Daniel W. Cunningham

A Logical Introduction to Proof

Springer

A Logical Introduction to Proof

Daniel W. Cunningham

A Logical Introduction to Proof

Springer

Daniel W. Cunningham
Mathematics Department
Buffalo State College
Buffalo, New York
USA

ISBN 978-1-4614-3630-0 ISBN 978-1-4614-3631-7 (eBook)
DOI 10.1007/978-1-4614-3631-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012939054

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

As an undergraduate, one of my most memorable and thrilling moments occurred when I first discovered how to prove an assigned theorem in one of my upper level mathematics courses. I also noticed that a few of my fellow students were having trouble proving this theorem. Even today, after being successful in calculus, many students have difficulty with proofs in their upper-division mathematics courses. To be successful in these more advanced courses, students must possess three essential skills: the ability to read, to understand, and to communicate in the language of mathematics. I wrote this book specifically to help students acquire these important skills and to enhance their ability to formulate and construct mathematical proofs.

When I was in college, what was it that helped me to discover and write mathematical proofs? Before enrolling in my first upper division mathematics course, I completed a beginning course in logic offered by the philosophy department. This course introduced me to formal proofs in a natural deduction system. During my first upper division mathematics course, I soon realized that I could apply the ideas that I learned in the logic course to help me write and find mathematical proofs. This book is intended to show students how basic logical principles can also help them to discover and compose mathematical proofs.

The core topics covered in the text are logic, sets, relations, functions and induction. Logic is covered, not as an end in itself, but as an instrument for analyzing the logical structure of mathematical assertions and as a tool for constructing valid mathematical proofs. I do not presume that the reader has an intuitive understanding of the principles of reasoning that many mathematicians take for granted.

Every theorem in mathematics must have a proof from a set of stated assumptions. The proof demonstrates that the conclusion of the theorem follows from the assumptions by the laws of logic alone. Can the notions of "laws of logic" and "proof" be made precise? This text is devoted to establishing a positive answer to this question. Students are given a plan of attack for finding and composing a correct proof of a given mathematical theorem. This is done by developing a method for

generating "proof diagrams."[1] For example, a proof of the statement $(\forall n \geq 1)\, P(n)$ by induction, uses the logical structure illustrated in the proof diagram

Prove $P(1)$.
Let $n \geq 1$ be an integer.
Assume $P(n)$.
Prove $P(n+1)$.

where indentation is used to display a proof's logical dependencies.

Diagramming a proof is a way of presenting the relationships between the various parts of a proof. The resulting proof diagram clearly demonstrates the structure of the proof and provides a tool for showing students how to write a correct mathematical proof. Analyzing a proof and portraying its *logical* structure with a consistent visual scheme can be helpful – both for proof beginners and for those trying to make sense of a proof at any level. Many students feel more confident using this well-structured approach for finding and writing proofs.

Each student of mathematics needs to learn how to find and write mathematical proofs. These are probably two of the most difficult skills that a mathematics major has to develop. Students often fail to construct a proof of a mathematical statement because they lack confidence or just do not know how to get started. This text is designed to increase students confidence by showing them how to first break up the statement into its logical parts and then use these parts as a guide for finding and writing a proof. Even with a guide, the work required to find a proof can be quite challenging. Professional mathematicians also have difficulty finding proofs; however, mathematicians know that persistence often pays off and thus, they do easily not give up.

> Patience and perseverance have a magical effect before which difficulties disappear and obstacles vanish. – John Quincy Adams
>
> Keep on going, and the chances are that you will stumble on something, perhaps when you are least expecting it. – Charles Kettering

What Is Covered?

The book is intended for students who want to learn how to prove theorems and be better prepared for the rigors required in their future mathematics courses. The first two chapters introduce students to logical connectives, truth tables, inference rules, deductions, quantifiers, variables, and truth sets. Chapter 1 focuses on propositional logic and Chapter 2 presents the logic of quantifiers. This initial emphasis on logic

[1] School children were once taught to diagram an English sentence as a means to analyze its grammatical structure.

is motivated only by the desire to show the reader how to discover, develop, and compose logically correct proofs.

Chapter 3 methodically presents the key strategies that are used in mathematical proofs. Each strategy is presented as a proof diagram and specifically responds to the logical form that a given mathematical statement may have. Furthermore, each proof strategy is carefully illustrated by a variety of mathematical theorems concerning the natural, rational and real numbers. The remaining chapters of the book presume the material presented in Chapter 3.

The attention in Chapter 4 is on proof by mathematical induction and concludes with a proof of the fundamental theorem of arithmetic.

Chapters 5–7 will introduce students to the essential concepts that appear in all branches of mathematics. Chapter 5 covers set theory and proofs about sets. In particular, I present the proof strategies that are used to prove set equality and to prove the subset relation. Functions are the main topic in Chapter 6, with a concentration on strategies for proving when functions are well-defined, one-to-one, and onto. With this foundation, Chapter 6 ends on the topic of countable and uncountable sets. Many of our proofs, on countability, presume the fundamental theorem of arithmetic. Chapter 7 presents equivalence relations, partitions, congruence modulo m relations, modular arithmetic, and partially ordered sets.

In the final two chapters, the objective is to better prepare students for abstract algebra and real analysis. Students will be introduced to some of the key topics that will be covered in their real analysis and abstract algebra courses; moreover, I do not just offer a preview of the basic concepts that will be addressed in these courses. The main goal is to give students some fundamental tools that will increase their likelihood of success.

A student's first course with a heavy emphasizes on proof is usually abstract algebra. The main aim of Chapter 8 is to prepare students for some of the important topics that will be covered in such a course. I first discuss algebraic structures and then move on to groups, subgroups, and normal subgroups. I provide proof strategies for dealing with these latter two concepts, as well. Using the results of Chapter 6 on one-to-one and onto functions, I also investigate permutation groups and the symmetric group. Chapter 8 also introduces rings and then ends on the topics of quotient algebras, quotient groups, and quotient rings. These final topics presume the material on equivalence relations and partitions covered in Chapter 7.

In real analysis, a facility for working with the supremum of a bounded set, the limit of a sequence, and the ε-δ definitions of continuity is essential for a student to be successful. Many students stumble when first asked to compose proofs using these core definitions. Chapter 9 is designed to better prepare students and allow them to overcome these initial hurdles. I present proof strategies that explicitly show students how to deal with the fundamental definitions that they will encounter in real analysis; followed by numerous examples of proofs that use these strategies.

Exercises are given at the end of each section in a chapter. Suggestions are also provided for those exercises that a newcomer to mathematical proof may find more challenging. The symbol Ⓢ marks the end of a solution, the symbol Ⓐ indicates the end of a proof analysis, and the symbol □ is used to identify the end of a proof.

Using This Text in a Transition Course

A standard transition course is designed to better prepare students for real analysis and abstract algebra. Such a course should cover Chapters 1–7. The basics of logic and proof are covered in Chapters 1–3. These first three chapters offer a basis for all of the material covered in the text. My experience has been that students easily grasp the topics covered in Chapters 1 and 2. Consequently, these first two chapters can usually be covered at a quick pace.

In Chapter 4 one could cover Sections 4.1–4.6 and then simply state the fundamental theorem of arithmetic, which is proven in Section 4.7. Furthermore, if students are already well versed in summation and factorial notation, Section 4.3 may be skipped. The topics covered in Chapters 5–6 are essential for any student of mathematics; however, Sections 5.4 (the axioms of set theory) and 6.5 (countable and uncountable sets) can be omitted as they are not used anywhere else in the book. Chapter 7 introduces equivalence relations, partitions, and congruence relations on the integers. Since the material in this chapter may be new to many students, these important topics should not be overlooked.

Chapters 8 and 9 are independent of each other. In Chapter 8, if time is limited, one could first introduce students to the notion of an algebraic structure and then focus their attention on groups by covering only Section 8.3, where proof strategies are presented that deal with subgroups and normal subgroups. Alternatively, after discussing algebraic structures, one could just introduce the ring concept by covering Section 8.5, as this section does not presume the group concept. Chapter 9 presents: (1) the supremum and infimum of a bounded set of real numbers, (2) the limit of a sequence, and (3) continuity of a function. Since these three topics are developed independently, one could discuss any combination of these concepts. In any case, it is hoped that students will view these final two chapters, and the entire book, as a useful resource in their future courses.

Acknowledgments

This book began as a set of notes for a course in discrete mathematics at Buffalo State College. The main focus of the course was to prepare students for upper division mathematics where they are expected to write and develop mathematical proofs. The two books that I used in this course were *Discrete Mathematics with Applications* by Epp [7] and *How to Prove It* by Velleman [12]. These excellent books inspired me to write my own book and their influence should be clear to anyone familiar with these textbooks.

I want to thank Marianne Foley, Michelle Schmitt, George Hole, Jon Lattanzio, and Tina Carter for their many helpful comments and corrections. The valuable recommendations made by the unnamed reviewers have allowed me to write a much better book and I am grateful to them all. Finally, I want to thank Elizabeth Loew, my editor at Springer, for her guidance and her patience.

Contents

Preface v

Acknowledgments ix

The Greek Alphabet xv

1 Propositional Logic 1
1.1 Logical Form and Logical Equivalence 1
1.1.1 Analyzing the Logical Form of English Statements 2
1.1.2 Truth Tables 4
1.1.3 Tautologies and Contradictions 7
1.1.4 Logical Equivalence 7
1.1.5 Propositional Logic Laws 9
1.1.6 Logic Laws and Substitution 10
1.2 The Conditional and Biconditional Connectives 12
1.2.1 Conditional Statements 12
1.2.2 Biconditional Statements 15
1.3 Valid and Invalid Arguments 19
1.3.1 Two Notorious Fallacies 22
1.3.2 Valid Arguments and Substitution 23
1.3.3 Inference Rules 24
1.3.4 Inference Rules and Substitution 25
1.3.5 Deductions 25

2 Predicate Logic 29
2.1 Variables, Predicates, and Truth Sets 29
2.1.1 Universe of Discourse 30
2.1.2 Sets Defined by a Predicate 30
2.1.3 Important Sets in Mathematics 31
2.2 Quantifiers 34
2.2.1 Analyzing the Logical Form of Statements 36
2.2.2 Bounded Quantifiers 39

2.3 Quantifiers and Negation 42
2.4 Statements Containing Multiple Quantifiers 46
2.4.1 Interpreting Adjacent Quantifiers 46
2.4.2 Interpreting Non-adjacent Quantifiers 49
2.4.3 Translating English Statements with Multiple Quantifiers ... 51
2.4.4 Negating Statements with More than One Quantifier 53
2.4.5 The Uniqueness Quantifier 54
2.5 Valid and Invalid Arguments 58

3 Proof Strategies and Diagrams 61
3.1 Conjecture and Proof in Mathematics 61
3.1.1 How to Prove an Algebraic Equation 63
3.1.2 How to Prove an Inequality 64
3.2 Using Proof Diagrams as Guides for Proving Theorems 66
3.3 Statements of the Form $P \to Q$ 67
3.4 Statements of the Form $\forall x P(x)$ and $\exists x P(x)$ 70
3.4.1 Working with Universal Statements 71
3.4.2 Working with Existential Statements 72
3.4.3 Working with Mixed Quantifier Statements 77
3.4.4 Uniqueness Proofs 78
3.5 Statements of the Form $P \wedge Q$ 80
3.6 Statements of the Form $P \vee Q$ 84
3.7 Statements of the Form $P \leftrightarrow Q$ 89
3.8 Indirect Proof 91
3.8.1 Proof by Contraposition 91
3.8.2 Proof by Contradiction 93

4 Mathematical Induction 99
4.1 The Well-Ordering Principle 99
4.2 Proof by Mathematical Induction 102
4.3 Sequences, Sums, and Factorials 106
4.3.1 Summation Notation 108
4.3.2 Evaluating Sums 109
4.3.3 Factorial Notation 113
4.4 Proving Equations by Mathematical Induction 116
4.5 More Proofs by Mathematical Induction 123
4.5.1 Recursive (Inductive) Definitions 124
4.6 Strong Mathematical Induction 127
4.6.1 Strong Induction with One Base Step 128
4.6.2 Strong Induction with Multiple Base Steps 131
4.7 Fundamental Theorem of Arithmetic 138

5 Set Theory 143
5.1 Basic Definitions of Set Theory 143
5.1.1 Set Operations 144

5.1.2 Cartesian Products 147
5.1.3 Partitions 147
5.2 Proofs in Set Theory 150
5.2.1 Strategy for Proving a Subset Relation 150
5.2.2 Strategies for Proving Set Equality 151
5.3 Indexed Families of Sets 156
5.3.1 Generalized Unions and Intersections 157
5.3.2 Unindexed Families of Sets 162
5.4 The Axioms of Set Theory 163
5.4.1 The Zermelo-Fraenkel Axioms 164
5.4.2 The Axiom of Choice 166

6 Functions 169
6.1 Functions Defined on General Sets 169
6.1.1 Is it a Function? 171
6.1.2 The Range of a Function 174
6.1.3 Equality of Functions 174
6.2 One-to-One, Onto, and Inverse Functions 177
6.2.1 One-to-One Functions 177
6.2.2 Onto Functions 179
6.2.3 Inverse Functions 181
6.3 Composition of Functions 184
6.3.1 Composing a Function with the Identity Function 185
6.3.2 Composing a Function with Its Inverse 186
6.3.3 Composing One-to-One Functions 186
6.3.4 Composing onto Functions 187
6.4 Functions Acting on Sets 189
6.5 On the Size of Infinite Sets 193
6.5.1 Countable Sets 195
6.5.2 Uncountable Sets 199
6.5.3 Cardinality 203

7 Relations 209
7.1 Relations on a Set 209
7.1.1 Reflexive, Symmetric, and Transitive Relations 210
7.2 Equivalence Relations and Partitions 214
7.3 Congruence Modulo m 218
7.3.1 Fundamental Properties 219
7.3.2 Congruence Classes 221
7.4 Modular Arithmetic 225
7.5 Order Relations 232

8 Core Concepts in Abstract Algebra 239
8.1 Binary Operations 239
8.1.1 Is it a Binary Operation? 241

8.2 Algebraic Structures 245
8.2.1 Substructures 246
8.3 Groups 247
8.3.1 Fundamental Properties of a Group 251
8.3.2 Subgroups 252
8.3.3 Normal Subgroups 255
8.3.4 The Order of an Element in a Group 257
8.4 Permutation Groups 260
8.4.1 The Symmetric Group 262
8.4.2 Cycle Products 264
8.4.3 Cycle Decomposition 266
8.5 Rings 273
8.5.1 Fundamental Properties of Rings 275
8.5.2 Subrings 276
8.5.3 Ideals 277
8.6 Quotient Algebras 281
8.6.1 Quotient Groups 284
8.6.2 Quotient Rings 287

9 Core Concepts in Real Analysis 293
9.1 Fields 293
9.1.1 Ordered Fields 295
9.2 The Real Field 298
9.3 The Completeness Axiom 300
9.3.1 Proofs on the Supremum of a Set 302
9.3.2 Proofs on the Infimum of a Set 303
9.3.3 Bounded Functions 306
9.3.4 Alternative Proof Strategies 307
9.4 Convergence of Sequences 311
9.4.1 Bounded Sequences 321
9.5 Limit Theorems for Sequences 323
9.6 Continuous Functions 329
9.6.1 Algebraic Operations on Functions 332
9.6.2 Preservation-of-Continuity Theorems 333

A Summary of Strategies 341

References 347

Index of Special Symbols 349

Index 351

The Greek Alphabet

A	α		alpha
B	β		beta
Γ	γ		gamma
Δ	δ		delta
E	ϵ	ε	epsilon
Z	ζ		zeta (“zay tuh”)[1]
H	η		eta (“ay tuh”)
Θ	θ	ϑ	theta (“thay tuh”)[2]
I	ι		iota
K	κ		kappa
Λ	λ		lambda
M	μ		mu
N	ν		nu
O	o		omicron
Ξ	ξ		xi (“zī”)[3]
Π	π		pi
P	ρ		rho (“row”)
Σ	σ		sigma
T	τ		tau
Υ	υ		upsilon
Φ	ϕ	φ	phi (“fī”)[4]
X	χ		chi (“kī”)
Ψ	ψ		psi (“sī”)
Ω	ω		omega

[1] “ay” is pronounced as in “say”.

[2] “th” is pronounced as in “thing”.

[3] ī is pronounced as i in hi.

[4] many people say “fee”.

CHAPTER 1

Propositional Logic

A mathematician establishes the truth of a mathematical statement by providing a proof. Such a proof often uses principles of reasoning that are best described within propositional logic. In this chapter we examine the basic tools in propositional logic that mathematicians use to demonstrate that their conclusions are valid. What is a proposition? A *proposition* is a declarative sentence or assertion that is either true or false. Here are two such propositions:

(1) Global warming is a serious problem.
(2) Paris is in France.

Propositional logic studies the results of combining propositions to form more complex statements. In particular, the following three sentences each contain the above two propositions (1) and (2) as components.

Global warming is a serious problem and Paris is in France.
Global warming is a serious problem or Paris is in France.
If Paris is in France, then global warming is not a serious problem.

We will pursue the meaning of assertions that are obtained by connecting statements using "and," "or," "not," "if–then," and "if and only if." These five expressions are frequently used in mathematics. For example, consider the following three mathematical statements:

$x \geq 2$ or $x \leq 2$,
if $x \geq 3$, then $x > 1$,
$x \geq 2$ if and only if $x + 5 \geq 7$,

where x is a real number. Are these three statements true for all real numbers x? If so, then they must be true for $x = 3$, $x = 2$, and for $x = 1$. We will address such issues here in this chapter.

1.1 Logical Form and Logical Equivalence

Symbols play a critical role in mathematics and science. When discussing the logic of propositional statements, we shall use symbols to represent these statements. Capital letters, for instance P, Q, R, will stand for propositional statements, or *propositional components*. As an example, we could use the letter G to represent the component "Global warming is a serious problem" and use the letter P to denote the component "Paris is in France."

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_1,

We shall identify symbols for each of the English connectors "and," "or," "not." We will use the symbol $\wedge$ to represent "and," the symbol $\vee$ to represent "or," and use the symbol $\neg$ to represent "not." In addition, the symbol $\rightarrow$ denotes the word "implies" and the symbol $\leftrightarrow$ represents the phrase "if and only if." The five symbols $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$ are called *logical connectives*. Using these logical connectives, we will be able to analyze the logical structure of an English sentence. For instance, the logical form of the sentence "Global warming is a serious problem and Paris is in France" can now be expressed by $G \wedge P$.

In this section we shall focus on the logical meaning of $\wedge, \vee, \neg$. We will then explore the connectives $\rightarrow$ and $\leftrightarrow$ in Section 1.2. We shall refer to the three logical connectives $\wedge, \vee, \neg$ as *conjunction, disjunction*, and *negation*, respectively. Given a list of propositional components $A, B, C, \ldots$, and the logical connectives $\wedge, \vee, \neg$, we can form *propositional sentences*. For example,

1. $P \wedge Q$ (means "P and Q").
2. $P \vee Q$ (means "P or Q").
3. $\neg P$ (means "not P" or "it is not the case that P").

Using propositional components as building blocks and the connectives as mortar, one can construct more complicated propositional sentences, for example,

$$(P \wedge \neg Q) \vee (\neg S \wedge R).$$

It is important to use parentheses so that our propositional sentences are clear and readable; however, we shall use the two conventions:

1. The *outermost* parentheses need not be explicitly written. Thus, we can write $A \wedge B$ to denote $(A \wedge B)$.
2. The negation symbol shall apply to as little as possible. We can therefore write $\neg A \wedge B$ to denote $(\neg A) \wedge B$.

1.1.1 Analyzing the Logical Form of English Statements

Virtually every English statement can be expressed as a propositional sentence. All one has to do is first identify the propositional components that appear in the English sentence and then identify the logical connectives that also appear in the sentence. There are times when these logical connectives are not explicitly stated and may be somewhat hidden by the words used in the sentence. We will see sentences that contain "hidden connectives" in our next example.

Example 1. Analyze the logical form of the following seven English statements. In other words write each statement symbolically and thereby reveal any hidden logical connectives.

1. Emily writes poetry and James doesn't write poetry.
2. Either Emily writes poetry and James doesn't, or James writes poetry and Emily doesn't.
3. Dan and Mary are both correct.
4. Dan and Mary are both not correct.
5. Dan and Mary are not both correct.
6. Neither Dan is correct nor Mary is correct.
7. Either Dan and Mary are both correct or neither of them is correct.

Solution. First we identify and symbolize the propositional components occurring in each English statement. Then we will express the English statement in logical form.

1. *Emily writes poetry and James doesn't write poetry.*

   Let E represent "Emily writes poetry." The statement "James doesn't write poetry" is a short form for the sentence "James does not write poetry." Let J represent "James does write poetry." Then the logical form of the English statement is $E \wedge \neg J$.

2. *Either Emily writes poetry and James doesn't, or James writes poetry and Emily doesn't.*

   The word 'either' can be thought of as a warning that an 'or' is coming. The expression 'either X or Y' means that X is true or Y is true. So, the given English sentence can be expressed as

   (Emily writes poetry and James doesn't) or
   (James writes poetry and Emily doesn't).

   Using the propositions E and J given in our solution to item 1, the logical form of the English sentence can be expressed as $(E \wedge \neg J) \vee (J \wedge \neg E)$.

3. *Dan and Mary are both correct.*

   Let D represent "Dan is correct" and let M represent "Mary is correct." Thus, the logical form of the given statement is $D \wedge M$.

4. *Dan and Mary are both not correct.*

   In this sentence, the expression "both not" means that Dan is not correct and Mary is not correct. Let D and M be the propositions used in our solution to item 3. We conclude that the logical form of the sentence is $\neg D \wedge \neg M$.

5. *Dan and Mary are not both correct.*

   The expression "not both," in the above sentence, means that it is not the case that Dan and Mary are both correct. Let D and M be the propositions used in our solution to item 3. Consequently, the logical form of the sentence is $\neg(D \wedge M)$.

6. *Neither Dan is correct nor Mary is correct.*

   The word 'neither' can be thought of as a warning that a 'nor' is coming. The expression 'neither X nor Y' means that X is false and Y is also false. Let D and M be the propositions used in our solution to item 3. Then the logical form of the English sentence in item 6 is $\neg D \wedge \neg M$.

7. *Either Dan and Mary are both correct or neither of them is correct.*

   Since this English sentence begins with an 'either', it can be expressed as

   (Dan and Mary are both correct) or (neither of them is correct).

   Let D and M be as in our solution to item 3. The logical form of the English sentence in item 7 is given by $(D \wedge M) \vee (\neg D \wedge \neg M)$.

This completes our solution. Ⓢ

As stated in the preface, we use the symbol Ⓢ to mark the end of a solution.

Example 2. Analyze the logical forms of the following mathematical statements using only the mathematical relations "$<$" and "$=$" (look out for any possible hidden logical connectives).

1. $x \leq 4$.
2. $\sqrt{3} \not< 4$.
3. $1 \leq x \leq 4$.

Solution.

1. The statement $x \leq 4$ means that x is less than or equal to 4. So there is a hidden 'or' in this statement. The logical form of $x \leq 4$ can be expressed by $x < 4 \vee x = 4$.
2. The statement $\sqrt{3} \not< 4$ means that $\sqrt{3}$ is not less than 4. Therefore, the logical form of $\sqrt{3} \not< 4$ is $\neg(\sqrt{3} < 4)$.
3. Finally, the statement $1 \leq x \leq 4$ means that $1 \leq x$ and $x \leq 4$. So this statement contains the hidden connective 'and.' The assertion $1 \leq x \leq 4$ has the logical form $(1 < x \vee 1 = x) \wedge (x < 4 \vee x = 4)$. Ⓢ

1.1.2 Truth Tables

Given a collection of propositional components, say P, Q, and R, we can assign truth values to these components. For example, we can assign the truth values of P, Q, R to be T, F, T respectively, where T means "true" while F means "false." The truth value of a sentence in propositional logic can be evaluated from the truth values assigned to its components. We shall explain what this "means" by using truth tables. The logical connectives $\wedge$, $\vee$, $\neg$ yield the natural truth values given by the following three truth tables, respectively, in Table 1.1.

Truth table (1) has four rows (not including the header). The columns beneath P and Q list all the possible pairs of truth values that can be assigned to the components P and Q. For each such pair, the corresponding truth value for $P \wedge Q$ appears to the right. For example, consider the second pair of truth values in this table, T F. Thus, when the propositional components P and Q are assigned the respective truth values T and F, we see that the truth value of $P \wedge Q$ is F.

Truth table (2) asserts that when P and Q are assigned the respective truth values F and T, then the truth value of $P \vee Q$ is T. Furthermore, when P and Q are assigned

Table 1.1 Basic truth tables

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

(1) Conjunction

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

(2) Disjunction

P	$\neg P$
T	F
F	T

(3) Negation

the truth values T and T, then the truth value of $P \vee Q$ is also T. In mathematics, the connective "or" has the same meaning as "and/or," that is, $P \vee Q$ is true if either P is true *or* Q is true, or both P *and* Q are true. Thus, the assertion "$x \geq 2$ or $x \leq 2$" is true when $x = 1$, $x = 2$, or when $x = 3$. Our truth table for $P \vee Q$ reflects the fact that we are working in mathematics. (The word "or" in every day English sometimes excludes the possibility that P and Q could both be true.) Finally, the truth table (3) shows that the negation of a statement reverses the truth value of the statement.

Now that we know how to build truth tables for the sentences $P \wedge Q$, $P \vee Q$, and $\neg P$, we will discuss how to build truth tables for more complicated propositional sentences such as $(P \vee R) \wedge (\neg Q \wedge S)$.

Constructing Truth Tables for More Complicated Sentences

In Table 1.1, truth table (2) allows us to determine the truth value of $P \vee Q$ whenever we know the truth value of P and Q. In other words, the truth value of $P \vee Q$ is presented as a function of the truth values assigned to the components P and Q. In this section, we illustrate a method that will allow us to construct a truth table for any propositional sentence.

Given a propositional sentence one can identify the "outside" connective, that is, the "last connective one needs to evaluate." When the outside connective in a propositional sentence has been identified, one can then break up the sentence into its "parts." For example, in the propositional sentence $\neg P \vee (Q \wedge P)$ the logical connective $\vee$ is the outside connective with parts $\neg P$ and $Q \wedge P$. For another example, consider the propositional sentence $\neg(P \vee (Q \wedge P))$. We see that $\neg$ is the outside connective with corresponding part $P \vee (Q \wedge P)$.

Example 3. Construct a truth table that can be used to evaluate the truth value of the sentence $\neg P \vee (Q \wedge P)$ as a function of the truth values assigned to its components P and Q.

Solution. Of course, the components P and Q will each need a column in our truth table. Since there are two components, there will be four possible combinations of truth values for P and Q. We will enter these combinations in the two left-most columns in the same order as that in Table 1.1(1). The outside connective of the propositional sentence $\neg P \vee (Q \wedge P)$ is $\vee$. We can break this sentence into two parts $\neg P$ and $Q \wedge P$. So these parts will also need a column in our truth table. Since we

can only break the sentences $\neg P$ and $Q \wedge P$ into components (namely, P and Q), we obtain the following truth table:

	P	Q	$\neg P$	$Q \wedge P$	$\neg P \vee (Q \wedge P)$
	T	T	F	T	T
	T	F	F	F	F
	F	T	T	F	T
	F	F	T	F	T
STEP #	1	1	2	3	4

We now describe in steps how we obtained the truth values in the above table. STEP 1: Specify all of the possible truth values that can be assigned to the components (resulting in four rows of truth values). STEP 2: In each row, use the truth value assigned to the component P to obtain the corresponding truth value for $\neg P$ by applying Table 1.1(3). STEP 3: In each row, use the truth values assigned to Q and P, to determine the corresponding truth value in the column under $Q \wedge P$, using Table 1.1(1). STEP 4: In each row, use the truth values in the columns under $\neg P$ and $Q \wedge P$ to obtain the matching truth value for the final column under the sentence $\neg P \vee (Q \wedge P)$ by employing Table 1.1(2). Ⓢ

In the construction of the above truth table, observe that whenever we broke up a sentence into parts, the columns for the parts appear to the left of the column for the sentence. We will do this again in our next example.

Example 4. Construct a truth table that can be used to evaluate the truth value of the sentence $P \wedge (Q \vee \neg R)$ as function of the truth values of the components P, Q, R.

Solution. We know that the components P, Q and R will each need a column in our truth table. Since there are three components, there will be eight possible truth value combinations for P, Q and R. The outside connective of the propositional sentence $P \wedge (Q \vee \neg R)$ is $\wedge$. We can break this sentence into two parts P and $Q \vee \neg R$. Since $Q \vee \neg R$ is not a component, it will need a column in our truth table. We now break up $Q \vee \neg R$ into the parts Q and $\neg R$. Because $\neg R$ is not a component, it will also require a column in our truth table. Thus, our desired truth table for $P \wedge (Q \vee \neg R)$ is

	P	Q	R	$\neg R$	$Q \vee \neg R$	$P \wedge (Q \vee \neg R)$
	T	T	T	F	T	T
	T	T	F	T	T	T
	T	F	T	F	F	F
	T	F	F	T	T	T
	F	T	T	F	T	F
	F	T	F	T	T	F
	F	F	T	F	F	F
	F	F	F	T	T	F
STEP #	1	1	1	2	3	4

We will now identify the steps that we used to obtain the truth values in the above table. STEP 1: Specify all of the possible truth values that can be assigned to the components (resulting in eight rows of truth values). STEP 2: In each row, use the truth value assigned to the component R to obtain the corresponding truth value for $\neg R$ by applying Table 1.1(3). STEP 3: In each row, use the truth values assigned to Q and $\neg R$, to determine the corresponding truth value in the column under $Q \vee \neg R$, using Table 1.1(2). STEP 4: In each row, use the truth values in the columns under P and $Q \vee \neg R$ to obtain the matching truth value for the final column under the sentence $P \wedge (Q \vee \neg R)$. Ⓢ

1.1.3 Tautologies and Contradictions

Suppose, after constructing a truth table for a propositional sentence, you see that each entry in the final column is true. This indicates a situation where the sentence is true no matter what truth values are assigned to its components. When this occurs, the sentence is called a tautology.

Definition 1.1.1. We shall say that a propositional sentence is a **tautology** when its truth value is true regardless of the truth values of its components.

Thus, a propositional sentence is a tautology if it is always true. For example, one can see from the following truth table that the sentence $P \vee \neg P$ is a tautology.

P	$\neg P$	$P \vee \neg P$
T	F	T
F	T	T

Definition 1.1.2. We shall say that a propositional sentence is a **contradiction** when its truth value is false regardless of the truth values of its components.

In other words, a propositional sentence is a contradiction if it is always false. One can easily show that the sentence $P \wedge \neg P$ is a contradiction.

1.1.4 Logical Equivalence

The following definition describes when two propositional sentences are logically equivalent, that is, when they mean the same thing. Mathematicians frequently take advantage of logical equivalence to simplify their proofs and we shall do the same in this book. We will use Greek letters (e.g., α, β, φ and ψ – see page xv) to represent propositional sentences.

Definition 1.1.3. Let ψ and φ be two sentences of propositional logic. We say that ψ and φ are **logically equivalent**, denoted by $\psi \Leftrightarrow \varphi$, when the following holds:

For every truth assignment applied to the components of ψ and φ, the resulting truth values of ψ and φ are identical.

Let φ and ψ be propositional sentences with the same components. Construct truth tables for φ and ψ so that each component has the same column in both tables. Suppose that these two truth tables also have the same final column. We can then conclude that φ and ψ are logically equivalent. Thus, φ and ψ are "both true at the same time and both false at the same time."

Example 5. Let ψ be the sentence $\neg(P \vee Q)$ and let φ be the sentence $\neg P \wedge \neg Q$. Show that ψ and φ are logically equivalent, that is, show $\neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$.

Solution. After constructing individual truth tables for the statements $\neg(P \vee Q)$ and $\neg P \wedge \neg Q$, we obtain

P	Q	$P \vee Q$	$\neg(P \vee Q)$
T	T	T	F
T	F	T	F
F	T	T	F
F	F	F	T

P	Q	$\neg P$	$\neg Q$	$\neg P \wedge \neg Q$
T	T	F	F	F
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

So each truth assignment applied to the components P and Q yields the same truth value for $\neg(P \vee Q)$ and $\neg P \wedge \neg Q$. Therefore, we have that $\neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$. In other words, since the final columns of the truth tables for $\neg(P \vee Q)$ and $\neg P \wedge \neg Q$ are the identical, we can conclude that they are logically equivalent. Ⓢ

When φ and ψ are logically equivalent, we will say that $\psi \Leftrightarrow \varphi$ is a *logic law*. We now present two important logic laws that are often used in mathematical proofs. These laws were first identified by Augustus De Morgan (see Example 5).

De Morgan's Laws (DML)

1. $\neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$.
2. $\neg(P \wedge Q) \Leftrightarrow \neg P \vee \neg Q$.

Let ψ and φ be two sentences of propositional logic. If one can apply a truth assignment to the components of ψ and φ so that the resulting truth values of ψ and φ disagree, then ψ and φ are *not logically equivalent*. We will use this fact in our next example which shows that the placement of parentheses in a propositional sentence is very important. A regrouping can change the meaning of the sentence.

Example 6. Show that $P \wedge (Q \vee R)$ and $(P \wedge Q) \vee R$ are not logically equivalent.

Solution. We shall use the truth table

P	Q	R	$P \wedge (Q \vee R)$	$(P \wedge Q) \vee R$
T	T	T	T	T
T	T	F	T	T
T	F	T	T	T
T	F	F	F	F
F	T	T	F	T
F	T	F	F	F
F	F	T	F	T
F	F	F	F	F

Since their final columns are not the identical, we see that $P \wedge (Q \vee R)$ and $(P \wedge Q) \vee R$ are not equivalent. In particular, the truth assignment to the components in row 5 yields different truth values for $P \wedge (Q \vee R)$ and $(P \wedge Q) \vee R$. Ⓢ

1.1.5 Propositional Logic Laws

Propositional logic will be used as a tool to help us develop both the structure and the presentation of mathematical proofs. Listed below are the important laws of logic that will allow us to simplify more complicated propositional sentences and to streamline the presentation of some mathematical proofs. In Section 1.1.6, we will also use these logic laws to derive new logic laws without the use of truth tables.

De Morgan's Laws (DML)

1. $\neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$.
2. $\neg(P \wedge Q) \Leftrightarrow \neg P \vee \neg Q$.

Commutative Laws

1. $P \wedge Q \Leftrightarrow Q \wedge P$.
2. $P \vee Q \Leftrightarrow Q \vee P$.

Associative Laws

1. $P \vee (Q \vee R) \Leftrightarrow (P \vee Q) \vee R$.
2. $P \wedge (Q \wedge R) \Leftrightarrow (P \wedge Q) \wedge R$.

Idempotent Laws

1. $P \wedge P \Leftrightarrow P$.
2. $P \vee P \Leftrightarrow P$.

Distributive Laws

1. $P \wedge (Q \vee R) \Leftrightarrow (P \wedge Q) \vee (P \wedge R)$.
2. $P \vee (Q \wedge R) \Leftrightarrow (P \vee Q) \wedge (P \vee R)$.

3. $(Q \vee R) \wedge P \Leftrightarrow (Q \wedge P) \vee (R \wedge P)$.
4. $(Q \wedge R) \vee P \Leftrightarrow (Q \vee P) \wedge (R \vee P)$.

Double Negation Law (DNL)

1. $\neg\neg P \Leftrightarrow P$.

Tautology Law

1. $P \wedge (\text{a tautology}) \Leftrightarrow P$.

Contradiction Law

1. $P \vee (\text{a contradiction}) \Leftrightarrow P$.

We now give examples of the above Tautology Law and Contradiction Law. First recall that $Q \vee \neg Q$ is a tautology. From the Tautology Law we obtain the following logical equivalence

$$P \wedge (Q \vee \neg Q) \Leftrightarrow P.$$

On the other hand, because $Q \wedge \neg Q$ is a contradiction, we conclude that

$$P \vee (Q \wedge \neg Q) \Leftrightarrow P$$

by the Contradiction Law.

1.1.6 Logic Laws and Substitution

Consider the algebraic identity $(x-y)(x+y) = x^2 - y^2$. If we replace x with ab, then we obtain another algebraic identity $(ab-y)(ab+y) = (ab)^2 - y^2$. Similarly, if a propositional component appears in a logic law and we replace all occurrences of this component with a propositional sentence, then we will obtain another logic law. For example, consider the distributive law

$$P \vee (Q \wedge R) \Leftrightarrow (P \vee Q) \wedge (P \vee R).$$

Let us replace P with $\neg A$ and replace Q with a propositional sentence ψ. We then obtain the following logical equivalence, which is an application of the Distributive Law:

$$\neg A \vee (\psi \wedge R) \Leftrightarrow (\neg A \vee \psi) \wedge (\neg A \vee R).$$

In addition, let α and β be propositional sentences that are logically equivalent, that is, $\alpha \Leftrightarrow \beta$. If α appears in a given propositional sentence Θ and we replace occurrences of α in Θ with β, then the resulting new sentence will be logically equivalent to Θ. For example, consider the sentence Θ given by $\neg Q \vee \alpha$ and suppose that $\alpha \Leftrightarrow \beta$. Upon replacing α with β, we obtain the new sentence $\neg Q \vee \beta$. We can conclude that $(\neg Q \vee \alpha) \Leftrightarrow (\neg Q \vee \beta)$, because $\alpha \Leftrightarrow \beta$.

Example 7. Using logic laws, find a simpler sentence equivalent to the formula $\neg Q \vee \neg(\neg P \vee \neg Q)$.

Solution. We start with $\neg Q \vee \neg(\neg P \vee \neg Q)$ and apply logic laws as follows:

$$\begin{aligned} \neg Q \vee \neg(\neg P \vee \neg Q) &\Leftrightarrow \neg Q \vee (\neg\neg P \wedge \neg\neg Q) && \text{by De Morgan's Law} \\ &\Leftrightarrow \neg Q \vee (P \wedge Q) && \text{by Double Negation Law} \\ &\Leftrightarrow (\neg Q \vee P) \wedge (\neg Q \vee Q) && \text{by Distributive Law} \\ &\Leftrightarrow (\neg Q \vee P) && \text{by Tautology Law.} \end{aligned}$$

Therefore, $\neg Q \vee \neg(\neg P \vee \neg Q) \Leftrightarrow \neg Q \vee P$. Thus, $\neg Q \vee P$ is a simplified version of $\neg Q \vee \neg(\neg P \vee \neg Q)$. Ⓢ

Example 8. Using propositional logic laws, show that $\neg(P \wedge \neg Q) \Leftrightarrow (\neg P \vee Q)$.

Solution. We start with the more complicated side $\neg(P \wedge \neg Q)$ and derive the simpler side as follows:

$$\begin{aligned} \neg(P \wedge \neg Q) &\Leftrightarrow (\neg P \vee \neg\neg Q) && \text{by De Morgan's Law} \\ &\Leftrightarrow (\neg P \vee Q) && \text{by Double Negation Law.} \end{aligned}$$

Therefore, $\neg(P \wedge \neg Q) \Leftrightarrow (\neg P \vee Q)$. Ⓢ

Exercises 1.1

1. Only one of the following is a tautology. Which one is it?

(a) $(P \vee \neg P) \wedge Q$.
(b) $(P \vee \neg P) \vee Q$.

2. Using one of De Morgan's Laws, write a negation of the statement: *Ron runs on Thursdays and Pete plays poker on Saturdays.* Express you answer in English.

3. Use one of De Morgan's Laws to write a negation, in English, of the statement: *My computer program has an error or the wrong value is assigned to a constant.*

4. Using propositional logic laws (see Section 1.1.5), supply a law justifying each step:

$$\begin{aligned} (P \vee \neg Q) \wedge (\neg P \vee \neg Q) &\Leftrightarrow (\neg Q \vee P) \wedge (\neg Q \vee \neg P) && \text{by } \_\_\_\_\_ \\ &\Leftrightarrow \neg Q \vee (\neg P \wedge P) && \text{by } \_\_\_\_\_ \\ &\Leftrightarrow \neg Q && \text{by } \_\_\_\_\_ \end{aligned}$$

5. Using the propositional logic laws in Section 1.1.5, find simpler sentences (see Example 7) that are equivalent to the following:

(a) $\neg(\neg P \wedge \neg Q)$.
(b) $\neg Q \wedge \neg(\neg P \wedge \neg Q)$.
(c) $\neg(\neg P \vee Q) \vee (P \wedge \neg R)$.

6. Which of the following statements are true and which are false?

(a) $(\pi^2 > 9) \wedge (\pi > 3)$.
(b) $(\pi^2 > 9) \vee (\pi > 3)$.
(c) $(\sin(2\pi) > 9) \vee (\sin(2\pi) < 0)$.
(d) $(\sin(\pi) > 9) \vee \neg(\sin(\pi) \leq 0)$.

7. Using truth tables, show that $(\neg P \vee Q) \vee (P \wedge \neg Q)$ is a tautology. What can you conclude about the sentence $\neg((\neg P \vee Q) \vee (P \wedge \neg Q))$?

8. Using propositional logic laws, show that $P \vee (Q \wedge \neg P) \Leftrightarrow P \vee Q$.

9. Using logic laws, show that $\neg(P \vee \neg Q) \vee (\neg P \wedge \neg Q) \Leftrightarrow \neg P$.

1.2 The Conditional and Biconditional Connectives

1.2.1 Conditional Statements

Many mathematical theorems have the form "if P, then Q" or, equivalently, "P implies Q." Here is one important example that you may have seen in your calculus course:

Theorem. *If f is differentiable at the point a, then f is continuous at a.*

Let D be the proposition "f is differentiable at the point a" and let C be the proposition "f is continuous at a." The theorem can now be expressed as

Theorem. *If D, then C.*

A *conditional statement* has the form "if P, then Q." The statement P is called the **hypothesis** and the statement Q is called the **conclusion**. Thus, a conditional statement asserts that the truth of the hypothesis "implies" the truth of the conclusion. This is such an important idea in mathematics that we will now introduce a logical connective which will capture the mathematical notion that the hypothesis implies the conclusion.

The Conditional Connective. Given propositions P and Q, the conditional connective $\rightarrow$ means "implies" and can be used to form the sentence $P \rightarrow Q$. The sentence $P \rightarrow Q$ can be read as "P implies Q" or "if P, then Q."

Question. At the beginning of a semester course, your teacher tells you:

"*If* you do your homework, *then* you will pass."

Your teacher, however, did not say anything about what would happen if you did not do your homework. After the final exam, you obtain your course grade. You may have done very well, or not so well, on the final exam. Under what conditions can you call your teacher a liar? A liar is one who clearly and knowingly tells a lie, that is, a falsehood. You can declare your teacher as being guilty of a lie only when there is clear evidence of such guilt. Without such evidence, we must presume that your teacher told the truth.

Let us define some propositional symbols that concisely express the teacher's statement. Let H stand for the assertion "you do your homework" and let P represent the statement "you will pass." Thus, $H \to P$ represents your teacher's statement, that is, "If you do your homework, then you will pass." Consider the truth table

H	P	$H \to P$
T	T	T
T	F	F
F	T	T
F	F	T

We will now give an explanation for the values in the final column of this truth table by starting with the first pair of truth values for H and P, and then continuing down until we reach the last such pair. For the first pair of truth values T and T, we see that if you did your homework and you passed the course, then your teacher told the truth (T). On the other hand, for the second pair of truth values T and F, we see that if you did your homework and you failed the course, then your teacher is a liar (F). Now, suppose that you did not do your homework, as in the third and fourth pair of truth values. Remember that your teacher did not tell you what would happen in this case. Consequently, you cannot call her a liar. So, if you did not do your homework, the conditional statement $H \to P$ must be viewed as being true (T), whether you passed or did not pass the course.[1] This is how conditional statements are viewed in mathematics and in courtrooms.

Conditional Truth Table. Given propositions P and Q, the sentence $P \to Q$ has the truth table

P	Q	$P \to Q$
T	T	T
T	F	F
F	T	T
F	F	T

[1] If you did very well on the final exam, then that would explain why you passed the class. If you failed the final exam, then that could account for failing the course.

The conditional $P \to Q$ can be read in several ways:

1. "If P, then Q."
2. "P implies Q."
3. "Q, if P."
4. "P only if Q."
5. "P is a sufficient condition for Q."
6. "Q is a necessary condition for P."

Beginners to logic and mathematics often have trouble with item 4. To avoid this trouble, one should interpret the expression "only if" as the arrow $\to$. Item 5 can be interpreted as stating "the truth of P is sufficient to guarantee the truth of Q." Item 6 can be interpreted as stating "the truth of Q is guaranteed, given that P is true." The words *sufficient* and *necessary* can also be confusing. One way to avoid this confusion is to think of the word "sufficient" as the arrow $\to$, and to think of the word "necessary" as the backward arrow $\leftarrow$.

Our next three logic laws involve conditional statements. The first law states that a conditional statement is equivalent to one that contains the connectives $\neg$ and $\vee$.

Conditional Laws

1. $(P \to Q) \Leftrightarrow (\neg P \vee Q)$.
2. $(P \to Q) \Leftrightarrow \neg(P \wedge \neg Q)$.
3. $\neg(P \to Q) \Leftrightarrow (P \wedge \neg Q)$.

Proof of Conditional Laws. We show that items 1 and 2 hold, by comparing the following truth tables

P	Q	$P \to Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	Q	$\neg P \vee Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	Q	$\neg(P \wedge \neg Q)$
T	T	T
T	F	F
F	T	T
F	F	T

Since all of the final columns agree, we see that $(P \to Q)$, $(\neg P \vee Q)$, and $\neg(P \wedge \neg Q)$ are logically equivalent. By constructing truth tables for $\neg(P \to Q)$ and $(P \wedge \neg Q)$, one can also show that $\neg(P \to Q) \Leftrightarrow (P \wedge \neg Q)$ (see Exercise 1). □

Definition 1.2.1. The **contrapositive** of a conditional statement $P \to Q$ is the conditional $\neg Q \to \neg P$.

The contrapositive of a conditional is obtained by first interchanging the hypothesis and conclusion of the original conditional, and then adding negations. A conditional statement and its contrapositive are logically equivalent.

Contrapositive Law

1. $(P \to Q) \Leftrightarrow (\neg Q \to \neg P)$.

Proof of Contrapositive Law. We compare the truth tables of $P \to Q$ and $\neg Q \to \neg P$,

P	Q	$P \to Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	Q	$\neg Q$	$\neg P$	$\neg Q \to \neg P$
T	T	F	F	T
T	F	T	F	F
F	T	F	T	T
F	F	T	T	T

Since the final columns are the same, we conclude that an implication is logically equivalent to its contrapositive. □

We shall now introduce the converse of a conditional statement, which is formed by just interchanging the hypothesis and conclusion of the original implication.

Definition 1.2.2. The **converse** of a conditional $P \to Q$ is the conditional statement $Q \to P$.

Important Note: The two statements $P \to Q$ and $Q \to P$ do not mean the same thing; that is, *a conditional and its converse are not logically equivalent.* This can be shown by comparing the final columns of their truth tables:

P	Q	$P \to Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	Q	$Q \to P$
T	T	T
T	F	T
F	T	F
F	F	T

In a proof one must never confuse a conditional with its converse, because they have different meanings. On the other hand, a conditional $P \to Q$ is logically equivalent to its contrapositive $\neg Q \to \neg P$. Thus, *a conditional and its contrapositive do mean the same thing.*

1.2.2 Biconditional Statements

In mathematics, it is highly valued when one can prove that a certain concept, say A, is equivalent to a seemingly different concept B. The resulting theorem will have the form:

Theorem. *A if and only if B.*

One uses the phrase "if and only if" to assert that two concepts are equivalent, that is, alternative ways of saying the same thing. We introduce a new logical connective that will convey the mathematical meaning of this phrase.

The Biconditional Connective. Given two propositions P and Q, the biconditional connective $\leftrightarrow$ means "if and only if" and can be used to form the sentence $P \leftrightarrow Q$.

Biconditional Truth Table. Given two propositions P and Q, the sentence $P \leftrightarrow Q$ has the truth table

P	Q	$P \leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

You can check that $P \leftrightarrow Q$ is equivalent to $(P \to Q) \wedge (Q \to P)$, by constructing a truth table. Thus, we have the following important logic law that is applied in many mathematical proofs (see Section 3.7).

Biconditional Law

1. $(P \leftrightarrow Q) \Leftrightarrow (P \to Q) \wedge (Q \to P)$.

Remark. The expression "if and only if" is often abbreviated as *iff*.

Example 1. Analyze the logical forms of the following statements.

1. The picnic will be canceled if and only if it is either windy or raining.
2. If the picnic is canceled, then John will watch a movie, and if the picnic is not canceled, then John will not watch a movie.

Solution. For sentence 1, let C, W and R represent the propositions:

C : "The picnic will be canceled."
W : "It is windy."
R : "It is raining."

Thus, the logical form of sentence 1 is $C \leftrightarrow (W \vee R)$. For sentence 2, let C and J denote the propositions:

C : "The picnic is canceled."
J : "John will watch a movie."

Then the logical form of sentence 2 is $(C \to J) \wedge (\neg C \to \neg J)$. Ⓢ

We now restate the laws of logic, presented earlier, that involve the conditional or biconditional. These laws are very useful and, as a result, they are implicitly used in many mathematical proofs.

Conditional Laws (CL)

1. $(P \to Q) \Leftrightarrow (\neg P \vee Q)$.
2. $(P \to Q) \Leftrightarrow \neg(P \wedge \neg Q)$.
3. $\neg(P \to Q) \Leftrightarrow (P \wedge \neg Q)$.

Contrapositive Law

1. $(P \to Q) \Leftrightarrow (\neg Q \to \neg P)$.

Biconditional Law

1. $(P \leftrightarrow Q) \Leftrightarrow (P \to Q) \wedge (Q \to P)$.

We can now use the above propositional logic laws, together with those listed in Section 1.1.5, to derive new logic laws.

Example 2. Show that $(P \to R) \wedge (Q \to R) \Leftrightarrow (P \vee Q) \to R$, by using propositional logic laws.

Solution. We first start with the more complicated side $(P \to R) \wedge (Q \to R)$ and derive the simpler side as follows:

$$\begin{aligned}
(P \to R) \wedge (Q \to R) &\Leftrightarrow (\neg P \vee R) \wedge (\neg Q \vee R) && \text{by Conditional Law(1)} \\
&\Leftrightarrow (\neg P \wedge \neg Q) \vee R && \text{by Distributive Law(4)} \\
&\Leftrightarrow \neg(P \vee Q) \vee R && \text{by De Morgan's Law(1)} \\
&\Leftrightarrow (P \vee Q) \to R && \text{by Conditional Law(1).}
\end{aligned}$$

Therefore, $(P \to R) \wedge (Q \to R) \Leftrightarrow (P \vee Q) \to R$. Ⓢ

Using a list of propositional components $A, B, C, \ldots$ and the logical connectives $\wedge, \vee, \neg, \to, \leftrightarrow$, we can form a variety of propositional sentences. For example,

$$(P \to R) \wedge \neg(Q \leftrightarrow (S \vee T)).$$

These connectives are also used to tie together a variety of mathematical statements. A good understanding of these logical connectives will allow us to more easily understand and construct mathematical proofs.

Exercises 1.2

1. Using truth tables, show that $\neg(P \to Q) \Leftrightarrow (P \wedge \neg Q)$.
2. Construct truth tables to show that $(P \leftrightarrow Q) \Leftrightarrow (P \to Q) \wedge (Q \to P)$.
3. Using truth tables, show that $P \Leftrightarrow (\neg P \to (Q \wedge \neg Q))$.
4. Which of the following statements are true and which are false?
   (a) $(\pi^2 > 9) \to (\pi > 3)$.
   (b) If $3 \geq 2$, then $3 \geq 1$.
   (c) If $3 \geq 2$, then $3 \leq 1$.
   (d) If $1 \geq 2$, then $1 \geq 1$.
   (e) $(1 + 5 = 2) \to$ (the author is a genius).

(f) $(1+5=2) \rightarrow$ (the author is an idiot).
(g) $(\sin(2\pi) > 9) \rightarrow (\sin(2\pi) < 0)$.
(h) $3 \geq 2$ if and only if $3+5 \geq 7$.
(i) $1 \geq 2$ if and only if $1+5 \geq 7$.

5. Let C, W and R represent the propositions:

C : "The picnic has been canceled."
W : "It is windy."
R : "It is raining."

Using the propositions C, W, R analyze the logical forms of the following three statements, that is, write each statement symbolically. Then determine which of the statements are logically equivalent (justify your answers).

(a) If it is either windy or raining, then the picnic has been canceled.
(b) If the picnic has not been canceled, then it's not windy and it's not raining.
(c) The picnic has been canceled only if it's either windy or raining.

Now form the converse of each of the logical forms you obtained and then express each result in English.

6. Consider the propositions:

P : "Pigs fly."
S : "The sky is polluted with pies."
M : "Math is a favorite subject."
F : "Food is in short supply."

Translate each of the following propositional sentences into English sentences:

(a) $\neg P \rightarrow (S \vee F)$
(b) $M \vee (S \wedge \neg F)$
(c) $P \rightarrow (M \rightarrow S)$
(d) $(P \rightarrow M) \rightarrow S$
(e) $(F \wedge S) \leftrightarrow P$
(f) $(F \wedge \neg S) \rightarrow (\neg P \vee M)$
(g) $\neg F \rightarrow (\neg M \leftrightarrow (P \vee S))$.

7. Using truth tables, show that $(P \vee Q)$ and $(\neg Q \rightarrow P)$ are logically equivalent.
8. Using propositional logic laws, show that $(P \rightarrow R) \wedge (Q \rightarrow R) \Leftrightarrow (P \vee Q) \rightarrow R$.
9. Using propositional logic laws, show that $(P \rightarrow R) \vee (Q \rightarrow R) \Leftrightarrow (P \wedge Q) \rightarrow R$.
10. Using propositional logic laws, show that $P \rightarrow (Q \rightarrow R) \Leftrightarrow (P \wedge Q) \rightarrow R$.
11. Show that $(P \rightarrow Q) \rightarrow R$ and $P \rightarrow (Q \rightarrow R)$ are not logically equivalent.
12. **True or False:** The negation of the statement *"If Sue is Pina's daughter, then Diane is Sue's cousin"* is logically equivalent to the assertion *"If Sue is Pina's daughter, then Diane is not Sue's cousin."*
13. Write a negation, in English, of the statement: *If n is an even number, then n is not odd.*

14. Write an English negation of the statement: *If n is prime, then n is odd or n is 2.*

15. Write the converse, in English, of the statement: *If Sam is a fast talker, then Sam will be invited to Sheila's party.*

16. Use the contrapositive to rewrite, in English, (and not change its meaning) the statement: *If Sam is a fast talker, then Sam will be invited to Sheila's party.*

17. Suppose you are going on a hike and you come to a point where the path branches to your left and also branches to your right. Only one of these branches will take you back to the campground and you do not know which branch to take. By chance, there is a camp ranger standing at this branching point. Unfortunately, he could be the notorious imposter who always lies to campers. On the other hand, we all know that a true ranger always tells the truth, that is, he never lies to campers. Suppose that you tell this possible ranger the following:

"You are a true ranger if and only if the branch to my right returns to camp."

