Let $A = \mathbf{R} - \{0\}$, the set of all nonzero real numbers, and consider the following relations on $A \times A$. Decide in each case whether R is an equivalence relation, and justify your answers.
 - a. $(a, b)R(c, d)$ if and only if $ad = bc$.
 - b. $(a, b)R(c, d)$ if and only if $ab = cd$.
 - c. $(a, b)R(c, d)$ if and only if $a^2 + b^2 = c^2 + d^2$.
 - d. $(a, b)R(c, d)$ if and only if $a - b = c - d$.
14. Let $A = \{1, 2, 3, 4\}$ and define R on $\mathcal{P}(A) - \{\emptyset\}$ by xRy if and only if $x \cap y \neq \emptyset$. Determine whether R is reflexive, symmetric, or transitive.
15. In each of the following parts, a relation R is defined on the power set $\mathcal{P}(A)$ of the nonempty set A . Determine in each case whether R is reflexive, symmetric, or transitive. Justify your answers.
 - a. xRy if and only if $x \cap y \neq \emptyset$.
 - b. xRy if and only if $x \subseteq y$.

16. Let $\mathcal{P}(A)$ be the power set of the nonempty set A , and let C denote a fixed subset of A . Define R on $\mathcal{P}(A)$ by xRy if and only if $x \cap C = y \cap C$. Prove that R is an equivalence relation on $\mathcal{P}(A)$.
17. For each of the following relations R defined on the set A of all triangles in a plane, determine whether R is reflexive, symmetric, or transitive. Justify your answers.
- aRb if and only if a is similar to b .
 - aRb if and only if a is congruent to b .
18. Give an example of a relation R on a nonempty set A that is symmetric and transitive, but not reflexive.
19. A relation R on a nonempty set A is called **irreflexive** if $x \not R x$ for all $x \in A$. Which of the relations in Exercise 2 are irreflexive?
20. A relation R on a nonempty set A is called **asymmetric** if, for x and y in A , xRy implies $y \not R x$. Which of the relations in Exercise 2 are asymmetric?
21. A relation R on a nonempty set A is called **antisymmetric** if, for x and y in A , xRy and yRx together imply $x = y$. (That is, R is antisymmetric if $x \neq y$ implies that either $x \not R y$ or $y \not R x$.) Which of the relations in Exercise 2 are antisymmetric?
22. For any relation R on the nonempty set A , the **inverse** of R is the relation R^{-1} defined by $xR^{-1}y$ if and only if yRx . Prove the following statements.
- R is symmetric if and only if $R = R^{-1}$.
 - R is antisymmetric if and only if $R \cap R^{-1}$ is a subset of $\{(a, a) | a \in A\}$.
 - R is asymmetric if and only if $R \cap R^{-1} = \emptyset$.
23. Let $\mathcal{L} = \{1, 2, 3\}$, $A_1 = \{a, b, c, d\}$, $A_2 = \{c, d, e, f\}$, and $A_3 = \{a, c, f, g\}$. Write out $\bigcup_{\lambda \in \mathcal{L}} A_\lambda$ and $\bigcap_{\lambda \in \mathcal{L}} A_\lambda$.
24. Let $\mathcal{L} = \{\alpha, \beta, \gamma\}$, $A_\alpha = \{1, 2, 3, \dots\}$, $A_\beta = \{-1, -2, -3, \dots\}$, and $A_\gamma = \{0\}$. Write out $\bigcup_{\lambda \in \mathcal{L}} A_\lambda$ and $\bigcap_{\lambda \in \mathcal{L}} A_\lambda$.
25. Suppose R is an equivalence relation on the nonempty set A . Prove that the distinct equivalence classes of R separate the elements of A into mutually disjoint subsets.
26. Let $A = \{1, 2, 3\}$, $B_1 = \{1, 2\}$, and $B_2 = \{2, 3\}$. Define the relation R on A by aRb if and only if there is a set B_i ($i = 1$ or 2) such that $a \in B_i$ and $b \in B_i$. Determine which of the properties of an equivalence relation hold for R , and give an example for each property that fails to hold.
27. Suppose $\{A_\lambda\}$, $\lambda \in \mathcal{L}$, represents a partition of the nonempty set A . Define R on A by xRy if and only if there is a subset A_λ such that $x \in A_\lambda$ and $y \in A_\lambda$. Prove that R is an equivalence relation on A and that the equivalence classes of R are the subsets A_λ .
28. Suppose that f is an onto mapping from A to B . Prove that if $\{B_\lambda\}$, $\lambda \in \mathcal{L}$, is a partition of B , then $\{f^{-1}(B_\lambda)\}$, $\lambda \in \mathcal{L}$, is a partition of A .

Key Words and Phrases

- addition of matrices, 43
- associative binary operation, 31
- associative property, 7, 20, 48
- bijective mapping, 18
- binary operation, 30
- binary relation, 55
- Cartesian product, 13
- closed subset, 32
- codomain, 14
- commutative binary operation, 31
- commutative property, 4
- complement, 5
- complex number, 6
- composite mapping, 19
- composition of mappings, 19
- conformable matrices, 46
- counterexample, 17
- De Morgan's Laws, 9
- diagonal matrix, 43
- dimension of a matrix, 42
- disjoint sets, 4
- distributive property, 8, 48
- domain, 14
- empty set, 4
- equal matrices, 43
- equivalence class, 57
- equivalence relation, 55
- even integer, 18
- identity element, 33
- identity mapping, 37
- identity matrix, 49
- image, 13, 15
- injective mapping, 16
- integers, 6
- intersection, 3
- inverse, 33
- inverse image, 15
- invertible element, 33
- invertible matrix, 50
- Kronecker delta, 49
- left identity element, 50
- left inverse, 33
- mapping, 13
- matrix, 42
- multiplication of matrices, 45
- odd integer, 18
- one-to-one correspondence, 18
- one-to-one mapping, 16
- onto mapping, 15
- partition, 7, 58
- permutation, 37
- positive integers, 6
- power set, 4
- product of matrices, 45
- proper subset, 3
- range, 14
- rational number, 6
- real numbers, 6
- reflexive property, 55
- relation, 55
- reverse order law, 54
- right identity element, 50
- right inverse, 33
- sigma notation, 47
- square matrix, 43
- subset, 2
- subtraction of matrices, 45
- surjective mapping, 15
- symmetric property, 55
- transitive property, 55
- union, 3
- universal set, 4
- Venn diagram, 4
- zero matrix, 45



SSPL/Image Works

A Pioneer in Mathematics Arthur Cayley (1821–1895)

The English mathematician Arthur Cayley, one of the three most prolific writers in mathematics, authored more than 200 mathematical papers. He founded the theory of matrices and was one of the first writers to describe abstract groups. According to mathematical historian Howard Eves, Cayley was one of the 19th-century algebraists who "opened the floodgates of modern abstract algebra."

Cayley displayed superior mathematical talent early in his life. At the age of 17 he studied at Trinity College of Cambridge University. Upon graduation, he accepted a position as assistant

tutor at the college. At the end of his third year as tutor, his appointment was not renewed because he declined to take the holy orders to become a parson. Cayley then turned to law and spent the next 15 years as a practicing lawyer. It was during this period that he wrote most of his mathematical papers, many of which are now classics.

Mathematics was not Cayley's only love, though. He was also an avid novel reader, a talented watercolor artist, an ardent mountain climber, and a passionate nature lover. However, even on his mountaineering trips, he spent a few hours each day on mathematics.

Cayley spent the last 32 years of his life as a professor of mathematics at Cambridge University. During this period, he campaigned successfully for the admission of women to the university.

This page intentionally left blank

The Integers

■ Introduction

It is unusual for a chapter to begin with an optional section, but there is an explanation for doing so here. Whether Section 2.1 is to be included or skipped is a matter of attitude or emphasis. If the approach is to emphasize the development of the basic properties of addition, multiplication, and ordering of integers from an initial list of postulates for the integers, then Section 2.1 should be included. As an alternative approach, these properties can be taken as known material from earlier experience, and Section 2.1 can be skipped. Whichever approach is taken, Section 2.1 summarizes the knowledge that is needed for the subsequent material in the chapter, and it separates “what we know” from “what we must prove.”

Although Section 2.2 on mathematical induction is not labeled as optional, this material may be familiar from calculus or previous algebra courses, and it might also be skipped.

The set \mathbf{Z}_n of congruence classes modulo n makes its first appearance in Section 2.5 as a set of equivalence classes. Binary operations of addition and multiplication are defined on \mathbf{Z}_n in Section 2.6. Both the additive and the multiplicative structures are drawn upon for examples in Chapters 3 and 4.

Sections 2.7 and 2.8 present optional introductions to coding theory and cryptography. The primary purpose of these sections is to demonstrate that the material in this text has usefulness other than as a foundation for mathematics courses at a higher level.

2.1

Postulates for the Integers (Optional)

The material in this chapter is concerned exclusively with integers. For this reason, we make a notational agreement that *all variables represent integers*. As our starting point, we shall take the system of integers as given and assume that the system of integers satisfies a certain list of basic axioms, or postulates. More precisely, we assume that there is a set \mathbf{Z} of elements, called the **integers**, that satisfies the following conditions.

Postulates for the Set \mathbf{Z} of Integers

1. **Addition postulates.** There is a binary operation defined in \mathbf{Z} that is called **addition**, is denoted by $+$, and has the following properties:
 - a. \mathbf{Z} is **closed** under addition.
 - b. Addition is **associative**.

- c. \mathbf{Z} contains an element 0 that is an **identity element for addition**.
 - d. For each $x \in \mathbf{Z}$, there is an **additive inverse** of x in \mathbf{Z} , denoted by $-x$, such that $x + (-x) = 0 = (-x) + x$.
 - e. Addition is **commutative**.
2. **Multiplication postulates.** There is a binary operation defined in \mathbf{Z} that is called **multiplication**, is denoted by \cdot , and has the following properties:
- a. \mathbf{Z} is **closed** under multiplication.
 - b. Multiplication is **associative**.
 - c. \mathbf{Z} contains an element 1 that is different from 0 and that is an **identity element for multiplication**.
 - d. Multiplication is **commutative**.
3. The **distributive law**,
- $$x \cdot (y + z) = x \cdot y + x \cdot z,$$
- holds for all elements $x, y, z \in \mathbf{Z}$.
4. \mathbf{Z} contains a subset \mathbf{Z}^+ , called the **positive integers**, that has the following properties:
- a. \mathbf{Z}^+ is **closed** under addition.
 - b. \mathbf{Z}^+ is **closed** under multiplication.
 - c. For each x in \mathbf{Z} , one and only one of the following statements is true.
 - i. $x \in \mathbf{Z}^+$
 - ii. $x = 0$
 - iii. $-x \in \mathbf{Z}^+$
5. **Induction postulate.** If S is a subset of \mathbf{Z}^+ such that
- a. $1 \in S$, and
 - b. $x \in S$ always implies $x + 1 \in S$,
- then $S = \mathbf{Z}^+$.

Note that we are taking the entire list of postulates as *assumptions* concerning \mathbf{Z} . This list is our set of basic properties. In this section we shall investigate briefly some of the consequences of this set of properties.

After the term *group* has been defined in Chapter 3, we shall see that the addition postulates state that \mathbf{Z} is a commutative group with respect to addition. Note that there is a major difference between the multiplication and the addition postulates, in that no inverses are required with respect to multiplication.

Postulate 3, the distributive law, is sometimes known as the **left distributive law**. The requirement that

$$(y + z) \cdot x = y \cdot x + z \cdot x$$

is known as the **right distributive law**. This property can be deduced from those in our list, as can all the familiar properties of addition and multiplication of integers.

Postulate 4c is referred to as the **law of trichotomy** because of its assertion that *exactly one of three possibilities* must hold. In case iii, where $-x \in \mathbf{Z}^+$, we say that x is a **negative integer** and that the set $\{x \mid -x \in \mathbf{Z}^+\}$ is the **set of all negative integers**.

The induction postulate is so named because it provides a basis for proofs by mathematical induction. Section 2.2 is devoted to the method of proof by induction, and the method is used from time to time throughout this book.

The right distributive law can be shown to follow from the set of postulates for the integers. We do this formally in the following theorem.

Theorem 2.1 ■ Right Distributive Law

The equality

$$(y + z) \cdot x = y \cdot x + z \cdot x$$

holds for all x, y, z in \mathbf{Z} .

Proof For arbitrary x, y, z in \mathbf{Z} , we have

$$\begin{aligned} (y + z) \cdot x &= x \cdot (y + z) && \text{by postulate 2d} \\ &= x \cdot y + x \cdot z && \text{by postulate 3} \\ &= y \cdot x + z \cdot x && \text{by postulate 2d.} \end{aligned}$$

The foregoing proof is admittedly trivial, but the point is that the usual manipulations involving integers are indeed consequences of our basic set of postulates. As another example, consider the statement[†] that $(-x)y = -(xy)$. In this equation, $-(xy)$ denotes the additive inverse of xy , just as $-x$ denotes the additive inverse of x .

Theorem 2.2 ■ Additive Inverse of a Product

For arbitrary x and y in \mathbf{Z} ,

$$(-x)y = -(xy).$$

Instead of attempting to prove this statement directly, we shall first prove a lemma.

Lemma 2.3 ■ Cancellation Law for Addition

If a, b , and c are integers and $a + b = a + c$, then $b = c$.

$p \Rightarrow q$ **Proof of the Lemma** Suppose $a + b = a + c$. Now $-a$ is in \mathbf{Z} , and hence

$$\begin{aligned} a + b &= a + c \Rightarrow (-a) + (a + b) = (-a) + (a + c) \\ &\Rightarrow [(-a) + a] + b = [(-a) + a] + c && \text{by postulate 1b} \\ &\Rightarrow 0 + b = 0 + c && \text{by postulate 1d} \\ &\Rightarrow b = c && \text{by postulate 1c.} \end{aligned}$$

This completes the proof of the lemma.

[†]We adopt the usual convention that the juxtaposition of x and y in xy indicates the operation of multiplication.

Proof of the Theorem Returning to the theorem, we see that we only need to show that $xy + (-x)y = xy + [-(xy)]$. That is, we need only show that $xy + (-x)y = 0$. We have

$$\begin{aligned}
 xy + (-x)y &= [x + (-x)]y && \text{by Theorem 2.1} \\
 &= 0 \cdot y && \text{by postulate 1d} \\
 &= 0 \cdot y + 0 && \text{by postulate 1c} \\
 &= 0 \cdot y + \{0 \cdot y + [-(0 \cdot y)]\} && \text{by postulate 1d} \\
 &= (0 \cdot y + 0 \cdot y) + [-(0 \cdot y)] && \text{by postulate 1b} \\
 &= (0 + 0)y + [-(0 \cdot y)] && \text{by Theorem 2.1} \\
 &= 0 \cdot y + [-(0 \cdot y)] && \text{by postulate 1c} \\
 &= 0 && \text{by postulate 1d.}
 \end{aligned}$$

We have shown that $xy + (-x)y = 0$, and the theorem is proved.

The proof of Theorem 2.2 would have been shorter if the fact that $0 \cdot y = 0$ had been available. However, our approach at present is to use in a proof only the basic postulates for \mathbf{Z} and those facts previously proved. Several statements similar to the last two theorems are given to be proved in the exercises at the end of this section. After this section, we assume the usual properties of addition and multiplication in \mathbf{Z} .

Postulate 4, which asserts the existence of the set \mathbf{Z}^+ of positive integers, can be used to introduce the order relation “less than” on the set of integers. We make the following definition.

Definition 2.4 ■ The Order Relation Less Than

For integers x and y ,

$$x < y \quad \text{if and only if} \quad y - x \in \mathbf{Z}^+$$

where $y - x = y + (-x)$.

The symbol $<$ is read “less than,” as usual. Here we have defined the relation, but we have not assumed any of its usual properties. However, they can be obtained by use of this definition and the properties of \mathbf{Z}^+ . Before illustrating this with an example, we note that $0 < y$ if and only if $y \in \mathbf{Z}^+$.

For an arbitrary $x \in \mathbf{Z}$ and a positive integer n , we define x^n as follows:

$$\begin{aligned}
 x^1 &= x \\
 x^{k+1} &= x^k \cdot x \quad \text{for any positive integer } k.
 \end{aligned}$$

Similarly, positive multiples nx of x are defined by

$$\begin{aligned}
 1x &= x \\
 (k+1)x &= kx + x \quad \text{for any positive integer } k.
 \end{aligned}$$

Strategy ■ Some proofs must be divided into different cases because the same argument does not apply to all elements under consideration. The proof of the next theorem separates naturally into two cases, based on the law of trichotomy (postulate 4c).

Theorem 2.5 ■ Squares of Nonzero Integers

For any $x \neq 0$ in \mathbf{Z} , $x^2 \in \mathbf{Z}^+$.

$p \Rightarrow q$ **Proof** Let $x \neq 0$ in \mathbf{Z} . By postulate 4, either $x \in \mathbf{Z}^+$ or $-x \in \mathbf{Z}^+$. If $x \in \mathbf{Z}^+$, then $x^2 = x \cdot x$ is in \mathbf{Z}^+ by postulate 4b. And if $-x \in \mathbf{Z}^+$, then $(-x)^2 = (-x) \cdot (-x)$ is in \mathbf{Z}^+ , by the same postulate. But

$$\begin{aligned} x^2 &= x \cdot x \\ &= (-x) \cdot (-x) \quad \text{by Exercise 5 in this section,} \end{aligned}$$

so x^2 is in \mathbf{Z}^+ if $-x \in \mathbf{Z}^+$. In each possible case, x^2 is in \mathbf{Z}^+ , and this completes the proof.

As a particular case of this theorem, $1 \in \mathbf{Z}^+$, since $1 = (1)^2$. That is, 1 must be a positive integer, a fact that may not be immediately evident in postulate 4. This in turn implies that $2 = 1 + 1$ is in \mathbf{Z}^+ , by postulate 4a. Repeated application of 4a gives $3 = 2 + 1 \in \mathbf{Z}^+$, $4 = 3 + 1 \in \mathbf{Z}^+$, $5 = 4 + 1 \in \mathbf{Z}^+$, and so on. It turns out that \mathbf{Z}^+ must necessarily be the set

$$\mathbf{Z}^+ = \{1, 2, 3, \dots, n, n + 1, \dots\}.$$

Although our discussion of order has been in terms of *less than*, the relations *greater than*, *less than or equal to*, and *greater than or equal to* can be introduced in \mathbf{Z} and similarly treated. We consider this treatment to be trivial and do not bother with it. At the same time, we accept terms such as *nonnegative* and *nonpositive* with their usual meanings and without formal definitions.

Exercises 2.1

True or False

Label each of the following statements as either true or false.

1. The set \mathbf{Z} of integers is closed with respect to subtraction.
2. The set $\mathbf{Z} - \mathbf{Z}^+$ is closed with respect to subtraction.
3. The set $\mathbf{Z} - \mathbf{Z}^+$ is closed with respect to multiplication.
4. If $xy = xz$ for all x, y , and z in \mathbf{Z} , then $y = z$.
5. Let A be a set of integers closed under subtraction. If x and y are elements of A then $x - ny$ is in A for any n in \mathbf{Z} .
6. $|x| \leq x$ for all x in \mathbf{Z} . (See the exercises for the definition of $|x|$, the **absolute value** of x .)

7. $|x + y|^2 \leq |x|^2 + |y|^2$ for all x and y in \mathbf{Z} .
 8. If $x < y$ then $x^2 < y^2$ for all x and y in \mathbf{Z} .
 9. If $x < y$ then $x^3 < y^3$ for all x and y in \mathbf{Z} .
 10. $\|x\| - \|y\| \leq |x - y|$ for all x and y in \mathbf{Z} .
-

Exercises

Prove that the equalities in Exercises 1–11 hold for all x , y , z , and w in \mathbf{Z} . Assume only the basic postulates for \mathbf{Z} and those properties proved in this section. **Subtraction** is defined by $x - y = x + (-y)$.

1. $x \cdot 0 = 0$
2. $-x = (-1) \cdot x$
3. $-(-x) = x$
4. $(-1)(-1) = 1$
5. $(-x)(-y) = xy$
6. $x - 0 = x$
7. $x(y - z) = xy - xz$
8. $(y - z)x = yx - zx$
9. $-(x + y) = (-x) + (-y)$
10. $(x - y) + (y - z) = x - z$
11. $(x + y)(z + w) = xz + xw + yz + yw$
12. Let A be a set of integers closed under subtraction.
 - a. Prove that if A is nonempty, then 0 is in A .
 - b. Prove that if x is in A then $-x$ is in A .

Sec. 2.2, #21 ≪

In Exercises 13–24, prove the statements concerning the relation $<$ on the set \mathbf{Z} of all integers.

13. If $x < y$, then $x + z < y + z$.
14. If $x < y$ and $z < w$, then $x + z < y + w$.
15. If $x = y$ and $0 < z$, then $y < x + z$.
16. If $x = y$ and $z < 0$, then $x + z < y$.
17. If $x < y$ and $y < z$, then $x < z$.
18. If $x < y$ and $0 < z$, then $xz < yz$.
19. If $x < y$ and $z < 0$, then $yz < xz$.

20. If $0 < x < y$, then $x^2 < y^2$.
21. If $0 < x < y$ and $0 < z < w$, then $xz < yw$.
22. If $0 < z$ and $xz < yz$, then $x < y$.
23. $z - x < z - y$ if and only if $y < x$.
24. If $x < y$, then $-y < -x$.
25. Prove that if x and y are integers and $xy = 0$, then either $x = 0$ or $y = 0$. (*Hint:* If $x \neq 0$, then either $x \in \mathbf{Z}^+$ or $-x \in \mathbf{Z}^+$, and similarly for y . Consider xy for the various cases.)
26. Prove that the cancellation law for multiplication holds in \mathbf{Z} . That is, if $xy = xz$ and $x \neq 0$, then $y = z$.
27. Let x and y be in \mathbf{Z} , not both zero, then $x^2 + y^2 \in \mathbf{Z}^+$.

For an integer x , the **absolute value** of x is denoted by $|x|$ and is defined by

$$|x| = \begin{cases} x & \text{if } 0 \leq x \\ -x & \text{if } x < 0. \end{cases}$$

Use this definition for the proofs in Exercises 28–30.

28. Prove that $-|x| \leq x \leq |x|$ for any integer x .
29. Prove that $|xy| = |x| \cdot |y|$ for all x and y in \mathbf{Z} .
30. Prove that $|x + y| \leq |x| + |y|$ for all x and y in \mathbf{Z} .
31. Prove that if a is positive and b is negative, then ab is negative.
32. Prove that if a is positive and ab is positive, then b is positive.
33. Prove that if a is positive and ab is negative, then b is negative.
34. Prove or disprove that $0 \leq x^2 - xy + y^2$ for all x and y in \mathbf{Z} .
35. Consider the set $\{0\}$ consisting of 0 alone, with $0 + 0 = 0$ and $0 \cdot 0 = 0$. Which of the postulates for \mathbf{Z} are satisfied?

Sec. 2.2, #44 <

2.2

Mathematical Induction

From this point on, full knowledge of the properties of addition, subtraction, and multiplication of integers is assumed. A study of divisibility begins in Section 2.3.

As mentioned in the last section, the induction postulate forms a basis for the method of proof known as mathematical induction. Some students may have encountered this method of proof in calculus or in other previous courses. In this case, it is possible to skip this section and continue with Section 2.3.

Strategy ■ **Proof by Mathematical Induction** In a typical proof by induction, there is a statement P_n to be proved true for every positive integer n . The proof consists of three steps:

1. The statement is verified for $n = 1$.
2. The statement is assumed true for $n = k$.
3. With this assumption made, the statement is then proved to be true for $n = k + 1$.

The assumption that is made in step 2 is called the **inductive assumption** or the **induction hypothesis**.

Principle of Mathematical Induction

The logic of the method is that

- a. if P_n is true for $n = 1$, and
- b. if the truth of P_k always implies that P_{k+1} is true,

then the statement P_n is true for all positive integers n . This logic fits the induction postulate perfectly if we let S be the set of all positive integers n for which P_n is true. When the induction postulate is used in this form, it is frequently called the **Principle of Mathematical Induction**.

Example 1 We shall prove that

$$\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \frac{1}{5 \cdot 7} + \cdots + \frac{1}{(2n-1)(2n+1)} = \frac{n}{2n+1}$$

for every positive integer n .

For each positive integer n , let P_n be the statement

$$\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \frac{1}{5 \cdot 7} + \cdots + \frac{1}{(2n-1)(2n+1)} = \frac{n}{2n+1}.$$

In an equation of this type, it is understood that $1/[(2n-1)(2n+1)]$ is the last term on the left side. When $n = 1$, there is only one term, and no addition is actually performed.

When $n = 1$, the value of the left side is

$$\frac{1}{[2(1)-1][2(1)+1]} = \frac{1}{1 \cdot 3} = \frac{1}{3}$$

and the value of the right side is

$$\frac{1}{2(1)+1} = \frac{1}{3}.$$

Thus P_1 is true.

Assume now that P_k is true. That is, assume that the equation

$$\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \frac{1}{5 \cdot 7} + \cdots + \frac{1}{(2k-1)(2k+1)} = \frac{k}{2k+1}$$

is true. With this assumption made, we need to prove that P_{k+1} is true. By adding

$$\frac{1}{[2(k+1)-1][2(k+1)+1]} = \frac{1}{(2k+1)(2k+3)}$$

to both sides of the assumed equality, we obtain

$$\begin{aligned} \frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \cdots + \frac{1}{(2k-1)(2k+1)} + \frac{1}{(2k+1)(2k+3)} \\ = \frac{k}{2k+1} + \frac{1}{(2k+1)(2k+3)} \\ = \frac{k(2k+3) + 1}{(2k+1)(2k+3)} \\ = \frac{2k^2 + 3k + 1}{(2k+1)(2k+3)} \\ = \frac{(2k+1)(k+1)}{(2k+1)(2k+3)} \\ = \frac{k+1}{2(k+1)+1}. \end{aligned}$$

The last expression matches exactly the fraction

$$\frac{n}{2n+1}$$

when n is replaced by $k+1$. Thus P_{k+1} is true whenever P_k is true.

It follows from the induction postulate that P_n is true for all positive integers n . ■

Example 2 We shall prove that any even positive power of a nonzero integer is positive. That is, if $x \neq 0$ in \mathbb{Z} , then x^{2n} is positive for every positive integer n .

The second formulation of the statement is suitable for a proof by induction on n . For $n = 1$, $x^{2n} = x^2$, and x^2 is positive by Theorem 2.5. Assume the statement is true for $n = k$; that is, x^{2k} is positive. For $n = k+1$, we have

$$\begin{aligned} x^{2n} &= x^{2(k+1)} \\ &= x^{2k+2} \\ &= x^{2k} \cdot x^2. \end{aligned}$$

Since x^{2k} and x^2 are positive, the product is positive by postulate 4b. Thus the statement is true for $n = k+1$. It follows from the induction postulate that the statement is true for all positive integers. ■

In Section 2.3 and in some of the exercises at the end of this section, we need to use the fact that 1 is the least positive integer. It might seem a bit strange to prove something so obvious, but the proof does reveal how this familiar fact is a consequence of the induction postulate.

Theorem 2.6 ■ Least Positive Integer

The integer 1 is the least positive integer. That is, $1 \leq x$ for all $x \in \mathbf{Z}^+$.

Induction Proof Let S be the set of all positive integers x such that $1 \leq x$. Then $1 \in S$. Suppose $k \in S$. Now $0 < 1$ implies $k = k + 0 < k + 1$, by Exercise 13 of Section 2.1, so we have $1 \leq k < k + 1$. Thus $k \in S$ implies $k + 1 \in S$, and $S = \mathbf{Z}^+$ by postulate 5.

Mathematical induction can sometimes be used in more complicated situations involving integers. Some statements that involve positive integers n are false for some values of the positive integer n but are true for all positive integers that are sufficiently large. Statements of this type can be proved by a modified form of mathematical induction. If a is a positive integer, and we wish to prove that a statement P_n is true for all positive integers $n \geq a$, we alter the three steps described in the strategy box of this section to the following form.

Strategy ■ Proof by Generalized Induction

1. The statement is verified for $n = a$.
2. The statement is assumed true for $n = k$, where $k \geq a$.
3. With this assumption made, the statement is then proved to be true for $n = k + 1$.

A proof of this type with $a = 4$ is given in Example 3.

Example 3

We shall prove that

$$1 + 3n < n^2$$

for every positive integer $n \geq 4$.

Note that the statement is actually false for $n = 1, 2$, and 3. For $n = 4$,

$$1 + 3n = 1 + 12 = 13 \quad \text{and} \quad n^2 = 4^2 = 16.$$

Since $13 < 16$, the statement is true for $n = 4$.

Assume now that the inequality is true for k where $k \geq 4$:

$$1 + 3k < k^2.$$

When $n = k + 1$, the left side of the inequality is $1 + 3(k + 1)$, and

$$\begin{aligned} 1 + 3(k + 1) &= 1 + 3k + 3 \\ &< k^2 + 3 \quad \text{since } 1 + 3k < k^2 \\ &= k^2 + 2 + 1 \\ &< k^2 + 2k + 1 \quad \text{since } 1 < k \text{ implies } 2 < 2k \\ &= (k + 1)^2. \end{aligned}$$

(In the steps involving $<$, we have used Exercises 13 and 18 of Section 2.1.) Since $(k + 1)^2$ is the right side of the inequality when $n = k + 1$, we have proved that

$$1 + 3n < n^2$$

is true when $n = k + 1$. Therefore, the inequality is true for all positive integers $n \geq 4$. ■

The modification of mathematical induction that is described just before Example 3 can be extended even more by allowing a to be 0 or a negative integer and using the same three steps listed in the strategy box to prove that a statement P_n is true for all integers $n \geq a$. This type of proof is requested in Exercise 23 of this section.

In some cases, it is more convenient to use yet another form of the induction postulate. This form is known by three different titles: It is called the **Second Principle of Finite Induction**, the method of proof by **Complete Induction**, and the method of **Strong Mathematical Induction**. In this form, a proof that a statement P_n is true for all integers $n \geq a$ consists of the following three steps.

Strategy ■ Proof by Complete Induction

1. The statement is proved true for $n = a$, where $a \in \mathbf{Z}$.
2. For an integer k , the statement is assumed true for all integers m such that $a \leq m < k$.
3. Under this assumption, the statement is proved to be true for $m = k$.

Our next example presents a proof by complete induction, and another example is provided by the proof of Theorem 2.18 in Section 2.4.

The fact stated in Example 4 is that every positive integer can be written as a sum of nonnegative powers of 2. This fact is a point of departure for developing the **binary representation** of real numbers, a representation that uses 2 as the number base instead of 10 as used in our familiar decimal system. Binary representations are used extensively in computer science.

Example 4 We shall use complete induction to prove the statement that every positive integer n can be expressed in the form

$$n = c_0 + c_1 \cdot 2 + c_2 \cdot 2^2 + \cdots + c_{j-1} \cdot 2^{j-1} + c_j \cdot 2^j,$$

where j is a nonnegative integer, $c_i \in \{0, 1\}$ for all $i < j$, and $c_j = 1$.

For $n = 1$, let $j = 0$ and $c_0 = 1$. Then

$$c_0 \cdot 2^0 = (1)(1) = 1,$$

and the statement is true for $n = 1$.

Assume now that $k > 1$ and the statement is true for all positive integers m such that $m < k$. We consider two cases: where k is even and where k is odd.

If k is even, then $k = 2p$ for some $p \in \mathbb{Z}^+$ with $p < k$. Since $p < k$, the induction hypothesis applies to p , and p can be written in the form

$$p = c_0 + c_1 \cdot 2 + c_2 \cdot 2^2 + \cdots + c_{j-1} \cdot 2^{j-1} + c_j \cdot 2^j,$$

where j is a nonnegative integer, $c_i \in \{0, 1\}$ for all i , and $c_j = 1$. Multiplying both sides of the equation for p by 2 gives

$$k = 2p = c_0 \cdot 2 + c_1 \cdot 2^2 + c_2 \cdot 2^3 + \cdots + c_{j-1} \cdot 2^j + c_j \cdot 2^{j+1},$$

and this is an equation for k that has the required form (when k is even).

Suppose now that k is odd, say, $k = 2p + 1$ for some $p \in \mathbb{Z}^+$. Since $k > 1$, this means that $k - 1 = 2p$ is in \mathbb{Z}^+ and

$$0 < p = \frac{k-1}{2} < \frac{k+k}{2} = k.$$

But $p < k$ implies that p can be written in the form

$$p = c_0 + c_1 \cdot 2 + c_2 \cdot 2^2 + \cdots + c_{j-1} \cdot 2^{j-1} + c_j \cdot 2^j$$

where $c_i \in \{0, 1\}$, and $c_j = 1$. Therefore,

$$2p = c_0 \cdot 2 + c_1 \cdot 2^2 + c_2 \cdot 2^3 + \cdots + c_{j-1} \cdot 2^j + c_j \cdot 2^{j+1}$$

and

$$\begin{aligned} k &= 2p + 1 \\ &= 1 + c_0 \cdot 2 + c_1 \cdot 2^2 + \cdots + c_{j-1} \cdot 2^j + c_j \cdot 2^{j+1}, \end{aligned}$$

which is an equation for k of the required form (when k is odd).

Combining the arguments for k even and k odd, we have proved that if $k > 1$ and the statement is true for all positive integers less than k , then it is also true for $n = k$. By the Second Principle of Finite Induction, the statement is true for all positive integers n . ■

Exercises 2.2

Prove that the statements in Exercises 1–14 are true for every positive integer n .

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

$$2. 1 + 3 + 5 + \cdots + (2n-1) = n^2$$

$$3. 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$4. 1^2 + 3^2 + 5^2 + \cdots + (2n-1)^2 = \frac{n(2n-1)(2n+1)}{3}$$

5. $2 + 2^2 + 2^3 + \cdots + 2^n = 2(2^n - 1)$

6. $1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$

7. $1^3 + 3^3 + 5^3 + \cdots + (2n-1)^3 = n^2(2n^2 - 1)$

8. $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) = \frac{n(n+1)(n+2)}{3}$

9. $1 \cdot 2 + 2 \cdot 2^2 + 3 \cdot 2^3 + \cdots + n \cdot 2^n = (n-1)2^{n+1} + 2$

10. $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}$

11. $\frac{1}{1 \cdot 4} + \frac{1}{4 \cdot 7} + \frac{1}{7 \cdot 10} + \cdots + \frac{1}{(3n-2)(3n+1)} = \frac{n}{3n+1}$

12. $\frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{2 \cdot 3 \cdot 4} + \frac{1}{3 \cdot 4 \cdot 5} + \cdots + \frac{1}{n(n+1)(n+2)} = \frac{n(n+3)}{4(n+1)(n+2)}$

13. $a + (a+d) + (a+2d) + \cdots + [a + (n-1)d] = \frac{n}{2}[2a + (n-1)d]$

14. $a + ar + ar^2 + \cdots + ar^{n-1} = a \frac{1-r^n}{1-r}$ if $r \neq 1$

Let x and y be integers, and let m and n be positive integers. Use mathematical induction to prove the statements in Exercises 15–20. (The definitions of x^n and nx are given before Theorem 2.5 in Section 2.1.)

15. $(xy)^n = x^n y^n$

16. $x^m \cdot x^n = x^{m+n}$

17. $(x^m)^n = x^{mn}$

18. $n(x+y) = nx+ny$

19. $(m+n)x = mx+nx$

20. $m(nx) = (mn)x$

Sec. 2.1, #12 ➤

21. Let A be a set of integers closed under subtraction. Prove that if x and y are in A , then $x - ny$ is in A for every positive integer n .

22. Let a and b be real numbers, and let n and r be integers with $0 \leq r \leq n$. The **binomial theorem** states that

$$\begin{aligned}(a+b)^n &= \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots \\ &\quad + \binom{n}{n-2}a^2b^{n-2} + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n \\ &= \sum_{r=0}^n \binom{n}{r}a^{n-r}b^r,\end{aligned}$$

where the **binomial coefficients** $\binom{n}{r}$ are defined by

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

with $r! = r(r - 1) \cdots (2)(1)$ for $r \geq 1$ and $0! = 1$. Prove that the binomial coefficients satisfy the equation

$$\text{Sec. 8.4, #35} \ll \binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r} \quad \text{for } 1 \leq r \leq n.$$

- $\text{Sec. 2.3, #48} \ll$ **23.** Use Exercise 22 and generalized induction to prove that $\binom{n}{r}$ is an integer for all integers n and r with $0 \leq r \leq n$.

- $\text{Sec. 6.3, #12} \ll$ **24.** Use the equation

$$\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r} \quad \text{for } 1 \leq r \leq n$$

and mathematical induction on n to prove the binomial theorem as it is stated in Exercise 22.

If B_1 , B_2 , and B_3 are matrices in $M_{n \times p}(\mathbf{R})$, part **b** of Theorem 1.30 implies that $B_1 + (B_2 + B_3) = (B_1 + B_2) + B_3$. For each positive integer $j \geq 3$, this associative property can be extended to the following generalized statement: Regardless of how symbols of grouping are introduced in the sum $B_1 + B_2 + \cdots + B_j$, the resulting value is the same matrix, and this justifies writing the sum without symbols of grouping. The generalized statement for sums is proved in Exercise 25 of Section 3.2 and for products in Theorem 3.7. Use these results in Exercises 25–27.

- 25.** Let A be an $m \times n$ matrix over \mathbf{R} , and let B_1, B_2, \dots, B_j be $n \times p$ matrices over \mathbf{R} . Use Theorem 1.33 and mathematical induction to prove that

$$A(B_1 + B_2 + \cdots + B_j) = AB_1 + AB_2 + \cdots + AB_j$$

for every positive integer j .

- 26.** Let C be a $p \times q$ matrix over \mathbf{R} , and let B_1, B_2, \dots, B_j be $n \times p$ matrices over \mathbf{R} . Use Theorem 1.33 and mathematical induction to prove that

$$(B_1 + B_2 + \cdots + B_j)C = B_1C + B_2C + \cdots + B_jC$$

for every positive integer j .

- $\text{Sec. 1.6, #31} \gg$ **27.** If A_1, A_2, \dots, A_n are square matrices of order m over \mathbf{R} and each A_i is invertible, then the product $A_1 A_2 \cdots A_n$ is invertible. Use the reverse order law for inverses and mathematical induction to prove

$$(A_1 A_2 \cdots A_n)^{-1} = A_n^{-1} \cdots A_2^{-1} A_1^{-1}$$

for all positive integers n .

In Exercises 28–32, use mathematical induction to prove that the given statement is true for all positive integers n .

- 28.** $4n > n + 2$

- 29.** $n < 2^n$

30. $1 + 2n \leq 3^n$

31. $x^n < y^n$, where x and y are integers with $0 < x < y$

32. $n! \leq n^n$

In Exercises 33–35, use mathematical induction on n to prove that the given statement is true.

Sec. 1.1, #10 ➤ **33.** If n is a nonnegative integer and the set A has n elements, then the power set $\mathcal{P}(A)$ has 2^n elements.

Sec. 1.1, #10 ➤ **34.** If $n \geq 2$ and the set A has n elements, then the number of elements of the power set $\mathcal{P}(A)$ containing exactly 2 elements is $\binom{n}{2} = \frac{n(n-1)}{2}$.

Sec. 1.1, #10 ➤ **35.** If $n \geq 3$ and the set A has n elements, then the number of elements of the power set $\mathcal{P}(A)$ containing exactly 3 elements is $\binom{n}{3} = \frac{n(n-1)(n-2)}{3!}$.

Sec. 1.1, #10 ➤ **36.** Exercises 33–35 can be generalized as follows: If $0 \leq k \leq n$ and the set A has n elements, then the number of elements of the power set $\mathcal{P}(A)$ containing exactly k elements is $\binom{n}{k}$.

- a.** Use this result to write an expression for the total number of elements in the power set $\mathcal{P}(A)$.
- b.** Use the binomial theorem as stated in Exercise 22 to evaluate the expression in part **a** and compare this result to Exercise 33. (*Hint:* Set $a = b = 1$ in the binomial theorem.)

In Exercises 37–41, use generalized induction to prove the given statement.

37. $1 + n < n^2$ for all integers $n \geq 2$

38. $1 + 2n < n^3$ for all integers $n \geq 2$

39. $1 + 2n < 2^n$ for all integers $n \geq 3$

40. $2^n < n!$ for all integers $n \geq 4$

41. $n^3 < n!$ for all integers $n \geq 6$

42. Use generalized induction and Exercise 37 to prove that $n^2 < n!$ for all integers $n \geq 4$.

43. Use generalized induction and Exercise 39 to prove that $n^2 < 2^n$ for all integers $n \geq 5$. (In connection with this result, see the discussion of *counterexamples* in the Appendix.)

Sec. 2.1, #30 ➤ **44.** Assume the statement from Exercise 30 in Section 2.1 that $|x + y| \leq |x| + |y|$ for all x and y in \mathbb{Z} . Use this assumption and mathematical induction to prove that

$$|a_1 + a_2 + \dots + a_n| \leq |a_1| + |a_2| + \dots + |a_n|$$

for all integers $n \geq 2$ and arbitrary integers a_1, a_2, \dots, a_n .

45. Show that if the statement

$$1 + 2 + 2^2 + \dots + 2^{n-1} = 2^n$$

is assumed to be true for $n = k$, then it can be proved to be true for $n = k + 1$. Is the statement true for all positive integers n ? Why?

- 46.** Show that if the statement

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

is assumed to be true for $n = k$, the same equation can be proved to be true for $n = k + 1$. Explain why this does not prove that the statement is true for all positive integers. Is the statement true for all positive integers? Why?

- 47.** Given the recursively defined sequence $a_1 = 1$, $a_2 = 4$, and $a_n = 2a_{n-1} - a_{n-2} + 2$, use complete induction to prove that $a_n = n^2$ for all positive integers n .
- 48.** Given the recursively defined sequence $a_1 = 1$, $a_2 = 3$, $a_3 = 9$, and $a_n = a_{n-1} + 3a_{n-2} + 9a_{n-3}$, use complete induction to prove that $a_n = 3^{n-1}$ for all positive integers n .
- 49.** Given the recursively defined sequence $a_1 = 0$, $a_2 = -30$, and $a_n = 8a_{n-1} - 15a_{n-2}$, use complete induction to prove that $a_n = 5 \cdot 3^n - 3 \cdot 5^n$ for all positive integers n .
- 50.** Given the recursively defined sequence $a_1 = 3$, $a_2 = 7$, $a_3 = 13$, and $a_n = 3a_{n-1} - 3a_{n-2} + a_{n-3}$, use complete induction to prove that $a_n = n^2 + n + 1$ for all positive integers n .
- 51.** The **Fibonacci[†] sequence** $\{f_n\} = 1, 1, 2, 3, 5, 8, 13, 21, \dots$ is defined recursively by

$$f_1 = 1, \quad f_2 = 1, \quad f_{n+2} = f_{n+1} + f_n \quad \text{for } n = 1, 2, 3, \dots$$

- a.** Prove $f_1 + f_2 + \cdots + f_n = f_{n+2} - 1$ for all positive integers n .
- b.** Use complete induction to prove that $f_n < 2^n$ for all positive integers n .
- c.** Use complete induction to prove that f_n is given by the explicit formula

$$f_n = \frac{(1 + \sqrt{5})^n - (1 - \sqrt{5})^n}{2^n \sqrt{5}}.$$

(This equation is known as **Binet's formula**, named after the 19th-century French mathematician Jacques Binet^{††}.)

- 52.** Let f_1, f_2, \dots, f_n be permutations on a nonempty set A . Prove that

$$(f_1 \circ f_2 \circ \cdots \circ f_n)^{-1} = f_n^{-1} \circ \cdots \circ f_2^{-1} \circ f_1^{-1}$$

for all positive integers n .

[†]The Fibonacci sequence was first introduced to the western world in 1202 by Leonardo of Pisa (c. 1170–c. 1250), who was posthumously given the nickname Fibonacci. Considered as one of the most talented mathematicians of the Middle Ages, Fibonacci appreciated the superiority of the Hindu-Arabic numeral system (as opposed to the Roman numeral system) for its ease in performing the basic arithmetic operations and is credited for introducing this system into Europe.

^{††}Jacques Binet (1786–1856) is credited for this formula for the n th term in the Fibonacci sequence (although it was known by Euler over a century earlier) and for developing the rule for matrix multiplication in 1812. Binet was also a noted physicist and astronomer.

53. Define powers of a permutation f on A by the following:

$$f^0 = I_A, \quad f^1 = f, \quad \text{and} \quad f^n = f^{n-1} \circ f \quad \text{for } n > 1.$$

Let f and g be permutations on a nonempty set A . Prove that

$$(f^{-1} \circ g \circ f)^n = f^{-1} \circ g^n \circ f$$

for all positive integers n .

2.3 Divisibility

We turn now to a study of divisibility in the set of integers. Our main goal in this section is to obtain the **Division Algorithm** (Theorem 2.10). To achieve this, we need an important consequence of the induction postulate, known as the **Well-Ordering Theorem**.

Theorem 2.7 ■ The Well-Ordering Theorem

Every nonempty set S of positive integers contains a least element. That is, there is an element $m \in S$ such that $m \leq x$ for all $x \in S$.

$p \Rightarrow q$ **Proof** Let S be a nonempty set of positive integers. If $1 \in S$, then $1 \leq x$ for all $x \in S$, by Theorem 2.6. In this case, $m = 1$ is the least element in S .

Consider now the case where $1 \notin S$, and let L be the set of all positive integers p such that $p < x$ for all $x \in S$. That is,

$$L = \{p \in \mathbf{Z}^+ \mid p < x \text{ for all } x \in S\}.$$

Since $1 \notin S$, Theorem 2.6 assures us that $1 \in L$. We shall show that there is a positive integer p_0 such that $p_0 \in L$ and $p_0 + 1 \notin L$. Suppose this is not the case. Then we have that $p \in L$ implies $p + 1 \in L$, and $L = \mathbf{Z}^+$ by the induction postulate. This contradicts the fact that S is nonempty (note that $L \cap S = \emptyset$). Therefore, there is a p_0 such that $p_0 \in L$ and $p_0 + 1 \notin L$.

We must show that $p_0 + 1 \in S$. We have $p_0 < x$ for all $x \in S$, so $p_0 + 1 \leq x$ for all $x \in S$ (see Exercise 28 at the end of this section). If $p_0 + 1 < x$ were always true, then $p_0 + 1$ would be in L . Hence $p_0 + 1 = x$ for some $x \in S$, and $m = p_0 + 1$ is the required least element in S .

Definition 2.8 ■ Divisor, Multiple

Let a and b be integers. We say that a **divides** b if there is an integer c such that $b = ac$.

If a divides b , we write $a|b$. Also, we say that b is a **multiple** of a , or that a is a **factor** of b , or that a is a **divisor** of b . If a does not divide b , we write $a\nmid b$.

It may come as a surprise that we can use our previous results to prove the following simple theorem.

Theorem 2.9 ■ Divisors of 1

The only divisors of 1 are 1 and -1 .

$p \Rightarrow (q \vee r)$ **Proof** Suppose a is a divisor of 1. Then $1 = ac$ for some integer c . The equation $1 = ac$ requires $a \neq 0$, so either $a \in \mathbf{Z}^+$ or $-a \in \mathbf{Z}^+$.

Consider first the case where $a \in \mathbf{Z}^+$. This requires $c \in \mathbf{Z}^+$ (see Exercise 32 of Section 2.1), so we have $1 \leq a$ and $1 \leq c$, by Theorem 2.6. Now

$$\begin{aligned} 1 < a &\Rightarrow 1 \cdot c < a \cdot c && \text{by Exercise 18 of Section 2.1} \\ &\Rightarrow c < 1 && \text{since } ac = 1, \end{aligned}$$

and this is a contradiction of $1 \leq c$. Thus $1 = a$ is the only possibility when $a \in \mathbf{Z}^+$.

Consider[†] now the case where $-a \in \mathbf{Z}^+$. By Exercise 5 of Section 2.1, we have

$$(-a)(-c) = ac = 1,$$

and $-a \in \mathbf{Z}^+$ implies that $-c \in \mathbf{Z}^+$ by Exercise 32 of Section 2.1. Therefore, $1 \leq -a$ and $1 \leq -c$ by Theorem 2.6. Now

$$\begin{aligned} 1 < -a &\Rightarrow (1)(-c) < (-a)(-c) && \text{by Exercise 18 of Section 2.1} \\ &\Rightarrow -c < 1 && \text{since } (-a)(-c) = 1, \end{aligned}$$

and $-c < 1$ is a contradiction to $1 \leq -c$. Therefore, $1 = -a$ is the only possibility when $-a \in \mathbf{Z}^+$, and we have

$$\begin{aligned} a &= -(-a) && \text{by Exercise 3 of Section 2.1} \\ &= -1 && \text{since } -a = 1. \end{aligned}$$

Combining the cases where $a \in \mathbf{Z}^+$ and where $-a \in \mathbf{Z}^+$, we have shown that either $a = 1$ or $a = -1$ if a is a divisor of 1.

Our next result is the basic theorem on divisibility.

Theorem 2.10 ■ The Division Algorithm

Let a and b be integers with $b > 0$. Then there exist unique integers q and r such that

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

Existence **Proof** Let S be the set of all integers x that can be written in the form $x = a - bn$ for $n \in \mathbf{Z}$, and let S' denote the set of all nonnegative integers in S . The set S' is nonempty.

[†]The proof for this case is similar to that where $a \in \mathbf{Z}^+$, but we include it here because it illustrates several uses of results from Section 2.1.

(See Exercise 29 at the end of this section.) If $0 \in S'$, we have $a - bq = 0$ for some q , and $a = bq + 0$. If $0 \notin S'$, then S' contains a least element $r = a - bq$, by the Well-Ordering Theorem, and

$$a = bq + r$$

where r is positive. Now

$$r - b = a - bq - b = a - b(q + 1),$$

so $r - b \in S$. Since r is the least element in S' and $r - b < r$, it must be true that $r - b$ is negative. That is, $r - b < 0$, and $r < b$. Combining the two cases (where $0 \in S'$ and where $0 \notin S'$), we have

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

Uniqueness

To show that q and r are unique, suppose $a = bq_1 + r_1$ and $a = bq_2 + r_2$, where $0 \leq r_1 < b$ and $0 \leq r_2 < b$. We may assume that $r_1 \leq r_2$ without loss of generality. This means that

$$0 \leq r_2 - r_1 \leq r_2 < b.$$

However, we also have

$$0 \leq r_2 - r_1 = (a - bq_2) - (a - bq_1) = b(q_1 - q_2).$$

That is, $r_2 - r_1$ is a nonnegative multiple of b that is less than b . For any positive integer n , $1 \leq n$ implies $b \leq bn$. Therefore, $r_2 - r_1 = 0$ and $r_1 = r_2$. It follows that $bq_1 = bq_2$, where $b \neq 0$. This implies that $q_1 = q_2$ (see Exercise 26 of Section 2.1). We have shown that $r_1 = r_2$ and $q_1 = q_2$, and this proves that q and r are unique.

The word *algorithm* in the heading of Theorem 2.10 may seem strange at first glance, since an algorithm is usually a repetitive procedure for obtaining a result. The use of the word here is derived from the fact that the elements $a - bn$ of S' in the proof may be found by repeated subtraction of b :

$$a - b, a - 2b, a - 3b,$$

and so on.

In the Division Algorithm, the integer q is called the **quotient** and r is called the **remainder** in the division of a by b . The conclusion of the theorem may be more familiar in the form

$$\frac{a}{b} = q + \frac{r}{b},$$

but we are restricting our work here so that only integers are involved.

Example 1 When a and b are both positive integers, the quotient q and remainder r can be found by the familiar routine of long division. For instance, if $a = 357$ and $b = 13$, long division gives

$$\begin{array}{r} 27 \\ 13 \overline{)357} \\ 26 \\ \hline 97 \\ 91 \\ \hline 6 \end{array}$$

so $q = 27$ and $r = 6$ in $a = bq + r$, with $0 \leq r < b$:

$$357 = (13)(27) + 6.$$

If a is negative, a minor adjustment (see Exercise 30 of this section) can be made to obtain the expression in the Division Algorithm. With $a = -357$ and $b = 13$, the preceding equation can be multiplied by -1 to obtain

$$-357 = (13)(-27) + (-6).$$

To obtain an expression with a positive remainder, we need only subtract and add 13 in the right member of the equation:

$$\begin{aligned} -357 &= (13)(-27) + (13)(-1) + (-6) + 13 \\ &= (13)(-28) + 7. \end{aligned}$$

Thus $q = -28$ and $r = 7$ in the Division Algorithm, with $a = -357$ and $b = 13$. ■

Exercises 2.3

True or False

Label each of the following statements as either true or false.

1. The Well-Ordering Theorem implies that the set of even integers contains a least element.
2. Let b be any integer. Then $0|b$.
3. Let b be any integer. Then $b|0$.
4. $0|b$ only if $b = 0$.
5. Let a and b be integers with $b > 0$. Then $b|a$ if and only if the remainder r in the Division Algorithm, when a is divided by b , is 0.
6. Let a and b be integers with $a \neq 0$, such that $a|b$. Then $a|-b$ and $-a|b$ and $-a|-b$.
7. Let a and b be integers. Then $2|ab(a + b)$.
8. If $a|c$ and $b|c$, then $ab|c$.
9. If $a|b$ and $b|a$, then $a = b$.

Exercises

1. List all divisors of the following integers.

a. 30	b. 42	c. 28	d. 45
e. 24	f. 40	g. 32	h. 210

 2. List all common divisors of each of the following pairs of integers.

a. 30, 28	b. 42, 45	c. 24, 32	d. 210, 40
e. -40, 24	f. -30, -50		
- With a and b as given in Exercises 3–16, find the q and r that satisfy the conditions in the Division Algorithm.
3. $a = 796, b = 26$
 4. $a = 512, b = 15$
 5. $a = 1149, b = 52$
 6. $a = 1205, b = 37$
 7. $a = -12, b = 5$
 8. $a = -27, b = 7$
 9. $a = -863, b = 17$
 10. $a = -921, b = 18$
 11. $a = 26, b = 796$
 12. $a = 15, b = 512$
 13. $a = -4317, b = 12$
 14. $a = -5316, b = 171$
 15. $a = 0, b = 3$
 16. $a = 0, b = 5$
17. Prove that if a , b , and c are integers such that $a|b$ and $a|c$, then $a|(b + c)$.
 18. Let R be the relation defined on the set of integers \mathbf{Z} by aRb if and only if $a|b$. Prove or disprove that R is an equivalence relation.
 19. Let a , b , c , m , and n be integers such that $a|b$ and $a|c$. Prove that $a|(mb + nc)$.
 20. Let a , b , c , and d be integers such that $a|b$ and $c|d$. Prove that $ac|bd$.
 21. Prove that if a and b are integers such that $a|b$ and $b|a$, then either $a = b$ or $a = -b$.
 22. Prove that if a and b are integers such that $b \neq 0$ and $a|b$, then $|a| \leq |b|$.
 23. Let a and b be integers such that $a|b$ and $|b| < |a|$. Prove that $b = 0$.
 24. Let a , b , and c be integers. Prove or disprove that $a|b$ implies $ac|bc$.
 25. Let a , b , and c be integers. Prove or disprove that $a|bc$ implies $a|b$ or $a|c$.
 26. Let a be an integer. Prove that $2|a(a + 1)$. (Hint: Consider two cases.)
 27. Let a be an integer. Prove that $3|a(a + 1)(a + 2)$. (Hint: Consider three cases.)
 28. Let m be an arbitrary integer. Prove that there is no integer n such that $m < n < m + 1$.
 29. Let S be as described in the proof of Theorem 2.10. Give a specific example of a positive element of S .
 30. Let a and b be integers with $b > 0$ and $a = bq + r$ with $0 \leq r < b$. Use this result to find the unique quotient and remainder as described by the Division Algorithm when $-a$ is divided by b .

31. Use the Division Algorithm to prove that if a and b are integers, with $b \neq 0$, then there exist unique integers q and r such that $a = bq + r$, with $0 \leq r < |b|$.
32. Prove that the Well-Ordering Theorem implies the induction postulate 5 in Section 2.1.
33. Assume that the Well-Ordering Theorem holds, and prove the second principle of finite induction.

In Exercises 34–47, use mathematical induction to prove that the given statement is true for all positive integers n .

34. 3 is a factor of $n^3 + 2n$
35. 3 is a factor of $n^3 - 7n$
36. 3 is a factor of $n^3 - n$
37. 3 is a factor of $n^3 + 5n$
38. 6 is a factor of $n^3 - n$
39. 6 is a factor of $n^3 + 5n$
40. 3 is a factor of $4^n - 1$
41. 8 is a factor of $9^n - 1$
42. 5 is a factor of $7^n - 2^n$
43. 4 is a factor of $9^n - 5^n$
44. 4 is a factor of $3^{2n} - 1$
45. 5 is a factor of $3^{2n} - 2^{2n}$
46. For all a and b in \mathbf{Z} , $a - b$ is a factor of $a^n - b^n$. (*Hint:* $a^{k+1} - b^{k+1} = a^k(a - b) + (a^k - b^k)b$)
47. For all a and b in \mathbf{Z} , $a + b$ is a factor of $a^{2n} - b^{2n}$.
48. a. The binomial coefficients $\binom{n}{r}$ are defined in Exercise 22 of Section 2.2. Use induction on r to prove that if p is a prime integer, then p is a factor of $\binom{p}{r}$ for $r = 1, 2, \dots, p - 1$. (From Exercise 23 of Section 2.2, it is known that $\binom{p}{r}$ is an integer.)
b. Use induction on n to prove that if p is a prime integer, then p is a factor of $n^p - n$.

Sec. 2.2, #23 ➤

2.4

Prime Factors and Greatest Common Divisor

In this section, we establish the existence of the greatest common divisor of two integers when at least one of them is nonzero. The **Unique Factorization Theorem**, also known as the **Fundamental Theorem of Arithmetic**, is obtained.

Definition 2.11 ■ Greatest Common Divisor

An integer d is a **greatest common divisor** of a and b if all these conditions are satisfied:

1. d is a positive integer.
2. $d|a$ and $d|b$.
3. $c|a$ and $c|b$ imply $c|d$.

The next theorem shows that the greatest common divisor d of a and b exists when at least one of them is not zero. Our proof also shows that d is a **linear combination** of a and b ; that is, $d = ma + nb$ for integers m and n .

Strategy ■ The technique of proof by use of the Well-Ordering Theorem in Theorem 2.12 should be compared to that used in the proof of the Division Algorithm (Theorem 2.10).

Theorem 2.12 ■ Greatest Common Divisor

Let a and b be integers, at least one of them not 0. Then there exists a unique greatest common divisor d of a and b . Moreover, d can be written as

$$d = am + bn$$

for integers m and n , and d is the smallest positive integer that can be written in this form.

Existence **Proof** Let a and b be integers, at least one of them not 0. If $b = 0$, then $a \neq 0$, so $|a| > 0$. It is easy to see that $d = |a|$ is a greatest common divisor of a and b in this case, and either $d = a \cdot (1) + b \cdot (0)$ or $d = a \cdot (-1) + b \cdot (0)$.

Suppose now that $b \neq 0$. Consider the set S of all integers that can be written in the form $ax + by$ for some integers x and y , and let S^+ be the set of all positive integers in S . The set S contains $b = a \cdot (0) + b \cdot (1)$ and $-b = a \cdot (0) + b \cdot (-1)$, so S^+ is not empty. By the Well-Ordering Theorem, S^+ has a least element d ,

$$d = am + bn.$$

We have d positive, and we shall show that d is a greatest common divisor of a and b .

By the Division Algorithm, there are integers q and r such that

$$a = dq + r \quad \text{with} \quad 0 \leq r < d.$$

From this equation,

$$\begin{aligned} r &= a - dq \\ &= a - (am + bn)q \\ &= a(1 - mq) + b(-nq). \end{aligned}$$

Thus r is in $S = \{ax + by\}$, and $0 \leq r < d$. By choice of d as the least element in S^+ , it must be true that $r = 0$, and $d|a$. Similarly, it can be shown that $d|b$.

If $c|a$ and $c|b$, then $a = ch$ and $b = ck$ for integers h and k . Therefore,

$$\begin{aligned} d &= am + bn \\ &= chm + ckn \\ &= c(hm + kn), \end{aligned}$$

and this shows that $c|d$. By Definition 2.11, $d = am + bn$ is a greatest common divisor of a and b . It follows from the choice of d as least element of S^+ that d is the smallest positive integer that can be written in this form.

Uniqueness

To show that the greatest common divisor of a and b is unique, assume that d_1 and d_2 are both greatest common divisors of a and b . Then it must be true that $d_1|d_2$ and $d_2|d_1$. Since d_1 and d_2 are positive integers, this means that $d_1 = d_2$ (see Exercise 21 of Section 2.3).

Whenever the greatest common divisor of a and b exists, we shall write (a, b) or $\gcd(a, b)$ to indicate the *unique greatest common divisor* of a and b .

When at least one of a and b is not 0, the proof of the last theorem establishes the existence of (a, b) , but looking for a smallest positive integer in $S = \{ax + by\}$ is not a very satisfactory method for finding this greatest common divisor. A procedure known as the **Euclidean Algorithm** furnishes a systematic method for finding (a, b) where $b > 0$. It can also be used to find integers m and n such that $(a, b) = am + bn$. This procedure consists of repeated applications of the Division Algorithm according to the following pattern, where a and b are integers with $b > 0$.

The Euclidean Algorithm

$$\begin{aligned} a &= bq_0 + r_1, & 0 \leq r_1 < b \\ b &= r_1 q_1 + r_2, & 0 \leq r_2 < r_1 \\ r_1 &= r_2 q_2 + r_3, & 0 \leq r_3 < r_2 \\ &\vdots & \vdots \\ r_k &= r_{k+1} q_{k+1} + r_{k+2}, & 0 \leq r_{k+2} < r_{k+1}. \end{aligned}$$

Since the integers r_1, r_2, \dots, r_{k+2} are decreasing and are all nonnegative, there is a smallest integer n such that $r_{n+1} = 0$:

$$r_{n-1} = r_n q_n + r_{n+1}, \quad 0 = r_{n+1}.$$

If we put $r_0 = b$, this last nonzero remainder r_n is always the greatest common divisor of a and b . The proof of this statement is left as an exercise.

As an example, we shall find the greatest common divisor of 1492 and 1776.

Example 1 Performing the arithmetic for the Euclidean Algorithm, we have

$$\begin{aligned} 1776 &= (1)(1492) + \mathbf{284} & (q_0 = 1, r_1 = 284) \\ 1492 &= (5)(\mathbf{284}) + \mathbf{72} & (q_1 = 5, r_2 = 72) \\ \mathbf{284} &= (3)(\mathbf{72}) + \mathbf{68} & (q_2 = 3, r_3 = 68) \\ \mathbf{72} &= (1)(\mathbf{68}) + \mathbf{4} & (q_3 = 1, r_4 = 4) \\ \mathbf{68} &= (4)(17) + 0 & (q_4 = 17, r_5 = 0). \end{aligned}$$

Thus the last nonzero remainder is $r_n = r_4 = 4$, and $(1776, 1492) = 4$. ■

As mentioned earlier, the Euclidean Algorithm can also be used to find integers m and n such that

$$(a, b) = am + bn.$$

We can obtain these integers by solving for the last nonzero remainder and substituting the remainders from the preceding equations successively until a and b are present in the equation. For example, the remainders in Example 1 can be expressed as

$$\begin{aligned} \mathbf{284} &= (1776)(1) + (1492)(-1) \\ \mathbf{72} &= (1492)(1) + (284)(-5) \\ \mathbf{68} &= (284)(1) + (72)(-3) \\ \mathbf{4} &= (72)(1) + (68)(-1). \end{aligned}$$

Substituting the remainders from the preceding equations successively, we have

$$\begin{aligned}
 4 &= (\mathbf{72})(1) + [(\mathbf{284})(1) + (\mathbf{72})(-3)](-1) \\
 &= (\mathbf{72})(1) + (\mathbf{284})(-1) + (\mathbf{72})(3) \\
 &= (\mathbf{72})(4) + (\mathbf{284})(-1) \quad \text{after the first substitution} \\
 &= [(\mathbf{1492})(1) + (\mathbf{284})(-5)](4) + (\mathbf{284})(-1) \\
 &= (\mathbf{1492})(4) + (\mathbf{284})(-20) + (\mathbf{284})(-1) \\
 &= (\mathbf{1492})(4) + (\mathbf{284})(-21) \quad \text{after the second substitution} \\
 &= (\mathbf{1492})(4) + [(\mathbf{1776})(1) + (\mathbf{1492})(-1)](-21) \\
 &= (\mathbf{1492})(4) + (\mathbf{1776})(-21) + (\mathbf{1492})(21) \\
 &= (\mathbf{1776})(-21) + (\mathbf{1492})(25) \quad \text{after the third substitution.}
 \end{aligned}$$

Thus $m = -21$ and $n = 25$ are integers such that

$$4 = 1776m + 1492n.$$

The remainders are printed in bold type in each of the preceding steps, and we carefully avoided performing a multiplication that involved a remainder.

The m and n are not unique in the equation

$$(a, b) = am + bn.$$

To see this, simply add and subtract the product ab :

$$\begin{aligned}
 (a, b) &= am + ab + bn - ab \\
 &= a(m + b) + b(n - a).
 \end{aligned}$$

Thus $m' = m + b$ and $n' = n - a$ are another pair of integers such that

$$(a, b) = am' + bn'.$$

Definition 2.13 ■ Relatively Prime Integers

Two integers a and b are **relatively prime** if their greatest common divisor is 1.

In the next two sections of this chapter, we prove some interesting results concerning those integers that are relatively prime to a given integer n . Theorem 2.14 is useful in the proofs of those results.

Theorem 2.14 ■

If a and b are relatively prime and $a|bc$, then $a|c$.

$(p \wedge q) \Rightarrow r$ **Proof** Assume that $(a, b) = 1$ and $a|bc$. Since $(a, b) = 1$, there are integers m and n such that $1 = am + bn$, by Theorem 2.12. Since $a|bc$, there exists an integer q such that $bc = aq$. Now,

$$\begin{aligned}
 1 &= am + bn \Rightarrow c = acm + bcn \\
 &\Rightarrow c = acm + aqn \quad \text{since } bc = aq \\
 &\Rightarrow c = a(cm + qn) \\
 &\Rightarrow a|c.
 \end{aligned}$$

Thus the theorem is proved.

Among the integers, there are those that have the fewest number of factors possible. Some of these are the *prime integers*.

Definition 2.15 ■ Prime Integer

An integer p is a **prime integer** if $p > 1$ and the only divisors of p are ± 1 and $\pm p$.

Note that the condition $p > 1$ makes p positive and ensures that $p \neq 1$. The exclusion of 1 from the set of primes makes possible the statement of the Unique Factorization Theorem. Before delving into that, we prove the important property of primes in Theorem 2.16.

Strategy ■ The conclusion in the next theorem has the form “ r or s .” One technique that can be used to prove an “or” statement such as this is to assume that one part (such as r) does not hold, and use this assumption to help prove that the other part must then hold.

Theorem 2.16 ■ Euclid's[†] Lemma

If p is a prime and $p|ab$, then either $p|a$ or $p|b$.

$(p \wedge q) \Rightarrow (r \vee s)$ **Proof** Assume p is a prime and $p|ab$. If $p|a$, the conclusion of the theorem is satisfied.

Suppose, then, that p does not divide a . This implies that $1 = (p, a)$, since the only positive divisors of p are 1 and p . Then Theorem 2.14 implies that $p|b$. Thus $p|b$ if p does not divide a , and the theorem is true in any case.

The following corollary generalizes Theorem 2.16 to products with more than two factors. Its proof is requested in the exercises. A direct result of this corollary is that if p is prime and $p|a^n$, then $p|a$.

Corollary 2.17 ■

If p is a prime and $p|(a_1a_2\cdots a_n)$, then p divides some a_j .

This brings us to the **Unique Factorization Theorem**, a result of such importance that it is frequently called the **Fundamental Theorem of Arithmetic**.

Strategy ■ Note the proof of the uniqueness part of Theorem 2.18: Two factorizations are assumed, and then it is proved that the two are equal.

[†]Euclid (c. 325 B.C.–c. 265 B.C.), a Greek mathematician considered to be the “Father of Geometry,” presented the principles of Euclidean geometry in his **Elements**, the most famous mathematics works in all of history.

Theorem 2.18 ■ Unique Factorization Theorem

Every positive integer n either is 1 or can be expressed as a product of prime integers, and this factorization is unique except for the order of the factors.

Complete
Induction

Proof In the statement of the theorem, the word *product* is used in an extended sense: The *product* may have just one factor.

Let P_n be the statement that either $n = 1$ or n can be expressed as a product of primes. We shall prove that P_n is true for all $n \in \mathbf{Z}^+$ by the Second Principle of Finite Induction.

Now P_1 is trivially true. Assume that P_m is true for all positive integers $m < k$. If k is a prime, then k is a product with one prime factor, and P_k is true. Suppose k is not a prime. Then $k = ab$, where neither a nor b is 1. Therefore, $1 < a < k$ and $1 < b < k$. By the induction hypothesis, P_a is true and P_b is true. That is,

$$a = p_1 p_2 \cdots p_r \quad \text{and} \quad b = q_1 q_2 \cdots q_s$$

for primes p_i and q_j . These factorizations give

$$k = ab = p_1 p_2 \cdots p_r q_1 q_2 \cdots q_s,$$

and k is thereby expressed as a product of primes. Thus P_k is true, and therefore P_n is true for all positive integers n .

Uniqueness

To prove that the factorization is unique, suppose that

$$n = p_1 p_2 \cdots p_t \quad \text{and} \quad n = q_1 q_2 \cdots q_v$$

are factorizations of n as products of prime factors p_i and q_j . Then

$$p_1 p_2 \cdots p_t = q_1 q_2 \cdots q_v,$$

so $p_1 | (q_1 q_2 \cdots q_v)$. By Corollary 2.17, $p_1 | q_j$ for some j , and there is no loss of generality if we assume $j = 1$. However, p_1 and q_1 are primes, so $p_1 | q_1$ implies $q_1 = p_1$. This gives

$$p_1 p_2 \cdots p_t = p_1 q_2 \cdots q_v,$$

and therefore

$$p_2 \cdots p_t = q_2 \cdots q_v$$

by the cancellation law. This argument can be repeated, removing one factor p_i with each application of the cancellation law, until we obtain

$$p_t = q_t \cdots q_v.$$

Since the only positive factors of p_t are 1 and p_t , and since each q_j is a prime, this means that there must be only one q_j on the right in this equation, and it is q_t . That is, $v = t$ and $q_t = p_t$. This completes the proof.

The Unique Factorization Theorem can be used to describe a standard form of a positive integer n . Suppose p_1, p_2, \dots, p_r are the *distinct* prime factors of n , arranged in order of magnitude so that

$$p_1 < p_2 < \cdots < p_r.$$

Then all repeated factors may be collected together and expressed by use of exponents to yield

$$n = p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r}$$

where each m_i is a positive integer. Each m_i is called the **multiplicity** of p_i , and this factorization is known as the **standard form** for n .

Example 2 The standard forms for two positive integers a and b can be used to find their greatest common divisor (a, b) and their least common multiple (see Exercises 28 and 29 at the end of this section). For instance, if

$$a = 31,752 = 2^3 \cdot 3^4 \cdot 7^2 \quad \text{and} \quad b = 126,000 = 2^4 \cdot 3^2 \cdot 5^3 \cdot 7,$$

then (a, b) can be found by forming the product of all the common prime factors, with each common factor raised to the least power to which it appears in either factorization:

$$(a, b) = 2^3 \cdot 3^2 \cdot 7 = 504.$$

■

From one point of view, the Unique Factorization Theorem says that the prime integers are building blocks for the integers, where the “building” is done by using multiplication and forming products. A natural question, then, is: How many blocks? Our next theorem states the answer given by the ancient Greek mathematician Euclid—that the number of primes is infinite. The proof is also credited to Euclid.

Theorem 2.19 ■ Euclid's Theorem on Primes

The number of primes is infinite.

Contradiction

Proof Assume there are only a finite number, n , of primes. Let these n primes be denoted by p_1, p_2, \dots, p_n , and consider the integer

$$m = p_1 p_2 \cdots p_n + 1.$$

It is clear that the remainder in the division of m by any prime p_i is 1, so each p_i is not a factor of m . Thus there are two possibilities: Either m is itself a prime, or it has a prime factor that is different from every one of the p_i . In either case, we have a prime integer that was not in the list p_1, p_2, \dots, p_n . Therefore, there are more than n primes, and this contradiction establishes the theorem.

Exercises 2.4

True or False

Label each of the following statements as either true or false.

1. The set of prime numbers is closed with respect to multiplication.
2. The set of prime numbers is closed with respect to addition.
3. The greatest common divisor is a binary operation from $\mathbf{Z} - \{0\} \times \mathbf{Z}$ to \mathbf{Z}^+ .

4. The least common multiple is a binary operation from $\mathbf{Z} - \{0\} \times \mathbf{Z} - \{0\}$ to \mathbf{Z}^+ .
 5. The greatest common divisor is unique, when it exists.
 6. Let a and b be integers, not both zero, such that $1 = (a, b)$. Then there exist integers x and y such that $1 = ax + by$ and $(x, y) = 1$.
 7. Let a and b be integers, not both zero, such that $d = ax + by$ for integers x and y . Then $d = (a, b)$.
 8. Let a and b be integers, not both zero, such that $d = (a, b)$. Then there exist unique integers x and y such that $d = ax + by$.
 9. Let a and b be integers, not both zero. Then $(a, b) = (-a, b)$.
 10. Let a be an integer, then $(a, a + 1) = 1$.
 11. Let a be an integer, then $(a, a + 2) = 2$.
 12. If $(a, b) = 1$ and $(a, c) = 1$, then $(b, c) = 1$.
-

Exercises

In this set of exercises, all variables represent integers.

1. List all the primes less than 100.
2. For each of the following pairs, write a and b in standard form and use these factorizations to find (a, b) .
 - a. $a = 1400, b = 980$
 - b. $a = 4950, b = 10,500$
 - c. $a = 3780, b = 16,200$
 - d. $a = 52,920, b = 25,200$
3. In each part, find the greatest common divisor (a, b) and integers m and n such that $(a, b) = am + bn$.

<ol style="list-style-type: none"> a. $a = 0, b = -3$ c. $a = 102, b = 66$ e. $a = 414, b = -33$ g. $a = 414, b = 693$ i. $a = 1197, b = 312$ k. $a = 6420, b = 132$ m. $a = 5088, b = -156$ 	<ol style="list-style-type: none"> b. $a = 65, b = -91$ d. $a = 52, b = 124$ f. $a = 252, b = -180$ h. $a = 382, b = 26$ j. $a = 3780, b = 1200$ l. $a = 602, b = 252$ n. $a = 8767, b = 252$
--	---
4. Find the smallest integer in the given set.
 - a. $\{x \in \mathbf{Z} | x > 0 \text{ and } x = 4s + 6t \text{ for some } s, t \text{ in } \mathbf{Z}\}$
 - b. $\{x \in \mathbf{Z} | x > 0 \text{ and } x = 6s + 15t \text{ for some } s, t \text{ in } \mathbf{Z}\}$
5. Prove that if p and q are distinct primes, then there exist integers m and n such that $pm + qn = 1$.

6. Show that $n^2 - n + 5$ is a prime integer when $n = 1, 2, 3, 4$ but that it is not true that $n^2 - n + 5$ is always a prime integer. Write out a similar set of statements for the polynomial $n^2 - n + 11$.
7. If $a > 0$ and $a|b$, then prove or disprove that $(a, b) = a$.
8. Let a, b , and c be integers such that $a \neq 0$. Prove that if $a|bc$, then $a|c \cdot (a, b)$.
9. Let a be a nonzero integer and b a positive integer. Prove or disprove that $(a, b) = (a, a + b)$.
10. Let $a|c$ and $b|c$, and $(a, b) = 1$, prove that ab divides c .
11. Prove that if $d = (a, b)$, $a|c$, and $b|c$, then $ab|cd$.
12. If $b > 0$ and $a = bq + r$, prove that $(a, b) = (b, r)$.
13. Let $r_0 = b > 0$. With the notation used in the description of the Euclidean Algorithm, use the result in Exercise 12 to prove that $(a, b) = r_n$, the last nonzero remainder.
14. Prove that every remainder r_j in the Euclidean Algorithm is a “linear combination” of a and b : $r_j = s_ja + t_jb$, for integers s_j and t_j .
15. Let a and b be integers, at least one of them not 0. Prove that an integer c can be expressed as a linear combination of a and b if and only if $(a, b)|c$.
16. Prove Corollary 2.17: If p is a prime and $p|(a_1a_2\cdots a_n)$, then p divides some a_j . (Hint: Use induction on n .)
17. Prove that if n is a positive integer greater than 1 such that n is not a prime, then n has a divisor d such that $1 < d \leq \sqrt{n}$.
18. Prove that $(ab, c) = 1$ if and only if $(a, c) = 1$ and $(b, c) = 1$.
19. Let $(a, b) = 1$ and $(a, c) = 1$. Prove or disprove that $(ac, b) = 1$.
20. Let $(a, b) = 1$. Prove $(a, bc) = (a, c)$, where c is any integer.
21. Let $(a, b) = 1$. Prove $(a^2, b^2) = 1$.
22. Let $(a, b) = 1$. Prove that $(a, b^n) = 1$ for all positive integers n .
23. Prove that if $m > 0$ and (a, b) exists, then $(ma, mb) = m \cdot (a, b)$.
24. Prove that if $d = (a, b)$, $a = a_0d$, and $b = b_0d$, then $(a_0, b_0) = 1$.
25. A *least common multiple* of two nonzero integers a and b is an integer m that satisfies all the following conditions:
 1. m is a positive integer.
 2. $a|m$ and $b|m$.
 3. $a|c$ and $b|c$ imply $m|c$.

Sec. 2.5, #28 <

26. Let a and b be positive integers. If $d = (a, b)$ and m is the least common multiple of a and b , prove that $dm = ab$. Note that it follows that the least common multiple of two positive relatively prime integers is their product.

Sec. 8.2, #35 <

Prove that the least common multiple of two nonzero integers exists and is unique.

27. Let a and b be positive integers. Prove that if $d = (a, b)$, $a = a_0d$, and $b = b_0d$, then the least common multiple of a and b is a_0b_0d .
28. Describe a procedure for using the standard forms of two positive integers to find their least common multiple.
29. For each pair of integers a, b in Exercise 2, find the least common multiple of a and b by using their standard forms.
30. Let a, b , and c be three nonzero integers.
- Use Definition 2.11 as a pattern to define a greatest common divisor of a, b , and c .
 - Use Theorem 2.12 and its proof as a pattern to prove the existence of a greatest common divisor of a, b , and c .
 - If d is the greatest common divisor of a, b , and c , show that $d = ((a, b), c)$.
 - Prove $((a, b), c) = (a, (b, c))$.
31. Find the greatest common divisor of a, b , and c and write it in the form $ax + by + cz$ for integers x, y , and z .
- $a = 14, b = 28, c = 35$
 - $a = 26, b = 52, c = 60$
 - $a = 143, b = 385, c = -65$
 - $a = 60, b = -84, c = 105$
32. Use the Second Principle of Finite Induction to prove that every positive integer n can be expressed in the form
- $$n = c_0 + c_1 \cdot 3 + c_2 \cdot 3^2 + \cdots + c_{j-1} \cdot 3^{j-1} + c_j \cdot 3^j,$$
- where j is a nonnegative integer, $c_i \in \{0, 1, 2\}$ for all $i < j$, and $c_j \in \{1, 2\}$.
33. Use the fact that 2 is a prime to prove that there do not exist nonzero integers a and b such that $a^2 = 2b^2$. Explain how this proves that $\sqrt{2}$ is not a rational number.
34. Use the fact that 3 is a prime to prove that there do not exist nonzero integers a and b such that $a^2 = 3b^2$. Explain how this proves that $\sqrt{3}$ is not a rational number.

2.5

Congruence of Integers

In Example 4 of Section 1.7, we defined the relation “congruence modulo 4” on the set \mathbf{Z} of all integers, and we proved this relation to be an equivalence relation on \mathbf{Z} . That example is a special case of **congruence modulo n** , as defined next.

Definition 2.20 ■ Congruence Modulo n

Let n be a positive integer, $n > 1$. For integers x and y , x is **congruent to y modulo n** if and only if $x - y$ is a multiple of n . We write

$$x \equiv y \pmod{n}$$

to indicate that x is congruent to y modulo n .

Thus $x \equiv y \pmod{n}$ if and only if n divides $x - y$, and this is equivalent to $x - y = nq$, or $x = y + nq$. Another way to describe this relation is to say that x and y yield the same remainder when each is divided by n . To see that this is true, let

$$x = nq_1 + r_1 \quad \text{with} \quad 0 \leq r_1 < n$$

and

$$y = nq_2 + r_2 \quad \text{with} \quad 0 \leq r_2 < n.$$

Then

$$x - y = n(q_1 - q_2) + (r_1 - r_2) \quad \text{with} \quad 0 \leq |r_1 - r_2| < n.$$

Thus $x - y$ is a multiple of n if and only if $r_1 - r_2 = 0$ —that is, if and only if $r_1 = r_2$. In particular, any integer x is congruent to its remainder when divided by n . This means that any x is congruent to one of

$$0, 1, 2, \dots, n - 1.$$

Congruence modulo n is an equivalence relation on \mathbf{Z} , and this fact is important enough to be stated as a theorem.

Theorem 2.21 ■ Equivalence Relation

The relation of congruence modulo n is an equivalence relation on \mathbf{Z} .

Proof We shall show that congruence modulo n is (1) reflexive, (2) symmetric, and (3) transitive. Let $n > 1$, and let x, y , and z be arbitrary in \mathbf{Z} .

Reflexive 1. $x \equiv x \pmod{n}$ since $x - x = (n)(0)$.

Symmetric 2. $x \equiv y \pmod{n} \Rightarrow x - y = nq$ for some $q \in \mathbf{Z}$
 $\Rightarrow y - x = n(-q)$ and $-q \in \mathbf{Z}$
 $\Rightarrow y \equiv x \pmod{n}$.

Transitive 3. $x \equiv y \pmod{n}$ and $y \equiv z \pmod{n}$
 $\Rightarrow x - y = nq$ and $y - z = nk$ and $q, k \in \mathbf{Z}$
 $\Rightarrow x - z = x - y + y - z$
 $= n(q + k)$, and $q + k \in \mathbf{Z}$
 $\Rightarrow x \equiv z \pmod{n}$.

As with any equivalence relation, the equivalence classes for congruence modulo n form a *partition* of \mathbf{Z} ; that is, they separate \mathbf{Z} into mutually disjoint subsets. These subsets are called **congruence classes** or **residue classes**. Referring to our discussion concerning

remainders, we see that there are n distinct congruence classes modulo n , given by

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

When $n = 4$, these classes appear as

$$\begin{aligned}[0] &= \{\dots, -8, -4, 0, 4, 8, \dots\} \\ [1] &= \{\dots, -7, -3, 1, 5, 9, \dots\} \\ [2] &= \{\dots, -6, -2, 2, 6, 10, \dots\} \\ [3] &= \{\dots, -5, -1, 3, 7, 11, \dots\}.\end{aligned}$$

Congruence classes are useful in connection with numerous examples, and we shall see more of them later.

Although $x \equiv y \pmod{n}$ is certainly not an equation, in many ways congruences can be handled in the same fashion as equations. The next theorem asserts that the same integer can be added to both members and that both members can be multiplied by the same integer.

Theorem 2.22 ■ Addition and Multiplication Properties

If $a \equiv b \pmod{n}$ and x is any integer, then

$$a + x \equiv b + x \pmod{n} \quad \text{and} \quad ax \equiv bx \pmod{n}.$$

$p \Rightarrow q$ **Proof** Let $a \equiv b \pmod{n}$ and $x \in \mathbf{Z}$. We shall prove that $ax \equiv bx \pmod{n}$ and leave the other part as an exercise. We have

$$\begin{aligned}a \equiv b \pmod{n} &\Rightarrow a - b = nq \quad \text{for } q \in \mathbf{Z} \\ &\Rightarrow (a - b)x = (nq)x \quad \text{for } q, x \in \mathbf{Z} \\ &\Rightarrow ax - bx = n(qx) \quad \text{for } qx \in \mathbf{Z} \\ &\Rightarrow ax \equiv bx \pmod{n}.\end{aligned}$$

Congruence modulo n also has substitution properties that are analogous to those possessed by equality. Suppose we wish to compute the product $(25)(17) \pmod{6}$. Since $25 \equiv 1 \pmod{6}$ and $17 \equiv 5 \pmod{6}$, the following theorem allows us to compute the product $(25)(17) \pmod{6}$ as $(1)(5) \equiv 5 \pmod{6}$ instead of $(25)(17) \equiv 425 \pmod{6} \equiv 5 \pmod{6}$.

Theorem 2.23 ■ Substitution Properties

Suppose $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$. Then

$$a + c \equiv b + d \pmod{n} \quad \text{and} \quad ac \equiv bd \pmod{n}.$$

$(p \wedge q) \Rightarrow r$ **Proof** Let $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$. By Theorem 2.22,

$$a \equiv b \pmod{n} \Rightarrow ac \equiv bc \pmod{n}$$

and

$$c \equiv d \pmod{n} \Rightarrow bc \equiv bd \pmod{n}.$$

But $ac \equiv bc \pmod{n}$ and $bc \equiv bd \pmod{n}$ imply $ac \equiv bd \pmod{n}$, by the transitive property.

The proof that $a + c \equiv b + d \pmod{n}$ is left as an exercise.

Example 1 Since exponentiation is just repeated multiplication, Theorem 2.23 can be used to evaluate powers modulo n . Consider $58^{23} \pmod{9}$. Since

$$58 \equiv 4 \pmod{9},$$

then by Theorem 2.23,

$$58^{23} \equiv 4^{23} \pmod{9}.$$

Also since

$$4^{23} = 4^2 \cdot (4^3)^7,$$

then

$$\begin{aligned} 58^{23} &\equiv 4^{23} \pmod{9} \\ &\equiv 4^2 \cdot (4^3)^7 \pmod{9} \\ &\equiv (16)(64)^7 \pmod{9} \\ &\equiv (7)(1)^7 \pmod{9} \\ &\equiv 7 \pmod{9}. \end{aligned}$$
■

It is easy to show that there is a “cancellation law” for addition that holds for congruences: $a + x \equiv a + y \pmod{n}$ implies $x \equiv y \pmod{n}$. This is not the case, however, with multiplication:

$$ax \equiv ay \pmod{n} \quad \text{and} \quad a \not\equiv 0 \pmod{n} \quad \text{do not imply} \quad x \equiv y \pmod{n}.$$

As an example,

$$(4)(6) \equiv (4)(21) \pmod{30} \quad \text{but} \quad 6 \not\equiv 21 \pmod{30}.$$

It is important to note here that $a = 4$ and $n = 30$ are not relatively prime. When the condition that a and n be relatively prime is imposed, we can obtain a cancellation law for multiplication.

Theorem 2.24 ■ Cancellation Law

If $ax \equiv ay \pmod{n}$ and $(a, n) = 1$, then

$$x \equiv y \pmod{n}.$$

$(p \wedge q) \Rightarrow r$ **Proof** Assume that $ax \equiv ay \pmod{n}$ and that a and n are relatively prime.

$$\begin{aligned} ax \equiv ay \pmod{n} &\Rightarrow n | (ax - ay) \\ &\Rightarrow n | a(x - y) \\ &\Rightarrow n | (x - y) \quad \text{by Theorem 2.14} \\ &\Rightarrow x \equiv y \pmod{n} \end{aligned}$$

This completes the proof.

We have seen that there are analogues for many of the manipulations that may be performed with equations. There are also techniques for obtaining solutions to congruence equations of certain types. The basic technique makes use of Theorem 2.23 and the Euclidean Algorithm. The use of the Euclidean Algorithm is illustrated in Example 2.

Theorem 2.25 ■ Linear Congruences

If a and n are relatively prime, the congruence $ax \equiv b \pmod{n}$ has a solution x in the integers, and any two solutions in \mathbf{Z} are congruent modulo n .

$p \Rightarrow q$ **Proof** Since a and n are relatively prime, there exist integers s and t such that

$$\begin{aligned} 1 &= as + nt \\ \Rightarrow b &= asb + nt \\ \Rightarrow a(sb) - b &= n(-tb) \\ \Rightarrow n | [a(sb) - b] & \\ \Rightarrow a(sb) &\equiv b \pmod{n}. \end{aligned}$$

Thus $x = sb$ is a solution to $ax \equiv b \pmod{n}$.

Uniqueness To complete the proof, suppose that both x and y are integers that are solutions to $ax \equiv b \pmod{n}$. Then we have

$$ax \equiv b \pmod{n} \quad \text{and} \quad ay \equiv b \pmod{n}.$$

Using the symmetric and transitive properties of congruence modulo n , we conclude that these relations imply

$$ax \equiv ay \pmod{n}.$$

Since $(a, n) = 1$, this requires that $x \equiv y \pmod{n}$, by Theorem 2.24. Hence any two solutions in \mathbf{Z} are congruent modulo n .

Strategy ■ We note that the “uniqueness” part of the proof of the theorem requires showing not that any two solutions to the system are “equal” but rather that they are congruent modulo n . This same type of proof is also used in Theorem 2.26.

Example 2 When $(a, n) = 1$, the Euclidean Algorithm can be used to find a solution x to $ax \equiv b \pmod{n}$. Consider the congruence

$$20x \equiv 14 \pmod{63}.$$

We first obtain s and t such that

$$1 = 20s + 63t.$$

Applying the Euclidean Algorithm, we have

$$\begin{aligned} 63 &= (20)(3) + 3 \\ 20 &= (3)(6) + 2 \\ 3 &= (2)(1) + 1 \\ 2 &= (1)(2). \end{aligned}$$

Solving for the nonzero remainders,

$$\begin{aligned} 3 &= 63 - (20)(3) \\ 2 &= 20 - (3)(6) \\ 1 &= 3 - (2)(1). \end{aligned}$$

Substituting the remainders in turn, we obtain

$$\begin{aligned} 1 &= 3 - (2)(1) \\ &= 3 - [20 - (3)(6)](1) \\ &= (3)(7) + (20)(-1) \\ &= [63 - (20)(3)](7) + (20)(-1) \\ &= (20)(-22) + (63)(7). \end{aligned}$$

Multiplying this equation by $b = 14$, we have

$$\begin{aligned} 14 &= (20)(-308) + (63)(98) \\ \Rightarrow 14 &\equiv (20)(-308) \pmod{63}. \end{aligned}$$

Thus $x = -308$ is a solution. However, any number is congruent modulo 63 to its remainder when divided by 63, and

$$-308 = (63)(-5) + 7.$$

Thus $x = 7$ is also a solution, one that is in the range $0 \leq x < 63$. ■

The preceding example illustrates the basic technique for obtaining a solution to $ax \equiv b \pmod{n}$ when a and n are relatively prime, but other methods are also very useful. Some of them make use of Theorems 2.23 and 2.24. Theorem 2.24 can be used to remove a factor c from both sides of the congruence, provided c and n are relatively prime. That is, c may be canceled from $crx \equiv ct \pmod{n}$ to obtain the equivalent congruence $rx \equiv t \pmod{n}$.

Example 3 Since 2 and 63 are relatively prime, the factor 2 in both sides of

$$20x \equiv 14 \pmod{63}$$

can be removed to obtain

$$10x \equiv 7 \pmod{63}.$$

Theorem 2.21 allows us to replace an integer by any other integer that is congruent to it modulo n . Now $7 \equiv 70 \pmod{63}$, and this substitution yields

$$10x \equiv 70 \pmod{63}.$$

Removing the factor 10 from both sides, we have

$$x \equiv 7 \pmod{63}.$$

Thus we have obtained the solution $x = 7$ much more easily than by the method of Example 1. However, this method is less systematic, and it requires more ingenuity. ■

Some systems of congruences can be solved using the result of the next theorem.

Theorem 2.26 ■ System of Congruences

Let m and n be relatively prime and a and b integers. There exists an integer x that satisfies the system of congruences

$$x \equiv a \pmod{m}$$

$$x \equiv b \pmod{n}.$$

Furthermore, any two solutions x and y are congruent modulo mn .

$p \Rightarrow q$ **Proof** Suppose $(m, n) = 1$. Let x be a solution to the first congruence in the system

$$x \equiv a \pmod{m}$$

$$x \equiv b \pmod{n}.$$

Thus $x = a + mk$ for some integer k , and this k must be such that

$$a + mk \equiv b \pmod{n}$$

or

$$mk \equiv b - a \pmod{n}.$$

Since $(m, n) = 1$, Theorem 2.25 guarantees the existence of such an integer k , and $x = a + mk$ satisfies the system.

Uniqueness

Now let y be another solution to the system of congruences; that is,

$$y \equiv a \pmod{m}$$

$$y \equiv b \pmod{n}.$$

By Theorem 2.21,

$$x \equiv y \pmod{m}$$

$$x \equiv y \pmod{n}$$

and

$$m|x - y \text{ and } n|x - y.$$

Then

$$mn|x - y$$

by Exercise 10 of Section 2.4. So $x \equiv y \pmod{mn}$.

Example 4 Since $(7, 5) = 1$, we use Theorem 2.26 to solve the system of congruences

$$x \equiv 5 \pmod{7}$$

$$x \equiv 3 \pmod{5}.$$

From the first congruence we write $x = 5 + 7k$ for some integer k and substitute this expression for x into the second congruence.

$$5 + 7k \equiv 3 \pmod{5}$$

or

$$7k \equiv -2 \pmod{5}$$

$$\Rightarrow 2k \equiv -2 \pmod{5}$$

$$\Rightarrow k \equiv -1 \pmod{5} \quad \text{since } (2, 5) = 1$$

$$\Rightarrow k \equiv 4 \pmod{5}.$$

Thus $x = 5 + 7(4) = 33$ satisfies the system and $x \equiv 33 \pmod{7 \cdot 5}$ or $x \equiv 33 \pmod{35}$ gives all solutions to the system of congruences. ■

An extension of Theorem 2.26 is the Chinese Remainder Theorem. In this theorem, we use the term “pairwise relatively prime” to mean that every pairing of integers n_i and n_j for all $i \neq j$ are relatively prime.

Theorem 2.27 ■ Chinese Remainder Theorem

Let n_1, n_2, \dots, n_m be pairwise relatively prime. There exists an integer x that satisfies the system of congruences

$$x \equiv a_1 \pmod{n_1}$$

$$x \equiv a_2 \pmod{n_2}$$

⋮

$$x \equiv a_m \pmod{n_m}.$$

Furthermore, any two solutions x and y are congruent modulo $n_1 n_2 \cdots n_m$.

The proof of the Chinese Remainder Theorem is requested in the exercises and we illustrate the technique in the next example.

Example 5 Consider the system of congruences

$$\begin{aligned}x &\equiv 5 \pmod{7} \\x &\equiv 3 \pmod{5} \\x &\equiv 2 \pmod{8} \\x &\equiv 2 \pmod{3}.\end{aligned}$$

Example 4 showed that $x \equiv 33 \pmod{35}$ is a solution to the first 2 congruences. Pairing this congruence with the third $x \equiv 2 \pmod{8}$ in the system gives

$$\begin{aligned}x &\equiv 33 \pmod{35} \\x &\equiv 2 \pmod{8}.\end{aligned}$$

So with $x = 33 + 35k$ for some $k \in \mathbf{Z}$ gives

$$\begin{aligned}33 + 35k &\equiv 2 \pmod{8} \\ \Rightarrow 35k &\equiv -31 \pmod{8} \\ \Rightarrow 3k &\equiv 1 \pmod{8} \\ \Rightarrow k &\equiv 3 \pmod{8} \\ \Rightarrow x &= 33 + 35 \cdot 3 \\ &= 138.\end{aligned}$$

Thus $x \equiv 138 \pmod{280}$ satisfies the first three congruences of the system. Pairing this with the last $x \equiv 2 \pmod{3}$ gives the system

$$\begin{aligned}x &\equiv 138 \pmod{280} \\x &\equiv 2 \pmod{3}.\end{aligned}$$

Setting $x = 138 + 280k$ for some integer k in the second congruence of the system gives

$$\begin{aligned}138 + 280k &\equiv 2 \pmod{3} \\ \Rightarrow 280k &\equiv -136 \pmod{3} \\ \Rightarrow k &\equiv 2 \pmod{3} \\ \Rightarrow x &= 138 + 280 \cdot 2 \\ &= 698.\end{aligned}$$

Thus $x \equiv 698 \pmod{280 \cdot 3} \equiv 698 \pmod{840}$ satisfies the original system. ■

Exercises 2.5

True or False

Label each of the following statements as either true or false.

1. $a \equiv b \pmod{n}$ implies $ac \equiv bc \pmod{nc}$ for $c \in \mathbf{Z}^+$.
2. $a \equiv b \pmod{n}$ and $c|n$ implies $a \equiv b \pmod{c}$ for $c \in \mathbf{Z}^+$.

3. $a^2 \equiv b^2 \pmod{n}$ implies $a \equiv b \pmod{n}$ or $a \equiv -b \pmod{n}$.
 4. a is congruent to b modulo n if and only if a and b yield the same remainder when each is divided by n .
 5. The congruence classes for congruence modulo n form a partition of \mathbf{Z} .
 6. If $ab \equiv 0 \pmod{n}$, then either $a \equiv 0 \pmod{n}$ or $b \equiv 0 \pmod{n}$.
 7. If $(a, n) = 1$, then $a \equiv 1 \pmod{n}$.
-

Exercises

In this exercise set, all variables are integers.

1. List the distinct congruence classes modulo 5, exhibiting at least three elements in each class.
2. Follow the instructions in Exercise 1 for the congruence classes modulo 6.

Find a solution $x \in \mathbf{Z}$, $0 \leq x < n$, for each of the congruences $ax \equiv b \pmod{n}$ in Exercises 3–24. Note that in each case, a and n are relatively prime.

- | | |
|---|-----------------------------------|
| 3. $2x \equiv 3 \pmod{7}$ | 4. $2x \equiv 3 \pmod{5}$ |
| 5. $3x \equiv 7 \pmod{13}$ | 6. $3x \equiv 4 \pmod{13}$ |
| 7. $8x \equiv 1 \pmod{21}$ | 8. $14x \equiv 8 \pmod{15}$ |
| 9. $11x \equiv 1 \pmod{317}$ | 10. $11x \equiv 3 \pmod{138}$ |
| 11. $8x \equiv 66 \pmod{79}$ | 12. $6x \equiv 14 \pmod{55}$ |
| 13. $8x + 3 \equiv 5 \pmod{9}$ | 14. $19x + 7 \equiv 27 \pmod{18}$ |
| 15. $13x + 19 \equiv 2 \pmod{23}$ | 16. $5x + 43 \equiv 15 \pmod{22}$ |
| 17. $25x \equiv 31 \pmod{7}$ | 18. $358x \equiv 17 \pmod{313}$ |
| 19. $55x \equiv 59 \pmod{42}$ | 20. $79x \equiv 83 \pmod{61}$ |
| 21. $92x + 17 \equiv 29 \pmod{37}$ | 22. $57x + 7 \equiv 78 \pmod{53}$ |
| 23. $35x + 14 \equiv 3 \pmod{27}$ | 24. $82x + 23 \equiv 2 \pmod{47}$ |
| 25. Complete the proof of Theorem 2.22: If $a \equiv b \pmod{n}$ and x is any integer, then $a + x \equiv b + x \pmod{n}$. | |
| 26. Complete the proof of Theorem 2.23: If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a + c \equiv b + d \pmod{n}$. | |
| 27. Prove that if $a + x \equiv a + y \pmod{n}$, then $x \equiv y \pmod{n}$. | |
| Sec. 2.4, #24 ➤ 28. If $ca \equiv cb \pmod{n}$ and $d = (c, n)$ where $n = dm$, prove that $a \equiv b \pmod{m}$. | |
| 29. Find the least positive integer that is congruent to the given sum, product, or power. | |
| a. $(3 + 19 + 23 + 52) \pmod{6}$ | b. $(2 + 17 + 43 + 117) \pmod{4}$ |
| c. $(14 + 46 + 65 + 92) \pmod{11}$ | d. $(9 + 25 + 38 + 92) \pmod{7}$ |
| e. $(7)(17)(32)(62) \pmod{5}$ | f. $(6)(16)(38)(118) \pmod{9}$ |

g. $(4)(9)(15)(59) \pmod{7}$

i. $43^{15} \pmod{4}$

k. $62^{33} \pmod{5}$

h. $(5)(11)(17)(65) \pmod{7}$

j. $25^{38} \pmod{7}$

l. $52^{26} \pmod{9}$

30. If $a \equiv b \pmod{n}$, prove that $a^m \equiv b^m \pmod{n}$ for every positive integer m .
31. Prove that if m is an integer, then either $m^2 \equiv 0 \pmod{4}$ or $m^2 \equiv 1 \pmod{4}$. (*Hint:* Consider the cases where m is even and where m is odd.)
32. Prove or disprove that if n is odd, then $n^2 \equiv 1 \pmod{8}$.
33. If m is an integer, show that m^2 is congruent modulo 8 to one of the integers 0, 1, or 4. (*Hint:* Use the Division Algorithm, and consider the possible remainders in $m = 4q + r$.)
34. Prove that $n^3 \equiv n \pmod{6}$ for every positive integer n .
35. Let x and y be integers. Prove that if there is an equivalence class $[a]$ modulo n such that $x \in [a]$ and $y \in [a]$, then $(x, n) = (y, n)$.
36. Prove that if p is a prime and $c \not\equiv 0 \pmod{p}$, then $cx \equiv b \pmod{p}$ has a unique solution modulo p . That is, a solution exists, and any two solutions are congruent modulo p .
37. Let $d = (a, n)$ where $n > 1$. Prove that if there is a solution to $ax \equiv b \pmod{n}$, then d divides b .
38. (See Exercise 37.) Suppose that $n > 1$ and that $d = (a, n)$ is a divisor of b . Let $a = a_0d$, $n = n_0d$, and $b = b_0d$, where a_0 , n_0 , and b_0 are integers. The following statements **a–e** lead to a proof that the congruence $ax \equiv b \pmod{n}$ has exactly d incongruent solutions modulo n , and they show how such a set of solutions can be found.
- Prove that $ax \equiv b \pmod{n}$ if and only if $a_0x \equiv b_0 \pmod{n_0}$.
 - Prove that if x_1 and x_2 are any two solutions to $a_0x \equiv b_0 \pmod{n_0}$, then it follows that $x_1 \equiv x_2 \pmod{n_0}$.
 - Let x_1 be a fixed solution to $a_0x \equiv b_0 \pmod{n_0}$, and prove that each of the d integers in the list

$$x_1, x_1 + n_0, x_1 + 2n_0, \dots, x_1 + (d - 1)n_0$$

is a solution to $ax \equiv b \pmod{n}$.

- Prove that no two of the solutions listed in part c are congruent modulo n .
- Prove that any solution to $ax \equiv b \pmod{n}$ is congruent to one of the numbers listed in part c.

In the congruences $ax \equiv b \pmod{n}$ in Exercises 39–50, a and n may not be relatively prime. Use the results in Exercises 37 and 38 to determine whether there are solutions. If there are, find d incongruent solutions modulo n .

39. $6x \equiv 33 \pmod{27}$

41. $8x \equiv 66 \pmod{78}$

43. $68x \equiv 36 \pmod{40}$

45. $24x + 5 \equiv 50 \pmod{348}$

40. $18x \equiv 33 \pmod{15}$

42. $35x \equiv 10 \pmod{20}$

44. $21x \equiv 18 \pmod{30}$

46. $36x + 1 \equiv 49 \pmod{270}$

47. $15x + 23 \equiv 153 \pmod{110}$

48. $20x + 13 \equiv 137 \pmod{76}$

49. $42x + 67 \equiv 23 \pmod{74}$

50. $38x + 54 \equiv 20 \pmod{60}$

Sec. 4.4, #20 <

Sec. 8.3, #11 <

51. Let p be a prime integer. Prove **Fermat's[†] Little Theorem**: For any positive integer a , $a^p \equiv a \pmod{p}$. (*Hint:* Use induction on a , with p held fixed.)

52. Prove the Chinese Remainder Theorem: Let n_1, n_2, \dots, n_m be pairwise relatively prime. There exists an integer x that satisfies the system of congruences

$$x \equiv a_1 \pmod{n_1}$$

$$x \equiv a_2 \pmod{n_2}$$

 \vdots

$$x \equiv a_m \pmod{n_m}.$$

Furthermore, any two solutions x and y are congruent modulo $n_1 n_2 \cdots n_m$.

53. Solve the following systems of congruences.

a. $x \equiv 2 \pmod{5}$

$x \equiv 3 \pmod{8}$

b. $x \equiv 4 \pmod{5}$

$x \equiv 2 \pmod{3}$

c. $x \equiv 4 \pmod{7}$

$3x + 2 \equiv 3 \pmod{8}$

d. $2x \equiv 5 \pmod{3}$

$5x + 4 \equiv 5 \pmod{7}$

e. $x \equiv 4 \pmod{5}$

$x \equiv 6 \pmod{8}$

$x \equiv 2 \pmod{3}$

f. $x \equiv 3 \pmod{4}$

$x \equiv 4 \pmod{5}$

$x \equiv 6 \pmod{7}$

g. $x \equiv 2 \pmod{3}$

$x \equiv 2 \pmod{5}$

$x \equiv 4 \pmod{7}$

$x \equiv 3 \pmod{8}$

h. $x \equiv 3 \pmod{5}$

$x \equiv 7 \pmod{8}$

$x \equiv 3 \pmod{9}$

$x \equiv 10 \pmod{11}$

54. a. Prove that $10^n \equiv 1 \pmod{9}$ for every positive integer n .

- b. Prove that a positive integer is divisible by 9 if and only if the sum of its digits is divisible by 9. (*Hint:* Any integer can be expressed in the form

$$a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_1 10 + a_0$$

where each a_i is one of the digits 0, 1, ..., 9.)

55. a. Prove that $10^n \equiv (-1)^n \pmod{11}$ for every positive integer n .

- b. Prove that a positive integer z is divisible by 11 if and only if 11 divides $a_0 - a_1 + a_2 - \cdots + (-1)^n a_n$, when z is written in the form as described in the previous problem.

[†]Pierre de Fermat (1601–1665) a French mathematician, is credited for work that led to modern calculus. He is most widely known for his famous Last Theorem: $x^n + y^n = z^n$ has no nonzero integral solutions for x, y , and z when $n > 2$. This unproven theorem was found by his son with a note by Fermat stating, “I have a truly marvelous demonstration of this proposition which this margin is too small to contain.” After many failed attempts by numerous mathematicians, a proof by Andrew Wiles and Richard Taylor was finally accepted as valid over 350 years later using techniques unknown to Fermat.

2.6 Congruence Classes

In connection with the relation of congruence modulo n , we have observed that there are n distinct congruence classes. Let \mathbf{Z}_n denote this set of classes:

$$\mathbf{Z}_n = \{[0], [1], [2], \dots, [n - 1]\}.$$

When addition and multiplication are defined in a natural and appropriate manner in \mathbf{Z}_n , these sets provide useful examples for our work in later chapters.

Theorem 2.28 ■ Addition in \mathbf{Z}_n

Consider the rule given by

$$[a] + [b] = [a + b].$$

- a. This rule defines an addition that is a binary operation on \mathbf{Z}_n .
- b. Addition is associative in \mathbf{Z}_n :

$$[a] + ([b] + [c]) = ([a] + [b]) + [c].$$

- c. Addition is commutative in \mathbf{Z}_n :

$$[a] + [b] = [b] + [a].$$

- d. \mathbf{Z}_n has the additive identity $[0]$.
- e. Each $[a]$ in \mathbf{Z}_n has $[-a]$ as its additive inverse in \mathbf{Z}_n .

Proof

- a. It is clear that the rule $[a] + [b] = [a + b]$ yields an element of \mathbf{Z}_n , but the uniqueness of this result needs to be verified. In other words, closure is obvious, but we need to show that the operation is well-defined. To do this, suppose that $[a] = [x]$ and $[b] = [y]$. Then

$$[a] = [x] \Rightarrow a \equiv x \pmod{n}$$

and

$$[b] = [y] \Rightarrow b \equiv y \pmod{n}.$$

By Theorem 2.23,

$$a + b \equiv x + y \pmod{n},$$

and therefore $[a + b] = [x + y]$.

- b. The associative property follows from

$$\begin{aligned} [a] + ([b] + [c]) &= [a] + [b + c] \\ &= [a + (b + c)] \\ &= [(a + b) + c] \\ &= [a + b] + [c] \\ &= ([a] + [b]) + [c]. \end{aligned}$$

Note that the key step here is the fact that addition is associative in \mathbf{Z} :

$$a + (b + c) = (a + b) + c.$$

- c. The commutative property follows from

$$\begin{aligned}[a] + [b] &= [a + b] \\ &= [b + a] \\ &= [b] + [a].\end{aligned}$$

- d. $[0]$ is the additive identity, since addition is commutative in \mathbf{Z}_n and

$$[a] + [0] = [a + 0] = [a].$$

- e. $[-a] = [n - a]$ is the additive inverse of $[a]$, since addition is commutative in \mathbf{Z}_n and

$$[-a] + [a] = [-a + a] = [0].$$

Example 1 Following the procedure described in Exercise 3 of Section 1.4, we can construct an addition table for $\mathbf{Z}_4 = \{[0], [1], [2], [3]\}$. In computing the entries for this table, $[a] + [b]$ is entered in the row with $[a]$ at the left and in the column with $[b]$ at the top. For instance,

$$[3] + [2] = [5] = [1]$$

is entered in the row with $[3]$ at the left and in the column with $[2]$ at the top. The complete addition table is shown in Figure 2.1.

+	[0]	[1]	[2]	[3]
[0]	[0]	[1]	[2]	[3]
[1]	[1]	[2]	[3]	[0]
[2]	[2]	[3]	[0]	[1]
[3]	[3]	[0]	[1]	[2]

■ Figure 2.1



In the following theorem, multiplication in \mathbf{Z}_n is defined in a natural way, and the basic properties for this operation are stated. The proofs of the various parts of the theorem are quite similar to those for the corresponding parts of Theorem 2.28, and are left as exercises.

Theorem 2.29 ■ Multiplication in \mathbf{Z}_n

Consider the rule for multiplication in \mathbf{Z}_n given by

$$[a][b] = [ab].$$

- a. Multiplication as defined by this rule is a binary operation on \mathbf{Z}_n .

- b. Multiplication is associative in \mathbf{Z}_n :

$$[a]([b][c]) = ([a][b])[c].$$

c. Multiplication is commutative in \mathbf{Z}_n :

$$[a][b] = [b][a].$$

d. \mathbf{Z}_n has the multiplicative identity [1].

When we compare the properties listed in Theorems 2.28 and 2.29, we see that the existence of multiplicative inverses, even for the nonzero elements, is conspicuously missing. The following example shows that this is appropriate because it illustrates a case where some of the nonzero elements of \mathbf{Z}_n do not have multiplicative inverses.

Example 2 A multiplication table for \mathbf{Z}_4 is shown in Figure 2.2. The third row of the table shows that [2] is a nonzero element of \mathbf{Z}_4 that has no multiplicative inverse; there is no $[x]$ in \mathbf{Z}_4 such that $[2][x] = [1]$. Another interesting point in connection with

\times	[0]	[1]	[2]	[3]
[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]
[2]	[0]	[2]	[0]	[2]
[3]	[0]	[3]	[2]	[1]

■ **Figure 2.2**

this table is that the equality $[2][2] = [0]$ shows that in \mathbf{Z}_4 , the product of nonzero factors may be zero. ■

Any nonzero element $[a]$ in \mathbf{Z}_n for which the equation $[a][x] = [0]$ has a nonzero solution $[x] \neq [0]$ in \mathbf{Z}_n is a **zero divisor**. The element [2] in \mathbf{Z}_4 is an example of a zero divisor.

The next theorem characterizes those elements of \mathbf{Z}_n that have multiplicative inverses.

Theorem 2.30 ■ Multiplicative Inverses in \mathbf{Z}_n

An element $[a]$ of \mathbf{Z}_n has a multiplicative inverse in \mathbf{Z}_n if and only if a and n are relatively prime.

$p \Rightarrow q$ **Proof** Suppose first that $[a]$ has a multiplicative inverse $[b]$ in \mathbf{Z}_n . Then

$$[a][b] = [1].$$

This means that

$$[ab] = [1] \quad \text{and} \quad ab \equiv 1 \pmod{n}.$$

Therefore,

$$ab - 1 = nq$$

for some integer q , and

$$a(b) + n(-q) = 1.$$

By Theorem 2.12, we have $(a, n) = 1$.

$p \Leftrightarrow q$ Conversely, if $(a, n) = 1$, then Theorem 2.25 guarantees the existence of a solution s to the congruence

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

Thus,

$$[a][s] = [1],$$

and $[a]$ has a multiplicative inverse $[s]$ in \mathbf{Z}_n .

Corollary 2.31 ■ Multiplicative Inverses in \mathbf{Z}_p

Every nonzero element of \mathbf{Z}_n has a multiplicative inverse if and only if n is a prime.

$p \Leftrightarrow q$ **Proof** The corollary follows from the fact that n is a prime if and only if every integer a such that $1 \leq a < n$ is relatively prime to n .

Example 3 The elements of \mathbf{Z}_{15} that have multiplicative inverses can be listed by writing down those $[a]$ that are such that $(a, 15) = 1$. These elements are

$$[1], [2], [4], [7], [8], [11], [13], [14].$$

■

Example 4 Suppose we wish to find the multiplicative inverse of $[13]$ in \mathbf{Z}_{191} . The modulus $n = 191$ is so large that it is not practical to test all of the elements in \mathbf{Z}_{191} , so we utilize the Euclidean Algorithm and proceed according to the last part of the proof of Theorem 2.30:

$$\begin{aligned} 191 &= (13)(14) + 9 \\ 13 &= (9)(1) + 4 \\ 9 &= (4)(2) + 1. \end{aligned}$$

Substituting the remainders in turn, we have

$$\begin{aligned} 1 &= 9 - (4)(2) \\ &= 9 - [13 - (9)(1)](2) \\ &= (9)(3) - (13)(2) \\ &= [191 - (13)(14)](3) - (13)(2) \\ &= (191)(3) + (13)(-44). \end{aligned}$$

Thus

$$(13)(-44) \equiv 1 \pmod{191}$$

or

$$[13][-44] = [1].$$

The desired inverse is

$$[13]^{-1} = [-44] = [147]. \quad \blacksquare$$

Since every element in \mathbf{Z}_n has an additive inverse, **subtraction** can be defined in \mathbf{Z}_n by the equation

$$\begin{aligned}[a] - [b] &= [a] + (-[b]) \\ &= [a] + [-b] \\ &= [a - b].\end{aligned}$$

We now have at hand the basic knowledge about addition, subtraction, multiplication, and multiplicative inverses in \mathbf{Z}_n . Utilizing this knowledge, we can successfully imitate many of the techniques that we use to solve equations in real numbers to solve equations involving elements of \mathbf{Z}_n . For example, Exercise 9 of this section states that $[x] = [a]^{-1}[b]$ is the unique solution to $[a][x] = [b]$ in \mathbf{Z}_n whenever $[a]^{-1}$ exists. In Exercise 19, some quadratic equations are to be solved by factoring. The next example shows how we can solve a simple system of linear equations in \mathbf{Z}_n by using the same kinds of steps that we use when working in \mathbf{R} .

Example 5 We shall solve the following system of linear equations in \mathbf{Z}_{26} .

$$\begin{aligned}[4][x] + [y] &= [22] \\ [19][x] + [y] &= [15]\end{aligned}$$

We can eliminate $[y]$ by subtracting the top equation from the bottom one:

$$[19][x] - [4][x] = [15] - [22].$$

This simplifies to

$$[15][x] = [-7]$$

or

$$[15][x] = [19].$$

Using the Euclidean Algorithm as we did in Example 4, we find that $[15]$ in \mathbf{Z}_{26} has the multiplicative inverse given by $[15]^{-1} = [7]$. Using the result in Exercise 9 of this section, we find that the solution $[x]$ to $[15][x] = [19]$ is

$$\begin{aligned}[x] &= [15]^{-1}[19] \\ &= [7][19] \\ &= [133] \\ &= [3].\end{aligned}$$

Solving for $[y]$ in the equation $[4][x] + [y] = [22]$, yields

$$\begin{aligned}[y] &= [22] - [4][x] \\ &= [22] - [4][3] \\ &= [22] - [12] \\ &= [10].\end{aligned}$$

It is easy to check that $[x] = [3], [y] = [10]$ is indeed a solution to the system. ■

Exercises 2.6

True or False

Label each of the following statements as either true or false.

1. Every element $[a]$ in \mathbf{Z}_n has an additive inverse.
2. Every element $[a] \neq [0]$ in \mathbf{Z}_n has a multiplicative inverse.
3. $[a][b] = [0]$ implies either $[a] = [0]$ or $[b] = [0]$.
4. $[a][x] = [a][y]$ and $[a] \neq [0]$ implies $[x] = [y]$.

Exercises

1. Perform the following computations in \mathbf{Z}_{12} .

a. $[8] + [7]$	b. $[10] + [9]$
c. $[8][11]$	d. $[6][9]$
e. $[6]([9] + [7])$	f. $[5]([8] + [11])$
g. $[6][9] + [6][7]$	h. $[5][8] + [5][11]$
2. a. Verify that $[1][2][3][4] = [4]$ in \mathbf{Z}_5 .
 b. Verify that $[1][2][3][4][5][6] = [6]$ in \mathbf{Z}_7 .
 c. Evaluate $[1][2][3]$ in \mathbf{Z}_4 .
 d. Evaluate $[1][2][3][4][5]$ in \mathbf{Z}_6 .
3. Make addition tables for each of the following.

a. \mathbf{Z}_2	b. \mathbf{Z}_3	c. \mathbf{Z}_5
d. \mathbf{Z}_6	e. \mathbf{Z}_7	f. \mathbf{Z}_8
4. Make multiplication tables for each of the following.

a. \mathbf{Z}_2	b. \mathbf{Z}_3	c. \mathbf{Z}_6
d. \mathbf{Z}_5	e. \mathbf{Z}_7	f. \mathbf{Z}_8
5. Find the multiplicative inverse of each given element.

a. $[3]$ in \mathbf{Z}_{13}	b. $[7]$ in \mathbf{Z}_{11}	c. $[17]$ in \mathbf{Z}_{20}
d. $[16]$ in \mathbf{Z}_{27}	e. $[17]$ in \mathbf{Z}_{42}	f. $[33]$ in \mathbf{Z}_{58}
g. $[11]$ in \mathbf{Z}_{317}	h. $[9]$ in \mathbf{Z}_{128}	

6. For each of the following \mathbf{Z}_n , list all the elements in \mathbf{Z}_n that have multiplicative inverses in \mathbf{Z}_n .
- a. \mathbf{Z}_6 b. \mathbf{Z}_8 c. \mathbf{Z}_{10}
 d. \mathbf{Z}_{12} e. \mathbf{Z}_{18} f. \mathbf{Z}_{20}
7. Find all zero divisors in each of the following \mathbf{Z}_n .
- a. \mathbf{Z}_6 b. \mathbf{Z}_8 c. \mathbf{Z}_{10}
 d. \mathbf{Z}_{12} e. \mathbf{Z}_{18} f. \mathbf{Z}_{20}
8. Whenever possible, find a solution for each of the following equations in the given \mathbf{Z}_n .
- a. $[4][x] = [2]$ in \mathbf{Z}_6 b. $[6][x] = [4]$ in \mathbf{Z}_{12}
 c. $[6][x] = [4]$ in \mathbf{Z}_8 d. $[10][x] = [6]$ in \mathbf{Z}_{12}
 e. $[8][x] = [6]$ in \mathbf{Z}_{12} f. $[4][x] = [6]$ in \mathbf{Z}_8
 g. $[8][x] = [4]$ in \mathbf{Z}_{12} h. $[4][x] = [10]$ in \mathbf{Z}_{14}
 i. $[10][x] = [4]$ in \mathbf{Z}_{12} j. $[9][x] = [3]$ in \mathbf{Z}_{12}
9. Let $[a]$ be an element of \mathbf{Z}_n that has a multiplicative inverse $[a]^{-1}$ in \mathbf{Z}_n . Prove that $[x] = [a]^{-1}[b]$ is the unique solution in \mathbf{Z}_n to the equation $[a][x] = [b]$.
10. Solve each of the following equations by finding $[a]^{-1}$ and using the result in Exercise 9.
- a. $[4][x] = [5]$ in \mathbf{Z}_{13} b. $[8][x] = [7]$ in \mathbf{Z}_{11}
 c. $[7][x] = [11]$ in \mathbf{Z}_{12} d. $[8][x] = [11]$ in \mathbf{Z}_{15}
 e. $[9][x] = [14]$ in \mathbf{Z}_{20} f. $[8][x] = [15]$ in \mathbf{Z}_{27}
 g. $[6][x] = [5]$ in \mathbf{Z}_{319} h. $[9][x] = [8]$ in \mathbf{Z}_{242}

In Exercises 11–14, solve the systems of equations in \mathbf{Z}_7 .

11. $[2][x] + [y] = [4]$
 $[2][x] + [4][y] = [5]$
12. $[4][x] + [2][y] = [1]$
 $[3][x] + [2][y] = [5]$
13. $[3][x] + [2][y] = [1]$
 $[5][x] + [6][y] = [5]$
14. $[2][x] + [5][y] = [6]$
 $[4][x] + [6][y] = [6]$
15. Prove Theorem 2.29.
16. Prove the following distributive property in \mathbf{Z}_n :

$$[a]([b] + [c]) = [a][b] + [a][c].$$

17. Prove the following equality in \mathbf{Z}_n :

$$([a] + [b])([c] + [d]) = [a][c] + [a][d] + [b][c] + [b][d].$$

18. Let p be a prime integer. Prove that if $[a][b] = [0]$ in \mathbf{Z}_p , then either $[a] = [0]$ or $[b] = [0]$.
19. Use the results in Exercises 16–18 and find all solutions $[x]$ to the following quadratic equations by the factoring method.
- a. $[x]^2 + [5][x] + [6] = [0]$ in \mathbf{Z}_7 b. $[x]^2 + [4][x] + [3] = [0]$ in \mathbf{Z}_5
 c. $[x]^2 + [x] + [5] = [0]$ in \mathbf{Z}_7 d. $[x]^2 + [x] + [3] = [0]$ in \mathbf{Z}_5
20. Let p be a prime integer. Prove that $[1]$ and $[p - 1]$ are the only elements in \mathbf{Z}_p that are their own multiplicative inverses.
21. Show that if n is not a prime, then there exist $[a]$ and $[b]$ in \mathbf{Z}_n such that $[a] \neq [0]$ and $[b] \neq [0]$, but $[a][b] = [0]$; that is, zero divisors exist in \mathbf{Z}_n if n is not prime.
22. Let p be a prime integer. Prove the following cancellation law in \mathbf{Z}_p : If $[a][x] = [a][y]$ and $[a] \neq [0]$, then $[x] = [y]$.
23. Show that if n is not a prime, the cancellation law stated in Exercise 22 does not hold in \mathbf{Z}_n .
24. Prove that a nonzero element $[a]$ in \mathbf{Z}_n is a zero divisor if and only if a and n are not relatively prime.

2.7

Introduction to Coding Theory (Optional)

In this section, we present some applications of congruence modulo n found in basic coding theory. When information is transmitted from one satellite to another or stored and retrieved in a computer or on a compact disc, the information is usually expressed in some sort of code. The ASCII code (American Standard Code for Information Interchange) of 256 characters used in computers is one example. However, errors can occur during the transmission or retrieval processes. The detection and correction of such errors are the fundamental goals of coding theory.

In binary coding theory, we omit the brackets on the elements in \mathbf{Z}_2 and call $\{0, 1\}$ the **binary alphabet**. A **bit**[†] is an element of the binary alphabet. A **word** (or **block**) is a sequence of bits, where all words in a message have the same **length**; that is, they contain the same number of bits. Thus a 2-bit word is an element of $\mathbf{Z}_2 \times \mathbf{Z}_2$. For notational convenience, we omit the comma and parentheses in the 2-bit word (a, b) and write ab , where $a \in \{0, 1\}$ and $b \in \{0, 1\}$. Thus

000	010	001	011
100	110	101	111

[†]Bit is an abbreviation for *binary digit*.

are all eight possible 3-bit words using the binary alphabet. There are thirty-two 5-bit words, so 5-bit words are frequently used to represent the 26 letters of our alphabet, along with 6 punctuation marks.

During the process of sending a message using k -bit words, one or more bits may be received incorrectly. It is essential that errors be detected and, if possible, corrected. The general idea is to generate a **code**, send the coded message, and then decode the coded message, as illustrated here:

$$\text{message} \xrightarrow{\text{encode}} \text{coded message} \xrightarrow{\text{send}} \text{received message} \xrightarrow{\text{decode}} \text{message}.$$

Ideally, the code is devised in such a way as to detect and/or correct any errors in the received message. Most codes require appending extra bits to each k -bit word, forming an n -bit code word. The next example illustrates an **error-detecting** scheme.

Example 1 Parity Check Consider 3-bit words of the form abc . One coding scheme maps abc onto $abcd$, where

$$d \equiv a + b + c \pmod{2}$$

is called the **parity check digit**. If $d = 0$, we say that the word abc has **even parity**. If $d = 1$, we say abc has **odd parity**. Thus the eight possible 3-bit words are mapped onto the eight 4-bit code words as follows:

$$\begin{aligned} \text{word} &\xrightarrow{\text{encode}} \text{code word} \\ 000 &\xrightarrow{\text{encode}} 0000 \\ 010 &\xrightarrow{\text{encode}} 0101 \\ 001 &\xrightarrow{\text{encode}} 0011 \\ 011 &\xrightarrow{\text{encode}} 0110 \\ 100 &\xrightarrow{\text{encode}} 1001 \\ 110 &\xrightarrow{\text{encode}} 1100 \\ 101 &\xrightarrow{\text{encode}} 1010 \\ 111 &\xrightarrow{\text{encode}} 1111. \end{aligned}$$

Note that each 4-bit code word has even parity. Therefore, a simple parity check on the code word will detect any single-bit error. For example, suppose that the coded message of five 4-bit code words

$$1101 \quad 1011 \quad 0000 \quad 0110 \quad 0011$$

is received. It is obvious that each of the first two code words 1101 and 1011 contains at least one error. This parity check scheme does not correct single-bit errors, nor will it detect which bit is in error. It also will not detect 2-bit errors. In this situation, the safest action is to request retransmission of the message, if retransmission is feasible. ■

Example 2 Repetition Codes Multiple errors can be detected (but not corrected) in a scheme in which a k -bit word is mapped onto a $2k$ -bit code word according to the following scheme:

$$x_1x_2 \cdots x_k \xrightarrow{\text{encode}} x_1x_2 \cdots x_kx_1x_2 \cdots x_k.$$

In the coded message with $k = 3$,

$$110110 \quad 010011 \quad 011011 \quad 101000,$$

errors occur in the second code word 010011 and in the last code word 101000. All other code words seem to be correct. If, upon retransmission, the coded message is received as

$$110110 \quad 011011 \quad 011011 \quad 100100,$$

it will be decoded as

$$110 \quad 011 \quad 011 \quad 100. \quad \blacksquare$$

Example 3 Maximum Likelihood Decoding Multiple errors can be detected and *corrected* if each k -bit word is mapped onto a $3k$ -bit code word according to the following scheme (called a **triple repetition code**):

$$x_1x_2 \cdots x_k \xrightarrow{\text{encode}} x_1x_2 \cdots x_kx_1x_2 \cdots x_kx_1x_2 \cdots x_k.$$

For example, if the 6-bit code word (for a 2-bit word)

$$010111$$

is received, then an error is detected. By separating the code word into three equal parts

$$01 \quad 01 \quad 11$$

and comparing bit by bit, we note that the first bits in each part do not agree. We correct the error by choosing the digit that occurs most often, in this case a 0. Thus the corrected code word is

$$010101,$$

and more than likely the correct message is 01. The main disadvantage of this type of coding is that each message requires three times as many bits as the decoded message, whereas with the parity check scheme, only one extra bit is needed for each word. \blacksquare

A combination of a parity check and a repetition code allows detection and correction of coded messages without requiring quite as many bits as in the maximum likelihood scheme. We illustrate this in the next example.

Example 4 Error Detection and Correction Suppose 4-bit words are mapped onto 9-bit code words using the scheme

$$x_1x_2x_3x_4 \xrightarrow{\text{encode}} x_1x_2x_3x_4x_1x_2x_3x_4P,$$

where p is the parity check digit

$$p \equiv x_1 + x_2 + x_3 + x_4 \pmod{2}.$$

For example, the 4-bit word 0110 is encoded as 011001100. Suppose, upon transmission, a code word 101011100 is received. Breaking 101011100 into three parts,

$$1010 \quad 1110 \quad 0,$$

indicates that an error occurs in the second bit. To have parity 0, the correct word must be 1010.

Errors might also occur in the parity digit. For example, if 001100111 is received, an error is detected, and more likely the error has been made in the parity check digit. Thus the correct word is 0011. ■

The last two examples bring up the question of probability of errors occurring in any one or more bits of an n -bit code word. We make the following assumptions:

1. The probability of any single bit being transmitted incorrectly is P .
2. The probability of any single bit being transmitted correctly or incorrectly is independent of the probability of any other single bit being transmitted correctly or incorrectly.

Thus the probability of transmitting a 5-bit code word with only one incorrect bit is $\binom{5}{1}P(1 - P)^4$. If it happens that $P = 0.01$ (approximately 1 of every 100 bits are transmitted incorrectly), then the probability of transmitting a 5-bit code word with only one incorrect bit is $\binom{5}{1}0.01(0.99)^4 = 0.04803$, and the probability of transmitting a 5-bit code word with no errors is $\binom{5}{0}(0.01)^0(0.99)^5 = 0.95099$. Hence the probability of transmitting a 5-bit code word with at most one error is $\binom{5}{1}0.01(0.99)^4 + \binom{5}{0}(0.01)^0(0.99)^5 = 0.99902$.

Up to this point, \mathbf{Z}_2 has been used in all of our examples. We next look at some instances in which other congruence classes play a role.

Example 5 Using Check Digits Many companies use **check digits** for security purposes or for error detection. For example, an 11th digit may be appended to a 10-bit identification number to obtain the 11-digit invoice number of the form

$$x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10}c,$$

where the 11th bit, c , is the check digit, computed as

$$x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10} \equiv c \pmod{n}.$$

If congruence modulo 9 is used, then the check digit for an identification number 3254782201 is 7, since $3254782201 \equiv 7 \pmod{9}$. Thus the complete correct invoice number would appear as 32547822017. If the invoice number 31547822017 were used instead and checked, an error would be detected, since $3154782201 \not\equiv 7 \pmod{9}$. [$3154782201 \equiv 6 \pmod{9}$.]

This particular scheme is not infallible in detecting errors. For example, if a transposition error (a common keyboarding error) occurred and the invoice number were erroneously

entered as 32548722017, an error would not be detected, since $3254872201 \equiv 7 \pmod{9}$. It can be shown that transposition errors will never be detected with this scheme (using congruence modulo 9) unless one of the digits is the check digit. (See Exercise 12.) ■

Even more sophisticated schemes for using check digits appear in such places as the ISBN numbers assigned to all books, the UPCs (Universal Product Codes) assigned to products in the marketplace, passport numbers, and the driver's licenses and license plate numbers in some states. Some of the schemes are very good at detecting errors, and others are surprisingly faulty. In these schemes, a *weighting vector* is used in conjunction with arithmetic on congruence classes modulo n (modular arithmetic). The **dot product** notation is useful in describing the situation. We define the dot product $(x_1, x_2, \dots, x_n) \cdot (y_1, y_2, \dots, y_n)$ of two ordered n -tuples (*vectors*) (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) by

$$(x_1, x_2, \dots, x_n) \cdot (y_1, y_2, \dots, y_n) = x_1y_1 + x_2y_2 + \dots + x_ny_n.$$

For example, $(1, 2, 3) \cdot (-3, 7, -1) = -3 + 14 - 3 = 8$. The next example describes the use of the dot product and weighting vector in bank identification numbers.