The purported ranger will respond with either 'true' or 'false.' Explain why if you receive the response 'true' then you know that the right branch returns to camp, and if you receive the response 'false' then you know that the left branch returns to camp.[2]

1.3 Valid and Invalid Arguments

To be a successful student of mathematics, it is very important to be able to think critically and be able to identify, construct and evaluate arguments. In mathematics, the word "argument" means something quite different from its meaning in ordinary English, where the word usually refers to a noisy or angry dispute between two parties. An argument in mathematics and logic is just a listing of statements, one of which is the conclusion and the others are the premises, or assumptions, of the argument.

Definition 1.3.1. An **argument** consists of a list of **premises** $\psi_1, \ldots, \psi_n$ followed by a **conclusion** φ. We can write this argument in the form:

$$\begin{array}{l} \psi_1 \\ \vdots \\ \underline{\psi_n} \\ \therefore \varphi \end{array} \qquad \text{(The symbol } \therefore \text{ means } \textit{therefore}\text{)}$$

[2]The imposter knows that he is not a true ranger.

Arguments are either valid or invalid. An argument is said to be valid if the conclusion follows from the premises. The concept of "follows from the premises" seems vague; however, there is a precise definition describing exactly what it means for a conclusion to follow from the premises.

Definition 1.3.2. An argument is **valid** if whenever the premises are true, then the conclusion is also true.

If an argument is valid and all of its premises are true, then we can be assured that the conclusion is also true. In other words, an argument is valid if it is impossible for the premises to be true and the conclusion to be false at the same time.

In our next example, we present an inference rule that allows you to conclude that Q is true if you know that P and $P \to Q$ are true.

Example 1. Show that the following argument is valid.

$$\begin{array}{l} P \to Q \\ P \\ \hline \therefore Q \end{array}$$

Solution. The following truth table shows that the argument is valid

		Premise 1	Premise 2	Conclusion
P	Q	$(P \to Q)$	P	Q
T	T	T	T	T
T	F	F	T	F
F	T	T	F	T
F	F	T	F	F

because whenever the premises are **all** true, the conclusion is also true. Ⓢ

The valid argument presented in Example 1 is called *modus ponens* and this argument is frequently applied in mathematical proofs. We now identify another argument, called *modus tollens*, which also regularly appears in proofs.

Example 2. Show that the following argument is valid.

$$\begin{array}{l} P \to Q \\ \neg Q \\ \hline \therefore \neg P \end{array}$$

Solution. In the following truth table we see that whenever all of the premises are true, then the conclusion is also true.

		Premise 1	Premise 2	Conclusion
P	Q	$(P \to Q)$	$\neg Q$	$\neg P$
T	T	T	F	F
T	F	F	T	F
F	T	T	F	T
F	F	T	T	T

Thus, the argument is valid. Ⓢ

An argument is **invalid** if there is a truth assignment that makes all of the premises true while making the conclusion false. So, to show that an argument is invalid we must find an assignment of truth values, to all the propositional components in the argument, that satisfies all of the premises and does not satisfy the conclusion.

Example 3. Is the argument

$$\begin{array}{l} P \vee Q \\ Q \\ \hline \therefore P \end{array}$$

a valid argument?

Solution. Consider the following truth table. Is it the case that whenever all of the premises are true, then the conclusion is true? No!

		Premise 1	Premise 2	Conclusion
P	Q	$(P \vee Q)$	Q	P
T	T	T	T	T
T	F	T	F	T
F	T	T	T	F
F	F	F	F	F

Observe that when P is false and Q is true, then all of the premises are true and yet, the conclusion is false. Thus, the argument is invalid. Ⓢ

Example 4. Analyze the logical form of the argument below. Identify the premises and the conclusion. Show that the argument is invalid.

Paula and Ernest will not both win the award for best poetry.
Ernest will win either the award for best fiction or for best nonfiction.
Paula will not win the award for best poetry.
Therefore, Ernest will win the award for best fiction.

Solution. First we shall symbolize the given argument. Let P represent the statement "Paula will win the award for best poetry" and let E represent "Ernest will win the award for best fiction." Finally, let N represent "Ernest will win the award for best nonfiction." We can now symbolize the argument and obtain

(a) $\neg(P \wedge E)$
(b) $E \vee N$
(c) $\neg P$
$\therefore E$

We will show that this argument is invalid. One could construct a truth table and then identify a truth assignment that makes all of the premises true and also makes the conclusion false; however, we shall take a slightly different approach which will save us some work. We will assign a truth value that makes the conclusion false and then try to continue in such a way as to make all the premises true.

So, we first assign the truth value F to the conclusion E. To make (c) true, we must assign the truth value F to the component P. To make (b) true, we must give N the truth value T. Since E has truth value F, we see that (a) is true. Thus, if we assign the truth values of E, P, N to be F, F, T (respectively), then the premises are true and the conclusion is false. Hence, the argument is not valid. Ⓢ

1.3.1 Two Notorious Fallacies

In everyday English the term "fallacy" is used to describe a false or mistaken belief. In logic the word "fallacy" refers to an invalid argument. There are two infamous fallacies that one must avoid in mathematics, the **converse error** and the **inverse error**, which have the form:

$P \to Q$ (Converse Error)
Q
$\therefore P$

$P \to Q$ (Inverse Error)
$\neg P$
$\therefore \neg Q$

Example 5 (Converse Error). Show that the following argument is invalid:

(a) If $x \geq 2$, then $x \geq 0$.
(b) $x \geq 0$.
Therefore, $x \geq 2$.

Solution. Assertion (a) is a true statement. Let $x = 1$. Thus (b) is also true, while the conclusion is false. So the argument is invalid. Ⓢ

Example 6 (Converse Error). Show that the argument

$P \to Q$
Q
$\therefore P$

is invalid.

Solution. Assign the truth values of P, Q to be F, T respectively. With this truth assignment we see that the conclusion is false while all of the premises are true. Hence, the argument is invalid. Ⓢ

Example 7 (Inverse Error). Verify that the following argument is invalid:

(a) If $x \geq 2$, then $x \geq 0$.
(b) $x \ngeq 2$.
Therefore, $x \ngeq 0$.

Solution. Assertion (a) is a true statement. Let $x = 1$. So, (b) is true while the conclusion is false. We conclude that the argument is invalid. Ⓢ

Example 8 (Inverse Error). Show that the following argument is invalid.

$$\begin{array}{l} P \to Q \\ \neg P \\ \hline \therefore \neg Q \end{array}$$

Solution. Assign the truth values of P, Q to be F, T respectively. With this truth assignment we see that the conclusion is false while all of the premises are true. Therefore, the argument is invalid. Ⓢ

1.3.2 Valid Arguments and Substitution

Suppose that a propositional component appears in a valid argument and we replace all occurrences of this component with a propositional sentence. Then we will obtain another valid argument. For example, consider the valid argument modus ponens (see Example 1):

$$\begin{array}{l} P \to Q \\ P \\ \hline \therefore Q \end{array} \tag{1.1}$$

Let us replace P and Q in (1.1) with any propositional sentences, say φ and ψ, respectively. We then obtain the following valid argument, which we also call modus ponens:

$$\begin{array}{l} \varphi \to \psi \\ \varphi \\ \hline \therefore \psi \end{array}$$

Similarly, let us replace P and Q in (1.1) with the propositional sentences $\neg(P \vee Q)$ and R, respectively. Thus, we have another application of modus ponens:

$$\begin{array}{l} \neg(P \vee Q) \to R \\ \neg(P \vee Q) \\ \hline \therefore R \end{array}$$

Table 1.2 Important inference rules

$P \to Q$ P $\therefore Q$	(Modus Ponens)	$P \to Q$ $\neg Q$ $\therefore \neg P$	(Modus Tollens)
P $\therefore P \vee Q$	(Disjunctive Addition)	$P \wedge Q$ $\therefore P$	(Conjunctive Simplification)
$P \vee Q$ $\neg P$ $\therefore Q$	(Disjunctive Syllogism)	P Q $\therefore P \wedge Q$	(Conjunctive Addition)

There is another kind of substitution that preserves valid arguments. Let α and β be logically equivalent. If α appears in a valid argument and we replace occurrences of α with β in this argument, then we obtain another valid argument. For example, suppose that the following argument is valid,

$$\begin{array}{l} \theta \to \alpha \\ \gamma \wedge \tau \\ \hline \therefore \alpha \end{array}$$

and that $\alpha \Leftrightarrow \beta$. Upon replacing α with β, we obtain the new valid argument

$$\begin{array}{l} \theta \to \beta \\ \gamma \wedge \tau \\ \hline \therefore \beta \end{array}$$

We will be using the ideas in our next section where we discuss inference rules.

1.3.3 Inference Rules

An *inference rule* is a valid argument that allows one to correctly derive a conclusion based solely on what one already knows. In Table 1.2 we identify some of the inference rules that are regularly employed in mathematical proofs. For example, if you are given $P \to Q$ as an assumption and you know that P holds, then modus ponens can be used to conclude that Q must be true. If you are given $P \vee Q$ as an assumption and you know that $\neg P$ holds, disjunctive syllogism can then be used to deduce that Q must be true.

The next two inference rules are alternative versions of Disjunctive Addition and Conjunctive Simplification.

$$\frac{Q}{\therefore P \vee Q} \quad \text{(Disjunctive Addition)} \qquad \frac{P \wedge Q}{\therefore Q} \quad \text{(Conjunctive Simplification)}$$

In Example 1 on page 20 we showed, using a truth table, that modus ponens is a valid argument. The argument in Table 1.2 identified as modus tollens was shown to be valid in Example 2. In a similar manner, one can show that all of the arguments presented in Table 1.2 are valid arguments.

1.3.4 Inference Rules and Substitution

In Section 1.3.2, we saw that certain substitutions preserve valid arguments and so, we can also apply these substitutions to inference rules. Thus, given an inference rule, we can form generalizations of this inference rule. For example, consider the two versions of disjunctive syllogism:

$$(1)\ \frac{P \vee Q \quad \neg P}{\therefore Q} \quad \text{(Disjunctive Syllogism)} \qquad (2)\ \frac{P \vee Q \quad \neg Q}{\therefore P} \quad \text{(Disjunctive Syllogism)}$$

In (1) and (2) let us replace P and Q with any propositional sentences φ and $\neg\psi$, respectively. The result is two new applications of disjunctive syllogism

$$(3)\ \frac{\varphi \vee \neg\psi \quad \neg\varphi}{\therefore \neg\psi} \quad \text{(Disjunctive Syllogism)} \qquad (4)\ \frac{\varphi \vee \neg\psi \quad \psi}{\therefore \varphi} \quad \text{(Disjunctive Syllogism)}$$

where in (4) we apply the double negation logic law $\neg\neg\psi \Leftrightarrow \psi$.

For another example, consider the inference rules modus ponens and modus tollens in Table 1.2. In these valid arguments, let us replace P and Q with $\neg\varphi$ and $\neg\psi$, respectively. We obtain another instance of modus ponens and modus tollens:

$$(5)\ \frac{\neg\varphi \rightarrow \neg\psi \quad \neg\varphi}{\therefore \neg\psi} \quad \text{(Modus Ponens)} \qquad (6)\ \frac{\neg\varphi \rightarrow \neg\psi \quad \psi}{\therefore \varphi} \quad \text{(Modus Tollens)}$$

where in (6) we are using the logical equivalences $\neg\neg\psi \Leftrightarrow \psi$ and $\neg\neg\varphi \Leftrightarrow \varphi$.

1.3.5 Deductions

Besides using truth tables, there is another way of showing that an argument is valid. An argument is valid if one can *deduce* the conclusion from the premises using a

collection of inference rules. A *formal deduction* includes a record of the inference rules that were used to derive the conclusion. We present a formal deduction in each of our next two examples.

Example 9. Using the inference rules in Table 1.2, formally deduce the conclusion from the premises in the argument:

(a) $(\neg P \vee Q) \rightarrow R$
(b) $S \vee \neg Q$
(c) $\neg T$
(d) $P \rightarrow T$
(e) $(\neg P \wedge R) \rightarrow \neg S$

$\therefore \neg Q$

Solution. We start making deductions from the premises (a)–(e). These deductions will allow us to make more deductions and, eventually, we will deduce the required $\neg Q$. Our formal deduction identifies, in the right hand column, the rules of inference that were used to draw a specific conclusion at each step:

(1) $\neg P$ by premises (c), (d) and modus tollens
(2) $\neg P \vee Q$ by (1) and disjunctive addition
(3) R by (2), premise (a) and modus ponens
(4) $\neg P \wedge R$ by (1), (3) and conjunctive addition
(5) $\neg S$ by (4), premise (e) and modus ponens
(6) $\neg Q$ by (5), premise (b) and disjunctive syllogism.

Thus, from the premises (a)–(e) we have deduced $\neg Q$. Ⓢ

Example 10. Using the inference rules in Table 1.2, formally deduce the conclusion from the premises in the argument:

(a) $\neg S \rightarrow D$
(b) $\neg S \vee (\neg D \rightarrow K)$
(c) $\neg D$

$\therefore K$

Solution. As in our solution to Example 9, we start making deductions from the premises (a)–(c) and deduce K as follows:

(1) S by premises (a), (c) and modus tollens
(2) $\neg D \rightarrow K$ by (1), premise (b) and disjunctive syllogism
(3) K by (2), premise (c) and modus ponens.

Thus, we have deduced K from premises (a)–(c). Ⓢ

A mathematical proof is also a deductive argument and mathematicians implicitly use the inference rules listed in Table 1.2 in their proofs.

Exercises 1.3

1. Using truth tables, show that all of the arguments in Table 1.2 are valid.

2. Using modus ponens or modus tollens, fill in the blanks so as to produce a valid argument.

(a) If π is rational, then $\pi = a/b$ for some integers a and b.
It is not true that $\pi = a/b$ for some integers a and b.
$\therefore$ _____

(b) If logic is easy, then I am a monkey's uncle.
I am not a monkey's uncle.
$\therefore$ _____

(c) If they were unsure of the address, then they would have telephoned.

_____
$\therefore$ They were sure of the address.

3. Let M, P and J represent the propositions:

M : "Mary does her homework."
P : "Peter does his homework."
J : "Jim does his homework."

Using the propositions M, P and J, analyze the logical form of the following argument. Identify the premises and the conclusion. Is the argument valid?

If Mary does her homework, then Peter will do his homework.
If Peter does his homework, then Jim will do his homework.
Mary does not do her homework.
Therefore, Jim does not do his homework.

4. Using the inference rules in Table 1.2, formally deduce the conclusion from the premises.

(a) $A \rightarrow (B \vee C)$
(b) $A \wedge \neg B$
$\therefore C$

5. Using the inference rules in Table 1.2, formally deduce the conclusion from the premises.

(a) $(P \wedge Q) \rightarrow R$
(b) $\neg(P \wedge Q) \rightarrow (\neg P \vee \neg Q)$
(c) $R \rightarrow S$
(d) $Q \wedge \neg S$
$\therefore \neg P$

6. Using the inference rules in Table 1.2, formally deduce the conclusion from the premises.

(a) $P \vee Q$
(b) $Q \rightarrow R$
(c) $(P \wedge S) \rightarrow T$
(d) $\neg R$
(e) $\neg Q \rightarrow (U \wedge S)$
$\therefore T$

7. Using the inference rules in Table 1.2, formally deduce the conclusion from the premises.

(a) $P \rightarrow Q$
(b) $R \vee S$
(c) $\neg S \rightarrow \neg T$
(d) $\neg Q \vee S$
(e) $\neg S$
(f) $(\neg P \wedge R) \rightarrow U$
(g) $W \vee T$
$\therefore U \wedge W$

CHAPTER 2

Predicate Logic

Mathematicians frequently use the expressions: *for all* and *there exists*. These two expressions are called quantifiers and are represented, respectively, by the symbols $\forall$ and $\exists$. The universal quantifier $\forall$ is applied when one wants to assert that everything satisfies a given property. The existential quantifier $\exists$ is employed when one wants to state that something satisfies a particular property. Many statements in mathematics involve quantifiers and to prove such statements, one needs to clearly understand the meaning of quantifiers. Before we go any further with our discussion of quantifiers, we must first investigate properties and predicates.

2.1 Variables, Predicates, and Truth Sets

Variables, for instance x, y and z, are used extensively in mathematics. They are used when we are interested in "properties" that may be true or false, depending on the values represented by the variables. A *predicate* is just a statement proclaiming that certain variables satisfy a property. For example, "x is tall" is a predicate and we can symbolize this predicate by $T(x)$. Of course, the truth or falsity of the expression $T(x)$ can be determined only when a value for x is given. Suppose Peter is 7 ft tall. Then the expression $T(\text{Peter})$, which means "Peter is tall," would be true.

Another predicate is "x evenly divides y" and we could symbolize this predicate by $D(x,y)$. Thus, the statement $D(9,27)$ is true and $D(9,5)$ is false. The *domain* of a predicate is just the collection of allowed values for the variable(s) in the predicate. So, the domain of the predicate $T(x)$ is the collection of all people. The domain of the predicate $D(x,y)$ is the collection of all integers (see Section 2.1.3).

Example 1. Consider the three predicates $P(x)$, $E(x)$, and $D(x,y)$ given by

- $P(x)$ symbolizes the statement "x is a prime number"
- $E(x)$ symbolizes the statement "x is even"
- $D(x,y)$ symbolizes the statement "x evenly divides y"

where x and y represent integers. Find some values for the variables that make the following logical formulas true, and others making them false.

1. $P(x) \wedge E(x)$.
2. $E(x) \vee D(x,y)$.
3. $\neg P(x) \wedge D(x,y)$.
4. $D(x,y) \rightarrow \neg P(x)$.

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_2,

Solution. We identify such values as follows:

1. $P(x) \wedge E(x)$. For $x = 2$, this statement is true. On the other hand, if we let $x = 3$, then the statement is false.
2. $E(x) \vee D(x,y)$. When $x = 2$ and y is any integer, then the statement is true. If we let $x = 3$ and $y = 4$, then the statement is false.
3. $\neg P(x) \wedge D(x,y)$. If $x = 4$ and $y = 8$, then the statement is true. When $x = 3$ and y is any integer, the statement is false.
4. $D(x,y) \rightarrow \neg P(x)$. When $x = 5$ and $y = 10$, the statement is true. If $x = 1$ and $y = 3$, then the statement is false. Ⓢ

Example 2. Analyze the logical forms of the following statements, that is, write each statement symbolically, using the predicates P, E, D defined in Example 1.

1. x is a prime number, and either y is even or z is divisible by x.
2. Exactly one of x and y is even.

Solution. The logical form of statement 1 is $P(x) \wedge (E(y) \vee D(x,z))$. In statement 2, the expression "Exactly one" means "one or the other, and not both." In other words statement 2 means that "x is even or y is even, and not both." We get $E(x) \vee E(y)$ for the "one or the other" part, and for the "not both" part we get $\neg(E(x) \wedge E(y))$. Thus, the logical form of the given statement is $(E(x) \vee E(y)) \wedge \neg(E(x) \wedge E(y))$. Ⓢ

2.1.1 Universe of Discourse

We say that A is a set when A is a collection of objects. The objects that belong to a set A are called the *elements* of A. We write $a \in A$ to mean that a is an *element*, or a *member*, of the set A. We write $a \notin A$ when a is *not* an element of the set A. A set is merely the result of collecting together objects of interest, and is usually identified by enclosing its elements with curly brackets. For example, the collection $A = \{3, 7, 86, 11, 99\}$ is a set where we see that $7 \in A$ and $8 \notin A$. In Section 5.1 we will cover the basics of set theory in greater depth (see pages 143–148).

When our attention is to be focused on just the elements in a particular set A, then we will say that A is our **universe of discourse**. In other words, a universe of discourse is just the set of all things we are considering during our discussion or study. For example, if we were just talking about students, then our universe of discourse would be the set of all students. In logic and in mathematics, the universe of discourse is a familiar concept.

2.1.2 Sets Defined by a Predicate

Given a property we will often form the collection of just those elements in a set that satisfy the property. For an example, let $A = \{0, 1, 2, 3, \ldots\}$ and suppose we

want to collect just those elements in A that are odd. We can easily describe this set by $\{n \in A : n \text{ is odd}\}$, that is,"the set of all $n \in A$ such that n is odd." Thus, $\{n \in A : n \text{ is odd}\} = \{1,3,5,7,\ldots\}$.

Definition 2.1.1. Given a **universe of discourse** (that is, a set of objects) U and a predicate $P(x)$ we can form the **truth set** $\{x \in U : P(x) \text{ is true}\} = \{x \in U : P(x)\}$. When the universe U is understood, we sometimes write the truth set as $\{x : P(x)\}$.

When constructing the set $\{x \in U : P(x)\}$, mathematicians will say "the set of all $x \in U$ such that $P(x)$." Some mathematicians write truth sets as $\{x \in U \,|\, P(x)\}$ using the vertical bar | rather than the colon. The colon : and the bar | can be thought of as an abbreviation for the expression "such that."

Example 3. Let $U = \{-20,-19,\ldots,19,20\}$. Then

1. $\{z \in U : z^2 = 9\} = \{-3,3\}$.
2. $\{z \in U : z^2 \leq 16\} = \{-3,-4,0,-1,-2,1,2,4,3\}$.
3. $\{x \in U : x^2 > 225\} = \{16,-16,-20,-19,-18,-17,19,20,18,17\}$.

2.1.3 Important Sets in Mathematics

The reader should be familiar with the natural numbers, the integers, and the real numbers. In elementary algebra, we learned many properties about the real numbers, including the commutative and associative laws for multiplication and addition. You may have also learned that when a real number can be expressed as the ratio of two integers, then it is called a *rational number*. Are there real numbers that are not rational? The answer is yes and we will verify this in Chapter 3.

Definition 2.1.2. A real number x is **rational** if and only if $x = \frac{a}{b}$ for some integers a,b where $b \neq 0$. If a real number is not rational, then it is called **irrational**.

Certain sets appear frequently in mathematics; namely, the sets of natural numbers, integers, rational and real numbers. These sets are usually denoted by:

1. $\mathbb{N} = \{1,2,3,\ldots\}$ is the set of natural numbers.
2. $\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3,\ldots\}$ is the set of integers.
3. $\mathbb{Q}$ is the set of rational numbers. Thus, $\frac{3}{2} \in \mathbb{Q}$.
4. $\mathbb{R}$ is the set of real numbers and so, $\pi \in \mathbb{R}$.

The set $\mathbb{N}$ is **closed** under the operations of addition and multiplication, that is, the sum and product of two natural numbers is a natural number. Moreover, the sets $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are closed under addition, multiplication, and subtraction. For example, if we add, multiply, or subtract any two rational numbers the result is again a rational number. Finally, recall that each nonzero element in $\mathbb{Q}$, and $\mathbb{R}$, has a **multiplicative inverse**. For example, if $x \in \mathbb{Q}$ and $x \neq 0$, then there is a $y \in \mathbb{Q}$ such that $x \cdot y = 1$. This element y is usually denoted by $\frac{1}{x}$.

For each of the sets $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$, we may add '$+$' or '$-$' as a superscript. The superscript $+$ indicates that only the positive numbers will be allowed. Similarly, the superscript $-$ means that only the negative numbers are permitted. For example,

1. $\mathbb{Q}^+ = \{x \in \mathbb{Q} : x > 0\}$.
2. $\mathbb{Z}^- = \{x \in \mathbb{Z} : x < 0\}$.
3. $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$.

For sets A and B we write $A \subseteq B$ to mean that the set A is a subset of the set B, that is, every element of A is also an element of B. Thus, $\mathbb{N} \subseteq \mathbb{Z}$.

Example 4. Consider the following three subsets of $\mathbb{Z}$:

1. $\{x \in \mathbb{Z} : x \text{ is a prime number}\} = \{2,3,5,7,11,\ldots\}$.
2. $\{x \in \mathbb{Z} : x \text{ is divisible by } 3\} = \{\ldots,-12,-9,-6,-3,0,3,6,9,12,\ldots\}$.
3. $\{z \in \mathbb{Z} : z^2 \leq 1\} = \{-1,0,1\}$.

Another set that appears in mathematics is the empty set $\varnothing$, a set that has no elements. Since $\varnothing$ has no elements, we see that $\varnothing = \{\ \}$. When first introduced to the empty set $\varnothing$, students sometimes have difficulty thinking of $\varnothing$ as a set. One can think of $\varnothing$ as a house in which nobody lives.

Equality of Rational Numbers

For integers m and n if $m = n$, then we know that m and n must be exactly the same number, in fact, they must look exactly the same. For instance, we know that $2 = 2$ and $5 = 5$; and we know that $2 \neq 5$.

A rational number is a number that can be written as a ratio $\frac{p}{q}$ where p and q are integers and $q \neq 0$. What does it mean to say two rational numbers $\frac{p}{q}$ and $\frac{a}{b}$ are equal when they may not look the same? For example, $\frac{1}{2}$ and $\frac{4}{8}$ are equal but they look different. Our next definition will answer this question.

Definition 2.1.3. Whenever $\frac{p}{q}$ and $\frac{a}{b}$ are rational numbers, we have that $\frac{p}{q} = \frac{a}{b}$ if and only if $pb = qa$ as integers.

Thus, we know that $\frac{1}{2}$ and $\frac{4}{8}$ are equal because $1 \cdot 8 = 2 \cdot 4$.

Example 5. Let $\frac{p}{q}$, $\frac{r}{s}$, $\frac{a}{b}$, $\frac{c}{d}$ be rational numbers. Suppose $\frac{p}{q} = \frac{a}{b}$ and $\frac{r}{s} = \frac{c}{d}$. Show that $\frac{pr}{qs} = \frac{ac}{bd}$.

Solution. Since $\frac{p}{q} = \frac{a}{b}$ and $\frac{r}{s} = \frac{c}{d}$, we have that (1) $pb = qa$ and (2) $rd = sc$, by Definition 2.1.3. To show that $\frac{pr}{qs} = \frac{ac}{bd}$, we must verify that $(pr)(bd) = (qs)(ac)$. We do this as follows:

$$\begin{aligned}(pr)(bd) &= (pb)(rd) && \text{by algebra}\\ &= (qa)(sc) && \text{by (1) and (2)}\\ &= (qs)(ac) && \text{by algebra.}\end{aligned}$$

Thus, $(pr)(bd) = (qs)(ac)$. Therefore, $\frac{pr}{qs} = \frac{ac}{bd}$ by Definition 2.1.3. Ⓢ

Interval Notation

In mathematics, an interval is a set consisting of all the real numbers that lie between two given real numbers a and b, where $a < b$. The numbers a and b are referred to as the endpoints of the interval. Furthermore, an interval may or may not include its endpoints.

1. The open interval (a,b) is defined to be $(a,b) = \{x \in \mathbb{R} : a < x < b\}$.
2. The closed interval $[a,b]$ is defined to be $[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$.
3. The left-closed interval $[a,b)$ is defined to be $[a,b) = \{x \in \mathbb{R} : a \leq x < b\}$.
4. The right-closed interval $(a,b]$ is defined to be $(a,b] = \{x \in \mathbb{R} : a < x \leq b\}$.

For each real number a we can also define intervals called rays or half-lines.

1. The interval (a,∞) is defined to be $(a,\infty) = \{x \in \mathbb{R} : a < x\}$.
2. The interval $[a,\infty)$ is defined to be $[a,\infty) = \{x \in \mathbb{R} : a \leq x\}$.
3. The interval $(-\infty,a)$ is defined to be $(-\infty,a) = \{x \in \mathbb{R} : x < a\}$.
4. The interval $(-\infty,a]$ is defined to be $(-\infty,a] = \{x \in \mathbb{R} : x \leq a\}$.

The symbol ∞ denotes 'infinity' and it does not represent a number. The notation ∞ is a useful symbol that allows us to represent an interval 'without a right endpoint.' Similarly, the notation $-\infty$ is used to denote an interval 'having no left endpoint.'

Example 6. Using interval notation, evaluate the truth sets:

(1) $\{x \in \mathbb{R} : x^2 - 1 < 3\}$.
(2) $\{x \in \mathbb{R}^+ : (x-1)^2 > 1\}$.
(3) $\{x \in \mathbb{R}^- : x > \frac{1}{x}\}$.

Solution. We will be using the standard properties of inequality that are typically reviewed in a calculus book (see 3.1.5 on page 64 of this book).

(1) We first solve the inequality $x^2 - 1 < 3$ for x^2 obtaining $x^2 < 4$. The solution to this latter inequality is $-2 < x < 2$. Thus, $\{x \in \mathbb{R} : x^2 - 1 < 3\} = (-2,2)$.
(2) We are looking for all the positive real numbers x that satisfy $(x-1)^2 > 1$. Since $(x-1)^2 = x^2 - 2x + 1$, we shall solve the inequality $x^2 - 2x + 1 > 1$. After subtracting 1 from both sides and factoring, we obtain $x(x-2) > 0$. Since $x > 0$, we conclude that $x > 2$. Therefore, $\{x \in \mathbb{R}^+ : (x-1)^2 > 1\} = (2,\infty)$.
(3) We need to find the negative real numbers x that satisfy $x > \frac{1}{x}$. From $x > \frac{1}{x}$, we conclude that $x^2 < 1$ because $x < 0$. So we have $-1 < x < 0$ and, as a result, we obtain $\{x \in \mathbb{R}^- : x > \frac{1}{x}\} = (-1,0)$. Ⓢ

Exercises 2.1

1. Let $\frac{p}{q}$ and $\frac{r}{s}$ be rational numbers where p,q,r,s are integers and q,s are nonzero. Suppose $\frac{p}{q} = \frac{r}{s}$ and $p \neq 0$. Using Definition 2.1.3, show that $\frac{ps}{qs} = \frac{pr}{ps}$.

2. Let $R(x)$ be the predicate $x > \frac{1}{x}$. Indicate whether the statements $R(2)$, $R(-2)$, $R(\frac{1}{2})$, $R(-\frac{1}{2})$ are true or false. Now evaluate the following truth sets:

(a) $\{x \in \mathbb{R}^+ : x > \frac{1}{x}\}$.
(b) $\{x \in \mathbb{R}^- : x^2 > \frac{1}{x}\}$.
(c) $\{x \in \mathbb{Z} : x > \frac{1}{x} \text{ and } x > 2\}$.
(d) $\{x \in \mathbb{Z} : x > \frac{1}{x} \text{ and } x \not> 2)\}$.

3. Let $L(x,y)$ be the predicate "$x < y$." Determine if the following statements are true or false:

(a) $L(2,3) \to L(4,9)$.
(b) $L(-3,-2) \to L(9,4)$.
(c) $L(-2,-3) \to L(9,4)$.

4. Evaluate the truth sets:

(a) $\{x \in \mathbb{R} : x^2 < 9\}$.
(b) $\{x \in \mathbb{Z} : x^2 < 9\}$.
(c) $\{x \in \mathbb{R} : 2x+9 \leq 5\}$.
(d) $\{x \in \mathbb{R} : x > 0 \text{ and } x^3 < \frac{16}{x}\}$.

5. Let $\frac{p}{q}$ and $\frac{r}{s}$ be rational numbers where p,q,r,s are integers and q,s are nonzero. Suppose $\frac{p}{q} = \frac{r}{s}$. Using Definition 2.1.3, show that $\frac{2p+q}{2q} = \frac{2r+s}{2s}$.

2.2 Quantifiers

Given a statement $P(x)$, which says something about the variable x, we want to express the fact that *every* element x in the universe makes $P(x)$ true. In addition, we may want to express the fact that *at least one* element x in the universe makes $P(x)$ true. To do this, we will form sentences using the quantifiers $\forall$ and $\exists$. The quantifier $\forall$ means "for all" and is called the *universal quantifier.* The quantifier $\exists$ means "there exists" and is called the *existential quantifier.* For example, we can form the sentences

1. $\forall x P(x)$ [means "for all x, $P(x)$"].
2. $\exists x P(x)$ [means "there exists an x such that $P(x)$"].

A statement of the form $\forall x P(x)$ is referred to as a *universal statement*. A statement having the form $\exists x P(x)$ is called an *existential statement*. Quantifiers offer us a valuable tool for clear thinking in mathematics, where many concepts begin with the expression "for every," or "there exists."

Example 1. What do the following formulas mean? Are they true or false?

1. $\forall x(x^2 \geq 0)$, where the universe of discourse is $\mathbb{R}$, the set of all real numbers.
2. $\forall x(x^2 > 0)$, where the universe is $\mathbb{R}$.
3. $\exists x(x^2 + x - 2 = 0)$, where the universe is $\mathbb{R}$.
4. $\exists x(x^2 + 1 = 0)$, where the universe is $\mathbb{R}$.
5. $\exists x(M(x) \wedge \neg B(x))$, where the universe is the set of all people, $M(x)$ means "x is a man," and $B(x)$ means "x has black hair."
6. $\forall x(M(x) \rightarrow B(x))$, where the universe is the set of all people, $M(x)$ means "x is a man," and $B(x)$ means "x has black hair."

Solution.

1. The formula $\forall x(x^2 \geq 0)$ means that for every real number x we have $x^2 \geq 0$, and this is true.
2. The formula $\forall x(x^2 > 0)$ means that for every real number x we have $x^2 > 0$. This not true for all real numbers, because $0^2 \not> 0$. So, the formula $\forall x(x^2 > 0)$ is false in the universe $\mathbb{R}$.
3. Is there is a real number x that satisfies the equation $x^2 + x - 2 = 0$? Since the number 1 satisfies the equation, we conclude that the formula $\exists x(x^2 + x - 2 = 0)$ is true in the universe $\mathbb{R}$.
4. Is there is a real number x that satisfies the equation $x^2 + 1 = 0$? Since no real number satisfies this equation, we see that the formula $\exists x(x^2 + 1 = 0)$ is false in the universe $\mathbb{R}$.
5. The formula $\exists x(M(x) \wedge \neg B(x))$ states that there is a person x such that x is a man and x does not have black hair. This is clearly true in the universe of all people.
6. The formula $\forall x(M(x) \rightarrow B(x))$ asserts that for every person x, if x is a man then x has black hair. This is clearly false in the universe of all people; because there are men who do not have black hair. Ⓢ

Remark 2.2.1. To show that a universal statement $\forall x P(x)$ is true in a particular universe U, one must show that $P(x)$ is true for every $x \in U$. On the other hand, to show that the statement $\forall x P(x)$ is false in a universe U, one must find at least one $x \in U$ for which $P(x)$ is false. A value $x \in U$ that verifies $P(x)$ is false is called a **counterexample** to the universal statement $\forall x P(x)$.

Example 2. Let the universe U be the set $\{-1, 0, 1, \ldots, 100\}$. Determine the truth value in the universe U of the following sentences:

1. $\exists y(y^2 = 9)$.
2. $\exists z(z^2 \neq 9)$.
3. $\forall x(x < 5 \rightarrow x^2 < 25)$.
4. $\exists x(x + x = x)$.

Solution. We shall translate each symbolic sentence into English. We will then determine whether, or not, the translation is true in the universe U.

1. The sentence $\exists y(y^2 = 9)$ means that "there is a $y \in U$ that satisfies $y^2 = 9$." This is true for $y = 3 \in U$. So, the sentence $\exists y(y^2 = 9)$ is true in the universe U.
2. Translating the sentence $\exists z(z^2 \neq 9)$ into English, we obtain "there is a $z \in U$ that satisfies $z^2 \neq 9$." This is true for $z = 0 \in U$. Hence, $\exists z(z^2 \neq 9)$ is true is U.
3. This sentence $\forall x(x < 5 \rightarrow x^2 < 25)$ states "for every $x \in U$, if $x < 5$ then $x^2 < 25$." All of the elements in U that are less than 5 are $-1, 0, 1, 2, 3, 4$. The square of each of these is less than 25. Thus, the sentence is true in U.
4. The sentence $\exists x(x + x = x)$ means that "there is an $x \in U$ satisfying $x + x = x$." This is true for $x = 0 \in U$.

Therefore, all of the logical sentences are true in the universe U. Ⓢ

Example 3. Find a new universe A, similar to the one in Example 2, in which the third sentence $\forall x(x < 5 \rightarrow x^2 < 25)$ is false.

Solution. By the result of Example 2, we have that the formula $\forall x(x < 5 \rightarrow x^2 < 25)$ is true in the universe $U = \{-1, 0, 1, \ldots, 100\}$. So, we shall try to add a new element to U that will make this formula false. Note that the conditional $(-5 < 5 \rightarrow 5^2 < 25)$ is false, as the hypothesis is true and the conclusion is false. By adding -5 to the set U, we get the set $A = \{-5, -1, 0, 1, \ldots, 100\}$ and conclude that $\forall x(x < 5 \rightarrow x^2 < 25)$ is false in the universe A. Ⓢ

Suppose that a variable x appears in a logical formula $P(x)$. In the statements $\forall x P(x)$ and $\exists x P(x)$, we will say that x is a *bound variable* and that x is bound by a quantifier. In other words, if a variable in a logical formula is immediately used by a quantifier, then that variable is referred to as a bound variable. If a variable in a statement is not bound by a quantifier then we shall say that the variable is a *free variable*. If a variable is free, then substitution may take place. In a given universe, when all of the free variables in a statement are replaced with values, then one can then determine the truth or falsity of the resulting statement.

2.2.1 Analyzing the Logical Form of Statements

Example 4. Analyze the logical forms of the following six sentences. In other words, identify the relevant predicates and then write each statement symbolically, revealing any possible hidden quantifiers or hidden logical connectives.

1. Every cat is an animal.
2. Some cat is an animal.
3. No cat is an animal.
4. Someone in this class does not do their homework.
5. Everyone in this class does their homework.
6. Nobody in this class does their homework.

Solution. We will express the given six sentences into logical form. We first identify the two predicates that appear in sentences 1–3. Let $C(x)$: "x is a cat" and $A(x)$: "x is an animal."

1. Sentence 1 means that "for all x, if x is a cat then x is an animal." In logical form, we have $\forall x(C(x) \rightarrow A(x))$.
2. Rephrasing sentence 2, we obtain "for some x, x is a cat and x is an animal." In logical form, we have $\exists x(C(x) \wedge A(x))$.
3. There are two equivalent ways to restate sentence 3. First, this sentence means that "it is false that some cat is an animal," that is, $\neg$(some cat is an animal). In logical form, we obtain $\neg\exists x(C(x) \wedge A(x))$. Secondly, the sentence also means that "every cat fails to be an animal;" that is, $\forall x(C(x) \rightarrow \neg A(x))$.

We now symbolize the predicates that appear in sentences 4–6. Let $C(x)$ represent the predicate "x is in this class" and let $H(x)$ identify the property "x does his/her homework."

4. Sentence 4 means that "there is a person x in this class and this person x does not do his homework," that is, $\exists x(C(x) \wedge \neg H(x))$.
5. In other words, sentence 5 states "for all x, if x is in this class then x does his homework." In logical form, we have $\forall x(C(x) \rightarrow H(x))$.
6. First, sentence 6 means that "it is not the case that there is a person in this class who does his homework," that is,

$$\neg(\text{there is a person in this class who does his homework}).$$

In logical form, we obtain $\neg\exists x(C(x) \wedge H(x))$. Alternatively, the sentence also means that "every person in this class fails to do his homework" and we get $\forall x(C(x) \rightarrow \neg H(x))$. Ⓢ

Our solutions in Example 4 demonstrate a theme concerning the quantifiers $\forall$ and $\exists$. Our solutions to items 1 and 5 indicate that the quantifier $\forall$ is frequently (but not always) followed by a statement that uses the conditional connective $\rightarrow$. In addition, our solutions to items 2 and 4 illustrate that the quantifier $\exists$ is frequently (but not always) followed by a statement that uses the conjunctive connective $\wedge$.

Quantifiers in a Tarskian World

The understanding of predicate logic is (one can argue) a requisite tool for the study of mathematics. To help students learn the language of predicate logic, Jon Barwise and John Etchemendy [1] created an innovative software program and book called Tarski's World. The program presents a visual display of geometric shapes sitting on a grid, referred to as a world (or universe). The shapes have a variety of colors and positions on the grid. The user can create logic formulas and then determine whether or not these formulas are true or false in the world. Tarski's World is named after Alfred Tarski, one of the early pioneers in mathematical logic.

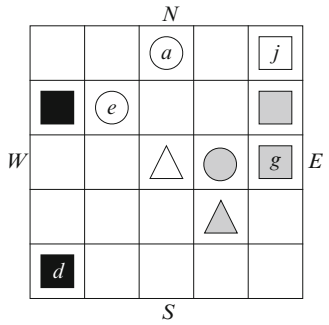


Fig. 2.1 A Tarskian World

In our next example we are given a Tarskian World and some English statements that are either true or false in this given world. We will be asked to translate these statements into logical form.

Example 5. The Tarskian World presented in Fig. 2.1 has a few individuals who are labeled with a name and some who are without a name. The universe consists of all the objects in this Tarskian world. Define the **Tarskian predicates**:

- $T(x)$ means "x is a triangle." $C(x)$ means "x is a circle." $S(x)$ means "x is a square."
- $I(x)$ means "x is white." $G(x)$ means "x is gray." $B(x)$ means "x is black."
- $N(x,y)$ means "x is on the northern side of y."
- $W(x,y)$ means "x is on the western side of y."
- $K(x,y)$ means "x has the same color as y."

Using these predicates, analyze the logical form of each of the following English statements; that is, write each statement symbolically, looking for possible hidden quantifiers and logical connectives.

1. There is a black square.
2. Every circle is white.
3. There are no black circles.
4. a is north of e.
5. a is not north of j.
6. Every circle is north of d.
7. There is a circle that has the same color as d.
8. d is west of every circle.

Solution. We first note that statements 2 and 7 are the only ones that are false in the Tarskian world depicted in Fig. 2.1. We shall now express, in order, each of the eight sentences in logical form.

1. The sentence means that "for some x, x is black and x is a square." Thus, the logical form of the sentence is $\exists x(B(x) \wedge S(x))$.

2. In other words, "for all x, if x is a circle then x is white." We thus obtain the logical form $\forall x(C(x) \to I(x))$.
3. The sentence can be stated in two equivalent ways. First, the sentence means that "it is false that some circle is black," that is, $\neg$(some circle is black). In logical form, we obtain $\neg\exists x(C(x) \wedge B(x))$. Secondly, the sentence also means that "every circle is not black" and we get $\forall x(C(x) \to \neg B(x))$.
4. In logical form, the sentence becomes $N(a,e)$.
5. The logical form of this sentence is $\neg N(a,j)$.
6. Rephrasing, we obtain "for all x, if x is a circle then x is north of d." In logical form, we have $\forall x(C(x) \to N(x,d))$.
7. The sentence asserts that "for some x, x is a circle and x has the same color as d." In logical form, we get $\exists x(C(x) \wedge K(x,d))$.
8. Stated more precisely, we obtain "for all x, if x is a circle then d is west of x." In logical form, we have $\forall x(C(x) \to W(d,x))$. Ⓢ

Remark 2.2.2. Two individuals that lie in the same column of a Tarskian world are not west of one another. Consequently, in Fig. 2.1 the assertions $W(i,g)$ and $W(g,i)$ are both false. Similarly, any two individuals who lie in the same row are not north of one another.

2.2.2 Bounded Quantifiers

Bounded quantifiers place *bounds* on the values that are to be considered. Bounded set quantifiers are often used in mathematics when one wants to put a restriction on the values under a quantifier. For example, to state that every real number x satisfies $P(x)$, we will write $(\forall x \in \mathbb{R})P(x)$. Similarly, to say that some real number x satisfies $P(x)$ we can write $(\exists x \in \mathbb{R})P(x)$.

Definition 2.2.3 (Bounded Set Quantifiers). For a set A, we write $(\forall x \in A)P(x)$ to mean that *for every x in A, $P(x)$ is true*. Similarly, we write $(\exists x \in A)P(x)$ to signify that *for some x in A, $P(x)$ is true*.

The assertion $(\forall x \in A)P(x)$ means that for every x, if $x \in A$ then $P(x)$ is true. Similarly, the statement $(\exists x \in A)P(x)$ means that there is an x such that $x \in A$ and $P(x)$ is true. Thus, we have the logical equivalences:

1. $(\forall x \in A)P(x) \Leftrightarrow \forall x(x \in A \to P(x))$.
2. $(\exists x \in A)P(x) \Leftrightarrow \exists x(x \in A \wedge P(x))$.

Bounded number quantifiers are also very useful when one wants to put some restriction on the numbers being quantified. To say that all numbers $x > 1$ satisfy $P(x)$, we shall write $(\forall x > 1)P(x)$. Similarly, to say that some number $x < 4$ satisfies $P(x)$ we can write $(\exists x < 4)P(x)$.

Definition 2.2.4 (Bounded Number Quantifiers). When our universe is a set of numbers and a is a specific number, we write $(\forall x < a)P(x)$ to mean that *for every number $x < a$, $P(x)$ is true*. Similarly, we write $(\exists x < a)P(x)$ to assert that *for some number $x < a$, $P(x)$ is true*.

Let a be a number. The assertion $(\forall x < a)P(x)$ means that for every number x, if $x < a$ then $P(x)$ is true. Similarly, the statement $(\exists x < a)P(x)$ means that there is a number x such that $x < a$ and $P(x)$ is true. Thus, we have the logical equivalences:

1. $(\forall x < a)P(x) \Leftrightarrow \forall x(x < a \rightarrow P(x))$.
2. $(\exists x < a)P(x) \Leftrightarrow \exists x(x < a \wedge P(x))$.

There are similar bounded number quantifiers for the inequalities $\leq$, $>$, $\geq$ as well; for example, the quantifiers in $(\forall x \leq a)P(x)$ and $(\exists x > a)P(x)$ are also referred to as bounded number quantifiers. The statement $(\forall x \leq a)P(x)$ means for every number $x \leq a$, the statement $P(x)$ is true. Similarly, the assertion $(\exists x > a)P(x)$ means that for some number $x > a$ the assertion $P(x)$ is true.

Exercises 2.2

1. Write the statements in logical form, using an appropriate bounded quantifier.
   (a) For every real number x, if $x > 1$ then $x > \frac{1}{x}$.
   (b) There exists a rational number y such that $y < 2$ and $y^2 > 4$.

2. Determine whether the statements are true or false in the universe $\mathbb{R}$.
   (a) $\forall x(x^2 + 1 > 0)$.
   (b) $\forall x(x^2 + x \geq 0)$.
   (c) $\forall x(x > \frac{1}{2} \rightarrow \frac{1}{x} < 3)$.
   (d) $\exists x(\frac{1}{x-1} = 3)$.
   (e) $\exists x(\frac{1}{x-1} = 0)$.

3. Consider the predicates:

   $C(x)$: "x is in the class."
   $M(x)$: "x is a mathematics major."

   Using these predicates, analyze the logical form of each of the sentences where the universe is the set of all college students.
   (a) Everyone in the class is a mathematics major.
   (b) Someone in the class is a mathematics major.
   (c) No one in the class is a mathematics major.
   (d) There a mathematics major who is not in the class.
   (e) Every mathematics major is in the class.

4. Let $D = \{-48, -14, -8, -2, 0, 1, 3, 7, 10, 12\}$. Determine which of the following statements are true. If a statement is false, then explain why.

(a) $(\forall x \in D)$(if x is odd, then $x > 0$).
(b) $(\forall x \in D)$(if $x > 12$, then $x < 0$).
(c) $(\exists x \in D)(x$ is a perfect cube). (An integer i is a *perfect cube* if $i = n^3$ for some integer n.)

5. Using the Tarskian predicates in Example 5 on page 38, translate the following English sentences into logical sentences.

(a) Something is white.
(b) Some circle is white.
(c) All squares are black.
(d) No squares are black.
(e) All triangles are west of d.
(f) A triangle is west of d.
(g) There is a triangle that is north of d but not west of a.
(h) Some triangle is not gray.
(i) Every triangle is either west of a or north of b.
(j) No square has the same color as b.

6. Using the Tarskian predicates in Example 5, translate the following five logical sentences into English sentences. Then determine the truth or falsity of each of these statements in the Tarskian world of Fig. 2.1.

(a) $\forall x(I(x) \to (T(x) \vee S(x)))$.
(b) $\forall x(B(x) \to (T(x) \vee S(x)))$.
(c) $\exists y(C(y) \wedge \neg N(y,d))$.
(d) $\exists y(C(y) \wedge N(y,d))$.
(e) $\exists y(C(y) \wedge \neg W(y,g))$.

7. Determine whether the sentences are true or false in the universe $\mathbb{R}$.

(a) $(\forall x > 2)(x < 4 \to x^2 < 16)$.
(b) $(\forall x < 2)(x < 4 \to x^2 < 16)$.
(c) $(\exists x < 2)(x < 4 \wedge x^2 < 4)$.
(d) $(\exists x > 2)(x < 4 \wedge x^2 < 4)$.

8. Determine whether the sentences are true or false.

(a) $(\forall x \in \mathbb{N})(x > 2 \to 3x < 2^x)$.
(b) $(\forall x \in \mathbb{N})(x > 4 \to 3x < 2^x)$.
(c) $(\exists x \in \mathbb{Z})(\frac{1}{5+x} \in \mathbb{N})$.
(d) $(\exists x \in \mathbb{N})(\frac{1}{5+x} \in \mathbb{Z})$.

9. Evaluate the truth sets:

(a) $\{x \in \mathbb{R} : (\exists y \in \mathbb{R})(x = y^2)\}$.
(b) $\{x \in \mathbb{R} : (\forall y \in \mathbb{R})(x < y^2)\}$.
(c) $\{x \in \mathbb{R} : (\forall y > 2)(x < y^2 + 1)\}$.
(d) $\{x \in \mathbb{R} : (\exists y > 2)(x < y^2 + 1)\}$.
(e) $\{x \in \mathbb{Z} : (\exists y \in \mathbb{Z})(x = y^2)\}$.
(f) $\{w \in \mathbb{Z} : (\exists x \in \mathbb{Z})(w = 3x)\}$.
(g) $\{q \in \mathbb{Q} : (\exists x \in \mathbb{Q}^+)(qx = 1)\}$.
(h) $\{q \in \mathbb{Q} : (\forall x \in \mathbb{Q})(qx = x)\}$.

2.3 Quantifiers and Negation

In this section we introduce laws that involve the negation of a quantified assertion. These laws are very useful when dealing with the denial of a complicated mathematical statement. For example, in advanced mathematics (e.g., real analysis) one may be given a mathematical statement, say ψ, and then be asked to work with its negation $\neg\psi$. Of course, $\neg\psi$ means "it not the case that ψ holds," but it may not be clear as to what such a negative statement really means. "Positive" assertions are just easier to understand. In this section we will show how one can rephrase a negative statement into an equivalent, positive statement that is more understandable. Having such a skill is very important when developing mathematical proofs.

Consider the sentence "Not everyone is rich," that is, "It is not the case that everyone is rich." We know that this sentence is true. Why is it true? Because there are some people who are not rich. Let us express the sentence symbolically. Let $R(x)$ be the predicate "x is rich." So, the sentence "Not everyone is rich" can be expressed as $\neg\forall x R(x)$. Since $\neg\forall x R(x)$ also means that "some people are not rich," we see that $\neg\forall x R(x) \Leftrightarrow \exists x \neg R(x)$. In other words, $\neg\forall x R(x)$ and $\exists x \neg R(x)$ are equivalent ways of saying the same thing.

Suppose now that a financial crisis has struck the world and, as a result, there are no longer any rich people; that is, it is not the case that there is a person who is rich. Thus, $\neg\exists x R(x)$. Note that $\neg\exists x R(x)$ also means that everyone fails to be rich, that is, $\forall x \neg R(x)$. Consequently, $\neg\exists x R(x) \Leftrightarrow \forall x \neg R(x)$ and, as a result, $\neg\exists x R(x)$ and $\forall x \neg R(x)$ mean the same thing.

We will now more formally state what was observed in the previous two paragraphs. Let $P(x)$ be any predicate. The assertion $\forall x P(x)$ means that "for every x, the statement $P(x)$ is true." So, the assertion $\neg\forall x P(x)$ means that "it is not the case that every x makes $P(x)$ true." Thus, there must be an x that does not make $P(x)$ true, which can be expressed as $\exists x \neg P(x)$. This reasoning is reversible as we will now show. The assertion $\exists x \neg P(x)$ means that "there is an x that makes $P(x)$ false." Hence, $P(x)$ is not true for all x, that is, $\neg\forall x P(x)$. Therefore, $\neg\forall x P(x)$ and $\exists x \neg P(x)$

are logically equivalent. Similar reasoning will show that $\neg\exists x P(x)$ and $\forall x \neg P(x)$ are equivalent. We now formally state these important new logic laws that connect quantifiers with negation.

Quantifier Negation Laws (QNL)

1. $\neg\forall x P(x) \Leftrightarrow \exists x \neg P(x)$.
2. $\neg\exists x P(x) \Leftrightarrow \forall x \neg P(x)$.

Remark 2.3.1. The quantifier negation law 1 states that the expression "not for all" is equivalent to the phrase "some are not." Similarly, law 2 asserts that the expression "not for some" is equivalent to "all are not."

Using quantifier negation laws, together with propositional logic laws, a negative statement can be translated into an equivalent *positive* statement. Roughly speaking, a positive statement is one that does not contain the negation symbol, or one that has the negation symbol appearing as far inside the statement as is possible.

Example 1. Using quantifier negation laws and propositional logic laws, express the *negations* of the following statements as positive statements.

1. Every cat is an animal. [Let $C(x)$: "x is a cat" and $A(x)$: "x is an animal."]
2. No cat is an animal.
3. Someone in this class does not do their homework. [Let $C(x)$: "x is in this class" and $H(x)$: "x does his/her homework."]

Solution. Using our solutions in Example 4 on page 36, we shall first translate each of the English statements into logical form. We will take negations of these logical forms and "push through" the negation symbol using quantifier negation laws and propositional logic laws. The result will then be expressed in English.

1. ENGLISH: "Every cat is an animal."

   LOGICAL FORM: $\forall x(C(x) \to A(x))$.

   ENGLISH NEGATION: "It is not the case that every cat is an animal."

   LOGICAL NEGATION: $\neg\forall x(C(x) \to A(x))$.

$$\begin{aligned} \neg\forall x(C(x) \to A(x)) &\Leftrightarrow \exists x \neg(C(x) \to A(x)) && \text{by Quantifier Negation Law} \\ &\Leftrightarrow \exists x(C(x) \wedge \neg A(x)) && \text{by Conditional Law.} \end{aligned}$$

   POSITIVE ENGLISH FORM: "There is a cat that is not an animal."

2. ENGLISH: "No cat is an animal."

   LOGICAL FORM: $\neg\exists x(C(x) \wedge A(x))$.

   ENGLISH NEGATION: "It is not the case that no cat is an animal."

   LOGICAL NEGATION: $\neg\neg\exists x(C(x) \wedge A(x))$.

$$\neg\neg\exists x(C(x) \wedge A(x)) \Leftrightarrow \exists x(C(x) \wedge A(x)) \quad \text{by Double Negation Law.}$$

   POSITIVE ENGLISH FORM: "Some cat is an animal."