Example 6 Bank Identification Numbers Identification numbers for banks have eight digits, x_1x_2, \dots, x_8 , and a check digit, x_9 , given by

$$(x_1, x_2, \dots, x_8) \cdot (7, 3, 9, 7, 3, 9, 7, 3) \equiv x_9 \pmod{10}.$$

The weighting vector is $(7, 3, 9, 7, 3, 9, 7, 3)$. Thus a bank with identification number 05320044 has check digit

$$\begin{aligned} (0, 5, 3, 2, 0, 0, 4, 4) \cdot (7, 3, 9, 7, 3, 9, 7, 3) &= 0 + 15 + 27 + 14 + 0 + 0 + 28 + 12 \\ &= 96 \\ &\equiv 6 \pmod{10} \end{aligned}$$

and appears as 053200446 at the bottom of the check. This particular scheme detects all one-digit errors. However, suppose that this same bank identification number is coded in as 503200446, with a transposition of the first and second digits. The check digit 6 does not detect the error:

$$\begin{aligned} (5, 0, 3, 2, 0, 0, 4, 4) \cdot (7, 3, 9, 7, 3, 9, 7, 3) &= 35 + 0 + 27 + 14 + 0 + 0 + 28 + 12 \\ &= 116 \\ &\equiv 6 \pmod{10}. \end{aligned}$$

Transposition errors of adjacent digits x_i and x_{i+1} will be detected by this scheme except when $|x_i - x_{i+1}| = 5$. (See Exercise 13.) ■

The next example illustrates the use of another weighting vector in Universal Product Codes.

Example 7 UPC Symbols UPC symbols consist of 12 digits, $x_1x_2 \cdots x_{12}$, with the last, x_{12} , being the check digit. The weighting vector used for the UPC symbols is the 11-tuple $(3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3)$. The check digit x_{12} can be computed as

$$-(x_1, x_2, \dots, x_{11}) \cdot (3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3) \equiv x_{12} \pmod{10}.$$

The computation

$$\begin{aligned} -(0, 2, 1, 2, 0, 0, 6, 9, 1, 1, 3) \cdot (3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3) &= -47 \\ &\equiv 3 \pmod{10} \end{aligned}$$

verifies the check digit 3 shown in the UPC symbol in Figure 2.3. As in the bank identification scheme, some transposition errors may go undetected.



■ **Figure 2.3**
UPC Symbol ■

In this section, we have attempted to introduce only the basic concepts of coding theory; more sophisticated coding schemes are constantly being developed. Much research is being done in this branch of mathematics, research based not only on group and field theory but also on linear algebra and probability theory.

Exercises 2.7

True or False

Label each of the following statements as either true or false.

1. Parity check schemes will always detect the position of an error.
2. All errors in a triple repetition code can be corrected by choosing the digit that occurs most often.
3. In parity check schemes, errors might occur in the parity check digit.
4. In a check digit scheme using congruence modulo 9, transposition errors will never be detected.

Exercises

1. Suppose 4-bit words $abcd$ are mapped onto 5-bit code words $abcde$, where e is the parity check digit. Detect any errors in the following six-word coded message.

11101 00101 00010 11100 00011 10100

2. Suppose 3-bit words abc are mapped onto 6-bit code words $abca\bar{b}c$ under a repetition scheme. Detect any errors in the following five-word coded message.

111011 101101 011110 001000 011011

3. Use maximum likelihood decoding to correct the following six-word coded message generated by a triple repetition code. Then decode the message.

101101101 110110101 110100101 101000111 110010011 011011011

4. Suppose 2-bit words ab are mapped onto 5-bit code words $ababc$, where c is the parity check digit. Correct the following seven-word coded message. Then decode the message.

11100 01011 01010 10101 00011 10111 11111

5. Suppose a coding scheme is devised that maps k -bit words onto n -bit code words. The **efficiency** of the code is the ratio k/n . Compute the efficiency of the coding scheme described in each of the following examples.

- a. Example 1
- b. Example 2
- c. Example 3
- d. Example 4

6. Suppose the probability of erroneously transmitting a single digit is $P = 0.03$. Compute the probability of transmitting a 4-bit code word with (a) at most one error, and (b) exactly four errors.

7. Suppose the probability of erroneously transmitting a single digit is $P = 0.0001$. Compute the probability of transmitting an 8-bit code word with (a) no errors, (b) exactly one error, (c) at most one error, (d) exactly two errors, and (e) at most two errors.

8. Suppose the probability of incorrectly transmitting a single bit is $P = 0.001$. Compute the probability of correctly receiving a 100-word coded message made up of 4-bit words.

9. Compute the check digit for the 8-digit identification number 41126450 if the check digit is computed using congruence modulo 7.

10. Is the identification number 11257402 correct if the last digit is the check digit computed using congruence modulo 7?

11. Show that the check digit x_9 in bank identification numbers satisfies the congruence equation

$$(x_1, x_2, \dots, x_8, x_9) \cdot (7, 3, 9, 7, 3, 9, 7, 3, 9) \equiv 0 \pmod{10}.$$

12. Suppose that the check digit is computed as described in Example 5. Prove that transposition errors of adjacent digits will not be detected unless one of the digits is the check digit.

13. Verify that transposition errors of adjacent digits x_i and x_{i+1} will be detected in a bank identification number except when $|x_i - x_{i+1}| = 5$.

14. Compute the check digit for the UPC symbols whose first 11 digits are given.



15. Verify that the check digit x_{12} in a UPC symbol satisfies the following congruence equation:

$$(x_1, x_2, \dots, x_{12}) \cdot (3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1) \equiv 0 \pmod{10}.$$

16. Show that transposition errors of the type

$$x_1 \dots x_{i-1} x_i x_{i+1} \dots x_{12} \rightarrow x_1 \dots x_{i+1} x_i x_{i-1} \dots x_{12}$$

($i = 2, 3, \dots, 11$) in a UPC symbol will not be detected by the check digit.

17. Passports contain identification codes of the following form.

passport number	check digit	birth date	check digit	date of expiry	check digit	final check
012345678	4	USA	480517	7 F 020721	2 <<<<<<<<<<<<	8

Each of the first three check digits is computed on the preceding identification numbers by using a weighting vector of the form

$$(7, 3, 1, 7, 3, 1, \dots)$$

in conjunction with congruence modulo 10. For example, in this passport identification code, the check digit 4 checks the *passport number*, the check digit 7 checks the *birth date*, and the check digit 2 checks the *date of expiry*. The final check digit is then computed by using the same type of weighting vector with all the digits (including check digits, excluding letters). Verify that this passport identification code is valid. Then check the validity of the following passport identification codes.

- a. 0987654326USA1512269F9901018 <<<<<<<<<<<< 4
- b. 0444555331USA4609205M0409131 <<<<<<<<<<<<< 8
- c. 0123987457USA7803012M9711219 <<<<<<<<<<<< 3
- d. 0246813570USA8301047F0312203 <<<<<<<<<<<< 6

- 18.** ISBN numbers are ten-digit numbers that identify books, where x_{10} is the check digit and $(x_1, x_2, \dots, x_{10}) \cdot (10, 9, 8, 7, 6, 5, 4, 3, 2, 1) \equiv 0 \pmod{11}$. Only digits 0 through 9 are used for the first nine digits, and if the check digit is required to be 10, then an X is used in place of the 10. If possible, detect any errors in the following ISBN numbers.
- ISBN 0-534-92888-9
 - ISBN 0-543-91568-X
 - ISBN 0-87150-334-X
 - ISBN 0-87150-063-4
- 19.** In the ISBN scheme, write the check digit x_{10} in the form
- $$(x_1, x_2, \dots, x_9) \cdot \mathbf{y} \equiv x_{10} \pmod{11},$$
- where \mathbf{y} is obtained from the weighting vector $(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)$.
- 20.** Suppose $\mathbf{x} = x_1x_2 \dots x_k$ and $\mathbf{y} = y_1y_2 \dots y_k$ are k -bit words. The **Hamming[†] distance** $d(\mathbf{x}, \mathbf{y})$ between \mathbf{x} and \mathbf{y} is defined to be the number of bits in which \mathbf{x} and \mathbf{y} differ. More precisely, $d(\mathbf{x}, \mathbf{y})$ is the number of indices in which $x_i \neq y_i$. Find the Hamming distance between the following pairs of words.
- 0011010 and 1011001
 - 01000 and 10100
 - 11110011 and 00110001
 - 011000 and 110111
- 21.** Let \mathbf{x} , \mathbf{y} , and \mathbf{z} be k -bit words. Prove the following properties of the Hamming distance.
- $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$
 - $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$
 - $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$
- 22.** The **Hamming weight** $wt(\mathbf{x})$ of a k -bit word is defined to be $wt(\mathbf{x}) = d(\mathbf{x}, \mathbf{0})$, where $\mathbf{0}$ is the k -bit word in which every bit is 0. Find the Hamming weight of each of the following words.
- 0011100
 - 11110
 - 10100001
 - 000110001
- 23.** The **minimum distance** of a code is defined to be the smallest distance between any pair of distinct code words in a code. Suppose a code consists of the following code words. This is the *repetition code on 2-bit words*.

0000 0101 1010 1111

Find the minimum distance of this code.

[†]This distance function is named in honor of Richard Hamming (1915–1998), who pioneered the development of error-correcting codes.

24. Repeat Exercise 23 for the code consisting of the following code words. This code is a *repetition code on 3-bit words with a parity check digit*.

0000000	0100101	0010011	0110110
1001001	1101100	1011010	1111111

25. Repeat Exercise 23 for the code consisting of the following code words.

0000000	0001011	0010111	0011100
0100101	0101110	0110010	0111001
1000110	1001101	1010001	1011010
1100011	1101000	1110100	1111111

This code is called the **Hamming (7,4) code**. Each code word $x_1x_2x_3x_4x_5x_6x_7$, with $x_i \in \{0, 1\}$, can be decoded by using the first four digits $x_1x_2x_3x_4$. The last three digits are parity check digits, where

$$\begin{aligned}x_5 &\equiv x_1 + x_2 + x_3 \pmod{2} \\x_6 &\equiv x_1 + x_3 + x_4 \pmod{2} \\x_7 &\equiv x_2 + x_3 + x_4 \pmod{2}.\end{aligned}$$

26. Write out the eight code words in the (5, 3) code where each code word $x_1x_2x_3x_4x_5$ is generated in the following way:

$$\begin{aligned}x_i &\in \{0, 1\} \\x_4 &\equiv x_1 + x_2 \pmod{2} \\x_5 &\equiv x_1 + x_3 \pmod{2}.\end{aligned}$$

2.8

Introduction to Cryptography (Optional)

An additional application of congruence modulo n is found in **cryptography**, the designing of secret codes. **Cryptanalysis** is the process of breaking the secret codes, and **cryptology** encompasses both cryptography and cryptanalysis. Cryptography differs from code theory in that code theory concentrates on the detection and correction of errors in messages, whereas cryptography concentrates on concealing a message from an unauthorized person.

History is rich with examples of secret writings, dating back as far as 1900 B.C. when an Egyptian master scribe altered hieroglyphic writing, thus forming “secret messages” in the tomb of the nobleman Khnumhotep II. Later, in 400 B.C., the Spartans used a device called a *skytable* to conceal messages. A ribbon was wound around a cylinder (the skytable); then a message was written on the ribbon. When the ribbon was removed, the message appeared scrambled. However, the recipient of the ribbon had a similar skytable upon which he wound the ribbon and then easily read the message. An early cryptological system, called the *Caesar cipher*, was employed by Julius Caesar in the Gallic wars. In this system,

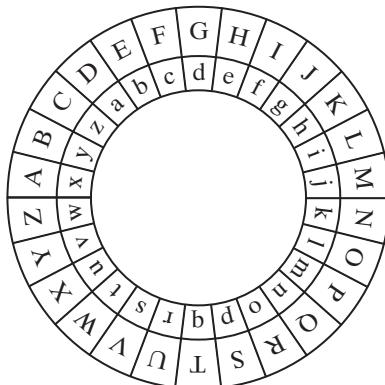
Caesar simply replaced (*substituted*) each letter of the alphabet (the *plaintext*) by the letter three positions to the right (the *ciphertext*). The complete substitution for our alphabet[†] would thus appear as

Plaintext:	a	b	c	d	e	f	g	...	t	u	v	w	x	y	z
Ciphertext:	D	E	F	G	H	I	J	...	W	X	Y	Z	A	B	C

and the plaintext message “attack at dawn” could easily be enciphered and deciphered using the substitution alphabet:

Plaintext:	a	t	t	a	c	k	a	t	d	a	w	n
Ciphertext:	D	W	W	D	F	N	D	W	G	D	Z	Q

The Caesar cipher is an example of an **additive cipher**, or **translation cipher**. All such translation ciphers can be illustrated in a cipher wheel made up of two concentric circles each containing the entire alphabet. One such cipher wheel is shown in Figure 2.4. The inner alphabet, representing the plaintext, is fixed, while the outer alphabet, representing the ciphertext, spins. One pair of plaintext/ciphertext letters determines the entire scheme. This **key** is all that is needed to decipher any message. Caesar’s plaintext/ciphertext key would appear as a/D.



■ **Figure 2.4**
Cipher Wheel

A translation cipher, as used by Caesar, and other, more sophisticated ciphers can be described mathematically. We first accept the following notational convention:

$a \bmod n$ is the remainder when a is divided by n ,

or, in symbols,

$$r = a \bmod n \Leftrightarrow a = nq + r \text{ where } q \text{ and } r \text{ are integers with } 0 \leq r < n.$$

Although this notation closely resembles the congruence notation defined in Section 2.5, the meaning is quite different and the distinction must be kept in mind. For a fixed y , the notation

$$x \equiv y \pmod{n}$$

[†]The letters j, u, and w were not in the Roman alphabet.

allows x to be *any integer* such that $x - y$ is a multiple of n , but the notation

$$x = y \bmod n$$

requires x to be *the unique integer* in the range $0 \leq x < n$ such that $x - y$ is a multiple of n . All of the statements

$$27 \equiv 19 \pmod{8}, \quad 11 \equiv 19 \pmod{8}, \quad \text{and} \quad 3 \equiv 19 \pmod{8}$$

are true, but the statement

$$x = 19 \bmod 8$$

is true if and only if $x = 3$.

Example 1

- a. $3 = 23 \bmod 5$ since $23 = 5(4) + 3$.
- b. $1 = 37 \bmod 4$ since $37 = 4(9) + 1$.
- c. $21 = 47 \bmod 26$ since $47 = 26(1) + 21$.
- d. $19 = -7 \bmod 26$ since $-7 = 26(-1) + 19$.

■

Next we describe a translation cipher in terms of congruence modulo n .

Example 2 Translation Cipher Associate the n letters of the “alphabet” with the integers $0, 1, 2, 3, \dots, n - 1$. Let $A = \{0, 1, 2, 3, \dots, n - 1\}$ and define the mapping $f: A \rightarrow A$ by

$$f(x) = x + k \bmod n$$

where k is the **key**, the number of positions from the plaintext to the ciphertext. If our alphabet consists of a through z , in natural order, followed by a blank, then we have 27 “letters” that we associate with the integers $0, 1, 2, \dots, 26$ as follows:

Alphabet:	a	b	c	d	e	f	...	v	w	x	y	z	“blank”
A:	0	1	2	3	4	5	...	21	22	23	24	25	26

Now if our key is $k = 12$, then the plaintext message “send money” translates into the ciphertext message “DQZPLY ZQJ” as follows:

send money	translate to A	18	4	13	3	26	12	14	13	4	24
	$f(x) = x + 12 \bmod 27$	3	16	25	15	11	24	26	25	16	9
	translate from A	D	Q	Z	P	L	Y		Z	Q	J

The mapping f , given by

$$f(x) = x + k \bmod n$$

can be shown to be one-to-one and onto, so the inverse exists and is given by

$$f^{-1}(x) = x - k \bmod n.$$

The mapping f^{-1} can then be used to decipher the ciphertext.

DQZPLY ZQJ	$\xrightarrow{\text{translate to } A}$	3	16	25	15	11	24	26	25	16	9
	$\xrightarrow{f(x)=x-12 \bmod 27}$	18	4	13	3	26	12	14	13	4	24
	$\xrightarrow{\text{translate from } A}$	s	e	n	d		m	o	n	e	y

A natural extension of the translation (or shift) cipher is found in a mapping of the form

$$f(x) = ax + b \bmod n$$

where a and b are fixed integers. This type of mapping is called an **affine mapping**. The ordered pair a, b of integers forms the key for this type of cipher. If $a = 1$, we simply have a translation cipher, whereas if $b = 0$, we have what's called a **multiplicative cipher**. It follows from Theorem 2.25 that an affine mapping $f:A \rightarrow A$ has an inverse $f^{-1}:A \rightarrow A$ if a and n are relatively prime. When $(a, n) = 1$, it can be shown that the inverse f^{-1} is given by

$$f^{-1}(x) = a'x + b' \bmod n$$

where a' is defined by

$$1 = a'a \bmod n, \text{ with } 0 < a' < n$$

and

$$b' = -a'b \bmod n.$$

Example 3 Affine Mapping We shall use an affine mapping with $a = 5$ and $b = 7$ as the key in our 27-letter alphabet. The mapping $f:A \rightarrow A$, where $A = \{0, 1, 2, \dots, 26\}$, is given by

$$f(x) = 5x + 7 \bmod 27.$$

The plaintext message “hi mom” is translated into the ciphertext “PUCNXN” as follows:

hi mom	$\xrightarrow{\text{translate to } A}$	7	8	26	12	14	12
	$\xrightarrow{f(x)=5x+7 \bmod 27}$	15	20	2	13	23	13
	$\xrightarrow{\text{translate from } A}$	P	U	C	N	X	N

Note that $(5, 27) = 1$, so the mapping f has an inverse given by

$$\begin{aligned} f^{-1}(x) &= 11x - 11(7) \bmod 27 \quad \text{since } 1 = 11 \cdot 5 \bmod 27 \\ &= 11x + 16(7) \bmod 27 \quad \text{since } 16 = -11 \bmod 27 \\ &= 11x + 112 \bmod 27 \\ &= 11x + 4 \bmod 27, \end{aligned}$$

which can then be used to decipher the ciphertext.

PUCNXN	$\xrightarrow{\text{translate to } A}$	15	20	2	13	23	13
	$f(x) = 11x + 4 \pmod{27}$	7	8	26	12	14	12
	$\xrightarrow{\text{translate from } A}$	h	i	m	o	m	■

Example 4 **Affine Mapping with Unknown Key** If a ciphertext message is relatively long, a frequency analysis of letters in a ciphertext can be used to “break the code” when the key to the affine mapping $f(x) = ax + b \pmod{n}$ is not known. Suppose we associate the letters a through z , in natural order, with the integers 0 through 25, respectively, to form the 26-“letter” alphabet $A = \{0, 1, 2, \dots, 25\}$. In the English language, with this alphabet the letter e occurs most often in a lengthy message, and the letters t , a , and o are the next most common. With this in mind, suppose that in a ciphertext message the letter W occurred most frequently, followed in frequency by P . It seems reasonable that the ciphertext letters W and P correspond to the plaintext letters e and t , respectively. Translating these into the set A , we have

	CIPHERTEXT	PLAINTEXT
most frequent:	$W \xrightarrow{\text{translate to } A} 22$	$e \xrightarrow{\text{translate to } A} 4$
next most frequent:	$P \xrightarrow{\text{translate to } A} 15$	$t \xrightarrow{\text{translate to } A} 19$

Therefore, we can determine the key from the solution of the following system of equations for a and b :

$$\begin{aligned} 22 &= a(4) + b \pmod{26} \\ 15 &= a(19) + b \pmod{26}. \end{aligned}$$

From Example 5 in Section 2.6, this solution is given by $a = 3$, $b = 10$. Thus we find the affine mapping $f: A \rightarrow A$ to be given by

$$f(x) = 3x + 10 \pmod{26},$$

with inverse $f^{-1}: A \rightarrow A$ defined by

$$f^{-1}(x) = 9x + 14 \pmod{26}. \quad ■$$

In each of the preceding examples, once the mapping f was known, finding the inverse mapping f^{-1} was not difficult. In other words, once the key is known, a message can easily be deciphered. If security is an important issue (which is usually the case in sending secret messages), then it would certainly be advantageous to devise a system that would be difficult to break even if the key were known. Such systems are called **Public Key Cryptosystems**. We examine the RSA[†] cryptosystem next. The RSA system is based on the difficulty of factoring large numbers.

[†]RSA comes from the initials of the last names of Ronald Rivest, Adi Shamir, and Len Adelman, who devised this system in 1977.

We begin by first choosing two distinct prime numbers, which we label as p and q . Then we form the product

$$m = pq.$$

The value of m can be made known to the public. However, the factorization of m as pq shall be kept secret. The larger the value of m , the more secure this system will be, since breaking the code relies on knowing the prime factors p and q of m . Next we choose e to be relatively prime to the product $(p - 1)(q - 1)$; that is, e is defined by

$$(e, (p - 1)(q - 1)) = 1.$$

Finally, we solve for d in the equation

$$1 = ed \pmod{(p - 1)(q - 1)}.$$

The public keys (the keys to be made known) are e and m , whereas the secret keys are p , q , and d .

Theorem 2.32 ■ RSA Public Key Cryptosystem

Suppose $A = \{0, 1, 2, \dots, m - 1\}$ is an alphabet, consisting of m “letters.” With m, p, q, e , and d as described in the preceding paragraph, let the mapping $f: A \rightarrow A$ be defined by

$$f(x) = x^e \pmod{m}.$$

Then f has the inverse mapping $g: A \rightarrow A$ given by

$$g(x) = x^d \pmod{m}.$$

$p \Rightarrow q$ **Proof** Let $y = x^e \pmod{m}$. Then

$$\begin{aligned} y^d &\equiv (x^e)^d \pmod{m} \\ &\equiv x^{ed} \pmod{m}. \end{aligned}$$

Since

$$1 = ed \pmod{(p - 1)(q - 1)},$$

then

$$ed = k(p - 1)(q - 1) + 1$$

for some integer k .

If $x \not\equiv 0 \pmod{p}$, then

$$\begin{aligned} x^{ed} &\equiv x^{k(p-1)(q-1)+1} \pmod{p} \\ &\equiv x^{k(p-1)(q-1)}x \pmod{p} \\ &\equiv (x^{p-1})^{k(q-1)}x \pmod{p} \\ &\equiv (1)^{k(q-1)}x \pmod{p} \\ &\equiv x \pmod{p} \end{aligned}$$

since $x^{p-1} \equiv 1 \pmod{p}$, from Exercise 51 and Theorem 2.24 in Section 2.5.

If $x \equiv 0 \pmod{p}$, it is clear that $x^{ed} \equiv 0^{ed} \pmod{p} \equiv 0 \pmod{p}$. Thus we have $x^{ed} \equiv x \pmod{p}$ in all cases.

Similarly,

$$x^{ed} \equiv x \pmod{q}.$$

Hence

$$p \mid (x^{ed} - x) \quad \text{and} \quad q \mid (x^{ed} - x).$$

By Exercise 10 in Section 2.4, this implies that

$$pq \mid (x^{ed} - x),$$

and since $m = pq$, we have

$$x^{ed} \equiv x \pmod{m}.$$

Thus $y^d \equiv x^{ed} \pmod{m} \equiv x \pmod{m}$, and it follows that $y^d \pmod{m} = x \pmod{m}$.

We have shown that $g(f(x)) = x$, and analogous steps can be used to verify that $f(g(x)) = x$. Therefore, g is the inverse mapping of f .

We illustrate the RSA cryptosystem with relatively small primes p and q . For the RSA system to be secure, it is recommended that the primes p and q be chosen so as to contain more than 100 digits.

Example 5 RSA Public Key Cryptosystem We first choose two primes (which are to be kept secret):

$$p = 17, \text{ and } q = 43.$$

Then we compute m (which is to be made public):

$$m = pq = 17 \cdot 43 = 731.$$

Next we choose e (which is to be made public), where e must be relatively prime to $(p-1)(q-1) = 16 \cdot 42 = 672$. Suppose we take $e = 205$. The Euclidean Algorithm can be used to verify that $(205, 672) = 1$. Then d is determined by the equation

$$1 = 205d \pmod{672}.$$

Using the Euclidean Algorithm, we find $d = 613$ (which is kept secret). The mapping $f:A \rightarrow A$, where $A = \{0, 1, 2, \dots, 730\}$, defined by

$$f(x) = x^{205} \pmod{731}$$

is used to encrypt a message. Then the inverse mapping $g:A \rightarrow A$, defined by

$$g(x) = x^{613} \pmod{731}$$

can be used to recover the original message.

Using the 27-letter alphabet as in Examples 2 and 3, the plaintext message “no problem” is translated into the message as follows:

plaintext:	n	o	p	r	o	b	l	e	m	
message:	13	14	26	15	17	14	01	11	04	12

The message becomes

$$13142615171401110412.$$

This message must be broken into blocks m_i , each of which is contained in A . If we choose three-digit blocks, each block $m_i < m = 731$.

$$\begin{aligned}m_i: & \quad 131 \quad 426 \quad 151 \quad 714 \quad 011 \quad 104 \quad 12 \\f(m_i) = m_i^{205} \bmod 731 = c_i: & \quad 082 \quad 715 \quad 376 \quad 459 \quad 551 \quad 593 \quad 320\end{aligned}$$

The enciphered message becomes

$$082 \quad 715 \quad 376 \quad 459 \quad 551 \quad 593 \quad 320$$

where we choose to report each c_i with three digits by appending any leading zeros as necessary.

To decipher the message, one must know the secret key $d = 613$ and apply the inverse mapping g to each enciphered message block $c_i = f(m_i)$:

$$\begin{aligned}c_i: & \quad 082 \quad 715 \quad 376 \quad 459 \quad 551 \quad 593 \quad 320 \\g(c_i) = c_i^{613} \bmod 731: & \quad 131 \quad 426 \quad 151 \quad 714 \quad 011 \quad 104 \quad 12\end{aligned}$$

Finally, by rebreaking the “message” back into two-digit blocks, one can translate it back into plaintext.

$$\begin{aligned}\text{three-digit block message: } & \quad 131 \quad 426 \quad 151 \quad 714 \quad 011 \quad 104 \quad 12 \\ \text{two-digit block message: } & \quad 13 \quad 14 \quad 26 \quad 15 \quad 17 \quad 14 \quad 01 \quad 11 \quad 04 \quad 12 \\ \text{plaintext: } & \quad n \quad o \quad p \quad r \quad o \quad b \quad l \quad e \quad m \quad \blacksquare\end{aligned}$$

The RSA Public Key Cipher is an example of an **exponentiation cipher**. As in coding theory, we have barely touched on the basics of cryptography. It is our hope that this short introduction may spark further interest in a topic whose basis lies in modern algebra.

Exercises 2.8

True or False

Label each of the following statements as either true or false.

1. The notation $x = y \bmod n$ is used to indicate the unique integer x in the range $0 \leq x < n$ such that $x - y$ is a multiple of n .
2. In order for an affine mapping $f(x) = ax + b \bmod n$ to have an inverse, a and n must be relatively prime.
3. An example of an exponentiation cipher is the RSA Public Key Cipher.

Exercises

1. In the 27-letter alphabet A described in Example 2, use the translation cipher with key $k = 8$ to encipher the following message.

the check is in the mail

What is the inverse mapping that will decipher the ciphertext?

2. Suppose the alphabet consists of a through z , in natural order, followed by a blank, a comma, a period, an apostrophe, and a question mark, in that order. Associate these “letters” with the numbers $0, 1, 2, \dots, 30$, respectively, thus forming a 31-letter alphabet B . Use the translation cipher with key $k = 21$ to encipher the following message.

what's up, doc?

What is the inverse mapping that will decipher the ciphertext?

3. In the 31-letter alphabet B as in Exercise 2, use the translation cipher with key $k = 11$ to decipher the following message.

?T R P. H G O Z G E Z A G. P L O G X P K

What is the inverse mapping that deciphers this ciphertext?

4. In the 27-letter alphabet A described in Example 2, use the translation cipher with key $k = 15$ to decipher the following message.

F X G T O P B S O G W X B T

What is the inverse mapping that deciphers this ciphertext?

5. In the 27-letter alphabet A described in Example 2, use the affine cipher with key $a = 7$ and $b = 5$ to encipher the following message.

all systems go

What is the inverse mapping that will decipher the ciphertext?

6. In the 31-letter alphabet B described in Exercise 2, use the affine cipher with key $a = 15$ and $b = 22$ to encipher the following message.

Houston, we have a problem.

What is the inverse mapping that will decipher the ciphertext?

7. Suppose the alphabet consists of a through z , in natural order, followed by a blank and then a period. Associate these “letters” with the numbers $0, 1, 2, \dots, 27$, respectively, thus forming a 28-letter alphabet, C . Use the affine cipher with key $a = 3$ and $b = 22$ to decipher the message

E E E T Z R I I Y U A I. G T A I C

and state the inverse mapping that deciphers this ciphertext.

8. Use the alphabet C from the preceding problem and the affine cipher with key $a = 11$ and $b = 7$ to decipher the message

ZZZ Y DJ BJ Y X M D

and state the inverse mapping that deciphers this ciphertext.

9. Suppose that in a long ciphertext message the letter x occurred most frequently, followed in frequency by c . Using the fact that in the 26-letter alphabet A , described in Example 4, e occurs most frequently, followed in frequency by t , read the portion of the message

R N C Y X R N C H F T

enciphered using an affine mapping on A . Write out the affine mapping f and its inverse.

10. Suppose that in a long ciphertext message the letter d occurred most frequently, followed in frequency by n . Using the fact that in the 27-letter alphabet A , described in Example 2, “blank” occurs most frequently, followed in frequency by e , read the portion of the message

G E N D O C F A A D O Q N I D P G M D C F E

enciphered using an affine mapping on A . Write out the affine mapping f and its inverse.

11. Suppose the alphabet consists of a through z , in natural order, followed by a blank and then the digits 0 through 9, in natural order. Associate these “letters” with the numbers 0, 1, 2, . . . , 36, respectively, thus forming a 37-letter alphabet, D . Use the affine cipher to decipher the message

x 0 1 9 1 6 r 9 1 6 5 4 6 m 9 c n 1 l 6 b 1 l l 6 x 0 r z 6 u i i

if you know that the plaintext message begins with “th”. Write out the affine mapping f and its inverse.

12. Suppose the alphabet consists of a through z , in natural order, followed by a blank, a comma, and a period, in that order. Associate these “letters” with the numbers 0, 1, 2, . . . , 28, respectively, thus forming a 29-letter alphabet, E . Use the affine cipher to decipher the message

B Z Z K, A U Z N Z G, R S K Z, A U W A O

if you know that the plaintext message begins with “b” and ends with “.”. Write out the affine mapping f and its inverse.

13. Let $f:A \rightarrow A$ be defined by $f(x) = ax + b \pmod{n}$. Show that $f^{-1}:A \rightarrow A$ exists if $(a, n) = 1$, and is given by $f^{-1}(x) = a'x + b' \pmod{n}$, where a' is defined by

$$1 = a'a \pmod{n}, \text{ with } 0 < a' < n$$

and

$$b' = -a'b \pmod{n}.$$

14. Suppose we encipher a plaintext message M using the mapping $f_1:A \rightarrow A$ resulting in the ciphertext C . Next we treat this ciphertext as plaintext and encipher it using the mapping $f_2: A \rightarrow A$ resulting in the ciphertext D . The composition mapping $f:A \rightarrow A$, where $f = f_2 \circ f_1$, could be used to encipher the plaintext message M resulting in the ciphertext D .
- Prove that if f_1 and f_2 are translation ciphers, then $f = f_2 \circ f_1$ is a translation cipher.
 - Prove that if f_1 and f_2 are affine ciphers, then $f = f_2 \circ f_1$ is an affine cipher.
15.
 - Excluding the identity cipher, how many different translation ciphers are there using an alphabet of n “letters”?
 - Excluding the identity cipher, how many different affine ciphers are there using an alphabet of n “letters,” where n is a prime?
16. Rework Example 5 by breaking the message into two-digit blocks instead of three-digit blocks. What is the enciphered message using the two-digit blocks?
17. Suppose that in an RSA Public Key Cryptosystem, the public key is $e = 13, m = 77$. Encrypt the message “go for it” using two-digit blocks and the 27-letter alphabet A from Example 2. What is the secret key d ?
18. Suppose that in an RSA Public Key Cryptosystem, the public key is $e = 35, m = 64$. Encrypt the message “pay me later” using two-digit blocks and the 27-letter alphabet A from Example 2. What is the secret key d ?
19. Suppose that in an RSA Public Key Cryptosystem, $p = 11, q = 13$, and $e = 7$. Encrypt the message “algebra” using the 26-letter alphabet from Example 4.
 - Use two-digit blocks.
 - Use three-digit blocks.
 - What is the secret key d ?
20. Suppose that in an RSA Public Key Cryptosystem, $p = 17, q = 19$, and $e = 19$. Encrypt the message “pascal” using the 26-letter alphabet from Example 4.
 - Use two-digit blocks.
 - Use three-digit blocks.
 - What is the secret key d ?
21. Suppose that in an RSA Public Key Cryptosystem, the public key is $e = 23, m = 55$. The ciphertext message

26 25 00 39 09 18 52 17 49 52 02

was intercepted. What was the message that was sent? Use the 27-letter alphabet from Example 2.

22. Suppose that in an RSA Public Key Cryptosystem, the public key is $e = 5, m = 51$. The ciphertext message

04 05 32 44 26 39 04 00 13 08 00 44 24 29 17 26 49 28 03

was intercepted. What was the message that was sent? Use the 27-letter alphabet from Example 2.

- 23.** The **Euler[†] phi-function** is defined for positive integers n as follows: $\phi(n)$ is the number of positive integers m such that $1 \leq m \leq n$ and $(m, n) = 1$. Evaluate each of the following and list each of the integers m relatively prime to the given n .

- a. $\phi(5)$
- b. $\phi(19)$
- c. $\phi(15)$
- d. $\phi(27)$
- e. $\phi(12)$
- f. $\phi(36)$

Sec. 3.4, #42 <

- 24.** Prove that the number of ordered pairs a, b that form a key for an affine cipher $f(x) = ax + b \pmod{n}$ is $\phi(n)n$.

- 25. a.** Evaluate each of the following.

i. $\phi(2 \cdot 3)$ ii. $\phi(2 \cdot 5)$ iii. $\phi(3 \cdot 5)$ iv. $\phi(3 \cdot 7)$

- b.** If p is a prime, then $\phi(p) = p - 1$, since all positive integers less than p are relatively prime to p . Prove that if p and q are distinct primes, then $\phi(pq) = (p - 1)(q - 1)$.

- 26. a.** Evaluate each of the following.

i. $\phi(2)$ ii. $\phi(2^2)$ iii. $\phi(2^3)$ iv. $\phi(2^4)$
 v. $\phi(3)$ vi. $\phi(3^2)$ vii. $\phi(3^3)$ viii. $\phi(3^4)$

- b.** If p is a prime and j is a positive integer, prove $\phi(p^j) = p^{j-1}(p - 1)$.

Key Words and Phrases

additive cipher, 124	cryptography, 123	induction postulate, 66
affine mapping, 126	cryptology, 123	key, 124
binary alphabet, 114	Division Algorithm, 81, 82	law of trichotomy, 66
binary representation, 75	dot product, 118	least common multiple, 94
bit, 114	efficiency, 120	length, 114
block, 114	error detecting, 115	less than, greater than, 68, 69
Caesar cipher, 123	Euclidean Algorithm, 88	linear combination, 86
cancellation law for addition, 67	Euler phi-function, 134	minimum distance of a code, 122
check digits, 115, 117	even parity, 115	multiplicative cipher, 126
ciphertext, 124	exponentiation cipher, 130	negative integer, 66
code, 115	Generalized Induction, 74	odd parity, 115
Complete Induction, 75	greatest common divisor, 86	parity check digit, 115
congruence classes, 96	Hamming (7,4) code, 123	plaintext, 124
congruence modulo n , 95	Hamming distance, 122	positive integer, 66
cryptanalysis, 123	Hamming weight, 122	prime integer, 90

[†]Leonhard Paul Euler (1707–1783) was a Swiss mathematician and physicist, who also worked in mechanics, optics, and astronomy. Euler is considered one of the greatest mathematicians of the 18th century and one of the best of all time. He has been featured on Swiss, German, and Russian postage stamps, a Swiss banknote, and has an asteroid named in his honor.

Principle of Mathematical Induction, 72
properties of addition in \mathbf{Z}_n , 107
properties of multiplication in \mathbf{Z}_n , 108
Public Key Cryptosystem, 127
relatively prime integers, 89
repetition codes, 116

residue classes, 96
RSA cryptosystem, 127
Second Principle of Finite Induction, 75
standard form of a positive integer, 92
Strong Mathematical Induction, 75

translation cipher, 124
triple repetition code, 116
Unique Factorization Theorem, 86, 91
Well-Ordering Theorem, 81
zero divisor, 109



Lebrecht/Image Works

A Pioneer in Mathematics

Blaise Pascal (1623–1662)

Blaise Pascal is most commonly associated with *Pascal's triangle*, a triangular-shaped pattern in which the binomial coefficients are generated. Although Pascal was not the first to discover this pattern, it was through his study of the pattern that he became the first writer to describe precisely the process of mathematical induction.

As a child, Pascal was frequently ill. His father, a mathematician himself, used to hide all his own mathematics books because he felt that his son's study of mathematics would be too strenuous. But when he was 12, Pascal was found in his playroom folding pieces of paper, doing an experiment by which he discovered that the sum of the angles in any triangle is equal to 180°. Pascal's father was so impressed that he gave his son Euclid's *Elements* to study, and Pascal soon discovered, on his own, many of the propositions of geometry.

At the age of 14, Pascal was allowed to participate actively in the gatherings of a group of French mathematicians. At 16, he had established significant results in projective geometry. Also at this time, he began developing a calculator to facilitate his father's work of auditing chaotic government tax records. Pascal perfected the machine over a period of ten years by building 50 various models, but ultimately it was too expensive to be practical.

Pascal made many contributions in the fields of mechanics and physics as well. The one-wheeled wheelbarrow is another of his inventions. Through his correspondence with the French mathematician Pierre de Fermat, he and Fermat laid the foundations of probability theory.

Pascal died in 1662 at the age of 39. His contributions to 17th-century mathematics were stunning, especially in view of his short life. Scholars wonder how much more mathematics would have issued from his gifted mind had he lived longer.

This page intentionally left blank

Groups

■ Introduction

Some of the standard topics in elementary group theory are treated in this chapter: subgroups, cyclic groups, isomorphisms, and homomorphisms.

In the development here, the topic of isomorphism appears before homomorphism. Some instructors prefer a different order and teach Section 3.6 (Homomorphisms) before Section 3.5 (Isomorphisms). Logic can be used to support either approach. Isomorphism is a special case of homomorphism, while homomorphism is a generalization of isomorphism. Isomorphisms were placed first in this book with the thought that “same structure” is the simpler idea.

Both the additive and the multiplicative structures in \mathbf{Z}_n serve as a basis for some of the examples in this chapter.

3.1

Definition of a Group

The fundamental notions of set, mapping, binary operation, and binary relation were presented in Chapter 1. These notions are essential for the study of an algebraic system. An **algebraic structure**, or **algebraic system**, is a nonempty set in which at least one equivalence relation (equality) and one or more binary operations are defined. The simplest structures occur when there is only one binary operation, as is the case with the algebraic system known as a *group*.

An introduction to the theory of groups is presented in this chapter, and it is appropriate to point out that this is only an introduction. Entire books have been devoted to the theory of groups; the group concept is extremely useful in both pure and applied mathematics.

A group may be defined as follows.

Definition 3.1 ■ Group

Suppose the binary operation $*$ is defined for elements of the set G . Then G is a **group** with respect to $*$ provided the following four conditions hold:

1. G is **closed** under $*$. That is, $x \in G$ and $y \in G$ imply that $x * y$ is in G .
2. $*$ is **associative**. For all x, y, z in G , $x * (y * z) = (x * y) * z$.

3. G has an **identity element** e . There is an e in G such that $x * e = e * x = x$ for all $x \in G$.
 4. G contains **inverses**. For each $a \in G$, there exists $b \in G$ such that $a * b = b * a = e$.
-

The phrase “with respect to $*$ ” should be noted. For example, the set \mathbf{Z} of all integers is a group with respect to addition but not with respect to multiplication (it has no inverses for elements other than ± 1). Similarly, the set $G = \{1, -1\}$ is a group with respect to multiplication but not with respect to addition. In most instances, however, only one binary operation is under consideration, and we say simply that “ G is a group.” If the binary operation is unspecified, we adopt the multiplicative notation and use the juxtaposition xy to indicate the result of combining x and y . Keep in mind, though, that the binary operation is not necessarily multiplication.

Definition 3.2 ■ Abelian Group

Let G be a group with respect to $*$. Then G is called a **commutative group**, or an **abelian**[†] **group**, if $*$ is commutative. That is, $x * y = y * x$ for all x, y in G .

Example 1 We can obtain some simple examples of groups by considering appropriate subsets of the familiar number systems.

- a. The set \mathbf{C} of all *complex numbers* is an abelian group with respect to addition.
- b. The set $\mathbf{Q} - \{0\}$ of all *nonzero rational numbers* is an abelian group with respect to multiplication.
- c. The set \mathbf{R}^+ of all *positive real numbers* is an abelian group with respect to multiplication, but it is not a group with respect to addition (it has no additive identity and no additive inverses). ■

The following examples give some indication of the great variety there is in groups.

Example 2 Recall from Chapter 1 that a permutation on a set A is a one-to-one mapping from A onto A and that $S(A)$ denotes the set of all permutations on A . We have seen that $S(A)$ is closed with respect to the binary operation \circ of mapping composition and that the operation \circ is associative. The identity mapping I_A is an identity element:

$$f \circ I_A = f = I_A \circ f$$

for all $f \in S(A)$, and each $f \in S(A)$ has an inverse in $S(A)$. Thus we may conclude from results in Chapter 1 that $S(A)$ is a group with respect to composition of mappings. However $S(A)$ is not abelian since mapping composition is not a commutative operation. ■

Example 3 We shall take $A = \{1, 2, 3\}$ and obtain an explicit example of $S(A)$. In order to define an element f of $S(A)$, we need to specify $f(1), f(2)$, and $f(3)$. There are three possible choices for $f(1)$. Since f is to be bijective, there are two choices for $f(2)$ after

[†]The term *abelian* is used in honor of Niels Henrik Abel (1802–1829). A biographical sketch of Abel appears on the last page of this chapter.

$f(1)$ has been designated, and then only one choice for $f(3)$. Hence there are $3! = 3 \cdot 2 \cdot 1$ different mappings f in $S(A)$. These are given by

$$e = I_A : \begin{cases} e(1) = 1 \\ e(2) = 2 \\ e(3) = 3 \end{cases} \quad \sigma : \begin{cases} \sigma(1) = 2 \\ \sigma(2) = 1 \\ \sigma(3) = 3 \end{cases}$$

$$\rho : \begin{cases} \rho(1) = 2 \\ \rho(2) = 3 \\ \rho(3) = 1 \end{cases} \quad \gamma : \begin{cases} \gamma(1) = 3 \\ \gamma(2) = 2 \\ \gamma(3) = 1 \end{cases}$$

$$\tau : \begin{cases} \tau(1) = 3 \\ \tau(2) = 1 \\ \tau(3) = 2 \end{cases} \quad \delta : \begin{cases} \delta(1) = 1 \\ \delta(2) = 3 \\ \delta(3) = 2. \end{cases}$$

Thus $\mathcal{S}(A) = \{e, \rho, \tau, \sigma, \gamma, \delta\}$. Following the same convention as in Exercise 3 of Section 1.4, we shall construct a “multiplication” table for $\mathcal{S}(A)$. As shown in Figure 3.1, the result of $f \circ g$ is entered in the row with f at the left and in the column with g at the top.

\circ		g
f	\cdots	$f \circ g$

■ Figure 3.1

In constructing the table for $\mathcal{S}(A)$, we list the elements of $\mathcal{S}(A)$ in a column at the left and in a row at the top, as shown in Figure 3.2. When the product $\rho^2 = \rho \circ \rho$ is computed, we have

$$\begin{aligned}\rho^2(1) &= \rho(\rho(1)) = \rho(2) = 3 \\ \rho^2(2) &= \rho(\rho(2)) = \rho(3) = 1 \\ \rho^2(3) &= \rho(\rho(3)) = \rho(1) = 2,\end{aligned}$$

so $\rho^2 = \tau$. Similarly, $\rho \circ \sigma = \gamma$, $\sigma \circ \rho = \delta$, and so on.

\circ	e	ρ	ρ^2	σ	γ	δ
e	e	ρ	ρ^2	σ	γ	δ
ρ	ρ	ρ^2	e	γ	δ	σ
ρ^2	ρ^2	e	ρ	δ	σ	γ
σ	σ	δ	γ	e	ρ^2	ρ
γ	γ	σ	δ	ρ	e	ρ^2
δ	δ	γ	σ	ρ^2	ρ	e

■ Figure 3.2

A table such as the one in Figure 3.2 is referred to in various texts as a **multiplication table**, a **group table**, or a **Cayley table**.[†] With such a table, it is easy to locate the identity

[†]The term *Cayley table* is in honor of Arthur Cayley (1821–1895). A biographical sketch of Cayley appears on the last page of Chapter 1.

and inverses of elements. An element e is a left identity if and only if the row headed by e at the left end reads exactly the same as the column headings in the table. Similarly, e is a right identity if and only if the column headed by e at the top reads exactly the same as the row headings in the table. If it exists, the inverse of a certain element a can be found by searching for the identity e in the row headed by a and again in the column headed by a .

If the elements in the row headings are listed in the same order from top to bottom as the elements in the column headings are listed from left to right, it is also possible to use the table to check for commutativity. The operation is commutative if and only if equal elements appear in all positions that are symmetrically placed relative to the diagonal from upper left to lower right. In Example 3, the group is not abelian since the table in Figure 3.2 is not symmetric. For example, $\gamma \circ \rho^2 = \delta$ is in row 5, column 3, and $\rho^2 \circ \gamma = \sigma$ is in row 3, column 5.

Example 4 Let G be the set of complex numbers given by $G = \{1, -1, i, -i\}$, where $i = \sqrt{-1}$, and consider the operation of multiplication of complex numbers in G . The table in Figure 3.3 shows that G is closed with respect to multiplication.

Multiplication in G is associative and commutative, since multiplication has these properties in the set of all complex numbers. We can observe from Figure 3.3 that 1 is the identity element and that all elements have inverses. Each of 1 and -1 is its own inverse, and i and $-i$ are inverses of each other. Thus G is an abelian group with respect to multiplication.

\times	1	-1	i	$-i$
1	1	-1	i	$-i$
-1	-1	1	$-i$	i
i	i	$-i$	-1	1
$-i$	$-i$	i	1	-1

■ Figure 3.3 ■

Example 5 It is an immediate corollary of Theorem 2.28 that the set

$$\mathbf{Z}_n = \{[0], [1], [2], \dots, [n-1]\}$$

of congruence classes modulo n forms an abelian group with respect to addition. ■

Example 6 Let $G = \{e, a, b, c\}$ with multiplication as defined by the table in Figure 3.4.

.	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

■ Figure 3.4 ■

From the table, we observe the following:

1. G is closed under this multiplication.
2. e is the identity element.

3. Each of e and b is its own inverse, and c and a are inverses of each other.
4. This multiplication is commutative.

This multiplication is also associative, but we shall not verify it here because it is a laborious task. It follows that G is an abelian group. ■

Example 7 The table in Figure 3.5 defines a binary operation $*$ on the set $S = \{A, B, C, D\}$.

*	A	B	C	D
A	B	C	A	B
B	C	D	B	A
C	A	B	C	D
D	A	B	D	D

■ **Figure 3.5**

From the table, we make the following observations:

1. S is closed under $*$.
2. C is an identity element.
3. D does not have an inverse since $DX = C$ has no solution.

Thus S is not a group with respect to $*$. ■

Definition 3.3 ■ Finite Group, Infinite Group, Order of a Group

If a group G has a finite number of elements, G is called a **finite group**, or a **group of finite order**. The number of elements in G is called the **order** of G and is denoted by either $o(G)$ or $|G|$. If G does not have a finite number of elements, G is called an **infinite group**.

Example 8 In Example 3, the group

$$G = \{e, \rho, \rho^2, \sigma, \gamma, \delta\}$$

has order $o(G) = 6$. In Example 5, $o(\mathbf{Z}_n) = n$. The set \mathbf{Z} of all integers is a group under addition, and this is an example of an infinite group. If A is an infinite set, then $\mathcal{S}(A)$ furnishes an example of an infinite group. ■

Exercises 3.1

True or False

Label each of the following statements as either true or false.

1. The identity element in a group G is its own inverse.
2. If G is an abelian group, then $x^{-1} = x$ for all x in G .

3. Let G be a group that is not abelian. Then $xy \neq yx$ for all x and y in G .
 4. The set of all nonzero elements in \mathbf{Z}_8 is an abelian group with respect to multiplication.
 5. The Cayley table for a group will always be symmetric with respect to the diagonal from upper left to lower right.
 6. If a set is closed with respect to the operation, then every element must have an inverse.
-

Exercises

In Exercises 1–12, decide whether each of the given sets is a group with respect to the indicated operation. If it is not a group, state a condition in Definition 3.1 that fails to hold.

1. The set of all rational numbers with operation addition.
2. The set of all irrational numbers with operation addition.
3. The set of all positive irrational numbers with operation multiplication.
4. The set of all positive rational numbers with operation multiplication.
5. The set of all real numbers x such that $0 < x \leq 1$, with operation multiplication.
6. For a fixed positive integer n , the set of all complex numbers x such that $x^n = 1$ (that is, the set of all n th roots of 1), with operation multiplication.
7. The set of all complex numbers x that have absolute value 1, with operation multiplication. Recall that the absolute value of a complex number x written in the form $x = a + bi$, with a and b real, is given by $|x| = |a + bi| = \sqrt{a^2 + b^2}$.
8. The set in Exercise 7 with operation addition.
9. The set \mathbf{E} of all even integers with operation addition.
10. The set \mathbf{E} of all even integers with operation multiplication.
11. The set of all multiples of a positive integer n with operation addition.
12. The set of all multiples of a positive integer n with operation multiplication.

In Exercises 13 and 14, the given table defines an operation of multiplication on the set $S = \{e, a, b, c\}$. In each case, find a condition in Definition 3.1 that fails to hold, and thereby show that S is not a group.

13. See Figure 3.6.
14. See Figure 3.7.

\times	e	a	b	c
e	e	a	b	c
a	a	b	a	b
b	b	c	b	c
c	c	e	c	e

■ Figure 3.6

\times	e	a	b	c
e	e	a	b	c
a	e	a	b	c
b	e	a	b	c
c	e	a	b	c

■ Figure 3.7

In Exercises 15–20, let the binary operation $*$ be defined on \mathbf{Z} by the given rule. Determine in each case whether \mathbf{Z} is a group with respect to $*$ and whether it is an abelian group. State which, if any, conditions fail to hold.

15. $x * y = x + y + 1$ **16.** $x * y = x + y - 1$

17. $x * y = x + xy$ **18.** $x * y = xy + y$

19. $x * y = x + xy + y$ **20.** $x * y = x - y$

In Exercises 21–26, decide whether each of the given sets is a group with respect to the indicated operation. If it is not a group, state all of the conditions in Definition 3.1 that fail to hold. If it is a group, state its order.

21. The set $\{[1], [3]\} \subseteq \mathbf{Z}_8$ with operation multiplication.

22. The set $\{[1], [2], [3], [4]\} \subseteq \mathbf{Z}_5$ with operation multiplication.

23. The set $\{[0], [2], [4]\} \subseteq \mathbf{Z}_8$ with operation multiplication.

24. The set $\{[0], [2], [4], [6], [8]\} \subseteq \mathbf{Z}_{10}$ with operation multiplication.

25. The set $\{[0], [2], [4], [6], [8]\} \subseteq \mathbf{Z}_{10}$ with operation addition.

26. The set $\{[0], [2], [4], [6]\} \subseteq \mathbf{Z}_8$ with operation addition.

27. a. Let $G = \{[a] \mid [a] \neq [0]\} \subseteq \mathbf{Z}_n$. Show that G is a group with respect to multiplication in \mathbf{Z}_n if and only if n is a prime. State the order of G .

b. Construct a multiplication table for the group G of all nonzero elements in \mathbf{Z}_7 , and identify the inverse of each element.

28. Let G be the set of eight elements $G = \{1, i, j, k, -1, -i, -j, -k\}$ with identity element 1 and noncommutative multiplication given by[†]

$$\begin{aligned} (-1)^2 &= 1, \\ i^2 &= j^2 = k^2 = -1, \\ ij &= -ji = k, \\ jk &= -kj = i, \\ ki &= -ik = j, \\ -x &= (-1)x = x(-1) \text{ for all } x \text{ in } G. \end{aligned}$$

Sec. 3.3, #18a, 27a ≪

Sec. 3.4, #2 ≪

Sec. 3.5, #11 ≪

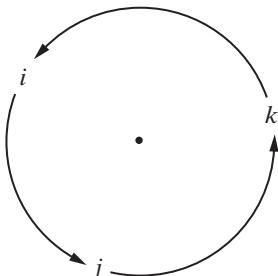
Sec. 4.4, #17 ≪

Sec. 4.5, #10 ≪

Sec. 4.6, #3, 11, 16 ≪

(The circular order of multiplication is indicated by the diagram in Figure 3.8.) Given that G is a group of order 8, write out the multiplication table for G . This group is known as the **quaternion group**.

[†]In a multiplicative group, a^2 is defined by $a^2 = a \cdot a$.



■ Figure 3.8

- 29.** A **permutation matrix** is a matrix that can be obtained from an identity matrix I_n by interchanging the rows one or more times (that is, by *permuting* the rows). For $n = 3$, the permutation matrices are I_3 and the five matrices

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad P_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$P_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad P_5 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Sec. 3.3, #18c, 27c ≪

Sec. 3.4, #5 ≪ Given that $G = \{I_3, P_1, P_2, P_3, P_4, P_5\}$ is a group of order 6 with respect to matrix multiplication, write out a multiplication table for G .

Sec. 4.2, #6 ≪

- 30.** Consider the matrices

$$R = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad V = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad T = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Sec. 3.3, #18b, 27b ≪

Sec. 4.1, #20 ≪ in $M_2(\mathbf{R})$, and let $G = \{I_2, R, R^2, R^3, H, D, V, T\}$. Given that G is a group of order 8 with respect to multiplication, write out a multiplication table for G .

- Sec. 4.6, #14 ≪ **31.** Prove or disprove that the set of all diagonal matrices in $M_n(\mathbf{R})$ forms a group with respect to addition.

- 32.** Let G be the set of all matrices in $M_3(\mathbf{R})$ that have the form

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

with all three numbers a , b , and c nonzero. Prove or disprove that G is a group with respect to multiplication.

33. Let G be the set of all matrices in $M_3(\mathbf{R})$ that have the form

$$\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

for arbitrary real numbers a , b , and c . Prove or disprove that G is a group with respect to multiplication.

34. Prove or disprove that the set G in Exercise 32 is a group with respect to addition.
 35. Prove or disprove that the set G in Exercise 33 is a group with respect to addition.
 36. For an arbitrary set A , the power set $\mathcal{P}(A)$ was defined in Section 1.1 by $\mathcal{P}(A) = \{X \mid X \subseteq A\}$, and addition in $\mathcal{P}(A)$ was defined by

$$\begin{aligned} X + Y &= (X \cup Y) - (X \cap Y) \\ &= (X - Y) \cup (Y - X). \end{aligned}$$

- a. Prove that $\mathcal{P}(A)$ is a group with respect to this operation of addition.
 b. If A has n distinct elements, state the order of $\mathcal{P}(A)$.

- Sec. 1.1, #7c ➤ 37. Write out the elements of $\mathcal{P}(A)$ for the set $A = \{a, b, c\}$, and construct an addition table for $\mathcal{P}(A)$ using addition as defined in Exercise 36.
 Sec. 1.1, #7c ➤ 38. Let $A = \{a, b, c\}$. Prove or disprove that $\mathcal{P}(A)$ is a group with respect to the operation of union.
 Sec. 1.1, #7c ➤ 39. Let $A = \{a, b, c\}$. Prove or disprove that $\mathcal{P}(A)$ is a group with respect to the operation of intersection.

3.2

Properties of Group Elements

Several consequences of the definition of a group are recorded in Theorem 3.4.

Strategy

- Parts **a** and **b** of the next theorem are statements about uniqueness, and they can be proved by the standard type of uniqueness proof: Assume that two such quantities exist, and then prove the two to be equal.

Theorem 3.4 ■ Properties of Group Elements

Let G be a group with respect to a binary operation that is written as multiplication.

- a. The identity element e in G is unique.
- b. For each $x \in G$, the inverse x^{-1} in G is unique.
- c. For each $x \in G$, $(x^{-1})^{-1} = x$.

d. Reverse order law. For any x and y in G , $(xy)^{-1} = y^{-1}x^{-1}$.

e. Cancellation laws. If a , x , and y are in G , then either of the equations $ax = ay$ or $xa = ya$ implies that $x = y$.

Uniqueness **Proof** We prove parts **b** and **d** and leave the others as exercises. To prove part **b**, let $x \in G$, and suppose that each of y and z is an inverse of x . That is,

$$xy = e = yx \quad \text{and} \quad xz = e = zx.$$

Then

$$\begin{aligned} y &= ey && \text{since } e \text{ is an identity} \\ &= (zx)y && \text{since } zx = e \\ &= z(xy) && \text{by associativity} \\ &= z(e) && \text{since } xy = e \\ &= z && \text{since } e \text{ is an identity.} \end{aligned}$$

Thus $y = z$, and this justifies the notation x^{-1} as the unique inverse of x in G .

$(p \wedge q) \Rightarrow r$ We shall use part **b** in the proof of part **d**. Specifically, we shall use the fact that the inverse $(xy)^{-1}$ is unique. This means that in order to show that $y^{-1}x^{-1} = (xy)^{-1}$, we need only to verify that $(xy)(y^{-1}x^{-1}) = e = (y^{-1}x^{-1})(xy)$. These calculations are straightforward:

$$(y^{-1}x^{-1})(xy) = y^{-1}(x^{-1}x)y = y^{-1}ey = y^{-1}y = e$$

and

$$(xy)(y^{-1}x^{-1}) = x(yy^{-1})x^{-1} = xex^{-1} = xx^{-1} = e.$$

The order of the factors y^{-1} and x^{-1} in the reverse order law $(xy)^{-1} = y^{-1}x^{-1}$ is crucial in a nonabelian group. An example where $(xy)^{-1} \neq x^{-1}y^{-1}$ is requested in Exercise 5 at the end of this section.

Part **e** of Theorem 3.4 implies that in the table for a finite group G , no element of G appears twice in the same row, and no element of G appears twice in the same column. These results can be extended to the statement in the following strategy box. The proof of this fact is requested in Exercise 10.

Strategy ■ In the multiplication table for a group G , each element of G appears exactly once in each row and also appears exactly once in each column.

Although our definition of a group is a standard one, alternative forms can be made. One of these is given in the next theorem.

Theorem 3.5 ■ Equivalent Conditions for a Group

Let G be a nonempty set that is closed under an associative binary operation called multiplication. Then G is a group if and only if the equations $ax = b$ and $ya = b$ have solutions x and y in G for all choices of a and b in G .

$p \Rightarrow (q \wedge r)$ **Proof** Assume first that G is a group, and let a and b represent arbitrary elements of G . Now a^{-1} is in G , and so are $x = a^{-1}b$ and $y = ba^{-1}$. With these choices for x and y , we have

$$ax = a(a^{-1}b) = (aa^{-1})b = eb = b$$

and

$$ya = (ba^{-1})a = b(a^{-1}a) = be = b.$$

Thus G contains solutions x and y to $ax = b$ and $ya = b$.

$(q \wedge r) \Rightarrow p$

Suppose now that the equations always have solutions in G . We first show that G has an identity element. Let a represent an arbitrary but fixed element in G . The equation $ax = a$ has a solution $x = u$ in G . We shall show that u is a right identity for every element in G . To do this, let b be arbitrary in G . With z a solution to $ya = b$, we have $za = b$ and

$$bu = (za)u = z(au) = za = b.$$

Thus u is a right identity for every element in G . In a similar fashion, there exists an element v in G such that $vb = b$ for all b in G . Then $vu = v$, since u is a right identity, and $vu = u$, since v is a left identity. That is, the element $e = u = v$ is an identity element for G .

Now for any a in G , let x be a solution to $ax = e$, and let y be a solution to $ya = e$. Combining these equations, we have

$$\begin{aligned} x &= ex \\ &= yax \\ &= ye \\ &= y, \end{aligned}$$

and $x = y$ is an inverse for a . This proves that G is a group.

In a group G , the associative property can be extended to products involving more than three factors. For example, if a_1, a_2, a_3 , and a_4 are elements of G , then applications of condition 2 in Definition 3.1 yield

$$[a_1(a_2a_3)]a_4 = [(a_1a_2)a_3]a_4$$

and

$$(a_1a_2)(a_3a_4) = [(a_1a_2)a_3]a_4.$$

These equalities suggest (but do not completely prove) that regardless of how symbols of grouping are introduced in a product $a_1a_2a_3a_4$, the resulting expression can be reduced to

$$[(a_1a_2)a_3]a_4.$$

With these observations in mind, we make the following definition.

Definition 3.6 ■ Product Notation

Let n be a positive integer, $n \geq 2$. For elements a_1, a_2, \dots, a_n in a group G , the expression $a_1 a_2 \cdots a_n$ is defined recursively by

$$a_1 a_2 \cdots a_k a_{k+1} = (a_1 a_2 \cdots a_k) a_{k+1} \quad \text{for } k \geq 1.$$

We can now prove the following generalization of the associative property.

Theorem 3.7 ■ Generalized Associative Law

Let $n \geq 2$ be a positive integer, and let a_1, a_2, \dots, a_n denote elements of a group G . For any positive integer m such that $1 \leq m < n$,

$$(a_1 a_2 \cdots a_m)(a_{m+1} \cdots a_n) = a_1 a_2 \cdots a_n.$$

Complete Induction **Proof** For $n \geq 2$, let P_n denote the statement of the theorem. With $n = 2$, the only possible value for m is $m = 1$, and P_2 asserts the trivial equality

$$(a_1)(a_2) = a_1 a_2.$$

Assume now that P_k is true: For any positive integer m such that $1 \leq m < k$,

$$(a_1 a_2 \cdots a_m)(a_{m+1} \cdots a_k) = a_1 a_2 \cdots a_k.$$

Consider the statement P_{k+1} , and let m be a positive integer such that $1 \leq m < k + 1$. We treat separately the cases where $m = k$ and where $1 \leq m < k$. If $m = k$, the desired equality is true at once from Definition 3.6, as follows:

$$(a_1 a_2 \cdots a_m)(a_{m+1} \cdots a_{k+1}) = (a_1 a_2 \cdots a_k) a_{k+1}.$$

If $1 \leq m < k$, then

$$a_{m+1} \cdots a_k a_{k+1} = (a_{m+1} \cdots a_k) a_{k+1}$$

by Definition 3.6, and consequently,

$$\begin{aligned} & (a_1 a_2 \cdots a_m)(a_{m+1} \cdots a_k a_{k+1}) \\ &= (a_1 a_2 \cdots a_m)[(a_{m+1} \cdots a_k) a_{k+1}] \\ &= [(a_1 a_2 \cdots a_m)(a_{m+1} \cdots a_k)] a_{k+1} \quad \text{by the associative property} \\ &= [a_1 a_2 \cdots a_k] a_{k+1} \quad \text{by } P_k \\ &= a_1 a_2 \cdots a_{k+1} \quad \text{by Definition 3.6.} \end{aligned}$$

Thus P_{k+1} is true whenever P_k is true, and the proof of the theorem is complete.

The material in Section 1.6 on matrices leads to some interesting examples of groups, both finite and infinite. This is pursued now in Examples 1 and 2.

Example 1 Theorem 1.30 translates directly into the statement that $M_{m \times n}(\mathbf{R})$ is an abelian group with respect to addition. This is an example of another infinite group.

When the proof of each part of Theorem 1.30 is examined, it becomes clear that each group property in $M_{m \times n}(\mathbf{R})$ derives in a natural way from the corresponding property in \mathbf{R} .

If the set \mathbf{R} is replaced by the set \mathbf{Z} of all integers, the steps in the proof of each part of Theorem 1.30 can be paralleled to prove the same group property for $M_{m \times n}(\mathbf{Z})$. Thus $M_{m \times n}(\mathbf{Z})$ is also a group under addition. The same reasoning is valid if \mathbf{R} is replaced by the set \mathbf{Q} of all rational numbers, by the set \mathbf{C} of all complex numbers, or by the set \mathbf{Z}_k of all congruence classes modulo k . That is, each of $M_{m \times n}(\mathbf{Q})$, $M_{m \times n}(\mathbf{C})$, and $M_{m \times n}(\mathbf{Z}_k)$ is a group with respect to addition.

We thus have a family of groups, with $M_{m \times n}(\mathbf{Z}_k)$ finite and all the others infinite. Some aspects of computation in $M_{m \times n}(\mathbf{Z}_k)$ may appear strange at first. For instance,

$$B = \begin{bmatrix} [1] & [3] & [0] \\ [2] & [4] & [2] \end{bmatrix}$$

is the additive inverse of

$$A = \begin{bmatrix} [4] & [2] & [0] \\ [3] & [1] & [3] \end{bmatrix}$$

in $M_{2 \times 3}(\mathbf{Z}_5)$, since

$$A + B = \begin{bmatrix} [0] & [0] & [0] \\ [0] & [0] & [0] \end{bmatrix} = B + A. \quad \blacksquare$$

In Example 4 of Section 1.6, it was shown that the matrix

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

in $M_2(\mathbf{R})$ does not have an inverse, so the nonzero elements of $M_2(\mathbf{R})$ do not form a group with respect to multiplication. This result generalizes to arbitrary $M_n(\mathbf{R})$ with $n > 1$; that is, the nonzero elements of $M_n(\mathbf{R})$ do not form a group with respect to multiplication. However, the next example shows that the invertible elements[†] of $M_n(\mathbf{R})$ form a group under multiplication.

Example 2 We shall show that the invertible elements of $M_n(\mathbf{R})$ form a group G with respect to matrix multiplication.

We have seen in Section 1.6 that matrix multiplication is a binary operation on $M_n(\mathbf{R})$, that this operation is associative (Theorem 1.32), and that $I_n = [\delta_{ij}]_{n \times n}$ is an identity element (Theorem 1.34). These properties remain valid when attention is restricted to the set G of invertible elements of $M_n(\mathbf{R})$, so we need only show that G is closed under multiplication. To this end, suppose that A and B are elements of $M_n(\mathbf{R})$ such that A^{-1} and B^{-1} exist. Using the associative property of matrix multiplication, we can write

$$\begin{aligned} (AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} \\ &= AI_nA^{-1} \\ &= AA^{-1} \\ &= I_n. \end{aligned}$$

[†]Recall that a square matrix A is called *invertible* if its multiplicative inverse, A^{-1} , exists.

Although matrix multiplication is not commutative, a similar simplification shows that

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

and it follows that $(AB)^{-1}$ exists and that $(AB)^{-1} = B^{-1}A^{-1}$. Thus G is a group.

As in Example 1, the discussion in the preceding paragraph can be extended by replacing \mathbf{R} with one of the systems \mathbf{Z} , \mathbf{Q} , \mathbf{C} , or \mathbf{Z}_k . That is, the invertible elements in each of the sets $M_n(\mathbf{Z})$, $M_n(\mathbf{Q})$, $M_n(\mathbf{C})$, and $M_n(\mathbf{Z}_k)$ form a group with respect to multiplication. Once again, the computations in $M_n(\mathbf{Z}_k)$ may seem strange. As an illustration, it can be verified by multiplication that

$$\begin{bmatrix} [3] & [1] \\ [5] & [2] \end{bmatrix} \text{ is the inverse of } \begin{bmatrix} [2] & [6] \\ [2] & [3] \end{bmatrix}$$

in the group of invertible elements of $M_2(\mathbf{Z}_7)$. ■

Exercises 3.2

True or False

Label each of the following statements as either true or false.

1. Let x , y , and z be elements of a group G . Then $(xyz)^{-1} = x^{-1}y^{-1}z^{-1}$.
2. In a Cayley table for a group, each element appears exactly once in each row.
3. The Generalized Associative Law applies to any group, no matter what the group operation is.
4. The nonzero elements of $M_{m \times n}(\mathbf{R})$ form a group with respect to matrix multiplication.
5. The nonzero elements of $M_n(\mathbf{R})$ form a group with respect to matrix multiplication.
6. The invertible elements of $M_n(\mathbf{R})$ with respect to matrix multiplication form an abelian group.

Exercises

1. Prove part **a** of Theorem 3.4.
2. Prove part **c** of Theorem 3.4.
3. Prove part **e** of Theorem 3.4.
4. An element x in a multiplicative group G is called **idempotent** if $x^2 = x$. Prove that the identity element e is the only idempotent element in a group G .
5. In Example 3 of Section 3.1, find elements a and b of $S(A)$ such that $(ab)^{-1} \neq a^{-1}b^{-1}$.
6. In Example 3 of Section 3.1, find elements a , b , and c of $S(A)$ such that $ab = bc$ but $a \neq c$.
7. In Example 3 of Section 3.1, find elements a and b of $S(A)$ such that $(ab)^2 \neq a^2b^2$.
8. Prove that in Theorem 3.5, the solutions to the equations $ax = b$ and $ya = b$ are actually unique.

9. Let G be a group.
- Prove that the relation R on G , defined by xRy if and only if there exists an $a \in G$ such that $y = a^{-1}xa$, is an equivalence relation.
 - Let $x \in G$. Find $[x]$, the equivalence class containing x , if G is abelian.
10. Suppose that G is a finite group. Prove that each element of G appears in the multiplication table for G exactly once in each row and exactly once in each column.

In Exercises 11 and 12, part of the multiplication table for the group $G = \{a, b, c, d\}$ is given. In each case, complete the table.

11. See Figure 3.9.

12. See Figure 3.10.

\times	a	b	c	d
a		d		
b				
c			c	
d				c

■ Figure 3.9

\times	a	b	c	d
a				
b			a	
c				a
d				

■ Figure 3.10

13. Prove that if $x = x^{-1}$ for all x in the group G , then G is abelian.
14. Let a and b be elements of a group G . Prove that G is abelian if and only if $(ab)^{-1} = a^{-1}b^{-1}$.
15. Let a and b be elements of a group G . Prove that G is abelian if and only if $(ab)^2 = a^2b^2$.
16. Use mathematical induction to prove that if a is an element of a group G , then $(a^{-1})^n = (a^n)^{-1}$ for every positive integer n .
17. Let a, b, c , and d be elements of a group G . Find an expression for $(abcd)^{-1}$ in terms of a^{-1}, b^{-1}, c^{-1} , and d^{-1} .
18. Use mathematical induction to prove that if a_1, a_2, \dots, a_n are elements of a group G , then $(a_1a_2 \cdots a_n)^{-1} = a_n^{-1}a_{n-1}^{-1} \cdots a_2^{-1}a_1^{-1}$. (This is the general form of the **reverse order law** for inverses.)
19. Let G be a group that has even order. Prove that there exists at least one element $a \in G$ such that $a \neq e$ and $a = a^{-1}$.
20. Prove or disprove that every group of order 3 is abelian.
21. Prove or disprove that every group of order 4 is abelian.
22. Suppose G is a finite set with n distinct elements given by $G = \{a_1, a_2, \dots, a_n\}$. Assume that G is closed under an associative binary operation $*$ and that the following

two cancellation laws hold for all a, x , and y in G :

$$\begin{aligned} a * x = a * y &\text{ implies } x = y; \\ x * a = y * a &\text{ implies } x = y. \end{aligned}$$

Prove that G is a group with respect to $*$.

23. Suppose that G is a nonempty set that is closed under an associative binary operation $*$ and that the following two conditions hold:

1. There exists a left identity e in G such that $e * x = x$ for all $x \in G$.
2. Each $a \in G$ has a left inverse a_l in G such that $a_l * a = e$.

Prove that G is a group by showing that e is in fact a two-sided identity for G and that a_l is a two-sided inverse of a .

24. Reword Definition 3.6 for a group with respect to addition.

25. State and prove Theorem 3.7 for an additive group.

26. Find the additive inverse of $\begin{bmatrix} [2] & [4] & [1] \\ [0] & [5] & [3] \end{bmatrix}$ in the given group.

a. $M_{2 \times 3}(\mathbf{Z}_6)$

b. $M_{2 \times 3}(\mathbf{Z}_7)$

27. Find the multiplicative inverse of $\begin{bmatrix} [1] & [2] \\ [3] & [4] \end{bmatrix}$ in the given group.

a. Invertible elements of $M_2(\mathbf{Z}_5)$

b. Invertible elements of $M_2(\mathbf{Z}_7)$

3.3 Subgroups

Among the nonempty subsets of a group G , there are some that themselves form a group with respect to the binary operation $*$ in G . That is, a subset $H \subseteq G$ may be such that H is also a group with respect to $*$. Such a subset H is called a *subgroup* of G .

Definition 3.8 ■ Subgroup

Let G be a group with respect to the binary operation $*$. A subset H of G is called a **subgroup** of G if H forms a group with respect to the binary operation $*$ that is defined in G .

The subsets $H = \{e\}$ and $H = G$ are always subgroups of the group G . They are referred to as **trivial subgroups**, and all other subgroups of G are called **nontrivial**.

Example 1 The set \mathbf{Z} of all integers is a group with respect to addition, and the set \mathbf{E} of all even integers is a nontrivial subgroup of \mathbf{Z} . (See Exercise 9 of Section 3.1.) ■

Example 2 The set of all nonzero complex numbers is a group under multiplication, and $G = \{1, -1, i, -i\}$ is a nontrivial subgroup of this group. (See Example 4 of Section 3.1.) ■

Example 3 From the discussion in Example 1 of Section 3.2, it is clear that for fixed m and n , each of the additive groups in the list

$$M_{m \times n}(\mathbf{Z}) \subseteq M_{m \times n}(\mathbf{Q}) \subseteq M_{m \times n}(\mathbf{R}) \subseteq M_{m \times n}(\mathbf{C})$$

is a subgroup of every listed group in which it is contained. ■

If G is a group with respect to $*$, then $*$ is an associative operation on any nonempty subset of G . A subset H of G is a subgroup, provided that

1. H contains the identity;
2. H is closed under $*$; and
3. H contains an inverse for each of its elements.

In connection with condition 1, consider the possibility that H might contain an identity e' for its elements that could be different from the identity e of G . Such an element e' would have the property that $e' * e' = e'$, and Exercise 4 of Section 3.2 then implies that $e' = e$. In connection with condition 3, we might consider the possibility that an element $a \in H$ might have one inverse as an element of the subgroup H and a different inverse as an element of the group G . In fact, this cannot happen because part **b** of Theorem 3.4 guarantees that the solution y to $a * y = y * a = e$ is unique in G . The following theorem gives a set of conditions that is slightly different from 1, 2, and 3.

Theorem 3.9 ■ Equivalent Set of Conditions for a Subgroup

A subset H of the group G is a subgroup of G if and only if these conditions are satisfied:

- a. H is nonempty;
- b. $x \in H$ and $y \in H$ imply $xy \in H$; and
- c. $x \in H$ implies $x^{-1} \in H$.

$p \Rightarrow q$ **Proof** If H is a subgroup of G , the conditions follow at once from Definitions 3.8 and 3.1.

$p \Leftarrow q$ Suppose that H is a subset of G that satisfies the conditions. Since H is nonempty, there is at least one $a \in H$. By condition c, $a^{-1} \in H$. But $a \in H$ and $a^{-1} \in H$ imply $aa^{-1} = e \in H$, by condition b. Thus H contains e , is closed, and contains inverses. Hence H is a subgroup.

Example 4 It follows from Example 5 of Section 3.1 that

$$G = \mathbf{Z}_8 = \{[0], [1], [2], [3], [4], [5], [6], [7]\}$$

forms an abelian group with respect to addition $[a] + [b] = [a + b]$. Consider the subset

$$H = \{[0], [2], [4], [6]\}$$

of G . An addition table for H is given in Figure 3.11. The subset H is nonempty, and it is evident from the table that H is closed and contains the inverse of each of its elements. Hence H is a nontrivial abelian subgroup of \mathbf{Z}_8 under addition.

+	[0]	[2]	[4]	[6]
[0]	[0]	[2]	[4]	[6]
[2]	[2]	[4]	[6]	[0]
[4]	[4]	[6]	[0]	[2]
[6]	[6]	[0]	[2]	[4]

Figure 3.11

Example 5 In Exercise 27 of Section 3.1, it was shown that

$$G = \{[1], [2], [3], [4], [5], [6]\} \subseteq \mathbf{Z}_7$$

is a group with respect to multiplication in \mathbf{Z}_7 . The multiplication table in Figure 3.12 shows that the nonempty subset

$$H = \{[1], [2], [4]\}$$

is closed and contains inverses and therefore is an abelian subgroup of G .

•	[1]	[2]	[4]
[1]	[1]	[2]	[4]
[2]	[2]	[4]	[1]
[4]	[4]	[1]	[2]

Figure 3.12

An even shorter set of conditions for a subgroup is given in the next theorem.

Theorem 3.10 ■ Equivalent Set of Conditions for a Subgroup

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

- a. H is nonempty, and
- b. $a \in H$ and $b \in H$ imply $ab^{-1} \in H$.

$p \Rightarrow q$ **Proof** Assume H is a subgroup of G . Then H is nonempty since $e \in H$. Let $a \in H$ and $b \in H$. Then $b^{-1} \in H$ since H contains inverses. Since $a \in H$ and $b^{-1} \in H$, the product $ab^{-1} \in H$ because H is closed. Thus conditions **a** and **b** are satisfied.

$p \Leftarrow q$ Suppose, conversely, that conditions **a** and **b** hold for H . There is at least one $a \in H$, and condition **b** implies that $aa^{-1} = e \in H$. For an arbitrary $x \in H$, we have $e \in H$ and $x \in H$, which implies that $ex^{-1} = x^{-1} \in H$. Thus H contains inverses. To show closure, let $x \in H$ and $y \in H$. Since H contains inverses, $y^{-1} \in H$. But $x \in H$ and $y^{-1} \in H$ imply $x(y^{-1})^{-1} = xy \in H$, by condition **b**. Hence H is closed; therefore, H is a subgroup of G .

When the phrase “ H is a subgroup of G ” is used, it indicates that H is a group with respect to the group operation in G . Consider the following example.

Example 6 The operation of multiplication is defined in \mathbf{Z}_{10} by

$$[a][b] = [ab].$$

This rule defines a binary operation that is associative, and \mathbf{Z}_{10} is closed under this multiplication. Also, $[1]$ is an identity element. However, \mathbf{Z}_{10} is *not* a group with respect to multiplication, since some of its elements do not have inverses. For example, the products

$$\begin{array}{ll} [2][0] = [0] & [2][1] = [2] \\ [2][2] = [4] & [2][3] = [6] \\ [2][4] = [8] & [2][5] = [0] \\ [2][6] = [2] & [2][7] = [4] \\ [2][8] = [6] & [2][9] = [8] \end{array}$$

show that $[2][x] = [1]$ has no solution in \mathbf{Z}_{10} .

Now let us examine the multiplication table for the subset $H = \{[2], [4], [6], [8]\}$ of \mathbf{Z}_{10} (see Figure 3.13). It is surprising, perhaps, but the table shows that $[6]$ is an identity element for H and that H actually forms a group with respect to multiplication. However, H is *not* a subgroup of \mathbf{Z}_{10} since \mathbf{Z}_{10} is not a group with respect to multiplication.

\times	[2]	[4]	[6]	[8]
[2]	[4]	[8]	[2]	[6]
[4]	[8]	[6]	[4]	[2]
[6]	[2]	[4]	[6]	[8]
[8]	[6]	[2]	[8]	[4]

■ **Figure 3.13**



Integral exponents can be defined for elements of a group as follows.

Definition 3.11 ■ Integral Exponents

Let G be a group with the binary operation written as multiplication. For any $a \in G$, we define **nonnegative integral exponents** by

$$a^0 = e, \quad a^1 = a,$$

and

$$a^{k+1} = a^k \cdot a \quad \text{for any positive integer } k.$$

Negative integral exponents are defined by

$$a^{-k} = (a^{-1})^k \quad \text{for any positive integer } k.$$

It is common practice to write the binary operation as addition in the case of abelian groups. When the operation is addition, the corresponding **multiples** of a are defined in a similar fashion. The following list shows how the notations correspond, where k is a positive integer.

Multiplicative Notation	Additive Notation
$a^0 = e$	$0a = 0$
$a^1 = a$	$1a = a$
$a^{k+1} = a^k \cdot a$	$(k + 1)a = ka + a$
$a^{-k} = (a^{-1})^k$	$(-k)a = k(-a)$

The notation ka in additive notation does not represent a *product* of k and a but, rather, a *sum*

$$ka = a + a + \cdots + a$$

with k terms. In $0a = 0$, the zero on the left is the zero integer, and the zero on the right represents the additive identity in the group.

Considering the rich variety of operations and sets that have been involved in our examples, it may be surprising and reassuring to find, in the next theorem, that the familiar **laws of exponents** hold in a group.

Theorem 3.12 ■ Laws of Exponents

Let x and y be elements of the group G , and let m and n denote integers. Then

- a. $x^n \cdot x^{-n} = e$
- b. $x^m \cdot x^n = x^{m+n}$
- c. $(x^m)^n = x^{mn}$
- d. If G is abelian, $(xy)^n = x^n y^n$.

Induction

Proof The proof of each statement involves the use of mathematical induction. It would be redundant, and even boring, to include a complete proof of the theorem, so we shall assume statement **a** and prove **b** for the case where m is a positive integer. Even then, the argument is lengthy. The proofs of the statements **a**, **c**, and **d** are left as exercises.

Let m be an arbitrary, but fixed, positive integer. There are three cases to consider for n :

- i. $n = 0$
- ii. n a positive integer
- iii. n a negative integer.

First, let $n = 0$ for case i. Then

$$x^m \cdot x^n = x^m \cdot x^0 = x^m \cdot e = x^m \quad \text{and} \quad x^{m+n} = x^{m+0} = x^m.$$

Thus $x^m \cdot x^n = x^{m+n}$ in the case where $n = 0$.

Second, we shall use induction on n for case ii where n is a positive integer. If $n = 1$, we have

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

and statement **b** of the theorem holds when $n = 1$. Assume that **b** is true for $n = k$. That is, assume that

$$x^m \cdot x^k = x^{m+k}.$$

Then, for $n = k + 1$, we have

$$\begin{aligned} x^m \cdot x^n &= x^m \cdot x^{k+1} \\ &= x^m \cdot (x^k \cdot x) \quad \text{by definition of } x^{k+1} \\ &= (x^m \cdot x^k) \cdot x \quad \text{by associativity} \\ &= x^{m+k} \cdot x \quad \text{by the induction hypothesis} \\ &= x^{m+k+1} \quad \text{by definition of } x^{(m+k)+1} \\ &= x^{m+n} \quad \text{since } n = k + 1. \end{aligned}$$

Thus **b** is true for $n = k + 1$, and it follows that it is true for all positive integers n .

Third, consider case iii where n is a negative integer. This means that $n = -p$, where p is a positive integer. We consider three possibilities for p : $p = m$, $p < m$, and $m < p$.

If $p = m$, then $n = -p = -m$, and we have

$$x^m \cdot x^n = x^m \cdot x^{-m} = e$$

by statement **a** of the theorem, and

$$x^{m+n} = x^{m-m} = x^0 = e.$$

We have $x^m \cdot x^n = x^{m+n}$ when $p = m$.

If $p < m$, let $m - p = q$, so that $m = q + p$ where q and p are positive integers. We have already proved statement **b** when m and n are positive integers, so we may use $x^{q+p} = x^q \cdot x^p$. This gives

$$\begin{aligned} x^m \cdot x^n &= x^{q+p} \cdot x^{-p} \\ &= x^q \cdot x^p \cdot x^{-p} \\ &= x^q \cdot e \quad \text{by statement a} \\ &= x^q \\ &= x^{q+p-p} \\ &= x^{m+n}. \end{aligned}$$

That is, $x^m \cdot x^n = x^{m+n}$ for the case where $p < m$.

Finally, suppose that $m < p$. Let $r = p - m$, so that r is a positive integer and $p = m + r$. By the definition of x^{-p} ,

$$\begin{aligned} x^{-p} &= (x^{-1})^p \\ &= (x^{-1})^{m+r} \\ &= (x^{-1})^m \cdot (x^{-1})^r \quad \text{since } m \text{ and } r \text{ are positive integers} \\ &= x^{-m} \cdot x^{-r}. \end{aligned}$$

Substituting this value for x^{-p} in $x^m \cdot x^n = x^m \cdot x^{-p}$, we have

$$\begin{aligned}x^m \cdot x^n &= x^m \cdot (x^{-m} \cdot x^{-r}) \\&= (x^m \cdot x^{-m}) \cdot x^{-r} \\&= e \cdot x^{-r} \\&= x^{-r}.\end{aligned}$$

We also have

$$\begin{aligned}x^{m+n} &= x^{m-p} \\&= x^{m-(m+r)} \\&= x^{-r},\end{aligned}$$

so $x^m \cdot x^n = x^{m+n}$ when $m < p$.

We have proved that $x^m \cdot x^n = x^{m+n}$ in the cases where m is a positive integer and n is any integer (zero, positive, or negative). Of course, this is not a complete proof of statement **b** of the theorem. A complete proof would require considering cases where $m = 0$ or where m is a negative integer. The proofs for these cases are similar to those given here, and we omit them entirely.

The laws of exponents in Theorem 3.12 translate into the following **laws of multiples** for an additive group G .

Laws of Multiples

- a. $nx + (-n)x = 0$
- b. $mx + nx = (m + n)x$
- c. $n(mx) = (nm)x$
- d. If G is abelian, $n(x + y) = nx + ny$.

In connection with integral exponents, consider the following example.

Example 7 Let G be a group, let a be an element of G , and let H be the set of all elements of the form a^n , where n is an integer. That is,

$$H = \{x \in G \mid x = a^n \text{ for } n \in \mathbf{Z}\}.$$

Then H is nonempty and actually forms a subgroup of G . For if $x = a^m \in H$ and $y = a^n \in H$, then $xy = a^{m+n} \in H$ and $x^{-1} = a^{-m} \in H$. It follows from Theorem 3.9 that H is a subgroup. ■

Definition 3.13 ■ Cyclic Subgroup

Let G be a group. For any $a \in G$, the subgroup

$$H = \{x \in G \mid x = a^n \text{ for } n \in \mathbf{Z}\}$$

is the **subgroup generated by a** and is denoted by $\langle a \rangle$. A given subgroup K of G is a **cyclic subgroup** if there exists an element b in G such that

$$K = \langle b \rangle = \{y \in G \mid y = b^n \text{ for some } n \in \mathbf{Z}\}.$$

In particular, G is a **cyclic group** if there is an element $a \in G$ such that $G = \langle a \rangle$.

Example 8

- a. The set \mathbf{Z} of integers is a cyclic group under addition. We have $\mathbf{Z} = \langle 1 \rangle$ and $\mathbf{Z} = \langle -1 \rangle$.
- b. The subgroup $\mathbf{E} \subseteq \mathbf{Z}$ of all even integers is a cyclic subgroup of the additive group \mathbf{Z} , generated by 2. Hence $\mathbf{E} = \langle 2 \rangle$.
- c. In Example 6, we saw that

$$H = \{[2], [4], [6], [8]\} \subseteq \mathbf{Z}_{10}$$

is an abelian group with respect to multiplication. Since

$$[2]^2 = [4], \quad [2]^3 = [8], \quad [2]^4 = [6],$$

then

$$H = \langle [2] \rangle.$$

- d. The group $\mathcal{S}(A) = \{e, \rho, \rho^2, \sigma, \gamma, \delta\}$ of Example 3 in Section 3.1 is not a cyclic group. This can be verified by considering $\langle a \rangle$ for all possible choices of a in $\mathcal{S}(A)$. ■

Exercises 3.3

True or False

Label each of the following statements as either true or false, where H is a subgroup of G .

1. Every group G contains at least two subgroups.
 2. The identity element in a subgroup H of a group G must be the same as the identity element in G .
 3. An element x in H has an inverse x^{-1} in H that may be different than its inverse in G .
 4. The generator of a cyclic group is unique.
 5. Any subgroup of an abelian group is abelian.
 6. If a subgroup H of a group G is abelian, then G must be abelian.
 7. The relation R on the set of all groups defined by HRK if and only if H is a subgroup of K is an equivalence relation.
 8. The empty set \emptyset is a subgroup of any group G .
 9. Any group of order 3 has no nontrivial subgroups.
 10. \mathbf{Z}_5 under addition modulo 5 is a subgroup of the group \mathbf{Z} under addition.
-

Exercises

1. Let $\mathcal{S}(A) = \{e, \rho, \rho^2, \sigma, \gamma, \delta\}$ be as in Example 3 in Section 3.1. Decide whether each of the following subsets is a subgroup of $\mathcal{S}(A)$. If a set is not a subgroup, give a reason why it is not. (*Hint:* Construct a multiplication table for each subset.)
 - a. $\{e, \sigma\}$
 - b. $\{e, \delta\}$
 - c. $\{e, \rho\}$
 - d. $\{e, \rho^2\}$
 - e. $\{e, \rho, \rho^2\}$
 - f. $\{e, \rho, \sigma\}$
 - g. $\{e, \sigma, \gamma\}$
 - h. $\{e, \sigma, \gamma, \delta\}$
2. Decide whether each of the following sets is a subgroup of $G = \{1, -1, i, -i\}$ under multiplication. If a set is not a subgroup, give a reason why it is not.
 - a. $\{1, -1\}$
 - b. $\{1, i\}$
 - c. $\{i, -i\}$
 - d. $\{1, -i\}$
3. Consider the group \mathbf{Z}_{16} under addition. List all the elements of the subgroup $\langle [6] \rangle$, and state its order.
4. List all the elements of the subgroup $\langle [8] \rangle$ in the group \mathbf{Z}_{18} under addition, and state its order.
5. Assume that the nonzero elements of \mathbf{Z}_{13} form a group G under multiplication $[a][b] = [ab]$.
 - a. List the elements of the subgroup $\langle [4] \rangle$ of G , and state its order.
 - b. List the elements of the subgroup $\langle [8] \rangle$ of G , and state its order.
6. Let G be the group of all invertible matrices in $M_2(\mathbf{R})$ under multiplication. List the elements of the subgroup $\langle A \rangle$ of G for the given A , and give $o(\langle A \rangle)$.

a. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$	b. $A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$
c. $A = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$	d. $A = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}$
7. Let G be the group $M_2(\mathbf{Z}_5)$ under addition. List the elements of the subgroup $\langle A \rangle$ of G for the given A , and give $o(\langle A \rangle)$.

a. $A = \begin{bmatrix} [2] & [0] \\ [0] & [3] \end{bmatrix}$	b. $A = \begin{bmatrix} [0] & [1] \\ [2] & [4] \end{bmatrix}$
--	--
8. Find a subset of \mathbf{Z} that is closed under addition but is not a subgroup of the additive group \mathbf{Z} .
9. Let G be the group of all nonzero real numbers under multiplication. Find a subset of G that is closed under multiplication but is not a subgroup of G .
10. Let $n > 1$ be an integer, and let a be a fixed integer. Prove or disprove that the set

$$H = \{x \in \mathbf{Z} \mid ax \equiv 0 \pmod{n}\}$$

is a subgroup of \mathbf{Z} under addition.

11. Let H be a subgroup of G , let a be a fixed element of G , and let K be the set of all elements of the form aha^{-1} , where $h \in H$. That is,

$$K = \{x \in G \mid x = aha^{-1} \text{ for some } h \in H\}.$$

Sec. 4.4, #8 <

Prove or disprove that K is a subgroup of G .

12. Prove or disprove that $H = \{h \in G \mid h^{-1} = h\}$ is a subgroup of the group G if G is abelian.

13. Prove that each of the following subsets H of $M_2(\mathbf{Z})$ is a subgroup of the group $M_2(\mathbf{Z})$ under addition.

a. $H = \left\{ \begin{bmatrix} x & y \\ z & w \end{bmatrix} \mid w = 0 \right\}$

c. $H = \left\{ \begin{bmatrix} x & y \\ 0 & 0 \end{bmatrix} \mid x = y \right\}$

b. $H = \left\{ \begin{bmatrix} x & y \\ z & w \end{bmatrix} \mid z = w = 0 \right\}$

d. $H = \left\{ \begin{bmatrix} x & y \\ z & w \end{bmatrix} \mid x + y + z + w = 0 \right\}$

14. Prove that each of the following subsets H of $M_2(\mathbf{R})$ is a subgroup of the group G of all invertible matrices in $M_2(\mathbf{R})$ under multiplication.

a. $H = \left\{ \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \mid a \in \mathbf{R} \right\}$

c. $H = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \mid a^2 + b^2 \neq 0 \right\}$

b. $H = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \mid a^2 + b^2 = 1 \right\}$

d. $H = \left\{ \begin{bmatrix} 1 & a \\ 0 & b \end{bmatrix} \mid b \neq 0 \right\}$

e. $H = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a + c = 1, b + d = 1, \text{ and } ad - bc \neq 0 \right\}$

f. $H = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \mid a \neq 0, b \neq 0 \right\}$

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

Sec. 3.5, #9 <

Sec. 1.6, #28 >

15. Prove that each of the following sets H is a subgroup of the group G of all invertible matrices in $M_2(\mathbf{C})$ under multiplication.

a. $H = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$

b. $H = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$

16. Consider the set of matrices $H = \{I_2, M_1, M_2, M_3, M_4, M_5\}$, where

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_1 = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix},$$

$$M_3 = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}, \quad M_4 = \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}, \quad M_5 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Sec. 3.5, #5 <

Show that H is a subgroup of the multiplicative group of all invertible matrices in $M_2(\mathbf{R})$.

Sec. 4.6, #15 <

- Sec. 3.2, #9 >> 17. a. For any group G , the set of all elements that commute with every element of G is called the **center** of G and is denoted by $Z(G)$:

Sec. 3.4, #31 <

Sec. 4.5, #22 <

Sec. 7.2, #39 <

$$Z(G) = \{a \in G \mid ax = xa \text{ for every } x \in G\}.$$

Prove that $Z(G)$ is a subgroup of G .

- b. Let R be the equivalence relation on G defined by xRy if and only if there exists an element a in G such that $y = a^{-1}xa$. If $x \in Z(G)$, find $[x]$, the equivalence class containing x .

18. (See Exercise 17.) Find the center $Z(G)$ for each of the following groups G .

Sec. 3.1, #28 >

- a. $G = \{1, i, j, k, -1, -i, -j, -k\}$ in Exercise 28 of Section 3.1.

Sec. 3.1, #30 >

- b. $G = \{I_2, R, R^2, R^3, H, D, V, T\}$ in Exercise 30 of Section 3.1.

Sec. 3.1, #29 >

- c. $G = \{I_3, P_1, P_2, P_3, P_4, P_5\}$ in Exercise 29 of Section 3.1.

- d. G is the group of all invertible matrices in $M_2(\mathbf{R})$ under multiplication.

19. Let G be a group and $Z(G)$ its center. Prove or disprove that if ab is in $Z(G)$, then a and b are in $Z(G)$.

20. Let G be a group and $Z(G)$ its center. Prove or disprove that if ab is in $Z(G)$, then $ab = ba$.

21. Let A be a given nonempty set. As noted in Example 2 of Section 3.1, $\mathcal{S}(A)$ is a group with respect to mapping composition. For a fixed element a in A , let H_a denote the set of all $f \in \mathcal{S}(A)$ such that $f(a) = a$. Prove that H_a is a subgroup of $\mathcal{S}(A)$.

22. (See Exercise 21.) Let A be an infinite set, and let H be the set of all $f \in \mathcal{S}(A)$ such that $f(x) = x$ for all but a finite number of elements x of A . Prove that H is a subgroup of $\mathcal{S}(A)$.

23. For each $n \in \mathbf{Z}$, define $f_n: \mathbf{Z} \rightarrow \mathbf{Z}$ by $f_n(x) = x + n$ for $x \in \mathbf{Z}$.

- a. Show that f_n is an element of $\mathcal{S}(\mathbf{Z})$.

- b. Let $H = \{f_n \in \mathcal{S}(\mathbf{Z}) \mid f_n(x) = x + n \text{ for each } n \in \mathbf{Z}\}$. Prove that H is a subgroup of $\mathcal{S}(\mathbf{Z})$ under mapping composition.

- c. Prove that H is abelian, even though $\mathcal{S}(\mathbf{Z})$ is not.

24. Let G be an abelian group. For a fixed positive integer n , let

$$G_n = \{a \in G \mid a = x^n \text{ for some } x \in G\}.$$

Prove that G_n is a subgroup of G .

25. For fixed integers a and b , let

$$S = \{ax + by \mid x \in \mathbf{Z} \text{ and } y \in \mathbf{Z}\}.$$

Prove that S is a subgroup of \mathbf{Z} under addition. (A special form of this S is used in proving the existence of a greatest common divisor in Theorem 2.12.)

26. For a fixed element a of a group G , the set $C_a = \{x \in G \mid ax = xa\}$ is the **centralizer** of a in G . Prove that for any $a \in G$, C_a is a subgroup of G .

Sec. 4.1, #22 <

- 27.** Find the centralizer for each element a in each of the following groups.
- The quaternion group $G = \{1, i, j, k, -1, -i, -j, -k\}$ in Exercise 28 of Section 3.1
 - $G = \{I_2, R, R^2, R^3, H, D, V, T\}$ in Exercise 30 of Section 3.1
 - $G = \{I_3, P_1, P_2, P_3, P_4, P_5\}$ in Exercise 29 of Section 3.1
- 28.** Prove that $C_a = C_{a^{-1}}$, where C_a is the centralizer of a in the group G .
- 29.** Suppose that H_1 and H_2 are subgroups of the group G . Prove that $H_1 \cap H_2$ is a subgroup of G .
- 30.** For an arbitrary n in \mathbf{Z} , the cyclic subgroup $\langle n \rangle$ of \mathbf{Z} , generated by n under addition, is the set of all multiples of n . Describe the subgroup $\langle m \rangle \cap \langle n \rangle$ for arbitrary m and n in \mathbf{Z} .
- 31.** Let $\{H_\lambda\}$, $\lambda \in \mathcal{L}$, be an arbitrary nonempty collection of subgroups H_λ of the group G , and let $K = \bigcap_{\lambda \in \mathcal{L}} H_\lambda$. Prove that K is a subgroup of G .
- 32.** If G is a group, prove that $Z(G) = \bigcap_{a \in G} C_a$, where $Z(G)$ is the center of G and C_a is the centralizer of a in G .
- 33.** Find subgroups H and K of the group $\mathcal{S}(A)$ in Example 3 of Section 3.1 such that $H \cup K$ is *not* a subgroup of $\mathcal{S}(A)$.
- 34.** Assume that H and K are subgroups of the abelian group G . Prove that the set of products $HK = \{g \in G \mid g = hk \text{ for } h \in H \text{ and } k \in K\}$ is a subgroup of G .
- 35.** Find subgroups H and K of the group $\mathcal{S}(A)$ in Example 3 of Section 3.1 such that the set HK defined in Exercise 34 is not a subgroup of $\mathcal{S}(A)$.
- 36.** Let G be a cyclic group, $G = \langle a \rangle$. Prove that G is abelian.
- 37.** Prove statement **a** of Theorem 3.12: $x^n \cdot x^{-n} = e$ for all integers n .
- 38.** Prove statement **c** of Theorem 3.12: $(x^m)^n = x^{mn}$ for all integers m and n .
- 39.** Prove statement **d** of Theorem 3.12: If G is abelian, $(xy)^n = x^n y^n$ for all integers n .
- 40.** Suppose that H is a nonempty subset of a group G . Prove that H is a subgroup of G if and only if $a^{-1}b \in H$ for all $a \in H$ and $b \in H$.
- 41.** Assume that G is a finite group, and let H be a nonempty subset of G . Prove that H is closed if and only if H is a subgroup of G .

3.4

Cyclic Groups

In the last section a group G was defined to be *cyclic* if there exists an element $a \in G$ such that $G = \langle a \rangle$. It may happen that there is more than one element $a \in G$ such that $G = \langle a \rangle$. For the additive group \mathbf{Z} , we have $\mathbf{Z} = \langle 1 \rangle$ and also $\mathbf{Z} = \langle -1 \rangle$, since any $n \in \mathbf{Z}$ can be written as $(-n)(-1)$. Here $(-n)(-1)$ does not indicate a product but rather a multiple of -1 , as described in Section 3.3.

Definition 3.14 ■ Generator

Any element a of the group G such that $G = \langle a \rangle$ is a **generator** of G .

If a is a generator of G , then a^{-1} is also, since any element $x \in G$ can be written as

$$x = a^n = (a^{-1})^{-n}$$

for some integer n .

Example 1

The additive group

$$\mathbf{Z}_n = \{[0], [1], \dots, [n - 1]\}$$

is a cyclic group with generator $[1]$, since any $[k]$ in \mathbf{Z}_n can be written as

$$[k] = k[1]$$

where $k[1]$ indicates a multiple of $[1]$ as described in Section 3.3. Elements other than $[1]$ may also be generators. To illustrate this, consider the particular case

$$\mathbf{Z}_6 = \{[0], [1], [2], [3], [4], [5]\}.$$

The element $[5]$ is also a generator of \mathbf{Z}_6 since $[5]$ is the additive inverse of $[1]$. The following list shows how \mathbf{Z}_6 is generated by $[5]$ —that is, how \mathbf{Z}_6 consists of multiples of $[5]$.

$$\begin{aligned} 1[5] &= [5] \\ 2[5] &= [5] + [5] = [4] \\ 3[5] &= [5] + [5] + [5] = [3] \\ 4[5] &= [2] \\ 5[5] &= [1] \\ 6[5] &= [0] \end{aligned}$$

The cyclic subgroups generated by the other elements of \mathbf{Z}_6 under addition are

$$\begin{aligned} \langle[0]\rangle &= \{[0]\} \\ \langle[2]\rangle &= \{[2], [4], [0]\} \\ \langle[3]\rangle &= \{[3], [0]\} \\ \langle[4]\rangle &= \{[4], [2], [0]\} = \langle[2]\rangle. \end{aligned}$$

Thus $[1]$ and $[5]$ are the only elements that are generators of the entire group. ■

Example 2

We saw in Example 8 of Section 3.3 that

$$H = \{[2], [4], [6], [8]\} \subseteq \mathbf{Z}_{10}$$

forms a cyclic group with respect to multiplication and that $[2]$ is a generator of H . The element $[8] = [2]^{-1}$ is also a generator of H , as the following computations confirm:

$$[8]^2 = [4], \quad [8]^3 = [2], \quad [8]^4 = [6].$$

Example 3

In the quaternion group $G = \{\pm 1, \pm i, \pm j, \pm k\}$, described in Exercise 28 of Section 3.1, we have

$$\begin{aligned} i^2 &= -1 \\ i^3 &= i^2 \cdot i = -i \\ i^4 &= i^3 \cdot i = -i^2 = 1. \end{aligned}$$

Thus i generates the cyclic subgroup of order 4 given by

$$\langle i \rangle = \{i, -1, -i, 1\},$$

although the group G itself is not cyclic. ■

Whether a group G is cyclic or not, each element a of G generates the cyclic subgroup $\langle a \rangle$, and

$$\langle a \rangle = \{x \in G \mid x = a^n \text{ for } n \in \mathbf{Z}\}.$$

We shall see that the structure of $\langle a \rangle$ depends entirely on whether or not $a^n = e$ for some positive integer n . The next two theorems state the possibilities for the structure of $\langle a \rangle$.

Strategy ■ The method of proof of the next theorem is by contradiction. A statement $p \Rightarrow q$ may be proved by assuming that p is true and q is false and then proving that this assumption leads to a situation where some statement is both true and false—a contradiction.

Theorem 3.15 ■ Infinite Cyclic Group

Let a be an element in the group G . If $a^n \neq e$ for every positive integer n , then $a^p \neq a^q$ whenever $p \neq q$ in \mathbf{Z} , and $\langle a \rangle$ is an infinite cyclic group.

Contradiction **Proof** Assume that a is an element of the group G such that $a^n \neq e$ for every positive integer n . Having made this assumption, suppose now that

$$\begin{aligned} & (p \wedge \sim q) \\ & \Rightarrow \sim p \end{aligned}$$

$$a^p = a^q$$

where $p \neq q$ in \mathbf{Z} . We may assume that $p > q$. Then

$$\begin{aligned} a^p = a^q & \Rightarrow a^p \cdot a^{-q} = a^q \cdot a^{-q} \\ & \Rightarrow a^{p-q} = e. \end{aligned}$$

Since $p - q$ is a positive integer, this result contradicts $a^n \neq e$ for every positive integer n . Therefore, it must be that $a^p \neq a^q$ whenever $p \neq q$. Thus all powers of a are distinct, and therefore $\langle a \rangle$ is an infinite cyclic group.

Corollary 3.16 ■

If G is a finite group and $a \in G$, then $a^n = e$ for some positive integer n .

$p \Rightarrow q$ **Proof** Suppose G is a finite group and $a \in G$. Since the cyclic subgroup

$$\langle a \rangle = \{x \in G \mid x = a^m \text{ for } m \in \mathbf{Z}\}$$

is a subset of G , $\langle a \rangle$ must also be finite. It must therefore happen that $a^p = a^q$ for some integers p and q with $p \neq q$. It follows from Theorem 3.15 that $a^n = e$ for some positive integer n .

If it happens that $a^n \neq e$ for every positive integer n , then Theorem 3.15 states that all the powers of a are distinct and that $\langle a \rangle$ is an infinite group. Of course, it may happen that $a^n = e$ for some positive integers n . In this case, Theorem 3.17 describes $\langle a \rangle$ completely.

Theorem 3.17 ■ Finite Cyclic Group

Let a be an element in a group G , and suppose $a^n = e$ for some positive integer n . If m is the least positive integer such that $a^m = e$, then

- a. $\langle a \rangle$ has order m , and $\langle a \rangle = \{a^0 = e = a^m, a^1, a^2, \dots, a^{m-1}\}$
- b. $a^s = a^t$ if and only if $s \equiv t \pmod{m}$.

$p \Rightarrow q$ **Proof** Assume that m is the least positive integer such that $a^m = e$. We first show that the elements

$$a^0 = e, a, a^2, \dots, a^{m-1}$$

are all distinct. Suppose

$$a^i = a^j \quad \text{where } 0 \leq i < m \quad \text{and} \quad 0 \leq j < m.$$

There is no loss of generality in assuming $i \geq j$. Then $a^i = a^j$ implies

$$a^{i-j} = a^i \cdot a^{-j} = e \quad \text{where } 0 \leq i - j < m.$$

Since m is the least positive integer such that $a^m = e$, and since $i - j < m$, it must be true that $i - j = 0$, and therefore $i = j$. Thus $\langle a \rangle$ contains the m distinct elements $a^0 = e, a, a^2, \dots, a^{m-1}$. The proof of part **a** will be complete if we can show that any power of a is equal to one of these elements. Consider an arbitrary a^k . By the Division Algorithm, there exist integers q and r such that

$$k = mq + r, \quad \text{with } 0 \leq r < m.$$

Thus

$$\begin{aligned} a^k &= a^{mq+r} \\ &= a^{mq} \cdot a^r \quad \text{by part b of Theorem 3.12} \\ &= (a^m)^q \cdot a^r \quad \text{by part c of Theorem 3.12} \\ &= e^q \cdot a^r \\ &= a^r \end{aligned}$$

where r is in the set $\{0, 1, 2, \dots, m - 1\}$. It follows that

$$\langle a \rangle = \{e, a, a^2, \dots, a^{m-1}\}, \quad \text{and } \langle a \rangle \text{ has order } m.$$

$p \Rightarrow (q \Leftrightarrow r)$

To obtain part **b**, we first observe that if $k = mq + r$, with $0 \leq r < m$, then $a^k = a^r$, where r is in the set $\{0, 1, 2, \dots, m - 1\}$. In particular, $a^k = e$ if and only if $r = 0$ —that is, if and only if $k \equiv 0 \pmod{m}$. Thus

$$\begin{aligned} a^s = a^t &\Leftrightarrow a^{s-t} = e \\ &\Leftrightarrow s - t \equiv 0 \pmod{m} \\ &\Leftrightarrow s \equiv t \pmod{m}, \end{aligned}$$

and the proof is complete.

We have defined the order $o(G)$ of a group G to be the number of elements in the group.

Definition 3.18 ■ Order of an Element

The **order** $o(a)$ of an element a of the group G is the order of the subgroup generated by a . That is, $o(a) = o(\langle a \rangle)$.

Part **a** of Theorem 3.17 immediately translates into the following corollary.

Corollary 3.19 ■ Finite Order of an Element

If $o(a)$ is finite, then $m = o(a)$ is the least positive integer such that $a^m = e$.

The next example illustrates the results of Theorem 3.17 and its corollary.

Example 4 It can be shown (see Exercise 16 at the end of this section) that

$$G = \{[1], [3], [5], [7], [9], [11], [13], [15]\} \subseteq \mathbf{Z}_{16}$$

is a group with respect to multiplication in \mathbf{Z}_{16} . The element $[3]$ of G generates a cyclic subgroup of order 4 since $[3]^4 = [1]$, and 4 is the least positive integer m such that $[3]^m = [1]$. Thus

$$\langle [3] \rangle = \{[3]^0 = [1], [3], [9], [11]\},$$

and the order of the element $[3]$ is 4. Also, powers larger than 4 of $[3]$ are easily computed using part **b** of Theorem 3.17. For example,

$$[3]^{191} = [3]^3 = [11]$$

since $191 \equiv 3 \pmod{4}$. ■

The multiplicative group $G = \{[1], [3], [5], [7], [9], [11], [13], [15]\} \subseteq \mathbf{Z}_{16}$ in Example 4 consists of all $[a]$ in \mathbf{Z}_{16} that have multiplicative inverses. This group is called the **group of units** in \mathbf{Z}_{16} and is designated by the symbol \mathbf{U}_{16} .

As might be expected, every subgroup of a cyclic group is also a cyclic group. It is even possible to predict a generator of the subgroup, as stated in Theorem 3.20.

Strategy ■ The conclusion of the next theorem has the form “either a or b .” To prove this statement, we can assume that a is false and prove that b must then be true.

Theorem 3.20 ■ Subgroup of a Cyclic Group

Let G be a cyclic group with $a \in G$ as a generator, and let H be a subgroup of G . Then either

- a. $H = \{e\} = \langle e \rangle$, or
- b. if $H \neq \{e\}$, then $H = \langle a^k \rangle$ where k is the least positive integer such that $a^k \in H$.

$(p \wedge q \wedge \sim r) \Rightarrow s$ **Proof** Let $G = \langle a \rangle$, and suppose H is a subgroup and $H \neq \{e\}$. Then H contains an element of the form a^j with $j \neq 0$. Since H contains inverses and $(a^j)^{-1} = a^{-j}$, both a^j and a^{-j} are in H . Thus H contains positive powers of a . Let k be the least positive integer such that $a^k \in H$.

Since H is closed and contains inverses, and since $a^k \in H$, all powers $(a^k)^t = a^{kt}$ are in H . We need to show that any element of H is a power of a^k . Let $a^n \in H$. There are integers q and r such that

$$n = kq + r \quad \text{with } 0 \leq r < k.$$

Now $a^{-kq} = (a^k)^{-q} \in H$ and $a^n \in H$ imply that

$$a^n \cdot a^{-kq} = a^{kq+r} \cdot a^{-kq} = a^r$$

is in H . Since $0 \leq r < k$ and k is the least positive integer such that $a^k \in H$, r must be zero and $a^n = a^{kq}$. Thus $H = \langle a^k \rangle$.

Corollary 3.21 ■

Any subgroup of a cyclic group is cyclic.

Note that Theorem 3.20 and Corollary 3.21 apply to infinite cyclic groups as well as to finite ones. The next theorem, however, applies only to finite groups.

Strategy In the proof of Theorem 3.22, we use the standard technique to prove that two sets A and B are equal: We show that $A \subseteq B$ and then that $B \subseteq A$.

Theorem 3.22 ■ Generators of Subgroups

Let G be a finite cyclic group of order n with $a \in G$ as a generator. For any integer m , the subgroup generated by a^m is the same as the subgroup generated by a^d , where $d = (m, n)$.

$p \Rightarrow q$ **Proof** Let $d = (m, n)$, and let $m = dp$. Since $a^m = a^{dp} = (a^d)^p$, then a^m is in $\langle a^d \rangle$, and therefore $\langle a^m \rangle \subseteq \langle a^d \rangle$. (See Exercise 27 at the end of this section.)

Similarly, to show that $\langle a^d \rangle \subseteq \langle a^m \rangle$, it is sufficient to show that a^d is in $\langle a^m \rangle$. By Theorem 2.12, there exist integers x and y such that

$$d = mx + ny.$$

Since a is a generator of G and $o(G) = n$, $a^n = e$. Using this fact, we have

$$\begin{aligned} a^d &= a^{mx+ny} \\ &= a^{mx} \cdot a^{ny} \\ &= (a^m)^x \cdot (a^n)^y \\ &= (a^m)^x \cdot (e)^y \\ &= (a^m)^x. \end{aligned}$$

Thus a^d is in $\langle a^m \rangle$, and the proof of the theorem is complete.

As an immediate corollary to Theorem 3.22, we have the following result.

Corollary 3.23 ■ Distinct Subgroups of a Finite Cyclic Group

Let G be a finite cyclic group of order n with $a \in G$ as a generator. The distinct subgroups of G are those subgroups $\langle a^d \rangle$ where d is a positive divisor of n .

Corollary 3.23 provides a systematic way to obtain all the subgroups of a cyclic group of order n . In the subgroup generated by a^d , the exponent d divides n , the order of G . Then there is a positive integer k such that $n = dk$ and $\langle a^d \rangle = \{a^d, a^{2d}, a^{3d}, \dots, a^{kd} = a^n = e\}$. Thus the order of $\langle a^d \rangle$ is k , and $o(\langle a^d \rangle) | o(G)$.

Example 5 Let $G = \langle a \rangle$ be a cyclic group of order 12. The divisors of 12 are 1, 2, 3, 4, 6, and 12, so the distinct subgroups of G are

$$\begin{aligned}\langle a \rangle &= G \\ \langle a^2 \rangle &= \{a^2, a^4, a^6, a^8, a^{10}, a^{12} = e\} \\ \langle a^3 \rangle &= \{a^3, a^6, a^9, a^{12} = e\} \\ \langle a^4 \rangle &= \{a^4, a^8, a^{12} = e\} \\ \langle a^6 \rangle &= \{a^6, a^{12} = e\} \\ \langle a^{12} \rangle &= \langle e \rangle = \{e\}.\end{aligned}$$

Thus Corollary 3.23 makes it easy to list all the distinct subgroups of a cyclic group. Theorem 3.22 itself makes it easy to determine which subgroup is generated by each element of the group. For our cyclic group of order 12,

$$\begin{aligned}\langle a^5 \rangle &= \langle a \rangle = G \quad \text{since } (5, 12) = 1 \\ \langle a^7 \rangle &= \langle a \rangle = G \quad \text{since } (7, 12) = 1 \\ \langle a^8 \rangle &= \langle a^4 \rangle \quad \text{since } (8, 12) = 4 \\ \langle a^9 \rangle &= \langle a^3 \rangle \quad \text{since } (9, 12) = 3 \\ \langle a^{10} \rangle &= \langle a^2 \rangle \quad \text{since } (10, 12) = 2 \\ \langle a^{11} \rangle &= \langle a \rangle = G \quad \text{since } (11, 12) = 1.\end{aligned}$$
■

The results in Example 5 lead us to a method for finding all generators of a finite cyclic group. This method is described in the next theorem.

Theorem 3.24 ■ Generators of a Finite Cyclic Group

Let $G = \langle a \rangle$ be a cyclic group of order n . Then a^m is a generator of G if and only if m and n are relatively prime.

$p \Leftarrow q$ **Proof** On the one hand, if m is such that m and n are relatively prime, then $d = (m, n) = 1$, and a^m is a generator of G by Theorem 3.22.

$p \Rightarrow q$ On the other hand, if a^m is a generator of G , then $a = (a^m)^p$ for some integer p . By part **b** of Theorem 3.17, this implies that $1 \equiv mp \pmod{n}$. That is,

$$1 - mp = nq$$

for some integer q . This gives

$$1 = mp + nq,$$

and it follows from Theorem 2.12 that $(m, n) = 1$.

The Euler phi-function $\phi(n)$ was defined for positive integers n in Exercise 23 of Section 2.8 as follows: $\phi(n)$ is the number of positive integers m such that $1 \leq m \leq n$ and $(m, n) = 1$. It follows, from Theorems 3.17 and 3.24, that the cyclic group $\langle a \rangle$ of order n has $\phi(n)$ distinct generators.

Example 6 Let $G = \langle a \rangle$ be a cyclic group of order 10. The positive integers less than 10 and relatively prime to 10 are 1, 3, 7, and 9. Therefore, all generators of G are included in the list

$$a, \quad a^3, \quad a^7, \quad \text{and} \quad a^9,$$

and G has $\phi(10) = 4$ distinct generators. ■

Example 7 Some other explicit uses of Theorem 3.24 can be demonstrated by using \mathbf{Z}_7 .

The generators of the additive group \mathbf{Z}_7 are those $[a]$ in \mathbf{Z}_7 such that a and 7 are relatively prime, and this includes all nonzero $[a]$. Thus every element of \mathbf{Z}_7 , except $[0]$, generates \mathbf{Z}_7 under addition.

The situation is quite different when we consider the group G of nonzero elements of \mathbf{Z}_7 under multiplication. It is easy to verify that $[3]$ is a generator:

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

According to Theorem 3.24, the only other generator of G is $[3]^5 = [5]$, since 2, 3, 4, and 6 are not relatively prime to 6. ■

Exercises 3.4

True or False

Label each of the following statements as either true or false.

1. The order of the identity element in any group is 1.
2. Every cyclic group is abelian.
3. Every abelian group is cyclic.
4. If a subgroup H of a group G is cyclic, then G must be cyclic.
5. Whether a group G is cyclic or not, each element a of G generates a cyclic subgroup.

6. Every subgroup of a cyclic group is cyclic.
 7. If there exists an $m \in \mathbf{Z}^+$ such that $a^m = e$, where a is an element of a group G , then $o(a) = m$.
 8. Any group of order 3 must be cyclic.
 9. Any group of order 4 must be cyclic.
 10. Let a be an element of a group G . Then $\langle a \rangle = \langle a^{-1} \rangle$.
-

Exercises

1. List all cyclic subgroups of the group $S(A)$ in Example 3 of Section 3.1.

Sec. 3.1, #28 >

2. Let $G = \{\pm 1, \pm i, \pm j, \pm k\}$ be the quaternion group. List all cyclic subgroups of G .

3. Find the order of each element of the group $S(A)$ in Example 3 of Section 3.1.

4. Find the order of each element of the group G in Exercise 2.

Sec. 3.1, #29 >

5. The elements of the multiplicative group G of 3×3 permutation matrices are given in Exercise 29 of Section 3.1. Find the order of each element of the group.

6. In the multiplicative group of invertible matrices in $M_4(\mathbf{R})$, find the order of the given element A .

$$\mathbf{a. } A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{b. } A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

7. Let a be an element of order 8 in a group G . Find the order of each of the following.

a. a^2 **b.** a^3 **c.** a^4 **d.** a^5 **e.** a^6 **f.** a^7 **g.** a^8

8. Let a be an element of order 9 in a group G . Find the order of each of the following.

a. a^2 **b.** a^3 **c.** a^4 **d.** a^5 **e.** a^6 **f.** a^7 **g.** a^8 **h.** a^9

9. For each of the following values of n , find all distinct generators of the cyclic group \mathbf{Z}_n under addition.

a. $n = 8$ **b.** $n = 12$ **c.** $n = 10$

d. $n = 15$ **e.** $n = 16$ **f.** $n = 18$

10. For each of the following values of n , find all subgroups of the cyclic group \mathbf{Z}_n under addition and state their order.

a. $n = 12$ **b.** $n = 8$ **c.** $n = 10$

d. $n = 15$ **e.** $n = 16$ **f.** $n = 18$

Sec. 3.1, #27 >

11. According to Exercise 27 of Section 3.1, the nonzero elements of \mathbf{Z}_n form a group G with respect to multiplication if n is a prime. For each of the following values of n , show that this group G is cyclic.

a. $n = 7$ **b.** $n = 5$ **c.** $n = 11$

d. $n = 13$ **e.** $n = 17$ **f.** $n = 19$

12. For each of the following values of n , find all distinct generators of the group G described in Exercise 11.
- a. $n = 7$ b. $n = 5$ c. $n = 11$
 d. $n = 13$ e. $n = 17$ f. $n = 19$
13. For each of the following values of n , find all subgroups of the group G described in Exercise 11, and state their order.
- a. $n = 7$ b. $n = 5$ c. $n = 11$
 d. $n = 13$ e. $n = 17$ f. $n = 19$
14. Prove that the set

$$H = \left\{ \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} \mid n \in \mathbf{Z} \right\}$$

Sec. 3.5, #7 <

is a cyclic subgroup of the group of all invertible matrices in $M_2(\mathbf{R})$.

15. a. Use trigonometric identities and mathematical induction to prove that

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}^n = \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}$$

for all integers n (positive, zero, or negative). Hence conclude that for a constant θ , the set

$$H = \left\{ \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix} \mid n \in \mathbf{Z} \right\}$$

is a cyclic subgroup of the group of all invertible matrices in $M_2(\mathbf{R})$.

- b. Evaluate each element of H for $\theta = 90^\circ$.

- c. Evaluate each element of H for $\theta = 120^\circ$.

16. For an integer $n > 1$, let $G = \mathbf{U}_n$, the group of units in \mathbf{Z}_n ; that is, the set of all $[a]$ in \mathbf{Z}_n that have multiplicative inverses. Prove that \mathbf{U}_n is a group with respect to multiplication.

Sec. 3.5, #3, 6 <

17. Let \mathbf{U}_n be the group of units as described in Exercise 16. Prove that $[a] \in \mathbf{U}_n$ if and only if a and n are relatively prime.

18. Let \mathbf{U}_n be the group of units as described in Exercise 16. For each value of n , write out the elements of \mathbf{U}_n and construct a multiplication table for \mathbf{U}_n .

Sec. 4.6, #7, 12 <

- a. $n = 20$ b. $n = 8$ c. $n = 24$ d. $n = 30$

19. Which of the groups in Exercise 18 are cyclic?

20. Consider the group \mathbf{U}_9 of all units in \mathbf{Z}_9 . Given that \mathbf{U}_9 is a cyclic group under multiplication, find all subgroups of \mathbf{U}_9 .

21. Suppose $G = \langle a \rangle$ is a cyclic group of order n . Determine the number of generators of G for each value of n and list all the distinct generators of G .

- a. $n = 8$ b. $n = 14$ c. $n = 18$
 d. $n = 24$ e. $n = 7$ f. $n = 13$

22. List all the distinct subgroups of each group in Exercise 21.

- 23.** Let $G = \langle a \rangle$ be a cyclic group of order 24. List all elements having each of the following orders in G .
- a. 2 b. 3 c. 4 d. 10
- 24.** Let $G = \langle a \rangle$ be a cyclic group of order 35. List all elements having each of the following orders in G .
- a. 2 b. 5 c. 7 d. 10
- 25.** Describe all subgroups of the group \mathbf{Z} under addition.
- 26.** Find all generators of an infinite cyclic group $G = \langle a \rangle$.
- 27.** Let a and b be elements of the group G . Prove that if $a \in \langle b \rangle$, then $\langle a \rangle \subseteq \langle b \rangle$.
- 28.** Let a and b be elements of a finite group G .
- Prove that a and a^{-1} have the same order.
 - Prove that a and bab^{-1} have the same order.
 - Prove that ab and ba have the same order.
- 29.** Let G be a group and define the relation R on G by aRb if and only if a and b have the same order. Prove that R is an equivalence relation.
- 30.** Prove that a subset H of a finite group G is a subgroup of G if and only if
- H is nonempty, and
 - $a \in H$ and $b \in H$ imply $ab \in H$.
- (Hint: Use Corollary 3.16.)

Sec. 3.3, #17 >

- 31.** In Exercise 17 of Section 3.3, the center $Z(G)$ is defined as

$$Z(G) = \{a \in G \mid ax = xa \text{ for every } x \in G\}.$$

Prove that if b is the only element of order 2 in G , then $b \in Z(G)$.

Sec. 4.6, #23 <

- 32.** If a is an element of order m in a group G and $a^k = e$, prove that m divides k .
- 33.** If G is a cyclic group, prove that the equation $x^2 = e$ has at most two distinct solutions in G .
- 34.** Let G be a finite cyclic group of order n . If d is a positive divisor of n , prove that the equation $x^d = e$ has exactly d distinct solutions in G .
- 35.** If G is a cyclic group of order p and p is a prime, how many elements in G are generators of G ?
- 36.** Suppose that a and b are elements of finite order in a group such that $ab = ba$ and $\langle a \rangle \cap \langle b \rangle = \{e\}$. Prove that $o(ab)$ is the least common multiple of $o(a)$ and $o(b)$.
- 37.** Suppose that a is an element of order m in a group G , and k is an integer. If $d = (k, m)$, prove that a^k has order m/d .
- 38.** Assume that $G = \langle a \rangle$ is a cyclic group of order n . Prove that if r divides n , then G has a subgroup of order r .
- 39.** Suppose a is an element of order mn in a group G , where m and n are relatively prime. Prove that a is the product of an element of order m and an element of order n .

Sec. 4.1, #12, 13 <

40. Prove or disprove: If every nontrivial subgroup of the group G is cyclic, then G is a cyclic group.

Sec. 4.5, #5 <
Sec. 4.6, #23 <

41. Let G be an abelian group. Prove that the set of all elements of finite order in G forms a subgroup of G . This subgroup is called the **torsion subgroup** of G .

Sec. 2.8, #23 >
42. Let d be a positive integer and $\phi(d)$ the Euler phi-function. Use Corollary 3.23 and the additive groups \mathbf{Z}_d to show that

$$n = \sum_{d|n} \phi(d)$$

where the sum has one term for each positive divisor d of n .

3.5 Isomorphisms

It turns out that the permutation groups can serve as models for all groups. For this reason, we examine permutation groups in great detail in the next chapter. In order to describe their relation to groups in general, we need the concept of an *isomorphism*. Before formally introducing this concept, however, we consider some examples.

Example 1 Consider a cyclic group of order 4. If G is a cyclic group of order 4, it must contain an identity element e and a generator $a \neq e$ in G . The proof of Theorem 3.17 shows that

$$G = \{e, a, a^2, a^3\}$$

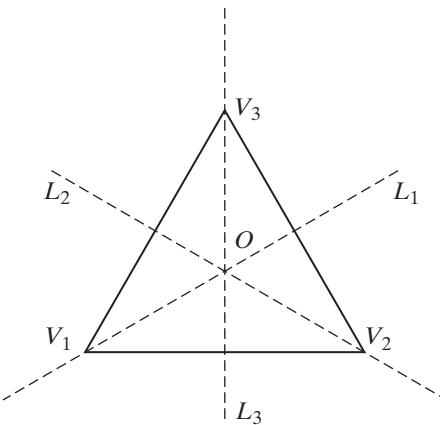
where $a^4 = e$. A multiplication table for G would have the form shown in Figure 3.14.

•	e	a	a^2	a^3
e	e	a	a^2	a^3
a	a	a^2	a^3	e
a^2	a^2	a^3	e	a
a^3	a^3	e	a	a^2

■ Figure 3.14

In a very definite way, then, the structure of G is determined. The details as to what the element a might be and what the operation in G might be may vary, but the basic structure of G fits the pattern in the table. ■

Example 2 Let us consider a group related to geometry. We begin with an equilateral triangle T with center point O and vertices labeled V_1 , V_2 , and V_3 (see Figure 3.15).



■ Figure 3.15

The equilateral triangle, of course, consists of the set of all points on the three sides of the triangle. By a **rigid motion** of the triangle, we mean a bijection of the set of points of the triangle onto itself that leaves the distance between any two points unchanged. In other words, a rigid motion of the triangle is a bijection that preserves distances. Such a rigid motion must map a vertex onto a vertex, and the entire mapping is determined by the images of the vertices V_1 , V_2 , and V_3 . These rigid motions (or **symmetries**, as they are often called) form a group with respect to mapping composition. (Verify this.) There are a total of six elements in the group, and they may be described as follows:

1. e , the identity mapping, that leaves all points unchanged.
2. r , a counterclockwise rotation through 120° about O in the plane of the triangle.
3. $r^2 = r \circ r$, a counterclockwise rotation through 240° about O in the plane of the triangle.
4. A reflection f about the line L_1 through V_1 and O .
5. A reflection g about the line L_2 through V_2 and O .
6. A reflection h about the line L_3 through V_3 and O .

These rigid motions can be described by indicating their values at the vertices as follows:

$$e: \begin{cases} e(V_1) = V_1 \\ e(V_2) = V_2 \\ e(V_3) = V_3 \end{cases} \quad h: \begin{cases} h(V_1) = V_2 \\ h(V_2) = V_1 \\ h(V_3) = V_3 \end{cases}$$

$$r: \begin{cases} r(V_1) = V_2 \\ r(V_2) = V_3 \\ r(V_3) = V_1 \end{cases} \quad g: \begin{cases} g(V_1) = V_3 \\ g(V_2) = V_2 \\ g(V_3) = V_1 \end{cases}$$

$$r^2: \begin{cases} r^2(V_1) = V_3 \\ r^2(V_2) = V_1 \\ r^2(V_3) = V_2 \end{cases} \quad f: \begin{cases} f(V_1) = V_1 \\ f(V_2) = V_3 \\ f(V_3) = V_2 \end{cases}$$

We have a group

$$G = \{e, r, r^2, h, g, f\},$$

and G has the multiplication table shown in Figure 3.16.

\circ	e	r	r^2	h	g	f
e	e	r	r^2	h	g	f
r	r	r^2	e	g	f	h
r^2	r^2	e	r	f	h	g
h	h	f	g	e	r^2	r
g	g	h	f	r	e	r^2
f	f	g	h	r^2	r	e

■ Figure 3.16

We shall compare this group G with the group $S(A)$ from Example 3 of Section 3.1, and we shall see that they are the same except for notation. Let the elements of G correspond to those of $S(A)$ according to the mapping $\phi: G \rightarrow S(A)$ given by

$$\begin{aligned}\phi(e) &= I_A & \phi(h) &= \sigma \\ \phi(r) &= \rho & \phi(g) &= \gamma \\ \phi(r^2) &= \rho^2 & \phi(f) &= \delta.\end{aligned}$$

This mapping is a one-to-one correspondence from G to $S(A)$. Moreover, ϕ has the property that

$$\phi(xy) = \phi(x) \cdot \phi(y)$$

for all x and y in G . This statement can be verified by using the multiplication tables for G and $S(A)$ in the following manner: In the entire multiplication table for G , we replace each element $x \in G$ by its image $\phi(x)$ in $S(A)$. This yields the table in Figure 3.17, which has $\phi(xy)$ in the row with $\phi(x)$ at the left and in the column with $\phi(y)$ at the top.

	I_A	ρ	ρ^2	σ	γ	δ
I_A	I_A	ρ	ρ^2	σ	γ	δ
ρ	ρ	ρ^2	I_A	γ	δ	σ
ρ^2	ρ^2	I_A	ρ	δ	σ	γ
σ	σ	δ	γ	I_A	ρ^2	ρ
γ	γ	σ	δ	ρ	I_A	ρ^2
δ	δ	γ	σ	ρ^2	ρ	I_A

■ Figure 3.17

The multiplication table for $S(A)$ given in Example 3 of Section 3.1 furnishes a table of values for $\phi(x) \cdot \phi(y)$, and the two tables agree in every position.[†] This means that $\phi(xy) = \phi(x) \cdot \phi(y)$ for all x and y in G . Thus G and $S(A)$ are the same except for notation. ■

A mapping such as ϕ in the preceding example is called an *isomorphism*.