3. ENGLISH: "Someone in this class does not do their homework."

   LOGICAL FORM: $\exists x(C(x) \wedge \neg H(x))$.

   ENGLISH NEGATION: "It is not the case that someone in this class does not do their homework."

   LOGICAL NEGATION: $\neg\exists x(C(x) \wedge \neg H(x))$.

$$\begin{aligned}
\neg\exists x(C(x) \wedge \neg H(x)) &\Leftrightarrow \forall x\neg(C(x) \wedge \neg H(x)) && \text{by Quantifier Negation Law}\\
&\Leftrightarrow \forall x(C(x) \rightarrow \neg\neg H(x)) && \text{by Conditional Law}\\
&\Leftrightarrow \forall x(C(x) \rightarrow H(x)) && \text{by Double Negation Law.}
\end{aligned}$$

   POSITIVE ENGLISH FORM: "Everyone in this class does their homework." Ⓢ

The reasoning used to justify the quantifier negation laws can also be used to verify negation laws for bounded quantifiers. For example, let $T(x)$ represent the predicate "$x^2 > 0$." The statement $\neg(\forall x \in \mathbb{R})T(x)$ asserts that $(\forall x \in \mathbb{R})T(x)$ is false. Is it false? Yes, because there is an $x \in \mathbb{R}$ (namely $x = 0$) that satisfies $\neg T(x)$ and so, we have that $(\exists x \in \mathbb{R})\neg T(x)$ holds. Thus, the statement $\neg(\forall x \in \mathbb{R})T(x)$ implies $(\exists x \in \mathbb{R})\neg T(x)$. Since this argument is reversible, we can conclude that

$$\neg(\forall x \in \mathbb{R})T(x) \Leftrightarrow (\exists x \in \mathbb{R})\neg T(x).$$

More generally, given any set A and predicate $P(x)$, the following logic laws connect bounded set quantifiers with negation.

Negation Laws for Bounded Set Quantifiers

1. $\neg(\forall x \in A)P(x) \Leftrightarrow (\exists x \in A)\neg P(x)$.
2. $\neg(\exists x \in A)P(x) \Leftrightarrow (\forall x \in A)\neg P(x)$.

In the negation laws for bounded set quantifiers, notice that when you push the negation symbol through a bounded set quantifier, the quantifier changes and the negation symbol passes over '$x \in A$'.

Example 2. We can express $\neg(\forall x \in A)(R(x) \rightarrow \neg S(x))$ into a positive form, as follows:

$$\begin{aligned}
\neg(\forall x \in A)(R(x) \rightarrow \neg S(x)) &\Leftrightarrow (\exists x \in A)\neg(R(x) \rightarrow \neg S(x)) && \text{by QNL}\\
&\Leftrightarrow (\exists x \in A)(R(x) \wedge \neg\neg S(x)) && \text{by CL(3)}\\
&\Leftrightarrow (\exists x \in A)(R(x) \wedge S(x)) && \text{by DNL.}
\end{aligned}$$

Negation laws, similar to those for bounded set quantifiers, also apply to the bounded number quantifiers. For example, let $\mathbb{R}$ be our universe and let $P(x)$ be a predicate. Note that the statement $\neg(\forall x > a)P(x)$ asserts that not every real number $x > a$ satisfies $P(x)$. This just means $(\exists x > a)\neg P(x)$; that is, there is a real number $x > a$ that fails to satisfy $P(x)$. Therefore, $\neg(\forall x > a)P(x) \Leftrightarrow (\exists x > a)\neg P(x)$.

Negation Laws for Bounded Number Quantifiers

1. $\neg(\forall x > a)P(x) \Leftrightarrow (\exists x > a)\neg P(x)$.
2. $\neg(\exists x > a)P(x) \Leftrightarrow (\forall x > a)\neg P(x)$.
3. $\neg(\forall x < a)P(x) \Leftrightarrow (\exists x < a)\neg P(x)$.
4. $\neg(\exists x < a)P(x) \Leftrightarrow (\forall x < a)\neg P(x)$.

Similar negation laws apply when the bounded number quantifiers involve the relations $\leq$ and $\geq$. In the above laws, if you move the negation symbol through a bounded number quantifier, then the quantifier changes and the negation symbol completely passes over the relations $x > a$ and $x < a$. For example, we can express $\neg(\forall x > 0)(x^2 > 1 \to x < 4)$ as a positive statement as follows:

$$\begin{aligned}
\neg(\forall x > 0)(x^2 > 1 \to x < 4) &\Leftrightarrow (\exists x > 0)\neg(x^2 > 1 \to x < 4) && \text{by QNL} \\
&\Leftrightarrow (\exists x > 0)(x^2 > 1 \wedge x \not< 4) && \text{by CL} \\
&\Leftrightarrow (\exists x > 0)(x^2 > 1 \wedge x \geq 4) && \text{by laws of inequality.}
\end{aligned}$$

Exercises 2.3

1. Using quantifier negation laws and propositional logic laws, translate each of the following assertions into positive statements. (The universe is $\mathbb{R}$.)

(a) $\neg(\forall x > 2)(x < 4 \to x^2 < 16)$.
(b) $\neg(\forall x < 2)(x < 4 \to x^2 < 16)$.
(c) $\neg(\exists x < 2)(x < 4 \wedge x^2 < 4)$.
(d) $\neg(\exists x > 2)(x < 4 \wedge x^2 < 4)$.
(e) $\neg(\forall x \in \mathbb{N})(x > 2 \to 3x < 2^x)$.
(f) $\neg(\forall x \in \mathbb{N})(x > 4 \to 3x < 2^x)$.

2. Express the negation of the following statement as a positive statement: *For all real numbers x, if x is rational and positive, then $\sqrt{x}$ is irrational.* State your result in English. [The square root operation $\sqrt{x}$ is defined on page 95.]

3. Consider the following statement and proposed negation of this statement.
Statement: *Every prime number is odd.*
Proposed Negation: *Every prime number is even.*
Determine whether the proposed negation is correct. If it is not correct, then write a correct negation.

4. Using quantifier negation laws and propositional logic laws, express each statement in as a positive statement. (The universe is the set of real numbers.)

(a) $\neg(\forall x > 3)(|x - 10| < \frac{1}{2} \to |x^2 - 100| < \frac{1}{3})$.
(b) $\neg(\exists x < -4)(|x + 6| < \frac{1}{100} \wedge |\sin(x) - 100| \geq \frac{1}{30})$.

5. Consider the Tarskian predicates in Example 5 on page 38. Using quantifier negation laws and propositional logic laws, express the *negation* of each of the following assertions as a positive statement. Then write your result in idiomatic English.

(a) $\forall x(I(x) \to (T(x) \vee S(x)))$.
(b) $\forall x(B(x) \to (T(x) \vee S(x)))$.
(c) $\exists y(C(y) \wedge \neg N(y,d))$.
(d) $\exists y(C(y) \wedge N(y,d))$.
(e) $\exists y(C(y) \wedge \neg W(y,g))$.

6. Consider the predicates:

$C(x)$: "x is in the class."
$M(x)$: "x is a mathematics major."

Using these predicates, translate the following English sentences into logical sentences. Then express the result as an equivalent positive logical statement. Write your final result as an English sentence.

(a) It is not the case that everyone in the class is a mathematics major.
(b) It is not the case that someone in the class is a mathematics major.
(c) It is not the case that no one in the class is a mathematics major.

2.4 Statements Containing Multiple Quantifiers

In the previous section we considered sentences that contain a single quantifier. Using both of the quantifiers $\forall$ and $\exists$, one can construct more intricate sentences. For example, $\forall x \exists y L(x,y)$ where $L(x,y)$ is a statement with free variables x and y. In this section we discuss how to determine the truth or falsity of a logical statement with multiple quantifiers. We do this by first translating the logical sentence into English. Such translations can be challenging. In any case, it is best to translate the quantifiers from "left to right" just as we read a sentence in English.

2.4.1 Interpreting Adjacent Quantifiers

Adjacent quantifiers have the form $\exists x \exists y$, $\forall x \forall y$, $\forall x \exists y$ and $\exists x \forall y$. In the next few examples, we will see how to interpret and understand statements with adjacent quantifiers. When a statement contains adjacent quantifiers, one should address the quantifiers, one at a time, in the order in which they are presented.

Example 1. Let the universe be a group of people and let $L(x,y)$ mean "x likes y." What do the formulas mean?

1. $\exists x \exists y L(x,y)$
2. $\exists y \exists x L(x,y)$.

Solution. Before we begin, we note that "x likes y" also means that "y is liked by x." We will translate these formulas from "left to right" as follows:

1. $\exists x \exists y L(x,y)$ means "there is a person x such that $\exists y L(x,y)$," that is, "there is a person x who likes some person y." Thus, $\exists x \exists y L(x,y)$ means that "someone likes someone."
2. $\exists y \exists x L(x,y)$ states that "there is a person y such that $\exists x L(x,y)$," that is, "there is a person y who is liked by some person x." So, $\exists y \exists x L(x,y)$ means that "someone is liked by someone."

We conclude that $\exists x \exists y L(x,y)$ and $\exists y \exists x L(x,y)$ both mean the same thing. Ⓢ

Example 2. Let the universe be a group of people and let $L(x,y)$ mean "x likes y." What do the formulas mean in English?

1. $\forall x \forall y L(x,y)$
2. $\forall y \forall x L(x,y)$.

Solution. We will work again from "left to right" as follows:

1. $\forall x \forall y L(x,y)$ means "for every person x, we have $\forall y L(x,y)$," that is, "for every person x, we have that x likes every person y." Hence, $\forall x \forall y L(x,y)$ means that "everyone likes everyone."
2. $\forall y \forall x L(x,y)$ states "for each person y, we have $\forall x L(x,y)$," that is, "for each person y, we have that y is liked by every person x." So, $\forall y \forall x L(x,y)$ means "everyone is liked by everyone."

We conclude that $\forall x \forall y L(x,y)$ and $\forall y \forall x L(x,y)$ both mean the same thing. Ⓢ

Example 3. Let the universe be a group of people and let $L(x,y)$ mean "x likes y." What do the formulas mean in English?

1. $\forall x \exists y L(x,y)$
2. $\exists y \forall x L(x,y)$.

Solution. We will translate the formulas as follows:

1. $\forall x \exists y L(x,y)$ means "for every person x we have $\exists y L(x,y)$," that is, "for every person x there is a person y that x likes." Thus, $\forall x \exists y L(x,y)$ means that "everyone likes someone."
2. $\exists y \forall x L(x,y)$ states that "there is a person y such that $\forall x L(x,y)$," that is, "there is a person y who is liked by every person x." In other words, $\exists y \forall x L(x,y)$ means "someone is liked by everyone."

We conclude that $\forall x \exists y L(x,y)$ and $\exists y \forall x L(x,y)$ do **not** mean the same thing. Ⓢ

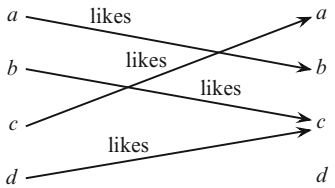


Fig. 2.2 A World where $\forall x \exists y L(x,y)$ is true, because everyone likes someone

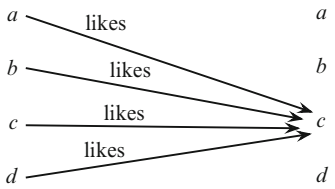


Fig. 2.3 A World where $\exists y \forall x L(x,y)$ is true, since someone is liked by everyone

To clarify the conclusion obtained in our solution of Example 3, consider the universe $U = \{a,b,c,d\}$ consisting of just four individuals with names as given. For this universe, Fig. 2.2 presents a world where $\forall x \exists y L(x,y)$ is true, where we portray the property $L(x,y)$ using the "arrow notation" $x \xrightarrow{\text{likes}} y$. Figure 2.2 thus illustrates a world where "everyone likes someone."

On the other hand, the statement $\exists y \forall x L(x,y)$ says that there is an individual who is extremely popular because everyone likes this person. Figure 2.3 gives a world where $\exists y \forall x L(x,y)$ is true. In this particular world, "someone is liked by everyone."

Let us focus our attention on Fig. 2.2. We observed that $\forall x \exists y L(x,y)$ is true in the world depicted in this figure. Furthermore, notice that $\exists y \forall x L(x,y)$ is actually false in this world. Thus, $\forall x \exists y L(x,y)$ is true and $\exists y \forall x L(x,y)$ is false in the world presented in Fig. 2.2. We can now conclude that $\forall x \exists y L(x,y)$ and $\exists y \forall x L(x,y)$ do not mean the same thing.

Our solution in Example 1 shows that $\exists x \exists y L(x,y)$ and $\exists y \exists x L(x,y)$ both mean "someone likes someone." This supports the true logical equivalence:

$$\exists x \exists y L(x,y) \Leftrightarrow \exists y \exists x L(x,y).$$

Similarly, Example 2 confirms the true logical equivalence:

$$\forall x \forall y L(x,y) \Leftrightarrow \forall y \forall x L(x,y).$$

Consequently, interchanging adjacent quantifiers of the same kind does not change the meaning. On other hand, Example 3 shows that $\forall y \exists x L(x,y)$ and $\exists x \forall y L(x,y)$ are not logically equivalent and thus, do not mean the same thing. Hence, interchanging unlike adjacent quantifiers can change the meaning.

Adjacent quantifiers of a different type are referred to as **mixed quantifiers**. We end this discussion with some good news and some bad news.

- Good News: **Adjacent quantifiers of the same type can be interchanged.**
- Bad News: **Adjacent quantifiers of a different type may not be interchanged.**

Example 4. Let the universe be $\mathbb{R}$, the set of real numbers. Determine the truth value of the following sentences:

1. $\exists y \forall x (x + y = x)$.
2. $\forall x \exists y (x + y = 0)$.
3. $\exists y \forall x (x + y = 0)$.
4. $\exists x \exists y (x + y = 0)$.
5. $\forall x \forall y (x + y = 0)$.

Solution.

1. $\exists y \forall x (x + y = x)$ states that there is a real number y such that $\forall x (x + y = x)$ is true. Since the real number $y = 0$ satisfies the formula $\forall x (x + y = x)$, we see that the statement $\exists y \forall x (x + y = x)$ is true.
2. $\forall x \exists y (x + y = 0)$ means that for every real number x the statement $\exists y (x + y = 0)$ is true. Since for every real number x the real number $y = -x$ makes the equation $x + y = 0$ true, we see that the sentence $\forall x \exists y (x + y = 0)$ is true.
3. $\exists y \forall x (x + y = 0)$ asserts that there is a real number y such that $\forall x (x + y = 0)$ is true. So, y is the additive inverse of *every* real number x. This is false; for example, when $x = (1 - y)$ we see that $x + y \neq 0$. Thus, the statement $\exists y \forall x (x + y = 0)$ is false.
4. $\exists x \exists y (x + y = 0)$ means that there is some real number x such that $\exists y (x + y = 0)$ is true. Since for $x = 3$ and $y = -3$, we have that the equation $x + y = 0$ is true. We conclude that the assertion $\exists x \exists y (x + y = 0)$ is true.
5. $\forall x \forall y (x + y = 0)$ declares that for every real number x the statement $\forall y (x + y = 0)$ is true. So, $\forall x \forall y (x + y = 0)$ means the equation $x + y = 0$ is true for *all* real numbers x and y. This is false, since $2 + 3 \neq 0$. Ⓢ

2.4.2 *Interpreting Non-adjacent Quantifiers*

When a logical statement contains multiple quantifiers, one should address all of the quantifiers one at a time, in the order presented. In our next example we are given a Tarskian World and some logical formulas with multiple quantifiers. We will be asked to determine if the formulas are true or false in this given world.

Example 5. Consider the Tarskian World in Fig. 2.4 where each individual is labeled with a name and recall the Tarskian predicates:

- $T(x)$ means "x is a triangle." $C(x)$ means "x is a circle." $S(x)$ means "x is a square."
- $I(x)$ means "x is white." $G(x)$ means "x is gray." $B(x)$ means "x is black."

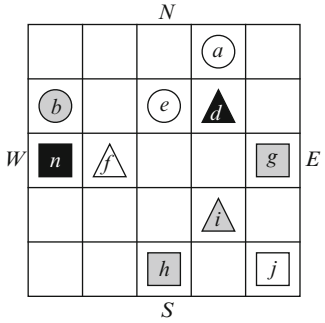


Fig. 2.4 A Tarskian World

- $N(x,y)$ means "x is on the northern side of y."
- $W(x,y)$ means "x is on the western side of y."
- $K(x,y)$ means "x has the same color as y."

Determine the truth or falsity of each of the following statements. The universe consists of all the objects in the world given in Fig. 2.4.

1. $\forall x(T(x) \rightarrow \exists y(S(y) \wedge K(x,y)))$.
2. $\exists x(S(x) \wedge \forall y(C(y) \rightarrow W(x,y)))$.
3. $\exists x(S(x) \wedge \forall y(C(y) \rightarrow W(y,x)))$.
4. $\forall y(T(y) \rightarrow \exists x N(x,y))$.
5. $\forall y(C(y) \rightarrow \exists x N(x,y))$.
6. $\forall y(C(y) \rightarrow \exists x N(y,x))$.
7. $\forall x(S(x) \rightarrow \exists y(C(y) \wedge K(x,y)))$.
8. $\exists y(T(y) \wedge \forall x(C(x) \rightarrow \neg K(x,y)))$.

Before we evaluate the truth value of these eight statements, we observe in our next remark that the concepts of north and west can be used, respectively, to also describe the notions of south and east. This will make it easier to translate some of the above logical sentences into English.

Remark 2.4.1. The predicate $N(x,y)$ means that "x is north of y." Thus, we can interpret $\exists x N(x,y)$ as "some one is north of y." Furthermore, the statement $N(x,y)$ also means that "y is south of x." So, we can interpret $\exists y N(x,y)$ as "some one is south of x." In addition, $W(x,y)$ means that "x is west of y," which also means that "y is east of x." Hence, we can translate $\forall x W(x,y)$ as "every one is west of y" and translate $\forall y W(x,y)$ as "every one is east of x."

Solution [for Example 5]. It turns out that statements given in 1, 3, 4 and 6 are the only ones that are true in Fig. 2.4. To verify this, we translate each logical formula into an English statement. Upon reading the English statement, we shall then determine if the original logical formula is true or false in the Tarskian world.

1. *For every triangle x there is a square of the same color as x.* TRUE.
2. *There is a square x who is west of all the circles.* FALSE.
3. *There is a square x who is east of all the circles.* TRUE.
4. *For every triangle y there is an individual who is north of y.* TRUE.
5. *For every circle y there is an individual who is north of y.* FALSE.
6. *For every circle y there is an individual who is south of y.* TRUE.
7. *For every square x there is a circle having the same color as x.* FALSE.
8. *There is a triangle y whose color is different than that of every circle.* TRUE. Ⓢ

2.4.3 Translating English Statements with Multiple Quantifiers

Many statements in English contain more than one quantifier, as do many mathematical statements. For example:

1. Some student in the algebra class is smarter than all of the students in the calculus class.
2. For every real number x there is a real number y such that $x+y=0$.

In the next two examples we will translate English sentences containing more than one quantifier into logical form. Translating such sentences can become quite perplexing unless we approach this task in a very systematic way. We shall be using a step-by-step method that can be summarized as follows: Start with the outer structure of the English sentence and then move inward. The intermediate steps will contain a mix of both English and logical notation. We illustrate this method in our solutions of the next two examples.

Example 6 Analyze the logical forms of the following statements, that is, identify the relevant predicates and write each statement symbolically, looking for possible hidden quantifiers and hidden logical connectives. The universe here is the set of all people.

1. Everybody in the office has a co-worker he does not like.
2. There is a voter who likes all of the candidates.

Solution We are working in the universe of all people. Once we introduce a variable that is to be associated with a quantifier, then for any new quantifier we must choose a new variable. We now translate, in order, the given sentences 1 and 2.

1. We first identify the predicates used in this sentence. Let $O(x)$: "x is in the office," $C(x,y)$: "x and y are co-workers," and $L(x,y)$: "x likes y." We shall translate this sentence in steps. 'Everybody' means 'for all x.' So, we can rewrite the sentence as: "For all x, if x is in the office, then there is a co-worker that x does not like." Therefore, we get

$$\forall x(O(x) \rightarrow \text{there is a co-worker that } x \text{ does not like}). \tag{2.1}$$

Now we translate "there is a co-worker that x does not like," which can be rephrased as "there is some person y where x and y are co-workers and x does not like y." This latter expression can be symbolized as $\exists y(C(x,y) \wedge \neg L(x,y))$ which will be substituted into (2.1). Thus, the logical form of the given sentence is $\forall x(O(x) \rightarrow \exists y(C(x,y) \wedge \neg L(x,y)))$.

2. The predicates in the sentence are $V(x)$: "x is a voter," $C(x)$: "x is a candidate," and $L(x,y)$: "x likes y." We can restate the sentence as: "There is an x such that x is a voter and x likes all of the candidates." So, the first step of our translation gives

$$\exists x(V(x) \wedge x \text{ likes all of the candidates}). \tag{2.2}$$

Next, we translate the inside expression "x likes all of the candidates." This expression means that "for all y, if y is a candidate then x likes y." This latter assertion translates to be

$$\forall y(C(y) \rightarrow L(x,y)).$$

So, (2.2) becomes $\exists x(V(x) \wedge \forall y(C(y) \rightarrow L(x,y)))$. Ⓢ

Example 7 Analyze the logical forms of the following mathematical statements about the universe $\mathbb{R}$ of real numbers:

(a) There is a positive number that is an identity element for multiplication.
(b) Every number, except 0, has a multiplicative inverse.
(c) Every positive number has a square root.

Solution We will be working in the universe $\mathbb{R}$, the set of real numbers.

(a) We translate the sentence in steps. The first step produces

$$\exists x(x > 0 \wedge x \text{ an identity element for multiplication}). \tag{2.3}$$

The phrase "x is an identity element for multiplication" means that $\forall y(x \cdot y = y)$. From (2.3), we obtain the translation $\exists x(x > 0 \wedge \forall y(x \cdot y = y))$.

(b) For our first step we obtain

$$\forall x(\text{if } x \neq 0, \text{ then } x \text{ has a multiplicative inverse}). \tag{2.4}$$

This expression "x has a multiplicative inverse" in (2.4) means "there is a y such that $x \cdot y = 1$." Thus, upon putting this expression into logical form, we obtain $\exists y(x \cdot y = 1)$. Hence, our final translation is $\forall x(x \neq 0 \rightarrow \exists y(x \cdot y = 1))$.

(c) We shall translate this sentence in steps. We first get

$$\forall x(\text{if } x > 0, \text{ then } x \text{ has a square root}). \tag{2.5}$$

The expression "x has a square root" means that "$x = y^2$ for some y." Therefore, (2.5) yields the translation $\forall x(x > 0 \rightarrow \exists y(x = y^2))$. Ⓢ

2.4.4 Negating Statements with More than One Quantifier

The rules for the negation of multiple quantifiers follow by repeating the rules for negating a single quantifier.

Example 8 Using the quantifier negation laws establish the following:

1. $\neg\exists x\forall y R(x,y) \Leftrightarrow \forall x\exists y\neg R(x,y)$.
2. $\neg\forall x\exists y R(x,y) \Leftrightarrow \exists x\forall y\neg R(x,y)$.

Solution We derive item 1 as follows:

$$\begin{aligned} \neg\exists x\forall y R(x,y) &\Leftrightarrow \forall x\neg\forall y R(x,y) && \text{by QNL} \\ &\Leftrightarrow \forall x\exists y\neg R(x,y) && \text{by QNL.} \end{aligned}$$

In a similar manner, one can derive item 2. Ⓢ

Example 9 Using the Tarskian predicates that were defined in Example 5 on page 49, translate each of the following statements into a logical formula. Then, using quantifier negation laws and propositional logic laws, express the *negation* of each formula as a positive statement. Finally, write your result in idiomatic English.

(1) For every circle there is a square of the same color.
(2) There is a square that is west of all the triangles.

Solution We shall first translate each English statement into logical form. We will take the negation of this logical form and then "push it through" using quantifier negation laws and propositional logic laws. The end result will then be written in English.

(1) ENGLISH: "For every circle there is a square of the same color."
LOGICAL FORM: $\forall x(C(x) \to \exists y(S(y) \wedge K(x,y)))$.
NEGATION: $\neg\forall x(C(x) \to \exists y(S(y) \wedge K(x,y)))$.

$$\begin{aligned} \neg\forall x(C(x) \to \exists y(S(y) \wedge K(x,y))) &\Leftrightarrow \exists x\neg(C(x) \to \exists y(S(y) \wedge K(x,y))) && \text{QNL} \\ &\Leftrightarrow \exists x(C(x) \wedge \neg\exists y(S(y) \wedge K(x,y))) && \text{CL} \\ &\Leftrightarrow \exists x(C(x) \wedge \forall y\neg(S(y) \wedge K(x,y))) && \text{QNL} \\ &\Leftrightarrow \exists x(C(x) \wedge \forall y(\neg S(y) \vee \neg K(x,y))) && \text{DML} \\ &\Leftrightarrow \exists x(C(x) \wedge \forall y(S(y) \to \neg K(x,y))) && \text{CL.} \end{aligned}$$

ENGLISH: "There is a circle of a different color than the color of every square."

(2) ENGLISH: "There is a square that is west of all the triangles."
LOGICAL FORM: $\exists x(S(x) \wedge \forall y(T(y) \rightarrow W(x,y)))$.
NEGATION: $\neg\exists x(S(x) \wedge \forall y(T(y) \rightarrow W(x,y)))$.

$$\begin{aligned}
\neg\exists x(S(x) \wedge \forall y(T(y) \rightarrow W(x,y))) &\Leftrightarrow \forall x\neg(S(x) \wedge \forall y(T(y) \rightarrow W(x,y))) && \text{QNL}\\
&\Leftrightarrow \forall x(\neg S(x) \vee \neg\forall y(T(y) \rightarrow W(x,y))) && \text{DML}\\
&\Leftrightarrow \forall x(S(x) \rightarrow \neg\forall y(T(y) \rightarrow W(x,y))) && \text{CL}\\
&\Leftrightarrow \forall x(S(x) \rightarrow \exists y\neg(T(y) \rightarrow W(x,y))) && \text{QNL}\\
&\Leftrightarrow \forall x(S(x) \rightarrow \exists y(T(y) \wedge \neg W(x,y))) && \text{CL.}
\end{aligned}$$

ENGLISH: "For each square there's a triangle that is not east of the square." Ⓢ

2.4.5 *The Uniqueness Quantifier*

We have introduced the existential quantifier $\exists$. For example, the expression $\exists xP(x)$ asserts that there is at least one value x that makes $P(x)$ true, but there could be more. Many times in mathematics, one needs to show that there is exactly one value that makes a property true. There is a third quantifier that is sometimes used, though not very often. It's called the *uniqueness quantifier*. This quantifier is written as $\exists!xP(x)$ and means that "there exists a unique x satisfying $P(x)$." This is in contrast with $\exists xP(x)$, which simply means that "at least one x satisfies $P(x)$."

As already noted, the quantifier $\exists!$ is rarely used. One reason for this is that the assertion $\exists!xP(x)$ can be expressed in terms of the other quantifiers $\exists$ and $\forall$. In fact, there are at least two ways of doing this. The two statements below, are equivalent to $\exists!xP(x)$:

$$\exists xP(x) \wedge \forall x\forall y([P(x) \wedge P(y)] \rightarrow x = y) \tag{2.6}$$

$$\exists x[P(x) \wedge \forall y(P(y) \rightarrow x = y)]. \tag{2.7}$$

The statement $\exists xP(x) \wedge \forall x\forall y([P(x) \wedge P(y)] \rightarrow x = y)$ in (2.6) is equivalent to $\exists!xP(x)$ because it means that

> "there is an x such that $P(x)$ holds, and any individuals x and y that satisfy $P(x)$ and $P(y)$ must be the same individual"

which is just another way of saying "there is exactly one value x that satisfies $P(x)$." Similarly, (2.7) is equivalent to $\exists!xP(x)$ as the assertion $\exists x[P(x) \wedge \forall y(P(y) \rightarrow x = y)]$ means that

> "there is an x such that $P(x)$ holds and any individual y that also satisfies $P(y)$ must be the same individual as x."

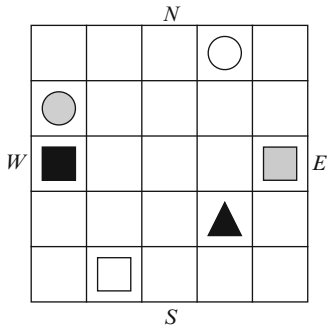


Fig. 2.5 A Tarskian World for Example 10

Example 10 Using the Tarskian predicates (as defined in Example 5 on page 49), determine the truth value of the following logical sentences in the Tarskian world in Fig. 2.5, below.

1. $\exists!x\,T(x)$.
2. $\exists!x\,C(x)$.
3. $\exists!x\,(G(x)\wedge C(x))$.
4. $\forall x\exists!y\,N(y,x)$.
5. $\exists!x\forall y\,(x\neq y\rightarrow W(y,x))$.
6. $\exists!x\forall y\,(x\neq y\rightarrow W(x,y))$.

Solution We first express each of the logical statements into English and then we will determine its truth value in the Tarskian world of Fig. 2.5.

1. $\exists!x\,T(x)$ means that "there is exactly one triangle." This is true!
2. $\exists!x\,C(x)$ states that "there is exactly one circle." This is false!
3. $\exists!x\,(G(x)\wedge C(x))$ declares that "there is exactly one grey circle," and this is true.
4. $\forall x\exists!y\,N(y,x)$ asserts that "for every individual x there is exactly one individual who is north of x." This is false in the given Tarskian world.
5. $\exists!x\forall y\,(x\neq y\rightarrow W(y,x))$ is translated to mean "there is exactly one individual x such that all the individuals who are different than x, are west of x." The grey square is this unique individual. So the statement is true.
6. $\exists!x\forall y\,(x\neq y\rightarrow W(x,y))$ is translated to mean "there is exactly one individual x such that all the individuals who are different than x, are east of x." The statement is false. Ⓢ

Exercises 2.4

1. Let the universe be a group of people and let $L(x,y)$ mean "x likes y." What do the following formulas mean in English?

(a) $\forall y \exists x L(x,y)$
(b) $\exists x \forall y L(x,y)$.

Show that these two statements are not logically equivalent by constructing a world, as in Example 3, where one statement is true while the other is false.

2. Write the following statement in logical form and then write a negation of this statement in English: *All even integers are twice some integer.* [Use the predicate $E(x)$ for "x is even," and let the universe be the set of integers.]

3. Determine whether the statements are true or false in the universe $\mathbb{R}$.

(a) $\forall a \exists x (x^2 = a)$.
(b) $\forall x \exists a (a + x = 0)$.
(c) $\exists a \forall x (a + x = 0)$.
(d) $\forall x \exists a (ax = 0)$.
(e) $\forall x \exists y (x < y)$.
(f) $\exists y \forall x (x < y)$.
(g) $\forall x \forall y (x = y \to x^2 = y^2)$.
(h) $\forall x \forall y (x^2 = y^2 \to x = y)$.

4. Using the given predicates, analyze the logical form of the following sentences.

(a) No one likes everyone. (Universe is a group of people.) [Let $L(x,y)$ mean "x likes y."]
(b) Someone likes no one. (Universe is a group of people.) [Let $L(x,y)$ mean "x likes y."]
(c) Every number is the cube of some number. (Universe is $\mathbb{R}$.)
(d) Someone in high school is smarter than everyone in college. (Universe is the set of all students.) [Let $H(x)$ mean "x is in high school," $C(x)$ mean "x is in college," and $S(x,y)$ mean "x is smarter than y."]

5. Using the Tarskian predicates given in Example 5, translate the following six English sentences into logical sentences.

(a) Every gray square is north of some triangle.
(b) Some circle is west of every square.
(c) Some circle is north of a white triangle.
(d) All squares are the same color as some triangle.
(e) All black squares are west of all gray circles.
(f) No square has the same color as any circle.

6. Using quantifier negation laws and propositional logic laws, express each of the following statements as a positive one. The universe is the set of real numbers.

(a) $\neg(\forall x > 2)(\exists y < 2)(x < 4 \to xy < 16)$.
(b) $\neg(\exists x > 2)(\forall y < 2)(x < 4 \to xy < 16)$.
(c) $\neg(\forall x \in \mathbb{N})(\exists y \in \mathbb{Z})(x > 2 \to x < y)$.
(d) $\neg(\exists x \in \mathbb{N})(\forall y \in \mathbb{N})(x < y)$.

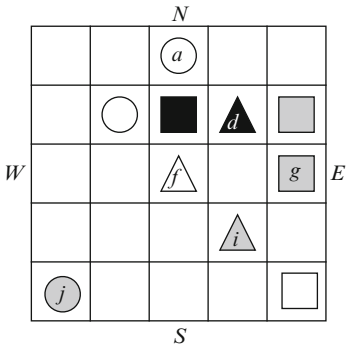


Fig. 2.6a Tarskian World for Exercise 8

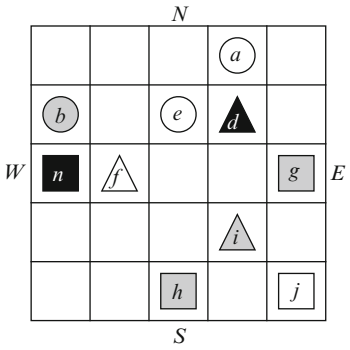


Fig. 2.6b Tarskian World for Exercise 10

7. Using quantifier negation laws and propositional logic laws, express each of the assertions as a positive statement. The universe is the set of real numbers where a and b are constant real numbers.

(a) $\neg(\forall\varepsilon > 0)(\exists\delta > 0)\forall x(|x-a| < \delta \rightarrow |x^2 - b| < \varepsilon)$.
(b) $\neg(\forall\varepsilon > 0)(\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(n > N \rightarrow \left|\frac{1}{n} - a\right| < \varepsilon)$.

8. Consider the Tarskian World in Fig. 2.6a. Determine if the following seven formulas are true or false in this Tarskian world. The Tarskian predicates are defined in Example 5.

(a) $\forall x(C(x) \rightarrow (G(x) \vee I(x)))$.
(b) $\forall y(T(y) \rightarrow N(y,j))$.
(c) $\forall y(S(y) \rightarrow \exists x N(y,x))$.
(d) $\forall y(S(y) \rightarrow \exists x N(x,y))$.
(e) $\forall x(S(x) \rightarrow \exists y(C(y) \wedge K(x,y)))$.
(f) $\exists y(T(y) \wedge \forall x(C(x) \rightarrow \neg K(x,y)))$.
(g) $\exists y \forall x(C(x) \rightarrow W(y,x))$.

9. Take the negation of the logical forms (a)–(e) in Exercise 8 and "push through" the negation using quantifier negation laws and propositional logic laws. Then write this final form of the negation in English.

10. Consider the Tarskian World in Fig. 2.6b. Determine the truth or falsity of each of the following statements in this world.

(a) $\exists! x S(x)$.
(b) $\exists! x(S(x) \wedge B(x))$.
(c) $\exists! x(C(x) \wedge \forall y(T(y) \rightarrow N(x,y)))$.
(d) $\exists! x(C(x) \wedge \forall y(S(y) \rightarrow N(x,y)))$.
(e) $\forall y \exists! x(x \neq y \wedge K(x,y))$.
(f) $\exists y \exists! x(x \neq y \wedge K(x,y))$.

2.5 Valid and Invalid Arguments

An argument in predicate logic is said to be valid if whenever all of the premises are true, then the conclusion is also true. In Table 2.1 we identify four valid arguments that are regularly used in mathematical proofs.

We now present an argument that illustrates the use of universal modus ponens and then an argument that applies universal modus tollens.

Example 1 Consider the classic argument, due to Aristotle:

All men are mortal.
Socrates is a man.
Therefore, Socrates is mortal.

We will show that this argument is valid by universal modus ponens. First we shall symbolize this argument. Let $M(x)$ represent the predicate "x is a man" and let $O(x)$ represent "x is mortal." Now let $\mathfrak{s}$ be a short name for Socrates. Our symbolization of the argument becomes

$$\begin{array}{l} \forall x(M(x) \to O(x)) \\ \underline{M(\mathfrak{s})} \\ \therefore O(\mathfrak{s}) \end{array}$$

We see that Aristotle's argument has the form of universal modus ponens and is therefore valid.

Example 2 We show that the argument below is valid by universal modus tollens.

All natural numbers are integers.
π is not an integer.
Therefore, π is not a natural number.

First we symbolize the argument. Let $N(x)$ represent the predicate "x is a natural number" and let $I(x)$ represent "x is an integer." Thus, we obtain

$$\begin{array}{l} \forall x(N(x) \to I(x)) \\ \underline{\neg I(\pi)} \\ \therefore \neg N(\pi) \end{array}$$

Since the argument has the form of universal modus tollens, it is valid.

Table 2.1 Important valid arguments

$\forall x(P(x) \to Q(x))$ $\underline{P(a)}$ $\therefore Q(a)$	(Universal Modus Ponens)	$\forall x(P(x) \to Q(x))$ $\underline{\neg Q(a)}$ $\therefore \neg P(a)$	(Universal Modus Tollens)
$\underline{\forall x P(x)}$ $\therefore P(a)$	(Universal Instantiation)	$\underline{\exists x P(x)}$ $\therefore P(c)$	(Existential Instantiation)

We now identify two errors in reasoning. These invalid arguments are referred to as *universal converse error* and *universal inverse error*. In mathematical proofs, one must never use these fallacies.

$$\begin{array}{l} \forall x(P(x) \to Q(x)) \\ \underline{Q(a)} \\ \therefore P(a) \end{array} \quad \text{(Converse Error)} \qquad \begin{array}{l} \forall x(P(x) \to Q(x)) \\ \underline{\neg P(a)} \\ \therefore \neg Q(a) \end{array} \quad \text{(Inverse Error)}$$

We shall present examples of the above invalid arguments. The first of which illustrates a universal converse error and the second invalid argument demonstrates a universal inverse error.

Example 3 Verify that the following argument is invalid.

All men are mortal.
Spot is mortal.
Therefore, Spot is a man.

Solution The argument implements a universal converse error. To verify this, let $M(x)$ represent the predicate "x is a man" and let $O(x)$ represent "x is mortal." Let $\mathfrak{s}$ be a short name for my dog Spot. Our symbolization of the argument is

$$\begin{array}{l} \forall x(M(x) \to O(x)) \\ \underline{O(\mathfrak{s})} \\ \therefore M(\mathfrak{s}) \end{array}$$

The premises are true while the conclusion is false. So the argument is invalid. Ⓢ

Example 4 Show that the following argument is not valid.

All natural numbers are integers.
-1 is not a natural number.
Therefore, -1 is not an integer.

Solution The argument employs a universal inverse error. Let $N(x)$ represent the predicate "x is a natural number" and let $I(x)$ represent "x is an integer." Thus, we can symbolize the argument as follows:

$$\begin{array}{l} \forall x(N(x) \to I(x)) \\ \underline{\neg N(-1)} \\ \therefore \neg I(-1) \end{array}$$

Since the premises are true and the conclusion is false, the argument is invalid. Ⓢ

Exercises 2.5

Some of the arguments in Exercises 1–5 are valid by universal modus ponens or modus tollens; others are invalid. State which are valid and which are invalid.

1. All healthy people eat an apple a day.
Johnny is not a healthy person.
$\therefore$ Johnny does not eat an apple a day.

2. All healthy people eat an apple a day.
Johnny eats an apple a day.
$\therefore$ Johnny is a healthy person.

3. All freshmen must take writing.
Dan is a freshman.
$\therefore$ Dan must take writing.

4. All natural numbers are integers.
π is not an integer.
$\therefore$ π is not a natural number.

5. All integers are natural numbers.
-5 is an integer.
$\therefore$ -5 is a natural number.

6. How is it possible for a valid argument to have a false conclusion?

CHAPTER 3

Proof Strategies and Diagrams

The main purpose of this book is to help you develop your mathematical reasoning ability and to help you learn how to use the language and notation of mathematics. In this chapter we will present a variety of proof and assumption strategies. Each proof strategy is motivated by the logical structure of the statement to be proven. A proof strategy will typically be followed by a corresponding assumption strategy. An assumption is an assertion that can be taken to be true.

3.1 Conjecture and Proof in Mathematics

A conjecture is a statement for which there is some evidence supporting the belief that the statement is true. We will now illustrate this idea. Consider the values of the elements in the sequence

$$1, 1+3, 1+3+5, 1+3+5+7, \ldots, 1+3+5+\cdots+(2n-1), \ldots$$

where the first entry is 1, the second entry is $1+3$, the third entry is $1+3+5$, and the n-th entry is $1+3+5+\cdots+(2n-1)$ (the sum of the first n odd numbers). The table of values below is obtained by evaluating the sums of the first six entries of the sequence.

n	$1+3+5+\cdots+(2n-1)$	value
1	1	1
2	1+3	4
3	1+3+5	9
4	1+3+5+7	16
5	1+3+5+7+9	25
6	1+3+5+7+9+11	36

Is there a pattern? Is there a general rule? It appears that the sum of the first n odd numbers is equal to n^2.

Conjecture 1. Let n be a natural number. Then $1+3+5+\cdots+(2n-1)=n^2$.

For another example, let us investigate the values of the elements in the sequence

$$1^2+1+41, 2^2+2+41, 3^2+3+41, 4^2+4+41, 5^2+5+41, \ldots, n^2+n+41, \ldots$$

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_3,

What could possibly be interesting about these numbers? Let us begin by evaluating a sample of the natural numbers that have the form n^2+n+41. By computing the first six values of this sequence, we obtain the table

n	n^2+n+41	value
1	1^2+1+41	43
2	2^2+2+41	47
3	3^2+3+41	53
4	4^2+4+41	61
5	5^2+5+41	71
6	6^2+6+41	83

What property do these values have? They cannot be factored! One observes that the formula n^2+n+41 seems to produce a prime number (see Definition 4.1.3).

Conjecture 2. Let n be a natural number. Then n^2+n+41 is a prime.

Conjecture + Proof = Theorem

Mathematicians state their results in a form called a *theorem* which is a mathematical statement that has been proven to be true. A **conjecture** is a statement that one thinks is plausible but whose truth has not been established. In mathematics one never accepts a conjecture as being true until a proof of the conjecture has been given. A **proof** is a logical argument that establishes the truth of the conjecture. Once a mathematical proof of the conjecture is produced, then the conjecture becomes a **theorem.** For example, one can give a proof of Conjecture 1 (see Exercise 1 on page 122) and thus, this conjecture will become a theorem.

On the other hand, to show that a conjecture is false one must find an assignment of values (an example) which makes all of the assumptions of the conjecture true while making the conclusion false. Such an assignment is called a **counterexample** to the conjecture. Consequently, a counterexample shows that the assumptions of the conjecture do not imply its conclusion. Actually, Conjecture 2 is false and to show this we will give a counterexample. Let $n=41$. So, $n \geq 1$ and

$$n^2+n+41 = (41)^2+41+41 = 41(41+1+1) = 41 \cdot 43,$$

which is not a prime. Hence, Conjecture 2 is false.

Example 1. Find a counterexample showing that the following conjecture is false:

Conjecture. Suppose x and y are real numbers satisfying $x > 2$ and $y < 3$. Then $x(1-y) > 2-x$.

Solution. We must find a counterexample that refutes the conjecture. Let $x = 3$ and $y = 2$. Since $x > 2$ and $y < 3$, the assumptions of the conjecture hold. Furthermore, $x(1-y) = -3$, $2-x = -1$ and $-3 \not> -1$. Thus, the conclusion $x(1-y) > 2-x$ is false. Therefore, the conjecture is false. Ⓢ

3.1.1 How to Prove an Algebraic Equation

Equations play a critical role in modern mathematics. In this text we will establish many theorems that will require us to correctly prove an equation. Because this knowledge is so important and fundamental, our first proof strategy presents two correct methods for proving an equation.

Proof Strategy 3.1.1. To prove an algebraic equation, there are two approaches:

(a) Transform one side of the equation into the other side of the equation.
(b) Derive the equation from any previously given, or assumed, equations.

We shall apply Strategy 3.1.1(a) to prove a well known algebraic identity.

Theorem 3.1.2. *Let a and b be real numbers. Then $(a+b)(a-b) = a^2 - b^2$.*

Proof. We[1] shall start with the left hand side $(a+b)(a-b)$ and transform it into the right hand side as follows:

$$\begin{aligned}
(a+b)(a-b) &= a(a-b)+b(a-b) && \text{by the distributive property}\\
&= a^2 - ab + ba - b^2 && \text{by the distributive property}\\
&= a^2 - ab + ab - b^2 && \text{by commutativity}\\
&= a^2 - b^2 && \text{because } -ab+ab = 0.
\end{aligned}$$

Thus, we have that $(a+b)(a-b) = a^2 - b^2$. □

In the proof of our next theorem, we apply Strategy 3.1.1(b) to prove a new equation from some given equations.

Theorem 3.1.3. *Suppose m, n, i, j are integers satisfying $m = 2i+5$ and $n = 3j$. Then $mn = 6ij + 15j$.*

Proof. We are given that $m = 2i+5$ and $n = 2j$. By multiplying corresponding sides of these two equations, we obtain $mn = (2i+5)(3j)$. Thus, $mn = 6ij + 15j$. □

We now establish that $1 = 0.999\cdots$, where the 9's repeat forever.

[1] Most mathematicians use the term "we" in their proofs. This is considered polite and is intended to include the reader in the discussion.

Example 2. Show that $1 = 0.999\cdots$.

Solution. Let $(\star)$ $x = .999\cdots$. We conclude that

$$\begin{aligned} 10x &= 10 \times .999\cdots && \text{multiplying both sides of } (\star) \text{ by } 10 \\ &= 9.999\cdots && \text{by arithmetic} \\ &= 9 + .999\cdots && \text{by arithmetic} \\ &= 9 + x && \text{by } (\star). \end{aligned}$$

Hence, $10x = 9 + x$. Solving for x, we obtain $x = 1$. Therefore, from $(\star)$ we conclude that $1 = .999\cdots$. Ⓢ

Remark 3.1.4. To prove that an equation $\varphi = \psi$ is true, it is not a correct method of proof to *assume* the equation $\varphi = \psi$ and then work on both sides of this equation to obtain an identity.

The method described in Remark 3.1.4 is a fallacious one and if applied, can produce false equations. For example, this fallacious method can be used to derive the equation $-1 = 1$. To illustrate this, let us assume the equation $-1 = 1$. Now square both sides, obtaining $(-1)^2 = 1^2$, which results in the true equation $1 = 1$. The method cited in Remark 3.1.4 would allow us to conclude that $-1 = 1$ is a true equation. This is complete nonsense. **We never want to apply a method that can produce false equations!**

The Proof Is Complete

It is convenient to have a mark that signals the end of a proof. Mathematicians in the past, would end a proof with letters Q.E.D., an abbreviation for the Latin expression "quod erat demonstrandum," which means "that which was to be proved." In current times, mathematicians typically use the symbol □ to let the reader know that the proof has been completed. In this book we shall do the same.

3.1.2 How to Prove an Inequality

To prove a new inequality from some given inequalities is a little more difficult than proving an equation. The key difference is that you have to correctly use the laws of inequality.

Laws of Inequality 3.1.5. *For all $a, b, c, d \in \mathbb{R}$ the following hold:*

1. *Exactly one of the following holds: $a < b$, $a = b$ or $a > b$.* (Trichotomy Law)
2. *If $a < b$ and $b < c$, then $a < c$.* (Transitivity Law)
3. *If $a < b$, then $a + c < b + c$.*

4. *If $a < b$ and $c > 0$, then $ac < bc$.*
5. *If $a < b$ and $c < 0$, then $ac > bc$.*

We write $a > b$ when $b < a$, and the inequality $a \leq b$ means that $a < b$ or $a = b$. Similarly, $a \geq b$ means that $a > b$ or $a = b$. The Trichotomy Law allows us to assert that if $a \not< b$, then $a \geq b$. It should be noted that one can actually prove law 5 from laws 1–4. Furthermore, using the above laws of inequality, one can prove that $0 < 1$ and $-1 < 0$ (see Proposition 9.1.5 on page 296).

Theorem 3.1.6. *Let a, b, c be real numbers where $a < b$. Then $a - c < b - c$.*

Proof. Let a, b, c be real numbers where $a < b$. From law 3 of 3.1.5, we obtain the inequality $a + (-c) < b + (-c)$. Thus, we infer that $a - c < b - c$. □

Theorem 3.1.7. *Let x be a real number such that $x > 2$. Then $x^2 > x + 1$.*

Proof. Let x be a real number satisfying $(\star)$ $x > 2$. We shall prove that $x^2 > x + 1$. Since $x > 2$, we see that $x > 0$. From $(\star)$ and law 4 of 3.1.5, we conclude that $xx > 2x$. Hence $x^2 > 2x$ and so, $x^2 > x + x$. Because $x > 1$, we obtain $x + x > x + 1$ using law 3 of 3.1.5. Therefore, $x^2 > x + 1$. □

Theorem 3.1.8. *Let a, b, c, d be real numbers and suppose that $a < b$ and $c < d$. Then $a + c < b + d$.*

Proof. Let a, b, c, d be real numbers satisfying (1) $a < b$ and (2) $c < d$. We shall prove that $a + c < b + d$. From (1) and law 3 of 3.1.5, we see that $a + c < b + c$. From (2) and law 3 again, we have that $b + c < b + d$. So, $a + c < b + c < b + d$. Therefore, $a + c < b + d$. □

Theorem 3.1.9. *Let a, b, c, d be positive real numbers satisfying $a < b$ and $c < d$. Then $ac < bd$.*

Proof. Let a, b, c, d be positive real numbers satisfying (1) $a < b$ and (2) $c < d$. We shall prove that $ac < bd$. From (1) and law 4 of the laws of inequality 3.1.5, we conclude that $ac < bc$ because $c > 0$. From (2) and law 4 again, we derive the inequality $bc < bd$ because $b > 0$. So, $ac < bc < bd$. Therefore, $ac < bd$. □

Exercises 3.1

1. Let x and y be real numbers. Prove that $(x - y)(x^2 + xy + y^2) = x^3 - y^3$.
2. Let x and y be real numbers. Prove that $(x + y)(x^2 - xy + y^2) = x^3 + y^3$.
3. Let x and y be real numbers. Prove that $(x + y)^2 = x^2 + 2xy + y^2$.
4. Let x and y be real numbers. Prove that $(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$, using Exercise 3.

5. Let $\varphi > 0$ be a real number satisfying $\varphi^2 - \varphi - 1 = 0$. Prove that $\varphi = \frac{1}{\varphi - 1}$.[2]

6. Let x be a real number such that $x > 1$. Prove that $x^2 > x$.

7. Let x be a real number where $x < 0$. Prove that $x^2 > 0$.

8. Let x be a real number where $x > 0$. Prove that $x^2 > 0$.

9. Let x be a real number where $x \neq 0$. Using Exercises 7 and 8, prove that $x^2 > 0$.

10. Let a and b be distinct real numbers. Using Exercise 9 prove that $a^2 + b^2 > 2ab$.

11. Let x and y be positive real numbers such that $x \neq y$. Using Exercise 10 prove that $\frac{x}{y} + \frac{y}{x} > 2$.

12. Let x be a real number such that $x^2 > x$. Must we conclude that $x > 1$?

13. Let x be a real number satisfying $0 < x < 1$. Prove that $x^2 < x$.

14. Let x be a real number where $x^2 < x$. Must we infer that $0 < x < 1$?

15. Let a and b are real numbers where $a < b$. Prove that $-a > -b$.

16. Let a, b be positive real numbers and let c, d be negative real numbers. Suppose $a < b$ and $c < d$. Prove that $ad > bc$.

17. Find a counterexample showing that the following conjecture is false: *Let a, b, c, d be natural numbers satisfying $\frac{a}{b} \leq \frac{c}{d}$. Then $a \leq c$ and $b \leq d$.*

18. Find a counterexample showing that the following conjecture is false: *Let $m \geq 0$ and $n \geq 0$ be integers. Then $m + n \leq m \cdot n$.*

19. Find a counterexample showing that the following conjecture is false: *Let $x \geq 0$ and $y \geq 0$ be real numbers. Then $\sqrt{x+y} = \sqrt{x} + \sqrt{y}$.*

20. Let a, b, c, d be real numbers. Suppose $a + b = c + d$ and $a \leq c$. Prove that $d \leq b$.

21. Show that $.045000\cdots = .044999\cdots$, where the 0's and the 9's repeat forever.

22. Let a, b, c be real numbers. Prove that $a^2 + b^2 + c^2 \geq ab + bc + ac$.

Exercise Notes: For Exercise 10, if $a \neq b$ then $a - b \neq 0$. For Exercise 20, note that $x \leq y$ if and only if $x - y \leq 0$, for real numbers x and y.

3.2 Using Proof Diagrams as Guides for Proving Theorems

In the previous section we developed techniques for proving theorems about equations and inequalities. For example, in the proof of Theorem 3.1.3 we assumed that $m = 2i + 5$ and $n = 3j$, and then deduced the equation $mn = 6ij + 15j$. Similarly, in our proof of Theorem 3.1.8 we assumed that $a < b$ and $c < d$, and then proved the inequality $a + c < b + d$. These two theorems illustrate the fact that a **mathematical proof** is a logical argument that demonstrates, under certain assumptions, that a

[2]The number φ is called the *golden ratio*. Many artists and architects use the golden ratio φ in their work. When a rectangle of length b and width a satisfies $\frac{b}{a} = \varphi$, it is considered to be more aesthetically pleasing.

particular conclusion must hold. The remainder of this chapter is devoted to showing you how to prove a rich variety of mathematical statements, many of which are more complicated than those we have previously studied. The key reason that we introduced logic in the first two chapters was to prepare you for the task of proving more challenging theorems.

It turns out that all mathematical proofs can be analyzed in terms of the correct use of the logical symbols $\wedge$, $\vee$, $\neg$, $\to$, $\leftrightarrow$, $\forall$, and $\exists$. In this chapter, we will present proof strategies that take advantage of these logical symbols. We will also show you how to "diagram" a mathematical proof. The construction of such a diagram is based upon a logical analysis of the statement to be proven. A generic example of a "general theorem" and a "general program" for proving this theorem follows.

General Theorem. Suppose $G_1, G_2, \ldots, G_n$. Then φ.

The "formal" proof of the theorem will have the form:

Proof. Assume $G_1, G_2, \ldots, G_n$. [The proof of φ goes here.] Therefore, φ. □

How will we find the proof of φ? We first identify the assumptions and conclusion, and then put these statements in the diagram

Assume G_1
$\vdots$
Assume G_n
Prove φ.

Indentation is used to indicate that the proof of the statement φ depends on the assumptions $G_1, G_2, \ldots, G_n$. Using the logical structure of the statement φ, we will be able to break down the proof into simpler parts. We shall then clarify these parts and discover how these parts can be used to form a logically correct proof. The proof will then just assemble these parts to form a coherent argument. In other words, the parts can be thought of as pieces of a jigsaw puzzle and the proof shows how these pieces fit together to form a clear and complete picture.