Definition 3.25 ■ Isomorphism, Automorphism

Let G be a group with respect to \circledast , and let G' be a group with respect to \boxtimes . A mapping $\phi: G \rightarrow G'$ is an **isomorphism** from G to G' if

1. ϕ is a one-to-one correspondence from G to G' , and
2. $\phi(x \circledast y) = \phi(x) \boxtimes \phi(y)$ for all x and y in G .

If an isomorphism from G to G' exists, we say that G is **isomorphic** to G' , and we use the notation $G \cong G'$ as shorthand for this phrase. An isomorphism from a group G to G itself is called an **automorphism** of G .

The use of \circledast and \boxtimes in Definition 3.25 is intended to emphasize the fact that the group operations may be different. Now that this point has been made, we revert to our convention of using the multiplicative notation for the group operation. An isomorphism is said to “preserve the operation,” since condition 2 of Definition 3.25 requires that the result be the same whether the group operation is performed before or after the mapping.

The notation \cong in Definition 3.25 is not standardized. The notations \simeq , \approx , and \approx are used for the same purpose in some other texts.

Because an isomorphism preserves the group operation between two groups, it is not surprising that the identity elements always correspond under an isomorphism and that inverses are always mapped onto inverses. These results are stated more precisely in the next theorem.

Theorem 3.26 ■ Images of Identities and Inverses

Suppose ϕ is an isomorphism from the group G to the group G' . If e denotes the identity in G and e' denotes the identity in G' , then

- a. $\phi(e) = e'$, and
- b. $\phi(x^{-1}) = [\phi(x)]^{-1}$ for all x in G .

$p \Rightarrow q$ **Proof** We have

$$\begin{aligned} e \cdot e = e &\Rightarrow \phi(e \cdot e) = \phi(e) \\ &\Rightarrow \phi(e) \cdot \phi(e) = \phi(e) \quad \text{since } \phi \text{ is an isomorphism} \\ &\Rightarrow \phi(e) \cdot \phi(e) = \phi(e) \cdot e' \quad \text{since } e' \text{ is an identity} \\ &\Rightarrow \phi(e) = e' \quad \text{by Theorem 3.4e.} \end{aligned}$$

[†]Note that the e in Example 3 of Section 3.1 stands for I_A .

$$(p \wedge q) \Rightarrow r \quad \text{For any } x \text{ in } G,$$

$$\begin{aligned} x \cdot x^{-1} &= e \Rightarrow \phi(x \cdot x^{-1}) = \phi(e) \\ &\Rightarrow \phi(x \cdot x^{-1}) = e' \quad \text{by part a} \\ &\Rightarrow \phi(x) \cdot \phi(x^{-1}) = e'. \end{aligned}$$

Similarly, $x^{-1} \cdot x = e$ implies $\phi(x^{-1}) \cdot \phi(x) = e'$, and therefore $\phi(x^{-1}) = [\phi(x)]^{-1}$.

The concept of isomorphism introduces the relation of being isomorphic on a collection \mathcal{G} of groups. This relation is an equivalence relation, as the following statements show.

1. Any group G in the collection \mathcal{G} is isomorphic to itself. The identity mapping I_G is an automorphism of G .
2. If G and G' are in \mathcal{G} and G is isomorphic to G' , then G' is isomorphic to G . In fact, if ϕ is an isomorphism from G to G' , then ϕ^{-1} is an isomorphism from G' to G . (See Exercise 1 at the end of this section.)
3. Suppose G_1, G_2, G_3 are in \mathcal{G} . If G_1 is isomorphic to G_2 and G_2 is isomorphic to G_3 , then G_1 is isomorphic to G_3 . It is left as an exercise to show that if ϕ_1 is an isomorphism from G_1 to G_2 and ϕ_2 is an isomorphism from G_2 to G_3 , then $\phi_2\phi_1$ is an isomorphism from G_1 to G_3 .

The fundamental idea behind isomorphisms is this: Groups that are isomorphic have the same structure relative to their respective group operation. They are algebraically the same, although details such as the appearance of the elements or the rule defining the operation may vary.

From our discussion at the beginning of this section, we see that any two cyclic groups of order 4 are isomorphic. In fact, any two cyclic groups of the same order are isomorphic (see Exercises 25 and 26 at the end of this section).

The next two examples emphasize the fact that the elements of two isomorphic groups and their group operations may be quite different from each other.

Example 3 Consider $G = \{1, i, -1, -i\}$ under multiplication and $G' = \mathbf{Z}_4 = \{[0], [1], [2], [3]\}$ under addition. Let $\phi: G \rightarrow G'$ be defined by

$$\phi(1) = [0], \quad \phi(i) = [1], \quad \phi(-1) = [2], \quad \phi(-i) = [3].$$

This defines a one-to-one correspondence ϕ from G to G' . To see that ϕ is an isomorphism from G to G' , we use the group tables for G and G' in the same way as in Example 2 of this section. Beginning with the multiplication table for G , we replace each x in the table with $\phi(x)$ (see Figures 3.18 and 3.19). Since the resulting table (Figure 3.19) agrees completely with the addition table for \mathbf{Z}_4 , we conclude that

$$\phi(xy) = \phi(x) + \phi(y)$$

for all $x \in G, y \in G$ and therefore that ϕ is an isomorphism from G to G' .

Multiplication Table for G

\cdot	1	i	-1	$-i$
1	1	i	-1	$-i$
i	i	-1	$-i$	1
-1	-1	$-i$	1	i
$-i$	$-i$	1	i	-1

**Table of $\phi(xy)$**

	[0]	[1]	[2]	[3]
[0]	[0]	[1]	[2]	[3]
[1]	[1]	[2]	[3]	[0]
[2]	[2]	[3]	[0]	[1]
[3]	[3]	[0]	[1]	[2]

Figure 3.18**Figure 3.19**

We conclude this section with an example involving matrices.

Example 4 The multiplicative group G of 3×3 permutation matrices was introduced in Exercise 29 of Section 3.1. This group G is given by $G = \{I_3, P_1, P_2, P_3, P_4, P_5\}$, where

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix},$$

$$P_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad P_5 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

We shall show that this group is isomorphic to the group $S(A) = \{I_A, \rho, \rho^2, \sigma, \gamma, \delta\}$ that appears in Example 2 of this section.

A multiplication table for G is needed as a guide in defining an isomorphism from G to $S(A)$. In constructing this table, we find that

$$P_3^2 = P_5, \quad P_3^3 = I_3, \quad P_3 P_1 = P_4, \quad \text{and} \quad P_3 P_4 = P_2.$$

Using the group table for $S(A)$ in Figure 3.17 as a pattern, we list the elements of G across the table in the order

$$I_3, P_3, P_3^2, P_1, P_4, P_2$$

and evaluate all the products as shown in Figure 3.20. A comparison of the group tables for G and $S(A)$ suggests that the one-to-one correspondence $\phi: G \rightarrow S(A)$ given by

$$\begin{aligned} \phi(I_3) &= I_A & \phi(P_3) &= \rho & \phi(P_3^2) &= \rho^2 \\ \phi(P_1) &= \sigma & \phi(P_4) &= \gamma & \phi(P_2) &= \delta \end{aligned}$$

might be an isomorphism. To verify the property $\phi(xy) = \phi(x)\phi(y)$, we replace each x in the table for G with its image $\phi(x)$ in $S(A)$. The resulting table is shown in Figure 3.21, and it agrees in every position with the group table for $S(A)$ in Figure 3.17. Thus ϕ is an isomorphism from G to $S(A)$.

Multiplication Table for G

\cdot	I_3	P_3	P_3^2	P_1	P_4	P_2
I_3	I_3	P_3	P_3^2	P_1	P_4	P_2
P_3	P_3	P_3^2	I_3	P_4	P_2	P_1
P_3^2	P_3^2	I_3	P_3	P_2	P_1	P_4
P_1	P_1	P_2	P_4	I_3	P_3^2	P_3
P_4	P_4	P_1	P_2	P_3	I_3	P_3^2
P_2	P_2	P_4	P_1	P_3^2	P_3	I_3

Table of $\phi(xy)$

	I_A	ρ	ρ^2	σ	γ	δ
I_A	I_A	ρ	ρ^2	σ	γ	δ
ρ	ρ	ρ^2	I_A	γ	δ	σ
ρ^2	ρ^2	I_A	ρ	δ	σ	γ
σ	σ	δ	γ	I_A	ρ^2	ρ
γ	γ	σ	δ	ρ	I_A	ρ^2
δ	δ	γ	σ	ρ^2	ρ	I_A

Figure 3.20**Figure 3.21**

Exercises 3.5

True or False

Label each of the following statements as either true or false.

1. Any two cyclic groups of the same order are isomorphic.
2. Any two abelian groups of the same order are isomorphic.
3. Any isomorphism is an automorphism.
4. Any automorphism is an isomorphism.
5. If two groups G and G' have order 3, then G and G' are isomorphic.
6. Any two groups of the same finite order are isomorphic.
7. Two groups can be isomorphic even though their group operations are different.
8. The relation of being isomorphic is an equivalence relation on a collection of groups.

Exercises

1. Prove that if ϕ is an isomorphism from the group G to the group G' , then ϕ^{-1} is an isomorphism from G' to G .
2. Let G_1 , G_2 , and G_3 be groups.
 - a. Prove that if ϕ_1 is an isomorphism from G_1 to G_2 and ϕ_2 is an isomorphism from G_2 to G_3 , then $\phi_2\phi_1$ is an isomorphism from G_1 to G_3 .
 - b. If ϕ_1 is an isomorphism from G_1 to G_3 and ϕ_2 is an isomorphism from G_2 to G_3 , find an isomorphism from G_1 to G_2 .
3. Find an isomorphism from the additive group[†] $\mathbf{Z}_4 = \{[0]_4, [1]_4, [2]_4, [3]_4\}$ to the multiplicative group of units $\mathbf{U}_5 = \{[1]_5, [2]_5, [3]_5, [4]_5\} \subseteq \mathbf{Z}_5$.

[†]For clarity, we are temporarily writing $[a]_n$ for $[a] \in \mathbf{Z}_n$.

4. Let $G = \{1, i, -1, -i\}$ under multiplication, and let $G' = \mathbf{Z}_4 = \{[0], [1], [2], [3]\}$ under addition. Find an isomorphism from G to G' that is different from the one given in Example 3 of this section.

Sec. 3.3, #16 ➤ 5. Let H be the group given in Exercise 16 of Section 3.3, and let $\mathcal{S}(A)$ be as given in Example 4 of this section. Find an isomorphism from H to $\mathcal{S}(A)$.

Sec. 3.4, #16 ➤ 6. Find an isomorphism from the additive group $\mathbf{Z}_6 = \{[a]_6\}$ to the multiplicative group of units $\mathbf{U}_7 = \{[a]_7 \in \mathbf{Z}_7 \mid [a]_7 \neq [0]_7\}$.

Sec. 3.4, #14 ➤ 7. Find an isomorphism ϕ from the additive group \mathbf{Z} to the multiplicative group

$$H = \left\{ \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} \mid n \in \mathbf{Z} \right\}$$

and prove that $\phi(x + y) = \phi(x)\phi(y)$.

Sec. 3.3, #15b ➤ 8. Find an isomorphism from the group $G = \{1, i, -1, -i\}$ in Example 3 of this section to the multiplicative group

$$H = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\}.$$

Sec. 3.3, #14c ➤ 9. Find an isomorphism ϕ from the multiplicative group G of nonzero complex numbers to the multiplicative group

$$H = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \mid a, b \in \mathbf{R} \text{ and } a^2 + b^2 \neq 0 \right\}$$

and prove that $\phi(xy) = \phi(x)\phi(y)$.

Sec. 3.3, #15a ➤ 10. Find an isomorphism from the multiplicative group

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

Sec. 4.2, #3 <

Sec. 4.4, #11 <
Sec. 4.6, #16 <
to the group $G = \{e, a, b, ab\}$ with multiplication table in Figure 3.22. This group is known as the **Klein[†] four group**.

•	e	a	b	ab
e	e	a	b	ab
a	a	e	ab	b
b	b	ab	e	a
ab	ab	b	a	e

■ **Figure 3.22**

[†]Felix Christian Klein (1849–1925) was a German mathematician known for his work on the connections between geometry and group theory. Klein successfully worked toward the admission of women to the University of Göttingen in Germany in 1893, and supervised the first Ph.D. thesis by a woman at Göttingen.

Sec. 3.1, #28 ➤

- 11.** The following set of matrices

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

$$\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}$$

forms a group H with respect to matrix multiplication. Find an isomorphism from H to the quaternion group.

- 12.** Let G be the additive group of all real numbers, and let G' be the group of all positive real numbers under multiplication. Verify that the mapping $\phi: G \rightarrow G'$ defined by $\phi(x) = 10^x$ is an isomorphism from G to G' .
- 13.** Let G and G' be as given in Exercise 12. Verify that the mapping $\theta: G' \rightarrow G$ defined by $\theta(x) = \log x$ is an isomorphism from G' to G .
- 14.** Assume that the nonzero complex numbers form a group G with respect to multiplication. If a and b are real numbers and $i = \sqrt{-1}$, the **conjugate** of the complex number $a + bi$ is defined to be $a - bi$. With this notation, let $\phi: G \rightarrow G$ be defined by $\phi(a + bi) = a - bi$ for all $a + bi$ in G . Prove that ϕ is an automorphism of G .
- 15.** Let G be a group. Prove that G is abelian if and only if the mapping $\phi: G \rightarrow G$ defined by $\phi(x) = x^{-1}$ for all x in G is an automorphism.
- 16.** Suppose $(m, n) = 1$ and let $\phi: \mathbf{Z}_n \rightarrow \mathbf{Z}_n$ be defined by $\phi([a]) = m[a]$. Prove or disprove that ϕ is an automorphism of the additive group \mathbf{Z}_n .

Sec. 3.1, #27a ➤

- 17.** According to Exercise 27a of Section 3.1, \mathbf{U}_n , the set of nonzero elements of \mathbf{Z}_n , forms a group with respect to multiplication if n is prime. Prove or disprove that the mapping $\phi: \mathbf{U}_n \rightarrow \mathbf{U}_n$ defined by the rule in Exercise 16 is an automorphism of \mathbf{U}_n .

Sec. 4.6, #32 ≈

- 18.** For each a in the group G , define a mapping $t_a: G \rightarrow G$ by $t_a(x) = axa^{-1}$. Prove that t_a is an automorphism of G .

Sec. 4.6, #32 ≈

- 19.** For a fixed group G , prove that the set of all automorphisms of G forms a group with respect to mapping composition.

- 20.** Assume G is a (not necessarily finite) cyclic group generated by a in G , and let ϕ be an automorphism of G . Prove that each element of G is equal to a power of $\phi(a)$; that is, prove that $\phi(a)$ is a generator of G .

- 21.** Let G be as in Exercise 20. Suppose also that a^r is a generator of G . Define f on G by $f(a) = a^r, f(a^i) = (a^r)^i = a^{ri}$. Prove that f is an automorphism of G .

- 22.** Let G be the multiplicative group of units \mathbf{U}_n . For each value of n , use the results of Exercises 20 and 21 to list all the automorphisms of G . For each automorphism ϕ , write out the images $\phi(x)$ for all x in G .

a. $n = 5$ **b.** $n = 7$

- 23.** Use the results of Exercises 20 and 21 to find the *number* of automorphisms of the additive group \mathbf{Z}_n for the given value of n .

a. $n = 3$ **b.** $n = 4$ **c.** $n = 8$ **d.** $n = 6$

24. Prove that any cyclic group of finite order n is isomorphic to \mathbf{Z}_n under addition.
25. For an arbitrary positive integer n , prove that any two cyclic groups of order n are isomorphic.
26. Prove that any infinite cyclic group is isomorphic to \mathbf{Z} under addition.
27. Let H be the group \mathbf{Z}_6 under addition. Find all isomorphisms from the multiplicative group \mathbf{U}_7 of units in \mathbf{Z}_7 to H .
28. Suppose that G and H are isomorphic groups. Prove that G is abelian if and only if H is abelian.
29. Prove that if G and H are two groups that contain exactly two elements each, then G and H are isomorphic.
30. Prove that any two groups of order 3 are isomorphic.
31. Exhibit two groups of the same finite order that are not isomorphic.
32. Let ϕ be an isomorphism from group G to group H . Let x be in G . Prove that $\phi(x^n) = (\phi(x))^n$ for every integer n .
33. If G and H are groups and $\phi: G \rightarrow H$ is an isomorphism, prove that a and $\phi(a)$ have the same order, for any $a \in G$.
34. Suppose that ϕ is an isomorphism from the group G to the group G' .
 - a. Prove that if H is any subgroup of G , then $\phi(H)$ is a subgroup of G' .
 - b. Prove that if K is any subgroup of G' , then $\phi^{-1}(K)$ is a subgroup of G .

3.6

Homomorphisms

We saw in the last section that an isomorphism between two groups provides a connection that shows that the two groups have the same structure relative to their group operations. It is for this reason that the concept of an isomorphism is extremely important in algebra.

The name *homomorphism* is given to another important type of mapping that is related to, but different from, the isomorphism. The basic differences are that a homomorphism is not required to be one-to-one and also not required to be onto. The formal definition is as follows.

Definition 3.27 ■ Homomorphism, Endomorphism, Epimorphism, Monomorphism

Let G be a group with respect to \circledast , and let G' be a group with respect to \boxtimes . A **homomorphism** from G to G' is a mapping $\phi: G \rightarrow G'$ such that

$$\phi(x \circledast y) = \phi(x) \boxtimes \phi(y)$$

for all x and y in G . If $G = G'$, the homomorphism ϕ is an **endomorphism**. A homomorphism ϕ is called an **epimorphism** if ϕ is onto, and a **monomorphism** if ϕ is one-to-one.

As we did with isomorphisms, we drop the special symbols \circledast and \boxtimes and simply write $\phi(xy) = \phi(x)\phi(y)$ for the given condition.

As already noted, a homomorphism ϕ from G to G' need not be one-to-one or onto. If ϕ is both (that is, if ϕ is a bijection), then ϕ is an isomorphism as defined in Definition 3.25.

Our first example of a homomorphism has a natural connection with our work in Chapter 2.

Example 1 For a fixed integer $n > 1$, consider the mapping ϕ from the additive group \mathbf{Z} to the additive group \mathbf{Z}_n defined by

$$\phi(x) = [x],$$

where $[x]$ is the congruence class in \mathbf{Z}_n that contains x . From the properties of addition in \mathbf{Z}_n (see Section 2.6), it follows that

$$\begin{aligned}\phi(x + y) &= [x + y] \\ &= [x] + [y] \\ &= \phi(x) + \phi(y).\end{aligned}$$

Thus ϕ is a homomorphism. It follows from the definition of \mathbf{Z}_n that ϕ is onto, so ϕ is, in fact, an epimorphism from \mathbf{Z} to \mathbf{Z}_n . Since $\phi(0) = \phi(n) = [0]$, then ϕ is not one-to-one and hence not a monomorphism. ■

Example 2 For two arbitrary groups G and G' , let e' denote the identity element in G' and define $\phi: G \rightarrow G'$ by $\phi(x) = e'$ for all $x \in G$. Then, for all x and y in G ,

$$\begin{aligned}\phi(x) \cdot \phi(y) &= e' \cdot e' \\ &= e' \\ &= \phi(xy),\end{aligned}$$

and ϕ is a homomorphism from G to G' . If G' has order greater than 1, then ϕ is not onto and hence not an epimorphism. Also ϕ is not one-to-one, since for any $x \neq y$, we have $\phi(x) = \phi(y) = e'$. Thus ϕ is not a monomorphism. ■

The two previous examples show that, unlike the situation with isomorphisms, the existence of a homomorphism from G to G' does not imply that G and G' have the same structure. However, we shall see that the existence of a homomorphism can reveal important and interesting information relating their structures. As with isomorphisms, we say that a homomorphism “preserves the group operation.” Two simple consequences of this condition are that identities must correspond and inverses must be mapped onto inverses. This is stated in our next theorem, and the proofs are requested in the exercises.

Theorem 3.28 ■ Images of Identities and Inverses

Let ϕ be a homomorphism from the group G to the group G' . If e denotes the identity in G , and e' denotes the identity in G' , then

- a. $\phi(e) = e'$, and
- b. $\phi(x^{-1}) = [\phi(x)]^{-1}$ for all x in G .

The following examples give some indication of the variety that is in homomorphisms. Other examples appear in the exercises for this section.

Example 3 Consider the group G of nonzero real numbers under multiplication and the additive group \mathbf{Z} . Define $\phi: \mathbf{Z} \rightarrow G$ by

$$\phi(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ -1 & \text{if } n \text{ is odd.} \end{cases}$$

Since every integer is either even or odd and not both, $\phi(n)$ is well-defined. The following table systematically checks the equality $\phi(m + n) = \phi(m) \cdot \phi(n)$.

	$m + n$	$\phi(m) \cdot \phi(n)$	$\phi(m + n)$
m, n both even	even	(1)(1)	1
one even, one odd	odd	(1)(-1)	-1
m, n both odd	even	(-1)(-1)	1

A comparison of the last two columns shows that ϕ is indeed a homomorphism from \mathbf{Z} to G . However since ϕ is not onto, it is not an epimorphism. Since $\phi(0) = \phi(2) = 1$, then ϕ is not one-to-one and hence not a monomorphism. ■

Example 4 Consider the additive group \mathbf{Z} and the mapping $\phi: \mathbf{Z} \rightarrow \mathbf{Z}$ defined by $\phi(x) = 5x$ for all $x \in \mathbf{Z}$. Since

$$\begin{aligned}\phi(x + y) &= 5(x + y) \\ &= 5x + 5y \\ &= \phi(x) + \phi(y),\end{aligned}$$

ϕ is an endomorphism. Clearly, ϕ is not an epimorphism since ϕ is not onto. However, since $\phi(x) = \phi(y)$ implies $5x = 5y$ and $x = y$, then ϕ is a monomorphism. ■

We saw in the last section that the relation of being isomorphic is an equivalence relation on a given collection \mathcal{G} of groups. The concept of homomorphism leads to a corresponding, but different, relation. If there exists an epimorphism from the group G to the group G' , then G' is called a **homomorphic image** of G . Example 1 in this section shows that the additive group \mathbf{Z}_n is a homomorphic image of the additive group \mathbf{Z} .

On a given collection \mathcal{G} of groups, the relation of being a homomorphic image is reflexive and transitive but may not be symmetric. These facts are brought out in the exercises for this section.

The real importance of homomorphisms will be much clearer at the end of Section 4.6 in the next chapter. The kernel of a homomorphism is one of the key concepts in that section.

Definition 3.29 ■ Kernel

Let ϕ be a homomorphism from the group G to the group G' . The **kernel** of ϕ is the set

$$\ker \phi = \{x \in G \mid \phi(x) = e'\}$$

where e' denotes the identity in G' .

Example 5 To illustrate Definition 3.29, we list the kernels of the homomorphisms from the preceding examples in this section.

The kernel of the homomorphism $\phi: \mathbf{Z} \rightarrow \mathbf{Z}_n$ defined by $\phi(x) = [x]$ in Example 1 is given by

$$\ker \phi = \{x \in \mathbf{Z} \mid x = kn \text{ for some } k \in \mathbf{Z}\},$$

since $\phi(x) = [x] = [0]$ if and only if x is a multiple of n .

The homomorphism $\phi: \mathbf{Z} \rightarrow G$ in Example 3 defined by

$$\phi(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ -1 & \text{if } n \text{ is odd} \end{cases}$$

has the set \mathbf{E} of all even integers as its kernel, since 1 is the identity in G .

For $\phi: \mathbf{Z} \rightarrow \mathbf{Z}$ defined by $\phi(x) = 5x$ in Example 4, we have $\ker \phi = \{0\}$, since $5x = 0$ if and only if $x = 0$. This kernel is an extreme case since part a of Theorem 3.28 assures us that the identity is always an element of the kernel.

At the other extreme, the homomorphism $\phi: G \rightarrow G'$ defined in Example 2 by $\phi(x) = e'$ for all $x \in G$ has $\ker \phi = G$. ■

Exercises 3.6**True or False**

Label each of the following statements as either true or false.

1. Every homomorphism is an isomorphism.
2. Every isomorphism is a homomorphism.
3. Every endomorphism is an epimorphism.
4. Every epimorphism is an endomorphism.
5. Every monomorphism is an isomorphism.
6. Every isomorphism is an epimorphism and a monomorphism.
7. The relation of being a homomorphic image is an equivalence relation on a collection of groups.
8. The kernel of a homomorphism is never empty.
9. It is possible to find at least one homomorphism from any group G to any group G' .
10. If there exists a homomorphism from group G to group G' , then G' is said to be a homomorphic image of G .

Exercises

1. Each of the following rules determines a mapping $\phi: G \rightarrow G$, where G is the group of all nonzero real numbers under multiplication. Decide in each case whether or not ϕ is an endomorphism. For those that are endomorphisms, state the kernel and decide whether ϕ is an epimorphism or a monomorphism.
 - a. $\phi(x) = |x|$
 - b. $\phi(x) = 1/x$
 - c. $\phi(x) = -x$
 - d. $\phi(x) = x^2$
 - e. $\phi(x) = \frac{|x|}{x}$
 - f. $\phi(x) = x^2 + 1$
 - g. $\phi(x) = \sqrt[3]{x}$
 - h. $\phi(x) = \frac{x}{2}$
2. Each of the following rules determines a mapping ϕ from the additive group \mathbf{Z}_4 to the additive groups \mathbf{Z}_2 . In each case prove or disprove that ϕ is a homomorphism. If ϕ is a homomorphism, find $\ker \phi$ and decide whether ϕ is an epimorphism or a monomorphism.
 - a. $\phi([x]) = \begin{cases} [0] & \text{if } x \text{ is even} \\ [1] & \text{if } x \text{ is odd} \end{cases}$
 - b. $\phi([x]) = [x + 2]$
3. Consider the additive groups of real numbers \mathbf{R} and complex numbers \mathbf{C} and define $\phi: \mathbf{R} \rightarrow \mathbf{C}$ by $\phi(x) = x + 0i$. Prove that ϕ is a homomorphism and find $\ker \phi$. Is ϕ an epimorphism? Is ϕ a monomorphism?
4. Consider the additive group \mathbf{Z} and the multiplicative group $G = \{1, i, -1, -i\}$ and define $\phi: \mathbf{Z} \rightarrow G$ by $\phi(n) = i^n$. Prove that ϕ is a homomorphism and find $\ker \phi$. Is ϕ an epimorphism? Is ϕ a monomorphism?
5. Consider the additive group \mathbf{Z}_{12} and define $\phi: \mathbf{Z}_{12} \rightarrow \mathbf{Z}_{12}$ by $\phi([x]) = [3x]$. Prove that ϕ is a homomorphism and find $\ker \phi$. Is ϕ an epimorphism? Is ϕ a monomorphism?
6. Consider the additive groups \mathbf{Z}_{12} and \mathbf{Z}_6 and define $\phi: \mathbf{Z}_{12} \rightarrow \mathbf{Z}_6$ by $\phi([x]_{12}) = [x]_6$. Prove that ϕ is a homomorphism and find $\ker \phi$. Is ϕ an epimorphism? Is ϕ a monomorphism?
7. Consider the additive groups \mathbf{Z}_8 and \mathbf{Z}_4 and define $\phi: \mathbf{Z}_8 \rightarrow \mathbf{Z}_4$ by $\phi([x]_8) = [x]_4$. Prove that ϕ is a homomorphism and find $\ker \phi$. Is ϕ an epimorphism? Is ϕ a monomorphism?
8. Consider the additive groups $M_2(\mathbf{Z})$ and \mathbf{Z} and define $\phi: M_2(\mathbf{Z}) \rightarrow \mathbf{Z}$ by $\phi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) =$
 - a. Prove that ϕ is a homomorphism and find $\ker \phi$. Is ϕ an epimorphism? Is ϕ a monomorphism?
9. Let G be the multiplicative group of invertible matrices in $M_2(\mathbf{R})$, and let G' be the group of nonzero real numbers under multiplication. Prove that the mapping $\phi: G \rightarrow G'$ defined by

$$\phi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$

Sec. 1.6, #28, 29 =>

9. Let G be the multiplicative group of invertible matrices in $M_2(\mathbf{R})$, and let G' be the group of nonzero real numbers under multiplication. Prove that the mapping $\phi: G \rightarrow G'$ defined by

is a homomorphism. Is ϕ an epimorphism? Is ϕ a monomorphism? (The value of this mapping is called the **determinant** of the matrix.)

Sec. 4.6, #14 <
Sec. 6.2, #15a <

10. Find an example of G , G' , and ϕ such that G is a nonabelian group, G' is an abelian group, and ϕ is an epimorphism from G to G' .
11. Let ϕ be a homomorphism from the group G to the group G' .
 - a. Prove part a of Theorem 3.28: If e denotes the identity in G and e' denotes the identity in G' , then $\phi(e) = e'$.
 - b. Prove part b of Theorem 3.28: $\phi(x^{-1}) = [\phi(x)]^{-1}$ for all x in G .
12. Prove that on a given collection \mathcal{G} of groups, the relation of being a homomorphic image has the reflexive property.
13. Suppose that G , G' , and G'' are groups. If G' is a homomorphic image of G , and G'' is a homomorphic image of G' , prove that G'' is a homomorphic image of G . (Thus the relation in Exercise 12 has the transitive property.)
14. Find two groups G and G' such that G' is a homomorphic image of G but G is not a homomorphic image of G' . (Thus the relation in Exercise 12 does not have the symmetric property.)
15. Suppose that ϕ is an epimorphism from the group G to the group G' . Prove that ϕ is an isomorphism if and only if $\ker \phi = \{e\}$, where e denotes the identity in G .
16. If G is an abelian group and the group G' is a homomorphic image of G , prove that G' is abelian.
17. Let a be a fixed element of the multiplicative group G . Define ϕ from the additive group \mathbf{Z} to G by $\phi(n) = a^n$ for all $n \in \mathbf{Z}$. Prove that ϕ is a homomorphism.
18. With ϕ as in Exercise 17, show that $\phi(\mathbf{Z}) = \langle a \rangle$, and describe the kernel of ϕ .
19. Assume that ϕ is a homomorphism from the group G to the group G' .
 - a. Prove that if H is any subgroup of G , then $\phi(H)$ is a subgroup of G' .
 - b. Prove that if K is any subgroup of G' , then $\phi^{-1}(K)$ is a subgroup of G .
20. Assume that the group G' is a homomorphic image of the group G .
 - a. Prove that G' is cyclic if G is cyclic.
 - b. Prove that $o(G')$ divides $o(G)$, whether G is cyclic or not.
21. Let ϕ be a homomorphism from the group G to the group G' , where $G = \langle a \rangle$, the cyclic group generated by a . Show that ϕ is completely determined by the image of the generator a of G .

Key Words and Phrases

abelian group, 138	finite group, 141	homomorphism, 183
automorphism, 177	generalized associative law, 148	idempotent element, 150
cyclic group, 159	generator of a group, 163	identity element, 138
endomorphism, 183	group, 137	infinite group, 141
epimorphism, 183	group of units, 167	integral exponents, 155
Euler phi-function, 170	homomorphic image, 185	integral multiples, 156

inverse, 138
isomorphism, 177
kernel of a homomorphism, 186
monomorphism, 183

nontrivial subgroup, 152
order of a group, 141
order of an element, 167
reverse order law, 146

rigid motion, 175
subgroup, 152
subgroup generated by a , 159



LC-USZ62-100653/Library of Congress Prints and Photographs Division

A Pioneer in Mathematics

Niels Henrik Abel (1802–1829)

Niels Henrik Abel was a leading 19th-century Norwegian mathematician. Although he died at the age of 27, his accomplishments were extraordinary, and he is Norway's most noted mathematician. His memory is honored in many ways. A monument to him was erected at Froland Church, his burial place, by his friend Baltazar Mathias Keilhau. History tells us that on his deathbed, Abel jokingly asked his friend to care for his fiancée after his death, perhaps by marrying her. (After Abel died, Keilhau did marry Abel's fiancée.) A statue of Abel stands in the Royal Park of Oslo, and Norway has issued five postage stamps in his honor. Many theorems of advanced mathematics bear his name. Probably the most lasting and significant recognition is in the term *abelian group*, coined around 1870.

Abel was one of seven children of a pastor. When he was 18 his father died, and supporting the family became his responsibility. In spite of this burden, Abel continued his study of mathematics and successfully solved a problem that had baffled mathematicians for more than 300 years: He proved that the general fifth-degree polynomial equation could not be solved using the four basic arithmetic operations and extraction of roots.

Although Abel never held an academic position, he continued to pursue his mathematical research, contributing not only to the groundwork for what later became known as abstract algebra but also to the theory of infinite series, elliptic functions, elliptic integrals, and abelian integrals.

In Berlin, Abel became friends with August Leopold Crelle (1780–1856), a civil engineer and founder of the first journal devoted entirely to mathematical research. It was only through Crelle's friendship and respect for Abel's talent that many of Abel's papers were published. In fact, Crelle finally obtained a faculty position for Abel at the University of Berlin, but unfortunately, the news reached Norway two days after Abel's death.

This page intentionally left blank

More on Groups

■ Introduction

The first two sections of this chapter present the standard material on permutation groups, and the optional Section 4.3 contains some real-world applications of such groups. The next section introduces cosets of a subgroup, a concept necessary to the study of normal subgroups and quotient groups in the next two sections. The chapter then concludes with two optional sections that present some results on finite abelian groups and give a sample of more advanced work.

The set \mathbf{Z}_n of congruence classes modulo n makes isolated appearances in this chapter.

4.1

Finite Permutation Groups

An appreciation of the importance of permutation groups must be based to some extent on a knowledge of their structures. The basic facts about finite permutation groups are presented in this section, and their importance is revealed in the next two sections.

Suppose A is a finite set of n elements—say,

$$A = \{a_1, a_2, \dots, a_n\}.$$

Any permutation f on A is determined by the choices for the n values

$$f(a_1), f(a_2), \dots, f(a_n).$$

In assigning these values, there are n choices for $f(a_1)$, then $n - 1$ choices of $f(a_2)$, then $n - 2$ choices of $f(a_3)$, and so on. Thus there are $n(n - 1) \cdots (2)(1) = n!$ different ways in which f can be defined, and $S(A)$ has $n!$ elements. Each element f in $S(A)$ can be represented by a matrix (rectangular array) in which the image of a_i is written under a_i :

$$f = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ f(a_1) & f(a_2) & \cdots & f(a_n) \end{bmatrix}.$$

Each permutation f on A can be made to correspond to a permutation f' on $B = \{1, 2, \dots, n\}$ by replacing a_k with k for $k = 1, 2, \dots, n$:

$$f' = \begin{bmatrix} 1 & 2 & \cdots & n \\ f'(1) & f'(2) & \cdots & f'(n) \end{bmatrix}.$$

The mapping $f \rightarrow f'$ is an isomorphism from $\mathcal{S}(A)$ to $\mathcal{S}(B)$, and the groups are the same except for notation. For this reason, we will henceforth consider a permutation on a set of n elements as being written on the set $B = \{1, 2, \dots, n\}$. The group $\mathcal{S}(B)$ is known as the **symmetric group** on n elements, and it is denoted by S_n .

Example 1 As an illustration of the matrix representation, the notation

$$f = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 4 & 2 \end{bmatrix}$$

indicates that f is an element of S_5 and that $f(1) = 3, f(2) = 5, f(3) = 1, f(4) = 4$, and $f(5) = 2$. ■

Definition 4.1 ■ Cycle

An element f of S_n is a **cycle** if there exists a set $\{i_1, i_2, \dots, i_r\}$ of distinct integers such that

$$f(i_1) = i_2, f(i_2) = i_3, \dots, f(i_{r-1}) = i_r, f(i_r) = i_1,$$

and f leaves all other elements fixed.

By this definition, f is a cycle if there are distinct integers i_1, i_2, \dots, i_r such that f maps these elements according to the cyclic pattern

$$i_1 \rightarrow i_2 \rightarrow i_3 \rightarrow \cdots \rightarrow i_{r-1} \rightarrow i_r,$$

and f leaves all other elements fixed. A cycle such as this can be written in the form

$$f = (i_1, i_2, \dots, i_r),$$

where it is understood that $f(i_k) = i_{k+1}$ for $1 \leq k < r$, and $f(i_r) = i_1$.

Example 2 The permutation

$$f = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 6 & 3 & 7 & 5 & 4 & 2 \end{bmatrix}$$

can be written simply as

$$f = (2, 6, 4, 7).$$

This expression is not unique, because

$$\begin{aligned} f &= (2, 6, 4, 7) \\ &= (6, 4, 7, 2) \\ &= (4, 7, 2, 6) \\ &= (7, 2, 6, 4). \end{aligned}$$

■

Example 3 It is easy to write the inverse of a cycle. Since $f(i_k) = i_{k+1}$ implies $f^{-1}(i_{k+1}) = i_k$, we only need to reverse the order of the cyclic pattern. For

$$f = (1, 2, 3, 4, 5, 6, 7, 8, 9),$$

we have

$$\begin{aligned} f^{-1} &= (9, 8, 7, 6, 5, 4, 3, 2, 1) \\ &= (1, 9, 8, 7, 6, 5, 4, 3, 2). \end{aligned} \quad \blacksquare$$

Not all elements of S_n are cycles, but every permutation can be written as a product of mutually disjoint cycles. As an example, consider the permutation

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

When we use the same representation scheme with $f(k)$ written beneath k , the result of a rearrangement of the columns in the matrix still represents f :

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

The columns have been arranged in a special way: If $f(p) = q$, the column with q at the top has been written next after the column with p at the top. This arranges the elements in the first row so that f maps them according to the following pattern:

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

Thus 1, 3, 2, and 8 are mapped in a circular pattern, and so are 4 and 6, and 5, 7, and 9. This procedure has led to a separation of the elements of $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ into disjoint subsets $\{1, 3, 2, 8\}$, $\{4, 6\}$, and $\{5, 7, 9\}$ according to the pattern determined by the following computations:[†]

$$\begin{array}{lll} f(1) = 3 & f(4) = 6 & f(5) = 7 \\ f^2(1) = f(3) = 2 & f^2(4) = f(6) = 4 & f^2(5) = f(7) = 9 \\ f^3(1) = f(2) = 8 & & f^3(5) = f(9) = 5. \\ f^4(1) = f(8) = 1 & & \end{array}$$

The disjoint subsets $\{1, 3, 2, 8\}$, $\{4, 6\}$, and $\{5, 7, 9\}$ are called the **orbits** of f .

For each orbit of f , we define a cycle that maps the elements in that orbit in the same way as does f :

$$\begin{aligned} g_1 &= (1, 3, 2, 8) \\ g_2 &= (4, 6) \\ g_3 &= (5, 7, 9). \end{aligned}$$

[†] $f^2 = f \circ f$, $f^3 = f \circ f^2 = f \circ f \circ f$ and so on.

These cycles are automatically on disjoint sets of elements since the orbits are disjoint, and we see that their product is f :

$$\begin{aligned} f &= g_1g_2g_3 \\ &= (1, 3, 2, 8)(4, 6)(5, 7, 9). \end{aligned}$$

Note that these cycles commute with each other because they are on disjoint sets of elements.

Example 4 The positive integral powers of a cycle f are easy to compute since f^m will map each integer in the cycle onto the integer located m places farther along in the cycle. For instance, if

$$f = (1, 2, 3, 4, 5, 6, 7, 8, 9),$$

then f^2 maps each element onto the element two places farther along, according to the pattern

$$\begin{gathered} \overbrace{1, 2, 3, 4, 5, 6, 7, \dots}^{\text{two places farther}} \\ f^2 = (1, 3, 5, 7, 9, 2, 4, 6, 8). \end{gathered}$$

Similarly, f^3 maps each element onto the element three places farther along, and so on for higher powers:

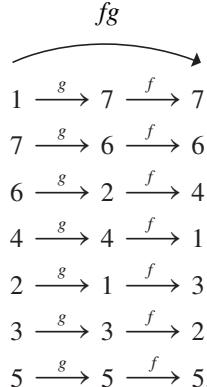
$$\begin{aligned} f^3 &= (1, 4, 7)(2, 5, 8)(3, 6, 9) \\ f^4 &= (1, 5, 9, 4, 8, 3, 7, 2, 6). \end{aligned}$$
■

In connection with Example 4, we note that the **order** of an r -cycle (a cycle with r elements) is r .

Ordinarily, cycles that are not on disjoint sets of elements will not commute, but their product is defined using mapping composition. For example, suppose $f = (1, 3, 2, 4)$ and $g = (1, 7, 6, 2)$. Then[†]

$$fg = (1, 3, 2, 4)(1, 7, 6, 2) = (1, 7, 6, 4)(2, 3),$$

since



[†]The product fg is computed from right to left, according to $f(g(x))$. Some texts multiply permutations from left to right.

The computation of fg may be easier to see in the following diagram:



$$fg = (1, 3, 2, 4)(1, 7, 6, 2) = (1, 7, 6, 4)(2, 3).$$

A similar diagram for gf appears as follows:



$$gf = (1, 7, 6, 2)(1, 3, 2, 4) = (1, 3)(2, 4, 7, 6).$$

Thus $gf \neq fg$. We adopt the notation that a 1-cycle such as (5) indicates that the element is left fixed. For example, gf could also be written as

$$gf = (1, 3)(2, 4, 7, 6)(5).$$

This allows expressions such as $e = (1)$ or $e = (1)(2)$ for the identity permutation.

Example 5 A product of cycles with any number of factors can be expressed as a product of disjoint cycles by the same procedure that we used in computing fg with $f = (1, 3, 2, 4)$ and $g = (1, 7, 6, 2)$. To illustrate, suppose we wish to express

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

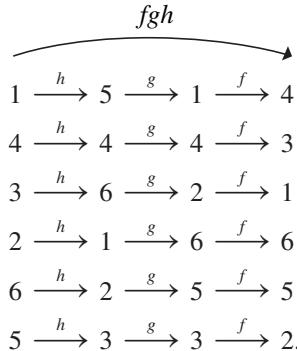
as a product of disjoint cycles. Let

$$f = (1, 4, 3, 2)$$

$$g = (1, 6, 2, 5)$$

$$h = (1, 5, 3, 6, 2).$$

The following computations can be done *mentally* to obtain fgh as a product of disjoint cycles:



Thus

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

When a permutation is written as a product of disjoint cycles, it is easy to find the order of the permutation if we use the result in Exercise 36 of Section 3.4: The order of the product is simply the least common multiple of the orders of the cycles. For example, the product $(1, 2, 3, 4)(5, 6, 7, 8, 9, 10)$ has order 12, the least common multiple of 4 and 6.

Example 6 The expression of permutations as products of cycles enables us to write the elements of S_n in a very compact form. The elements of S_3 are given by

$$\begin{aligned} e &= (1) & \sigma &= (1, 2) \\ \rho &= (1, 2, 3) & \gamma &= (1, 3) \\ \rho^2 &= (1, 3, 2) & \delta &= (2, 3). \end{aligned}$$
■

A 2-cycle such as $(3, 7)$ is called a **transposition**. Every permutation can be written as a product of transpositions, for every permutation can be written as a product of cycles, and any cycle (i_1, i_2, \dots, i_r) can be written as

$$(i_1, i_2, \dots, i_r) = (i_1, i_r)(i_1, i_{r-1}) \cdots (i_1, i_3)(i_1, i_2).$$

For example,

$$(1, 3, 2, 4) = (1, 4)(1, 2)(1, 3).$$

The factorization into a product of transpositions is not unique, as the next example shows.

Example 7 Consider the product fg , where $f = (1, 3, 2, 4)$ and $g = (1, 7, 6, 2)$. This product can be written as

$$(1, 3, 2, 4)(1, 7, 6, 2) = (1, 4)(1, 2)(1, 3)(1, 2)(1, 6)(1, 7)$$

and also as

$$\begin{aligned} (1, 3, 2, 4)(1, 7, 6, 2) &= (1, 7, 6, 4)(2, 3) \\ &= (1, 4)(1, 6)(1, 7)(2, 3). \end{aligned}$$
■

Although the expression of a permutation as a product of transpositions is not unique, the number of transpositions used for a certain permutation is either *always odd* or else *always even*. Our proof of this fact takes us somewhat astray from our main course in this chapter. It involves consideration of a polynomial P in n variables x_1, x_2, \dots, x_n that is the product of all factors of the form $(x_i - x_j)$ with $1 \leq i < j \leq n$:

$$P = \prod_{i < j}^n (x_i - x_j).$$

(The symbol \prod indicates a product in the same way that \sum is used to indicate sums.) For example, if $n = 3$, then

$$\begin{aligned} P &= \prod_{i < j}^3 (x_i - x_j) \\ &= (x_1 - x_2)(x_1 - x_3)(x_2 - x_3). \end{aligned}$$

For $n = 4$, P is given by

$$\begin{aligned} P &= \prod_{i < j}^4 (x_i - x_j) \\ &= (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4), \end{aligned}$$

and similarly for larger values of n .

If f is any permutation on $\{1, 2, \dots, n\}$, then f is applied to P by the rule

$$f(P) = \prod_{i < j}^n (x_{f(i)} - x_{f(j)}).$$

As an illustration, let us apply the transposition $t = (2, 4)$ to the polynomial

$$\begin{aligned} P &= \prod_{i < j}^4 (x_i - x_j) \\ &= (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4). \end{aligned}$$

We have

$$t(P) = (x_1 - x_4)(x_1 - x_3)(x_1 - x_2)(x_4 - x_3)(x_4 - x_2)(x_3 - x_2),$$

since 2 and 4 are interchanged by t . Analyzing this result, we observe the following:

1. The factor $(x_2 - x_4)$ in P is changed to $(x_4 - x_2)$ in $t(P)$, so this factor changes sign.
2. The factor $(x_1 - x_3)$ is unchanged.
3. The remaining factors in $t(P)$ may be grouped in pairs as

$$(x_1 - x_4)(x_1 - x_2) \text{ and } (x_4 - x_3)(x_3 - x_2) = (x_3 - x_4)(x_2 - x_3).$$

The products of these pairs are unchanged by t .

Thus $t(P) = (-1)P$, in this particular case. The sort of analysis we have used here can be used to prove the following lemma.

Lemma 4.2

If $t = (r, s)$ is any transposition on $\{1, 2, \dots, n\}$ and $P = \prod_{i < j}^n (x_i - x_j)$, then

$$t(P) = (-1)P.$$

$(u \wedge v) \Rightarrow w$ **Proof** Since $t = (r, s) = (s, r)$, we may assume that $r < s$. We have

$$t(P) = \prod_{i < j}^n (x_{t(i)} - x_{t(j)}).$$

The factors of $t(P)$ may be analyzed as follows:

1. The factor $(x_r - x_s)$ in P is changed to $(x_s - x_r)$ in $t(P)$, so this factor changes sign.
2. The factors $(x_i - x_j)$ in P with both subscripts different from r and s are unchanged by t .

3. The remaining factors in P have exactly one subscript k different from r and s and may be grouped into pairs according to the following plan.
- If $k < r < s$, the pair $(x_k - x_r)(x_k - x_s)$ becomes $(x_k - x_s)(x_k - x_r)$, and their product is unchanged by the transposition t .
 - Similarly, if $r < s < k$, the product $(x_r - x_k)(x_s - x_k)$ is also unchanged by t .
 - Finally, if $r < k < s$, then the pair $(x_r - x_k)(x_k - x_s)$ is unchanged by t since

$$\begin{aligned}(x_s - x_k)(x_k - x_r) &= [-(x_k - x_s)][-(x_r - x_k)] \\ &= (x_k - x_s)(x_r - x_k).\end{aligned}$$

Thus $t(P) = (-1)P$, and the proof of the lemma is complete.

Strategy ■ The conclusion in the next theorem has the form “ r or s .” In previous conclusions of this type, we have assumed that r was false and proved that s must then be true. It is interesting to note that this time, our technique is different and uses no negative assumption.

Theorem 4.3 ■ Products of Transpositions

If a certain permutation f is expressed as a product of p transpositions and also as a product of q transpositions, then either p and q are both even, or else p and q are both odd.

$(u \wedge v)$ **Proof** Suppose
 $\Rightarrow (r \vee s)$

$$f = t_1 t_2 \cdots t_p \text{ and } f = t'_1 t'_2 \cdots t'_q$$

where each t_i and each t'_j are transpositions. With the first factorization, the result of applying f to

$$P = \prod_{i < j}^n (x_i - x_j)$$

can be obtained by successive application of the transpositions $t_p, t_{p-1}, \dots, t_2, t_1$. By Lemma 4.2, each t_i changes the sign of P , so

$$f(P) = (-1)^p P.$$

Repeating this same line of reasoning with the second factorization, we obtain

$$f(P) = (-1)^q P.$$

This means that

$$(-1)^p P = (-1)^q P,$$

and consequently,

$$(-1)^p = (-1)^q.$$

Therefore, either p or q are both even, or p and q are both odd.

Theorem 4.3 assures us that when a particular permutation is expressed in different ways as a product of transpositions, the number of transpositions used either will always be an even number or else will always be an odd number. This fact allows us to make the following definition.

Definition 4.4 ■ Even, Odd Permutations

A permutation that can be expressed as a product of an even number of transpositions is called an **even permutation**, and a permutation that can be expressed as a product of an odd number of transpositions is called an **odd permutation**.

The product fg in Example 7 was written as a product of six transpositions and then as a product of four transpositions, and fg is an even permutation.

The factorization of an r -cycle (i_1, i_2, \dots, i_r) as

$$(i_1, i_2, \dots, i_r) = (i_1, i_r)(i_1, i_{r-1}) \cdots (i_1, i_3)(i_1, i_2)$$

uses $r - 1$ transpositions. This shows that *an r -cycle is an even permutation if r is odd and an odd permutation if r is even*. The identity is an even permutation since $e = (1, 2)(1, 2)$. The product of two even permutations is clearly an even permutation. Since any permutation can be written as a product of disjoint cycles, and since the inverse of an r -cycle is an r -cycle, the inverse of an even permutation is an even permutation. These remarks show that the set A_n of all even permutations in S_n is a subgroup of S_n . It is called the *alternating group* on n elements.

Definition 4.5 ■ Alternating Group

The **alternating group** A_n is the subgroup of S_n that consists of all even permutations in S_n .

Example 8 The elements of the group A_4 are as follows:

$$\begin{array}{cccc} (1) & (1, 2, 4) & (1, 4, 2) & (1, 2)(3, 4) \\ (1, 2, 3) & (1, 4, 3) & (2, 3, 4) & (1, 3)(2, 4) \\ (1, 3, 2) & (1, 3, 4) & (2, 4, 3) & (1, 4)(2, 3). \end{array}$$

■

The concept of conjugate elements in a group is basic to the study of normal subgroups. This concept is defined as follows.

Definition 4.6 ■ Conjugate Elements

If a and b are elements of the group G , the **conjugate** of a by b is the element bab^{-1} . We say that $c \in G$ is a **conjugate** of a if and only if $c = bab^{-1}$ for some b in G .

We should point out that this concept is trivial in an abelian group G , because $bab^{-1} = bb^{-1}a = ea = a$ for all $b \in G$.

There is a procedure by which conjugates of elements in a permutation group may be computed with ease. To see how this works, suppose that f and g are permutations on $\{1, 2, \dots, n\}$ that have been written as products of disjoint cycles, and consider gfg^{-1} . If i_1 and i_2 are integers such that $f(i_1) = i_2$, then gfg^{-1} maps $g(i_1)$ to $g(i_2)$, as the following diagram shows:

$$g(i_1) \xrightarrow{g^{-1}} i_1 \xrightarrow{f} i_2 \xrightarrow{g} g(i_2).$$

This means that if

$$(i_1, i_2, \dots, i_r)$$

is one of the disjoint cycles in f , then

$$(g(i_1), g(i_2), \dots, g(i_r))$$

is a corresponding cycle in gfg^{-1} . Thus, if

$$f = (i_1, i_2, \dots, i_r)(j_1, j_2, \dots, j_s) \cdots (k_1, k_2, \dots, k_t),$$

then

$$gfg^{-1} = (g(i_1), g(i_2), \dots, g(i_r))(g(j_1), \dots, g(j_s)) \cdots (g(k_1), \dots, g(k_t)).$$

Example 9

If

$$f = (1, 3, 6, 9, 5)(2, 4, 7),$$

and

$$g = (1, 2, 8)(3, 6)(4, 5, 7),$$

then gfg^{-1} may be obtained from f as follows:

$$\begin{aligned} f &= (1, 3, 6, 9, 5)(2, 4, 7) \\ &\quad \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \\ gfg^{-1} &= (2, 6, 3, 9, 7)(8, 5, 4) \\ &= (2, 6, 3, 9, 7)(4, 8, 5), \end{aligned}$$

where the arrows indicate replacement of i by $g(i)$. This result may be verified by direct computation of g^{-1} and the product gfg^{-1} . ■

The procedure for computing conjugates described just before Example 9 shows that any conjugate of a given permutation f has the same type of factorization into disjoint cycles as f does. If suitable permutations f and h are given, the procedure also indicates how g may be found so that $gfg^{-1} = h$. This is illustrated in Example 10.

Example 10

Suppose $f = (1, 4, 2)(3, 5)$, $h = (6, 8, 9)(5, 7)$, and we wish to find g such that $gfg^{-1} = h$. Using arrows to indicate replacements in the same way as in Example 9, we wish to obtain $gfg^{-1} = h$ from f as follows:

$$\begin{aligned} f &= (1, 4, 2)(3, 5) \\ &\quad \downarrow \downarrow \downarrow \downarrow \downarrow \\ gfg^{-1} &= (6, 8, 9)(5, 7). \end{aligned}$$

From this diagram it is easy to see that

$$g = (1, 6)(4, 8)(2, 9)(3, 5, 7)$$

is a solution to our problem. It is also easy to see that g is not unique. For example,

$$(1, 6, 4, 8, 2, 9)(3, 5, 7)$$

is another value of g that works just as well. ■

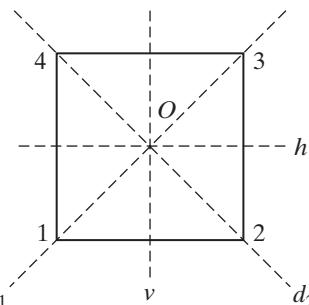
In Example 2 of Section 3.5, we considered the group of all *rigid motions*, or *symmetries*, of an equilateral triangle. Every geometric figure has an associated group of rigid motions. (We are considering only rigid motions in space here. For a plane figure, one can similarly consider rigid motions of the figure in that plane.) For simple figures such as a square, a regular pentagon, or a cube, a rigid motion is completely determined by the images of the vertices. If the vertices are labeled 1, 2, 3, . . . rather than V_1, V_2, V_3, \dots , the rigid motions may be represented by permutation notation. In Example 2 of Section 3.5, the mappings

$$h: \begin{cases} h(V_1) = V_2 \\ h(V_2) = V_1 \\ h(V_3) = V_3 \end{cases} \quad \text{and} \quad r: \begin{cases} r(V_1) = V_2 \\ r(V_2) = V_3 \\ r(V_3) = V_1 \end{cases}$$

can be written simply as

$$h = (1, 2) \quad \text{and} \quad r = (1, 2, 3).$$

Example 11 Using the notational convention described in the preceding paragraph, we shall write out the (space) group G of rigid motions of a square (see Figure 4.1).



■ **Figure 4.1**

The elements of the group G are as follows:

1. the identity mapping $e = (1)$
2. the counterclockwise rotation $\alpha = (1, 2, 3, 4)$ through 90° about the center O
3. the counterclockwise rotation $\alpha^2 = (1, 3)(2, 4)$ through 180° about the center O
4. the counterclockwise rotation $\alpha^3 = (1, 4, 3, 2)$ through 270° about the center O
5. the reflection $\beta = (1, 4)(2, 3)$ about the horizontal line h

6. the reflection $\gamma = (2, 4)$ about the diagonal d_1
7. the reflection $\Delta = (1, 2)(3, 4)$ about the vertical line v
8. the reflection $\theta = (1, 3)$ about the diagonal d_2 .

The group $G = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$ of rigid motions of the square is known as the **octic group**. The multiplication table for G is requested in Exercise 18 of this section. ■

Exercises 4.1

True or False

Label each of the following statements as either true or false.

1. Every permutation can be written as a product of transpositions.
2. A permutation can be uniquely expressed as a product of transpositions.
3. The product of cycles under mapping composition is a commutative operation.
4. Disjoint cycles commute under mapping composition.
5. The identity permutation can be expressed in more than one way.
6. Every permutation can be expressed as a product of disjoint cycles.
7. An r -cycle is an even permutation if r is even and an odd permutation if r is odd.
8. The set of all odd permutations in S_n is a subgroup of S_n .
9. The symmetric group S_n on n elements has order n .
10. A transposition leaves all elements except two fixed.
11. The order of an r -cycle is r .
12. The mutually disjoint cycles of a permutation are the same as its orbits.

Exercises

1. Express each permutation as a product of disjoint cycles and find the orbits of each permutation.

a.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 3 & 1 & 2 \end{bmatrix}$$

b.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 5 & 4 \end{bmatrix}$$

c.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 5 & 2 \end{bmatrix}$$

d.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 4 & 1 \end{bmatrix}$$

e.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 & 1 & 2 & 7 \end{bmatrix}$$

f.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 1 & 3 & 7 & 2 & 6 & 4 \end{bmatrix}$$

g.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 4 & 1 & 5 \end{bmatrix}$$

h.
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 4 & 2 \end{bmatrix}$$

2. Express each permutation as a product of disjoint cycles and find the orbits of each permutation.
- (1, 9, 2, 3)(1, 9, 6, 5)(1, 4, 8, 7)
 - (1, 2, 9)(3, 4)(5, 6, 7, 8, 9)(4, 9)
 - (1, 4, 8, 7)(1, 9, 6, 5)(1, 5, 3, 2, 9)
 - (1, 4, 2, 3, 5)(1, 3, 4, 5)
 - (1, 3, 5, 4, 2)(1, 4, 3, 5)
 - (1, 9, 2, 4)(1, 7, 6, 5, 9)(1, 2, 3, 8)
 - (2, 3, 7)(1, 2)(3, 5, 7, 6, 4)(1, 4)
 - (4, 9, 6, 7, 8)(2, 6, 4)(1, 8, 7)(3, 5)
3. In each part of Exercise 1, decide whether the permutation is even or odd.
4. In each part of Exercise 2, decide whether the permutation is even or odd.
5. Find the order of each permutation in Exercise 1.
6. Find the order of each permutation in Exercise 2.
7. Express each permutation in Exercise 1 as a product of transpositions.
8. Express each permutation in Exercise 2 as a product of transpositions.
9. Compute f^2 , f^3 , and f^{-1} for each of the following permutations.
- | | |
|-----------------------------------|--------------------------------------|
| a. $f = (1, 5, 2, 4)$ | b. $f = (2, 7, 4, 3, 5)$ |
| c. $f = (1, 6, 2)(3, 4, 5)$ | d. $f = (1, 2)(3, 5, 7, 4)$ |
| e. $f = (1, 2, 8)(3, 4, 7, 5, 6)$ | f. $f = (1, 3, 7, 4)(2, 5, 9, 8, 6)$ |
10. Compute gfg^{-1} , the conjugate of f by g , for each pair f, g .
- | | |
|------------------------------|--------------------------|
| a. $f = (1, 2, 4, 3);$ | $g = (1, 3, 2)$ |
| b. $f = (1, 3, 5, 6);$ | $g = (2, 5, 4, 6)$ |
| c. $f = (2, 3, 5, 4);$ | $g = (1, 3, 2)(4, 5)$ |
| d. $f = (1, 4)(2, 3);$ | $g = (1, 2, 3)$ |
| e. $f = (1, 3, 5)(2, 4);$ | $g = (2, 5)(3, 4)$ |
| f. $f = (1, 3, 5, 2)(4, 6);$ | $g = (1, 3, 6)(2, 4, 5)$ |
11. For the given permutations, f and h , find a permutation g such that h is the conjugate of f by g —that is, such that $h = gfg^{-1}$.
- | | |
|------------------------------|--------------------------|
| a. $f = (1, 5, 9);$ | $h = (2, 6, 4)$ |
| b. $f = (1, 3, 5, 7);$ | $h = (3, 4, 6, 8)$ |
| c. $f = (1, 3, 5)(2, 4);$ | $h = (2, 4, 3)(1, 5)$ |
| d. $f = (1, 2, 3)(4, 5);$ | $h = (2, 3, 4)(1, 6)$ |
| e. $f = (1, 4, 7)(2, 5, 8);$ | $h = (1, 5, 4)(2, 3, 6)$ |
| f. $f = (1, 3, 5)(2, 4, 6);$ | $h = (1, 2, 4)(3, 5, 6)$ |

- Sec. 3.4, #39 ➤ **12.** Write the permutation $f = (1, 2, 3, 4, 5, 6)$ as a product of a permutation g of order 2 and a permutation h of order 3.
- Sec. 3.4, #39 ➤ **13.** Write the permutation $f = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$ as a product of a permutation g of order 3 and h of order 4.
- 14.** List all the elements of the alternating group A_3 , written in cyclic notation.
- 15.** List all the elements of S_4 , written in cyclic notation.
- 16.** Find all the distinct cyclic subgroups of A_4 .
- 17.** Find cyclic subgroups of S_4 that have three different orders.
- 18.** Construct a multiplication table for the octic group described in Example 11 of this section.
- 19.** Find all the distinct cyclic subgroups of the octic group in Exercise 18.
- Sec. 3.1, #30 ➤ **20.** Find an isomorphism from the octic group G in Example 11 of this section to the group $G' = \{I_2, R, R^2, R^3, H, D, V, T\}$ in Exercise 30 of Section 3.1.
- 21.** Prove that in any group, the relation “ x is a conjugate of y ” is an equivalence relation.
- Sec. 3.3, #26 ➤ **22.** As stated in Exercise 26 of Section 3.3, the **centralizer** of an element a in the group G is the subgroup given by $C_a = \{x \in G \mid ax = xa\}$. Use the multiplication table constructed in Exercise 18 to find the centralizer C_a for each element a of the octic group.
- 23.** A subgroup H of the group S_n is called **transitive** on $B = \{1, 2, \dots, n\}$ if for each pair i, j of elements of B there exists an element $h \in H$ such that $h(i) = j$. Show that there exists a cyclic subgroup H of S_n that is transitive on B .
- Sec. 4.4, #25 ⇐ **24.** Let ϕ be the mapping from S_n to the additive group \mathbf{Z}_2 defined by
- $$\phi(f) = \begin{cases} [0] & \text{if } f \text{ is an even permutation} \\ [1] & \text{if } f \text{ is an odd permutation.} \end{cases}$$
- a.** Prove that ϕ is a homomorphism.
b. Find the kernel of ϕ .
c. Prove or disprove that ϕ an epimorphism.
d. Prove or disprove that ϕ an isomorphism.
- 25.** Let f and g be disjoint cycles in S_n . Prove that $fg = gf$.
- 26.** Prove that the order of A_n is $\frac{n!}{2}$.

4.2**Cayley's[†] Theorem**

At the opening of Section 3.5, we stated that permutation groups can serve as models for all groups. A more precise statement is that every group is isomorphic to a group of permutations; this is the reason for the fundamental importance of permutation groups in algebra.

Theorem 4.7 ■ **Cayley's Theorem**

Every group is isomorphic to a group of permutations.

$p \Rightarrow q$ **Proof** Let G be a given group. The permutations that we use in the proof will be mappings defined on the set of all elements in G .

For each element a in G , we define a mapping $f_a: G \rightarrow G$ by

$$f_a(x) = ax \text{ for all } x \text{ in } G.$$

That is, the image of each x in G is obtained by multiplying x on the left by a . Now f_a is one-to-one since

$$\begin{aligned} f_a(x) = f_a(y) &\Rightarrow ax = ay \\ &\Rightarrow x = y. \end{aligned}$$

To see that f_a is onto, let b be arbitrary in G . Then $x = a^{-1}b$ is in G , and for this particular x we have

$$\begin{aligned} f_a(x) &= ax \\ &= a(a^{-1}b) = b. \end{aligned}$$

Thus f_a is a permutation on the set of elements of G .

We shall show that the set

$$G' = \{f_a \mid a \in G\}$$

actually forms a group of permutations. Since mapping composition is always associative, we need only show that G' is closed, has an identity, and contains inverses.

For any f_a and f_b in G' , we have

$$f_a f_b(x) = f_a(f_b(x)) = f_a(bx) = a(bx) = (ab)(x) = f_{ab}(x)$$

for all x in G . Thus $f_a f_b = f_{ab}$, and G' is closed. Since

$$f_e(x) = ex = x$$

for all x in G , f_e is the identity permutation, $f_e = I_G$. Using the result $f_a f_b = f_{ab}$, we have

$$f_a f_{a^{-1}} = f_{aa^{-1}} = f_e$$

[†]A biographical sketch of Author Cayley (1821–1895) is given at the end of Chapter 1.

and

$$f_{a^{-1}} f_a = f_{a^{-1}a} = f_e.$$

Thus $(f_a)^{-1} = f_{a^{-1}}$ is in G' , and G' is a group of permutations.

All that remains is to show that G is isomorphic to G' . The mapping $\phi: G \rightarrow G'$ defined by

$$\phi(a) = f_a$$

is clearly onto. It is one-to-one since

$$\begin{aligned}\phi(a) = \phi(b) &\Rightarrow f_a = f_b \\ &\Rightarrow f_a(x) = f_b(x) \quad \text{for all } x \in G \\ &\Rightarrow ax = bx \quad \text{for all } x \in G \\ &\Rightarrow a = b.\end{aligned}$$

Finally, ϕ is an isomorphism since

$$\phi(a)\phi(b) = f_a f_b = f_{ab} = \phi(ab)$$

for all a, b in G .

Note that the group $G' = \{f_a \mid a \in G\}$ is a subgroup of the group $S(G)$ of all permutations on G , and $G' \neq S(G)$ in most cases.

Example 1 We shall follow the proof of Cayley's Theorem with the group $G = \{1, i, -1, -i\}$ to obtain a group of permutations G' that is isomorphic to G and an isomorphism from G to G' .

With $f_a: G \rightarrow G$ defined by $f_a(x) = ax$ for each $a \in G$, we obtain the following permutations on the set of elements of G :

$$\begin{array}{ll}f_1: \begin{cases} f_1(1) = 1 \\ f_1(i) = i \\ f_1(-1) = -1 \\ f_1(-i) = -i \end{cases} & f_i: \begin{cases} f_i(1) = i \\ f_i(i) = -1 \\ f_i(-1) = -i \\ f_i(-i) = 1 \end{cases} \\ f_{-1}: \begin{cases} f_{-1}(1) = -1 \\ f_{-1}(i) = -i \\ f_{-1}(-1) = 1 \\ f_{-1}(-i) = i \end{cases} & f_{-i}: \begin{cases} f_{-i}(1) = -i \\ f_{-i}(i) = 1 \\ f_{-i}(-1) = i \\ f_{-i}(-i) = -1 \end{cases}\end{array}$$

In a more compact form, we write

$$\begin{array}{ll}f_1 = (1) & f_i = (1, i, -1, -i) \\ f_{-1} = (1, -1)(i, -i) & f_{-i} = (1, -i, -1, i).\end{array}$$

According to the proof of Cayley's Theorem, the set

$$G' = \{f_1, f_i, f_{-1}, f_{-i}\}$$

is a group of permutations, and the mapping $\phi: G \rightarrow G'$ defined by

$$\phi: \begin{cases} \phi(1) = f_1 \\ \phi(i) = f_i \\ \phi(-1) = f_{-1} \\ \phi(-i) = f_{-i} \end{cases}$$

is an isomorphism from G to G' . ■

Exercises 4.2

True or False

Label the following statement as either true or false.

1. Every finite group G of order n is isomorphic to a subgroup of order n of the group $S(G)$ of all permutations on G .

Exercises

In Exercises 1–7, let G be the given group. Write out the elements of a group of permutations that is isomorphic to G , and exhibit an isomorphism from G to this group.

1. Let G be the additive group \mathbf{Z}_3 .
2. Let G be the cyclic group $\langle a \rangle$ of order 5.
3. Let G be the Klein four group $\{e, a, b, ab\}$ with its multiplication table given in Figure 4.2.

.	e	a	b	ab
e	e	a	b	ab
a	a	e	ab	b
b	b	ab	e	a
ab	ab	b	a	e

Figure 4.2

4. Let G be the multiplicative group of units $\mathbf{U}_5 = \{[1], [2], [3], [4]\} \subseteq \mathbf{Z}_5$.
5. Let G be the multiplicative group $\{[2], [4], [6], [8]\} \subseteq \mathbf{Z}_{10}$.
6. Let G be the group of permutations matrices $\{I_3, P_1, P_2, P_3, P_4, P_5\}$ as given in Exercise 29 of Section 3.1.
7. Let G be the octic group $\{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$.

8. For each a in the group G , define a mapping $h_a: G \rightarrow G$ by $h_a(x) = xa$ for all x in G .
 - a. Prove that each h_a is a permutation on the set of elements in G .
 - b. Prove that $H = \{h_a | a \in G\}$ is a group with respect to mapping composition.
 - c. Define $\phi: G \rightarrow H$ by $\phi(a) = h_a$ for each a in G . Determine whether ϕ is always an isomorphism.
9. For each element a in the group G , define a mapping $k_a: G \rightarrow G$ by $k_a(x) = xa^{-1}$ for all x in G .
 - a. Prove that each k_a is a permutation on the set of elements of G .
 - b. Prove that $K = \{k_a | a \in G\}$ is a group with respect to mapping composition.
 - c. Define $\phi: G \rightarrow K$ by $\phi(a) = k_a$ for each a in G . Determine whether ϕ is always an isomorphism.
10. For each a in the group G , define a mapping $m_a: G \rightarrow G$ by $m_a(x) = a^{-1}x$ for all x in G .
 - a. Prove that each m_a is a permutation on the set of elements of G .
 - b. Prove that $M = \{m_a | a \in G\}$ is a group with respect to mapping composition.
 - c. Define $\phi: G \rightarrow M$ by $\phi(a) = m_a$ for each a in G . Determine whether ϕ is always an isomorphism.

4.3

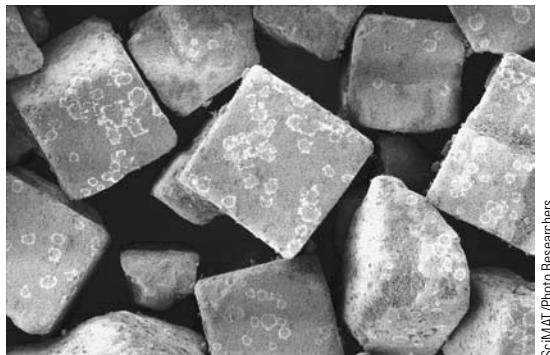
Permutation Groups in Science and Art (Optional)

Often, the usefulness of some particular knowledge in mathematics is neither obvious nor simple. So it is with permutation groups. Their applications in the real world come about through connections that are somewhat involved. Nevertheless, we shall indicate here some of their uses in both science and art.

Most of the scientific applications of permutation groups are in physics and chemistry. One of the most impressive applications occurred in 1962. In that year, physicists Murray Gell-Mann and Yuval Ne'eman used group theory to predict the existence of a new particle, which was designated the *omega minus particle*. It was not until 1964 that the existence of this particle was confirmed in laboratory experiments.

One of the most extensive uses made of permutation groups has been in the science of crystallography. As mentioned in Section 4.1, every geometric figure in two or three dimensions has its associated rigid motions, or *symmetries*. This association provides a natural connection between permutation groups and many objects in the real world. One of the most fruitful of these connections has been made in the study of the structure of crystals. Crystals are classified according to geometric symmetry based on a structure with a balanced arrangement of faces. One of the simplest and most common examples of such a structure is provided by the fact that a common table salt (NaCl) crystal is in the shape of a cube. (See photo on the next page.)

In this section, we examine some groups related to the rigid motions of a plane figure. We have already seen two examples of this type of group. The first was the group of symmetries of an equilateral triangle in Example 2 of Section 3.5, and the other was the group of symmetries of a square in Example 11 of Section 4.1.



Salt crystals are in the form of cubes.

It is not hard to see that the symmetries of any plane figure F form a group under mapping composition. We already know that the permutations on the set F form a group $\mathcal{S}(F)$ with respect to mapping composition. The identity permutation I_F preserves distances and consequently is a symmetry of F . If two permutations on F preserve distances, their composition does also, and if a given permutation preserves distances, its inverse does also. Thus the symmetries of F form a subgroup of $\mathcal{S}(F)$.

Before we consider some other specific plane figures F , a discussion of the term *symmetry* is in order. In agreement with conventional terminology in algebra, we have used the word *symmetry* to refer to a rigid motion of a geometric figure. However, the term is commonly used in another way. For example, the pentagon shown in Figure 4.3 is said to have *symmetry* with respect to the vertical line ℓ through the center O and the vertex at the top, or to be *symmetric* with respect to ℓ . To make a distinction between the two uses, we shall use the phrase *geometric symmetry* for the latter type of symmetry.

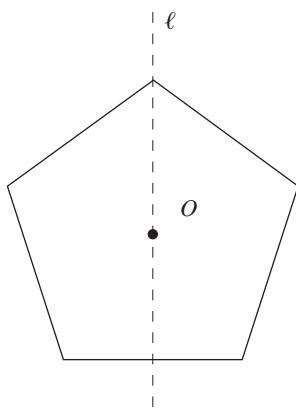


Figure 4.3

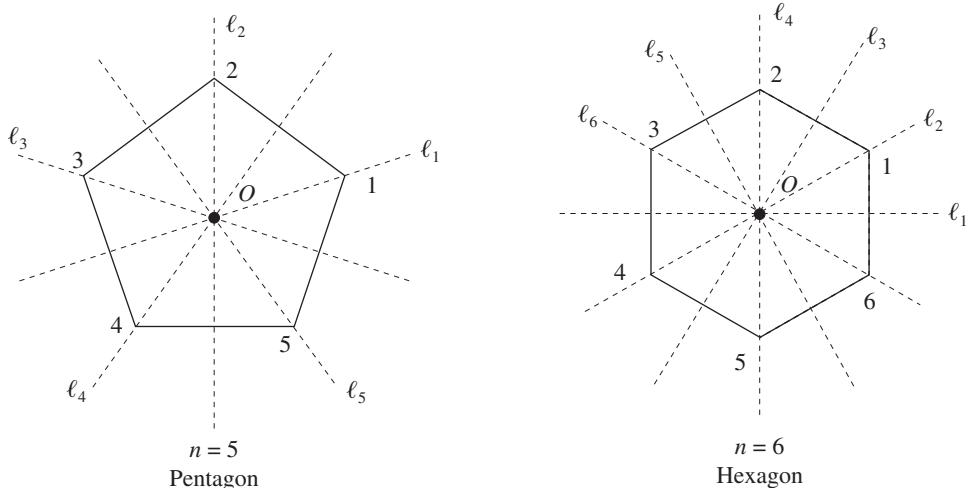
The groups of symmetries for regular polygons with three or four sides generalize to a regular polygon P with n sides, for any positive integer $n > 4$. Any symmetry f of P is determined by the images of the vertices of P . Let the vertices be numbered $1, 2, \dots, n$,

and consider the mapping that makes the symmetry f of P correspond to the permutation on $\{1, 2, \dots, n\}$ that has the matrix form

$$\begin{bmatrix} 1 & 2 & \dots & n \\ f(1) & f(2) & \dots & f(n) \end{bmatrix}.$$

Since f is completely determined by the images of the vertices, this mapping is clearly a bijection between the rigid motions of P and a subset D_n of the symmetric group S_n of all permutations on $\{1, 2, \dots, n\}$. This mapping is in fact an isomorphism, D_n is a subgroup (called the **dihedral group**) of S_n , and we identify the rigid motions of P with the elements of D_n in the same way that we did in Example 11 of Section 4.1.

Regular polygons with $n = 5$ (a pentagon) and $n = 6$ (a hexagon) are shown in Figure 4.4. Bearing in mind that a symmetry is determined by the images of the vertices, it can be seen that D_n consists of n counterclockwise rotations and n reflections about a line through the center O of P . If n is odd, each reflection is about a line through a vertex and the midpoint of the opposite side. If n is even, half of the reflections are about lines through pairs of opposite vertices, and the other half are about lines through midpoints of opposite sides. Thus D_n has order $2n$.



■ Figure 4.4

Example 1 Consider the pentagon in Figure 4.4. If we let R denote the rotation of $\frac{360^\circ}{5} = 72^\circ$ counterclockwise about the center O , then all possible rotations in D_5 are found in the following list:

$$\begin{aligned} R &= (1, 2, 3, 4, 5), \quad R^2 = (1, 3, 5, 2, 4), \quad R^3 = (1, 4, 2, 5, 3), \\ R^4 &= (1, 5, 4, 3, 2), \quad R^5 = (1). \end{aligned}$$

If we let L_k denote the reflection about line ℓ_k for $k = 1, 2, 3, 4, 5$, then the reflections in D_5 appear as follows in cyclic notation:

$$\begin{aligned} L_1 &= (2, 5)(3, 4), \quad L_2 = (1, 3)(4, 5), \quad L_3 = (1, 5)(2, 4), \\ L_4 &= (1, 2)(3, 5), \quad L_5 = (1, 4)(2, 3). \end{aligned}$$

Direct computations verify that

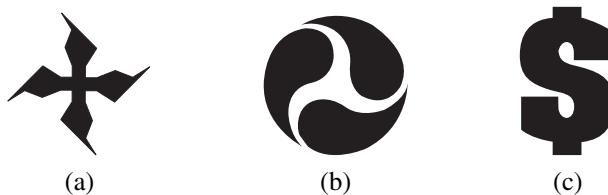
$$L_1R = L_3, \quad L_1R^2 = L_5, \quad L_1R^3 = L_2, \quad \text{and} \quad L_1R^4 = L_4.$$

Thus the elements of D_5 can be listed in the form

$$D_5 = \{I, R, R^2, R^3, R^4, L_1, L_1R, L_1R^2, L_1R^3, L_1R^4\}. \quad \blacksquare$$

All the symmetries in our examples have been either rotations or reflections about a line. This is no accident because these are the only kinds of symmetries that exist for a bounded nonempty set. If the group of symmetries of a certain figure contains a rotation different from the identity mapping, then the figure is said to possess **rotational symmetry**. A figure with a group of symmetries that includes a reflection about a line is said to have **reflective symmetry**.

Example 2 Each part of Figure 4.5 has a group of symmetries that consists entirely of rotations, and each possesses only rotational symmetry. In contrast, the group of symmetries of the pentagon contains both reflections and rotations, and the pentagon has both reflective symmetry and rotational symmetry.



■ Figure 4.5

We have barely touched on the subject of symmetries in this section, concentrating primarily on bounded nonempty sets in the plane. When attention is extended to unbounded sets in the plane, there are two more types of symmetries that can be considered: translations and glide reflections.

A **translation** is simply a sliding (or glide) of the entire object through a certain distance in a fixed direction. A **glide reflection** consists of a translation (or glide) followed by a reflection about a line parallel to the direction of the translation. These types of symmetries are treated in detail in more advanced books than this one, and it can be shown that there are only four kinds of symmetries for plane figures: *rotations, reflections, translations, and glide reflections*.

As our final example in this section, we consider the group of symmetries of an unbounded set.

Example 3 The unbounded set shown in Figure 4.6 is composed of a horizontal string of copies of the letter **R**, equally spaced one unit from the beginning of one **R** to the beginning of the next **R**, and endless in both directions.



■ Figure 4.6

If t denotes a translation of the set in Figure 4.6 one unit to the right, then t^2 is a translation two units to the right and t^n is a translation n units to the right for any positive integer n . Thus all positive integral powers of t are symmetries on the set of \mathbf{R} 's. The inverse mapping t^{-1} is a translation of the set one unit to the left, and t^{-n} is a translation n units to the left for any positive integer n . Thus all integral powers of t are symmetries on the set of \mathbf{R} 's, and the set

$$\{\cdots, t^{-2}, t^{-1}, t^0 = I, t, t^2, \cdots\}$$

is the (infinite) group of symmetries of this set. ■

Translations and glide reflections are common in the group of symmetries for wallpaper patterns, textile patterns, pottery, ribbons, and all sorts of decorative art. The interested reader can find an excellent exposition of the applications that we have touched on in Tannenbaum and Arnold's *Excursions in Modern Mathematics*, 3rd ed. (Englewood Cliffs, NJ: Prentice Hall, 1998).

The outstanding connection between permutation groups and art is provided by the famous works of the great Dutch artist M. C. Escher.[†] Concerning Escher, J. Taylor Hollist said, "Mathematicians continue to use his periodic patterns of animal figures as clever illustrations of translation, rotation and reflection symmetry. Psychologists use his optical illusions and distorted views of life as enchanting examples in the study of vision."^{††}

Exercises 4.3

True or False

Label each of the following statements as either true or false.

1. The symmetries of any plane figure form a group under mapping composition.
2. The regular pentagon possesses only rotational symmetry.
3. The regular hexagon possesses both rotational and reflective symmetry.
4. The group D_n of symmetries for a regular polygon with n sides has order n .
5. The symmetric group S_3 on 3 elements is the same as the group D_3 of symmetries for an equilateral triangle. That is, $S_3 = D_3$.
6. The symmetric group S_4 on 4 elements is the same as the group D_4 of symmetries for a square. That is, $S_4 = D_4$.
7. The alternating group A_4 on 4 elements is the same as the group D_4 of symmetries for a square. That is, $A_4 = D_4$.

[†]Maurits Cornelis Escher (1898–1972) was a Dutch graphic artist. He is known for his explorations of infinity in his mathematically inspired art. Some of his original works are housed in leading public and private collections. The asteroid 4444 is named in his honor.

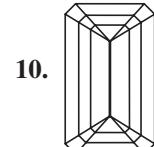
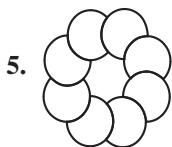
^{††}J. Taylor Hollist, "Escher Correspondence in the Roosevelt Collection," *Leonardo*, Vol. 24, No. 3 (1991), p. 329.

Exercises

List all elements in the group of symmetries of the given set.

1. The letter **T**
2. The letter **M**
3. The letter **S**
4. The letter **H**

Determine whether the given figure has rotational symmetry or reflective symmetry.



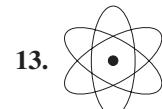
Describe the elements in the group of symmetries of the given bounded figure.



Recycle



Crafted With Pride



Atom



Biohazard



Radiation

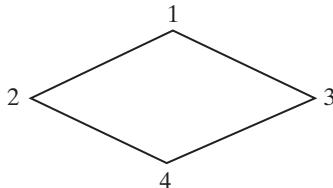


Do Not Dry Clean

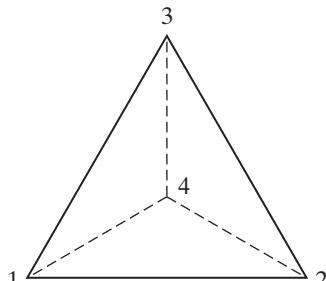
Describe the elements in the groups of symmetries of the given unbounded figures.



21. Show that the group of symmetries in Example 3 of this section is isomorphic to the group of integers under addition.
22. Construct a multiplication table for the group G of rigid motions of an isosceles triangle with vertices 1, 2, 3 if the isosceles triangle is not an equilateral triangle.
23. Construct a multiplication table for the group G of rigid motions of a rectangle with vertices 1, 2, 3, 4 if the rectangle is not a square.
24. Construct a multiplication table for the group G of rigid motions of the rhombus in Figure 4.7 with vertices 1, 2, 3, 4 if the rhombus is not a square.

**Figure 4.7**

- Sec. 4.6, #4 <
25. Construct a multiplication table for the group G of rigid motions of a regular pentagon with vertices 1, 2, 3, 4, 5.
 26. List the elements of the group G of rigid motions of a regular hexagon with vertices 1, 2, 3, 4, 5, 6.
 27. Let G be the group of rigid motions of a cube. Find the order $o(G)$.
 28. Let G be the group of rigid motions of a regular tetrahedron (see Figure 4.8). Find the order $o(G)$.

**Figure 4.8**

- Sec. 3.3, #15a > **29.** Find an isomorphism from the group G in Exercise 23 of this exercise set to the multiplicative group

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

4.4**Cosets of a Subgroup**

The binary operation in a given group can be used in a natural way to define a product between subsets of the group. The importance of this product is difficult to appreciate at this point in our development. It leads to the definition of *cosets*; cosets in turn lead to *quotient groups*; and quotient groups provide a systematic description of all *homomorphic images* of a group in Section 4.6.

Definition 4.8 ■ Product of Subsets

Let A and B be nonempty subsets of the group G . The **product** AB is defined by

$$AB = \{x \in G \mid x = ab \text{ for some } a \in A, b \in B\}.$$

This product is formed by using the group operation in G . A more precise formulation would be

$$A * B = \{x \in G \mid x = a * b \text{ for some } a \in A, b \in B\},$$

where $*$ is the group operation in G .

Several properties of this product are worth mentioning. For nonempty subsets A , B , and C of the group G ,

$$\begin{aligned} A(BC) &= \{a(bc) \mid a \in A, b \in B, c \in C\} \\ &= \{(ab)c \mid a \in A, b \in B, c \in C\} \\ &= (AB)C. \end{aligned}$$

It is obvious that

$$B = C \Rightarrow AB = AC \text{ and } BA = CA,$$

but we must be careful about the order because AB and BA may be different sets.

Example 1 Consider the subsets $A = \{(1, 2, 3), (1, 2)\}$ and $B = \{(1, 3), (2, 3)\}$ in $G = S_3$. We have

$$\begin{aligned} AB &= \{(1, 2, 3)(1, 3), (1, 2)(1, 3), (1, 2, 3)(2, 3), (1, 2)(2, 3)\} \\ &= \{(2, 3), (1, 3, 2), (1, 2), (1, 2, 3)\} \end{aligned}$$

and

$$\begin{aligned} BA &= \{(1, 3)(1, 2, 3), (2, 3)(1, 2, 3), (1, 3)(1, 2), (2, 3)(1, 2)\} \\ &= \{(1, 2), (1, 3), (1, 2, 3), (1, 3, 2)\}, \end{aligned}$$

so $AB \neq BA$. ■

For a nonabelian group G , we would probably expect AB and BA to be different. A fact that is not quite so “natural” is that

$$AB = AC \not\Rightarrow B = C.$$

Example 2 An example where $AB = AC$ but $B \neq C$ is provided by $A = \{(1, 2, 3), (1, 3, 2)\}$, $B = \{(1, 3), (2, 3)\}$, and $C = \{(1, 2), (1, 3)\}$ in $G = S_3$. Straightforward calculations show that

$$AB = \{(2, 3), (1, 2), (1, 3)\} = AC,$$

but $B \neq C$. ■

If $B = \{g\}$ consists of a single element g of a group G , then AB is written simply as Ag instead of as $A\{g\}$:

$$Ag = \{x \in G \mid x = ag \text{ for some } a \in A\}.$$

Similarly,

$$gA = \{x \in G \mid x = ga \text{ for some } a \in A\}.$$

This is one instance in which a cancellation law does hold:

$$gA = gB \Rightarrow A = B.$$

This is true because

$$\begin{aligned} gA = gB &\Rightarrow g^{-1}(gA) = g^{-1}(gB) \\ &\Rightarrow (g^{-1}g)A = (g^{-1}g)B \\ &\Rightarrow eA = eB \\ &\Rightarrow A = B. \end{aligned}$$

For convenience of reference, we summarize these results in a theorem.

Theorem 4.9 ■ Properties of the Product of Subsets

Let A , B , and C denote nonempty subsets of the group G , and let g denote an element of G . Then the following statements hold:

- a. $A(BC) = (AB)C$.
- b. $B = C$ implies $AB = AC$ and $BA = CA$.
- c. The product AB is not commutative.
- d. $AB = AC$ does not imply $B = C$.
- e. $gA = gB$ implies $A = B$.

Statements **d** and **e** have obvious duals in which the common factor is on the right side.

We shall be concerned mainly with products of subsets in which one of the factors is a subgroup. The cosets of a subgroup are of special importance.

Definition 4.10 ■ Cosets

Let H be a subgroup of the group G . For any a in G ,

$$aH = \{x \in G \mid x = ah \text{ for some } h \in H\}$$

is a **left coset** of H in G . Similarly, Ha is called a **right coset** of H in G .

The left coset aH and the right coset Ha are never disjoint, since $a = ae = ea$ is in both sets. In spite of this, aH and Ha may happen to be different sets, as the next example shows.

Example 3 Consider the subgroup

$$K = \{(1), (1, 2)\}$$

of

$$G = S_3 = \{(1), (1, 2, 3), (1, 3, 2), (1, 2), (1, 3), (2, 3)\}.$$

For $a = (1, 2, 3)$, we have

$$\begin{aligned} aK &= \{(1, 2, 3), (1, 2, 3)(1, 2)\} \\ &= \{(1, 2, 3), (1, 3)\} \end{aligned}$$

and

$$\begin{aligned} Ka &= \{(1, 2, 3), (1, 2)(1, 2, 3)\} \\ &= \{(1, 2, 3), (2, 3)\}. \end{aligned}$$

In this case, $aK \neq Ka$. ■

Although a left coset of H and a right coset of H may be neither equal nor disjoint, this cannot happen with two left cosets of H . This fact is fundamental to the proof of Lagrange's Theorem (Theorem 4.13), so we designate it as a lemma.

Strategy ■ The proof of this lemma is by use of the **contrapositive**. The contrapositive of $p \Rightarrow q$ is $\sim q \Rightarrow \sim p$. As shown in the Appendix to this book, any statement and its contrapositive are logically equivalent.

The following proof illustrates a case where it is easier to prove the contrapositive than the original statement.

Lemma 4.11 ■ Left Coset Partition

Let H be a subgroup of the group G . The distinct left cosets of H in G form a partition of G ; that is, they separate the elements of G into mutually disjoint subsets.

$\sim q \Rightarrow \sim p$ **Proof** It is sufficient to show that any two left cosets of H that are not disjoint must be the same left coset.

Suppose aH and bH have at least one element in common—say, $z \in aH \cap bH$. Then $z = ah_1$ for some $h_1 \in H$, and $z = bh_2$ for some $h_2 \in H$. This means that $ah_1 = bh_2$ and $a = bh_2h_1^{-1}$. We have that $h_2h_1^{-1}$ is in H since H is a subgroup, so $a = bh_3$ where $h_3 = h_2h_1^{-1} \in H$. Now, for every $h \in H$,

$$\begin{aligned} ah &= bh_3h \\ &= bh_4 \end{aligned}$$

where $h_4 = h_3 \cdot h$ is in H . That is, $ah \in bH$ for all $h \in H$. This proves that $aH \subseteq bH$. A similar argument shows that $bH \subseteq aH$, and thus $aH = bH$.

The distinct right cosets of a subgroup H of a group G also form a partition of G . That is, Lemma 4.11 can be restated in terms of right cosets (see Exercise 7).

Example 4 Consider again the subgroup

$$K = \{(1), (1, 2)\}$$

of

$$G = S_3 = \{(1), (1, 2, 3), (1, 3, 2), (1, 2), (1, 3), (2, 3)\}.$$

In Example 3 of this section, we saw that

$$(1, 2, 3)K = \{(1, 2, 3), (1, 3)\}.$$

Since $(1, 3)$ is in this left coset, it follows from Lemma 4.11 that

$$(1, 3)K = (1, 2, 3)K = \{(1, 2, 3), (1, 3)\}.$$

Straightforward computations show that

$$(1)K = (1, 2)K = \{(1), (1, 2)\} = K$$

and

$$(2, 3)K = (1, 3, 2)K = \{(1, 3, 2), (2, 3)\}.$$

Thus the distinct left cosets of K in G are given by

$$K, (1, 2, 3)K, (1, 3, 2)K$$

and a partition of G is

$$G = K \cup (1, 2, 3)K \cup (1, 3, 2)K.$$

■

Definition 4.12 ■ Index

Let H be a subgroup of G . The number of distinct left cosets of H in G is called the **index** of H in G and is denoted by $[G : H]$.

In the proof of the next theorem, we show that if $o(G)$ is finite, then the order of any subgroup of G must divide the order of the group G .

Theorem 4.13 ■ Lagrange's[†] Theorem

If G is a finite group and H is a subgroup of G , then

$$\text{order of } G = (\text{order of } H) \cdot (\text{index of } H \text{ in } G).$$

$(p \wedge q) \Rightarrow r$ **Proof** Let G be a finite group of order n , and let H be a subgroup of G with order k . We shall show that k divides n .

From Lemma 4.11, we know that the left cosets of H in G separate the elements of G into mutually disjoint subsets. Let m be the index of H in G ; that is, there are m distinct left cosets of H in G . We shall show that each left coset has exactly k elements.

Let aH represent an arbitrary left coset of H . The mapping $\phi: H \rightarrow aH$ defined by

$$\phi(h) = ah$$

is one-to-one, because the left cancellation law holds in G . It is also onto, since any x in aH can be written as $x = ah$ for $h \in H$. Thus ϕ is a one-to-one correspondence from H to aH , and this means that aH has the same number of elements as does H .

We have the n elements of G separated into m disjoint subsets, and each subset has k elements. Therefore, $n = km$, and

$$o(G) = o(H) \cdot [G : H].$$

Lagrange's Theorem is of great value if we are interested in finding all the subgroups of a finite group. In connection with this task, it is worthwhile to record this immediate corollary.

Corollary 4.14 ■ $o(a) | o(G)$

The order of an element of a finite group must divide the order of the group.

Example 5 To illustrate the usefulness of the foregoing results, we shall exhibit all of the subgroups of S_3 . Any subgroup of S_3 must be of order 1, 2, 3, or 6, since $o(S_3) = 6$. An element in a subgroup of order 3 must have order dividing 3, and therefore any subgroup of order 3 is cyclic. Similarly, any subgroup of order 2 is cyclic. The following list is thus a complete list of the subgroups of S_3 :

$$\begin{array}{ll} H_1 = \{(1)\} & H_4 = \{(1), (2, 3)\} \\ H_2 = \{(1), (1, 2)\} & H_5 = \{(1), (1, 2, 3), (1, 3, 2)\} \\ H_3 = \{(1), (1, 3)\} & H_6 = S_3. \end{array}$$

■

[†]Joseph-Louis Lagrange (1736–1813) made significant contributions to analysis, number theory, ordinary and partial differential equations, calculus, analytical geometry, theory of equations, and to classical and celestial mechanics. Lagrange was responsible for the metric system, which resulted from his tenure on a commission for the reform of weights and measures. Napoleon designated Lagrange a count, and the crater Lagrange is so named in his honor.

It can be shown that if p is a prime, then any group of order p must be cyclic (see Exercise 21 at the end of this section). This means that, up to an isomorphism, there is only one group of order p , if p is a prime. In particular, the only groups of order 2, 3, or 5 are the cyclic groups.

By examination of the possible orders of the elements and the possible multiplication tables, it can be shown that a group of order 4 either is cyclic or is isomorphic to the Klein four group

$$G = \{e, a, b, ab = ba\}$$

of Exercise 10 in Section 3.5. Hence every group of order 4 is abelian.

Exercises 4.4

True or False

Label each of the following statements as either true or false.

1. $aH \cap Ha \neq \emptyset$ where H is any subgroup of a group G and $a \in G$.
2. Let H be any subgroup of a group G . Then H is a left coset of H in G .
3. Let H be any subgroup of a group G and $a \in G$. Then $aH = Ha$.
4. The elements of G can be separated into mutually disjoint subsets using either left cosets or right cosets of a subgroup H of G .
5. The order of an element of a finite group divides the order of the group.
6. The order of any subgroup of a finite group divides the order of the group.
7. Let H be a subgroup of a finite group G . The index of H in G must divide the order of G .
8. Every left coset of a group G is a subgroup of G .

Exercises

In Exercises 1 and 2, let G be the octic group $\{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$ in Example 11 of Section 4.1, with its multiplication table requested in Exercise 18 of the same section.

- Sec. 4.5, #1 <<
1. Let H be the subgroup $\{e, \beta\}$ of the octic group G .
 - a. Find the distinct left cosets of H in G , write out their elements, and partition G into left cosets of H .
 - b. Find the distinct right cosets of H in G , write out their elements, and partition G into right cosets of H .
 2. Let H be the subgroup $\{e, \Delta\}$ of the octic group G .
 - a. Find the distinct left cosets of H in G , write out their elements, and partition G into left cosets of H .
 - b. Find the distinct right cosets of H in G , write out their elements, and partition G into right cosets of H .
- Sec. 4.5, #1 <<

3. Let H be the subgroup $\{(1), (1, 2)\}$ of S_3 .

a. Find the distinct left cosets of H in S_3 , write out their elements, and partition S_3 into left cosets of H .

b. Find the distinct right cosets of H in S_3 , write out their elements, and partition S_3 into right cosets of H .

- Sec. 4.5, #1 ≪ 4. Let H be the subgroup $\{(1), (2, 3)\}$ of S_3 .

a. Find the distinct left cosets of H in S_3 , write out their elements, and partition S_3 into left cosets of H .

b. Find the distinct right cosets of H in S_3 , write out their elements, and partition S_3 into right cosets of H .

In Exercises 5 and 6, let G be the multiplicative group of permutation matrices $\{I_3, P_3, P_3^2, P_1, P_4, P_2\}$ in Example 4 of Section 3.5.

5. Let H be the subgroup of G given by

$$H = \{I_3, P_4\} = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right\}.$$

a. Find the distinct left cosets of H in G , write out their elements, and partition G into left cosets of H .

b. Find the distinct right cosets of H in G , write out their elements, and partition G into right cosets of H .

- Sec. 4.5, #1 ≪ 6. Let H be the subgroup of G given by

$$H = \{I_3, P_3, P_3^2\} = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \right\}.$$

a. Find the distinct left cosets of H in G , write out their elements, and partition G into left cosets of H .

b. Find the distinct right cosets of H in G , write out their elements, and partition G into right cosets of H .

7. Let H be a subgroup of the group G . Prove that if two right cosets Ha and Hb are not disjoint, then $Ha = Hb$ —that is, the distinct right cosets of H in G form a partition of G .

- Sec. 3.3, #11 ≫ 8. Let H be a subgroup of a group G .

a. Prove that gHg^{-1} is a subgroup of G for any $g \in G$. We say that gHg^{-1} is a **conjugate** of H and that H and gHg^{-1} are **conjugate subgroups**.

b. Prove that if H is abelian, then gHg^{-1} is abelian.

c. Prove that if H is cyclic, then gHg^{-1} is cyclic.

d. Prove that H and gHg^{-1} are isomorphic.

9. For an arbitrary subgroup H of the group G , define the mapping θ from the set of left cosets of H in G to the set of right cosets of H in G by $\theta(aH) = Ha^{-1}$. Prove that θ is a bijection.
10. Let H be a subgroup of the group G . Prove that the index of H in G is the number of distinct right cosets of H in G .
- Sec. 3.5, #10 ➤ 11. Show that a group of order 4 either is cyclic or is isomorphic to the Klein four group $\{e, a, b, ab = ba\}$.
- Sec. 4.6, #21 < 12. Let G be a group of finite order n . Prove that $a^n = e$ for all a in G .
- Sec. 3.1, #27 ➤ 13. Find the order of each of the following elements in the multiplicative group of units \mathbf{U}_p .
- a. [2] for $p = 13$
 - b. [5] for $p = 13$
 - c. [3] for $p = 17$
 - d. [8] for $p = 17$
14. Find all subgroups of the octic group.
15. Find all subgroups of the alternating group A_4 .
16. Lagrange's Theorem states that the order of a subgroup of a finite group must divide the order of the group. Prove or disprove its converse: If k divides the order of a finite group G , then there must exist a subgroup of G having order k .
- Sec. 3.1, #28 ➤ 17. Find all subgroups of the quaternion group.
18. Find two groups of order 6 that are not isomorphic.
19. If H and K are arbitrary subgroups of G , prove that $HK = KH$ if and only if HK is a subgroup of G .
- Sec. 2.5, #51 ➤ 20. Let p be prime and G the multiplicative group of units $\mathbf{U}_p = \{[a] \in \mathbf{Z}_p | [a] \neq [0]\}$. Use Lagrange's Theorem in G to prove **Fermat's Little Theorem** in the form $[a]^p = [a]$ for any $a \in \mathbf{Z}$. (Compare with Exercise 51 in Section 2.5.)
- Sec. 3.1, #27 ➤ 21. Prove that any group with prime order is cyclic.
- Sec. 8.3, #11 < 22. Let G be a group of order pq , where p and q are primes. Prove that any nontrivial subgroup of G is cyclic.
23. Let G be a group of order pq , where p and q are distinct prime integers. If G has only one subgroup of order p and only one subgroup of order q , prove that G is cyclic.
- Sec. 3.2, #19 ➤ 24. Let G be an abelian group of order $2n$, where n is odd. Use Lagrange's Theorem to prove that G contains exactly one element of order 2.
- Sec. 4.1, #23 ➤ 25. A subgroup H of the group S_n is called **transitive** on $B = \{1, 2, \dots, n\}$ if for each pair i, j of elements of B there exists an element $h \in H$ such that $h(i) = j$. Suppose G is a group that is transitive on $\{1, 2, \dots, n\}$, and let H_i be the subgroup of G that leaves i fixed:

$$H_i = \{g \in G | g(i) = i\}$$

for $i = 1, 2, \dots, n$. Prove that $o(G) = n \cdot o(H_i)$.

26. (See Exercise 25.) Suppose G is a group that is transitive on $\{1, 2, \dots, n\}$, and let K_i be the subgroup that leaves each of the elements $1, 2, \dots, i$ fixed:

$$K_i = \{g \in G \mid g(k) = k \text{ for } k = 1, 2, \dots, i\}$$

for $i = 1, 2, \dots, n$. Prove that $G = S_n$ if and only if $H_i \neq H_j$ for all pairs i, j such that $i \neq j$ and $i < n - 1$.

4.5 Normal Subgroups

Among the subgroups of a group are those known as the *normal subgroups*. The significance of the normal subgroups is revealed in the next section.

Definition 4.15 ■ Normal Subgroup

Let H be a subgroup of G . Then H is a **normal** (or **invariant**) **subgroup** of G if $xH = Hx$ for all $x \in G$.

Note that the condition $xH = Hx$ is an equality of sets, and it does not require that $xh = hx$ for all $h \in H$.

Example 1 Let

$$H = A_3 = \{(1), (1, 2, 3), (1, 3, 2)\} = \langle(1, 2, 3)\rangle$$

and

$$G = S_3 = \{(1), (1, 2, 3), (1, 3, 2), (1, 2), (1, 3), (2, 3)\}.$$

For $x = (1, 2)$ we have

$$\begin{aligned} xH &= \{(1, 2)(1), (1, 2)(1, 2, 3), (1, 2)(1, 3, 2)\} \\ &= \{(1, 2), (2, 3), (1, 3)\} \end{aligned}$$

and

$$\begin{aligned} Hx &= \{(1)(1, 2), (1, 2, 3)(1, 2), (1, 3, 2)(1, 2)\} \\ &= \{(1, 2), (1, 3), (2, 3)\}. \end{aligned}$$

We have $xH = Hx$, but $xh \neq hx$ when $h = (1, 2, 3) \in H$. Similar computations show that

$$\begin{aligned} (1)H &= (1, 2, 3)H = (1, 3, 2)H = \{(1), (1, 2, 3), (1, 3, 2)\} = H \\ H(1) &= H(1, 2, 3) = H(1, 3, 2) = \{(1), (1, 2, 3), (1, 3, 2)\} = H \\ (1, 2)H &= (1, 3)H = (2, 3)H = \{(1, 2), (1, 3), (2, 3)\} \\ H(1, 2) &= H(1, 3) = H(2, 3) = \{(1, 2), (1, 3), (2, 3)\}. \end{aligned}$$

Thus H is a normal subgroup of G . Additionally we note that G can be expressed as

$$G = H \cup (1, 2)H.$$



In Example 1, we have $hH = H = Hh$ for all $h \in H$. These equalities hold for all subgroups, as stated in the following theorem.

Theorem 4.16 ■ A Special Coset H

If H is any subgroup of a group G , then $hH = H = Hh$ for all $h \in H$.

$p \Rightarrow q$ **Proof** Let h be an arbitrary element in the subgroup H of the group G .

If $x \in hH$, then $x = hy$ for some $y \in H$. But $h \in H$ and $y \in H$ imply $hy = x$ is in H , since H is closed. Thus $hH \subseteq H$.

For any $x \in H$, the element $h^{-1}x$ is in H since H contains the inverse of h and H is closed. But

$$h^{-1}x \in H \Rightarrow h(h^{-1}x) = x \in hH,$$

and this proves that $H \subseteq hH$. It follows that $hH = H$.

The proof of the equality $Hh = H$ is similar.

The proof of the following corollary is left as an exercise.

Corollary 4.17 ■ The Square of a Subgroup

For any subgroup H of a group G , $H^2 = H$, where H^2 denotes the product HH as defined in Definition 4.8.

Example 2 As an example of a subgroup that is *not* normal, let $K = \{(1), (1, 2)\}$ in S_3 . With $x = (1, 2, 3)$, we have

$$\begin{aligned} xK &= \{(1, 2, 3), (1, 2, 3)(1, 2)\} \\ &= \{(1, 2, 3), (1, 3)\} \\ Kx &= \{(1, 2, 3), (1, 2)(1, 2, 3)\} \\ &= \{(1, 2, 3), (2, 3)\}. \end{aligned}$$

Thus $xK \neq Kx$, and K is not a normal subgroup of S_3 . ■

The definition of a normal subgroup can be formulated in several different ways. For instance, we can write

$$\begin{aligned} xH = Hx \quad \text{for all } x \in G &\Leftrightarrow xHx^{-1} = H \quad \text{for all } x \in G \\ &\Leftrightarrow x^{-1}Hx = H \quad \text{for all } x \in G. \end{aligned}$$

Other formulations can be made. One that is frequently taken as the definition is given in Theorem 4.18.

Theorem 4.18 ■ Normal Subgroups and Conjugates

Let H be a subgroup of G . Then H is a normal subgroup of G if and only if $xhx^{-1} \in H$ for every $h \in H$ and every $x \in G$.

$p \Rightarrow q$ **Proof** If H is a normal subgroup of G , then the condition follows easily, since H normal requires

$$\begin{aligned} xHx^{-1} = H \text{ for all } x \in G &\Rightarrow xHx^{-1} \subseteq H \text{ for all } x \in G \\ &\Rightarrow xhx^{-1} \in H \text{ for all } h \in H \text{ and all } x \in G. \end{aligned}$$

$p \Leftarrow q$ Suppose now that the condition holds. For any $x \in G$, $xHx^{-1} \subseteq H$ follows immediately, and we need only show that $H \subseteq xHx^{-1}$. Let h be arbitrary in H , and let $x \in G$. Now x^{-1} is an element in G , and the condition implies that

$$(x^{-1})(h)(x^{-1})^{-1} = x^{-1}hx$$

is in H ; that is,

$$\begin{aligned} x^{-1}hx = h_1 \text{ for some } h_1 \in H &\Rightarrow h = xh_1x^{-1} \text{ for some } h_1 \in H \\ &\Rightarrow h \in xHx^{-1}. \end{aligned}$$

Thus $H \subseteq xHx^{-1}$, and we have $xHx^{-1} = H$ for all $x \in G$. It follows that H is a normal subgroup of G .

The concept of generators can be extended from cyclic subgroups $\langle a \rangle$ to more complicated situations where a subgroup is generated by more than one element. We only touch on this topic here, but it is a fundamental idea in more advanced study of groups.

Definition 4.19 ■ Set Generated by A

If A is a nonempty subset of the group G , then the **set generated by A** , denoted by $\langle A \rangle$, is the set defined by

$$\langle A \rangle = \{x \in G \mid x = a_1a_2 \cdots a_n \text{ with either } a_i \in A \text{ or } a_i^{-1} \in A\}.$$

In other words, $\langle A \rangle$ is the set of all products that can be formed with a finite number of factors, each of which either is an element of A or has an inverse that is an element of A .

Theorem 4.20 ■ Subgroup Generated by A

For any nonempty subset A of a group G , the set $\langle A \rangle$ is a subgroup of G called the **subgroup of G generated by A** .

$p \Rightarrow q$ **Proof** There exists at least one $a \in A$, since $A \neq \emptyset$. Then $e = aa^{-1} \in \langle A \rangle$, so $\langle A \rangle$ is nonempty.

If $x \in \langle A \rangle$ and $y \in \langle A \rangle$, then

$$x = x_1x_2 \cdots x_n \text{ with either } x_i \in A \text{ or } x_i^{-1} \in A$$

and

$$y = y_1 y_2 \cdots y_k \text{ with either } y_j \in A \text{ or } y_j^{-1} \in A.$$

Thus

$$xy = x_1 x_2 \cdots x_n y_1 y_2 \cdots y_k,$$

where each factor on the right either is in A or has an inverse that is an element of A . Also,

$$x^{-1} = x_n^{-1} \cdots x_2^{-1} x_1^{-1} \text{ with either } x_i^{-1} \in A \text{ or } x_i \in A.$$

The nonempty set $\langle A \rangle$ is closed and contains inverses, and therefore it is a subgroup of G .

In work with *finite groups*, the result in Exercise 41 of Section 3.3 is extremely helpful in finding $\langle A \rangle$, since it implies that $\langle A \rangle$ is the smallest subset of G that contains A and is closed under the operation. (This is true *only for finite groups*.) The subgroup $\langle A \rangle$ can be constructed systematically by starting a multiplication table using the elements of A and enlarging the table by adjoining additional elements until closure is obtained. A practical first step in this direction is to begin the table using all the elements of A and all their distinct powers. This is illustrated in the next example.

Example 3 Let $A = \{(1, 2, 3, 4), (1, 4)(2, 3)\}$, and consider the problem of finding $\langle A \rangle$ in S_4 . We begin by computing the distinct powers of the elements of A :

$$\begin{array}{ll} \alpha = (1, 2, 3, 4) & \alpha^2 = (1, 3)(2, 4) \\ \alpha^3 = \alpha^{-1} = (1, 4, 3, 2) & \alpha^4 = e = (1) \\ \beta = (1, 4)(2, 3) & \beta^2 = e. \end{array}$$

Starting a multiplication table using $e, \alpha, \alpha^2, \alpha^3, \beta$, we find the following new elements of $\langle A \rangle$:

$$\begin{aligned} \alpha\beta &= (1, 2, 3, 4)(1, 4)(2, 3) = (2, 4) = \gamma \\ \alpha^2\beta &= (1, 3)(2, 4)(1, 4)(2, 3) = (1, 2)(3, 4) = \Delta \\ \alpha^3\beta &= (1, 4, 3, 2)(1, 4)(2, 3) = (1, 3) = \theta. \end{aligned}$$

We then enlarge the table so as to use all eight elements

$$e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta.$$

Proceeding to fill out the enlarged table, we obtain the table in Figure 4.9, which shows that the set

$$G = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$$

is the subgroup of S_4 generated by $A = \{\alpha, \beta\}$. This group G is the **octic group** that was presented in Example 11 of Section 4.1.

\circ	e	α	α^2	α^3	β	γ	Δ	θ
e	e	α	α^2	α^3	β	γ	Δ	θ
α	α	α^2	α^3	e	γ	Δ	θ	β
α^2	α^2	α^3	e	α	Δ	θ	β	γ
α^3	α^3	e	α	α^2	θ	β	γ	Δ
β	β	θ	Δ	γ	e	α^3	α^2	α
γ	γ	β	θ	Δ	α	e	α^3	α^2
Δ	Δ	γ	β	θ	α^2	α	e	α^3
θ	θ	Δ	γ	β	α^3	α^2	α	e

Figure 4.9

Exercises 4.5

True or False

Label each of the following statements as either true or false.

- Let H be any subgroup of a group G and $a \in G$. Then $aH = Ha$ implies $ah = ha$ for all h in H .
- The trivial subgroups $\{e\}$ and G are both normal subgroups of the group G .
- The trivial subgroups $\{e\}$ and G are the only normal subgroups of a nonabelian group G .
- Let H be a subgroup of a group G . If $hH = H = Hh$ for all $h \in H$, then H is normal in G .
- If a group G contains a normal subgroup, then every subgroup of G must be normal.
- Let A be a nonempty subset of a group G . Then $A \in \langle A \rangle$.
- Let A be a nonempty subset of a group G . Then $\langle A \rangle$ is closed under the group operation if and only if A is closed under the same operation.

Exercises

Sec. 4.4, #1–6 ➤

- Let G be the group and H the subgroup given in each of the following exercises of Section 4.4. In each case, is H normal in G ?
 - Exercise 1
 - Exercise 2
 - Exercise 3
 - Exercise 4
 - Exercise 5
 - Exercise 6
- Show that

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

is a normal subgroup of the multiplicative group G of invertible matrices in $M_2(\mathbf{R})$.

3. For any subgroup H of the group G , let H^2 denote the product $H^2 = HH$ as defined in Definition 4.8. Prove Corollary 4.17: $H^2 = H$.

4. Let H be a normal cyclic subgroup of a finite group G . Prove that every subgroup K of H is normal in G .

Sec. 3.4, #41 ➤ 5. Let H be a torsion subgroup of an abelian group G . That is, H is the set of all elements of finite order in G . Prove that H is normal in G .

6. Show that every subgroup of an abelian group is normal.

7. Consider the octic group G of Example 3.

a. Find a subgroup of G that has order 2 and is a normal subgroup of G .

b. Find a subgroup of G that has order 2 and is *not* a normal subgroup of G .

Sec. 4.6, #9 < 8. Find all normal subgroups of the octic group.

Sec. 4.6, #10 < 9. Find all normal subgroups of the alternating group A_4 .

Sec. 3.1, #28 ➤ 10. Find all normal subgroups of the quaternion group.

Sec. 4.6, #11 <

11. Exercise 6 states that every subgroup of an abelian group is normal. Give an example of a nonabelian group for which every subgroup is normal.

12. Find groups H and G such that $H \subseteq G \subseteq A_4$ and the following conditions are satisfied:

a. H is a normal subgroup of G .

b. G is a normal subgroup of A_4 .

c. H is not a normal subgroup of A_4 .

(Thus the statement “A normal subgroup of a normal subgroup is a normal subgroup” is false.)

13. Find groups H and K such that the following conditions are satisfied:

a. H is a normal subgroup of K .

b. K is a normal subgroup of the octic group.

c. H is not a normal subgroup of the octic group.

14. Let H be a subgroup of G and assume that every left coset aH of H in G is equal to a right coset Hb of H in G . Prove that H is a normal subgroup of G .

15. If $\{H_\lambda\}$, $\lambda \in \mathcal{L}$, is a collection of normal subgroups H_λ of G , prove that $\bigcap_{\lambda \in \mathcal{L}} H_\lambda$ is a normal subgroup of G .

Sec. 4.6, #36 < 16. If H is a subgroup of G , and K is a normal subgroup of G , prove that $HK = KH$.

Sec. 4.6, #36 < 17. With H and K as in Exercise 16, prove that HK is a subgroup of G .

Sec. 4.6, #36 < 18. With H and K as in Exercise 16, prove that $H \cap K$ is a normal subgroup of H .

Sec. 4.6, #36 < 19. With H and K as in Exercise 16, prove that K is a normal subgroup of HK .

20. If H and K are both normal subgroups of G , prove that HK is a normal subgroup of G .

- 21.** Prove that if H and K are normal subgroups of G such that $H \cap K = \{e\}$, then $hk = kh$ for all $h \in H$, $k \in K$.

Sec. 4.6, #34 ≪ **22.** The **center** $Z(G)$ of a group G is defined as

$$Z(G) = \{a \in G \mid ax = xa \text{ for all } x \in G\}.$$

Sec. 4.6, #30, 33 ≪ Prove that $Z(G)$ is a normal subgroup of G .

- 23.** (See Exercise 22.) Find the center of the octic group.
- 24.** (See Exercise 22.) Find the center of A_4 .
- 25.** Suppose H is a normal subgroup of order 2 of a group G . Prove that H is contained in $Z(G)$, the center of G .
- 26.** For an arbitrary subgroup H of the group G , the **normalizer** of H in G is the set $\mathcal{N}(H) = \{x \in G \mid xHx^{-1} = H\}$.
- a.** Prove that $\mathcal{N}(H)$ is a subgroup of G .
 - b.** Prove that H is a normal subgroup of $\mathcal{N}(H)$.
 - c.** Prove that if K is a subgroup of G that contains H as a normal subgroup, then $K \subseteq \mathcal{N}(H)$.
- 27.** Find the normalizer of the subgroup $\{(1), (1, 3)(2, 4)\}$ of the octic group.
- 28.** Find the normalizer of the subgroup $\{(1), (1, 4)(2, 3)\}$ of the octic group.
- 29.** Let H be a subgroup of G . Define the relation “congruence modulo H ” on G by

$$a \equiv b \pmod{H} \quad \text{if and only if} \quad a^{-1}b \in H.$$

Prove that congruence modulo H is an equivalence relation on G .

- 30.** Describe the equivalence classes in Exercise 29.
- 31.** Let $n > 1$ in the group of integers under addition, and let $H = \langle n \rangle$. Prove that
- $$a \equiv b \pmod{H} \quad \text{if and only if} \quad a \equiv b \pmod{n}.$$
- 32.** Let H be a subgroup of G with index 2.
- a.** Prove that H is a normal subgroup of G .
 - b.** Prove that $g^2 \in H$ for all $g \in G$.
- 33.** Show that A_n has index 2 in S_n , and thereby conclude that A_n is always a normal subgroup of S_n .
- 34.** Let A be a nonempty subset of a group G . Prove that $A \subseteq \langle A \rangle$.
- 35.** Find the subgroup of S_n that is generated by the given set.
- | | |
|--------------------------------------|--|
| a. $\{(1, 2), (1, 3)\}$ | b. $\{(1, 3), (1, 2, 3, 4)\}$ |
| c. $\{(1, 2, 4), (2, 3, 4)\}$ | d. $\{(1, 2), (1, 3), (1, 4)\}$ |
- 36.** Let n be a positive integer, $n > 1$. Prove by induction that the set of transpositions $\{(1, 2), (1, 3), \dots, (1, n)\}$ generates the entire group S_n .

4.6 Quotient Groups

If H is a normal subgroup of G , then $xH = Hx$ for all x in G , so there is no distinction between left and right cosets of H in G . In this case, we refer simply to the cosets of H in G .

If H is any subgroup of G , then $hH = H = Hh$ for all h in H , according to Theorem 4.16. Corollary 4.17 states that $H^2 = H \cdot H = H$ for all subgroups H . We use this fact in proving the next theorem.

Theorem 4.21 ■ Group of Cosets

Let H be a normal subgroup of G . Then the cosets of H in G form a group with respect to the product of subsets as given in Definition 4.8.

$p \Rightarrow q$ **Proof** Let H be a normal subgroup of G . We shall denote the set of all distinct cosets of H in G by G/H . Multiplication in G/H is associative, by part **a** of Theorem 4.9.

We need to show that the cosets of H in G are closed under the given product. Let aH and bH be arbitrary cosets of H in G . Using the associative property freely, we have

$$\begin{aligned}(aH)(bH) &= a(Hb)H \\ &= a(bH)H \quad \text{since } H \text{ is normal} \\ &= (ab)H \cdot H \\ &= abH \quad \text{since } H^2 = H.\end{aligned}$$

Thus G/H is closed and $(aH)(bH) = abH$.

The coset $H = eH$ is an identity element, since $(aH)(eH) = aeH = aH$ and $(eH)(aH) = eaH = aH$ for all aH in G/H .

The inverse of aH is $a^{-1}H$, since

$$(aH)(a^{-1}H) = aa^{-1}H = eH = H$$

and

$$(a^{-1}H)(aH) = a^{-1}aH = eH = H.$$

This completes the proof.

Definition 4.22 ■ Quotient Group

If H is a normal subgroup of G , the group G/H that consists of the cosets of H in G is called the **quotient group** or **factor group** of G by H .

If the group G is abelian, then so is the quotient group G/H . Let a and b be elements of G , then

$$\begin{aligned}aHbH &= abH \quad \text{since } H \text{ is normal} \\ &= baH \quad \text{since } G \text{ is abelian} \\ &= bHaH \quad \text{since } H \text{ is normal}\end{aligned}$$

and G/H is abelian.

Suppose the group G has finite order n and the normal subgroup H has order m . Then by Lagrange's Theorem, we have

$$o(G) = o(H) \cdot o(G/H)$$

or

$$n = m \cdot o(G/H),$$

and the order of the quotient group is $o(G/H) = n/m$.

Example 1 Let G be the octic group as given in Example 3 of Section 4.5:

$$G = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}.$$

It can be readily verified that $H = \{e, \gamma, \theta, \alpha^2\}$ is a normal subgroup of G . The distinct cosets of H in G are

$$H = eH = \gamma H = \theta H = \alpha^2 H = \{e, \gamma, \theta, \alpha^2\}$$

and

$$\alpha H = \alpha^3 H = \beta H = \Delta H = \{\alpha, \alpha^3, \beta, \Delta\}.$$

Thus $G/H = \{H, \alpha H\}$, and a multiplication table for G/H is as follows.

.	H	αH
H	H	αH
αH	αH	H

■

There is a very important and natural relation between the quotient groups of a group G and the epimorphisms from G to another group G' . Our next theorem shows that every quotient group G/H is a homomorphic image of G .

Theorem 4.23 ■ Quotient Group \Rightarrow Homomorphic Image

Let G be a group, and let H be a normal subgroup of G . The mapping $\phi: G \rightarrow G/H$ defined by

$$\phi(a) = aH$$

is an epimorphism from G to G/H .

$p \Rightarrow q$ **Proof** The rule $\phi(a) = aH$ clearly defines a mapping from G to G/H . For any a and b in G ,

$$\begin{aligned} \phi(a) \cdot \phi(b) &= (aH)(bH) \\ &= abH \quad \text{since } H \text{ is normal in } G \\ &= \phi(ab). \end{aligned}$$

Thus ϕ is a homomorphism. Every element of G/H is a coset of H in G that has the form aH for some a in G . For any such a , we have $\phi(a) = aH$. Therefore, ϕ is an epimorphism.

Example 2 Consider the octic group

$$G = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$$

and its normal subgroup

$$H = \{e, \gamma, \theta, \alpha^2\}.$$

We saw in Example 1 that $G/H = \{H, \alpha H\}$. Theorem 4.23 assures us that the mapping $\phi: G \rightarrow G/H$ defined by

$$\phi(a) = aH$$

is an epimorphism. The values of ϕ are given in this case by

$$\begin{aligned}\phi(e) &= \phi(\gamma) = \phi(\theta) = \phi(\alpha^2) = H \\ \phi(\alpha) &= \phi(\alpha^3) = \phi(\beta) = \phi(\Delta) = \alpha H.\end{aligned}$$
■

Theorem 4.23 says that every quotient G/H is a homomorphic image of G . We shall see that, up to an isomorphism, these quotient groups give all of the homomorphic images of G . In order to prove this, we need the following result about the kernel of a homomorphism.

Theorem 4.24 ■ Kernel of a Homomorphism

For any homomorphism ϕ from the group G to the group G' , $\ker \phi$ is a normal subgroup of G .

p \Rightarrow q **Proof** The identity e is in $\ker \phi$ since $\phi(e) = e'$, so $\ker \phi$ is always nonempty. If $a \in \ker \phi$ and $b \in \ker \phi$, then $\phi(a) = e'$ and $\phi(b) = e'$. Also, by Theorem 3.28, $\phi(b^{-1}) = [\phi(b)]^{-1}$, so

$$\begin{aligned}\phi(ab^{-1}) &= \phi(a)\phi(b^{-1}) \\ &= \phi(a)[\phi(b)]^{-1} \\ &= e' \cdot (e')^{-1} \\ &= e',\end{aligned}$$

and therefore $ab^{-1} \in \ker \phi$. Thus, by Theorem 3.10, $\ker \phi$ is a subgroup of G .

To show that $\ker \phi$ is normal, let $x \in G$ and $a \in \ker \phi$. Then

$$\begin{aligned}\phi(xax^{-1}) &= \phi(x)\phi(a)\phi(x^{-1}) && \text{since } \phi \text{ is a homomorphism} \\ &= \phi(x) \cdot e' \cdot \phi(x^{-1}) && \text{since } a \in \ker \phi \\ &= \phi(x) \cdot \phi(x^{-1}) \\ &= e' && \text{by part b of Theorem 3.28.}\end{aligned}$$

Thus xax^{-1} is in $\ker \phi$, and $\ker \phi$ is a normal subgroup by Theorem 4.18.

The mapping ϕ in Theorem 4.23 has H as its kernel, and this shows that every normal subgroup of G is the kernel of a homomorphism. Combining this fact with Theorem 4.24, we see that the normal subgroups of G and the kernels of the homomorphisms from G to another group are the same subgroups of G .

We can now prove that every homomorphic image of G is isomorphic to a quotient group of G .

Theorem 4.25 ■ Homomorphic Image \Rightarrow Quotient Group

Let G and G' be groups with G' a homomorphic image of G . Then G' is isomorphic to a quotient group of G .

$p \Rightarrow q$ **Proof** Let ϕ be an epimorphism from G to G' , and let $K = \ker \phi$. For each aK in G/K , define $\theta(aK)$ by

$$\theta(aK) = \phi(a).$$

First we need to prove that this rule defines a mapping. For any aK and bK in G/K ,

$$\begin{aligned} aK = bK &\Leftrightarrow b^{-1}aK = K \\ &\Leftrightarrow b^{-1}a \in K \\ &\Leftrightarrow \phi(b^{-1}a) = e' \\ &\Leftrightarrow \phi(b^{-1})\phi(a) = e' \\ &\Leftrightarrow [\phi(b)]^{-1}\phi(a) = e' \\ &\Leftrightarrow \phi(a) = \phi(b) \\ &\Leftrightarrow \theta(aK) = \theta(bK). \end{aligned}$$

Thus θ is a well-defined mapping from G/K to G' , and the \Leftarrow parts of the \Leftrightarrow statements show that θ is one-to-one as well.

We shall show that θ is an isomorphism from G/K to G' . Since

$$\begin{aligned} \theta(aK \cdot bK) &= \theta(abK) \\ &= \phi(ab) \\ &= \phi(a) \cdot \phi(b) \\ &= \theta(aK) \cdot \theta(bK), \end{aligned}$$

θ is a homomorphism. To show that θ is onto, let a' be arbitrary in G' . Since ϕ is an epimorphism, there exists an element a in G such that $\phi(a) = a'$. Then aK is in G/K , and

$$\theta(aK) = \phi(a) = a'.$$

Thus every element in G' is an image under θ , and this proves that θ is an isomorphism.

Theorem 4.26 ■ Fundamental Theorem of Homomorphisms

If ϕ is an epimorphism from the group G to the group G' , then G' is isomorphic to $G/\ker \phi$.

The Fundamental Theorem follows at once from the proof of Theorem 4.25.

In order to give nontrivial illustrations of Theorem 4.24 and 4.25, we need an example of a homomorphism that is somewhat involved. This homomorphism is presented in the next example.

Example 3 Consider the permutation group

$$G = S_3 = \{(1), (1, 2, 3), (1, 3, 2), (1, 2), (1, 3), (2, 3)\}$$

and the multiplicative group

$$G' = \{[1], [2]\} \subseteq \mathbf{Z}_3.$$

The mapping $\phi: G \rightarrow G'$ defined by

$$\begin{aligned}\phi(1) &= \phi(1, 2, 3) = \phi(1, 3, 2) = [1] \\ \phi(1, 2) &= \phi(1, 3) = \phi(2, 3) = [2]\end{aligned}$$

can be shown by direct computation to be an epimorphism from G to G' , but it is tedious to verify $\phi(xy) = \phi(x)\phi(y)$ for all 36 choices of the pair of factors x, y in S_3 . As an alternative to this chore, we shall obtain another description of ϕ . We first note that if $\alpha = (1, 2, 3)$ and $\beta = (1, 2)$, the elements of S_3 can be written as

$$\begin{aligned}(1) &= \alpha^0\beta^0 & (1, 2, 3) &= \alpha\beta^0 & (1, 3, 2) &= \alpha^2\beta^0 \\ (1, 2) &= \alpha^0\beta & (1, 3) &= \alpha\beta & (2, 3) &= \alpha^2\beta.\end{aligned}$$

We then make the following observations concerning S_3 :

1. Any element of S_3 can be written in the form $\alpha^i\beta^k$, with $i \in \{0, 1, 2\}$ and $k \in \{0, 1\}$.
2. $\beta\alpha^i = \alpha^{-i}\beta$.
3. Any $x \in S_3$ is either of the form $x = \alpha^i$ or of the form $x = \alpha^i\beta$.

Routine calculations will confirm that our mapping ϕ can be described by the rule

$$\phi(\alpha^r\beta^k) = [2]^k \text{ for any integer } r.$$

Having made these observations, we can now verify the equation $\phi(x)\phi(y) = \phi(xy)$ with a reasonable amount of work. For arbitrary x and y in S_3 , we write either $x = \alpha^i$ or $x = \alpha^i\beta$, and $y = \alpha^m\beta^n$ where $m \in \{0, 1, 2\}$ and $n \in \{0, 1\}$.

If $x = \alpha^i$, we have

$$\phi(xy) = \phi(\alpha^i\alpha^m\beta^n) = \phi(\alpha^{i+m}\beta^n) = [2]^n$$

and

$$\phi(x)\phi(y) = \phi(\alpha^i)\phi(\alpha^m\beta^n) = [2]^0[2]^n = [2]^n.$$

If $x = \alpha^i\beta$, we have

$$\begin{aligned}\phi(xy) &= \phi(\alpha^i\beta\alpha^m\beta^n) \\ &= \phi(\alpha^i\alpha^{-m}\beta\beta^n) \\ &= \phi(\alpha^{i-m}\beta^{n+1}) \\ &= [2]^{n+1}\end{aligned}$$

and

$$\begin{aligned}\phi(x)\phi(y) &= \phi(\alpha'\beta)\phi(\alpha''\beta'') \\ &= [2][2]^n \\ &= [2]^{n+1}.\end{aligned}$$

Thus $\phi(xy) = \phi(x)\phi(y)$ in all cases, and ϕ is a homomorphism (an epimorphism, actually) from G to G' . ■

Example 4 To illustrate Theorems 4.24 and 4.25, consider the groups $G = S_3$ and $G' = \{[1], [2]\}$ in the previous example. We see that the kernel of the epimorphism $\phi: G \rightarrow G'$ is the normal subgroup

$$\begin{aligned}K &= \ker \phi \\ &= \{(1), (1, 2, 3), (1, 3, 2)\}\end{aligned}$$

of G . The quotient group G/K is given by

$$G/K = \{K, (1, 2)K\}$$

where

$$(1, 2)K = \{(1, 2), (2, 3), (1, 3)\}.$$

The isomorphism $\theta: G/K \rightarrow G'$ has values

$$\begin{aligned}\theta(K) &= \phi(1) = [1] \\ \theta((1, 2)K) &= \phi(1, 2) = [2].\end{aligned}$$

Using the results of this section, we can systematically find all of the homomorphic images of a group G . We now know that the homomorphic images of G are the same (in the sense of isomorphism) as the quotient groups of G .

Example 5 Let $G = S_3$, the symmetric group on three elements. In order to find all the homomorphic images of G , we need only find all of the normal subgroups H of G and form all possible quotient groups G/H . As we saw in Section 4.4, a complete list of the subgroups of G is

$$\begin{array}{ll}H_1 = \{(1)\} & H_4 = \{(1), (2, 3)\} \\ H_2 = \{(1), (1, 2)\} & H_5 = \{(1), (1, 2, 3), (1, 3, 2)\} \\ H_3 = \{(1), (1, 3)\} & H_6 = S_3.\end{array}$$

Of these, H_1 , H_5 , and H_6 are the only normal subgroups. The possible homomorphic images of G , then, are

$$\begin{aligned}G/H_1 &= \{H_1, (1, 2)H_1, (1, 3)H_1, (2, 3)H_1, (1, 2, 3)H_1, (1, 3, 2)H_1\} \\ G/H_5 &= \{H_5, (1, 2)H_5\} \\ G/G &= \{G\}.\end{aligned}$$

Thus any homomorphic image of S_3 is isomorphic to S_3 , to a cyclic group of order 2, or to a group with only the identity element. ■

Exercises 4.6

True or False

Label each of the following statements as either true or false.