In this chapter we shall present the key strategies that are used in mathematical proofs. These strategies depend on the logical form of the statement that is to be proven. We will begin by presenting a strategy for proving mathematical statements that have the logical form "If —, then —."

3.3 Statements of the Form $P \to Q$

Consider the conditional statement "If P, then Q." Suppose we want to show that this conditional is true. We know that the conditional is true whenever P is false (see the conditional truth table on page 13). So we only need to show that when P is true,

then Q is also true. Thus, under the assumption that P is true we must verify that Q is also true. We can now introduce a proof strategy for conditional statements.

Proof Strategy 3.3.1. Given a diagram containing the form

Prove $P \to Q$

replace this form with

Assume P
  Prove Q.

Proof strategy 3.3.1 is called a **direct proof**. To assume P means to take P as given. So, when we apply the conditional strategy, we get an additional assumption to use in our proof.

Theorem 3.3.2. *Suppose a and b are positive real numbers. If $a < b$, then $a^2 < b^2$.*

Proof Analysis. The assumption is "a and b are positive real numbers." The conclusion of the theorem is "If $a < b$, then $a^2 < b^2$." We shall construct proof diagrams, using the Conditional Proof Strategy 3.3.1 to get the second diagram:

Assume a and b are positive real numbers.
  Prove $a < b \to a^2 < b^2$.

Assume a and b are positive real numbers.
  Assume $a < b$.
    Prove $a^2 < b^2$.

From the assumptions we have that a and b are positive and that $a < b$. We have to derive the inequality $a^2 < b^2$. To do this, we shall use the given inequality $a < b$ and introduce a^2 and b^2. We first multiply both sides of $a < b$ by the positive real number a, and then multiply both sides of $a < b$ by the positive b. The desired inequality $a^2 < b^2$ will follow. The final diagram will guide our composition of a well-structured proof of the theorem. Ⓐ

Proof. Suppose a and b are positive real numbers. Assume $a < b$. We must prove that $a^2 < b^2$. Multiplying both sides of the inequality $a < b$ by the positive a gives the inequality (i) $a^2 < ab$, and multiplying both sides of the inequality $a < b$ by the positive b yields the inequality (ii) $ab < b^2$. From (i) and (ii) we conclude that $a^2 < ab < b^2$. Therefore, $a^2 < b^2$. □

As stated in the preface, we shall use the symbol Ⓐ to mark the end of a *proof analysis* and we use the symbol □ to identify the end of a *proof*. The proof analysis is not part of the proof and is presented only to help the reader understand how we arrived at the proof.

Let a, p, x, y be real numbers. Suppose that you have the expression $a + x$ and you know that $x = y$. Then, by substitution, you can conclude that $a + x = a + y$. Similarly, if you have the expression px and you know that $x = y$, then you can conclude that $px = py$, again by substitution. These two substitution properties of

equality are frequently used to derive equations. Can these substitution properties be extended to the inequality relation? That is, suppose that you working with the expressions $a+x$ and px and you know that $x < y$. Then, by substitution, can you conclude that $a+x < a+y$ and $px < py$?

We now identify four substitution properties of inequality that we will use to derive inequalities. These substitution properties offer a slightly different way of viewing the laws of inequality given in 3.1.5. Note that a summand is a value that is to be added; for example, in $1+2$ the summands are 1 and 2.

Substitution Properties of Inequality 3.3.3. *Let a, p, x, y real numbers with $p > 0$. Then the following hold:*

(1) *Given the sum $a+x$, if $x < y$, then you can conclude that $a+x < a+y$.*
(Replacing a summand with a larger value yields a larger sum.)

(2) *Given the sum $a+x$, if $x > y$, then you can conclude that $a+x > a+y$.*
(Replacing a summand with a smaller value yields a smaller sum.)

(3) *Given the product px, if $x < y$, then you can conclude that $px < py$.*
(Replacing a factor with a larger value yields a larger product.)

(4) *Given the product px, if $x > y$, then you can conclude that $px > py$.*
(Replacing a factor with a smaller value yields a smaller product.)

Before applying (3) or (4) in the above substitution properties of inequality, one must ensure that $p > 0$. Similar properties hold if we replace each occurrence of $<$ with $\leq$ in (1) and (3), and replace each occurrence of $>$ in (2) and (4) with $\geq$. For example, given the sum $a+x$, if $x \leq y$, then you can conclude that $a+x \leq a+y$. Similarly, given the product px, if $x \geq y$, then you can conclude that $px \geq py$ when $p > 0$. In our next example, we apply properties (2) and (4) of 3.3.3 to derive an inequality.

Example 1. Prove that if $4^n > 4n+1$, then $4^{n+1} > 4(n+1)+1$, when $n \geq 2$ is a natural number.

Solution. Let $n \geq 2$ be a natural number and assume $4^n > 4n+1$. We prove that $4^{n+1} > 4(n+1)+1$ as follows:

$$\begin{aligned}
4^{n+1} &= 4 \cdot 4^n && \text{by property of exponents} \\
&> 4(4n+1) && \text{by (4) of 3.3.3 because } 4^n > 4n+1 \\
&= 16n+4 && \text{by distributivity} \\
&= 4n+4+12n && \text{because } 16n = 4n+12n \\
&= 4(n+1)+12n && \text{by distributivity} \\
&> 4(n+1)+1 && \text{by (2) of 3.3.3 because } 12n \geq 12 \cdot 2 > 1.
\end{aligned}$$

Therefore, $4^{n+1} > 4(n+1)+1$ and this completes the proof. Ⓢ

Suppose that a statement of the form $P \to Q$ is an **assumption** and not the conclusion you are trying to prove. How can you **use** this assumption in your proof? Recall the inference rules modus ponens and modus tollens discussed in Section 1.3.3.

Assumption Strategy 3.3.4. Given a diagram containing the form

$$\text{Assume } P \to Q$$

there are two approaches:

(a) If you are assuming or can prove P, then you can conclude Q by modus ponens.
(b) If you are assuming or can prove $\neg Q$, then you can deduce $\neg P$ by modus tollens.

Exercises 3.3

1. Find a counterexample showing that the following conjecture is false: *Let a and b are real numbers. If $a < b$, then $a^2 < b^2$.*
2. Let x and y be real numbers satisfying $4x + 5y \geq 6$. Prove if $x < 4$, then $y > -2$.
3. Suppose a and b are negative real numbers. Prove that if $a < b$, then $a^2 > b^2$.
4. Let a be a real number. Prove that if $a > 0$, then $(a+4)^2 > a^2 + 16$.
5. Let x be a real number. Prove that if $x \geq 4$, then $x^2 > 2x + 1$.
6. Let n be an integer. Prove that if $n > 3$, then $2^{n-1} + 2^{n-2} + 2^{n-3} \leq 2^n$.
7. Prove that if $(1+x)^n \geq 1 + nx$, then $(1+x)^{n+1} \geq 1 + (n+1)x$, when $x > -1$ is a real number and n is a natural number.
8. Let $n \geq 2$ be an integer. Prove that if $2^n > n$, then $2^{n+1} > n + 1$.
9. Show that Theorem 3.1.9 implies Theorem 3.3.2.

Exercise Notes: For Exercises 4–8, one should review the Substitution Properties of Inequality 3.3.3 and Example 1.

3.4 Statements of the Form $\forall x P(x)$ and $\exists x P(x)$

To prove that a statement is true for all x, we must prove that the statement is true for every element x in our universe. On the other hand, to prove that a statement is true for some x, we may have to first find such an individual and then prove that this individual satisfies the statement.

3.4.1 Working with Universal Statements

Consider the universal statement $\forall x P(x)$. To show that this statement is true, we can let x be completely *arbitrary* and then show that the statement $P(x)$ is true. This idea inspires our next proof strategy that incorporates a strategy for the 'for all' quantifier and the 'bounded for all' quantifier, in parallel.

Proof Strategy 3.4.1. Given a diagram containing one of the forms

replace the form with the corresponding lower diagram, as follows:

If the letter x is already being used in the proof, then use another letter, say y, in the lower diagram.

Theorem 3.4.2. *For all $x \in \mathbb{R}$, if $x > 3$, then $x^2 + 5x + 2 > 25$.*

Proof Analysis. Observe that the statement of the theorem has the logical form

$$(\forall x \in \mathbb{R})(x > 3 \to x^2 + 5x + 2 > 25).$$

We can now construct the following proof diagrams:

Prove $(\forall x \in \mathbb{R})(x > 3 \to x^2 + 5x + 2 > 25)$.

Let $x \in \mathbb{R}$ be arbitrary.
  Prove $(x > 3 \to x^2 + 5x + 2 > 25)$.

Let $x \in \mathbb{R}$ be arbitrary.
  Assume $x > 3$.
    Prove $x^2 + 5x + 2 > 25$.

We applied the $\forall$-Proof Strategy 3.4.1 to get the second diagram. The last diagram was obtained by applying the Conditional Proof Strategy 3.3.1.

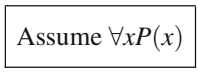
From the assumption $x > 3$ we must prove the inequality $x^2 + 5x + 2 > 25$. So, starting with $x > 3$ we need to derive some information about x^2 and $5x$. Since $x > 3$, we can deduce that (1) $x^2 > 3^2$ by Theorem 3.3.2. Because $x > 3$, we can also conclude that (2) $5x > 5 \cdot 3$. By adding the correspond sides of (1) and (2), the desired inequality $x^2 + 5x + 2 > 25$ will follow. The above final proof diagram will guide our composition of a well-structured proof of the theorem. Ⓐ

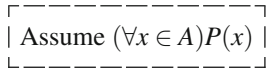
Proof. Let x be an arbitrary real number. Assume $x > 3$. We prove that the inequality $x^2 + 5x + 2 > 25$ holds. Since $x > 3$, we have that $x > 0$ and thus, (1) $x^2 > 9$ by Theorem 3.3.2. Because $x > 3$, we also have that (2) $5x > 15$. From (1) and (2) we obtain $x^2 + 5x > 24$ and hence, $x^2 + 5x + 2 > 26$. Since $26 > 25$, we see that $x^2 + 5x + 2 > 25$. This completes the proof. □

Remark 3.4.3. Theorem 3.4.2 has the logical form $(\forall x \in \mathbb{R})P(x)$ and we started the proof of this theorem with the expression "let x be an arbitrary real number." On the other hand, many mathematicians would begin this proof with just the expression "let x be a real number." Such a proof is then completed under the implicit understanding that x is to be considered as an *arbitrary* real number. In this book, we will also prove statements of the form $(\forall x \in A)P(x)$ by starting the proof with the expression "let $x \in A$." The reader should then consider x to be taken as completely arbitrary.

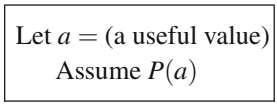
Suppose a statement of the form $\forall x P(x)$ is an **assumption**. How can you **use** this assumption in your proof? You may plug in *any useful* value for x, say a, and then use $P(a)$ in your proof. The expression 'useful value' is ambiguous; however, a useful value is usually one that appears in your proof (see Exercise 16).

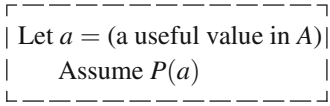
Assumption Strategy 3.4.4. Given a diagram containing one of the forms





replace the form with the corresponding lower diagram





3.4.2 Working with Existential Statements

Consider the mathematical statement:

There is a natural number n such that $2n^2 - 5n + 2$ is a prime number.

This statement can be stated symbolically as

$$(\exists n \in \mathbb{N})(2n^2 - 5n + 2 \text{ is a prime number}).$$

How would you prove this statement? One could begin by trying to find such an n. After looking at $n = 0, 1, 2, 3$, you would discover that when $n = 3$, the number $2n^2 - 5n + 2 = 5$ is a prime. So the statement is true.

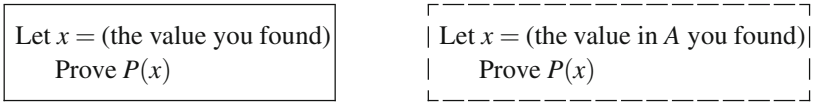
In mathematics one is often required to prove an existential statement having the form $\exists x P(x)$. Our next proof strategy presents two methods for proving the existence

of a mathematical object c satisfying $P(c)$. The first method, Proof Strategy 3.4.5(a), requires one to specifically find such an object and then prove that it satisfies the desired property. The second method (b) is more indirect and presumes that this desired object will be discovered during the proof.

Proof Strategy 3.4.5. Given a diagram containing one of the forms

there are two methods, (a) and (b):

(a) First find a value x, or respectively $x \in A$, such that $P(x)$ is true. Then replace the form with the corresponding next diagram, as follows:



(b) Replace the form with the corresponding next diagram, as follows:

When using Proof Strategy 3.4.5(a) you must first identify the value for x that will satisfy $P(x)$ and, in your resulting proof, you must prove that your value for x satisfies $P(x)$. Furthermore, in the proof you do not have to explain how this value was obtained. On the other hand, when applying method (b), you should expect the correct value for x to "fall out of your proof" (see Remark 3.4.12).

Both of the approaches (a) and (b) in 3.4.5 are referred to as a *constructive proof* strategy because they demonstrate the existence of a certain mathematical object by first identifying or constructing such an object. This is in contrast to a nonconstructive proof which shows that a particular mathematical object must exist, but does not provide a specific example or a means for producing the object.

Theorem 3.4.6. *Let n be an integer. Then there is an $x \in \mathbb{Q}$ satisfying $(n + \frac{1}{2})x = 1$.*

Proof Analysis. First we construct the following proof diagrams (we apply the bounded $\exists$-Proof Strategy 3.4.5(a) to get the second diagram):

Let n be an integer.
  Prove $(\exists x \in \mathbb{Q})[(n + \frac{1}{2})x = 1]$.

Let n be an integer.
  Let $x =$ (the value in $\mathbb{Q}$ you found).
    Prove $(n + \frac{1}{2})x = 1$.

Let n be an integer.
  Let $x = \frac{2}{2n+1}$.
    Prove $(n + \frac{1}{2})x = 1$.

In the last diagram we just identify the value for x in $\mathbb{Q}$ that we found by solving the equation $(n+\frac{1}{2})x = 1$ for x as follows

$$x = \frac{1}{n+\frac{1}{2}} = \frac{1}{\frac{2n+1}{2}} = \frac{2}{2n+1}.$$

Observe that $n+\frac{1}{2} \neq 0$ because n is an integer and thus $2n+1 \neq 0$. Since $x = \frac{2}{2n+1}$ is a ratio of two integers, we see that x is rational. In our proof we must show that this value for x actually satisfies the equation. The last proof diagram will guide our composition of a well-structured proof. *We will not mention, in our proof, how we found our value for x.* In our proof we will just state "Let $x = \frac{2}{2n+1}$," and then prove that this value for x satisfies the equation $(n+\frac{1}{2})x = 1$. Ⓐ

Proof. Let n be an integer. Let $x = \frac{2}{2n+1}$ where $2n+1 \neq 0$ because n is an integer. Clearly, $x \in \mathbb{Q}$. We now show that $(n+\frac{1}{2})x = 1$ as follows:

$$\left(n+\frac{1}{2}\right)x = \left(n+\frac{1}{2}\right)\left(\frac{2}{2n+1}\right) = \left(\frac{2n+1}{2}\right)\left(\frac{2}{2n+1}\right) = 1.$$

Therefore, $(n+\frac{1}{2})x = 1$. □

Now, suppose that a statement of the form $\exists x P(x)$ is an **assumption**. To use this assumption in your proof, you just assign a name, say x_0, to represent an element satisfying $P(x_0)$.

Assumption Strategy 3.4.7. If a diagram contains one of the forms

introduce a **new** variable, say x_0, representing an object that makes $P(x_0)$ true and replace the form with the corresponding lower diagram

Assumption Strategy 3.4.7 is used when one wants to "clarify" the assumption $\exists x P(x)$. Suppose in a proof you are assuming that $\exists x P(x)$ and you want to use this assumption in your proof. To do this, you just assign a variable x_0 that represents a value satisfying $P(x_0)$. A word of warning should be noted here. When you continue with your proof, you must not assign this same variable x_0 again. In other words, you must use a different variable to clarify another assumption in your proof (see Remark 3.4.9).

We shall apply both Proof Strategy 3.4.5(b) and Assumption Strategy 3.4.7 in the proof of our next theorem. Before doing this, we define what it means for an integer

to be even and what it means for an integer to be odd. These fundamental concepts can be described in terms of some very simple equations.

Definition 3.4.8. An integer n is **even** if and only if $n = 2i$ for some $i \in \mathbb{Z}$. An integer n is **odd** if and only if $n = 2k + 1$ for some $k \in \mathbb{Z}$.

One can prove that every integer is either even or odd (see Exercise 3 on page 106) and in this chapter we will implicitly use this fact (see the proofs of Theorems 3.6.2 and 3.8.2).

Remark 3.4.9. *One must use different letters to express different things!* Suppose, for example, that we have that m and n are both even integers. If we write $m = 2i$ and $n = 2i$, then we must have that $m = n$ and one could then conclude that any two even numbers are equal! Of course, this is not the case. For another example, suppose that we have that m is even and n is odd. If we write $m = 2i$ and $n = 2i + 1$, then we must deduce that $n = m + 1$. Using this, one could then say that when $m = 6$ and $n = 11$, it follows that $11 = 6 + 1$. To avoid such absurdities, **one must never use the same letter to clarify two different assertions.**

Theorem 3.4.10. *Suppose n is an integer. If n is even, then $n^2 + 1$ is odd.*

Proof Analysis. First we present the proof diagrams, where the second diagram is the result of applying the Conditional Proof Strategy 3.3.1. By clarifying what it means to be even and what it means to be odd, we obtain the third diagram. Notice that the last line of the final diagram utilizes Proof Strategy 3.4.5(b) because we expect the value for j will be made clear as a result of our proof.

Assume n is an integer.
  Prove (n is even) $\rightarrow$ ($n^2 + 1$ is odd).

Assume n is an integer.
  Assume n is even.
    Prove $n^2 + 1$ is odd.

Assume n is an integer.
  Assume $n = 2i$ for some $i \in \mathbb{Z}$.
    Prove $n^2 + 1 = 2j + 1$ for some $j \in \mathbb{Z}$.

Using the equation $n = 2i$, we need to prove the equation ($\star$) $n^2 + 1 = 2j + 1$ and find the value for j. To prove equation ($\star$), we start with the right hand side $n^2 + 1$ and derive the left hand side of this equation (see 3.1.1(a)). We will use our proof diagrams as a guide in our composition of the following proof. Ⓐ

Proof. Let n be an integer. Assume n is even, that is, (a) $n = 2i$ for some $i \in \mathbb{Z}$. We shall prove that $n^2 + 1$ is odd. From (a) we obtain $n^2 + 1 = (2i)^2 + 1 = 2(2i^2) + 1$. Hence, $n^2 + 1 = 2j + 1$ where $j = 2i^2$ is an integer. Therefore, $n^2 + 1$ is odd. □

Recall that a real number x is rational if and only if $r = \frac{a}{b}$ for integers a, b where $b \neq 0$. In our proof of the next theorem, we will use Assumption Strategy 3.4.7 when we clarify the assumption that "x is rational."

Theorem 3.4.11. *For every real number x, if x is rational then x^2 is also rational.*

Proof Analysis. First we construct the proof diagrams. The $\forall$-Proof Strategy 3.4.1 is used to obtain our second diagram and the final line of the last diagram applies the $\exists$-Proof Strategy 3.4.5(b).

Prove $(\forall x \in \mathbb{R})[(x \text{ is rational}) \rightarrow (x^2 \text{ is rational})]$.

Let x be a real number.
  Prove $(x \text{ is rational}) \rightarrow (x^2 \text{ is rational})$.

Let x be a real number.
  Assume x is rational.
    Prove x^2 is rational.

Let x be a real number.
  Assume $x = \frac{p}{q}$ for some $p, q \in \mathbb{Z}$ where $q \neq 0$.
    Prove $x^2 = \frac{a}{b}$ for some $a, b \in \mathbb{Z}$ where $b \neq 0$.

The last diagram was obtained by writing out what it means for x and x^2 to be rational (see Remark 3.4.9). In the end, we must prove that $x^2 = \frac{a}{b}$ for some $a, b \in \mathbb{Z}$ where $b \neq 0$. Observe that we are applying Proof Strategy 3.4.5(b), as we expect that the correct values for a and b will "fall out of our proof." We are assuming the equation (i) $x = \frac{p}{q}$ where we consider p and q to be known. We must derive an equation of the form (ii) $x^2 = \frac{a}{b}$ where a and b are not known. Starting with x^2 and performing a substitution using (i), we can easily derive the right hand side of (ii) and find the unknowns a and b. The above proof diagrams will guide the composition of the following proof. Ⓐ

Proof. Let x be a real number. Assume that x is rational, that is, $x = \frac{p}{q}$ for some $p, q \in \mathbb{Z}$ where $q \neq 0$. We will prove that x^2 is rational. Since $x = \frac{p}{q}$, using a little algebra, we get $x^2 = (\frac{p}{q})^2 = \frac{p^2}{q^2}$ where $q^2 \neq 0$ because $q \neq 0$. Thus, $x^2 = \frac{a}{b}$ where $a = p^2$, $b = q^2$ are integers and $b \neq 0$. Therefore, x^2 is rational. □

Remark 3.4.12. How does one know when to apply option (a) or (b) of the $\exists$-Proof Strategy 3.4.5? In general, if the assumptions can be expressed as equations and the conclusion can also be expressed as a similar equation, then option (b) is the one to choose first.

3.4.3 Working with Mixed Quantifier Statements

To prove a statement with mixed quantifiers, one must combine Proof Strategies 3.4.1 and 3.4.5. Since the method of proof depends on the order in which the quantifiers are presented, we have the following two different proof strategies:

Proof Strategy 3.4.13. Given a diagram of the form

Prove $\forall x \exists y P(x,y)$

use the diagram

Let x be arbitrary.
  Let $y =$ (the value you found).
    Prove $P(x,y)$.

Proof Strategy 3.4.14. Given a diagram of the form

Prove $\exists y \forall x P(x,y)$

use the diagram

Let $y =$ (the value you found).
  Let x be arbitrary.
    Prove $P(x,y)$.

When using Proof Strategy 3.4.13, the value y (that you must find) typically involves x. On the other hand, when using Proof Strategy 3.4.14, the value y will not involve x. We will apply these strategies, respectively, in our proofs of the next two theorems.

Theorem 3.4.15. *For every real number $x > 0$ there exists a real number $y < 0$ such that $yx^2 + 2x = x$.*

Proof Analysis. Using Proof Strategy 3.4.13, we first construct the following proof diagrams:

Let $x > 0$ be a real number.
  Let $y =$ (the negative real number you found).
    Prove $yx^2 + 2x = x$.

Let $x > 0$ be a real number.
  Let $y = -\frac{1}{x}$.
    Prove $yx^2 + 2x = x$.

In the last diagram we identified the value for $y = -\frac{1}{x}$ that we found by solving the equation $yx^2 + 2x = x$ for y. Since $x > 0$, it follows that $y < 0$. We will use this latter diagram as a guide for our proof. Ⓐ

Proof. Let $x > 0$ be a real number. Let $y = -\frac{1}{x}$ and observe that $y < 0$. We show that $yx^2 + 2x = x$ as follows: $yx^2 + 2x = \left(-\frac{1}{x}\right)x^2 + 2x = -x + 2x = x$. Therefore, $yx^2 + 2x = x$. □

Theorem 3.4.16. *There exists a real number $y > 0$ such that for every real number x we have $y^2x^2 + y^2 - 5yx^2 - 5y = 0$.*

Proof Analysis. Using proof strategy 3.4.14, we first construct the following proof diagram:

Let $y =$ (the positive real number you found).
  Let x be a real number.
    Prove $y^2x^2 + y^2 - 5yx^2 - 5y = 0$.

Let $y = 5$.
  Let x be a real number.
    Prove $y^2x^2 + y^2 - 5yx^2 - 5y = 0$.

In the last diagram we identified the value for $y = 5$ that we found by solving the equation $y^2x^2 + y^2 - 5yx^2 - 5y = 0$ for y. We will use this latter diagram as a guide for our proof. Ⓐ

Proof. Let $y = 5 > 0$ and let x be any real number. We shall establish the equation $y^2x^2 + y^2 - 5yx^2 - 5y = 0$ as follows:

$$y^2x^2 + y^2 - 5yx^2 - 5y = 5^2x^2 + 5^2 - 5 \cdot 5x^2 - 5 \cdot 5 = 25x^2 + 25 - 25x^2 - 25 = 0.$$

Therefore, $y^2x^2 + y^2 - 5yx^2 - 5y = 0$. □

3.4.4 Uniqueness Proofs

Theorems asserting that there is a "unique" element that satisfies a particular property are known as uniqueness theorems. Such theorems pervade mathematics. For example, in linear algebra there are theorems concerning the uniqueness of solutions to certain linear systems of equations. Typically the proof of a uniqueness theorem applies the strategy below.

Proof Strategy 3.4.17. Given a diagram containing the form

Prove $\exists! x P(x)$

replace this form with the diagram

Prove $\exists x P(x)$
Prove $\forall x \forall y[(P(x) \wedge P(y)) \rightarrow x = y]$.

In other words, one can use the abbreviated diagram

Existence: Prove $\exists x P(x)$
Uniqueness: Assume $P(x) \wedge P(y)$
  Prove $x = y$.

Theorem 3.4.18. *Let a, b be real numbers where $a \neq 0$. Then there exists a unique real number x satisfying $ax + b = 0$.*

Proof Analysis. We are given real numbers a and b with $a \neq 0$. First we need to prove that there exists a real number x that satisfies the equation $ax + b = 0$; that is, we need to prove that $(\exists x \in \mathbb{R})(ax + b = 0)$. Afterwards, we must prove that there is only one such solution. We use the uniqueness–Proof Strategy 3.4.17 to obtain the following proof diagram where $P(x)$ is the assertion that $ax + b = 0$:

Assume $a, b \in \mathbb{R}$ where $a \neq 0$.
Existence: Prove $(\exists x \in \mathbb{R})(ax + b = 0)$.
Uniqueness: Assume $(ax + b = 0) \wedge (ay + b = 0)$.
Prove $x = y$.

We apply the $\exists$-Proof Strategy 3.4.5(b) to obtain the diagram

Assume $a, b \in \mathbb{R}$ where $a \neq 0$.
Existence: Let $x =$ (the value in $\mathbb{R}$ you found).
Prove $ax + b = 0$.
Uniqueness: Assume $(ax + b = 0) \wedge (ay + b = 0)$.
Prove $x = y$.

To find this x, we simply solve the equation $ax + b = 0$ for x and obtain $x = -\frac{b}{a}$. We have our final proof diagram

Assume $a, b \in \mathbb{R}$ where $a \neq 0$.
Existence: Let $x = -\frac{b}{a}$.
Prove $ax + b = 0$.
Uniqueness: Assume $(ax + b = 0) \wedge (ay + b = 0)$.
Prove $x = y$.

This final diagram will guide our composition of the following proof. Ⓐ

Proof. Let a, b be real numbers where $a \neq 0$. First we prove that there exists a real number x satisfying $ax + b = 0$. Then we will prove that such an x is unique.

Existence: Let $x = -\frac{b}{a}$. Since $a \neq 0$, we see that x is a real number. Now, since $x = -\frac{b}{a}$, using a little algebra we get $ax + b = a(-\frac{b}{a}) + b = 0$. Therefore, there is an x satisfying $ax + b = 0$.

Uniqueness: Suppose (1) $ax + b = 0$ and (2) $ay + b = 0$. We prove that $x = y$. From (1) and (2) we see that $ax + b = ay + b$. Using algebra, we conclude that $x = y$. □

Exercises 3.4

Prove the following theorems:

1. **Theorem.** Let $c \neq 1$ be a real number. There exists a unique real number x satisfying $\frac{x+1}{x-2} = c$.

2. **Theorem.** Let m be an integer. If m is odd, then m^2 is odd.
3. **Theorem.** Let m be an integer. If m is even, then $m+5$ is odd.
4. **Theorem.** Let m and n be integers. If n is even, then mn is even.
5. **Theorem.** For all integers m and n, if $m-n$ is even, then m^2-n^2 is even.
6. **Theorem.** There exists a real number x such that $|3x-2| = -7x$.
7. **Theorem.** For every real number $a > 3$, there is a real number $x < 0$ such that $|3x-2| = -ax$.
8. **Theorem.** For every real number $y > 0$, there is a real number $x < 0$ such that $y^2 + 2xy = -x^2$.
9. **Theorem.** For each real number x, there is a real number y that satisfies the equation $y^2 - 2xy = 2$.
10. **Theorem.** There is a real number $d > 0$ such that for all real numbers x, if $|x-1| < d$, then $|3x-3| < \frac{1}{2}$.
11. **Theorem.** For every integer n, if n is odd, then $\frac{n-1}{2}$ is an integer.
12. **Theorem.** Let a, b be integers where $a \neq 0$ or $b \neq 0$. There is an integer $n \geq 1$ of the form $n = sa + tb$ for some integers s, t.
13. **Theorem.** For every integer i there is a unique integer j such that $3j + 9i = 6$.
14. **Theorem.** For every real number x there is a unique real number y such that $yx^2 - 3x = -2y$.
15. **Theorem.** There is a unique real number y such that $yx + 6 = 2x + 3y$ for every real number x.
16. **Theorem.** Let $c \leq 2$ be a real number. Suppose $x + y \leq xy$ for all real numbers x and y satisfying $x \geq c$ and $y \geq c$. Then $c = 2$.

Exercise Notes: For Exercises 6 and 7, recall that $|y| = \pm y$ (see Definition 3.6.6). For Exercise 15, to prove uniqueness, note that if the equation $yx+6 = 2x+3y$ holds for every real number x, then the equation holds for $x = 1$. For Exercise 16, note that $2 \geq c$ and $c \geq c$.

3.5 Statements of the Form $P \wedge Q$

Consider the statement "P and Q." To show that this statement is true, we must show that P is true and show that Q is also true. We can now introduce a proof strategy for such "and" statements.

Proof Strategy 3.5.1. Given a diagram containing the form

Prove $P \wedge Q$

replace this form with

Prove P
Prove Q.

The following **divisibility** concept is thoroughly explored in number theory.[3]

Definition 3.5.2. Let m and n be integers. We write $m \mid n$ if and only if $n = mk$ for some $k \in \mathbb{Z}$.

When $m \mid n$ we may say the "m divides n" or "m evenly divides n." To assert that m does not evenly divide n, we write $m \nmid n$.

Theorem 3.5.3. *For all integers n, if $12 \mid n$ then $3 \mid n$ and $4 \mid n$.*

Proof Analysis. First we construct the proof diagrams (the second diagram is the result of applying the $\wedge$-Proof Strategy 3.5.1):

Let n be an integer.
  Assume $12 \mid n$.
    Prove $3 \mid n$ and $4 \mid n$.

Let n be an integer.
  Assume $12 \mid n$.
    Prove $3 \mid n$.
    Prove $4 \mid n$.

Let n be an integer.
  Assume $n = 12k$ for some $k \in \mathbb{Z}$.
    Prove $n = 3i$ for some $i \in \mathbb{Z}$.
    Prove $n = 4j$ for some $j \in \mathbb{Z}$.

The last diagram was obtained by writing out what it means to be divisible by 12, 3, and 4 (respectively). These diagrams guide the following proof. Ⓐ

Proof. Let n be an integer. Assume $12 \mid n$, that is, assume (1) $n = 12k$ for some $k \in \mathbb{Z}$. We prove that $3 \mid n$ and $4 \mid n$. We shall first prove that $3 \mid n$. Using some algebra on (1), we deduce that $n = 12k = 3(4k)$. Thus, $n = 3i$ for some $i \in \mathbb{Z}$, namely $i = 4k$. Hence, $3 \mid n$. We now prove that $4 \mid n$. Again, by applying algebra on (1), we conclude that $n = 12k = 4(3k)$. So, $n = 4j$ where $j = 3k$ is an integer. Therefore, $4 \mid n$. □

Suppose that a statement of the form $P \wedge Q$ is an **assumption** and not the conclusion you are trying to prove. In this case you can assume that P is true and assume that Q is also true.

[3]Number theory is the branch of mathematics that is principally concerned with the integers and their properties.

Assumption Strategy 3.5.4. Given a diagram containing the form

Assume $P \wedge Q$

replace this form with the diagram

Assume P
Assume Q.

Theorem 3.5.5. *The product of two odd integers is odd.*

Proof Analysis. First we construct three proof diagrams (Assumption Strategy 3.5.4 yields the second diagram):

Let m and n be integers.
  Assume m and n are odd.
    Prove mn is odd.

Let m and n be integers.
  Assume m is odd.
  Assume n is odd.
    Prove mn is odd.

Let m and n be integers.
  Assume $m = 2i+1$ for some $i \in \mathbb{Z}$.
  Assume $n = 2j+1$ for some $j \in \mathbb{Z}$.
    Prove $mn = 2k+1$ for some $k \in \mathbb{Z}$.

We will use the above diagrams to direct our proof. Ⓐ

Proof. Let m and n be integers. Assume m and n are odd, that is, $m = 2i+1$ and $n = 2j+1$ for some $i, j \in \mathbb{Z}$. We will prove that mn is odd. To do this, note that $mn = (2i+1)(2j+1)$ by the assumptions. Using some algebra, we deduce that

$$mn = (2i+1)(2j+1) = 4ij + 2(i+j) + 1 = 2(2ij+i+j) + 1.$$

Thus, $mn = 2k+1$ where $k = (2ij+i+j)$ is in $\mathbb{Z}$. Therefore, mn is odd. □

We will now show that divisibility is *transitive*, that is, for all integers a, b, and c, if a divides b and b divides c, then a divides c.

Theorem 3.5.6 (Transitivity of Divisibility). *For all integers a, b and c, if $a \mid b$ and $b \mid c$, then $a \mid c$.*

Proof Analysis. We build the proof diagrams:

Let a, b, and c be integers.
  Assume $a \mid b$ and $b \mid c$.
    Prove $a \mid c$.

Let a, b and c be integers.
  Assume $a \mid b$.
  Assume $b \mid c$.
    Prove $a \mid c$.

Let a, b and c be integers.
  Assume $b = ak$ for some $k \in \mathbb{Z}$.
  Assume $c = bi$ for some $i \in \mathbb{Z}$.
    Prove $c = aj$ for some $j \in \mathbb{Z}$.

In our proof we shall assume (1) $b = ak$ and (2) $c = bi$ where we consider k, i to be known integers. We have to derive an equation of the form (3) $c = aj$ where the integer j is not known. Substituting the value for b given by (1) into (2), we will derive equation (3) and find the unknown j. Ⓐ

Proof. Let a, b, and c be integers. Assume $a \mid b$ and $b \mid c$, that is, $b = ak$ for some $k \in \mathbb{Z}$ and $c = bi$ for some $i \in \mathbb{Z}$. We prove that $a \mid c$. Since $b = ak$ and $c = bi$, we deduce that $c = bi = (ak)i = a(ki)$. Thus, $c = aj$ for some $j \in \mathbb{Z}$, namely $j = ki$. Therefore, $a \mid c$. □

Our next theorem shows that if $m \mid a$ and $m \mid b$, then m evenly divides any "linear combination" of a and b, that is, $m \mid (sa + tb)$ for all $s, t \in \mathbb{Z}$.

Theorem 3.5.7. *Let m, a, and b be integers. If $m \mid a$ and $m \mid b$, then $m \mid (sa + tb)$ for all integers s and t.*

Proof. Let m, a, and b be integers. Assume $m \mid a$ and $m \mid b$, that is, assume that

$$a = mi \tag{3.1}$$

$$b = mj \tag{3.2}$$

for some $i, j \in \mathbb{Z}$. Let $s, t \in \mathbb{Z}$. We shall prove that $m \mid (sa + tb)$. From (3.1) and (3.2) we obtain

$$sa + tb = s(mi) + t(mj) = m(si + tj).$$

Thus, $sa + tb = mk$ where $k = si + tj$ is an integer. Therefore, $m \mid (sa + tb)$. □

Exercises 3.5

Prove the following theorems:

1. **Theorem.** Let a, b be real numbers. If $a > 0$ and $b > 0$, then $(a+b)^2 > a^2 + b^2$.
2. **Theorem.** Let a, b be real numbers. If $a < 0$ and $b < 0$, then $(a+b)^2 > a^2 + b^2$.
3. **Theorem.** For all $x \in \mathbb{R}$ and $y \in \mathbb{R}$, if x and y are rational, then $x+y$ is rational.
4. **Theorem.** For all integers a, b, and c, if $c \mid a$ and $c \mid b$, then $c \mid (a+b)$, $c \mid (a-b)$, and $c \mid (ai)$ for any integer i.
5. **Theorem.** Let n be an integer. If $21 \mid n$, then $3 \mid n$ and $7 \mid n$.
6. **Theorem.** Suppose n is an integer. If $3 \mid n$ and $7 \mid n$, then $21 \mid n$.
7. **Theorem.** For every integer n, if n is odd, then $4 \mid (n^2 - 1)$.
8. **Theorem.** Suppose m, n are positive integers. If $m \mid n$, then $m \leq n$.
9. **Theorem.** For all positive integers a and b, if $a \mid b$ and $b \mid a$, then $a = b$.
10. **Theorem.** Let a, b, x, y be negative integers. If $a < b$ and $x < y$, then $ax > by$.
11. **Theorem.** Let $a > 0$ and $b < -4$ be real numbers. Then $ab + b < -4(a+1)$.
12. **Theorem.** For all integers a and b, if $a \mid b$, then $a^2 \mid b^2$.
13. **Theorem.** Suppose m, a, b, c, d are integers. If $m \mid (a-b)$ and $m \mid (c-d)$, then $m \mid ((a+c)-(b+d))$.
14. **Theorem.** Let m, a, b, c, d be integers. If $m \mid (a-b)$ and $m \mid (c-d)$, then $m \mid (ac - bd)$.
15. **Theorem.** Let a, b, d be real numbers. If $0 \leq a < d$ and $0 \leq b < d$, then $a - b < d$ and $b - a < d$.
16. **Theorem.** If $0 \leq a < d$ and $0 \leq b < d$, then $-d < a - b < d$ where $a, b, d \in \mathbb{R}$.
17. **Theorem.** For all integers a, b, c, d, if $a \neq c$ and $ad \neq bc$, then there exists a unique rational number x such that $\frac{ax+b}{cx+d} = 1$.

Exercise Notes: For Exercises 1 and 2, one should review the substitution properties of inequality 3.3.3. For Exercise 6, use the identity $n = 7n - 6n$. For Exercise 17, after you identify x in your proof, you must prove that $cx + d \neq 0$.

3.6 Statements of the Form $P \vee Q$

Consider the statement "P or Q." To show that this statement is true, we must verify that either P is true or that Q is true. So we can try to prove P or try to prove Q. This direct approach can sometimes be difficult, as we may then have to work with an inadequate set of assumptions. Fortunately, logic offers us an easier approach. We know that $(P \vee Q)$, $(\neg P \to Q)$, and $(\neg Q \to P)$ are all logically equivalent. Thus, to prove $(P \vee Q)$, we can either prove $(\neg P \to Q)$ or prove $(\neg Q \to P)$. In either case, we obtain a new assumption that we can use in our proof. We can now introduce a proof strategy for such "or" statements.

Proof Strategy 3.6.1. Given a diagram containing the form

Prove $P \vee Q$

there are three approaches:

(a) Replace the form with the diagram

Assume $\neg P$
Prove Q.

(b) Replace the form with the diagram

Assume $\neg Q$
Prove P.

(c) When using a *division by cases*,[4] in each case, prove P or prove Q.

Theorem 3.6.2. *Let m and n be integers. If mn is even, then m is even or n is even.*

Proof Analysis. We first construct the proof diagrams. The second diagram is obtained by applying Strategy 3.6.1(a).

Assume m and n are integers.
  Assume mn is even.
    Prove m is even or n is even.

Assume m and n are integers.
  Assume mn is even.
    Assume m is not even.
      Prove n is even.

Assume m and n are integers.
  Assume $mn = 2i$ for some $i \in \mathbb{Z}$.
    Assume $m = 2j + 1$ for some $j \in \mathbb{Z}$.
      Prove $n = 2k$ for some $k \in \mathbb{Z}$.

To say that m is not even means that m is odd. The last diagram was obtained by writing out what it means for mn and n to be even, and what it means for m to be odd. We will be assuming the equations (1) $mn = 2i$ and (2) $m = 2j + 1$. Upon substituting the value for m given by (2) into (1), we will obtain an 'isolated' occurrence of n for which we can solve. These diagrams will guide our proof. Ⓐ

Proof. Let m and n be integers. Suppose that mn is even and so, (1) $mn = 2i$ for some $i \in \mathbb{Z}$. We shall prove that either m is even or n is even. Assume that m is not even. Then $m = 2j + 1$ for some $j \in \mathbb{Z}$. Substituting $m = 2j + 1$ into equation (1), we obtain the equation $(2j + 1)n = 2i$. After some algebra, we see that $n = 2i - 2jn$. Hence, $n = 2k$ where $k = i - jn$ is an integer. Therefore, n is even. □

[4] See the proof of Theorem 3.6.5.

It will often happen that you want to prove that some property R holds when you know that one thing or another is true, say P or Q, but you do not know which one is true. Our next assumption strategy will addresses this situation.

Assumption Strategy 3.6.3. Given a diagram containing the form

Assume $P \vee Q$
Prove R

there are three approaches:

(a) Use a *proof by cases*; that is, replace the form with

Case 1: Assume P
Prove R
Case 2: Assume Q
Prove R.

(b) If you are assuming or can prove $\neg P$, then you can deduce Q. Now prove R.
(c) If you are assuming or can prove $\neg Q$, then you can infer P. Now prove R.

Item (b), in Assumption Strategy 3.6.3, is an application of disjunctive syllogism. In other words, if you are given $P \vee Q$ as an assumption and you know that $\neg P$ holds, then you can conclude that Q must be true. Now, using Q, try to prove R. Similarly, for item (c) of 3.6.3.

Theorem 3.6.4. *If one of two integers is even, then their product is even.*

Proof Analysis. We construct the following proof diagrams (the third diagram is obtained by applying the division into cases Strategy 3.6.3(a)):

Assume n and m are integers.
    Prove (m is even or n is even) $\rightarrow$ (mn is even).

Assume n and m are integers.
    Assume m is even or n is even.
        Prove mn is even.

Assume n and m are integers.
    Case 1: Assume m is even.
        Prove mn is even.
    Case 2: Assume n is even.
        Prove mn is even.

Assume n and m are integers.
    Case 1: Assume $m = 2i$ for some $i \in \mathbb{Z}$.
        Prove $mn = 2k$ for some $k \in \mathbb{Z}$.
    Case 2: Assume $n = 2i$ for some $i \in \mathbb{Z}$.
        Prove $mn = 2k$ for some $k \in \mathbb{Z}$.

The diagrams will guide our proof of the theorem. Because the proof under Case 2 is essentially identical to that of Case 1, we will just consider Case 1 in our proof. The expression "without loss of generality," or W.L.O.G., is a common way of signaling to the reader that the proof will be treating just one of the cases as the argument for the other case (or cases) is very similar. Ⓐ

Proof. Suppose m and n are integers. Assume that either m is even or n is even. Without loss of generality, we shall only consider the case that m is even and then prove that mn is even. So, suppose that m is even, that is, (1) $m = 2i$ for some $i \in \mathbb{Z}$. By multiplying both sides of (1) by n, we get $mn = (2i)n$. Hence, $mn = (2i)n = 2(in)$ and $mn = 2k$ for some $k \in \mathbb{Z}$, namely $k = in$. Thus, mn is even. □

Assumption Strategy 3.6.3(a) asserts that when you assume a statement of the form $P \vee Q$ and you need to prove a statement R, then you should break up the proof into two cases. Suppose you are assuming a statement of the form $P_1 \vee P_2 \vee \cdots \vee P_n$ and you need to prove R. Then you could use a division of cases and break up your proof into n many cases. In the first case you assume P_1 and prove R. In the second case, you assume P_2 and prove R. You must then continue to do this for all of the remaining statements $P_3, \ldots, P_n$, as well. We will demonstrate this idea in the proof of our next theorem, where we break up the proof into three cases.

Theorem 3.6.5. *Let x and q be integers such that $x = 3q + i$ where i is either 0, 1, or 2. Then x^2 has the form $3k$ or $3k+1$ for some integer k.*

Proof Analysis. We construct a division by cases diagram:

Let x and q be integers.
  Case 1: Assume $x = 3q$.
    Prove $x^2 = 3k$ or $x^2 = 3k+1$ for some $k \in \mathbb{Z}$.
  Case 2: Assume $x = 3q+1$.
    Prove $x^2 = 3k$ or $x^2 = 3k+1$ for some $k \in \mathbb{Z}$.
  Case 3: Assume $x = 3q+2$.
    Prove $x^2 = 3k$ or $x^2 = 3k+1$ for some $k \in \mathbb{Z}$.

In each of the three cases, we need to prove an "or" statement and thus, our proof will illustrate an application of Proof Strategy 3.6.1(c). Our diagram guides the following well-structured composition of the proof. Ⓐ

Proof. Let x and q be integers satisfying $x = 3q + i$ where i is 0, 1, or 2. We prove that x^2 has the form $3k$ or $3k+1$ for some integer k. We shall use a division by cases.

CASE 1: Assume $x = 3q$. Then $x^2 = (3q)^2 = 3(3q^2)$. Therefore, $x^2 = 3k$ where $k = 3q^2$ is an integer.

CASE 2: Assume $x = 3q+1$. Then $x^2 = (3q+1)^2 = 9q^2+6q+1 = 3(3q^2+2q)+1$. Therefore, $x^2 = 3k+1$ where $k = 3q^2+2q$ is an integer.

CASE 3: Assume $x = 3q+2$. Then

$$x^2 = (3q+2)^2 = 9q^2+12q+4 = 9q^2+12q+3+1 = 3\left(3q^2+4q+1\right)+1.$$

Therefore, $x^2 = 3k+1$ where $k = 3q^2+4q+1$ is an integer. □

To use a division by cases, in a proof, you must first identify all the possibilities, that is, all the cases. Then you must prove the conclusion under each case. The definition of absolute value, below, is one that is given in most calculus books. Observe that the definition of $|x|$ is based on two cases; namely, $x \geq 0$ and $x < 0$. Consequently, proofs about the absolute value function often use a division by these two cases.

Definition 3.6.6. Given a real number x, the **absolute value** of x, denoted by $|x|$, is defined by

$$|x| = \begin{cases} x, & \text{if } x \geq 0; \\ -x, & \text{if } x < 0. \end{cases}$$

One of the most commonly used functions in mathematics is the absolute value function. The absolute value $|x|$ is simply the distance from x to the origin. In more advanced mathematics, especially in real analysis, the absolute value function is used extensively (see Chapter 9 and Theorems 9.2.3–9.2.4).

Exercises 3.6

Prove the following theorems:

1. **Theorem.** If x is an integer, then x^2 has the form $4k$ or $4k+1$ for an integer k.
2. **Theorem.** Let n and m be integers. If mn is even, then m is even or n is even.
3. **Theorem.** Let n and m be integers. If $m+n$ is odd, then m is odd or n is odd.
4. **Theorem.** If $a > 0$ is a real number, then $1 < a + \frac{1}{a}$.
5. **Theorem.** Let a and b be real numbers. If $0 \leq a \leq b$, then $a^2 \leq b^2$.
6. **Theorem.** Let a, b, x, y be non-negative real numbers. If $a \leq b$ and $x \leq y$, then $ax \leq by$.
7. **Theorem.** Let n and d be integers where $d \geq 1$. There exists an integer k such that $n - dk \geq 0$.
8. **Theorem.** For all real numbers x we have that $x^2 \geq 0$.
9. **Theorem.** For all real numbers x and y, if $x \geq 2$ and $y \geq 2$, then $xy \geq x+y$
10. **Theorem.** Let x be a real number. Then $|x| \geq 0$.
11. **Theorem.** Let x be a real number. Then $x \leq |x|$.
12. **Theorem.** Let x, y be real numbers. Then $|xy| = |x|\,|y|$.
13. **Theorem.** Let x be a real number. Then $x^2 = |x|^2$.

Exercise Notes: For Exercise 1, use the fact that x is either even or odd. For Exercises 2 and 3, recall that every integer is either even or odd. For Exercise 4, since either $a < 1$ or $a = 1$ or $1 < a$, use a division by cases. More specifically, use the proof diagram

Assume $a > 0$ is a real number.
Case 1: Assume $a < 1$.
Prove $1 < a + \frac{1}{a}$.
Case 2: Assume $a = 1$.
Prove $1 < a + \frac{1}{a}$.
Case 3: Assume $1 < a$.
Prove $1 < a + \frac{1}{a}$.

Note: $0 < \frac{1}{a}$. For Exercise 5, there are four cases to consider: (1) $0 = a < b$; (2) $0 = a = b$; (3) $0 < a = b$; (4) $0 < a < b$. Recall Theorem 3.3.2. For Exercise 6, use a division by cases and Exercise 10 on page 84. For Exercise 7, there are two cases to consider: $n \leq 0$ and $n > 0$. For Exercise 9, consider the cases $x \leq y$ and $y \leq x$. Note that when $y \geq 2$ and $x > 0$, one can conclude that $xy \geq 2x = x + x$ (review the Laws of Inequality 3.1.5). For Exercises 10 and 11, there are two cases (1) $x \geq 0$; (2) $x < 0$. Also, note that if $x < 0$ then $|x| = -x$ by Definition 3.6.6. For Exercise 12, there are four cases (1) $x, y \geq 0$; (2) $x < 0$ and $y \geq 0$; (3) $x \geq 0$ and $y < 0$; (4) $x, y < 0$.

3.7 Statements of the Form $P \leftrightarrow Q$

Consider the biconditional statement "P if and only if Q." Suppose we want to show that such a statement is true. Recall that this biconditional is equivalent to the conjunction of the two conditional statements "if P then Q" and "if Q then P" (see the biconditional law on page 17). We arrive at the following proof strategy.

Proof Strategy 3.7.1. Given a diagram containing the form

Prove $P \leftrightarrow Q$

replace this form with

Prove $P \rightarrow Q$
Prove $Q \rightarrow P$.

In other words one has to prove $P \rightarrow Q$ and prove $Q \rightarrow P$, separately.

Theorem 3.7.2. *Let n be an integer. Then $6 \mid n$ if and only if $2 \mid n$ and $3 \mid n$.*

Proof. Let n be an integer. We will prove that $6 \mid n$ if and only if $2 \mid n$ and $3 \mid n$.

($\Rightarrow$). First we prove that if $6 \mid n$, then $2 \mid n$ and $3 \mid n$. Assume $6 \mid n$. Thus, $n = 6i$ for some $i \in \mathbb{Z}$. So, $n = 6i = 2(3i) = 3(2i)$ and hence, $2 \mid n$ and $3 \mid n$.

($\Leftarrow$). Now we prove that if $2 \mid n$ and $3 \mid n$, then $6 \mid n$. Assume $2 \mid n$ and $3 \mid n$. Thus there are integers i and j such that $n = 2i$ and $n = 3j$. Therefore,

$$n = 3n - 2n = 3(2i) - 2(3j) = 6i - 6j = 6(i - j).$$

Hence, $n = 6k$ where $k = i - j$ is an integer. Therefore, $6 \mid n$. □

In the above proof of Theorem 3.7.2, the annotations ($\Rightarrow$) and ($\Leftarrow$) are added as a courtesy to the reader. The arrow $\Rightarrow$ is used to abbreviate "implies." In other words, ($\Rightarrow$) and ($\Leftarrow$) are added to make it clear where each conditional is being established in the proof.

Remark 3.7.3. Another way to prove some statements of the form $P \leftrightarrow Q$ is to write a string of equivalences starting with P and ending with Q.

If more than two statements $P_1, P_2, \ldots, P_n$ are all equivalent, then the statements are either *all false at the same time*, or are *all true at the same time*. Furthermore, to prove that statements $P_1, P_2, \ldots, P_n$ are all equivalent, we can prove

$$P_1 \rightarrow P_2 \rightarrow \cdots \rightarrow P_n \rightarrow P_1.$$

That is, first we prove $P_1 \rightarrow P_2$, then we prove $P_2 \rightarrow P_3$ and we continue in this manner until we prove $P_{n-1} \rightarrow P_n$. For the final step, we must "complete the circle" by proving $P_n \rightarrow P_1$. This proof technique is sometimes called a *round-robin* proof.

Suppose that you know that a certain biconditional statement holds and you want to use it in a proof. Since the biconditional statement "P if and only if Q" is equivalent to the assertion that "if P then Q" and "if Q then P" both hold, the following assumption strategy can be used.

Assumption Strategy 3.7.4. Given a diagram containing the form

Assume $P \leftrightarrow Q$

replace this form with

Assume $P \rightarrow Q$
Assume $Q \rightarrow P$.

Remark 3.7.5. When *assuming* $P \leftrightarrow Q$, if you know that P holds, then you can conclude Q. Also, if you know that Q is true, then you can conclude P. On the other hand, if you know that $\neg P$ holds, then you can deduce $\neg Q$. Similarly, if you have that $\neg Q$ is true, then you can deduce $\neg P$.

Exercises 3.7

Prove the following theorems:

1. **Theorem.** Let n be an integer. Then n is even if and only if $n + 1$ is odd.

2. **Theorem.** Let n and k be integers where k is odd. Then n is odd if and only if $n = 2i + k$ for some integer i.
3. **Theorem.** Let n be an integer. Then $3 \mid n$ if and only if $3 \mid (5n + 6)$.
4. **Theorem.** Let n be an integer. Then $15 \mid n$ if and only if $3 \mid n$ and $5 \mid n$.
5. **Theorem.** Let x be a real number. Then $|x| = 0$ if and only if $x = 0$.
6. **Theorem.** Let $c > 0$. For $x \in \mathbb{R}$, $|x| = c$ if and only if $x = -c$ or $x = c$.
7. **Theorem.** Let $c > 0$. For all real numbers x, $|x| < c$ if and only if $-c < x < c$.

Exercise Notes: For Exercise 2, in your proof of the direction ($\Rightarrow$) use the algebraic identity $n = n - k + k$. For Exercise 3, in your proof of the direction ($\Leftarrow$) use the algebraic identity $n = 6n - 5n$. For Exercise 4, the proof is very similar to the proof of Theorem 3.7.2. Note that $n = 6n - 5n$. For Exercises 6 and 7, in your proof of both directions ($\Leftarrow$) and ($\Rightarrow$) there are two cases: (1) $x \geq 0$ and (2) $x < 0$.

3.8 Indirect Proof

We are now familiar with "direct proofs" of mathematical statements. A direct proof establishes that a statement is true by using the definitions and previous results to logically derive the statement. An indirect proof of a statement can take two different forms: proof by contraposition and proof by contradiction. Proof by contraposition establishes the truth of an alternative statement whose truth implies the truth of the original statement. Proof by contradiction argues that the original statement cannot possibly be false and therefore, it must be true. In mathematics, indirect proofs are very common. The first indirect proof that we investigate is proof by contraposition, and then we shall pursue proof by contradiction.

3.8.1 Proof by Contraposition

There may be times when it is not easy to prove a conditional statement, say $\psi \to \varphi$; that is, using the assumption ψ it may be difficult to prove φ. In this case, logic can come to the rescue. Since $\psi \to \varphi$ and its contrapositive $\neg\varphi \to \neg\psi$ are logically equivalent, we can prove the contrapositive instead. Consequently, we can assume $\neg\varphi$ and try to prove $\neg\psi$. This alternative approach is called *proof by contraposition*.

Proof Strategy 3.8.1. Given a diagram containing the form

Prove $P \to Q$

to apply proof by contraposition, replace this form with

Assume $\neg Q$
Prove $\neg P$.

Let n be an integer. Our next theorem will show that if n^2 is even, then n is even. The proof of this theorem uses proof by contraposition. Why not apply the conditional strategy? Such a direct proof leads to some difficulties as we will now show. Assuming n^2 is even, let us try to prove that n is even. Since n^2 is even, we can write $n^2 = 2i$ for some integer i. To prove that n is even, we must derive an equation of the form $n = 2j$ where j **must be an integer**. Let us try to do this! From the equation $n^2 = 2i$, we divide both sides by n and obtain $n = 2(\frac{i}{n})$. There are two difficulties with this equation. First n must be different from 0. The second difficulty is more serious. Clearly $\frac{i}{n}$ is a rational number; but how do we know that $\frac{i}{n}$ is an integer? To avoid these problems, we use proof by contraposition.

Theorem 3.8.2. *Let n be an integer. If n^2 is even, then n is even.*

Proof Analysis. First we construct the following proof diagrams where the second diagram results by applying the proof by contraposition Strategy 3.8.1:

Assume n is an integer.
  Prove (n^2 is even) $\rightarrow$ (n is even).

Assume n is an integer.
  Assume n is not even.
    Prove n^2 is not even.

Assume n is an integer.
  Assume $n = 2k+1$ for some $k \in \mathbb{Z}$.
    Prove $n^2 = 2j+1$ for some $j \in \mathbb{Z}$.

If an integer is not even, then it must be odd. We applied this fact[5] to obtain our last diagram. These diagrams will guide our proof of the theorem. Ⓐ

Proof. Let n be an integer. We will prove that if n^2 is even, then n is even. We shall use proof by contraposition. Assume n is not even, that is, n is odd. Let $k \in \mathbb{Z}$ be so that $n = 2k+1$. We prove that n^2 is not even. Since $n = 2k+1$, using some algebra, we see that $n^2 = (2k+1)^2 = 4k^2+4k+1 = 2(2k^2+2k)+1$. Hence, $n^2 = 2j+1$ where $j = 2k^2+2k$ is an integer. Therefore, n^2 is not even. □

Theorem 3.8.3. *Suppose a, b, c are real numbers with $a > b$. If $ac \leq bc$, then $c \leq 0$.*

Proof Analysis. The assumptions are "a, b, c are real numbers" and "$a > b$." We have to prove the statement "If $ac \leq bc$, then $c \leq 0$." Now we construct the proof diagrams using the contraposition Strategy 3.8.1 to obtain the second diagram:

[5] See Exercise 3 on page 106.

Assume a, b, c are real numbers.
Assume $a > b$.
    Prove $ac \leq bc \to c \leq 0$.

Assume a and b are real numbers.
Assume $a > b$.
    Assume $\neg(c \leq 0)$.
        Prove $\neg(ac \leq bc)$.

Assume a and b are real numbers.
Assume $a > b$.
    Assume $c > 0$.
        Prove $ac > bc$.

Since $\neg(c \leq 0)$ means that $c > 0$ and $\neg(ac \leq bc)$ means $ac > bc$, our last diagram follows from the trichotomy law of inequality. These diagrams will now guide our composition of the following proof. Ⓐ

Proof. Suppose a, b, c are real numbers and $a > b$. Assume $c > 0$. We must prove that $ac > bc$. Since $a > b$ and c is a positive number, we can multiply both sides of the inequality $a > b$ by the positive c and conclude that $ac > bc$. Therefore, if $ac \leq bc$, then $c \leq 0$. □

3.8.2 Proof by Contradiction

There are times when is not easy to see how to prove a mathematical statement, say ψ. When this happens one should try the strategy called proof by contradiction. This strategy is perhaps the strangest method of proof. Here is how it works: To prove that ψ is true, we first assume that $\neg\psi$ is true. Then working with $\neg\psi$, we derive something that is false (for example, $0 = 1$ or $\varphi \wedge \neg\varphi$). Since the assumption $\neg\psi$ has led us to a fallacy, we must conclude that ψ is true. Another way to justify proof by contradiction is to observe that ψ and $(\neg\psi \to (\varphi \wedge \neg\varphi))$ are logically equivalent (see Exercise 3 on page 17). Thus, if one proves $\neg\psi \to (\varphi \wedge \neg\varphi)$, then ψ is true.

Proof Strategy 3.8.4. Given a diagram containing the form

Prove P

to apply proof by contradiction, use the form

To prove P:
    Assume $\neg P$
        Derive "a contradiction."

That is, to prove P, assume $\neg P$ and derive a contradiction.

Proof by contradiction is typically used when a direct proof is difficult to find or just "does not seem to work." Moreover, it is often not clear what the contradiction will be until you "get there."

In the proof analysis of our next theorem, we will first be working with the proof diagram

Assume m and n are integers.
Assume mn is odd.
Prove n is odd.

In other words, we have integers m and n where (1) $mn = 2i+1$ for an $i \in \mathbb{Z}$ and we want to prove that (2) $n = 2j+1$ for some $j \in \mathbb{Z}$. So, using equation (1) we must derive equation (2). We can solve equation (1) for n and obtain $n = 2(\frac{i}{m}) + \frac{1}{m}$; but this would require $\frac{i}{m}$ to be an integer and $\frac{1}{m} = 1$. Clearly, when $m > 1$, we see that $\frac{1}{m} \neq 1$. How can we get around these serious difficulties? *Proof by contradiction!*

Theorem 3.8.5. *Let m and n be integers. If mn is odd, then n is odd.*

Proof Analysis. We construct the following proof diagrams (Proof Strategy 3.8.4 gives us the second diagram):

Let m and n be integers.
  Assume mn is odd.
    Prove n is odd.

Let m and n be integers.
  Assume mn is odd.
    To prove n is odd:
      Assume n is even.
        Derive "a contradiction."

Let m and n be integers.
  Assume mn is odd.
    To prove n is odd:
      Assume $n = 2k$ for some $k \in \mathbb{N}$.
        Derive "a contradiction."

To say that n is not odd means that n even. In our proof by contradiction, we will be working with the assumptions (1) mn is odd, and (2) $n = 2k$ for some $k \in \mathbb{Z}$. From these two assumptions, we must derive a contradiction, that is, we must derive some nonsense! These diagrams will guide our composition of a well-structured proof of the theorem. Ⓐ

Proof. Let m and n be integers. Assume mn is odd. We must prove that n is odd. Suppose, for a contradiction, that n is even, that is, $n = 2k$ for some $k \in \mathbb{Z}$. Using some algebra, we deduce that $mn = m(2k) = 2(mk)$. Hence, mn is even and this contradicts our assumption that mn is odd. Therefore, n is odd. □

In the proof of Theorem 3.8.5, we added the phrase "suppose, for a contradiction, that …". This is done as a courtesy to the reader. We have thus told the reader that we are using proof by contradiction.

In elementary mathematics and in calculus we learned that for any real number $x \geq 0$ there is a unique real number $y \geq 0$ such that $y^2 = x$. We write $y = \sqrt{x}$ and say that y is the *square root* of x. Consequently, $(\sqrt{x})^2 = x$. Our next theorem shows that the square root operation preserves the inequality relation $<$ for positive numbers.

Theorem 3.8.6. *Suppose a and b are positive real numbers. If $a < b$, then $\sqrt{a} < \sqrt{b}$.*

Proof Analysis. We first construct proof diagrams (Proof Strategy 3.8.4 yields the second diagram):

Assume $a > 0$ and $b > 0$.
  Assume $a < b$.
    Prove $\sqrt{a} < \sqrt{b}$.

Assume $a > 0$ and $b > 0$.
  Assume $a < b$.
    To prove $\sqrt{a} < \sqrt{b}$:
      Assume $\sqrt{a} \geq \sqrt{b}$.
        Derive "a contradiction."

Assume $a > 0$ and $b > 0$.
  Assume $a < b$.
    To prove $\sqrt{a} < \sqrt{b}$:
      Assume $\sqrt{a} > \sqrt{b}$ or $\sqrt{a} = \sqrt{b}$.
        Derive "a contradiction."

In the last diagram we are assuming an "or" statement. Thus, we will implicitly use the "division by cases" Strategy 3.6.3(a) in our proof, where in each case we shall derive a contradiction. Ⓐ

Proof. Suppose a and b are positive real numbers. Assume $(\star)$ $a < b$. We will prove that $\sqrt{a} < \sqrt{b}$. Suppose, for a contradiction, that $\sqrt{a} \geq \sqrt{b}$. So, either $\sqrt{a} > \sqrt{b}$ or $\sqrt{a} = \sqrt{b}$. If $\sqrt{a} > \sqrt{b}$, then $a = (\sqrt{a})^2 > (\sqrt{b})^2 = b$ by Theorem 3.3.2 and so $a > b$, which contradicts $(\star)$. If $\sqrt{a} = \sqrt{b}$, then $a = (\sqrt{a})^2 = (\sqrt{b})^2 = b$ and so $a = b$, which also contradicts $(\star)$. Therefore, $\sqrt{a} < \sqrt{b}$. □

Definition 3.8.7. A positive rational number $\frac{m}{n}$, where m and n are natural numbers, is in **reduced form** if m and n have no common factors greater than 1.

Example 1. The rational number $\frac{4}{3}$ is in reduced form, $\frac{12}{9}$ is not in reduced form. Clearly every positive rational number can be put into reduced form.

The ancient Greeks believed that all numbers were rational. Consequently, they were convinced that the length of the diagonal in the unit square (which is $\sqrt{2}$)[6] must be rational. Thus, by the Pythagorean Theorem, these ancient mathematicians also believed that there is a rational number whose square is 2. At some point, one of these early mathematicians made a revolutionary discovery and established the following theorem.

Theorem 3.8.8. *There is no rational number x such that $x^2 = 2$.*

Proof Analysis. In logical form, the statement of the theorem can be expressed as

$$\neg(\exists x \in \mathbb{Q})(x^2 = 2).$$

We build the following proof diagram by applying the proof-by-contradiction Strategy 3.8.4.

To prove $\neg(\exists x \in \mathbb{Q})(x^2 = 2)$:
Assume $(\exists x \in \mathbb{Q})(x^2 = 2)$.
Derive "a contradiction."

This diagram will guide our composition of a well-structured proof of the theorem by contradiction. Ⓐ

Proof. Suppose, for a contradiction, that $x^2 = 2$ for some rational number x. So, we can consider x to be positive. Since x is rational, we can write $x = \frac{m}{n}$ where m and n are natural numbers. We can presume that the ratio $\frac{m}{n}$ has been put into reduced form. It follows that

$$m \text{ and } n \text{ have no common factors greater than 1.} \tag{3.3}$$

Since $x^2 = 2$ and $x = \frac{m}{n}$, we obtain $\frac{m^2}{n^2} = 2$ and thus, $m^2 = 2n^2$. Since m^2 is even, we conclude that m is even by Theorem 3.8.2. So, $m = 2k$ for some $k \in \mathbb{N}$. Substituting $m = 2k$ into our equation $2n^2 = m^2$, we obtain

$$\begin{aligned} 2n^2 &= (2k)^2 \\ 2n^2 &= 4k^2 \\ n^2 &= 2k^2. \end{aligned}$$

Therefore, n^2 is even. So n is must be even as well. Hence, m and n have a common factor of 2, which contradicts (3.3). This completes the proof. □

Theorem 3.8.8 was a very important discovery in the history of mathematics. Courant and Robbins [3, p. 59], in their book *What is Mathematics?*, state that:

[6] The ancient Greeks did not use the notation $\sqrt{2}$.

> This revelation was a scientific event of the highest importance. Quite possibly it marked the origin of what we consider the specifically Greek contribution to rigorous procedure [mathematical proof] in mathematics. Certainly it has profoundly affected mathematics and philosophy from the time of the Greeks to the present day.

Since $(\sqrt{2})^2 = 2$, Theorem 3.8.8 implies that $\sqrt{2}$ is irrational. Thus, we now know that there is at least one irrational number. Are there any others? Yes, there are infinitely many irrational numbers. In fact, in Section 6.5 it will be shown that there are more irrational numbers than there are rational ones (see Exercise 18 on page 206). Consequently, one can prove that most irrational numbers cannot be obtained by performing algebraic operations on rational numbers. In particular, the vast majority of irrational numbers cannot be realized by taking the square root of a rational number.

Exercises 3.8

Prove the following theorems and corollaries. A corollary is a statement that follows from a previously established theorem. Each corollary below follows from the theorem above it.

1. **Theorem.** Let x and y be real numbers. If $x^2 = y^2$, then $|x| = |y|$.
2. **Theorem.** Let a and b be natural numbers. If $ab = 1$, then $a = 1$ and $b = 1$.
3. **Corollary.** Let a and b be natural numbers. If $a \,|\, b$ and $b \,|\, a$, then $a = b$.
4. **Corollary.** Let n be an integer. Then n is not both even and odd.
5. **Theorem.** Let r and $s \neq 0$ be rational numbers. If x is irrational, then $r + sx$ is irrational.
6. **Theorem.** Let $a > 0$ be a real number. Then $\frac{1}{a} > 0$.
7. **Corollary.** Let x and y be real numbers where $x > 0$. If $xy > 0$, then $y > 0$.
8. **Corollary.** Suppose a, b, c, d are all positive real numbers. If $ab = cd$ and $a \leq c$, then $d \leq b$.
9. **Corollary.** Suppose a, b, c are positive real numbers. If $a < c$, then $\frac{a}{b} < \frac{c}{b}$.
10. **Corollary.** Let a, b, d be positive real numbers. If $d < b$, then $\frac{a}{b} < \frac{a}{d}$.
11. **Corollary.** Let a, b, d be positive real numbers. If $a < c$ and $d < b$, then $\frac{a}{b} < \frac{c}{d}$.
12. **Theorem.** For all real numbers a and b, if $a^2 < b^2$, then $|a| < |b|$.
13. **Theorem.** For all real numbers x, if $x > 1$ then $0 < \frac{1}{x} < 1$.
14. **Theorem.** Let $x > 0$ be a real number. If x is irrational, then $\sqrt{x}$ is irrational.
15. **Theorem.** The real number $\sqrt{2} + \sqrt{3}$ is an irrational number.
16. **Theorem.** Let a and b be positive real numbers. Then $\sqrt{a} + \sqrt{b} > \sqrt{a+b}$.
17. **Theorem.** Let m and n be integers. Then mn is even if and only if m is even or n is even.
18. **Theorem.** Let a, b, c be integers. If $a + b = c$, then at least one of a, b, and c must be even.

19. Theorem. Let x and y be positive real numbers. Then $\frac{x}{y} + \frac{y}{x} \geq 2$.

20. Theorem. Let x and y be positive real numbers. Then $\frac{x+y}{2} \geq \sqrt{xy}$.

21. Theorem (Triangle Inequality). Let x, y be real numbers. Then $|x+y| \leq |x| + |y|$.

22. Theorem. Let x be a real number. Then $x^2 < 1$ if and only if $-1 < x < 1$.

Exercise Notes: For Exercise 4, if $2i = 2j + 1$ and $i, j \in \mathbb{Z}$, then $2(i - j) = 1$. For Exercise 12, see Exercise 13 on page 88. For Exercise 15, assume $\sqrt{2} + \sqrt{3} = \frac{a}{b}$ for nonzero integers a and b. Solve for $\sqrt{3}$ and then square both sides to obtain a contradiction. For Exercises 16 and 19, use contradiction. For Exercise 20, use contradiction and Theorem 3.3.2. For Exercise 21, use contradiction, Theorem 3.3.2 and Exercises 10, 11 and 13 of Section 3.6. For Exercise 22, see Exercise 3 on page 70 and Theorem 3.3.2.

CHAPTER 4

Mathematical Induction

Mathematical induction is a method of proof that is frequently used to establish that certain statements are true for every natural number. Before we introduce this method of proof, we must first discuss the well-ordering principle.

4.1 The Well-Ordering Principle

An important property of the set of natural numbers $\mathbb{N}$ is that any nonempty subset of $\mathbb{N}$ has a least element. A set is *nonempty* if it contains at least one element.

Well-Ordering Principle 4.1.1. Let S be a nonempty set of natural numbers. Then S has a least element, that is, there is a $k \in S$ such that $k \leq n$ for all $n \in S$.

For example, let S be the set of integers greater than 20 which are divisible by 9. Then 27 is the least element in S. In this section we will use the well-ordering principle to establish a theorem on the number of primes. First, we shall show that the Well-Ordering Principle 4.1.1 implies a slightly more general principle.

Theorem 4.1.2 (General Well-Ordering Principle). *Let b be an integer and let S be a nonempty set of integers all of which are greater than or equal to b. Then S has a least element.*

Before we prove Theorem 4.1.2, consider the set $S = \{-3,-1,3,5,7,9,\ldots\}$. Observe that S is a nonempty set of integers satisfying $-4 \leq n$ for every $n \in S$. Let S^* be the set obtained by adding $-(-4)+1 = 5$ to each integer in the set S. Thus, S^* can be described by $S^* = \{n-(-4)+1 : n \in S\} = \{2,4,8,10,12,14,\ldots\}$ which is a subset of $\mathbb{N}$. For every $n^* \in S^*$ there is an $n \in S$ such that $n^* = n-(-4)+1$. For example, $8 \in S^*$ and $8 = 3-(-4)+1$ where $3 \in S$. We will apply these ideas in the following proof.

Proof. Let b be an integer and let S be a nonempty set of integers such that $b \leq n$ for all $n \in S$. For any integer n, it follows (by properties of inequality) that

$$b \leq n \text{ if and only if } 1 \leq n-b+1. \tag{4.1}$$

Let $S^* = \{n-b+1 : n \in S\}$. It follows from (4.1) that S^* is a nonempty set of natural numbers. By the Well-Ordering Principle 4.1.1, S^* has a least element k^*. Since $k^* \in S^*$, there is a $k \in S$ such that $k^* = k-b+1$. Hence, $k = k^*+b-1$.

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_4,

We now show that k is the least element in S. Let $n \in S$. Then $n-b+1$ is in S^*. Thus, $k^* \leq n-b+1$ and we conclude that $k^*+b-1 \leq n$. Therefore, $k \leq n$. □

Definition 4.1.3. A natural number $p > 1$ is a **prime number** if and only if for all natural numbers a and b, if $p = ab$ then either $a = 1$ or $b = 1$. A natural number $n > 1$ is a **composite number** if and only if there are natural numbers a and b such that $n = ab$ where $a \neq 1$ and $b \neq 1$.

The first 25 prime numbers are:

$$2,3,5,7,11,13,17,19,23,29,31,37,41,43,47,53,59,61,67,71,73,79,83,89,97.$$

Is there a prime number larger than all those in this list? The ancient Greeks were fascinated by the prime numbers and it was initially not known how many prime numbers there were among the natural numbers. The oldest known proof asserting that there are infinitely many prime numbers is given by the Greek mathematician Euclid in his *Elements*.[1] Before we present this proof, we will identify the proof strategy that was used by Euclid. This strategy, which is still in use today, employs the well-ordering principle and proof by contradiction.

Remark 4.1.4. In this chapter we will be letting $P(n)$ represent a statement in which the free variable n represents an integer. If a variable is free, then substitution may take place. So, we can replace n with any integer; for example, we can replace n with 5 and obtain the statement $P(5)$. Given an integer n, we shall say that $P(n)$ is *defined* when the resulting statement $P(n)$ is meaningful and it is either true or false. For example, let $P(n)$ be the statement "n^2+2n is even." Clearly, $P(2)$ is true and $P(3)$ is false. So the statements $P(2)$ and $P(3)$ are defined. In fact, $P(n)$ is defined for every integer n. For another example, let $P(n)$ represent the statement "the sum of the first n positive integers is odd." Note that $P(3)$ is false and $P(5)$ is true. On the other hand, the statement $P(-10)$ is not meaningful and thus, it is neither true nor false. Hence, $P(-10)$ is undefined.

Let b is an integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. To prove statements of the form $(\forall n \geq b)P(n)$ mathematicians sometimes use proof by contradiction. The proof typically begins by assuming $\neg(\forall n \geq b)P(n)$ and then concludes that $(\exists n \geq b)\neg P(n)$. Thus, the set $S = \{n \in \mathbb{Z} : n \geq b \wedge \neg P(n)\}$ is nonempty. By the Well-Ordering Principle 4.1.2 the set S has a least element, say N. Hence, $N \geq b$ and $\neg P(N)$. Because N is the least element in S, it follows that for all integers k, if $b \leq k < N$ then $P(k)$. This latter statement then leads to a contradiction. The idea motivates our next proof strategy.

Proof Strategy 4.1.5 (Well-ordering proof strategy). Let b be a fixed integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. To prove a statement of the form $(\forall n \geq b)P(n)$ by the well-ordering principle, use the diagram

[1] The *Elements* is a mathematical work consisting of 13 books written by Euclid around 300 BC.

Assume that $\neg P(n)$ holds for some integer $n \geq b$.
Let $N \geq b$ be the least integer satisfying $\neg P(N)$.
Derive "a contradiction."

Euclid used Proof Strategy 4.1.5 to prove our next lemma.[2]

Lemma 4.1.6. *Every natural number $n \geq 2$ is divisible by a prime number.*

Proof. Suppose, for a contradiction, that there is a natural number $n \geq 2$ that is not divisible by a prime. By the well-ordering principle, there is a least natural number $N \geq 2$ that is not divisible by a prime. Thus, we have that:

(1) N is not divisible by a prime.
(2) If $2 \leq k < N$, then k is divisible by a prime, when k is a natural number.

We consider two cases: either N is a prime or it is not a prime.

CASE 1: N is a prime. Since N is divisible by itself, this contradicts (1).

CASE 2: N is not a prime. So, N is a composite number and $N = ab$ for some natural numbers a and b with $2 \leq a < N$ and $2 \leq b < N$. Since $2 \leq a < N$, assertion (2) implies that $p \,|\, a$ for some prime p. Because $a \,|\, N$, Theorem 3.5.6 implies that $p \,|\, N$. Hence, N is divisible by a prime. This contradicts (1) and completes the proof. □

Lemma 4.1.7. *For any integer n and any prime number p, if $p \,|\, n$ then $p \nmid (n+1)$.*

Proof. Let n be an integer and let p be a prime. Assume $p \,|\, n$, that is, assume $n = pi$ for some $i \in \mathbb{Z}$. We prove that $p \nmid (n+1)$. Suppose, for a contradiction, that $p \,|\, (n+1)$ and so, $(*)$ $n + 1 = pk$ for some $k \in \mathbb{Z}$. Substituting $n = pi$ into equation $(*)$, we obtain $ip + 1 = kp$. Thus, $1 = (k-i)p$ with $p > 1$ and hence, $(k-i) \geq 1$ as $(k-i)$ is an integer. Since $(k-i) \geq 1$ and $p > 1$, we conclude that $(k-i)p > (k-i) \geq 1$. Thus, $(k-i)p > 1$. This contradiction completes the proof. □

We now present Euclid's ingenious proof that there are infinitely many primes.

Theorem 4.1.8 (Infinitude of the Primes). *The set of prime numbers is infinite.*

Proof. Suppose, for a contradiction, that the set of prime numbers is finite. Then we can list all of the prime numbers as a finite list in ascending order, namely,

$$p_1, p_2, p_3, \ldots, p_n \tag{4.2}$$

where p_1 is the first prime and p_n is the last prime. Let $N = p_1 p_2 p_3 \cdots p_n + 1$. Since $N \geq 2$ and N is a natural number, Lemma 4.1.6 implies that $p \,|\, N$ for some prime p. Since every prime appears in the list (4.2), p must be an entry in this list. Therefore, $p \,|\, (p_1 p_2 p_3 \cdots p_n)$ and thus, $p \nmid (p_1 p_2 p_3 \cdots p_n + 1)$ by Lemma 4.1.7. So $p \nmid N$. Hence, $p \,|\, N$ and $p \nmid N$. This contradiction completes the proof. □

[2] In mathematics, a lemma is a 'little theorem' that is usually used in the proof of a more important theorem.

The proof of Theorem 4.1.8 shows that when you have a list $(\star)$ $p_1, p_2, p_3, \ldots, p_n$ of the first n primes, then any prime number that evenly divides $p_1 p_2 p_3 \cdots p_n + 1$ is not in the list $(\star)$. We illustrate this result with an example using the first six primes. Observe that $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30{,}031$ which is divisible only by the two primes 59 and 509, both of which are not in the list $2, 3, 5, 7, 11, 13$.

Exercises 4.1

1. Explain why Theorem 4.1.2 implies the Well-Ordering Principle 4.1.1.
2. Let $(*)$ $q_1, q_2, \ldots, q_m$ be *any* finite list of prime numbers. Let p be a prime that evenly divides $q_1 q_2 \cdots q_m + 1$. Prove that p is not in the list $(*)$.
3. Prove the statement: For all integers $n \geq 1$, we can write $n = 2^k \cdot m$ for some integers k and m where $k \geq 0$ and m is odd.
4. Prove that for every integer $n \geq 0$, either n is even or n is odd.
5. Prove that for every integer $n \geq 3$, we have $2^n \geq 2n + 1$.
6. Let $P(n)$ be a statement that is defined for all integers $n \geq 1$. Suppose the following two conditions hold: (a) $P(1)$ is true, and (b) for all integers $n \geq 1$, if $P(n)$ holds then $P(n+1)$ also holds. Use the Well-Ordering Principle 4.1.1 to prove that $P(n)$ must be true for all integers $n \geq 1$.

Exercise Notes: For Exercise 3, use the well-ordering Proof Strategy 4.1.5 where $P(n)$ is the statement "$n = 2^k \cdot m$ for some integers k and m where $k \geq 0$ and m is odd." There are two cases to consider about N. If N is even, then $N = 2i$ for some integer i where $1 \leq i < N$. If N is odd, then note that $N = 2^0 N$. For Exercise 4, use Proof Strategy 4.1.5. Show that $N > 0$. So, $0 \leq N - 1 < N$ and thus $N - 1$ is either even or odd. For Exercise 5, using Proof Strategy 4.1.5, one obtains the assumption $2^N < 2N + 1$. Observe that $N > 3$ and thus, $3 \leq N - 1 < N$. So, you can conclude that $2^{N-1} \geq 2(N-1) + 1$. Multiply both sides of this latter inequality by 2 to derive a contradiction. For Exercise 6, if $N \geq 1$ is the least such integer satisfying $\neg P(N)$, then explain why $N = n + 1$ for some $n \geq 1$ and observe that $n < N$.

4.2 Proof by Mathematical Induction

Mathematical induction is a powerful method for proving theorems about the natural numbers. Suppose you have a statement $P(n)$ that you want to prove is true for every integer n greater than or equal to the integer b. How can you prove this statement by mathematical induction? First you prove that the statement definitely holds for b. Then you have to prove that whenever the statement holds for an integer $n \geq b$,

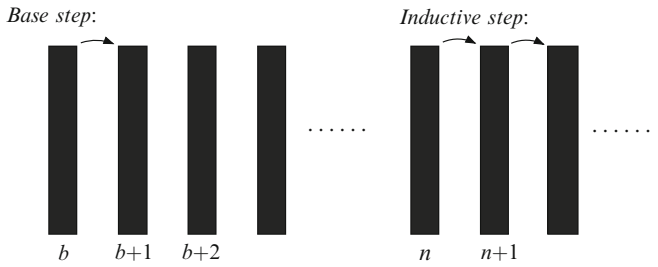


Fig. 4.1 The base step and inductive step force the dominoes to fall, one and all

then it must hold for the next integer $n+1$ as well. In other words, mathematical induction is a method of proof that works by first proving the statement $P(b)$ is true for the starting value b, which is called the base step. Then one must prove the inductive step, which shows that the truth of the statement $P(n)$ implies the truth of the statement $P(n+1)$. If the base step and the inductive step are both proven, then the statement $P(n)$ is true for all the natural numbers $n \geq b$.

It may be helpful to think of the domino effect where one is presented with an infinite row of dominoes, each standing on end, as pictured in Fig. 4.1. The base step shows that the first domino will fall. The inductive step ensures that each domino is perfectly aligned with the one ahead of it. Thus, we know that the first domino will fall (base step) and whenever the n-th domino falls, the $n+1$ domino will also fall (inductive step). Therefore, all of the dominoes must fall.

The well-ordering principle implies the following related principle (see Exercise 6 on page 102 and Remark 4.1.4).

Principle of Mathematical Induction. Let b be a given integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. Suppose that

1. $P(b)$ is true, and
2. For all $n \geq b$, if $P(n)$ then $P(n+1)$.

Then for *all* integers $n \geq b$, the assertion $P(n)$ is true.

Many mathematical statements have the form: *For every integer $n \geq b$, "something about n happens,"* where b is a fixed integer. The Principle of Mathematical Induction motivates our next proof strategy called *proof by mathematical induction.*

Proof Strategy 4.2.1. Let b be a given integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. To prove $(\forall n \geq b)P(n)$ by mathematical induction, use the diagram

Prove $P(b)$.
Prove $(\forall n \geq b)[P(n) \rightarrow P(n+1)]$.

In other words, use the diagram

Base step: Prove $P(b)$.
Inductive step: Let $n \geq b$ be an integer.
Assume $P(n)$.
Prove $P(n+1)$.

In a proof of a statement $(\forall n \geq b)P(n)$ by mathematical induction, b is referred to as the *base value*. The proof of $P(b)$ is called the *base step* and the proof of $(\forall n \geq b)[P(n) \to P(n+1)]$ is called the *inductive step*. In the second proof diagram of Proof Strategy 4.2.1, the assumption $P(n)$ is called the *induction hypothesis* (IH) and the statement to be proven, $P(n+1)$, is called the *induction conclusion* (IC).

In the *base step* you must show that the statement $P(b)$ is true. To do so, simply replace n by b everywhere in $P(n)$ and verify that $P(b)$ holds.

The *inductive step* is more challenging. It requires you to reach the conclusion that $P(n+1)$ is true after assuming $P(n)$ is true. To prove that $P(n+1)$ is true, you should somehow try to rewrite the statement $P(n+1)$ in terms that relate to the assumption $P(n)$ (as will soon be illustrated), for then you will be able to make use of the assumption $P(n)$. Appealing to the assumption $P(n)$ is referred to as using the induction hypothesis. After establishing that $P(n+1)$ is true, the proof will be complete. A proof that uses Strategy 4.2.1 is also called an *induction proof* or *proof by induction*.

Our next theorem can be proven in more than one way; however, we shall provide a proof that uses mathematical induction. We will first present a "proof analysis" and then we will prove the theorem by induction.

Theorem 4.2.2. *For every integer $n \geq 1$, the number n^2+n+2 is even.*

Proof Analysis. When you are looking for a proof by induction, it is helpful to write down the statement $P(n)$. In this case we have

$$P(n): \ n^2+n+2 \text{ is even.}$$

In this theorem our base value is $b = 1$. Next, we will construct a proof diagram by letting $b = 1$ in the second diagram of Proof Strategy 4.2.1:

Base step: Prove $P(1)$.
Inductive step: Let $n \geq 1$ be an integer.
Assume $P(n)$.
Prove $P(n+1)$.

Now, we carefully write out the statements $P(1)$ and $P(n+1)$. Upon replacing n everywhere in the statement

$$P(n): \quad n^2+n+2 \text{ is even}$$

with 1, we obtain

$$P(1): \quad 1^2+1+2 \text{ is even.}$$

By replacing n with $n+1$ everywhere in $P(n)$, we obtain

$$P(n+1): \quad (n+1)^2+(n+1)+2 \text{ is even.}$$

Thus, we can rewrite our proof diagram as

Base step: Prove 1^2+1+2 is even.
Inductive step: Let $n \geq 1$ be an integer.
Assume n^2+n+2 is even.
Prove $(n+1)^2+(n+1)+2$ is even.

We write out what it means for these values to be even as follows:

Base step: Prove $1^2+1+2=2k$ for some $k \in \mathbb{Z}$.
Inductive step: Let $n \geq 1$ be an integer.
Assume $n^2+n+2=2i$ for some $i \in \mathbb{Z}$.
Prove $(n+1)^2+(n+1)+2=2j$ for $j \in \mathbb{Z}$.

For the base step we see that $1^2+1+2=2\cdot 2$, which is clearly even. For the inductive step, we must use the induction hypothesis

$$n^2+n+2=2i \text{ where } i \in \mathbb{Z} \tag{IH}$$

to deduce the induction conclusion

$$(n+1)^2+(n+1)+2=2j \text{ for some } j \in \mathbb{Z}. \tag{IC}$$

To prove this induction conclusion, we will begin with the left hand side of the equality in (IC) and, using algebra, will make some changes so that the left hand side of the equation in (IH) appears. From the induction hypothesis (IH) we have that $n^2+n+2=2i$. Using this equation, we will be able to prove (IC) and thus conclude that the integer $(n+1)^2+(n+1)+2$ is even. Ⓐ

Our proof diagrams and analysis will guide the composition of a well-structured proof of the Theorem 4.2.2 by mathematical induction.

Proof. We prove, by mathematical induction, that n^2+n+2 is even for all $n \geq 1$.

Base step: For $n=1$, we see that $1^2+1+2=2\cdot 2$ is even.

Inductive step: Let $n \geq 1$ be an integer and assume the induction hypothesis that the integer n^2+n+2 is even, that is, assume

$$n^2+n+2=2i \text{ where } i \in \mathbb{Z}. \tag{IH}$$

We show that $(n+1)^2+(n+1)+2$ is even as follows:

$$\begin{aligned}(n+1)^2+(n+1)+2 &= n^2+2n+1+n+1+2 && \text{because } (n+1)^2=n^2+2n+1\\ &= n^2+n+1+2n+2 && \text{by regrouping}\end{aligned}$$

$$
\begin{aligned}
&= (n^2+n+1)+2(n+1) && \text{by distributivity}\\
&= 2i+2(n+1) && \text{by induction hypothesis (IH)}\\
&= 2(i+n+1) && \text{by distributivity.}
\end{aligned}
$$

Hence, $(n+1)^2+(n+1)+2 = 2j$ where $j = i+n+1$ is an integer. Therefore, $(n+1)^2+(n+1)+2$ is even and this completes the proof. □

Exercises 4.2

1. Using mathematical induction, prove that for every integer $n \geq 0$, the number n^2+3n+5 is odd.
2. Prove, by induction, that for every integer $n \geq 0$, either n is even or n is odd.
3. Using Exercise 2, prove that every integer n is either even or odd.
4. Using Exercise 3, prove that for every integer n we have that $n(n+1)$ is even.
5. For any integer $n \geq 1$, let $P(n)$ be the statement: $1+2+\cdots+n = \frac{n(n+1)}{2}$. What is statement $P(3)$? What is statement $P(1)$? What is statement $P(n+1)$?
6. Prove that $1+2+\cdots+n = \frac{n(n+1)}{2}$ for all integers $n \geq 1$, by induction.

Exercise Notes: For Exercise 2, use the induction proof strategy

Base step: Prove 0 is even or odd.
Inductive step: Let $n \geq 0$ be an integer.
Assume n is even or n is odd.
Prove $n+1$ is even or $n+1$ is odd.

In the inductive step, since you are assuming an 'or' statement, use a proof by cases. For Exercise 3, there are two cases: (1) $n \geq 0$; (2) $n < 0$. For Exercise 4, there are two cases: (1) n is odd; (2) n is even.

4.3 Sequences, Sums, and Factorials

In mathematics one often works with regular patterns or repeated processes. The main tool used to study repeated processes is the *sequence*, a fundamental concept with a rich history in mathematics. Sequences are interesting mathematical objects with lots of surprising properties, many of which can be verified by mathematical induction.

In this section we shall introduce the notation and terminology of sequences. A sequence is just a list of real numbers; for example $1,3,5,7,9,\ldots$ is an infinite sequence. In general, an infinite sequence is usually written in the form

$$a_1, a_2, a_3, \ldots, a_n, a_{n+1}, \ldots$$

where the sequence starts with $n = 1$. Furthermore, for an integer m, a sequence having the form

$$a_m, a_{m+1}, a_{m+2}, \ldots, a_n, a_{n+1}, \ldots$$

starts with $n = m$. Many times a sequence is described by means of a formula $f(n)$ where $a_n = f(n)$.

Example 1. Write out the first few terms of the sequences determined by the given formula:

1. $a_n = 2n - 1$ starting with $n = 1$.
2. $a_n = 2n - 1$ starting with $n = 3$.
3. $a_n = (-1)^n$ starting with $n = 0$.
4. $a_n = (-1)^n$ starting with $n = 1$.
5. $a_n = (-1)^{n+1}$ starting with $n = 1$.

Solution. For each formula, we evaluate some initial terms a_n, for $n = 1,2,3,\ldots$, and obtain:

1. $1,3,5,7,9,\ldots$
2. $5,7,9,11,\ldots$
3. $1,-1,1,-1,1,-1,1,\ldots$
4. $-1,1,-1,1,-1,1,-1,\ldots$
5. $1,-1,1,-1,1,-1,1,\ldots.$ Ⓢ

The formulas given in items 3–5 of Example 1 can be used to generate sequences having alternating signs. Items 3 and 5 start with a plus sign and item 4 starts with a negative sign. We now show how to take advantage of these alternating sequences.

Example 2. Find a formula that generates the given sequence:

(a) $1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}, \frac{1}{36}, \ldots$ starting with $n = 1$.
(b) $1, -\frac{1}{4}, \frac{1}{9}, -\frac{1}{16}, \frac{1}{25}, -\frac{1}{36}, \ldots$ starting with $n = 1$.
(c) $-1, \frac{1}{4}, -\frac{1}{9}, \frac{1}{16}, -\frac{1}{25}, \frac{1}{36}, \ldots$ starting with $n = 1$.

Solution. For item (a), we see that the formula $f(n) = \frac{1}{n^2}$ for $n \geq 1$ generates the given sequence. The sequence in (b) is an alternating form of the sequence in (a) that starts with a plus sign. Using our solution to item 5 of Example 1, we obtain the formula $g(n) = (-1)^{n+1}\frac{1}{n^2}$ which generates the sequence in (b). Similarly, the sequence in (c) is also an alternating form of the sequence in (a) which starts with a negative sign. The formula $h(n) = (-1)^n\frac{1}{n^2}$ generates the sequence in (c). Ⓢ

4.3.1 Summation Notation

Σ-notation, or summation notation, is used to write the sum of many terms in a concise and compact way. Given a finite sequence $a_1, a_2, \ldots, a_n$ we write $\sum_{k=1}^{n} a_k$ as shorthand for the sum $a_1 + a_2 + \cdots + a_n$, that is,

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n.$$

For example, $\sum_{k=1}^{n} 2^k = 2^1 + 2^2 + 2^3 + \cdots + 2^{n-1} + 2^n$. More generally, given integers $m \leq n$ and a finite sequence $a_m, a_{m+1}, \ldots, a_n$ we define

$$\sum_{k=m}^{n} a_k = a_m + a_{m+1} + a_{m+2} + \cdots + a_{n-1} + a_n. \tag{4.3}$$

In other words, $\sum_{k=m}^{n} a_k$ is the **summation of every** a_k **from** k **equals** m **to** n**.**

The capital Greek letter Σ is called sigma and we use this symbol to denote the words *sum* or *summation*. Given the sum $\sum_{k=m}^{n} a_k$, we shall call k the **index** of the summation. We will also say that m is the **lower limit** and that n is the **upper limit** of the summation. We shall refer to a_k as the **summand**.

Three of the most useful formulas for dealing with summations are given below. One can prove these formulas using mathematical induction (see Theorem 4.4.3, and Exercise 7 on page 122).

Theorem 4.3.1 (Linearity Properties). *Let c is a fixed real number. For integers $m \leq n$, we have the following three identities:*

1. $\sum_{k=m}^{n} (a_k + b_k) = \sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k$
2. $\sum_{k=m}^{n} (a_k - b_k) = \sum_{k=m}^{n} a_k - \sum_{k=m}^{n} b_k$
3. $\sum_{k=m}^{n} ca_k = c \sum_{k=m}^{n} a_k.$

Linearity properties 1 and 2 allow us to split summations into simpler sums, and property 3 allows us to factor out any constant multiple that does not involve the summation index k.

We will also be working with sums of the form $\sum_{k=1}^{n} f(k)$ where $n \geq 1$ and $f(k)$ is a formula involving an integer variable k. There is nothing new here, since this form satisfies the equation

$$\sum_{k=1}^{n} f(k) = f(1) + f(2) + \cdots + f(n). \tag{4.4}$$

Items 1 and 2 of our next example are identities that are often used in proofs by induction. Item 3 will be applied in the proof of Theorem 4.3.4. Item 4 can be used to prove the binomial theorem (see Exercise 20 on page 116 and Exercise 16 on page 122).

Example 3. Let $f(k)$ be a formula involving an integer variable k. Show that:

1. $\sum_{k=1}^{1} f(k) = f(1)$.
2. $\sum_{k=1}^{n+1} f(k) = \left(\sum_{k=1}^{n} f(k)\right) + f(n+1)$, for $n \geq 1$.
3. $\sum_{k=1}^{n} f(k) = \sum_{k=1}^{i} f(k) + \sum_{k=i+1}^{n} f(k)$, when $1 < i < n$.
4. $\sum_{k=0}^{n+1} f(k) = f(0) + \left(\sum_{k=1}^{n} f(k)\right) + f(n+1)$, when $n \geq 1$.

Solution. When $n = 1$ in (4.4), we obtain $\sum_{k=1}^{1} f(k) = f(1)$, which is the equation in item 1. We justify equation 2 as follows:

$$\begin{aligned}
\sum_{k=1}^{n+1} f(k) &= f(1) + f(2) + \cdots + f(n) + f(n+1) && \text{by def. of } \Sigma \text{ notation} \\
&= (f(1) + f(2) + \cdots + f(n)) + f(n+1) && \text{regrouping} \\
&= \left(\sum_{k=1}^{n} f(k)\right) + f(n+1) && \text{by def. of } \Sigma \text{ notation.}
\end{aligned}$$

Equations 3 and 4 can be justified in a similar manner. Ⓢ

Just as in the solution to Example 3, one can derive the identity $\sum_{k=m}^{m} f(k) = f(m)$ and the identity $\sum_{k=m}^{n+1} f(k) = \left(\sum_{k=m}^{n} f(k)\right) + f(n+1)$ for $n \geq m$, where m is any integer.

4.3.2 Evaluating Sums

One way to evaluate the sum $\sum_{k=m}^{n} a_k$ is to expand the sigma notation, as in (4.3), and then add up all of the resulting terms. We shall call this an 'open-form solution.' Is there another way to evaluate this sum? Many times there is a formula, which does not involve the summation symbol, that can be used to get the correct answer.

Definition 4.3.2. Let $n \geq m$ be integers and let $a_m, a_{m+1}, a_{m+2}, \ldots, a_n$ be a sequence of real numbers. Since $\sum_{k=m}^{n} a_k = a_m + a_{m+1} + a_{m+2} + \cdots + a_n$, we call this an **open-form solution**, or an **open sum**, for $\sum_{k=m}^{n} a_k$. A formula $g(n)$ satisfying $\sum_{k=m}^{n} a_k = g(n)$ for every $n \geq m$, is called a **closed-form solution** for $\sum_{k=m}^{n} a_k$.

Example 4 (some open-form solutions)**.** For all integers $n \geq 1$,

1. $\sum_{k=1}^{n} k = 1 + 2 + \cdots + n$
2. $\sum_{k=1}^{n} k^2 = 1^2 + 2^2 + \cdots + n^2$
3. $\sum_{k=1}^{n} k^3 = 1^3 + 2^3 + \cdots + n^3$.

Example 5 (some closed-form solutions)**.** For all integers $n \geq 1$,

1. $\sum_{k=1}^{n} 1 = n$
2. $\sum_{k=0}^{n} 1 = n + 1$
3. $\sum_{k=1}^{n} 3 = 3n$.

To see why item 1 holds, notice that $\sum_{k=1}^{n} 1$ has the form $\sum_{k=1}^{n} a_k$, where $a_k = 1$ for $k = 1, 2, \ldots, n$. So, expressing the sum $\sum_{k=1}^{n} 1$ as an open-form solution, we obtain the identity $\sum_{k=1}^{n} 1 = \underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = n$. Items 2–3 are left as exercises.

Our next theorem presents closed-form solutions for the sums in Example 4 (see Theorem 4.4.1, and also Exercises 6 and 11 starting on page 122).

Theorem 4.3.3. *For all integers $n \geq 1$,*

1. $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$
2. $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$
3. $\sum_{k=1}^{n} k^3 = \left[\frac{n(n+1)}{2}\right]^2$.

Example 6. Using Theorems 4.3.1 and 4.3.3, find a closed-form solution for the sum $\sum_{k=1}^{n} (k+3)^2$.

Solution. We have that

$$\begin{aligned}
\sum_{k=1}^{n}(k+3)^2 &= \sum_{k=1}^{n}(k^2+6k+9) && \text{as } (k+3)^2 = k^2+6k+9\\
&= \sum_{k=1}^{n}k^2 + 6\sum_{k=1}^{n}k + 9\sum_{k=1}^{n}1 && \text{by Theorem 4.3.1}\\
&= \frac{n(n+1)(2n+1)}{6} + 6\cdot\frac{n(n+1)}{2} + 9\sum_{k=1}^{n}1 && \text{by Theorem 4.3.3}\\
&= \frac{n(n+1)(2n+1)}{6} + 6\cdot\frac{n(n+1)}{2} + 9n && \text{by Example 5(1)}\\
&= \frac{n(n+1)(2n+1)}{6} + 3n(n+4) && \text{by algebra.}
\end{aligned}$$

Thus, $\sum_{k=1}^{n}(k+3)^2 = \frac{n(n+1)(2n+1)}{6} + 3n(n+4)$ is our closed-form solution. Ⓢ

Theorem 4.3.4. *If* $\sum_{k=m}^{n} a_k = g(n)$ *for all integers* $n \geq m$, *then* $\sum_{k=i}^{n} a_k = g(n) - g(i-1)$ *whenever* i *is an integer satisfying* $m < i \leq n$.

Proof. Assume $\sum_{k=m}^{n} a_k = g(n)$ for all integers $n \geq m$. Let i be an integer satisfying $m < i \leq n$. Hence, $\sum_{k=m}^{i-1} a_k = g(i-1)$ because $i-1 \geq m$. Since $\sum_{k=m}^{n} a_k = \sum_{k=m}^{i-1} a_k + \sum_{k=i}^{n} a_k$, we conclude that $g(n) = g(i-1) + \sum_{k=i}^{n} a_k$. Therefore, $\sum_{k=i}^{n} a_k = g(n) - g(i-1)$. □

Example 7. Find a closed-form solution for $\sum_{k=5}^{n} k^2$.

Solution. Theorem 4.3.3 states that $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$ for every $n \geq 1$. When $n \geq 5$ we have that $1 < 5 \leq n$. Since $5-1=4$, Theorem 4.3.4 implies that

$$\sum_{k=5}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} - \frac{4(4+1)(2\cdot 4+1)}{6} = \frac{n(n+1)(2n+1)}{6} - 30.$$

Therefore, $\sum_{k=5}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} - 30$. Ⓢ

Changing the Index Variable

Consider the sums $\sum_{j=1}^{n} a_j$ and $\sum_{k=1}^{n} a_k$ and notice that the only difference between these sums is the use of the different index variables j and k. Observe that

$$\sum_{j=1}^{n} a_j = a_1 + a_2 + \cdots + a_n$$

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n.$$

and thus $\sum_{j=1}^{n} a_j = \sum_{k=1}^{n} a_k$. For this reason, the index variable of a sum is sometimes called a "dummy" variable. More generally, we have that $\sum_{j=m}^{n} a_j = \sum_{k=m}^{n} a_k$.

Shifting the Lower Limit of a Sum

To find a closed-form solution for a particular sum, one may need to transform the sum to one that has a different lower limit. The shift rule, below, will allow us to change the lower limit of a sum without changing the value of the sum.

The Shift Rule. Consider the sum $\sum_{k=m}^{n} h(k)$ with lower limit m. To rewrite this sum as one with the lower limit s, compute $d = s - m$ and then

$$\sum_{k=m}^{n} h(k) = \sum_{k=m+d}^{n+d} h(k-d) = \sum_{k=s}^{n+d} h(k-d) \tag{4.5}$$

which is easy to verify.

In our next example, we will apply the shift rule to show that $\sum_{k=3}^{n} \frac{1}{k^5} = \sum_{k=7}^{n+4} \frac{1}{(k-4)^5}$.

Example 8. Using (4.5), let us rewrite the sum $\sum_{k=3}^{n} \frac{1}{k^5}$ as an equal sum with lower limit 7. Compute $d = 7 - 3 = 4$. Then shift the upper and lower limits **up** by 4 and modify the summand by shifting k **down** by 4, obtaining $\sum_{k=3}^{n} \frac{1}{k^5} = \sum_{k=7}^{n+4} \frac{1}{(k-4)^5}$.

Example 9. Let us shift the sum $\sum_{k=3}^{n} \frac{1}{k^5}$ to an equal sum with lower limit 1. We obtain $d = 1 - 3 = -2$. We now shift the upper and lower limit values **down** by 2 and modify the summand by shifting k **up** by 2, to obtain $\sum_{k=3}^{n} \frac{1}{k^5} = \sum_{k=1}^{n-2} \frac{1}{(k+2)^5}$.

The next example shows that if the upper limit value n appears in the summand, then we must leave the value n in the summand alone when applying the shift rule.

Example 10. We shall shift the sum $\sum_{k=1}^{n} \frac{n}{k^5}$ to an equal sum with lower limit 4. First, $d = 4 - 1 = 3$. After shifting the upper and lower limits **up** by 3 and shifting the variable k in the summand **down** by 3, we obtain $\sum_{k=1}^{n} \frac{n}{k^5} = \sum_{k=4}^{n+3} \frac{n}{(k-3)^5}$.

4.3.3 Factorial Notation

In mathematics, the factorial of a natural number n is the product of all the natural less than or equal to n. This is written as $n!$ and is called "n factorial."

Definition 4.3.5. For each natural number n the value $n!$ is defined to be

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

By convention, we define $0! = 1$ (which is used to simplify mathematical formulas).

Example 11. Note that

1. $8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$.
2. $7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$.
3. $8! = 8 \cdot 7!$.
4. $n! = n(n-1)!$ for $n \geq 1$.
5. $(n+1)! = (n+1)n!$ for $n \geq 0$.

Example 12. Using Example 11, we simplify the following expressions.

1. $\dfrac{8!}{7!} = \dfrac{8 \cdot 7!}{7!} = 8.$
2. $\dfrac{(n+1)!}{(n-1)!} = \dfrac{(n+1)(n)(n-1)!}{(n-1)!} = n(n+1).$
3. $\dfrac{6!}{2! \cdot 3!} = \dfrac{6 \cdot 5 \cdot 4 \cdot 3!}{2! \cdot 3!} = \dfrac{6 \cdot 5 \cdot 4}{2!} = 60.$

We now introduce the *binomial coefficient* $\binom{n}{k}$, an important tool that is used in combinatorics (a branch of mathematics with a concentration on techniques of counting).

Definition 4.3.6. Let $n \geq k \geq 0$ be integers. Then $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Remark. Since $0! = 1$, we obtain $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$ for all integers $n \geq 0$.

The binomial coefficient $\binom{n}{k}$ is used in many areas of mathematics, for instance, algebra and probability theory. In combinatorics, the number $\binom{n}{k}$ is used to count the number of different subsets of a set that one can choose, when each subset must have k many elements and the set has n elements. Hence, $\binom{n}{k}$ is often read as "n choose k." For example, the set $A = \{a, b, c, d, e\}$ has five elements. How many subsets of A are there that have exactly three elements? The answer is $\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$.

Exercises 4.3

1. For each sequence find a formula that generates the sequence:

(a) $\frac{1}{3}, \frac{2}{9}, \frac{3}{27}, \frac{4}{81}, \frac{5}{243}, \ldots$ starting with $n = 1$.

(b) $\frac{1}{3}, \frac{2}{9}, \frac{3}{27}, \frac{4}{81}, \frac{5}{243}, \ldots$ starting with $n = 0$.

(c) $-\frac{1}{3}, \frac{2}{9}, -\frac{3}{27}, \frac{4}{81}, -\frac{5}{243}, \ldots$ starting with $n = 1$.

(d) $\frac{1}{3}, -\frac{2}{9}, \frac{3}{27}, -\frac{4}{81}, \frac{5}{243}, \ldots$ starting with $n = 0$.

2. Justify equations 3 and 4 of Example 3 on page 109.

3. Justify the identities 2 and 3 presented in Example 5 on page 110.

4. Rewrite $\sum_{k=1}^{n+1} \frac{1}{k^2}$ by separating the last term.

5. Rewrite $\sum_{k=1}^{n+1} 2^k$ by separating the last two terms.

6. Let $r \neq 0$ be a real number. Justify the identities:

(a) $\sum_{k=0}^{0} r^k = 1$.

(b) $\sum_{k=0}^{n+1} r^k = \left(\sum_{k=0}^{n} r^k\right) + r^{n+1}$ when $n \geq 0$.

(c) $\left(\sum_{k=0}^{n+3} r^k\right) - \left(\sum_{k=0}^{n} r^k\right) = r^n(r^3 + r^2 + r)$ when $n \geq 0$.

7. Find a formula $h(k)$ and a formula $f(k)$ that generates the respective sum:

(a) $\sum_{k=1}^{5} h(k) = (1 - \frac{1}{2}) + (\frac{1}{2} - \frac{1}{3}) + (\frac{1}{3} - \frac{1}{4}) + (\frac{1}{4} - \frac{1}{5}) + (\frac{1}{5} - \frac{1}{6})$.

(b) $\sum_{k=1}^{5} f(k) = (1 - \frac{1}{2}) - (\frac{1}{2} - \frac{1}{3}) + (\frac{1}{3} - \frac{1}{4}) - (\frac{1}{4} - \frac{1}{5}) + (\frac{1}{5} - \frac{1}{6})$.

8. Let $h(k)$ be any formula and let m, n, d be integers with $m \leq n$. Verify the following equality

$$\sum_{k=m}^{n} h(k) = \sum_{k=m+d}^{n+d} h(k-d)$$

(given in the shift rule) by expressing each side of this equality as an open sum.

9. Transform the following sums as requested:

(a) Shift the sum $\sum_{k=1}^{n} \frac{(k-1)^2}{k!}$ to an equal sum with starting value $k = 7$.