1. Every normal subgroup of a group is the kernel of a homomorphism.
2. The kernel of any homomorphism from group G to group G' is a normal subgroup of G' .
3. $aHbH = abH$ for any subgroup H of a group G and for all a, b in G .
4. Every homomorphic image of a group G is isomorphic to a quotient group of G .
5. The homomorphic images of a group G are the same (up to an isomorphism) as the quotient groups of G .

Exercises

In Exercises 1–6, H is a normal subgroup of the group G . Find the order of the quotient group G/H . Write out the distinct elements of G/H and construct a multiplication table for G/H .

1. The octic group $G = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$; $H = \{e, \alpha^2\}$
2. The octic group $G = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$; $H = \{e, \beta, \Delta, \alpha^2\}$

Sec. 3.1, #28 ➤

3. The quaternion group $G = \{\pm 1, \pm i, \pm j, \pm k\}$; $H = \{\pm 1\}$

Sec. 4.3, #25 ➤

4. The group of rigid motions of a regular pentagon $G = \{e, \alpha, \alpha^2, \alpha^3, \alpha^4, \beta, \gamma, \Delta, \theta, \sigma\} = D_5$; $H = \{e, \alpha, \alpha^2, \alpha^3, \alpha^4\}$, where $\alpha = (1, 2, 3, 4, 5)$, $\beta = (2, 5)(3, 4)$, $\gamma = (1, 2)(3, 4)$, $\Delta = (1, 3)(4, 5)$, $\theta = (1, 4)(2, 3)$, and $\sigma = (1, 5)(2, 4)$.

5. The alternating group $G = A_4$; $H = \{(1), (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}$
6. The symmetric group $G = S_4$; $H = \{(1), (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}$

Sec. 3.4, #18 ➤

7. Let G be the multiplicative group of units \mathbf{U}_{20} consisting of all $[a]$ in \mathbf{Z}_{20} that have multiplicative inverses. Find a normal subgroup H of G that has order 2 and construct a multiplication table for G/H .

8. Suppose G_1 and G_2 are groups with normal subgroups H_1 and H_2 , respectively, and with G_1/H_1 isomorphic to G_2/H_2 . Determine the possible orders of H_1 and H_2 under the following conditions.

- a. $o(G_1) = 24$ and $o(G_2) = 18$
- b. $o(G_1) = 32$ and $o(G_2) = 40$

Sec. 4.5, #8 ➤

9. Find all homomorphic images of the octic group.

Sec. 4.5, #9 ➤

10. Find all homomorphic images of A_4 .

Sec. 3.1, #28 ➤

11. Find all homomorphic images of the quaternion group.

Sec. 4.5, #10 ➤

Sec. 3.4, #18 ➤

12. Find all homomorphic images of each group G in Exercise 18 of Section 3.4.

- 13.** Let $G = S_3$. For each H that follows, show that the set of all left cosets of H in G does *not* form a group with respect to a product defined by $(aH)(bH) = abH$.

- a. $H = \{(1), (1, 2)\}$
- b. $H = \{(1), (1, 3)\}$
- c. $H = \{(1), (2, 3)\}$

- Sec. 3.1, #30 ➤ **14.** Let $G = \{I_2, R, R^2, R^3, H, D, V, T\}$ be the multiplicative group of matrices in Exercise 30 of Section 3.1, let $G' = \{1, -1\}$ under multiplication, and define $\phi: G \rightarrow G'$ by
Sec. 3.6, #9 ➤

$$\phi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc.$$

- a. Assume that ϕ is an epimorphism, and find the elements of $K = \ker \phi$.
- b. Write out the distinct elements of G/K .
- c. Let $\theta: G/K \rightarrow G'$ be the isomorphism described in the proof of Theorem 4.25, and write out the values of θ .

- Sec. 3.3, #16 ➤ **15.** Repeat Exercise 14 with $G = \{I_2, M_1, M_2, M_3, M_4, M_5\}$, the multiplicative group of matrices in Exercise 16 of Section 3.3.

- Sec. 3.1, #28 ➤ **16.** Repeat Exercise 14 with the quaternion group $G = \{1, i, j, k, -1, -i, -j, -k\}$, the Klein four group $G' = \{e, a, b, ab\}$, and $\phi: G \rightarrow G'$ defined by
Sec. 3.5, #10 ➤

$$\begin{aligned} \phi(1) &= \phi(-1) = e & \phi(i) &= \phi(-i) = a \\ \phi(j) &= \phi(-j) = b & \phi(k) &= \phi(-k) = ab. \end{aligned}$$

- 17.** Repeat Exercise 14 where G is the multiplicative group of units \mathbf{U}_{20} and G' is the cyclic group of order 4. That is,

$$\begin{aligned} G &= \{[1], [3], [7], [9], [11], [13], [17], [19]\}, \\ G' &= \langle a \rangle = \{e, a, a^2, a^3\}. \end{aligned}$$

Define $\phi: G \rightarrow G'$ by

$$\begin{aligned} \phi([1]) &= \phi([11]) = e & \phi([3]) &= \phi([13]) = a \\ \phi([9]) &= \phi([19]) = a^2 & \phi([7]) &= \phi([17]) = a^3. \end{aligned}$$

- 18.** If H is a subgroup of the group G such that $(aH)(bH) = abH$ for all left cosets aH and bH of H in G , prove that H is normal in G .

- 19.** Let H be a subgroup of the group G . Prove that H is normal in G if and only if $(Ha)(Hb) = Hab$ for all right cosets Ha and Hb of H in G .

- 20.** If H is a normal subgroup of the group G , prove that $(aH)^n = a^nH$ for every positive integer n .

- Sec. 4.4, #12 ➤ **21.** Let H be a normal subgroup of finite group G . If the order of the quotient group G/H is m , prove that g^m is in H for all g in G .

- 22.** Let H be a normal subgroup of the group G . Prove that G/H is abelian if and only if $a^{-1}b^{-1}ab \in H$ for all $a, b \in G$.

- Sec. 3.4, #32, 41 ➤
- 23. Let G be a torsion group, as defined in Exercise 41 of Section 3.4, and H a normal subgroup of G . Prove that the quotient group G/H is a torsion group.
 - 24. Let G be a cyclic group. Prove that for every normal subgroup H of G , G/H is a cyclic group.
 - 25. Prove or disprove that if a group G has a cyclic quotient group G/H , then G must be cyclic.
 - 26. Prove or disprove that if a group G has an abelian quotient group G/H , then G must be abelian.
 - 27.
 - a. Show that a cyclic group of order 8 has a cyclic group of order 4 as a homomorphic image.
 - b. Show that a cyclic group of order 6 has a cyclic group of order 2 as a homomorphic image.
- Sec. 3.6, #19 ➤
- 28. Assume that ϕ is an epimorphism from the group G to the group G' .
 - a. Prove that the mapping $H \rightarrow \phi(H)$ is a bijection from the set of all subgroups of G that contain $\ker \phi$ to the set of all subgroups of G' .
 - b. Prove that if K is a normal subgroup of G' , then $\phi^{-1}(K)$ is a normal subgroup of G .
 - 29. Suppose ϕ is an epimorphism from the group G to the group G' . Let H be a normal subgroup of G containing $\ker \phi$, and let $H' = \phi(H)$.
 - a. Prove that H' is a normal subgroup of G' .
 - b. Prove that G/H is isomorphic to G'/H' .
- Sec. 4.5, #22 ➤
- 30. Let G be a group with center $Z(G) = C$. Prove that if G/C is cyclic, then G is abelian.
 - 31. (See Exercise 30.) Prove that if p and q are primes and G is a nonabelian group of order pq , then the center of G is the trivial subgroup $\{e\}$.
- Sec. 3.5, #18, 19 ➤
- 32. Let a be a fixed element of the group G . According to Exercise 18 of Section 3.5, the mapping $t_a: G \rightarrow G$ defined by $t_a(x) = axa^{-1}$ is an automorphism of G . Each of these automorphisms t_a is called an **inner automorphism** of G . Prove that the set $\text{Inn}(G) = \{t_a | a \in G\}$ forms a normal subgroup of the group of all automorphisms of G .
- Sec. 4.5, #22 ➤
- 33. (See Exercise 32.) Let G be a group with center $Z(G) = C$. Prove that $\text{Inn}(G)$ is isomorphic to G/C .
- Sec. 4.5, #21 ➤
- 34. If H and K are normal subgroups of the group G such that $G = HK$ and $H \cap K = \{e\}$, then G is said to be the **internal direct product** of H and K , and we write $G = H \times K$ to denote this. If $G = H \times K$, prove that $\phi: H \rightarrow G/K$ defined by $\phi(h) = hK$ is an isomorphism from H to G/K .
 - 35. (See Exercise 34.) If $G = H \times K$, prove that each element $g \in G$ can be written uniquely as $g = hk$ with $h \in H$ and $k \in K$.
- Sec. 4.5, #16–19 ➤
- 36. Let H be a subgroup of G and let K be a normal subgroup of G .
 - a. Prove that the mapping $\phi: H \rightarrow HK/K$ defined by $\phi(h) = hK$ is an epimorphism from H to HK/K .
 - b. Prove that $\ker \phi = H \cap K$.
 - c. Prove that $H/H \cap K$ is isomorphic to HK/K .
- Sec. 6.2, #27 ≪

- 37.** Let H and K be arbitrary groups and let $H \otimes K$ denote the Cartesian product of H and K :

$$H \otimes K = \{(h, k) | h \in H \text{ and } k \in K\}.$$

Equality in $H \otimes K$ is defined by $(h, k) = (h', k')$ if and only if $h = h'$ and $k = k'$. Multiplication in $H \otimes K$ is defined by

$$(h_1, k_1)(h_2, k_2) = (h_1 h_2, k_1 k_2).$$

- a. Prove that $H \otimes K$ is a group. This group is called the **external direct product** of H and K .
- b. Suppose that e_1 and e_2 are the identity elements of H and K , respectively. Show that $H' = \{(h, e_2) | h \in H\}$ is a normal subgroup of $H \otimes K$ that is isomorphic to H and, similarly, that $K' = \{(e_1, k) | k \in K\}$ is a normal subgroup isomorphic to K .
- c. Prove that $H \otimes K / H'$ is isomorphic to K and that $H \otimes K / K'$ is isomorphic to H .

Sec. 4.7, #16 <

- 38.** (See Exercise 37.) Let a and b be fixed elements of a group G , and let $\mathbf{Z} \otimes \mathbf{Z}$ be the external direct product of the additive group \mathbf{Z} with itself. Prove that the mapping $\phi: \mathbf{Z} \otimes \mathbf{Z} \rightarrow G$ defined by $\phi(m, n) = a^m b^n$ is a homomorphism if and only if $ab = ba$ in G .

4.7

Direct Sums (Optional)

The overall objective of this and the next section is to present some of the basic material on abelian groups. A tremendous amount of work has been done on the subject. One of the concepts fundamental to abelian groups is a *direct sum*, to be defined in this section. Throughout this section we write all abelian groups in additive notation.

We begin by defining the sum of a finite number of subgroups in an abelian group and showing that this sum is a subgroup.

Definition 4.27 ■ Sum of Subgroups

Let H_1, H_2, \dots, H_n be subgroups of the abelian group G . The **sum** $H_1 + H_2 + \dots + H_n$ of these subgroups is defined by

$$H_1 + H_2 + \dots + H_n = \{x \in G | x = h_1 + h_2 + \dots + h_n \text{ with } h_i \in H_i\}.$$

Theorem 4.28 ■ Sum of Subgroups

If H_1, H_2, \dots, H_n are subgroups of the abelian group G , then $H_1 + H_2 + \dots + H_n$ is a subgroup of G .

$p \Rightarrow q$ **Proof** The sum $H_1 + H_2 + \dots + H_n$ is clearly nonempty. For arbitrary

$$x = h_1 + h_2 + \dots + h_n$$

with $h_i \in H_i$, the inverse

$$-x = (-h_1) + (-h_2) + \cdots + (-h_n)$$

is in the sum $H_1 + H_2 + \cdots + H_n$, since $-h_i \in H_i$ for each i . Also, if

$$y = h'_1 + h'_2 + \cdots + h'_n$$

with $h'_i \in H_i$, then

$$x + y = (h_1 + h'_1) + (h_2 + h'_2) + \cdots + (h_n + h'_n)$$

is in the sum of the H_i , since $h_i + h'_i \in H_i$ for each i . Thus $H_1 + H_2 + \cdots + H_n$ is a subgroup of G .

The contents of Definition 4.19 and Theorem 4.20 may be restated as follows, with addition as the binary operation:

If A is a nonempty subset of the group G , then the *subgroup of G generated by A* is the set

$$\langle A \rangle = \{x \in G \mid x = a_1 + a_2 + \cdots + a_n \text{ with } a_i \in A \text{ or } -a_i \in A\}.$$

It is left as an exercise to prove that if H_1, H_2, \dots, H_n are subgroups of an abelian group G , then $G = H_1 + H_2 + \cdots + H_n$ if and only if G is generated by $\bigcup_{i=1}^n H_i$.

Example 1 Let G be the group $G = \mathbf{Z}_{12}$ under addition, and consider the following sums of subgroups in G .

a. If

$$H_1 = \langle [3] \rangle = \{[3], [6], [9], [0]\}$$

and

$$H_2 = \langle [2] \rangle = \{[2], [4], [6], [8], [10], [0]\},$$

then

$$\begin{aligned} H_1 + H_2 &= \{r[3] + s[2] \mid r, s \in \mathbf{Z}\} \\ &= \{[3r + 2s] \mid r, s \in \mathbf{Z}\} \end{aligned}$$

is a subgroup. Since $[3(1) + 2(11)] = [25] = [1]$ in \mathbf{Z}_{12} and $[1]$ generates \mathbf{Z}_{12} under addition, we have

$$H_1 + H_2 = G.$$

b. Now let

$$\begin{aligned} K_1 &= H_1 = \langle [3] \rangle, \\ K_2 &= \langle [4] \rangle = \{[4], [8], [0]\}. \end{aligned}$$

The sum $K_1 + K_2$ is given by

$$\begin{aligned} K_1 + K_2 &= \{u[3] + v[4] \mid u, v \in \mathbf{Z}\} \\ &= \{[3u + 4v] \mid u, v \in \mathbf{Z}\}. \end{aligned}$$

Since $[3(-1) + 4(1)] = [1]$, $[1] \in K_1 + K_2$, and hence

$$K_1 + K_2 = G.$$

c. With the same notation as in parts **a** and **b**,

$$H_2 + K_2 = H_2,$$

since $K_2 \subseteq H_2$. ■

We now consider the definition of a direct sum.

Definition 4.29 ■ Direct Sum

If H_1, H_2, \dots, H_n are subgroups of the abelian group G , then $H_1 + H_2 + \dots + H_n$ is a **direct sum** if and only if the expression for each x in the sum as

$$x = h_1 + h_2 + \dots + h_n$$

with $h_i \in H_i$ is **unique**. We write

$$H_1 \oplus H_2 \oplus \dots \oplus H_n$$

to indicate a direct sum.

The next theorem gives a simple fact about direct sums that can be very useful when we work with finite groups.

Theorem 4.30 ■ Order of a Direct Sum

If H_1, H_2, \dots, H_n are finite subgroups of the abelian group G such that their sum is direct, then the order of $H_1 \oplus H_2 \oplus \dots \oplus H_n$ is the product of the orders of the subgroups H_i :

$$o(H_1 \oplus H_2 \oplus \dots \oplus H_n) = o(H_1)o(H_2) \cdots o(H_n).$$

$p \Rightarrow q$ **Proof** With $h_i \in H_i$ in the expression

$$x = h_1 + h_2 + \dots + h_n,$$

there are $o(H_i)$ choices for each h_i . Any change in one of the h_i produces a different element x , by the uniqueness property stated in Definition 4.29. Hence there are

$$o(H_1)o(H_2) \cdots o(H_n)$$

distinct elements x of the form $x = h_1 + h_2 + \dots + h_n$, and the theorem follows.

There are several equivalent ways to formulate the definition of direct sum. One of these is presented in the following theorem.

Theorem 4.31 ■ Equivalent Condition for a Direct Sum

If each H_i is a subgroup of the abelian group G , then the sum $H_1 + H_2 + \cdots + H_n$ is direct if and only if the following condition holds: Any equation of the form

$$h_1 + h_2 + \cdots + h_n = 0$$

with $h_i \in H_i$ implies that all $h_i = 0$.

$p \Leftarrow q$ **Proof** Assume first that the condition holds. If an element x in the sum of the H_i is written as

$$x = h_1 + h_2 + \cdots + h_n$$

and also as

$$x = h'_1 + h'_2 + \cdots + h'_n$$

with h_i and $h'_i \in H_i$ for each i , then

$$h_1 + h_2 + \cdots + h_n = h'_1 + h'_2 + \cdots + h'_n$$

and

$$(h_1 - h'_1) + (h_2 - h'_2) + \cdots + (h_n - h'_n) = 0.$$

The condition implies that $h_i - h'_i = 0$, and hence $h_i = h'_i$ for each i . Thus the sum $H_1 + H_2 + \cdots + H_n$ is direct.

$p \Rightarrow q$ Conversely, suppose the sum $H_1 + H_2 + \cdots + H_n$ is direct. Then the identity element 0 in the sum can be written *uniquely* as

$$0 = 0 + 0 + \cdots + 0$$

where the sum on the right indicates a choice of 0 as the term from each H_i . From the uniqueness property,

$$h_1 + h_2 + \cdots + h_n = 0$$

with $h_i \in H_i$ requires that all $h_i = 0$.

Some intuitive feeling for the concept of a direct sum is provided by considering the special case where the sum has only two terms.

Theorem 4.32 ■ Direct Sum of Two Subgroups

Let H_1 and H_2 be subgroups of the abelian group G . Then $G = H_1 \oplus H_2$ if and only if $G = H_1 + H_2$ and $H_1 \cap H_2 = \{0\}$.

$p \Rightarrow (q \wedge r)$ **Proof** Assume first that $G = H_1 \oplus H_2$, and let $x \in H_1 \cap H_2$. Then $x = h_1$ for some $h_1 \in H_1$. Also, $x \in H_2$, and therefore $-x \in H_2$. Let $h_2 = -x$. Then

$$\begin{aligned} h_1 + h_2 &= x + (-x) \\ &= 0 \end{aligned}$$

where $h_i \in H_i$. This implies that $x = h_1 = h_2 = 0$, by Theorem 4.31.

$p \Leftarrow (q \wedge r)$ Assume now that $G = H_1 + H_2$ and $H_1 \cap H_2 = \{0\}$. If

$$h_1 + h_2 = 0$$

with $h_i \in H_i$, then $h_1 = -h_2 \in H_1 \cap H_2$. Therefore, $h_1 = 0$ and $h_2 = 0$. By Theorem 4.31, $G = H_1 \oplus H_2$.

Example 2 In Example 1, we saw that the equations $H_1 + H_2 = G$ and $K_1 + K_2 = G$ were both valid. Since $H_1 \cap H_2 = \{[0], [6]\}$, the sum $H_1 + H_2$ is not direct. However, $K_1 \cap K_2 = \{[0]\}$, so $G = K_1 \oplus K_2$ in Example 1. ■

Theorem 4.32 can be generalized to the results stated in the next theorem. A proof is requested in the exercises.

Theorem 4.33 ■ Direct Sum of n Subgroups

Let H_1, H_2, \dots, H_n be subgroups of the abelian group G . The sum $H_1 + H_2 + \dots + H_n$ is direct if and only if the intersection of each H_j with the subgroup generated by $\bigcup_{i=1, i \neq j}^n H_i$ is the identity subgroup $\{0\}$.

Example 3 Consider the following subgroups of the abelian group \mathbf{Z}_{42} under addition:

$$H_1 = \{[0], [21]\} = \langle [21] \rangle$$

$$H_2 = \{[0], [14], [28]\} = \langle [14] \rangle$$

$$H_3 = \{[0], [6], [12], [18], [24], [30], [36]\} = \langle [6] \rangle.$$

Since each of the orders of H_1 , which is 2, of H_2 , which is 3, and of H_3 , which is 7, must divide the order of the group generated by $G = H_1 \cup H_2 \cup H_3$, then G must have order at least 42. The sum $G = H_1 + H_2 + H_3$ is direct since $\{[0]\} = H_1 \cap (H_2 \cup H_3) = H_2 \cap (H_1 \cup H_3) = H_3 \cap (H_1 \cup H_2)$. Since

$$1[21] + (-1)[14] + (-1)[6] = [1]$$

and $[1]$ generates \mathbf{Z}_{42} under addition, then

$$H_1 \oplus H_2 \oplus H_3 = \mathbf{Z}_{42}. \quad \blacksquare$$

As a final result for this section, we prove the following theorem.

Theorem 4.34 ■ Direct Sums and Isomorphisms

Let H_1 and H_2 be subgroups of the abelian group G such that $G = H_1 \oplus H_2$. Then G/H_2 is isomorphic to H_1 .

$p \Rightarrow q$ **Proof** The rule $\phi(h_1) = h_1 + H_2$ defines a mapping ϕ from H_1 to G/H_2 . This mapping is a homomorphism, since

$$\begin{aligned}\phi(h_1 + h'_1) &= (h_1 + h'_1)H_2 \\ &= (h_1 + H_2) + (h'_1 + H_2) \\ &= \phi(h_1) + \phi(h'_1).\end{aligned}$$

Now

$$\begin{aligned}h_1 \in \ker \phi &\Leftrightarrow \phi(h_1) = H_2 \\ &\Leftrightarrow h_1 + H_2 = H_2 \\ &\Leftrightarrow h_1 \in H_2 \\ &\Leftrightarrow h_1 = 0 \quad \text{since } H_1 \cap H_2 = \{0\}.\end{aligned}$$

Thus ϕ is one-to-one. Let $g + H_2$ be arbitrary in G/H_2 . Since $G = H_1 \oplus H_2$, g can be written as $g = h_1 + h_2$ with $h_i \in H_i$.

Then

$$\begin{aligned}g + H_2 &= (h_1 + h_2) + H_2 \\ &= h_1 + H_2 \quad \text{since } h_2 + H_2 = H_2 \\ &= \phi(h_1),\end{aligned}$$

and this shows that ϕ is onto. Thus ϕ is an isomorphism from H_1 to G/H_2 .

Exercises 4.7

True or False

Label each of the following statements as either true or false.

1. Let H_1, H_2 be finite groups of an abelian group G . Then $o(H_1 \oplus H_2) = o(H_1) + o(H_2)$.
 2. Let H_1, H_2 be finite groups of an abelian group G . If $G = H_1 + H_2$, then $G = \langle H_1 \cup H_2 \rangle$.
-

Exercises

1. Let $H_1 = \{[0], [6]\}$ and $H_2 = \{[0], [3], [6], [9]\}$ be subgroups of the abelian group \mathbf{Z}_{12} under addition. Find $H_1 + H_2$ and determine if the sum is direct.
2. Let $H_1 = \{[0], [6], [12]\}$ and $H_2 = \{[0], [3], [6], [9], [12], [15]\}$ be subgroups of the abelian group \mathbf{Z}_{18} under addition. Find $H_1 + H_2$ and determine if the sum is direct.
3. Let $H_1 = \{[0], [5]\}$ and $H_2 = \{[0], [2], [4], [6], [8]\}$ be subgroups of the abelian group \mathbf{Z}_{10} under addition. Show that $\mathbf{Z}_{10} = H_1 \oplus H_2$.
4. Let $H_1 = \{[0], [7], [14]\}$ and $H_2 = \{[0], [3], [6], [9], [12], [15], [18]\}$ be subgroups of the abelian group \mathbf{Z}_{21} under addition. Show that $\mathbf{Z}_{21} = H_1 \oplus H_2$.

5. Let $H_1 = \{[0], [15]\}$, $H_2 = \{[0], [10], [20]\}$ and $H_3 = \{[0], [6], [12], [18], [24]\}$ be subgroups of the abelian group \mathbf{Z}_{30} under addition. Show that $\mathbf{Z}_{30} = H_1 \oplus H_2 \oplus H_3$.
6. Let $H_1 = \{[0], [10], [20], [30], [40], [50], [60]\}$, $H_2 = \{[0], [14], [28], [42], [56]\}$ and $H_3 = \{[0], [35]\}$ be subgroups of the abelian group \mathbf{Z}_{70} under addition. Show that $\mathbf{Z}_{70} = H_1 \oplus H_2 \oplus H_3$.
7. Write \mathbf{Z}_{20} as the direct sum of two of its nontrivial subgroups.
8. Write \mathbf{Z}_{24} as the direct sum of two of its nontrivial subgroups.
9. Suppose that H_1 and H_2 are subgroups of the abelian group G such that $H_1 \subseteq H_2$. Prove that $H_1 + H_2 = H_2$.
10. Suppose that H_1 and H_2 are subgroups of the abelian group G such that $G = H_1 \oplus H_2$. If K is a subgroup of G such that $K \supseteq H_1$, prove that $K = H_1 \oplus (K \cap H_2)$.
11. Assume that H_1, H_2, \dots, H_n are subgroups of the abelian group G such that the sum $H_1 + H_2 + \dots + H_n$ is direct. If K_i is a subgroup of H_i for $i = 1, 2, \dots, n$, prove that $K_1 + K_2 + \dots + K_n$ is a direct sum.
12. Assume that H_1, H_2, \dots, H_n are subgroups of the abelian group G . Prove that $H_1 + H_2 + \dots + H_n$ is the smallest subgroup of G that contains all the subgroups H_i .
13. Assume that H_1, H_2, \dots, H_n are subgroups of the abelian group G . Prove that $G = H_1 + H_2 + \dots + H_n$ if and only if G is generated by $\bigcup_{i=1}^n H_i$.
14. Let G be an abelian group of order mn , where m and n are relatively prime. If $H_1 = \{x \in G \mid mx = 0\}$ and $H_2 = \{x \in G \mid nx = 0\}$, prove that $G = H_1 \oplus H_2$.
15. Let H_1 and H_2 be cyclic subgroups of the abelian group G , where $H_1 \cap H_2 = \{0\}$. Prove that $H_1 \oplus H_2$ is cyclic if and only if $o(H_1)$ and $o(H_2)$ are relatively prime.
16. (This is the additive version of Exercise 37 in section 4.6, with proofs the same except for notation.) Let H and K be arbitrary abelian groups with addition as the group operation, and let $H \oplus K$ denote the Cartesian product of H and K :

$$H \oplus K = \{(h, k) \mid h \in H \text{ and } k \in K\}.$$

Equality in $H \oplus K$ is defined by $(h, k) = (h', k')$ if and only if $h = h'$ and $k = k'$. Addition in $H \oplus K$ is defined by

$$(h_1, k_1) + (h_2, k_2) = (h_1 + h_2, k_1 + k_2).$$

- a. Prove that $H \oplus K$ is a group. This group is called the **external direct sum** of H and K .
- b. For simplicity, we denote the additive identity in both H and K by 0. Show that $H' = \{(h, 0) \mid h \in H\}$ is a normal subgroup of $H \oplus K$ that is isomorphic to H , and that $K' = \{(0, k) \mid k \in K\}$ is a normal subgroup isomorphic to K .
- c. Prove that $H \oplus K/H'$ is isomorphic to K and $H \oplus K/K'$ is isomorphic to H .
17. (See Exercise 16.) Find the order of each of the following elements.

a. $([2], [3])$ in $\mathbf{Z}_4 \oplus \mathbf{Z}_6$ c. $([2], [3])$ in $\mathbf{Z}_3 \oplus \mathbf{Z}_6$	b. $([2], [6])$ in $\mathbf{Z}_4 \oplus \mathbf{Z}_{12}$ d. $([2], [3])$ in $\mathbf{Z}_6 \oplus \mathbf{Z}_9$
--	---

Sec. 4.6, #37 ➤

- 18.** **a.** Find all subgroups of $\mathbf{Z}_2 \oplus \mathbf{Z}_4$.
b. Find all subgroups of $\mathbf{Z}_2 \oplus \mathbf{Z}_6$.
- 19.** **a.** Show that \mathbf{Z}_{15} is isomorphic to $\mathbf{Z}_3 \oplus \mathbf{Z}_5$, where the group operation in each of \mathbf{Z}_{15} , \mathbf{Z}_3 , and \mathbf{Z}_5 is addition.
b. Show that \mathbf{Z}_{12} is isomorphic to $\mathbf{Z}_3 \oplus \mathbf{Z}_4$, where all group operations are addition.
- 20.** Suppose that G and G' are abelian groups such that $G = H_1 \oplus H_2$ and $G' = H'_1 \oplus H'_2$. If H_1 is isomorphic to H'_1 and H_2 is isomorphic to H'_2 , prove that G is isomorphic to G' .
- 21.** Suppose a is an element of order rs in an abelian group G . Prove that if r and s are relatively prime, then a can be written in the form $a = b_1 + b_2$, where b_1 has order r and b_2 has order s .
- 22.** (See Exercise 21.) Assume that a is an element of order $r_1 r_2 \cdots r_n$ in an abelian group, where r_i and r_j are relatively prime if $i \neq j$. Prove that a can be written in the form $a = b_1 + b_2 + \cdots + b_n$, where each b_i has order r_i .
- 23.** Prove that if r and s are relatively prime positive integers, then any cyclic group of order rs is the direct sum of a cyclic group of order r and a cyclic group of order s .
- 24.** Prove Theorem 4.33: If H_1, H_2, \dots, H_n are subgroups of the abelian group G , then the sum $H_1 + H_2 + \cdots + H_n$ is direct if and only if the intersection of each H_j with the subgroup generated by $\bigcup_{i=1, i \neq j}^n H_i$ is the identity subgroup $\{0\}$.

4.8

Some Results on Finite Abelian Groups (Optional)

The aim of this section is to sample the flavor of more advanced work in groups while maintaining an acceptable level of rigor in the presentation. We attempt to achieve this balance by restricting our attention to proofs of results for abelian groups. There are instances where more general results hold, but their proofs are beyond the level of this text. In most instances of this sort, the more general results are stated informally and without proof.

The following definition of a p -group is fundamental to this entire section.

Definition 4.35 ■ p -Group

If p is a prime, then a group G is called a **p -group** if and only if each of its elements has an order that is a power of p .

A p -group can be finite or infinite. Although we do not prove it here, a finite group is a p -group if and only if its order is a power of p . Whether or not a group is abelian has nothing at all to do with whether it is a p -group. This is brought out in the following example.

Example 1 With $p = 2$, we can easily exhibit three p -groups of order 8.

- a. Consider first the cyclic group $C_8 = \langle a \rangle$ of order 8 generated by the permutation $a = (1, 2, 3, 4, 5, 6, 7, 8)$:

Each of a, a^3, a^5 , and a^7 has order 8.

a^2 and a^6 have order 4.

a^4 has order 2.

The identity e has order 1.

Thus C_8 is a 2-group.

- b. Consider now the quaternion group $G = \{\pm 1, \pm i, \pm j, \pm k\}$ of Exercise 28 in Section 3.1:

Each of the elements $\pm i, \pm j, \pm k$ has order 4.

-1 has order 2.

1 has order 1.

Hence G is another 2-group of order 8.

- c. Last, consider the octic group $G' = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$ of Example 3 in Section 4.5:

Each of α and α^3 has order 4.

Each of $\alpha^2, \beta, \gamma, \Delta, \theta$ has order 2.

The identity e has order 1.

Thus G' is also a 2-group of order 8.

Of these three p -groups, C_8 is abelian and both G and G' are nonabelian. ■

It may happen that G is not a p -group, yet some of its subgroups are p -groups. In connection with that possibility, we make the following definition.

Definition 4.36 ■ The Set G_p

If G is a finite abelian group that has order divisible by the prime p , then G_p is the set of all elements of G that have orders that are powers of p .

As might be expected, the set G_p turns out to be a subgroup. For the remainder of this section, we write all abelian groups in additive notation.

Theorem 4.37 ■ p -Subgroups

The set G_p defined in Definition 4.36 is a subgroup of G .

$u \Rightarrow v$ **Proof** The identity 0 has order $1 = p^0$, so $0 \in G_p$. If $a \in G_p$, then a has order p^r for some nonnegative integer r . Since a and its inverse $-a$ have the same order, $-a$ is also in the

set G_p . Let b be another element of the set G_p . Then b has order p^s for a nonnegative integer s . If t is the larger of r and s , then

$$\begin{aligned} p^t(a + b) &= p^t a + p^t b \\ &= 0 + 0 \\ &= 0. \end{aligned}$$

This implies that the order of $a + b$ divides p^t and is therefore a power of p since p is a prime. Thus $a + b \in G_p$, and set G_p is a subgroup of G .

Example 2 Consider the additive group $G = \mathbf{Z}_6$. The order of \mathbf{Z}_6 is 6, which is divisible by the primes 2 and 3. In this group:

Each of [1] and [5] has order 6.

Each of [2] and [4] has order 3.

[3] has order 2.

[0] has order 1.

For $p = 2$ or $p = 3$, the subgroups G_p are given by

$$\begin{aligned} G_2 &= \{[3], [0]\} \\ G_3 &= \{[2], [4], [0]\}. \end{aligned}$$

The group G is not a p -group, but G_2 is a 2-subgroup of G , and G_3 is a 3-subgroup of G . ■

If a group G has p -subgroups, certain of them are given special names, as described in the following definition.

Definition 4.38 ■ Sylow[†] p -Subgroup

If p is a prime and m is a positive integer such that $p^m|o(G)$ and $p^{m+1}\nmid o(G)$, then a subgroup of G that has order p^m is called a **Sylow p -subgroup** of G .

Example 3 In Example 2, G_2 is a Sylow 2-subgroup of G , and G_3 is a Sylow 3-subgroup of G . As a less trivial example, consider the octic group from Example 3 of Section 4.5:

$$H = \{e, \alpha, \alpha^2, \alpha^3, \beta, \gamma, \Delta, \theta\}$$

where

$$\begin{aligned} e &= (1) & \alpha &= (1, 2, 3, 4) & \alpha^2 &= (1, 3)(2, 4) & \alpha^3 &= (1, 4, 3, 2) \\ \beta &= (1, 4)(2, 3) & \gamma &= (2, 4) & \Delta &= (1, 2)(3, 4) & \theta &= (1, 3). \end{aligned}$$

The group H is a subgroup of order 2^3 in the symmetric group $G = S_4$, which has order $4! = 24$. Since $2^3|o(S_4)$ and $2^4\nmid o(S_4)$, the octic group is a Sylow 2-subgroup of S_4 . ■

[†]Peter Ludwig Mejdell Sylow (1832–1918) was a Norwegian mathematician who worked in group theory, publishing his Sylow theorems in a 10-page paper in 1872.

Theorem 4.39 ■ Cauchy's[†] Theorem for Abelian Groups

If G is an abelian group of order n and p is a prime such that $p \mid n$, then G has at least one element of order p .

Induction Proof The proof is by induction on the order n of G and uses the Second Principle of Finite Induction. For $n = 1$, the theorem holds by default.

Now let k be a positive integer, assume that the theorem is true for all positive integers $n < k$, and let G be an abelian group of order k . Also, suppose that the prime p is a divisor of k .

Consider first the case where G has only the trivial subgroups $\{0\}$ and G . Then any $a \neq 0$ in G must be a generator of G , $G = \langle a \rangle$. It follows from Exercise 38 of Section 3.4 that the order k of G must be a prime. Since p divides this order, p must equal k , and G actually has $p - 1$ elements of order p , by Theorem 3.22.

Now consider the case where G has a nontrivial subgroup H ; that is, $H \neq \{0\}$ and $H \neq G$, so that $1 < o(H) < k$. If $p \mid o(H)$, then H contains an element of order p by the induction hypothesis, and the theorem is true for G . Suppose then that $p \nmid o(H)$. Since G is abelian, H is normal in G , and the quotient group G/H has order

$$o(G/H) = \frac{o(G)}{o(H)}.$$

We have

$$o(G) = o(H)o(G/H),$$

so p divides the product $o(H)o(G/H)$. Since p is a prime and $p \nmid o(H)$, p must divide $o(G/H) < o(G) = k$. Applying the induction hypothesis, we see that the abelian group G/H has an element $b + H$ of order p . Then

$$H = p(b + H) = pb + H,$$

and therefore $pb \in H$, where $b \notin H$. Let $r = o(H)$. The order of pb must be a divisor of r so that $r(pb) = 0$ and $p(rb) = 0$. Since p is a prime and $p \nmid r$, p and r are relatively prime. Hence there exist integers u and v such that $pu + rv = 1$.

The contention now is that the element $c = rb$ has order p . We have $pc = 0$, and we need to show that $c = rb \neq 0$. Assume the contrary, that $rb = 0$. Then

$$\begin{aligned} b &= 1b \\ &= (pu + rv)b \\ &= u(pb) + v(rb) \\ &= u(pb) + 0 \\ &= u(pb). \end{aligned}$$

Now $pb \in H$, and therefore $u(pb) \in H$. But $b \notin H$, so we have a contradiction. Thus $c = rb \neq 0$ is an element of order p in G , and the proof is complete.

Cauchy's Theorem also holds for nonabelian groups, but we do not prove it here. The next theorem applies only to abelian groups.

[†]A biographical sketch of Augustin Louis Cauchy (1789–1857) is given at the end of this chapter.

Theorem 4.40 ■ Sylow p -Subgroup

If G is a finite abelian group and p is a prime such that $p \mid o(G)$, then G_p is a Sylow p -subgroup.

$(u \wedge v) \Rightarrow w$ **Proof** Assume that G is a finite abelian group such that p^m divides $o(G)$ but p^{m+1} does not divide $o(G)$. Then $o(G) = p^m k$, where p and k are relatively prime. We need to prove that G_p has order p^m .

We first argue that $o(G_p)$ is a power of p . If $o(G_p)$ had a prime factor q different from p , then G_p would have to contain an element of order q , according to Cauchy's Theorem. This would contradict the very definition of G_p , so we conclude that $o(G_p)$ is a power of p . Let $o(G_p) = p^t$.

Suppose now that $o(G_p) < p^m$ —that is, that $t < m$. Then the quotient group G/G_p has order $p^m k / p^t = p^{m-t} k$, which is divisible by p . Hence G/G_p contains an element $a + G_p$ of order p , by Theorem 4.39. Then

$$G_p = p(a + G_p) = pa + G_p,$$

and this implies that $pa \in G_p$. Thus pa has order that is a power of p . This implies that a has order a power of p , and therefore $a \in G_p$; that is, $a + G_p = G_p$. This is a contradiction to the fact that $a + G_p$ has order p . Therefore, $o(G_p) = p^m$, and G_p is a Sylow p -subgroup of G .

The next theorem shows the true significance of the Sylow p -subgroups in the structure of abelian groups.

Theorem 4.41 ■ Direct Sum of Sylow p -Subgroups

Let G be an abelian group of order $n = p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r}$ where the p_i are distinct primes and each m_i is a positive integer. Then

$$G = G_{p_1} \oplus G_{p_2} \oplus \cdots \oplus G_{p_r}$$

where G_{p_i} is the Sylow p_i -subgroup of G that corresponds to the prime p_i .

$u \Rightarrow v$ **Proof** Assume the hypothesis of the theorem. For each prime p_i , G_{p_i} is a Sylow p -subgroup of G by Theorem 4.40. Suppose an element $a_1 \in G_{p_1}$ is also in the subgroup generated by $G_{p_2}, G_{p_3}, \dots, G_{p_r}$. Then

$$a_1 = a_2 + a_3 + \cdots + a_r$$

where $a_i \in G_{p_i}$. Since G_{p_i} has order $p_i^{m_i}$, $p_i^{m_i} a_i = 0$ for $i = 2, \dots, r$. Hence

$$p_2^{m_2} p_3^{m_3} \cdots p_r^{m_r} a_1 = 0.$$

Since the order of any $a_1 \in G_{p_1}$ is a power of p_1 , and p_1 is relatively prime to $p_2^{m_2} p_3^{m_3} \cdots p_r^{m_r}$, this requires that $a_1 = 0$. A similar argument shows that the intersection of any G_{p_i} with the subgroup generated by the remaining subgroups

$$G_{p_1}, G_{p_2}, \dots, G_{p_{i-1}}, G_{p_{i+1}}, \dots, G_{p_r}$$

is the identity subgroup $\{0\}$. Hence the sum

$$G_{p_1} \oplus G_{p_2} \oplus \cdots \oplus G_{p_r}$$

is direct and has order equal to the product of the orders $p_i^{m_i}$:

$$o(G_{p_1} \oplus G_{p_2} \oplus \cdots \oplus G_{p_r}) = p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r} = o(G).$$

Therefore,

$$G = G_{p_1} \oplus G_{p_2} \oplus \cdots \oplus G_{p_r}.$$

Example 4 In Example 2, $G = G_2 \oplus G_3$. ■

Our next theorem is concerned with a class that is more general than finite abelian groups, the *finitely generated abelian groups*. An abelian group G is said to be **finitely generated** if there exists a set of elements $\{a_1, a_2, \dots, a_n\}$ in G such that every $x \in G$ can be written in the form

$$x = z_1 a_1 + z_2 a_2 + \cdots + z_n a_n$$

where each z_i is an integer. The elements a_i are called **generators** of G , and the set $\{a_1, a_2, \dots, a_n\}$ is called a **generating set** for G . A finite abelian group G is surely a finitely generated group, since G itself is a generating set.

In a finitely generated group, the Well-Ordering Principle assures us that there are generating sets that have the smallest possible number of elements. Such sets are called **minimal generating sets**. The number of elements in a minimal generating set for G is called the **rank** of G .

Theorem 4.42 ■ Direct Sum of Cyclic Groups

Any finitely generated abelian group G (and therefore any finite abelian group) is a direct sum of cyclic groups.

Induction **Proof** The proof is by induction on the rank of G . If G has rank 1, then G is cyclic and the theorem is true.

Assume that the theorem is true for any group of rank $k - 1$, and let G be a group of rank k . We consider two cases.

Case 1 Suppose there exists a minimal generating set $\{a_1, a_2, \dots, a_k\}$ for G such that any relation of the form

$$z_1 a_1 + z_2 a_2 + \cdots + z_k a_k = 0$$

with $z_i \in \mathbf{Z}$ implies that $z_1 a_1 = z_2 a_2 = \cdots = z_k a_k = 0$. Then

$$G = \langle a_1 \rangle + \langle a_2 \rangle + \cdots + \langle a_k \rangle,$$

and the theorem is true for this case.

Case 2 Suppose that Case 1 does not hold. That is, for any minimal generating set $\{a_1, a_2, \dots, a_k\}$ of G , there exists a relation of the form

$$z_1a_1 + z_2a_2 + \cdots + z_ka_k = 0$$

with $z_i \in \mathbf{Z}$ such that some of the $z_i a_i \neq 0$. Among all the minimal generating sets and all the relations of this form, there exists a smallest positive integer \bar{z}_i that occurs as a coefficient in one of these relations. Suppose this \bar{z}_i occurs in a relation with the generating set $\{b_1, b_2, \dots, b_k\}$. If necessary, the elements in $\{b_1, b_2, \dots, b_k\}$ can be rearranged so that this smallest positive coefficient occurs as \bar{z}_1 with b_1 in

$$\bar{z}_1b_1 + \bar{z}_2b_2 + \cdots + \bar{z}_kb_k = 0. \quad (1)$$

Now let s_1, s_2, \dots, s_k be any set of integers that occur as coefficients in a relation of the form

$$s_1b_1 + s_2b_2 + \cdots + s_kb_k = 0 \quad (2)$$

with these generators b_i . We shall show that \bar{z}_1 divides s_1 . By the Division Algorithm, $s_1 = \bar{z}_1q_1 + r_1$, where $0 \leq r_1 < \bar{z}_1$. Multiplying equation (1) by q_1 and subtracting the result from equation (2), we have

$$r_1b_1 + (s_2 - \bar{z}_2q_1)b_2 + \cdots + (s_k - \bar{z}_kq_1)b_k = 0.$$

The condition $0 \leq r_1 < \bar{z}_1$ forces $r_1 = 0$ by choice of \bar{z}_1 as the smallest positive integer in a relation of this form. Thus \bar{z}_1 is a factor of s_1 .

We now show that $\bar{z}_1 | \bar{z}_i$ for $i = 2, \dots, k$. Consider \bar{z}_2 , for example. By the Division Algorithm, $\bar{z}_2 = \bar{z}_1q_2 + r_2$, where $0 \leq r_2 < \bar{z}_1$. If we let $b'_1 = b_1 + q_2b_2$, then $\{b'_1, b_2, \dots, b_k\}$ is a minimal generating set for G , and

$$\begin{aligned} & \bar{z}_1b_1 + \bar{z}_2b_2 + \cdots + \bar{z}_kb_k = 0 \\ \Rightarrow & \bar{z}_1(b'_1 - q_2b_2) + \bar{z}_2b_2 + \cdots + \bar{z}_kb_k = 0 \\ \Rightarrow & \bar{z}_1b'_1 + (\bar{z}_2 - \bar{z}_1q_2)b_2 + \cdots + \bar{z}_kb_k = 0 \\ \Rightarrow & \bar{z}_1b'_1 + r_2b_2 + \cdots + \bar{z}_kb_k = 0. \end{aligned}$$

Now $r_2 \neq 0$ and $0 \leq r_2 < \bar{z}_1$ would contradict the choice of \bar{z}_1 , so it must be that $r_2 = 0$ and $\bar{z}_1 | \bar{z}_2$. The same sort of argument can be applied to each of $\bar{z}_3, \dots, \bar{z}_k$, so we have $\bar{z}_i = \bar{z}_1q_i$ for $i = 2, \dots, k$. Substituting in equation (1), we obtain

$$\bar{z}_1b_1 + \bar{z}_1q_2b_2 + \cdots + \bar{z}_1q_kb_k = 0.$$

Let $c_1 = b_1 + q_2b_2 + \cdots + q_kb_k$, and consider the set $\{c_1, b_2, \dots, b_k\}$. This set generates G , and we have

$$\begin{aligned} \bar{z}_1c_1 &= \bar{z}_1b_1 + \bar{z}_1q_2b_2 + \cdots + \bar{z}_1q_kb_k \\ &= \bar{z}_1b_1 + \bar{z}_2b_2 + \cdots + \bar{z}_kb_k \\ &= 0. \end{aligned}$$

If H denotes the subgroup of G that is generated by the set $\{b_2, \dots, b_k\}$, then $G = \langle c_1 \rangle + H$ since the set $\{c_1, b_2, \dots, b_k\}$ is a generating set for G . We shall show that the sum is direct.

If s_1, s_2, \dots, s_k are any integers such that

$$s_1c_1 + s_2b_2 + \cdots + s_kb_k = 0,$$

then substitution for c_1 yields

$$s_1b_1 + (s_1q_2 + s_2)b_2 + \cdots + (s_1q_k + s_k)b_k = 0.$$

This implies that \bar{z}_1 divides s_1 , and therefore $s_1c_1 = 0$ since $\bar{z}_1c_1 = 0$. Hence the sum is direct, and

$$G = \langle c_1 \rangle \oplus H.$$

Since H has rank $k - 1$, the induction hypothesis applies to H , and H is a direct sum of cyclic groups. Therefore, G is a direct sum of cyclic groups, and the theorem follows by induction.

We can now give a complete description of the structure of any finite abelian group G . As in Theorem 4.41,

$$G = G_{p_1} \oplus G_{p_2} \oplus \cdots \oplus G_{p_r}$$

where G_{p_i} is the Sylow p_i -subgroup of order $p_i^{m_i}$ corresponding to the prime p_i . Each G_{p_i} can in turn be decomposed into a direct sum of cyclic subgroups $\langle a_{i,j} \rangle$, each of which has order a power of p_i :

$$G_{p_i} = \langle a_{i,1} \rangle \oplus \langle a_{i,2} \rangle \oplus \cdots \oplus \langle a_{i,t_i} \rangle$$

where the product of the orders of the subgroups $\langle a_{i,j} \rangle$ is $p_i^{m_i}$. This description is frequently referred to as the **Fundamental Theorem on Finite Abelian Groups**. It can be used to systematically describe all the abelian groups of a given finite order, up to isomorphism.

Example 5 For n a positive integer, let C_n denote a cyclic group of order n . If G is an abelian group of order $72 = 2^3 \cdot 3^2$, then G is the direct sum of its Sylow p -subgroups G_2 of order 2^3 and G_3 of order 3^2 :

$$G = G_2 \oplus G_3.$$

Each of G_2 and G_3 is a sum of cyclic groups as described in the preceding paragraph. By considering all possibilities for the decompositions of G_2 and G_3 , we deduce that any abelian group of order 72 is isomorphic to one of the following direct sums of cyclic groups:

$C_{2^3} \oplus C_{3^2}$	$C_{2^3} \oplus C_3 \oplus C_3$
$C_2 \oplus C_{2^2} \oplus C_{3^2}$	$C_2 \oplus C_{2^2} \oplus C_3 \oplus C_3$
$C_2 \oplus C_2 \oplus C_2 \oplus C_{3^2}$	$C_2 \oplus C_2 \oplus C_2 \oplus C_3 \oplus C_3$

■

The main emphasis of this section has been on finite abelian groups, but the results presented here hardly scratch the surface. As an example of the interesting and important work that has been done on finite groups in general, we state the following theorem without proof.

Theorem 4.43 ■ Sylow's Theorem

Let G be a finite group, and let p be a prime integer.

- If m is a positive integer such that $p^m \mid o(G)$ and $p^{m+1} \nmid o(G)$, then G has a subgroup of order p^m .
- For the same prime p , any two Sylow p -subgroups of G are conjugate subgroups.
- If $p \mid o(G)$, the number n_p of distinct Sylow p -subgroups of G satisfies $n_p \equiv 1 \pmod{p}$.

The result in part **a** of Theorem 4.43 can be generalized to state that if $p^m \mid o(G)$ and $p^{m+1} \nmid o(G)$, then G has a subgroup of order p^k for any $k \in \mathbf{Z}$ such that $0 \leq k \leq m$.

Exercises 4.8**True or False**

Label each of the following statements as either true or false.

1. A p -group can be finite or infinite.
2. Every p -group is abelian.
3. Every p -group is cyclic.
4. Every subgroup of a p -group is a p -group.
5. Every Sylow p -subgroup of a group G is cyclic.
6. If every nontrivial subgroup of a group G is a p -group, then G must be a p -group.

Exercises

1. Give an example of a p -group of order 9.
2. Find two p -groups of order 4 that are not isomorphic.
3. a. Find all Sylow 3-subgroups of the alternating group A_4 .
b. Find all Sylow 2-subgroups of A_4 .
4. Find all Sylow 3-subgroups of the symmetric group S_4 .
5. For each of the following \mathbf{Z}_n , let G be the additive group $G = \mathbf{Z}_n$, and write G as a direct sum of cyclic groups.

a. \mathbf{Z}_{10}	b. \mathbf{Z}_{15}	c. \mathbf{Z}_{12}	d. \mathbf{Z}_{18}
----------------------	----------------------	----------------------	----------------------
6. For each of the following values of n , describe all the abelian groups of order n , up to isomorphism.

a. $n = 6$	b. $n = 10$	c. $n = 12$
d. $n = 18$	e. $n = 36$	f. $n = 100$
- Sec. 4.4, #8 ➤ 7. Let G be a group and $g \in G$. Prove that if H is a Sylow p -group of G , then so is gHg^{-1} .

8. Let G be a finite group, p prime, and H a Sylow p -group. Prove that H is normal in G if and only if H is the only Sylow p -group in G .
9. Determine which of the Sylow p -groups in each part of Exercise 3 are normal.
10. Determine which of the Sylow 3-groups in Exercise 4 are normal.
11. Show that $\{a_1, a_2, \dots, a_n\}$ is a generating set for the additive abelian group G if and only if $G = \langle a_1 \rangle + \langle a_2 \rangle + \dots + \langle a_n \rangle$.
12. Give an example where G is a finite *nonabelian* group with order that is divisible by a prime p , and where the set of all elements that have orders that are powers of p is *not* a subgroup of G .
13. If p_1, p_2, \dots, p_r are distinct primes, prove that any two abelian groups that have order $n = p_1 p_2 \cdots p_r$ are isomorphic.
14. Suppose that the abelian group G can be written as the direct sum $G = C_{2^2} \oplus C_3 \oplus C_3$, where C_n is a cyclic group of order n .
 - a. Prove that G has elements of order 12 but no element of order greater than 12.
 - b. Find the number of distinct elements of G that have order 12.
15. Assume that G can be written as the direct sum $G = C_2 \oplus C_2 \oplus C_3 \oplus C_3$, where C_n is a cyclic group of order n .
 - a. Prove that G has elements of order 6 but no element of order greater than 6.
 - b. Find the number of distinct elements of G that have order 6.
16. Suppose that G is a *cyclic* group of order p^m , where p is a prime. If k is any integer such that $0 \leq k \leq m$, prove that G has a subgroup of order p^k .
17. Prove the result in Exercise 16 for an arbitrary *abelian* group G of order p^m , where G is not necessarily cyclic.
18. Prove that if G is an abelian group of order n and s is an integer that divides n , then G has a subgroup of order s .

Key Words and Phrases

alternating group, 199

Cauchy's Theorem for Abelian Groups, 249

Cayley's Theorem, 205

conjugate, 199, 221

cycle, 192

dihedral group, 210

direct product, 238, 239

direct sum of subgroups, 241

even permutation, 199

Fundamental Theorem of Homomorphisms, 233

generating set, 225, 251

glide reflection, 211

index of a subgroup, 218

Klein four group, 207

Lagrange's Theorem, 219

left coset, 217

minimal generating set, 251

normal (invariant) subgroup, 223

octic group, 202, 226

odd permutation, 199

orbit, 193

p -group, 246

product of subsets, 215

quotient (factor) group, 230

rank, 251

reflective symmetry, 211

right coset, 217

rotational symmetry, 211

subgroup generated by A , 225

sum of subgroups, 239

Sylow p -subgroup, 248

translation, 211

transposition, 196



SSPL/Image Works

A Pioneer in Mathematics

Augustin Louis Cauchy (1789–1857)

Augustin Louis Cauchy, a 19th-century French mathematician, has the distinction of being a major contributor to the development of modern calculus. The calculus that we know today is based substantially on his clear and precise definition of limits, which changed the whole complexion of the field. Cauchy's attention was not confined to calculus, though. In 1814, he began to develop the theory of functions of complex variables. He made significant contributions in the areas of differential equations, infinite series, probability, determinants, and mathematical physics, as well as abstract algebra. The current notation and terminology used for permutations are credited to Cauchy. A major theorem in the study of abelian groups (Theorem 4.39) was proved by Cauchy and thus was named for him.

Cauchy was born in Paris on August 21, 1789. By the time he was 11 years old, French mathematicians recognized his rare talent. He went on to study civil engineering and spent the first few years of his career as an engineer in Napoleon's army, pursuing mathematical research on the side. For health reasons, he gave up engineering and began a teaching career that was mathematically fruitful in spite of political unrest in France. In 1830, Cauchy, an ardent supporter of King Charles X, refused to swear allegiance to the new government after the exile of the king. He lost his professorship and was forced to leave France for eight years. He subsequently taught in church schools and produced so many papers that the Academy of Sciences, alarmed at the printing bills that resulted, passed a rule limiting each paper to four pages. After the February Revolution of 1848, Cauchy was appointed professor of celestial mechanics at the École Polytechnique, a position he retained for the rest of his career.

Rings, Integral Domains, and Fields

■ Introduction

Rings, integral domains, and fields are introduced in this chapter. The field of quotients of an integral domain is constructed, and ordered integral domains are considered. The development of \mathbf{Z}_n continues in Section 5.1, where it appears for the first time in its proper context as a ring.

5.1

Definition of a Ring

A group is one of the simpler algebraic systems because it has only one binary operation. A step upward in the order of complexity is the *ring*. A ring has two binary operations called *addition* and *multiplication*. Conditions are made on both binary operations, but fewer are made on multiplication. A full list of the conditions is in our formal definition.

Definition 5.1a ■ Definition of a Ring

Suppose R is a set in which a relation of equality, denoted by $=$, and operations of addition and multiplication, denoted by $+$ and \cdot , respectively, are defined. Then R is a **ring** (with respect to these operations) if the following conditions are satisfied:

1. R is **closed** under addition: $x \in R$ and $y \in R$ imply $x + y \in R$.
2. Addition in R is **associative**: $x + (y + z) = (x + y) + z$ for all x, y, z in R .
3. R contains an **additive identity** 0 : $x + 0 = 0 + x = x$ for all $x \in R$.
4. R contains **additive inverses**: For x in R , there exists $-x$ in R such that $x + (-x) = (-x) + x = 0$.
5. Addition in R is **commutative**: $x + y = y + x$ for all x, y in R .
6. R is **closed** under multiplication: $x \in R$ and $y \in R$ imply $x \cdot y \in R$.
7. Multiplication in R is **associative**: $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ for all x, y, z in R .
8. Two **distributive laws** hold in R : $x \cdot (y + z) = x \cdot y + x \cdot z$ and $(x + y) \cdot z = x \cdot z + y \cdot z$ for all x, y, z in R .

The notation xy will be used interchangeably with $x \cdot y$ to indicate multiplication.

The additive identity of a ring is denoted by 0 and referred to as the **zero** of the ring. The additive inverse $-a$ is called the **negative** of a or the **opposite** of a , and **subtraction** in a ring is defined by

$$x - y = x + (-y).$$

As in elementary algebra, we adhere to the convention that *multiplication takes precedence over addition*. That is, it is understood that in any expression involving multiplication and addition, multiplications are performed first. Thus $xy + xz$ represents $(x \cdot y) + (x \cdot z)$, not $x(y + x)z$.

The statement of the definition can be shortened to a form that is easier to remember if we note that the first five conditions amount to the requirement that R be an abelian group under addition.

Definition 5.1b ■ Alternative Definition of a Ring

Suppose R is a set in which a relation of equality, denoted by $=$, and operations of addition and multiplication, denoted by $+$ and \cdot , respectively, are defined. Then R is a **ring** (with respect to these operations) if these conditions hold:

1. R forms an **abelian group** with respect to **addition**.
 2. R is **closed** with respect to an **associative multiplication**.
 3. Two **distributive laws** hold in R : $x \cdot (y + z) = x \cdot y + x \cdot z$ and $(x + y) \cdot z = x \cdot z + y \cdot z$ for all x, y, z in R .
-

Example 1 Some simple examples of rings are provided by the familiar number systems with their usual operations of addition and multiplication:

- a. the set **Z** of all integers
- b. the set **Q** of all rational numbers
- c. the set **R** of all real numbers
- d. the set **C** of all complex numbers.

■

Example 2 We shall verify that the set **E** of all even integers is a ring with respect to the usual addition and multiplication in **Z**. The following conditions of Definition 5.1a are satisfied automatically since they hold throughout the ring **Z**, which contains **E**.

2. Addition in **E** is associative.
5. Addition in **E** is commutative.
7. Multiplication in **E** is associative.
8. The two distributive laws in Definition 5.1a hold in **E**.

The remaining conditions in Definition 5.1a may be checked as follows:

1. If $x \in \mathbf{E}$ and $y \in \mathbf{E}$, then $x = 2m$ and $y = 2n$ with m and n in **Z**. For the sum, we have $x + y = 2m + 2n = 2(m + n)$, which is in **E**. Thus **E** is closed under addition.
3. **E** contains the additive identity, since $0 = (2)(0)$.

4. For any $x = 2k$ in \mathbf{E} , the additive inverse of x is in \mathbf{E} , since $-x = 2(-k)$.
6. For $x = 2m$ and $y = 2n$ in \mathbf{E} , the product $xy = 2(2mn)$ is in \mathbf{E} , so \mathbf{E} is closed under multiplication. ■

Definition 5.2 ■ Subring

Whenever a ring R_1 is a subset of a ring R_2 and has addition and multiplication as defined in R_2 , we say that R_1 is a **subring** of R_2 .

Thus the ring \mathbf{E} of even integers is a subring of the ring \mathbf{Z} of all integers. From Example 1, we see that the ring \mathbf{Z} is a subring of the rational numbers, the rational numbers form a subring of the real numbers, and the real numbers form a subring of the complex numbers.

Generalizing from Example 2, we may observe that conditions 2, 5, 7, and 8 of Definition 5.1a are automatically satisfied in any subset of a ring, leaving only conditions 1, 3, 4, and 6 to be verified for the subset to form a subring. A slightly more efficient characterization of subrings is given in the following theorem, the proof of which is left as an exercise.

Theorem 5.3 ■ Equivalent Set of Conditions for a Subring

A subset S of the ring R is a subring of R if and only if these conditions are satisfied:

- a. S is nonempty.
- b. $x \in S$ and $y \in S$ imply that $x + y$ and xy are in S .
- c. $x \in S$ implies $-x \in S$.

An even more efficient characterization of subrings is provided by the next theorem. The proof of this theorem is left as an exercise.

Theorem 5.4 ■ Characterization of a Subring

A subset S of the ring R is a subring of R if and only if these conditions are satisfied:

- a. S is nonempty.
- b. $x \in S$ and $y \in S$ imply that $x - y$ and xy are in S .

Example 3 Using Theorem 5.3 or Theorem 5.4, it is not difficult to verify the following examples of subrings.

- a. The set of all real numbers of the form $m + n\sqrt{2}$, with $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$, is a subring of the ring of all real numbers.
- b. The set of all real numbers of the form $a + b\sqrt[3]{2}$, with a and b rational numbers, is a subring of the real numbers.
- c. The set of all real numbers of the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$, with a , b , and c rational numbers, is a subring of the real numbers. ■

The preceding examples of rings are all drawn from the number systems. The next example exhibits a class of rings with a different flavor: They are **finite rings** (that is, rings with a finite number of elements). The next example is also important because it presents the set \mathbf{Z}_n of congruence classes modulo n for the first time in its proper context as a *ring*.

Example 4 For $n > 1$, let \mathbf{Z}_n denote the congruence classes of the integers modulo n :

$$\mathbf{Z}_n = \{[0], [1], [2], \dots, [n - 1]\}.$$

We have previously seen that the rules

$$[a] + [b] = [a + b] \quad \text{and} \quad [a] \cdot [b] = [ab]$$

define binary operations of addition and multiplication in \mathbf{Z}_n . We have seen that \mathbf{Z}_n forms an abelian group under addition, with $[0]$ as the additive identity and $[-a]$ as the additive inverse of $[a]$. It has also been noted that \mathbf{Z}_n is closed with respect to multiplication and that this multiplication is associative. For arbitrary $[a], [b], [c]$ in \mathbf{Z}_n , we have

$$\begin{aligned} [a] \cdot ([b] + [c]) &= [a] \cdot [b + c] \\ &= [a(b + c)] \\ &= [ab + ac] \\ &= [ab] + [ac] \\ &= [a] \cdot [b] + [a] \cdot [c], \end{aligned}$$

so the left distributive law holds in \mathbf{Z}_n . The right distributive law can be verified in a similar way, and \mathbf{Z}_n is a ring with respect to these operations. ■

Making use of some results from Chapter 1, we can obtain an example of a ring quite different from any of those previously discussed.

Example 5 Let U be a nonempty universal set, and let $\mathcal{P}(U)$ denote the collection of all subsets of U .

For arbitrary subsets A and B of U , let $A + B$ be defined as in Exercise 40 of Section 1.1:

$$A + B = (A \cup B) - (A \cap B).$$

This rule defines an operation of addition on the subsets of U , $\mathcal{P}(U)$ is closed with respect to this addition, and this operation is associative, by Exercise 40b of Section 1.1. This addition is commutative, since $A \cup B = B \cup A$ and $A \cap B = B \cap A$. The empty set \emptyset is an additive identity because

$$\begin{aligned} \emptyset + A &= A + \emptyset \\ &= (A \cup \emptyset) - (A \cap \emptyset) \\ &= A - \emptyset \\ &= A. \end{aligned}$$

An unusual feature here is that each subset A of U is its own additive inverse:

$$\begin{aligned} A + A &= (A \cup A) - (A \cap A) \\ &= A - A \\ &= \emptyset. \end{aligned}$$

We define multiplication in $\mathcal{P}(U)$ by

$$A \cdot B = A \cap B,$$

and $\mathcal{P}(U)$ is closed with respect to this multiplication. Also multiplication is associative since

$$\begin{aligned} A \cdot (B \cdot C) &= A \cap (B \cap C) \\ &= (A \cap B) \cap C \\ &= (A \cdot B) \cdot C. \end{aligned}$$

The left distributive law $A \cap (B + C) = (A \cap B) + (A \cap C)$ is part c of Exercise 40, Section 1.1, and the right distributive law follows from this one since forming intersections of sets is a commutative operation. Thus $\mathcal{P}(U)$ is a ring with respect to the operations $+$ and \cdot as we have defined them. ■

Definition 5.5 ■ Ring with Unity, Commutative Ring

Let R be a ring. If there exists an element e in R such that $x \cdot e = e \cdot x = x$ for all x in R , then e is called a **unity**, and R is a **ring with unity**. If multiplication in R is commutative, then R is called a **commutative ring**.

A ring may have one of the properties in Definition 5.5 without the other, it may have neither, or it may have both of the properties. These possibilities are illustrated in the following examples.

Example 6 The ring \mathbf{Z} of all integers has both properties, so \mathbf{Z} is a commutative ring with a unity. As other examples of this type, \mathbf{Z}_n is a commutative ring with unity [1], and $\mathcal{P}(U)$ is a commutative ring with the subset U as unity. ■

Example 7 The ring \mathbf{E} of all even integers is a commutative ring, but \mathbf{E} does not have a unity. ■

Example 8 It follows from our work in Sections 1.6 and 3.3 that if $n \geq 2$, then each of the sets in the list

$$M_n(\mathbf{Z}) \subseteq M_n(\mathbf{Q}) \subseteq M_n(\mathbf{R}) \subseteq M_n(\mathbf{C})$$

is a noncommutative ring with unity I_n . Each of these four rings is a subring of every listed ring in which it is contained. ■

Example 9 The set

$$M_2(\mathbf{E}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a, b, c, \text{ and } d \text{ are in } \mathbf{E} \right\}$$

of all 2×2 matrices over the ring \mathbf{E} of even integers is a noncommutative ring that does not have a unity. ■

The definition of a unity allows the possibility of more than one unity in a ring. However, this possibility cannot happen.

Theorem 5.6 ■ Uniqueness of the Unity

If R is a ring that has a unity, the unity is unique.

Uniqueness **Proof** Suppose that both e and e' are unity elements in a ring R . Consider the product $e \cdot e'$ in R . On the one hand, we have $e \cdot e' = e$, since e' is a unity. On the other hand, $e \cdot e' = e'$, since e is a unity. Thus

$$e = e \cdot e' = e',$$

and the unity is unique.

In general discussions, we shall denote a unity by e . When a ring R has a unity, it is in order to consider the existence of multiplicative inverses.

Definition 5.7 ■ Multiplicative Inverse

Let R be a ring with unity e , and let $a \in R$. If there is an element x in R such that $ax = xa = e$, then x is a **multiplicative inverse** of a and a is called a **unit** (or an **invertible element**) in R .

As with the unity, a multiplicative inverse of an element is unique whenever it exists. The proof of this is left as an exercise.

Theorem 5.8 ■ Uniqueness of the Multiplicative Inverse

Suppose R is a ring with unity e . If an element $a \in R$ has a multiplicative inverse, the multiplicative inverse of a is unique.

We shall use the standard notation a^{-1} to denote the multiplicative inverse of a , if the inverse exists.

Example 10 Some elements in a ring R may have multiplicative inverses whereas others do not. In the ring \mathbf{Z}_{10} , [1] and [9] are their own multiplicative inverses, whereas [3] and [7] are inverses of each other. All other elements of \mathbf{Z}_{10} do not have multiplicative inverses. ■

Since every ring R forms an abelian group with respect to addition, many of our results for groups have immediate applications concerning addition in a ring. For example, Theorem 3.4 gives these results:

1. The zero element in R is unique.
2. For each x in R , $-x$ is unique.
3. For each x in R , $-(-x) = x$.
4. For any x and y in R , $-(x + y) = -y - x$.
5. If a , x , and y are in R and $a + x = a + y$, then $x = y$.

Whenever both addition and multiplication are involved, the results are not so direct, but they turn out much as we might expect. One basic result of this type is that a product is 0 if one of the factors is 0.

Theorem 5.9 ■ Zero Product

If R is a ring, then

$$a \cdot 0 = 0 \cdot a = 0$$

for all $a \in R$.

Proof Let a be arbitrary in R . We reduce $a \cdot 0$ to 0 by using various conditions in Definition 5.1a, as indicated:

$$\begin{aligned} a \cdot 0 &= a \cdot 0 + 0 && \text{by condition 3} \\ &= a \cdot 0 + \{a \cdot 0 + [-(a \cdot 0)]\} && \text{by condition 4} \\ &= (a \cdot 0 + a \cdot 0) + [-(a \cdot 0)] && \text{by condition 2} \\ &= [a \cdot (0 + 0)] + [-(a \cdot 0)] && \text{by condition 8} \\ &= a \cdot 0 + [-(a \cdot 0)] && \text{by condition 3} \\ &= 0 && \text{by condition 4.} \end{aligned}$$

Similar steps can be used to reduce $0 \cdot a$ to 0.

Theorem 5.9 says that a product is 0 if one of the factors is 0. Note that the converse is not true: A product may be 0 when neither factor is 0. An illustration is provided by $[2] \cdot [5] = [0]$ in \mathbf{Z}_{10} .

Definition 5.10 ■ Zero Divisor

Let R be a ring and let $a \in R$. If $a \neq 0$, and if there exists an element $b \neq 0$ in R such that either $ab = 0$ or $ba = 0$, then a is called a **proper divisor of zero**, or a **zero divisor**.

If we compare the steps used in the proof of Theorem 5.9 to the last part of the proof of Theorem 2.2, we see that they are much the same. In the same fashion, the proof of the first part of the next theorem is parallel to another part of the proof of Theorem 2.2. The same sort of similarity exists between Exercises 1–10 of Section 2.1 and the remaining parts of Theorem 5.11. Because of this similarity, their proofs are left as exercises.

Theorem 5.11 ■ Additive Inverses and Products

For arbitrary x , y , and z in a ring R , the following equalities hold:

- a. $(-x)y = -(xy)$
- b. $x(-y) = -(xy)$

- c. $(-x)(-y) = xy$
- d. $x(y - z) = xy - xz$
- e. $(x - y)z = xz - yz.$

Proof of a Since the additive inverse $-(xy)$ of the element xy is unique, we only need to show that $xy + (-x)y = 0$. We have

$$\begin{aligned} xy + (-x)y &= [x + (-x)]y \quad \text{by the right distributive law} \\ &= 0 \cdot y \quad \text{by the definition of } -x \\ &= 0 \quad \text{by Theorem 5.9.} \end{aligned}$$

Even though a ring does not form a group with respect to multiplication, both associative laws in a ring R can be generalized by the procedure followed in Definition 3.6 and Theorem 3.7. For any integer $n \geq 2$, the expressions $a_1 + a_2 + \cdots + a_n$ and $a_1a_2 \cdots a_n$ are defined recursively by

$$a_1 + a_2 + \cdots + a_k + a_{k+1} = (a_1 + a_2 + \cdots + a_k) + a_{k+1}$$

and

$$a_1a_2 \cdots a_k a_{k+1} = (a_1a_2 \cdots a_k)a_{k+1}.$$

The details are too repetitive to present here, so we accept the following theorem without proof.

Theorem 5.12 ■ Generalized Associative Laws

Let $n \geq 2$ be a positive integer, and let a_1, a_2, \dots, a_n denote elements of a ring R . For any positive integer m such that $1 \leq m < n$,

$$(a_1 + a_2 + \cdots + a_m) + (a_{m+1} + \cdots + a_n) = a_1 + a_2 + \cdots + a_n$$

and

$$(a_1a_2 \cdots a_m)(a_{m+1} \cdots a_n) = a_1a_2 \cdots a_n.$$

Generalized distributive laws also hold in an arbitrary ring. This fact is stated in the following theorem, with the proofs left as exercises.

Theorem 5.13 ■ Generalized Distributive Laws

Let $n \geq 2$ be a positive integer, and let b, a_1, a_2, \dots, a_n denote elements of a ring R . Then we have

- a. $b(a_1 + a_2 + \cdots + a_n) = ba_1 + ba_2 + \cdots + ba_n$, and
 - b. $(a_1 + a_2 + \cdots + a_n)b = a_1b + a_2b + \cdots + a_nb.$
-

Exercises 5.1

True or False

Label each of the following statements as either true or false.

1. Every ring is an abelian group with respect to the operations of addition and multiplication.
2. Let R be a ring. The set $\{0\}$ is a subring of R with respect to the operations in R .
3. Let R be a ring. Then R is a subring of itself.
4. Both \mathbf{E} , the set of even integers, and $\mathbf{Z} - \mathbf{E}$, the set of odd integers, are subrings of the set \mathbf{Z} of all integers.
5. If one element in a ring R has a multiplicative inverse, then all elements in R must have multiplicative inverses.
6. Let x and y be elements in a ring R . If $xy = 0$, then either $x = 0$ or $y = 0$.
7. Let R be a ring with unity and S a subring (with unity) of R . Then R and S must have the same unity elements.
8. A unity exists in any commutative ring.
9. Any ring with unity must be commutative.

Exercises

1. Confirm the statements made in Example 3 by proving that the following sets are subrings of the ring of all real numbers.

Sec. 5.2, #1a ≪

a. the set of all real numbers of the form $m + n\sqrt{2}$, with $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$

Sec. 5.2, #1b ≪

b. the set of all real numbers of the form $a + b\sqrt{2}$, with a and b rational numbers

Sec. 5.2, #1d ≪

c. the set of all real numbers of the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$, with a , b , and c rational numbers

2. Decide whether each of the following sets is a ring with respect to the usual operations of addition and multiplication. If it is not a ring, state at least one condition in Definition 5.1a that fails to hold.

a. the set of all integers that are multiples of 5

b. the set of all real numbers of the form $m + n\sqrt{3}$ with $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$

c. the set of all real numbers of the form $a + b\sqrt[3]{5}$, where a and b are rational numbers

d. the set of all real numbers of the form $a + b\sqrt[3]{5} + c\sqrt[3]{25}$, where a , b , and c are rational numbers

e. the set of all positive real numbers

f. the set of all complex numbers of the form $m + ni$, where $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$ (This set is known as the **Gaussian integers**.)

Sec. 6.4, #11, 12 ≪

g. the set of all real numbers of the form $m + n\sqrt{2}$, where $m \in \mathbf{E}$ and $n \in \mathbf{Z}$

Sec. 5.2, #1h ≪

h. the set of all real numbers of the form $m + n\sqrt{2}$, where $m \in \mathbf{Z}$ and $n \in \mathbf{E}$

- Sec. 5.2, #4, 5 <<
3. Let $U = \{a, b\}$. Using addition and multiplication as they are defined in Example 5, construct addition and multiplication tables for the ring $\mathcal{P}(U)$ that consists of the elements $\emptyset, A = \{a\}, B = \{b\}, U$.

4. Follow the instructions in Exercise 3, and use the universal set $U = \{a, b, c\}$.
5. Let $U = \{a, b\}$. Define addition and multiplication in $\mathcal{P}(U)$ by $C + D = C \cup D$ and $CD = C \cap D$. Decide whether $\mathcal{P}(U)$ is a ring with respect to these operations. If it is not, state a condition in Definition 5.1a that fails to hold.
6. Work Exercise 5 using $U = \{a\}$.
7. Find all zero divisors in \mathbf{Z}_n for the following values of n .
- | | |
|--------------------|-------------------------------|
| a. $n = 6$ | b. $n = 8$ |
| c. $n = 10$ | d. $n = 12$ |
| e. $n = 14$ | f. n a prime integer |
8. For the given value of n , find all the units in \mathbf{Z}_n .
- | | |
|--------------------|-------------------------------|
| a. $n = 6$ | b. $n = 8$ |
| c. $n = 16$ | d. $n = 12$ |
| e. $n = 14$ | f. n a prime integer |
9. Prove Theorem 5.3: A subset S of the ring R is a subring of R if and only if these conditions are satisfied:
- S is nonempty.
 - $x \in S$ and $y \in S$ imply that $x + y$ and xy are in S .
 - $x \in S$ implies $-x \in S$.
10. Prove Theorem 5.4: A subset S of the ring R is a subring of R if and only if these conditions are satisfied:
- S is nonempty.
 - $x \in S$ and $y \in S$ imply that $x - y$ and xy are in S .
11. Assume R is a ring with unity e . Prove Theorem 5.8: If $a \in R$ has a multiplicative inverse, the multiplicative inverse of a is unique.
12. (See Example 4.) Prove the right distributive law in \mathbf{Z}_n :
- $$([a] + [b]) \cdot [c] = [a] \cdot [c] + [b] \cdot [c].$$
13. Complete the proof of Theorem 5.9 by showing that $0 \cdot a = 0$ for any a in a ring R .
14. Let R be a ring, and let x, y , and z be arbitrary elements of R . Complete the proof of Theorem 5.11 by proving the following statements.
- | | |
|--------------------------------|--------------------------------|
| a. $x(-y) = -(xy)$ | b. $(-x)(-y) = xy$ |
| c. $x(y - z) = xy - xz$ | d. $(x - y)z = xz - yz$ |
15. Let a and b be elements of a ring R . Prove that the equation $a + x = b$ has a unique solution.

16. Suppose that G is an abelian group with respect to addition, with identity element 0. Define a multiplication in G by $ab = 0$ for all $a, b \in G$. Show that G forms a ring with respect to these operations.
17. If R_1 and R_2 are subrings of the ring R , prove that $R_1 \cap R_2$ is a subring of R .
18. Find subrings R_1 and R_2 of \mathbf{Z} such that $R_1 \cup R_2$ is not a subring of \mathbf{Z} .
19. Find a specific example of two elements a and b in a ring R such that $ab = 0$ and $ba \neq 0$.
20. Find a specific example of two nonzero elements a and b in a ring R such that the equations $ax = b$ and $ya = b$ have solutions $x \neq y$.
21. Define a new operation of addition in \mathbf{Z} by $x \oplus y = x + y - 1$ with a new multiplication in \mathbf{Z} by $x \odot y = x + y - xy$. Verify that \mathbf{Z} forms a ring with respect to these operations.
22. Let R be a ring with unity and S be the set of all units in R .
 - a. Prove or disprove that S is a subring of R .
 - b. Prove or disprove that S is a group with respect to multiplication in R .
23. Prove that if a is a unit in a ring R with unity, then a is not a zero divisor.
24. (See Exercise 8.) Describe the units of \mathbf{Z}_n .
25. Suppose that a , b , and c are elements of a ring R such that $ab = ac$. Prove that if a has a multiplicative inverse, then $b = c$.
26. Let R be a ring with no zero divisors. Prove that if a , b , c , and d are elements in R such that $ab = c \neq 0$ and $ad = c \neq 0$, then $b = d$.
27. For a fixed element a of a ring R , prove that the set $\{x \in R \mid ax = 0\}$ is a subring of R .
28. For a fixed element a of a ring R , prove that the set $\{xa \mid x \in R\}$ is a subring of R .
29. Let R be a ring. Prove that the set $S = \{x \in R \mid xa = ax \text{ for all } a \in R\}$ is a subring of R . This subring is called the **center** of R .
30. Consider the set $R = \{[0], [2], [4], [6], [8]\} \subseteq \mathbf{Z}_{10}$.
 - a. Construct addition and multiplication tables for R , using the operations as defined in \mathbf{Z}_{10} .
 - b. Observe that R is a commutative ring with unity [6], and compare this unity with the unity in \mathbf{Z}_{10} .
 - c. Is R a subring of \mathbf{Z}_{10} ? If not, give a reason.
 - d. Does R have zero divisors?
 - e. Which elements of R have multiplicative inverses?
31. Consider the set $S = \{[0], [2], [4], [6], [8], [10], [12], [14], [16]\} \subseteq \mathbf{Z}_{18}$. Using addition and multiplication as defined in \mathbf{Z}_{18} , consider the following questions.
 - a. Is S a ring? If not, give a reason.
 - b. Is S a commutative ring with unity? If not, give a reason.
 - c. Is S a subring of \mathbf{Z}_{18} ? If not, give a reason.
 - d. Does S have zero divisors?
 - e. Which elements of S have multiplicative inverses?

- Sec. 6.2, #19 ≪
32. The addition table and part of the multiplication table for the ring $R = \{a, b, c\}$ are given in Figure 5.1. Use the distributive laws to complete the multiplication table.

+	a	b	c	
a	a	b	c	
b	b	c	a	
c	c	a	b	

·	a	b	c	
a	a	a	a	
b	a	c		
c	a		a	

Figure 5.1

- Sec. 6.2, #20 ≪
33. The addition table and part of the multiplication table for the ring $R = \{a, b, c, d\}$ are given in Figure 5.2. Use the distributive laws to complete the multiplication table.

+	a	b	c	d	
a	a	b	c	d	
b	b	c	d	a	
c	c	d	a	b	
d	d	a	b	c	

·	a	b	c	d	
a	a	a	a	a	
b	a	c			
c	a		a		
d	a		a	c	

Figure 5.2

34. Give an example of a zero divisor in the ring $M_2(\mathbf{Z})$.
35. Let a and b be elements in a ring R . If ab is a zero divisor, prove that either a or b is a zero divisor.
- Sec. 3.2, #4 ≫
Sec. 5.2, #18 ≪
Sec. 6.2, #3 ≪
36. An element x in a ring is called **idempotent** if $x^2 = x$. Find two different idempotent elements in $M_2(\mathbf{Z})$.
37. (See Exercise 36.) Show that the set of all idempotent elements of a commutative ring is closed under multiplication.
38. Let a be idempotent in a ring with unity. Prove $e - a$ is also idempotent.
39. Decide whether each of the following sets S is a subring of the ring $M_2(\mathbf{Z})$. If a set is not a subring, give a reason why it is not. If it is a subring, determine if S is commutative and find the unity, if one exists. For those that have a unity, which elements in S have multiplicative inverses in S ?

a. $S = \left\{ \begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix} \mid x \in \mathbf{Z} \right\}$

Sec. 6.1, #21 ≪

c. $S = \left\{ \begin{bmatrix} x & y \\ x & y \end{bmatrix} \mid x, y \in \mathbf{Z} \right\}$

e. $S = \left\{ \begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix} \mid x \in \mathbf{Z} \right\}$

g. $S = \left\{ \begin{bmatrix} x & 0 \\ 0 & 2x \end{bmatrix} \mid x \in \mathbf{Z} \right\}$

b. $S = \left\{ \begin{bmatrix} x & x \\ 0 & 0 \end{bmatrix} \mid x \in \mathbf{Z} \right\}$

d. $S = \left\{ \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} \mid x, y, z \in \mathbf{Z} \right\}$

f. $S = \left\{ \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \mid x, y \in \mathbf{Z} \right\}$

h. $S = \left\{ \begin{bmatrix} x & 0 \\ 0 & x^2 \end{bmatrix} \mid x \in \mathbf{Z} \right\}$

- 40.** Let $S = \left\{ \begin{bmatrix} a+bi & c+di \\ -c+di & a-bi \end{bmatrix} \mid a, b, c, d \in \mathbf{R} \right\}$.
- Show that S is a noncommutative subring of $M_2(\mathbf{C})$.
 - Find the unity element, if it exists.
- 41.** Consider the set T of all 2×2 matrices of the form $\begin{bmatrix} a & a \\ b & b \end{bmatrix}$, where a and b are real numbers, with the same rules for addition and multiplication as in $M_2(\mathbf{R})$.
- Show that T is a ring that does not have a unity.
 - Show that T is not a commutative ring.
- 42.** Prove the following equalities in an arbitrary ring R .
- $(x+y)(z+w) = (xz + xw) + (yz + yw)$
 - $(x+y)(z-w) = (xz + yz) - (xw + yw)$
 - $(x-y)(z-w) = (xz + yw) - (xw + yz)$
 - $(x+y)(x-y) = (x^2 - y^2) + (yx - xy)$
- 43.** Let R be a set of elements containing the unity e , that satisfy all of the conditions in Definition 5.1a, except condition 5: Addition is commutative. Prove that condition 5 must also hold.
- 44.** Prove Theorem 5.13a.
- 45.** Prove Theorem 5.13b.
- 46.** An element a of a ring R is called **nilpotent** if $a^n = 0$ for some positive integer n . Prove that the set of all nilpotent elements in a commutative ring R forms a subring of R .
- 47.** Let R and S be arbitrary rings. In the Cartesian product $R \times S$ of R and S , define
- Sec. 5.2, #21 ≪ $(r, s) = (r', s') \text{ if and only if } r = r' \text{ and } s = s'$,
 - Sec. 6.2, #21 ≪ $(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2)$,
 - Sec. 6.3, #2, 6, 7 ≪ $(r_1, s_1) \cdot (r_2, s_2) = (r_1 r_2, s_1 s_2)$.
 - Sec. 6.4, #24, 25 ≪
- Prove that the Cartesian product is a ring with respect to these operations. It is called the **direct sum** of R and S and is denoted by $R \oplus S$.
 - Prove that $R \oplus S$ is commutative if both R and S are commutative.
 - Prove that $R \oplus S$ has a unity element if both R and S have unity elements.
- 48.** (See Exercise 47.) Write out the elements of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ and construct addition and multiplication tables for this ring (*Suggestion:* Write 0 for [0], 1 for [1] in \mathbf{Z}_2 .)
- 49.**
 - Show that $S_1 = \{[0], [2]\}$ is a subring of \mathbf{Z}_4 , and $S_2 = \{[0], [3]\}$ is a subring of \mathbf{Z}_6 .
 - Write out the elements of $S_1 \oplus S_2$, and construct addition and multiplication tables for this ring.
 - Is $S_1 \oplus S_2$ a commutative ring?
 - Find the unity in $S_1 \oplus S_2$ if one exists.
- 50.** Suppose R is a ring in which all elements x satisfy $x^2 = x$. (Such a ring is called a **Boolean ring**.)
- Prove that $x = -x$ for each $x \in R$. (*Hint:* Consider $(x + x)^2$.)
 - Prove that R is commutative. (*Hint:* Consider $(x + y)^2$.)

5.2 Integral Domains and Fields

In the preceding section we defined the terms *ring with unity*, *commutative ring*, and *zero divisors*. All three of these terms are used in defining an *integral domain*.

Definition 5.14 ■ Integral Domain

Let D be a ring. Then D is an **integral domain** provided these conditions hold:

1. D is a commutative ring.
2. D has a unity e , and $e \neq 0$.
3. D has no zero divisors.

Note that the requirement $e \neq 0$ means that an integral domain must have at least two elements.

Example 1 The ring \mathbf{Z} of all integers is an integral domain, but the ring \mathbf{E} of all even integers is not an integral domain, because it does not contain a unity. As familiar examples of integral domains, we can list the set of all rational numbers, the set of all real numbers, and the set of all complex numbers—all of these with their usual operations. ■

Example 2 The ring \mathbf{Z}_{10} is a commutative ring with a unity, but the presence of zero divisors such as $[2]$ and $[5]$ prevents \mathbf{Z}_{10} from being an integral domain. Considered as a possible integral domain, the ring M of all 2×2 matrices with real numbers as elements fails on two counts: Multiplication is not commutative, and it has zero divisors. ■

In Example 4 of Section 5.1, we saw that \mathbf{Z}_n is a ring for every value of $n > 1$. Moreover, \mathbf{Z}_n is a commutative ring since

$$[a] \cdot [b] = [ab] = [ba] = [b] \cdot [a]$$

for all $[a], [b]$ in \mathbf{Z}_n . Since \mathbf{Z}_n has $[1]$ as the unity, \mathbf{Z}_n is an integral domain if and only if it has no zero divisors. The following theorem characterizes these \mathbf{Z}_n , and it provides us with a large class of *finite integral domains* (that is, integral domains that have a finite number of elements).

Theorem 5.15 ■ The Integral Domain \mathbf{Z}_n When n Is a Prime

For $n > 1$, \mathbf{Z}_n is an integral domain if and only if n is a prime.

Proof From the previous discussion, it is clear that we need only prove that \mathbf{Z}_n has no zero divisors if and only if n is a prime.

Suppose first that n is a prime. Let $[a] \neq [0]$ in \mathbf{Z}_n , and suppose $[a][b] = [0]$ for some $[b]$ in \mathbf{Z}_n . Now $[a][b] = [0]$ implies that $[ab] = [0]$, and therefore, $n|ab$. However, $[a] \neq [0]$ means that $n \nmid a$. Thus $n|ab$ and $n \nmid a$. Since n is a prime, this implies that $n|b$, by Theorem 2.16;

$\sim p \Leftarrow \sim q$

that is, $[b] = [0]$. We have shown that if $[a] \neq [0]$, the only way that $[a][b]$ can be $[0]$ is for $[b]$ to be $[0]$. Therefore, \mathbf{Z}_n has no zero divisors and is an integral domain.

Suppose now that n is not a prime. Then n has divisors other than ± 1 and $\pm n$, so there are integers a and b such that

$$n = ab \quad \text{where } 1 < a < n \text{ and } 1 < b < n.$$

This means that $[a] \neq [0]$, $[b] \neq [0]$, but

$$[a][b] = [ab] = [n] = [0].$$

Therefore, $[a]$ is a zero divisor in \mathbf{Z}_n , and \mathbf{Z}_n is not an integral domain.

Combining the two cases, we see that n is a prime if and only if \mathbf{Z}_n is an integral domain.

One direct consequence of the absence of zero divisors in an integral domain is that the cancellation law for multiplication must hold.

Theorem 5.16 ■ Cancellation Law for Multiplication

If a , b , and c are elements of an integral domain D such that $a \neq 0$ and $ab = ac$, then $b = c$.

$(p \wedge q) \Rightarrow r$ **Proof** Suppose a , b , and c are elements of an integral domain D such that $a \neq 0$ and $ab = ac$. Now

$$\begin{aligned} ab = ac &\Rightarrow ab - ac = 0 \\ &\Rightarrow a(b - c) = 0. \end{aligned}$$

Since $a \neq 0$ and D has no zero divisors, it must be true that $b - c = 0$, and hence $b = c$.

It can be shown that if the cancellation law holds in a commutative ring, then the ring cannot have zero divisors. The proof of this is left as an exercise.

To require that a ring has no zero divisors is equivalent to requiring that a product of nonzero elements must always be different from 0. Or, stated another way, a product that is 0 must have at least one factor equal to 0.

A *field* is another special type of ring, and we shall examine the relationship between a field and an integral domain. We begin with a definition.

Definition 5.17 ■ Field

Let F be a ring. Then F is a **field** provided these conditions hold:

1. F is a commutative ring.
 2. F has a unity e , and $e \neq 0$.
 3. Every nonzero element of F has a multiplicative inverse.
-

The rational numbers, the real numbers, and the complex numbers are familiar examples of fields. We shall see in Corollary 5.20 that if p is a prime, then \mathbf{Z}_p is a field. Other and less familiar examples of fields are found in the exercises for this section.

Part of the relation between fields and integral domains is stated in the following theorem.

Theorem 5.18 ■ Fields and Integral Domains

Every field is an integral domain.

$p \Rightarrow q$ **Proof** Let F be a field. To prove that F is an integral domain, we need only show that F has no zero divisors. Suppose a and b are elements of F such that $ab = 0$. If $a \neq 0$, then $a^{-1} \in F$ and

$$\begin{aligned} ab = 0 &\Rightarrow a^{-1}(ab) = a^{-1} \cdot 0 \\ &\Rightarrow (a^{-1}a)b = 0 \\ &\Rightarrow eb = 0 \\ &\Rightarrow b = 0. \end{aligned}$$

Similarly, if $b \neq 0$, then $a = 0$. Therefore, F has no zero divisors and is an integral domain.

It is certainly not true that every integral domain is a field. For example, the set \mathbf{Z} of all integers forms an integral domain, and the integers 1 and -1 are the only elements of \mathbf{Z} that have multiplicative inverses. It is perhaps surprising, but an integral domain with a finite number of elements is always a field. This is the other part of the relationship between a field and an integral domain.

Theorem 5.19 ■ Finite Integral Domains and Fields

Every finite integral domain is a field.

$p \Rightarrow q$ **Proof** Assume that D is a finite integral domain. Let n be the number of distinct elements in D ; say,

$$D = \{d_1, d_2, \dots, d_n\},$$

where the d_i are the distinct elements of D . Now let a be any nonzero element of D , and consider the set of products

$$\{ad_1, ad_2, \dots, ad_n\}.$$

These products are all distinct, for $a \neq 0$ and $ad_r = ad_s$ would imply $d_r = d_s$, by Theorem 5.16, and the d_i are all distinct. These n products are all contained in D , and no two of them are equal. Hence they are the same as the elements of D , except possibly for order. This means that every element of D appears somewhere in the list

$$ad_1, ad_2, \dots, ad_n.$$

In particular, the unity e is one of these products. That is, $ad_k = e$ for some d_k . Since multiplication is commutative in D , we have $d_k a = ad_k = e$, and d_k is a multiplicative inverse of a . Thus D is a field.

Corollary 5.20 ■ The Field \mathbf{Z}_n When n Is a Prime

\mathbf{Z}_n is a field if and only if n is a prime.

Proof This follows at once from Theorems 5.15, 5.18, and 5.19.

We have seen that the elements of a ring form an abelian group with respect to addition. A similar comparison can be made for the nonzero elements of a field. It is readily seen that the nonzero elements form an abelian group with respect to multiplication. The definition of a field can thus be reformulated as follows: A **field** is a set of elements in which equality, addition, and multiplication are defined such that the following conditions hold.

1. F forms an abelian group with respect to addition.
2. The nonzero elements of F form an abelian group with respect to multiplication.
3. The distributive law $x(y + z) = xy + xz$ holds for all x, y, z in F .

The last example in this section points out that some of our most familiar rings do not form integral domains.

Example 3 For $n \geq 2$, each of the rings

$$M_n(\mathbf{Z}), \quad M_n(\mathbf{Q}), \quad M_n(\mathbf{R}), \quad M_n(\mathbf{C})$$

is not an integral domain, since multiplication in each of them is not commutative. It is also true that each of them contains zero divisors if $n \geq 2$. For $n = 2$, the product

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

illustrates this statement. Similar examples can easily be constructed for $n > 2$. ■

Exercises 5.2

True or False

Label each of the following statements as either true or false.

1. An integral domain contains at least 2 elements.
 2. Every field is an integral domain.
 3. Every integral domain is a field.
 4. If a set S is not an integral domain, then S is not a field.
-

Exercises

1. Decide which of the following are integral domains and which are fields with respect to the usual operations of addition and multiplication. For each one that fails to be an integral domain or a field, state a reason.
 - a. the set of all real numbers of the form $m + n\sqrt{2}$, where m and n are integers
 - b. the set of all real numbers of the form $a + b\sqrt{2}$, where a and b are rational numbers
 - c. the set of all real numbers of the form $a + b\sqrt[3]{2}$, where a and b are rational numbers
 - d. the set of all real numbers of the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$, where a, b , and c are rational numbers
 - e. the **Gaussian integers**—that is, the set of all complex numbers of the form $m + ni$, where $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$
 - f. the set of all complex numbers of the form $m + ni$, where $m \in \mathbf{E}$ and $n \in \mathbf{E}$ (\mathbf{E} is the ring of all even integers.)
 - g. the set of all complex numbers of the form $a + bi$, where a and b are rational numbers
 - h. the set of all real numbers of the form $m + n\sqrt{2}$, where $m \in \mathbf{Z}$ and $n \in \mathbf{E}$
2. Consider the set $R = \{[0], [2], [4], [6], [8]\} \subseteq \mathbf{Z}_{10}$, with addition and multiplication as defined in \mathbf{Z}_{10} .
 - a. Is R an integral domain? If not, give a reason.
 - b. Is R a field? If not, give a reason.
3. Consider the set $S = \{[0], [2], [4], [6], [8], [10], [12], [14], [16]\} \subseteq \mathbf{Z}_{18}$, with addition and multiplication as defined in \mathbf{Z}_{18} .
 - a. Is S an integral domain? If not, give a reason.
 - b. Is S a field? If not, give a reason.

Examples 5 and 6 of Section 5.1 showed that $\mathcal{P}(U)$ is a commutative ring with unity. In Exercises 4 and 5, let $U = \{a, b\}$.

- Sec. 5.1, #3 ➤ 4. Is $\mathcal{P}(U)$ an integral domain? If not, find all zero divisors in $\mathcal{P}(U)$.
- Sec. 5.1, #3 ➤ 5. Is $\mathcal{P}(U)$ a field? If not, find all nonzero elements that do not have multiplicative inverses.
6. Let $S = \{(0, 0), (1, 1), (0, 1), (1, 0)\}$, where $0 = [0]$ and $1 = [1]$ are the elements of \mathbf{Z}_2 . Equality, addition, and multiplication are defined in S as follows:

$$(a, b) = (c, d) \quad \text{if and only if } a = c \text{ and } b = d \text{ in } \mathbf{Z}_2,$$

$$(a, b) + (c, d) = (a + c, b + d),$$

$$(a, b) \cdot (c, d) = (ad + bc + bd, ad + bc + ac).$$

- a. Prove that multiplication in S is associative.

Assume that S is a ring and consider these questions, giving a reason for any negative answers.

- b. Is S a commutative ring?
- c. Does S have a unity?

Sec. 5.1, #1a ➤
Sec. 5.3, #16 <=

Sec. 5.1, #1b ➤
Sec. 5.1, #1c ➤
Sec. 5.1, #1d ➤
Sec. 5.1, #1e ➤
Sec. 5.1, #1f ➤
Sec. 5.3, #15 <=

Sec. 5.1, #1h ➤

Sec. 5.1, #3 ➤

Sec. 5.1, #3 ➤

- d.** Is S an integral domain?
e. Is S a field?
7. Let W be the set of all ordered pairs (x, y) of integers x and y . Equality, addition, and multiplication are defined as follows:

$$(x, y) = (z, w) \quad \text{if and only if } x = z \text{ and } y = w \text{ in } \mathbf{Z},$$

$$(x, y) + (z, w) = (x + z, y + w),$$

$$(x, y) \cdot (z, w) = (xz - yw, xw + yz).$$

Given that W is a ring, determine whether W is commutative and whether W has a unity. Justify your decisions.

- Sec. 5.3, #9 <
8. Let S be the set of all 2×2 matrices of the form $\begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix}$, where x is a real number. Assume that S is a ring with respect to matrix addition and multiplication. Answer the following questions, and give a reason for any negative answers.
- a.** Is S a commutative ring?
 - b.** Does S have a unity? If so, identify the unity.
 - c.** Is S an integral domain?
 - d.** Is S a field?
9. Work Exercise 8 using S as the set of all 2×2 matrices of the form $\begin{bmatrix} x & x \\ 0 & 0 \end{bmatrix}$, where x is a real number.

- Sec. 1.6, #23 > Sec. 5.3, #9 <
10. Let R be the set of all matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, where a and b are integers. Assume that R is a ring with respect to matrix addition and multiplication. Determine whether R is commutative, and identify the unity if R has one.

11. Let R be the set of all matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, where a and b are real numbers. Assume that R is a ring with respect to matrix addition and multiplication. Answer the following questions and give a reason for any negative answers.
- a.** Is R a commutative ring?
 - b.** Does R have a unity? If so, identify the unity.
 - c.** Is R an integral domain?
 - d.** Is R a field?

Sec. 5.3, #10 <

12. Consider the set $S = \{a + bi \mid a, b \in \mathbf{Z}_3\} = \{0, 1, 2, i, 1+i, 2+i, 2i, 1+2i, 2+2i\}$, where we write 0 for [0], 1 for [1], and 2 for [2] in \mathbf{Z}_3 . Addition and multiplication are as in the complex numbers except that the coefficients are added and multiplied as in \mathbf{Z}_3 . Thus $i^2 = -1$ as in the complex numbers and $-1 = 2$ in \mathbf{Z}_3 .
- a.** Is S a commutative ring?
 - b.** Does S have a unity?
 - c.** Is S an integral domain?
 - d.** Is S a field?

13. Work Exercise 12 using $S = \{a + bi \mid a, b \in \mathbf{Z}_5\}$.
14. Let R be a commutative ring with unity in which the cancellation law for multiplication holds. That is, if a , b , and c are elements of R , then $a \neq 0$ and $ab = ac$ always imply $b = c$. Prove that R is an integral domain.
15. Prove or disprove that every commutative ring with no zero divisors is an integral domain.
16. Prove that if a subring R of an integral domain D contains the unity element of D , then R is an integral domain.
17. If e is the unity in an integral domain D , prove that $(-e)a = -a$ for all $a \in D$.
18. Prove that the only idempotent elements in an integral domain are 0 and e .
19. a. Give an example where a and b are not zero divisors in a ring R , but the sum $a + b$ is a zero divisor.
b. Give an example where a and b are zero divisors in a ring R with $a + b \neq 0$, and $a + b$ is not a zero divisor.
c. Prove that the set of all elements in a ring R that are not zero divisors is closed under multiplication.

20. Find the multiplicative inverse of the given element. (See Example 4 of Section 2.6.)

- a. [11] in \mathbf{Z}_{317} b. [11] in \mathbf{Z}_{138} c. [9] in \mathbf{Z}_{242} d. [6] in \mathbf{Z}_{319}

Sec. 5.1, #47 ➤
21. Prove that if R and S are integral domains, then the direct sum $R \oplus S$ is *not* an integral domain.

Sec. 5.1, #50 ➤
22. Let R be a Boolean ring with unity e . Prove that every element of R except 0 and e is a zero divisor.
23. If $a \neq 0$ in a field F , prove that for every $b \in F$ the equation $ax = b$ has a unique solution x in F .
24. Suppose S is a subset of a field F that contains at least two elements and satisfies both of the following conditions: $x \in S$ and $y \in S$ imply $x - y \in S$, and $x \in S$ and $y \neq 0 \in S$ imply $xy^{-1} \in S$. Prove that S is a field.

5.3

The Field of Quotients of an Integral Domain

The example of an integral domain that is most familiar to us is the set \mathbf{Z} of all integers, and the most familiar example of a field is the set of all rational numbers. There is a very natural and intimate relationship between these two systems. In fact, a rational number is by definition a quotient a/b of integers a and b , with $b \neq 0$; that is, the set of rational numbers is the set of all quotients of integers with nonzero denominators. For this reason, the set of rational numbers is frequently referred to as “the quotient field of the integers.” In this section, we shall see that an analogous field of quotients can be constructed for an arbitrary integral domain.

Before we present this construction, let us review the basic definitions of equality, addition, and multiplication in the rational numbers. We recall that for rational numbers $\frac{a}{b}$ and $\frac{c}{d}$,

$$\frac{a}{b} = \frac{c}{d} \quad \text{if and only if } ad = bc,$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd},$$

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

Note that the definitions of equality, addition, and multiplication for rational numbers are based on the corresponding definitions for the integers. These definitions guide our construction of the quotient field for an arbitrary integral domain D .

Our first step in this construction is the following definition.

Definition 5.21 ■ A Relation on Ordered Pairs

Let D be an integral domain and let S be the set of all ordered pairs (a, b) of elements of D with $b \neq 0$:

$$S = \{(a, b) \mid a, b \in D \text{ and } b \neq 0\}.$$

The relation \sim is defined on S by

$$(a, b) \sim (c, d) \quad \text{if and only if } ad = bc.$$

The relation \sim is an obvious imitation of the equality of rational numbers, and we can show that it is indeed an equivalence relation on S .

Lemma 5.22 ■ The Equivalence Relation \sim

The relation \sim in Definition 5.21 is an equivalence relation on S .

Proof We shall show that \sim is reflexive, symmetric, and transitive. Let (a, b) , (c, d) , and (f, g) be arbitrary elements of S .

Reflexive 1. $(a, b) \sim (a, b)$, since the commutative multiplication in D implies that $ab = ba$.

Symmetric 2. $(a, b) \sim (c, d) \Rightarrow ad = bc$ by definition of \sim
 $\Rightarrow da = cb$ or $cb = da$ since multiplication is commutative in D
 $\Rightarrow (c, d) \sim (a, b)$ by definition of \sim .

Transitive 3. Assume that $(a, b) \sim (c, d)$ and $(c, d) \sim (f, g)$.

$$\left. \begin{aligned} (a, b) \sim (c, d) &\Rightarrow ad = bc \Rightarrow adg = bcd \\ (c, d) \sim (f, g) &\Rightarrow cg = df \Rightarrow bcd = bdf \end{aligned} \right\} \Rightarrow adg = bdf$$

Using the commutative property of multiplication in D once again, we have[†]

$$dag = dbf$$

where $d \neq 0$, and therefore

$$ag = bf$$

by Theorem 5.16. According to Definition 5.21, this implies that $(a, b) \sim (f, g)$. Thus \sim is an equivalence relation on S .

The next definition reveals the basic plan for our construction of the quotient field of D .

Definition 5.23 ■ The Set of Quotients

Let D , S , and \sim be the same as in Definition 5.21 and Lemma 5.22. For each (a, b) in S , let $[a, b]$ denote the equivalence class in S that contains (a, b) , and let Q denote the set of all equivalence classes $[a, b]$:

$$Q = \{[a, b] \mid (a, b) \in S\}.$$

The set Q is called the **set of quotients** for D .

We shall at times need the fact that for any $x \neq 0$ in D and any $[a, b]$ in Q ,

$$[a, b] = [ax, bx].$$

This follows at once from the equality $a(bx) = b(ax)$ in the integral domain D .

Lemma 5.24 ■ Addition and Multiplication in Q

The following rules define binary operations on Q . **Addition** in Q is defined by

$$[a, b] + [c, d] = [ad + bc, bd],$$

and **multiplication** in Q is defined by

$$[a, b] \cdot [c, d] = [ac, bd].$$

Proof We shall verify that the rule stated for addition defines a binary operation on Q . For arbitrary $[a, b]$ and $[c, d]$ in Q , we have $b \neq 0$ and $d \neq 0$ in D . Since D is an integral domain, $b \neq 0$ and $d \neq 0$ imply $bd \neq 0$, so $[a, b] + [c, d] = [ad + bc, bd]$ is an element of Q .

To show that the sum of two elements is unique (or well-defined), suppose that $[a, b] = [x, y]$ and $[c, d] = [z, w]$ in Q . We need to show $[a, b] + [c, d] = [x, y] + [z, w]$. Now

$$[a, b] + [c, d] = [ad + bc, bd]$$

and

$$[x, y] + [z, w] = [xw + yz, yw].$$

[†]It is tempting here to use $ad = bc$ and $cg = df$ to obtain $(ad)(cg) = (bc)(df)$, but this would not imply that $ag = bf$, because c might be zero.

To prove these elements equal, we need

$$(ad + bc)yw = bd(xw + yz)$$

or

$$adyw + bcyw = bdxw + bdyz.$$

We have

$$\begin{aligned} [a, b] = [x, y] &\Rightarrow ay = bx \\ &\Rightarrow (ay)(dw) = (bx)(dw) \\ &\Rightarrow adyw = bdxw \end{aligned}$$

and

$$\begin{aligned} [c, d] = [z, w] &\Rightarrow cw = dz \\ &\Rightarrow (cw)(by) = (dz)(by) \\ &\Rightarrow bcyw = bdyz. \end{aligned}$$

By adding corresponding sides of equations, we obtain

$$adyw + bcyw = bdxw + bdyz.$$

Thus $[a, b] + [c, d] = [x, y] + [z, w]$.

It can be similarly shown that multiplication as defined by the given rule is a binary operation on Q .

It is important to note that the set of all ordered pairs of the form $(0, x)$, where $x \neq 0$, forms a complete equivalence class that can be written as $[0, b]$ for any nonzero element b of D . With these preliminaries out of the way, we can now state our theorem.

Theorem 5.25 ■ The Quotient Field

Let D be an integral domain. The set Q as given in Definition 5.23 is a field, called the **quotient field** of D with respect to the operations defined in Lemma 5.24.

Proof We first consider the postulates for addition. It is left as an exercise to prove that addition is associative. The zero element of Q is the class $[0, b]$, since

$$[x, y] + [0, b] = [x \cdot b + y \cdot 0, y \cdot b] = [xb, yb] = [x, y],$$

and similar steps show that

$$[0, b] + [x, y] = [x, y].$$

The equality $[xb, yb] = [x, y]$ follows from the fact that $b \neq 0$, as was pointed out just after Definition 5.23. Routine calculations show that $[-a, b]$ is the additive inverse of $[a, b]$ in Q and that addition in Q is commutative. The verification of the associative property for multiplication is left as an exercise.

We shall verify the left distributive property and leave the other as an exercise. Let $[x, y]$, $[z, w]$, and $[u, v]$ denote arbitrary elements of \mathcal{Q} . We have

$$\begin{aligned}[x, y] \cdot ([z, w] + [u, v]) &= [x, y][zv + wu, wv] \\ &= [xzv + xwu, ywv]\end{aligned}$$

and

$$\begin{aligned}[x, y] \cdot [z, w] + [x, y] \cdot [u, v] &= [xz, yw] + [xu, yv] \\ &= [xyzv + xywu, y^2wv] \\ &= [y(xzv + xwu), y(ywv)].\end{aligned}$$

Comparing the results of these two calculations, we see that the last one differs from the first only in that both elements in the pair have been multiplied by y . Since $[x, y]$ in \mathcal{Q} requires $y \neq 0$, these results are equal.

Since multiplication in D is commutative, we have

$$\begin{aligned}[a, b] \cdot [c, d] &= [ac, bd] \\ &= [ca, db] \\ &= [c, d] \cdot [a, b].\end{aligned}$$

Thus \mathcal{Q} is a commutative ring.

Let $b \neq 0$ in D , and consider the element $[b, b]$ in \mathcal{Q} . For any $[x, y]$ in \mathcal{Q} we have

$$\begin{aligned}[x, y] \cdot [b, b] &= [xb, yb] \\ &= [x, y],\end{aligned}$$

so $[b, b]$ is a right identity for multiplication. Since multiplication is commutative, $[b, b]$ is a nonzero unity for \mathcal{Q} .

We have seen that the zero element of \mathcal{Q} is the class $[0, b]$. Thus any nonzero element has the form $[c, d]$, with both c and d nonzero. But then $[d, c]$ is also in \mathcal{Q} , and

$$\begin{aligned}[c, d] \cdot [d, c] &= [cd, dc] \\ &= [d, d],\end{aligned}$$

so $[d, c]$ is the multiplicative inverse of $[c, d]$ in \mathcal{Q} . This completes the proof that \mathcal{Q} is a field.

Note that in the proof of Theorem 5.25, the unity e in D did not appear explicitly anywhere. In fact, the construction yields a field if we start with a commutative ring that has no zero divisors instead of with an integral domain. However, we make use of the unity of D in Theorem 5.27.

The concept of an isomorphism can be applied to rings as well as to groups. The definition is a very natural extension of the concept of a group isomorphism. Since there are two binary operations involved in the definition of a ring, we simply require that both operations be preserved.

Definition 5.26 ■ Ring Isomorphism

Let R and R' denote two rings. A mapping $\phi: R \rightarrow R'$ is a **ring isomorphism** from R to R' provided the following conditions hold:

1. ϕ is a one-to-one correspondence from R to R' .
2. $\phi(x + y) = \phi(x) + \phi(y)$ for all x and y in R .
3. $\phi(x \cdot y) = \phi(x) \cdot \phi(y)$ for all x and y in R .

If an isomorphism from R to R' exists, we say that R is **isomorphic** to R' .

Of course, the term *ring isomorphism* may be applied to systems that are more than a ring; that is, there may be a ring isomorphism that involves integral domains or fields. The relation of being isomorphic is reflexive, symmetric, and transitive on rings, just as it was with groups.

The field of quotients Q of an integral domain D has a significant feature that has not yet been brought to light. In the sense of isomorphism, it contains the integral domain D . More precisely, Q contains a subring D' that is isomorphic to D .

Theorem 5.27 ■ Subring of Q Isomorphic to D

Let D and Q be as given in Definition 5.23, and let e denote the unity of D . The set D' that consists of all elements of Q that have the form $[x, e]$ is a subring of Q , and D is isomorphic to D' .

Proof Referring to Definition 5.1a, we see that conditions 2, 5, 7, and 8 are automatically satisfied in D' , and we need only check conditions 1, 3, 4, and 6.

For arbitrary $[x, e]$ and $[y, e]$ in D' , we have

$$\begin{aligned}[x, e] + [y, e] &= [x \cdot e + y \cdot e, e \cdot e] \\ &= [x + y, e],\end{aligned}$$

and D' is closed under addition. The element $[0, e]$ is in D' , so D' contains the zero element of Q . For $[x, e]$ in D' , the additive inverse is $[-x, e]$, an element of D' . Finally, the calculation

$$[x, e] \cdot [y, e] = [xy, e]$$

shows that D' is closed under multiplication. Thus D' is a subring of Q .

To prove that D is isomorphic to D' , we use the natural mapping $\phi: D \rightarrow D'$ defined by

$$\phi(x) = [x, e].$$

The mapping ϕ is obviously a one-to-one correspondence. Since

$$\begin{aligned}\phi(x + y) &= [x + y, e] \\ &= [x, e] + [y, e] \\ &= \phi(x) + \phi(y)\end{aligned}$$

and

$$\begin{aligned}\phi(x \cdot y) &= [xy, e] \\ &= [x, e] \cdot [y, e] \\ &= \phi(x) \cdot \phi(y),\end{aligned}$$

ϕ is a ring isomorphism from D to D' .

Thus the quotient field Q contains D in the sense of isomorphism. We say that D is **embedded** in Q or that Q is an **extension** of D . More generally, if S is a ring that contains a subring R' that is isomorphic to a given ring R , we say that R is **embedded** in S or that S is an **extension** of R .

There is one more observation about Q that should be made. For any nonzero $[b, e]$ in D' , the multiplicative inverse of $[b, e]$ in Q is $[b, e]^{-1} = [e, b]$, and every element of Q can be written in the form

$$[a, b] = [a, e] \cdot [e, b] = [a, e] \cdot [b, e]^{-1}.$$

If the isomorphism ϕ in the proof of Theorem 5.27 is used to identify x in D with $[x, e]$ in D' , then every element of Q can be identified as a quotient ab^{-1} of elements a and b of D , with $b \neq 0$.

From this, it follows that any field F that contains the integral domain D must also contain Q because F must contain b^{-1} for each $b \neq 0$ in D and must also contain the product ab^{-1} for all $a \in D$. Thus Q is the smallest field that “contains” D .

If the construction presented in this section is carried out beginning with $D = \mathbf{Z}$, the field \mathbf{Q} of rational numbers is obtained, with the elements written as $[a, b]$ instead of a/b . The isomorphism ϕ in the proof of Theorem 5.27 maps an integer x onto $[x, 1]$, which is playing the role of $x/1$ in the notation, and we end up with the integers embedded in the rational numbers. The construction of the rational numbers from the integers is in this way a special case of the procedure described here.

Exercises 5.3

True or False

Label each of the following statements as either true or false.

1. The field \mathbf{Q} of rational numbers is an extension of the integral domain \mathbf{Z} of integers.
 2. The field \mathbf{R} of real numbers is an extension of the integral domain \mathbf{Z} of integers.
 3. The field of quotients Q of an integral domain D contains D .
 4. The field of quotients Q of an integral domain D contains a subring $D' = \{[x, e] \mid x \in D, e \text{ is the unity in } D\}$.
 5. A field of quotients can be constructed from an arbitrary integral domain.
-

Exercises

1. Prove that the multiplication defined in Lemma 5.24 is a binary operation on Q .
2. Prove that addition is associative in Q .
3. Show that $[-a, b]$ is the additive inverse of $[a, b]$ in Q .
4. Prove that addition is commutative in Q .
5. Prove that multiplication is associative in Q .
6. Prove the right distributive property in Q :

$$([x, y] + [z, w]) \cdot [u, v] = [x, y] \cdot [u, v] + [z, w] \cdot [u, v].$$

7. Prove that on a given set of rings, the relation of being isomorphic has the reflexive, symmetric, and transitive properties.
8. Assume that the ring R is isomorphic to the ring R' . Prove that if R is commutative, then R' is commutative.

Sec. 5.2, #7, 10 ➤

9. Let W be the ring in Exercise 7 of Section 5.2, and let R be the ring in Exercise 10 of the same section. Given that W and R are isomorphic rings, define an isomorphism from W to R and prove that your mapping is an isomorphism.

Sec. 5.2, #11 ➤

10. Assume that the set R in Exercise 11 of Section 5.2 is a field, and let \mathbf{C} be the field of all complex numbers $a + bi$, where a and b are real numbers and $i^2 = -1$. Given that R and \mathbf{C} are isomorphic fields, define an isomorphism from \mathbf{C} to R and prove that your mapping is an isomorphism.
11. Since this section presents a method for constructing a field of quotients for an arbitrary integral domain D , we might ask what happens if D is already a field. As an example, consider the situation when $D = \mathbf{Z}_3$.
 - a. With $D = \mathbf{Z}_3$, write out all the elements of S , sort these elements according to the relation \sim , and then list all the distinct elements of Q .
 - b. Exhibit an isomorphism from D to Q .

12. Work Exercise 11 with $D = \mathbf{Z}_5$.

13. Prove that if D is a field to begin with, then the field of quotients Q is isomorphic to D .
14. Just after the end of the proof of Theorem 5.25, we noted that the construction in the proof yields a field if we start with a commutative ring that has no zero divisors. Assume this is true, and let F denote the field of quotients of the ring \mathbf{E} of all even integers. Prove that F is isomorphic to the field of rational numbers.

Sec. 5.2, #1e ➤

15. Let D be the set of all complex numbers of the form $m + ni$, where $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$. Carry out the construction of the quotient field Q for this integral domain, and show that this quotient field is isomorphic to the set of all complex numbers of the form $a + bi$, where a and b are rational numbers.

Sec. 5.2, #1a ➤

16. Let D be the set of all real numbers of the form $m + n\sqrt{2}$, where $m, n \in \mathbf{Z}$. Carry out the construction of the quotient field Q for this integral domain, and show that this

quotient field is isomorphic to the set of real numbers of the form $a + b\sqrt{2}$ where a and b are rational numbers.

17. Prove that any field that contains an integral domain D must contain a subfield isomorphic to the quotient field Q of D .
18. Assume R is a ring, and let S be the set of all ordered pairs (m, x) where $m \in \mathbf{Z}$ and $x \in R$. Equality in S is defined by

$$(m, x) = (n, y) \text{ if and only if } m = n \text{ and } x = y.$$

Addition and multiplication in S are defined by

$$(m, x) + (n, y) = (m + n, x + y)$$

and

$$(m, x) \cdot (n, y) = (mn, my + nx + xy),$$

where my and nx are *multiples* of y and x in the ring R .

- a. Prove that S is a ring with unity.
- b. Prove that $\phi: R \rightarrow S$ defined by $\phi(x) = (0, x)$ is an isomorphism from R to a subring R' of S . This result shows that any ring can be embedded in a ring that has a unity.
19. Let T be the smallest subring of the field \mathbf{Q} of rational numbers that contains $\frac{1}{2}$. Find a description for a typical element of T .

5.4

Ordered Integral Domains

In Section 2.1 we assumed that the set \mathbf{Z} of all integers satisfied a list of five postulates. The last two of these postulates led to the introduction of the order relation “greater than” in \mathbf{Z} , and to the proof of the Well-Ordering Theorem (Theorem 2.7). In this section, we follow a development along similar lines in a more general setting.

Definition 5.28 ■ Ordered Integral Domain

An integral domain D is an **ordered integral domain** if D contains a subset D^+ that has the following properties:

1. D^+ is closed under addition.
2. D^+ is closed under multiplication.
3. For each $x \in D$, one and only one of the following statements is true:

$$x \in D^+, \quad x = 0, \quad -x \in D^+.$$

Such a subset D^+ is called a **set of positive elements** for D .

Analogous to the situation in \mathbf{Z} , condition 3 in Definition 5.28 is referred to as the **law of trichotomy**, and an element $x \in D$ such that $-x \in D^+$ is called a **negative element** of D .

Example 1 The integral domain \mathbf{Z} is, of course, an example of an ordered integral domain. With their usual sets of positive elements, the set of all rational numbers and the set of all real numbers furnish two other examples of ordered integral domains. ■

Later, we shall see that not all integral domains are ordered integral domains.

Following the same sort of procedure that we followed with the integers, we can use the set of positive elements in an ordered integral domain D to define the order relation “greater than” in D .

Definition 5.29 ■ Greater Than

Let D be an ordered integral domain with D^+ as the set of positive elements. The relation **greater than**, denoted by $>$, is defined on elements x and y of D by

$$x > y \quad \text{if and only if} \quad x - y \in D^+.$$

The symbol $>$ is read “greater than.” Similarly, $<$ is read “is less than.” We define $x < y$ if and only if $y > x$. As direct consequences of the definition, we have

$$x > 0 \quad \text{if and only if} \quad x \in D^+$$

and

$$x < 0 \quad \text{if and only if} \quad -x \in D^+.$$

The three properties of D^+ in Definition 5.28 translate at once into the following properties of $>$ in D .

1. If $x > 0$ and $y > 0$, then $x + y > 0$.
2. If $x > 0$ and $y > 0$, then $xy > 0$.
3. For each $x \in D$, one and only one of the following statements is true:

$$x > 0, \quad x = 0, \quad x < 0.$$

The other basic properties of $>$ are stated in the next theorem. We prove the first two and leave the proofs of the others as exercises.

Theorem 5.30 ■ Properties of $>$

Suppose that D is an ordered integral domain. The relation $>$ has the following properties, where x , y , and z are arbitrary elements of D .

- a. If $x > y$, then $x + z > y + z$.
- b. If $x > y$ and $z > 0$, then $xz > yz$.
- c. If $x > y$ and $y > z$, then $x > z$.
- d. One and only one of the following statements is true:

$$x > y, \quad x = y, \quad x < y.$$

$p \Rightarrow q$ **Proof of a** If $x > y$, then $x - y \in D^+$, by Definition 5.29. Since

$$\begin{aligned}(x + z) - (y + z) &= x + z - y - z \\ &= x - y,\end{aligned}$$

this means that $(x + z) - (y + z) \in D^+$, and therefore $x + z > y + z$.

$(p \wedge q) \Rightarrow r$ **Proof of b** Suppose $x > y$ and $z > 0$. Then $x - y \in D^+$ and $z \in D^+$. Condition 2 of Definition 5.28 requires that D^+ be closed under multiplication, so the product $(x - y)z$ must be in D^+ . Since $(x - y)z = xz - yz$, we have $xz - yz \in D^+$, and therefore $xz > yz$.

Our main goal in this section is to characterize the integers as an ordered integral domain that has a certain type of set of positive elements. As a first step in this direction, we prove the following simple theorem, which may be compared to Theorem 2.5.

Theorem 5.31 ■ Square of a Nonzero Element

For any $x \neq 0$ in an ordered integral domain D , $x^2 \in D^+$.

$p \Rightarrow q$ **Proof** Suppose $x \neq 0$ in D . By condition 3 of Definition 5.28, either $x \in D^+$ or $-x \in D^+$. If $x \in D^+$, then $x^2 = x \cdot x$ is in D^+ since D^+ is closed under multiplication. If $-x \in D^+$, then $x^2 = x \cdot x = (-x)(-x)$ is in D^+ , again by closure of D^+ under multiplication. In either case, we have $x^2 \in D^+$.

Corollary 5.32 ■ The Unity Element

In any ordered integral domain, $e \in D^+$.

Proof This follows from the fact that $e = e^2$.

The preceding theorem and its corollary can be used to show that the set \mathbf{C} of all complex numbers does not form an ordered integral domain. Suppose, to the contrary, that \mathbf{C} does contain a set \mathbf{C}^+ of positive elements. By Corollary 5.32, $1 \in \mathbf{C}^+$, and therefore $-1 \notin \mathbf{C}^+$ by the law of trichotomy. Theorem 5.31 requires, however, that $i^2 = -1$ be in \mathbf{C}^+ , and we have a contradiction. Therefore, \mathbf{C} does not contain a set of positive elements. In other words, *it is impossible to impose an order relation on the set of complex numbers*.

In the next definition, we use the symbol \leq with its usual meaning. Similarly, we later use the symbol \geq with its usual meaning and without formal definition.

Definition 5.33 ■ Well-Ordered Subset

A nonempty subset S of an ordered integral domain D is **well-ordered** if for every nonempty subset T of S , there is an element $m \in T$ such that $m \leq x$ for all $x \in T$. Such an element m is called a **least element** of T .

Thus $S \neq \emptyset$ in D is well-ordered if every nonempty subset of S contains a least element. We proved in Theorem 2.7 that the set of all positive integers is well-ordered.

The next step toward our characterization of the integers is the following theorem.

Theorem 5.34 ■ Well-Ordered D^+

If D is an ordered integral domain in which the set D^+ of positive elements is well-ordered, then

- a. e is the least element of D^+ , and
- b. $D^+ = \{ne \mid n \in \mathbf{Z}^+\}$.

p \Rightarrow q **Proof** We have $e \in D^+$ by Corollary 5.32. To prove that e is the least element of D^+ , let T be the set of all $x \in D^+$ such that $e > x > 0$, and assume that T is nonempty. Since D^+ is well-ordered, T has a least element m , and

$$e > m > 0.$$

Using Theorem 5.30b and multiplying by m , we have

$$m \cdot e > m^2 > m \cdot 0.$$

That is,

$$m > m^2 > 0,$$

and this contradicts the choice of m as the least element of T . Therefore, T is empty and e is the least element of D^+ .

p \Rightarrow r Now let S be the set of all $n \in \mathbf{Z}^+$ such that $ne \in D^+$. We have $1 \in S$ since $1e = e \in D^+$. Assume that $k \in S$. Then $ke \in D^+$, and this implies that

$$(k + 1)e = ke + e$$

is in S , since D^+ is closed under addition. Thus $k \in S$ implies $k + 1 \in S$, and $S = \mathbf{Z}^+$ by the induction postulate for the positive integers. This proves that

$$D^+ \supseteq \{ne \mid n \in \mathbf{Z}^+\}.$$

In order to prove that $D^+ \subseteq \{ne \mid n \in \mathbf{Z}^+\}$, let L be the set of all elements of D^+ that are not of the form ne with $ne \in \mathbf{Z}^+$, and suppose that L is nonempty. Since D^+ is well-ordered, L has a least element ℓ . It must be true that

$$\ell > e,$$

since e is the least element of D^+ , and therefore $\ell - e > 0$. Now

$$\begin{aligned} e > 0 &\Rightarrow e + (-e) > 0 + (-e) && \text{by Theorem 5.30a} \\ &\Rightarrow 0 > -e \\ &\Rightarrow \ell > \ell - e && \text{by Theorem 5.30a.} \end{aligned}$$

We thus have $\ell > \ell - e > 0$. By choice of ℓ as least element of L , $\ell - e \notin L$, so

$$\ell - e = pe \quad \text{for some } p \in \mathbf{Z}^+.$$

This implies that

$$\begin{aligned}\ell &= pe + e \\ &= (p + 1)e, \quad \text{where } p + 1 \in \mathbf{Z}^+, \end{aligned}$$

and we have a contradiction to the fact that ℓ is an element that cannot be written in the form ne with $n \in \mathbf{Z}^+$. Therefore, $L = \emptyset$, and

$$D^+ = \{ne \mid n \in \mathbf{Z}^+\}.$$

We can now give the characterization of the integers toward which we have been working.

Theorem 5.35 ■ Isomorphic Images of \mathbf{Z}

If D is an ordered integral domain in which the set D^+ of positive elements is well-ordered, then D is isomorphic to the ring \mathbf{Z} of all integers.

$(p \wedge q) \Rightarrow r$ **Proof** We first show that

$$D = \{ne \mid n \in \mathbf{Z}\}.$$

For an arbitrary $x \in D$, the law of trichotomy requires that exactly one of the following holds:

$$x \in D^+, \quad x = 0, \quad -x \in D^+.$$

If $x \in D^+$, then $x = ne$ for some $n \in \mathbf{Z}^+$, by Theorem 5.34b. If $x = 0$, then $x = 0e$. Finally, if $-x \in D^+$, then $-x = me$ for $m \in \mathbf{Z}^+$, and therefore,[†] $x = -(me) = (-m)e$, where $-m \in \mathbf{Z}$. Hence $D = \{ne \mid n \in \mathbf{Z}\}$.

Consider now the rule defined by

$$\phi(ne) = n,$$

for any ne in D . To demonstrate that this rule is well-defined, it is sufficient to show that each element of D can be written as ne in only one way. To do this, suppose $me = ne$. Without loss of generality, we may assume that $m \geq n$. Now

$$\begin{aligned}me &= ne \Rightarrow me - ne = 0 \\ &\Rightarrow (m - n)e = 0. \end{aligned}$$

If $m - n > 0$, then $(m - n)e \in D^+$ by Theorem 5.34b. Therefore, it must be that $m - n = 0$ and $m = n$. This shows that the rule $\phi(ne) = n$ defines a mapping ϕ from D to \mathbf{Z} .

If $\phi(me) = \phi(ne)$, then $m = n$, so $me = ne$. Hence ϕ is one-to-one. An arbitrary $n \in \mathbf{Z}$ is the image of $ne \in D$ under ϕ , so ϕ is an onto mapping.

To show that ϕ is a ring isomorphism, we need to verify that

$$\phi(me + ne) = \phi(me) + \phi(ne)$$

[†]The equality $-(me) = (-m)e$ is the additive form of the familiar property of exponents $(a^m)^{-1} = a^{-m}$ in a group.

and also that

$$\phi(me \cdot ne) = \phi(me) \cdot \phi(ne).$$

From the laws of multiples in Section 3.3, we know that $me + ne = (m + n)e$, and it follows that

$$\begin{aligned}\phi(me + ne) &= \phi((m + n)e) \\ &= m + n \\ &= \phi(me) + \phi(ne).\end{aligned}$$

To show that ϕ preserves multiplication, we need the fact that $me \cdot ne = (mn)e$. This fact is a consequence of the generalized distributive laws stated in Theorem 5.13 and other results from Section 5.1. We leave the details of this proof as Exercise 9 at the end of this section. Using $me \cdot ne = (mn)e$, we have

$$\begin{aligned}\phi(me \cdot ne) &= \phi[(mn)e] \\ &= mn \\ &= \phi(me) \cdot \phi(ne).\end{aligned}$$

Exercises 5.4

True or False

Label each of the following statements as either true or false.

1. Every integral domain contains a set of positive elements.
 2. It is impossible to impose an order relation on the set \mathbf{C} of complex numbers.
 3. In any ordered integral domain, the unity element e is a positive element.
 4. The set \mathbf{R} of real numbers is an ordered integral domain.
 5. The set of all integers is well-ordered.
-

Exercises

1. Complete the proof of Theorem 5.30 by proving the following statements, where x, y , and z are arbitrary elements of an ordered integral domain D .
 - a. If $x > y$ and $y > z$, then $x > z$.
 - b. One and only one of the following statements is true:

$$x > y, \quad x = y, \quad x < y.$$
2. Prove the following statements for arbitrary elements x, y, z of an ordered integral domain D .
 - a. If $x > y$ and $z < 0$, then $xz < yz$.
 - b. If $x > y$ and $z > w$, then $x + z > y + w$.
 - c. If $x > y > 0$, then $x^2 > y^2$.

- d.** If $x \neq 0$ in D , then $x^{2n} > 0$ for every positive integer n .
- e.** If $x > 0$ and $xy > 0$, then $y > 0$.
- f.** If $x > 0$ and $xy > xz$, then $y > z$.
3. Prove the following statements for arbitrary elements in an ordered integral domain.
- $a > b$ implies $-b > -a$.
 - $a > e$ implies $a^2 > a$.
 - If $a > b$ and $c > d$, where a, b, c , and d are all positive elements, then $ac > bd$.
4. Suppose a and b have multiplicative inverses in an ordered integral domain. Prove each of the following statements.
- If $a > b > 0$, then $b^{-1} > a^{-1} > 0$.
 - If $a < 0$, then $a^{-1} < 0$.
5. Prove that the equation $x^2 + e = 0$ has no solution in an ordered integral domain.
6. Prove that if a is any element of an ordered integral domain D , then there exists an element $b \in D$ such that $b > a$. (Thus D has no greatest element, and *no finite integral domain can be an ordered integral domain*.)
7. For an element x of an ordered integral domain D , the **absolute value** $|x|$ is defined by
- Sec. 7.3, #28 ≪
- $$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } 0 > x. \end{cases}$$
- Prove that $|-x| = |x|$ for all $x \in D$.
 - Prove that $-|x| \leq x \leq |x|$ for all $x \in D$.
 - Prove that $|xy| = |x| \cdot |y|$ for all $x, y \in D$.
 - Prove that $|x + y| \leq |x| + |y|$ for all $x, y \in D$.
 - Prove that $\|x| - |y\| \leq |x - y|$ for all $x, y \in D$.
8. If x and y are elements of an ordered integral domain D , prove the following inequalities.
- $x^2 - 2xy + y^2 \geq 0$
 - $x^2 + y^2 \geq xy$
 - $x^2 + y^2 \geq -xy$
9. If e denotes the unity element in an integral domain D , prove that $me \cdot ne = (mn)e$ for all $m, n \in \mathbf{Z}$.
10. An **ordered field** is an ordered integral domain that is also a field. In the quotient field Q of an ordered integral domain D , define Q^+ by

$$Q^+ = \{[a, b] \mid ab \in D^+\}.$$

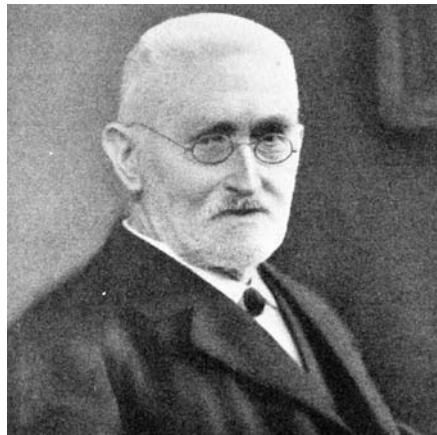
Prove that Q^+ is a set of positive elements for Q and hence, that Q is an ordered field.

11. (See Exercise 10.) According to Definition 5.29, $>$ is defined in Q by $[a, b] > [c, d]$ if and only if $[a, b] - [c, d] \in Q^+$. Show that $[a, b] > [c, d]$ if and only if $abd^2 - cbd^2 \in D^+$.

12. (See Exercises 10 and 11.) If each $x \in D$ is identified with $[x, e]$ in Q , prove that $D^+ \subseteq Q^+$. (This means that the order relation defined in Exercise 10 coincides in D with the original order relation in D . We say that the ordering in Q is an **extension** of the ordering in D .)
13. Prove that if x and y are rational numbers such that $x > y$, then there exists a rational number z such that $x > z > y$. (This means that between any two distinct rational numbers there is another rational number.)
14.
 - a. If D is an ordered integral domain, prove that each element in the quotient field Q of D can be written in the form $[a, b]$ with $b > 0$ in D .
 - b. If $[a, b] \in Q$ with $b > 0$ in D , prove that $[a, b] \in Q^+$ if and only if $a > 0$ in D .
15. (See Exercise 14.) If $[a, b]$ and $[c, d] \in Q$ with $b > 0$ and $d > 0$ in D , prove that $[a, b] > [c, d]$ if and only if $ad > bc$ in D .
16. If x and y are positive rational numbers, prove that there exists a positive integer n such that $nx > y$. This property is called the **Archimedean Property** of the rational numbers. (*Hint:* Write $x = a/b$ and $y = c/d$ with each of $a, b, c, d \in \mathbf{Z}^+$.)

Key Words and Phrases

commutative ring, 261	isomorphic rings, 281	ring isomorphism, 281
embedded, 282	law of trichotomy, 285	ring with unity, 261
extension, 282, 291	least element, 286	set of positive elements, 284
field, 271, 273	multiplicative inverse, 262	subring, 259
finite ring, 260	negative element, 285	unit, 262
generalized associative laws, 264	negative of an element, 258	well-ordered subset, 286
generalized distributive laws, 264	ordered integral domain, 284	zero divisor, 263
greater than, less than, 285	proper divisor of zero, 263	zero of a ring, 258
integral domain, 270	quotient field, 279	
invertible element, 262	ring, 257	



Public Domain Images

A Pioneer in Mathematics

Richard Dedekind (1831–1916)

Julius Wilhelm Richard Dedekind, born on October 6, 1831, in Brunswick, Germany, has been called “the effective founder of abstract algebra” by the mathematics historian Morris Kline. Dedekind introduced the concepts of a ring and an ideal; in fact, he coined the terms *ring*, *ideal*, and *field*. His *Dedekind cuts* provided a technique for construction of the real numbers. Far ahead of his time, he built a foundation for further developments in ring and ideal theory by the famous algebraist Emmy Noether (1882–1935).

At the age of 21, Dedekind earned his doctorate in mathematics working under Carl Friedrich Gauss (1777–1855) at the University of Göttingen. He taught at the university for a few years and presented the first formal lectures on Galois theory to an audience of two students. For four years, beginning in 1858, he was a professor in Zurich, Switzerland. Dedekind spent the next 50 years of his life in Brunswick, teaching in a technical high school that he had once attended. He died on February 12, 1916.

More on Rings

■ Introduction

The basic theorems on quotient rings and ring homomorphisms are presented in this chapter, along with a section on the characteristic of a ring and a section on maximal ideals. The development of \mathbf{Z}_n culminates in Section 6.1 with the final description of \mathbf{Z}_n as a quotient ring of the integers by the principal ideal (n) .

6.1

Ideals and Quotient Rings

In this chapter we develop some theory of rings that parallels the theory of groups presented in Chapters 3 and 4. We shall see that the concept of an *ideal* in a ring is analogous to that of a *normal subgroup* in a group.

Definition 6.1a ■ Definition of an Ideal

The subset I of a ring R is an **ideal** of R if the following conditions hold:

1. I is a subring of R .
2. $x \in I$ and $r \in R$ imply that xr and rx are in I .

Note that the second condition in this definition requires more than closure of I under multiplication. It requires that I “absorbs” multiplication by arbitrary elements of R , both on the right and on the left.

In more advanced study of rings, the type of subring described in Definition 6.1a is referred to as a “two-sided” ideal, and terms that are more specialized are introduced: A **right ideal** of R is a subring S of R such that $xr \in S$ for all $x \in S$, $r \in R$, and a **left ideal** of R is a subring S of R such that $rx \in S$ for all $x \in S$, $r \in R$. Here we only mention these terms in passing, and observe that these distinctions cannot be made in a commutative ring.

The subrings $I = \{0\}$ and $I = R$ are always ideals of a ring R . These ideals are labeled **trivial**.

If R is a ring with unity e and I is an ideal of R that contains e , then it can be shown that it must be true that $I = R$ (see Exercise 11).

Example 1 In Section 5.1, we saw that the set \mathbf{E} of all even integers is a subring of the ring \mathbf{Z} of all integers. To show that condition 2 of Definition 6.1a holds, let $x \in \mathbf{E}$ and $m \in \mathbf{Z}$. Since $x \in \mathbf{E}$, $x = 2k$ for some integer k . We have

$$xm = mx = m(2k) = 2(mk),$$

so $xm = mx$ is in \mathbf{E} . Thus \mathbf{E} is an ideal of \mathbf{Z} .

It is worth noting that \mathbf{E} is also a subring of the ring \mathbf{Q} of all rational numbers, but \mathbf{E} is not an ideal of \mathbf{Q} . Condition 2 fails with $x = 4 \in \mathbf{E}$ and $r = \frac{1}{3} \in \mathbf{Q}$, but $xr = \frac{4}{3}$ is not in \mathbf{E} . ■

In combination with Theorem 5.3, Definition 6.1a provides the following checklist of conditions that must be satisfied in order that a subset I of a ring R be an ideal:

1. I is nonempty.
2. $x \in I$ and $y \in I$ imply that $x + y$ and xy are in I .
3. $x \in I$ implies $-x \in I$.
4. $x \in I$ and $r \in R$ imply that xr and rx are in I .

The multiplicative closure in the second condition is implied by the fourth condition, so it may be deleted to obtain an alternative form of the definition of an ideal.

Definition 6.1b ■ Alternative Definition of an Ideal

A subset I of a ring R is an **ideal** of R provided the following conditions are satisfied:

1. I is nonempty.
2. $x \in I$ and $y \in I$ imply $x + y \in I$.
3. $x \in I$ implies $-x \in I$.
4. $x \in I$ and $r \in R$ imply that xr and rx are in I .

A more efficient checklist is given in Exercise 1 at the end of this section.

Example 2 In Exercise 39d of Section 5.1, we saw that the set

$$S = \left\{ \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \mid a, b, c \in \mathbf{Z} \right\}$$

forms a noncommutative ring with respect to the operations of matrix addition and multiplication. In this ring S , consider the subset

$$I = \left\{ \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} \mid b \in \mathbf{Z} \right\},$$

which is clearly nonempty. Since

$$\begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & y \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & x+y \\ 0 & 0 \end{bmatrix},$$

I is closed under addition. And since

$$-\begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -b \\ 0 & 0 \end{bmatrix},$$

I contains the additive inverse of each of its elements. For arbitrary $\begin{bmatrix} x & y \\ 0 & z \end{bmatrix}$ in S , we have

$$\begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} = \begin{bmatrix} 0 & bz \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & xb \\ 0 & 0 \end{bmatrix},$$

and both of these products are in I . Thus I is an ideal of S . ■

Example 3 Example 8 of Section 5.1 introduced the ring $M = M_2(\mathbf{R})$ of all 2×2 matrices over the real numbers \mathbf{R} , and Exercise 41 of Section 5.1 introduced the subring T of M , given by

$$T = \left\{ \begin{bmatrix} a & a \\ b & b \end{bmatrix} \mid a, b \in \mathbf{R} \right\}.$$

For arbitrary $\begin{bmatrix} a & a \\ b & b \end{bmatrix} \in T$, $\begin{bmatrix} x & y \\ z & w \end{bmatrix} \in M$, the product

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix} \begin{bmatrix} a & a \\ b & b \end{bmatrix} = \begin{bmatrix} xa + yb & xa + yb \\ za + wb & za + wb \end{bmatrix}$$

is in T , so T absorbs multiplication on the left by elements of M . However, the product

$$\begin{bmatrix} a & a \\ b & b \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} ax + az & ay + aw \\ bx + bz & by + bw \end{bmatrix}$$

is *not* always in T , and T does not absorb multiplication on the right by elements of M . This failure keeps T from being an ideal[†] of M . ■

Example 1 may be generalized to the set of all multiples of any fixed integer n . That is, the set $\{nk \mid k \in \mathbf{Z}\}$ of all multiples of n is an ideal of \mathbf{Z} . Instead of proving this fact, we establish the following more general result.

Example 4 Let R be a commutative ring with unity e . For any fixed $a \in R$, we shall show that the set

$$(a) = \{ar \mid r \in R\}$$

is an ideal of R .

[†] T could be said to be a *left ideal* of M .

This set is nonempty, since $a = ae$ is in (a) . Let $x = ar$ and $y = as$ be arbitrary elements of (a) , where $r \in R$, $s \in R$. Then

$$x + y = ar + as = a(r + s),$$

where $r + s \in R$, so (a) is closed under addition. We also have

$$-x = -(ar) = a(-r),$$

where $-r \in R$, so (a) contains additive inverses. For arbitrary $t \in R$,

$$tx = xt = (ar)t = a(rt),$$

where $rt \in R$. Thus $tx = xt$ is in (a) for arbitrary $x \in (a)$, $t \in R$, and (a) is an ideal of R . ■

This example leads to the following definition.

Definition 6.2 ■ Principal Ideal

If a is a fixed element of the commutative ring R with unity, the ideal

$$(a) = \{ar \mid r \in R\},$$

which consists of all multiples of a by elements r of R , is called the **principal ideal** generated by a in R .

The next theorem gives an indication of the importance of principal ideals.

Theorem 6.3 ■ Ideals in \mathbf{Z}

In the ring \mathbf{Z} of integers, every ideal is a principal ideal.

$p \Rightarrow q$ **Proof** The trivial ideal $\{0\}$ is certainly a principal ideal, $\{0\} = (0)$. Consider then an ideal I of \mathbf{Z} such that $I \neq \{0\}$. Since $I \neq \{0\}$, I contains an integer $m \neq 0$. And since I contains both m and $-m$, it must contain some positive integers. Let n be the least positive integer in I . (Such an n exists, by the Well-Ordering Theorem.) For an arbitrary $k \in I$, the Division Algorithm asserts that there are integers q and r such that

$$k = nq + r \quad \text{with} \quad 0 \leq r < n.$$

Solving for r , we have

$$r = k - nq,$$

and this equation shows that $r \in I$, since k and n are in I and I is an ideal. That is, r is an element of I such that $0 \leq r < n$, where n is the *least positive element* of I . This forces the equality $r = 0$, and therefore, $k = nq$. It follows that every element of I is a multiple of n , and therefore $I = (n)$.

Part of the analogy between ideals of a ring and normal subgroups of a group lies in the fact that ideals form the basis for a quotient structure much like the quotient group formed from the cosets of a normal subgroup.

To begin with, a ring R is an abelian group under addition, and any ideal I of R is a normal subgroup of this additive group. Thus we may consider the additive quotient group R/I that consists of all the cosets

$$r + I = I + r = \{r + x \mid x \in I\}$$

of I in R . From our work in Chapter 4, we know that

$$a + I = b + I \quad \text{if and only if} \quad a - b \in I,$$

that

$$(a + I) + (b + I) = (a + b) + I,$$

and that R/I is an abelian group with respect to this operation of addition.

Strategy

- If the defining rule for a possible binary operation is stated in terms of a certain type of representation for the elements, then the rule does not define a binary operation unless the result is independent of the representation for the elements—that is, unless the rule is well-defined.

In order to make a ring from the cosets in R/I , we consider a multiplication defined by

$$(a + I)(b + I) = ab + I.$$

We must show that this multiplication is well-defined. That is, we need to show that if

$$a + I = a' + I \quad \text{and} \quad b + I = b' + I,$$

then

$$ab + I = a'b' + I.$$

Now

$$\begin{aligned} a + I = a' + I &\Rightarrow a = a' + x \quad \text{where } x \in I \\ b + I = b' + I &\Rightarrow b = b' + y \quad \text{where } y \in I. \end{aligned}$$

Thus

$$ab = (a' + x)(b' + y) = a'b' + a'y + xb' + xy.$$

Since $x \in I$, $y \in I$, and I is an ideal, each of $a'y$, xb' , and xy is in I . Therefore, their sum

$$z = a'y + xb' + xy$$

is in I , and $z + I = I$. This gives

$$ab + I = a'b' + z + I = a'b' + I$$

and our product is well-defined.

Theorem 6.4 ■ The Ring of Cosets

Let I be an ideal of the ring R . Then the set R/I of additive cosets $r + I$ of I in R forms a ring with respect to coset addition

$$(a + I) + (b + I) = (a + b) + I$$

and coset multiplication

$$(a + I)(b + I) = ab + I.$$

Proof Assume I is an ideal of R . We noted earlier that the additive quotient group R/I is an abelian group with respect to addition.

We have already proved that the product

$$(a + I)(b + I) = ab + I$$

is well-defined in R/I , and closure under multiplication is automatic from the definition of this product. That the product is associative follows from

$$\begin{aligned} (a + I)[(b + I)(c + I)] &= (a + I)(bc + I) \\ &= a(bc) + I \\ &= (ab)c + I \text{ since multiplication is associative in } R \\ &= (ab + I)(c + I) \\ &= [(a + I)(b + I)](c + I). \end{aligned}$$

Verifying the left distributive law, we have

$$\begin{aligned} (a + I)[(b + I) + (c + I)] &= (a + I)[(b + c) + I] \\ &= a(b + c) + I \\ &= (ab + ac) + I \text{ from the left distributive law in } R \\ &= (ab + I) + (ac + I) \\ &= (a + I)(b + I) + (a + I)(c + I). \end{aligned}$$

The proof of the right distributive law is similar. Leaving that as an exercise, we conclude that R/I is a ring.

Definition 6.5 ■ Quotient Ring

If I is an ideal of the ring R , the ring R/I described in Theorem 6.4 is called the **quotient ring** of R by I .[†]

[†] R/I is also known as “the ring of residue classes modulo the ideal I .”

Example 5 In the ring \mathbf{Z} of integers, consider the principal ideal

$$(4) = \{4k \mid k \in \mathbf{Z}\}.$$

The distinct elements of the ring $\mathbf{Z}/(4)$ are

$$\begin{aligned}(4) &= \{\dots, -8, -4, 0, 4, 8, \dots\} \\ 1 + (4) &= \{\dots, -7, -3, 1, 5, 9, \dots\} \\ 2 + (4) &= \{\dots, -6, -2, 2, 6, 10, \dots\} \\ 3 + (4) &= \{\dots, -5, -1, 3, 7, 11, \dots\}.\end{aligned}$$

We see, then, that these cosets are the same as the elements of \mathbf{Z}_4 :

$$(4) = [0], \quad 1 + (4) = [1], \quad 2 + (4) = [2], \quad 3 + (4) = [3].$$

Moreover, the addition

$$\{a + (4)\} + \{b + (4)\} = \{a + b\} + (4)$$

agrees exactly with

$$[a] + [b] = [a + b]$$

in \mathbf{Z}_4 , and the multiplication

$$\{a + (4)\}\{b + (4)\} = ab + (4)$$

agrees exactly with

$$[a][b] = [ab]$$

in \mathbf{Z}_4 . Thus $\mathbf{Z}/(4)$ is our old friend \mathbf{Z}_4 . Put another way, \mathbf{Z}_4 is the quotient ring of the integers \mathbf{Z} by the ideal (4) . ■

The specific case in Example 5 generalizes at once to an arbitrary integer $n > 1$, and we see that \mathbf{Z}_n is the quotient ring of \mathbf{Z} by the ideal (n) . This is our final and best description of \mathbf{Z}_n .

As a final remark to this section, we note that

$$\begin{aligned}(a + I)(b + I) &= ab + I \\ &\neq \{xy \mid x \in a + I \text{ and } y \in b + I\}.\end{aligned}$$

As a particular instance, consider $I = (4)$ as in Example 5. We have

$$(0 + I)(0 + I) = 0 + I = I.$$

However,

$$\{xy \mid x \in 0 + I \text{ and } y \in 0 + I\} = \{16r \mid r \in \mathbf{Z}\},$$

since $x = 4p$ and $y = 4q$ for $p, q \in \mathbf{Z}$ imply $xy = 16pq$.

Exercises 6.1

True or False

Label each of the following statements as either true or false.

1. Every ideal of a ring R is a subring of R .
 2. Every subring of a ring R is an ideal of R .
 3. The only ideal of a ring R that contains the unity e is the ring R itself.
 4. Any ideal of a ring R is a normal subgroup of the additive group R .
 5. The only ideals of the set of real numbers \mathbf{R} are the trivial ideals.
 6. Every ideal of \mathbf{Z} is a principal ideal.
 7. For $n > 1$, the quotient ring of \mathbf{Z} by the ideal (n) is \mathbf{Z}_n .
 8. If I is an ideal of S where S is a subring of a ring R , then I is an ideal of R .
-

Exercises

1. Let I be a subset of the ring R . Prove that I is an ideal of R if and only if I is nonempty and $x - y$, xr , and rx are in I for all x and $y \in I$, $r \in R$.
2.
 - a. Complete the proof of Theorem 6.4 by proving the right distributive law in R/I .
 - b. Prove that R/I is commutative if R is commutative.
 - c. Prove that R/I has a unity if R has a unity.
3. Prove or disprove each of the following statements.
 - a. The set \mathbf{Q} of rational numbers is an ideal of the set \mathbf{R} of real numbers.
 - b. The set \mathbf{Z} of integers is an ideal of the set \mathbf{Q} of rational numbers.
4. If I_1 and I_2 are two ideals of the ring R , prove that $I_1 \cap I_2$ is an ideal of R .
5. If $\{I_\lambda\}$, $\lambda \in \mathcal{L}$, is an arbitrary collection of ideals I_λ of the ring R , prove that $\bigcap_{\lambda \in \mathcal{L}} I_\lambda$ is an ideal of R .
6. Find two ideals I_1 and I_2 of the ring \mathbf{Z} such that
 - a. $I_1 \cup I_2$ is *not* an ideal of \mathbf{Z} .
 - b. $I_1 \cup I_2$ is an ideal of \mathbf{Z} .
7. Let I be an ideal of a ring R , and let S be a subring of R . Prove that $I \cap S$ is an ideal of S .
8. If I_1 and I_2 are two ideals of the ring R , prove that the set

$$I_1 + I_2 = \{x + y \mid x \in I_1, y \in I_2\}$$

is an ideal of R that contains each of I_1 and I_2 .

9. Let I_1 and I_2 be ideals of the ring R . Prove that the set

$$I_1 I_2 = \{a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \mid a_i \in I_1, b_i \in I_2, n \in \mathbf{Z}^+\}$$

is an ideal of R .

10. Prove that if R is a field, then R has only the trivial ideals $\{0\}$ and R .
11. Let I be an ideal in a ring R with unity e . Prove that if $e \in I$, then $I = R$.
12. Let I be an ideal in a ring R with unity. Prove that if I contains an element a that has a multiplicative inverse, then $I = R$.
13. In the ring \mathbf{Z} of integers, prove that every subring is an ideal.
14. Let $a \neq 0$ in the ring of integers \mathbf{Z} . Find $b \in \mathbf{Z}$ such that $a \neq b$ but $(a) = (b)$.
15. Let m and n be nonzero integers. Prove that $(m) \subseteq (n)$ if and only if n divides m .
16. If a and b are nonzero integers and m is the least common multiple of a and b , prove that $(a) \cap (b) = (m)$.

Sec. 6.2, #23 <

17. Prove that every ideal of \mathbf{Z}_n is a principal ideal. (*Hint:* See Corollary 3.23.)
18. Let $[a] \in \mathbf{Z}_n$. Prove $([a]) = ([n - a])$.
19. Find all distinct principal ideals of \mathbf{Z}_n for the given value of n .

a. $n = 7$	b. $n = 11$	c. $n = 12$
d. $n = 18$	e. $n = 20$	f. $n = 24$
20. If R is a commutative ring and a is a fixed element of R , prove that the set $I_a = \{x \in R \mid ax = 0\}$ is an ideal of R . (The set I_a is called the **annihilator** of a in the ring R .)

Sec. 5.1, #39d >

21. Given that the set

$$S = \left\{ \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} \middle| x, y, z \in \mathbf{Z} \right\}$$

is a ring with respect to matrix addition and multiplication, show that

$$I = \left\{ \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} \middle| a, b \in \mathbf{Z} \right\}$$

Sec. 6.2, #6 <

is an ideal of S .

22. Show that the set

$$I = M_2(\mathbf{E}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a, b, c, \text{ and } d \text{ are in } \mathbf{E} \right\}$$

of all 2×2 matrices over the ring \mathbf{E} of even integers is an ideal of the ring $M_2(\mathbf{Z})$.

23. With S as in Exercise 21, decide whether or not the set

$$U = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} \middle| a, b \in \mathbf{Z} \right\}$$

is an ideal of S , and justify your answer.

24. a. Show that the set

$$R = \left\{ \begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} \middle| x, y \in \mathbf{Z} \right\}$$

is a ring with respect to matrix addition and multiplication.

- b.** Is R commutative?
- c.** Does R have a unity?
- d.** Decide whether or not the set

$$U = \left\{ \begin{bmatrix} 0 & 0 \\ a & 0 \end{bmatrix} \mid a \in \mathbf{Z} \right\}$$

Sec. 6.2, #7 << is an ideal of R , and justify your answer.

Sec. 5.1, #1f >> **25.** Let G be the set of Gaussian integers $\{m + ni \mid m, n \in \mathbf{Z}\}$. Let

$$I = \{a + bi \mid a \in \mathbf{Z}, b \in \mathbf{E}\}.$$

- a.** Prove or disprove that I is a subring of G .
 - b.** Prove or disprove that I is an ideal of G .
- 26.** **a.** For a fixed element a of a commutative ring R , prove that the set $I = \{ar \mid r \in R\}$ is an ideal of R . (*Hint:* Compare this with Example 4, and note that the element a itself may not be in this set I .)
- b.** Give an example of a commutative ring R and an element $a \in R$ such that $a \notin (a) = \{ar \mid r \in R\}$.
- 27.** Let R be a commutative ring that does not have a unity. For a fixed $a \in R$, prove that the set

$$(a) = \{na + ra \mid n \in \mathbf{Z}, r \in R\}$$

Sec. 6.4, #2–4 <<

Sec. 6.4, #16, 17 << is an ideal of R that contains the element a . (This ideal is called the **principal ideal** of R that is **generated** by a .)

Sec. 6.4, #19, 20 <<

- 28. a.** Let I be an ideal of the commutative ring R and $a \in R$. Prove that the set

$$S = \{ar + s \mid r \in R, s \in I\}$$

is an ideal of R containing I .

- b.** If $e \in R$ and $a \notin I$, show that $I \subset S$.

Sec. 5.1, #46 >> **29.** An element a of a ring R is called **nilpotent** if $a^n = 0$ for some positive integer n . Show that the set of all nilpotent elements in a commutative ring R forms an ideal of R . (This ideal is called the **radical** of R .)

Sec. 6.2, #18b <<

- 30.** If I is an ideal of R , prove that the set

$$K_I = \{x \in R \mid xa = 0 \text{ for all } a \in I\}$$

is an ideal of R . (The set K_I is called the **annihilator** of the ideal I .)

- 31.** Let R be a commutative ring with unity whose only ideals are $\{0\}$ and R itself. Prove that R is a field. (*Hint:* See Exercise 26.)
- 32.** Suppose that R is a commutative ring with unity and that I is an ideal of R . Prove that the set of all $x \in R$ such that $x^n \in I$ for some positive integer n is an ideal of R .

6.2**Ring Homomorphisms**

We turn our attention now to *ring homomorphisms* and their relations to ideals and quotient rings.

Definition 6.6 ■ **Ring Homomorphism**

If R and R' are rings, a **ring homomorphism** from R to R' is a mapping $\theta: R \rightarrow R'$ such that

$$\theta(x + y) = \theta(x) + \theta(y) \quad \text{and} \quad \theta(xy) = \theta(x)\theta(y)$$

for all x and y in R .

That is, a ring homomorphism is a mapping from one ring to another that preserves both ring operations. This situation is analogous to the one where a homomorphism from one group to another preserves the group operation, and it explains the use of the term *homomorphism* in both situations. It is sometimes desirable to use either the term *group homomorphism* or the term *ring homomorphism* for clarity, but in many cases, the context makes the meaning clear for the single word *homomorphism*. If only groups are under consideration, then *homomorphism* means group homomorphism, and if rings are under consideration, *homomorphism* means ring homomorphism.

Some terminology for a special type of homomorphism is given in the following definition.

Definition 6.7 ■ **Ring Epimorphism, Isomorphism**

Let θ be a homomorphism from the ring R to the ring R' .

1. If θ is onto, then θ is called an **epimorphism** and R' is called a **homomorphic image** of R .
2. If θ is a one-to-one correspondence (both onto and one-to-one), then θ is an **isomorphism**.

Example 1 Consider the mapping $\theta: \mathbf{Z} \rightarrow \mathbf{Z}_n$ defined by

$$\theta(a) = [a].$$

Since

$$\theta(a + b) = [a + b] = [a] + [b] = \theta(a) + \theta(b)$$

and

$$\theta(ab) = [ab] = [a][b] = \theta(a)\theta(b)$$

for all a and b in \mathbf{Z} , θ is a homomorphism from \mathbf{Z} to \mathbf{Z}_n . In fact, θ is an *epimorphism* and \mathbf{Z}_n is a *homomorphic image* of \mathbf{Z} . ■

Example 2 Consider $\theta: \mathbf{Z}_6 \rightarrow \mathbf{Z}_6$ defined by

$$\theta([a]) = 4[a].$$

It follows from

$$\begin{aligned}\theta([a] + [b]) &= 4([a] + [b]) \\ &= 4[a] + 4[b] \\ &= \theta([a]) + \theta([b])\end{aligned}$$

that θ preserves addition. For multiplication, we have

$$\theta([a][b]) = \theta([ab]) = 4[ab] = [4ab]$$

and

$$\theta([a])\theta([b]) = (4[a])(4[b]) = 16[ab] = [16ab] = [4ab],$$

since $[16] = [4]$ in \mathbf{Z}_6 . Thus θ is a homomorphism. It can be verified that $\theta(\mathbf{Z}_6) = \{[0], [2], [4]\}$, and we see that θ is neither onto nor one-to-one. ■

Theorem 6.8 ■ Images of Zero and Additive Inverses

If θ is a homomorphism from the ring R to the ring R' , then

- a. $\theta(0) = 0$, and
- b. $\theta(-r) = -\theta(r)$ for all $r \in R$.

$p \Rightarrow q$ **Proof** The statement in part a follows from

$$\begin{aligned}\theta(0) &= \theta(0) + 0 \\ &= \theta(0) + \theta(0) - \theta(0) \\ &= \theta(0 + 0) - \theta(0) \\ &= \theta(0) - \theta(0) \\ &= 0.\end{aligned}$$

$(p \wedge q) \Rightarrow r$ To prove part b, we observe that

$$\begin{aligned}\theta(r) + \theta(-r) &= \theta[r + (-r)] \\ &= \theta(0) \\ &= 0.\end{aligned}$$

Since the additive inverse is unique in the additive group of R' ,

$$-\theta(r) = \theta(-r).$$

Under a ring homomorphism, images of subrings are subrings, and inverse images of subrings are also subrings. This is the content of the next theorem.

Theorem 6.9 ■ Images and Inverse Images of Subrings

Suppose θ is a homomorphism from the ring R to the ring R' .

- a. If S is a subring of R , then $\theta(S)$ is a subring of R' .
- b. If S' is a subring of R' , then $\theta^{-1}(S')$ is a subring of R .

$(p \wedge q) \Rightarrow r$ **Proof** To prove part **a**, suppose S is a subring of R . We shall verify that the conditions of Theorem 5.3 are satisfied by $\theta(S)$. The element $\theta(0) = 0$ is in $\theta(S)$, so $\theta(S)$ is nonempty. Let x' and y' be arbitrary elements of $\theta(S)$. Then there exist elements $x, y \in S$ such that $\theta(x) = x'$ and $\theta(y) = y'$. Since S is a subring, $x + y$ and xy are in S . Therefore,

$$\begin{aligned}\theta(x + y) &= \theta(x) + \theta(y) \\ &= x' + y'\end{aligned}$$

and

$$\theta(xy) = \theta(x)\theta(y) = x'y'$$

are in $\theta(S)$, and $\theta(S)$ is closed under addition and multiplication. Since $-x$ is in S and

$$\theta(-x) = -\theta(x) = -x',$$

we have $-x' \in \theta(S)$, and it follows that $\theta(S)$ is a subring of R' .

$(p \wedge q) \Rightarrow r$ To prove part **b**, assume that S' is a subring of R' . We have 0 in $\theta^{-1}(S')$ since $\theta(0) = 0$, so $\theta^{-1}(S')$ is nonempty. Let $x \in \theta^{-1}(S')$ and $y \in \theta^{-1}(S')$. This implies that $\theta(x) \in S'$ and $\theta(y) \in S'$. Hence $\theta(x) + \theta(y) = \theta(x + y)$ and $\theta(x)\theta(y) = \theta(xy)$ are in S' , since S' is a subring. Now

$$\theta(x + y) \in S' \Rightarrow x + y \in \theta^{-1}(S')$$

and

$$\theta(xy) \in S' \Rightarrow xy \in \theta^{-1}(S').$$

We also have

$$\begin{aligned}\theta(x) \in S' &\Rightarrow -\theta(x) = \theta(-x) \in S' \\ &\Rightarrow -x \in \theta^{-1}(S'),\end{aligned}$$

and $\theta^{-1}(S')$ is a subring of R by Theorem 5.3.

Definition 6.10 ■ Kernel

If θ is a homomorphism from the ring R to the ring R' , the **kernel** of θ is the set

$$\ker \theta = \{x \in R \mid \theta(x) = 0\}.$$

Example 3 In Example 1, the epimorphism $\theta: \mathbf{Z} \rightarrow \mathbf{Z}_n$ is defined by $\theta(a) = [a]$. Now $\theta(a) = [0]$ if and only if a is a multiple of n , so

$$\ker \theta = \{\dots, -2n, -n, 0, n, 2n, \dots\}$$

for this θ .

In Example 2, the homomorphism $\theta: \mathbf{Z}_6 \rightarrow \mathbf{Z}_6$ defined by $\theta([a]) = 4[a]$ has kernel given by

$$\ker \theta = \{[0], [3]\}. \quad \blacksquare$$

In these two examples, $\ker \theta$ is an ideal of the domain of θ . This is true in general for homomorphisms, according to the following theorem.

Theorem 6.11 ■ Kernel of a Ring Homomorphism

If θ is any homomorphism from the ring R to the ring R' , then $\ker \theta$ is an ideal of R , and $\ker \theta = \{0\}$ if and only if θ is one-to-one.

$p \Rightarrow q$ **Proof** Under the hypothesis, we know that $\ker \theta$ is a subring of R from Theorem 6.9. For any $x \in \ker \theta$ and $r \in R$, we have

$$\begin{aligned}\theta(xr) &= \theta(x)\theta(r) \\ &= 0 \cdot \theta(r) = 0,\end{aligned}$$

and similarly $\theta(rx) = 0$. Thus xr and rx are in $\ker \theta$, and $\ker \theta$ is an ideal of R .

$u \Leftarrow v$ Suppose θ is one-to-one. Then $x \in \ker \theta$ implies $\theta(x) = 0 = \theta(0)$, and therefore $x = 0$. Hence $\ker \theta = \{0\}$ if θ is one-to-one.

$u \Rightarrow v$ Conversely, if $\ker \theta = \{0\}$, then

$$\begin{aligned}\theta(x) = \theta(y) &\Rightarrow \theta(x) - \theta(y) = 0 \\ &\Rightarrow \theta(x - y) = 0 \\ &\Rightarrow x - y = 0 \\ &\Rightarrow x = y.\end{aligned}$$

This means that θ is one-to-one if $\ker \theta = \{0\}$, and the proof is complete.

Example 4 This example illustrates the last part of Theorem 6.11 and provides a nice example of a ring isomorphism.

For the set $U = \{a, b\}$, the power set of U is $\mathcal{P}(U) = \{\emptyset, A, B, U\}$, where $A = \{a\}$ and $B = \{b\}$. With addition defined by

$$X + Y = (X \cup Y) - (X \cap Y)$$

and multiplication by

$$X \cdot Y = X \cap Y,$$

$\mathcal{P}(U)$ forms a ring, as we saw in Example 5 of Section 5.1. Addition and multiplication tables for $\mathcal{P}(U)$ are given in Figure 6.1.

+	\emptyset	A	B	U
\emptyset	\emptyset	A	B	U
A	A	\emptyset	U	B
B	B	U	\emptyset	A
U	U	B	A	\emptyset

.	\emptyset	A	B	U
\emptyset	\emptyset	\emptyset	\emptyset	\emptyset
A	\emptyset	A	\emptyset	A
B	\emptyset	\emptyset	B	B
U	\emptyset	A	B	U

Figure 6.1

The ring $R = \mathbf{Z}_2 \oplus \mathbf{Z}_2$ was introduced in Exercises 47 and 48 of Section 5.1. If we write 0 for [0] and 1 for [1] in \mathbf{Z}_2 , the set R is given by $R = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. Addition and multiplication tables for R are displayed in Figure 6.2.

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

.	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
(1, 0)	(0, 0)	(1, 0)	(0, 0)	(1, 0)
(0, 1)	(0, 0)	(0, 0)	(0, 1)	(0, 1)
(1, 1)	(0, 0)	(1, 0)	(0, 1)	(1, 1)

Figure 6.2

Consider the mapping $\theta: \mathcal{P}(U) \rightarrow R$ defined by

$$\theta(\emptyset) = (0, 0), \quad \theta(A) = (1, 0), \quad \theta(B) = (0, 1), \quad \theta(U) = (1, 1).$$

If each element x in the tables for $\mathcal{P}(U)$ is replaced by $\theta(x)$, the resulting tables agree completely with those in Figure 6.2. Thus θ is an isomorphism. We note that the kernel of θ consists of the zero element in $\mathcal{P}(U)$. ■

We know now that every kernel of a homomorphism from a ring R is an ideal of R . The next theorem shows that every ideal of R is a kernel of a homomorphism from R . This means that the ideals of R and the kernels of the homomorphisms from R to another ring are the same subrings of R .

Theorem 6.12 ■ Quotient Ring \Rightarrow Homomorphic Image

If I is an ideal of the ring R , the mapping $\theta: R \rightarrow R/I$ defined by

$$\theta(r) = r + I$$

is an epimorphism from R to R/I with kernel I .

$p \Rightarrow q$ **Proof** It is clear that the rule $\theta(r) = r + I$ defines an onto mapping θ from R to R/I and that $\ker \theta = I$. Since

$$\begin{aligned}\theta(x + y) &= (x + y) + I \\ &= (x + I) + (y + I) \\ &= \theta(x) + \theta(y)\end{aligned}$$

and

$$\begin{aligned}\theta(xy) &= xy + I \\ &= (x + I)(y + I) \\ &= \theta(x)\theta(y),\end{aligned}$$

θ is indeed an epimorphism from R to R/I .

The last theorem shows that every quotient ring of a ring R is a homomorphic image of R . A result in the opposite direction is given in the next theorem.

Strategy ■ In the proof of Theorem 6.13, it is shown that a certain rule defines a mapping ϕ . When the defining rule for a possible mapping is stated in terms of a certain type of representation for the elements, the rule does not define a mapping unless the result is independent of the representation of the elements—that is, unless the rule is well-defined.

Theorem 6.13 ■ Homomorphic Image \Rightarrow Quotient Ring

If a ring R' is a homomorphic image of the ring R , then R' is isomorphic to a quotient ring of R .

$p \Rightarrow q$ **Proof** Suppose θ is an epimorphism from R to R' , and let $K = \ker \theta$. For each $a + K$ in R/K , define $\phi(a + K)$ by

$$\phi(a + K) = \theta(a).$$

To prove that this rule defines a mapping, let $a + K$ and $b + K$ be arbitrary elements of R/K . Then

$$\begin{aligned}a + K = b + K &\Leftrightarrow a - b \in K \\ &\Leftrightarrow \theta(a - b) = 0 \\ &\Leftrightarrow \theta(a) = \theta(b) \\ &\Leftrightarrow \phi(a + K) = \phi(b + K).\end{aligned}$$

This shows that ϕ is well-defined and one-to-one as well. From the definition of ϕ , it follows that $\phi(R/K) = \theta(R)$. But $\theta(R) = R'$, since θ is an epimorphism. Thus ϕ is onto and, consequently, is a one-to-one correspondence from R/K to R' .

For arbitrary $a + K$ and $b + K$ in R/K ,

$$\begin{aligned}\phi[(a+K)+(b+K)] &= \phi[(a+b)+K] \\ &= \theta(a+b) \\ &= \theta(a)+\theta(b) \text{ since } \theta \text{ is an epimorphism} \\ &= \phi(a+K)+\phi(b+K)\end{aligned}$$

and

$$\begin{aligned}\phi[(a+K)(b+K)] &= \phi(ab+K) \\ &= \theta(ab) \\ &= \theta(a)\theta(b) \text{ since } \theta \text{ is an epimorphism} \\ &= \phi(a+K)\phi(b+K).\end{aligned}$$

Thus ϕ is an isomorphism from R/K to R' .

As an immediate consequence of the proof of this theorem, we have the following **Fundamental Theorem of Ring Homomorphisms**.

Theorem 6.14 ■ Fundamental Theorem of Ring Homomorphisms

If θ is an epimorphism from the ring R to the ring R' , then R' is isomorphic to $R/\ker \theta$.

We now see that, in the sense of isomorphism, the homomorphic images of a ring R are the same as the quotient rings of R . This gives a systematic way to search for all the homomorphic images of a given ring. To illustrate the usefulness of this method, we shall find all the homomorphic images of the ring \mathbf{Z} of integers.

Example 5 In order to find all homomorphic images of \mathbf{Z} , we shall find all possible ideals of \mathbf{Z} and form all possible quotient rings. According to Theorem 6.3, every ideal of \mathbf{Z} is a principal ideal.

For the trivial ideal $(0) = \{0\}$, we obtain the quotient ring $\mathbf{Z}/(0)$, which is isomorphic to \mathbf{Z} , since $a+(0) = b+(0)$ if and only if $a = b$. For the other trivial ideal $(1) = \mathbf{Z}$, we obtain the quotient ring \mathbf{Z}/\mathbf{Z} , which has only one element and is isomorphic to $\{0\}$. As shown in the proof of Theorem 6.3, any nontrivial ideal I of \mathbf{Z} has the form $I = (n)$ for some positive integer $n > 1$. For these ideals, we obtain the quotient rings[†] $\mathbf{Z}/(n) = \mathbf{Z}_n$. Thus the homomorphic images of \mathbf{Z} are \mathbf{Z} itself, $\{0\}$, and the rings \mathbf{Z}_n . ■

Exercises 6.2

True or False

Label each of the following statements as either true or false.

1. A ring homomorphism from a ring R to a ring R' must preserve both ring operations.
2. If a homomorphism exists from a ring R to a ring R' , then R' is called a homomorphic image of R .

[†]See the paragraph immediately following Example 5 in Section 6.1.

3. The ideals of a ring R and the kernels of the homomorphisms from R to another ring are the same subrings of R .
 4. Every quotient ring of a ring R is a homomorphic image of R .
 5. A ring homomorphism from R to R' is a group homomorphism from the additive group R to the additive group R' .
-

Exercises

Unless otherwise stated, R and R' denote arbitrary rings throughout this set of exercises. In Exercises 1–4, suppose R and R' are isomorphic rings.

1. Prove that R is commutative if and only if R' is commutative.
2. Prove that R has a unity if and only if R' has a unity.
3. Prove that R contains an idempotent element if and only if R' does.
4. Prove that R contains a zero divisor if and only if R' does.

Sec. 5.1, #36 ➤

5. (See Exercise 2.) Suppose that θ is an epimorphism from R to R' and that R has a unity. Prove that if a^{-1} exists for $a \in R$, then $[\theta(a)]^{-1}$ exists, and $[\theta(a)]^{-1} = \theta(a^{-1})$.

Sec. 6.1, #21 ➤

6. Assume that the set

$$S = \left\{ \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} \mid x, y, z \in \mathbf{Z} \right\}$$

is a ring with respect to matrix addition and multiplication.

- a. Verify that the mapping $\theta: S \rightarrow \mathbf{Z}$ defined by $\theta\left(\begin{bmatrix} x & y \\ 0 & z \end{bmatrix}\right) = z$ is an epimorphism from S to \mathbf{Z} .
- b. Describe $\ker \theta$, and exhibit an isomorphism from $S/\ker \theta$ to \mathbf{Z} .

Sec. 6.1, #24 ➤

7. Assume that the set

$$R = \left\{ \begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} \mid x, y \in \mathbf{Z} \right\}$$

is a ring with respect to matrix addition and multiplication.

- a. Verify that the mapping $\theta: R \rightarrow \mathbf{Z}$ defined by $\theta\left(\begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix}\right) = x$ is an epimorphism from R to \mathbf{Z} .
- b. Describe $\ker \theta$ and exhibit an isomorphism from $R/\ker \theta$ to \mathbf{Z} .
8. For any $a \in \mathbf{Z}$, let $[a]_6$ denote $[a]$ in \mathbf{Z}_6 and let $[a]_2$ denote $[a]$ in \mathbf{Z}_2 .
 - a. Prove that the mapping $\theta: \mathbf{Z}_6 \rightarrow \mathbf{Z}_2$ defined by $\theta([a]_6) = [a]_2$ is a homomorphism.
 - b. Find $\ker \theta$.
9. Let $\theta: \mathbf{Z}_3 \rightarrow \mathbf{Z}_{12}$ be defined by $\theta([x]_3) = 4[x]_{12}$ using the same notational convention as in Exercise 8.
 - a. Prove that θ is a ring homomorphism.
 - b. Is $\theta(e) = e'$ where e is the unity in \mathbf{Z}_3 and e' is the unity in \mathbf{Z}_{12} ?

10. Let R be a ring with unity e . Verify that the mapping $\theta: \mathbf{Z} \rightarrow R$ defined by $\theta(x) = x \cdot e$ is a homomorphism.
11. In the field \mathbf{C} of complex numbers, show that the mapping θ that maps each complex number onto its conjugate, $\theta(a + bi) = a - bi$, is an isomorphism from \mathbf{C} to \mathbf{C} .
12. (See Example 3 of Section 5.1.) Let S denote the subring of the real numbers that consists of all real numbers of the form $m + n\sqrt{2}$, with $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$. Prove that $\theta(m + n\sqrt{2}) = m - n\sqrt{2}$ defines an isomorphism from S to S .
13. Define $\theta: M_2(\mathbf{Z}) \rightarrow M_2(\mathbf{Z}_2)$ by

$$\theta\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} [a] & [b] \\ [c] & [d] \end{bmatrix}.$$

Prove that θ is a homomorphism, and describe $\ker \theta$.

14. Assume that

$$R = \left\{ \begin{bmatrix} m & 2n \\ n & m \end{bmatrix} \mid m, n \in \mathbf{Z} \right\}$$

and

$$R' = \{m + n\sqrt{2} \mid m, n \in \mathbf{Z}\}$$

are rings with respect to their usual operations, and prove that R and R' are isomorphic rings.

15. Let $\theta: M_2(\mathbf{Z}) \rightarrow \mathbf{Z}$ where $M_2(\mathbf{Z})$ is the ring of 2×2 matrices over the integers \mathbf{Z} . Prove or disprove that each of the following mappings is a homomorphism.

Sec. 3.6, #9 ➤

a. $\theta\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$

b. $\theta\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + d$ (This mapping is called the **trace** of the matrix.)

16. Consider the mapping $\theta: \mathbf{Z}_{12} \rightarrow \mathbf{Z}_{12}$ defined by $\theta([a]) = 4[a]$. Decide whether θ is a homomorphism, and justify your answer.

17. Let R, R', R'' be rings and $\theta_1: R \rightarrow R'$ and $\theta_2: R' \rightarrow R''$ be homomorphisms. Prove that $\theta_2\theta_1: R \rightarrow R''$ is a homomorphism.

18. Suppose θ is a homomorphism from R to R' .

- a. Let $x \in R$. Prove that $\theta(x^n) = (\theta(x))^n$ for all positive integers n .

- b. Prove that if $x \in R$ is nilpotent, then $\theta(x)$ is nilpotent in R' .

Sec. 6.1, #29 ➤

- Sec. 5.1, #32 ➤ 19. Figure 6.3 gives addition and multiplication tables for the ring $R = \{a, b, c\}$ in Exercise 32 of Section 5.1. Use these tables, together with addition and multiplication tables for \mathbf{Z}_3 , to find an isomorphism from R to \mathbf{Z}_3 .

+	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

.	a	b	c
a	a	a	a
b	a	c	b
c	a	b	c

Figure 6.3

Sec. 5.1, #33 >

- 20.** Figure 6.4 gives addition and multiplication tables for the ring $R = \{a, b, c, d\}$ in Exercise 33 of Section 5.1. Construct addition and multiplication tables for the subring $R' = \{[0], [2], [4], [6]\}$ of \mathbf{Z}_8 , and find an isomorphism from R to R' .

+	a	b	c	d
a	a	b	c	d
b	b	c	d	a
c	c	d	a	b
d	d	a	b	c

.	a	b	c	d
a	a	a	a	a
b	a	c	a	c
c	a	a	a	a
d	a	c	a	c

Figure 6.4

Sec. 5.1, #47 >

- 21.** Let R_1 be the subring of $R \oplus R'$ that consists of all elements of the form $(r, 0)$, where $r \in R$. Prove that R_1 is isomorphic to R .
- 22.** Each of the following rules determines a mapping $\theta: \mathbf{R} \rightarrow \mathbf{R}$, where \mathbf{R} is the field of real numbers. Decide in each case whether θ preserves addition, whether θ preserves multiplication, and whether θ is a homomorphism.
- a.** $\theta(x) = |x|$
 - b.** $\theta(x) = 2x$
 - c.** $\theta(x) = -x$
 - d.** $\theta(x) = x^2$
 - e.** $\theta(x) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{1}{x} & \text{if } x \neq 0 \end{cases}$
 - f.** $\theta(x) = x + 1$

Sec. 6.1, #17 >

- 23.** For each given value of n , find all homomorphic images of \mathbf{Z}_n .
- a.** $n = 6$
 - b.** $n = 10$
 - c.** $n = 12$
 - d.** $n = 18$
 - e.** $n = 8$
 - f.** $n = 20$
- 24.** Suppose F is a field and θ is an epimorphism from F to a ring S such that $\ker \theta \neq F$. Prove that θ is an isomorphism and that S is a field.
- 25.** Assume that θ is an epimorphism from R to R' . Prove the following statements.
- a.** If I is an ideal of R , then $\theta(I)$ is an ideal of R' .
 - b.** If I' is an ideal of R' , then $\theta^{-1}(I')$ is an ideal of R .
 - c.** The mapping $I \rightarrow \theta(I)$ is a bijection from the set of ideals I of R that contain $\ker \theta$ to the set of all ideals of R' .

26. In the ring \mathbf{Z} of integers, let new operations of addition and multiplication be defined by

$$x \oplus y = x + y + 1 \quad \text{and} \quad x \odot y = xy + x + y,$$

where x and y are arbitrary integers and $x + y$ and xy denote the usual addition and multiplication in \mathbf{Z} .

- a. Prove that the integers form a ring R' with respect to \oplus and \odot .
- b. Identify the zero element and unity of R' .
- c. Prove that \mathbf{Z} is isomorphic to R' .

Sec. 4.6, #36 ➤ 27. Let K and I be ideals of the ring R . Prove that $K/K \cap I$ is isomorphic to $(K + I)/I$.

6.3

The Characteristic of a Ring

In this section, we focus on the fact that the elements of a ring R form an abelian group under addition.

When the binary operation in a group G is multiplication, each element a of G generates a cyclic group $\langle a \rangle$ that consists of all integral powers of a . If there are positive integers n such that $a^n = e$ and m is the smallest such positive integer, then m is the (multiplicative) *order* of a .

When the binary operation in a group is addition, the cyclic subgroup $\langle a \rangle$ consists of all integral multiples ka of a . If there are positive integers n such that $na = 0$ and m is the smallest such positive integer, then m is the (additive) *order* of a . In a sense, the characteristic of a ring is a generalization from this idea.

Definition 6.15 ■ Characteristic

If there are positive integers n such that $nx = 0$ for *all* x in the ring R , then the smallest positive integer m such that $mx = 0$ for all $x \in R$ is called the **characteristic** of R . If no such positive integer exists, then R is said to be of **characteristic zero**.

It is logical in the last case to call zero the characteristic of R since $n = 0$ is the only integer such that $nx = 0$ for all $x \in R$.

Example 1 The ring \mathbf{Z} of integers has characteristic zero since $nx = 0$ for all $x \in \mathbf{Z}$ requires that $n = 0$. For the same reason, the field \mathbf{R} of real numbers and the field \mathbf{C} of complex numbers both have characteristic zero. ■

Example 2 Consider the ring \mathbf{Z}_6 . For the various elements of \mathbf{Z}_6 , we have

$$\begin{array}{lll} 1[0] = [0] & 6[1] = [0] & 3[2] = [0] \\ 2[3] = [0] & 3[4] = [0] & 6[5] = [0]. \end{array}$$

Although smaller positive integers work for some individual elements of \mathbf{Z}_6 , the smallest positive integer m such that $m[a] = [0]$ for all $[a] \in \mathbf{Z}_6$ is $m = 6$. Thus \mathbf{Z}_6 has characteristic 6. This example generalizes readily, and we see that \mathbf{Z}_n has characteristic n . ■

Theorem 6.16 ■ Characteristic of a Ring

Let R be a ring with unity e . If e has finite additive order m , then m is the characteristic of R .

$p \Rightarrow q$ **Proof** Suppose R is a ring with unity e and that e has finite additive order m . Then m is the least positive integer such that $me = 0$. For arbitrary $x \in R$,

$$mx = m(ex) = (me)x = 0 \cdot x = 0.$$

Thus $mx = 0$ for all $x \in R$, and m is the smallest positive integer for which this is true. By Definition 6.15, R has characteristic m .

In connection with the last theorem, we note that if R has a unity e and e does not have finite additive order, then R has characteristic zero. In either case, the characteristic can be determined simply by investigating the additive order of e .

Theorem 6.17 ■ Characteristic of an Integral Domain

The characteristic of an integral domain is either zero or a prime integer.

$\sim p \Leftarrow (\sim q \wedge \sim r)$ **Proof** Let D be an integral domain. As mentioned before, D has characteristic zero if the additive order of the unity e is not finite. Suppose, then, that e has finite additive order m . By Theorem 6.16, D has characteristic m , and we need only show that m is a prime integer. Assume, to the contrary, that m is not a prime and $m = rs$ for positive integers r and s such that $1 < r < m$ and $1 < s < m$. Then we have $re \neq 0$ and $se \neq 0$, but

$$(re)(se) = (rs)e^2 = (rs)e = me = 0.$$

This is a contradiction to the fact that D is an integral domain. Therefore, m is a prime integer, and the proof is complete.

If the characteristic of a ring R is zero, it follows that R has an infinite number of elements. However, the converse is not true. R may have an infinite number of elements and not have characteristic zero. This is illustrated in the next example.

Example 3 Consider the ring $\mathcal{P}(\mathbf{Z})$ of all subsets of the integers \mathbf{Z} , with operations

$$X + Y = (X \cup Y) - (X \cap Y)$$

$$X \cdot Y = X \cap Y$$

for all X, Y in $\mathcal{P}(\mathbf{Z})$. The ring $\mathcal{P}(\mathbf{Z})$ has an infinite number of elements, yet

$$\begin{aligned} X + X &= (X \cup X) - (X \cap X) \\ &= X - X \\ &= \emptyset, \end{aligned}$$

where \emptyset is the zero element for $\mathcal{P}(\mathbf{Z})$. Thus $\mathcal{P}(\mathbf{Z})$ has characteristic 2. ■

Theorem 6.18 ■ Integral Domains, \mathbf{Z} , and \mathbf{Z}_p

An integral domain with characteristic zero contains a subring that is isomorphic to \mathbf{Z} , and an integral domain with positive characteristic p contains a subring that is isomorphic to \mathbf{Z}_p .

Proof Let D be an integral domain with unity e . Define the mapping $\theta: \mathbf{Z} \rightarrow D$ by

$$\theta(n) = ne$$

for each $n \in \mathbf{Z}$. Since

$$\theta(m+n) = (m+n)e = me + ne = \theta(m) + \theta(n)$$

and

$$\theta(mn) = (mn)e = mne^2 = (me)(ne) = \theta(m)\theta(n),$$

θ is a homomorphism from \mathbf{Z} to D . By Theorem 6.9a, $\theta(\mathbf{Z})$ is a subring of D .

$r \Rightarrow s$ Suppose D has characteristic zero. Then $ne = 0$ if and only if $n = 0$, and it follows that $\ker \theta = \{0\}$. According to Theorem 6.11, this means that θ is one-to-one and therefore an isomorphism from \mathbf{Z} to the subring $\theta(\mathbf{Z})$ of D .

$u \Rightarrow v$ Suppose now that D has characteristic p . Then p is the additive order of e , and $ne = 0$ if and only if $p|n$, by Theorem 3.17b. In this case, we have $\ker \theta = (p)$, the set of all multiples of p in \mathbf{Z} . By Theorem 6.14, the subring $\theta(\mathbf{Z})$ of D is isomorphic to $\mathbf{Z}/(p) = \mathbf{Z}_p$.

The terms *embedded* and *extension* were introduced in connection with quotient fields in Section 5.3. Stated in these terms, Theorem 6.18 says that any integral domain with characteristic zero has \mathbf{Z} embedded in it, and any integral domain with characteristic p has \mathbf{Z}_p embedded in it.

In Exercise 18 of Section 5.3, a construction was given by which an arbitrary ring can be embedded in a ring with unity. The next theorem is an improvement on that statement.

Theorem 6.19 ■ Embedding a Ring in a Ring with Unity

Any ring R can be embedded in a ring S with unity that has the same characteristic as R .

$u \Rightarrow (v \wedge w)$

Proof If R has characteristic zero, Exercise 18 of Section 5.3 gives a construction whereby R can be embedded in a ring S with unity. To see that the ring S has characteristic zero, we observe that

$$n(1, 0) = (n, 0) = (0, 0)$$

if and only if $n = 0$.

Suppose now that R has characteristic n . We follow the same type of construction as before, with \mathbf{Z} replaced by \mathbf{Z}_n . Let S be the set of all ordered pairs $([m], x)$, where $[m] \in \mathbf{Z}_n$ and $x \in R$. Equality in S is defined by

$$([m], x) = ([k], y) \quad \text{if and only if} \quad [m] = [k] \quad \text{and} \quad x = y.$$

Addition and multiplication are defined by

$$([m], x) + ([k], y) = ([m+k], x+y)$$

and

$$([m], x) \cdot ([k], y) = ([mk], my + kx + xy).$$

It is straightforward to show that S forms an abelian group with respect to addition, the zero element being $([0], 0)$. This is left as an exercise (see Exercise 23 at the end of this section).

The rule for multiplication yields an element of S , but we need to show that this element is unique. To do this, let $([m_1], x_1) = ([m_2], x_2)$ and $([k_1], y_1) = ([k_2], y_2)$. Then $[m_1] = [m_2]$, $x_1 = x_2$, $[k_1] = [k_2]$, and $y_1 = y_2$ from the definition of equality. Using the definition of multiplication and these equalities, we get

$$([m_1], x_1) \cdot ([k_1], y_1) = ([m_1k_1], m_1y_1 + k_1x_1 + x_1y_1)$$

and

$$\begin{aligned} ([m_2], x_2) \cdot ([k_2], y_2) &= ([m_2k_2], m_2y_2 + k_2x_2 + x_2y_2) \\ &= ([m_1k_1], m_2y_1 + k_2x_1 + x_1y_1). \end{aligned}$$

Comparing the results of these two computations, we see that we need

$$m_2y_1 + k_2x_1 = m_1y_1 + k_1x_1$$

to conclude that the results are equal. Now

$$\begin{aligned} [m_1] = [m_2] &\Rightarrow m_2 - m_1 = pn \quad \text{for some } p \in \mathbf{Z} \\ &\Rightarrow m_2 = m_1 + pn. \end{aligned}$$

Therefore,

$$\begin{aligned} m_2y_1 &= (m_1 + pn)y_1 \\ &= m_1y_1 + npy_1 \\ &= m_1y_1, \end{aligned}$$

since py_1 is in R and R has characteristic n . Similarly, $k_2x_1 = k_1x_1$, and we conclude that the product is well-defined.

Verifying that multiplication is associative, we have

$$\begin{aligned} ([m], x)\{([k], y)([r], z)\} &= ([m], x)([kr], kz + ry + yz) \\ &= ([mkr], mkz + mry + myz + krx + kxz \\ &\quad + rxy + xyz) \\ &= ([mk], my + kx + xy) \cdot ([r], z) \\ &= \{([m], x)([k], y)\}([r], z). \end{aligned}$$

The left distributive law follows from

$$\begin{aligned} ([m], x)\{([k], y) + ([r], z)\} &= ([m], x)([k+r], y+z) \\ &= ([mk+mr], my+mz+kx+rx+xy+xz) \\ &= ([mk], my+kx+xy) + ([mr], mz+rx+xz) \\ &= ([m], x)([k], y) + ([m], x)([r], z). \end{aligned}$$

The verification of the right distributive law is similar to this and is left as an exercise.

The argument up to this point shows that S is a ring. Since each of \mathbf{Z}_n and R has characteristic n ,

$$n([m], x) = (n[m], nx) = ([0], 0)$$

for all $([m], x)$ in S , and n is the least positive integer for which this is true. Thus S has characteristic n .

Consider now the mapping $\theta: R \rightarrow S$ defined by $\theta(x) = ([0], x)$ for all $x \in R$. Since

$$\theta(x) = \theta(y) \Leftrightarrow ([0], x) = ([0], y) \Leftrightarrow x = y,$$

θ is a one-to-one correspondence from R to $\theta(R)$. Now

$$\theta(x+y) = ([0], x+y) = ([0], x) + ([0], y) = \theta(x) + \theta(y)$$

and

$$\theta(xy) = ([0], xy) = ([0], x)([0], y) = \theta(x)\theta(y),$$

so θ is an isomorphism from R to $\theta(R)$, and $\theta(R)$ is a subring of S by Theorem 6.9a. This shows that R is embedded in S .

Exercises 6.3

True or False

Label each of the following statements as either true or false.

1. The characteristic of a ring R is the positive integer n such that $nx = 0$ for all x in R .
 2. The characteristic of a ring R is the smallest positive integer n such that $nx = 0$ for some x in R .
 3. The characteristic of a ring R is zero if $n = 0$ is the only integer such that $nx = 0$ for all x in R .
 4. If a ring R has characteristic zero, then R must have an infinite number of elements.
 5. If a ring R has an infinite number of elements, then R must have characteristic zero.
-

Exercises

1. Find the characteristic of each of the following rings:

- | | | |
|----------------------|------------------------|------------------------|
| a. E | b. Q | c. $M_2(\mathbf{Z})$ |
| d. $M_2(\mathbf{R})$ | e. $M_2(\mathbf{Z}_2)$ | f. $M_2(\mathbf{Z}_3)$ |

- Sec. 5.1, #47 ➤
2. Find the characteristic of the following rings. ($R \oplus S$ is defined in Exercise 47 of Section 5.1.)
- | | | |
|---------------------------------------|---------------------------------------|---------------------------------------|
| a. $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ | b. $\mathbf{Z}_3 \oplus \mathbf{Z}_3$ | c. $\mathbf{Z}_2 \oplus \mathbf{Z}_3$ |
| d. $\mathbf{Z}_2 \oplus \mathbf{Z}_4$ | e. $\mathbf{Z}_4 \oplus \mathbf{Z}_6$ | |
3. Let D be an integral domain with positive characteristic. Prove that all nonzero elements of D have the same additive order.
 4. Show by example that the statement in Exercise 3 is no longer true if “an integral domain” is replaced by “a ring.”
 5. Let R be a ring with unity of characteristic $m > 0$. Prove that $k \cdot e = 0$ if and only if m divides k .
- Sec. 5.1, #47 ➤
6. Suppose that R and S are rings with positive characteristics m and n , respectively. If k is the least common multiple of m and n , prove that $R \oplus S$ has characteristic k .
- Sec. 5.1, #47 ➤
7. Prove that if both R and S in Exercise 6 are integral domains, then $R \oplus S$ has characteristic mn if $m \neq n$.
 8. Prove that the characteristic of a field is either 0 or a prime.
 9. Let D be an integral domain with four elements, $D = \{0, e, a, b\}$, where e is the unity.
 - a. Prove that D has characteristic 2.
 - b. Construct an addition table for D .
 10. Let R be a commutative ring with characteristic 2. Show that each of the following are true for all $x, y \in R$.

a. $(x + y)^2 = x^2 + y^2$	b. $(x + y)^4 = x^4 + y^4$
----------------------------	----------------------------
 11. a. Give an example of a ring R of characteristic 4, and elements x, y in R such that $(x + y)^4 \neq x^4 + y^4$.
 b. Give an example of a noncommutative ring R with characteristic 4, and elements x, y in R such that $(x + y)^4 \neq x^4 + y^4$.
- Sec. 2.2, #23 ➤
12. Let R be a commutative ring with prime characteristic p . Prove, for any x, y in R , that
- $$(x + y)^{p^n} = x^{p^n} + y^{p^n}$$
- for every positive integer n .
13. Prove that \mathbf{Z}_n has a nonzero element whose additive order is less than n if and only if n is not a prime integer.
 14. Let R be a ring with more than one element that has no zero divisors. Prove that the characteristic of R is either zero or a prime integer.
 15. In a commutative ring R of characteristic 2, prove that the idempotent elements form a subring of R .
- Sec. 5.1, #50 ➤
16. A **Boolean ring** is a ring in which all elements x satisfy $x^2 = x$. Prove that every Boolean ring has characteristic 2.

17. Suppose R is a ring with positive characteristic n . Prove that if I is any ideal of R , then n is a multiple of the characteristic of I .
18. If F is a field with positive characteristic p , prove that the set
- $$\{0e = 0, e, 2e, 3e, \dots, (p - 1)e\}$$
- of multiples of the unity e forms a subfield of F .
19. If p is a positive prime integer, prove that any field with p elements is isomorphic to \mathbf{Z}_p .
20. Let I be the set of all elements of a ring R that have finite additive order. Prove that I is an ideal of R .
21. Prove that if a ring R has a finite number of elements, then the characteristic of R is a positive integer.
22. Let R be a ring with a finite number n of elements. Show that the characteristic of R divides n .
23. As in the proof of Theorem 6.19, let $S = \{([m], x) | [m] \in \mathbf{Z}_n \text{ and } x \in R\}$. Prove that S forms an abelian group with respect to addition.
24. With S as in Exercise 23, prove that the right distributive law holds in S .
25. With S as in Exercise 23, prove that the set $R' = \{([0], x) | x \in R\}$ is an ideal of S .
26. Prove that every ordered integral domain has characteristic zero.

6.4

Maximal Ideals (Optional)

We conclude this chapter with a brief study of certain ideals that yield very special quotient rings. We are interested primarily in commutative rings R with unity, and we consider the question of when a quotient ring R/I is a field. (The question of when R/I is an integral domain is treated very briefly in the exercises for this section.)

Definition 6.20 ■ Maximal Ideal

Let M be an ideal of the commutative ring R . Then M is a **maximal ideal** of R if M is not a proper subset[†] of any ideal except R itself.

Thus an ideal M is a maximal ideal of R if and only if $M \subset I \subseteq R$ where I is an ideal, implies $I = R$.

Example 1 Consider the commutative ring $R = \mathbf{Z}$. According to Theorem 6.3, every ideal of \mathbf{Z} is a principal ideal (n) . We shall show that if $n \neq 1$, then (n) is a maximal ideal of \mathbf{Z} if and only if n is a prime.

[†]The term *proper subset* is defined in Definition 1.3.

Suppose first that $n = p$, where p is a prime integer, and let I be an ideal of \mathbf{Z} such that $(p) \subset I \subseteq \mathbf{Z}$. Then there exists an integer k in I such that $k \notin (p)$. That is, k is not a multiple of p . Since p is a prime, this implies that k and p are relatively prime and there exist integers u and v such that

$$1 = uk + vp.$$

Now $uk \in I$, since $k \in I$. We also have $vp \in I$, since $p \in I$. Therefore, $uk + vp = 1$ is in I , since I is an ideal. But $1 \in I$ implies immediately that $I = \mathbf{Z}$, and this proves that (p) is a maximal ideal if p is a prime.

Suppose now that n is not a prime integer. Since $n \neq 1$, there are integers a and b such that

$$n = ab \quad \text{where} \quad 1 < a < n \quad \text{and} \quad 1 < b < n.$$

Consider the ideal $I = (a)$. We have $(n) \subset I$, since $a < n$. Also, we have $I \subset \mathbf{Z}$, since $1 < a$. Thus $(n) \subset I \subset \mathbf{Z}$, and (n) is not a maximal ideal if n is not a prime. ■

Example 2 Example 1 shows that the ideal (4) is not maximal in \mathbf{Z} . However, (4) is a maximal ideal of the ring \mathbf{E} of all even integers. To see that this is true, let I be an ideal of \mathbf{E} such that $(4) \subset I \subseteq \mathbf{E}$. Let x be any element of I that is not in (4) . Then x has the form

$$x = 4k + 2 = 2(2k + 1),$$

where $k \in \mathbf{Z}$. Since I is an ideal,

$$x \in I \quad \text{and} \quad 4k \in I \Rightarrow x - 4k = 2 \in I.$$

But $2 \in I$ implies $I = \mathbf{E}$. Thus (4) is a maximal ideal of \mathbf{E} . ■

The importance of maximal ideals is evident from the result of the following theorem.

Theorem 6.21 ■ Quotient Rings That Are Fields

Let R be a commutative ring with unity, and let M be an ideal of R . Then R/M is a field if and only if M is a maximal ideal of R .

Proof Let R be a commutative ring with unity e , and let M be an ideal of R . It follows immediately from Theorem 6.4 that R/M is a commutative ring with unity $e + M$. Thus R/M is a field if and only if every nonzero element of R/M has a multiplicative inverse in R/M .

$p \Leftarrow q$ Assume first that M is a maximal ideal, and let $a + M$ be a nonzero element of R/M . That is, $a + M \neq M$ and $a \notin M$. Let

$$I = \{ar + m \mid r \in R, m \in M\}.$$

It is clear that each element $a \cdot 0 + m = m$ of M is in I and that $a = ae + 0$ is in I but not in M . Thus $M \subset I$. We shall show that I is an ideal of R .

Let $x = ar_1 + m_1$ and $y = ar_2 + m_2$ be arbitrary elements of I with $r_i \in R$ and $m_i \in M$. Then

$$x + y = a(r_1 + r_2) + (m_1 + m_2),$$

where $r_1 + r_2 \in R$ and $m_1 + m_2 \in M$, since M is an ideal. Thus $x + y \in I$. Also,

$$-x = a(-r_1) + (-m_1)$$

is in I , since $-r_1 \in R$ and $-m_1 \in M$. For any element r of R ,

$$rx = xr = a(r_1r) + (m_1r)$$

is in I , since $r_1r \in R$ and $m_1r \in M$. Thus I is an ideal of R .

Since M is a maximal ideal and $M \subset I$, it must be true that $I = R$. Therefore, there exist $r \in R$ and $m \in M$ such that

$$ar + m = e.$$

Hence

$$\begin{aligned} e + M &= (ar + m) + M \\ &= ar + M \quad \text{since } m \in M \\ &= (a + M)(r + M), \end{aligned}$$

and this means that $r + M$ is the multiplicative inverse of $a + M$ in R/M . We have thus shown that R/M is a field if M is a maximal ideal.

$p \Rightarrow q$ Assume now that R/M is a field, and let I be an ideal of R such that $M \subset I \subseteq R$. Since $M \subset I$, there exists an element $a \in I$ such that $a \notin M$.

We shall show that $I = R$. To this end, let b be an arbitrary element of R . Since R/M is a field and $a + M$ is not zero in R/M , there exists[†] an element $x + M$ in R/M such that

$$(a + M)(x + M) = b + M$$

or

$$ax + M = b + M.$$

Therefore, $ax - b = m$ for some $m \in M$, and

$$b = ax - m.$$

Now $ax \in I$, since $a \in I$, $x \in R$, and I is an ideal of R . Also, $m \in I$ since $M \subset I$. Hence $b = ax - m \in I$. Since b was an arbitrary element of R , we have proved that $R \subseteq I$, and therefore, $I = R$. It follows that M is a maximal ideal of R .

Example 3 We showed in Example 1 of this section that (n) is a maximal ideal of \mathbf{Z} if and only if n is a prime. It follows from Theorem 6.21 that $\mathbf{Z}/(n)$ is a field if and only if n is a prime. However, this fact is not new to us. In connection with Example 5 of Section 6.1, we saw that \mathbf{Z}_n was the same as $\mathbf{Z}/(n)$, and we know from Corollary 5.20 that \mathbf{Z}_n is a field if and only if n is a prime. ■

[†]See Exercise 23 of Section 5.2.

Example 4 We saw in Example 2 of this section that (4) is a maximal ideal of the ring \mathbf{E} of all even integers. The distinct elements of the quotient ring $\mathbf{E}/(4)$ are given by

$$(4) = \{\dots, -8, -4, 0, 4, 8, \dots\}$$

$$2 + (4) = \{\dots, -6, -2, 2, 6, 10, \dots\}.$$

Now $\mathbf{E}/(4)$ is not a field, since $2 + (4)$ is not zero in $\mathbf{E}/(4)$, but

$$[2 + (4)][2 + (4)] = 4 + (4) = (4),$$

and (4) is the zero in $\mathbf{E}/(4)$. At first glance, this seems to contradict Theorem 6.21. However, \mathbf{E} does not have the unity that is required in the hypothesis of Theorem 6.21. ■

Exercises 6.4

True or False

Label each of the following statements as either true or false.

1. The only ideal of a ring R that properly contains a maximal ideal is the trivial ideal R .
2. Only one maximal ideal exists for a given ring R .

Exercises

1. According to part a of Example 3 in Section 5.1, the set

$$R = \{m + n\sqrt{2} \mid m \in \mathbf{Z}, n \in \mathbf{Z}\}$$

is a ring. Assume that the set

$$I = \{a + b\sqrt{2} \mid a \in \mathbf{E}, b \in \mathbf{E}\}$$

is an ideal of R , and show that I is not a maximal ideal of R .

Sec. 6.1, #27 ➤

2. Let R be as in Exercise 1, and show that the principal ideal

$$I = (\sqrt{2}) = \{2n + m\sqrt{2} \mid n \in \mathbf{Z}, m \in \mathbf{Z}\}$$

is a maximal ideal of R .

Sec. 6.1, #27 ➤

3. Show that the ideal $I = (6)$ is a maximal ideal of \mathbf{E} .

Sec. 6.1, #27 ➤

4. Show that the ideal $I = (10)$ is a maximal ideal of \mathbf{E} .

5. Let R and I be as in Exercise 1, and write out the distinct elements of R/I .

6. Let R and I be as in Exercise 2, and write out the distinct elements of R/I .

7. With I as in Exercise 3, write out the distinct elements of \mathbf{E}/I .

8. With I as in Exercise 4, write out the distinct elements of \mathbf{E}/I .

9. Find all maximal ideals of \mathbf{Z}_{12} .

10. Find all maximal ideals of \mathbf{Z}_{18} .

Sec. 5.1, #2f ➤ 11. Let R be the ring of Gaussian integers $\{m + ni \mid m, n \in \mathbf{Z}\}$. Let

$$M = \{a + bi \mid 3 \text{ divides } a \text{ and } 3 \text{ divides } b\}.$$

- a. Show that M is an ideal of R .
- b. Show that M is a maximal ideal of R .

Sec. 5.1, #2f ➤ 12. Let R be the ring of Gaussian integers as in Exercise 11, and let

$$I = \{a + bi \mid 2 \text{ divides } a \text{ and } 2 \text{ divides } b\}.$$

- a. Show that I is an ideal of R .
- b. Show that I is not a maximal ideal of R .
- 13. An ideal I of a commutative ring R is a **prime ideal** if $I \neq R$ and if $ab \in I$ implies either $a \in I$ or $b \in I$. Let R be a commutative ring with unity, and suppose that I is an ideal of R such that $I \neq R$ and $I \neq \{0\}$. Prove that R/I is an integral domain if and only if I is a prime ideal.
- 14. Prove that for $n \neq 1$ and $(n) \neq \{0\}$, an ideal (n) of \mathbf{Z} is a prime ideal if and only if n is a prime integer.
- 15. Show that the ideal I in Exercise 1 is not a prime ideal of R .

Sec. 6.1, #27 ➤ 16. Show that the ideal (4) of \mathbf{E} is not a prime ideal of \mathbf{E} .

Sec. 6.1, #27 ➤ 17. Show that the ideal (6) in Exercise 3 is a prime ideal of \mathbf{E} .

18. Show that the ideal I in Exercise 2 is a prime ideal of R .

Sec. 6.1, #27 ➤ 19. Show that (10) is a prime ideal of \mathbf{E} .

Sec. 6.1, #27 ➤ 20. Show that (14) is a prime ideal of \mathbf{E} .

21. Find all prime ideals of \mathbf{Z}_{12} .

22. Find all prime ideals of \mathbf{Z}_{18} .

23. Give an example of two prime ideals such that their intersection is not prime.

Sec. 5.1, #47 ➤ 24. Show that $\mathbf{Z} \oplus \mathbf{E}$ is a maximal ideal of $\mathbf{Z} \oplus \mathbf{Z}$.

Sec. 5.1, #47 ➤ 25. Show that $\mathbf{Z} \oplus \{0\}$ is a prime ideal of $\mathbf{Z} \oplus \mathbf{Z}$ but is not a maximal ideal of $\mathbf{Z} \oplus \mathbf{Z}$.

26. a. Let $R = M_2(\mathbf{R})$, and $M = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right\}$. Show that M and $M_2(\mathbf{R})$ are the only ideals of $M_2(\mathbf{R})$ and hence M is a maximal ideal.

b. Show that R/M is not a field. Hence Theorem 6.21 is not true if the condition that R is commutative is removed.

27. If R is a commutative ring with unity, prove that any maximal ideal of R is also a prime ideal.

28. If R is a finite commutative ring with unity, prove that every prime ideal of R is a maximal ideal of R .

Key Words and Phrases

characteristic of a ring, 313
 epimorphism, 303
 Fundamental Theorem of Ring Homomorphisms, 309
 homomorphic image, 303

ideal, 293
 isomorphism, 303
 kernel, 305
 maximal ideal, 319
 prime ideal, 323

principal ideal, 296, 302
 quotient ring, 298
 ring homomorphism, 303
 trivial ideals, 293



Hulton Archive/Getty Images

A Pioneer in Mathematics **Amalie Emmy Noether (1882–1935)**

Amalie Emmy Noether, born on March 23, 1882, in Erlangen, Germany, is considered the foremost female mathematician up to her time. She overcame numerous obstacles to receive her education and to be permitted to work as a mathematician in a university environment. Yet her contributions revolutionized abstract algebra and subsequently influenced mathematics as a whole.

Even though university policy stated that admission of women would “overthrow all academic order,”[†] in 1900, Noether and one other woman were given special permission to audit classes at the

University of Erlangen along with one thousand regularly enrolled male students. It wasn’t until 1904 that Noether was allowed to enroll formally and enjoy the same privileges as her male counterparts. Three years later, she completed her doctoral dissertation.

Between 1908 and 1915, Noether was allowed only to substitute teach at Erlangen whenever her father was ill. In 1915, she was brought to the University of Göttingen by David Hilbert (1862–1943) to help in his study of the mathematics involved in the general theory of relativity. Hilbert tried to secure a teaching position for Noether but met strong opposition from the faculty to his request to hire a woman. According to David M. Burton, Hilbert, in a faculty senate meeting held to discuss her appointment, exploded in frustration, “I do not see that the sex of the candidate is an argument against her admission as a Privatdozent. After all we are a university, not a bathhouse.” Her appointment was voted down, but Hilbert allowed her to lecture in courses that were listed under his own name.

At Göttingen, Noether eventually became a lecturer in algebra and earned a modest salary. Göttingen was an international center of mathematics during this time. From her students, the “Noether boys,” came some of the brightest mathematical talents of the era.

Noether, a Jew, was forced to leave Germany in 1933 when Hitler came into power. She fled to the United States, where she accepted a position as visiting professor at Bryn Mawr College in Pennsylvania. She also worked at the Institute for Advanced Study in Princeton, New Jersey. Eighteen months later, at the height of her creative career, she died unexpectedly after an operation.

[†]David M. Burton, *Abstract Algebra* (Cincinnati: William C. Brown, 1988), p. 242.

Real and Complex Numbers

■ Introduction

The material in this chapter is included for the benefit of those who would not see it in some other course. However, it may be skipped by some instructors. It is possible to cover Chapter 8 before this one, and some instructors use this option.

7.1

The Field of Real Numbers

At this point it is possible to fit some of the familiar number systems into the structures developed in the preceding chapters.

In Theorem 5.35, the ring \mathbf{Z} of all integers was characterized as an ordered integral domain in which the set of positive elements is well-ordered. By “characterized,” we mean that any ordered integral domain in which the set of positive elements is well-ordered must be isomorphic to the ring \mathbf{Z} of all integers.

At the end of Section 5.3 we noted that the construction of the rational numbers from \mathbf{Z} is a special case of the procedure described in that section. That is, the set \mathbf{Q} of all rational numbers is the quotient field of \mathbf{Z} and therefore, is the smallest field that contains \mathbf{Z} . From a more abstract point of view, the field of rational numbers can be characterized as the smallest ordered field. That is, any ordered field must contain a subfield that is isomorphic to \mathbf{Q} . (See Exercises 22–24 at the end of this section.)

The main goal of this section is to present a similar characterization for the field of real numbers. The following definition is essential.

Definition 7.1 ■ Upper Bound, Least Upper Bound

Let S be a nonempty subset of an ordered field F . An element u of F is an **upper bound** of S if $u \geq x$ for all $x \in S$. An element u of F is a **least upper bound** of S if these conditions are satisfied:

1. u is an upper bound of S .
2. If $b \in F$ is an upper bound of S , then $b \geq u$.

The phrase *least upper bound* is abbreviated l.u.b.

Example 1 Let $F = \mathbf{Q}$ be the field of rational numbers, and let S be the set of all negative rational numbers.

If a is any negative rational number, then there exists $b \in \mathbf{Q}$ such that $0 > b > a$, by Exercise 13 of Section 5.4. Thus no negative number is an upper bound of S . However, any positive rational number u is an upper bound of S , since

$$u > 0 > x \quad \text{for all } x \in S.$$

The rational number 0 is also an upper bound of S , since $0 > x$ for all $x \in S$. In fact, 0 is a least upper bound of S in \mathbf{Q} . ■

If $u \in F$ and $v \in F$ are both least upper bounds of the nonempty subset S of an ordered field F , then the second condition in Definition 7.1 requires both $v \geq u$ and $u \geq v$. Therefore, $u = v$ and the least upper bound of S in F is unique whenever it exists.

Later we shall exhibit a nonempty subset of \mathbf{Q} that has an upper bound in \mathbf{Q} but does not have a least upper bound in \mathbf{Q} . The following theorem will be needed.

Theorem 7.2 ■ $\sqrt{2}$ Is Not Rational

There is no rational number x such that $x^2 = 2$.

Contradiction

Proof Assume that the theorem is false. That is, assume a rational number x exists such that $x^2 = 2$. We may assume, without loss of generality, that $x = p/q$ is expressed in *lowest terms* as a quotient of integers p and q . That is,

$$\left(\frac{p}{q}\right)^2 = 2$$

with 1 as the greatest common divisor of p and q . This implies that

$$p^2 = 2q^2.$$

Hence 2 divides p^2 , and since 2 is a prime, this implies that 2 divides p , by Theorem 2.16. Let $p = 2r$, where $r \in \mathbf{Z}$. Then we have

$$(2r)^2 = 2q^2$$

$$4r^2 = 2q^2$$

and therefore,

$$2r^2 = q^2.$$

This implies, however, that 2 divides q , by another application of Theorem 2.16. Thus 2 is a common divisor of p and q , and we have a contradiction to the fact that 1 is the greatest common divisor of p and q . This contradiction establishes the theorem.

Example 2 Let

$$S = \{x \in \mathbf{Q} \mid x > 0 \text{ and } x^2 \leq 2\}.$$

We shall show that S is a nonempty subset of \mathbf{Q} that has an upper bound in \mathbf{Q} but does not have a least upper bound (l.u.b.) in \mathbf{Q} .

The set S is nonempty since 1 is in S . The rational number 3 is an upper bound of S in \mathbf{Q} since $x \geq 3$ requires $x^2 \geq 9$ by Exercise 2c of Section 5.4.

It is not so easy to show that S does not have a l.u.b. in \mathbf{Q} . As a start, we shall prove the following two statements for *positive* $u \in \mathbf{Q}$:

1. If u is not an upper bound of S , then $u^2 < 2$.
2. If $u^2 < 2$, then u is not an upper bound of S .

Consider statement 1. If $u \in \mathbf{Q}$ is not an upper bound of S , then there exists $x \in S$ such that $0 < u < x$. By Exercise 2c of Section 5.4, this implies that $u^2 < x^2$. Since $x^2 \leq 2$ for all $x \in S$, we have $u^2 < 2$.

To prove statement 2, suppose that $u \in \mathbf{Q}$ is positive and $u^2 < 2$. Then $\frac{2-u^2}{2u+1}$ is a positive rational number. By Exercise 13 of Section 5.4, there exists a rational number d such that

$$0 < d < \min \left\{ 1, \frac{2-u^2}{2u+1} \right\},$$

where $\min \left\{ 1, \frac{2-u^2}{2u+1} \right\}$ denotes the smaller of the two numbers in braces. If we now put $v = u + d$, then v is a positive rational number, $v > u$, and

$$\begin{aligned} v^2 &= u^2 + 2ud + d^2 \\ &< u^2 + 2ud + d && \text{since } 0 < d < 1 \text{ implies } 0 < d^2 < d \\ &= u^2 + (2u+1)d \\ &< u^2 + (2u+1) \cdot \frac{2-u^2}{2u+1} && \text{since } d < \frac{2-u^2}{2u+1} \\ &= 2. \end{aligned}$$

Thus v is an element of S such that $v > u$, and hence u is not an upper bound of S .

Having established statements 1 and 2, we may combine them with Theorem 7.2 and obtain the following statement:

3. A positive $u \in \mathbf{Q}$ is an upper bound of S if and only if $u^2 > 2$.

With this fact at hand, we can now show that S does not have a l.u.b. in \mathbf{Q} .

Suppose $u \in \mathbf{Q}$ is an upper bound of S . Then u is positive, since all elements of S are positive, and $u^2 > 2$ by statement 3. Let

$$\begin{aligned} w &= u - \frac{u^2 - 2}{2u} \\ &= \frac{u^2 + 2}{2u} \\ &= \frac{u}{2} + \frac{1}{u}. \end{aligned}$$

Then w is a positive rational number. We also have $w < u$, since $\frac{u^2 - 2}{2u}$ is positive. Now

$$\begin{aligned} w^2 &= \left(u - \frac{u^2 - 2}{2u}\right)^2 \\ &= u^2 - (u^2 - 2) + \left(\frac{u^2 - 2}{2u}\right)^2 \\ &= 2 + \left(\frac{u^2 - 2}{2u}\right)^2 \\ &> 2, \end{aligned}$$

so w is an upper bound of S by statement 3. Since $w < u$, we have that u is not a least upper bound of S . Since u was an arbitrary upper bound of S in \mathbf{Q} , this proves that S does not have a l.u.b. in \mathbf{Q} . ■

Example 2 establishes a very significant deficiency in the field \mathbf{Q} of rational numbers: Some nonempty sets of rational numbers have an upper bound in \mathbf{Q} but fail to have a least upper bound in \mathbf{Q} . The next definition gives a designation for those ordered fields that do not have this deficiency.

Definition 7.3 ■ Complete Ordered Field

Let F be an ordered field. Then F is **complete** if every nonempty subset of F that has an upper bound in F has a least upper bound in F .

The basic difference between the field of rational numbers and the field of real numbers is that the real number field is complete. It is possible to construct the field of real numbers from the field of rational numbers, but this construction is too lengthy and difficult to be included here. It is more properly a part of that area of mathematics known as *analysis*. The method of construction most commonly used is one that is credited to Richard Dedekind (1831–1916) and utilizes what are called *Dedekind cuts*. In our treatment, we shall assume the validity of the following theorem.

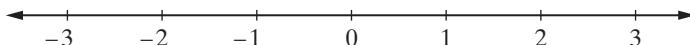
Theorem 7.4 ■ The Field of Real Numbers

There exists a field \mathbf{R} , called the **field of real numbers**, that is a complete ordered field. Any complete ordered field F has the following properties:

- a. F is isomorphic to \mathbf{R} .
 - b. F contains a subfield that is isomorphic to the field \mathbf{Q} of **rational numbers**, and the ordering in F is an extension of the ordering in this subfield.
-

The set of all real numbers may be represented geometrically by setting up a one-to-one correspondence between real numbers and the points on a straight line. To begin, we select a point on a horizontal line, designate it as the *origin*, and let this point correspond to

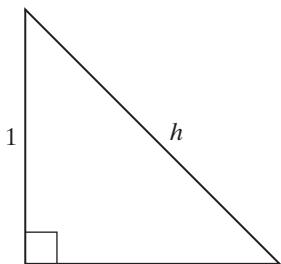
the number 0. A second point is now chosen to the right of the origin, and we let this point correspond to the number 1. The distance between the two points corresponding to 0 and 1 is now taken as one unit of measure. Points on the line located successively one unit farther to the right are made to correspond to the positive integers 2, 3, 4, . . . in succession. With the same unit of measure and beginning at the origin, points on the line located successively one unit farther to the left are made to correspond to the negative integers $-1, -2, -3, \dots$ (see Figure 7.1). This sets up a one-to-one correspondence between the set \mathbf{Z} of all integers and some of the points on the line.



■ **Figure 7.1**

Points on the line that correspond to nonintegral rational numbers are now located by using distances proportional to their expressions as quotients a/b of integers a and b and by using directions to the right for positive numbers and to the left for negative numbers. For example, the point corresponding to $\frac{3}{2}$ is located midway between the points that correspond to 1 and 2, whereas the point corresponding to $-\frac{3}{2}$ is located midway between those that correspond to -1 and -2 . In this manner, a one-to-one correspondence is established between the set \mathbf{Q} of rational numbers and a subset of the points on the line.

It is not very difficult to demonstrate that there are points on the line that do not correspond to any rational number. This can be done by considering a right triangle with each leg one unit in length (see Figure 7.2). By the Pythagorean Theorem, the length h of the hypotenuse of the triangle in Figure 7.2 satisfies the equation $h^2 = 2$. There is a point on the line located at a distance h units to the right of the origin, but by Theorem 7.2, this point cannot correspond to a rational number.



■ **Figure 7.2**

The foregoing demonstration shows that there are gaps in the rational numbers, even though any two distinct rational numbers have another rational number located between them (see Exercise 13 of Section 5.4). We *assume* now that the one-to-one correspondence that we have set up between the rational numbers and points on the line can be extended to the set of all real numbers and the set of all points on the line. The points that do not correspond to rational numbers are assumed to correspond to real numbers that are

not rational—that is, to **irrational** numbers. For example, the discussion in the preceding paragraph located the point that corresponds to the irrational number $h = \sqrt{2}$.

One more aspect of the real numbers is worthy of mention: the **decimal representation** of real numbers. Here we assume that each real number can be represented by a decimal expression that either terminates, as does

$$\frac{9}{8} = 1.125,$$

or continues without end, as do the repeating decimal[†]

$$\frac{14}{11} = 1.272727 \cdots = 1.\overline{27}$$

and the nonrepeating decimal

$$\sqrt{2} = 1.41421356 \cdots.$$

The decimal expression for a rational number a/b may be found by long division. For example, for the rational number $\frac{14}{11}$, long division yields the following.

$$\begin{array}{r} 1.27 \\ 11 \overline{)14.00} \\ 11 \\ \hline 30 \\ 22 \\ \hline 80 \\ 77 \\ \hline 3 \end{array}$$

The repetition of the remainder 3 at this point makes it clear that we have the repeating decimal expression

$$\frac{14}{11} = 1.272727 \cdots = 1.\overline{27}.$$

A terminating decimal expression may be regarded as a repeating pattern where zeros repeat endlessly. For example,

$$\frac{9}{8} = 1.125000 \cdots = 1.125\bar{0}.$$

With this point of view, the decimal expression for any rational number a/b will always have a repeating pattern. This can be seen from the long-division algorithm: Each remainder satisfies $0 \leq r < b$, so there are only b distinct possibilities for the remainders, and the expression starts repeating whenever a remainder occurs for the second time.

[†]The bar above 27 indicates that the digits 27 repeat endlessly.

Rational numbers that have a terminating decimal expression can be represented in another way by changing the range on the remainders in the long division from $0 \leq r < b$ to $0 < r \leq b$. If we perform the long division for $\frac{9}{8}$ in this way, it appears as follows.

$$\begin{array}{r} 1.1249 \\ 8 \overline{)9.0000} \\ 8 \\ \hline 10 \\ 8 \\ \hline 20 \\ 16 \\ \hline 40 \\ 32 \\ \hline 80 \\ 72 \\ \hline 8 \end{array}$$

At this point, the remainder 8 has occurred twice, and the repeating pattern is seen to be

$$\frac{9}{8} = 1.124999 \cdots = 1.12\bar{4}.$$

It is shown in calculus that if $a \neq 0$, then the infinite geometric series

$$\sum_{n=1}^{\infty} ar^{n-1}$$

diverges for $|r| \geq 1$ and converges to $a/(1 - r)$ when $|r| < 1$. Thus every nonterminating repeating decimal expression represents a rational number, since it is the sum of an infinite geometric series with $r = 10^{-k}$, for some positive integer k . The next example illustrates this situation.

Example 3 We shall express $2.\overline{134}$ as a quotient of integers. We have

$$\begin{aligned} 2.\overline{134} &= 2.1343434 \cdots \\ &= 2.1 + 0.034 + 0.00034 + 0.0000034 + \cdots \end{aligned}$$

and the terms $0.034 + 0.00034 + 0.0000034 + \cdots$ form an infinite geometric series with $a = 0.034$ and $r = 10^{-2} = 0.01$. Since $|r| < 1$, this geometric series converges to $0.034/(1 - 0.01)$ and

$$\begin{aligned} 2.\overline{134} &= 2.1 + \sum_{n=1}^{\infty} (0.034)(0.01)^{n-1} \\ &= \frac{21}{10} + \frac{0.034}{1 - 0.01} \\ &= \frac{21}{10} + \frac{0.034}{0.99} \\ &= \frac{21}{10} + \frac{34}{990} \\ &= \frac{2113}{990}. \end{aligned}$$

■

This discussion of decimal representations is not intended to be a rigorous presentation. Its purpose is to make the following remarks appear plausible:

1. Each real number can be represented by a decimal expression.
2. Decimal expressions that repeat or terminate represent rational numbers.
3. Decimal expressions that do not repeat and do not terminate represent irrational numbers.

Exercises 7.1

True or False

Label each of the following statements as either true or false.

1. Every least upper bound of a nonempty set S is an upper bound.
2. Every upper bound of a nonempty set S is a least upper bound.
3. The least upper bound of a nonempty set S is unique.
4. Every upper bound of a nonempty set S must be an element of S .
5. If a nonempty set S contains an upper bound, then a least upper bound must exist in S .
6. The field of real numbers is complete.
7. The field of rational numbers is complete.
8. Every decimal representation of a real number that terminates represents a rational number.
9. Every decimal representation of a real number that does not terminate represents an irrational number.

Exercises

Find the decimal representation for each of the numbers in Exercises 1–6.

- | | | |
|--------------------------|--------------------------|---------------------------|
| 1. $\frac{5}{9}$ | 2. $\frac{7}{33}$ | 3. $\frac{80}{81}$ |
| 4. $\frac{16}{7}$ | 5. $\frac{22}{7}$ | 6. $\frac{19}{11}$ |

Express each of the numbers in Exercises 7–12 as a quotient of integers, reduced to lowest terms.

- | | | |
|------------------------------|------------------------------|---------------------------------|
| 7. $3.\bar{4}$ | 8. $1.\bar{6}$ | 9. $0.\bar{1}\bar{2}$ |
| 10. $0.\overline{63}$ | 11. $2.\overline{51}$ | 12. $3.21\overline{321}$ |

- 13.** Prove that $\sqrt{3}$ is irrational. (That is, prove there is no rational number x such that $x^2 = 3$.)
- 14.** Prove that $\sqrt[3]{2}$ is irrational.
- 15.** Prove that if p is a prime integer, then \sqrt{p} is irrational.
- 16.** Prove that if a is rational and b is irrational, then $a + b$ is irrational.
- 17.** Prove that if a is a nonzero rational number and b is irrational, then ab is irrational.
- 18.** Prove that if a is an irrational number, then a^{-1} is an irrational number.

19. Prove that if a is a nonzero rational number and ab is irrational, then b is irrational.
20. Give counterexamples for the following statements.
 - a. If a and b are irrational, then $a + b$ is irrational.
 - b. If a and b are irrational, then ab is irrational.
21. Let S be a nonempty subset of an ordered field F .
 - a. Write definitions for **lower bound** of S and **greatest lower bound** of S .
 - b. Prove that if F is a complete ordered field and the nonempty subset S has a lower bound in F , then S has a greatest lower bound in F .
22. Prove that if F is an ordered field with F^+ as its set of positive elements, then $F^+ \supseteq \{ne | n \in \mathbf{Z}^+\}$, where e denotes the multiplicative identity in F . (*Hint:* See Theorem 5.34 and its proof.)
23. If F is an ordered field, prove that F contains a subring that is isomorphic to \mathbf{Z} . (*Hint:* See Theorem 5.35 and its proof.)
24. Prove that any ordered field must contain a subfield that is isomorphic to the field \mathbf{Q} of rational numbers.
25. If a and b are positive real numbers, prove that there exists a positive integer n such that $na > b$. This property is called the **Archimedean[†] Property** of the real numbers. (*Hint:* If $ma \leq b$ for all $m \in \mathbf{Z}^+$, then b is an upper bound for the set $S = \{ma | m \in \mathbf{Z}^+\}$. Use the completeness property of \mathbf{R} to arrive at a contradiction.)
26. Prove that if a and b are real numbers such that $a > b$, then there exists a rational number m/n such that $a > m/n > b$. (*Hint:* Use Exercise 25 to obtain $n \in \mathbf{Z}^+$ such that $a - b > 1/n$. Then choose m to be the least integer such that $m > nb$. With these choices of m and n , show that $(m - 1)/n \leq b$ and then that $a > m/n > b$.)

7.2

Complex Numbers and Quaternions

The fact that negative real numbers do not have square roots in \mathbf{R} is a serious deficiency of the field of real numbers, but it is one that can be overcome by the introduction of complex numbers.

Although we do not present a characterization of the field of complex numbers until Section 8.4, it is possible to construct the complex numbers from the real numbers. Such a construction is the main purpose of this section.

In our construction, complex numbers appear first as ordered pairs (a, b) and later in the more familiar form $a + bi$. The operations given in the following definition will seem more natural if they are compared with the usual operations on complex numbers in the form $a + bi$.

[†] Archimedes (c. 287 B.C.–c. 212 B.C.) was a Greek mathematician, physicist, engineer, and astronomer. He is regarded as the leading scientist of his time and as one of the greatest mathematicians ever. He is famous for his innovative machine designs, including the screw pump. He is honored with a lunar crater and a lunar mountain range named after him. California adopted his famous *Eureka!* as its state motto.

Definition 7.5 ■ Complex Numbers

Let \mathbf{C} be the set of all ordered pairs (a, b) of real numbers a and b . Equality, addition, and multiplication are defined in \mathbf{C} by

$$\begin{aligned}(a, b) &= (c, d) \quad \text{if and only if} \quad a = c \quad \text{and} \quad b = d \\ (a, b) + (c, d) &= (a + c, b + d) \\ (a, b)(c, d) &= (ac - bd, ad + bc).\end{aligned}$$

The elements of \mathbf{C} are called **complex numbers**.

It is easy to see that the stated rules for addition and multiplication do in fact define binary operations on \mathbf{C} .

Theorem 7.6 ■ The Field of Complex Numbers

With addition and multiplication as given in Definition 7.5, \mathbf{C} is a field. The set of all elements of the form $(a, 0)$ in \mathbf{C} forms a subfield of \mathbf{C} that is isomorphic to the field \mathbf{R} of real numbers.

Proof Closure of \mathbf{C} under addition follows at once from the fact that \mathbf{R} is closed under addition. It is left for the exercises to prove that addition is associative and commutative, that $(0, 0)$ is the additive identity in \mathbf{C} , and that the additive inverse of $(a, b) \in \mathbf{C}$ is $(-a, -b) \in \mathbf{C}$.

Since \mathbf{R} is closed under multiplication and addition, each of $ac - bd$ and $ad + bc$ is in \mathbf{R} whenever (a, b) and (c, d) are in \mathbf{C} . Thus \mathbf{C} is closed under multiplication.

For the remainder of the proof, let (a, b) , (c, d) , and (e, f) represent arbitrary elements of \mathbf{C} . The associative property of multiplication is verified by the following computations:

$$\begin{aligned}(a, b)[(c, d)(e, f)] &= (a, b)(ce - df, cf + de) \\ &= [a(ce - df) - b(cf + de), a(cf + de) + b(ce - df)] \\ &= (ace - adf - bcf - bde, acf + ade + bce - bdf) \\ &= [(ac - bd)e - (ad + bc)f, (ac - bd)f + (ad + bc)e] \\ &= (ac - bd, ad + bc)(e, f) \\ &= [(a, b)(c, d)](e, f).\end{aligned}$$

Before considering the distributive laws, we shall show that multiplication is commutative in \mathbf{C} . This follows from

$$\begin{aligned}(c, d)(a, b) &= (ca - db, cb + da) \\ &= (ca - db, da + cb) \\ &= (ac - bd, ad + bc) \\ &= (a, b)(c, d).\end{aligned}$$

We shall verify the left distributive property and leave the proof of the right distributive property as an exercise:

$$\begin{aligned}
 (a, b)[(c, d) + (e, f)] &= (a, b)(c + e, d + f) \\
 &= [a(c + e) - b(d + f), a(d + f) + b(c + e)] \\
 &= (ac + ae - bd - bf, ad + af + bc + be) \\
 &= (ac - bd, ad + bc) + (ae - bf, af + be) \\
 &= (a, b)(c, d) + (a, b)(e, f).
 \end{aligned}$$

To this point, we have established that \mathbf{C} is a commutative ring.

The computation

$$(1, 0)(a, b) = (1 \cdot a - 0 \cdot b, 1 \cdot b + 0 \cdot a) = (a, b)$$

shows that $(1, 0)$ is a left identity for multiplication in \mathbf{C} . Since multiplication in \mathbf{C} is commutative, it follows that $(1, 0)$ is a nonzero unity in \mathbf{C} .

If $(a, b) \neq (0, 0)$ in \mathbf{C} , then at least one of the real numbers a or b is nonzero, and it follows that $a^2 + b^2$ is a positive real number. Hence

$$\left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right)$$

is an element of \mathbf{C} . The multiplication

$$(a, b) \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right) = \left(\frac{a^2 + b^2}{a^2 + b^2}, \frac{-ab + ba}{a^2 + b^2} \right) = (1, 0)$$

shows that

$$(a, b)^{-1} = \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right),$$

since multiplication is commutative in \mathbf{C} . This completes the proof that \mathbf{C} is a field.

Consider now the set R' that consists of all elements of \mathbf{C} that have the form $(a, 0)$:

$$R' = \{(a, 0) | a \in \mathbf{R}\}.$$

The proof that R' is a subfield of \mathbf{C} is left as an exercise. The mapping $\theta: \mathbf{R} \rightarrow R'$ defined by

$$\theta(a) = (a, 0)$$

is clearly onto, and is one-to-one, since $(a, 0) = (b, 0)$ if and only if $a = b$. For arbitrary a and b in \mathbf{R} ,

$$\begin{aligned}
 \theta(a + b) &= (a + b, 0) \\
 &= (a, 0) + (b, 0) \\
 &= \theta(a) + \theta(b)
 \end{aligned}$$

and

$$\begin{aligned}
 \theta(ab) &= (ab, 0) \\
 &= (a, 0)(b, 0) \\
 &= \theta(a)\theta(b).
 \end{aligned}$$

Thus θ preserves both operations and is an isomorphism from \mathbf{R} to R' .

We shall use the isomorphism θ in the preceding proof to identify $a \in \mathbf{R}$ with $(a, 0)$ in R' . We write a instead of $(a, 0)$ and consider \mathbf{R} to be a subset of \mathbf{C} . The calculation

$$\begin{aligned}(0, 1)(0, 1) &= (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) \\ &= (-1, 0) \\ &= -1\end{aligned}$$

shows that the equation $x^2 = -1$ has a solution $x = (0, 1)$ in \mathbf{C} .

To obtain the customary notation for complex numbers, we define the number i by

$$i = (0, 1).$$

This makes i a number such that $i^2 = -1$. We now note that any $(a, b) \in \mathbf{C}$ can be written in the form

$$\begin{aligned}(a, b) &= (a, 0) + (0, b) \\ &= (a, 0) + b(0, 1) \\ &= a + bi,\end{aligned}$$

and this gives us the familiar form for complex numbers.

Using the field properties freely, we may rewrite the rules for addition and multiplication in \mathbf{C} as follows:

$$\begin{aligned}(a + bi) + (c + di) &= a + c + bi + di \\ &= (a + c) + (b + d)i\end{aligned}$$

and

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

where the last step was obtained by replacing i^2 with -1 .

The fact that $i^2 = -1$ was used in Section 5.4 to prove that it is impossible to impose an order relation on \mathbf{C} . Hence \mathbf{C} is not an ordered field.

It is easy to show that all negative real numbers have square roots in \mathbf{C} . For any positive real number a , the negative real number $-a$ has both \sqrt{ai} and $-\sqrt{ai}$ as square roots, since

$$(\sqrt{ai})^2 = (\sqrt{a})^2 i^2 = a(-1) = -a$$

and

$$(-\sqrt{ai})^2 = (-\sqrt{a})^2 i^2 = a(-1) = -a.$$

We shall see later in this chapter that every nonzero complex number has two distinct square roots in \mathbf{C} .

Example 1 The following results illustrate some calculations with complex numbers.

- a. $(1 + 2i)(3 - 5i) = 3 + 6i - 5i - 10i^2 = 13 + i$
- b. $(2 + 3i)(2 - 3i) = 4 - 9i^2 = 13$
- c. $(-3 + 4i)(3 + 4i) = -9 + 16i^2 = -25$

d. $(1 - i)^2 = 1 - 2i + i^2 = -2i$

e. $i^4 = (i^2)^2 = (-1)^2 = 1$ ■

In connection with part b of Example 1, we note that

$$\begin{aligned}(a + bi)(a - bi) &= a^2 - b^2i^2 \\ &= a^2 + b^2\end{aligned}$$

for any complex number $a + bi$. The number $a^2 + b^2$ is always real, and it is positive if $a + bi$ is nonzero.

Definition 7.7 ■ Conjugate

For any a, b in \mathbf{R} , the **conjugate** of the complex number $a + bi$ is the number $a - bi$. The notation \bar{z} indicates the conjugate of z : If $z = a + bi$ with a and b real, then $\bar{z} = a - bi$.

Using the bar notation of Definition 7.7, we can write

$$\bar{z}\bar{z} = z\bar{z} = a^2 + b^2,$$

and the multiplicative inverse of a nonzero z is given by

$$z^{-1} = \left(\frac{1}{\bar{z}z} \right) \bar{z}.$$

Division of complex numbers may be accomplished by multiplying the numerator and denominator of a quotient by the conjugate of the denominator.

Example 2 We have the following illustrations of division.

a. $\frac{3 + 7i}{2 - 3i} = \frac{3 + 7i}{2 - 3i} \cdot \frac{2 + 3i}{2 + 3i} = \frac{6 + 23i - 21}{4 + 9} = -\frac{15}{13} + \frac{23}{13}i$

b. $\frac{1}{2 + i} = \frac{1}{2 + i} \cdot \frac{2 - i}{2 - i} = \frac{2 - i}{5} = \frac{2}{5} - \frac{1}{5}i$ ■

By using the techniques illustrated in Examples 1 and 2, we can write the result of any calculation involving the field operations with complex numbers in the form $a + bi$, with a and b real numbers. This form is called the **standard form** of the complex number. If $b \neq 0$, the number is called **imaginary**. If $a = 0$ and $b \neq 0$, the number is called **pure imaginary**.

The construction of the complex numbers by use of ordered pairs was first accomplished by Hamilton (see the biographical section at the end of this chapter). Eventually, he was able to use ordered quadruples (x, y, z, w) of real numbers to extend the complex numbers to a larger set that he called the **quaternions**. His quaternions satisfy all the postulates for a field except the requirement that multiplication be commutative. A system with these properties is called a **division ring**, or a **skew field**, and Hamilton's quaternions were the first known example of a division ring.

Example 3 In this example we outline the development of the quaternions as the set

$$H = \{(x, y, z, w) | x, y, z, w \in \mathbf{R}\},$$

with most of the details left as exercises.

Equality and addition are defined in H by

$$(x, y, z, w) = (r, s, t, u) \quad \text{if and only if } x = r, y = s, z = t, \text{ and } w = u;$$

$$(x, y, z, w) + (r, s, t, u) = (x + r, y + s, z + t, w + u).$$

It is easy to see that this addition is a binary operation on H , and $(0, 0, 0, 0)$ in H is the additive identity. Also, each (x, y, z, w) in H has an additive inverse $(-x, -y, -z, -w)$ in H . The proofs that addition is associative and commutative are left as exercises. Thus H forms an abelian group with respect to addition.

When the definition of multiplication in H is presented in the same manner as multiplication of complex numbers in Definition 7.5, it has the following complicated appearance:

$$(x, y, z, w)(r, s, t, u) = (xr - ys - zt - wu, xs + yr + zu - wt, \\ xt - yu + zr + ws, xu + yt - zs + wr).$$

This multiplication is a binary operation on H , and it is easy to verify that $(1, 0, 0, 0)$ is a unity in H . Laborious computations will show that multiplication is associative in H and that both distributive laws hold. These verifications are left as exercises and lead to the conclusion that H is a ring.

At this point, it can be shown that the set

$$R' = \{(a, 0, 0, 0) | a \in \mathbf{R}\}$$

is a field contained in H and that the mapping $\theta: \mathbf{R} \rightarrow R'$ defined by

$$\theta(a) = (a, 0, 0, 0)$$

is an isomorphism. In a manner similar to the identification of a with $(a, 0)$ in \mathbf{C} , we can identify a in \mathbf{R} with $(a, 0, 0, 0)$ in R' and consider \mathbf{R} to be a subring of H .

Some other notational changes can be used to give the elements of H a more natural appearance. We let

$$i = (0, 1, 0, 0), \quad j = (0, 0, 1, 0), \quad \text{and} \quad k = (0, 0, 0, 1).$$

Then an arbitrary element (x, y, z, w) in H can be written as

$$(x, y, z, w) = (x, 0, 0, 0) + (y, 0, 0, 0)i + (z, 0, 0, 0)j + (w, 0, 0, 0)k \\ = x + yi + zj + wk.$$

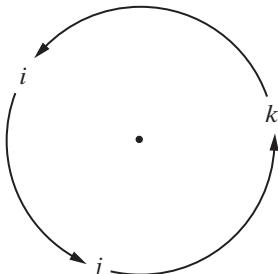
Routine calculations confirm the equations

$$\begin{array}{ll} (-1)^2 = 1 & ij = -ji = k \\ i^2 = j^2 = k^2 = -1 & jk = -kj = i \\ (-1)a = a(-1) = -a & \text{for all } a \in \{\pm 1, \pm i, \pm j, \pm k\} \quad ki = -ik = j. \end{array}$$

In fact, this multiplication agrees with the table constructed for the quaternion group in Exercise 28 of Section 3.1. The circular order of multiplication observed previously is also

valid in H (see Figure 7.3). With a positive (counterclockwise) rotation, the product of two consecutive elements is the third one on the circle, and the sign changes with a negative (clockwise) rotation.

Computations such as $ij = k$ and $ji = -k$ show that multiplication in H is not commutative, and H is not a field.



■ **Figure 7.3**

With the i, j, k notation, H can be written in the form

$$H = \{x + yi + zj + wk \mid x, y, z, w \in \mathbf{R}\},$$

with addition and multiplication appearing as

$$(x + yi + zj + wk) + (r + si + tj + uk) = (x + r) + (y + s)i + (z + t)j + (w + u)k;$$

$$\begin{aligned} (x + yi + zj + wk)(r + si + tj + uk) &= (xr - ys - zt - wu) \\ &\quad + (xs + yr + zu - wt)i \\ &\quad + (xt - yu + zr + ws)j \\ &\quad + (xu + yt - zs + wr)k. \end{aligned}$$

Multiplication can thus be performed by using the distributive laws and other natural ring properties, with two exceptions:

1. Multiplication is not commutative.
2. Products of i, j , or k are simplified using the equations on the preceding page.

The most outstanding feature of H is that each nonzero element has a multiplicative inverse. For each $q = x + yi + zj + wk$ in H , we imitate conjugates in \mathbf{C} and write

$$\bar{q} = x - yi - zj - wk.$$

It is left as an exercise to verify that

$$\bar{q}q = q\bar{q} = x^2 + y^2 + z^2 + w^2.$$

If $q \neq 0$, then $\bar{q}q \neq 0$, and

$$q^{-1} = \left(\frac{1}{\bar{q}q} \right) \bar{q}.$$

Thus H has all the field properties except commutative multiplication. ■

Exercises 7.2

True or False

Label each of the following statements as either true or false.

1. It is possible to impose an order relation on \mathbf{C} , the set of complex numbers.
2. Negative real numbers have two distinct square roots in the field of complex numbers.
3. The inverse of any nonzero complex number can be expressed in terms of its conjugate.
4. The complex numbers form a field.
5. The quaternions form a field.
6. Every field is a division ring.
7. Every division ring is a field.

Exercises

Perform the computations in Exercises 1–12 and express the results in standard form $a + bi$.

1. $(2 - 3i)(-1 + 4i)$
2. $(5 - 3i)(2 - 4i)$
3. i^{15}
4. i^{87}
5. $(2 - i)^3$
6. $i(2 + i)^2$
7. $\frac{1}{2 - i}$
8. $\frac{1}{3 + i}$
9. $\frac{2 - i}{8 - 6i}$
10. $\frac{1 - i}{1 + 3i}$
11. $\frac{5 + 2i}{5 - 2i}$
12. $\frac{4 - 3i}{4 + 3i}$
13. Find two square roots of each given number.
 - a. -9
 - b. -16
 - c. -25
 - d. -36
 - e. -13
 - f. -8
14. With addition as given in Definition 7.5, prove the following statements.
 - a. Addition is associative in \mathbf{C} .
 - b. Addition is commutative in \mathbf{C} .
 - c. $(0, 0)$ is the additive identity in \mathbf{C} .
 - d. The additive inverse of $(a, b) \in \mathbf{C}$ is $(-a, -b) \in \mathbf{C}$.
15. With addition and multiplication as in Definition 7.5, prove that the right distributive property holds in \mathbf{C} .
16. Show that $i^n = i^m$ for all integers n , where $n \equiv m \pmod{4}$.
17. With \mathbf{C} given in Definition 7.5, prove that $R' = \{(a, 0) \mid a \in \mathbf{R}\}$ is a subfield of \mathbf{C} .

18. Let $B = \{bi \mid b \in \mathbf{R}\}$ be a subset of \mathbf{C} with the usual operations of addition and multiplication of complex numbers.

- Prove or disprove that B is an abelian group with respect to addition.
- Prove or disprove that B is a ring.

19. Let θ be the mapping $\theta: \mathbf{C} \rightarrow \mathbf{C}$ defined for $z = a + bi$ in standard form by

$$\theta(z) = a - bi.$$

Prove that θ is a ring isomorphism.

20. It follows from Exercise 19 that $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ and that $\overline{z_1 z_2} = \overline{z_1} \overline{z_2}$ for all z_1, z_2 in \mathbf{C} . Prove the following statements concerning conjugates of complex numbers.

- $\overline{(z)} = z$
- If $z \neq 0$, then $(\overline{z})^{-1} = \overline{(z^{-1})}$.
- $z + \bar{z} \in \mathbf{R}$
- $z = \bar{z}$ if and only if $z \in \mathbf{R}$.
- $\bar{z} = -z$ if and only if z is pure imaginary or $\bar{z} = 0$.

21. a. Show that $x + yi$ satisfies the equation $z^2 = a + bi$ where

$$x = \frac{b}{2y}, \text{ and } y = \pm \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}}.$$

- b. Find two square roots of each of the following complex numbers.

- $3 - 4i$
- $4 + 3i$
- $5 + 12i$
- $-12 + 5i$

22. Assume that $\theta: \mathbf{C} \rightarrow \mathbf{C}$ is an isomorphism and $\theta(a) = a$ for all $a \in \mathbf{R}$. Prove that if θ is not the identity mapping, then $\theta(z) = \bar{z}$ for all $z \in \mathbf{C}$.

23. (See Example 8 of Section 5.1.) Show that the mapping θ defined by

$$\theta(a + bi) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \quad \text{for } a, b \in \mathbf{R}$$

is an isomorphism from \mathbf{C} to a subring of the ring of all 2×2 matrices over \mathbf{R} .

24. With addition as given in Example 3 of this section, prove the following statements.

- Addition is associative in H .
- Addition is commutative in H .

25. Prove that multiplication in the set H of Example 3 has the associative property.

26. With addition and multiplication as defined in Example 3, prove that both distributive laws hold in H .

Exercises 27–31 are stated using the notation in the last paragraph of Example 3.

- Prove that $\overline{(q)} = q$ for all $q \in H$.
- Prove that $\overline{q_1 + q_2} = \overline{q_1} + \overline{q_2}$ for all $q_1, q_2 \in H$.
- Prove that $\overline{q_1 q_2} = \overline{q_2} \overline{q_1}$ for all $q_1, q_2 \in H$.

30. Prove or disprove: $\overline{q_1 q_2} = \overline{q_1} \overline{q_2}$ for all $q_1, q_2 \in H$.
31. Verify that $\bar{q}q = q\bar{q} = x^2 + y^2 + z^2 + w^2$ for arbitrary $q = x + yi + zj + wk$ in H .
32. (See Exercise 31.) For arbitrary $q = x + yi + zj + wk$ in H , we define the **absolute value** of q by $|q| = \sqrt{x^2 + y^2 + z^2 + w^2}$. Verify that $|q_1 q_2| = |q_1| \cdot |q_2|$.
33. Let $q_1, q_2 \in H$. Prove $q_1 q_2 = 0$ implies $q_1 = 0$ or $q_2 = 0$.
34. Show that the equation $x^2 = -1$ has an infinite number of solutions in the quaternions.
35. a. With H as defined in Example 3, prove that the set

$$R' = \{(a, 0, 0, 0) | a \in \mathbf{R}\}$$

is a field that is contained in H .

- b. Prove that the mapping $\theta: \mathbf{R} \rightarrow R'$ defined by $\theta(a) = (a, 0, 0, 0)$ is an isomorphism.

36. Assume that

$$C' = \{(a, b, 0, 0) | a, b \in \mathbf{R}\}$$

is a subring of the quaternions in Example 3 when H is regarded as a set of quadruples. Prove that the mapping $\theta: \mathbf{C} \rightarrow C'$ defined by $\theta(a + bi) = (a, b, 0, 0)$ is an isomorphism from the field \mathbf{C} of complex numbers to C' . (Thus we can consider \mathbf{C} to be a subring of H .)

37. Suppose the mapping f is defined on the set H of quaternions by $f(q) = \bar{q}$ for all $q \in H$. Show that f is one-to-one, onto, and satisfies the following properties.

$$f(q_1 + q_2) = f(q_1) + f(q_2) \text{ and } f(q_1 q_2) = f(q_2) f(q_1) \text{ for all } q_1, q_2 \in H$$

38. Let S be the subset of $M_2(\mathbf{C})$ given by

$$S = \left\{ \begin{bmatrix} x & y \\ -\bar{y} & \bar{x} \end{bmatrix} \mid x, y \in \mathbf{C} \right\}.$$

- a. Prove that S is a subring of $M_2(\mathbf{C})$.

- b. Prove that the mapping $\theta: H \rightarrow S$ defined by

$$\theta(a + bi + cj + dk) = \begin{bmatrix} a + bi & c + di \\ -(c - di) & a - bi \end{bmatrix}$$

is an isomorphism from the ring of quaternions H to S . [Note that $a + bi + cj + dk = (a + bi) + (c + di)j$.]

- Sec. 3.3, #17 ➤ 39. Let K be the group of nonzero quaternions with the operation of multiplication. Show that the center of K is $\{x = x + 0i + 0j + 0k \mid x \in \mathbf{R}, x \neq 0\}$.

- Sec. 5.2, #18 ➤ 40. An element a in a ring R is **idempotent** if $a^2 = a$. Prove that a division ring must contain exactly two idempotent elements.

41. Prove that a finite ring R with unity and no zero divisors is a division ring.

7.3**De Moivre's[†] Theorem and Roots of Complex Numbers**

We have seen that real numbers may be represented geometrically by the points on a straight line. In much the same way, it is possible to represent complex numbers by the points in a plane. We begin with a conventional rectangular coordinate system in the plane (see Figure 7.4). With each complex number $x + yi$ in standard form, we associate the point that has coordinates (x, y) . This association establishes a one-to-one correspondence from the set \mathbf{C} of complex numbers to the set of all points in the plane.

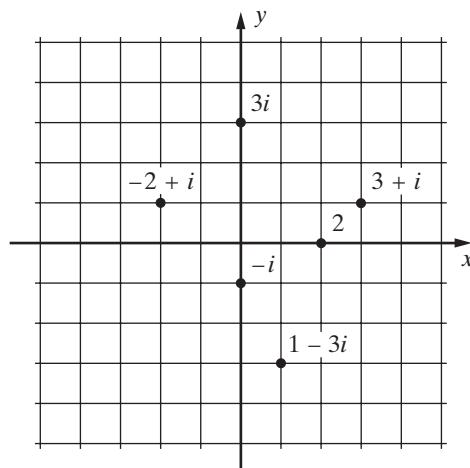
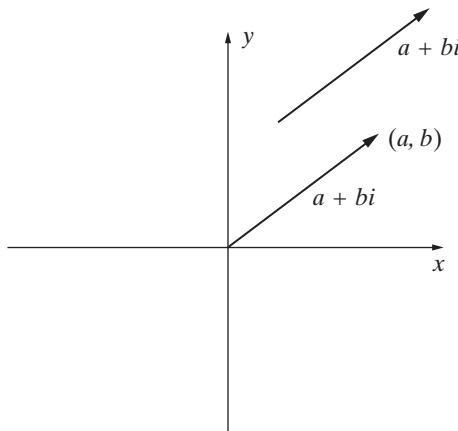


Figure 7.4

The point in the plane that corresponds to a complex number is called the **graph** of the number, and the complex number that corresponds to a point in the plane is called the **coordinate** of the point. Points on the horizontal axis have coordinates $a + 0i$ that are real numbers. Consequently, the horizontal axis is referred to as the **real axis**. Points, other than the origin, on the vertical axis have coordinates $0 + bi$ that are pure imaginary numbers, so the vertical axis is called the **imaginary axis**. Several points are labeled with their coordinates in Figure 7.4.

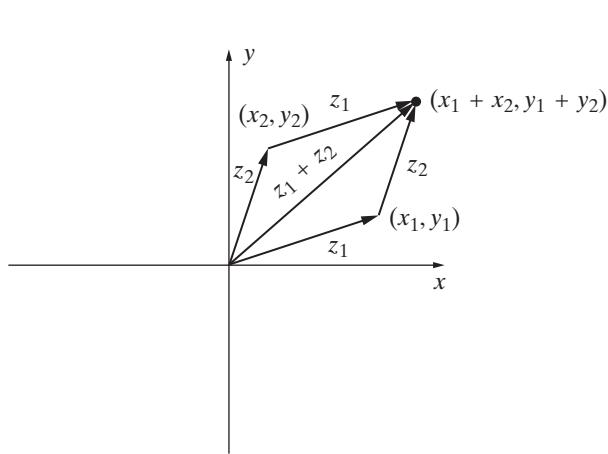
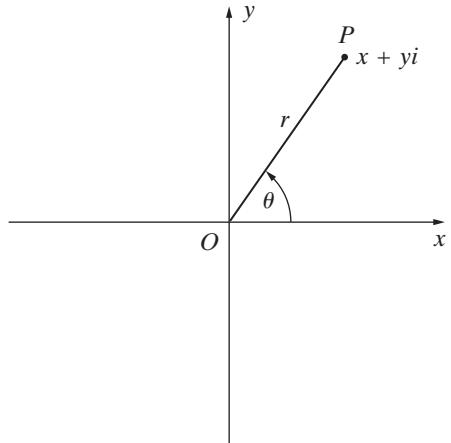
Complex numbers are sometimes represented geometrically by directed line segments called **vectors**. In this approach, the complex number $a + bi$ is represented by the directed line segment from the origin of the coordinate system to the point with rectangular

[†] Abraham de Moivre (1667–1754) was a French mathematician famous for his book on probability theory, *The Doctrine of Chances*. It is rumored that de Moivre predicted the date of his own death, and he was the first to discover Binet's formula for the n th term in the Fibonacci sequence, although Binet is given credit for it.

**Figure 7.5**

coordinates (a, b) or by any directed line segment with the same length and direction as this one. This is shown in Figure 7.5.

In this book we have little use for the vector representation of complex numbers. We simply note that in this interpretation, addition of complex numbers corresponds to the usual “parallelogram rule” for adding vectors. This is illustrated in Figure 7.6.

**Figure 7.6****Figure 7.7**

Returning now to the representation of complex numbers by points in the plane, we observe that any point P in the plane can be located by designating its *distance* r from the origin O and an *angle* θ in standard position that has OP as its terminal side. Figure 7.7 shows r and θ for a complex number $x + yi$ in standard form.

From Figure 7.7, we see that r and θ are related to x and y by the equations

$$x = r \cos \theta, \quad y = r \sin \theta, \quad r = \sqrt{x^2 + y^2}.$$

The complex number $x + yi$ can thus be written in the form

$$x + yi = r(\cos \theta + i \sin \theta).$$

Definition 7.8 ■ Trigonometric Form

When a complex number in standard form $x + yi$ is written as

$$x + yi = r(\cos \theta + i \sin \theta),$$

the expression[†] $r(\cos \theta + i \sin \theta)$ is called the **trigonometric form** (or **polar form**) of $x + yi$. The number $r = \sqrt{x^2 + y^2}$ is called the **absolute value** (or **modulus**) of $x + yi$, and the angle θ is called the **argument** (or **amplitude**) of $x + yi$.

The usual notation is used for the absolute value of a complex number:

$$|x + yi| = r = \sqrt{x^2 + y^2}.$$

The absolute value, r , is unique, but the angle θ is not unique since there are many angles in standard position with P on their terminal side. This is illustrated in the next example.

Example 1 Expressing the complex number $-1 - i$ in trigonometric form,^{††} we have

$$\begin{aligned}-1 - i &= \sqrt{2} \left(-\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i \right) \\&= \sqrt{2} \left(\cos \frac{5\pi}{4} + i \sin \frac{5\pi}{4} \right) \\&= \sqrt{2} \left[\cos \left(-\frac{3\pi}{4} \right) + i \sin \left(-\frac{3\pi}{4} \right) \right] \\&= \sqrt{2} \left(\cos \frac{13\pi}{4} + i \sin \frac{13\pi}{4} \right).\end{aligned}$$

Many other such expressions are possible. ■

Although the argument θ is not unique, an equation of the form

$$r_1(\cos \theta_1 + i \sin \theta_1) = r_2(\cos \theta_2 + i \sin \theta_2)$$

does require that $r_1 = r_2$ and that θ_1 and θ_2 be coterminal. Hence

$$\theta_2 = \theta_1 + k(2\pi)$$

for some integer k .

The next theorem gives a hint as to the usefulness of the trigonometric form of complex numbers. In proving the theorem, we shall use the following identities from trigonometry:

$$\begin{aligned}\cos(A + B) &= \cos A \cos B - \sin A \sin B \\ \sin(A + B) &= \sin A \cos B + \cos A \sin B.\end{aligned}$$

[†]The expression $\cos \theta + i \sin \theta$ sometimes abbreviated as $\text{cis } \theta$.

^{††}We choose to use radian measure for angles. Degree measure could also be used.

Theorem 7.9 ■ Product of Complex Numbers

If

$$z_1 = r_1(\cos \theta_1 + i \sin \theta_1)$$

and

$$z_2 = r_2(\cos \theta_2 + i \sin \theta_2)$$

are arbitrary complex numbers in trigonometric form, then

$$z_1 z_2 = r_1 r_2 [\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)].$$

In words, the absolute value of the product of two complex numbers is the product of their absolute values, and an argument of the product is the sum of their arguments.

Proof The statement of the theorem follows from

$$\begin{aligned} z_1 z_2 &= [r_1(\cos \theta_1 + i \sin \theta_1)][r_2(\cos \theta_2 + i \sin \theta_2)] \\ &= r_1 r_2 [(\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) \\ &\quad + i(\cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2)] \\ &= r_1 r_2 [\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)]. \end{aligned}$$

The preceding result leads to the next theorem, which begins to reveal the true usefulness of the trigonometric form.

Theorem 7.10 ■ De Moivre's Theorem

If n is a positive integer and

$$z = r(\cos \theta + i \sin \theta)$$

is a complex number in trigonometric form, then

$$z^n = r^n(\cos n\theta + i \sin n\theta).$$

Induction **Proof** For $n = 1$, the statement is trivial. Assume that it is true for $n = k$ —that is, that

$$z^k = r^k(\cos k\theta + i \sin k\theta).$$

Using Theorem 7.9, we have

$$\begin{aligned} z^{k+1} &= z^k \cdot z \\ &= [r^k(\cos k\theta + i \sin k\theta)][r(\cos \theta + i \sin \theta)] \\ &= r^{k+1}[\cos(k\theta + \theta) + i \sin(k\theta + \theta)] \\ &= r^{k+1}[\cos(k+1)\theta + i \sin(k+1)\theta]. \end{aligned}$$

Thus the theorem is true for $n = k + 1$, and it follows by induction that the theorem is true for all positive integers.

Example 2 Some applications of De Moivre's Theorem are shown in the following computations.

$$\begin{aligned}\text{a. } (-2 + 2i)^4 &= \left[2\sqrt{2} \left(-\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i \right) \right]^4 \\ &= \left[2\sqrt{2} \left(\cos \frac{3\pi}{4} + i \sin \frac{3\pi}{4} \right) \right]^4 \\ &= 64(\cos 3\pi + i \sin 3\pi) \\ &= 64(-1 + 0i) = -64\end{aligned}$$

$$\begin{aligned}\text{b. } \left(\frac{\sqrt{3}}{2} + \frac{1}{2}i \right)^{40} &= \left[1 \left(\cos \frac{\pi}{6} + i \sin \frac{\pi}{6} \right) \right]^{40} \\ &= 1^{40} \left(\cos \frac{20\pi}{3} + i \sin \frac{20\pi}{3} \right) \\ &= \cos \left(\frac{2\pi}{3} + 6\pi \right) + i \sin \left(\frac{2\pi}{3} + 6\pi \right) \\ &= \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3} \\ &= -\frac{1}{2} + \frac{\sqrt{3}}{2}i\end{aligned}$$

■

If n is a positive integer greater than 1 and $u^n = z$ for complex numbers u and z , then u is called an **n th root** of z . We shall prove that every nonzero complex number has exactly n n th roots in \mathbf{C} .

Theorem 7.11 ■ n th Roots of a Complex Number

For each integer $n \geq 1$, any nonzero complex number

$$z = r(\cos \theta + i \sin \theta)$$

has exactly n distinct n th roots in \mathbf{C} , and these are given by

$$r^{1/n} \left(\cos \frac{\theta + 2k\pi}{n} + i \sin \frac{\theta + 2k\pi}{n} \right), \quad k = 0, 1, 2, \dots, n-1,$$

where $r^{1/n} = \sqrt[n]{r}$ denotes the positive real n th root of r .

Proof For an arbitrary integer k , let

$$v = r^{1/n} \left(\cos \frac{\theta + 2k\pi}{n} + i \sin \frac{\theta + 2k\pi}{n} \right).$$

Then

$$\begin{aligned} v^n &= (r^{1/n})^n \left(\cos \frac{n(\theta + 2k\pi)}{n} + i \sin \frac{n(\theta + 2k\pi)}{n} \right) \\ &= r[\cos(\theta + 2k\pi) + i \sin(\theta + 2k\pi)] \\ &= r(\cos \theta + i \sin \theta) \\ &= z, \end{aligned}$$

and v is an n th root of z . The n angles

$$\frac{\theta}{n}, \quad \frac{\theta + 2\pi}{n}, \quad \frac{\theta + 2(2\pi)}{n}, \dots, \quad \frac{\theta + (n-1)(2\pi)}{n}$$

are equally spaced $\frac{2\pi}{n}$ radians apart, so no two of them have the same terminal side. Thus the n values of v obtained by letting $k = 0, 1, 2, \dots, n-1$ are distinct, and we have shown that z has at least n distinct n th roots in \mathbf{C} .

To show there are no other n th roots of z in \mathbf{C} , suppose $v = t(\cos \phi + i \sin \phi)$ is the trigonometric form of a complex number v such that $v^n = z$. Then

$$t^n(\cos n\phi + i \sin n\phi) = r(\cos \theta + i \sin \theta),$$

by De Moivre's Theorem. It follows from this that

$$t^n = r, \quad \cos n\phi = \cos \theta, \quad \text{and} \quad \sin n\phi = \sin \theta.$$

Since r and t are positive, it must be true that $t = r^{1/n}$. The other equations require that $n\phi$ and θ be coterminal, and hence they differ by a multiple of 2π :

$$n\phi = \theta + m(2\pi)$$

for some integer m . By the Division Algorithm,

$$m = qn + k,$$

where $k \in \{0, 1, 2, \dots, n-1\}$. Thus

$$n\phi = \theta + (qn + k)(2\pi)$$

and

$$\phi = \frac{\theta + 2k\pi}{n} + q(2\pi).$$

This equation shows that ϕ is coterminal with the angle $\frac{(\theta + 2k\pi)}{n}$, and hence v is one of the n th roots listed in the statement of the theorem.