(b) Shift $\sum_{k=2}^{n} \frac{(k-1)^2}{k!}$ to an equal sum with starting value $k = -7$.

(c) Shift the summation $\sum_{k=-2}^{n+2} \frac{(k+1)^n}{(k+3)!}$ to an equal sum with starting value $k = 2$.

10. Suppose $\sum_{k=1}^{n} h(k)$ yields the values (4.6). Find formulas $h(k)$ and $g(n)$ so that $\sum_{k=1}^{n} h(k) = g(n)$ for all $n \geq 1$, where $g(n)$ is a proposed closed-form solution.

$$\begin{aligned}
\sum_{k=1}^{1} h(k) &= \frac{1}{1\cdot 3} = \frac{1}{3} \\
\sum_{k=1}^{2} h(k) &= \frac{1}{1\cdot 3} + \frac{1}{3\cdot 5} = \frac{2}{5} \\
\sum_{k=1}^{3} h(k) &= \frac{1}{1\cdot 3} + \frac{1}{3\cdot 5} + \frac{1}{5\cdot 7} = \frac{3}{7} \\
\sum_{k=1}^{4} h(k) &= \frac{1}{1\cdot 3} + \frac{1}{3\cdot 5} + \frac{1}{5\cdot 7} + \frac{1}{7\cdot 9} = \frac{4}{9}.
\end{aligned} \tag{4.6}$$

11. Suppose $\sum_{k=1}^{n} h(k)$ yields the values (4.7). Find formulas $h(k)$ and $g(n)$ so that $\sum_{k=1}^{n} h(k) = g(n)$ for all $n \geq 1$, where $g(n)$ is a proposed closed-form solution.

$$\begin{aligned}
\sum_{k=1}^{1} h(k) &= 1\cdot 1! = 1 \\
\sum_{k=1}^{2} h(k) &= 1\cdot 1! + 2\cdot 2! = 5 \\
\sum_{k=1}^{3} h(k) &= 1\cdot 1! + 2\cdot 2! + 3\cdot 3! = 23 \\
\sum_{k=1}^{4} h(k) &= 1\cdot 1! + 2\cdot 2! + 3\cdot 3! + 4\cdot 4! = 119.
\end{aligned} \tag{4.7}$$

12. Find a closed-form solution for the sum $\sum_{k=1}^{n} (4k^2 - 1)$.

13. Write $\left(\sum_{k=1}^{n} (6k-3)\right) + \left(\sum_{k=1}^{n} (4-5k)\right)$ as a single sum.

14. Write $2 \cdot \left(\sum_{k=1}^{n} (3k^2-4)\right) + 3 \cdot \left(\sum_{k=1}^{n} (3k^2+1)\right)$ as a single sum and obtain a closed-form solution for this single sum.

15. Find closed-form solutions for $\sum_{k=10}^{n} k$ and $\sum_{k=11}^{n} k^3$ using Theorems 4.3.4 and 4.3.3.

16. Using the shift rule, find a closed-form solution for $\sum_{k=-29}^{n} (k+28)^2$.

17. Show that $\sum_{k=1}^{n} (2^k - 2^{k-1}) = 2^n - 1$.

18. Let $0 \leq k \leq n$ be integers. Prove that $\binom{n}{k} = \binom{n}{n-k}$.

19. Let $1 \leq k \leq n$ be integers. Prove that $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$.

20. Justify the following sequence of equalities:

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

Exercise Notes: For Exercise 17, express $\sum_{k=1}^{n} (2^k - 2^{k-1})$ in open-form (see (4.4) on page 108) with the first four terms and the last two terms appearing in the open sum. For Exercise 11, compute 2!, 3!, 4!, 5!. For Exercise 20, review Exercise 19, Example 3 on page 109, and the shift rule. This sequence of equalities can be used to prove the Binomial Theorem (see Exercise 16 on page 122).

4.4 Proving Equations by Mathematical Induction

The key mathematical tool for verifying patterns and properties of a sequence or summation is mathematical induction. Consider the sum of the natural numbers in the sequence $1, 2, 3, \ldots, n$. The next theorem establishes a closed-form solution for this sum. We will first perform a "proof analysis" and then we will prove the theorem by induction.

Theorem 4.4.1. *For every integer $n \geq 1$, the equation $1+2+3+\cdots+n = \frac{n(n+1)}{2}$ holds.*

Proof Analysis. When you are using proof by induction, it is helpful to write down the statement $P(n)$. In this case we have

$$P(n):\ 1+2+3+\cdots+n = \frac{n(n+1)}{2}.$$

In this theorem our base value is $b = 1$. Next, we construct an induction proof diagram by first starting with

Base step: Prove $P(1)$.
Inductive step: Let $n \geq 1$ be an integer.
Assume $P(n)$.
Prove $P(n+1)$.

We now write out the statements $P(1)$ and $P(n+1)$. Replacing n everywhere in the statement

$$P(n):\quad 1+2+3+\cdots+n = \tfrac{n(n+1)}{2}$$

with 1, we obtain

$$P(1):\quad 1 = \tfrac{1(1+1)}{2}.$$

By replacing n with $n+1$ everywhere in $P(n)$, we obtain

$$P(n+1):\quad 1+2+3+\cdots+(n+1) = \tfrac{(n+1)(n+2)}{2}.$$

Thus, we obtain our desired last proof diagram:

Base step: Prove $1 = \frac{1(1+1)}{2}$.
Inductive step: Let $n \geq 1$ be an integer.
Assume $1+2+3+\cdots+n = \frac{n(n+1)}{2}$.
Prove $1+2+3+\cdots+(n+1) = \frac{(n+1)(n+2)}{2}$.

For the base step we must verify $1 = \frac{1(1+1)}{2}$. This is easy to verify, using simple arithmetic. For the inductive step, we must use the induction hypothesis

$$1+2+3+\cdots+n = \frac{n(n+1)}{2} \tag{IH}$$

to prove that the induction conclusion

$$1+2+3+\cdots+(n+1) = \frac{(n+1)(n+2)}{2} \tag{IC}$$

is true. To prove this conclusion, we begin with the left hand side of the equality in (IC) and make some changes so that the left hand side of the equality in (IH)

appears. Using the induction hypothesis (IH), we can replace $(1+2+3+\cdots+n)$ with $\frac{n(n+1)}{2}$ and obtain $\frac{(n+1)(n+2)}{2}$ which will complete our proof. Ⓐ

Our last proof diagram and analysis will guide the composition of a well-structured proof of Theorem 4.4.1 by mathematical induction.

Proof. We prove, by mathematical induction, that $1+2+3+\cdots+n=\frac{n(n+1)}{2}$ for all integers $n \geq 1$.

Base step: For $n=1$, we see that $1=\frac{1(1+1)}{2}$.

Inductive step: Let $n \geq 1$ be an integer and assume the induction hypothesis

$$1+2+3+\cdots+n=\frac{n(n+1)}{2}. \tag{IH}$$

We prove the equation $1+2+3+\cdots+(n+1)=\frac{(n+1)(n+2)}{2}$ as follows:

$$\begin{aligned}
1+2+3+\cdots+(n+1) &= (1+2+3+\cdots+n)+(n+1) && \text{by regrouping}\\
&= \frac{n(n+1)}{2}+(n+1) && \text{by ind. hypothesis (IH)}\\
&= \frac{n(n+1)}{2}+\frac{2(n+1)}{2} && \text{common denominator}\\
&= \frac{n(n+1)+2(n+1)}{2} && \text{by algebra}\\
&= \frac{(n+1)(n+2)}{2} && \text{by commutativity and distributivity.}
\end{aligned}$$

Hence, $1+2+3+\cdots+(n+1)=\frac{(n+1)(n+2)}{2}$ and the proof is complete. □

Our next theorem establishes an important formula in mathematics. A *geometric sequence* has the form

$$1, r, r^2, r^3, \ldots, r^n, r^{n+1}, \ldots$$

for some real number r. Note that $r^0=1$ when $r \neq 0$. Consider the open sum

$$\sum_{k=0}^{n} r^k = r^0 + r^1 + r^2 + \cdots + r^n.$$

We will show, when $r \neq 1$ and r is nonzero, that this sum has a closed-form solution. This closed-form solution is used throughout mathematics and has applications in physics, biology, economics, and computer science.

Theorem 4.4.2 (Sum of a geometric sequence). *Let $r \neq 1$ be a nonzero real number. For every integer $n \geq 0$, we have $\sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}$.*

Proof Analysis. We first write down the statement $P(n)$. In this case we have

$$P(n): \quad \sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}.$$

In this theorem our base value is $b = 0$. We generate an induction proof diagram by first starting with

Base step:	Prove $P(0)$.
Inductive step:	Let $n \geq 0$ be an integer.
	Assume $P(n)$.
	Prove $P(n+1)$.

We now carefully write out the statements $P(0)$ and $P(n+1)$. By replacing n everywhere in the statement

$$P(n): \sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}$$

with 0, we obtain

$$P(0): \sum_{k=0}^{0} r^k = \frac{r^{0+1}-1}{r-1}.$$

By replacing n with $(n+1)$ everywhere in $P(n)$ and then simplifying, we obtain

$$P(n+1): \sum_{k=0}^{n+1} r^k = \frac{r^{n+2}-1}{r-1}.$$

Thus, we obtain our desired final proof diagram:

Base step:	Prove $\sum_{k=0}^{0} r^k = \frac{r^{0+1}-1}{r-1}$.
Inductive step:	Let $n \geq 0$ be an integer.
	Assume $\sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}$.
	Prove $\sum_{k=0}^{n+1} r^k = \frac{r^{n+2}-1}{r-1}$.

For the base step, we must verify the equation $\sum_{k=0}^{0} r^k = \frac{r^{0+1}-1}{r-1}$. For the inductive step, we shall assume the induction hypothesis

$$\sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1} \tag{IH}$$

and use it to prove the induction conclusion

$$\sum_{k=0}^{n+1} r^k = \frac{r^{n+2}-1}{r-1}. \tag{IC}$$

To prove this conclusion, we begin with the left hand side of the equality in (IC) and make some changes so that the left hand side of the equality in (IH) appears.[3] Then, using the induction hypothesis (IH), we can replace $\left(\sum_{k=0}^{n} r^k\right)$ with $\frac{r^{n+1}-1}{r-1}$ and obtain $\frac{r^{n+2}-1}{r-1}$. Thus, we will have derived (IC). Ⓐ

Our analysis and final proof diagram motivates our proof of the Theorem 4.4.2 by mathematical induction.

Proof. Let $r \neq 1$ be a nonzero real number. We prove that $\sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}$ for all $n \geq 0$, by induction.

Base step: For $n = 0$, we have $\sum_{k=0}^{0} r^k = r^0 = 1$ and $\frac{r^{0+1}-1}{r-1} = 1$. So $\sum_{k=0}^{0} r^k = \frac{r^{0+1}-1}{r-1}$.

Inductive step: Let $n \geq 0$ be an integer and assume the induction hypothesis

$$\sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}. \tag{IH}$$

We show that $\sum_{k=0}^{n+1} r^k = \frac{r^{n+2}-1}{r-1}$ as follows:

$$\begin{aligned}
\sum_{k=0}^{n+1} r^k &= \left(\sum_{k=0}^{n} r^k\right) + r^{n+1} && \text{by property of } \Sigma \text{ notation} \\
&= \frac{r^{n+1}-1}{r-1} + r^{n+1} && \text{by induction hypothesis (IH)} \\
&= \frac{r^{n+1}-1}{r-1} + \frac{(r-1)r^{n+1}}{r-1} && \text{common denominator} \\
&= \frac{r^{n+1}-1+(r-1)r^{n+1}}{r-1} && \text{by algebra} \\
&= \frac{r^{n+1}-1+r^{n+2}-r^{n+1}}{r-1} && \text{by distributivity} \\
&= \frac{r^{n+2}-1}{r-1} && \text{by algebra.}
\end{aligned}$$

Hence, $\sum_{k=0}^{n+1} r^k = \frac{r^{n+2}-1}{r-1}$ and the proof is complete. □

[3] See item 2 of Example 3 on page 109.

The formula for the sum of a geometric sequence, given in Theorem 4.4.2, is valid for any nonzero $r \neq 1$. We conclude that $\sum_{k=0}^{n+1} 2^k = \frac{2^{n+2}-1}{2-1} = 2^{n+2} - 1$.

Before reading the proof of our next theorem, one should review the sentence presented after the solution of Example 3 on page 109.

Theorem 4.4.3. *Let $a_m, a_{m+1}, \ldots$ and $b_m, b_{m+1}, \ldots$ be sequences where m is an integer. Then $\sum_{k=m}^{n} (a_k + b_k) = \sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k$, for all $n \geq m$.*

Proof. Let $a_m, a_{m+1}, \ldots$ and $b_m, b_{m+1}, \ldots$ be sequences where m is an integer. We prove that $\sum_{k=m}^{n} (a_k + b_k) = \sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k$, for all $n \geq m$.

Base step: For $n = m$, we have $\sum_{k=m}^{m} (a_k + b_k) = a_m + b_m$. Since $\sum_{k=m}^{m} a_k = a_m$ and $\sum_{k=m}^{m} b_k = b_m$, we conclude that $\sum_{k=m}^{m} (a_k + b_k) = \sum_{k=m}^{m} a_k + \sum_{k=m}^{m} b_k$.

Inductive step: Let $n \geq m$ be an integer and assume the induction hypothesis

$$\sum_{k=m}^{n} (a_k + b_k) = \sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k. \tag{IH}$$

We show that $\sum_{k=m}^{n+1} (a_k + b_k) = \sum_{k=m}^{n+1} a_k + \sum_{k=m}^{n+1} b_k$ as follows:

$$\begin{aligned}
\sum_{k=m}^{n+1} (a_k + b_k) &= \left(\sum_{k=m}^{n} (a_k + b_k) \right) + a_{n+1} + b_{n+1} && \text{property of } \Sigma \text{ notation} \\
&= \left(\sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k \right) + a_{n+1} + b_{n+1} && \text{by ind. hypothesis (IH)} \\
&= \left(\sum_{k=m}^{n} a_k \right) + a_{n+1} + \left(\sum_{k=m}^{n} b_k \right) + b_{n+1} && \text{by commutativity and associativity} \\
&= \sum_{k=m}^{n+1} a_k + \sum_{k=m}^{n+1} b_k && \text{property of } \Sigma \text{ notation.}
\end{aligned}$$

Hence, $\sum_{k=m}^{n+1} (a_k + b_k) = \sum_{k=m}^{n+1} a_k + \sum_{k=m}^{n+1} b_k$ and the proof is complete. □

Exercises 4.4

1. Prove, for every integer $n \geq 1$, that $1+3+5+\cdots+(2n-1) = n^2$.

2. Prove, for every integer $n \geq 0$, that $1+2+2^2+\cdots+2^n = 2^{n+1}-1$.

3. Prove that $2+6+18+\cdots+2\cdot 3^{n-1} = 3^n - 1$, for every integer $n \geq 1$.

4. Prove that $\left(1-\frac{1}{4}\right)\left(1-\frac{1}{9}\right)\cdots\left(1-\frac{1}{n^2}\right) = \frac{n+1}{2n}$, for all integers $n \geq 2$.

5. Prove that $\sum_{k=1}^{n}(2^k - 2^{k-1}) = 2^n - 1$ for all integers $n \geq 1$, by mathematical induction.

6. Prove that $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$ holds, for every integer $n \geq 1$.

7. Let c be a real number and let $a_m, a_{m+1}, \ldots$ be a sequence where m is an integer. Prove that $\sum_{k=m}^{n} ca_k = c\sum_{k=m}^{n} a_k$, for all $n \geq m$.

8. Prove, for every integer $n \geq 1$, that $\sum_{k=1}^{n} \frac{1}{4k^2-1} = \frac{n}{2n+1}$.

9. Prove that $\sum_{k=1}^{n} k\cdot k! = (n+1)! - 1$, for all integers $n \geq 1$.

10. Prove that $\sum_{k=1}^{n}(-1)^k k^2 = (-1)^n \frac{n(n+1)}{2}$, for all integers $n \geq 1$.

11. Prove, for every integer $n \geq 1$, that $\sum_{k=1}^{n} k^3 = \left[\frac{n(n+1)}{2}\right]^2$.

12. Prove for all integers $n \geq 1$ that $\binom{n}{k}$ is a natural number whenever k is an integer satisfying $0 \leq k \leq n$.

13. Find a closed-form solution for the sum $\sum_{k=0}^{n}(2^k - 1)(3^k + 1)$.

14. Prove that $\sum_{k=1}^{n} \frac{1}{(2k-1)(2k+1)} = \frac{n}{2n+1}$, for all integers $n \geq 1$.

15. Let $r \neq 1$ be nonzero. Find a closed-form solution for the sum $\sum_{k=-9}^{n} r^k$.

16. (The Binomial Theorem) Let x and y be variables, representing real numbers. Prove for every integer $n \geq 1$ that $\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k = (x+y)^n$.

Exercise Notes: For Exercise 8, note that $4(n+1)^2 - 1 = (2n+1)(2n+3)$. For Exercise 12, use induction on n and Exercise 19 on page 116. For Exercise 16, in the inductive step use Exercise 20 on page 116.

4.5 More Proofs by Mathematical Induction

In this section we use mathematical induction to prove statements about inequality, divisibility, and recursively defined sequences. The basic steps in each proof are the same as before; one must identify the statement $P(n)$ and then prove the base step and inductive step. Before reading the proofs of our next lemma and theorem, the reader should review the substitution properties of inequality 3.3.3 presented on page 69. The following Lemma 4.5.1 will be used in the proof of Theorem 4.5.2.

Lemma 4.5.1. *Let n be an integer. If $n \geq 3$, then $n^2 > 2n+1$.*

Proof. Assume $n \geq 3$. We prove that $n^2 > 2n+1$ as follows:

$$\begin{aligned} n^2 &= n \cdot n && \text{by property of exponents} \\ &\geq 3n && \text{because } n \geq 3 \\ &= 2n+n && \text{as } 3n = 2n+n \\ &> 2n+1 && \text{because } n > 1. \end{aligned}$$

Therefore, $n^2 > 2n+1$. □

Theorem 4.5.2. *For every integer $n \geq 5$, the inequality $2^n > n^2$ holds.*

Proof. We prove, by mathematical induction, that $2^n > n^2$ for all $n \geq 5$.

Base step: For $n = 5$, we see that $2^n = 32$ and $n^2 = 25$. Thus, $2^5 > 5^2$.

Inductive step: Let $n \geq 5$ be an integer. We shall show, assuming the induction hypothesis (IH) $2^n > n^2$, that $2^{n+1} > (n+1)^2$ as follows:

$$\begin{aligned} 2^{n+1} &= 2 \cdot 2^n && \text{property of exponents} \\ &> 2n^2 && \text{by the induction hypothesis (IH)} \\ &= n^2 + n^2 && \text{because } 2a = a+a \\ &> n^2 + 2n + 1 && \text{because } n^2 > 2n+1, \text{ by Lemma 4.5.1} \\ &= (n+1)^2 && \text{by factoring.} \end{aligned}$$

Therefore, $2^{n+1} > (n+1)^2$ and the proof is complete. □

Theorem 4.5.3. *For every integer $n \geq 1$, we have $3 \mid (n^3 - n)$.*

Proof. We prove, by mathematical induction, that $3 \mid (n^3 - n)$ for all $n \geq 1$.

Base step: For $n = 1$, we see that $(1^3 - 1) = 0$. Since $3 \mid 0$, we see that $3 \mid (1^3 - 1)$.

Inductive step: Let $n \geq 1$ be an integer. Assume the induction hypothesis $3 \mid (n^3 - n)$, that is, assume

$$n^3 - n = 3j \text{ for some } j \in \mathbb{N}. \qquad \text{(IH)}$$

We show that $(n+1)^3-(n+1)$ is evenly divisible by 3 as follows:

$$\begin{aligned}(n+1)^3-(n+1) &= n^3+3n^2+3n+1-n-1 && \text{by expanding } (n+1)^3\\ &= (n^3-n)+3n^2+3n && \text{by regrouping}\\ &= 3j+3n^2+3n && \text{by (IH)}\\ &= 3(j+n^2+n) && \text{by distributivity.}\end{aligned}$$

Thus, $(n+1)^3-(n+1)=3k$ where $k=j+n^2+n$ is an integer and therefore, $3\,|\,((n+1)^3-(n+1))$. □

4.5.1 Recursive (Inductive) Definitions

A sequence is defined recursively if some its initial terms are first specified and each of the remaining terms are then defined using one (or more) of the earlier terms. For example, suppose that the first term of a sequence is given to be $a_1=1$, and to form the second term we add 3 to the first term to obtain $a_2=a_1+3=4$. Similarly, to get the third tern we must add 3 to a_2 and so, $a_3=a_2+3=7$. By continuing in this manner we can construct every term of the sequence. A more succinct way of describing this sequence is to first define $a_1=1$ and then define the remaining terms by $a_{n+1}=a_n+3$ for all natural numbers $n\geq 1$. Actually, one can prove by induction that each term in this sequence is given by $a_n=3n-2$.

We now offer a general description of a recursively defined sequence where the first term is identified and thereafter, each successive term is defined in terms of the previous term.

Defining Sequences by Recursion. Suppose that $M[x]$ is a "method" or formula for computing a number m using another number x. We write this as $m=M[x]$. Let b be a given number and let i be a fixed integer. We define an infinite sequence $a_i,a_{i+1},a_{i+2},\ldots,a_n,\ldots$ of numbers, starting at $n=i$, *by recursion* as follows:

(1) $a_i=b$
(2) $a_{n+1}=M[a_n]$ for all $n\geq i$.

A definition involving the above steps (1) and (2) is called a **recursive definition** or an **inductive definition**. Step (1) identifies the first term a_i of the sequence and step (2) allows us to compute the value a_{n+1}, if we already know the value of a_n.

Example 1. Define a sequence $a_1,a_2,a_3,\ldots$ by recursion by using $b=1$ and using the "method" $M[a_n]=2a_n+1$. Determine the values a_1, a_2, a_3, and a_4.

Solution. The sequence is defined by the recursion

1. $a_1=1$
2. $a_{n+1}=2a_n+1$ for all $n\geq 1$.

So, $a_1 = 1$, $a_2 = 2a_1 + 1 = 3$, $a_3 = 2a_2 + 1 = 7$, and $a_4 = 2a_3 + 13 = 15$. Is there a formula $f(n)$ such that $a_n = f(n)$ for all $n \geq 1$? Ⓢ

Example 2 (Factorial function $n!$). Define a sequence $a_0, a_1, a_2, \ldots$ by recursion by using $b = 1$ and using the "method" $M[a_n, n] = (n+1) \cdot a_n$. Determine the values of a_0, a_1, a_2, a_3 and a_4.

Solution. The sequence is defined by the recursion

1. $a_0 = 1$
2. $a_{n+1} = (n+1)a_n$ for all $n \geq 0$.

We see that $a_0 = 1$, $a_1 = 1 \cdot a_0 = 1$, $a_2 = 2a_1 = 2$, $a_3 = 3a_2 = 6$ and $a_4 = 4a_3 = 24$. Observe that $a_n = n!$ for all $n \geq 0$. Ⓢ

If a sequence has been defined recursively, then proofs of statements about this sequence often use "proof by induction." Example 2 shows that $n!$ can be defined recursively. Thus, the proof of our next theorem is by mathematical induction.

Theorem 4.5.4. *For every integer $n \geq 4$, we have $n! > 2^n$.*

Proof. We prove, by mathematical induction, that $n! > 2^n$ for all $n \geq 4$.

Base step: For $n = 4$, we see that $n! = 24$ and $2^n = 16$. So $4! > 2^4$.

Inductive step: Let $n \geq 4$ be an integer and assume (IH) $n! > 2^n$. We prove that $(n+1)! > 2^{n+1}$ as follows:

$$\begin{aligned}
(n+1)! &= (n+1)n! \\
&> (n+1)2^n && \text{by induction hypothesis (IH)} \\
&> 2 \cdot 2^n && \text{because } n+1 > 2, \text{ as } n \geq 4 \\
&= 2^{n+1} && \text{by property of exponents.}
\end{aligned}$$

Hence, $(n+1)! > 2^{n+1}$ and this completes the proof. □

When a sequence is defined by recursion, it can be very difficult or impossible to find an explicit formula for the sequence. If you discover such a formula, then you must prove that the formula is correct by using mathematical induction.

Example 3. Consider the sequence $a_1, a_2, a_3, \ldots$ defined recursively by

1. $a_1 = 1$
2. $a_{n+1} = 2a_n + 1$ for all $n \geq 1$.

Find a formula for a_n and prove (by induction) that your formula is correct.

Solution. Using $a_1 = 1$ and $a_n = 2a_{n-1} + 1$ for $n \geq 2$, we obtain the following table of values: One makes the conjecture that $a_n = 2^n - 1$, for all $n \geq 1$. Ⓢ

n	a_n
1	1
2	3
3	7
4	15
5	31
6	63
7	127

Theorem 4.5.5. *Consider the sequence $a_1, a_2, a_3, \ldots$ defined recursively by*

1. $a_1 = 1$
2. $a_{n+1} = 2a_n + 1$ *for all* $n \geq 1$.

Then, $a_n = 2^n - 1$ *for all* $n \geq 1$.

Proof. We prove, by mathematical induction, that $a_n = 2^n - 1$ for all $n \geq 1$.

Base step: For $n = 1$, we see that $a_n = 1$ and $2^1 - 1 = 1$. Thus, $a_1 = 2^1 - 1$.

Inductive step: Let $n \geq 1$ be an integer and assume that (IH) $a_n = 2^n - 1$. We show that $a_{n+1} = 2^{n+1} - 1$ as follows:

$$\begin{aligned} a_{n+1} &= 2a_n + 1 && \text{by definition of sequence} \\ &= 2(2^n - 1) + 1 && \text{by induction hypothesis (IH)} \\ &= 2^{n+1} - 1 && \text{by arithmetic.} \end{aligned}$$

Therefore, $a_{n+1} = 2^{n+1} - 1$ and the proof is complete. □

Exercises 4.5

1. Prove for each integer $n \geq 1$, that $4^n - 1$ is divisible by 3.
2. Prove that $8 \mid (9^n - 1)$, for every integer $n \geq 1$.
3. Prove that the inequality $2^n > n$ holds for every integer $n \geq 2$.
4. Prove, for every integer $n \geq 6$, that $n! > 2^{n+2}$.
5. Let a, b be positive real numbers where $a < b$. Prove that $a^n < b^n$, for every integer $n \geq 1$.
6. Prove that $6 \mid n(n^2 + 5)$, for every integer $n \geq 1$.
7. Let $x \geq -1$ be a real number. Prove that $(1 + x)^n \geq 1 + nx$ for all integers $n \geq 1$.
8. Consider the sequence $a_1, a_2, a_3, \ldots$ defined recursively by
   (a) $a_1 = 1$
   (b) $a_{n+1} = 5a_n + 1$ for all $n \geq 1$.
   Prove, by induction, that $a_n = \frac{5^n - 1}{4}$ for all integers $n \geq 1$.

9. Consider the sequence $a_1, a_2, a_3, \ldots$ defined recursively by

(a) $a_1 = 3$
(b) $a_{n+1} = 2a_n + 3$ for all $n \geq 1$.

Prove, by induction, that $a_n = 3(2^n - 1)$ for all integers $n \geq 1$.

10. Consider the sequence $a_1, a_2, a_3, \ldots$ defined recursively by

(a) $a_1 = 2$
(b) $a_{n+1} = (n+1)a_n$ for all $n \geq 1$.

Prove, by induction, that $a_n \geq 2^n$ for all integers $n \geq 1$.

11. Consider the sequence $s_1, s_2, \ldots$ defined recursively by

(a) $s_1 = 2$
(b) $s_{n+1} = \frac{1}{4}(s_n + 5)$ for all $n \geq 1$.

Prove, by induction, that $1 \leq s_{n+1} \leq s_n \leq 2$ for all integers $n \geq 1$.

12. Consider the sequence $a_1, a_2, a_3, \ldots$ defined recursively by

(a) $a_1 = 1$
(b) $a_{n+1} = \frac{a_n}{n+1}$ for all $n \geq 1$.

Prove, by induction, that $a_n = \frac{1}{n!}$ for all integers $n \geq 1$.

13. Consider the sequence $a_1, a_2, a_3, \ldots$ defined recursively by

(a) $a_1 = 3$
(b) $a_{n+1} = a_n + 2n + 3$ for all $n \geq 1$.

Find a formula for a_n and prove (by induction) that your formula is correct.

14. Let $c < d$ be real numbers. Suppose that $a_1, a_2, a_3, \ldots$ is an infinite sequence of real numbers satisfying $c \leq a_n \leq d$ for all $n \geq 1$. Prove that $c \leq \frac{a_1 + a_2 + \cdots + a_n}{n} \leq d$ for all integers $n \geq 1$.

Exercise Notes: Exercise 7 is called Bernoulli's inequality. For Exercise 11, in the inductive step start with $s_{n+1} \leq s_n$ and derive the inequality $s_{n+2} \leq s_{n+1}$; also start with $1 \leq s_{n+1}$ and derive $1 \leq s_{n+2}$. For Exercise 14, multiply both sides of the induction hypothesis inequality by n and then use the assumption $c \leq a_{n+1} \leq d$, together with Theorem 3.1.8.

4.6 Strong Mathematical Induction

A variation of "proof by induction" arises when in the inductive step you are having difficulty relating $P(n+1)$ to $P(n)$, or when the relationship you discover does not prove fruitful. Suppose, however, that you can relate $P(n)$ to $P(k)$ for some (or all) k satisfying $b \leq k < n$, where b is the base value. In this case, you would like to

use the fact that $P(k)$ is true to conclude that $P(n)$ is true; but, can you assume that $P(k)$ is true? The answer is yes and, in fact, you can assume all the statements $P(b), P(b+1), \ldots, P(n-1)$ are true.

4.6.1 Strong Induction with One Base Step

Principle of Strong Induction I. Let b be a given integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. Suppose that

1. $P(b)$ is true, and
2. For all $n > b$, if $P(b)$, $P(b+1)$, …, $P(n-1)$ are true, then $P(n)$ is true.

Then for *all* integers $n \geq b$, the statement $P(n)$ is true.

Remark. Another way to assert that all the statements $P(b), P(b+1), \ldots, P(n-1)$ hold is to write $\forall k(b \leq k < n \rightarrow P(k))$.

The principle of strong induction justifies our next proof strategy, which is often used in mathematical proofs.

Proof Strategy 4.6.1. Let b be a fixed integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. To prove $(\forall n \geq b)P(n)$ by **strong** induction, use the diagram

Prove $P(b)$.
Prove $(\forall n > b)[(\forall k(b \leq k < n \rightarrow P(k)) \rightarrow P(n)]$.

In other words, use the diagram

Base step: Prove $P(b)$.
Inductive step: Let $n > b$ be an integer.
Assume $P(k)$ whenever $b \leq k < n$.
Prove $P(n)$.

So to prove a statement $(\forall n \geq b)P(n)$ by strong induction, one would use the above final diagram in Proof Strategy 4.6.1. One way to interpret this final diagram is, once again, to think of the domino effect as pictured in Fig. 4.2. The base step verifies that the first domino will fall. The above inductive step can be thought of asserting that each domino is perfectly aligned with *all* of the dominos that come before it. Thus, we can be sure that the first domino will fall (base step) and whenever all of the dominoes that precede the n-th domino fall, we can be certain that the n-th domino will also fall (inductive step). Therefore, all of the dominoes must fall.

A proof of a statement of the form $(\forall n \geq b)P(n)$ by strong induction is very much like a proof by mathematical induction. Strong induction just changes the focus slightly. Rather than trying to prove $P(n+1)$, one tries to prove the statement

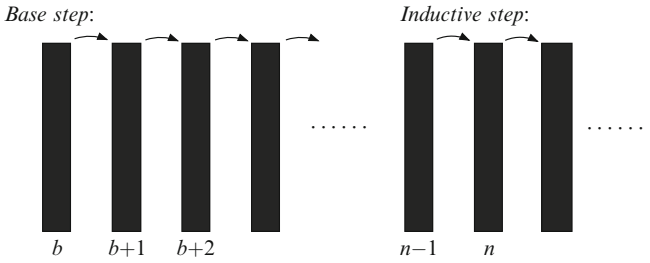


Fig. 4.2 Strong induction forces all of the dominoes to fall

$P(n)$ by relating $P(n)$ to one (or all) of the statements $P(k)$ where $k < n$. Our next theorem applies the strong induction Proof Strategy 4.6.1 to establish a result on the uniqueness of certain types of decimal expansions.

One can prove that every real number has an infinite decimal expansion. For example, $\frac{1}{4} = 0.2500000\cdots$, $\frac{4}{3} = 1.3333333\cdots$, and $\pi = 3.1415926\cdots$. Some real numbers have two infinite decimal expansions. It is easy to show (see Example 2 on page 63) that $1.00000\cdots = 0.99999\cdots$ and thus, the number 1 has more than one decimal expansion. One can also show that $0.450000000\cdots = 0.449999999\cdots$. So, decimal expansions are not necessarily unique. Let us now consider decimal expansions where 9 is not allowed as a digit. Can a real number have two different decimal expansions, neither of which contain any 9's? Does $0.8888888\cdots$ have a second decimal expansion that also does not contain any 9's? Our next theorem shows that decimal expansions of the form $0.z_1z_2z_3\cdots$ with $0 \le z_i \le 8$ are unique.

Theorem 4.6.2. *Suppose* $0.x_1x_2x_3\cdots = 0.y_1y_2y_3\cdots$ *where* $0 \le x_i \le 8$ *and* $0 \le y_i \le 8$ *for each natural number i. Then* $x_n = y_n$ *for all* $n \ge 1$.

Proof Analysis. We write down the statement $P(n)$, which is $x_n = y_n$. Next, we construct a proof diagram using the strong induction Proof Strategy 4.6.1:

Base step:	Prove $x_1 = y_1$.
Inductive step:	Let $n > 1$ be an integer.
	Assume $x_k = y_k$ whenever $1 \le k < n$.
	Prove $x_n = y_n$.

This diagram will guide our proof. Ⓐ

Proof. We have that

$$0.x_1x_2x_3\cdots = 0.y_1y_2y_3\cdots \tag{4.8}$$

where $0 \le x_i, y_i \le 8$ for each $i \ge 1$. We prove that $x_n = y_n$ for all $n \ge 1$.

Base step: Let $n = 1$. We will prove that $x_1 = y_1$. By multiplying both sides of (4.8) by 10, we obtain the equations

$$\begin{aligned} x_1.x_2x_3x_4\cdots &= y_1.y_2y_3y_4\cdots \\ x_1 + 0.x_2x_3x_4\cdots &= y_1 + 0.y_2y_3y_4\cdots . \end{aligned} \tag{4.9}$$

Let $a = 0.x_2x_3x_4\cdots$ and $b = 0.y_2y_3y_4\cdots$. Since $0 \le a \le 0.888\cdots < 1$, it follows that $0 \le a < 1$. Similarly, we conclude that $0 \le b < 1$. Hence, $-1 < b - a < 1$. So, if $b - a$ is an integer, then $b - a = 0$. We can now rewrite (4.9) as $x_1 + a = y_1 + b$ and, by algebra, obtain the equation $x_1 - y_1 = b - a$. Since $x_1 - y_1$ is an integer, we see that $b - a$ is also an integer. Therefore $b - a = 0$ and consequently, $x_1 = y_1$.

Inductive step: Let $n > 1$ be a natural number and assume the strong induction hypothesis

$$x_1 = y_1,\ x_2 = y_2,\ \ldots,\ x_{n-1} = y_{n-1}. \tag{IH}$$

We shall prove that $x_n = y_n$. By multiplying both sides of (4.8) by 10, we obtain

$$x_1.x_2x_3x_4\cdots = y_1.y_2y_3y_4\cdots$$

and thus,

$$x_1 + 0.x_2x_3x_4\cdots = y_1 + 0.y_2y_3y_4\cdots. \tag{4.10}$$

Since $x_1 = y_1$, we see from (4.10) that

$$0.x_2x_3x_4\cdots = 0.y_2y_3y_4\cdots. \tag{4.11}$$

By multiplying equation (4.11) by 10 we obtain

$$x_2 + 0.x_3x_4x_5\cdots = y_2 + 0.y_3y_4y_5\cdots$$

and, since $x_2 = y_2$, we see that $0.x_3x_4x_5\cdots = 0.y_3y_4y_5\cdots$. Because $x_3 = y_3$, $\ldots$, $x_{n-1} = y_{n-1}$, we can continue this reasoning and conclude that

$$0.x_nx_{n+1}x_{n+2}\cdots = 0.y_ny_{n+1}y_{n+2}\cdots. \tag{4.12}$$

The argument used in the base step, applied to (4.12), shows that $x_n = y_n$. Therefore, $x_n = y_n$ for all $n \ge 1$. □

Remark 4.6.3. We shall say that a decimal expansion of a real number ends with a *string* of 9's if it ends with an infinite repeating sequence of 9's; for example, the decimal expansion $0.12327739999999\cdots$ ends with a string of 9's. Similarly, we will say the a decimal expansion ends with a *string* of 0's if it ends with an infinite repeating sequence of 0's. Exercises 2 and 3 of this section imply that if a real number has two different decimal expansions, then one of the expansions must end with a string of 9's and the other expansion must end with a string of 0's. Consequently, the real number $0.989898\cdots$ has no other decimal expansions. Furthermore, one can prove that a real number which has a decimal expansion ending in a string of 9's also has a decimal expansion that ends with a string of 0's.

4.6.2 Strong Induction with Multiple Base Steps

Suppose you are trying to prove a statement of the form $(\forall n \geq b)P(n)$ by strong induction and, in the inductive step, you see how to relate $P(n)$ to some $P(k)$ where $k < n$. In addition, you also discover that it may be the case that $k < b$. You know that when $k < b$ one cannot assume $P(k)$ as part of the induction hypothesis. Is there a way to around this difficulty? To handle such unusual cases, we introduce another version of strong induction in which there can be multiple base steps.

Principle of Strong Induction II. Let b be an integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. Suppose for an integer $c > b$ we have that

1. $P(b), P(b+1), \ldots, P(c)$ are all true, and
2. For all $n > c$, if $P(b), P(b+1), \ldots, P(n-1)$ are true, then $P(n)$ is true.

Then for *all* integers $n \geq b$, the statement $P(n)$ is true.

The above item 2 is similar to the corresponding item in the Strong Induction Principle I (see page 128); however, item 1 is different and states that the property must be true for more than one initial value. Our next proof strategy shows how to put all of the pieces of this principle together to obtain a proof by strong induction. This strategy is essentially the same as that for strong induction with one base step. The new ingredient is that one has to know, beforehand, the number of base steps that are required in the proof. We will address this issue in Remark 4.6.5.

Proof Strategy 4.6.4. Let b be an integer and let $P(n)$ be a statement that is defined for all integers $n \geq b$. To prove $(\forall n \geq b)P(n)$ by **strong** induction, identify the integer $c > b$ and then use the diagram

Prove $P(b)$.
Prove $P(b+1)$.
$\vdots$
Prove $P(c)$.
Prove $(\forall n > c)[(\forall k(b \leq k < n \to P(k)) \to P(n)]$.

In other words, use the diagram

Base steps: Prove $P(i)$ for each i satisfying $b \leq i \leq c$.
Inductive step: Let $n > c$ be an integer.
Assume $P(k)$ whenever $b \leq k < n$.
Prove $P(n)$.

Remark 4.6.5. In this latter proof diagram, the value one needs for c depends on the proof of $P(n)$ in the inductive step. Suppose the proof of $P(n)$ requires $P(n_1), P(n_2), \ldots, P(n_j)$ to hold in the induction hypothesis where $n_1, n_2, \ldots, n_j < n$.

Then c is chosen so that if $n > c$, we will be assured that each $n_i \geq b$ when $1 \leq i \leq j$. We shall illustrate this idea in the proof of our next theorem which involves three base steps.

Theorem 4.6.6. *Let $a_1, a_2, \ldots$ be the sequence recursively defined by*

1. $a_1 = 1$
2. $a_2 = 2$
3. $a_3 = 3$
4. $a_n = a_{n-1} + a_{n-2} + a_{n-3}$ *for all integers* $n > 3$.

Then $a_n \leq 2^n$ for all integers $n \geq 1$.

Proof Analysis. We write down the statement $P(n)$. In this case we have

$$P(n) : a_n \leq 2^n.$$

Next, we construct a proof diagram using the strong induction Proof Strategy 4.6.1. Our base value is $b = 1$. In our inductive step, we will need to prove that $a_n \leq 2^n$. Since $a_n = a_{n-1} + a_{n-2} + a_{n-3}$, our induction hypothesis must hold for $n-1$, $n-2$ and $n-3$. Thus we must have $n - 3 \geq 1$, that is, we need $n \geq 4$. So, we take $c = 3$ and obtain the following proof diagram:

Base step:	Prove $a_1 \leq 2^1$.
Base step:	Prove $a_2 \leq 2^2$.
Base step:	Prove $a_3 \leq 2^3$.
Inductive step:	Let $n > 3$ be an integer.
	Assume $a_k \leq 2^k$ when $1 \leq k < n$.
	Prove $a_n \leq 2^n$.

This diagram will guide our proof. Ⓐ

Proof. Let $a_1, a_2, \ldots$ be the sequence defined recursively in the statement of the theorem. We prove, by strong induction, that $a_n \leq 2^n$ for all $n \geq 1$.

Base step: We see that $a_1 = 1 \leq 2^1$.
Base step: We see that $a_2 = 2 \leq 2^2$.
Base step: We see that $a_3 = 3 \leq 2^3$.

Inductive step: Let $n > 3$ be an integer and assume

$$a_k \leq 2^k \text{ whenever } 1 \leq k < n. \tag{IH}$$

We will show that $a_n \leq 2^n$. Note that $a_n = a_{n-1} + a_{n-2} + a_{n-3}$ by item 4 of the definition of the sequence. Thus

$$\begin{aligned} a_n &= a_{n-1} + a_{n-2} + a_{n-3} && \text{by definition of sequence} \\ &\leq 2^{n-1} + 2^{n-2} + 2^{n-3} && \text{by induction hypothesis (IH)} \end{aligned}$$

$$\begin{aligned} &= 2^{n-3}(2^2 + 2^1 + 1) && \text{by distributivity} \\ &\leq 2^{n-3} \cdot 2^3 && \text{because } 2^2 + 2^1 + 1 = 7 \leq 2^3 \\ &= 2^n && \text{by property of exponents.} \end{aligned}$$

Therefore, $a_n \leq 2^n$ and the proof is complete. □

When we divide a natural number by 4 we expect a remainder between 0 and 3. We will now give a proof establishing that this is the case.

Theorem 4.6.7. *Every integer $n \geq 0$ can be expressed as $n = 4q + r$ for integers q and r where $0 \leq r < 4$.*

Proof Analysis. We write down the statement $P(n)$. In this case we have

$$P(n) : \text{“}n = 4q + r \text{ for some integers } q \text{ and } 0 \leq r < 4.\text{”}$$

Next, we construct a proof diagram using the strong induction Proof Strategy 4.6.1. Our base value is $b = 0$. In our proof we will need the induction hypothesis to hold for $n - 4$. Thus, we require that $n - 4 \geq 0$ and so, we need $n \geq 4$. So, we take $c = 3$ (see Remark 4.6.5). We use $b = 0$ and $c = 3$ to obtain the proof diagram:

Base step: Prove $0 = 4q + r$ for some integers q and $0 \leq r < 4$.
Base step: Prove $1 = 4q + r$ for some integers q and $0 \leq r < 4$.
Base step: Prove $2 = 4q + r$ for some integers q and $0 \leq r < 4$.
Base step: Prove $3 = 4q + r$ for some integers q and $0 \leq r < 4$.
Induct. step: Let $n > 3$ be an integer.
Assume $k = 4i + j$ for integers i and $0 \leq j < 4$, if $0 \leq k < n$.
Prove $n = 4q + r$ for some integers q and $0 \leq r < 4$.

We will use this diagram as a guide for our proof. Ⓐ

Proof. We prove, by strong mathematical induction, that every integer $n \geq 0$ can be written as $n = 4q + r$ for some integers q and r where $0 \leq r < 4$.

Base step: We see that $0 = 4(0) + 0$.
Base step: We see that $1 = 4(0) + 1$.
Base step: We see that $2 = 4(0) + 2$.
Base step: We see that $3 = 4(0) + 3$.

Inductive step: Let $n > 3$ be an integer and assume

$$k = 4i + j \text{ for some integers } i \text{ and } j \text{ where } 0 \leq j < 4, \text{ whenever } 0 \leq k < n. \quad \text{(IH)}$$

We will show that $n = 4q + r$ for some integers q and $0 \leq r < 4$. Since $n > 3$, we have that $0 \leq n - 4 < n$. The induction hypothesis (IH) implies that $(*)$ $n - 4 = 4i + r$ for some integers i and r where $0 \leq r < 4$. Solving $(*)$ for n, we obtain $n = 4(i + 1) + r$. Thus, $n = 4q + r$ where $q = i + 1$ and $0 \leq r < 4$ are integers. This completes the proof. □

Our next theorem, Theorem 4.6.8, has many applications in number theory and abstract algebra. Theorem 4.6.8 generalizes Theorem 4.6.7 and precisely expresses the outcome of the usual process of division by an integer $d \geq 1$. Our proof of this theorem is very similar to the proof of Theorem 4.6.7. In particular, the base steps show that the result holds for each of the integers $0, 1, 2, \ldots, d-1$.

Theorem 4.6.8. *Let n and d be integers, where $n \geq 0$ and $d \geq 1$. Then there exist unique integers q and r such that $n = dq + r$ and $0 \leq r < d$.*

Proof. Let $d \geq 1$ be an integer. We first prove, by strong mathematical induction, that every integer $n \geq 0$ can be written as $n = dq + r$ for some integers q and r where $0 \leq r < d$.

Base steps: When $0 \leq \ell \leq d-1$, we see that $\ell = d \cdot 0 + \ell$ where $0 \leq \ell < d$.

Inductive step: Let $n > d-1$ be an integer and assume the induction hypothesis

$$k = di + j \text{ for some integers } i \text{ and } j \text{ where } 0 \leq j < d, \text{ whenever } 0 \leq k < n. \quad \text{(IH)}$$

We show that $n = dq + r$ for some integers q and r where $0 \leq r < d$. Clearly, we have that $0 \leq n - d < n$, since $n \geq d > 0$. So, by the induction hypothesis (IH), we conclude that $n - d = di + r$ for some integers i and r where $0 \leq r < d$. Solving for n, we obtain $n = d(i+1) + r$. Thus, letting $q = i + 1$, we have that $n = dq + r$ and $0 \leq r < d$. This completes the induction proof.

To show uniqueness, let n and d be integers where $n \geq 0$ and $d \geq 1$. Suppose that $n = dq + r$ for integers q and r where $0 \leq r < d$, and suppose i and j are also integers satisfying $n = di + j$ and $0 \leq j < d$. We shall show that $q = i$ and $r = j$. Since $n = dq + r$ and $n = di + j$, it follows that $dq + r = di + j$ and so, $(\star)$ $(q-i)d = j - r$. First we prove that $r = j$. Suppose, for a contradiction, that $r \neq j$. Without loss of generality, we shall presume that $r < j$. Hence, $0 < j - r$. Because $0 \leq r < j < d$, we obtain $0 < j - r < d$. We now conclude from $(\star)$ that $0 < (q-i)d < d$. But then $0 < (q-i) < 1$; however, since $(q-i)$ is an integer, the inequality $0 < (q-i) < 1$ cannot hold. This contradiction shows that we must have $r = j$ and equation $(\star)$ implies that $q = i$. □

Example 1. Let $n = 173$ and $d = 9$. To find q and r satisfying $n = dq + r$ with $0 \leq r < 9$, one performs the long division $9\overline{)173}$ to obtain the quotient $q = 19$ and the remainder $r = 2$. Thus, $173 = 9 \cdot 19 + 2$.

Theorem 4.6.8 applies to all non-negative integers n. We shall extend this theorem to any integer n. This extension is called the *division algorithm* because, as noted earlier, if $|n| \geq d$, then one can obtain the quotient q and remainder r by performing long division.

Theorem 4.6.9 (Division Algorithm). *Let n and d be integers, where $d \geq 1$. Then there exist unique integers q and r such that $n = dq + r$ and $0 \leq r < d$.*

Proof. Let n and d be integers where $d \geq 1$. If $n \geq 0$, then the result follows from Theorem 4.6.8. Suppose $n < 0$. Then $-n > 0$. By Theorem 4.6.8, there are integers

i and j such that $-n = di + j$ and $0 \leq j < d$. So $(\star)$ $n = d(-i) - j$. Since $0 \leq j < d$, there are two cases to consider; namely, $0 = j$ and $0 < j$. If $0 = j$, then $(\star)$ becomes $n = d(-i) + 0$ and we can let $q = -i$ and $r = 0$. If $0 < j$ then, since $0 < j < d$, we see that $0 < d - j < d$. We can rewrite $(\star)$ as $n = d(-i) - d + d - j = d(-i-1) + (d-j)$ and so, we can let $q = (-i-1)$ and $r = (d-j)$. We conclude that $n = dq + r$ for integers q and r where $0 \leq r < d$. The proof that q and r are unique follows just as in the proof of Theorem 4.6.8. □

The greatest common divisor of two nonzero integers m and n is the largest natural number that evenly divides both m and n. We now give a precise definition.

Definition 4.6.10 (Greatest Common Divisor). For integers m and n where $m \neq 0$ or $n \neq 0$, the **greatest common divisor** of m and n, denoted by $\gcd(m,n)$, is the integer d satisfying:

(1) $d \geq 1$.
(2) $d \mid m$ and $d \mid n$.
(3) For all $c \in \mathbb{Z}$, if $c \mid m$ and $c \mid n$, then $c \mid d$.[4]

Let m and n be integers with at least one being nonzero. The next important theorem shows that there is a natural number d that satisfies conditions (1)–(3) of Definition 4.6.10. The proof of this theorem uses the well–ordering principle.

Theorem 4.6.11. *For integers m and n where $m \neq 0$ or $n \neq 0$, the greatest common divisor $d = \gcd(m,n)$ exists and can be written as $d = sm + tn$ for some integers s and t.*

Proof. See Exercise 11. □

Theorem 4.6.11 asserts that d, the greatest common divisor of m and n, can be written in the form $d = sm + tn$ for some integers s and t. This equation is frequently used in number theory and abstract algebra.

Definition 4.6.12. Let a and b be integers. We say that a and b are **relatively prime** when 1 is the only natural number that evenly divides both a and b.

In other words, the integers a and b are relatively prime if $\gcd(a,b) = 1$. For example, 6 and 35 are relatively prime. Whereas, 6 and 27 are not relatively prime because they are both divisible by 3. We can now derive the following corollary.[5]

Corollary 4.6.13. *Let m and n be integers with either $m \neq 0$ or $n \neq 0$. Then $\gcd(m,n) = 1$ if and only if $sm + tn = 1$ for some integers s and t.*

Proof. Let m and n be integers where either $m \neq 0$ or $n \neq 0$. If $\gcd(m,n) = 1$, then Theorem 4.6.11 implies that $sm + tn = 1$ for some integers s and t. To prove

[4] Consequently, $c \leq d$.

[5] A corollary is a statement that follows from a previously proven theorem.

the converse, assume $sm+tn=1$ where s and t are integers. We will prove that $\gcd(m,n)=1$. Let $d=\gcd(m,n)$. Since $d\,|\,m$ and $d\,|\,n$, Theorem 3.5.7 implies that $d\,|\,1$. Therefore, $d=1$. □

Theorem 4.6.14. *Let m,a,b be nonzero integers. If $m\,|\,(ab)$ and $\gcd(m,a)=1$, then $m\,|\,b$.*

Proof. See Exercise 12. □

Exercises 4.6

1. Let n be a natural number. Under the assumptions of Theorem 4.6.2, explicitly prove that the equality

$$0.x_nx_{n+1}x_{n+2}\cdots = 0.y_ny_{n+1}y_{n+2}\cdots$$

implies $x_n=y_n$.

2. Suppose $0.x_1x_2x_3\cdots$ and $0.y_1y_2y_3\cdots$ are two infinite decimal expansions, where $0\le x_i\le 9$ and $0\le y_i\le 9$ for each $i\in\mathbb{N}$, that satisfy the two conditions:

(a) $0.x_1x_2x_3\cdots = 0.y_1y_2y_3\cdots$
(b) $0.x_ix_{i+1}x_{i+1}\cdots < 1$ and $0.y_iy_{i+1}y_{i+2}\cdots < 1$ for all $i\in\mathbb{N}$.

Prove that $x_n=y_n$ for all $n\ge 1$. (Remark: Condition (b) is another way of saying that neither decimal expansion ends with a string of 9's (see Remark 4.6.3).)

3. Suppose $0.x_1x_2x_3\cdots$ and $0.y_1y_2y_3\cdots$ are two infinite decimal expansions, where $0\le x_i\le 9$ and $0\le y_i\le 9$ for each $i\in\mathbb{N}$, that satisfy the two conditions:

(a) $0.x_1x_2x_3\cdots = 0.y_1y_2y_3\cdots$
(b) $0.x_ix_{i+1}x_{i+1}\cdots > 0$ and $0.y_iy_{i+1}y_{i+2}\cdots > 0$ for all $i\in\mathbb{N}$.

Prove that $x_n=y_n$ for all $n\ge 1$. (Remark: Condition (b) means that neither decimal expansion ends with a string of 0's (see Remark 4.6.3).)

4. For each pair of integers n and d, find q and r such that $n=dq+r$ where $0\le r<d$.

(a) $n=335$ and $d=17$.
(b) $n=-335$ and $d=17$.
(c) $n=121$ and $d=13$.
(d) $n=-121$ and $d=13$.

5. Use the division algorithm with $d=3$ to prove that the square of every integer n has the form $3k$ or $3k+1$ for some integer k.

6. Define a sequence $a_1,a_2,\ldots$ recursively as follows:

(a) $a_1=1$

(b) $a_2 = 2$

(c) $a_n = a_{n-1} + a_{n-2}$ for all integers $n > 2$.

Use strong induction to prove that $a_n \leq \left(\frac{5}{3}\right)^n$ for all integers $n \geq 1$.

7. Consider the sequence $a_1, a_2, \ldots$ defined recursively as follows:

(a) $a_1 = 6$

(b) $a_2 = 3$

(c) $a_n = 4a_{n-1} + 5a_{n-2}$ for all integers $n > 2$.

Prove that $3 \,|\, a_n$ for all integers $n \geq 1$.

8. Let m and n be nonzero integers that are relatively prime. Suppose $m \,|\, \ell$ and $n \,|\, \ell$ for an integer ℓ. Prove that $mn \,|\, \ell$.

9. Let $s \in \mathbb{Q}$ and $d \in \mathbb{N}$ be such that $s \geq 0$ and $d \geq 1$. Prove that there exist unique numbers $q \in \mathbb{N}$ and $r \in \mathbb{Q}$ such that $s = dq + r$ and $0 \leq r < d$.