Example 3 We shall find the three cube roots of $8i$ and express each in standard form $a + bi$. Expressing $8i$ in trigonometric form, we have

$$8i = 8 \left(\cos \frac{\pi}{2} + i \sin \frac{\pi}{2} \right).$$

By the formula in Theorem 7.11, the cube roots of $z = 8i$ are given by

$$8^{1/3} \left(\cos \frac{\frac{\pi}{2} + 2k\pi}{3} + i \sin \frac{\frac{\pi}{2} + 2k\pi}{3} \right), \quad k = 0, 1, 2.$$

Each of these roots has absolute value $8^{1/3} = 2$, and they are equally spaced $\frac{2\pi}{3}$ radians apart, with the first one at $\frac{\pi}{6}$. Thus the three cube roots of $8i$ are

$$2 \left(\cos \frac{\pi}{6} + i \sin \frac{\pi}{6} \right) = 2 \left(\frac{\sqrt{3}}{2} + \frac{1}{2}i \right) = \sqrt{3} + i$$

$$2 \left(\cos \frac{5\pi}{6} + i \sin \frac{5\pi}{6} \right) = 2 \left(-\frac{\sqrt{3}}{2} + \frac{1}{2}i \right) = -\sqrt{3} + i$$

$$2 \left(\cos \frac{3\pi}{2} + i \sin \frac{3\pi}{2} \right) = 2(0 - i) = -2i.$$

These results may be checked by direct multiplication. ■

Exercises 7.3

True or False

Label each of the following statements as either true or false.

1. There is a one-to-one correspondence between the standard form and the trigonometric form of a complex number.
2. Every nonzero complex number has exactly n n th roots in \mathbf{C} .
3. In order for two trigonometric forms to represent the same complex number, the absolute values must be equal and the arguments must be equal.
4. The n n th roots of any complex number are equally spaced around a circle with center at the origin.

Exercises

1. Graph each of the following complex numbers, and express each in trigonometric form.

a. $-2 + 2\sqrt{3}i$ c. $3 - 3i$ e. $1 + \sqrt{3}i$ g. -4	b. $2 + 2i$ d. $\sqrt{3} + i$ f. $-1 - i$ h. $-5i$
--	---

2. Find each of the following products. Write each result in both trigonometric and standard form.

a. $[4(\cos \frac{\pi}{8} + i \sin \frac{\pi}{8})][\cos \frac{5\pi}{8} + i \sin \frac{5\pi}{8}]$
 b. $[3(\cos \frac{7\pi}{6} + i \sin \frac{7\pi}{6})][\cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3}]$
 c. $[2(\cos \frac{5\pi}{6} + i \sin \frac{5\pi}{6})]\{3[\cos(-\frac{\pi}{6}) + i \sin(-\frac{\pi}{6})]\}$
 d. $[6(\cos \frac{5\pi}{3} + i \sin \frac{5\pi}{3})]\{5[\cos(-\frac{\pi}{3}) + i \sin(-\frac{\pi}{3})]\}$

3. Use De Moivre's Theorem to find the value of each of the following. Leave your answers in standard form $a + bi$.

a. $(\sqrt{3} + i)^7$ b. $\left(\frac{\sqrt{3}}{2} + \frac{1}{2}i\right)^{21}$
 c. $(-\sqrt{3} + i)^{10}$ d. $\left(\frac{\sqrt{3}}{2} - \frac{1}{2}i\right)^{18}$
 e. $\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)^8$ f. $(\sqrt{2} + \sqrt{2}i)^9$
 g. $(1 - \sqrt{3}i)^8$ h. $\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)^{12}$

4. Show that the n distinct n th roots of 1 are equally spaced around a circle with center at the origin and radius 1.

5. If $\omega = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}$, show that the distinct n th roots of 1 are given by $\omega, \omega^2, \dots, \omega^{n-1}, \omega^n = 1$.

6. Find the indicated roots of 1 in standard form $a + bi$, and graph them on a unit circle with center at the origin.

a. cube roots of 1 b. fourth roots of 1
 c. eighth roots of 1 d. sixth roots of 1

7. Find all the indicated roots of the given number. Leave your results in trigonometric form.

a. cube roots of $\frac{\sqrt{3}}{2} + \frac{1}{2}i$ b. cube roots of $-1 + i$
 c. fourth roots of $-\frac{\sqrt{3}}{2} + \frac{1}{2}i$ d. fourth roots of $\frac{1}{2} - \frac{\sqrt{3}}{2}i$
 e. fifth roots of $-16\sqrt{2} - 16\sqrt{2}i$ f. sixth roots of $32\sqrt{3} - 32i$

8. Find all complex numbers that are solutions of the given equation. Leave your answers in standard form $a + bi$.

a. $z^3 + 27 = 0$ b. $z^8 - 16 = 0$
 c. $z^3 - i = 0$ d. $z^3 + 8i = 0$

e. $z^4 + \frac{1}{2} - \frac{\sqrt{3}}{2}i = 0$

f. $z^4 + 1 - \sqrt{3}i = 0$

g. $z^4 + \frac{1}{2} + \frac{\sqrt{3}}{2}i = 0$

h. $z^4 + 8 + 8\sqrt{3}i = 0$

9. If $\omega = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}$, and u is any n th root of $z \in \mathbf{C}$, show that the n th roots of z are given by $\omega u, \omega^2 u, \dots, \omega^{n-1} u, \omega^n u = u$.
10. Prove that for a fixed value of n , the set U_n of all n th roots of 1 forms a group with respect to multiplication.

In Exercises 11–14, take U_n to be the group in Exercise 10.

- a. Find all elements of the subgroup $\langle a \rangle$ generated by the given a . Leave your answers in trigonometric form.
- b. State the order of $\langle a \rangle$.
- c. Find all the generators of $\langle a \rangle$.

11. $a = \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3}$ in U_6

12. $a = \cos \frac{3\pi}{2} + i \sin \frac{3\pi}{2}$ in U_8

13. $a = \cos \frac{5\pi}{3} + i \sin \frac{5\pi}{3}$ in U_6

14. $a = \cos \frac{5\pi}{4} + i \sin \frac{5\pi}{4}$ in U_8

15. Prove that the group in Exercise 10 is cyclic, with $\omega = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}$ as a generator.

16. Any generator of the group in Exercise 10 is called a **primitive n th root of 1**. Prove that

$$\cos \frac{2k\pi}{n} + i \sin \frac{2k\pi}{n}$$

is a primitive n th root of 1 if and only if k and n are relatively prime.

17. a. Find all primitive sixth roots of 1.
b. Find all primitive eighth roots of 1.
18. Let $\omega_k = \cos \frac{2k\pi}{n} + i \sin \frac{2k\pi}{n}$ be a primitive n th root of unity. Prove that if r is a positive integer such that $(n, r) = d$, then ω_k^r is a primitive (n/d) th root of unity.
19. Prove that the set of *all* roots of 1 forms a group with respect to multiplication.
20. Prove that the sum of all the distinct n th roots of 1 is 0.
21. Prove that the product of all the distinct n th roots of 1 is $(-1)^{n+1}$.
22. Prove the following statements concerning absolute values of complex numbers. (As in Definition 7.7, \bar{z} denotes the conjugate of z .)
- a. $|\bar{z}| = |z|$
b. $z\bar{z} = |z|^2$
- c. If $z \neq 0$, then $z^{-1} = \frac{\bar{z}}{|z|^2}$.
d. If $z_2 \neq 0$, then $\left| \frac{z_1}{z_2} \right| = \frac{|z_1|}{|z_2|}$.
- e. $|z_1 + z_2| \leq |z_1| + |z_2|$

23. Prove that the set of all complex numbers that have absolute value 1 forms a group with respect to multiplication.
24. Prove that if $z = r(\cos \theta + i \sin \theta)$ is a nonzero complex number in trigonometric form, then $z^{-1} = r^{-1}[\cos(-\theta) + i \sin(-\theta)]$.
25. Prove that if n is a positive integer and $z = r(\cos \theta + i \sin \theta)$ is a nonzero complex number in trigonometric form, then $z^{-n} = r^{-n}[\cos(-n\theta) + i \sin(-n\theta)]$.
26. Prove that if $z_1 = r_1(\cos \theta_1 + i \sin \theta_1)$ and $z_2 = r_2(\cos \theta_2 + i \sin \theta_2)$ are complex numbers in trigonometric form and z_2 is nonzero, then

$$\frac{z_1}{z_2} = \frac{r_1}{r_2} [\cos(\theta_1 - \theta_2) + i \sin(\theta_1 - \theta_2)].$$

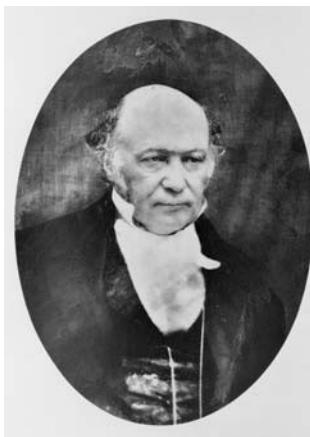
27. Let u be an n th root of unity.
- Show that u^{-1} is also an n th root of unity.
 - Show that \bar{u} is also an n th root of unity.
- Sec. 5.4, #7 ➤ 28. In the ordered field \mathbf{R} , absolute value is defined according to Exercise 7 of Section 5.4 by

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0. \end{cases}$$

For $a \in \mathbf{R}$, show that the absolute value of $a + 0i$ according to Definition 7.8 agrees with the definition from Chapter 5. (Keep in mind, however, that \mathbf{C} is not an ordered field, as was shown in Section 5.4.)

Key Words and Phrases

absolute value, 342, 345	division ring, 337	rational numbers, 328
amplitude, 345	imaginary number, 337	real numbers, 328
argument, 345	irrational numbers, 330	skew field, 337
complete ordered field, 328	least upper bound, 325	standard form of a complex number, 337
complex numbers, 334	modulus, 345	trigonometric form, 345
conjugate of a complex number, 337	n th roots, 347	upper bound, 325
decimal representation, 330	polar form, 345	
De Moivre's Theorem, 346	pure imaginary number, 337	
	quaternions, 337	



LC-USZ62-100657/Library of Congress Prints and Photographs Division

A Pioneer in Mathematics

William Rowan Hamilton (1805–1865)

William Rowan Hamilton, born in Dublin, Ireland, on August 3, 1805, became Ireland's greatest mathematician. He was the fourth of nine children and did not attend school. Instead, he was tutored by an uncle. By the age of 3, he showed amazing ability in reading and arithmetic; he had mastered 13 languages by age 13. His interest turned to mathematics in 1813, when he placed only second in a public contest of arithmetic skills. This humbling incident led him to a study of the classical mathematics texts in their original languages of Greek, Latin, and French. In 1823, he was the top student entering Trinity College, Dublin. He was knighted in 1835 for obtaining significant results in the field of optics.

In 1833, Hamilton initiated a new line of thought about complex numbers by treating them as ordered pairs. He spent the next ten years of his life trying to generalize this treatment of ordered pairs to ordered triples. One day, while walking and chatting with his wife along the Royal Canal on the way to a meeting, he became preoccupied with his own thoughts about the ordered triples and suddenly made a dramatic discovery. He realized that if he considered quadruples (the "quaternions") instead of triples and compromised the commutative law for multiplication, he would have the generalization that he had been seeking for several years. Hamilton became so excited about his discovery that he recorded it in a pocket book and impulsively carved it in a stone on the Brougham Bridge. A tablet there marks the spot of Hamilton's discovery of the quaternions.

Hamilton's approach to complex numbers and their four-dimensional generalization, the quaternions, revolutionized algebraic thought. He spent the last 22 years of his life studying the theory of quaternions and reporting his results.

This page intentionally left blank

Polynomials

■ Introduction

The elementary theory of polynomials over a field is presented in this chapter. Topics included are the division algorithm, the greatest common divisor, factorization theorems, simple algebraic extensions, and splitting fields for polynomials. This chapter may be studied independently of Chapter 7.

8.1

Polynomials over a Ring

Starting with beginning algebra courses, a great deal of time is devoted to developing skills in various manipulations with polynomials. Procedures are learned for the basic operations of addition, subtraction, multiplication, and division of polynomials. By the time a student begins an abstract algebra course, polynomials are a very familiar topic.

Much of this prior experience involved polynomials in a single letter, such as $5 + 4t + t^2$, where the letter usually represented a variable with domain a subset of the real numbers. In this section our point of view is very different. We wish to start with a commutative ring R with unity[†] 1 and *construct* a ring that contains both R and a given element x . More precisely, we want to construct a *smallest* ring that contains R and x in this sense: Any ring that contains both R and x would necessarily contain the constructed ring. We assume that x is *not* an element of R , but nothing more than this. For the time being, the letter x will be a formal symbol subject only to the definitions that are made as we proceed. The letter x is referred to as an **indeterminate** in order to emphasize its role here. Later, we shall consider other possible roles for x .

Definition 8.1 ■ Polynomial in x over R

Let R be a commutative ring with unity 1, and let x be an indeterminate. A **polynomial in x with coefficients in R** , or a **polynomial in x over R** , is an expression of the form

$$a_0x^0 + a_1x^1 + a_2x^2 + \cdots + a_nx^n$$

where n is a nonnegative integer and each a_i is an element of R . The set of all polynomials in x over R is denoted by $R[x]$.

[†]Throughout this chapter, the unity is denoted by 1 rather than e . A similar construction can be made with fewer restrictions on R , but such generality results in complications that are avoided here.

The construction that we shall carry out will be guided by our previous experience with polynomials. Consistent with this, we adopt the familiar language of elementary algebra and refer to the parts $a_i x^i$ of the expression in Definition 8.1 as **terms** of the polynomial and to a_i as the **coefficient** of x^i in the term $a_i x^i$. As a notational convenience, we shall use functional notations such as $f(x)$ for shorthand names of polynomials. That is, we shall write things such as

$$f(x) = a_0 x^0 + a_1 x^1 + \cdots + a_n x^n,$$

but this indicates only that $f(x)$ is a symbolic name for the polynomial. It does *not* indicate a function or a function value.

Example 1 Some examples of polynomials in x over the ring \mathbf{Z} of integers are listed here.

- a. $f(x) = 2x^0 + (-4)x^1 + 0x^2 + 5x^3$
- b. $g(x) = 1x^0 + 2x^1 + (-1)x^2$
- c. $h(x) = (-5)x^0 + 0x^1 + 0x^2$

■

We have not yet defined equality of polynomials. (The preceding use of $=$ only indicated that certain polynomials had been given shorthand names.) To be consistent with prior experience, it is desirable to define equality of polynomials so that terms with zero coefficients can be deleted with equality retained. With this goal in mind, we make the following (somewhat cumbersome) definition.

Definition 8.2a ■ Equality of Polynomials

Suppose that R is a commutative ring with unity, that x is an indeterminate, and that

$$f(x) = a_0 x^0 + a_1 x^1 + \cdots + a_n x^n$$

and

$$g(x) = b_0 x^0 + b_1 x^1 + \cdots + b_m x^m$$

are polynomials in x over R . Then $f(x)$ and $g(x)$ are **equal polynomials**, $f(x) = g(x)$, if and only if the following conditions hold for all i that occur as a subscript on a coefficient in either $f(x)$ or $g(x)$:

1. If one of a_i, b_i is zero, then the other either is omitted or is also zero.
 2. If one of a_i, b_i is not zero, then the other is not omitted, and $a_i = b_i$.
-

Example 2 According to Definition 8.2a, the following equalities are valid in the set $\mathbf{Z}[x]$ of all polynomials in x over \mathbf{Z} .

- a. $2x^0 + (-4)x^1 + 0x^2 + 5x^3 = 2x^0 + (-4)x^1 + 5x^3$
- b. $(-5)x^0 + 0x^1 + 0x^2 = (-5)x^0$

■

The compact sigma notation is useful when we work with polynomials. The polynomial

$$f(x) = a_0x^0 + a_1x^1 + \cdots + a_nx^n$$

may be written compactly using the sigma notation as

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

After the convention concerning zero coefficients has been clarified and agreed upon as stated in conditions 1 and 2 of Definition 8.2a, the definition of *equality of polynomials* may be shortened as follows.

Definition 8.2b ■ Alternative Definition, Equality of Polynomials

If R is a commutative ring with unity, and $f(x) = \sum_{i=0}^n a_i x^i$ and $g(x) = \sum_{i=0}^m b_i x^i$ are polynomials in x over R , then $f(x) = g(x)$ if and only if $a_i = b_i$ for all i .

It is understood, of course, that any polynomial over R has only a finite number of nonzero terms. The notational agreements that have been made allow us to make concise definitions of addition and multiplication in $R[x]$.

Definition 8.3 ■ Addition and Multiplication of Polynomials

Let R be a commutative ring with unity. For any $f(x) = \sum_{i=0}^n a_i x^i$ and $g(x) = \sum_{i=0}^m b_i x^i$ in $R[x]$, we define **addition** in $R[x]$ by

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

where k is the larger of the two integers n, m . We define **multiplication** in $R[x]$ by

$$f(x)g(x) = \sum_{i=0}^{n+m} c_i x^i,$$

where $c_i = \sum_{j=0}^i a_j b_{i-j}$.

The expanded expression for c_i appears as

$$c_i = a_0 b_i + a_1 b_{i-1} + a_2 b_{i-2} + \cdots + a_{i-2} b_2 + a_{i-1} b_1 + a_i b_0.$$

We shall see presently that this formula agrees with previous experience in the multiplication of polynomials.

To introduce some novelty in our next example, we consider the sum and product of two polynomials over the ring \mathbf{Z}_6 .

Example 3 We shall follow a convention that has been used on some earlier occasions and write a for $[a]$ in \mathbf{Z}_6 . Let

$$f(x) = \sum_{i=0}^3 a_i x^i = 1x^0 + 5x^1 + 3x^3$$

and

$$g(x) = \sum_{i=0}^1 b_i x^i = 4x^0 + 2x^1$$

in $\mathbf{Z}_6[x]$. According to our agreement regarding zero coefficients, these polynomials may be written as

$$\begin{aligned} f(x) &= 1x^0 + 5x^1 + 0x^2 + 3x^3 \\ g(x) &= 4x^0 + 2x^1 + 0x^2 + 0x^3, \end{aligned}$$

and the definition of addition yields

$$\begin{aligned} f(x) + g(x) &= \sum_{i=0}^3 (a_i + b_i)x^i \\ &= (1 + 4)x^0 + (5 + 2)x^1 + (0 + 0)x^2 + (3 + 0)x^3 \\ &= 5x^0 + 1x^1 + 0x^2 + 3x^3 \\ &= 5x^0 + 1x^1 + 3x^3, \end{aligned}$$

since $5 + 2 = 1$ in \mathbf{Z}_6 . The definition of multiplication gives

$$f(x)g(x) = \sum_{i=1}^4 c_i x^i,$$

where

$$\begin{aligned} c_0 &= a_0 b_0 = 1 \cdot 4 = 4 \\ c_1 &= a_0 b_1 + a_1 b_0 = 1 \cdot 2 + 5 \cdot 4 = 2 + 2 = 4 \\ c_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0 = 1 \cdot 0 + 5 \cdot 2 + 0 \cdot 4 = 4 \\ c_3 &= a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0 = 1 \cdot 0 + 5 \cdot 0 + 0 \cdot 2 + 3 \cdot 4 = 0 \\ c_4 &= a_0 b_4 + a_1 b_3 + a_2 b_2 + a_3 b_1 + a_4 b_0 \\ &= 1 \cdot 0 + 5 \cdot 0 + 0 \cdot 0 + 3 \cdot 2 + 0 \cdot 4 = 0. \end{aligned}$$

Thus

$$\begin{aligned} f(x)g(x) &= (1x^0 + 5x^1 + 3x^3)(4x^0 + 2x^1) \\ &= 4x^0 + 4x^1 + 4x^2 + 0x^3 + 0x^4 \\ &= 4x^0 + 4x^1 + 4x^2 \end{aligned}$$

in $\mathbf{Z}_6[x]$. This product, obtained by using Definition 8.3, agrees with the result obtained by the usual multiplication procedure based on the distributive laws:

$$\begin{aligned} f(x)g(x) &= (1x^0 + 5x^1 + 3x^3)(4x^0) + (1x^0 + 5x^1 + 3x^3)(2x^1) \\ &= (4x^0 + 2x^1 + 0x^3) + (2x^1 + 4x^2 + 0x^4) \\ &= 4x^0 + 4x^1 + 4x^2. \end{aligned}$$

■

The expanded forms of the c_i in Example 3 illustrate how the coefficient of x^i in the product is the sum of all products of the form $a_p b_q$ with $p + q = i$. In general, it is true that

$$\begin{aligned} c_i &= \sum_{j=0}^i a_j b_{i-j} \\ &= a_0 b_i + a_1 b_{i-1} + a_2 b_{i-2} + \cdots + a_{i-1} b_1 + a_i b_0 \\ &= \sum_{p+q=i} a_p b_q. \end{aligned}$$

This observation is useful in the proof of our next theorem.

Theorem 8.4 ■ The Ring of Polynomials over R

Let R be a commutative ring with unity. With addition and multiplication as given in Definition 8.3, $R[x]$ forms a commutative ring with unity.

Proof Let

$$f(x) = \sum_{i=0}^n a_i x^i, \quad g(x) = \sum_{i=0}^m b_i x^i, \quad h(x) = \sum_{i=0}^k c_i x^i$$

represent arbitrary elements of $R[x]$, and let s be the greatest of the integers n , m , and k .

It follows immediately from Definition 8.3 that the sum $f(x) + g(x)$ is a well-defined element of $R[x]$, and $R[x]$ is closed under addition. Addition in $R[x]$ is associative since

$$\begin{aligned} f(x) + [g(x) + h(x)] &= \sum_{i=0}^n a_i x^i + \sum_{i=0}^s (b_i + c_i) x^i \\ &= \sum_{i=0}^s [a_i + (b_i + c_i)] x^i \\ &= \sum_{i=0}^s [(a_i + b_i) + c_i] x^i \text{ since addition is associative in } R \\ &= \sum_{i=0}^s (a_i + b_i) x^i + \sum_{i=0}^k c_i x^i \\ &= [f(x) + g(x)] + h(x). \end{aligned}$$

The polynomial $0x^0$ is an additive identity in $R[x]$ since

$$f(x) + 0x^0 = 0x^0 + f(x) = f(x)$$

for all $f(x)$ in $R[x]$. The additive inverse of $f(x)$ is $\sum_{i=0}^n (-a_i)x^i$ since

$$f(x) + \sum_{i=0}^n (-a_i)x^i = \sum_{i=0}^n [a_i + (-a_i)]x^i = 0x^0,$$

and $\sum_{i=0}^n (-a_i)x^i + f(x) = 0x^0$ in similar fashion. Addition in $R[x]$ is commutative since

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

Thus $R[x]$ is an abelian group with respect to addition.

It is clear from Definition 8.3 that $R[x]$ is closed under the binary operation of multiplication. To see that multiplication is associative in $R[x]$, we first note that the coefficient of x^i in $f(x)[g(x)h(x)]$ is given by

$$\sum_{p+q+r=i} a_p(b_q c_r),$$

the sum of all products $a_p(b_q c_r)$ of coefficients a_p, b_q, c_r such that the subscripts sum to i . Similarly, in $[f(x)g(x)]h(x)$, the coefficient of x^i is

$$\sum_{p+q+r=i} (a_p b_q) c_r.$$

Now $a_p(b_q c_r) = (a_p b_q)c_r$ since multiplication is associative in R , and therefore $f(x)[g(x)h(x)] = [f(x)g(x)]h(x)$.

Before considering the distributive laws, we shall establish that multiplication in $R[x]$ is commutative. This follows from the equalities

$$\begin{aligned} f(x)g(x) &= \sum_{i=0}^{n+m} \left(\sum_{p+q=i} a_p b_q \right) x^i \\ &= \sum_{i=0}^{m+n} \left(\sum_{q+p=i} b_q a_p \right) x^i \text{ since multiplication is commutative in } R \\ &= g(x)f(x). \end{aligned}$$

Let t be the greater of the integers m and k , and consider the left distributive law. We have

$$\begin{aligned} f(x)[g(x) + h(x)] &= \sum_{i=0}^n a_i x^i \left[\sum_{i=0}^t (b_i + c_i) x^i \right] \\ &= \sum_{i=0}^{n+t} \left[\sum_{p+q=i} a_p (b_q + c_q) \right] x^i \\ &= \sum_{i=0}^{n+t} \left[\sum_{p+q=i} (a_p b_q + a_p c_q) \right] x^i \\ &= \sum_{i=0}^{n+t} \left(\sum_{p+q=i} a_p b_q \right) x^i + \sum_{i=0}^{n+t} \left(\sum_{p+q=i} a_p c_q \right) x^i \\ &= \sum_{i=0}^{n+m} \left(\sum_{p+q=i} a_p b_q \right) x^i + \sum_{i=0}^{n+k} \left(\sum_{p+q=i} a_p c_q \right) x^i \\ &= f(x)g(x) + f(x)h(x), \end{aligned}$$

and the left distributive property is established. The right distributive property is now easy to prove:

$$\begin{aligned} [f(x) + g(x)]h(x) &= h(x)[f(x) + g(x)] && \text{since multiplication is} \\ &&& \text{commutative in } R[x] \\ &= h(x)f(x) + h(x)g(x) && \text{by the left distributive law} \\ &= f(x)h(x) + g(x)h(x) && \text{since multiplication is} \\ &&& \text{commutative in } R[x]. \end{aligned}$$

The element $1x^0$ is a unity in $R[x]$ since

$$1x^0 \cdot f(x) = f(x) \cdot 1x^0 = \sum_{i=0}^n (a_i \cdot 1)x^i = \sum_{i=0}^n a_i x^i = f(x).$$

This completes the proof that $R[x]$ is a commutative ring with unity.

Theorem 8.4 justifies referring to $R[x]$ as the **ring of polynomials over R** or as the **ring of polynomials with coefficients in R** .

Theorem 8.5 ■ Subring of $R[x]$ Isomorphic to R

For any commutative ring R with unity, the ring $R[x]$ of polynomials over R contains a subring R' that is isomorphic to R .

Proof Let R' be the subset of $R[x]$ that consists of all elements of the form ax^0 . We shall show that R' is a subring by utilizing Theorem 5.4.

The subset R' contains elements such as the additive identity $0x^0$ and the unity $1x^0$ of $R[x]$. For arbitrary ax^0 and bx^0 in R' ,

$$ax^0 - bx^0 = (a - b)x^0$$

and

$$(ax^0)(bx^0) = (ab)x^0$$

are in R' , and therefore R' is a subring of $R[x]$ by Theorem 5.4.

Guided by our previous experience with polynomials, we define $\theta: R \rightarrow R'$ by

$$\theta(a) = ax^0$$

for all $a \in R$. This rule defines a one-to-one correspondence since θ is onto and

$$\theta(a) = \theta(b) \Leftrightarrow ax^0 = bx^0 \Leftrightarrow a = b.$$

Moreover, θ is an isomorphism, since

$$\theta(a + b) = (a + b)x^0 = ax^0 + bx^0 = \theta(a) + \theta(b)$$

and

$$\theta(ab) = (ab)x^0 = (ax^0)(bx^0) = \theta(a)\theta(b).$$

Thus R is embedded in $R[x]$. We can use the isomorphism θ to identify $a \in R$ with ax^0 in $R[x]$, and from now on we shall write a in place of ax^0 . In particular, 0 may denote the zero polynomial $0x^0$, and 1 may denote the unity $1x^0$ in $R[x]$. We write an arbitrary polynomial

$$f(x) = a_0x^0 + a_1x^1 + a_2x^2 + \cdots + a_nx^n$$

as

$$f(x) = a_0 + a_1x^1 + a_2x^2 + \cdots + a_nx^n.$$

Actually, we want to carry this notational simplification a bit further, writing x for x^1 , x^i for $1x^i$, and $-a_i x^i$ for $(-a_i)x^i$. This allows us to use all the conventional polynomial notations for the elements of $R[x]$. Also, we can now regard each term $a_i x^i$ with $i \geq 1$ as a product:

$$a_i x^i = a_i \cdot x \cdot x \cdot \cdots \cdot x$$

with i factors of x in the product.

Having made the agreements described in the last paragraph, we may observe that our major goal for this section has been achieved. We have constructed a “smallest” ring $R[x]$ that contains R and x . It is “smallest” because any ring that contained both R and x would have to contain all polynomials

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

as a consequence of the closure properties.

It is now appropriate to pick up some more of the language that is customarily used in work with polynomials.

Definition 8.6 ■ Degree, Leading Coefficient, Constant Term

Let R be a commutative ring with unity, and let

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

be a *nonzero element* of $R[x]$. Then the **degree** of $f(x)$ is the largest integer k such that the coefficient of x^k is not zero, and this coefficient a_k is called the **leading coefficient** of $f(x)$. The term a_0 of $f(x)$ is called the **constant term** of $f(x)$, and elements of R are referred to as **constant polynomials**.

The degree of $f(x)$ will be abbreviated $\deg f(x)$. Note that degree is *not defined* for the zero polynomial. (The reason for this will be clear later.) Note also that the *polynomials of degree zero* are the same as the *nonzero elements* of R .

Example 4 The polynomials $f(x)$ and $g(x)$ in Example 3 can now be written as

$$\begin{aligned} f(x) &= 1 + 5x + 3x^3 = 3x^3 + 5x + 1 \\ g(x) &= 4 + 2x = 2x + 4. \end{aligned}$$

- a. The constant term of $f(x)$ is 1, and the leading coefficient of $f(x)$ is 3.
- b. The polynomial $g(x)$ has constant term 4 and leading coefficient 2.
- c. $\deg f(x) = 3$ and $\deg g(x) = 1$.
- d. In Example 3, we found that

$$f(x)g(x) = 4 + 4x + 4x^2,$$

so $\deg(f(x)g(x)) = 2$. In connection with the next theorem, we note that

$$\deg(f(x)g(x)) \neq \deg f(x) + \deg g(x)$$

in this instance. ■

Theorem 8.7 ■ Degree of a Product

If R is an integral domain and $f(x)$ and $g(x)$ are nonzero elements of $R[x]$, then

$$\deg(f(x)g(x)) = \deg f(x) + \deg g(x).$$

$(p \wedge q) \Rightarrow r$ **Proof** Let R be an integral domain, and suppose that

$$f(x) = \sum_{i=0}^n a_i x^i \text{ has degree } n$$

and

$$g(x) = \sum_{i=0}^m b_i x^i \text{ has degree } m$$

in $R[x]$. Then $a_n \neq 0$ and $b_m \neq 0$, and this implies that $a_n b_m \neq 0$ since R is an integral domain. But $a_n b_m$ is the leading coefficient in $f(x)g(x)$ since

$$f(x)g(x) = \sum_{i=0}^{n+m} \left(\sum_{j=0}^i a_j b_{i-j} \right) x^i$$

by Definition 8.3. Therefore,

$$\deg(f(x)g(x)) = n + m = \deg f(x) + \deg g(x).$$

Corollary 8.8 ■ Polynomials over an Integral Domain

$R[x]$ is an integral domain if and only if R is an integral domain.

$p \Leftarrow q$ **Proof** Assume that R is an integral domain. If $f(x)$ and $g(x)$ are arbitrary nonzero elements of $R[x]$, then both $f(x)$ and $g(x)$ have degrees. According to Theorem 8.7, $f(x)g(x)$ has a degree that is the sum of $\deg f(x)$ and $\deg g(x)$. Therefore, $f(x)g(x)$ is not the zero polynomial, and this shows that $R[x]$ is an integral domain.

$p \Rightarrow q$ If $R[x]$ is an integral domain, however, then R must also be an integral domain since R is a commutative ring with unity and $R \subseteq R[x]$.

We make some final observations concerning Theorem 8.7. Since the product of the zero polynomial and any polynomial always yields the zero polynomial, the equation in Theorem 8.7 cannot hold when one of the factors is a zero polynomial. This is justification for not defining degree for the zero polynomial. We also note that the reason why the conclusion of Theorem 8.7 fails to hold in Example 4 is that \mathbf{Z}_6 is *not* an integral domain.

Exercises 8.1

True or False

Label each of the following statements as either true or false where R represents a commutative ring with unity.

1. A polynomial in x over R is made up of sums of terms of the form $a_i x^i$ where each $a_i \in R$ and $i \in \mathbf{Z}$.
2. The zero polynomial has degree zero.
3. Polynomials of degree zero over R are the same as the nonzero elements of R .
4. The degree of the sum of any two polynomials $f(x)$ and $g(x)$ over R is always the sum of the degrees of $f(x)$ and $g(x)$.
5. The degree of the product of any two polynomials $f(x)$ and $g(x)$ over R is always the product of the degrees of $f(x)$ and $g(x)$.
6. The degree of the product of any two polynomials $f(x)$ and $g(x)$ over R is always the sum of the degrees of $f(x)$ and $g(x)$.
7. The degree of the product of any two polynomials $f(x)$ and $g(x)$ over an integral domain R always is the sum of the degrees of $f(x)$ and $g(x)$.

Exercises

1. Write the following polynomials in expanded form.

a. $\sum_{i=0}^3 c_i x^i$

c. $\sum_{k=1}^3 a_k x^k$

b. $\sum_{j=0}^4 d_j x^j$

d. $\sum_{k=2}^4 x^k$

2. Express the following polynomials by using sigma notation.

a. $c_0 x^0 + c_1 x^1 + c_2 x^2$

c. $x + x^2 + x^3 + x^4$

b. $d_2 x^2 + d_3 x^3 + d_4 x^4$

d. $x^3 + x^4 + x^5$

3. Consider the following polynomials over \mathbf{Z}_8 , where a is written for $[a]$ in \mathbf{Z}_8 :

$$f(x) = 2x^3 + 7x + 4, \quad g(x) = 4x^2 + 4x + 6, \quad h(x) = 6x^2 + 3.$$

Find each of the following polynomials with all coefficients in \mathbf{Z}_8 .

a. $f(x) + g(x)$

c. $f(x)g(x)$

e. $f(x)g(x) + h(x)$

g. $f(x)g(x) + f(x)h(x)$

b. $g(x) + h(x)$

d. $g(x)h(x)$

f. $f(x) + g(x)h(x)$

h. $f(x)h(x) + g(x)h(x)$

4. Consider the following polynomials over \mathbf{Z}_9 , where a is written for $[a]$ in \mathbf{Z}_9 :

$$f(x) = 2x^3 + 7x + 4, \quad g(x) = 4x^2 + 4x + 6, \quad h(x) = 6x^2 + 3.$$

Find each of the following polynomials with all coefficients in \mathbf{Z}_9 .

- a. $f(x) + g(x)$
- b. $g(x) + h(x)$
- c. $f(x)g(x)$
- d. $g(x)h(x)$
- e. $f(x)g(x) + h(x)$
- f. $f(x) + g(x)h(x)$
- g. $f(x)g(x) + f(x)h(x)$
- h. $f(x)h(x) + g(x)h(x)$

5. Decide whether each of the following subsets is a subring of $R[x]$, and justify your decision in each case.
 - a. the set of all polynomials with zero constant term
 - b. the set of all polynomials that have zero coefficients for all *even* powers of x
 - c. the set of all polynomials that have zero coefficients for all *odd* powers of x
 - d. the set consisting of the zero polynomial together with all polynomials that have degree 2 or less

6. Determine which of the subsets in Exercise 5 are ideals of $R[x]$ and which are principal ideals. Justify your choices.

7. a. Prove that

$$I[x] = \{a_0 + a_1x + \cdots + a_nx^n \mid a_0 = 2k \text{ for } k \in \mathbf{Z}\},$$

the set of all polynomials in $\mathbf{Z}[x]$ with even constant term, is an ideal of $\mathbf{Z}[x]$.

- b. Show that $I[x] = \{x \cdot f(x) + 2 \cdot g(x) \mid f(x), g(x) \in \mathbf{Z}[x]\}$.

8. a. Prove that

$$I[x] = \{a_0 + a_1x + \cdots + a_nx^n \mid a_i = 2k_i \text{ for } k_i \in \mathbf{Z}\},$$

the set of all polynomials with even coefficients, is an ideal of $\mathbf{Z}[x]$.

- b. Prove or disprove that $I[x]$ is a principal ideal.

9. a. Let F be a field. Prove that

$$I[x] = \{a_0 + a_1x + \cdots + a_nx^n \mid a_i \in F \text{ and } a_0 + a_1 + \cdots + a_n = 0\},$$

the set of all polynomials in $F[x]$ such that the sum of the coefficients is zero, is an ideal of $F[x]$.

- b. Prove or disprove that $I[x]$ is a principal ideal.

10. Let R be a commutative ring with unity. Prove that

$$\deg(f(x)g(x)) \leq \deg f(x) + \deg g(x)$$

for all nonzero $f(x), g(x)$ in $R[x]$, even if R is not an integral domain.

11. a. List all the polynomials in $\mathbf{Z}_3[x]$ that have degree 2.

- b. Determine which of the polynomials in part a are units. If none exists, state so.

12. a. Find a nonconstant polynomial in $\mathbf{Z}_4[x]$, if one exists, that is a unit.

- b. Find a nonconstant polynomial in $\mathbf{Z}_3[x]$, if one exists, that is a unit.

- c. Prove or disprove that there exist nonconstant polynomials in $\mathbf{Z}_p[x]$ that are units if p is prime.

13. a. How many polynomials of degree 2 are there in $\mathbf{Z}_n[x]$?
 b. If m is a positive integer, how many polynomials of degree m are there in $\mathbf{Z}_n[x]$?
14. Prove or disprove that $R[x]$ is a field if R is a field.
15. Prove that if I is an ideal in a commutative ring R with unity, then $I[x]$ is an ideal in $R[x]$.
16. a. If R is a commutative ring with unity, show that the characteristic of $R[x]$ is the same as the characteristic of R .
 b. State the characteristic of $\mathbf{Z}_n[x]$.
 c. State the characteristic of $\mathbf{Z}[x]$.
17. a. Suppose that R is a commutative ring with unity, and define $\theta: R[x] \rightarrow R$ by
- $$\theta(a_0 + a_1x + \cdots + a_nx^n) = a_0$$
- for all $a_0 + a_1x + \cdots + a_nx^n$ in $R[x]$. Prove that θ is an epimorphism from $R[x]$ to R .
 b. Describe the kernel of the epimorphism in part a.
18. Let R be a commutative ring with unity, and let I be the principal ideal $I = (x)$ in $R[x]$. Prove that $R[x]/I$ is isomorphic to R .
19. In the integral domain $\mathbf{Z}[x]$, let $(\mathbf{Z}[x])^+$ denote the set of all $f(x)$ in $\mathbf{Z}[x]$ that have a positive integer as a leading coefficient. Prove that $\mathbf{Z}[x]$ is an ordered integral domain by proving that $(\mathbf{Z}[x])^+$ is a set of positive elements for $\mathbf{Z}[x]$.
20. Consider the mapping $\phi: \mathbf{Z}[x] \rightarrow \mathbf{Z}_k[x]$ defined by

$$\phi(a_0 + a_1x + \cdots + a_nx^n) = [a_0] + [a_1]x + \cdots + [a_n]x^n,$$

where $[a_i]$ denotes the congruence class of \mathbf{Z}_k that contains a_i . Prove that ϕ is an epimorphism from $\mathbf{Z}[x]$ to $\mathbf{Z}_k[x]$.

21. Describe the kernel of the epimorphism ϕ in Exercise 20.
22. Assume that each of R and S is a commutative ring with unity and that $\theta: R \rightarrow S$ is an epimorphism from R to S . Let $\phi: R[x] \rightarrow S[x]$ be defined by

$$\phi(a_0 + a_1x + \cdots + a_nx^n) = \theta(a_0) + \theta(a_1)x + \cdots + \theta(a_n)x^n.$$

Prove that ϕ is an epimorphism from $R[x]$ to $S[x]$.

23. Describe the kernel of the epimorphism ϕ in Exercise 22.
24. For each $f(x) = \sum_{i=0}^n a_i x^i$ in $R[x]$, the **formal derivative** of $f(x)$ is the polynomial

$$f'(x) = \sum_{i=1}^n i a_i x^{i-1}.$$

(For $n = 0$, $f'(x) = 0$ by definition.)

- a. Prove that $[f(x) + g(x)]' = f'(x) + g'(x)$.
 b. Prove that $[f(x)g(x)]' = f(x)g'(x) + f'(x)g(x)$.

8.2**Divisibility and Greatest Common Divisor**

If a ring R is not an integral domain, the division of polynomials over R is not a very satisfactory subject for study, because of the possible presence of zero divisors. In order for us to obtain the results we need on division of polynomials, the ring of coefficients actually must be a field. For this reason, with a few exceptions in the exercises, we confine our attention for the rest of this chapter to rings of polynomials $F[x]$ where F is a *field*. This assures us that $F[x]$ is an integral domain (Corollary 8.8) and that every nonzero element of F has a multiplicative inverse.

The definition, the theorems and even proofs in this section are very similar to corresponding statements in Chapter 2 about division in the integral domain \mathbf{Z} .

Definition 8.9 ■ **Divisor, Multiple**

If $f(x)$ and $g(x)$ are in $F[x]$, then $f(x)$ **divides** $g(x)$ if there exists $h(x)$ in $F[x]$ such that $g(x) = f(x)h(x)$.

If $f(x)$ divides $g(x)$, we write $f(x) | g(x)$, and we say that $g(x)$ is a **multiple** of $f(x)$, that $f(x)$ is a **factor** of $g(x)$, or that $f(x)$ is a **divisor** of $g(x)$. We write $f(x) \nmid g(x)$ to indicate that $f(x)$ does not divide $g(x)$.

Polynomials of degree zero (the nonzero elements of F) have two special properties that are worth noting. First, any nonzero element a of F is a factor of every $f(x) \in F[x]$, because $a^{-1}f(x)$ is in $F[x]$ and

$$f(x) = a[a^{-1}f(x)].$$

Second, if $f(x) | g(x)$, then $af(x) | g(x)$ for all nonzero $a \in F$, since the equation

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

implies that

$$g(x) = [af(x)][a^{-1}h(x)].$$

The Division Algorithm for integers has the following analogue in $F[x]$.

Theorem 8.10 ■ **The Division Algorithm**

Let $f(x)$ and $g(x)$ be elements of $F[x]$, with $f(x)$ a nonzero polynomial. There exist unique elements $q(x)$ and $r(x)$ in $F[x]$ such that

$$g(x) = f(x)q(x) + r(x)$$

with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$.

Existence **Proof** We postpone the proof of uniqueness until existence of the required $q(x)$ and $r(x)$ in $F[x]$ has been proved. There are two trivial cases that we shall dispose of first.

1. If $g(x) = 0$ or if $\deg g(x) < \deg f(x)$, then we see from the equality

$$g(x) = f(x) \cdot 0 + g(x)$$

that $q(x) = 0$ and $r(x) = g(x)$ satisfy the required conditions.

2. If $\deg f(x) = 0$, then $f(x) = c$ for some nonzero constant c . The equality

$$g(x) = c[c^{-1}g(x)] + 0$$

shows that $q(x) = c^{-1}g(x)$ and $r(x) = 0$ satisfy the required conditions.

Complete Induction

Suppose now that $g(x) \neq 0$ and $1 \leq \deg f(x) \leq \deg g(x)$. The proof is by induction on $n = \deg g(x)$, using the second principle of finite induction. For each positive integer n , let S_n be the statement that if $g(x) \in F[x]$ has degree n and $1 \leq \deg f(x) \leq \deg g(x)$, then there exist $q(x)$ and $r(x) \in F[x]$ such that $g(x) = f(x)q(x) + r(x)$, with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$.

If $n = 1$, then the condition $1 \leq \deg f(x) \leq \deg g(x) = n$ requires that both $f(x)$ and $g(x)$ have degree 1—say,

$$f(x) = ax + b, \quad g(x) = cx + d,$$

where $a \neq 0$ and $c \neq 0$. The equality

$$cx + d = (ax + b)(ca^{-1}) + (d - bca^{-1})$$

shows that $q(x) = ca^{-1}$ and $r(x) = d - bca^{-1}$ satisfy the required conditions, and S_1 is true.

Now assume that k is a positive integer such that S_m is true for all positive integers $m < k$. To prove that S_k is true, let $g(x) \in F[x]$ with $\deg g(x) = k$ and $f(x) \in F[x]$ with $1 \leq \deg f(x) \leq \deg g(x)$. Then

$$f(x) = ax^j + \dots, \quad g(x) = cx^k + \dots$$

with $a \neq 0$, $c \neq 0$, and $j \leq k$. The first step in the usual long division of $g(x)$ by $f(x)$ is shown in Figure 8.1.

$$\begin{array}{r} ca^{-1}x^{k-j} \\ \hline ax^j + \dots & | cx^k + \dots \\ & \underline{ca^{-1}x^{k-j}f(x)} \\ & g(x) - ca^{-1}x^{k-j}f(x) \end{array}$$

Figure 8.1

This first step in long division yields

$$g(x) = ca^{-1}x^{k-j}f(x) + [g(x) - ca^{-1}x^{k-j}f(x)].$$

Let $h(x) = g(x) - ca^{-1}x^{k-j}f(x)$. Then the coefficient of x^k in $h(x)$ is zero, and $\deg h(x) < k$. By the induction hypothesis, there exist polynomials $q_0(x)$ and $r(x)$ such that

$$h(x) = f(x)q_0(x) + r(x)$$

with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$. This gives the equality

$$\begin{aligned} g(x) &= ca^{-1}x^{k-j}f(x) + h(x) \\ &= ca^{-1}x^{k-j}f(x) + f(x)q_0(x) + r(x) \\ &= f(x)[ca^{-1}x^{k-j} + q_0(x)] + r(x), \end{aligned}$$

which shows that $q(x) = ca^{-1}x^{k-j} + q_0(x)$ and $r(x)$ are polynomials that satisfy the required conditions. Therefore, S_k is true, and the existence part of the theorem follows from the second principle of finite induction.

Uniqueness

To prove uniqueness, suppose that $g(x) = f(x)q_1(x) + r_1(x)$ and $g(x) = f(x)q_2(x) + r_2(x)$, where either $r_i(x) = 0$ or $\deg r_i(x) < \deg f(x)$ for $i = 1, 2$. Then

$$\begin{aligned} r_1(x) - r_2(x) &= [g(x) - f(x)q_1(x)] - [g(x) - f(x)q_2(x)] \\ &= f(x)[q_2(x) - q_1(x)]. \end{aligned}$$

The right member of this equation, $f(x)[q_2(x) - q_1(x)]$, either is zero or has degree greater than or equal to $\deg f(x)$, by Theorem 8.7. However, the left member, $r_1(x) - r_2(x)$, either is zero or has degree less than $\deg f(x)$, since $\deg r_1(x) < \deg f(x)$ and $\deg r_2(x) < \deg f(x)$. Therefore, both members must be zero, and this requires that $r_1(x) = r_2(x)$ and $q_1(x) = q_2(x)$ since $f(x)$ is nonzero. Therefore, $q(x)$ and $r(x)$ are unique and the proof is complete.

In the Division Algorithm, the polynomial $q(x)$ is called the **quotient** and $r(x)$ is called the **remainder** in the division of $g(x)$ by $f(x)$. For any field F , the quotient and remainder in $F[x]$ can be found by the familiar long-division procedures. An illustration is given in the next example.

Example 1 Let $f(x) = 3x^2 + 2$ and $g(x) = 4x^4 + 2x^3 + 6x^2 + 4x + 5$ in $\mathbf{Z}_7[x]$. We shall find $q(x)$ and $r(x)$ by the long-division procedure. Referring to Figure 8.1, we have $a = 3$ in $f(x)$, $c = 4$ in $g(x)$, and $ca^{-1} = 3(4^{-1}) = 3(2) = 6$ in the first step.

$$\begin{array}{r} 6x^2 + 3x + 5 \\ 3x^2 + 2 \overline{)4x^4 + 2x^3 + 6x^2 + 4x + 5} \\ 4x^4 + \quad\quad\quad 5x^2 \\ \hline 2x^3 + \quad\quad\quad x^2 \\ 2x^3 + \quad\quad\quad 6x \\ \hline x^2 + 5x \\ x^2 + \quad\quad\quad + 3 \\ \hline 5x + 2 \end{array}$$

Thus the quotient is $q(x) = 6x^2 + 3x + 5$ and the remainder is $r(x) = 5x + 2$ in the division of $g(x)$ by $f(x)$. ■

Our next objective in this section is to prove that any two nonzero polynomials over F have a greatest common divisor in $F[x]$. We saw earlier that if $f(x)$ is a divisor of $g(x)$, then $af(x)$ is also a divisor of $g(x)$ for every nonzero $a \in F$. By choosing a to be the multiplicative inverse of the leading coefficient of $f(x)$, we can make the leading coefficient in $af(x)$ equal to 1. This means that when we consider common divisors of two polynomials, there is no loss of generality if we restrict our attention to polynomials that have 1 as their leading coefficient.

Definition 8.11 ■ Monic Polynomial

A polynomial with 1 as its leading coefficient is called a **monic** polynomial.

One of the conditions that we place on a greatest common divisor of two polynomials is that it be monic. Without this condition, the greatest common divisor of two polynomials would not be unique.

Definition 8.12 ■ Greatest Common Divisor

Let $f(x)$ and $g(x)$ be nonzero polynomials in $F[x]$. A polynomial $d(x)$ in $F[x]$ is a **greatest common divisor** of $f(x)$ and $g(x)$ if these conditions are satisfied:

1. $d(x)$ is a monic polynomial.
2. $d(x) | f(x)$ and $d(x) | g(x)$.
3. $h(x) | f(x)$ and $h(x) | g(x)$ imply that $h(x) | d(x)$.

The next theorem shows that any two nonzero elements $f(x), g(x)$ of $F[x]$ have a unique greatest common divisor $d(x)$.

Strategy ■ The proof of Theorem 8.13 is obtained by making minor adjustments in the proof of Theorem 2.12, and it shows that $d(x)$ is a **linear combination** of $f(x)$ and $g(x)$; that is, $d(x)$ can be written in the form

$$d(x) = f(x)s(x) + g(x)t(x)$$

for some $s(x), t(x) \in F[x]$.

Theorem 8.13 ■ Greatest Common Divisor

Let $f(x)$ and $g(x)$ be nonzero polynomials over F . Then there exists a unique greatest common divisor $d(x)$ of $f(x)$ and $g(x)$ in $F[x]$. Moreover, $d(x)$ can be expressed as

$$d(x) = f(x)s(x) + g(x)t(x)$$

for $s(x)$ and $t(x)$ in $F[x]$, and $d(x)$ is the monic polynomial of least degree that can be written in this form.

Existence **Proof** Consider the set S of all polynomials in $F[x]$ that can be written in the form

$$f(x)u(x) + g(x)v(x)$$

with $u(x)$ and $v(x)$ in $F[x]$. Since $f(x) = f(x) \cdot 1 + g(x) \cdot 0 \neq 0$, the set of nonzero polynomials in S is nonempty. Let

$$d_1(x) = f(x)u_1(x) + g(x)v_1(x)$$

be a polynomial of least degree among the nonzero elements of S . If c is the leading coefficient of $d_1(x)$, then

$$d(x) = c^{-1}d_1(x) = f(x)[c^{-1}u_1(x)] + g(x)[c^{-1}v_1(x)]$$

is a monic polynomial of least degree in S . Letting $s(x) = c^{-1}u_1(x)$ and $t(x) = c^{-1}v_1(x)$, we have a polynomial

$$d(x) = f(x)s(x) + g(x)t(x),$$

which is expressed in the required form and satisfies the first condition in Definition 8.12.

We shall show that $d(x) \mid f(x)$. By the Division Algorithm, there are elements $q(x)$ and $r(x)$ of $F[x]$ such that

$$f(x) = d(x)q(x) + r(x)$$

with either $r(x) = 0$ or $\deg r(x) < \deg d(x)$. Since

$$\begin{aligned} r(x) &= f(x) - d(x)q(x) \\ &= f(x) - [f(x)s(x) + g(x)t(x)]q(x) \\ &= f(x)[1 - s(x)q(x)] + g(x)[-t(x)q(x)], \end{aligned}$$

$r(x)$ is an element of S . By choice of $d(x)$ as having smallest possible degree among the nonzero elements of S , it cannot be true that $\deg r(x) < \deg d(x)$. Therefore, $r(x) = 0$ and $d(x) \mid f(x)$. A similar argument shows that $d(x) \mid g(x)$, and hence $d(x)$ satisfies condition 2 in Definition 8.12.

If $h(x) \mid f(x)$ and $h(x) \mid g(x)$, then $f(x) = h(x)p_1(x)$ and $g(x) = h(x)p_2(x)$ for $p_i(x) \in F[x]$. Therefore,

$$\begin{aligned} d(x) &= f(x)s(x) + g(x)t(x) \\ &= h(x)p_1(x)s(x) + h(x)p_2(x)t(x) \\ &= h(x)[p_1(x)s(x) + p_2(x)t(x)], \end{aligned}$$

and this shows that $h(x) \mid d(x)$. By Definition 8.12, $d(x)$ is a greatest common divisor of $f(x)$ and $g(x)$.

Uniqueness

To show uniqueness, suppose that $d_1(x)$ and $d_2(x)$ are both greatest common divisors of $f(x)$ and $g(x)$. Then $d_1(x) \mid d_2(x)$ and also $d_2(x) \mid d_1(x)$. Since both $d_1(x)$ and $d_2(x)$ are monic polynomials, this means that $d_1(x) = d_2(x)$. (See Exercise 26 at the end of this section.)

If $f(x)$ and $g(x)$ are nonzero polynomials such that $f(x) \mid g(x)$, then the greatest common divisor of $f(x)$ and $g(x)$ is simply the product of $f(x)$ and the multiplicative inverse of its leading coefficient. If $f(x) \nmid g(x)$, the Euclidean Algorithm extends readily to polynomials, furnishing a systematic method for finding the greatest common divisor of $f(x)$ and $g(x)$ and for finding $s(x)$ and $t(x)$ in the equation

$$d(x) = f(x)s(x) + g(x)t(x).$$

The Euclidean Algorithm consists of repeated application of the Division Algorithm to yield the following sequence, where $r_n(x)$ is the last nonzero remainder.

Euclidean Algorithm

$$\begin{aligned}
 g(x) &= f(x)q_0(x) + r_1(x), & \deg r_1(x) &< \deg f(x) \\
 f(x) &= r_1(x)q_1(x) + r_2(x), & \deg r_2(x) &< \deg r_1(x) \\
 r_1(x) &= r_2(x)q_2(x) + r_3(x), & \deg r_3(x) &< \deg r_2(x) \\
 &\vdots & &\vdots \\
 r_{n-2}(x) &= r_{n-1}(x)q_{n-1}(x) + r_n(x), & \deg r_n(x) &< \deg r_{n-1}(x) \\
 r_{n-1}(x) &= r_n(x)q_n(x)
 \end{aligned}$$

Suppose that a is the leading coefficient of the last nonzero remainder, $r_n(x)$. It is left as an exercise to prove that $a^{-1}r_n(x)$ is the greatest common divisor of $f(x)$ and $g(x)$.

Example 2 We shall find the greatest common divisor of $f(x) = 3x^3 + 5x^2 + 6x$ and $g(x) = 4x^4 + 2x^3 + 6x^2 + 4x + 5$ in $\mathbf{Z}_7[x]$. Long division of $g(x)$ by $f(x)$ yields a quotient of $q_0(x) = 6x$ and a remainder of $r_1(x) = 5x^2 + 4x + 5$, so we have

$$g(x) = f(x) \cdot (6x) + (5x^2 + 4x + 5).$$

Dividing $f(x)$ by $r_1(x)$, we obtain

$$f(x) = r_1(x) \cdot (2x + 5) + (4x + 3),$$

so $q_1(x) = 2x + 5$ and $r_2(x) = 4x + 3$ in the Euclidean Algorithm. Division of $r_1(x)$ by $r_2(x)$ then yields

$$r_1(x) = r_2(x) \cdot (3x + 4).$$

Thus $r_2(x) = 4x + 3$ is the last nonzero remainder, and the greatest common divisor of $f(x)$ and $g(x)$ in $\mathbf{Z}_7[x]$ is

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

As mentioned earlier, the Euclidean Algorithm can also be used to find polynomials $s(x)$ and $t(x)$ such that

$$d(x) = f(x)s(x) + g(x)t(x).$$

This is illustrated in Example 3.

Example 3 As in Example 2, let $f(x) = 3x^3 + 5x^2 + 6x$ and $g(x) = 4x^4 + 2x^3 + 6x^2 + 4x + 5$ in $\mathbf{Z}_7[x]$. From Example 2, the greatest common divisor of $f(x)$ and $g(x)$ is $d(x) = x + 6$. To find polynomials $s(x)$ and $t(x)$ such that

$$d(x) = f(x)s(x) + g(x)t(x),$$

we first solve for the remainders in the Euclidean Algorithm (see Example 2) as follows:

$$\begin{aligned}
 r_2(x) &= f(x) - r_1(x)(2x + 5) \\
 r_1(x) &= g(x) - f(x)(6x).
 \end{aligned}$$

Substituting for $r_1(x)$ in the first equation, we have

$$\begin{aligned} r_2(x) &= f(x) - [g(x) - f(x)(6x)](2x + 5) \\ &= f(x) + f(x)(6x)(2x + 5) - g(x)(2x + 5) \\ &= f(x)[1 + (6x)(2x + 5)] + g(x)(-2x - 5) \\ &= f(x)(5x^2 + 2x + 1) + g(x)(5x + 2). \end{aligned}$$

To express $d(x) = 4^{-1} r_2(x) = 2r_2(x)$ as a linear combination of $f(x)$ and $g(x)$, we multiply both members of the last equation by $4^{-1} = 2$:

$$\begin{aligned} d(x) &= 2r_2(x) = f(x)(2)(5x^2 + 2x + 1) + g(x)(2)(5x + 2) \\ d(x) &= f(x)(3x^2 + 4x + 2) + g(x)(3x + 4). \end{aligned}$$

The desired polynomials are given by $s(x) = 3x^2 + 4x + 2$ and $t(x) = 3x + 4$. ■

Exercises 8.2

True or False

Label each of the following statements as either true or false.

1. Every $f(x)$ in $F[x]$, where F is a field, can be factored.
2. Any two nonzero polynomials over a field F have a unique greatest common divisor.
3. The greatest common divisor of two polynomials $f(x)$ and $g(x)$ over a field F may not be monic if at least one of $f(x)$ or $g(x)$ is not monic.

Exercises

For $f(x)$, $g(x)$, and $\mathbf{Z}_n[x]$ given in Exercises 1–6, find $q(x)$ and $r(x)$ in $\mathbf{Z}_n[x]$ that satisfy the conditions in the Division Algorithm.

1. $f(x) = 3x + 1$, $g(x) = 2x^3 + 3x^2 + 4x + 1$, in $\mathbf{Z}_5[x]$
2. $f(x) = 2x + 2$, $g(x) = x^3 + 2x^2 + 2$, in $\mathbf{Z}_3[x]$
3. $f(x) = x^3 + x^2 + 2x + 2$, $g(x) = x^4 + 2x^2 + x + 1$, in $\mathbf{Z}_3[x]$
4. $f(x) = x^3 + 2x^2 + 2$, $g(x) = 2x^5 + 2x^4 + x^2 + 2$, in $\mathbf{Z}_3[x]$
5. $f(x) = 3x^2 + 2$, $g(x) = x^4 + 5x^2 + 2x + 2$, in $\mathbf{Z}_7[x]$
6. $f(x) = 3x^2 + 2$, $g(x) = 4x^4 + 2x^3 + 6x^2 + 4x + 5$, in $\mathbf{Z}_7[x]$

For $f(x)$, $g(x)$, and $\mathbf{Z}_n[x]$ given in Exercises 7–10, find the greatest common divisor $d(x)$ of $f(x)$ and $g(x)$ in $\mathbf{Z}_n[x]$.

7. $f(x) = x^3 + x^2 + 2x + 2$, $g(x) = x^4 + 2x^2 + x + 1$, in $\mathbf{Z}_3[x]$
8. $f(x) = x^3 + 2x^2 + 2$, $g(x) = 2x^5 + 2x^4 + x^2 + 2$, in $\mathbf{Z}_3[x]$

9. $f(x) = 3x^2 + 2$, $g(x) = x^4 + 5x^2 + 2x + 2$, in $\mathbf{Z}_7[x]$

Sec. 8.3, #27 < 10. $f(x) = 3x^2 + 2$, $g(x) = 4x^4 + 2x^3 + 6x^2 + 4x + 5$, in $\mathbf{Z}_7[x]$

For $f(x)$, $g(x)$ and $\mathbf{Z}_n[x]$ given in Exercises 11–14, find $s(x)$ and $t(x)$ in $\mathbf{Z}_n[x]$ such that $d(x) = f(x)s(x) + g(x)t(x)$ is the greatest common divisor of $f(x)$ and $g(x)$.

11. $f(x) = 2x^3 + 2x^2 + x + 1$, $g(x) = x^4 + 2x^2 + x + 1$, in $\mathbf{Z}_3[x]$

12. $f(x) = 2x^3 + x^2 + 1$, $g(x) = x^5 + x^4 + 2x^2 + 1$, in $\mathbf{Z}_3[x]$

13. $f(x) = 3x^2 + 2$, $g(x) = x^4 + 5x^2 + 2x + 2$, in $\mathbf{Z}_7[x]$

14. $f(x) = 3x^2 + 2$, $g(x) = 4x^4 + 2x^3 + 6x^2 + 4x + 5$, in $\mathbf{Z}_7[x]$

15. a. Factor x as a product of two polynomials of degree 1 in $\mathbf{Z}_{12}[x]$.

b. Factor x as a product of two polynomials of degree 1 in $\mathbf{Z}_{18}[x]$.

16. Factor each of the following polynomials as the product of two polynomials of degree 1 in $\mathbf{Z}_{12}[x]$.

a. $x + 2$

b. $x + 3$

17. Factor each of the following polynomials as the product of two polynomials of degree 1 in $\mathbf{Z}_{10}[x]$.

a. $x + 7$

b. $x + 9$

18. Prove or disprove that the polynomial x can be factored as the product of two polynomials of degree 1 in $F[x]$, where F is a field.

19. Let I be the principal ideal $(x^2 + 1) = \{(x^2 + 1)f(x) | f(x) \in \mathbf{Z}[x]\}$. Determine whether each of the following polynomials are elements of I .

a. $x^4 - 3x^3 + 3x^2 - 3x + 2$

b. $x^4 + x^3 - 2x^2 + x + 1$

20. Let I be the principal ideal $(x^2 + 1) = \{(x^2 + 1)f(x) | f(x) \in \mathbf{Z}_5[x]\}$. Determine whether each of the following polynomials are elements of I .

a. $2x^4 + 4x^3 + 4x + 3$

b. $3x^5 + x^4 + 2x^3 + 3x^2 + 4x + 1$

21. Let I be the principal ideal $(x + 2) = \{(x + 2)f(x) | f(x) \in \mathbf{Z}_7[x]\}$. Determine whether each of the following polynomials are elements of I .

a. $4x^4 + x^2 + x + 2$

b. $5x^4 + 5x^3 + 3x^2 + 2x + 1$

22. Let I be the principal ideal $(2x + 7) = \{(2x + 7)f(x) | f(x) \in \mathbf{Z}_{11}[x]\}$. Determine whether each of the following polynomials are elements of I .

a. $4x^4 + 6x^3 + x^2 + 7x + 4$

b. $9x^4 + x^3 + 8x^2 + 2x + 10$

23. Let $f(x), g(x) \in F[x]$ where $f(x) | g(x)$. Prove $(g(x)) \subseteq (f(x))$.

24. Let $f(x) = a_nx^n + a_{n-1}x^{n-1} + \dots + a_0$ where $a_n \neq 0$. Find the greatest common divisor of $f(x)$ and the zero polynomial.

25. Prove that if $f(x)$ and $g(x)$ are nonzero elements of $F[x]$ such that $f(x) | g(x)$ and $g(x) | f(x)$, then $f(x) = ag(x)$ for some nonzero $a \in F$.

26. Prove that if $d_1(x)$ and $d_2(x)$ are monic polynomials over the field F such that $d_1(x) \mid d_2(x)$ and $d_2(x) \mid d_1(x)$, then $d_1(x) = d_2(x)$.

27. Show that the polynomials $s(x)$ and $t(x)$ in the expression

$$d(x) = f(x)s(x) + g(x)t(x)$$

in Theorem 8.13 are not unique.

28. Prove that if $h(x) \mid f(x)$ and $h(x) \mid g(x)$ in $F[x]$, then $h(x)$ divides $f(x)u(x) + g(x)v(x)$ for all $u(x)$ and $v(x)$ in $F[x]$.

29. Let $f(x), g(x), h(x) \in F[x]$. Prove that if $f(x) \mid g(x)$ and $g(x) \mid h(x)$ then $f(x) \mid h(x)$.

30. In the statement of the Division Algorithm (Theorem 8.10), prove that the greatest common divisor of $g(x)$ and $f(x)$ is equal to the greatest common divisor of $f(x)$ and $r(x)$.

31. With the notation used in the description of the Euclidean Algorithm, prove that $a^{-1}r_n(x)$ is the greatest common divisor of $f(x)$ and $g(x)$.

32. Prove that every nonzero remainder $r_j(x)$ in the Euclidean Algorithm is a linear combination of $f(x)$ and $g(x)$: $r_j(x) = f(x)s_j(x) + g(x)t_j(x)$ for some $s_j(x)$ and $t_j(x)$ in $F[x]$.

33. Prove that the only elements of $F[x]$ that have multiplicative inverses are the nonzero elements of the field F . (Hence $F[x]$ is *not* a field.)

34. Prove that every ideal in $F[x]$, where F is a field, is a principal ideal.

- Sec. 2.4, #25 ➤
Sec. 8.3, #27 ⇐
35. Follow the pattern in Exercise 25 of Section 2.4 to define the least common multiple of two nonzero polynomials $f(x)$ and $g(x)$ over the field F .

8.3

Factorization in $F[x]$

Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ denote an arbitrary polynomial over the field F . For any $c \in F$, $f(c)$ is defined by the equation

$$f(c) = a_0 + a_1c + a_2c^2 + \cdots + a_nc^n.$$

That is, $f(c)$ is obtained by replacing the indeterminate x in $f(x)$ by the element c . For each $c \in F$, this replacement rule yields a unique value $f(c) \in F$, and hence the pairing $(c, f(c))$ defines a mapping from F to F . A mapping obtained in this manner is called a **polynomial mapping**, or a **polynomial function**, from F to F .

Definition 8.14 ■ Zero, Root, Solution

Let $f(x)$ be a polynomial over the field F . If c is an element of F such that $f(c) = 0$, then c is called a **zero** of $f(x)$, and we say that c is a **root**, or a **solution**, of the equation $f(x) = 0$.

Example 1 Consider $f(x) = x^2 + 1$ in $\mathbf{Z}_5[x]$. Since

$$f(2) = 2^2 + 1 = 0$$

in \mathbf{Z}_5 , 2 is a *zero* of $x^2 + 1$. Also, 2 is a *root*, or a *solution*, of $x^2 + 1 = 0$ over \mathbf{Z}_5 . ■

For arbitrary polynomials $f(x)$ and $g(x)$ over a field F , let $h(x) = f(x) + g(x)$ and $p(x) = f(x)g(x)$. Two consequences of the definitions of addition and multiplication in $F[x]$ are that

$$h(c) = f(c) + g(c) \quad \text{and} \quad p(c) = f(c)g(c)$$

for all c in F . We shall use these results quite freely, with their justifications left as exercises.

The difference in the roles of the letters x and c in the preceding paragraph should be emphasized. Beginning with the second paragraph of Section 8.1, the indeterminate x has been used as a formal symbol which represents an element that is not in R (or F) and subject only to the definitions that we have made since that paragraph. However, the symbol c represents a variable element in the field F , and $f(c)$ represents the value of the polynomial function f at the element c .

The next example shows that $f(x)$ and $g(x)$ may be *different* polynomials in $F[x]$ that define the *same* polynomial function from F to F . That is, we may have $f(c) = g(c)$ for all c in F while the polynomials $f(x)$ and $g(x)$ are *not equal*.

Example 2 Consider the polynomials $f(x) = 3x^5 - 4x^2$ and $g(x) = x^2 + 3x$ in $\mathbf{Z}_5[x]$. By direct computation, we find that

$$\begin{aligned} f(0) &= 0 = g(0) & f(1) &= 4 = g(1) & f(2) &= 0 = g(2) \\ f(3) &= 3 = g(3) & f(4) &= 3 = g(4). \end{aligned}$$

Thus $f(c) = g(c)$ for all c in \mathbf{Z}_5 , but $f(x) \neq g(x)$ in $\mathbf{Z}_5[x]$. ■

The next two theorems are two of the simplest and most useful results on factorization in $F[x]$.

Theorem 8.15 ■ The Remainder Theorem

If $f(x)$ is a polynomial over the field F and $c \in F$, then the remainder in the division of $f(x)$ by $x - c$ is $f(c)$.

($u \wedge v \Rightarrow w$) **Proof** Since $x - c$ has degree 1, the remainder r in

$$f(x) = (x - c)q(x) + r$$

is a constant. Replacing x with c , we obtain

$$\begin{aligned} f(c) &= (c - c)q(c) + r \\ &= 0 \cdot q(c) + r \\ &= r. \end{aligned}$$

Thus $r = f(c)$.

Theorem 8.16 ■ The Factor Theorem

A polynomial $f(x)$ over the field F has a factor $x - c$ in $F[x]$ if and only if $c \in F$ is a zero of $f(x)$.

$p \Leftrightarrow q$ **Proof** From the Remainder Theorem, we have

$$f(x) = (x - c)q(x) + f(c).$$

Thus $x - c$ is a factor of $f(x)$ if and only if $f(c) = 0$.

The Factor Theorem can be extended as follows.

Theorem 8.17 ■ Factorization of $f(x)$ with Distinct Zeros

Let $f(x)$ be a polynomial over the field F that has positive degree n and leading coefficient a . If c_1, c_2, \dots, c_n are n distinct zeros of $f(x)$ in F , then

$$f(x) = a(x - c_1)(x - c_2) \cdots (x - c_n).$$

Induction **Proof** The proof is by induction on $n = \deg f(x)$. For each positive integer n , let S_n be the statement of the theorem.

For $n = 1$, suppose that $f(x)$ has degree 1 and leading coefficient a , and let c_1 be a zero of $f(x)$ in F . Then $f(x) = ax + b$, where $a \neq 0$ and $f(c_1) = 0$. This implies that $ac_1 + b = 0$ and $b = -ac_1$. Therefore, $f(x) = ax - ac_1 = a(x - c_1)$, and S_1 is true.

Assume now that S_k is true, and let $f(x)$ be a polynomial with leading coefficient a and degree $k + 1$ that has $k + 1$ distinct zeros $c_1, c_2, \dots, c_k, c_{k+1}$ in F . Since c_{k+1} is a zero of $f(x)$,

$$f(x) = (x - c_{k+1})q(x)$$

by the Factor Theorem. By Theorem 8.7, $q(x)$ must have degree k . Since the factor $x - c_{k+1}$ is monic, $q(x)$ and $f(x)$ have the same leading coefficient. For $i = 1, 2, \dots, k$, we have

$$(c_i - c_{k+1})q(c_i) = f(c_i) = 0,$$

where $c_i - c_{k+1} \neq 0$, since the zeros $c_1, c_2, \dots, c_k, c_{k+1}$ are distinct. Therefore, $q(c_i) = 0$ for $i = 1, 2, \dots, k$. That is, c_1, c_2, \dots, c_k are k distinct zeros of $q(x)$ in F . By the induction hypothesis,

$$q(x) = a(x - c_1)(x - c_2) \cdots (x - c_k).$$

Substitution of this factored expression for $q(x)$ in $f(x) = (x - c_{k+1})q(x)$ yields

$$f(x) = a(x - c_1)(x - c_2) \cdots (x - c_k)(x - c_{k+1}).$$

Therefore, S_{k+1} is true whenever S_k is true, and it follows by induction that S_n is true for all positive integers n .

The proof of the following corollary is left as an exercise.

Corollary 8.18 ■ Number of Distinct Zeros

A polynomial of positive degree n over the field F has at most n distinct zeros in F .

In the factorization of polynomials over a field F , the concept of an irreducible polynomial is analogous to the concept of a prime integer in the factorization of integers.

Definition 8.19 ■ Irreducible, Prime, Reducible

A polynomial $f(x)$ in $F[x]$ is **irreducible** (or **prime**) over F if $f(x)$ has positive degree and if $f(x)$ *cannot* be expressed as a product $f(x) = g(x)h(x)$ with both $g(x)$ and $h(x)$ of positive degree in $F[x]$. If $f(x)$ is not irreducible, then $f(x)$ is said to be **reducible**.

Example 3 Note that whether or not a given polynomial is irreducible over F depends on the field F . For instance, $x^2 + 1$ is irreducible over the field of real numbers, but it is reducible over the field \mathbf{C} of complex numbers, since $x^2 + 1$ can be factored as

$$x^2 + 1 = (x - i)(x + i)$$

in $\mathbf{C}[x]$. ■

If $g(x)$ and $h(x)$ are polynomials of positive degree, their product $g(x)h(x)$ has degree at least 2. Therefore, all polynomials of degree 1 are irreducible. Constant polynomials, however, are never irreducible because they do not have positive degree.

It is usually not easy to decide whether or not a given polynomial is irreducible over a certain field. However, the following theorem is sometimes quite helpful for polynomials with degree less than 4.

Theorem 8.20 ■ Polynomials of Degree 2 or 3

If $f(x)$ is a polynomial of degree 2 or 3 over the field F , then $f(x)$ is irreducible over F if and only if $f(x)$ has no zeros in F .

Proof Let $f(x)$ be a polynomial of degree 2 or 3 over the field F .

We shall prove the theorem in this form: $f(x)$ is reducible over F if and only if $f(x)$ has at least one zero in F .

$\sim p \Leftrightarrow \sim q$ Suppose first that $f(x)$ has a zero c in F . By the Factor Theorem,

$$f(x) = (x - c)q(x),$$

where $q(x)$ has degree one less than that of $f(x)$ by Theorem 8.7. This factorization shows that $f(x)$ is reducible over F .

$\sim p \Rightarrow \sim q$ Assume, conversely, that $f(x)$ is reducible over F . That is, there are polynomials $g(x)$ and $h(x)$ in $F[x]$ such that $f(x) = g(x)h(x)$, with both $g(x)$ and $h(x)$ of positive degree. By Theorem 8.7,

$$\deg f(x) = \deg g(x) + \deg h(x).$$

Since $\deg f(x)$ is either 2 or 3, one of the factors $g(x)$ and $h(x)$ must have degree 1. Without loss of generality, we may assume that this factor is $g(x)$, and we have

$$f(x) = (ax + b)h(x),$$

where $a \neq 0$. It follows at once from this equation that $-a^{-1}b$ is a zero of $f(x)$ in F , and the proof is complete.

Example 4 Let us determine whether each of the following polynomials is irreducible over \mathbf{Z}_5 .

a. $f(x) = x^3 + 2x^2 - 3x + 4$

b. $g(x) = x^2 + 3x + 4$

Routine computations show that

$$f(0) = 4, \quad f(1) = 4, \quad f(2) = 4, \quad f(3) = 0, \quad f(4) = 3.$$

Thus 3 is a zero of $f(x)$ in \mathbf{Z}_5 , and $f(x)$ is reducible over \mathbf{Z}_5 . However, $g(x)$ is irreducible over \mathbf{Z}_5 since $g(x)$ has no zeros in \mathbf{Z}_5 :

$$g(0) = 4, \quad g(1) = 3, \quad g(2) = 4, \quad g(3) = 2, \quad g(4) = 2. \quad \blacksquare$$

Irreducible polynomials play a role in the factorization of polynomials corresponding to the role that prime integers play in the factorization of integers. This is illustrated by the next theorem.

Theorem 8.21 ■ Irreducible Factors

If $p(x)$ is an irreducible polynomial over the field F and $p(x)$ divides $f(x)g(x)$ in $F[x]$, then either $p(x) | f(x)$ or $p(x) | g(x)$ in $F[x]$.

($u \wedge v \Rightarrow (w \vee z)$) **Proof** Assume that $p(x)$ is irreducible over F and that $p(x)$ divides $f(x)g(x)$; say,

$$f(x)g(x) = p(x)q(x)$$

for some $q(x)$ in $F[x]$. If $p(x) | f(x)$, the conclusion is satisfied. Suppose, then, that $p(x)$ does not divide $f(x)$. This means that 1 is the greatest common divisor of $f(x)$ and $p(x)$, since the only divisors of $p(x)$ with positive degree are constant multiples of $p(x)$. By Theorem 8.13, there exist $s(x)$ and $t(x)$ in $F[x]$ such that

$$1 = f(x)s(x) + p(x)t(x),$$

and this implies that

$$\begin{aligned} g(x) &= g(x)[f(x)s(x) + p(x)t(x)] \\ &= f(x)g(x)s(x) + p(x)g(x)t(x) \\ &= p(x)q(x)s(x) + p(x)g(x)t(x), \end{aligned}$$

since $f(x)g(x) = p(x)q(x)$. Factoring $p(x)$ from the two terms in the right member, we see that $p(x) \mid g(x)$:

$$g(x) = p(x)[q(x)s(x) + g(x)t(x)].$$

Thus $p(x)$ divides $g(x)$ if it does not divide $f(x)$.

A comparison of Theorem 8.21 with Theorem 2.16 provides an indication of how closely the theory of divisibility in $F[x]$ resembles the theory of divisibility in the integers. This analogy carries over to the proofs as well. For this reason, the proofs of the remaining results in this section are left as exercises.

Theorem 8.22 ■

Suppose $p(x)$ is an irreducible polynomial over the field F such that $p(x)$ divides a product $f_1(x)f_2(x) \cdots f_n(x)$ in $F[x]$, then $p(x)$ divides some $f_j(x)$.

Just as with integers, two nonzero polynomials $f(x)$ and $g(x)$ over the field F are called **relatively prime** over F if their greatest common divisor in $F[x]$ is 1.

Theorem 8.23 ■

If $f(x)$ and $g(x)$ are relatively prime polynomials over the field F and if $f(x) \mid g(x)h(x)$ in $F[x]$, then $f(x) \mid h(x)$ in $F[x]$.

Theorem 8.24 ■ Unique Factorization Theorem

Every polynomial of positive degree over the field F can be expressed as a product of its leading coefficient and a finite number of monic irreducible polynomials over F . This factorization is unique except for the order of the factors.

Of course, the monic irreducible polynomials involved in the factorization of $f(x)$ over F may not all be distinct. If $p_1(x), p_2(x), \dots, p_r(x)$ are the *distinct* monic irreducible factors of $f(x)$, then all repeated factors may be collected together and expressed by use of exponents to yield

$$f(x) = a[p_1(x)]^{m_1}[p_2(x)]^{m_2} \cdots [p_r(x)]^{m_r},$$

where each m_i is a positive integer.

In the last expression for $f(x)$, m_i is called the **multiplicity** of the factor $p_i(x)$. More generally, if $g(x)$ is an arbitrary polynomial of positive degree such that $[g(x)]^m$ divides $f(x)$ and no higher power of $g(x)$ divides $f(x)$ in $F[x]$, then $g(x)$ is said to be a factor of $f(x)$ over $F[x]$ with **multiplicity m**. Also, if c is an element of the field F such that $(x - c)^m$ divides $f(x)$ for some positive integer m but no higher power of $x - c$ divides $f(x)$, then c is called a **zero of multiplicity m**.

Example 5 We shall find the factorization that is described in the Unique Factorization Theorem for the polynomial

$$f(x) = 2x^4 + x^3 + 3x^2 + 2x + 4$$

over the field \mathbf{Z}_5 .

We first determine the zeros of $f(x)$ in \mathbf{Z}_5 :

$$f(0) = 4, \quad f(1) = 2, \quad f(2) = 0, \quad f(3) = 1, \quad f(4) = 1.$$

Thus 2 is the only zero of $f(x)$ in \mathbf{Z}_5 , and the Factor Theorem assures us that $x - 2$ is a factor of $f(x)$. Dividing by $x - 2$, we get

$$f(x) = (x - 2)(2x^3 + 3x + 3).$$

By Exercise 16 at the end of this section, the zeros of $f(x)$ are 2 and the zeros of $g(x) = 2x^3 + 3x + 3$. We therefore need to determine the zeros of $g(x)$, and the only possibility is 2, since this is the only zero of $f(x)$ in \mathbf{Z}_5 . We find that $g(2) = 0$, and this indicates that $x - 2$ is a factor of $g(x)$. Performing the required division, we obtain

$$2x^3 + 3x + 3 = (x - 2)(2x^2 + 4x + 1)$$

and

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

We now find that $2x^2 + 4x + 1$ is irreducible over \mathbf{Z}_5 , since it has no zeros in \mathbf{Z}_5 . To arrive at the desired factorization, we need only factor the leading coefficient of $f(x)$ from the factor $2x^2 + 4x + 1$:

$$\begin{aligned} f(x) &= (x - 2)^2(2x^2 + 4x + 1) \\ &= (x - 2)^2[2x^2 + 4x + (2)(3)] \\ &= 2(x - 2)^2(x^2 + 2x + 3). \end{aligned}$$
■

Exercises 8.3

True or False

Label each of the following statements as either true or false.

- For each c in a field F , the value $f(c) \in F$ is unique, where $f(x) = a_0 + a_1x + \cdots + a_nx^n \in F[x]$.
- We say that $c \in F$ is a solution to the polynomial equation $f(x) = 0$ if and only if $f(c) = 0$ in F .
- Let $f(x)$ and $g(x)$ be arbitrary polynomials over a field F . If $f(c) = g(c)$ for all $c \in F$, then $f(x) = g(x)$.
- Any polynomial of positive degree n over the field F has exactly n distinct zeros in F .
- There are nonzero elements in a field F that can be considered as irreducible polynomials in $F[x]$.

6. Since the polynomial $ax + b$ of degree 1 over a field F can be factored as $a(x + a^{-1}b)$, then $ax + b$ is not irreducible.
 7. Whether or not a given polynomial is irreducible over a field F depends on F .
 8. Any polynomial $f(x)$ of positive degree that is reducible over a field F has at least one zero in F .
-

Exercises

1. Determine the remainder r when $f(x)$ is divided by $x - c$ over the field F for the given $f(x)$, c , and F , where **R** denotes the field of real numbers and **C** the field of complex numbers.
 - a. $f(x) = x^4 - 7x^2 - 3x + 9$, $c = 2$, $F = \mathbf{R}$
 - b. $f(x) = 3x^5 - 2x^4 + 5x^2 + 2x - 1$, $c = -1$, $F = \mathbf{R}$
 - c. $f(x) = x^4 - ix^3 + 3x^2 - 3ix$, $c = i$, $F = \mathbf{C}$
 - d. $f(x) = x^4 - ix^3 + 3x^2 - 3ix$, $c = -i$, $F = \mathbf{C}$
 - e. $f(x) = x^5 + x^3 + x + 1$, $c = 1$, $F = \mathbf{Z}_3$
 - f. $f(x) = x^4 + x^3 + 2x^2 + 1$, $c = 2$, $F = \mathbf{Z}_3$
 - g. $f(x) = x^3 + 4x^2 + 2x + 1$, $c = 3$, $F = \mathbf{Z}_5$
 - h. $f(x) = 2x^4 + 3x^3 + 4x^2 + 3$, $c = 2$, $F = \mathbf{Z}_5$
 - i. $f(x) = x^4 + 5x^3 + 2x^2 + 6x + 2$, $c = 4$, $F = \mathbf{Z}_7$
 - j. $f(x) = x^3 + 6x^2 + 2x + 2$, $c = 5$, $F = \mathbf{Z}_7$
2. Let **Q** denote the field of rational numbers, **R** the field of real numbers, and **C** the field of complex numbers. Determine whether each of the following polynomials is irreducible over each of the indicated fields, and state all the zeros in each of the fields.
 - a. $x^2 - 2$ over **Q**, **R**, and **C**
 - b. $x^2 + 1$ over **Q**, **R**, and **C**
 - c. $x^2 + x - 2$ over **Q**, **R**, and **C**
 - d. $x^2 + 2x + 2$ over **Q**, **R**, and **C**
 - e. $x^2 + x + 2$ over \mathbf{Z}_3 , \mathbf{Z}_5 , and \mathbf{Z}_7
 - f. $x^2 + 2x + 2$ over \mathbf{Z}_3 , \mathbf{Z}_5 , and \mathbf{Z}_7
 - g. $x^3 - x^2 + 2x + 2$ over \mathbf{Z}_3 , \mathbf{Z}_5 , and \mathbf{Z}_7
 - h. $x^4 + 2x^2 + 1$ over \mathbf{Z}_3 , \mathbf{Z}_5 , and \mathbf{Z}_7
3. Find all monic irreducible polynomials of degree 2 over \mathbf{Z}_3 .
4. Write each of the following polynomials as a product of its leading coefficient and a finite number of monic irreducible polynomials over \mathbf{Z}_5 . State their zeros and the multiplicity of each zero.

a. $2x^3 + 1$	b. $3x^3 + 2x^2 + x + 2$
c. $3x^3 + x^2 + 2x + 4$	d. $2x^3 + 4x^2 + 3x + 1$

- e. $2x^4 + x^3 + 3x + 2$
 g. $x^4 + x^3 + x^2 + 2x + 3$
 i. $x^4 + 2x^3 + 3x + 4$

- f. $3x^4 + 3x^3 + x + 3$
 h. $x^4 + x^3 + 2x^2 + 3x + 2$
 j. $x^5 + x^4 + 3x^3 + 2x^2 + 4x$

5. Let F be a field and $f(x) = a_0 + a_1x + \dots + a_nx^n \in F[x]$.
- Prove that $x - 1$ is a factor of $f(x)$ if and only if $a_0 + a_1 + \dots + a_n = 0$.
 - Prove that $x + 1$ is a factor of $f(x)$ if and only if $a_0 - a_1 + \dots + (-1)^n a_n = 0$.
6. Prove Corollary 8.18: A polynomial of positive degree n over the field F has at most n distinct zeros in F .
7. Corollary 8.18 requires that F be a field. Show that each of the following polynomials of positive degree n has more than n zeros over F where F is not a field.
- $4x^2 + 4$ over \mathbf{Z}_8
 - $5x^3 + 3$ over \mathbf{Z}_{10}
8. Let $f(x)$ be an irreducible polynomial over a field F . Prove that $af(x)$ is irreducible over F for all nonzero a in F .
9. Let F be a field. Prove that if c is a zero of $f(x) = a_0 + a_1x + \dots + a_nx^n \in F[x]$, then c^{-1} is a zero of $a_n + a_{n-1}x + \dots + a_0x^n$.
10. Let $f(x)$ and $g(x)$ be two polynomials over the field F , both of degree n or less. Prove that if $m > n$ and if there exist m distinct elements c_1, c_2, \dots, c_m of F such that $f(c_i) = g(c_i)$ for $i = 1, 2, \dots, m$, then $f(x) = g(x)$.
11. Let p be a prime integer, and consider the polynomials $f(x) = x^p$ and $g(x) = x$ over the field \mathbf{Z}_p . Prove that $f(c) = g(c)$ for all c in \mathbf{Z}_p . (This result is another form of **Fermat's Little Theorem**: $n^p \equiv n \pmod{p}$. To prove it, consider the multiplicative group of nonzero elements of \mathbf{Z}_p .)
12. Find all the zeros of each of the following polynomials over the indicated fields.
- $x^5 - x$ over \mathbf{Z}_5
 - $x^{11} - x$ over \mathbf{Z}_{11}
13. Give an example of a polynomial of degree 4 over the field \mathbf{R} of real numbers that is reducible over \mathbf{R} and yet has no zeros in the real numbers.
14. If $f(x)$ and $g(x)$ are polynomials over the field F , and $h(x) = f(x) + g(x)$, prove that $h(c) = f(c) + g(c)$ for all c in F .
15. If $f(x)$ and $g(x)$ are polynomials over the field F , and $p(x) = f(x)g(x)$, prove that $p(c) = f(c)g(c)$ for all c in F .
16. Let $f(x)$ be a polynomial of positive degree n over the field F , and assume that $f(x) = (x - c)q(x)$ for some $c \in F$ and $q(x)$ in $F[x]$. Prove that
- c and the zeros of $q(x)$ in F are zeros of $f(x)$
 - $f(x)$ has no other zeros in F .
17. Suppose that $f(x)$, $g(x)$, and $h(x)$ are polynomials over the field F , each of which has positive degree, and that $f(x) = g(x)h(x)$. Prove that the zeros of $f(x)$ in F consist of the zeros of $g(x)$ in F together with the zeros of $h(x)$ in F .

Sec. 2.5, #51 ➤
 Sec. 4.4, #20 ➤

18. Prove that a polynomial $f(x)$ of positive degree n over the field F has at most n (not necessarily distinct) zeros in F .
19. Prove Theorem 8.22: Suppose $p(x)$ is an irreducible polynomial over the field F such that $p(x)$ divides a product $f_1(x)f_2(x) \cdots f_n(x)$ in $F[x]$, then $p(x)$ divides some $f_j(x)$.
20. Prove Theorem 8.23: If $f(x)$ and $g(x)$ are relatively prime polynomials over the field F and if $f(x) | g(x)h(x)$ in $F[x]$, then $f(x) | h(x)$ in $F[x]$.
21. Prove the Unique Factorization Theorem in $F[x]$ (Theorem 8.24).
22. Let $a \neq b$ in a field F . Show that $x + a$ and $x + b$ are relatively prime in $F[x]$.
23. Let $f(x), g(x), h(x) \in F[x]$ where $f(x)$ and $g(x)$ are relatively prime. If $h(x) | f(x)$, prove that $h(x)$ and $g(x)$ are relatively prime.
24. Let $f(x), g(x), h(x) \in F[x]$ where $f(x)$ and $g(x)$ are relatively prime. If $f(x) | h(x)$ and $g(x) | h(x)$, prove that $f(x)g(x) | h(x)$.
25. Let $f(x), g(x), h(x) \in F[x]$ where $f(x)$ and $g(x)$ are relatively prime and $f(x)$ and $h(x)$ are relatively prime. Prove that $f(x)$ and $g(x)h(x)$ are relatively prime.
26. Let $f(x), g(x) \in F[x]$ and $d(x)$ the greatest common divisor of $f(x)$ and $g(x)$ where $f(x) = h(x)d(x)$ and $g(x) = k(x)d(x)$ for some $h(x), k(x) \in F[x]$. Prove that $h(x)$ and $k(x)$ are relatively prime.

Sec. 8.2, #7–10, 35 ➤ 27. Find the least common multiple of each pair of polynomials given in Exercises 7–10 of Section 8.2.

8.4

Zeros of a Polynomial

We now focus our interest on polynomials that have their coefficients in the field **C** of complex numbers, the field **R** of real numbers, or the field **Q** of rational numbers. Our results are concerned with the zeros of these polynomials and the related property of irreducibility over these fields.

The statement in Theorem 8.25 is so important that it is known as the Fundamental Theorem of Algebra. It was first proved in 1799 by the great German mathematician Carl Friedrich Gauss (1777–1855). Unfortunately, all known proofs of this theorem require theories that we do not have at our disposal, so we are forced to accept the theorem without proof.

Theorem 8.25 ■ The Fundamental Theorem of Algebra

If $f(x)$ is a polynomial of positive degree over the field of complex numbers, then $f(x)$ has a zero in the complex numbers.

The Fundamental Theorem opens the door to a complete decomposition of any polynomial over **C**, as described in the following theorem.

Theorem 8.26 ■ Factorization over **C**

If $f(x)$ is a polynomial of positive degree n over the field **C** of complex numbers, then $f(x)$ can be factored as

$$f(x) = a(x - c_1)(x - c_2) \cdots (x - c_n),$$

where a is the leading coefficient of $f(x)$ and c_1, c_2, \dots, c_n are n (not necessarily distinct) complex numbers that are zeros of $f(x)$.

Induction Proof For each positive integer n , let S_n be the statement of the theorem.

If $n = 1$, then $f(x) = ax + b$, where $a \neq 0$. The complex number $c_1 = -a^{-1}b$ is a zero of $f(x)$, and

$$f(x) = ax + b = ax - ac_1 = a(x - c_1).$$

Thus S_1 is true.

Assume that S_k is true, and let $f(x)$ be a polynomial of degree $k + 1$ over **C**. By the Fundamental Theorem of Algebra, $f(x)$ has a zero c_1 in the complex numbers, and the Factor Theorem asserts that

$$f(x) = (x - c_1)q(x)$$

for some polynomial $q(x)$ over **C**. Since $x - c_1$ is monic, $q(x)$ has the same leading coefficient as $f(x)$, and Theorem 8.7 implies that $q(x)$ has degree k . By the induction hypothesis, $q(x)$ can be factored as the product of its leading coefficient and k factors of the form $x - c_i$:

$$q(x) = a(x - c_2)(x - c_3) \cdots (x - c_{k+1}).$$

Therefore,

$$\begin{aligned} f(x) &= (x - c_1)q(x) \\ &= a(x - c_1)(x - c_2) \cdots (x - c_{k+1}), \end{aligned}$$

and S_{k+1} is true. It follows that the theorem is true for all positive integers n .

As noted in the statement of Theorem 8.26, the zeros c_i are not necessarily distinct in the factorization of $f(x)$ that is described there. If the repeated factors are collected together, we have

$$f(x) = a(x - c_1)^{m_1}(x - c_2)^{m_2} \cdots (x - c_r)^{m_r}$$

as a standard form for the unique factorization of a polynomial over the complex numbers. In particular, we observe that *the only irreducible polynomials over C are the first-degree polynomials*.

With such a simple description of the irreducible polynomials over **C**, it is natural to ask which polynomials are irreducible over the real numbers. For polynomials of degree 2 (quadratic polynomials), an answer to this question is readily available from the *quadratic formula*. According to the **quadratic formula**, the zeros of a polynomial

$$f(x) = ax^2 + bx + c$$

with real coefficients[†] and $a \neq 0$ are given by

$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

These zeros are not real numbers if and only if the *discriminant*, $b^2 - 4ac$, is negative. Thus a quadratic polynomial is irreducible over the real numbers if and only if it has a negative discriminant.

If we introduce some appropriate terminology, a meaningful characterization of the field of complex numbers can now be formulated. If F and E are fields such that $F \subseteq E$, then E is called an **extension** of F . An element $a \in E$ is called **algebraic** over F if a is the zero of a polynomial $f(x)$ with coefficients in F , and E is an **algebraic extension** of F if every element of E is algebraic over F . E is **algebraically closed** if every polynomial of positive degree over E has a zero in E .

The field **C** of complex numbers can be characterized as a field with the following properties:

1. **C** is an algebraic extension of the field **R** of real numbers.
2. **C** is algebraically closed.

If $z = a + bi$ with $a, b \in \mathbf{R}$, then z is a zero of the polynomial

$$\begin{aligned} f(x) &= [x - (a + bi)][x - (a - bi)] \\ &= x^2 - 2ax + (a^2 + b^2) \end{aligned}$$

over **R**. Thus z is algebraic over **R**, and property 1 is established. The Fundamental Theorem of Algebra (Theorem 8.25) asserts that **C** is algebraically closed. It can be proved that any field that is an algebraic extension of **R** and is algebraically closed must be isomorphic to **C**. The proof of this assertion is beyond the scope of this text.

If a and b are real numbers, the *conjugate* of the complex number $z = a + bi$ is the complex number $\bar{z} = a - bi$. Note that the zeros r_1 and r_2 given by the quadratic formula are conjugates of each other when the coefficients are real and $b^2 - 4ac < 0$.

In the exercises at the end of this section, proofs are requested for the following facts concerning conjugates:

$$\begin{aligned} \overline{z_1 + z_2 + \cdots + z_n} &= \bar{z}_1 + \bar{z}_2 + \cdots + \bar{z}_n \\ \overline{z_1 \cdot z_2 \cdot \cdots \cdot z_n} &= \bar{z}_1 \cdot \bar{z}_2 \cdot \cdots \cdot \bar{z}_n. \end{aligned}$$

That is, the conjugate of a sum of terms is the sum of the conjugates of the individual terms, and the conjugate of a product of factors is the product of the conjugates of the individual factors. As a special case for products,

$$\overline{(z^n)} = (\bar{z})^n.$$

These properties of conjugates are used in the proof of the next theorem.

[†]The quadratic formula is also valid if the coefficients are complex numbers, but at the moment we are interested only in the real case.

Theorem 8.27 ■ Conjugate Zeros

Suppose that $f(x)$ is a polynomial that has all its coefficients in the real numbers. If the complex number z is a zero of $f(x)$, then its conjugate \bar{z} is also a zero of $f(x)$.

$p \Rightarrow q$ **Proof** Let $f(x) = \sum_{i=0}^n a_i x^i$, where all a_i are real, and assume that z is a zero of $f(x)$. Then $f(z) = 0$, and therefore,

$$\begin{aligned} 0 &= \overline{f(z)} \\ &= \overline{a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n} \\ &= \bar{a}_0 + \overline{a_1 z} + \overline{a_2 z^2} + \cdots + \overline{a_n z^n} \\ &= \bar{a}_0 + \bar{a}_1 \bar{z} + \bar{a}_2 (\bar{z})^2 + \cdots + \bar{a}_n (\bar{z})^n \\ &= a_0 + a_1 \bar{z} + a_2 (\bar{z})^2 + \cdots + a_n (\bar{z})^n, \end{aligned}$$

where the last equality follows from the fact that each a_i is a real number. We thus have $f(\bar{z}) = 0$, and the theorem is proved.

Example 1 The monic polynomial of least degree over the complex numbers that has $1 - i$ and $2i$ as zeros is

$$\begin{aligned} f(x) &= [x - (1 - i)][x - 2i] \\ &= x^2 - (1 + i)x + 2 + 2i. \end{aligned}$$

However, a polynomial with *real coefficients* that has $1 - i$ and $2i$ as zeros must also have $1 + i$ and $-2i$ as zeros. Thus the monic polynomial of least degree with real coefficients that has $1 - i$ and $2i$ as zeros is

$$\begin{aligned} g(x) &= [x - (1 - i)][x - (1 + i)][x - 2i][x + 2i] \\ &= (x^2 - 2x + 2)(x^2 + 4) \\ &= x^4 - 2x^3 + 6x^2 - 8x + 8. \end{aligned}$$

■

Example 2 Suppose that it is known that $1 - 2i$ is a zero of the fourth-degree polynomial $f(x) = x^4 - 3x^3 + x^2 + 7x - 30$ and that we wish to find all the zeros of $f(x)$. From Theorem 8.27, we know that $1 + 2i$ is also a zero of $f(x)$. The Factor Theorem then assures us that $x - (1 - 2i)$ and $x - (1 + 2i)$ are factors of $f(x)$:

$$f(x) = [x - (1 - 2i)][x - (1 + 2i)]q(x).$$

To find $q(x)$, we divide $f(x)$ by the polynomial

$$[x - (1 - 2i)][x - (1 + 2i)] = x^2 - 2x + 5$$

and obtain $q(x) = x^2 - x - 6$. Thus

$$\begin{aligned} f(x) &= [x - (1 - 2i)][x - (1 + 2i)](x^2 - x - 6) \\ &= [x - (1 - 2i)][x - (1 + 2i)](x - 3)(x + 2). \end{aligned}$$

It is now evident that the zeros of $f(x)$ are $1 - 2i$, $1 + 2i$, 3 , and -2 .

■

The results obtained thus far prepare for the next theorem, which describes a standard form for the unique factorization of a polynomial over the real numbers. The proof of this theorem is left as an exercise.

Theorem 8.28 ■ Factorization over \mathbf{R}

Every polynomial of positive degree over the field \mathbf{R} of real numbers can be factored as the product of its leading coefficient and a finite number of monic irreducible polynomials over \mathbf{R} , each of which is either quadratic or of first degree.

We restrict our attention now to the rational zeros of polynomials with rational coefficients and to the irreducibility of such polynomials. Neither the zeros of a polynomial nor its irreducibility are changed when it is multiplied by a nonzero constant, so we lose no generality by restricting our attention to polynomials with coefficients that are all integers.

Theorem 8.29 ■ Rational Zeros

Let

$$f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$$

be a polynomial of positive degree n with coefficients that are all integers, and let p/q be a rational number that has been written in lowest terms. If p/q is a zero of $f(x)$, then p divides a_0 and q divides a_n .

$u \Rightarrow w$ **Proof** Suppose that p/q is a rational number expressed in lowest terms that is a zero of $f(x) = \sum_{i=0}^n a_i x^i$. Then

$$a_0 + a_1\left(\frac{p}{q}\right) + \cdots + a_{n-1}\left(\frac{p}{q}\right)^{n-1} + a_n\left(\frac{p}{q}\right)^n = 0.$$

Multiplying both sides of this equality by q^n gives

$$a_0q^n + a_1pq^{n-1} + \cdots + a_{n-1}p^{n-1}q + a_np^n = 0.$$

Subtracting a_np^n from both sides, we have

$$a_0q^n + a_1pq^{n-1} + \cdots + a_{n-1}p^{n-1}q = -a_np^n,$$

and hence,

$$q(a_0q^{n-1} + a_1pq^{n-2} + \cdots + a_{n-1}p^{n-1}) = -a_np^n.$$

This shows that q divides a_np^n , and therefore $q|a_n$, since q and p are relatively prime.

$u \Rightarrow v$ Similarly, the equation

$$a_1pq^{n-1} + \cdots + a_{n-1}p^{n-1}q + a_np^n = -a_0q^n$$

can be used to show that $p|a_0$.

It is important to note that Theorem 8.29 only restricts the possibilities of the rational zeros. It does not guarantee that any of these possibilities is actually a zero of $f(x)$.

It may happen that when some of the rational zeros of a polynomial have been found, the remaining zeros may be obtained by use of the quadratic formula. This is illustrated in the next example.

Example 3 We shall obtain all zeros of the polynomial

$$f(x) = 2x^4 - 5x^3 + 3x^2 + 4x - 6$$

by first finding the rational zeros of $f(x)$. According to Theorem 8.29, any rational zero p/q of $f(x)$ that is in lowest terms must have a numerator p that divides the constant term and a denominator q that divides the leading coefficient. This means that

$$p \in \{\pm 1, \pm 2, \pm 3, \pm 6\}$$

$$q \in \{\pm 1, \pm 2\}$$

$$\frac{p}{q} \in \left\{ \pm \frac{1}{2}, \pm 1, \pm \frac{3}{2}, \pm 2, \pm 3, \pm 6 \right\}.$$

Testing the positive possibilities systematically, we get

$$f\left(\frac{1}{2}\right) = -\frac{15}{4}, f(1) = -2, f\left(\frac{3}{2}\right) = 0.$$

We could continue to test the remaining possibilities, but chances are that it is worthwhile to divide $f(x)$ by $x - (3/2)$ and then work with the quotient. Performing the division, we obtain

$$\begin{aligned} f(x) &= \left(x - \frac{3}{2}\right)(2x^3 - 2x^2 + 4) \\ &= (2x - 3)(x^3 - x^2 + 2). \end{aligned}$$

From this factorization, we see that the other zeros of $f(x)$ are the zeros of the factor $g(x) = x^3 - x^2 + 2$. Since this factor is monic, the only possible rational zeros are the divisors of 2. We already know that 1 is not a zero, since $f(1) = -2$. Thus the remaining possibilities are 2, -1, and -2. We find that

$$q(2) = 6, q(-1) = 0.$$

Therefore, $x + 1$ is a factor of $x^3 - x^2 + 2$. Division by $x + 1$ yields

$$x^3 - x^2 + 2 = (x + 1)(x^2 - 2x + 2)$$

and

$$f(x) = (2x - 3)(x + 1)(x^2 - 2x + 2).$$

The remaining zeros of $f(x)$ can be found by using the quadratic formula on the factor $x^2 - 2x + 2$:

$$x = \frac{2 \pm \sqrt{4 - 8}}{2} = 1 \pm i.$$

Thus the zeros of $f(x)$ are $3/2, -1, 1 + i$, and $1 - i$. ■

The results concerning irreducibility over the field **Q** of rational numbers are not nearly as neat or complete as those we have obtained for the fields **C** and **R**. The best-known result

for **Q** is a theorem that states what is known as *Eisenstein's Irreducibility Criterion*. To establish this result is the goal of the rest of this section. We need the following definition and two intermediate theorems to reach our objective.

Definition 8.30 ■ Primitive Polynomial

Let $f(x) = \sum_{i=0}^n a_i x^i$ be a polynomial in which all coefficients are integers. Then $f(x)$ is a **primitive** polynomial if the greatest common divisor of a_0, a_1, \dots, a_n is 1.

That is, a polynomial is primitive if and only if there is no prime integer that divides all of its coefficients.

Our first intermediate result simply asserts that the product of two primitive polynomials is primitive.

Theorem 8.31 ■ Product of Primitive Polynomials

$(p \wedge q) \Rightarrow r$ If $g(x)$ and $h(x)$ are primitive polynomials, then $g(x)h(x)$ is a primitive polynomial.

$(p \wedge q \wedge \sim r) \Rightarrow$ **Proof** We shall assume that the theorem is false and arrive at a contradiction.

$(\sim p \vee \sim q)$ Suppose that $g(x)$ and $h(x)$ are primitive polynomials, but the product $f(x) = g(x)h(x)$ is not primitive. Then there is a prime integer p that divides every coefficient of $f(x) = \sum_{i=0}^n a_i x^i$. The mapping $\phi: \mathbf{Z}[x] \rightarrow \mathbf{Z}_p[x]$ defined by

$$\phi(a_0 + a_1 x + \cdots + a_n x^n) = [a_0] + [a_1]x + \cdots + [a_n]x^n$$

is an epimorphism from $\mathbf{Z}[x]$ to $\mathbf{Z}_p[x]$, by Exercise 20 of Section 8.1. Since every coefficient of $f(x)$ is a multiple of p , $\phi(f(x)) = [0]$ in $\mathbf{Z}_p[x]$. Therefore,

$$\begin{aligned}\phi(g(x)) \cdot \phi(h(x)) &= \phi(g(x)h(x)) \\ &= \phi(f(x)) \\ &= [0]\end{aligned}$$

in $\mathbf{Z}_p[x]$. Since p is a prime, $\mathbf{Z}_p[x]$ is an integral domain, and either $\phi(g(x)) = [0]$ or $\phi(h(x)) = [0]$. Consequently, either p divides every coefficient of $g(x)$, or p divides every coefficient of $h(x)$. In either case, we have a contradiction to the supposition that $g(x)$ and $h(x)$ are primitive polynomials. This contradiction establishes the theorem.

The following theorem is credited to the same mathematician who first proved the Fundamental Theorem of Algebra.

Theorem 8.32 ■ Gauss's[†] Lemma

Let $f(x)$ be a primitive polynomial. If $f(x)$ can be factored as $f(x) = g(x)h(x)$, where $g(x)$ and $h(x)$ have rational coefficients and positive degree, then $f(x)$ can be factored as $f(x) = G(x)H(x)$, where $G(x)$ and $H(x)$ have integral coefficients and positive degree.

[†]A biographical sketch of Carl Friedrich Gauss (1777–1855) is given at the end of this chapter.

$p \Rightarrow q$ **Proof** Suppose that $f(x) = g(x)h(x)$ as described in the hypothesis. Let b be the least common denominator of the coefficients of $g(x)$, so that $g(x)$ can be expressed as $g(x) = \frac{1}{b}g_1(x)$, where $g_1(x)$ has integral coefficients. Now let a be the greatest common divisor of the coefficients of $g_1(x)$, so that $g_1(x) = aG(x)$, where $G(x)$ is a primitive polynomial. Then we have $g(x) = \frac{a}{b}G(x)$, where a and b are integers and $G(x)$ is primitive and of the same degree as $g(x)$. Similarly, we may write $h(x) = \frac{c}{d}H(x)$, where c and d are integers and $H(x)$ is primitive and of the same degree as $h(x)$. Substituting these expressions for $g(x)$ and $h(x)$, we obtain

$$f(x) = \frac{a}{b}G(x) \cdot \frac{c}{d}H(x),$$

and therefore,

$$bdf(x) = acG(x)H(x).$$

Since $f(x)$ is primitive, the greatest common divisor of the coefficients of the left member of this equation is bd . By Theorem 8.31, $G(x)H(x)$ is primitive, and therefore the greatest common divisor of the coefficients of the right member is ac . Hence $bd = ac$, and this implies that $f(x) = G(x)H(x)$, where $G(x)$ and $H(x)$ have integral coefficients and positive degrees.

Example 4 The polynomial $f(x) = x^5 + 2x^4 - 10x^3 - 9x^2 + 30x - 12$ is a primitive polynomial in $\mathbf{Z}[x]$ that can be factored as

$$f(x) = \left(\frac{2}{3}x^3 - 4x + 2\right)\left(\frac{3}{2}x^2 + 3x - 6\right),$$

where the factors on the right have rational coefficients and positive degrees. Using the same technique as in the proof of Gauss's Lemma, we can write

$$\begin{aligned} f(x) &= \frac{2}{3}(x^3 - 6x + 3)\left(\frac{3}{2}(x^2 + 2x - 4)\right) \\ &= (x^3 - 6x + 3)(x^2 + 2x - 4). \end{aligned}$$

Thus $f(x) = G(x)H(x)$, where $G(x) = x^3 - 6x + 3$ and $H(x) = x^2 + 2x - 4$ have integral coefficients and positive degree. ■

We are now in a position to prove Eisenstein's result.

Theorem 8.33 ■ Eisenstein's[†] Irreducibility Criterion

Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial of positive degree with integral coefficients. If there exists a prime integer p such that $p \mid a_i$ for $i = 0, 1, \dots, n-1$ but $p \nmid a_n$ and $p^2 \nmid a_0$, then $f(x)$ is irreducible over the field of rational numbers.

[†]Ferdinand Gotthold Max Eisenstein (1823–1852) was a German mathematician inspired to do mathematical research by Abel's proof of the impossibility of solving fifth-degree polynomials using only the operations of addition, subtraction, multiplication, division, and the extraction of roots. He experienced health problems throughout his life and died of tuberculosis at the age of 29.

Contradiction **Proof** Dividing out the greatest common divisor of the coefficients of a polynomial would have no effect on whether or not the criterion was satisfied by a prime p because of the requirement that $p \nmid a_n$. Therefore, we may restrict our attention to the case where $f(x)$ is a primitive polynomial.

Let $f(x) = \sum_{i=0}^n a_i x^i$ be a primitive polynomial, and assume there exists a prime integer p that satisfies the hypothesis. At the same time, assume that the conclusion is false, so that $f(x)$ factors over the rational numbers as a product of two polynomials of positive degree. Then $f(x)$ can be factored as the product of two polynomials of positive degree that have integral coefficients, by Theorem 8.32. Suppose that

$$f(x) = (b_0 + b_1 x + \cdots + b_r x^r)(c_0 + c_1 x + \cdots + c_s x^s),$$

where all the coefficients are integers and $r > 0, s > 0$. Then $a_0 = b_0 c_0$, and hence $p \mid b_0 c_0$, but $p^2 \nmid b_0 c_0$ by the hypothesis. This implies that either $p \mid b_0$ or $p \mid c_0$, but p does not divide both b_0 and c_0 . Without loss of generality, we may assume that $p \mid b_0$ and $p \nmid c_0$. If all of the b_i were divisible by p , then p would divide all the coefficients in the product, $f(x)$. Since $p \nmid a_n$, some of the b_i are not divisible by p . Let k be the smallest subscript such that $p \nmid b_k$, and consider

$$a_k = b_0 c_k + b_1 c_{k-1} + \cdots + b_{k-1} c_1 + b_k c_0.$$

By the choice of k , p divides each of b_0, b_1, \dots, b_{k-1} , and therefore,

$$p \mid (b_0 c_k + b_1 c_{k-1} + \cdots + b_{k-1} c_1).$$

Also, $p \mid a_k$, since $k < n$. Hence p divides the difference:

$$p \mid [a_k - (b_0 c_k + b_1 c_{k-1} + \cdots + b_{k-1} c_1)].$$

That is, $p \mid b_k c_0$. This is impossible, however, since $p \nmid b_k$ and $p \nmid c_0$. We have arrived at a contradiction, and therefore $f(x)$ is irreducible over the rational numbers.

Example 5 Consider the polynomial

$$f(x) = 10 - 15x + 25x^2 - 7x^4.$$

The prime integer $p = 5$ divides all of the coefficients in $f(x)$ except the leading coefficient $a_n = -7$, and 5^2 does not divide the constant term $a_0 = 10$. Therefore, $f(x)$ is irreducible over the rational numbers, by Eisenstein's Criterion. ■

Sometimes when Eisenstein's Irreducibility Criterion does not apply to a given polynomial, a change of variable will result in a polynomial for which Eisenstein's Irreducibility Criterion does apply, as shown in Example 6.

Example 6 Consider the polynomial

$$f(x) = x^4 + x^3 + 6x^2 - 14x + 16.$$

Eisenstein's Irreducibility Criterion does not apply to this polynomial. However, if we replace x by $x + 1$ in $f(x)$, we obtain

$$\begin{aligned} f(x+1) &= (x+1)^4 + (x+1)^3 + 6(x+1)^2 - 14(x+1) + 16 \\ &= x^4 + 5x^3 + 15x^2 + 5x + 10. \end{aligned}$$

Now 5 is prime and divides all the coefficients of $f(x+1)$ except the leading coefficient and $5^2 \nmid 10 = a_0$. Thus $f(x+1) = x^4 + 5x^3 + 15x^2 + 5x + 10$ is irreducible and hence, $f(x) = x^4 + x^3 + 6x^2 - 14x + 16$ is irreducible (see Exercise 33 at the end of this section). ■

We end this section with another technique for determining if a polynomial is irreducible over the field \mathbf{Q} of rational numbers.

Theorem 8.34 ■ Irreducibility of $f(x)$ in $\mathbf{Q}[x]$

Suppose $f(x) = a_0 + a_1x + \cdots + a_nx^n$ is a polynomial of positive degree with integral coefficients and p is a prime integer that does not divide a_n . Let

$$f_p(x) = [a_0] + [a_1]x + \cdots + [a_n]x^n$$

where $[a_i] \in \mathbf{Z}_p$ for $i = 0, 1, \dots, n$. If $f_p(x)$ is irreducible in $\mathbf{Z}_p[x]$, then $f(x)$ is irreducible in $\mathbf{Q}[x]$.

$\sim q \Rightarrow \sim p$ **Proof** Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial of positive degree with integral coefficients and define

$$f_p(x) = [a_0] + [a_1]x + \cdots + [a_n]x^n$$

where p is a prime integer that does not divide a_n . Assume $f(x)$ is reducible over \mathbf{Q} , that is, there exists polynomials $g(x), h(x)$ of positive degree in $\mathbf{Z}[x]$ such that $f(x) = g(x)h(x)$. The leading coefficient of the product $g(x)h(x)$ is the leading coefficient a_n of $f(x)$. Since p does not divide a_n , then p does not divide the leading coefficient of either $g(x)$ or of $h(x)$. Hence the leading coefficients of $g_p(x)$ and $h_p(x)$ are nonzero elements in \mathbf{Z}_p . Therefore the $\deg g_p(x) = \deg g(x) \geq 1$ and $\deg h_p(x) = \deg h(x) \geq 1$.

Now let $\phi: \mathbf{Z}[x] \rightarrow \mathbf{Z}_p[x]$ defined by $\phi(f(x)) = f_p(x)$. This mapping is an epimorphism (see Exercise 20 in Section 8.1). Thus

$$\begin{aligned} f_p(x) &= \phi(f(x)) \\ &= \phi(g(x)h(x)) \\ &= \phi(g(x))\phi(h(x)) \\ &= g_p(x)h_p(x), \end{aligned}$$

and $f_p(x)$ is reducible over \mathbf{Z}_p .

We illustrate the use of Theorem 8.34 in the last two examples of this section.

Example 7 Consider $f(x) = x^4 + 7x^3 - 4x^2 + 12x + 9$. Now $p = 2$ is a prime integer that does not divide $a_n = 1$ and

$$\begin{aligned}f_2(x) &= [1]x^4 + [7]x^3 - [4]x^2 + [12]x + [9] \\&= x^4 + x^3 + 1\end{aligned}$$

where we are writing a for $[a]$ in \mathbf{Z}_2 . Since $f_2(0) = 1$ and $f_2(1) = 1$, then $f_2(x)$ has no zeros and hence no first-degree factors in \mathbf{Z}_2 .

The only possible second-degree factors in \mathbf{Z}_2 are x^2 , $x^2 + x$, $x^2 + 1$ and $x^2 + x + 1$. Now $x^2 = x \cdot x$, $x^2 + x = x(x + 1)$ and $x^2 + 1 = (x + 1)^2$ are not factors of $f_2(x)$, since $f_2(x)$ has no first-degree factors. Long division shows that $x^2 + x + 1$ is not a factor of $f_2(x)$. Thus $f_2(x)$ is irreducible in \mathbf{Z}_2 and hence $f(x) = x^4 + 7x^3 - 4x^2 + 12x + 9$ is irreducible by Theorem 8.34. ■

Example 8 The polynomial $f(x) = x^3 + 3x + 5$ is irreducible since $f_2(x) = x^3 + x + 1$ is irreducible over \mathbf{Z}_2 . However $p = 3$ is also prime and $f_3(x) = x^3 + 2$ is not irreducible, since $x = 1$ is a zero of $f_3(x)$. Thus Theorem 8.34 does not require that $f_p(x)$ be irreducible for all positive primes. So finding a prime p such that $f_p(x)$ is reducible leads to no conclusion. ■

Exercises 8.4

True or False

Label each of the following statements as either true or false.

1. Every polynomial of positive degree over the complex numbers has a zero in the complex numbers.
2. The only irreducible polynomials over the complex numbers are of degree 1.
3. The field of complex numbers is an algebraic extension of the field of real numbers.
4. The field of real numbers is algebraically closed.
5. If $z = a + bi$ is a zero of a polynomial $f(x)$ with coefficients in the field \mathbf{C} , then \bar{z} is also a zero of $f(x)$ over \mathbf{C} .
6. Every polynomial of positive degree over the field \mathbf{R} of real numbers can be factored as the product of its leading coefficient and a finite number of monic irreducible polynomials of first degree over \mathbf{R} .
7. A polynomial is primitive if and only if there is no prime integer that divides all its coefficients.
8. The product of two primitive polynomials is primitive.
9. The sum of two primitive polynomials is primitive.
10. Every monic polynomial is primitive.
11. Every primitive polynomial is monic.
12. Every primitive polynomial is irreducible.

13. Every irreducible polynomial is primitive.
 14. A polynomial with real coefficients may have no real zeros.
 15. If z is a zero of multiplicity m of a polynomial $f(x)$ with coefficients in the field \mathbf{R} of real numbers, then \bar{z} is a zero of $f(x)$ of multiplicity m .
-

Exercises

1. Find a monic polynomial $f(x)$ of least degree over \mathbf{C} that has the given numbers as zeros, and a monic polynomial $g(x)$ of least degree with real coefficients that has the given numbers as zeros.

a. $2i, 3$ c. $2, 1 - i$ e. $3i, 1 + 2i$ g. $2 + i, -i$, and 1	b. $-3i, 4$ d. $3, 2 - i$ f. $i, 2 - i$ h. $3 - i, i$, and 2
--	--
2. One of the zeros is given for each of the following polynomials. Find the other zeros in the field of complex numbers.

a. $x^3 - 4x^2 + 6x - 4$; $1 - i$ is a zero. b. $x^3 + x^2 - 4x + 6$; $1 - i$ is a zero. c. $x^4 + x^3 + 2x^2 + x + 1$; $-i$ is a zero. d. $x^4 + 3x^3 + 6x^2 + 12x + 8$; $2i$ is a zero.
--

Find all rational zeros of each of the polynomials in Exercises 3–6.

3. $2x^3 - x^2 - 8x - 5$ 5. $2x^4 - x^3 - x^2 - x - 3$	4. $3x^3 + 19x^2 + 30x + 8$ 6. $2x^4 + x^3 - 8x^2 + x - 10$
---	--

In Exercises 7–12, find all zeros of the given polynomial.

7. $x^3 + x^2 - x + 2$ 9. $3x^3 + 2x^2 - 7x + 2$ 11. $6x^3 + 11x^2 + x - 4$	8. $3x^3 - 7x^2 + 8x - 2$ 10. $3x^3 - 2x^2 - 7x - 2$ 12. $9x^3 + 27x^2 + 8x - 20$
--	--

Factor each of the polynomials in Exercises 13–16 as a product of its leading coefficient and a finite number of monic irreducible polynomials over the field of rational numbers.

- | | |
|--|---|
| 13. $x^4 - x^3 - 2x^2 + 6x - 4$
15. $2x^4 + 5x^3 - 7x^2 - 10x + 6$
17. Show that each of the following polynomials is irreducible over the field of rational numbers. | 14. $2x^4 - x^3 - 13x^2 + 5x + 15$
16. $6x^4 + x^3 + 3x^2 - 14x - 8$
a. $3 + 9x + x^3$
c. $3 - 27x^2 + 2x^5$ |
|--|---|

- 18.** Show that the converse of Eisenstein's Irreducibility Criterion is not true by finding an irreducible $f(x) \in \mathbf{Q}[x]$ such that there is no p that satisfies the hypothesis of Eisenstein's Irreducibility Criterion.
- 19.** Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial of positive degree with integral coefficients. If there exists a prime integer p such that $p|a_i$ for $i = 1, 2, \dots, n$ but $p \nmid a_0$ and $p^2 \nmid a_n$, prove that $f(x)$ is irreducible over the field of rational numbers.
- 20.** Show that each of the following polynomials is irreducible over the field \mathbf{Q} of rational numbers.
- $1 + 2x + 6x^2 - 4x^3 + 2x^4$
 - $4 + 9x^2 - 15x^3 + 12x^4$
 - $6 - 35x + 14x^2 + 7x^5$
 - $12 + 22x - 55x^2 + 11x^4 + 33x^6$
- 21.** Use Theorem 8.34 to show that each of the following polynomials is irreducible over the field \mathbf{Q} of rational numbers.
- $f(x) = 27x^3 - 16x^2 + 3x - 25$
 - $f(x) = 8x^3 - 2x^2 - 5x + 10$
 - $f(x) = 12x^3 - 2x^2 + 15x - 2$
 - $f(x) = 30x^3 + 11x^2 - 2x + 8$
 - $f(x) = 3x^4 + 9x^3 - 7x^2 + 15x + 25$
 - $f(x) = 9x^5 - x^4 + 6x^3 + 5x^2 - x + 21$
- 22.** Show that the converse of Theorem 8.34 is not true by finding an irreducible $f(x)$ in $\mathbf{Q}[x]$, different from the $f(x)$ given in Example 8, such that $f_p(x)$ in $\mathbf{Z}_p[x]$ is reducible for a prime p that does not divide the leading coefficient of $f(x)$.
- 23.** Prove that $\overline{z_1 + z_2 + \cdots + z_n} = \overline{z_1} + \overline{z_2} + \cdots + \overline{z_n}$ for complex numbers z_1, z_2, \dots, z_n .
- 24.** Prove that $\overline{z_1 \cdot z_2 \cdot \cdots \cdot z_n} = \overline{z_1} \cdot \overline{z_2} \cdot \cdots \cdot \overline{z_n}$ for complex numbers z_1, z_2, \dots, z_n .
- 25.** Prove that for every positive integer n there exist polynomials of degree n that are irreducible over the rational numbers. (*Hint:* Consider $x^n - 2$.)
- 26.** Let $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n$ be a monic polynomial of positive degree n with coefficients that are all integers. Prove that any rational zero of $f(x)$ is an integer that divides the constant term a_0 .
- 27.** Derive the quadratic formula for the zeros of $ax^2 + bx + c$, where a , b , and c are complex numbers and $a \neq 0$.
- 28.** Prove Theorem 8.28. (*Hint:* In the factorization described in Theorem 8.26, pair those factors of the form $x - (a + bi)$ and $x - (a - bi)$.)
- 29.** Prove that any polynomial of odd degree that has real coefficients must have a zero in the field of real numbers.

30. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ in $\mathbf{R}[x]$. Prove that if $a_i \geq 0$ for all $i = 0, 1, \dots, n$ or if $a_i \leq 0$ for all $i = 0, 1, \dots, n$, then $f(x)$ has no positive zeros.
31. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ in $\mathbf{R}[x]$. Prove that if the coefficients a_i alternate in sign, where a zero coefficient can be considered as positive or negative to establish an alternating pattern, then $f(x)$ has no negative zeros.
32. Let a be in the field F . Define the mapping $\phi: F[x] \rightarrow F[x]$ by $\phi(f(x)) = f(x + a)$. Prove that ϕ is an automorphism.
33. Let $f(x) \in F[x]$ where F is a field and let $a \in F$. Prove that if $f(x + a)$ is irreducible over F , then $f(x)$ is irreducible over F .
34. Show that each of the following polynomials is irreducible over the field of rational numbers by making the appropriate change of variable and applying Eisenstein's Irreducibility Criterion.
- a. $x^3 + 3x + 8$ b. $x^3 + 5x^2 - 9x + 13$
- Sec. 2.2, #22 ➤ 35. Prove that $f(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible over \mathbf{Q} for any prime p .
(Hint: Note that $f(x) = (x^p - 1)/(x - 1)$ and consider $f(x + 1) = ((x + 1)^p - 1)/((x + 1) - 1)$. Use the Binomial Theorem and Eisenstein's Irreducibility Criterion.)

8.5

Solution of Cubic and Quartic Equations by Formulas (Optional)

In this section we focus on polynomials that have their coefficients in the field \mathbf{R} of real numbers. Up to this point, results have been stated with emphasis on the zeros of polynomials or on the related property of irreducibility.

We now place emphasis on a different point of view. Finding the zeros of a polynomial

$$f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$$

is equivalent to finding the solutions of the equation

$$a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n = 0.$$

Historically, mathematics developed with emphasis on the solution of equations.

The solution of linear equations

$$ax + b = 0$$

by the formula

$$x = -\frac{b}{a}$$

and the solution of quadratic equations

$$ax^2 + bx + c = 0$$

by the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

long ago prompted mathematicians to seek similar formulas for equations of higher degree with real coefficients.

In the 16th century, Italian mathematicians named Ferro, Tartaglia, Ferrari, and Cardano developed methods for solving third- and fourth-degree equations with real coefficients by the use of formulas that involved only the operations of addition, subtraction, multiplication, division, and the extraction of roots. For more than two hundred years afterward, mathematicians struggled to obtain similar formulas for equations with degree higher than 4 or to prove that such formulas did not exist. It was in the early 19th century that the Norwegian mathematician Abel[†] proved that it was impossible to obtain such formulas for equations with degree greater than 4.

The proof of Abel's result is beyond the level of this text, but the formulas for cubic and quartic (third- and fourth-degree) equations with real coefficients are within our reach.

We consider first the solution of the general cubic equation

$$a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

where the coefficients are real numbers and $a_3 \neq 0$. There is no loss of generality in assuming that the cubic polynomial is monic since division of both sides of the equation by a_3 yields an equivalent equation. Thus we assume an equation of the form

$$x^3 + ax^2 + bx + c = 0.$$

As would be expected, cube roots of complex numbers play a major role in the development. For this reason, some remarks on cube roots are in order.

An easy application of Theorem 7.11 yields the fact that the cube roots of 1 are given by

$$\begin{aligned} \cos 0 + i \sin 0 &= 1, \\ \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3} &= \frac{-1 + i\sqrt{3}}{2}, \\ \cos \frac{4\pi}{3} + i \sin \frac{4\pi}{3} &= \frac{-1 - i\sqrt{3}}{2}. \end{aligned}$$

If we let $\omega = (-1 + i\sqrt{3})/2$, the distinct cube roots of 1 are ω , ω^2 , and $\omega^3 = 1$. For an arbitrary nonzero complex number z , let $\sqrt[3]{z}$ denote any fixed cube root of z . Then each of the numbers $\sqrt[3]{z}$, $\omega\sqrt[3]{z}$, and $\omega^2\sqrt[3]{z}$ is a cube root of z , and they are clearly distinct. Thus the three cube roots of z are given by

$$\sqrt[3]{z}, \omega\sqrt[3]{z}, \omega^2\sqrt[3]{z},$$

where $\omega = (-1 + i\sqrt{3})/2$. This result is used in solving the cubic equation in Theorem 8.36.

[†]See the biographical sketch of Niels Henrik Abel at the end of Chapter 3.

The following two theorems lead to formulas for the solutions of the general cubic equation

$$x^3 + ax^2 + bx + c = 0.$$

Theorem 8.35 ■ Change of Variable in the Cubic

The change of variable

$$x = y - \frac{a}{3}$$

in $x^3 + ax^2 + bx + c = 0$ yields the equation

$$y^3 + py + q = 0,$$

where

$$p = b - \frac{a^2}{3}, \quad q = c - \frac{ab}{3} + \frac{2a^3}{27}.$$

$u \Rightarrow v$ **Proof** The theorem can be proved by direct substitution, but the details are neater if we first consider a substitution of the form $x = y + h$, where h is unspecified at this point. This substitution yields

$$(y + h)^3 + a(y + h)^2 + b(y + h) + c = 0.$$

When this equation is simplified, it appears as

$$y^3 + (3h + a)y^2 + (3h^2 + 2ah + b)y + (h^3 + ah^2 + bh + c) = 0.$$

If we let $h = -\frac{a}{3}$, the coefficients then simplify as follows:

$$\begin{aligned} 3h + a &= 3\left(-\frac{a}{3}\right) + a = 0 \\ 3h^2 + 2ha + b &= 3\left(\frac{a^2}{9}\right) + 2a\left(-\frac{a}{3}\right) + b = b - \frac{a^2}{3} \\ h^3 + ah^2 + bh + c &= -\frac{a^3}{27} + \frac{a^3}{9} - \frac{ab}{3} + c = c - \frac{ab}{3} + \frac{2a^3}{27}. \end{aligned}$$

This establishes the theorem.

Theorem 8.36 ■ Solutions to the Cubic Equation

Consider the equation $y^3 + py + q = 0$, and let

$$\omega = \frac{-1 + i\sqrt{3}}{2}, \quad A = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}, \quad B = -\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}.$$

The solutions to $y^3 + py + q = 0$ are given by

$$\sqrt[3]{A} + \sqrt[3]{B}, \quad \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B}, \quad \text{and} \quad \omega^2\sqrt[3]{A} + \omega\sqrt[3]{B},$$

where $\sqrt[3]{A}$ and $\sqrt[3]{B}$ denote (real or complex) cube roots of A and B chosen so that $\sqrt[3]{A}\sqrt[3]{B} = -\frac{p}{3}$.

$u \Rightarrow v$ **Proof** For an efficient proof, we resort to a “trick” substitution: We let

$$y = z - \frac{p}{3z}$$

in $y^3 + py + q = 0$. This substitution yields

$$\left(z - \frac{p}{3z}\right)^3 + p\left(z - \frac{p}{3z}\right) + q = 0.$$

This equation then simplifies to

$$z^3 - \frac{p^3}{27z^3} + q = 0$$

and then to

$$z^6 + qz^3 - \frac{p^3}{27} = 0.$$

This is a quadratic equation in z^3 , and we can use the quadratic formula to obtain

$$z^3 = \frac{-q \pm \sqrt{q^2 + \frac{4p^3}{27}}}{2} = -\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}.$$

With A and B as given in the statement of the theorem, we have

$$z^3 = A \quad \text{or} \quad z^3 = B.$$

Noting that

$$\begin{aligned} AB &= \left(-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}\right)\left(-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}\right) \\ &= \left(\frac{q}{2}\right)^2 - \left(\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3\right) \\ &= -\frac{p^3}{27}, \end{aligned}$$

we see that $\sqrt[3]{A}$ and $\sqrt[3]{B}$ need to be chosen so that

$$\sqrt[3]{A}\sqrt[3]{B} = -\frac{p}{3}.$$

With these choices made, the six solutions for z are given by

$$\sqrt[3]{A}, \quad \omega\sqrt[3]{A}, \quad \omega^2\sqrt[3]{A}, \quad \sqrt[3]{B}, \quad \omega\sqrt[3]{B}, \quad \omega^2\sqrt[3]{B}.$$

Substituting these values in

$$y = z - \frac{p}{3z} = z + \frac{\sqrt[3]{A}\sqrt[3]{B}}{z},$$

and using $\frac{1}{\omega} = \omega^2$ or $\frac{1}{\omega^2} = \omega$, we obtain the following three solutions for y :

$$\sqrt[3]{A} + \sqrt[3]{B}, \quad \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B}, \quad \text{and} \quad \omega^2\sqrt[3]{A} + \omega\sqrt[3]{B}.$$

Example 1 We shall use the formulas in Theorem 8.36 to solve the equation

$$y^3 - 9y - 12 = 0.$$

We have $p = -9$ and $q = -12$. Thus

$$A = \frac{12}{2} + \sqrt{(-6)^2 + (-3)^3} = 6 + \sqrt{9} = 9,$$

$$B = 6 - \sqrt{9} = 3,$$

and the real cube roots $\sqrt[3]{9}$ and $\sqrt[3]{3}$ satisfy $\sqrt[3]{A}\sqrt[3]{B} = -\frac{p}{3}$. The solutions are given by

$$\sqrt[3]{9} + \sqrt[3]{3},$$

$$\begin{aligned} \omega\sqrt[3]{9} + \omega^2\sqrt[3]{3} &= \left(\frac{-1 + i\sqrt{3}}{2}\right)\sqrt[3]{9} + \left(\frac{-1 - i\sqrt{3}}{2}\right)\sqrt[3]{3} \\ &= -\frac{1}{2}(\sqrt[3]{9} + \sqrt[3]{3}) + \frac{i\sqrt{3}}{2}(\sqrt[3]{9} - \sqrt[3]{3}), \\ \omega^2\sqrt[3]{9} + \omega\sqrt[3]{3} &= \left(\frac{-1 - i\sqrt{3}}{2}\right)\sqrt[3]{9} + \left(\frac{-1 + i\sqrt{3}}{2}\right)\sqrt[3]{3} \\ &= -\frac{1}{2}(\sqrt[3]{9} + \sqrt[3]{3}) - \frac{i\sqrt{3}}{2}(\sqrt[3]{9} - \sqrt[3]{3}). \end{aligned}$$
■

The results of Theorems 8.35 and 8.36 combine to yield the following theorem. The formulas in the theorem are known as **Cardano's Formulas**.

Theorem 8.37 ■ Cardano's[†] Formulas

The solutions to the cubic equation

$$x^3 + ax^2 + bx + c = 0$$

[†]Gerolamo Cardano (1501–1576) was an Italian Renaissance mathematician, physician, astrologer, and gambler who used his gambling expertise as a source of needed income. One of his books (published after his death) was an early treatment of probability that included information on cheating techniques for gambling. Cardano is credited with several inventions and he also published two natural science encyclopedias as well as several other works on a wide variety of subjects.

are given by

$$\sqrt[3]{A} + \sqrt[3]{B} - \frac{a}{3}, \quad \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B} - \frac{a}{3}, \quad \text{and} \quad \omega^2\sqrt[3]{A} + \omega\sqrt[3]{B} - \frac{a}{3},$$

where

$$\omega = \frac{-1 + i\sqrt{3}}{2}, \quad p = b - \frac{a^2}{3}, \quad q = c - \frac{ab}{3} + \frac{2a^3}{27},$$

$$A = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}, \quad B = -\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}$$

with $\sqrt[3]{A}$ and $\sqrt[3]{B}$ chosen so that

$$\sqrt[3]{A}\sqrt[3]{B} = -\frac{p}{3}.$$

The use of Theorem 8.37 is demonstrated in the following example.

Example 2 For the equation

$$x^3 - 3x^2 - 6x - 4 = 0,$$

we have $a = -3$, $b = -6$, and $c = -4$. The formulas in Theorem 8.37 yield

$$p = -6 - \frac{9}{3} = -9,$$

$$q = -4 - \frac{18}{3} - \frac{54}{27} = -12,$$

$$A = 6 + \sqrt{(-6)^2 + (-3)^3} = 9,$$

$$B = 6 - \sqrt{(-6)^2 + (-3)^3} = 3.$$

The real cube roots $\sqrt[3]{9}$ and $\sqrt[3]{3}$ satisfy $\sqrt[3]{A}\sqrt[3]{B} = -\frac{p}{3}$, and the solutions are given by

$$\sqrt[3]{9} + \sqrt[3]{3} + 1,$$

$$\omega\sqrt[3]{9} + \omega^2\sqrt[3]{3} + 1 = -\frac{1}{2}(\sqrt[3]{9} + \sqrt[3]{3} - 2) + \frac{i\sqrt{3}}{2}(\sqrt[3]{9} - \sqrt[3]{3}),$$

$$\omega^2\sqrt[3]{9} + \omega\sqrt[3]{3} + 1 = -\frac{1}{2}(\sqrt[3]{9} + \sqrt[3]{3} - 2) - \frac{i\sqrt{3}}{2}(\sqrt[3]{9} - \sqrt[3]{3}). \quad \blacksquare$$

We turn our attention now to the solution of quartic equations. As in the case of the cubic equation, there is no loss of generality in assuming that the equation is monic. Thus we assume an equation of the form

$$x^4 + ax^3 + bx^2 + cx + d = 0.$$

We find again that an appropriate substitution will remove the term of second-highest degree.

Theorem 8.38 ■ Change of Variable in the Quartic

The change of variable

$$x = y - \frac{a}{4}$$

in $x^4 + ax^3 + bx^2 + cx + d = 0$ yields an equation of the form

$$y^4 + py^2 + qy + r = 0.$$

Theorem 8.38 can be proved by direct substitution, and this proof is left as an exercise. In contrast to Theorem 8.35, we are not interested in formulas for p , q , and r at this time.

Consider now an equation of the form

$$y^4 + py^2 + qy + r = 0,$$

which can be written as

$$y^4 = -py^2 - qy - r.$$

The basic idea of our method, which was devised by Ferrari, is to add an expression to each side of the last equation that will make *both sides* perfect squares (squares of binomials). With this idea in mind, we add

$$ty^2 + \frac{t^2}{4}$$

to both sides, where t is yet to be determined. This gives

$$y^4 + ty^2 + \frac{t^2}{4} = -py^2 - qy - r + ty^2 + \frac{t^2}{4},$$

or

$$\left(y^2 + \frac{t}{2}\right)^2 = (t - p)y^2 - qy + \left(\frac{t^2}{4} - r\right).$$

We recall that a quadratic polynomial $Ay^2 + By + C$ is the square of a binomial

$$Ay^2 + By + C = (Dy + E)^2$$

if and only if $B^2 - 4AC = 0$. Thus

$$(t - p)y^2 - qy + \left(\frac{t^2}{4} - r\right) = (Dy + E)^2$$

if and only if

$$(-q)^2 - 4(t - p)\left(\frac{t^2}{4} - r\right) = 0.$$

This equation simplifies to the equation

$$t^3 - pt^2 - 4rt + 4rp - q^2 = 0,$$

which is known as the **resolvent equation** for $y^4 + py^2 + qy + r = 0$.

The resolvent equation can be solved for t by Cardano's method. Any one of the three solutions for t may be used in

$$\left(y^2 + \frac{t}{2}\right)^2 = (t - p)y^2 - qy + \left(\frac{t^2}{4} - r\right)$$

to obtain an equation of the form

$$\left(y^2 + \frac{t}{2}\right)^2 = (Dy + E)^2.$$

The solutions to the original equation can then be found by solving the two quadratic equations

$$y^2 + \frac{t}{2} = Dy + E \quad \text{and} \quad y^2 + \frac{t}{2} = -Dy - E.$$

The method is illustrated in the following example.

Example 3 We illustrate the preceding discussion by solving the equation

$$y^4 + y^2 - 2y + 6 = 0.$$

We have $p = 1$, $q = -2$, and $r = 6$. The resolvent equation is given by

$$t^3 - t^2 - 24t + 20 = 0.$$

We find that $t = 5$ is a solution to the resolvent equation, and the equation

$$\left(y^2 + \frac{t}{2}\right)^2 = (t - p)y^2 - qy + \left(\frac{t^2}{4} - r\right)$$

becomes

$$\left(y^2 + \frac{5}{2}\right)^2 = 4y^2 + 2y + \frac{1}{4} = \left(2y + \frac{1}{2}\right)^2.$$

Equating square roots, we obtain

$$y^2 + \frac{5}{2} = 2y + \frac{1}{2} \quad \text{or} \quad y^2 + \frac{5}{2} = -\left(2y + \frac{1}{2}\right)$$

and then

$$y^2 - 2y + 2 = 0 \quad \text{or} \quad y^2 + 2y + 3 = 0.$$

The quadratic formula then yields

$$y = 1 \pm i \quad \text{and} \quad y = -1 \pm i\sqrt{2}$$

as the solutions of the original equation. ■

We can now describe a method of solution for an arbitrary quartic equation

$$x^4 + ax^3 + bx^2 + cx + d = 0.$$

We first make the substitution

$$x = y - \frac{a}{4}$$

and obtain an equation for the form

$$y^4 + py^2 + qy + r = 0.$$

We next use the method of Example 3 to find the four solutions y_1, y_2, y_3 , and y_4 of the equation in y . Then the solutions to the original equation are given by

$$x_j = y_j - \frac{a}{4} \quad \text{for } j = 1, 2, 3, 4.$$

This is illustrated in Example 4.

Example 4 Consider the equation

$$x^4 + 4x^3 + 7x^2 + 4x + 6 = 0.$$

The substitution formula $x = y - \frac{a}{4}$ yields $x = y - 1$ and the resulting equation

$$(y - 1)^4 + 4(y - 1)^3 + 7(y - 1)^2 + 4(y - 1) + 6 = 0.$$

This equation simplifies to

$$y^4 + y^2 - 2y + 6 = 0.$$

From Example 3, the solutions to the last equation are

$$y_1 = 1 + i, \quad y_2 = 1 - i, \quad y_3 = -1 + i\sqrt{2}, \quad \text{and} \quad y_4 = -1 - i\sqrt{2}.$$

Hence the solutions $x_i = y_i - 1$ are given by

$$x_1 = i, \quad x_2 = -i, \quad x_3 = -2 + i\sqrt{2}, \quad \text{and} \quad x_4 = -2 - i\sqrt{2}. \quad \blacksquare$$

Just as the discriminant $b^2 - 4ac$ can be used to characterize the solutions of the quadratic equation $ax^2 + bx + c = 0$, the discriminant of a polynomial equation can be used to characterize its solutions. In particular, we will see that a cubic equation will have either exactly one real solution or exactly three real solutions. We begin with the next definition.

Definition 8.39 ■ Discriminant of a Cubic Polynomial

Let $f(y) = y^3 + py + q$ have zeros c_1, c_2 , and c_3 . The **discriminant** of $f(y)$ is D^2 where

$$D = \prod_{i < j} (c_i - c_j) = (c_1 - c_2)(c_1 - c_3)(c_2 - c_3).$$

The reason for defining the discriminant as D^2 , rather than as D , is because the sign of D depends on the order of the zeros. However, the sign of D^2 is independent on the order of the zeros.

Theorem 8.40 ■ Discriminant of a Cubic Polynomial

The discriminant of $f(y) = y^3 + py + q$ is $D^2 = -27q^2 - 4p^3$.

$p \Rightarrow q$ **Proof** Let c_1 , c_2 , and c_3 be zeros of $f(y) = y^3 + py + q$. Then we can write $f(y) = (y - c_1)(y - c_2)(y - c_3)$ where

$$c_1 = \sqrt[3]{A} + \sqrt[3]{B}$$

$$c_2 = \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B}$$

$$c_3 = \omega^2\sqrt[3]{A} + \omega\sqrt[3]{B}$$

and

$$A = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}$$

$$B = -\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}$$

$$\omega = \frac{-1 + i\sqrt{3}}{2}.$$

The discriminant is

$$D^2 = (c_1 - c_2)^2(c_1 - c_3)^2(c_2 - c_3)^2,$$

and using $\omega^3 = 1$, we have

$$\begin{aligned} c_1 - c_2 &= (\sqrt[3]{A} + \sqrt[3]{B}) - (\omega\sqrt[3]{A} + \omega^2\sqrt[3]{B}) \\ &= \sqrt[3]{A} + \sqrt[3]{B} - \omega\sqrt[3]{A} - \omega^2\sqrt[3]{B} \\ &= (1 - \omega)(\sqrt[3]{A} - \omega^2\sqrt[3]{B}) \end{aligned}$$

$$\begin{aligned} c_1 - c_3 &= (\sqrt[3]{A} + \sqrt[3]{B}) - (\omega^2\sqrt[3]{A} + \omega\sqrt[3]{B}) \\ &= \sqrt[3]{A} + \sqrt[3]{B} - \omega^2\sqrt[3]{A} - \omega\sqrt[3]{B} \\ &= -\omega^2(1 - \omega)(\sqrt[3]{A} - \omega\sqrt[3]{B}) \end{aligned}$$

$$\begin{aligned} c_2 - c_3 &= (\omega\sqrt[3]{A} + \omega^2\sqrt[3]{B}) - (\omega^2\sqrt[3]{A} + \omega\sqrt[3]{B}) \\ &= \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B} - \omega^2\sqrt[3]{A} - \omega\sqrt[3]{B} \\ &= \omega(1 - \omega)(\sqrt[3]{A} - \sqrt[3]{B}). \end{aligned}$$

Then

$$\begin{aligned}
 D &= -\omega^3(1 - \omega)^3(\sqrt[3]{A} - \omega^2\sqrt[3]{B})(\sqrt[3]{A} - \omega\sqrt[3]{B})(\sqrt[3]{A} - \sqrt[3]{B}) \\
 &= 3i\sqrt{3}(\sqrt[3]{A} - \omega^2\sqrt[3]{B})(\sqrt[3]{A} - \omega\sqrt[3]{B})(\sqrt[3]{A} - \sqrt[3]{B}) \\
 &= 3i\sqrt{3}(\sqrt[3]{A^3} - \omega^3\sqrt[3]{B^3} + \omega^3\sqrt[3]{A}\sqrt[3]{B^2} - \sqrt[3]{A^2}\sqrt[3]{B} \\
 &\quad - \omega\sqrt[3]{A^2}\sqrt[3]{B} + \omega\sqrt[3]{A}\sqrt[3]{B^2} - \omega^2\sqrt[3]{A^2}\sqrt[3]{B} + \omega^2\sqrt[3]{A}\sqrt[3]{B^2}) \\
 &= 3i\sqrt{3}(A - B + \sqrt[3]{A}\sqrt[3]{B^2} - \sqrt[3]{A^2}\sqrt[3]{B} \\
 &\quad - \omega(\sqrt[3]{A^2}\sqrt[3]{B} - \sqrt[3]{A}\sqrt[3]{B^2}) - \omega^2(\sqrt[3]{A^2}\sqrt[3]{B} - \sqrt[3]{A}\sqrt[3]{B^2})) \\
 &= 3i\sqrt{3}(A - B + (-1 - \omega - \omega^2)(\sqrt[3]{A^2}\sqrt[3]{B} - \sqrt[3]{A}\sqrt[3]{B^2})) \\
 &= 3i\sqrt{3}(A - B) \\
 &= 3i\sqrt{3}\sqrt{q^2 + \frac{4}{27}p^3}
 \end{aligned}$$

since $-1 - \omega - \omega^2 = 0$ and

$$\begin{aligned}
 A - B &= -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} - \left(-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}\right) \\
 &= 2\sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} \\
 &= \sqrt{q^2 + \frac{4}{27}p^3}.
 \end{aligned}$$

Thus

$$\begin{aligned}
 D^2 &= \left(3i\sqrt{3}\sqrt{q^2 + \frac{4}{27}p^3}\right)^2 \\
 &= -27q^2 - 4p^3.
 \end{aligned}$$

The result of Theorem 8.40 can be used to characterize the solutions to the polynomial equation $y^3 + py + q = 0$.

Theorem 8.41 ■ Real Solutions of a Cubic Equation

The equation $y^3 + py + q = 0$ has exactly three real solutions if and only if $D^2 \geq 0$; that is, if and only if $-27q^2 - 4p^3 \geq 0$.

$p \Rightarrow q$ **Proof** Let c_1 , c_2 , and c_3 be real solutions to $y^3 + py + q = 0$. Then

$$D = (c_1 - c_2)(c_1 - c_3)(c_2 - c_3)$$

is real, and the discriminant $D^2 \geq 0$.

$\sim p \Rightarrow \sim q$ Now assume that there are exactly one real solution c_1 and two nonreal solutions c_2 and c_3 . We know that the nonreal solutions must be conjugates, so let $c_2 = z = a + bi$ and $c_3 = \bar{z} = a - bi$. Then

$$\begin{aligned} D &= (c_1 - z)(c_1 - \bar{z})(z - \bar{z}) \\ &= (c_1 - (a + bi))(c_1 - (a - bi))(a + bi - (a - bi)) \\ &= 2bi((a - c_1)^2 + b^2) \end{aligned}$$

and

$$\begin{aligned} D^2 &= (2bi((a - c_1)^2 + b^2))^2 \\ &= -4b^2((a - c_1)^2 + b^2)^2 \\ &< 0 \end{aligned}$$

since $b \neq 0$. Thus, if there is a nonreal solution, then the discriminant is negative. It follows that if the discriminant is nonnegative, then the solutions must all be real.

We note that the discriminant for the polynomial $y^3 - 9y - 12$ in Example 1 with two nonreal zeros is $D^2 = -27q^2 - 4p^3 = -27(-12)^2 - 4(-9)^3 = -972 < 0$.

Exercises 8.5

True or False

Label each of the following statements as either true or false.

1. Every cubic equation over the reals has at least one real solution.
 2. Every quartic equation over the reals has at least one real solution.
 3. If the discriminant is positive for a quadratic or cubic polynomial over the reals, then all the zeros must be real.
 4. If the discriminant is negative for a quadratic or cubic polynomial over the reals, then all the zeros must be nonreal.
-

Exercises

In Exercises 1–18, use the techniques presented in this section to find all solutions of the given equation.

- | | |
|---|---|
| <ol style="list-style-type: none"> 1. $x^3 - 15x - 30 = 0$ 3. $x^3 - 12x - 20 = 0$ 5. $x^3 - 6x - 6 = 0$ 7. $x^3 + 9x + 6 = 0$ 9. $2x^3 + 6x - 3 = 0$ 11. $x^3 - 6x^2 + 33x - 92 = 0$ | <ol style="list-style-type: none"> 2. $x^3 - 9x + 12 = 0$ 4. $x^3 + 15x - 20 = 0$ 6. $x^3 + 6x - 2 = 0$ 8. $x^3 + 9x - 6 = 0$ 10. $2x^3 - 6x - 5 = 0$ 12. $x^3 + 3x^2 + 21x + 13 = 0$ |
|---|---|

13. $8x^3 + 12x^2 + 150x + 25 = 0$

15. $x^4 + x^2 - 2x + 6 = 0$

17. $x^4 + 4x^3 + 3x^2 + 4x + 2 = 0$

14. $8x^3 - 12x^2 + 54x - 9 = 0$

16. $x^4 - 2x^2 + 8x - 3 = 0$

18. $x^4 - 4x^3 + 4x^2 - 8x + 4 = 0$

In Exercises 19–24, characterize the solutions to the following equations by evaluating the discriminant D^2 .

19. $x^3 - 91x + 90 = 0$

21. $x^3 - 55x - 72 = 0$

23. $x^3 - 47x - 136 = 0$

25. Prove Theorem 8.38: The change of variable $x = y - \frac{a}{4}$ in

$$x^4 + ax^3 + bx^2 + cx + d = 0$$

yields an equation of the form

$$y^4 + py^2 + qy + r = 0.$$

26. Show that the change of variable $x = y - \frac{1}{n}a_{n-1}$ in

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 = 0$$

yields an equation of the form $y^n + 0 \cdot y^{n-1} + b_{n-2}y^{n-2} + \cdots + b_1y + b_0 = 0$ or

$$y^n + b_{n-2}y^{n-2} + \cdots + b_1y + b_0 = 0.$$

27. Derive the quadratic formula by using the change in variable $x = y - \frac{1}{2}\left(\frac{b}{a}\right)$ to transform the quadratic equation $x^2 + \frac{b}{a}x + \frac{c}{a} = 0$ into one involving the difference of two squares and solve the resulting equation.

28. Use the definition of the discriminant

$$D^2 = \prod_{i < j} (c_i - c_j)^2$$

to show that the discriminant of $x^2 + \left(\frac{b}{a}\right)x + \frac{c}{a}$ is $\left(\frac{b}{a}\right)^2 - 4\left(\frac{c}{a}\right)$.

8.6

Algebraic Extensions of a Field

Some of the results in Chapter 6 concerning ideals and quotient rings are put to good use in this section. Starting with an irreducible polynomial $p(x)$ over a field F , these results are used in the construction of a field which is an extension of F that contains a zero of $p(x)$.

As a special case of Definition 6.2, if $p(x)$ is a fixed polynomial over the field F , the *principal ideal* generated by $p(x)$ in $F[x]$ is the set

$$P = (p(x)) = \{f(x)p(x) \mid f(x) \in F[x]\},$$

which consists of all multiples of $p(x)$ by elements $f(x)$ of $F[x]$. Most of our work in this section is related to quotient rings of the form $F[x]/(p(x))$.

Theorem 8.42 ■ The Quotient Rings $F[x]/(p(x))$

Let $p(x)$ be a polynomial of positive degree over the field F . Then the quotient ring $F[x]/(p(x))$ is a commutative ring with unity that contains a subring that is isomorphic to F .

Proof For a fixed polynomial $p(x)$ in $F[x]$, let $P = (p(x))$. According to Theorem 6.4, the set $F[x]/P$ forms a ring with respect to addition defined by

$$[f(x) + P] + [g(x) + P] = (f(x) + g(x)) + P$$

and multiplication defined by

$$[f(x) + P][g(x) + P] = f(x)g(x) + P.$$

The ring $F[x]/P$ is commutative, since $f(x)g(x) = g(x)f(x)$ in $F[x]$, and $1 + P$ is the unity in $F[x]$.

Consider the nonempty subset F' of $F[x]/P$ that consists of all cosets of the form $a + P$ with $a \in F$:

$$F' = \{a + P \mid a \in F\}.$$

For arbitrary elements $a + P$ and $b + P$ of F' , the elements

$$(a + P) - (b + P) = (a - b) + P$$

and

$$(a + P)(b + P) = ab + P$$

are in F' since $a - b$ and ab are in F . Thus F' is a subring of $F[x]/P$, by Theorem 5.4. The unity $1 + P$ is in F' , and every nonzero element $a + P$ of F' has the multiplicative inverse $a^{-1} + P$ in F' . Hence F' is a field.

The mapping $\theta: F \rightarrow F'$ defined by

$$\theta(a) = a + P$$

is a homomorphism, since

$$\begin{aligned}\theta(a + b) &= (a + b) + P \\ &= (a + P) + (b + P) \\ &= \theta(a) + \theta(b)\end{aligned}$$

and

$$\begin{aligned}\theta(ab) &= ab + P \\ &= (a + P)(b + P) \\ &= \theta(a)\theta(b).\end{aligned}$$

It follows from the definition of F' that θ is an epimorphism. Since $p(x)$ has positive degree, 0 is the only element of F that is contained in P , and therefore,

$$\begin{aligned}\theta(a) = \theta(b) &\Leftrightarrow a + P = b + P \\ &\Leftrightarrow a - b \in P \\ &\Leftrightarrow a = b.\end{aligned}$$

Thus θ is an isomorphism from F to the subring F' of $F[x]/(p(x))$.

As we have done in similar situations in the past, we can now use the isomorphism θ in the preceding proof to identify $a \in F$ with $a + P$ in $F[x]/(p(x))$. This identification allows us to regard F as a subset of $F[x]/(p(x))$. This point of view is especially advantageous when the quotient ring $F[x]/(p(x))$ is a field.

Theorem 8.43 ■ $F[x]/(p(x))$ with $p(x)$ Irreducible

Let $p(x)$ be a polynomial of positive degree over the field F . Then the ring $F[x]/(p(x))$ is a field if and only if $p(x)$ is an irreducible polynomial over F .

u \Leftarrow v **Proof** As in the proof of Theorem 8.42, let $P = (p(x))$. Assume first that $p(x)$ is an irreducible polynomial over F . In view of Theorem 8.42, we need only show that any nonzero element $f(x) + P$ in $F[x]/P$ has a multiplicative inverse in $F[x]/P$. If $f(x) + P \neq P$, then $f(x)$ is not a multiple of $p(x)$, and this means that the greatest common divisor of $f(x)$ and $p(x)$ is 1, since $p(x)$ is irreducible. By Theorem 8.13, there exist $s(x)$ and $t(x)$ in $F[x]$ such that

$$f(x)s(x) + p(x)t(x) = 1.$$

Now $p(x)t(x) \in P$, so $p(x)t(x) + P = 0 + P$, and hence

$$\begin{aligned}1 + P &= [f(x)s(x) + p(x)t(x)] + P \\ &= [f(x)s(x) + P] + [p(x)t(x) + P] \\ &= [f(x)s(x) + P] + [0 + P] \\ &= f(x)s(x) + P \\ &= [f(x) + P][s(x) + P].\end{aligned}$$

Thus $s(x) + P = [f(x) + P]^{-1}$, and we have proved that $F[x]/P$ is a field.

~u \Leftarrow ~v Suppose now that $p(x)$ is reducible over F . Then there exist polynomials $g(x)$ and $h(x)$ of positive degree in $F[x]$ such that $p(x) = g(x)h(x)$. Since $\deg p(x) = \deg g(x) + \deg h(x)$ and all these degrees are positive, it must be true that $\deg g(x) < \deg p(x)$ and $\deg h(x) < \deg p(x)$. Therefore, neither $g(x)$ nor $h(x)$ is a multiple of $p(x)$. That is,

$$g(x) + P \neq P \quad \text{and} \quad h(x) + P \neq P,$$

but

$$\begin{aligned}[g(x) + P][h(x) + P] &= g(x)h(x) + P \\ &= p(x) + P \\ &= P.\end{aligned}$$

We have $g(x) + P$ and $h(x) + P$ as two nonzero elements of $F[x]/P$ whose product is zero. Hence $F[x]/P$ is not a field in this case, and the proof is complete.

If F and E are fields such that $F \subseteq E$, then E is called an **extension field** of F . With the identification that we have made between F and F' , the preceding theorem shows that $F[x]/(p(x))$ is an extension field of F if and only if $p(x)$ is an irreducible polynomial over F . The main significance of all this becomes clear in the proof of the next theorem, which is credited to the German mathematician Leopold Kronecker (1823–1891).

Theorem 8.44 ■ Extension Field Containing a Zero

If $p(x)$ is an irreducible polynomial over the field F , there exists an extension field of F that contains a zero of $p(x)$.

$u \Rightarrow v$ **Proof** For a given irreducible polynomial

$$p(x) = p_0 + p_1x + p_2x^2 + \cdots + p_nx^n$$

over the field F , let $P = (p(x))$ in $F[x]$ and let $\alpha = x + P$ in $F[x]/P$. From the definition of multiplication in $F[x]/P$, it follows that

$$\alpha^2 = (x + P)(x + P) = x^2 + P$$

and that

$$\alpha^i = x^i + P$$

for every positive integer i . By using the identification of $a \in F$ with $a + P$ in $F[x]/P$, we can write the polynomial

$$p(x) = p_0 + p_1x + p_2x^2 + \cdots + p_nx^n$$

in the form

$$p(x) = (p_0 + P) + (p_1 + P)x + (p_2 + P)x^2 + \cdots + (p_n + P)x^n.$$

Hence

$$\begin{aligned} p(\alpha) &= (p_0 + P) + (p_1 + P)\alpha + (p_2 + P)\alpha^2 + \cdots + (p_n + P)\alpha^n \\ &= (p_0 + P) + (p_1 + P)(x + P) + (p_2 + P)(x^2 + P) \\ &\quad + \cdots + (p_n + P)(x^n + P) \\ &= (p_0 + P) + (p_1x + P) + (p_2x^2 + P) + \cdots + (p_nx^n + P) \\ &= (p_0 + p_1x + p_2x^2 + \cdots + p_nx^n) + P \\ &= p(x) + P \\ &= 0 + P. \end{aligned}$$

Thus $p(\alpha)$ is the zero element of $F[x]/P$, and α is a zero of $p(x)$ in $F[x]/P$.

For a particular polynomial $p(x)$, explicit standard forms for the elements of the ring $F[x]/(p(x))$ can be given. Before going into this, we note that the ring $F[x]/(p(x))$ is unchanged if $p(x)$ is replaced by a multiple of the form $cp(x)$, with $c \neq 0$ in F . This follows from the fact that the ideal $P = (p(x))$, which consists of the set of all multiples of $p(x)$ in $F[x]$, is the same as the set of all multiples of $cp(x)$ in $F[x]$. In particular, we can choose c to

be the multiplicative inverse of the leading coefficient of $p(x)$, thereby obtaining a monic polynomial that gives the same ring $F[x]/P$ as $p(x)$ does. Thus there is no loss of generality in assuming from now on that $p(x)$ is a *monic* polynomial over F .

Before considering the general situation, we examine some particular cases in the following examples.

Example 1 Consider the monic irreducible polynomial

$$p(x) = x^2 + 2x + 2$$

over the field \mathbf{Z}_3 . We shall determine all the elements of the field $\mathbf{Z}_3[x]/(p(x))$ and, at the same time, construct addition and multiplication tables for this field.

Let $P = (p(x))$ and $\alpha = x + P$ in $\mathbf{Z}_3[x]/P$. We start construction of the addition table for $\mathbf{Z}_3[x]/P$ with the elements $0 = 0 + P$, $1 = 1 + P$, $2 = 2 + P$, and α . Filling out the table until closure is obtained, we pick up the new elements $\alpha + 1$, $\alpha + 2$, 2α , $2\alpha + 1$, and $2\alpha + 2$. The completed table in Figure 8.2 shows that the set

$$\{0, 1, 2, \alpha, \alpha + 1, \alpha + 2, 2\alpha, 2\alpha + 1, 2\alpha + 2\}$$

is closed under addition.

+	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
0	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
1	1	2	0	$\alpha + 1$	$\alpha + 2$	α	$2\alpha + 1$	$2\alpha + 2$	2α
2	2	0	1	$\alpha + 2$	α	$\alpha + 1$	$2\alpha + 2$	2α	$2\alpha + 1$
α	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$	0	1	2
$\alpha + 1$	$\alpha + 1$	$\alpha + 2$	α	$2\alpha + 1$	$2\alpha + 2$	2α	1	2	0
$\alpha + 2$	$\alpha + 2$	α	$\alpha + 1$	$2\alpha + 2$	2α	$2\alpha + 1$	2	0	1
2α	2α	$2\alpha + 1$	$2\alpha + 2$	0	1	2	α	$\alpha + 1$	$\alpha + 2$
$2\alpha + 1$	$2\alpha + 1$	$2\alpha + 2$	2α	1	2	0	$\alpha + 1$	$\alpha + 2$	α
$2\alpha + 2$	$2\alpha + 2$	2α	$2\alpha + 1$	2	0	1	$\alpha + 2$	α	$\alpha + 1$

■ **Figure 8.2**

Turning now to multiplication, we start with the same nine elements that occur in the addition table. In constructing this table, we make use of the fact that α is a zero of $p(x) = x^2 + 2x + 2$ in the following manner:

$$\alpha^2 + 2\alpha + 2 = 0 \Rightarrow \alpha^2 = -2\alpha - 2 = \alpha + 1.$$

That is, whenever α^2 occurs in a product, it is replaced by $\alpha + 1$. As an illustration, we have

$$\begin{aligned} (2\alpha + 1)(\alpha + 2) &= 2\alpha^2 + 2\alpha + 2 \\ &= 2(\alpha + 1) + 2\alpha + 2 \\ &= 2\alpha + 2 + 2\alpha + 2 \\ &= \alpha + 1. \end{aligned}$$

The completed table is shown in Figure 8.3.

.	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
2	0	2	1	2α	$2\alpha + 2$	$2\alpha + 1$	α	$\alpha + 2$	$\alpha + 1$
α	0	α	2α	$\alpha + 1$	$2\alpha + 1$	1	$2\alpha + 2$	2	$\alpha + 2$
$\alpha + 1$	0	$\alpha + 1$	$2\alpha + 2$	$2\alpha + 1$	2	α	$\alpha + 2$	2α	1
$\alpha + 2$	0	$\alpha + 2$	$2\alpha + 1$	1	α	$2\alpha + 2$	2	$\alpha + 1$	2α
2α	0	2α	α	$2\alpha + 2$	$\alpha + 2$	2	$\alpha + 1$	1	$2\alpha + 1$
$2\alpha + 1$	0	$2\alpha + 1$	$\alpha + 2$	2	2α	$\alpha + 1$	1	$2\alpha + 2$	α
$2\alpha + 2$	0	$2\alpha + 2$	$\alpha + 1$	$\alpha + 2$	1	2α	$2\alpha + 1$	α	2

■ Figure 8.3



Example 2 The polynomial $p(x) = x^2 + 1$ is not irreducible over the field \mathbf{Z}_2 , since $p(1) = 0$. We follow the same procedure as in Example 1 and construct addition and multiplication tables for the ring $\mathbf{Z}_2[x]/(p(x))$.

As before, let $P = (p(x))$ and $\alpha = x + P$ in $\mathbf{Z}_2[x]/P$. Extending an addition table until closure is obtained, we arrive at the table shown in Figure 8.4.

+	0	1	α	$\alpha + 1$
0	0	1	α	$\alpha + 1$
1	1	0	$\alpha + 1$	α
α	α	$\alpha + 1$	0	1
$\alpha + 1$	$\alpha + 1$	α	1	0

■ Figure 8.4

In making the multiplication table shown in Figure 8.5, we use the fact that $p(\alpha) = 0$ in this way:

$$\begin{aligned}\alpha^2 + 1 &= 0 \Rightarrow \alpha^2 = -1 \\ &\Rightarrow \alpha^2 = 1.\end{aligned}$$

.	0	1	α	$\alpha + 1$
0	0	0	0	0
1	0	1	α	$\alpha + 1$
α	0	α	1	$\alpha + 1$
$\alpha + 1$	0	$\alpha + 1$	$\alpha + 1$	0

■ Figure 8.5

Theorem 8.43 assures us that $\mathbf{Z}_2[x]/P$ is not a field, and the multiplication table confirms this fact by showing that $\alpha + 1$ does not have a multiplicative inverse. ■

The next theorem and its corollary set forth the standard forms for the elements of the ring $F[x]/(p(x))$ that we referred to earlier.

Theorem 8.45 ■ Elements of $F[x]/(p(x))$

Let $p(x)$ be a polynomial of positive degree n over the field F , and let $P = (p(x))$ in $F[x]$. Then each element of the ring $F[x]/P$ can be expressed uniquely in the form

$$(a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}) + P.$$

$u \Rightarrow v$ **Proof** Assume the hypothesis and let $f(x) + P$ be an arbitrary element in $F[x]/P$. By the Division Algorithm, there exist $q(x)$ and $r(x)$ in $F[x]$ such that

$$f(x) = p(x)q(x) + r(x),$$

where either $r(x) = 0$ or $\deg r(x) < n = \deg p(x)$. In either case, we may write

$$r(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}.$$

Since $p(x)q(x)$ is in P , $p(x)q(x) + P = 0 + P$, and therefore,

$$\begin{aligned} f(x) + P &= [p(x)q(x) + P] + [r(x) + P] \\ &= [0 + P] + [r(x) + P] \\ &= r(x) + P \\ &= (a_0 + a_1x + \cdots + a_{n-1}x^{n-1}) + P. \end{aligned}$$

Uniqueness

To show uniqueness, suppose that $f(x) + P = r(x) + P$ as before and also that $f(x) + P = g(x) + P$, where

$$g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_{n-1}x^{n-1}.$$

Then $r(x) + P = g(x) + P$, and therefore $r(x) - g(x)$ is in P . Each of $r(x)$ and $g(x)$ either is zero or has degree less than n , and this implies that the difference $r(x) - g(x)$ either is zero or has degree less than n . Since $P = (p(x))$ contains no polynomials with degree less than n , it must be true that $r(x) - g(x) = 0$, and $r(x) = g(x)$.

Corollary 8.46 ■ Elements of $F[x]/P$ as Polynomials

For a polynomial $p(x)$ of positive degree n over the field F , let $P = (p(x))$ in $F[x]$ and let $\alpha = x + P$ in $F[x]/P$. Then each element of the ring $F[x]/P$ can be expressed uniquely in the form

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1}.$$

$u \Rightarrow v$ **Proof** From the theorem, each $f(x) + P$ in $F[x]/P$ can be expressed uniquely in the form

$$\begin{aligned} f(x) + P &= (a_0 + a_1x + \cdots + a_{n-1}x^{n-1}) + P \\ &= (a_0 + P) + (a_1 + P)(x + P) + \cdots + (a_{n-1} + P)(x^{n-1} + P) \\ &= (a_0 + P) + (a_1 + P)\alpha + \cdots + (a_{n-1} + P)\alpha^{n-1} \\ &= a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}, \end{aligned}$$

where the last equality follows from the identification of a_i in F with $a_i + P$ in $F[x]/P$.

In Example 1, the polynomials $f(x)$ in $\mathbf{Z}_3[x]$ and the cosets $f(x) + P$ in $\mathbf{Z}_3[x]/P$ receded into the background once the notation $\alpha = x + P$ was introduced, and we ended up with a field whose elements had the form $a_0 + a_1\alpha$, with $a_i \in \mathbf{Z}_3$. This field $\mathbf{Z}_3(\alpha)$ of nine elements, given by

$$\mathbf{Z}_3(\alpha) = \{0, 1, 2, \alpha, \alpha + 1, \alpha + 2, 2\alpha, 2\alpha + 1, 2\alpha + 2\},$$

is called *the field obtained by adjoining a zero α of $x^2 + 2x + 2$ to \mathbf{Z}_3* .

In general, if $p(x)$ is an irreducible polynomial over the field F , the smallest field that contains both F and a zero α of $p(x)$ is denoted by $F(\alpha)$ and is referred to as the **field[†] obtained by adjoining α to the field F** . A field $F(\alpha)$ of this type is called a **simple algebraic extension** of F , and F is referred to as the **ground field**. Corollary 8.46 describes the standard form for the elements of $F(\alpha)$.

Example 3 The polynomial $p(x) = x^3 + 2x^2 + 4x + 2$ is irreducible over \mathbf{Z}_5 , since

$$p(0) = 2, \quad p(1) = 4, \quad p(2) = 1, \quad p(3) = 4, \quad p(4) = 4.$$

In the field $\mathbf{Z}_5(\alpha)$ obtained by adjoining a zero α of $p(x)$ to \mathbf{Z}_5 , we shall obtain a formula for the product of two arbitrary elements $a_0 + a_1\alpha + a_2\alpha^2$ and $b_0 + b_1\alpha + b_2\alpha^2$.

In order to accomplish this objective, we first express α^3 and α^4 as polynomials in α with degrees less than 3. Since $p(\alpha) = 0$, we have

$$\begin{aligned} \alpha^3 + 2\alpha^2 + 4\alpha + 2 = 0 &\Rightarrow \alpha^3 = -2\alpha^2 - 4\alpha - 2 \\ &= 3\alpha^2 + \alpha + 3. \end{aligned}$$

Hence

$$\begin{aligned} \alpha^4 &= \alpha(3\alpha^2 + \alpha + 3) \\ &= 3\alpha^3 + \alpha^2 + 3\alpha \\ &= 3(-2\alpha^2 - 4\alpha - 2) + \alpha^2 + 3\alpha \\ &= 4\alpha^2 + 3\alpha + 4 + \alpha^2 + 3\alpha \\ &= \alpha + 4. \end{aligned}$$

Using these results, we get

$$\begin{aligned} (a_0 + a_1\alpha + a_2\alpha^2)(b_0 + b_1\alpha + b_2\alpha^2) &= a_0b_0 + (a_0b_1 + a_1b_0)\alpha + (a_0b_2 + a_1b_1 + a_2b_0)\alpha^2 \\ &\quad + (a_1b_2 + a_2b_1)\alpha^3 + a_2b_2\alpha^4 \\ &= a_0b_0 + (a_0b_1 + a_1b_0)\alpha + (a_0b_2 + a_1b_1 + a_2b_0)\alpha^2 \\ &\quad + (a_1b_2 + a_2b_1)(3\alpha^2 + \alpha + 3) + a_2b_2(\alpha + 4) \\ &= (a_0b_0 + 3a_1b_2 + 3a_2b_1 + 4a_2b_2) \\ &\quad + (a_0b_1 + a_1b_0 + a_1b_2 + a_2b_1 + a_2b_2)\alpha \\ &\quad + (a_0b_2 + a_1b_1 + a_2b_0 + 3a_1b_2 + 3a_2b_1)\alpha^2. \end{aligned}$$
■

[†]The existence of such a field $F(\alpha)$ is ensured by Theorem 8.44.

Example 4 With $\mathbf{Z}_5(\alpha)$ as in Example 3, suppose that we wish to find the multiplicative inverse of the element $\alpha^2 + 3\alpha + 1$ in the field $\mathbf{Z}_5(\alpha)$.

The polynomials $f(x) = x^2 + 3x + 1$ and $p(x) = x^3 + 2x^2 + 4x + 2$ are relatively prime over \mathbf{Z}_5 , so there exist $s(x)$ and $t(x)$ in $\mathbf{Z}_5[x]$ such that

$$f(x)s(x) + p(x)t(x) = 1,$$

by Theorem 8.13. Since $p(\alpha) = 0$, this means that

$$f(\alpha)s(\alpha) = 1$$

and that $(\alpha^2 + 3\alpha + 1)^{-1} = [f(\alpha)]^{-1} = s(\alpha)$. In order to find $s(x)$ and $t(x)$, we use the Euclidean Algorithm:

$$\begin{aligned} p(x) &= f(x)(x + 4) + (x + 3) \\ f(x) &= (x + 3)(x) + 1. \end{aligned}$$

Thus

$$\begin{aligned} 1 &= f(x) - x(x + 3) \\ &= f(x) - x[p(x) - f(x)(x + 4)] \\ &= f(x)[1 + x(x + 4)] + p(x)(-x) \\ &= f(x)(x^2 + 4x + 1) + p(x)(-x), \end{aligned}$$

so we have $s(x) = x^2 + 4x + 1$ and $t(x) = -x$. Therefore,

$$(\alpha^2 + 3\alpha + 1)^{-1} = s(\alpha) = \alpha^2 + 4\alpha + 1.$$

The result may be checked by computing the product

$$(\alpha^2 + 3\alpha + 1)(\alpha^2 + 4\alpha + 1)$$

in $\mathbf{Z}_5(\alpha)$. ■

It is of some interest to consider an example similar to Example 4 but in a more familiar setting.

Example 5 The polynomial $p(x) = x^2 - 2$ is irreducible over the field \mathbf{Q} of rational numbers. In the field $\mathbf{Q}(\sqrt{2})$ obtained by adjoining a zero $\alpha = \sqrt{2}$ of $p(x)$ to \mathbf{Q} , let us find the multiplicative inverse of the element $4 + 3\sqrt{2}$ by the method employed in Example 4. The polynomials $f(x) = 3x + 4$ and $p(x) = x^2 - 2$ are relatively prime over \mathbf{Q} . To find $s(x)$ and $t(x)$ such that

$$f(x)s(x) + p(x)t(x) = 1,$$

we need only one step in the Euclidean Algorithm:

$$p(x) = f(x) \cdot \left(\frac{1}{3}x - \frac{4}{9} \right) + \left(-\frac{2}{9} \right).$$

Multiplying by 9/2 and rewriting this equation, we obtain

$$f(x) \cdot \left(\frac{3}{2}x - 2 \right) + p(x) \left(-\frac{9}{2} \right) = 1.$$

Since $p(\sqrt{2}) = 0$, this gives

$$f(\sqrt{2}) \cdot \left(\frac{3}{2}\sqrt{2} - 2 \right) = 1$$

and

$$(4 + 3\sqrt{2})^{-1} = [f(\sqrt{2})]^{-1} = \frac{3}{2}\sqrt{2} - 2.$$

This agrees with the result obtained by the usual procedure of rationalizing the denominator:

$$\begin{aligned} \frac{1}{4 + 3\sqrt{2}} &= \frac{(1)(4 - 3\sqrt{2})}{(4 + 3\sqrt{2})(4 - 3\sqrt{2})} = \frac{4 - 3\sqrt{2}}{-2} \\ &= \frac{3}{2}\sqrt{2} - 2. \end{aligned}$$
■

The result in Theorem 8.44 generalizes to the following theorem.

Theorem 8.47 ■ Splitting Field

If $p(x)$ is a polynomial of positive degree n over the field F , there exists an extension field E of F that contains n zeros of $p(x)$.

Complete
Induction

Proof The proof is by induction on the degree n of $p(x)$. If $n = 1$, then $p(x)$ has the form $p(x) = ax + b$, with $a \neq 0$. Since $p(x)$ has the unique zero $-a^{-1}b$ in F , the theorem is true for $n = 1$.

Assume the theorem is true for all polynomials of degree less than k , and let $p(x)$ be a polynomial of degree k . We consider two cases, depending on whether $p(x)$ is irreducible.

If $p(x)$ is irreducible, then there exists an extension field E_1 of F that contains a zero α of $p(x)$, by Theorem 8.44. By the Factor Theorem,

$$p(x) = (x - \alpha)q(x),$$

where $q(x)$ must have degree $k - 1$, according to Theorem 8.7. Since $q(x)$ is a polynomial over E_1 that has degree less than k , the induction hypothesis applies to $q(x)$ over E_1 , and there exists an extension field E of E_1 such that $q(x)$ has $k - 1$ zeros in E . By Exercise 16 of Section 8.3, the zeros of $p(x)$ in E consist of α and the zeros of $q(x)$ in E . Thus $p(x)$ has k zeros in E .

If $p(x)$ is reducible, then $p(x)$ can be factored as a product $p(x) = g(x)h(x)$, where $n_1 = \deg g(x)$ and $n_2 = \deg h(x)$ are positive integers such that $n_1 + n_2 = k$. Since $n_1 < k$, the induction hypothesis applies to $g(x)$ over F , and there exists an extension field E_1 of F that contains n_1 zeros of $g(x)$. Now $h(x)$ is a polynomial of degree $n_2 < k$ over E_1 , so the induction hypothesis applies again to $h(x)$ over E_1 , and there exists an extension field E of E_1 such that $h(x)$ has n_2 zeros in E . By Exercise 17 of Section 8.3, the zeros of $p(x)$ in E

consist of the zeros of $g(x)$ in E together with the zeros of $h(x)$ in E . There are altogether $n_1 + n_2 = k$ of these zeros in E .

In either case, we have proved the existence of an extension field of F that contains k zeros of $p(x)$, and the theorem follows by induction.

If E is a field that contains all the zeros of a polynomial $p(x)$, and if no proper subfield of E contains all of these zeros, then E is called the **splitting field** of $p(x)$ because it is the “smallest” field over which $p(x)$ “splits” into first-degree factors. When considering $x^2 + 1$ as a polynomial over the field **R** of real numbers, then **R**(i), or the field **C** of complex numbers, is the splitting field for $x^2 + 1$ where

$$x^2 + 1 = (x - i)(x + i).$$

However, if $x^2 + 1$ is considered a polynomial over the field **Q** of rational numbers, then the splitting field of $x^2 + 1$ is **Q**(i), a proper subset of **C**.

The basic facts about zeros of polynomials have been presented in this chapter. The two most important facts are found in Theorems 8.26 and 8.47. Theorem 8.26 asserts that for any polynomial $p(x)$ of positive degree n over **C**, the field **C** contains n zeros of $p(x)$. Theorem 8.47 states that for an arbitrary field F and any polynomial $p(x)$ of positive degree n over F , there exists an extension field of F that contains n zeros of $p(x)$.

Important as it is, the material in this chapter is only a small part of the knowledge about extension fields. The study of extension fields leads into the area of mathematics known as *Galois*[†] theory. Interesting results concerning some ancient problems lie in this direction. One of these results is that it is impossible to trisect an arbitrary angle using only a straightedge and a compass. Another is that it is impossible to express the zeros of the general equation of degree 5 or more by formulas that use only the four basic arithmetic operations and extraction of roots.

The end of this book is actually a beginning. It is a gateway to higher mathematics courses in several directions, especially those in abstract algebra and linear algebra. These higher-level courses are more theoretical and stimulating intellectually, and they might well lead to a lifelong interest in mathematics.

Exercises 8.6

True or False

Label each of the following statements as either true or false.

1. Every polynomial equation of degree n over a field F can be solved over an extension field E of F .
2. If $p(x)$ is an irreducible polynomial over a field F , then the largest field that contains both F and a zero α of $p(x)$ is $F(\alpha)$.
3. Let F be a field. If $p(x)$ is reducible over F , the quotient ring $F[x]/(p(x))$ is also a field.

[†]Évariste Galois (1811–1832) was a French mathematician who solved the problem of finding a necessary and sufficient condition for solving polynomials by radicals and laid the foundations for Galois theory. He died at the age of 20 from wounds suffered in a duel.

Exercises

1. Each of the following polynomials $p(x)$ is irreducible over \mathbf{Z}_3 . For each of these polynomials, find all the elements of $\mathbf{Z}_3[x]/(p(x))$ and construct addition and multiplication tables for this field.
 - a. $p(x) = x^2 + x + 2$
 - b. $p(x) = x^2 + 1$
2. In each of the following parts, a polynomial $p(x)$ over a field F is given. Construct addition and multiplication tables for the ring $F[x]/(p(x))$ in each case and decide whether this ring is a field.
 - a. $p(x) = x^2 + x + 1$ over $F = \mathbf{Z}_2$
 - b. $p(x) = x^3 + 1$ over $F = \mathbf{Z}_2$
 - c. $p(x) = x^3 + x + 1$ over $F = \mathbf{Z}_2$
 - d. $p(x) = x^3 + x^2 + 1$ over $F = \mathbf{Z}_2$
 - e. $p(x) = x^2 + x + 1$ over $F = \mathbf{Z}_3$
 - f. $p(x) = x^2 + 2$ over $F = \mathbf{Z}_3$

In Exercises 3–6, a field F , a polynomial $p(x)$ over F , and an element of the field $F(\alpha)$ obtained by adjoining a zero α of $p(x)$ to F are given. In each case:

- a. Verify that $p(x)$ is irreducible over F .
- b. Write out a formula for the product of two arbitrary elements $a_0 + a_1\alpha + a_2\alpha^2$ and $b_0 + b_1\alpha + b_2\alpha^2$ of $F(\alpha)$.
- c. Find the multiplicative inverse of the given element of $F(\alpha)$.
3. $F = \mathbf{Z}_3$, $p(x) = x^3 + 2x^2 + 1$, $\alpha^2 + \alpha + 2$
4. $F = \mathbf{Z}_3$, $p(x) = x^3 + x^2 + 2x + 1$, $\alpha^2 + 2\alpha + 1$
5. $F = \mathbf{Z}_5$, $p(x) = x^3 + x + 1$, $\alpha^2 + 4\alpha$
6. $F = \mathbf{Z}_5$, $p(x) = x^3 + x^2 + 1$, $\alpha^2 + 2\alpha + 3$
7. For the given irreducible polynomial $p(x)$ over \mathbf{Z}_3 , list all elements of the field $\mathbf{Z}_3(\alpha)$ that is obtained by adjoining a zero α of $p(x)$ to \mathbf{Z}_3 .
 - a. $p(x) = x^3 + 2x^2 + 1$
 - b. $p(x) = x^3 + x^2 + 2x + 1$
8. If F is a finite field with k elements, and $p(x)$ is a polynomial of positive degree n over F , find a formula for the number of elements in the ring $F[x]/(p(x))$.
9. Construct a field having the following number of elements.
 - a. 2^4
 - b. 5^2
 - c. 3^3
 - d. 7^2
10. Find the multiplicative inverse of $\sqrt[3]{4} - 2\sqrt[3]{2} - 2$ in $\mathbf{Q}(\sqrt[3]{2})$, where \mathbf{Q} is the field of rational numbers.
11. Find the multiplicative inverse of $\sqrt[3]{9} - \sqrt[3]{3} + 2$ in $\mathbf{Q}(\sqrt[3]{3})$, where \mathbf{Q} is the field of rational numbers.
12. An element u of a field F is a perfect square in F if there exists an element v in F such that $u = v^2$. The quadratic formula can be generalized in the following way: Suppose that $1 + 1 \neq 0$ in F , and let $p(x) = ax^2 + bx + c$, $a \neq 0$, be a quadratic polynomial over F .

- a.** Prove that $p(x)$ has a zero in F if and only if $b^2 - 4ac$ is a perfect square in F .
b. If $b^2 - 4ac$ is a perfect square in F , show that the zeros of $p(x)$ in F are given by

$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \text{ and } r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

and that these zeros are distinct if $b^2 - 4ac \neq 0$.

- 13.** Determine whether each of the following polynomials has a zero in the given field F . If a polynomial has zeros in the field, use the quadratic formula to find them.
- a.** $x^2 + 3x + 2, \quad F = \mathbf{Z}_5$
b. $x^2 + 3x + 3, \quad F = \mathbf{Z}_5$
c. $x^2 + 2x + 6, \quad F = \mathbf{Z}_7$
d. $x^2 + 3x + 1, \quad F = \mathbf{Z}_7$
e. $2x^2 + x + 1, \quad F = \mathbf{Z}_7$
f. $3x^2 + 2x - 1, \quad F = \mathbf{Z}_7$
- 14. a.** Find the value of c that will cause the polynomial $f(x) = x^2 + 3x + c$ to have 3 as a zero in the field \mathbf{Z}_7 .
b. Find the other zero of $f(x)$ in \mathbf{Z}_7 .

Each of the polynomials $p(x)$ in Exercises 15–18 is irreducible over the given field F . Find all zeros of $p(x)$ in the field $F(\alpha)$ obtained by adjoining a zero of $p(x)$ to F . (In Exercises 17 and 18, $p(x)$ has three zeros in $F(\alpha)$.)

- 15.** $p(x) = x^2 + 2x + 2, \quad F = \mathbf{Z}_3$
16. $p(x) = x^2 + x + 2, \quad F = \mathbf{Z}_3$
17. $p(x) = x^3 + x^2 + 1, \quad F = \mathbf{Z}_5$
18. $p(x) = x^3 + 2x^2 + 4x + 2, \quad F = \mathbf{Z}_5$

Key Words and Phrases

addition of polynomials, 357	factorization of polynomials over \mathbf{C} , 385	ground field, 416
algebraically closed field, 386	factorization of polynomials over \mathbf{R} , 388	indeterminate, 355
algebraic element, 386	factor (or divisor) of a polynomial, 367	irreducible (prime) polynomial, 378
algebraic extension, 386	Factor Theorem, 377	linear combination, 370
Cardano's Formulas, 401	Fundamental Theorem of Algebra, 384	monic polynomial, 370
degree of a polynomial, 362	Gauss's Lemma, 390	multiple of a polynomial, 367
discriminant, 405	greatest common divisor, 370	multiplication of polynomials, 357
Division Algorithm, 367		multiplicity of a factor, 380
Eisenstein's Irreducibility Criterion, 391		multiplicity of a zero, 380
equality of polynomials, 356		

polynomial in x over R , 355
 polynomial mapping, 375
 primitive polynomial, 390
 quadratic formula, 385
 quotient, remainder, 369
 rational zeros, 388

reducible polynomial, 378
 relatively prime polynomials, 380
 Remainder Theorem, 376
 resolvent equation, 403
 ring of polynomials
 over R , 359, 361

root (solution) of a polynomial
 equation, 375
 simple algebraic extension, 416
 splitting field, 419
 zero of a polynomial, 375, 380



Image Works

A Pioneer in Mathematics

Carl Friedrich Gauss (1777–1855)

Carl Friedrich Gauss was born in Brunswick, Germany, on April 30, 1777. He is regarded as the greatest mathematician of the 19th century and has been called the Prince of Mathematics. Part of Gauss's greatness is due to the fact that his interest spanned all mathematics known in his time. Since then, the volume of knowledge in mathematics has become so large that no one person could ever hope to master the whole field. In this sense, he may have been the last complete mathematician.

The world was almost deprived of Gauss's genius when, as a child, he fell into an overflowing canal near his home. It is said that he surely would have drowned had he not been rescued by a passerby.

His mathematical genius became evident early in his life. He often said that he could reckon before he could talk. In school, his precocity attracted the attention of the Duke of Brunswick. The Duke decided to finance the education of the young prodigy and granted him a fixed pension so that he could devote himself to work without financial considerations.

Gauss made some of the greatest contributions to mathematics when he was a young man. He developed the method of least squares while preparing for university studies at Collegium Carolinum. Two years later, he solved a 2000-year-old problem by proving that a regular 17-sided polygon can be constructed with only a straightedge and a compass. In his doctoral dissertation, Gauss proved the Fundamental Theorem of Algebra, a result that had been accepted without proof for many years. In 1801, at the age of 24, he published the monumental work *Disquisitiones Arithmeticae*, in which he laid the foundations of the area of mathematics called *number theory*.

Also in 1801, when Gauss turned his attention to astronomy, he accomplished an extraordinary achievement. Using a scanty amount of data, he was able to predict accurately the orbit of the asteroid Ceres. For this achievement, he garnered international acclaim. In 1807, he was appointed director of the astronomical observatory of Göttingen.

The Basics of Logic

In any mathematical system, just as in any language, there must be some undefined terms. For example, the words *set* and *element* are undefined terms. We think of a set as a collection of objects, and the individual objects as elements of the set. We need to understand the word *set* to describe the word *element*, and vice versa. Hence we must rely on our intuition to understand these undefined terms and feel comfortable using them to define new terms.

A **statement**, or **proposition**, is a declarative sentence that is either true or false, but not both. **Postulates** are statements (often expressed using undefined terms) that are assumed to be true. Postulates and definitions are used to prove statements called **theorems**. Once a theorem is proved to be true, it can be used to establish the truth of subsequent theorems. A **lemma** is itself a theorem whose major importance lies not in its own statement but in its role as a stepping stone toward the statement or proof of a theorem. Finally, a **corollary** is also a theorem but is not so named because it is usually either a direct consequence of or a special case of a preceding theorem. To avoid “stealing the thunder” of the more important theorem, it is labeled a corollary.

We now briefly discuss the basic concepts of logic that are essential to the mathematician for constructing proofs. We use the letters p , q , r , s , and so on, to represent statements. Consider the following statements:

- p : The sum of the angles in a triangle is 180° .
- q : $2^2 + 3^2 = (2 + 3)^2$
- r : $x^2 + 1 = 0$
- s : Beckie is pretty.

The statement p is a true proposition from plane geometry. The statement q is a false proposition, when we consider the usual multiplication and addition in the set of real numbers. The statement r is not a proposition, since its truth or falsity cannot be determined unless the value of x is known. The statement s is not a proposition, since its truth or falsity “is in the eyes of the beholder” and also depends on which “Beckie” is under consideration.

The statement r in the preceding paragraph can be clarified by placing restrictions on the variable x , such as “for every x ,” “for each x ,” “for all x ,” “for some x ,” “for at least one x ,” or “there exists an x .” The phrases “for every x ,” “for all x ,” and “for each x ” mean the same thing and are often abbreviated by the symbol \forall , which is called the **universal quantifier**. Similarly, the phrases “for some x ,” “for at least one x ,” and “there exists an x ” mean the same thing and are abbreviated by the symbol \exists , which is called the **existential quantifier**. Another commonly used symbol is \exists , which is read “such that.”

Thus the statement

$$\forall x, x > 0$$

is read

“For every x , $x > 0$.”

Similarly, the statement

$$\exists y \exists y^2 + 1 = 0$$

is read

“There exists a y such that $y^2 + 1 = 0$.”

A statement about the variable x may be true for some values of x and false for other values of x . Some such statements can be proved by furnishing an example, but others cannot. The quantifier used in the statement determines the type of proof required.

If the statement has an existential quantifier, then one example where the statement is true will establish the statement as a theorem. Consider the statement

“There exists an integer x such that $x^2 + 2x = 24$.”

If the value 4 is assigned to x , and it is then verified that $4^2 + 2(4) = 16 + 8 = 24$, this proves that the statement is true. The phrase “there exists an integer x ” requires only one value of x that works to make the statement true.

If the statement has a universal quantifier, a specific example does not make a proof. Consider the statement

“For any integer n , $n - 1$ is a factor of $n^2 - 4n + 3$.”

If the value 7 is assigned to n , and it is then verified that $n - 1 = 6$ is indeed a factor of $7^2 - 4(7) + 3 = 24 = 6(4)$, this illustrates a case where the statement is true, but *it does not prove that the statement is true for any value of n other than 7* and thus does not constitute a proof. The phrase “for any integer n ” requires an argument that can be applied independently of the value of n . In this case, a proof can be supplied by demonstrating that

$$(n - 1)(n - 3) = n^2 - 4n + 3,$$

since this shows that $n - 1$ is always a factor of $n^2 - 4n + 3$.

If a statement about x with a universal quantifier is not true for at least one value of x , the statement is declared to be false (and therefore is not a theorem). Consider the statement

“ $x^2 < 2^x$ for all real numbers x .”

For $x = 3$,

$$3^2 < 2^3$$

is false. Therefore, the statement

“ $x^2 < 2^x$ for all real numbers x ”

is false.

A demonstration in which a statement is shown to be false for a certain value of the variable is called a **counterexample**. A statement with a universal quantifier can be proved false by finding just one counterexample, as we did in the last paragraph.

If p is a proposition, then the **negation of p** is denoted by $\sim p$ and is read “not p .” If p is a true proposition, then $\sim p$ must be false, and vice versa. We illustrate the idea using a truth table (see Figure A.1), where T stands for true and F stands for false.

**Truth Table
for $\sim p$**

p	$\sim p$
T	F
F	T

■ **Figure A.1**

The negation of statements involving the universal quantifier and the existential quantifier are given next. We use $p(x)$ to represent a statement involving the variable x . Then the statement

$$\sim(\forall x, p(x)) \text{ is } \exists x \ni \sim p(x)$$

is read

“The negation of ‘For every x , $p(x)$ is true’

is

‘There exists an x such that $p(x)$ is false.’ ”

We also write

$$\sim(\exists x \ni p(x)) \text{ is } \forall x, \sim p(x)$$

and read

“The negation of ‘There exists an x such that $p(x)$ is true’

is

‘For every x , $p(x)$ is false.’ ”

Example 1 The negation of the statement

“All the students in the class are female”

is

“There exists at least one student in the class who is not female.” ■

Example 2 The negation of the statement

“There is at least one student who passed the course”

is

“All the students failed the course.” ■

Connectives are used to join propositions to make compound statements. Propositions p and q can be joined with the connective “and,” which is commonly symbolized by \wedge and called **conjunction**. We define $p \wedge q$ to be true only when both p is true and q is true. The corresponding truth table for $p \wedge q$ is given in Figure A.2.

**Truth Table
for $p \wedge q$**

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

■ **Figure A.2**

Similarly, propositions p and q can be joined with the connective “or,” symbolized by \vee and called **disjunction**. We define $p \vee q$ to be true when either p is true or q is true, or both p and q are true. The truth table for $p \vee q$ is given in Figure A.3.

**Truth Table
for $p \vee q$**

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

■ **Figure A.3**

Probably the most important connective is **implication**, denoted by \Rightarrow . Suppose p and q are propositions. Then

$$p \Rightarrow q$$

is read in several ways:

- “ p implies q ”
- “if p then q ”
- “ p only if q ”
- “ p is sufficient for q ”
- “ q is necessary for p .”

In each of these statements, p is called the **hypothesis** and q is called the **conclusion**.

Let us consider the following situations. Algebra class meets only three days a week, on Monday, Wednesday, and Friday. Let p and q be the following propositions:

- p : Today is Monday.
 q : Algebra class meets today.

Consider the implication

$$p \Rightarrow q.$$

This implication is true if both p and q are true:

$$\text{Today is Monday} \Rightarrow \text{Algebra class meets today.}$$

Suppose p is true and q is false. Then the implication

$$\text{Today is Monday} \Rightarrow \text{Algebra class meets today}$$

is false. Next suppose that p is false. The falsity of p does not affect the truth or falsity of q . That is,

$$\text{Today is not Monday}$$

does not give any information about whether algebra class meets today. Thus we conclude that

$$p \Rightarrow q$$

is false only when p is true and q is false. We record these results in the truth table in Figure A.4.

**Truth Table
for $p \Rightarrow q$**

p	q	$p \Rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

■ **Figure A.4**

Another prominent connective is the **biconditional**, which is denoted by

$$p \Leftrightarrow q$$

and is read in any one of three ways:

- “ p if and only if q ”
 “ p is necessary and sufficient for q ”
 “ p is equivalent to q .”

The biconditional statement

$$p \Leftrightarrow q$$

can be expressed as the conjunction of two statements:

$$(p \Rightarrow q) \wedge (q \Rightarrow p).$$

The truth table in Figure A.5 illustrates that the statement $p \Leftrightarrow q$ is true when p and q are both true or both false; otherwise, $p \Leftrightarrow q$ is false.

Truth Table for $p \Leftrightarrow q$

p	q	$p \Rightarrow q$	$q \Rightarrow p$	$(p \Rightarrow q) \wedge (q \Rightarrow p)$ $p \Leftrightarrow q$
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

■ **Figure A.5**

If the truth tables for two propositions are identical, then the two propositions are said to be **logically equivalent**, and we use the \Leftrightarrow symbol to designate this.

Example 3 To show that

$$\sim(p \wedge q) \Leftrightarrow (\sim p) \vee (\sim q),$$

we examine the two columns headed by $\sim(p \wedge q)$ and by $(\sim p) \vee (\sim q)$ in the truth table in Figure A.6 and note that they are identical.

Truth Table for $\sim(p \wedge q) \Leftrightarrow (\sim p) \vee (\sim q)$

p	q	$p \wedge q$	$\sim(p \wedge q)$	$\sim p$	$\sim q$	$(\sim p) \vee (\sim q)$
T	T	T	F	F	F	F
T	F	F	T	F	T	T
F	T	F	T	T	F	T
F	F	F	T	T	T	T

■ **Figure A.6**



The statement in Example 3 is the logical form of one of **De Morgan's Laws**. The corresponding form for sets is given at the end of Section 1.1. The next example illustrates a truth table involving three propositions.

Example 4 To show that

$$r \wedge (p \vee q) \Leftrightarrow (r \wedge p) \vee (r \wedge q),$$

we need eight rows in our truth table, since there are 2^3 different ways to assign true and false to the three different statements (see Figure A.7).

Truth Table for $r \wedge (p \vee q) \Leftrightarrow (r \wedge p) \vee (r \wedge q)$

r	p	q	$p \vee q$	$r \wedge (p \vee q)$	$r \wedge p$	$r \wedge q$	$(r \wedge p) \vee (r \wedge q)$
T	T	T	T	T	T	T	T
T	T	F	T	T	T	F	T
T	F	T	T	T	F	T	T
T	F	F	F	F	F	F	F
F	T	T	T	F	F	F	F
F	T	F	T	F	F	F	F
F	F	T	T	F	F	F	F
F	F	F	F	F	F	F	F

Figure A.7

In this text, we see some theorems whose statements involve an implication

$$p \Rightarrow q.$$

In some instances, it is more convenient to prove a statement that is logically equivalent to the implication $p \Rightarrow q$. The truth table in Figure A.8 shows that the implication

$$p \Rightarrow q \text{ (implication)}$$

is logically equivalent to the statement

$$\sim q \Rightarrow \sim p \text{ (contrapositive),}$$

which is called the **contrapositive** of $p \Rightarrow q$.

Truth Table for $(p \Rightarrow q) \Rightarrow (\sim q \Rightarrow \sim p)$

p	q	$p \Rightarrow q$	$\sim q$	$\sim p$	$\sim q \Rightarrow \sim p$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Figure A.8

Two other variations of the implication $p \Rightarrow q$ are given special names. They are

$$q \Rightarrow p \text{ is the } \mathbf{converse} \text{ of } p \Rightarrow q$$

and

$$\sim p \Rightarrow \sim q \text{ is the } \mathbf{inverse} \text{ of } p \Rightarrow q.$$

We note that the converse and the inverse are logically equivalent; that is,

$$(q \Rightarrow p) \Leftrightarrow (\sim p \Rightarrow \sim q).$$

Example 5 Let p and q be the following statements:

p : x is an even integer.

q : x is an integer.

In Figure A.9, we describe the implication $p \Rightarrow q$ and its variations.

<i>Logically equivalent</i>		<i>Logically equivalent</i>	
Implication $p \Rightarrow q$ x is an even integer. \Rightarrow x is an integer. TRUE	Contrapositive $\sim q \Rightarrow \sim p$ x is not an integer. \Rightarrow x is not an even integer. TRUE	Converse $q \Rightarrow p$ x is an integer. \Rightarrow x is an even integer. FALSE	Inverse $\sim p \Rightarrow \sim q$ x is not an even integer. \Rightarrow x is not an integer. FALSE

■ **Figure A.9**



Example 6 Suppose p and q are the following statements:

p : The Packers win this week.

q : The Packers are in the playoffs next week.

Suppose the only way the Packers go to the playoffs is if they win this week. Hence, if they do not win this week, they will not go to the playoffs next week. In Figure A.10, we examine the implication $p \Rightarrow q$ and its variations.

<i>Logically equivalent</i>		<i>Logically equivalent</i>	
Implication $p \Rightarrow q$ Packers win this week. \Rightarrow Packers are in the playoffs next week. TRUE	Contrapositive $\sim q \Rightarrow \sim p$ Packers are not in the playoffs next week. \Rightarrow Packers do not win this week. TRUE	Converse $q \Rightarrow p$ Packers are in the playoffs next week. \Rightarrow Packers win this week. TRUE	Inverse $\sim p \Rightarrow \sim q$ Packers do not win this week. \Rightarrow Packers are not in the playoffs next week. TRUE

■ **Figure A.10**



Since the implication and its converse are true, we write

$$p \Leftrightarrow q.$$

The method of **proof by contradiction** is sometimes useful in proving statements of the form “ p implies q .” As shown in Figure A.4, the statement “ p implies q ” is true in all cases except when p is true and q is false. In a proof by contradiction, we assume that p is true and that q is false and then reach a contradiction (an impossible situation).

To provide a simple example, consider the following propositions:[†]

p : x is an integer and x^2 is even.

q : x is an even integer.

We shall use a proof by contradiction to prove that $p \Rightarrow q$.

Assume that p is true and q is false. Since x is not an even integer, x must be an odd integer. That is, $x = 2n + 1$ for some integer n . This implies that

$$\begin{aligned}x^2 &= (2n + 1)(2n + 1) \\&= 4n^2 + 4n + 1 \\&= 2(2n^2 + 2n) + 1,\end{aligned}$$

and therefore x^2 is an odd integer. This directly contradicts proposition p . Therefore, q must be true when p is true, and this means that p implies q .

Appendix Exercises

Prove that each of the statements in Exercises 1–6 is false.

1. For every real number x , $x^2 > 0$.
2. For any real number x , $x^2 \geq x$.
3. For each real number a , there is a real number b such that $ab = 1$.
4. $2^x < 3^x$ for all real numbers x .
5. $-x < |x|$ for all real numbers x .
6. If x is a real number such that $x < 1$, then $x^2 < x$.

Prove that each of the statements in Exercises 7–12 is true.

7. There is an integer n such that $n^2 + 2n = 48$.
8. There is a real number x such that $x + \frac{1}{x} = \frac{13}{6}$.
9. $n^2 < 2^n$ for some integer n .
10. $1 + 3n < 2^n$ for some integer n .
11. There exists an integer n such that $n^2 + n$ is an even integer.
12. There exists an integer n such that $n^2 + 2n$ is a multiple of 5.

Write the negation of each of the statements in Exercises 13–36.

13. All the children received a Valentine card.
14. Every house has a fireplace.
15. Every senior graduated and received a job offer.
16. All the cheerleaders are tall and athletic.

[†]An integer m is defined to be an *even integer* if $m = 2k$ for some integer k , and m is defined to be an *odd integer* if $m = 2q + 1$ for some integer q . More details may be found in Section 1.2.

17. There is a rotten apple in the basket.
18. There is a snake that is nonpoisonous.
19. There is a politician who is honest and trustworthy.
20. There is a cold medication that is safe and effective.
21. For every $x \in A$, $x \in B$. (The notation $x \in A$ is defined in Section 1.1.)
22. For every real number r , the square of r is nonnegative.
23. For every right triangle with sides a and b and hypotenuse c , we have $c^2 = a^2 + b^2$.
24. For any two rational numbers r and s , there is an irrational number j between them.
25. Every complex number has a multiplicative inverse.
26. For all 2×2 matrices A and B over the real numbers, we have $AB = BA$. (The product of two matrices is given in Definition 1.31 of Section 1.6.)
27. For all sets A and B , their Cartesian products satisfy the equation $A \times B = B \times A$. (The Cartesian product is defined in Definition 1.8 of Section 1.2.)
28. For any real number c , $x < y \Rightarrow cx < cy$.
29. There exists a complex number x such that $x^2 + 1 = 0$.
30. There exists a 2×2 matrix A over the real numbers such that $A^2 = I$ where $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $A^2 = A \cdot A$. (The product of two matrices is given in Definition 1.31 of Section 1.6.)
31. There exists a set A such that $A \subseteq A \cap B$. (The notation $A \subseteq A \cap B$ is defined in Section 1.1.)
32. There exists a complex number z such that $\bar{z} = z$. (The notation \bar{z} is given in Definition 7.7 of Section 7.2.)
33. There exists a triangle with angles α , β , and γ such that $\alpha + \beta + \gamma > 180^\circ$.
34. There exists an angle θ such that $\sin \theta = 2.1$.
35. There exists a real number x such that $2^x \leq 0$.
36. There exists an even integer x such that x^2 is odd.

Construct truth tables for each of the statements in Exercises 37–52.

- | | |
|---|--|
| 37. $p \Leftrightarrow \sim(\sim p)$ | 38. $p \vee (\sim p)$ |
| 39. $\sim(p \wedge (\sim p))$ | 40. $p \Rightarrow (p \vee q)$ |
| 41. $(p \wedge q) \Rightarrow p$ | 42. $\sim(p \vee q) \Leftrightarrow (\sim p) \wedge (\sim q)$ |
| 43. $(p \wedge (p \Rightarrow q)) \Rightarrow q$ | 44. $(p \Rightarrow q) \Leftrightarrow \sim(p \wedge \sim q)$ |
| 45. $(p \Rightarrow q) \Leftrightarrow ((\sim p) \vee q)$ | 46. $(\sim(p \Rightarrow q)) \Leftrightarrow (p \wedge (\sim q))$ |
| 47. $(p \Rightarrow q) \Leftrightarrow (p \wedge (\sim q) \Rightarrow (\sim p))$ | 48. $r \vee (p \wedge q) \Leftrightarrow (r \vee p) \wedge (r \vee q)$ |
| 49. $(p \wedge q \wedge r) \Rightarrow ((p \vee q) \wedge r)$ | 50. $((p \Rightarrow q) \wedge (q \Rightarrow r)) \Rightarrow (p \Rightarrow r)$ |
| 51. $(p \Rightarrow (q \wedge r)) \Leftrightarrow ((p \Rightarrow q) \wedge (p \Rightarrow r))$ | |
| 52. $((p \wedge q) \Rightarrow r) \Leftrightarrow (p \Rightarrow (q \Rightarrow r))$ | |

In Exercises 53–68, examine the implication $p \Rightarrow q$ and its variations (contrapositive, inverse, and converse) by writing each in English. Determine the truth or falsity of each.

- 53.** p : My grade for this course is A.
 q : I can enroll in the next course.
- 54.** p : My car ran out of gas.
 q : My car won't start.
- 55.** p : The Saints win the Super Bowl.
 q : The Saints are the champion football team.
- 56.** p : I have completed all the requirements for a bachelor's degree.
 q : I can graduate with a bachelor's degree.
- 57.** p : My pet has four legs.
 q : My pet is a dog.
- 58.** p : I am within 30 miles of home.
 q : I am within 20 miles of home.
- 59.** p : Quadrilateral $ABCD$ is a square.
 q : Quadrilateral $ABCD$ is a rectangle.
- 60.** p : Triangle ABC is isosceles.
 q : Triangle ABC is equilateral.
- 61.** p : x is a positive real number.
 q : x is a nonnegative real number.
- 62.** p : x is a positive real number.
 q : x^2 is a positive real number.
- 63.** p : $5x$ is odd.
 q : x is odd.
- 64.** p : $5 + x$ is odd.
 q : x is even.
- 65.** p : xy is even.
 q : x is even or y is even.
- 66.** p : x is even and y is even.
 q : $x + y$ is even.
- 67.** p : $x^2 > y^2$
 q : $x > y$
- 68.** p : $\frac{x}{y} > 0$
 q : $xy > 0$

State the contrapositive, converse, and inverse of each of the implications in Exercises 69–74.

- | | |
|--|--|
| 69. $p \Rightarrow (q \vee r)$
71. $p \Rightarrow \sim q$
73. $(p \vee q) \Rightarrow (r \wedge s)$ | 70. $p \Rightarrow (q \wedge r)$
72. $(p \wedge \sim q) \Rightarrow \sim p$
74. $(p \wedge q) \Rightarrow (r \wedge s)$ |
|--|--|

This page intentionally left blank

Answers to True/False and Selected Computational Exercises

Exercises 1.1 Pages 9–12

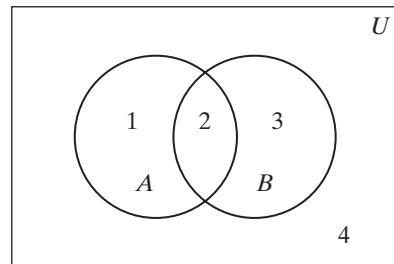
True or False

- | | | | | |
|----------|---------|----------|----------|-----------|
| 1. true | 2. true | 3. false | 4. true | 5. true |
| 6. false | 7. true | 8. true | 9. false | 10. false |
-

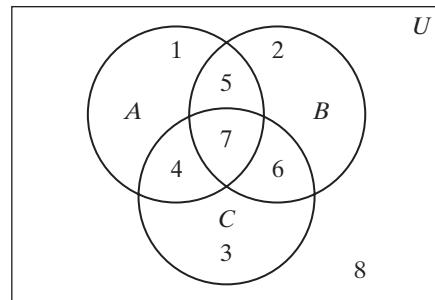
Exercises

1. a. $A = \{x \mid x \text{ is a nonnegative even integer less than } 12\}$
c. $A = \{x \mid x \text{ is a negative integer}\}$
2. a. false c. false e. false
3. a. false c. true e. true g. false i. false
4. a. true c. false e. false g. false
5. a. $\{0, 1, 2, 3, 4, 5, 6, 8, 10\}$ c. $\{0, 2, 4, 6, 7, 8, 9, 10\}$ e. \emptyset
g. $\{0, 2, 3, 4, 5\}$ i. $\{1, 3, 5\}$ k. $\{1, 2, 3, 5\}$ m. $\{3, 5\}$
6. a. A c. \emptyset e. A g. A i. U k. U m. A
7. a. $\{\emptyset, A\}$ c. $\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, A\}$
e. $\{\emptyset, \{1\}, \{\{1\}\}, A\}$ g. $\{\emptyset, A\}$
8. a. One possible partition is $X_1 = \{x \mid x \text{ is a negative integer}\}$ and $X_2 = \{x \mid x \text{ is a nonnegative integer}\}$. Another partition is $X_1 = \{x \mid x \text{ is a negative integer}\}$, $X_2 = \{0\}$, $X_3 = \{x \mid x \text{ is a positive integer}\}$.
c. One partition is $X_1 = \{1, 5, 9\}$ and $X_2 = \{11, 15\}$. Another partition is $X_1 = \{1, 15\}$, $X_2 = \{11\}$, and $X_3 = \{5, 9\}$.
9. a. $X_1 = \{1\}$, $X_2 = \{2\}$, $X_3 = \{3\}$; $X_1 = \{1\}$, $X_2 = \{2, 3\}$; $X_1 = \{2\}$, $X_2 = \{1, 3\}$; $X_1 = \{3\}$, $X_2 = \{1, 2\}$
11. a. $A \subseteq B$ c. $B \subseteq A$ e. $A = B = U$ g. $A = U$
35. Let $A = \{a\}$, $B = \{a, b\}$, and $C = \{a, c\}$. Then $A \cap B = \{a\} = A \cap C$ but $B \neq C$.
39. $(A \cap B') \cup (A' \cap B) = (A \cup B) \cap (A' \cup B')$

40. a.

 $A \cup B$: Regions 1, 2, 3 $A \cap B$: Region 2 $(A \cup B) - (A \cap B)$: Regions 1, 3 $A - B$: Region 1 $B - A$: Region 3 $A + B$: Regions 1, 3Each of $A + B$ and $(A - B) \cup (B - A)$ consists of Regions 1, 3.

c.

 A : Regions 1, 4, 5, 7 $B + C$: Regions 2, 3, 4, 5 $A \cap (B + C)$: Regions 4, 5 $A \cap B$: Regions 5, 7 $A \cap C$: Regions 4, 7 $(A \cap B) + (A \cap C)$: Regions 4, 5Each of $A \cap (B + C)$ and $(A \cap B) + (A \cap C)$ consists of Regions 4, 5.41. a. $A + A = (A \cup A) - (A \cap A) = A - A = A \cap A' = \emptyset$

Exercises 1.2 Pages 21–25

True or False

- | | | | | |
|----------|---------|----------|----------|----------|
| 1. false | 2. true | 3. false | 4. false | 5. false |
| 6. true | 7. true | 8. false | 9. true | |

Exercises

1. a. $\{(a, 0), (a, 1), (b, 0), (b, 1)\}$ c. $\{(2, 2), (4, 2), (6, 2), (8, 2)\}$
e. $\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$

2. a. domain = \mathbf{E} , codomain = \mathbf{Z} , range = \mathbf{Z}
 c. domain = \mathbf{E} , codomain = \mathbf{Z} ,
 range = $\{y \mid y \text{ is a nonnegative even integer}\} = (\mathbf{Z}^+ \cap \mathbf{E}) \cup \{0\}$
3. a. $f(S) = \{1, 3, 5, \dots\} = \mathbf{Z}^+ - \mathbf{E}$, $f^{-1}(T) = \{-4, -3, -1, 1, 3, 4\}$
 c. $f(S) = \{0, 1, 4\}$, $f^{-1}(T) = \emptyset$
4. a. The mapping f is not onto, since there is no $x \in \mathbf{Z}$ such that $f(x) = 1$. The mapping f is one-to-one.
 c. The mapping f is onto and one-to-one.
 e. The mapping f is not onto, since there is no $x \in \mathbf{Z}$ such that $f(x) = -1$. It is not one-to-one, since $f(1) = f(-1)$ and $1 \neq -1$.
 g. The mapping f is not onto, since there is no $x \in \mathbf{Z}$ such that $f(x) = 3$. It is one-to-one.
 i. The mapping f is onto. It is not one-to-one, since $f(9) = f(4)$ and $9 \neq 4$.
5. a. The mapping f is both onto and one-to-one.
 c. The mapping f is both onto and one-to-one.
 e. The mapping f is not onto, since there is no $x \in \mathbf{R}$ such that $f(x) = -1$. It is not one-to-one, since $f(1) = f(-1)$ and $1 \neq -1$.
6. a. The mapping f is onto and one-to-one.
7. a. The mapping f is onto. The mapping is not one-to-one, since $f(-1) = f(1)$ and $-1 \neq 1$.
 c. The mapping f is onto and one-to-one.
8. a. The mapping f is not onto, since there is no $x \in \mathbf{Z}$ such that $|x + 4| = -1$.
 The mapping f is not one-to-one, since $f(1) = f(-9) = 5$ but $1 \neq -9$.
9. a. The mapping f is not onto, since there is no $x \in \mathbf{Z}^+$ such that $2^x = 3$. The mapping f is one-to-one.
10. a. Let $f: \mathbf{E} \rightarrow \mathbf{E}$ where $f(x) = x$.
 c. Let $f: \mathbf{E} \rightarrow \mathbf{E}$ where
- $$f(x) = \begin{cases} x/2 & \text{if } x \text{ is a multiple of 4} \\ x & \text{if } x \text{ is not a multiple of 4.} \end{cases}$$
11. a. For arbitrary $a \in \mathbf{Z}$, $2a$ is even and $f(2a) = \frac{2a}{2} = a$. Thus f is onto. But f is not one-to-one, since $f(1) = f(-1) = 0$.
 c. For arbitrary $a \in \mathbf{Z}$, $2a - 1$ is odd, and therefore
- $$f(2a - 1) = \frac{(2a - 1) + 1}{2} = a.$$
- Thus, f is onto. But f is not one-to-one, since $f(2) = 5$ and also $f(9) = 5$.
- e. The mapping f is not onto, since there is no $x \in \mathbf{Z}$ such that $f(x) = 4$. Since $f(2) = 6$ and $f(3) = 6$, then f is not one-to-one.

- 12. a.** The mapping f is not onto, since there is no $x \in \mathbf{R} - \{0\}$ such that $f(x) = 1$.
 If $a_1, a_2 \in \mathbf{R} - \{0\}$,

$$\begin{aligned} f(a_1) = f(a_2) &\Rightarrow \frac{a_1 - 1}{a_1} = \frac{a_2 - 1}{a_2} \\ &\Rightarrow a_2(a_1 - 1) = a_1(a_2 - 1) \\ &\Rightarrow a_2a_1 - a_2 = a_1a_2 - a_1 \\ &\Rightarrow -a_2 = -a_1 \\ &\Rightarrow a_2 = a_1. \end{aligned}$$

Thus f is one-to-one.

- c.** The mapping f is not onto, since there is no $x \in \mathbf{R} - \{0\}$ such that $f(x) = 0$. It is not one-to-one, since $f(2) = \frac{2}{5}$ and $f(\frac{1}{2}) = \frac{2}{5}$.
- 13. a.** The mapping f is onto, since for every $(y, x) \in B = \mathbf{Z} \times \mathbf{Z}$ there exists an $(x, y) \in A = \mathbf{Z} \times \mathbf{Z}$ such that $f(x, y) = (y, x)$.

To show that f is one-to-one, we assume $(a, b) \in A = \mathbf{Z} \times \mathbf{Z}$ and $(c, d) \in A$ and

$$f(a, b) = f(c, d)$$

or

$$(b, a) = (d, c).$$

This means $b = d$ and $a = c$ and

$$(a, b) = (c, d).$$

- c.** Since for every $x \in B = \mathbf{Z}$ there exists an $(x, y) \in A = \mathbf{Z} \times \mathbf{Z}$ such that $f(x, y) = x$, the mapping f is onto. However, f is not one-to-one, since $f(1, 0) = f(1, 1)$ and $(1, 0) \neq (1, 1)$.
- e.** The mapping f is not onto, since there is no (x, y) in $\mathbf{Z}^+ \times \mathbf{Z}^+$ such that $f(x, y) = \frac{x}{y} = 0$. The mapping f is not one-to-one, since $f(2, 1) = f(4, 2) = 2$.
- 15. a.** The mapping f is not onto, since there is no $a \in A$ such that $f(a) = 9 \in B$. It is not one-to-one, since $f(-2) = f(2)$ and $-2 \neq 2$.
- c.** With $T = \{4, 9\}$, $f^{-1}(T) = \{-2, 2\}$, and $f(f^{-1}(T)) = f(\{-2, 2\}) = \{4\} \neq T$.

- 16. a.** $g(S) = \{2, 4\}$, $g^{-1}(g(S)) = \{2, 3, 4, 7\}$

- 17. a.** $f(S) = \{-1, 2, 3\}$, $f^{-1}(f(S)) = S$

- 18. a.** $(f \circ g)(x) = \begin{cases} 2x & \text{if } x \text{ is even} \\ 2(2x - 1) & \text{if } x \text{ is odd} \end{cases}$

- c.** $(f \circ g)(x) = \begin{cases} \frac{x + |x|}{2} & \text{if } x \text{ is even} \\ |x| - x & \text{if } x \text{ is odd} \end{cases}$

- e.** $(f \circ g)(x) = (x - |x|)^2$

19. a. $(g \circ f)(x) = 2x$ c. $(g \circ f)(x) = \frac{x + |x|}{2}$ e. $(g \circ f)(x) = 0$
 21. $n!$

Exercises 1.3 Pages 28–30

True or False

1. false 2. true 3. false 4. false 5. false 6. false

Exercises

1. a. The mapping $f \circ g$ is not onto, since there is no $x \in \mathbf{Z}$ such that $(f \circ g)(x) = 1$. The mapping $f \circ g$ is one-to-one.
 c. The mapping $f \circ g$ is not onto, since there is no $x \in \mathbf{Z}$ such that $(f \circ g)(x) = 1$. It is not one-to-one, since $(f \circ g)(-2) = (f \circ g)(0)$ and $-2 \neq 0$.
 e. The mapping $f \circ g$ is not onto, since there is no $x \in \mathbf{Z}$ such that $(f \circ g)(x) = -1$. It is not one-to-one, since $(f \circ g)(1) = (f \circ g)(2)$ and $1 \neq 2$.
2. a. The mapping $g \circ f$ is not onto, since there is no $x \in \mathbf{Z}$ such that $(g \circ f)(x) = 1$. The mapping $g \circ f$ is one-to-one.
 c. The mapping $g \circ f$ is not onto, since there is no $x \in \mathbf{Z}$ such that $(g \circ f)(x) = -1$. It is not one-to-one, since $(g \circ f)(-1) = (g \circ f)(-2)$ and $-1 \neq -2$.
 e. The mapping $g \circ f$ is not onto, since there is no $x \in \mathbf{Z}$ such that $(g \circ f)(x) = 1$. It is not one-to-one, since $(g \circ f)(0) = (g \circ f)(1)$ and $0 \neq 1$.
3. Let $A = \{0, 1\}$, $B = \{-2, 1, 2\}$, $C = \{1, 4\}$. Let $g: A \rightarrow B$ be defined by $g(x) = x + 1$ and $f: B \rightarrow C$ be defined by $f(x) = x^2$. Then g is not onto, since $-2 \notin g(A)$. The mapping f is onto. Also, $f \circ g$ is onto, since $(f \circ g)(0) = f(1) = 1$ and $(f \circ g)(1) = f(2) = 4$.
5. a. Let $f: \mathbf{Z} \rightarrow \mathbf{Z}$ and $g: \mathbf{Z} \rightarrow \mathbf{Z}$ be defined by

$$f(x) = x \quad g(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ x & \text{if } x \text{ is odd.} \end{cases}$$

The mapping f is one-to-one and the mapping g is onto, but the composition $f \circ g = g$ is not one-to-one, since $(f \circ g)(1) = (f \circ g)(2)$ and $1 \neq 2$.

6. a. Let $f: \mathbf{Z} \rightarrow \mathbf{Z}$ and $g: \mathbf{Z} \rightarrow \mathbf{Z}$ be defined by

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ x & \text{if } x \text{ is odd} \end{cases} \quad g(x) = x.$$

The mapping f is onto and the mapping g is one-to-one, but the composition $f \circ g = f$ is not one-to-one, since $(f \circ g)(1) = (f \circ g)(2)$ and $1 \neq 2$.

8. a. Let $f(x) = x$, $g(x) = x^2$, and $h(x) = |x|$, for all $x \in \mathbf{Z}$.

Exercises 1.4 Pages 34–37**True or False**

1. false 2. true 3. true 4. false 5. true 6. true
 7. true 8. true 9. true
-

Exercises

1. **a.** The set B is not closed, since $-1 \in B$ and $-1 * -1 = 1 \notin B$.
c. The set B is closed.
e. The set B is not closed, since $1 \in B$ and $1 * 1 = 0 \notin B$.
g. The set B is closed.
2. **a.** not commutative; not associative; no identity element
c. not commutative; not associative; no identity element
e. commutative; associative; no identity element
g. Commutative; associative; 0 is an identity element; 0 is the only invertible element, and its inverse is 0.
i. not commutative; not associative; no identity element
k. not commutative; not associative; no identity element
m. not commutative; not associative; no identity element
3. **a.** The binary operation $*$ is not commutative, since $B * C \neq C * B$.
b. There is no identity element.
5. **a.** The binary operation $*$ is not commutative, since $D * A \neq A * D$.
b. C is an identity element.
c. The elements A and B are inverses of each other, and C is its own inverse.
7. The set of nonzero integers is not closed with respect to division, since 1 and 2 are nonzero integers but $1 \div 2$ is not a nonzero integer.

Exercises 1.5 Pages 41–42**True or False**

1. true 2. false 3. false
-

Exercises

1. **a.** A right inverse does not exist, since f is not onto.
c. A right inverse $g: \mathbf{Z} \rightarrow \mathbf{Z}$ is $g(x) = x - 2$.
e. A right inverse does not exist, since f is not onto.

g. A right inverse does not exist, since f is not onto.

i. A right inverse does not exist, since f is not onto.

k. A right inverse $g: \mathbf{Z} \rightarrow \mathbf{Z}$ is $g(x) = \begin{cases} x & \text{if } x \text{ is even} \\ 2x + 1 & \text{if } x \text{ is odd.} \end{cases}$

m. A right inverse $g: \mathbf{Z} \rightarrow \mathbf{Z}$ is $g(x) = \begin{cases} 2x & \text{if } x \text{ is even} \\ x - 2 & \text{if } x \text{ is odd.} \end{cases}$

2. a. A left inverse $g: \mathbf{Z} \rightarrow \mathbf{Z}$ is $g(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ 1 & \text{if } x \text{ is odd.} \end{cases}$

c. A left inverse $g: \mathbf{Z} \rightarrow \mathbf{Z}$ is $g(x) = x - 2$.

e. A left inverse $g: \mathbf{Z} \rightarrow \mathbf{Z}$ is $g(x) = \begin{cases} y & \text{if } x = y^3 \text{ for some } y \in \mathbf{Z} \\ 0 & \text{if } x \neq y^3 \text{ for some } y \in \mathbf{Z}. \end{cases}$

g. A left inverse $g: \mathbf{Z} \rightarrow \mathbf{Z}$ is $g(x) = \begin{cases} x & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$

i. There is no left inverse, since f is not one-to-one.

k. There is no left inverse, since f is not one-to-one.

m. There is no left inverse, since f is not one-to-one.

3. $n!$

5. Let $f: A \rightarrow A$, where A is nonempty.

f has a right inverse $\Leftrightarrow f$ is onto, by Lemma 1.25

$\Leftrightarrow f(f^{-1}(T)) = T$ for every subset T of A , by
Exercise 28 of Section 1.2

Exercises 1.6 Pages 51–54

True or False

1. true **2.** false **3.** false **4.** false **5.** false **6.** false

7. true **8.** false **9.** false **10.** false **11.** true **12.** true

Exercises

1. a. $A = \begin{bmatrix} 1 & 0 \\ 3 & 2 \\ 5 & 4 \end{bmatrix}$

c. $B = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$

e. $C = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 4 & 0 \\ 4 & 5 & 6 \\ 5 & 6 & 7 \end{bmatrix}$

2. a. $\begin{bmatrix} 3 & 0 & -4 \\ 8 & -8 & 6 \end{bmatrix}$

c. not possible

3. a. $\begin{bmatrix} -5 & 7 \\ 8 & -1 \end{bmatrix}$

c. not possible

e. $\begin{bmatrix} 4 & 2 \\ 3 & 7 \end{bmatrix}$

g. not possible

i. [4]

4. $c_{ij} = \sum_{k=1}^3 (i+k)(2k-j)$

$$= (i+1)(2-j) + (i+2)(4-j) + (i+3)(6-j)$$

$$= 12i - 6j - 3ij + 28$$

7. a. n

c. 12

8.

.	I	A	B	C
I	I	A	B	C
A	A	B	C	I
B	B	C	I	A
C	C	I	A	B

9. (answer not unique) $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

11. (answer not unique) $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} -6 & -6 \\ 3 & 3 \end{bmatrix}$

13. $(A - B)(A + B) = \begin{bmatrix} 10 & 1 \\ 2 & 1 \end{bmatrix}$ and $A^2 - B^2 = \begin{bmatrix} 2 & 6 \\ -4 & 9 \end{bmatrix}$, $(A - B)(A + B) \neq A^2 - B^2$

15. $X = A^{-1}B$

22. b. For each x in G of the form $\begin{bmatrix} a & a \\ 0 & 0 \end{bmatrix}$, then $y = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. For each x in G of the form $\begin{bmatrix} 0 & 0 \\ a & a \end{bmatrix}$, then $y = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$.

25. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix}$. Then the product $AB = \begin{bmatrix} 2 & 7 \\ 2 & 7 \end{bmatrix}$ is not diagonal even though B is diagonal.

27. c. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. Then the product $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is upper triangular, but neither A nor B is upper triangular.

Exercises 1.7 Pages 58–61

True or False

1. true 2. false 3. true 4. false 5. true 6. false
-

Exercises

1. **a.** This is a mapping, since for every $a \in A$ there is a unique $b \in A$ such that (a, b) is an element of the relation.
- c.** This is not a mapping, since the element 1 is related to three different values; $1R1$, $1R3$, and $1R5$.
- e.** This is a mapping, since for every $a \in A$ there is a unique $b \in A$ such that (a, b) is an element of the relation.
2. **a.** The relation R is not reflexive, since $x \neq 2x$ for $x \neq 0$, $x \in \mathbf{Z}$. It is not symmetric, since $x = 2y \not\Rightarrow y = 2x$ for nonzero x and $y \in \mathbf{Z}$. It is not transitive, since $x = 2y$ and $y = 2z$ do not imply that $x = 2z$, for nonzero x, y , and z in \mathbf{Z} .
- c.** The relation R is reflexive and transitive, but it is not symmetric, since for arbitrary x, y , and z in \mathbf{Z} , we have
 - (1) $x = x + 1$ with $1 \in \mathbf{Z}$,
 - (2) $6 = 3(2)$ with $2 \in \mathbf{Z}$ but $3 \neq 6k$ where $k \in \mathbf{Z}$,
 - (3) $y = xk_1$ for some $k_1 \in \mathbf{Z}$ and $z = yk_2$ for some $k_2 \in \mathbf{Z}$ imply $z = yk_2 = x(k_1 k_2)$ with $k_1 k_2 \in \mathbf{Z}$.
- e.** The relation R is reflexive since $x \geq x$ for all $x \in \mathbf{Z}$. It is not symmetric since $5R3$ but $3 \not R 5$. It is transitive, since $x \geq y$ and $y \geq z$ imply $x \geq z$ for all x, y, z in \mathbf{Z} .
- g.** The relation R is not reflexive, since $| -6 | \not\leq | -6 + 1 |$. It is not symmetric, since $| 3 | \leq | 5 + 1 |$, but $| 5 | \not\leq | 3 + 1 |$. It is not transitive, since $| 4 | \leq | 3 + 1 |$ and $| 3 | \leq | 2 + 1 |$, but $| 4 | \not\leq | 2 + 1 |$.
- i.** The relation R is not reflexive, since $2 \not R 2$. It is symmetric, since $xy \leq 0$ implies $yx \leq 0$ for all $x, y \in \mathbf{Z}$. It is not transitive, since $-1R2$ and $2R(-3)$, but $(-1) \not R (-3)$.
- k.** The relation R is reflexive, symmetric, and transitive, since for arbitrary x, y , and z in \mathbf{Z} , we have
 - (1) $| x - x | = | 0 | < 1$,
 - (2) $| x - y | < 1 \Rightarrow | y - x | < 1$,
 - (3) $| x - y | < 1$ and $| y - z | < 1 \Rightarrow x = y$ and $y = z \Rightarrow | x - z | < 1$.
- 3. a.** $\{-3, 3\}$
- 5. b.** $[0] = \{\dots, -14, -7, 0, 7, 14, \dots\}$, $[1] = \{\dots, -13, -6, 1, 8, 15, \dots\}$,
 $[3] = \{\dots, -11, -4, 3, 10, 17, \dots\}$, $[9] = [2] = \{\dots, -12, -5, 2, 9, 16, \dots\}$,
 $[-2] = [5] = \{\dots, -9, -2, 5, 12, 19, \dots\}$
- 7.** $[0] = \{0, \pm 5, \pm 10, \dots\}$, $\{\pm 1, \pm 4, \pm 6, \pm 9, \dots\} \subseteq [1]$,
 $\{\pm 2, \pm 3, \pm 7, \pm 8, \dots\} \subseteq [2]$

9. $[0] = \{\dots, -7, 0, 7, 14, \dots\}$, $[1] = \{\dots, -13, -6, 1, 8, \dots\}$,
 $[2] = \{\dots, -12, -5, 2, 9, \dots\}$, $[3] = \{\dots, -11, -4, 3, 10, \dots\}$
 $[4] = \{\dots, -10, -3, 4, 11, \dots\}$, $[5] = \{\dots, -9, -2, 5, 12, \dots\}$
 $[6] = \{\dots, -8, -1, 6, 13, \dots\}$

11. a. The relation R is reflexive and transitive but not symmetric, since for arbitrary nonempty subsets x , y , and z of A , we have the following:
- (1) x is a subset of x .
 - (2) x is a subset of y does not imply that y is a subset of x .
 - (3) x is a subset of y and y is a subset of z imply that x is a subset of z .
- c. The relation R is reflexive, symmetric, and transitive, since for arbitrary nonempty subsets x , y , and z of A , we have the following:
- (1) x and x have the same number of elements.
 - (2) If x and y have the same number of elements, then y and x have the same number of elements.
 - (3) If x and y have the same number of elements and y and z have the same number of elements, then x and z have the same number of elements.
12. a. The relation is reflexive and symmetric but not transitive, since if x , y , and z are human beings, we have the following:
- (1) x lives within 400 miles of x .
 - (2) x lives within 400 miles of y implies that y lives within 400 miles of x .
 - (3) x lives within 400 miles of y and y lives within 400 miles of z do not imply that x lives within 400 miles of z .
- c. The relation is symmetric but not reflexive and not transitive. Let x , y , and z be human beings, and we have the following:
- (1) x is a first cousin of x is not a true statement.
 - (2) x is a first cousin of y implies that y is a first cousin of x .
 - (3) x is a first cousin of y and y is a first cousin of z do not imply that x is a first cousin of z .
- e. The relation is reflexive, symmetric, and transitive, since if x , y , and z are human beings, we have the following:
- (1) x and x have the same mother.
 - (2) x and y have the same mother implies that y and x have the same mother.
 - (3) x and y have the same mother and y and z have the same mother imply that x and z have the same mother.
13. a. The relation R is an equivalence relation on $A \times A$. Let a , b , c , d , p , and q be arbitrary elements of A .
- (1) $(a, b)R(a, b)$ since $ab = ba$
 - (2) $(a, b)R(c, d) \Rightarrow ad = bc \Rightarrow cb = da \Rightarrow (c, d)R(a, b)$
 - (3) $(a, b)R(c, d)$ and $(c, d)R(p, q) \Rightarrow ad = bc$ and $cq = dp$
 $\Rightarrow adcq = bcdp$
 $\Rightarrow aq = bp$ since $c \neq 0$ and $d \neq 0$
 $\Rightarrow (a, b)R(p, q)$

- c. The relation R is an equivalence relation on $A \times A$. Let a, b, c, d, p , and q be arbitrary elements of A .

$$\begin{aligned} (1) \quad & (a, b)R(a, b) \text{ since } a^2 + b^2 = a^2 + b^2 \\ (2) \quad & (a, b)R(c, d) \Rightarrow a^2 + b^2 = c^2 + d^2 \Rightarrow c^2 + d^2 = a^2 + b^2 \Rightarrow (c, d)R(a, b) \\ (3) \quad & (a, b)R(c, d) \text{ and } (c, d)R(p, q) \Rightarrow a^2 + b^2 = c^2 + d^2 \text{ and } c^2 + d^2 = p^2 + q^2 \\ & \Rightarrow a^2 + b^2 = p^2 + q^2 \\ & \Rightarrow (a, b)R(p, q) \end{aligned}$$

14. The relation R is reflexive and symmetric but not transitive.

15. a. The relation is symmetric but not reflexive and not transitive. Let x, y , and z be arbitrary elements of the power set $\mathcal{P}(A)$ of the nonempty set A .

$$\begin{aligned} (1) \quad & x \cap x \neq \emptyset \text{ is not true if } x = \emptyset. \\ (2) \quad & x \cap y \neq \emptyset \text{ implies that } y \cap x \neq \emptyset. \\ (3) \quad & x \cap y \neq \emptyset \text{ and } y \cap z \neq \emptyset \text{ do not imply that } x \cap z \neq \emptyset. \text{ For example, let} \\ & A = \{a, b, c, d\}, \quad x = \{b, c\}, \quad y = \{c, d\}, \quad \text{and} \quad z = \{d, a\}. \quad \text{Then } x \cap y = \\ & \{c\} \neq \emptyset, \quad y \cap z = \{d\} \neq \emptyset \text{ but } x \cap z = \emptyset. \end{aligned}$$

16. The relation is reflexive, symmetric, and transitive. Let x, y , and z be arbitrary elements of the power set $\mathcal{P}(A)$ and C a fixed subset of A .

$$\begin{aligned} (1) \quad & xRx \text{ since } x \cap C = x \cap C \\ (2) \quad & xRy \Rightarrow x \cap C = y \cap C \Rightarrow y \cap C = x \cap C \Rightarrow yRx \\ (3) \quad & xRy \text{ and } yRz \Rightarrow x \cap C = y \cap C \text{ and } y \cap C = z \cap C \\ & \Rightarrow x \cap C = z \cap C \\ & \Rightarrow xRz \end{aligned}$$

Thus R is an equivalence relation on $\mathcal{P}(A)$.

17. a. The relation is reflexive, symmetric, and transitive. Let a, b , and c represent arbitrary triangles in the plane. Then

$$\begin{aligned} (1) \quad & a \text{ is similar to } a \text{ is true.} \\ (2) \quad & a \text{ is similar to } b \text{ implies that } b \text{ is similar to } a. \\ (3) \quad & a \text{ is similar to } b \text{ and } b \text{ is similar to } c \text{ imply that } a \text{ is similar to } c. \end{aligned}$$

19. d, j

21. a, d, e, f, k

$$23. \bigcup_{\lambda \in \mathcal{L}} A_\lambda = A_1 \cup A_2 \cup A_3 = \{a, b, c, d, e, f, g\}, \quad \bigcap_{\lambda \in \mathcal{L}} A_\lambda = A_1 \cap A_2 \cap A_3 = \{c\}$$

Exercises 2.1 Pages 69–71

True or False

- | | | | | |
|----------|----------|----------|----------|----------|
| 1. true | 2. false | 3. false | 4. false | 5. true |
| 6. false | 7. false | 8. false | 9. true | 10. true |

Exercises

- 35.** All the addition postulates and all the multiplication postulates except 2c are satisfied. Postulate 2c is not satisfied, since $\{0\}$ does not contain an element different from 0. The set $\{0\}$ has the properties required in postulate 4, and postulate 5 is satisfied vacuously (that is, there is no counterexample). Thus all postulates except 2c are satisfied.

Exercises 2.3 Pages 84–86

True or False

- | | | | | |
|----------|----------|----------|----------|---------|
| 1. false | 2. false | 3. true | 4. true | 5. true |
| 6. true | 7. true | 8. false | 9. false | |
-

Exercises

- | | |
|---|---|
| 1. a. $\pm 1, \pm 2, \pm 3, \pm 5, \pm 6, \pm 10, \pm 15, \pm 30$ | c. $\pm 1, \pm 2, \pm 4, \pm 7, \pm 14, \pm 28$ |
| e. $\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 8, \pm 12, \pm 24$ | g. $\pm 1, \pm 2, \pm 4, \pm 8, \pm 16, \pm 32$ |
| 2. a. $\pm 1, \pm 2$ | c. $\pm 1, \pm 2, \pm 4, \pm 8$ |
| 3. $q = 30, r = 16$ | 5. $q = 22, r = 5$ |
| 9. $q = -51, r = 4$ | 11. $q = 0, r = 26$ |
| 15. $q = 0, r = 0$ | 13. $q = -360, r = 3$ |
25. Counterexample: Let $a = 6, b = 8$, and $c = 9$.
29. If $a = 0$, then $n = -1$ makes $a - bn = 0 - b(-1) = b > 0$, and we have a positive element of S in this case. If $a \neq 0$, the choice $n = -2|a|$ gives $a - bn = a + 2b|a|$ as a specific example of a positive element of S . The problem does not explicitly require a proof that our element is positive, but this can be done as follows.

Since $b > 0$, we have $b \geq 1$ by Theorem 2.6. This implies $b|a| \geq |a|$ by Exercise 18 of Section 2.1. It follows from the definition of absolute value that $|a| \geq -a$. Now

$$b|a| \geq |a| \text{ and } |a| \geq -a \Rightarrow b|a| \geq -a.$$

Since $a \neq 0$, $|a| > 0$, and therefore, $|a| \geq 1$ by Theorem 2.6. Hence $b|a| \geq b$ by Exercise 18 of Section 2.1.

$$b|a| \geq b \text{ and } b > 0 \Rightarrow b|a| > 0$$

We have $b|a| \geq -a$ and $b|a| > 0$. By Exercise 14 of Section 2.1,

$$b|a| + b|a| > -a + 0,$$

$$2b|a| > -a, \text{ and}$$

$$a + 2b|a| > 0.$$

This shows that $a + 2b|a|$ is positive.

Exercises 2.4 Pages 92–95

True or False

- | | | | | | |
|----------|----------|---------|----------|-----------|-----------|
| 1. false | 2. false | 3. true | 4. true | 5. true | 6. true |
| 7. false | 8. false | 9. true | 10. true | 11. false | 12. false |
-

Exercises

1. $2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97$
2. a. $1400 = 2^3 \cdot 5^2 \cdot 7$; $980 = 2^2 \cdot 5 \cdot 7^2$; $(1400, 980) = 2^2 \cdot 5 \cdot 7 = 140$
c. $3780 = 2^2 \cdot 3^3 \cdot 5 \cdot 7$; $16,200 = 2^3 \cdot 3^4 \cdot 5^2$; $(3780, 16,200) = 2^2 \cdot 3^3 \cdot 5 = 540$
3. a. $(a, b) = 3, m = 0, n = -1$
c. $(a, b) = 6, m = 2, n = -3$
e. $(a, b) = 3, m = 2, n = 25$
g. $(a, b) = 9, m = -5, n = 3$
i. $(a, b) = 3, m = -49, n = 188$
k. $(a, b) = 12, m = -3, n = 146$
m. $(a, b) = 12, m = 5, n = 163$
4. a. $(4, 6) = 2$
19. Let $a = 2$ and $b = c = 3$. Then $(a, b) = (a, c) = (2, 3) = 1$, and $(ac, b) = (6, 3) = 3 \neq 1$.
28. After a and b are written in their standard forms, the least common multiple of a and b can be found by forming the product of all the distinct prime factors that appear in the standard form of either a or b , with each factor raised to the greatest power to which it appears in either standard form.
29. a. The least common multiple of $1400 = 2^3 \cdot 5^2 \cdot 7$ and $980 = 2^2 \cdot 5 \cdot 7^2$ is $2^3 \cdot 5^2 \cdot 7^2 = 9800$.
c. The least common multiple of $3780 = 2^2 \cdot 3^3 \cdot 5 \cdot 7$ and $16,200 = 2^3 \cdot 3^4 \cdot 5^2$ is $2^3 \cdot 3^4 \cdot 5^2 \cdot 7 = 113,400$.
30. a. An integer d is a **greatest common divisor** of a, b , and c if these conditions are satisfied:
 - (1) d is a positive integer.
 - (2) $d|a, d|b$, and $d|c$.
 - (3) If $n|a, n|b$, and $n|c$, then $n|d$.
31. a. $7 = 14(-2) + 28(0) + 35(1)$
c. $1 = 143(-53) + 385(18) + (-65)(-10)$

Exercises 2.5 Pages 103–106**True or False**

1. true 2. true 3. false 4. true 5. true 6. false 7. false
-

Exercises

1. $[0] = \{\dots, -5, 0, 5, \dots\}$, $[1] = \{\dots, -4, 1, 6, \dots\}$,
 $[2] = \{\dots, -3, 2, 7, \dots\}$, $[3] = \{\dots, -2, 3, 8, \dots\}$,
 $[4] = \{\dots, -1, 4, 9, \dots\}$
3. $x = 5$ 5. $x = 11$ 7. $x = 8$ 9. $x = 173$ 11. $x = 28$
13. $x = 7$ 15. $x = 4$ 17. $x = 6$ 19. $x = 11$ 21. $x = 13$
23. $x = 2$ 29. a. 1 c. 8 e. 1 g. 3 i. 3 k. 2
39. $d = (6, 27) = 3$ and 3 divides 33; $x = 1, x = 10, x = 19$ are solutions.
41. $d = (8, 78) = 2$ and 2 divides 66; $x = 18$ and $x = 57$ are solutions.
43. $d = (68, 40) = 4$ and 4 divides 36; $x = 7, x = 17, x = 27, x = 37$ are solutions.
45. $d = (24, 348) = 12$ and 12 does not divide 45; therefore, there are no solutions.
47. $d = (15, 110) = 5$ and 5 divides 130; $x = 16, x = 38, x = 60, x = 82$, and $x = 104$ are solutions.
49. $d = (42, 74) = 2$ and 2 divides 30; $x = 6$ and $x = 43$ are solutions.
53. a. $x = 27$ or $x \equiv 27 \pmod{40}$ c. $x = 11$ or $x \equiv 11 \pmod{56}$
e. $x = 14$ or $x \equiv 14 \pmod{120}$ g. $x = 347$ or $x \equiv 347 \pmod{840}$

Exercises 2.6 Pages 112–114**True or False**

1. true 2. false 3. false 4. false
-

Exercises

1. a. $[3]$ c. $[4]$ e. $[6][4] = [0]$ g. $[6] + [6] = [0]$
2. a. $[1][2][3][4] = [24] = [4]$ c. $[1][2][3] = [6] = [2]$
3. a.

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

c.

$+$	[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]

e.

$+$	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[1]	[1]	[2]	[3]	[4]	[5]	[6]	[0]
[2]	[2]	[3]	[4]	[5]	[6]	[0]	[1]
[3]	[3]	[4]	[5]	[6]	[0]	[1]	[2]
[4]	[4]	[5]	[6]	[0]	[1]	[2]	[3]
[5]	[5]	[6]	[0]	[1]	[2]	[3]	[4]
[6]	[6]	[0]	[1]	[2]	[3]	[4]	[5]

4. a.

\times	[0]	[1]
[0]	[0]	[0]
[1]	[0]	[1]

c.

\times	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

e.	\times	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
[2]	[0]	[2]	[4]	[6]	[1]	[3]	[5]	
[3]	[0]	[3]	[6]	[2]	[5]	[1]	[4]	
[4]	[0]	[4]	[1]	[5]	[2]	[6]	[3]	
[5]	[0]	[5]	[3]	[1]	[6]	[4]	[2]	
[6]	[0]	[6]	[5]	[4]	[3]	[2]	[1]	

5. a. [9] c. [13] e. [5] g. [173]
 6. a. [1], [5] c. [1], [3], [7], [9] e. [1], [5], [7], [11], [13], [17]
 7. a. [2], [3], [4] c. [2], [4], [5], [6], [8]
 e. [2], [3], [4], [6], [8], [9], [10], [12], [14], [15], [16]
 8. a. $[x] = [2]$ or $[x] = [5]$ c. $[x] = [2]$ or $[x] = [6]$
 e. No solution exists.
 g. $[x] = [2]$, $[x] = [5]$, $[x] = [8]$, or $[x] = [11]$
 i. $[x] = [4]$ or $[x] = [10]$
 10. a. $[x] = [4]^{-1}[5] = [10][5] = [11]$ c. $[x] = [7]^{-1}[11] = [7][11] = [5]$
 e. $[x] = [9]^{-1}[14] = [9][14] = [6]$ g. $[x] = [6]^{-1}[5] = [266][5] = [54]$
 11. $[x] = [3]$, $[y] = [5]$
 13. $[x] = [3]$, $[y] = [3]$
 19. a. $[x] = [4]$ or $[x] = [5]$ c. $[x] = [1]$ or $[x] = [5]$

Exercises 2.7 Pages 119–123

True or False

1. false 2. false 3. true 4. false

Exercises

1. Errors occur in 00010 and 11100.

3. Correct coded message:

101101101 110110110 100100100 101101101 010010010 011011011

Decoded message: 101 110 100 101 010 011

5. a. $\frac{3}{4}$ c. $\frac{2}{6} = \frac{1}{3}$
6. a. $(0.97)^4 + 4(0.97)^3(0.03) = 0.9948136$
7. a. $(0.9999)^8 = 0.9992003$
c. $(0.9999)^8 + 8(0.9999)^7(0.0001) = 0.9999997$
e. 1.000000
9. 1 14. a. 7 c. 1 17. a. valid c. not valid
18. a. No error is detected. c. An error is detected.
19. $y = -(10, 9, 8, 7, 6, 5, 4, 3, 2)$ 20. a. 3 c. 3
22. a. 3 c. 3 23. 2 25. 3

Exercises 2.8 Pages 130–134

True or False

1. true 2. true 3. true

Exercises

1. Ciphertext: APMHKPMKSHQ HQVHAPMHUIQT

$$f^{-1}(x) = x + 19 \bmod 27$$

3. Plaintext: “tiger, do you read me?”

$$f^{-1}(x) = x + 20 \bmod 31$$

5. Ciphertext: FBBZXLXDGIXZUW

$$f^{-1}(x) = 4x + 7 \bmod 27$$

7. Plaintext: www.brookscole.com

$$f^{-1}(x) = 19x + 2 \bmod 28$$

9. Plaintext: mathematics

$$f(x) = 9x + 13 \bmod 26$$

$$f^{-1}(x) = 3x + 13 \bmod 26$$

11. Plaintext: there are 25 primes less than 100

$$f(x) = 12x + 17 \bmod 37$$

$$f^{-1}(x) = 34x + 14 \bmod 37$$

15. a. $n - 1$

b. $(n - 1)n - 1 = n^2 - n - 1$

$$\{1p, 2p, 3p, \dots, (p^{j-1} - 1)p, p^{j-1}p\}.$$

Since this set contains p^{j-1} elements,

$$\phi(p^j) = p^j - p^{j-1} = p^{j-1}(p - 1).$$

Exercises 3.1 Pages 141–145

True or False

- 1.** true **2.** false **3.** false **4.** false **5.** false **6.** false

Exercises

1. group
 3. The set of all positive irrational numbers with the operation of multiplication does not form a group. The set is not closed with respect to multiplication. For example, $\sqrt{2}$ is a positive irrational number, but $\sqrt{2} \sqrt{2} = 2$ is not. Also, there is no identity element.
 5. The set of all real numbers x such that $0 < x \leq 1$ is not a group with respect to multiplication because not all elements have inverses.
 7. group 9. group 11. group
 13. The operation \times is not associative, since

$$a \times (c \times a) = a \times e = a,$$

whereas

$$(a \times c) \times a = b \times a = c.$$

Also, there are no inverses for the elements a and b .

15. The set \mathbf{Z} is an abelian group with respect to $*$. The identity element is -1 . The element $-x - 2$ is the inverse of the element $x \in \mathbf{Z}$.
17. The set \mathbf{Z} is not a group and hence, not an abelian group with respect to the operation $*$. The operation is not associative. There is no identity element and hence no inverse elements.
19. The set \mathbf{Z} is not a group and hence, not an abelian group with respect to $*$. The identity element is 0 , but 1 does not have an inverse in \mathbf{Z} .
21. group, 2
23. The set is not a group with respect to multiplication, since it does not have an identity element and hence has no inverse elements.
25. group, 5

27. a. $n = 1$

b.

\times	[1]	[2]	[3]	[4]	[5]	[6]
[1]	[1]	[2]	[3]	[4]	[5]	[6]
[2]	[2]	[4]	[6]	[1]	[3]	[5]
[3]	[3]	[6]	[2]	[5]	[1]	[4]
[4]	[4]	[1]	[5]	[2]	[6]	[3]
[5]	[5]	[3]	[1]	[6]	[4]	[2]
[6]	[6]	[5]	[4]	[3]	[2]	[1]

$[1]^{-1} = [1]$, $[2]$ and $[4]$ are inverses of each other,
 $[3]$ and $[5]$ are inverses of each other, and $[6]^{-1} = [6]$.

29.

\times	I_3	P_1	P_2	P_3	P_4	P_5
I_3	I_3	P_1	P_2	P_3	P_4	P_5
P_1	P_1	I_3	P_3	P_2	P_5	P_4
P_2	P_2	P_5	I_3	P_4	P_3	P_1
P_3	P_3	P_4	P_1	P_5	P_2	I_3
P_4	P_4	P_3	P_5	P_1	I_3	P_2
P_5	P_5	P_2	P_4	I_3	P_1	P_3

35. The set G is not a group with respect to addition, since it does not contain an identity element.

36. b. 2^n

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

+	\emptyset	$\{a\}$	$\{b\}$	$\{c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	A
\emptyset	\emptyset	$\{a\}$	$\{b\}$	$\{c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	A
$\{a\}$	$\{a\}$	\emptyset	$\{a, b\}$	$\{a, c\}$	$\{b\}$	$\{c\}$	A	$\{b, c\}$
$\{b\}$	$\{b\}$	$\{a, b\}$	\emptyset	$\{b, c\}$	$\{a\}$	A	$\{c\}$	$\{a, c\}$
$\{c\}$	$\{c\}$	$\{a, c\}$	$\{b, c\}$	\emptyset	A	$\{a\}$	$\{b\}$	$\{a, b\}$
$\{a, b\}$	$\{a, b\}$	$\{b\}$	$\{a\}$	A	\emptyset	$\{b, c\}$	$\{a, c\}$	$\{c\}$
$\{a, c\}$	$\{a, c\}$	$\{c\}$	A	$\{a\}$	$\{b, c\}$	\emptyset	$\{a, b\}$	$\{b\}$
$\{b, c\}$	$\{b, c\}$	A	$\{c\}$	$\{b\}$	$\{a, c\}$	$\{a, b\}$	\emptyset	$\{a\}$
A	A	$\{b, c\}$	$\{a, c\}$	$\{a, b\}$	$\{c\}$	$\{b\}$	$\{a\}$	\emptyset

- 39.** The set A is an identity element. But the set $\mathcal{P}(A)$ is not a group with respect to the operation of intersection, since A is the only element that has an inverse.

Exercises 3.2 Pages 150–152

True or False

- 1.** false **2.** true **3.** true **4.** false **5.** false **6.** false

Exercises

- 5.** One possible choice is $a = \rho$ and $b = \sigma$. Then $(ab)^{-1} = (\rho \circ \sigma)^{-1} = \gamma^{-1} = \gamma$ and $a^{-1}b^{-1} = \rho^{-1} \circ \sigma^{-1} = \rho^2 \circ \sigma = \delta$, so $(ab)^{-1} \neq a^{-1}b^{-1}$.
- 7.** One possible choice is $a = \rho$ and $b = \delta$. Then $(ab)^2 = (\rho \circ \delta)^2 = \sigma^2 = e$ and $a^2b^2 = \rho^2 \circ \delta^2 = \rho^2 \circ e = \rho^2$, so $(ab)^2 \neq a^2b^2$.
- 9. b.** $\{x\}$

11.

\times	a	b	c	d
a	c	d	a	b
b	d	c	b	a
c	a	b	c	d
d	b	a	d	c

17. $(abcd)^{-1} = ((d^{-1}c^{-1})b^{-1})a^{-1}$

- 19.** Consider the set $S = \{a \in G | a \neq a^{-1}\}$. Now $a \in S$ if and only if $a^{-1} \in S$, so S has an even number of elements. Since both G and S have an even number of elements, the complement $G - S = \{a \in G | a = a^{-1}\}$ must also have an even number of elements. The element e is in $G - S$ since $e = e^{-1}$ and therefore there is at least one $a \neq e$ such that $a = a^{-1}$.

26. a. $\begin{bmatrix} [4] & [2] & [5] \\ [0] & [1] & [3] \end{bmatrix}$

27. a. $\begin{bmatrix} [3] & [1] \\ [4] & [2] \end{bmatrix}$

Exercises 3.3 Pages 159–163

True or False

- | | | | | |
|-----------------|-----------------|-----------------|-----------------|------------------|
| 1. true | 2. true | 3. false | 4. false | 5. true |
| 6. false | 7. false | 8. false | 9. true | 10. false |

Exercises

- 1. a.** The set $\{e, \sigma\}$ is a subgroup of $\mathcal{S}(A)$.
The multiplication table is

◦	e	σ
e	e	σ
σ	σ	e

- c.** The set $\{e, \rho\}$ is not a subgroup of $\mathcal{S}(A)$, since it is not closed. We have $\rho \circ \rho = \rho^2 \notin \{e, \rho\}$. The multiplication table is

◦	e	ρ
e	e	ρ
ρ	ρ	ρ^2

- e.** The set $\{e, \rho, \rho^2\}$ is a subgroup of $\mathcal{S}(A)$. The multiplication table is

◦	e	ρ	ρ^2
e	e	ρ	ρ^2
ρ	ρ	ρ^2	e
ρ^2	ρ^2	e	ρ

- g.** The set $\{e, \sigma, \gamma\}$ is not a subgroup of $\mathcal{S}(A)$, since it is not closed. We have $\gamma \circ \sigma = \rho \notin \{e, \sigma, \gamma\}$. The multiplication table is

◦	e	σ	γ
e	e	σ	γ
σ	σ	e	ρ^2
γ	γ	ρ	e

2. a. subgroup

c. The set $\{i, -i\}$ is not a subgroup of G , since it is not closed. We have $i \cdot i = -1 \notin \{i, -i\}$.

3. $\langle [6] \rangle = \{[0], [2], [4], [6], [8], [10], [12], [14]\}$, $o(\langle [6] \rangle) = 8$

5. a. $\{[1], [3], [4], [9], [10], [12]\}$, $o(\langle [4] \rangle) = 6$

6. a. $\langle A \rangle = \left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$, $o(\langle A \rangle) = 4$

c. $\langle A \rangle = \left\{ \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$, $o(\langle A \rangle) = 3$

7. a. $\langle A \rangle = \left\{ \begin{bmatrix} [2] & [0] \\ [0] & [3] \end{bmatrix}, \begin{bmatrix} [4] & [0] \\ [0] & [1] \end{bmatrix}, \begin{bmatrix} [1] & [0] \\ [0] & [4] \end{bmatrix}, \begin{bmatrix} [3] & [0] \\ [0] & [2] \end{bmatrix}, \begin{bmatrix} [0] & [0] \\ [0] & [0] \end{bmatrix} \right\}$,
 $o(\langle A \rangle) = 5$

9. The set of all real numbers that are greater than 1 is closed under multiplication but is not a subgroup of G , since it does not contain inverses. (If $x > 1$, then $x^{-1} < 1$.)

17. b. $[x] = \{x\}$

18. a. $\{1, -1\}$

c. $\{I_3\}$

27. a. $C_1 = C_{-1} = G$, $C_i = C_{-i} = \{1, i, -1, -i\}$, $C_j = C_{-j} = \{1, j, -1, -j\}$,
 $C_k = C_{-k} = \{1, k, -1, -k\}$

c. $C_{I_3} = G$, $C_{P_1} = \{I_3, P_1\}$, $C_{P_2} = \{I_3, P_2\}$, $C_{P_3} = C_{P_5} = \{I_3, P_3, P_5\}$, $C_{P_4} = \{I_3, P_4\}$

30. The subgroup $\langle m \rangle \cap \langle n \rangle$ is the set of all multiples of the least common multiple of m and n .

33. Let $H = \{e, \sigma\}$ and $K = \{e, \gamma\}$.

35. Let $H = \{e, \sigma\}$ and $K = \{e, \gamma\}$.

Exercises 3.4 Pages 170–174

True or False

1. true

2. true

3. false

4. false

5. true

6. true

7. false

8. true

9. false

10. true

Exercises

1. $\langle e \rangle = \{e\}$, $\langle \rho \rangle = \{e, \rho, \rho^2\}$, $\langle \sigma \rangle = \{e, \sigma\}$, $\langle \gamma \rangle = \{e, \gamma\}$, $\langle \delta \rangle = \{e, \delta\}$

3. The element e has order 1. Each of the elements σ , γ , and δ has order 2. Each of the elements ρ and ρ^2 has order 3.

5. $o(I_3) = 1$, $o(P_1) = o(P_2) = o(P_4) = 2$, $o(P_3) = o(P_5) = 3$

6. a. $o(A) = 2$

7. a. 4 c. 2 e. 4 g. 1

8. a. 9 c. 9 e. 3 g. 9

9. a. [1], [3], [5], [7]

c. [1], [3], [7], [9]

e. [1], [3], [5], [7], [9], [11], [13], [15]

10. a. $\{[0]\}, 1; \{[0], [6]\}, 2; \{[0], [4], [8]\}, 3; \{[0], [3], [6], [9]\}, 4;$
 $\{[0], [2], [4], [6], [8], [10]\}, 6; \mathbf{Z}_{12}, 12$

c. $\{[0]\}, 1; \{[0], [5]\}, 2; \{[0], [2], [4], [6], [8]\}, 5; \mathbf{Z}_{10}, 10$

e. $\{[0]\}, 1; \{[0], [8]\}, 2; \{[0], [4], [8], [12]\}, 4;$
 $\{[0], [2], [4], [6], [8], [10], [12], [14]\}, 8; \mathbf{Z}_{16}, 16$

11. a. $G = \langle [3] \rangle = \langle [5] \rangle$

c. $G = \langle [2] \rangle = \langle [6] \rangle = \langle [7] \rangle = \langle [8] \rangle$

e. $G = \langle [3] \rangle = \langle [5] \rangle = \langle [6] \rangle = \langle [7] \rangle = \langle [10] \rangle = \langle [11] \rangle = \langle [12] \rangle = \langle [14] \rangle$

12. a. [3], [5]

c. [2], [6], [7], [8]

e. [3], [5], [6], [7], [10], [11], [12], [14]

13. a. $\{[1]\}, 1; \{[1], [6]\}, 2; \{[1], [2], [4]\}, 3; G, 6$

c. $\{[1]\}, 1; \{[1], [10]\}, 2; \{[1], [3], [4], [5], [9]\}, 5; G, 10$

e. $\{[1]\}, 1; \{[1], [16]\}, 2; \{[1], [4], [13], [16]\}, 4;$
 $\{[1], [2], [4], [8], [9], [13], [15], [16]\}, 8; G, 16$

15. c. $H = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \right\}$

18. a. $\mathbf{U}_{20} = \{[1], [3], [7], [9], [11], [13], [17], [19]\}$

.	[1]	[3]	[7]	[9]	[11]	[13]	[17]	[19]
[1]	[1]	[3]	[7]	[9]	[11]	[13]	[17]	[19]
[3]	[3]	[9]	[1]	[7]	[13]	[19]	[11]	[17]
[7]	[7]	[1]	[9]	[3]	[17]	[11]	[19]	[13]
[9]	[9]	[7]	[3]	[1]	[19]	[17]	[13]	[11]
[11]	[11]	[13]	[17]	[19]	[1]	[3]	[7]	[9]
[13]	[13]	[19]	[11]	[17]	[3]	[9]	[1]	[7]
[17]	[17]	[11]	[19]	[13]	[7]	[1]	[9]	[3]
[19]	[19]	[17]	[13]	[11]	[9]	[7]	[3]	[1]

c. $\mathbf{U}_{24} = \{[1], [5], [7], [11], [13], [17], [19], [23]\}$

.	[1]	[5]	[7]	[11]	[13]	[17]	[19]	[23]
[1]	[1]	[5]	[7]	[11]	[13]	[17]	[19]	[23]
[5]	[5]	[1]	[11]	[7]	[17]	[13]	[23]	[19]
[7]	[7]	[11]	[1]	[5]	[19]	[23]	[13]	[17]
[11]	[11]	[7]	[5]	[1]	[23]	[19]	[17]	[13]
[13]	[13]	[17]	[19]	[23]	[1]	[5]	[7]	[11]
[17]	[17]	[13]	[23]	[19]	[5]	[1]	[11]	[7]
[19]	[19]	[23]	[13]	[17]	[7]	[11]	[1]	[5]
[23]	[23]	[19]	[17]	[13]	[11]	[7]	[5]	[1]

19. a. not cyclic

c. not cyclic

21. a. $\phi(8) = 4; a, a^3, a^5, a^7$

c. $\phi(18) = 6; a, a^5, a^7, a^{11}, a^{13}, a^{17}$

e. $\phi(7) = 6; a, a^2, a^3, a^4, a^5, a^6$

22. a. $\langle a \rangle = G$

$\langle a^2 \rangle = \langle a^6 \rangle = \{a^2, a^4, a^6, a^8 = e\}$

$\langle a^4 \rangle = \{a^4, a^8 = e\}$

$\langle a^8 \rangle = \langle e \rangle = \{e\}$

c. $\langle a \rangle = G$

$\langle a^2 \rangle = \langle a^4 \rangle = \langle a^8 \rangle = \langle a^{10} \rangle = \langle a^{14} \rangle = \langle a^{16} \rangle = \{a^2, a^4, a^6, a^8, a^{10}, a^{12}, a^{14}, a^{16}, a^{18} = e\}$

$\langle a^3 \rangle = \langle a^{15} \rangle = \{a^3, a^6, a^9, a^{12}, a^{15}, a^{18} = e\}$

$\langle a^6 \rangle = \langle a^{12} \rangle = \{a^6, a^{12}, a^{18} = e\}$

$\langle a^9 \rangle = \{a^9, a^{18} = e\}$

$\langle a^{18} \rangle = \langle e \rangle = \{e\}$

e. $\langle a \rangle = G, \langle a^7 \rangle = \langle e \rangle = \{e\}$

23. a. a^{12}

c. a^6, a^{18}

24. a. none

c. $a^5, a^{10}, a^{15}, a^{20}, a^{25}, a^{30}$

25. All subgroups of \mathbf{Z} are of the form $\langle n \rangle$, n a fixed integer.

35. $p - 1$

Exercises 3.5 Pages 180–183

True or False

1. true 2. false 3. false 4. true 5. true
 6. false 7. true 8. true
-

Exercises

3. Let $\phi: \mathbf{Z}_4 \rightarrow \mathbf{U}_5$ be defined by

$$\phi([0]_4) = [1]_5, \quad \phi([1]_4) = [2]_5, \quad \phi([2]_4) = [4]_5, \quad \phi([3]_4) = [3]_5.$$

5. Let $\phi: H \rightarrow \mathcal{S}(A)$ be defined by

$$\phi(I_2) = I_A, \quad \phi(M_1) = \sigma, \quad \phi(M_2) = \rho, \quad \phi(M_3) = \rho^2, \quad \phi(M_4) = \gamma, \quad \phi(M_5) = \delta.$$

7. Let $\phi: \mathbf{Z} \rightarrow H$ be defined by $\phi(n) = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}, n \in \mathbf{Z}$. Then

$$\phi(n+m) = \begin{bmatrix} 1 & n+m \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix} = \phi(n) \cdot \phi(m)$$

for all $n, m \in \mathbf{Z}$.