10. The following offers another proof of Theorem 4.6.9. This proof uses the well-ordering principle 4.1.2 and does not use strong induction. Let $d \geq 1$ and n be integers.

(a) Prove that there exists an integer k such that $n - dk \geq 0$.

(b) Let $A = \{m \in \mathbb{Z} : m \geq 0 \text{ and } m = n - dk \text{ for some } k \in \mathbb{Z}\}$. By (a), the set A is nonempty. Thus, A has a least element r by the well-ordering principle. Prove that $n = dq + r$ for some $q \in \mathbb{Z}$ and that $0 \leq r < d$.

11. Let $A = \{k \in \mathbb{N} : k = sm + tn \text{ for some integers } s, t\}$ where m, n are integers with $m \neq 0$ or $n \neq 0$. Exercise 12 on page 80 implies that $A \neq \varnothing$. By the well-ordering principle, A has a least element $d \geq 1$. Prove the following statements in the order given.

(1) There are integers s, t satisfying $d = sm + tn$.

(2) For all $c \in \mathbb{Z}$, if $c \,|\, m$ and $c \,|\, n$, then $c \,|\, d$.

(3) $d \,|\, m$ and $d \,|\, n$.

12. Let $m, a, b \in \mathbb{Z}$ where $\gcd(m, a) = 1$. By Corollary 4.6.13 there are integers s and t such that $1 = sa + tm$. Using this equation, prove Theorem 4.6.14.

13. Let $n, a, b \in \mathbb{Z}$ be nonzero and let $d = \gcd(a, b)$. Suppose that $d \,|\, n$. Show that there exist integers x and y such that $n = xa + yb$.

14. Prove that every integer $n \geq 0$ can be expressed as

$$n = a_i 3^i + a_{i-1} 3^{i-1} + \cdots + a_2 3^2 + a_1 3 + a_0$$

for some integers $i \geq 0$ and $0 \leq a_j \leq 2$ for $j = 0, 1, 2, \ldots, i$.

15. Define the Fibonacci sequence $f_0, f_1, \ldots$ recursively as follows:

(a) $f_0 = 0$

(b) $f_1 = 1$

(c) $f_n = f_{n-1} + f_{n-2}$ for all integers $n > 1$.

Prove that $f_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^n$ for all integers $n \geq 0$.

Exercise Notes: For Exercise 4, when $n > 0$ see Example 1; when $n < 0$ read the proof of Theorem 4.6.9. For Exercise 8, since $\gcd(m,n) = 1$, there are integers s and t such that $1 = sm + tn$. Now multiply both sides of this equation by ℓ. Fore Exercise 11(3), to prove that $d \mid m$, assume $d \nmid m$ and obtain a contradiction from the following: By Theorem 4.6.9 there are integers q and r where $qd = m - r$ with $1 \leq r < d$. Multiply both sides of the equation in (1) by q to obtain a new equation. In this new equation replace qd with $m - r$ and solve for r. For Exercise 12, multiply both sides of the equation $1 = sa + tm$ by b. For Exercise 14, use the strong induction Strategy 4.6.1. By Theorem 4.6.8 every integer $n \geq 0$ can be written as $n = 3q + r$ for integers q and $0 \leq r \leq 2$.

4.7 Fundamental Theorem of Arithmetic

In number theory, Euclid's fundamental theorem of arithmetic (or unique factorization theorem) states that every natural number greater than 1 can be written as a unique product of prime numbers. For example, $23{,}456{,}700 = 2^2 \cdot 3^3 \cdot 5^2 \cdot 67 \cdot 389$ and there is no other factorization of 23,456,700 into prime numbers. Because multiplication is commutative and associative, the order of the prime factors is usually written in ascending order, that is, from least to greatest.

We shall prove the fundamental theorem of arithmetic in two steps. First we prove Theorem 4.7.1 which states that every natural number $n \geq 2$ can be written as a product of primes. In the second step we prove that there is only one such prime factorization (see Theorem 4.7.6). Both of these proofs are by strong induction.

Theorem 4.7.1 (Existence of Prime Factorization). *Every natural number $n \geq 2$ can be expressed as a product of primes.*

Proof Analysis. We shall apply Proof Strategy 4.6.1. First we write down the statement $P(n)$. In this case we have

$$P(n) : \text{“}n \text{ is a product of primes.”}$$

Next, we construct a strong induction proof diagram with base value $b = 2$:

Base step: Prove 2 is a product of primes.
Inductive step: Let $n > 2$ be an integer.
Assume k is a product of primes, whenever $2 \leq k < n$.
Prove n is a product of primes.

This diagram will guide the composition of the following proof. Ⓐ

Proof. We prove, by strong mathematical induction, that every natural number $n \geq 2$ is a product of primes.

Base step: For $n = 2$, we see that 2 is clearly a ‘product’ of one prime.

Inductive step: Let $n > 2$ be a natural number. Assume the induction hypothesis

$$k \text{ is a product of primes, whenever } 2 \leq k < n. \tag{IH}$$

We show that n can be expressed as a product of primes. Clearly, either n is a prime or n is not a prime. Thus, there are two cases to consider:

CASE 1: n is a prime p. Since $n = p$ is a prime, n is a 'product' of one prime.

CASE 2: n is not a prime. Since n is not a prime, then n can be expressed as a product $n = ij$ where $2 \leq i < n$ and $2 \leq j < n$. By the induction hypothesis (IH), i and j can expressed as a product of primes, say $i = p_1 p_2 \cdots p_k$ and $j = q_1 q_2 \cdots q_\ell$. Therefore,

$$n = ij = (p_1 p_2 \cdots p_k)(q_1 q_2 \cdots q_\ell) = p_1 p_2 \cdots p_k q_1 q_2 \cdots q_\ell$$

and so, n can be expressed as a product of primes. This completes the proof. □

Euclid was the first to prove our next lemma which shows that if a prime number p evenly divides the product of two natural numbers ab, then either p divides a or p divides b. Euclid's lemma is used to prove the uniqueness of a prime factorization.

Lemma 4.7.2 (Euclid's Lemma). *Let a and b be natural numbers and let p be a prime. If $p \mid (ab)$, then $p \mid a$ or $p \mid b$.*

Proof. Let a and b be natural numbers and let p be a prime so that $p \mid (ab)$. Suppose $p \nmid a$. Since p is a prime and $p \nmid a$, it follows that $\gcd(a, p) = 1$. Theorem 4.6.14 now implies that $p \mid b$. □

Corollary 4.7.3. *Let a be a natural number and p be a prime. If $p \mid a^2$, then $p \mid a$.*

Our next theorem can be proven by induction on n, using Lemma 4.7.2 in the inductive step.

Theorem 4.7.4. *Let $a_1, a_2, \ldots, a_n$ be natural numbers and let p be a prime. If $p \mid (a_1 a_2 \cdots a_n)$, then $p \mid a_i$ for some i where $1 \leq i \leq n$.*

Definition 4.7.5. A prime factorization $n = p_1 p_2 \cdots p_k$, for a natural number $n > 1$, is in **ascending order** if $p_i \leq p_j$ when $1 \leq i \leq j \leq k$. Such a factorization shall be referred to as an **ascending prime factorization**.

Example 1. Here are four examples of ascending prime factorizations: $10 = 2 \cdot 5$, $20 = 2 \cdot 2 \cdot 5$, $13 = 13$, $84 = 2 \cdot 3 \cdot 3 \cdot 7$.

Theorem 4.7.6 (Uniqueness of Prime Factorization). *Let $n \geq 2$ be a natural number. Suppose $n = p_1 p_2 \cdots p_r$ and $n = q_1 q_2 \cdots q_s$ are ascending prime factorizations of n. Then $r = s$ and $p_1 = q_1, p_2 = q_2, \ldots, p_r = q_s$.*

Proof. We prove, by strong mathematical induction, that for all natural numbers $n \geq 2$, there is only one ascending prime factorization of n.

Base step: For $n = 2$, we see that $n = 2$ is the only prime factorization of n.

Inductive step: Let $n > 2$ be a natural number. Assume that

there is only one ascending prime factorization of k, whenever $2 \leq k < n$. (IH)

If n is a prime number p, then $n = p$ is the only prime factorization of n. So we will now consider the case when n is not a prime. Suppose that

$$n = p_1 p_2 \cdots p_r \text{ and } n = q_1 q_2 \cdots q_s. \tag{4.13}$$

are ascending prime factorizations of n where $r > 1$ and $s > 1$. We must prove that these factorizations are exactly the same. We first prove that $p_r = q_s$. Suppose, for a contradiction, that $p_r \neq q_s$. Thus, either $p_r < q_s$ or $p_r > q_s$. Suppose that $(\star)$ $p_r < q_s$. Since $n = q_1 q_2 \cdots q_s$, it follows that $q_s \,|\, n$. Moreover, because $n = p_1 p_2 \cdots p_r$, we have that $q_s \,|\, (p_1 p_2 \cdots p_r)$. By Theorem 4.7.4, there is an i with $1 \leq i \leq r$ such that $q_s \,|\, p_i$. Since q_s and p_i are both primes, we see that $q_s = p_i$. Since $p_i \leq p_r$, we conclude that $q_s \leq p_r$. Since the inequality $q_s \leq p_r$ contradicts $(\star)$, we infer that $p_r < q_s$ is impossible. A similar argument (see Exercise 9) shows that $p_r > q_s$ is also impossible. Therefore, $p_r = q_s$. Now since $p_r = q_s$, let $\overline{p} = p_r = q_s$ be this common value. From (4.13) we see that

$$n = p_1 p_2 \cdots p_{r-1} \overline{p} = q_1 q_2 \cdots q_{s-1} \overline{p}.$$

By canceling $\overline{p}$, we obtain

$$k = p_1 p_2 \cdots p_{r-1} = q_1 q_2 \cdots q_{s-1}$$

with $2 \leq k < n$. Since $2 \leq k < n$, our induction hypothesis (IH) implies that

$$r - 1 = s - 1 \text{ and } p_1 = q_1,\, p_2 = q_2, \ldots, p_{r-1} = q_{s-1}.$$

Thus, $r = s$ and because $p_r = q_s$, the ascending prime factorizations in (4.13) are exactly the same. This completes the proof. □

If a prime appears more than once in an ascending prime factorization, then we can simplify the factorization by using exponents, e.g., $882 = 2 \cdot 3 \cdot 3 \cdot 7 \cdot 7 = 2 \cdot 3^2 \cdot 7^2$. So, each ascending prime factorization can be written as $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ where $p_1, p_2, \ldots, p_k$ are distinct ascending primes and $a_1, a_2, \ldots, a_k$ are natural numbers. Theorems 4.7.1 and 4.7.6 easily imply our next result.

Theorem 4.7.7 (Fundamental Theorem of Arithmetic). *Let $n > 1$ be a natural number. There exist primes $p_1 < p_2 < \cdots < p_k$ and natural numbers $a_1, a_2, \ldots, a_k$ such that $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$. Furthermore, if $n = q_1^{b_1} q_2^{b_2} \cdots q_\ell^{b_\ell}$ is any ascending prime factorization into distinct primes, then $\ell = k$, $p_1 = q_1$, $p_2 = q_2$, …, $p_k = q_\ell$ and $a_1 = b_1$, $a_2 = b_2$, …, $a_k = b_\ell$.*

Example 2. If $2^i5^k = 2^45^3$ where i, j are natural numbers, then Theorem 4.7.7 implies that $i = 4$ and $k = 3$. In addition, if p_1, p_2, p_3 are distinct ascending prime numbers and a, b, c are natural numbers satisfying $p_1^a p_2^b p_3^c = 2^{10}5^7 11^4$, then $p_1 = 2$, $p_2 = 5$, $p_3 = 11$ and $a = 10$, $b = 7$, and $c = 4$.

Exercises 4.7

1. Prove Theorem 4.7.4.

2. Let a and b be integers and let p be a prime. Prove that if $p \mid a$ and $p \mid (a^2 + b^2)$, then $p \mid b$.

3. Let p be a prime. Prove that $\sqrt{p}$ is irrational.

4. Let p be a prime and let $m > 1$ be a natural number. Prove that if $p \nmid m$, then $\sqrt{mp}$ is irrational.

5. Let $a, b \in \mathbb{N}$ and p be a prime. Prove for all natural numbers n, if $p^n \mid (ab)$ and $p \nmid a$, then $p^n \mid b$.

6. Let a, b, c, d are natural numbers where a, b are relatively prime and c, d are relatively prime. Suppose $\frac{a}{b} = \frac{c}{d}$. Prove $a = c$ and $b = d$.

7. In the proof of Theorem 4.7.6 it was stated that if n is a prime number p, then $n = p$ is the only prime factorization of n. Explain why this is true.

8. The following begins another proof of Theorem 4.7.1. You are asked to complete this proof.

Proof. Suppose, for a contradiction, that there are natural numbers greater than 1 that cannot be expressed as a product of primes. By the well-ordering principle, there is a smallest such natural number. Let N be this smallest natural number. [Complete the proof!]

9. In our proof of Theorem 4.7.6 we said that "A similar argument shows that $p_r > q_s$ is also impossible." Under the assumptions used in our proof of Theorem 4.7.6, present this similar argument.

Exercise Notes: For Exercise 5, use induction on n. For Exercise 6, see Theorem 4.6.14 and Exercise 9 on page 84.

CHAPTER 5

Set Theory

In modern mathematics, many of the most important ideas are expressed in terms of sets. A *set* is a collection of objects, which can be numbers, words, other sets, functions, etc. In general, the objects that are in a set are referred to as the *elements* of the set. We are already familiar with the set of natural numbers $\mathbb{N}$ and the set of real numbers $\mathbb{R}$. In this chapter we will view sets more abstractly and investigate the operations and relations on sets that are commonly used in mathematics. Furthermore, we will learn how to prove theorems about sets.

5.1 Basic Definitions of Set Theory

An object x may or may not belong to a given set A. If x belongs to the set A, then we say that x is an *element of* A and we write $x \in A$. Otherwise, x is not an element of A and we write $x \notin A$.

Definition 5.1.1. The following set notation is used throughout mathematics.

1. For sets A and B we write $A = B$ when both sets have exactly the same elements.
2. For sets A and B we write $A \subseteq B$ to mean that the set A is a **subset** of the set B, that is, every element of A is also an element of B.
3. We say that the set A is a **proper** subset of the set B when $A \subseteq B$ and $A \neq B$, that is, when every element of A is an element of B but there is at least one element in B that is not in A.
4. We write $\varnothing$ for the empty the set, that is, the set with no elements.
5. If A is a finite set, then $|A|$ represents the number of elements in A.
6. Two sets A and B are **disjoint** if they have no elements in common.

Venn diagrams are geometric shapes that are used to depict sets and their relationships. In Fig. 5.1a we present a Venn diagram that illustrates the subset relation, a very important concept in set theory and mathematics. Figure 5.1b portrays two sets that are disjoint.

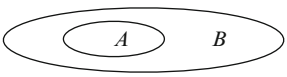


Fig. 5.1a Venn diagram of $A \subseteq B$

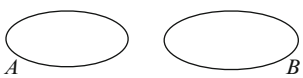


Fig. 5.1b Disjoint sets A and B

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_5, 143

Remark 5.1.2. Since $A \subseteq B$ means that every element in A is also an element of B, we can express this relationship in logical form. By taking negations, we can also represent the assertion $A \not\subseteq B$ in logical form as well. The relation $A = B$ means that an element belongs to A if and only if it belongs to B. Thus, we have the following logical forms:

1. $A \subseteq B$ iff $\forall x(x \in A \to x \in B)$; that is, for all x, if $x \in A$ then $x \in B$.
2. $A \not\subseteq B$ iff $\exists x(x \in A \wedge x \notin B)$; that is, there is an x such that $x \in A$ and $x \notin B$.
3. $A = B$ iff $\forall x(x \in A \leftrightarrow x \in B)$; that is, for all x, we have $x \in A$ if and only if $x \in B$.

It follows that $A \subseteq A$ and $\varnothing \subseteq A$, for any set A. To see why $\varnothing \subseteq A$, suppose that $\varnothing \not\subseteq A$. Then, by item 2 of Remark 5.1.2, there exists an $x \in \varnothing$ such that $x \notin A$. Because there is no x such that $x \in \varnothing$, we arrive at a contradiction. Therefore, we must have that $\varnothing \subseteq A$.

Recalling Definition 2.1.1, given a set A and a property $P(x)$ we can form the truth set $\{x \in A : P(x)\}$ which is a subset of A.

Example 1. Evaluate each of the truth sets.

1. $A = \{x \in \mathbb{N} : 3 < x < 12\}$
2. $B = \{y \in \mathbb{Z} : y^2 = 4\}$
3. $C = \{z \in \mathbb{N} : 3 \,|\, z\}$
4. $D = \{y \in \mathbb{R}^+ : 1 \leq y^2 \leq 4\}$.

Solution. $A = \{4,5,6,7,8,9,10,11\}$, $B = \{2,-2\}$, $C = \{3,6,9,12,15,\ldots\}$, and $D = [1,2]$ using interval notation. Ⓢ

Example 2. Let $A = \{1,8,27,64,\ldots\}$ and $B = \{\ldots,-12,-8,-4,0,4,8,12,\ldots\}$. Express each of these sets as a truth set.

Solution. We obtain $A = \{n \in \mathbb{N} : n = k^3 \text{ for some } k \in \mathbb{N}\}$, and $B = \{n \in \mathbb{Z} : 4 \,|\, n\}$ or $B = \{n \in \mathbb{Z} : n = 4i \text{ for some } i \in \mathbb{Z}\}$. Ⓢ

Definition 5.1.3. Let A be a set. The **power set** of A, denoted by $\mathcal{P}(A)$, is the set whose elements are all of the subsets of A. That is, $\mathcal{P}(A) = \{X : X \subseteq A\}$.

Thus, $X \in \mathcal{P}(A)$ if and only if $X \subseteq A$. One can show that if A is a finite set with n many elements, then the set $\mathcal{P}(A)$ has 2^n many elements. The set $A = \{1,2,3\}$ has three elements and thus $\mathcal{P}(A)$ has eight elements, namely,

$$\mathcal{P}(A) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

5.1.1 Set Operations

For a pair of sets A and B, there are three important and fundamental operations that we can perform on these sets: the *union*, the *intersection* and the *difference*. A fourth

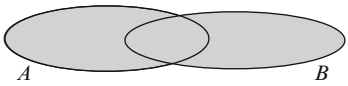


Fig. 5.2a Venn diagram of $A \cup B$

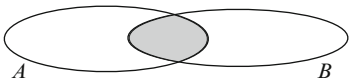


Fig. 5.2b Venn diagram of $A \cap B$

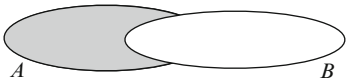


Fig. 5.2c Venn diagram of $A \setminus B$

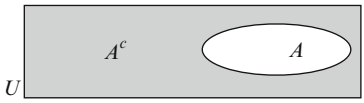


Fig. 5.2d Venn diagram of A^c

operation is called the *complement* of a set. The complement of a set A is performed when A is completely contained in a particular set U.

Definition 5.1.4. Given sets A and B, we can build new sets using the following **set operations**:

(a) $A \cup B = \{x : x \in A \text{ or } x \in B\}$ is the **union** of A and B.
(b) $A \cap B = \{x : x \in A \text{ and } x \in B\}$ is the **intersection** of A and B.
(c) $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$ is the **set difference** of A and B (also stated in English as A "minus" B).
(d) Given a universe of objects U and $A \subseteq U$, the set $A^c = U \setminus A = \{x \in U : x \notin A\}$ is called the **complement** of A.

The four set operations defined in Definition 5.1.4 are illustrated in Fig. 5.2a–d. Shading is used to identify the result of each set operation.

Remark 5.1.5. The definition of $A \cup B$ states that for an object to be an element of $A \cup B$, it must be in A or it must be in B. For an object to belong to $A \cap B$ it must be in A and it must be in B. An element in $A \setminus B$ belongs to A and does not belong to B. Thus, we can express these set operations in logical form:

(a) $x \in A \cup B$ iff $x \in A \vee x \in B$; that is, $x \in A$ or $x \in B$.
(b) $x \in A \cap B$ iff $x \in A \wedge x \in B$; that is, $x \in A$ and $x \in B$.
(c) $x \in A \setminus B$ iff $x \in A \wedge x \notin B$; that is, $x \in A$ and $x \notin B$.
(d) $x \in A^c$ iff $x \notin A$ (when $x \in U$ is understood).

When the elements of A and B are clearly presented, then one can easily evaluate the operations of union, intersection, and difference.

Example 3. Let $A = \{1,2,3,4,5,6\}$ and $B = \{2,4,6,8,10,12\}$. Then

- $A \cup B = \{1,2,3,4,5,6,8,10,12\}$.
- $A \cap B = \{2,4,6\}$.
- $A \setminus B = \{1,3,5\}$.
- $B \setminus A = \{8,10,12\}$.

- $(A \setminus B) \cup (B \setminus A) = \{1,3,5,8,10,12\}$.
- $(A \setminus B) \cap (B \setminus A) = \varnothing$.

Example 4. Recalling the notation (see page 33) for intervals on the real line, evaluate the result of the following set operations:

1. $(-3,2) \cap (1,3)$.
2. $(-3,4) \cup (0,\infty)$.
3. $(-3,2) \setminus [1,3)$.

Solution. While reading the solution to each of these items, it may be helpful to sketch the relevant intervals on the real line.

1. Since $x \in (-3,2) \cap (1,3)$ if and only if $x \in (-3,2)$ and $x \in (1,3)$, we see that x is in this intersection when x satisfies both (a) $-3 < x < 2$ and (b) $1 < x < 3$. We infer that the only values for x that satisfies both (a) and (b) are those such that $1 < x < 2$. So, $(-3,2) \cap (1,3) = (1,2)$.
2. Since $x \in (-3,4) \cup (0,\infty)$ if and only if $x \in (-3,4)$ or $x \in (0,\infty)$, we conclude that x is in this union precisely when x satisfies either (a) $-3 < x < 4$ or (b) $0 < x$. Hence, the only values for x that satisfies either (a) or (b) are those such that $-3 < x$. Consequently, $(-3,4) \cup (0,\infty) = (-3,\infty)$.
3. Since $x \in (-3,2) \setminus [1,3)$ if and only if $x \in (-3,2)$ and $x \notin [1,3)$, we observe that x is in this set difference precisely when x satisfies (a) $-3 < x < 2$ and (b) $\neg(1 \leq x < 3)$. Therefore, the only values for x that satisfies both (a) and (b) are those such that $-3 < x < 1$. Thus, $(-3,2) \setminus [1,3) = (-3,1)$. Ⓢ

In our next example, we will perform set operations on sets two C and D which will produce a pair of disjoint sets. Recall that two sets A and B are disjoint if they have no elements in common, that is, $A \cap B = \varnothing$.

Example 5. Let $C = \{1,2,3,4,5,6\}$ and $D = \{2,4,6,8,10,12\}$. Consider the sets $C \setminus D = \{1,3,5\}$, $D \setminus C = \{8,10,12\}$, and $C \cap D = \{2,4,6\}$. Since

$$(C \setminus D) \cap (D \setminus C) = \varnothing,$$

we see that $C \setminus D$ and $D \setminus C$ are disjoint sets. Because $(C \cap D) \cap (D \setminus C) = \varnothing$, the sets $C \cap D$ and $D \setminus C$ are also disjoint.

Remark 5.1.6. What does it mean to say that an object is not in $A \cup B$? This means that the object is not in A and it is not in B. Similarly, when we say that an object does not belong to $A \cap B$, then we are stating that either it is not in A or it is not in B. We can now put these observations in logical form.

(a) $x \notin A \cup B$ iff $x \notin A \wedge x \notin B$; that is, $x \notin A$ and $x \notin B$.
(b) $x \notin A \cap B$ iff $x \notin A \vee x \notin B$; that is, $x \notin A$ or $x \notin B$.
(c) $x \notin A \setminus B$ iff $x \notin A \vee x \in B$; that is, $x \notin A$ or $x \in B$.

Items (a)–(c) in Remark 5.1.6 follow from the corresponding items in Remark 5.1.5, using De Morgan's Laws, as follows:

$$x \notin A \cup B \text{ iff } \neg(x \in A \cup B) \text{ iff } \neg(x \in A \vee x \in B) \text{ iff } (x \notin A \wedge x \notin B) \quad \text{(a)}$$

$$x \notin A \cap B \text{ iff } \neg(x \in A \cap B) \text{ iff } \neg(x \in A \wedge x \in B) \text{ iff } (x \notin A \vee x \notin B) \quad \text{(b)}$$

$$x \notin A \setminus B \text{ iff } \neg(x \in A \setminus B) \text{ iff } \neg(x \in A \wedge x \notin B) \text{ iff } (x \notin A \vee x \in B). \quad \text{(c)}$$

5.1.2 Cartesian Products

Definition 5.1.7. An **ordered pair** has the form (a,b), where a is called the first component and b is called the second component.

Let (x,y) and (a,b) be ordered pairs. Then $(x,y) = (a,b)$ if and only if $x = a$ and $y = b$. For example, $(2,3)$ is an ordered pair, and so is $(3,2)$. Note that $(2,3) \neq (3,2)$.

Definition 5.1.8. Let A and B be sets. The **Cartesian product** $A \times B$ is defined to be the set

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

In other words, $A \times B$ is the set of *all* ordered pairs with first component in A and second component in B. Thus, $(a,b) \in A \times B$ if and only if $a \in A$ and $b \in B$.

Example 6. Let $A = \{1,2,c\}$ and $B = \{c,d\}$. We evaluate $A \times B$ and $B \times A$ to obtain

$$A \times B = \{(1,c), (1,d), (2,c), (2,d), (c,c), (c,d)\}$$
$$B \times A = \{(c,1), (c,2), (c,c), (d,1), (d,2), (d,c)\}.$$

Thus, $(1,d) \in A \times B$ and $(1,d) \notin B \times A$.

Thus, for sets A and B, the Cartesian product $A \times B$ is the set of all possible ordered pairs whose first component is a member of A and whose second component is a member of B. We are already familiar with the Cartesian coordinate system $\mathbb{R} \times \mathbb{R}$, or $\mathbb{R}^2$ for short, which is just the set of all ordered pairs of real numbers. In elementary algebra we were taught how to plot the points $(1,2)$ and $(-3,4)$ in this coordinate system, also called the xy-plane.

5.1.3 Partitions

A *partition* of a set is a way of breaking up the set into disjoint subsets. For example, Fig. 5.3 depicts a set A that is broken up into four disjoint subsets X,Y,U,V. We can then say that $P = \{X,Y,U,V\}$ is a partition of the set A because every element of A is in one of the sets X,Y,U,V and any two of these sets are disjoint. The following definition formalizes this notion of breaking up a set into disjoint pieces.

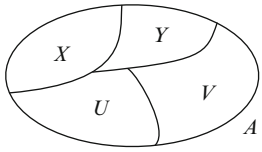


Fig. 5.3 A partition of A

Definition 5.1.9. Let A be a set. Let P be a collection of nonempty subsets of A. We say that P is a **partition** of A if the following two conditions are true:

1. For every element $a \in A$ there is a set $S \in P$ such that $a \in S$.
2. For all $S, T \in P$ if $S \neq T$, then $S \cap T = \varnothing$.

Item 1 of Definition 5.1.9 asserts that every element in A belongs to a set in the partition P. Item 2 states that any two different sets in P are disjoint. When this occurs we say that the sets in P are *pairwise disjoint.*

Example 7. Consider the subsets of $\mathbb{Z}$ defined by

$$S_0 = \{n \in \mathbb{Z} : n = 3k \text{ for some } k \in \mathbb{Z}\} = \{\ldots, -6, -3, 0, 3, 6, \ldots\}$$
$$S_1 = \{n \in \mathbb{Z} : n = 3k+1 \text{ for some } k \in \mathbb{Z}\} = \{\ldots, -5, -2, 1, 4, 7, \ldots\}$$
$$S_2 = \{n \in \mathbb{Z} : n = 3k+2 \text{ for some } k \in \mathbb{Z}\} = \{\ldots, -4, -1, 2, 5, 8, \ldots\}.$$

Then the set $P = \{S_0, S_1, S_2\}$ forms a *partition* of the set of integers $\mathbb{Z}$, as illustrated by the figure

$$\mathbb{Z} = \begin{array}{|c|c|c|} \hline \vdots & \vdots & \vdots \\ 6 & 7 & 8 \\ 3 & 4 & 5 \\ 0 & 1 & 2 \\ -3 & -2 & -1 \\ -6 & -5 & -4 \\ \vdots & \vdots & \vdots \\ \hline \uparrow & \uparrow & \uparrow \\ S_0 & S_1 & S_2 \end{array}$$

In conclusion, a partition of a set A divides the set into non-overlapping parts that cover all of A.

Exercises 5.1

1. Recalling our discussion on interval notation on page 33, evaluate the following set operations:

(a) $(-2, 0) \cap (-\infty, 2)$.

(b) $(-2,4) \cup (-\infty, 2)$.
(c) $(-\infty, 0] \setminus (-\infty, 2]$.
(d) $\mathbb{R} \setminus (2, \infty)$.
(e) $(\mathbb{R} \setminus (-\infty, 2]) \cup (1, \infty)$.

2. Express the following sets as truth sets.

(a) $A = \{1, 4, 9, 16, 25, \dots\}$
(b) $B = \{\dots, -15, -10, -5, 0, 5, 10, 15, \dots\}$.

3. Evaluate the truth sets.

(a) $A = \{x \in \mathbb{N} : 0 < x^2 < 24\}$
(b) $B = \{y \in \mathbb{Z} : y \,|\, 12\}$
(c) $C = \{z \in \mathbb{N} : 4 \,|\, z\}$
(d) $D = \{y \in \mathbb{R}^- : 1 \leq y^2 \leq 4\}$.

4. Let A, B, and C be the sets in Exercise 3. Evaluate the following sets: $A \cup B$, $A \cap C$, $A \setminus B$, $B \setminus A$, and $C \setminus (A \cup B)$.

5. Find two elements in the set $\mathbb{R} \setminus \mathbb{Q}$. Explain why $\mathbb{Q} \setminus \mathbb{R} = \varnothing$.

6. Let $A = \{2, 3\}$ and $B = \{a, b, c\}$. Evaluate $A \times A$, $A \times B$, $B \times A$, and $B \times B$.

7. Let $A = \{2, 3\}$ and $B = \{3, a\}$. Evaluate $\mathcal{P}(A \cup B)$ and $\mathcal{P}(A) \cup \mathcal{P}(B)$.

8. Find $\mathcal{P}(\varnothing)$ and $\mathcal{P}(\mathcal{P}(\varnothing))$.

9. Let $A = \{2, 3\}$, $B = \{a, b\}$ and $C = \{x, y\}$. Evaluate $(A \times B) \times C$ and $\mathcal{P}(A \times B)$.

10. Let $A = \{2, 3\}$, $B = \{3, 4\}$ and $C = \{3, y\}$. Is $A \times (B \cup C) = (A \times B) \cup (A \times C)$?

11. Let A, B, and C be sets. Determine which of the following statements are always true and which are not always true.

(a) If $x \in A$, then $x \in A \cup B$.
(b) If $x \in A \cup B$, then $x \in A$.
(c) If $x \in B$ and $A \subseteq B$, then $x \in A$.
(d) If $x \notin B$ and $A \subseteq B$, then $x \notin A$.
(e) If $x \in A$ and $A \not\subseteq B$, then $x \notin B$.
(f) If $x \in C$ and $A = C$, then $x \in A$.
(g) If $x \in A \cap B$, then $x \in A \cup B$.
(h) If $x \notin A \cap B$, then $x \notin A \cup B$.
(i) If $x \notin A \setminus B$, then $x \notin A$ or $x \in B$.
(j) If $(x, y) \in A \times B$, then $x \in A$ and $y \in B$.
(k) If $(x, y) \notin A \times B$, then $y \in A$ and $x \in B$.
(l) If $A \in \mathcal{P}(B)$, then $A \subseteq B$.

12. Let E be the set of even integers and let O be the set of odd integers. Is $\{E, O\}$ a partition of $\mathbb{Z}$? Justify your answer.

13. Find a partition $P = \{S_0, S_1, S_2, S_3\}$ of $\mathbb{Z}$, similar to the one in Example 7 on page 148, that breaks $\mathbb{Z}$ up into 4 disjoint subsets.

14. Find a partition $P = \{S_0, S_1, S_2, S_3, S_4\}$ of $\mathbb{Z}$, similar to the one in Example 7, that breaks $\mathbb{Z}$ up into 5 disjoint subsets.

15. Is the given collection P a partition of the set A? Justify your answers.

(a) $A = \{1,2,\ldots,10\}$ and $P = \{\{2,4,6,8\},\{1,3,5,7,9\}\}$.
(b) $A = \mathbb{Z}$ and $P = \{\{n \in \mathbb{Z} : n > 0\},\{n \in \mathbb{Z} : n < 0\}\}$.
(c) $A = \mathbb{Z}$ and $P = \{\{n \in \mathbb{Z} : n \geq 0\},\{n \in \mathbb{Z} : n \leq 0\}\}$.
(d) $A = \mathbb{Q}$ and $P = \{\mathbb{Q}^+,\{0\},\mathbb{Q}^-\}$.
(e) $A = \mathbb{R}$ and $P = \{[n,n+1) : n \in \mathbb{Z}\}$, where $[n,n+1)$ is the half-open interval.

5.2 Proofs in Set Theory

In this section we shall offer strategies for proving that a set is a subset of another set and for proving that two sets are equal. You will encounter such proofs in your future mathematics courses.

In the real number system there are many things that one can prove about the operations $+$, $\cdot$, $-$ and the relations $=$ and $\leq$. For example, one can prove that $a^2 + b^2 \leq (a+b)^2$ when a and b are non-negative. Similarly, there are many things in set theory that can be proven concerning the operations $\cup$, $\cap$, $\setminus$ and the relations $=$ and $\subseteq$. In particular, one can prove that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ when A, B and C are sets. In this section, and in Section 5.3, we will investigate the fundamental properties of sets that are necessary for advanced mathematics.

5.2.1 Strategy for Proving a Subset Relation

Given two sets A and B, to prove the statement $A \subseteq B$, one takes an arbitrary $x \in A$ and shows that $x \in B$. Thus, we have our first set theoretic proof strategy.

Proof Strategy 5.2.1. Given a diagram containing the form

Prove $A \subseteq B$

replace this form with

Prove $\forall x(x \in A \to x \in B)$.

In other words, use the diagram

Let $x \in A$.
  Prove $x \in B$.

The above strategy is used frequently in mathematical proofs. We will be applying this strategy in this chapter and in each of the remaining chapters of this book. In particular, it will be used to prove our next theorem.

Theorem 5.2.2. *Given any two sets A and B, we have that $(A \cup B) \setminus B \subseteq A$.*

Proof Analysis. The only assumption is "A and B are sets." We have to prove that $(A \cup B) \setminus B \subseteq A$. Now we construct proof diagrams where we apply Strategy 5.2.1 to obtain the final proof diagram:

Assume A and B are sets.
Prove $(A \cup B) \setminus B \subseteq A$.

Assume A and B are sets.
Let $x \in (A \cup B) \setminus B$.
Prove $x \in A$.

These diagrams will guide our composition of a proof of the theorem. Ⓐ

Proof. Let A and B be sets. We shall prove that $(A \cup B) \setminus B \subseteq A$. Let $x \in (A \cup B) \setminus B$. We show that $x \in A$. Since $x \in (A \cup B) \setminus B$, it follows that $x \in A \cup B$ and $x \notin B$. So $x \in A$ or $x \in B$, and $x \notin B$. Therefore, $x \in A$.[1] □

In mathematics one often assumes that a set is a subset of another set, and then wants to use this assumption to establish something new; for example, to show that another subset relation holds. We now offer a simple assumption strategy that can be used in such a proof.

Assumption Strategy 5.2.3. Given a diagram containing the form

Assume $A \subseteq B$

there are two approaches:

(a) If you are assuming or can prove $x \in A$, then you can conclude $x \in B$.
(b) If you are assuming or can prove $x \notin B$, then you can conclude $x \notin A$.

5.2.2 Strategies for Proving Set Equality

Definition 5.2.4 (Set Equality). For sets A and B we write $A = B$ when these sets have exactly the same elements.

For sets A and B, there are two alternative ways of asserting that $A = B$:

1. $A \subseteq B$ and $B \subseteq A$;
2. For all x, we have $x \in A$ if and only if $x \in B$.

Thus, there are two strategies that we can employ to prove that two sets are equal.

Proof Strategy 5.2.5. Given a diagram containing the form

Prove $A = B$

[1] We used disjunctive syllogism.

where A and B are sets, there are two approaches:

(a) Replace the form with

Prove $A \subseteq B$
Prove $B \subseteq A$.

(b) Replace this form with

Prove $\forall x(x \in A \leftrightarrow x \in B)$.

That is, use the diagram

Let x be arbitrary.
    Prove $x \in A \leftrightarrow x \in B$.

We will refer to Strategy 5.2.5(a) as the "double-subset" strategy, and we shall describe strategy (b) as the "iff" strategy. One efficient way of executing strategy (b) is to derive a string of *equivalences* starting with $x \in A$ and ending with $x \in B$. This is usually accomplished by citing appropriate logic laws (see Section 1.1.5).

To illustrate the difference between the "double-subset" strategy and the "iff" strategy, we will give two proofs of our next theorem. The first proof will employ the proof diagram (a) given in 5.2.5 and the second proof will apply proof diagram (b) also given in 5.2.5.

Theorem 5.2.6. *Suppose A, B and C are sets. Then $A \cap (B \setminus C) = (A \cap B) \setminus C$.*

Proof Analysis. We must prove that $A \cap (B \setminus C) = (A \cap B) \setminus C$. To see how we can prove this set equality, we shall first construct proof diagrams using the double-subset Strategy 5.2.5(a):

Assume A, B and C are sets.
    Prove $A \cap (B \setminus C) = (A \cap B) \setminus C$.

Assume A, B and C are sets.
    Prove $A \cap (B \setminus C) \subseteq (A \cap B) \setminus C$
    Prove $(A \cap B) \setminus C \subseteq A \cap (B \setminus C)$.

These proof diagrams will guide our first proof of Theorem 5.2.6. Ⓐ

First Proof. Let A, B and C be sets. We prove that $A \cap (B \setminus C) = (A \cap B) \setminus C$.

($\subseteq$). Let $x \in A \cap (B \setminus C)$. We shall prove that $x \in (A \cap B) \setminus C$. Since $x \in A \cap (B \setminus C)$, it follows that $x \in A$ and $x \in B \setminus C$. Thus, $x \in A$ and, $x \in B$ and $x \notin C$. Because $x \in A$ and $x \in B$, we have that $x \in A \cap B$. Since we also have that $x \notin C$, we conclude that $x \in (A \cap B) \setminus C$.

($\supseteq$). Let $x \in (A \cap B) \setminus C$. We will prove that $x \in A \cap (B \setminus C)$. Since $x \in (A \cap B) \setminus C$, we see that $x \in A \cap B$ and $x \notin C$. Because $x \in A \cap B$, we have that $x \in A$ and $x \in B$. Since we also have that $x \notin C$, we see that $x \in B \setminus C$. Furthermore, we know that $x \in A$. Hence, $x \in A \cap (B \setminus C)$.

Therefore, $A \cap (B \setminus C) = (A \cap B) \setminus C$. □

In our first proof of Theorem 5.2.6, the annotations ($\subseteq$) and ($\supseteq$) are added as a courtesy to the reader. The notation ($\subseteq$) is used to make it clear to the reader that we are proving that the first set[2] is a subset of the second set. The notation ($\supseteq$) indicates that we are proving that the second set is a subset of the first set.

We shall now reprove Theorem 5.2.6 using the iff strategy.

Proof Analysis. The only assumption is "A, B and C are sets." We must prove that $A \cap (B \setminus C) = (A \cap B) \setminus C$. To see how to prove this set equality, we construct proof diagrams using Strategy 5.2.5(b) as follows:

Assume A, B and C are sets.
  Prove $A \cap (B \setminus C) = (A \cap B) \setminus C$.

Assume A, B and C are sets.
  Let x be arbitrary.
    Prove $x \in A \cap (B \setminus C) \leftrightarrow x \in (A \cap B) \setminus C$.

These proof diagrams will be used to direct our second proof of Theorem 5.2.6. Ⓐ

Second Proof. Suppose that A, B and C are sets. Let x be arbitrary. We prove that $x \in A \cap (B \setminus C)$ iff $x \in (A \cap B) \setminus C$, as follows:

$x \in A \cap (B \setminus C)$ iff	$x \in A \wedge x \in (B \setminus C)$	by the definition of $\cap$
iff	$x \in A \wedge (x \in B \wedge x \notin C)$	by the definition of $\setminus$
iff	$(x \in A \wedge x \in B) \wedge x \notin C$	by logical associativity
iff	$x \in A \cap B \wedge x \notin C$	by the definition of $\cap$
iff	$x \in (A \cap B) \setminus C$	by the definition of $\setminus$.

Therefore, $A \cap (B \setminus C) = (A \cap B) \setminus C$. □

In our second proof of Theorem 5.2.6, we translated the statement

$$x \in A \cap (B \setminus C) \tag{5.1}$$

into another statement that involved logical connectives and no set operations. This was done, in steps, by using parentheses to identify the order in which the set operations are to be performed; namely, one must translate from the "outside-in." For example, $\cap$ is the "outer most" set operation in (5.1). Thus, we first translated $\cap$. After identifying the "next outer most" operation to be $\setminus$, we translated it to obtain the statement

$$x \in A \wedge (x \in B \wedge x \notin C)$$

[2]When proving $X = Y$ for sets X and Y, we shall say that X is the first set and Y is the second set.

which involves logical connectives and no set operations. We then applied a relevant logic law to obtain the logically equivalent statement

$$(x \in A \wedge x \in B) \wedge x \notin C. \tag{5.2}$$

Afterwards, we translated the expression (5.2) into a statement involving only set operations. One must perform this translation in a particular order, as well; namely, one must translate from the "inside-out." We shall refer to the logical connective $\wedge$ appearing within the parentheses in (5.2) as an "inner most" logical connective. We shall refer to the other occurrence of $\wedge$ in (5.2) as the "next inner most" logical connective. In the second proof of Theorem 5.2.6, we translated (5.2) by first translating the "inner most" logical connective $\wedge$ into the set operation $\cap$. This was followed by translating the "next inner most" logical connective into a set operation. We were able to conclude that (5.2) is equivalent to the assertion $x \in (A \cap B) \setminus C$, which involves only set operations. We also follow this procedure in the proof of our next theorem.

Theorem 5.2.7 (Distributive Laws). *Suppose A, B and C are sets. Then*

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

Proof. To prove (1), let x be arbitrary. We prove the equivalence $x \in A \cup (B \cap C)$ if and only if $x \in (A \cup B) \cap (A \cup C)$, as follows:

$x \in A \cup (B \cap C)$ iff	$x \in A \vee x \in (B \cap C)$	by the definition of $\cup$
iff	$x \in A \vee (x \in B \wedge x \in C)$	by the definition of $\cap$
iff	$(x \in A \vee x \in B) \wedge (x \in A \vee x \in C)$	by logical distributivity
iff	$x \in A \cup B \wedge x \in A \cup C$	by the definition of $\cup$
iff	$x \in (A \cup B) \cap (A \cup C)$	by the definition of $\cap$.

Therefore, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. The proof of (2) is left as an exercise. □

Theorem 5.2.8 (Associative Laws). *Suppose A, B and C are sets. Then*

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

Proof. See Exercise 6. □

Assumption Strategy 5.2.9. When *assuming* $A = B$, if you are also assuming or can prove $x \in A$, then you can deduce that $x \in B$. If you are assuming or can prove $x \notin A$, then you can conclude $x \notin B$.

We will apply Assumption Strategy 5.2.9 in the proof of our next theorem, where we shall assume the set equality $A \cap B = A$.

Theorem 5.2.10. *Suppose A and B are sets. If $A \cap B = A$, then $A \subseteq B$.*

Proof. Let A and B be sets and assume that $A \cap B = A$. To prove that $A \subseteq B$, let $x \in A$. We shall show that $x \in B$. Since $x \in A$ and $A = A \cap B$, it follows that $x \in A \cap B$. That is, $x \in A$ and $x \in B$. Therefore, $x \in B$ and the proof is complete. □

Theorem 5.2.11. *Let A, B, and C be sets. Then*

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

In our proof of Theorem 5.2.11, we shall use the double-subset strategy to prove item (1) and then the iff strategy to prove (2).

Proof. Let A, B, and C be sets. We first prove that $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

($\subseteq$). Let $(x,y) \in A \times (B \cup C)$. So $x \in A$ and $y \in B \cup C$. Thus, $x \in A$ and either $y \in B$ or $y \in C$. If $y \in B$, then $(x,y) \in A \times B$. If $y \in C$, then $(x,y) \in A \times C$. In each case we can conclude that $(x,y) \in (A \times B) \cup (A \times C)$.

($\supseteq$). Let $(x,y) \in (A \times B) \cup (A \times C)$. Thus, either $(x,y) \in (A \times B)$ or $(x,y) \in (A \times C)$. If $(x,y) \in A \times B$, then $x \in A$ and $y \in B$. Hence $x \in A$ and $y \in B \cup C$. Consequently, $(x,y) \in A \times (B \cup C)$. If $(x,y) \in A \times C$, then $x \in A$ and $y \in C$. So $x \in A$ and $y \in B \cup C$, and thus $(x,y) \in A \times (B \cup C)$. Therefore, in either case $(x,y) \in A \times (B \cup C)$.

Now we prove that $A \times (B \cap C) = (A \times B) \cap (A \times C)$. Let x and y be arbitrary. We show that $(x,y) \in A \times (B \cap C)$ if and only if $(x,y) \in (A \times B) \cap (A \times C)$, as follows:

$(x,y) \in A \times (B \cap C)$	iff $x \in A \wedge y \in (B \cap C)$	by the definition of $\times$
	iff $x \in A \wedge (y \in B \wedge y \in C)$	by the definition of $\cap$
	iff $x \in A \wedge y \in B \wedge y \in C$	by logical associativity
	iff $x \in A \wedge x \in A \wedge y \in B \wedge y \in C$	by the idempotent law
	iff $(x \in A \wedge y \in B) \wedge (x \in A \wedge y \in C)$	by logical commutativity
	iff $(x,y) \in A \times B \wedge (x,y) \in A \times C$	by the definition of $\times$
	iff $(x,y) \in (A \times B) \cap (A \times C)$	by the definition of $\cap$.

This completes the proof. □

Theorem 5.2.12. *Let A and B be sets. Then $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cap \mathcal{P}(B)$.*

Proof. Let A and B be sets. We shall prove that $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cap \mathcal{P}(B)$.

($\subseteq$). Let $X \in \mathcal{P}(A \cap B)$. So, $X \subseteq A \cap B$. Thus, $X \subseteq A$ and $X \subseteq B$ (see Exercise 3). Hence, $X \in \mathcal{P}(A)$ and $X \in \mathcal{P}(B)$. Therefore, $X \in \mathcal{P}(A) \cap \mathcal{P}(B)$.

($\supseteq$). Let $X \in \mathcal{P}(A) \cap \mathcal{P}(B)$. Thus, $X \in \mathcal{P}(A)$ and $X \in \mathcal{P}(B)$. So, $X \subseteq A$ and $X \subseteq B$. Therefore, $X \subseteq A \cap B$ (see Exercise 3) and we conclude that $X \in \mathcal{P}(A \cap B)$. □

Theorem 5.2.12 motivates the following question: Can one prove the equality $\mathcal{P}(A \cup B) = \mathcal{P}(A) \cup \mathcal{P}(B)$ for any two sets A and B? The answer is no. Let $A = \{1,2\}$

and $B = \{2,3\}$. Clearly, the set $X = \{1,3\}$ is subset of $A \cup B$ and thus, $X \in \mathcal{P}(A \cup B)$. Since X is not a subset A and is also not a subset of B, we see that $X \notin \mathcal{P}(A) \cup \mathcal{P}(B)$. So $X \in \mathcal{P}(A \cup B)$ and $X \notin \mathcal{P}(A) \cup \mathcal{P}(B)$. Therefore, $\mathcal{P}(A \cup B) \neq \mathcal{P}(A) \cup \mathcal{P}(B)$.

Exercises 5.2

Prove the following theorems, where A, B, C, and D are sets.

1. **Theorem.** If $A \subseteq B$, then $A \subseteq A \cup B$ and $A \cap B \subseteq A$.
2. **Theorem.** If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.
3. **Theorem.** $C \subseteq A$ and $C \subseteq B$ if and only if $C \subseteq A \cap B$.
4. **Theorem.** $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
5. **Theorem.** $(A \setminus B) \cap (C \setminus B) = (A \cap C) \setminus B$.
6. **Theorem.** $A \cap (B \cap C) = (A \cap B) \cap C$ and $A \cup (B \cup C) = (A \cup B) \cup C$.
7. **Theorem.** $(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$.
8. **Theorem.** If $A \setminus B \subseteq C$, then $A \setminus C \subseteq B$.
9. **Theorem.** If $A \subseteq B$ and $B \cap C = \varnothing$, then $A \subseteq B \setminus C$.
10. **Theorem.** If $A \setminus B \subseteq C$ and $A \not\subseteq C$, then $A \cap B \neq \varnothing$.
11. **Theorem.** $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$.
12. **Theorem.** $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cup D)$.
13. **Theorem.** $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.
14. **Theorem.** $A \subseteq B$ if and only if $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.
15. **Theorem.** $\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$.

Exercise Notes: For Exercises 4–6: Use Proof Strategy 5.2.5(b) and review the propositional logic laws in Section 1.1.5. For Exercise 7, one may want to use Proof Strategy 5.2.5(a). For Exercise 8, to prove that $x \in B$, use proof by contradiction. For Exercise 10, review Remark 5.1.2(2).

5.3 Indexed Families of Sets

Given a property $P(x)$ we can form the truth set $\{x : P(x)\}$ when the universe is understood. There is another way to build sets. For example, consider the set S of all perfect squares, that is, the set of all numbers of the form n^2 for some natural number n. We can define S in two ways:

1. $S = \{x : (\exists n \in \mathbb{N})\,(x = n^2)\} = \{1,4,9,16,25,\ldots\}$.
2. $S = \{n^2 : n \in \mathbb{N}\} = \{1,4,9,16,25,\ldots\}$.

In item 1 we have expressed S as a truth set. Item 2 offers an alternative method for constructing the same set S. For each $n \in \mathbb{N}$ we obtain the new number n^2. So S is just the set of all these new numbers. This alternative method is a special case of the following technique used to build a set from the set $\mathbb{N}$ of natural numbers. Suppose for each $n \in \mathbb{N}$ we have some object a_n. Then we can form the set $A = \{a_n : n \in \mathbb{N}\}$ of all such objects. In this case, the set $\mathbb{N}$ is called the index set and the set A is called an indexed set. Since this concept is used so often in mathematics, we will now formulate this idea in terms of a general definition.

Definition 5.3.1. Let I be *any* nonempty set and for each $i \in I$, let x_i be some *object.* Then we can form the set $S = \{x_i : i \in I\}$. The set I is called the **index set** and the set S is called an **indexed set**.

Consider an indexed set $S = \{x_i : i \in I\}$. Since $S = \{x : (\exists i \in I)(x = x_i)\}$, we see that S can also be defined as a truth set.

Example 1. Explain what the following statements mean.

1. $y \in \{\sin(x) : x \in \mathbb{Q}\}$.
2. $\{x_i : i \in I\} \subseteq A$.
3. $\{x_i : i \in I\} \not\subseteq A$.

Solution. The first statement $y \in \{\sin(x) : x \in \mathbb{Q}\}$ means that $y = \sin(x)$ for some $x \in \mathbb{Q}$. The second statement $\{x_i : i \in I\} \subseteq A$ means that $x_i \in A$ for every $i \in I$. Finally, the third statement $\{x_i : i \in I\} \not\subseteq A$ means that $x_i \notin A$ for some $i \in I$. Ⓢ

Definition 5.3.2. A set $\mathcal{F}$, whose elements are sets, is called a **family of sets**.

Definition 5.3.3. Let I be any nonempty set and for each $i \in I$, let C_i be a set. Then we can form the set $\mathcal{F} = \{C_i : i \in I\}$. The set I is called the **index set** and $\mathcal{F}$ is called an **indexed family of sets**.

Example 2. Let $A_n = \{1, 2, \ldots, n\}$ for each natural number n. Thus, can construct the indexed family of sets $\mathcal{F} = \{A_n : n \in \mathbb{N}\} = \{A_1, A_2, A_3, \ldots\}$ where the set $\mathbb{N}$ of natural numbers is the index set.

Example 3. For each real number $x > 0$, let $B_x = \{y \in \mathbb{R} : -x < y < x + 1\}$, that is, $B_x = (-x, x+1)$. Define the indexed family of sets by $\mathcal{F} = \{B_x : x \in \mathbb{R}^+\}$, where $\mathbb{R}^+$ is the index set. Note that $B_2 \cap B_{\frac{5}{2}} = (-2, 3) \cap (-\frac{5}{2}, \frac{7}{2}) = (-2, 3)$.

Example 4. Consider the index set $I = \{i \in \mathbb{R} : i > 1\}$. For each real number $i \in I$, let $B_i = [-i, \frac{1}{i}]$, that is, $B_i = \{x \in \mathbb{R} : -i \leq x \leq \frac{1}{i}\}$. Define the indexed family of sets by $\mathcal{F} = \{B_i : i \in I\}$. Note that $B_2 \cap B_{\frac{5}{2}} = [-2, \frac{1}{2}] \cap [-\frac{5}{2}, \frac{2}{5}] = [-2, \frac{2}{5}]$.

5.3.1 Generalized Unions and Intersections

Given two sets A and B we can form the union $A \cup B$ and the intersection $A \cap B$ of these sets. In mathematics we often need to form the union and intersection of

many more than just two sets. To see how this is done, we need to generalize the operations of union and intersection so that they will apply to more than two sets. We do this by first extending these operations to a finite number of sets, and then to an infinite number of sets.