9. Define $\phi: G \rightarrow H$ by $\phi(a+bi) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ for $a+bi \in G$. Let $x = a+bi \in G$ and $y = c+di \in G$. Then

$$\begin{aligned} \phi(xy) &= \phi((a+bi)(c+di)) \\ &= \phi((ac-bd)+(bc+ad)i) \\ &= \begin{bmatrix} ac-bd & -bc-ad \\ bc+ad & ac-bd \end{bmatrix} \\ &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} c & -d \\ d & c \end{bmatrix} \\ &= \phi(a+bi)\phi(c+di) \\ &= \phi(x)\phi(y). \end{aligned}$$

11. Define $\phi: H \rightarrow G$ by

$$\begin{array}{ll} \phi\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = 1 & \phi\left(\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\right) = -1 \\ \phi\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right) = i & \phi\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) = -i \\ \phi\left(\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}\right) = j & \phi\left(\begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}\right) = -j \\ \phi\left(\begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}\right) = k & \phi\left(\begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}\right) = -k. \end{array}$$

- 22. a.** For notational convenience we let a represent $[a]$. The elements 2 and $2^3 = 3$ are generators of \mathbf{U}_5 . The automorphisms of \mathbf{U}_5 are ϕ_1 and ϕ_2 defined by

$$\phi_1: \begin{cases} \phi_1(1) = 1 \\ \phi_1(2) = 2 \\ \phi_1(3) = 3 \\ \phi_1(4) = 4 \end{cases} \quad \phi_2: \begin{cases} \phi_2(2) = 3 \\ \phi_2(2^2) = \phi_2(4) = 3^2 = 4 \\ \phi_2(2^3) = \phi_2(3) = 3^3 = 2 \\ \phi_2(2^4) = \phi_2(1) = 3^4 = 1 \end{cases}$$

- 23. a. 2**

- c. 4**

- 27.** Suppose $[a]_7$ represents $[a]$ in \mathbf{U}_7 and $[a]_6$ represents $[a]$ in \mathbf{Z}_6 . Let $\phi_1: \mathbf{U}_7 \rightarrow \mathbf{Z}_6$ and $\phi_2: \mathbf{U}_7 \rightarrow \mathbf{Z}_6$ be defined by

$$\phi_1: \begin{cases} \phi_1([1]_7) = [0]_6 \\ \phi_1([2]_7) = [2]_6 \\ \phi_1([3]_7) = [1]_6 \\ \phi_1([4]_7) = [4]_6 \\ \phi_1([5]_7) = [5]_6 \\ \phi_1([6]_7) = [3]_6 \end{cases} \quad \text{and} \quad \phi_2: \begin{cases} \phi_2([1]_7) = [0]_6 \\ \phi_2([2]_7) = [4]_6 \\ \phi_2([3]_7) = [5]_6 \\ \phi_2([4]_7) = [2]_6 \\ \phi_2([5]_7) = [1]_6 \\ \phi_2([6]_7) = [3]_6 \end{cases}$$

- 31.** The cyclic group of order 4 and the Klein 4-group.

Exercises 3.6 Pages 186–188

True or False

- | | | | | |
|----------|----------|----------|----------|-----------|
| 1. false | 2. true | 3. false | 4. false | 5. false |
| 6. true | 7. false | 8. true | 9. true | 10. false |
-

Exercises

- 1. a.** ϕ is an endomorphism and $\ker \phi = \{\pm 1\}$. ϕ is not an epimorphism nor a monomorphism.
c. ϕ is not an endomorphism.
e. ϕ is an endomorphism and $\ker \phi = \mathbf{R}^+$. ϕ is not an epimorphism nor a monomorphism.
g. ϕ is an endomorphism and $\ker \phi = \{1\}$. ϕ is an epimorphism and a monomorphism.
- 2. a.** ϕ is a homomorphism and $\ker \phi = \{[0], [2]\}$. ϕ is an epimorphism but not a monomorphism.
- 3.** $\ker \phi = \{0\}$, ϕ is not an epimorphism, and ϕ is a monomorphism.
- 5.** $\ker \phi = \{[0], [4], [8]\}$, ϕ is not an epimorphism, and ϕ is not a monomorphism.

7. $\ker \phi = \{[0]_8, [4]_8\}$, ϕ is an epimorphism, and ϕ is not a monomorphism.
 9. ϕ is an epimorphism but not a monomorphism.

Exercises 4.1 Pages 202–204

True or False

- | | | | | | |
|----------|----------|----------|----------|----------|-----------|
| 1. true | 2. false | 3. false | 4. true | 5. true | 6. true |
| 7. false | 8. false | 9. false | 10. true | 11. true | 12. false |
-

Exercises

1. a. $(1, 4)(2, 5); \{1, 4\}, \{2, 5\}$
 c. $(1, 4, 5, 2); \{1, 4, 5, 2\}$
 e. $(1, 3, 5)(2, 4, 6); \{1, 3, 5\}, \{2, 4, 6\}$
 g. $(1, 4)(2, 3, 5); \{1, 4\}, \{2, 3, 5\}$
2. a. $(1, 4, 8, 7, 2, 3)(5, 9, 6); \{1, 4, 8, 7, 2, 3\}, \{5, 9, 6\}$
 c. $(1, 4, 8, 7)(2, 6, 5, 3); \{1, 4, 8, 7\}, \{2, 6, 5, 3\}$
 e. $(1, 2)(3, 4, 5); \{1, 2\}, \{3, 4, 5\}$
 g. $(1, 7, 6, 4, 3, 5, 2); \{1, 7, 6, 4, 3, 5, 2\}$
3. a. even c. odd e. even g. odd
4. a. odd c. even e. odd g. even
5. a. two c. four e. three g. six
6. a. six c. four e. six g. seven
7. a. $(1, 4)(2, 5)$ b. $(1, 3)(1, 2)(1, 7)(1, 8)(1, 4)(5, 6)(5, 9)$
 c. $(1, 2)(1, 5)(1, 4)$ d. $(1, 7)(1, 8)(1, 4)(2, 3)(2, 5)(2, 6)$
 e. $(1, 5)(1, 3)(2, 6)(2, 4)$ f. $(1, 2)(3, 5)(3, 4)$
 g. $(1, 4)(2, 5)(2, 3)$ h. $(1, 2)(1, 5)(1, 3)(1, 4)(1, 6)(1, 7)$
9. a. $f^2 = (1, 2)(4, 5), f^3 = f^{-1} = (1, 4, 2, 5)$
 c. $f^2 = f^{-1} = (1, 2, 6)(3, 5, 4), f^3 = (1)$
 e. $f^2 = (1, 8, 2)(3, 7, 6, 4, 5), f^3 = (3, 5, 4, 6, 7), f^{-1} = (1, 8, 2)(3, 6, 5, 7, 4)$
10. a. $(3, 1, 4, 2) = (1, 4, 2, 3)$ b. $(1, 2)(4, 9)(5, 6)$
 c. $(1, 2, 4, 5)$ d. $(1, 2)(3, 4, 5)$
 e. $(1, 4, 2)(5, 3) = (1, 4, 2)(3, 5)$ f. $(3, 7, 4, 5)(6, 8)$
13. $g = f^4 = (1, 5, 9)(2, 6, 10)(3, 7, 11)(4, 8, 12),$
 $h = f^9 = (1, 10, 7, 4)(2, 11, 8, 5)(3, 12, 9, 6)$

- 15.** $(1, 2, 3, 4)$ $(1, 2, 3)$ $(1, 2)$ $(1, 2)(3, 4)$
 $(1, 2, 4, 3)$ $(1, 3, 2)$ $(1, 3)$ $(1, 3)(2, 4)$
 $(1, 3, 2, 4)$ $(1, 2, 4)$ $(1, 4)$ $(1, 4)(2, 3)$
 $(1, 3, 4, 2)$ $(1, 4, 2)$ $(2, 3)$ (1)
 $(1, 4, 2, 3)$ $(1, 3, 4)$ $(2, 4)$
 $(1, 4, 3, 2)$ $(1, 4, 3)$ $(3, 4)$
 $(2, 3, 4)$
 $(2, 4, 3)$

- 17.** $\langle(1, 2)\rangle = \{(1), (1, 2)\}$ has order 2.

$\langle(1, 2, 3)\rangle = \{(1), (1, 2, 3), (1, 3, 2)\}$ has order 3.

$\langle(1, 2, 3, 4)\rangle = \{(1), (1, 2, 3, 4), (1, 3)(2, 4), (1, 4, 3, 2)\}$ has order 4.

- 19.** $\{e\}, \{e, \beta\}, \{e, \gamma\}, \{e, \Delta\}, \{e, \theta\}, \{e, \alpha^2\}, \{e, \alpha, \alpha^2, \alpha^3\}$

Exercises 4.2 Pages 207–208

True or False

- 1.** true

Exercises

- 1.** For notational convenience, we let a represent $[a]$ in this solution. With $f_g: \mathbf{Z}_3 \rightarrow \mathbf{Z}_3$ defined by $f_g(x) = g + x$ for each $g \in \mathbf{Z}_3$, we obtain the following permutations on the set of elements in \mathbf{Z}_3 :

$$f_0 = (0), \quad f_1 = (0, 1, 2), \quad f_2 = (0, 2, 1).$$

The set $G' = \{f_0, f_1, f_2\}$ is a group of permutations, and the mapping $\phi: \mathbf{Z}_3 \rightarrow G'$ defined by

$$\phi: \begin{cases} \phi(0) = f_0 \\ \phi(1) = f_1 \\ \phi(2) = f_2 \end{cases}$$

is an isomorphism from \mathbf{Z}_3 to G' .

- 3.** With $f_g: G \rightarrow G$ defined by $f_g(x) = gx$ for each $g \in G$, we obtain the following permutations on the set of elements of G :

$$f_e = (e), \quad f_a = (e, a)(b, ab), \quad f_b = (e, b)(a, ab), \quad f_{ab} = (e, ab)(a, b).$$

The set $G' = \{f_e, f_a, f_b, f_{ab}\}$ is a group of permutations, and the mapping $\phi: G \rightarrow G'$ defined by

$$\phi: \begin{cases} \phi(e) = f_e \\ \phi(a) = f_a \\ \phi(b) = f_b \\ \phi(ab) = f_{ab} \end{cases}$$

is an isomorphism from G to G' .

5. For notational convenience, we let a represent $[a]$ in this solution.

Let $f_a: G \rightarrow G$ be defined by $f_a(x) = ax$ for each $x \in G$. Then we have the following permutations:

$$f_2 = (2, 4, 8, 6), \quad f_4 = (2, 8)(4, 6), \quad f_6 = (6), \quad f_8 = (2, 6, 8, 4).$$

The set $G' = \{f_2, f_4, f_6, f_8\}$ is a group of permutations, and the mapping $\phi: G \rightarrow G'$ defined by

$$\phi: \begin{cases} \phi(2) = f_2 \\ \phi(4) = f_4 \\ \phi(6) = f_6 \\ \phi(8) = f_8 \end{cases}$$

is an isomorphism from G to G' .

7. With $f_g: G \rightarrow G$ defined by $f_g(x) = gx$ for each $g \in G$, we obtain the following permutations on the set of elements in G :

$$\begin{array}{ll} f_e = (e) & f_\alpha = (e, \alpha, \alpha^2, \alpha^3)(\beta, \gamma, \Delta, \theta) \\ f_{\alpha^2} = (e, \alpha^2)(\alpha, \alpha^3)(\beta, \Delta)(\gamma, \theta) & f_{\alpha^3} = (e, \alpha^3, \alpha^2, \alpha)(\beta, \theta, \Delta, \gamma) \\ f_\beta = (e, \beta)(\alpha, \theta)(\alpha^2, \Delta)(\alpha^3, \gamma) & f_\gamma = (e, \gamma)(\alpha, \beta)(\alpha^2, \theta)(\alpha^3, \Delta) \\ f_\Delta = (e, \Delta)(\alpha, \gamma)(\alpha^2, \beta)(\alpha^3, \theta) & f_\theta = (e, \theta)(\alpha, \Delta)(\alpha^2, \gamma)(\alpha^3, \beta) \end{array}$$

The set $G' = \{f_e, f_\alpha, f_{\alpha^2}, f_{\alpha^3}, f_\beta, f_\gamma, f_\Delta, f_\theta\}$ is a group of permutations, and the mapping $\phi: G \rightarrow G'$ defined by

$$\phi: \begin{cases} \phi(e) = f_e \\ \phi(\alpha) = f_\alpha \\ \phi(\alpha^2) = f_{\alpha^2} \\ \phi(\alpha^3) = f_{\alpha^3} \\ \phi(\beta) = f_\beta \\ \phi(\gamma) = f_\gamma \\ \phi(\Delta) = f_\Delta \\ \phi(\theta) = f_\theta \end{cases}$$

is an isomorphism from G to G' .

9. c. The mapping is an isomorphism.

Exercises 4.3 Pages 212–214

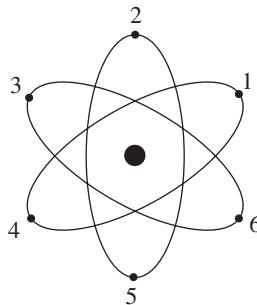
True or False

1. true 2. false 3. true 4. false 5. true 6. false 7. false

Exercises

1. $\{I, V\}$, where I is the identity mapping and V is the reflection about the vertical axis of symmetry

3. $\{I, R\}$, where I is the identity mapping and R is the counterclockwise rotation through 180° about the center of symmetry
5. rotational symmetry only
7. reflective symmetry only
9. both rotational symmetry and reflective symmetry
11. $\{R, R^2, R^3 = I\}$, where I is the identity mapping and R is the counterclockwise rotation through 120° about the center of the triangle determined by the arrow tips
13. Let the vertices of the ellipses be numbered as in the following figure.



Then any symmetry of the figure can be identified with the corresponding permutation on $\{1, 2, 3, 4, 5, 6\}$, and the group G of symmetries of the figure can be described with the notation

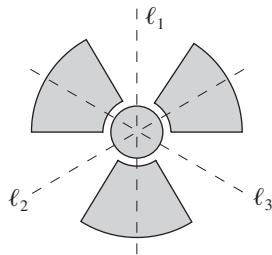
$$G = \{R, R^2, R^3, R^4, R^5, R^6 = I, L, LR, LR^2, LR^3, LR^4, LR^5\},$$

where

$$\begin{array}{ll} I = (1) & L = (2, 6)(3, 5) \\ R = (1, 2, 3, 4, 5, 6) & LR = (1, 6)(2, 5)(3, 4) \\ R^2 = (1, 3, 5)(2, 4, 6) & LR^2 = (1, 5)(2, 4) \\ R^3 = (1, 4)(2, 5)(3, 6) & LR^3 = (1, 4)(2, 3)(5, 6) \\ R^4 = (1, 5, 3)(2, 6, 4) & LR^4 = (1, 3)(4, 6) \\ R^5 = (1, 6, 5, 4, 3, 2) & LR^5 = (1, 2)(3, 6)(4, 5). \end{array}$$

This is the same permutation group as the one in the answer to Exercise 26 of this exercise set.

15. Let the axes of symmetry be labeled as in the following figure.



Then the group G of symmetries of the figure can be described as

$$G = \{I, R, R^2, L, LR, LR^2\},$$

where

I is the identity mapping,

R is the rotation through 120° counterclockwise about the center,

R^2 is the rotation through 240° counterclockwise about the center,

L is the reflection about the vertical axis ℓ_1 ,

LR is the reflection about the axis ℓ_2 , and

LR^2 is the reflection about the axis ℓ_3 .

17. Let I denote the identity mapping, and let t denote a translation of the set of **E**'s one unit to the right. Then t^{-1} is a translation of the set of **E**'s one unit to the left, and the collection

$$\{\dots, t^{-2}, t^{-1}, t^0 = I, t, t^2, \dots\}$$

are elements of the (infinite) group of symmetries of the figure. Let r denote the reflection of the figure about the horizontal axis of symmetry through the **E**'s. Then $r^2 = I = r^0$, $rt = tr$, and the group of symmetries consists of all products of the form $r^i t^j$, where i is either 0 or 1 and j is an integer.

19. Let I denote the identity mapping, and let t denote a translation of the set of **T**'s one unit to the right. Then t^{-1} is a translation of the set of **T**'s one unit to the left. There is a vertical axis of symmetry through each copy of the letter **T** and a corresponding reflection of the figure about that vertical axis. Each of these reflections is its own inverse. The group of symmetries consists of this infinite collection of reflections (one for each copy of the letter **T**) together with the identity I and all the integral powers of the translation t .

23. Using the same notational convention as in Example 11 of Section 4.1, the elements of G are as follows:

$$e = (1), \quad \alpha = (1, 3)(2, 4), \quad \beta = (1, 4)(2, 3), \quad \Delta = (1, 2)(3, 4).$$

With this notation, we obtain the following multiplication table for G .

\circ	e	α	β	Δ
e	e	α	β	Δ
α	α	e	Δ	β
β	β	Δ	e	α
Δ	Δ	β	α	e

- 25.** Using the same notational convention as in Example 11 of Section 4.1, the elements of G are as follows:

$$\begin{array}{ll} e = (1) & \beta = (2, 5)(3, 4) \\ \alpha = (1, 2, 3, 4, 5) & \gamma = \alpha\beta = \beta\alpha^4 = (1, 2)(3, 5) \\ \alpha^2 = (1, 3, 5, 2, 4) & \Delta = \alpha^2\beta = \beta\alpha^3 = (1, 3)(4, 5) \\ \alpha^3 = (1, 4, 2, 5, 3) & \theta = \alpha^3\beta = \beta\alpha^2 = (1, 4)(2, 3) \\ \alpha^4 = (1, 5, 4, 3, 2) & \sigma = \alpha^4\beta = \beta\alpha = (1, 5)(2, 4). \end{array}$$

With this notation, we obtain the following multiplication table for G .

\circ	e	α	α^2	α^3	α^4	β	γ	Δ	θ	σ
e	e	α	α^2	α^3	α^4	β	γ	Δ	θ	σ
α	α	α^2	α^3	α^4	e	γ	Δ	θ	σ	β
α^2	α^2	α^3	α^4	e	α	Δ	θ	σ	β	γ
α^3	α^3	α^4	e	α	α^2	θ	σ	β	γ	Δ
α^4	α^4	e	α	α^2	α^3	σ	β	γ	Δ	θ
β	β	σ	θ	Δ	γ	e	α^4	α^3	α^2	α
γ	γ	β	σ	θ	Δ	α	e	α^4	α^3	α^2
Δ	Δ	γ	β	σ	θ	α^2	α	e	α^4	α^3
θ	θ	Δ	γ	β	σ	α^3	α^2	α	e	α^4
σ	σ	θ	Δ	γ	β	α^4	α^3	α^2	α	e

- 27.** 48

- 29.** Using the same notational convention as in Example 11 of Section 4.1, the elements of $G = \{e, \alpha, \beta, \Delta\}$ are $e = (1)$, $\alpha = (1, 3)(2, 4)$, $\beta = (1, 4)(2, 3)$, $\Delta = (1, 2)(3, 4)$. Let $\phi: G \rightarrow H$ be defined by

$$\begin{aligned} \phi(e) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & \phi(\alpha) &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \\ \phi(\beta) &= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, & \phi(\Delta) &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \end{aligned}$$

Exercises 4.4 Pages 220–223

True or False

1. true 2. true 3. false 4. true 5. true
 6. true 7. true 8. false
-

Exercises

- 1.** **a.** $eH = \beta H = H = \{e, \beta\}$; $\alpha H = \gamma H = \{\alpha, \gamma\}$;
 $\alpha^2 H = \Delta H = \{\alpha^2, \Delta\}$; $\alpha^3 H = \theta H = \{\alpha^3, \theta\}$;
 $G = H \cup \alpha H \cup \alpha^2 H \cup \alpha^3 H$
- b.** $He = H\beta = H = \{e, \beta\}$; $H\alpha = H\theta = \{\alpha, \theta\}$;
 $H\alpha^2 = H\Delta = \{\alpha^2, \Delta\}$; $H\alpha^3 = H\gamma = \{\alpha^3, \gamma\}$;
 $G = H \cup H\alpha \cup H\alpha^2 \cup H\alpha^3$
- 3.** **a.** $(1)H = (1, 2)H = H = \{(1), (1, 2)\}$; $(1, 2, 3)H = (1, 3)H = \{(1, 2, 3), (1, 3)\}$;
 $(1, 3, 2)H = (2, 3)H = \{(1, 3, 2), (2, 3)\}$;
 $G = H \cup (1, 2, 3)H \cup (1, 3, 2)H$
- b.** $H(1) = H(1, 2) = H = \{(1), (1, 2)\}$;
 $H(1, 2, 3) = H(2, 3) = \{(1, 2, 3), (2, 3)\}$;
 $H(1, 3, 2) = H(1, 3) = \{(1, 3, 2), (1, 3)\}$;
 $G = H \cup H(1, 2, 3) \cup H(1, 3, 2)$
- 5.** **a.** $I_3 H = P_4 H = H = \{I_3, P_4\}$; $P_1 H = P_3^2 H = \{P_1, P_3^2\}$; $P_2 H = P_3 H = \{P_2, P_3\}$;
 $G = H \cup P_1 H \cup P_2 H$
- b.** $HI_3 = HP_4 = H = \{I_3, P_4\}$; $HP_1 = HP_3 = \{P_1, P_3\}$; $HP_2 = HP_3^2 = \{P_2, P_3^2\}$;
 $G = H \cup HP_1 \cup HP_2$
- 13. a.** 12 **c.** 16
- 15.** Order 1: $\{(1)\}$
 Order 2: $\{(1), (1, 2)(3, 4)\}, \{(1), (1, 3)(2, 4)\}, \{(1), (1, 4)(2, 3)\}$
 Order 3: $\{(1), (1, 2, 3), (1, 3, 2)\}, \{(1), (1, 2, 4), (1, 4, 2)\},$
 $\{(1), (1, 4, 3), (1, 3, 4)\}, \{(1), (2, 3, 4), (2, 4, 3)\}$
 Order 4: $\{(1), (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}$
 Order 12: A_4 , as given in Example 8 of Section 4.1
- 17.** Order 1: $\{(1)\}$
 Order 2: $\{-1, 1\}$
 Order 4: $\{i, -1, -i, 1\}, \{j, -1, -j, 1\}, \{k, -1, -k, 1\}$
 Order 8: $\{1, -1, i, -i, j, -j, k, -k\}$

Exercises 4.5 Pages 227–229**True or False**

1. false 2. true 3. false 4. false 5. false 6. false 7. false
-

Exercises

1. a. no c. no e. no
 7. a. $\{e, \alpha^2\}$ b. $\{e, \Delta\}$
 9. Order 1: $\{(1)\}$
 Order 4: $\{(1), (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}$
 Order 12: A_4 , as given in Example 8 of Section 4.1
 11. Every subgroup of the nonabelian quaternion group is normal.
 13. $H = \{e, \Delta\}, K = \{e, \beta, \Delta, \alpha^2\}$
 23. $\{e, \alpha^2\}$
 27. For $H = \{(1), (1, 3)(2, 4)\}$, $N(H)$ is the octic group G since H is normal in G .
 35. a. $S_3 = \{(1), (1, 2), (1, 3), (2, 3), (1, 2, 3), (1, 3, 2)\}$
 c. A_4 , as given in Example 8 of Section 4.1

Exercises 4.6 Pages 236–239**True or False**

1. true 2. false 3. false 4. true 5. true
-

Exercises

1. $o(G/H) = 4; G/H = \{H, \alpha H, \beta H, \gamma H\}$, where

$$\begin{aligned} H &= \{e, \alpha^2\}, & \alpha H &= \{\alpha, \alpha^3\}, \\ \beta H &= \{\beta, \Delta\}, & \gamma H &= \{\gamma, \theta\}. \end{aligned}$$

.	H	αH	βH	γH
H	H	αH	βH	γH
αH	αH	H	γH	βH
βH	βH	γH	H	αH
γH	γH	βH	αH	H

3. $o(G/H) = 4$; $G/H = \{H, iH, jH, kH\}$, where

$$\begin{aligned} H &= \{1, -1\}, & iH &= \{i, -i\}, \\ jH &= \{j, -j\}, & kH &= \{k, -k\}. \end{aligned}$$

.	H	iH	jH	kH
H	H	iH	jH	kH
iH	iH	H	kH	jH
jH	jH	kH	H	iH
kH	kH	jH	iH	H

5. $o(G/H) = 3$; $G/H = \{H, (1, 2, 3)H, (1, 3, 2)H\}$, where

$$\begin{aligned} (1, 2, 3)H &= \{(1, 2, 3), (1, 3, 4), (2, 4, 3), (1, 4, 2)\}, \\ (1, 3, 2)H &= \{(1, 3, 2), (2, 3, 4), (1, 2, 4), (1, 4, 3)\}. \end{aligned}$$

.	H	$(1, 2, 3)H$	$(1, 3, 2)H$
H	H	$(1, 2, 3)H$	$(1, 3, 2)H$
$(1, 2, 3)H$	$(1, 2, 3)H$	$(1, 3, 2)H$	H
$(1, 3, 2)H$	$(1, 3, 2)H$	H	$(1, 2, 3)H$

7. $H = \{[1], [9]\}$, $\mathbf{U}_{20}/H = \{H, [3]H, [11]H, [13]H\}$, where

$$[3]H = \{[3], [7]\},$$

$$[11]H = \{[11], [19]\},$$

$$[13]H = \{[13], [17]\}.$$

.	H	$[3]H$	$[11]H$	$[13]H$
H	H	$[3]H$	$[11]H$	$[13]H$
$[3]H$	$[3]H$	H	$[13]H$	$[11]H$
$[11]H$	$[11]H$	$[13]H$	H	$[3]H$
$[13]H$	$[13]H$	$[11]H$	$[3]H$	H

8. a. $o(H_1) = 4n$ when $o(H_2) = 3n$, $n = 1, 2, 3, 6$

9. The normal subgroups of the octic group G are $H_1 = \{e\}$, $H_2 = \{e, \alpha^2\}$, $H_3 = \{e, \alpha, \alpha^2, \alpha^3\}$, $H_4 = \{e, \beta, \Delta, \alpha^2\}$, $H_5 = \{e, \gamma, \theta, \alpha^2\}$, and $H_6 = G$. We consider the possible quotient groups.

- (1) G/H_1 is isomorphic to G .
- (2) $G/H_2 = \{H_2, \alpha H_2, \beta H_2, \gamma H_2\}$ is isomorphic to the Klein four group. (See Exercise 3 of Section 4.2.)
- (3) Each of G/H_3 , G/H_4 , and G/H_5 is a cyclic group of order 2.
- (4) $G/G = \{G\}$ is a group of order 1.

Thus the homomorphic images of the octic group G are G itself, a Klein four group, a cyclic group of order 2, and a group with only the identity element.

- 11.** The normal subgroups of the quaternion group G are $H_1 = \{1\}$, $H_2 = \{-1, 1\}$, $H_3 = \{i, -1, -i, 1\}$, $H_4 = \{j, -1, -j, 1\}$, $H_5 = \{k, -1, -k, 1\}$, and $H_6 = G$. We consider the quotient groups.

- (1) G/H_1 is isomorphic to G .
- (2) $G/H_2 = \{H_2, iH_2, jH_2, kH_2\}$ is isomorphic to the Klein four group. (See Exercise 3 of Section 4.2.)
- (3) Each of G/H_3 , G/H_4 , and G/H_5 is a cyclic group of order 2.
- (4) $G/G = \{G\}$ is a group of order 1.

Thus the homomorphic images of the quaternion group G are G itself, a Klein four group, a cyclic group of order 2, and a group with only the identity element.

- 13. a.** The left cosets of $H = \{(1), (1, 2)\}$ in $G = S_3$ are given by

$$(1)H = (1, 2)H = \{(1), (1, 2)\}$$

$$(1, 3)H = (1, 2, 3)H = \{(1, 3), (1, 2, 3)\}$$

$$(2, 3)H = (1, 3, 2)H = \{(2, 3), (1, 3, 2)\}.$$

The rule $aHbH = abH$ leads to

$$(1, 3)H(2, 3)H = (1, 3)(2, 3)H = (1, 3, 2)H$$

and also to

$$(1, 2, 3)H(1, 3, 2)H = (1, 2, 3)(1, 3, 2)H = (1)H.$$

We have $(1, 3)H = (1, 2, 3)H$ and $(2, 3)H = (1, 3, 2)H$, but

$$(1, 3)H(2, 3)H \neq (1, 2, 3)H(1, 3, 2)H.$$

Thus the rule $aHbH = abH$ does not define a binary operation on the left cosets of H in G . (That is, the result is not well-defined.)

- c.** The left cosets of $H = \{(1), (2, 3)\}$ in $G = S_3$ are given by

$$(1)H = (2, 3)H = \{(1), (2, 3)\}$$

$$(1, 2)H = (1, 2, 3)H = \{(1, 2), (1, 2, 3)\}$$

$$(1, 3)H = (1, 3, 2)H = \{(1, 3), (1, 3, 2)\}.$$

The rule $aHbH = abH$ leads to

$$(1, 2)H(1, 3)H = (1, 2)(1, 3)H = (1, 3, 2)H$$

and also to

$$(1, 2, 3)H(1, 3, 2)H = (1, 2, 3)(1, 3, 2)H = (1)H.$$

We have $(1, 2)H = (1, 2, 3)H$ and $(1, 3)H = (1, 3, 2)H$, but

$$(1, 2)H(1, 3)H \neq (1, 2, 3)H(1, 3, 2)H.$$

Thus the rule $aHbH = abH$ does not define a binary operation on the left cosets of H in G . (That is, the result is not well-defined.)

15. a. $K = \{I_2, M_2, M_3\}$ b. $K = \{I_2, M_2, M_3\}$, $M_1K = \{M_1, M_4, M_5\}$
c. $\theta(K) = 1$, $\theta(M_1K) = -1$
17. a. $K = \{[1], [11]\}$
b. $K = \{[1], [11]\}$, $[3]K = \{[3], [13]\}$, $[7]K = \{[7], [17]\}$, $[9]K = \{[9], [19]\}$
c. $\theta(K) = e$, $\theta([3]K) = a$, $\theta([7]K) = a^3$, $\theta([9]K) = a^2$
25. S_3 is not cyclic. However $H = \{(1), (1, 2, 3), (1, 3, 2)\}$ is normal, and $S_3/H = \{H, (1, 2)H\}$ is cyclic.
27. a. Let $G = \{a, a^2, a^3, a^4, a^5, a^6, a^7, a^8 = e\}$ be a cyclic group of order 8. The subgroup $H = \{a^2, a^4, a^6, a^8 = e\}$ of G is a cyclic group of order 4, and the mapping $\phi: G \rightarrow H$ defined by $\phi(x) = x^2$ is a homomorphism, since

$$\begin{aligned}\phi(xy) &= (xy)^2 \\ &= x^2y^2 \text{ since } G \text{ is abelian} \\ &= \phi(x)\phi(y).\end{aligned}$$

The mapping ϕ is an epimorphism, since

$$\begin{aligned}\phi(G) &= \{\phi(a), \phi(a^2), \phi(a^3), \phi(a^4), \phi(a^5), \phi(a^6), \phi(a^7), \phi(e)\} \\ &= \{a^2, a^4, a^6, a^8 = e, a^{10} = a^2, a^{12} = a^4, a^{14} = a^6, e\} \\ &= \{a^2, a^4, a^6, a^8 = e\} \\ &= H.\end{aligned}$$

Thus G has H as a homomorphic image.

Exercises 4.7 Pages 244–246

True or False

1. false 2. true

Exercises

1. $H_1 + H_2 = H_2$ The sum is not direct.
7. $\mathbf{Z}_{20} = \langle [4] \rangle \oplus \langle [5] \rangle$
17. a. 2 c. 6

- 18. a.** $\{([0], [0])\}$,
 $\{([0], [0]), ([0], [2])\}$,
 $\{([0], [0]), ([1], [0])\}$,
 $\{([0], [0]), ([1], [2])\}$,
 $\{([0], [0]), ([0], [1]), ([0], [2]), ([0], [3])\}$,
 $\{([0], [0]), ([0], [2]), ([1], [0]), ([1], [2])\}$,
 $\{([0], [0]), ([0], [2]), ([1], [1]), ([1], [3])\}$,
 $\mathbf{Z}_2 \oplus \mathbf{Z}_4$

- 19. a.** $\mathbf{Z}_{15} = \langle [5] \rangle \oplus \langle [3] \rangle$, where $\langle [5] \rangle = \{[5], [10], [0]\}$ is a cyclic group of order 3, and $\langle [3] \rangle = \{[3], [6], [9], [12], [0]\}$ is a cyclic group of order 5. From this, it is intuitively clear that \mathbf{Z}_{15} is isomorphic to $\mathbf{Z}_3 \oplus \mathbf{Z}_5$. The idea can be formalized as follows. For each $a \in \mathbf{Z}$, let $[a]_{15}$, $[a]_3$, and $[a]_5$ denote the congruence class of a modulo 15, 3, and 5, respectively. Any $[a]_{15}$ and $[b]_{15}$ in \mathbf{Z}_{15} can be written as

$$[a]_{15} = r[5]_{15} + s[3]_{15} \quad \text{and} \quad [b]_{15} = p[5]_{15} + q[3]_{15}$$

with r, s, p , and q integers. Since

$$\begin{aligned} [a]_{15} = [b]_{15} &\Leftrightarrow r[5]_{15} + s[3]_{15} = p[5]_{15} + q[3]_{15} \\ &\Leftrightarrow (r - p)[5]_{15} = (q - s)[3]_{15} \\ &\Leftrightarrow (r - p)[5]_{15} = [0]_{15} = (q - s)[3]_{15} \\ &\Leftrightarrow r - p \equiv 0 \pmod{3} \quad \text{and} \quad q - s \equiv 0 \pmod{5} \\ &\Leftrightarrow [r]_3 = [p]_3 \quad \text{and} \quad [q]_5 = [s]_5, \end{aligned}$$

the rule

$$\phi([a]_{15}) = ([r]_3, [s]_5)$$

defines a one-to-one mapping from \mathbf{Z}_{15} to the external direct sum $\mathbf{Z}_3 \oplus \mathbf{Z}_5$. ϕ is clearly onto, and ϕ is a homomorphism, since

$$\begin{aligned} \phi([a]_{15} + [b]_{15}) &= \phi((r + p)[5]_{15} + (s + q)[3]_{15}) \\ &= ([r + p]_3, [s + q]_5) \\ &= ([r]_3 + [p]_3, [s]_5 + [q]_5) \\ &= ([r]_3, [s]_5) + ([p]_3, [q]_5) \\ &= \phi([a]_{15}) + \phi([b]_{15}). \end{aligned}$$

Thus ϕ is an isomorphism from \mathbf{Z}_{15} to $\mathbf{Z}_3 \oplus \mathbf{Z}_5$.

Exercises 4.8 Pages 254–255**True or False**

1. true 2. false 3. false 4. true 5. false 6. false
-

Exercises

1. The cyclic group $C_9 = \langle a \rangle$ of order 9 is a p -group with $p = 3$.
3. a. $\langle (1, 2, 3) \rangle, \langle (1, 2, 4) \rangle, \langle (1, 3, 4) \rangle, \langle (2, 3, 4) \rangle$
5. a. $\mathbf{Z}_{10} = \langle [5] \rangle \oplus \langle [2] \rangle$
= $\{[5], [0]\} \oplus \{[2], [4], [6], [8], [0]\}$
= $C_2 \oplus C_5$
- c. $\mathbf{Z}_{12} = \langle [3] \rangle \oplus \langle [4] \rangle$
= $\{[3], [6], [9], [0]\} \oplus \{[4], [8], [0]\}$
= $C_4 \oplus C_3$
6. a. Any abelian group of order 6 is isomorphic to $C_3 \oplus C_2$, where C_n is a cyclic group of order n .
- c. Any abelian group of order 12 is isomorphic to either $C_4 \oplus C_3$ or $C_2 \oplus C_2 \oplus C_3$.
- e. Any abelian group of order 36 is isomorphic to one of the direct sums $C_4 \oplus C_9$, $C_2 \oplus C_2 \oplus C_9$, $C_4 \oplus C_3 \oplus C_3$, $C_2 \oplus C_2 \oplus C_3 \oplus C_3$.
9. a. none
15. b. There are 24 distinct elements of G that have order 6.

Exercises 5.1 Pages 265–269**True or False**

1. false 2. true 3. true 4. false 5. false
6. false 7. false 8. false 9. false
-

Exercises

2. a. ring
- c. Not a ring. The set is not closed with respect to multiplication. For example, $\sqrt[3]{5}$ is in the set, but the product $\sqrt[3]{5} \cdot \sqrt[3]{5} = \sqrt[3]{25}$ is not in the set.
- e. Not a ring. The set of positive real numbers does not contain an additive identity.
- g. ring

3.

+	\emptyset	A	B	U
\emptyset	\emptyset	A	B	U
A	A	\emptyset	U	B
B	B	U	\emptyset	A
U	U	B	A	\emptyset

.	\emptyset	A	B	U
\emptyset	\emptyset	\emptyset	\emptyset	\emptyset
A	\emptyset	A	\emptyset	A
B	\emptyset	\emptyset	B	B
U	\emptyset	A	B	U

5. The set $\mathcal{P}(A)$ is not a ring with respect to the operations of addition and multiplication as defined, since the set does not contain additive inverse elements.
7. a. [2], [3], [4]
 c. [2], [4], [5], [6], [8]
 e. [2], [4], [6], [7], [8], [10], [12]
8. a. $[1]^{-1} = [1]$, $[5]^{-1} = [5]$
 c. $[1]^{-1} = [1]$, $[3]^{-1} = [11]$, $[5]^{-1} = [13]$, $[7]^{-1} = [7]$, $[9]^{-1} = [9]$,
 $[11]^{-1} = [3]$, $[13]^{-1} = [5]$, $[15]^{-1} = [15]$
 e. $[1]^{-1} = [1]$, $[3]^{-1} = [5]$, $[5]^{-1} = [3]$, $[9]^{-1} = [11]$, $[11]^{-1} = [9]$,
 $[13]^{-1} = [13]$
19. In the ring $M_2(\mathbb{Z})$, let $a = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $b = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. Then $ab = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ but
 $ba = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.
31. a. yes
 b. The set S is a commutative ring, and it contains the unity [10].
 c. yes
 d. yes, [6] and [12]
 e. [2], [4], [8], [10], [14], [16]

33.

.	a	b	c	d
a	a	a	a	a
b	a	c	a	c
c	a	a	a	a
d	a	c	a	c

39. a. S is a commutative subring of $M_2(\mathbb{Z})$ with unity $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. All $\begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix}$ for $x = \pm 1$ have multiplicative inverses.
 c. S is a noncommutative subring of $M_2(\mathbb{Z})$ without unity.

- e.** S is a commutative subring of $M_2(\mathbb{Z})$ without unity.
g. S is not a subring of $M_2(\mathbb{Z})$, since it is not closed with respect to multiplication.

49. For notational convenience, we let a represent $[a]$.

b. $S_1 \oplus S_2 = \{(0, 0), (0, 3), (2, 0), (2, 3)\}$

+	(0, 0)	(0, 3)	(2, 0)	(2, 3)		.	(0, 0)	(0, 3)	(2, 0)	(2, 3)
(0, 0)	(0, 0)	(0, 3)	(2, 0)	(2, 3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	
(0, 3)	(0, 3)	(0, 0)	(2, 3)	(2, 0)	(0, 3)	(0, 0)	(0, 3)	(0, 0)	(0, 3)	
(2, 0)	(2, 0)	(2, 3)	(0, 0)	(0, 3)	(2, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	
(2, 3)	(2, 3)	(2, 0)	(0, 3)	(0, 0)	(2, 3)	(0, 0)	(0, 3)	(0, 0)	(0, 3)	

Exercises 5.2 Pages 273–276

True or False

1. true 2. true 3. false 4. true

Exercises

1. a. The set of real numbers of the form $m + n\sqrt{2}$, where m and n are integers, is an integral domain. It is not a field, since not every element (for example, $2 + 0\sqrt{2}$) has a multiplicative inverse.

c. The set of real numbers of the form $a + b\sqrt[3]{2}$, where a and b are rational numbers, is neither an integral domain nor a field, since it is not a ring. The set is not closed with respect to multiplication. For example, $\sqrt[3]{2} \cdot \sqrt[3]{2} = \sqrt[3]{4}$ is not in the set.

e. The set of all complex numbers of the form $m + ni$, where $m \in \mathbf{Z}$ and $n \in \mathbf{Z}$, is an integral domain. It is not a field, since not every element (for example, $2 + 0i$) has a multiplicative inverse.

g. The set of all complex numbers of the form $a + bi$, where a and b are rational numbers, is both an integral domain and a field.

3. a. The set S is not an integral domain, since the elements [6] and [12] are zero divisors.

b. The set S is not a field, since [6] and [12] do not have multiplicative inverses.

5. $\mathcal{P}(U)$ is not a field. $A = \{a\}$ and $B = \{b\}$ do not have multiplicative inverses.

7. The ring W is commutative, since if (x, y) and (z, w) are elements of W , we have

$$\begin{aligned}(x, y) \cdot (z, w) &= (xz - yw, xw + yz) \\&= (zx - wy, zy + wx) \\&= (z, w) \cdot (x, y).\end{aligned}$$

The element $(1, 0)$ in W is the unity element, since for (x, y) in W we have

$$\begin{aligned}(x, y) \cdot (1, 0) &= (1, 0) \cdot (x, y) \\ &= (1x - 0y, 1y + 0x) \\ &= (x, y).\end{aligned}$$

- 9.** **a.** S is a commutative ring. **b.** S has the unity element $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$.
c. S is an integral domain. **d.** S is a field.
- 11.** **a.** R is a commutative ring. **b.** R has the unity element $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
c. R is an integral domain. **d.** R is a field.
- 13.** **a.** yes **b.** yes, [1] **c.** No, since $(2+i)(2+4i) \equiv 0 \pmod{5}$. **d.** no
- 15.** The set of even integers is a commutative ring with no zero divisors but not an integral domain, since it has no unity.
- 19.** **a.** Consider the ring \mathbf{Z}_{10} . The elements [1] and [3] are not zero divisors, but the sum $[1] + [3] = [4]$ is a zero divisor.
- 20.** **a.** [173] **c.** [27]

Exercises 5.3 Pages 282–284

True or False

- 1.** true **2.** false **3.** false **4.** true **5.** true

Exercises

- 9.** Define $\phi: W \rightarrow R$ by

$$\phi((x, y)) = \begin{bmatrix} x & -y \\ y & x \end{bmatrix}.$$

The mapping ϕ is clearly a one-to-one correspondence from W to R .

$$\begin{aligned}\phi((x, y) + (z, w)) &= \phi((x+z, y+w)) \\ &= \begin{bmatrix} x+z & -y-w \\ y+w & x+z \end{bmatrix} = \begin{bmatrix} x & -y \\ y & x \end{bmatrix} + \begin{bmatrix} z & -w \\ w & z \end{bmatrix} \\ &= \phi((x, y)) + \phi((z, w))\end{aligned}$$

$$\begin{aligned}\phi((x, y) \cdot (z, w)) &= \phi((xz - yw, xw + yz)) \\ &= \begin{bmatrix} xz - yw & -xw - yz \\ xw + yz & xz - yw \end{bmatrix} = \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \cdot \begin{bmatrix} z & -w \\ w & z \end{bmatrix} \\ &= \phi((x, y)) \cdot \phi((z, w))\end{aligned}$$

Thus, ϕ is an isomorphism.

- 11. a.** For notational convenience in this solution, we write 0 for [0], 1 for [1], and 2 for [2] in \mathbf{Z}_3 . Then

$$S = \{(0, 1), (0, 2), (1, 1), (1, 2), (2, 1), (2, 2)\}.$$

Since $(0, 1) \sim (0, 2)$, $(1, 1) \sim (2, 2)$, and $(1, 2) \sim (2, 1)$ in S , the distinct elements of Q are $[0, 1]$, $[1, 1]$, and $[2, 1]$.

- b.** Define $\phi: D \rightarrow Q$ by

$$\phi(0) = [0, 1]$$

$$\phi(1) = [1, 1]$$

$$\phi(2) = [2, 1].$$

- 15.** The set of all quotients for D is the set Q of all equivalence classes $[m + ni, r + si]$, where $m + ni \in D$ and $r + si \in D$ with not both r and s equal to 0. To show that Q is isomorphic to the set C of all complex numbers of the form $a + bi$, where a and b are rational numbers, we define $\phi: Q \rightarrow C$ by

$$\phi([m + ni, r + si]) = \frac{m + ni}{r + si}.$$

This rule does define a mapping from Q into C , since for $[m + ni, r + si] \in Q$ we can write

$$\frac{m + ni}{r + si} = \frac{mr + ns}{r^2 + s^2} + \frac{nr - ms}{r^2 + s^2} i,$$

which is an element in C .

To show that ϕ is onto, let $a + bi$ be an arbitrary element in C . Since a and b are both rational numbers, there exist integers p, q, t , and u such that

$$a = \frac{p}{q} \quad \text{and} \quad b = \frac{t}{u}.$$

Then the element $[pu + qt, qu + 0i]$ is in Q , and

$$\begin{aligned} \phi([pu + qt, qu + 0i]) &= \frac{pu + qt}{qu + 0i} \\ &= \frac{p}{q} + \frac{t}{u} i \\ &= a + bi. \end{aligned}$$

To show that ϕ is one-to-one, let $[m + ni, r + si]$ and $[x + yi, z + wi]$ be elements of Q such that

$$\phi([m + ni, r + si]) = \phi([x + yi, z + wi]).$$

Then

$$\frac{m + ni}{r + si} = \frac{x + yi}{z + wi},$$

and this implies that

$$(m + ni)(z + wi) = (r + si)(x + yi).$$

By the definition of equality in Q , we have

$$[m + ni, r + si] = [x + yi, z + wi],$$

and therefore ϕ is one-to-one. Since

$$\begin{aligned} \phi([m + ni, r + si] + [x + yi, z + wi]) \\ &= \phi([(m + ni)(z + wi) + (r + si)(x + yi), (r + si)(z + wi)]) \\ &= \frac{(m + ni)(z + wi) + (r + si)(x + yi)}{(r + si)(z + wi)} \\ &= \frac{m + ni}{r + si} + \frac{x + yi}{z + wi} \\ &= \phi([m + ni, r + si]) + \phi([x + yi, z + wi]) \end{aligned}$$

and

$$\begin{aligned} \phi([m + ni, r + si] \cdot [x + yi, z + wi]) \\ &= \phi([(m + ni)(x + yi), (r + si)(z + wi)]) \\ &= \frac{(m + ni)(x + yi)}{(r + si)(z + wi)} \\ &= \frac{m + ni}{r + si} \cdot \frac{x + yi}{z + wi} \\ &= \phi([m + ni, r + si]) \cdot \phi([x + yi, z + wi]), \end{aligned}$$

ϕ is an isomorphism from Q to C .

19. $\frac{m}{2^n}$ for $m, n \in \mathbf{Z}$.

Exercises 5.4 Pages 289–291

True or False

-
1. false 2. true 3. true 4. true 5. false
-

Exercises 6.1 Pages 300–302

True or False

-
1. true 2. false 3. true 4. true 5. true
 6. true 7. true 8. false
-

Exercises

3. a. \mathbf{Q} is a subring of \mathbf{R} , and $1 \in \mathbf{Q}$, $\sqrt{2} \in \mathbf{R}$, but $\sqrt{2} \cdot 1 \notin \mathbf{Q}$. Thus \mathbf{Q} is not an ideal of \mathbf{R} .
6. a. Let $I_1 = (2)$ and $I_2 = (3)$. Then 2 and 3 are in $I_1 \cup I_2$, but the sum $2 + 3 = 5$ is not in $I_1 \cup I_2$. Hence $I_1 \cup I_2$ is not an ideal of \mathbf{Z} .
19. a. $\{[0]\}, \mathbf{Z}_7$
 c. $\{[0]\}$
 $([6]) = \{[0], [6]\}$
 $([4]) = \{[0], [4], [8]\}$
 $([3]) = \{[0], [3], [6], [9]\}$
 $([2]) = \{[0], [2], [4], [6], [8], [10]\}$
 \mathbf{Z}_{12}
 e. $\{[0]\}$
 $([10]) = \{[0], [10]\}$
 $([5]) = \{[0], [5], [10], [15]\}$
 $([4]) = \{[0], [4], [8], [12], [16]\}$
 $([2]) = \{[0], [2], [4], [6], [8], [10], [12], [14], [16], [18]\}$
 \mathbf{Z}_{20}
23. The set U is not an ideal of S . $X = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ is in U , and $R = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$ is in S , but $XR = \begin{bmatrix} 1 & 8 \\ 0 & 3 \end{bmatrix}$ is not in U .
25. b. Since $1 + 2i \in I$, $1 + i \in G$, but $(1 + 2i)(1 + i) = -1 + 3i \notin I$, then I is not an ideal of G .
26. b. \mathbf{E} is a commutative ring, and $2 \in \mathbf{E}$ but $2 \notin (2) = \{2n \mid n \in \mathbf{E}\}$.

Exercises 6.2 Pages 309–313

True or False

1. true 2. false 3. true 4. true 5. true

Exercises

7. b. $\ker \theta = \left\{ \begin{bmatrix} 0 & 0 \\ y & 0 \end{bmatrix} \middle| y \in \mathbf{Z} \right\}$, $\phi \left(\begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} + \ker \theta \right) = \theta \left(\begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} \right) = x$
9. b. no
13. $\ker \theta = \left\{ \begin{bmatrix} 2m & 2n \\ 2p & 2q \end{bmatrix} \middle| m, n, p, q \in \mathbf{Z} \right\}$

- 15.** **a.** θ does not preserve addition.
b. θ does not preserve multiplication.
- 19.** The mapping $\phi: R \rightarrow \mathbf{Z}_3$ given by

$$\phi(a) = [0], \quad \phi(b) = [2], \quad \phi(c) = [1]$$

is an isomorphism.

- 22.** **a.** θ does not preserve addition; θ preserves multiplication; θ is not a homomorphism.
c. θ preserves addition; θ does not preserve multiplication; θ is not a homomorphism.
e. θ does not preserve addition; θ preserves multiplication; θ is not a homomorphism.
- 23.** **a.** The ideals of \mathbf{Z}_6 are $I_1 = \{[0]\}$, $I_2 = \{[0], [3]\}$, $I_3 = \{[0], [2], [4]\}$, and $I_4 = \mathbf{Z}_6$. We consider the quotient rings:

- (1) \mathbf{Z}_6/I_1 is isomorphic to \mathbf{Z}_6 .
- (2) $\mathbf{Z}_6/I_2 = \{I_2, [1] + I_2, [2] + I_2\}$ is isomorphic to \mathbf{Z}_3 .
- (3) $\mathbf{Z}_6/I_3 = \{I_3, [1] + I_3\}$ is isomorphic to \mathbf{Z}_2 .
- (4) $\mathbf{Z}_6/\mathbf{Z}_6 = \{\mathbf{Z}_6\}$ is a ring with only the zero element.

Thus, the homomorphic images of \mathbf{Z}_6 are (isomorphic to) \mathbf{Z}_6 , \mathbf{Z}_3 , \mathbf{Z}_2 , and $\{0\}$.

- c.** The ideals of \mathbf{Z}_{12} are $I_1 = \{[0]\}$, $I_2 = \{[0], [6]\}$, $I_3 = \{[0], [4], [8]\}$, $I_4 = \{[0], [3], [6], [9]\}$, $I_5 = \{[0], [2], [4], [6], [8], [10]\}$, and $I_6 = \mathbf{Z}_{12}$. The quotient rings are as follows:

- (1) \mathbf{Z}_{12}/I_1 is isomorphic to \mathbf{Z}_{12} .
- (2) $\mathbf{Z}_{12}/I_2 = \{I_2, [1] + I_2, [2] + I_2, [3] + I_2, [4] + I_2, [5] + I_2\}$ is isomorphic to \mathbf{Z}_6 .
- (3) $\mathbf{Z}_{12}/I_3 = \{I_3, [1] + I_3, [2] + I_3, [3] + I_3\}$ is isomorphic to \mathbf{Z}_4 .
- (4) $\mathbf{Z}_{12}/I_4 = \{I_4, [1] + I_4, [2] + I_4\}$ is isomorphic to \mathbf{Z}_3 .
- (5) $\mathbf{Z}_{12}/I_5 = \{I_5, [1] + I_5\}$ is isomorphic to \mathbf{Z}_2 .
- (6) $\mathbf{Z}_{12}/\mathbf{Z}_{12} = \{\mathbf{Z}_{12}\}$ is a ring with only the zero element.

The homomorphic images of \mathbf{Z}_{12} are (isomorphic to) \mathbf{Z}_{12} , \mathbf{Z}_6 , \mathbf{Z}_4 , \mathbf{Z}_3 , \mathbf{Z}_2 , and $\{0\}$.

- e.** The ideals of \mathbf{Z}_8 are $I_1 = \{[0]\}$, $I_2 = \{[0], [4]\}$, $I_3 = \{[0], [2], [4], [6]\}$, and $I_4 = \mathbf{Z}_8$. The quotient rings are as follows:

- (1) \mathbf{Z}_8/I_1 is isomorphic to \mathbf{Z}_8 .
- (2) $\mathbf{Z}_8/I_2 = \{I_2, [1] + I_2, [2] + I_2, [3] + I_2\}$ is isomorphic to \mathbf{Z}_4 .
- (3) $\mathbf{Z}_8/I_3 = \{I_3, [1] + I_3\}$ is isomorphic to \mathbf{Z}_2 .
- (4) $\mathbf{Z}_8/\mathbf{Z}_8 = \{\mathbf{Z}_8\}$ is a ring with only the zero element.

The homomorphic images of \mathbf{Z}_8 are (isomorphic to) \mathbf{Z}_8 , \mathbf{Z}_4 , \mathbf{Z}_2 , and $\{0\}$.

Exercises 6.3 Pages 317–319

True or False

1. false 2. false 3. true 4. true 5. false

Exercises

1. a. 0

c. 0

e. 2

2. a. 2

c. 6

e. 12

9. b. Exercise 3 assures us that e , a , and b all have additive order 2. The other entries in the table can be determined by using the fact that D forms a group with respect to addition. For example, $e + a = a$ would imply $e = 0$, so $e + a = b$ must be true.

+	0	e	a	b
0	0	e	a	b
e	e	0	b	a
a	a	b	0	e
b	b	a	e	0

11. a. Let $R = \{[0], [5], [10], [15]\} \subseteq \mathbf{Z}_{20}$. Take $x = y = [5]$. Then

$$([5] + [5])^4 = [10]^4 = [0]$$

and

$$[5]^4 + [5]^4 = [5] + [5] = [10].$$

Thus

$$([5] + [5])^4 \neq [5]^4 + [5]^4.$$

Exercises 6.4 Pages 322–323

True or False

1. true 2. false

Exercises

5. $R/I = \{I, 1 + I, \sqrt{2} + I, 1 + \sqrt{2} + I\}$

7. $E/I = \{I, 2 + I, 4 + I\}$

9. $\{[0], [3], [6], [9]\}$ and $\{[0], [2], [4], [6], [8], [10]\}$

21. $\{[0], [3], [6], [9]\}$ and $\{[0], [2], [4], [6], [8], [10]\}$

23. $I_1 = \{[0], [3], [6], [9]\}$ and $I_2 = \{[0], [2], [4], [6], [8], [10]\}$ are prime ideals of \mathbf{Z}_{12} , but $I_1 \cap I_2 = \{[0], [6]\}$ is not a prime ideal of \mathbf{Z}_{12} , since $[2][3] \in I_1 \cap I_2$, but $[2] \notin I_1 \cap I_2$ and $[3] \notin I_1 \cap I_2$.

Exercises 7.1 Pages 332–333**True or False**

1. true 2. false 3. true 4. false 5. false
 6. true 7. false 8. true 9. false
-

Exercises

1. $0.\bar{5}$ 3. $0.\overline{987654320}$ 5. $3.\overline{142857}$
 7. $31/9$ 9. $4/33$ 11. $83/33$
20. a. $a = \sqrt{2}$ and $b = -\sqrt{2}$ are irrational, but $a + b = 0$ is rational.
21. a. An element ν of F is a *lower bound* of S if $\nu \leq x$ for all $x \in S$. An element ν of F is a *greatest lower bound* of S if these conditions are satisfied:
 (1) ν is a lower bound of S .
 (2) If $b \in F$ is a lower bound of S , then $b \leq \nu$.

Exercises 7.2 Pages 340–342**True or False**

1. false 2. true 3. true 4. true 5. false 6. true 7. false
-

Exercises

1. $10 + 11i$
 3. $-i$
 5. $2 - 11i$
 7. $\frac{2}{5} + \frac{1}{5}i$
 9. $\frac{11}{50} + \frac{1}{25}i$
 11. $\frac{21}{29} + \frac{20}{29}i$
13. a. $3i, -3i$
 c. $5i, -5i$
 e. $\sqrt{13}i, -\sqrt{13}i$
21. b. i. $-2 + i, 2 - i$ iii. $3 + 2i, -3 - 2i$

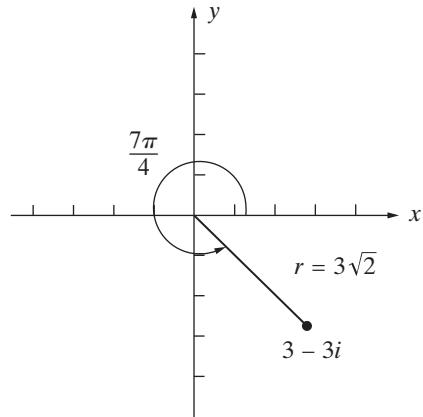
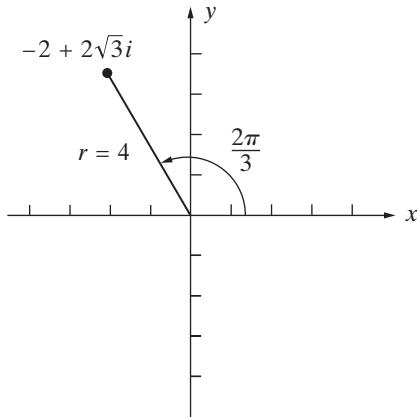
Exercises 7.3 Pages 349–352

True or False

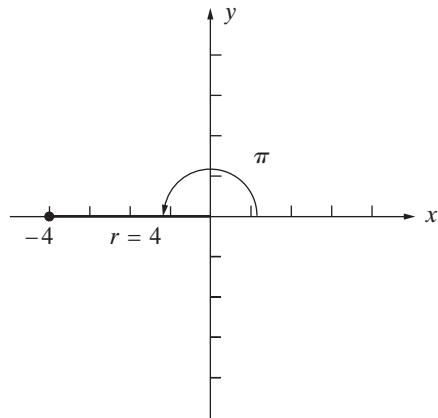
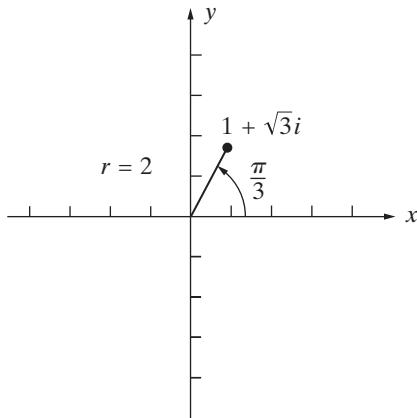
1. false 2. true 3. false 4. true

Exercises

1. a. $-2 + 2\sqrt{3}i = 4(\cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3})$ c. $3 - 3i = 3\sqrt{2}(\cos \frac{7\pi}{4} + i \sin \frac{7\pi}{4})$



e. $1 + \sqrt{3}i = 2(\cos \frac{\pi}{3} + i \sin \frac{\pi}{3})$ g. $-4 = 4(\cos \pi + i \sin \pi)$

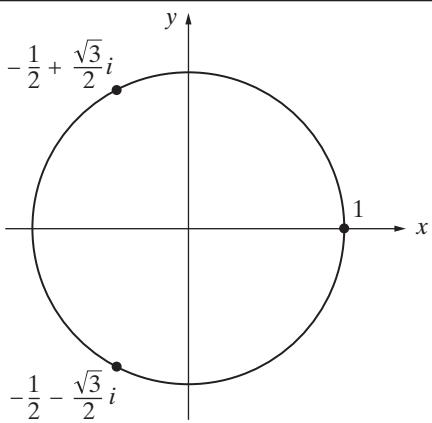


2. a. $4(\cos \frac{3\pi}{4} + i \sin \frac{3\pi}{4}) = -2\sqrt{2} + 2\sqrt{2}i$

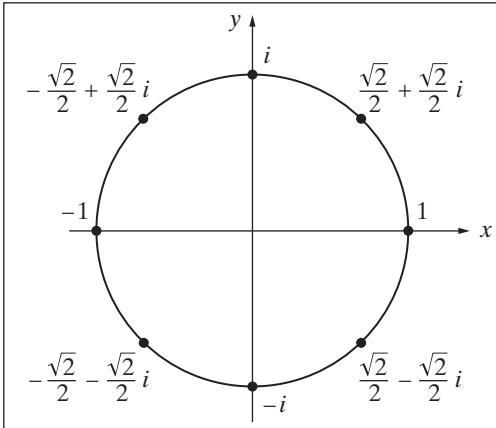
c. $6(\cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3}) = -3 + 3\sqrt{3}i$

3. a. $-64\sqrt{3} - 64i$ c. $512 + 512\sqrt{3}i$ e. 1 g. $-128 - 128\sqrt{3}i$

6. a.



c.



7. a. $\cos \frac{\pi}{18} + i \sin \frac{\pi}{18}$, $\cos \frac{13\pi}{18} + i \sin \frac{13\pi}{18}$, $\cos \frac{25\pi}{18} + i \sin \frac{25\pi}{18}$
 c. $\cos \frac{5\pi}{24} + i \sin \frac{5\pi}{24}$, $\cos \frac{17\pi}{24} + i \sin \frac{17\pi}{24}$, $\cos \frac{29\pi}{24} + i \sin \frac{29\pi}{24}$, $\cos \frac{41\pi}{24} + i \sin \frac{41\pi}{24}$
 e. $2(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4})$, $2(\cos \frac{13\pi}{20} + i \sin \frac{13\pi}{20})$, $2(\cos \frac{21\pi}{20} + i \sin \frac{21\pi}{20})$,
 $2(\cos \frac{29\pi}{20} + i \sin \frac{29\pi}{20})$, $2(\cos \frac{37\pi}{20} + i \sin \frac{37\pi}{20})$
8. a. $\frac{3}{2} + \frac{3\sqrt{3}}{2}i$, $-3, \frac{3}{2} - \frac{3\sqrt{3}}{2}i$ c. $\frac{\sqrt{3}}{2} + \frac{1}{2}i, -\frac{\sqrt{3}}{2} + \frac{1}{2}i, -i$
 e. $\frac{\sqrt{3}}{2} + \frac{1}{2}i, -\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{\sqrt{3}}{2} - \frac{1}{2}i, \frac{1}{2} - \frac{\sqrt{3}}{2}i$
 g. $\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{\sqrt{3}}{2} + \frac{1}{2}i, -\frac{1}{2} - \frac{\sqrt{3}}{2}i, \frac{\sqrt{3}}{2} - \frac{1}{2}i$
11. a. $\langle \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3} \rangle = \{\cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3}, \cos \frac{4\pi}{3} + i \sin \frac{4\pi}{3}, \cos 0 + i \sin 0\}$
 b. $o\langle a \rangle = 3$
 c. $\cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3}, \cos \frac{4\pi}{3} + i \sin \frac{4\pi}{3}$
13. a. $\langle \cos \frac{5\pi}{3} + i \sin \frac{5\pi}{3} \rangle = \{\cos \frac{5\pi}{3} + i \sin \frac{5\pi}{3}, \cos \frac{4\pi}{3} + i \sin \frac{4\pi}{3}, \cos \pi + i \sin \pi,$
 $\cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3}, \cos \frac{\pi}{3} + i \sin \frac{\pi}{3}, \cos 0 + i \sin 0\}$

- b. $o\langle a \rangle = 6$
- c. $\cos \frac{5\pi}{3} + i \sin \frac{5\pi}{3}, \cos \frac{\pi}{3} + i \sin \frac{\pi}{3}$
17. a. $\cos \frac{\pi}{3} + i \sin \frac{\pi}{3} = \frac{1}{2} + \frac{\sqrt{3}}{2}i, \cos \frac{5\pi}{3} + i \sin \frac{5\pi}{3} = \frac{1}{2} - \frac{\sqrt{3}}{2}i$

Exercises 8.1 Pages 364–366

True or False

1. false 2. false 3. true 4. false 5. false 6. false 7. true

Exercises

1. a. $c_0x^0 + c_1x^1 + c_2x^2 + c_3x^3$, or $c_0 + c_1x + c_2x^2 + c_3x^3$
 c. $a_1x^1 + a_2x^2 + a_3x^3$, or $a_1x + a_2x^2 + a_3x^3$
2. a. $\sum_{j=0}^2 c_jx^j$
 c. $\sum_{i=1}^4 x^i$
3. a. $2x^3 + 4x^2 + 3x + 2$
 e. $2x^2 + 2x + 3$
 g. $4x^5 + 4x^2 + 7x + 4$
4. a. $2x^3 + 4x^2 + 2x + 1$
 e. $8x^5 + 8x^4 + 4x^3 + 8x^2 + 4x + 6$
 g. $2x^5 + 8x^4 + 7x^3 + 5x^2 + 7x$
5. a. The set S of all polynomials with zero constant term is nonempty, since it contains the zero polynomial. Both the sum and the product of polynomials with zero constant term are again polynomials with zero constant term, so S is closed under addition and multiplication. The additive inverse of a polynomial with zero constant term is also a polynomial with zero constant term, so S is a subring of $R[x]$.
 c. Let S be the set of all polynomials that have zero coefficients for all odd powers of x . Then x^2 is in S , so S is nonempty. For arbitrary

$$f(x) = \sum_{i=0}^n a_{2i}x^{2i} \quad \text{and} \quad g(x) = \sum_{i=0}^m b_{2i}x^{2i}$$

in S , let k be the larger of n and m . Then

$$f(x) + g(x) = \sum_{i=0}^k (a_{2i} + b_{2i})x^{2i}$$

has zero coefficients for all odd powers of x and therefore is in S . Also,

$$f(x)g(x) = \sum_{i=0}^{m+n} \left(\sum_{p+q=i} a_{2p}b_{2q} \right) x^{2i}$$

is in S , and

$$-f(x) = \sum_{i=0}^n (-a_{2i})x^{2i}$$

is in S . Thus S is a subring of $R[x]$.

- 6.** **a.** Since a product of a polynomial with zero constant term and any other polynomial always has zero constant term, S is an ideal of $R[x]$. Also S is a principal ideal where $S = (x) = \{x \cdot f(x) | f(x) \in R[x]\}$.
- c.** The polynomial $x^2 \in S$ and $x \in R[x]$, but the product $x(x^2) = x^3 \notin S$. Thus, S is not an ideal of $R[x]$, and hence S is not a principal ideal.
- 9.** **b.** $I[x]$ is a principal ideal where $I[x] = (1 - x)$.
- 11.** **a.** $x^2, x^2 + 1, x^2 + 2, x^2 + x, x^2 + x + 1, x^2 + x + 2, x^2 + 2x, x^2 + 2x + 1, x^2 + 2x + 2, 2x^2, 2x^2 + 1, 2x^2 + 2, 2x^2 + x, 2x^2 + x + 1, 2x^2 + x + 2, 2x^2 + 2x, 2x^2 + 2x + 1, 2x^2 + 2x + 2$
- b. none**
- 12.** **a.** We write a for $[a]$ in \mathbb{Z}_4 . The polynomial $2x + 1$ is a unit, since $(2x + 1)(2x + 1) = 4x^2 + 4x + 1 = 1$ in $\mathbb{Z}_4[x]$.
- 13.** **a.** $n^2(n - 1)$ **b.** $n^m(n - 1)$
- 16.** **b.** n **c.** 0
- 17.** **b.** $\ker \theta$ is the set of all polynomials in $R[x]$ that have zero constant term. (That is, $\ker \theta$ is the principal ideal (x) generated by x in $R[x]$.)
- 21.** $\ker \phi$ is the set of all polynomials in $\mathbb{Z}[x]$ that are multiples of k . (That is, $\ker \phi$ is the principal ideal (k) generated by k in $R[x]$.)
- 23.** $\ker \phi$ is the set of all polynomials $f(x) = a_0 + a_1x + \cdots + a_nx^n$ in $R[x]$ such that all the coefficients a_i are in $\ker \theta$.

Exercises 8.2 Pages 373–375

True or False

- 1.** true **2.** true **3.** false

Exercises

- 1.** $q(x) = 4x^2 + 3x + 2, r(x) = 4$
- 3.** $q(x) = x + 2, r(x) = x^2 + x$
- 5.** $q(x) = 5x^2 + 3, r(x) = 2x + 3$
- 7.** $d(x) = x + 1$
- 9.** $d(x) = x + 5$
- 11.** $s(x) = x^2 + 2x + 1, t(x) = x$
- 13.** $s(x) = x^2 + 2, t(x) = 4$
- 15.** **a.** $(3x + 4)(4x + 3) = x$

16. a. $(3x - 2)(4x - 1) = (3x + 10)(4x + 11)$
17. a. $(2x - 1)(5x - 7) = (2x + 9)(5x + 3)$
19. a. yes
20. a. yes
21. a. no
22. a. no
35. A *least common multiple* of two nonzero polynomials $f(x)$ and $g(x)$ in $F[x]$ is a polynomial $m(x)$ in $F[x]$ that satisfies the following conditions:
1. $m(x)$ is monic.
 2. $f(x) | m(x)$ and $g(x) | m(x)$.
 3. $f(x) | k(x)$ and $g(x) | k(x)$ imply $m(x) | k(x)$.

Exercises 8.3 Pages 381–384

True or False

1. true 2. true 3. false 4. false 5. false
6. false 7. true 8. false

Exercises

1. a. -9 c. 0 e. 1 g. 0 i. 4
2. a. $x^2 - 2$ is irreducible over \mathbf{Q} , reducible over \mathbf{R} and \mathbf{C} , since $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$. $\sqrt{2}$ and $-\sqrt{2}$ are zeros in \mathbf{R} and \mathbf{C} .
c. $x^2 + x - 2 = (x + 2)(x - 1)$ is reducible over the fields \mathbf{Q} , \mathbf{R} , and \mathbf{C} with zeros -2 and 1 in \mathbf{Q} , \mathbf{R} , and \mathbf{C} .
e. $x^2 + x + 2$ is irreducible over \mathbf{Z}_3 and \mathbf{Z}_5 ; $x^2 + x + 2 = (x + 4)^2$ is reducible over \mathbf{Z}_7 , and 3 is a zero of multiplicity 2 in \mathbf{Z}_7 .
g. $x^3 - x^2 + 2x + 2$ is irreducible over \mathbf{Z}_3 ; $x^3 - x^2 + 2x + 2 = (x + 3)^3$ is reducible over \mathbf{Z}_5 , and 2 is a zero of multiplicity 3 in \mathbf{Z}_5 ; Also $x^3 - x^2 + 2x + 2 = (x + 2)(x^2 + 4x + 1)$ is reducible over \mathbf{Z}_7 , and 5 is a zero in \mathbf{Z}_7 .
3. $x^2 + 1, x^2 + x + 2, x^2 + 2x + 2$
4. a. $2x^3 + 1 = 2(x + 2)(x^2 + 3x + 4)$, and 3 is a zero of multiplicity 1 .
c. $3x^3 + x^2 + 2x + 4 = 3(x + 1)(x + 2)(x + 4)$, and $4, 3$, and 1 are zeros, each of multiplicity 1 .
e. $2x^4 + x^3 + 3x + 2 = 2(x + 1)(x + 2)(x^2 + 3)$ with zeros 4 and 3 , each of multiplicity 1 .

- g.** $x^4 + x^3 + x^2 + 2x + 3 = (x + 3)^2(x^2 + 2)$, and 2 is a zero of multiplicity 2.
- i.** $x^4 + 2x^3 + 3x + 4 = (x + 4)(x + 1)^3$, 1 is a zero of multiplicity 1, and 4 is a zero of multiplicity 3.
- 7. a.** $4x^2 + 4$ has degree 2 and has 4 zeros: 1, 3, 5, and 7.
- 12. a.** 0, 1, 2, 3, and 4 in \mathbf{Z}_5
- 13.** $x^4 + 5x^2 + 4 = (x^2 + 1)(x^2 + 4)$ is reducible over \mathbf{R} and has no zeros in the field of real numbers.
- 27.** Exercise 7: $(x + 1)^2(x + 2)(x^3 + 2x^2 + 1) = x^6 + x^4 + x^3 + 2x^2 + 2x + 2$
 Exercise 9: $(x + 2)(x + 5)(x^3 + 2x^2 + 2x + 6) = x^5 + 2x^4 + 5x^3 + 5x^2 + 6x + 4$

Exercises 8.4 Pages 394–397

True or False

- | | | | | |
|------------------|------------------|------------------|-----------------|-----------------|
| 1. true | 2. true | 3. true | 4. false | 5. false |
| 6. false | 7. true | 8. true | 9. false | 10. true |
| 11. false | 12. false | 13. false | 14. true | 15. true |
-

Exercises

- 1. a.** $f(x) = x^2 - (3 + 2i)x + 6i$, $g(x) = x^3 - 3x^2 + 4x - 12$
c. $f(x) = x^2 - (3 - i)x + (2 - 2i)$, $g(x) = x^3 - 4x^2 + 6x - 4$
e. $f(x) = x^2 - (1 + 5i)x - (6 - 3i)$, $g(x) = x^4 - 2x^3 + 14x^2 - 18x + 45$
g. $f(x) = x^3 - 3x^2 + (3 - 2i)x - (1 - 2i)$,
 $g(x) = x^5 - 5x^4 + 10x^3 - 10x^2 + 9x - 5$
- 2. a.** $1 + i, 2$
c. $i, (-1 + i\sqrt{3})/2, (-1 - i\sqrt{3})/2$
- 3.** $5/2, -1$
- 5.** $3/2, -1$
- 7.** $-2, (1 + i\sqrt{3})/2, (1 - i\sqrt{3})/2$
- 9.** $1, 1/3, -2$
- 11.** $-1, 1/2, -4/3$
- 13.** $x^4 - x^3 - 2x^2 + 6x - 4 = (x - 1)(x + 2)(x^2 - 2x + 2)$
- 15.** $2x^4 + 5x^3 - 7x^2 - 10x + 6 = 2(x - \frac{1}{2})(x + 3)(x^2 - 2)$

- 17.** **a.** Let $f(x) = 3 + 9x + x^3$. The prime integer 3 divides all the coefficients of $f(x)$ except the leading coefficient $a_n = 1$, and 3^2 does not divide $a_0 = 3$. Thus $f(x)$ is irreducible by Eisenstein's Criterion.
- c.** Let $f(x) = 3 - 27x^2 + 2x^5$. The prime integer 3 divides all the coefficients of $f(x)$ except the leading coefficient $a_n = 2$, and 3^2 does not divide $a_0 = 3$. Thus $f(x)$ is irreducible by Eisenstein's Criterion.
- 20.** **a.** Let $f(x) = 1 + 2x + 6x^2 - 4x^3 + 2x^4$. The prime integer 2 divides all the coefficients of $f(x)$ except the constant term $a_0 = 1$, and 2^2 does not divide $a_n = 2$. Thus $f(x)$ is irreducible by Exercise 19.
- c.** Let $f(x) = 6 - 35x + 14x^2 + 7x^5$. The prime integer 7 divides all the coefficients of $f(x)$ except the constant term $a_0 = 6$, and 7^2 does not divide $a_n = 7$. Thus $f(x)$ is irreducible by Exercise 19.
- 21.** **a.** $f_2(x) = x^3 + x + 1$ has no zeros in \mathbf{Z}_2 .
- c.** $f_5(x) = 2x^3 + 3x^2 + 3$ has no zeros in \mathbf{Z}_5 .
- e.** $f_2(x) = x^4 + x^3 + x^2 + x + 1$ has no zeros in \mathbf{Z}_2 and hence no first-degree factors in \mathbf{Z}_2 . The only possible second-degree factors in \mathbf{Z}_2 are x^2 , $x^2 + x$, $x^2 + 1$, and $x^2 + x + 1$. Now $x^2 = x \cdot x$, $x^2 + x = x(x + 1)$, and $x^2 + 1 = (x + 1)^2$ are not factors of $f_2(x)$, since $f_2(x)$ has no first-degree factors. Long division shows that $x^2 + x + 1$ is not a factor of $f_2(x)$. Thus $f_2(x)$ is irreducible in \mathbf{Z}_2 , and hence $f(x) = 3x^4 + 9x^3 - 7x^2 + 15x + 25$ is irreducible by Theorem 8.34.
- 34.** **a.** Let $f(x) = x^3 + 3x + 8$. Then $f(x + 1) = x^3 + 3x^2 + 6x + 12$ is irreducible by the Eisenstein Irreducibility Criterion implies $f(x)$ is irreducible over \mathbf{Q} .

Exercises 8.5 Pages 408–409

True or False

1. true 2. false 3. true 4. false

Exercises

1. $\sqrt[3]{25} + \sqrt[3]{5}$, $-\frac{\sqrt[3]{25} + \sqrt[3]{5}}{2} \pm i\sqrt{3}\frac{\sqrt[3]{25} - \sqrt[3]{5}}{2}$
3. $\sqrt[3]{16} + \sqrt[3]{4}$, $-\frac{\sqrt[3]{16} + \sqrt[3]{4}}{2} \pm i\sqrt{3}\frac{\sqrt[3]{16} - \sqrt[3]{4}}{2}$
5. $\sqrt[3]{4} + \sqrt[3]{2}$, $-\frac{\sqrt[3]{4} + \sqrt[3]{2}}{2} \pm i\sqrt{3}\frac{\sqrt[3]{4} - \sqrt[3]{2}}{2}$
7. $\sqrt[3]{3} - \sqrt[3]{9}$, $\frac{\sqrt[3]{9} - \sqrt[3]{3}}{2} \pm i\sqrt{3}\frac{\sqrt[3]{9} + \sqrt[3]{3}}{2}$
9. $\frac{2\sqrt[3]{2} - \sqrt[3]{4}}{2}$, $\frac{\sqrt[3]{4} - 2\sqrt[3]{2}}{4} \pm i\sqrt{3}\frac{\sqrt[3]{4} + 2\sqrt[3]{2}}{4}$

11. $\sqrt[3]{49} - \sqrt[3]{7} + 2, \quad \frac{\sqrt[3]{7} - \sqrt[3]{49} + 4}{2} \pm i\sqrt{3} \frac{\sqrt[3]{7} + \sqrt[3]{49}}{2}$

13. $\sqrt[3]{18} - \sqrt[3]{12} - \frac{1}{2}, \quad \frac{\sqrt[3]{12} - \sqrt[3]{18} - 1}{2} \pm i\sqrt{3} \frac{\sqrt[3]{12} + \sqrt[3]{18}}{2}$

15. $1 \pm i, \quad -1 \pm i\sqrt{2}$

17. $\pm i, \quad -2 \pm \sqrt{2}$

19. Since $D^2 = -27(90)^2 - 4(-91)^3 = 2,795,584 > 0$, all three solutions are real.

21. Since $D^2 = -27(-72)^2 - 4(-55)^3 = 525,532 > 0$, all three solutions are real.

23. Since $D^2 = -27(-136)^2 - 4(-47)^3 = -84,100 < 0$, there is one real solution and one pair of complex conjugates.

Exercises 8.6 Pages 419–421

True or False

1. true 2. false 3. false

Exercises

1. a. Let $P = (p(x))$ and $\alpha = x + P$ in $\mathbf{Z}_3[x]/P$. The elements of $\mathbf{Z}_3[x]/P$ are

$$\{0, 1, 2, \alpha, \alpha + 1, \alpha + 2, 2\alpha, 2\alpha + 1, 2\alpha + 2\},$$

where $0 = 0 + P$, $1 = 1 + P$, and $2 = 2 + P$. Addition and multiplication tables are as follows:

+	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
0	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
1	1	2	0	$\alpha + 1$	$\alpha + 2$	α	$2\alpha + 1$	$2\alpha + 2$	2α
2	2	0	1	$\alpha + 2$	α	$\alpha + 1$	$2\alpha + 2$	2α	$2\alpha + 1$
α	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$	0	1	2
$\alpha + 1$	$\alpha + 1$	$\alpha + 2$	α	$2\alpha + 1$	$2\alpha + 2$	2α	1	2	0
$\alpha + 2$	$\alpha + 2$	α	$\alpha + 1$	$2\alpha + 2$	2α	$2\alpha + 1$	2	0	1
2α	2α	$2\alpha + 1$	$2\alpha + 2$	0	1	2	α	$\alpha + 1$	$\alpha + 2$
$2\alpha + 1$	$2\alpha + 1$	$2\alpha + 2$	2α	1	2	0	$\alpha + 1$	$\alpha + 2$	α
$2\alpha + 2$	$2\alpha + 2$	2α	$2\alpha + 1$	2	0	1	$\alpha + 2$	α	$\alpha + 1$

.	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
2	0	2	1	2α	$2\alpha + 2$	$2\alpha + 1$	α	$\alpha + 2$	$\alpha + 1$
α	0	α	2α	$2\alpha + 1$	1	$\alpha + 1$	$\alpha + 2$	$2\alpha + 2$	2
$\alpha + 1$	0	$\alpha + 1$	$2\alpha + 2$	1	$\alpha + 2$	2α	2	α	$2\alpha + 1$
$\alpha + 2$	0	$\alpha + 2$	$2\alpha + 1$	$\alpha + 1$	2α	2	$2\alpha + 2$	1	α
2α	0	2α	α	$\alpha + 2$	2	$2\alpha + 2$	$2\alpha + 1$	$\alpha + 1$	1
$2\alpha + 1$	0	$2\alpha + 1$	$\alpha + 2$	$2\alpha + 2$	α	1	$\alpha + 1$	2	2α
$2\alpha + 2$	0	$2\alpha + 2$	$\alpha + 1$	2	$2\alpha + 1$	α	1	2α	$\alpha + 2$

2. a. $\mathbf{Z}_2[x]/(p(x)) = \{0, 1, \alpha, \alpha + 1\}$ is a field.

+	0	1	α	$\alpha + 1$
0	0	1	α	$\alpha + 1$
1	1	0	$\alpha + 1$	α
α	α	$\alpha + 1$	0	1
$\alpha + 1$	$\alpha + 1$	α	1	0

.	0	1	α	$\alpha + 1$
0	0	0	0	0
1	0	1	α	$\alpha + 1$
α	0	α	$\alpha + 1$	1
$\alpha + 1$	0	$\alpha + 1$	1	α

c. $\mathbf{Z}_2[x]/(p(x)) = \{0, 1, \alpha, \alpha + 1, \alpha^2, \alpha^2 + 1, \alpha^2 + \alpha, \alpha^2 + \alpha + 1\}$ is a field.

+	0	1	α	$\alpha + 1$	α^2	$\alpha^2 + 1$	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$
0	0	1	α	$\alpha + 1$	α^2	$\alpha^2 + 1$	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$
1	1	0	$\alpha + 1$	α	$\alpha^2 + 1$	α^2	$\alpha^2 + \alpha + 1$	$\alpha^2 + \alpha$
α	α	$\alpha + 1$	0	1	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$	α^2	$\alpha^2 + 1$
$\alpha + 1$	$\alpha + 1$	α	1	0	$\alpha^2 + \alpha + 1$	$\alpha^2 + \alpha$	$\alpha^2 + 1$	α^2
α^2	α^2	$\alpha^2 + 1$	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$	0	1	α	$\alpha + 1$
$\alpha^2 + 1$	$\alpha^2 + 1$	α^2	$\alpha^2 + \alpha + 1$	$\alpha^2 + \alpha$	1	0	$\alpha + 1$	α
$\alpha^2 + \alpha$	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$	α^2	$\alpha^2 + 1$	α	$\alpha + 1$	0	1
$\alpha^2 + \alpha + 1$	$\alpha^2 + \alpha + 1$	$\alpha^2 + \alpha$	$\alpha^2 + 1$	α^2	$\alpha + 1$	α	1	0

.	0	1	α	$\alpha + 1$	α^2	$\alpha^2 + 1$	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$
0	0	0	0	0	0	0	0	0
1	0	1	α	$\alpha + 1$	α^2	$\alpha^2 + 1$	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$
α	0	α	α^2	$\alpha^2 + \alpha$	$\alpha + 1$	1	$\alpha^2 + \alpha + 1$	$\alpha^2 + \alpha + 1$
$\alpha + 1$	0	$\alpha + 1$	$\alpha^2 + \alpha$	$\alpha^2 + 1$	$\alpha^2 + \alpha + 1$	α^2	1	α
α^2	0	α^2	$\alpha + 1$	$\alpha^2 + \alpha + 1$	$\alpha^2 + \alpha$	α	$\alpha^2 + 1$	1
$\alpha^2 + 1$	0	$\alpha^2 + 1$	1	α^2	α	$\alpha^2 + \alpha + 1$	$\alpha + 1$	$\alpha^2 + \alpha$
$\alpha^2 + \alpha$	0	$\alpha^2 + \alpha$	$\alpha^2 + \alpha + 1$	1	$\alpha^2 + 1$	$\alpha + 1$	α	α^2
$\alpha^2 + \alpha + 1$	0	$\alpha^2 + \alpha + 1$	$\alpha^2 + 1$	α	1	$\alpha^2 + \alpha$	α^2	$\alpha + 1$

e. The elements of $\mathbf{Z}_3[x]/(p(x))$ are given by

$$\{0, 1, 2, \alpha, \alpha + 1, \alpha + 2, 2\alpha, 2\alpha + 1, 2\alpha + 2\}.$$

This ring is not a field, since $\alpha + 2$ does not have a multiplicative inverse.

+	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
0	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
1	1	2	0	$\alpha + 1$	$\alpha + 2$	α	$2\alpha + 1$	$2\alpha + 2$	2α
2	2	0	1	$\alpha + 2$	α	$\alpha + 1$	$2\alpha + 2$	2α	$2\alpha + 1$
α	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$	0	1	2
$\alpha + 1$	$\alpha + 1$	$\alpha + 2$	α	$2\alpha + 1$	$2\alpha + 2$	2α	1	2	0
$\alpha + 2$	$\alpha + 2$	α	$\alpha + 1$	$2\alpha + 2$	2α	$2\alpha + 1$	2	0	1
2α	2α	$2\alpha + 1$	$2\alpha + 2$	0	1	2	α	$\alpha + 1$	$\alpha + 2$
$2\alpha + 1$	$2\alpha + 1$	$2\alpha + 2$	2α	1	2	0	$\alpha + 1$	$\alpha + 2$	α
$2\alpha + 2$	$2\alpha + 2$	2α	$2\alpha + 1$	2	0	1	$\alpha + 2$	α	$\alpha + 1$

.	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	α	$\alpha + 1$	$\alpha + 2$	2α	$2\alpha + 1$	$2\alpha + 2$
2	0	2	1	2α	$2\alpha + 2$	$2\alpha + 1$	α	$\alpha + 2$	$\alpha + 1$
α	0	α	2α	$2\alpha + 2$	2	$\alpha + 2$	$\alpha + 1$	$2\alpha + 1$	1
$\alpha + 1$	0	$\alpha + 1$	$2\alpha + 2$	2	α	$2\alpha + 1$	1	$\alpha + 2$	2α
$\alpha + 2$	0	$\alpha + 2$	$2\alpha + 1$	$\alpha + 2$	$2\alpha + 1$	0	$2\alpha + 1$	0	$\alpha + 2$
2α	0	2α	α	$\alpha + 1$	1	$2\alpha + 1$	$2\alpha + 2$	$\alpha + 2$	2
$2\alpha + 1$	0	$2\alpha + 1$	$\alpha + 2$	$2\alpha + 1$	$\alpha + 2$	0	$\alpha + 2$	0	$2\alpha + 1$
$2\alpha + 2$	0	$2\alpha + 2$	$\alpha + 1$	1	2α	$\alpha + 2$	2	$2\alpha + 1$	α

3. a. We have $p(0) = 1$, $p(1) = 1$, and $p(2) = 2$. Therefore, $p(x)$ is irreducible by Theorem 8.20.

$$\begin{aligned}
 \mathbf{b.} \quad & (a_0 + a_1\alpha + a_2\alpha^2)(b_0 + b_1\alpha + b_2\alpha^2) \\
 &= (a_0b_0 + 2a_1b_2 + 2a_2b_1 + 2a_2b_2) \\
 &\quad + (a_0b_1 + a_1b_0 + 2a_2b_2)\alpha \\
 &\quad + (a_0b_2 + a_1b_1 + a_1b_2 + a_2b_0 + a_2b_1 + a_2b_2)\alpha^2
 \end{aligned}$$

$$\mathbf{c.} \quad (\alpha^2 + \alpha + 2)^{-1} = \alpha + 1$$

5. a. Since $p(0) = 1$, $p(1) = 3$, $p(2) = 1$, $p(3) = 1$, and $p(4) = 4$, Theorem 8.20 assures us that $p(x)$ is irreducible.

$$\begin{aligned}
 \mathbf{b.} \quad & (a_0 + a_1\alpha + a_2\alpha^2)(b_0 + b_1\alpha + b_2\alpha^2) \\
 &= (a_0b_0 + 4a_1b_2 + 4a_2b_1) \\
 &\quad + (a_0b_1 + a_1b_0 + 4a_1b_2 + 4a_2b_1 + 4a_2b_2)\alpha \\
 &\quad + (a_0b_2 + a_1b_1 + a_2b_0 + 4a_2b_2)\alpha^2
 \end{aligned}$$

$$\mathbf{c.} \quad (\alpha^2 + 4\alpha)^{-1} = 4\alpha^2 + 3\alpha + 2$$

7. a. $0, 1, 2, \alpha, \alpha + 1, \alpha + 2, 2\alpha, 2\alpha + 1, 2\alpha + 2, \alpha^2, \alpha^2 + 1, \alpha^2 + 2, 2\alpha^2, 2\alpha^2 + 1, 2\alpha^2 + 2, \alpha^2 + \alpha, \alpha^2 + \alpha + 1, \alpha^2 + \alpha + 2, 2\alpha^2 + \alpha, 2\alpha^2 + \alpha + 1, 2\alpha^2 + \alpha + 2, \alpha^2 + 2\alpha, \alpha^2 + 2\alpha + 1, \alpha^2 + 2\alpha + 2, 2\alpha^2 + 2\alpha, 2\alpha^2 + 2\alpha + 1, 2\alpha^2 + 2\alpha + 2$

9. a. The polynomial $p(x) = x^4 + x^2 + 1$ is irreducible over \mathbf{Z}_2 . Let α be a zero of $p(x)$ in $\mathbf{Z}_2[x]/(p(x))$. The quotient ring

$$\begin{aligned}
 \mathbf{Z}_2[x]/(p(x)) = \{ & 0, 1, \alpha, \alpha + 1, \alpha^2, \alpha^2 + 1, \alpha^2 + \alpha, \alpha^2 + \alpha + 1, \alpha^3, \\
 & \alpha^3 + 1, \alpha^3 + \alpha, \alpha^3 + \alpha + 1, \alpha^3 + \alpha^2, \alpha^3 + \alpha^2 + 1, \\
 & \alpha^3 + \alpha^2 + \alpha, \alpha^3 + \alpha^2 + \alpha + 1 \}
 \end{aligned}$$

containing 2^4 elements is a field.

- c. The polynomial $p(x) = x^3 + 2x^2 + x + 1$ is irreducible over \mathbf{Z}_3 . Let α be a zero of $p(x)$ in $\mathbf{Z}_3[x]/(p(x))$. The quotient ring

$$\begin{aligned}\mathbf{Z}_3[x]/(p(x)) = \{0, 1, 2, \alpha, \alpha + 1, \alpha + 2, 2\alpha, 2\alpha + 1, 2\alpha + 2, \alpha^2, \alpha^2 + 1, \\ \alpha^2 + 2, 2\alpha^2, 2\alpha^2 + 1, 2\alpha^2 + 2, \alpha^2 + \alpha, \alpha^2 + \alpha + 1, \alpha^2 + \alpha + 2, \\ \alpha^2 + 2\alpha, \alpha^2 + 2\alpha + 1, \alpha^2 + 2\alpha + 2, 2\alpha^2 + \alpha, 2\alpha^2 + \alpha + 1, \\ 2\alpha^2 + \alpha + 2, 2\alpha^2 + 2\alpha, 2\alpha^2 + 2\alpha + 1, 2\alpha^2 + 2\alpha + 2\}\end{aligned}$$

containing 3^3 elements is a field.

11. $\frac{1}{32}(-\sqrt[3]{9} + 5\sqrt[3]{3} + 7)$
 13. a. 3, 4 c. 2, 3 e. 5, 5
 15. $\alpha, 2\alpha + 1$ 17. $\alpha, 2\alpha^2 + 3\alpha, 3\alpha^2 + \alpha + 4$

Appendix Exercises Pages 431–433

1. For $x = 0$, the statement $0^2 > 0$ is false.
3. For $a = 0$ and any real number b , the statement $0 \cdot b = 1$ is false.
5. For $x = -4$, the statement $-(-4) < |-4|$ is false.
7. For $n = 6$, the statement $6^2 + 2(6) = 48$ is true.
9. For $n = 5$, the statement $5^2 < 2^5$ is true.
11. For $n = 3$, the integer $3^2 + 3$ is an even integer.
13. There is at least one child who did not receive a Valentine card.
15. There is at least one senior who either did not graduate or did not receive a job offer.
17. All of the apples in the basket are not rotten.
19. All of the politicians are dishonest or untrustworthy.
21. There is at least one $x \in A$ such that $x \notin B$.
23. There exists a right triangle with sides a and b and hypotenuse c such that $c^2 \neq a^2 + b^2$.
25. Some complex number does not have a multiplicative inverse.
27. There are sets A and B such that the Cartesian products $A \times B$ and $B \times A$ are not equal.
29. For every complex number x , $x^2 + 1 \neq 0$.
31. For all sets A and B , the set A is not a subset of $A \cap B$.
33. For any triangle with angles α, β , and γ , the inequality $\alpha + \beta + \gamma \leq 180^\circ$ holds.

35. For every real number x , $2^x > 0$.

37. TRUTH TABLE for $p \Leftrightarrow \sim(\sim p)$

p	$\sim p$	$\sim(\sim p)$
T	F	T
F	T	F

We examine the two columns headed by p and $\sim(\sim p)$ and note that they are identical.

39. TRUTH TABLE for $\sim(p \wedge (\sim p))$

p	$\sim p$	$p \wedge (\sim p)$	$\sim(p \wedge (\sim p))$
T	F	F	T
F	T	F	T

41. TRUTH TABLE for $(p \wedge q) \Rightarrow p$

p	q	$p \wedge q$	$(p \wedge q) \Rightarrow p$
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	T

43. TRUTH TABLE for $(p \wedge (p \Rightarrow q)) \Rightarrow q$

p	q	$p \Rightarrow q$	$p \wedge (p \Rightarrow q)$	$(p \wedge (p \Rightarrow q)) \Rightarrow q$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

45. TRUTH TABLE for $(p \Rightarrow q) \Leftrightarrow ((\sim p) \vee q)$

p	q	$p \Rightarrow q$	$\sim p$	$(\sim p) \vee q$
T	T	T	F	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

We examine the two columns headed by $p \Rightarrow q$ and $(\sim p) \vee q$ and note that they are identical.

47. TRUTH TABLE for $(p \Rightarrow q) \Leftrightarrow ((p \wedge (\sim q)) \Rightarrow (\sim p))$

p	q	$p \Rightarrow q$	$\sim q$	$p \wedge (\sim q)$	$\sim p$	$(p \wedge (\sim q)) \Rightarrow (\sim p)$
T	T	T	F	F	F	T
T	F	F	T	T	F	F
F	T	T	F	F	T	T
F	F	T	T	F	T	T

We examine the two columns headed by $p \Rightarrow q$ and $(p \wedge (\sim q)) \Rightarrow (\sim p)$ and note that they are identical.

49. TRUTH TABLE for $(p \wedge q \wedge r) \Rightarrow ((p \vee q) \wedge r)$

p	q	r	$p \wedge q \wedge r$	$p \vee q$	$(p \vee q) \wedge r$	$(p \wedge q \wedge r) \Rightarrow ((p \vee q) \wedge r)$
T	T	T	T	T	T	T
T	T	F	F	T	F	T
T	F	T	F	T	T	T
T	F	F	F	T	F	T
F	T	T	F	T	T	T
F	T	F	F	T	F	T
F	F	T	F	F	F	T
F	F	F	F	F	F	T

51. TRUTH TABLE for $(p \Rightarrow (q \wedge r)) \Leftrightarrow ((p \Rightarrow q) \wedge (p \Rightarrow r))$

p	q	r	$q \wedge r$	$p \Rightarrow (q \wedge r)$	$p \Rightarrow q$	$p \Rightarrow r$	$(p \Rightarrow q) \wedge (p \Rightarrow r)$
T	T	T	T	T	T	T	T
T	T	F	F	F	T	F	F
T	F	T	F	F	F	T	F
T	F	F	F	F	F	F	F
F	T	T	T	T	T	T	T
F	T	F	F	T	T	T	T
F	F	T	F	T	T	T	T
F	F	F	F	T	T	T	T

We examine the two columns headed by $p \Rightarrow (q \wedge r)$ and $(p \Rightarrow q) \wedge (p \Rightarrow r)$ and note that they are identical.

- 53.** The implication $(p \Rightarrow q)$ is true: My grade for this course is A implies that I can enroll in the next course.

The contrapositive $(\sim q \Rightarrow \sim p)$ is true: I cannot enroll in the next course implies that my grade for this course is not A.

The inverse $(\sim p \Rightarrow \sim q)$ is false: My grade for this course is not A implies that I cannot enroll in the next course.

The converse $(q \Rightarrow p)$ is false: I can enroll in the next course implies that my grade for this course is A.

- 55.** The implication $(p \Rightarrow q)$ is true: The Saints win the Super Bowl implies that the Saints are the champion football team.

The contrapositive $(\sim q \Rightarrow \sim p)$ is true: The Saints are not the champion football team implies that the Saints did not win the Super Bowl.

The inverse $(\sim p \Rightarrow \sim q)$ is true: The Saints did not win the Super Bowl implies that the Saints are not the champion football team.

The converse $(q \Rightarrow p)$ is true: The Saints are the champion football team implies that the Saints did win the Super Bowl.

- 57.** The implication $(p \Rightarrow q)$ is false: My pet has four legs implies that my pet is a dog.

The contrapositive $(\sim q \Rightarrow \sim p)$ is false: My pet is not a dog implies that my pet does not have four legs.

The inverse $(\sim p \Rightarrow \sim q)$ is true: My pet does not have four legs implies that my pet is not a dog.

The converse $(q \Rightarrow p)$ is true: My pet is a dog implies that my pet has four legs.

- 59.** The implication ($p \Rightarrow q$) is true: Quadrilateral $ABCD$ is a square implies that quadrilateral $ABCD$ is a rectangle.

The contrapositive ($\sim q \Rightarrow \sim p$) is true: Quadrilateral $ABCD$ is not a rectangle implies that quadrilateral $ABCD$ is not a square.

The inverse ($\sim p \Rightarrow \sim q$) is false: Quadrilateral $ABCD$ is not a square implies that quadrilateral $ABCD$ is not a rectangle.

The converse ($q \Rightarrow p$) is false: Quadrilateral $ABCD$ is a rectangle implies that quadrilateral $ABCD$ is a square.

- 61.** The implication ($p \Rightarrow q$) is true: x is a positive real number implies that x is a nonnegative real number.

The contrapositive ($\sim q \Rightarrow \sim p$) is true: x is a negative real number implies that x is a nonpositive real number.

The inverse ($\sim p \Rightarrow \sim q$) is false: x is a nonpositive real number implies that x is a negative real number.

The converse ($q \Rightarrow p$) is false: x is a nonnegative real number implies that x is a positive real number.

- 63.** The implication ($p \Rightarrow q$) is true: $5x$ is odd implies that x is odd.

The contrapositive ($\sim q \Rightarrow \sim p$) is true: x is not odd implies that $5x$ is not odd.

The inverse ($\sim p \Rightarrow \sim q$) is true: $5x$ is not odd implies that x is not odd.

The converse ($q \Rightarrow p$) is true: x is odd implies that $5x$ is odd.

- 65.** The implication ($p \Rightarrow q$) is true: xy is even implies that x is even or y is even.

The contrapositive ($\sim q \Rightarrow \sim p$) is true: x is odd and y is odd implies that xy is odd.

The inverse ($\sim p \Rightarrow \sim q$) is true: xy is odd implies that x is odd and y is odd.

The converse ($q \Rightarrow p$) is true: x is even or y is even implies that xy is even.

- 67.** The implication ($p \Rightarrow q$) is false: $x^2 > y^2$ implies that $x > y$.

The contrapositive ($\sim q \Rightarrow \sim p$) is false: $x \leq y$ implies that $x^2 \leq y^2$.

The inverse ($\sim p \Rightarrow \sim q$) is false: $x^2 \leq y^2$ implies that $x \leq y$.

The converse ($q \Rightarrow p$) is false: $x > y$ implies that $x^2 > y^2$.

- 69.** Contrapositive: $\sim(q \vee r) \Rightarrow \sim p$, or $((\sim q) \wedge (\sim r)) \Rightarrow \sim p$

Converse: $(q \vee r) \Rightarrow p$

Inverse: $\sim p \Rightarrow \sim(q \vee r)$, or $\sim p \Rightarrow ((\sim q) \wedge (\sim r))$

- 71.** Contrapositive: $q \Rightarrow \sim p$

Converse: $\sim q \Rightarrow p$

Inverse: $\sim p \Rightarrow q$

- 73.** Contrapositive: $\sim(r \wedge s) \Rightarrow \sim(p \vee q)$, or $((\sim r) \vee (\sim s)) \Rightarrow ((\sim p) \wedge (\sim q))$

Converse: $(r \wedge s) \Rightarrow (p \vee q)$

Inverse: $\sim(p \vee q) \Rightarrow \sim(r \wedge s)$, or $((\sim p) \wedge (\sim q)) \Rightarrow ((\sim r) \vee (\sim s))$

Bibliography

- Ames, Dennis B. *An Introduction to Abstract Algebra*. Scranton, PA: International Textbook, 1969.
- Anderson, Marlow, and Todd Feil. *A First Course in Abstract Algebra*. 2nd ed. Boca Raton: Chapman & Hall/CRC, 2005.
- Ball, Richard W. *Principles of Abstract Algebra*. New York: Holt, Rinehart and Winston, 1963.
- Ball, W. W. Rouse. *Mathematical Recreations & Essays*. 13th ed. New York: Dover, 1987.
- Beker, Henry. *Cipher Systems: The Protection of Communications*. New York: Wiley, 1982.
- Birkhoff, Garrett, and Saunders MacLane. *A Survey of Modern Algebra*. 4th ed. New York: A. K. Peters Limited, 1997.
- Bland, Paul. *The Basics of Abstract Algebra*. San Francisco: Freeman, 2001.
- Bloch, Norman J. *Abstract Algebra with Applications*. Englewood Cliffs, NJ: Prentice Hall, 1986.
- Bondi, Christine (editor). *New Applications of Mathematics*. New York: Penguin Books, 1991.
- Bourbaki, Nicolas. *Elements of Mathematics, Algebra*. Chapters 4–7. New York: Springer-Verlag, 2003.
- Buchthal, David C., and Douglas E. Cameron. *Modern Abstract Algebra*. Boston: PWS-Kent, 1987.
- Burdick, Charles M., and John J. Leeson. *Essentials of Abstract Algebra*. Monterey, CA: Brooks/Cole, 1972.
- Burton, David M. *Abstract Algebra*. Dubuque, IA: Wm. C. Brown, 1988.
- _____. *The History of Mathematics*. 4th ed. Boston: WCB McGraw-Hill, 1999.
- Childs, Lindsay N. *A Concrete Introduction to Higher Algebra*, 2nd ed. New York: Springer-Verlag, 2000.
- Clark, Allan. *Elements of Abstract Algebra*. New York: Dover, 1984.
- Cohn, P. M. *Algebra*. 3rd ed. 2 vols. New York: Wiley, 2000.
- _____. *Basic Algebra*. New York: Springer-Verlag, 2003.
- _____. *Classic Algebra*. New York: Wiley, 2001.
- Connell, I. *Modern Algebra: A Constructive Introduction*. New York: North Holland, 1982.
- Crown, G., M. Fenrick, and R. Valenza. *Abstract Algebra*. New York: Marcel Dekker, 1986.
- Dean, R. A. *Elements of Abstract Algebra*. New York: Wiley, 1966.
- Dubisch, Roy. *Introduction to Abstract Algebra*. New York: Wiley, 1985.
- Dummit, David S., and Richard M. Foote. *Abstract Algebra*. 3rd ed. Hoboken, NJ: Wiley, 2004.
- Durbin, John R. *Modern Algebra*. 4th ed. New York: Wiley, 1999.
- Eves, Howard. *Great Moments in Mathematics (After 1650)*. Washington, DC: Mathematical Association of America, 1981.
- _____. *Great Moments in Mathematics (Before 1650)*. Washington, DC: Mathematical Association of America, 1983.
- _____. *An Introduction to the History of Mathematics*. 6th ed. Philadelphia: Saunders, 1990.
- _____. *In Mathematical Circles, Quadrants I and II*. Boston: PWS, 1969.
- _____. *In Mathematical Circles, Quadrants III and IV*. Boston: PWS, 1969.
- Fraleigh, John B. *A First Course in Abstract Algebra*. 7th ed. Reading, MA: Addison-Wesley, 2003.
- Fuchs, Laszlo. *Infinite Abelian Groups*. 2 vols. New York: Academic Press, 1973.
- Gallian, Joseph A. *Contemporary Abstract Algebra*. 5th ed. Houghton Mifflin, 2002.

- Gilbert, W., and W. Keith Nicholson. *Modern Algebra with Applications*. 2nd ed. Hoboken, NJ: Wiley-Interscience, 2004.
- Goldstein, L. J. *Abstract Algebra: A First Course*. Englewood Cliffs, NJ: Prentice Hall, 1973.
- Goodman, Frederick M. *Algebra*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 2003.
- Grillet, Pierre Antoine. *Algebra*. New York: Wiley, 1999.
- Hall, F. M. *Introduction to Abstract Algebra*. 2nd ed. Vol. 1. New Rochelle, NY: Cambridge University Press, 1980.
- _____. *Introduction to Abstract Algebra*. 2nd ed. Vol. 2. New Rochelle, NY: Cambridge University Press, 1980.
- Hall, Marshall, Jr. *The Theory of Groups*. 2nd ed. New York: Chelsea, 1976.
- Hardy, Darel W., and Carol L. Walker. *Applied Algebra*. Englewood Cliffs, NJ: Prentice Hall, 2003.
- Herstein, I. N. *Abstract Algebra*. 3rd ed. New York: Wiley, 1999.
- Hillman, Abraham P., and Gerald L. Alexanderson. *A First Undergraduate Course in Abstract Algebra*. 5th ed. Boston: PWS, 1994.
- Hungerford, T. W. *Abstract Algebra—An Introduction*. 2nd ed. Pacific Grove, CA: Brooks/Cole, 1997.
- _____. *Algebra*. New York: Springer-Verlag, 1989.
- Jacobson, N. *Basic Algebra I*. 2nd ed. San Francisco: Freeman, 1985.
- _____. *Lectures in Abstract Algebra*. 3 vols. New York: Springer-Verlag, 1981, 1984, 1997.
- Jones, Burton W. *An Introduction to Modern Algebra*. New York: Macmillan, 1967.
- Kahn, David. *The Codebreakers: The Story of Secret Writing*. 2nd ed. New York: Scribner, 1996.
- _____. *Kahn on Codes: Secrets of the New Cryptology*. New York: Macmillan, 1983.
- Keesee, John W. *Elementary Abstract Algebra*. Lexington, MA: D.C. Heath, 1965.
- Kline, Morris. *Mathematical Thought from Ancient to Modern Times*. 3 vols. New York: Oxford University Press, 1990.
- Koblitz, Neal. *A Course in Number Theory and Cryptography*. 2nd ed. New York: Springer-Verlag, 1994.
- Konheim, Alan G. *Cryptography, A Primer*. New York: Wiley, 1981.
- Kuczkowski, J., and J. Gersting. *Abstract Algebra: A First Look*. New York: Marcel Dekker, 1977.
- Kurosh, A. *Theory of Groups*. 2 vols. Translated by K. A. Hirsch. New York: Chelsea, 1979.
- Lang, S. *Algebra*. 3rd ed. New York: Springer-Verlag, 2002.
- _____. *Undergraduate Algebra*. 3rd ed. New York: Springer-Verlag, 2005.
- Larney, V. C. *Abstract Algebra: A First Course*. Boston: PWS, 1975.
- Larsen, Max D. *Introduction to Modern Algebraic Concepts*. Reading, MA: Addison-Wesley, 1969.
- Lauritzen, Niels. *Concrete Abstract Algebra*. New Rochelle, NY: Cambridge University Press, 2003.
- Lax, Robert R. *Modern Algebra and Discrete Structures*. Reading, MA: Addison-Wesley, 1991.
- Lederman, Walter. *Introduction to the Theory of Finite Groups*. 4th ed. New York: Interscience, 1961.
- Lidi, Rudolf, and Guenter Pilz. *Applied Abstract Algebra*. 2nd ed. New York: Springer-Verlag, 1998.
- McCoy, Neal H. *Fundamentals of Abstract Algebra*. Boston: Allyn and Bacon, 1972.
- _____. *Rings and Ideals* (Carus Mathematical Monograph No. 8). Washington, DC: The Mathematical Association of America, 1968.
- _____. *The Theory of Rings*. New York: Chelsea, 1973.
- McCoy, N. H., and T. R. Berger. *Algebra: Groups, Rings, and Other Topics*. Boston: Allyn and Bacon, 1977.
- McCoy, Neal H., and Gerald Janusz. *Introduction to Modern Algebra*. 6th ed. New York: McGraw-Hill, 2000.
- Mackiw, George. *Applications of Abstract Algebra*. New York: Wiley, 1985.
- Marcus, M. *Introduction to Modern Algebra*. New York: Marcel Dekker, 1978.
- Maxfield, John E., and Margaret W. Maxfield. *Abstract Algebra and Solution by Radicals*. New York: Dover, 1992.

- Mitchell, A. Richard, and Roger W. Mitchell. *An Introduction to Abstract Algebra*. Belmont, CA: Brooks/Cole, 1974.
- Moore, J. T. *Introduction to Abstract Algebra*. New York: Academic Press, 1975.
- Mostow, George D., Joseph H. Sampson, and Jean-Pierre Meyer. *Fundamental Structures of Algebra*. New York: McGraw-Hill, 1963.
- Newman, James R. *The World of Mathematics*. Vol. 1. Scranton, PA: Harper & Row, 1988.
- Nicholson, W. Keith. *Introduction to Abstract Algebra*. 3rd ed. Hoboken, NJ: Wiley-Interscience, 2007.
- Niven, Ivan, and Herbert S. Zuckerman. *An Introduction to the Theory of Numbers*. 5th ed. New York: Wiley, 1991.
- Paley, H., and P. Weichsel. *A First Course in Abstract Algebra*. New York: Holt, Rinehart and Winston, 1966.
- Papantonopoulou, Aigili. *Algebra: Pure and Applied*. Englewood Cliffs, NJ: Prentice Hall, 2002.
- Pinter, C. C. *A Book of Abstract Algebra*. 2nd ed. New York: McGraw-Hill, 1989.
- Redfield, Robert H. *Abstract Algebra: A Concrete Introduction*. Reading, MA: Addison-Wesley, 2001.
- Rotman, Joseph J. *A First Course in Abstract Algebra*. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.
- _____. *The Theory of Groups: An Introduction*. 3rd ed. Dubuque, IA: Wm. C. Brown, 1984.
- Saracino, Dan. *Abstract Algebra: A First Course*. Reading, MA: Addison-Wesley, 1980.
- Schilling, Otto F. G., and W. Stephen Piper. *Basic Abstract Algebra*. Boston: Allyn and Bacon, 1975.
- Schneier, Bruce. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*. 2nd ed. New York: Wiley, 1996.
- Scott, W. R. *Group Theory*. 2nd ed. New York: Dover, 1987.
- Seberry, Jennifer. *Cryptography: An Introduction to Computer Security*. Englewood Cliffs, NJ: Prentice Hall, 1989.
- Shapiro, Louis. *Introduction to Abstract Algebra*. New York: McGraw-Hill, 1975.
- Sierpinski, W., and A. Schinzel. *Elementary Theory of Numbers*. New York: Elsevier, 1988.
- Smith, Laurence Dwight. *Cryptography: The Science of Secret Writing*. New York: Dover Publications, 1955.
- Solomon, Ronald. *Abstract Algebra*. Pacific Grove, CA: Brooks/Cole, 2003.
- Spence, Lawrence E., and Charles Vanden Eynden. *Elementary Abstract Algebra*. New York: HarperCollins, 1993.
- Stahl, Saul. *Introductory Modern Algebra: A Historical Approach*. New York: Wiley, 1996.
- Tannenbaum, Peter, and Robert Arnold. *Excursions in Modern Mathematics*. 4th ed. Upper Saddle River, NJ: Prentice Hall, 2001.
- Van der Waerden, Bartel. *Algebra*. Vol. 1. New York: Springer-Verlag, 2003.
- Walker, Elbert A. *Introduction to Abstract Algebra*. New York: Random House, 1987.
- Weiss, Marie J., and Roy Dubisch. *Higher Algebra for the Undergraduate*. 2nd ed. New York: Wiley, 1962.
- Welsh, Dominic. *Codes and Cryptography*. New York: Oxford University Press, 1989.

This page intentionally left blank

Index

- Abel, Niels Henrik, 189
Abelian group, 138
Absolute value
of a complex number, 142, 345
of an integer, 71
of a quaternion, 342
in an ordered integral domain, 290
Addition
of complex numbers, 334
of cosets, 298
of ideals, 300
of matrices, 43
of polynomials, 357
postulates for \mathbb{Z} , 65
properties in \mathbb{Z}_n , 107
of quaternions, 338, 339
of rational numbers, 277
Additive cipher, 124
Affine mapping, 126
Algebraic
element, 386
extension, 386, 416
system, 137
Algebraically closed field, 386
Alternating group, 199
Amplitude of a complex number, 345
Antisymmetric relation, 61
Archimedean property, 291, 333
Argument of a complex number, 345
Array, 42
Associative binary operation, 31
Associative property, 7, 20, 48
generalized, 148, 264
Asymmetric relation, 61
Automorphism, 177
inner, 238
Axiom of Choice, 39
Axis
imaginary, 343
real, 343
Biconditional, 427
Bijection, 18
Bijective mapping, 18
Binary
alphabet, 114
digit, 114
operation, 30
relation, 55
representation, 75
Binet's formula, 80
Binomial
coefficients, 77
theorem, 77
Bit, 114
Block, 114
Boolean ring, 269, 318
Caesar cipher, 123
Cancellation law
for addition, 67
in a group, 146
Cardano's Formulas, 401
Cartesian product, 13
Cauchy, Augustin Louis, 256
Cauchy's Theorem, 249
Cayley, Arthur, 62
Cayley table, 139
Cayley's Theorem, 205
Center
of a group, 162, 229
of a ring, 267
Centralizer, 162
Characteristic, 313
Check digit, 115, 117
Cipher
additive, 124
Caesar, 123
exponentiation, 130
multiplicative, 126
translation, 124
Ciphertext, 124
Closed set with respect to an operation, 32
Code, 115
error-detecting, 115
Hamming, 123
repetition, 116
triple repetition, 116
Codomain, 14
Coefficient, 356
leading, 362
Commutative
binary operation, 31
group, 138
property, 4, 31
ring, 261
Complement, 5
of one set in another, 5
Complete
induction, 75
ordered field, 328
Complex numbers, 6, 334
in polar form, 345
in standard form, 337
in trigonometric form, 345
Composite mapping, 19
Composition of mappings, 19
Conclusion, 426
Conformable matrices, 46
Congruence
class, 96
modulo n , 95
modulo a subgroup, 229
Conjugate
of a complex number, 182, 337
of an element, 199
of a subgroup, 221
zeros, 387
Conjunction, 426
Connectives, 426
Constant
polynomial, 362
term, 362
Contradiction, 430
Contrapositive, 217, 429
Converse, 429
Coordinate, 343
Corollary, 423
Coset, 217
Counterexample, 17, 425
Cryptoanalysis, 123
Cryptography, 123
Cryptology, 123
Cycle, 192
Cyclic
group, 159
subgroup, 159
Decimal representation, 330
Dedekind, Richard, 292, 328
Dedekind cuts, 292, 328
Degree of a polynomial, 362

- De Moivre's Theorem, 346
 De Morgan's Laws, 9, 428
 Determinant, 187
 Diagonal matrix, 43
 Dihedral group, 210
 Dimension of a matrix, 42
 Direct product of groups, 238, 239
 Direct sum
 of rings, 269
 of subgroups, 241
 Discriminant, 386, 405
 Disjoint
 cycles, 194
 sets, 4
 Disjunction, 426
 Distributive property, 8, 48, 66, 67, 264
 Division Algorithm
 for integers, 81, 82
 for polynomials, 367
 Division ring, 337
 Divisor, 81, 367
 greatest common, 86, 370
 zero, 109, 263
 Domain, 14
 integral, 270
 Dot product, 118
 Efficiency, 120
 Eisenstein's Irreducibility Criterion, 391
 Embedded, 282, 315
 Empty set, 4
 Endomorphism, 183
 Epimorphism, 183, 303
 Equality
 of complex numbers, 334
 of mappings, 13
 of matrices, 43
 of polynomials, 356
 of quaternions, 338
 of rational numbers, 277
 of sets, 2
 Equivalence class, 57
 relation, 55
 Error detection, 115
 Escher, M. C., 212
 Euclidean Algorithm, 88, 372
 Euclid's Lemma, 90
 Euclid's Theorem on Primes, 92
 Euler phi-function, 134, 170
 Even integer, 18
 Even parity, 115
 Even permutation, 199
 Existential quantifier, 423
 Exponentiation cipher, 130
 Exponents, 68, 155
 Extension, 282, 291, 386, 416
 algebraic, 386, 416
 field, 412
 External direct product, 239
 Factor, 81, 367
 group, 230
 theorem, 377
 Fermat's Little Theorem, 106,
 222, 383
 Fibonacci sequence, 80
 Field, 271, 273
 algebraically closed, 386
 complete ordered, 328
 of complex numbers, 334
 extension, 282, 386, 412
 ground, 416
 of quotients, 279
 of rational numbers, 282
 of real numbers, 328
 skew, 337
 splitting, 419
 Finite
 group, 141
 integral domain, 270
 ring, 260
 Finitely generated group, 251
 Four group, 181, 207
 Function, 12
 polynomial, 375
 Fundamental theorem
 of algebra, 384
 of arithmetic, 90
 on finite abelian groups, 253
 of group homomorphisms, 233
 of ring homomorphisms, 309
 Gauss, Carl Friedrich, 292, 422
 Gaussian integers, 265, 274
 Gauss's Lemma, 390
 Generalized
 associative laws, 148, 264
 distributive laws, 264
 induction, 74
 Generating set, 225, 251
 minimal, 251
 Generator, 163, 225, 251
 Geometric symmetry, 209
 Glide reflection, 211
 Graph of a complex number, 343
 Greater than, 69, 285
 Greatest common divisor, 86, 370
 Greatest lower bound, 333
 Ground field, 416
 Group, 137
 abelian, 138
 alternating, 199
 center of, 162, 229
 commutative, 138
 cyclic, 159
 dihedral, 210
 factor, 230
 finite, 141
 finitely generated, 251
 four, 181, 207
 generator of, 163
 infinite, 141
 isomorphic, 177
 Klein four, 181, 207
 octic, 202, 226
 of units, 167
 order of, 141
 quaternion, 143
 quotient, 230
 symmetric, 192
 table, 139
 Hamilton, William Rowan, 337, 353
 Hamming code, 123
 distance, 122
 weight, 122
 Hilbert, David, 324
 Homomorphic image, 185, 303
 Homomorphism, 183
 kernel of, 186, 305
 ring, 303
 Hypothesis, 426
 Ideal, 293, 294
 left, 293
 maximal, 319
 prime, 323
 principal, 296, 302
 right, 293
 trivial, 293
 Idempotent element, 150, 268, 342
 Identity element, 33, 66, 138
 left, 49
 of a group, 138
 mapping, 37
 matrix, 49
 right, 49
 two-sided, 50
 Image, 13, 15
 homomorphic, 185, 303
 Imaginary axis, 343
 number, 337
 Implication, 426
 Indeterminate, 355
 Index
 of a subgroup, 218
 of summation, 47
 Indexed collection, 58
 Induction
 complete, 75
 generalized, 74
 mathematical, 72
 postulate, 66
 strong mathematical, 75

- Infinite group, 141
- Injective mapping, 16
- Inner automorphism, 238
- Integers, 6, 65
 - even, 18
 - Gaussian, 265, 274
 - negative, 66
 - odd, 18
 - positive, 6, 66
 - postulates for, 65
 - prime, 90
 - relatively prime, 89
- Integral
 - exponents, 68, 155
 - multiples, 68, 156
- Integral domain, 270
 - finite, 270
 - ordered, 284
- Internal direct product, 238
- Intersection, 3
- Invariant subgroup, 223
- Inverse, 33, 66, 138
 - image, 15
 - implication, 429
 - of a mapping, 40
 - of a matrix, 50
 - of a relation, 61
 - multiplicative, 50, 262
- Invertible
 - element, 33, 262
 - mapping, 40
 - matrix, 50
- Irrational number, 330
- Irreducible polynomial, 378
- Irreflexive relation, 61
- Isomorphic
 - groups, 177
 - rings, 281, 303
- Isomorphism
 - of groups, 177
 - of rings, 281, 303
- Kernel, 186, 305
- Key, 124
- Klein four group, 181, 207
- Kronecker delta, 49
- Lagrange's Theorem, 219
- Law of trichotomy, 66, 285
- Laws
 - of exponents, 156
 - of multiples, 158
- Leading coefficient, 362
- Least common multiple, 94
- Least element, 286
- Least upper bound (l.u.b.), 325
- Left coset, 217
- Left distributive law, 66
- Left ideal, 293
- Left identity element, 50
- Left inverse, 33
- Lemma, 423
- Length of a word, 114
- Less than, 68, 285
- Linear combination, 86, 370
- Logical equivalence, 428
- Lower bound, 333
- Mapping, 13
 - affine, 126
 - bijective, 18
 - codomain of, 14
 - composition, 19
 - domain of, 14
 - equality of, 13
 - identity, 37
 - injective, 16
 - one-to-one, 16
 - onto, 15
 - range of, 14
 - surjective, 15
- Mathematical induction, 72
- Matrix, 42
 - addition, 43
 - diagonal, 43
 - dimension of, 42
 - equality of, 43
 - identity, 49
 - invertible, 50
 - multiplication, 45
 - multiplicative inverse of, 50
 - permutation, 144
 - square, 43
 - subtraction, 45
 - sum, 43
 - zero, 45
- Maximal ideal, 319
- Maximum likelihood
 - decoding, 116
- Minimal generating set, 251
- Minimum distance, 122
- Modulus of a complex
 - number, 345
- Monic polynomial, 370
- Monomorphism, 183
- Morgan Saucier, 506
- Multiple, 68, 81, 156, 367
- Multiplication
 - of complex numbers, 334, 346
 - of cosets, 298
 - of matrices, 45
 - of polynomials, 357
 - postulates for \mathbf{Z} , 66
 - properties in \mathbf{Z}_n , 108
 - of quaternions, 338, 339
 - of rational numbers, 277
 - table, 139
- Multiplicative
 - cipher, 126
 - inverse, 50, 262
- Multiplicity, 92, 380
- Negation of a statement, 425
- Negative, 258
 - element, 285
 - integer, 66
 - integral exponents, 155
- Nilpotent element, 269, 302
- Noether, Amalie Emmy, 324
- Nontrivial subgroup, 152
- Normal subgroup, 223
- Normalizer of a subgroup, 229
- n th root, 347
 - primitive, 351
- Octic group, 202, 226
- Odd integer, 18
- Odd parity, 115
- Odd permutation, 199
- One-to-one
 - correspondence, 18
 - mapping, 16
- Onto mapping, 15
- Opposite, 258
- Orbits, 193
- Order
 - of a group, 141
 - of an element, 167, 194
 - relation, 68
- Ordered
 - field, 290
 - integral domain, 284
 - pair, 13
- Parallelogram rule, 344
- Parity
 - check digit, 115
 - even, 115
 - odd, 115
- Partition, 7, 58
- Pascal, Blaise, 135
- Permutation, 37
 - even, 199
 - matrix, 144
 - odd, 199
- p -group, 246
- Plaintext, 124
- Polar form of a complex
 - number, 345
- Polynomial(s), 355
 - addition of, 357
 - coefficient of, 356
 - constant, 362
 - degree of, 362
 - equality of, 356
 - irreducible, 378
 - monic, 370
 - multiplication of, 357
 - prime, 378
 - primitive, 390
 - reducible, 378
 - terms of, 356
 - zero of, 375, 380

- Polynomial**
 function, 375
 mapping, 375
Positive elements, 284
Positive integer, 6, 66
Postulate, 65, 423
 Power set, 4
Prime ideal, 323
Prime integer, 90
Prime polynomial, 378
Primitive
 n th root, 351
 polynomial, 390
Principal ideal, 296, 302
Principal of mathematical induction, 72
Product
 Cartesian, 13
 dot, 118
 external direct, 239
 of complex numbers, 334, 346
 of cosets, 298
 of matrices, 45
 of polynomials, 357
 of quaternions, 338, 339
 of rational numbers, 277
 of subsets, 215
 internal direct, 238
 notation, 148
Proof by contradiction, 430
Proper
 divisor of zero, 263
 subset, 3
Properties
 of addition in \mathbb{Z}_n , 107
 of multiplication in \mathbb{Z}_n , 108
Proposition, 423
Public Key Cryptosystem, 127
Pure imaginary number, 337

Quadratic formula, 385
Quantifier
 existential, 423
 universal, 423
Quaternion group, 143
Quaternions, 337
Quotient, 83, 369
 field, 279
 group, 230
 ring, 298, 410
 set, 278

Range, 14
Rank, 251
Rational numbers, 6, 328
Rational zeros, 388
Real axis, 343
Real numbers, 6, 328

Reducible polynomial, 378
Reflective symmetry, 211
Reflexive property, 55
Relation, 55
 antisymmetric, 61
 asymmetric, 61
 equivalence, 55
 inverse of, 61
 irreflexive, 61
 order, 68
Relatively prime
 integers, 89
 polynomials, 380
Remainder, 83, 369
 Theorem, 376
Repetition codes, 116
Resolvent equation, 403
Reverse Order Law, 54, 146, 151
Residue classes, 96
Right coset, 217
Right distributive law, 66, 67
Right ideal, 293
Right identity element, 50
Right inverse, 33
Rigid motion, 175, 201, 208
Ring, 257
 Boolean, 269, 318
 characteristic of, 313
 commutative, 261
 division, 337
 finite, 260
 homomorphism, 303
 of integers modulo n , 260
 isomorphism, 281, 303
 of polynomials over R , 359, 361
 quotient, 298
 with unity, 261
Root of a polynomial equation, 375
Rotational symmetry, 211
RSA cryptosystem, 127

Saucier, Morgan, 505
Second principle of finite induction, 75
Set
 of positive elements, 284
 of quotients, 278
 power, 4
Set-builder notation, 2
Sets, 1
 disjoint, 4
 empty, 4
 equal, 2
 intersection of, 3
 union of, 3
 universal, 4
Sigma notation, 47

Simple algebraic extension, 416
Skew field, 337
Solution, 375
Splitting field, 419
Square matrix, 43
Standard form
 of a complex number, 337
 of a positive integer, 92
Statement, 423
Strong mathematical induction, 75
Subgroup, 152
 conjugate, 221
 cyclic, 159
 generated by an element, 159
 generated by a subset, 225
 index of, 218
 invariant, 223
 nontrivial, 152
 normal, 223
 normalizer of, 229
 sum of, 239
 Sylow p -, 248
 transitive, 204, 222
 torsion, 174, 228, 238
 trivial, 152
Subring, 259
Subset, 2
 product of, 215
 proper, 3
Subtraction
 of integers, 70
 of matrices, 45
Sum
 of complex numbers, 334
 of cosets, 298
 direct, 269
 of ideals, 300
 of matrices, 43
 of polynomials, 357
 of quaternions, 338, 339
 of subgroups, 239
Surjective mapping, 15
Sylow p -subgroup, 248
Sylow's Theorem, 254
Symmetric
 group, 192
 property, 55
Symmetries, 175, 201, 208
Symmetry, 208
 geometric, 209
 reflective, 211
 rotational, 211

Terms of a polynomial, 356
Theorem, 423
Torsion subgroup, 174, 228, 238
Transformation, 12
Transitive
 property, 55
subgroup, 204, 222

- Translation, 211
cipher, 124
Transposition, 196
Trichotomy law, 66, 285
Trigonometric form of a complex number, 345
Triple repetition code, 116
Trivial ideal, 293
Trivial subgroup, 152
Truth table, 425
Two-sided identity, 50
Two-sided inverse, 33
- Union of sets, 3
Unique Factorization Theorem, 86, 91, 380
Unit, 262
Unity, 261
Universal quantifier, 423
Universal set, 4
UPC symbol, 119
Upper bound, 325
- Vector, 118, 343
Venn diagram, 4
- Well-ordered, 286
Well-Ordering Theorem, 81
Word, 114
- Zero
characteristic, 313
divisor, 109, 263
matrix, 45
of multiplicity m , 380
of a polynomial, 375, 380
of a ring, 258