We know that $x \in A \cup B$ means that x is in at least one of the two sets A and B. We can generalize this notion of union to any number of sets. For finitely many sets, say $A_1, A_2, \ldots, A_n$, we will say that x is in the union

$$A_1 \cup A_2 \cup \cdots \cup A_n$$

when x is in *at least one of the sets* $A_1, A_2, \ldots, A_n$; that is, $x \in A_i$ for *some* natural number i between 1 and n. There is a more compact way to denote this union. Using $I = \{1, 2, \ldots, n\}$ as an index set, we define $\bigcup_{i \in I} A_i$ to be

$$\bigcup_{i \in I} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$

and so, $x \in \bigcup_{i \in I} A_i$ means that $x \in A_i$ for some $i \in I$. For example, let $I = \{1, 2, 3, 4\}$ and let $A_1 = \{0, 2, 4, 11\}$, $A_2 = \{0, 1, 4, 9\}$, $A_3 = \{0, 3, 4, 7, 10\}$, $A_4 = \{6, 5, 8, 10, 11\}$. Then

$$\bigcup_{i \in I} A_i = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}.$$

Recall that $x \in A \cap B$ means that x is in both of the sets A and B. We can also generalize the intersection operation to more than two sets. For finitely many sets, say $A_1, A_2, \ldots, A_n$, we shall say that x is in the intersection

$$A_1 \cap A_2 \cap \cdots \cap A_n$$

when x is in *every one of the sets* $A_1, A_2, \ldots, A_n$; that is, $x \in A_i$ for *every* natural number i between 1 and n. There is easier way to express this intersection. Using $I = \{1, 2, \ldots, n\}$ as an index set, we define $\bigcap_{i \in I} A_i$ to be

$$\bigcap_{i \in I} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$$

and so, $x \in \bigcap_{i \in I} A_i$ means that $x \in A_i$ for every $i \in I$. For example, let $I = \{1, 2, 3, 4\}$ and let $A_1 = \{0, 2, 4, 9\}$, $A_2 = \{0, 1, 4, 9\}$, $A_3 = \{0, 3, 4, 7, 9\}$, $A_4 = \{0, 1, 4, 5, 8, 9\}$. Then

$$\bigcap_{i \in I} A_i = \{0, 4, 9\}.$$

Similarly, we can form the union and intersection of an indexed family of sets $\{C_i : i \in I\}$, where I can be a finite or an infinite set. We now generalize the union and intersection operations to any indexed collection of sets.

Definition 5.3.4. Let $\{C_i : i \in I\}$ be an indexed family of sets. The **union** $\bigcup_{i\in I} C_i$ is the set of elements x such that $x \in C_i$ for at least one $i \in I$; that is,

$$\bigcup_{i\in I} C_i = \{x : x \in C_i \text{ for some } i \in I\}.$$

The union $\bigcup_{i\in I} C_i$ contains just those elements that are in at least one of the sets C_i. In other words, $\bigcup_{i\in I} C_i$ is the set obtained by putting together the elements that belong to any one of the C_i's.

Definition 5.3.5. Let $\{C_i : i \in I\}$ be an indexed family of sets. The **intersection** $\bigcap_{i\in I} C_i$ is the set of elements x such that $x \in C_i$ for all $i \in I$; that is,

$$\bigcap_{i\in I} C_i = \{x : x \in C_i \text{ for every } i \in I\}.$$

The intersection $\bigcap_{i\in I} C_i$ consists of those elements that belong to each and every one of the sets C_i. Thus, $\bigcap_{i\in I} C_i$ is the result of collecting the elements that are common to all of the C_i's and then putting them together to form a set.

Example 5. For each $n \in \mathbb{N}$, let C_n be the closed interval $C_n = \left[1, 1+\frac{1}{n}\right]$. Then $\{C_n : n \in \mathbb{N}\}$ is an indexed family of sets. Evaluate the sets $\bigcup_{n\in\mathbb{N}} C_n$ and $\bigcap_{n\in\mathbb{N}} C_n$.

Solution. First observe that $C_1 = [1,2]$, $C_2 = \left[1, \frac{3}{2}\right]$, $C_3 = \left[1, \frac{4}{3}\right]$, and $C_4 = \left[1, \frac{5}{4}\right]$. We evaluate the union $\bigcup_{n\in\mathbb{N}} C_n$ as follows:

$$\begin{aligned}
x \in \bigcup_{n\in\mathbb{N}} C_n \text{ iff } & x \in C_n \text{ for some } n \in \mathbb{N} && \text{by def. of } \bigcup \\
\text{iff } & x \in \left[1, 1+\frac{1}{n}\right] \text{ for some } n \in \mathbb{N} && \text{by def. of } C_n \\
\text{iff } & 1 \le x \le 1+\frac{1}{n} \text{ for some } n \in \mathbb{N} && \text{by interval notation} \\
\text{iff } & 1 \le x \le 2 && \text{by Exercise 3.}
\end{aligned}$$

Hence, $\bigcup_{n\in\mathbb{N}} C_n = [1,2]$. We now evaluate the intersection $\bigcap_{n\in\mathbb{N}} C_n$:

$$\begin{aligned}
x \in \bigcap_{n\in\mathbb{N}} C_n \text{ iff } & x \in C_n \text{ for every } n \in \mathbb{N} && \text{by def. of } \bigcap \\
\text{iff } & x \in \left[1, 1+\frac{1}{n}\right] \text{ for every } n \in \mathbb{N} && \text{by def. of } C_n
\end{aligned}$$

$$\text{iff } 1 \le x \le 1 + \frac{1}{n} \text{ for every } n \in \mathbb{N} \qquad \text{by interval notation}$$

$$\text{iff } x = 1 \qquad \text{by Exercise 4.}$$

Thus, $\bigcap_{n\in\mathbb{N}} C_n = \{1\}$. Since $1 + \frac{1}{n}$ gets closer and closer to 1 as n gets larger and larger, it follows that 1 is the only number that is in every C_n. Ⓢ

We now know what it means to say that an object belongs to an indexed union or to an indexed intersection of sets. What does it mean to say that an object does *not* belong to an indexed union or intersection of sets?

Example 6. Suppose that $\{C_i : i \in I\}$ is an indexed family of sets. Explain why the following statements are true.

(1) $x \in \bigcup_{i\in I} C_i$ means that $x \in C_i$ for some $i \in I$.
(2) $x \notin \bigcup_{i\in I} C_i$ means that $x \notin C_i$ for every $i \in I$.
(3) $x \in \bigcap_{i\in I} C_i$ means that $x \in C_i$ for every $i \in I$.
(4) $x \notin \bigcap_{i\in I} C_i$ means that $x \notin C_i$ for some $i \in I$.

Solution. We first note that the assertion $x \notin \bigcup_{i\in I} C_i$ in (2) is the negation of statement in (1). Similarly, the statement $x \notin \bigcap_{i\in I} C_i$ in (4) is the negation of that in (3).

(1) Clearly, $x \in \bigcup_{i\in I} C_i$ means $x \in C_i$ for some $i \in I$, by Definition 5.3.4. We conclude that $x \in \bigcup_{i\in I} C_i$ iff $(\exists i \in I)(x \in C_i)$.

(2) In our solution to (1), we observed that $x \in \bigcup_{i\in I} C_i$ iff $(\exists i \in I)(x \in C_i)$. Thus, using a bounded quantifier negation law, we obtain

$$x \notin \bigcup_{i\in I} C_i \text{ iff } \neg(\exists i \in I)(x \in C_i) \text{ iff } (\forall i \in I)(x \notin C_i).$$

So, $x \notin \bigcup_{i\in I} C_i$ means $(\forall i \in I)(x \notin C_i)$, that is, $x \notin C_i$ for every $i \in I$.

(3) From Definition 5.3.5, we see that $x \in \bigcap_{i\in I} C_i$ means $x \in C_i$ for every $i \in I$. Consequently, $x \in \bigcap_{i\in I} C_i$ iff $(\forall i \in I)(x \in C_i)$.

(4) In our solution to (3) we noted that $x \in \bigcap_{i\in I} C_i$ iff $(\forall i \in I)(x \in C_i)$. Hence, using a bounded quantifier negation law, we obtain

$$x \notin \bigcap_{i\in I} C_i \text{ iff } \neg(\forall i \in I)(x \in C_i) \text{ iff } (\exists i \in I)(x \notin C_i).$$

So, $x \notin \bigcap_{i\in I} C_i$ means $(\exists i \in I)(x \notin C_i)$, that is, $x \notin C_i$ for some $i \in I$. Ⓢ

Example 6 shows that the definitions of indexed unions and intersections, and their negations, can be expressed in terms of quantifiers over the index set. One is advised to have a clear understanding of the four items presented in this example.

De Morgan's Laws for Families of Sets

We will apply the "double-subset" Proof Strategy 5.2.5(a) in the proof of our next theorem. Before reading this proof, one should review items (1)–(4) of Example 6.

Theorem 5.3.6. *Suppose that A is a set and that $\{B_i : i \in I\}$ is an indexed family of sets. Then* (1) $A \setminus \bigcup_{i\in I} B_i = \bigcap_{i\in I}(A \setminus B_i)$ *and* (2) $A \setminus \bigcap_{i\in I} B_i = \bigcup_{i\in I}(A \setminus B_i)$.

Proof. We shall prove that $A \setminus \bigcup_{i\in I} B_i = \bigcap_{i\in I}(A \setminus B_i)$ and leave (2) as an exercise.

($\subseteq$). Let $x \in A \setminus \bigcup_{i\in I} B_i$. We prove that $x \in \bigcap_{i\in I}(A \setminus B_i)$ as follows[3]:

$$
\begin{aligned}
x \in A \setminus \bigcup_{i\in I} B_i &\Rightarrow x \in A \text{ and } x \notin \bigcup_{i\in I} B_i && \text{by the definition of } \setminus \\
&\Rightarrow x \in A \text{ and } x \notin B_i \text{ for every } i \in I && \text{by the definition of } \bigcup \\
&\Rightarrow x \in A \setminus B_i \text{ for every } i \in I && \text{by the definition of } \setminus \\
&\Rightarrow x \in \bigcap_{i\in I}(A \setminus B_i) && \text{by the definition of } \bigcap.
\end{aligned}
$$

Therefore, $A \setminus \bigcup_{i\in I} B_i \subseteq \bigcap_{i\in I}(A \setminus B_i)$.

($\supseteq$). Let $x \in \bigcap_{i\in I}(A \setminus B_i)$. We prove that $x \in A \setminus \bigcup_{i\in I} B_i$ as follows:

$$
\begin{aligned}
x \in \bigcap_{i\in I}(A \setminus B_i) &\Rightarrow x \in A \setminus B_i \text{ for every } i \in I && \text{by the definition of } \bigcap \\
&\Rightarrow x \in A \text{ and } x \notin B_i \text{ for every } i \in I && \text{by the definition of } \setminus \\
&\Rightarrow x \in A \text{ and } x \notin \bigcup_{i\in I} B_i && \text{by the definition of } \bigcup \\
&\Rightarrow x \in A \setminus \bigcup_{i\in I} B_i && \text{by the definition of } \setminus.
\end{aligned}
$$

Therefore, $\bigcap_{i\in I}(A \setminus B_i) \subseteq A \setminus \bigcup_{i\in I} B_i$. The proof of (1) is complete. □

[3]The arrow $\Rightarrow$ will be used to abbreviate the word "implies."

5.3.2 Unindexed Families of Sets

Indexed families of sets occur frequently in mathematics. Moreover, mathematicians also deal with families of sets that are not described as an indexed set. Fortunately, by a simple change in notation, every family of sets can be expressed as an indexed set. Let $\mathcal{F}$ be a family of sets. Then $\mathcal{F} = \{C_A : A \in \mathcal{F}\}$ where $\mathcal{F}$ is the index set and $C_A = A$ for each $A \in \mathcal{F}$.

Since every family $\mathcal{F}$ of sets can be expressed as an indexed family of sets, it follows that all of the operations and theorems we presented in Section 5.3.1 also apply to families of sets. When $\mathcal{F}$ is a family of sets, the **union** $\bigcup\mathcal{F}$ is the set of elements x such that $x \in C$ for some $C \in \mathcal{F}$; that is,

$$\bigcup\mathcal{F} = \{x : x \in C \text{ for some } C \in \mathcal{F}\}.$$

The **intersection** $\bigcap\mathcal{F}$ is the set of elements x such that $x \in C$ for all $C \in \mathcal{F}$; that is,

$$\bigcap\mathcal{F} = \{x : x \in C \text{ for every } C \in \mathcal{F}\}.$$

For example, let $\mathcal{F}$ be the family of sets defined by $\mathcal{F} = \{\{1,2,9\},\{2,9\},\{4,9\}\}$. Then $\bigcup\mathcal{F} = \{1,2,4,9\}$ and $\bigcap\mathcal{F} = \{9\}$. Furthermore, letting $A = \{1,4,9,10,11\}$, we can construct the following new family of sets

$$\{A \setminus B : B \in \mathcal{F}\} = \{\{4,10,11\},\{1,4,10,11\},\{1,10,11\}\}.$$

We have the following "unindexed" version of De Morgan's Theorem 5.3.6.

Theorem 5.3.7. *Suppose that A is a set and that $\mathcal{F}$ is a family of sets. Then*

(1) $A \setminus \bigcup\mathcal{F} = \bigcap\{A \setminus B : B \in \mathcal{F}\}$,
(2) $A \setminus \bigcap\mathcal{F} = \bigcup\{A \setminus B : B \in \mathcal{F}\}$.

Exercises 5.3

1. Let $I = \{2,3,4,5\}$ and for each $i \in I$, let $C_i = \{i, i+1, i-1, 2i\}$. For each $i \in I$, explicitly list the elements of C_i. Then find $\bigcap\limits_{i\in I} C_i$ and $\bigcup\limits_{i\in I} C_i$.

2. For each $n \in \mathbb{N}$, let $A_n = \{0,1,2,\ldots,n,n+1\}$. Evaluate $\bigcup\limits_{n\in\mathbb{N}} A_n$ and $\bigcap\limits_{n\in\mathbb{N}} A_n$.

3. Let $x \in \mathbb{R}$. Prove that $1 \leq x \leq 2$ if and only if $1 \leq x \leq 1 + \frac{1}{n}$ for some $n \in \mathbb{N}$.

4. Let $x \in \mathbb{R}$. Prove that $x = 1$ if and only if $1 \leq x \leq 1 + \frac{1}{n}$ for all $n \in \mathbb{N}$.

5. For each $n \in \mathbb{N}$, let O_n be the open interval $O_n = \left(1, 1 + \frac{1}{n}\right)$. Then $\{O_n : n \in \mathbb{N}\}$ is an indexed family of sets. Evaluate $\bigcap\limits_{n\in\mathbb{N}} O_n$ and $\bigcup\limits_{n\in\mathbb{N}} O_n$.

6. Let $I = \{i \in \mathbb{R} : 1 \leq i\} = [1, \infty)$ and let $A_i = \left\{x \in \mathbb{R} : -\frac{1}{i} \leq x \leq 2 - \frac{1}{i}\right\}$, for each $i \in I$. Express $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$ in interval notation, if possible.

7. Let p and q be prime numbers. Define $A = \{p^i : i \in \mathbb{N}\}$ and $B = \{q^i : i \in \mathbb{N}\}$. Prove that if $A \cap B \neq \varnothing$, then $A = B$.

8. In our proof of Theorem 5.3.6(1) we applied the "double-subset" Proof Strategy 5.2.5(a). Reprove Theorem 5.3.6(1) using the "iff" Strategy 5.2.5(b).

9. Prove the following theorems:

(a) **Theorem.** Let $\{A_i : i \in I\}$ and $\{B_i : i \in I\}$ be indexed families of sets with the same indexed set I. Suppose $A_i \subseteq B_i$ for all $i \in I$. Then $\bigcup_{i \in I} A_i \subseteq \bigcup_{i \in I} B_i$.

(b) **Theorem.** Let $\{A_i : i \in I\}$ and $\{B_i : i \in I\}$ be indexed families of sets with the same indexed set I. Suppose $A_i \subseteq B_i$ for all $i \in I$. Then $\bigcap_{i \in I} A_i \subseteq \bigcap_{i \in I} B_i$.

(c) **Theorem.** Let $\{A_i : i \in I\}$ and $\{B_j : j \in J\}$ be indexed families of sets. If there is an $i_0 \in I$ such that $A_{i_0} \subseteq B_j$ for all $j \in J$, then $\bigcap_{i \in I} A_i \subseteq \bigcap_{j \in J} B_j$.

(d) **Theorem.** Suppose that A is a set and that $\{B_i : i \in I\}$ is an indexed family of sets. Then $A \cap \bigcup_{i \in I} B_i = \bigcup_{i \in I} (A \cap B_i)$.

(e) **Theorem.** Suppose that A is a set and that $\{B_i : i \in I\}$ is an indexed family of sets. Then $A \cup \bigcap_{i \in I} B_i = \bigcap_{i \in I} (A \cup B_i)$.

(f) **Theorem.** Suppose that A is a set and that $\{B_i : i \in I\}$ is an indexed family of sets. Then $A \setminus \bigcap_{i \in I} B_i = \bigcup_{i \in I} (A \setminus B_i)$.

10. Let $\{B_x : x \in \mathbb{R}^+\}$ be the family of sets in Example 3 on page 157. Evaluate $\bigcap_{x \in \mathbb{R}^+} B_x$ and $\bigcup_{x \in \mathbb{R}^+} B_x$.

11. Let $\{B_i : i \in I\}$ be the family of sets in Example 4. Evaluate $\bigcap_{i \in I} B_i$ and $\bigcup_{i \in I} B_i$.

12. Prove Theorem 5.3.7.

13. Let $\mathcal{F}$ and $\mathcal{G}$ be two families of sets. Prove that $\bigcup(\mathcal{F} \cup \mathcal{G}) = (\bigcup \mathcal{F}) \cup (\bigcup \mathcal{G})$.

5.4 The Axioms of Set Theory

Albert Einstein devoted much of his professional life to the search for a unified theory of physics, that is, a theory that fully explains and links together all known physical phenomena. Einstein was not successful in his quest to find such a theory. Since then one of the most engaging goals for researchers in physics has been to construct a unifying theory for physics. Stephen Hawking concludes his book *A Brief History of Time* with the hope that someone will discover a unified theory and observes that if such a theory can be realized, then "it would be the ultimate triumph – for then we would know the mind of God."

Georg Cantor spent much of his professional life in the development of a new branch of mathematics: Set Theory. Little did he know that his work would lead to a unifying theory for mathematics. In his earlier work, Cantor started with a set P of real numbers and then formed the *derived set* P' of all limit points of P. Then he repeated this operation and obtained further derived sets P'', P''', ... Using these derived sets he was able to prove a theorem on trigonometric series. This work lead Cantor to investigate sets in a more general setting and then to develop an abstract theory of sets which would change the whole course of mathematics.

5.4.1 The Zermelo-Fraenkel Axioms

Throughout this chapter we have been informally using set theoretic concepts in the same way that they are used by most contemporary mathematicians. Cantor also used an informal approach in his development of set theory. For example, Cantor regularly used the **Comprehension Principle**: *The collection of all mathematical objects that share a property forms a set.* In other words, given a property $P(x)$, the comprehension principle asserts that the collection $\{x : P(x)\}$ is a set. Such a principle, unfortunately, leads to contradictions. The most well-known of which is called Russel's paradox and is due to Bertrand Russell. Consider the property $x \notin x$, where x is understood to represent a set. The comprehension principle would allow us to conclude that $A = \{x : x \notin x\}$ is a set. Thus, $(\star)$ *the set A consists of all the sets x that satisfy* $x \notin x$. Clearly, either $A \in A$ or $A \notin A$. Suppose $A \in A$. Then, as noted in $(\star)$, A must satisfy the property $A \notin A$ which is a contradiction. Suppose $A \notin A$. Since A satisfies $A \notin A$, we infer from $(\star)$ that $A \in A$ which is again a contradiction. After Russell's paradox appeared, it became clear that the comprehension principle needed to be restricted in some way.

Ernst Zermelo resolved the problems discovered with the comprehension principle by producing a collection of axioms for set theory. Shortly afterward, Abraham Fraenkel amended Zermelo's axioms to obtain the Zermelo-Fraenkel axioms for set theory. These axioms have now become the accepted formulation of Cantor's ideas about the nature of sets.

After years of effort, mathematicians have shown that virtually all mathematical concepts and results can be formalized within set theory. This has been recognized as one of the greatest achievements of modern mathematics and, as a result, we can now say that "set theory is a unifying theory for mathematics."

We now present the Zermelo-Fraenkel axioms. Each of these axioms is first stated in English and then written in logical form. After the presentation, we will discuss these axioms and some of their consequences.

1. **Extensionality Axiom.** *Two sets are equal if and only if they have the same elements.*
$$\forall A \forall B(A = B \leftrightarrow \forall x(x \in A \leftrightarrow x \in B)).$$

2. **Empty Set Axiom.** *There is a set with no elements.*

$$\exists A \forall x (x \notin A).$$

3. **Pairing Axiom.** *For every u and v there is a set that consists of just u and v.*

$$\forall u \forall v \exists A \forall x (x \in A \leftrightarrow (x = u \vee x = v)).$$

4. **Union Axiom.** *For every $\mathcal{F}$ there is a set U that consists of all the elements that belong to at least one set in $\mathcal{F}$.*

$$\forall \mathcal{F} \exists U \forall x (x \in U \leftrightarrow \exists C (x \in C \wedge C \in \mathcal{F})).$$

5. **Power Set Axiom.** *For every set A there is a set P that consists of all the sets that are subsets of A.*

$$\forall A \exists P \forall x (x \in P \leftrightarrow \forall y (y \in x \rightarrow y \in A)).$$

6. **Subset Axiom.** *Let $P(x)$ be a formula with one free variable x. For every set A there is a set S that consists of all the elements $x \in A$ such that $P(x)$ holds.*

$$\forall A \exists S \forall x (x \in S \leftrightarrow (y \in A \wedge P(x))).$$

7. **Infinity Axiom.** *There is a set I that contains the empty set as an element and whenever $x \in I$, then $x \cup \{x\} \in I$.*

$$\exists I (\varnothing \in I \wedge \forall x (x \in I \rightarrow x \cup \{x\} \in I)).$$

8. **Replacement Axiom.** *Let $P(x,y)$ be a formula with two free variables x and y. For every set A if for each $x \in A$ there is a unique y such that $P(x,y)$, then there is a set S that consists of all the elements y such that $P(x,y)$ for some $x \in A$.*

$$\forall A ((\forall x \in A) \exists ! y P(x,y) \rightarrow \exists S \forall y (y \in S \leftrightarrow (\exists x \in A) P(x,y))).$$

9. **Regularity Axiom.** *Any nonempty set A has an element that is disjoint from A.*

$$\forall A (A \neq \varnothing \rightarrow \exists x (x \in A \wedge x \cap A = \varnothing)).$$

The extensionality axiom is a restatement of Definition 5.1.1(1). The empty set axiom asserts that there is a set with no elements. Since the extensionality axiom implies that this set is unique, we let $\varnothing$ denote the empty set. The pairing axiom just states that the set $\{u,v\}$ exists for any two elements u and v. Since $\{u,u\} = \{u\}$, it follows that the set $\{u\}$ also exists for each u.

The union axiom proclaims the union of a family of sets $\mathcal{F}$ exists, that is, the set $\bigcup \mathcal{F}$ exists. In particular, the union axiom and the pairing axiom can be used to

show that the union of two sets exists. Let A and B be two sets. The pairing axiom yields the set $\{A,B\}$. Thus, by the union axiom, the set $\bigcup\{A,B\}$ exists, which is the set $A \cup B$. The power set axiom states that for any set A the set $\mathcal{P}(A)$ exists (see Definition 5.1.3).

The subset axiom affirms that whenever we have a property $P(x)$ and a set A, we can then form the set $\{x \in A : P(x)\}$ which is a subset of A. Of course, the subset axiom is a restricted form of the comprehension principle but it does not lead to the contradiction that we encountered in Russell's paradox. The subset axiom allows us to show that the intersection of two sets exists. Let A and B be two sets and let $P(x)$ be the property $x \in B$. Then the set $\{x \in A : P(x)\}$ is just $A \cap B$. Similarly, one can conclude that the intersection of a family of sets also exists (see Exercise 2).

The infinity axiom declares that there is a set I such that $\varnothing \in I$ and whenever $x \in I$, then $x \cup \{x\} \in I$. Since $\varnothing \in I$, we conclude that $\varnothing \cup \{\varnothing\} = \{\varnothing\} \in I$. Since $\{\varnothing\} \in I$, we have that $\{\varnothing\} \cup \{\{\varnothing\}\} = \{\varnothing, \{\varnothing\}\} \in I$. Continuing in this manner, we see that the set I must contain all of the sets

$$\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, \ldots$$

One can show that all of the sets in the above list are distinct. Thus, the set I contains an infinite number of elements; that is, I is an infinite set.

The replacement axiom asserts that for any set A if for every $x \in A$ there is a unique element y that is "directly connected" to x, then we can *replace* each $x \in A$ with its connection y and the result forms a new set. In the words of Paul Halmos [8], "anything intelligent that one can do to the elements of a set yields a set."

Can a set belong to itself? The regularity axiom rules out this possibility. Using this axiom one can prove that $A \notin A$ for all sets A (see Exercise 3). This completes our discussion of the individual axioms introduced by Zermelo and Fraenkel.

In Sections 5.1–5.3 we presented the basic concepts and results in set theory that you will need for your future mathematics courses. In this section we wanted to let you know that there is another approach to set theory, that is, one that employs the axiomatic method. One advantage of working with a set of axioms is that the initial assumptions that one can use in a proof are made explicit.

As noted earlier, it is a remarkable fact that essentially all mathematical objects can be defined as sets. For example, the natural numbers and the real numbers can be constructed within set theory. Consequently, the theorems of mathematics can be viewed as statements about sets. Moreover, one can prove these theorems using just the axioms of set theory. Thus, it has been said that "mathematics can be embedded in set theory."

5.4.2 The Axiom of Choice

Suppose that we have a family of nonempty sets. Is it possible to uniformly select exactly one element from each set in the family? Alternatively, can we choose one

element from each of these sets and then form a set consisting of just the chosen elements? The following set theoretic principle answers these questions.

Axiom of Choice. *Let $\{A_i : i \in I\}$ be an indexed family of nonempty sets. Then there is an indexed set $\{x_i : i \in I\}$ such that $x_i \in A_i$ for all $i \in I$.*

Definition 5.4.1. Suppose $\{A_i : i \in I\}$ is an indexed family of nonempty sets. A set $\{x_i : i \in I\}$ shall be called a **choice set** if $x_i \in A_i$ for all $i \in I$.

Given an indexed family $\{A_i : i \in I\}$ of nonempty sets, if there is a method for identifying a single element in A_i and if this method works for each $i \in I$, then we can construct a choice set without using the axiom of choice. On the other hand, sometimes the only way to obtain a choice set is by appealing to the axiom of choice. In Examples 1 and 2 below, we will construct a choice set without the axiom of choice. Example 3 presents an indexed family of sets where it is not clear how to define a choice set and thus, one must use the axiom of choice.

Example 1. Suppose $\{A_i : i \in \mathbb{Z}\}$ is an indexed family of nonempty sets of natural numbers. Define a choice set without using the axiom of choice.

Solution. By the Well-Ordering Principle 4.1.1, we know that every set A_i has a least element. Let n_i be the least element in A_i for each $i \in \mathbb{Z}$. Thus, $\{n_i : i \in \mathbb{Z}\}$ is a choice set for the family $\{A_i : i \in \mathbb{Z}\}$. So, we have defined the desired choice set. Ⓢ

Example 2. Consider the indexed family of open intervals $\{(a_i, b_i) : i \in \mathbb{N}\}$ where $a_i < b_i$ are real numbers for each $i \in \mathbb{N}$. Define a choice set without employing the axiom of choice.

Solution. For each $i \in \mathbb{N}$, let $x_i = \frac{a_i + b_i}{2}$ and note that $a_i < x_i < b_i$. So $\{x_i : i \in \mathbb{N}\}$ is a choice set for the family $\{(a_i, b_i) : i \in \mathbb{N}\}$. We have thus defined a choice set and did not use the axiom of choice. Ⓢ

Example 3. Suppose $\{A_i : i \in \mathbb{N}\}$ is an indexed family of nonempty sets of real numbers. Can you define a choice set without applying the axiom of choice?

Solution. All we know is that each set A_i is a nonempty set of real numbers. It is not clear how to uniformly define, for every $i \in \mathbb{N}$, a distinct real number $x_i \in A_i$. So we must use the axiom of choice to obtain a choice set for the family $\{A_i : i \in \mathbb{N}\}$. Ⓢ

If the index set is finite, then one can prove that there is a choice set without using the axiom of choice. Mathematicians often use the axiom of choice when the index set is infinite and it is not clear how to construct a choice set. We will be using the axiom of choice, for this very reason, in our proof of Theorem 6.5.18 in Chapter 6.

The Zermelo-Fraenkel system of axioms is frequently denoted by ZF and the axiom of choice is abbreviated by AC. The axiom of choice was formulated by Ernst Zermelo; however, Zermelo and Fraenkel did not include this axiom in ZF. For this reason, the result of adding the axiom of choice to ZF is denoted by ZFC. There were some early attempts to prove the axiom of choice using just the axioms in ZF; however, these attempts were not successful. Mathematicians then began to

doubt the possibility of proving the axiom of choice and, eventually, this was shown to be the case. The combined work of Paul Cohen and Kurt Gödel proved that the axiom of choice is independent of the Zermelo-Fraenkel axioms. In other words, AC cannot be proven or refuted using only the axioms in ZF.[4] Nevertheless, the axiom of choice is a powerful tool in mathematics and there are important theorems that cannot be proven without it. Consequently, mathematicians typically assume the axiom of choice and usually cite the axiom when they use it in a proof.

Exercises 5.4

1. Let u, v and w be sets. By the pairing axiom, we know that $\{u\}$ and $\{v,w\}$ are sets. Using the pairing axiom again and the union axiom, show that the set $\{u,v,w\}$ exists.

2. Let $\mathcal{F}$ be a family of sets. By the union axiom we know that the set $\bigcup\mathcal{F}$ exists. Let $P(x)$ be the property $(\forall C \in \mathcal{F})(x \in C)$. So, by the subset axiom, the set $\{x \in \bigcup\mathcal{F} : P(x)\}$ exists. Show that $\{x \in \bigcup\mathcal{F} : P(x)\}$ is just the set $\bigcap\mathcal{F}$.

3. Let A be a set. Thus, the pairing axiom implies that the set $\{A\}$ exists. Using the regularity axiom, show that $A \cap \{A\} = \varnothing$. Conclude that $A \notin A$.

4. Given sets A and B, the set $\{A,B\}$ exists by the pairing axiom. Suppose $A \in B$. Using the regularity axiom, show that $A \cap \{A,B\} = \varnothing$. Conclude that $B \notin A$.

5. Let A, B, and C be sets. Suppose that $A \in B$ and $B \in C$. Using the regularity axiom, show that $C \notin A$.

6. Let $\mathcal{F}$ be a family of nonempty sets. Using the axiom of choice, show that there is a set $\{x_A : A \in \mathcal{F}\}$ such that $x_A \in A$ for all $A \in \mathcal{F}$.

7. Define a choice set, without using the axiom of choice, for each of the following family of sets:

(a) $\{A_i : i \in \mathbb{N}\}$ is a family of nonempty subsets of $\mathbb{R}$, where each A_i has exactly three elements.

(b) $\{A_i : i \in \mathbb{R}\}$ is a family of nonempty subsets of $\mathbb{Z}^-$.

(c) $\{A_i : i \in \mathbb{Z}\}$ is a family of nonempty subsets of $\mathbb{Q}^+$.

8. Let $A \subseteq \mathbb{R}$ and suppose that for all $n \in \mathbb{N}$ there is an $x \in A$ such that $0 < x < \frac{1}{n}$. Using the axiom of choice, show that there is an indexed set $\{x_n : n \in \mathbb{N}\}$ such that $0 < x_n < \frac{1}{n}$ and $x_n \in A$, for all $n \in \mathbb{N}$.

[4] Presuming that ZF does not lead to a contradiction.

CHAPTER 6

Functions

One of the most essential ideas in modern mathematics is the concept of a *function*. A function is a way of associating each element of a set A with exactly one element of another set B. The set A will be referred to as the *domain* of the function, and B will be called the *co-domain*. Functions should be familiar to you from calculus and other mathematics courses you have taken; for example, $f(x) = x^2$ is a function that one can differentiate and integrate. The functions studied in your first-year calculus course have the set $\mathbb{R}$ of real numbers, or subsets of $\mathbb{R}$, as domain and co-domain. In this chapter, we will look at functions in a more general context and examine some important properties that functions may possess. To do this, we must first give a precise set-theoretic definition of a function.

6.1 Functions Defined on General Sets

Let A and B be sets. A *function f from A to B* is a subset of $A \times B$ such that for each $x \in A$ there is exactly one $y \in B$ so that $(x,y) \in f$. We now express this notion in terms of a formal definition.

Definition 6.1.1. Let A and B be sets, and let $f \subseteq A \times B$. Then f is said to be a **function from** A **to** B if the following two conditions hold:

(1) For each $x \in A$ there is a $y \in B$ such that $(x,y) \in f$.
(2) If $(x,y) \in f$ and $(x,z) \in f$, then $y = z$.

The set A is called the **domain** of f and the B is called the **co-domain** of f.

Example 1. Let $A = \{a,b,c,d,e\}$ and $B = \{5,6,7,8,9\}$. Then

$$f = \{(a,8),(b,7),(c,9),(d,6),(e,5)\}$$

is a function from A to B because for each $x \in A$ there is exactly one $y \in B$ such that $(x,y) \in f$. On the other hand, the set of ordered pairs

$$g = \{(a,8),(b,7),(c,9),(d,6),(b,8),(e,5)\}$$

is not a function from A to B because $(b,7) \in g$ and $(b,8) \in g$, but $7 \neq 8$. Hence, item (2) of Definition 6.1.1 fails to hold. In addition, the set

$$h = \{(a,8),(b,7),(c,9),(e,5)\}$$

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_6,

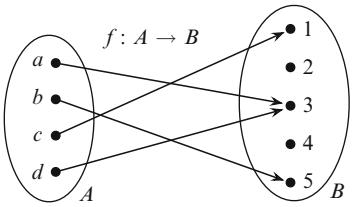


Fig. 6.1 Arrow diagram of the function given in Example 2

is not a function from A to B because $d \in A$ and there is no $y \in B$ such that $(d,y) \in h$. Thus, item (1) of Definition 6.1.1 does not hold.

We write $f: A \to B$ to indicate that f is a function from the set A to the set B. Thus, for each $x \in A$ there is exactly one $y \in B$ such that $(x,y) \in f$. This unique y is called "the value of f at x" and is denoted by $f(x)$. Thus, $(x,y) \in f$ if and only if $f(x) = y$. The value $f(x)$ is called "f of x," or "the image of x under f." In addition, we shall say that $x \in A$ is an *input* for the function f and that $f(x)$ is the resulting *output*. We will also say that x gets *mapped* to $f(x)$.

Remark 6.1.2. Technically speaking, when f is a set of ordered pairs, one can use the notation $f(x)$ only when it is known, or it is clear, that f is a function.

Consider the set of ordered pairs $f = \{(x,y) \in \mathbb{R} \times \mathbb{R} : y = 3x^2 - 1\}$. One can easily show that f satisfies the conditions of Definition 6.1.1. Thus, f is a function and we can write $f(x) = 3x^2 - 1$ for all $x \in \mathbb{R}$.

Remark 6.1.3. Given a function $f: A \to B$, we know that each $x \in A$ is mapped to exactly one element $f(x)$ in B. Consequently, we shall say that f is **single-valued**.

When A and B are finite sets, then a function $f: A \to B$ can be represented by drawing an arrow from each element $x \in A$ to the corresponding element $f(x) \in B$ (see Fig. 6.1). Such a drawing is called an *arrow diagram*. These diagrams can help us gain a better understanding of the concept of a function together with its domain and co-domain.

Example 2. Let $A = \{a,b,c,d\}$ and $B = \{1,2,3,4,5\}$. Now consider the function $f: A \to B$ defined by

$$f = \{(a,3),(b,5),(c,1),(d,3)\}.$$

Thus,

$$f(a) = 3,\ f(b) = 5,\ f(c) = 1, \text{ and } f(d) = 3.$$

Since A and B are finite, we can illustrate the function f by means of the arrow diagram in Fig. 6.1. Clearly, each element $x \in A$ is mapped to exactly one element $f(x)$ in B. Observe that $a \neq d$ and $f(a) = f(d)$. So it is possible for distinct elements in the domain to produce the same value under a function f. In fact, many functions have this "repeated value" property (see Example 7 on page 174).

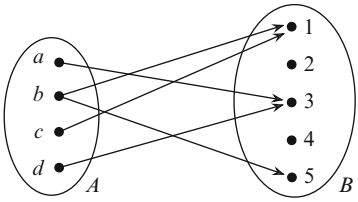


Fig. 6.2a Why is this not a function?

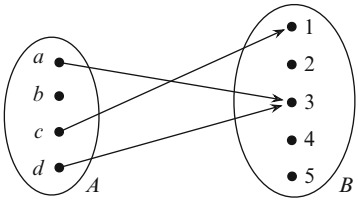


Fig. 6.2b Why is this not a function?

6.1.1 Is it a Function?

In this section we will discuss certain ways in which one may fail to have a function. First we recall a key property that a function must possess. A function $f: A \to B$ must be single valued; that is, it must satisfy the property:

$$\textit{For every element of } x \in A \textit{ there is exactly one value } f(x) \textit{ in } B. \qquad (\star)$$

If this property is not satisfied, then we do not have a function. Figure 6.2a, b presents two arrow diagrams from a set A to a set B that fail to satisfy ($\star$).

Putting the Cart Before the Horse

In mathematics, the expression **well-defined** means that a particular object, which has been described, satisfies the required properties.[1] Usually this object is specified without ambiguity and it is clear that the object satisfies the necessary properties. On the other hand, sometimes it may not be obvious that the object satisfies each of the critical properties and then one must verify that these properties in fact do hold. These issues commonly arise in the definition of a function.

Suppose that A and B are sets and $R(x,y)$ is a predicate. Let

$$f = \{(x,y) \in A \times B : R(x,y)\} \qquad (6.1)$$

and suppose that f satisfies conditions (1) and (2) of Definition 6.1.1. Thus f is a function and for $x \in A$ and $y \in B$ we can conclude that

$$f(x) = y \text{ if and only if } R(x,y). \qquad (6.2)$$

[1] "Putting the horse before the cart" is an expression that is used when the order of certain facts or ideas have been reversed.

Before introducing the notation $f(x)$ in (6.2), we stated that f in (6.1) was a function; however, mathematicians often put the "cart" before the "horse." That is, a mathematician will typically introduce $f(x)$, as in (6.2), without first proving that the set f in (6.1) is a function. This rarely creates a problem, but it can sometimes lead to erroneous conclusions.

In calculus, we often introduce a function f by specifying a formula for $f(x)$. For example, we say "let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3 - x$" which is another way of stating that

$$f(x) = y \text{ if and only if } y = x^3 - x.$$

This is a situation where the set $f = \{(x,y) \in \mathbb{R} \times \mathbb{R} : y = x^3 - x\}$ is a function, as it clearly satisfies the properties given in Definition 6.1.1.

Example 3. Suppose a fellow student wants to know if the description (or rule)

$$f(x) = y \text{ if and only if } y^2 = x \tag{6.3}$$

produces a function $f\colon \mathbb{R}^+ \to \mathbb{R}$. Show that f is not a function.

Solution. Let $f = \{(x,y) \in \mathbb{R}^+ \times \mathbb{R} : y^2 = x\}$. Note that $(4,2) \in f$ as $2^2 = 4$, and $(4,-2) \in f$ because $(-2)^2 = 4$. Since $(4,2) \in f$ and $(4,-2) \in f$, we see that f is not a function. Ⓢ

After changing the "co-domain" of the alleged function in Example 3, we will prove in our next proposition that the description (6.3) will then yield a function f. Such a proof is often referred to as *proving that the function is well-defined*.

Proposition 6.1.4. *Consider the proposed function $f\colon \mathbb{R}^+ \to \mathbb{R}^+$ described by*

$$f(x) = y \textit{ if and only if } y^2 = x. \tag{6.4}$$

Then f is a function.

Proof. Let $f = \{(x,y) \in \mathbb{R}^+ \times \mathbb{R}^+ : y^2 = x\}$. For each $x \in \mathbb{R}^+$, we know that $\sqrt{x}$ is a positive real number (see page 95). Since $(\sqrt{x})^2 = x$, we see that $(x, \sqrt{x}) \in f$. Suppose that $(x,y) \in f$ and $(x,z) \in f$. Then $y^2 = x$ and $z^2 = x$. So $y^2 = z^2$ and thus, $y = \sqrt{y^2} = \sqrt{z^2} = z$ because $y > 0$ and $z > 0$. Hence, $y = z$ and f is a function. □

Many mathematical objects have multiple representations; for instance, 2 can be represented by $\frac{2}{1}$ and $\frac{10}{5}$. Thus, two things may look different when, in fact, they are the same. This multiplicity is one of the main reasons why an apparent function may not be well defined.

Example 4. Consider the purported function $f\colon \mathbb{Q} \to \mathbb{Q}$ described by the rule

$$f\left(\frac{m}{n}\right) = \frac{m-1}{n}, \tag{6.5}$$

where m and $n \neq 0$ are integers. Show that f is not a function.

Solution. The rule in (6.5) is an abbreviation for the more formal description

$$f(x) = y \text{ if and only if } (\exists m \in \mathbb{Z})(\exists n \in \mathbb{Z}) \left(x = \frac{m}{n} \text{ and } y = \frac{m-1}{n} \right)$$

for rational numbers x and y. Using the above description, we see that the intended subset f of $\mathbb{Q} \times \mathbb{Q}$ satisfies

$$(x,y) \in f \text{ if and only if } (\exists m \in \mathbb{Z})(\exists n \in \mathbb{Z}) \left(x = \frac{m}{n} \text{ and } y = \frac{m-1}{n} \right).$$

Observe that $(\frac{1}{2},0) \in f$ because $\frac{1}{2} = \frac{1}{2}$ and $0 = \frac{1-1}{2}$. Also, $(\frac{1}{2},\frac{1}{4}) \in f$ since $\frac{1}{2} = \frac{2}{4}$ and $\frac{1}{4} = \frac{2-1}{4}$. Since $(\frac{1}{2},0) \in f$ and $(\frac{1}{2},\frac{1}{4}) \in f$, we see that f is not a function. Ⓢ

Proposition 6.1.5. *Let $\mathbb{Q}^*$ be the set of nonzero rational numbers and let the putative function $f\colon \mathbb{Q}^* \to \mathbb{Q}^*$ be described by the rule*

$$f\left(\frac{m}{n}\right) = \frac{n}{2m} \tag{6.6}$$

where m and n are nonzero integers. Then f is a function.

Proof. Let $\mathbb{Z}^*$ be the set of nonzero integers. The rule (6.6) informally describes the subset f of $\mathbb{Q}^* \times \mathbb{Q}^*$ defined by

$$(x,y) \in f \text{ if and only if } (\exists m \in \mathbb{Z}^*)(\exists n \in \mathbb{Z}^*) \left(x = \frac{m}{n} \text{ and } y = \frac{n}{2m} \right).$$

We shall prove that f is a function. First we prove that f satisfies property (1) of Definition 6.1.1. Let $x \in \mathbb{Q}^*$. So there are nonzero integers m and n so that $x = \frac{m}{n}$. Since $y = \frac{n}{2m}$ is also in $\mathbb{Q}^*$, we see that $(x,y) \in f$. Now we prove that f satisfies property (2) of Definition 6.1.1. Suppose that $(x,y) \in f$ and $(x,z) \in f$. We must prove that $y = z$. Since $(x,y) \in f$, we have that $x = \frac{m}{n}$ and $y = \frac{n}{2m}$ for some nonzero integers m and n. Similarly, as $(x,z) \in f$, there are nonzero integers a and b such that $x = \frac{a}{b}$ and $z = \frac{b}{2a}$. Since $x = \frac{m}{n}$ and $x = \frac{a}{b}$, we have that $\frac{m}{n} = \frac{a}{b}$. To prove that $y = z$ we shall show that

$$\frac{m}{n} = \frac{a}{b} \text{ implies } \frac{n}{2m} = \frac{b}{2a}. \tag{6.7}$$

Given that $\frac{m}{n} = \frac{a}{b}$, we have that $mb = na$ by Definition 2.1.3. Thus $na = mb$ and so, $n(2a) = b(2m)$. Hence, $\frac{n}{2m} = \frac{b}{2a}$. Therefore, $y = z$ and f is a function. □

In Example 4 and Proposition 6.1.5, we encountered two proposed functions that were described by rules having the form $f(x) = y$ where x and y were represented as a ratio of integers. Each rational number has many such representations. We showed that the rule given in Example 4 does not define a function. On the other hand, we were able to prove that the rule presented in Proposition 6.1.5 does produce a function. Such a proof is also referred to as *proving that the function is well-defined.*

Remark 6.1.6. The proof given in Proposition 6.1.5 can be shortened as follows: Let $m,n,a,b \in \mathbb{Z}^*$. To prove property (1) of Definition 6.1.1, it is enough to observe that $\frac{n}{2m} \in \mathbb{Q}^*$ when $\frac{m}{n} \in \mathbb{Q}^*$. To prove property (2) of Definition 6.1.1, it is sufficient to just prove implication (6.7) which shows that the rule (6.6) is independent of the representation used for an input $x \in \mathbb{Q}^*$.

For the remainder of this chapter (and book) we will be working with a variety of functions. Each such function f will be introduced by defining $f(x)$ directly, or indirectly as in (6.2). This will not cause a problem as each of these definitions will clearly produce a function.

6.1.2 The Range of a Function

Given a function $f: A \to B$ and an element $x \in A$, we know that the value $f(x)$ is an element in B. We now consider the set of all such values of the function f.

Definition 6.1.7. For a function $f: A \to B$ the **range** of f, denoted by $\text{ran}(f)$, is the subset of B defined by

$$\text{ran}(f) = \{f(a) : a \in A\} = \{b \in B : b = f(a) \text{ for some } a \in A\}.$$

The range of a function is the set of all "output" values produced by the function. For the function f in Fig. 6.3a, we see that $\text{ran}(f) = \{1,3,5\}$.

Remark 6.1.8. Let $h: X \to Y$ be a function. Then $b \in \text{ran}(h)$ means that $b = f(x)$ for some $x \in X$; that is, b is a *value* of the function f obtained by some input.

Example 5. For any set A, the **identity function** $i_A: A \to A$ is defined by $i_A(x) = x$ for all $x \in A$. Thus, $\text{ran}(i_A) = \{i_A(x) : x \in A\} = \{x : x \in A\} = A$.

Example 6. Let A and B be nonempty sets. For each $c \in B$, the **constant function** $g: A \to B$ is defined by $g(x) = c$ for all $x \in A$. So, $\text{ran}(g) = \{g(x) : x \in A\} = \{c\}$.

Example 7. Let $f: \mathbb{R} \to \mathbb{R}$ be the function in Fig. 6.3b defined by the formula $f(x) = x^2 - x$. Then $\text{ran}(f) = \{f(x) : x \in \mathbb{R}\} = \{x^2 - x : x \in \mathbb{R}\} = [-\frac{1}{4}, \infty)$. Observe that $f(-2) = 6$ and $f(3) = 6$. Thus, this function has a "repeated value."

For the function f in Fig. 6.3b, note that $-1 \notin \text{ran}(f)$. Similarly, there are elements in the co-domain of the function in Fig. 6.3a that are not in the range of this function. Thus, the range of a function may not equal its co-domain.

6.1.3 Equality of Functions

In your ensuing mathematics courses, you will be asked (at some point) to prove that two functions are equal. Before we address this issue, we present and prove a

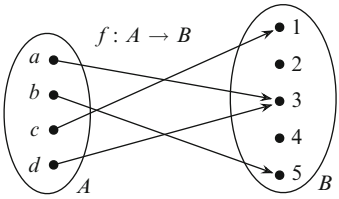


Fig. 6.3a Arrow diagram of a function

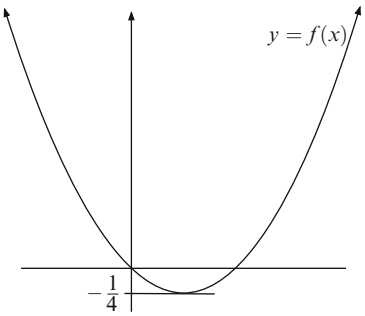


Fig. 6.3b Graph of function in Example 7

theorem stating that two functions f and g are equal (i.e., $f = g$) when they have the same domain and the same value $f(x) = g(x)$ for *all* x in their common domain. In Section 5.2.2 we discussed strategies for proving, and assuming, that two sets are equal. Since a function is a *set* of ordered pairs, these ideas will be used in the proof of our next theorem.

Theorem 6.1.9 (Function Equality). *Let $f: A \to B$ and $g: A \to B$ be functions. Then $f = g$ if and only if $f(x) = g(x)$ for all $x \in A$.*

Proof. Let $f: A \to B$ and $g: A \to B$ be functions. We shall prove that $f = g$ if and only if $f(x) = g(x)$ for all $x \in A$.

($\Rightarrow$). Assume that $f = g$. Let $x \in A$. Since f is a function from A to B, there is a $y \in B$ such that $(x, y) \in f$. Thus, $f(x) = y$. In addition, because $f = g$, it follows that $(x, y) \in g$. Hence, $g(x) = y$. Therefore, $f(x) = g(x)$.

($\Leftarrow$). Assume that $f(x) = g(x)$ for all $x \in A$. We shall prove that $f = g$. Let $(x, y) \in A \times B$. We shall prove that $(x, y) \in f$ if and only if $(x, y) \in g$ as follows:

$$\begin{aligned} (x,y) \in f \text{ iff } & f(x) = y && \text{because } f \text{ is a function} \\ \text{iff } & g(x) = y && \text{because } f(x) = g(x) \\ \text{iff } & (x,y) \in g && \text{because } g \text{ is a function.} \end{aligned}$$

Therefore, $f = g$ and this completes the proof. □

We can now identify a key strategy for proving that two functions are equal.

Proof Strategy 6.1.10. Given a diagram containing the form

Prove $f = g$

where $f: A \to B$ and $g: A \to B$, use the diagram:

Let $x \in A$.
    Prove $f(x) = g(x)$.

Proposition 6.1.11. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = (x-1)^2$ and let $g\colon \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = x^2 - 2x + 1$. Prove that $f = g$.*

Proof. Let $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ be defined as stated in the proposition. Let $x \in \mathbb{R}$. We prove that $f(x) = g(x)$ as follows:

$$\begin{aligned} f(x) &= (x-1)^2 && \text{by the definition of } f \\ &= x^2 - 2x + 1 && \text{by algebra} \\ &= g(x) && \text{by definition of } g. \end{aligned}$$

Therefore, $f = g$. □

Proposition 6.1.12. *Let $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ be functions. Define $s\colon \mathbb{R} \to \mathbb{R}$ and $t\colon \mathbb{R} \to \mathbb{R}$ by*

$$s(x) = f(x) \cdot g(x) \text{ for all } x \in \mathbb{R} \tag{6.8}$$

$$t(x) = g(x) \cdot f(x) \text{ for all } x \in \mathbb{R}. \tag{6.9}$$

Then $s = t$.

Proof. Let $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ be functions and let $s\colon \mathbb{R} \to \mathbb{R}$ and $t\colon \mathbb{R} \to \mathbb{R}$ be defined by (6.8) and (6.9). Let x be a real number. We shall prove that $s(x) = t(x)$ as follows:

$$\begin{aligned} s(x) &= f(x) \cdot g(x) && \text{by (6.8)} \\ &= g(x) \cdot f(x) && \text{by the commutative law of multiplication} \\ &= t(x) && \text{by (6.9).} \end{aligned}$$

Therefore, $s = t$. □

Remark 6.1.13. Given functions $f\colon A \to B$ and $g\colon A \to B$, to show that $f \neq g$ you must find at least one element $x \in A$ and show that $f(x) \neq g(x)$.

Exercises 6.1

1. Let $f\colon \mathbb{Z} \to \mathbb{Z}$ be defined by $f(n) = 4n + 1$. Determine the range of f.
2. Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = -x^2 + 4x$. Determine the range of f.
3. Consider the functions $f\colon \mathbb{R}^+ \to \mathbb{R}$ and $g\colon \mathbb{R}^+ \to \mathbb{R}$ defined by $f(x) = \frac{16x^2-1}{4x+1}$ and $g(x) = 4x - 1$ for all $x \in \mathbb{R}^+$. Prove that $f = g$.
4. Let $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ be functions. Define $s\colon \mathbb{R} \to \mathbb{R}$ and $t\colon \mathbb{R} \to \mathbb{R}$ by

$$s(x) = 2f(x) + 3g(x) \quad \text{for all } x \in \mathbb{R} \tag{6.10}$$

$$t(x) = 6f(x) - g(x) \quad \text{for all } x \in \mathbb{R}. \tag{6.11}$$

Prove that if $s = t$, then $f = g$.

5. Consider the purported function $f\colon \mathbb{Q} \to \mathbb{R}$ defined by $f(\frac{m}{n}) = 2^m 3^n$ where $m, n \in \mathbb{Z}$ and $n \neq 0$. Show that f is not well defined.

6. Consider the purported function $f\colon \mathbb{Q} \to \mathbb{Q}$ defined by $f(\frac{i}{j}) = \frac{ij+i^2}{j^2}$ whenever $i, j \in \mathbb{Z}$ and $j \neq 0$. After reviewing Remark 6.1.6, prove that f is well-defined.

7. Consider the purported functions $f\colon \mathbb{Q} \to \mathbb{Q}$ and $g\colon \mathbb{Q} \to \mathbb{Z}$ defined by

(a) $f(\frac{i}{j}) = \frac{i+j}{j^2}$

(b) $g(\frac{i}{j}) = 6i - j$

where $i, j \in \mathbb{Z}$ and $j \neq 0$. Are f and g functions? Provide a proof, or give a counterexample, to verify your answers. (Review Remark 6.1.6.)

8. Consider the purported function $g\colon \mathbb{N} \to \mathbb{N}$ defined by $g(n) = i$ if and only if $n = ik$ for some $k \in \mathbb{N}$. Show that g is not a function.

9. Consider the purported function $f\colon (0,1) \to \{0,1,2,\ldots,9\}$ defined by

$$f(x) = x_3 \text{ where } x = 0.x_1x_2x_3\cdots \text{ is an infinite decimal expansion of } x.$$

Show that f is not a function. Now read Remark 4.6.3 and then change the definition of f, slightly, so that your new definition will produce a function.

6.2 One-to-One, Onto, and Inverse Functions

In this section, we will examine two of the most useful properties that a function may have; namely, the property of being *one-to-one* and the property of being *onto*. Roughly speaking, a function is one-to-one if it has no repeated values, and a function is onto when every element in its co-domain is a value of the function. We will present formal mathematical definitions and proof strategies that deal with these important concepts. A function that is both one-to-one and onto will allow us to construct a new function, called the *inverse function*.

6.2.1 One-to-One Functions

Some functions (see Fig. 6.3a, b) may have two inputs that are assigned to the same output and thus, such functions have a repeated value. When a function never produces a repeated value, then we will say that the function is *one-to-one*. For example, it is easy to see that the function in Fig. 6.4 is one-to-one. Unfortunately, the vast majority of mathematical functions cannot be represented by an arrow diagram. Without an arrow diagram, it is more difficult to determine whether or

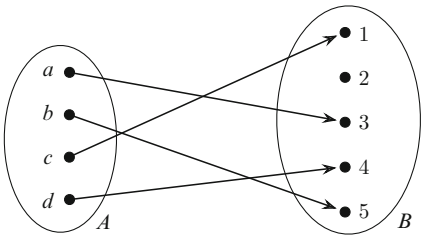


Fig. 6.4 Arrow diagram of a one-to-one function $f: A \to B$

not a function is one-to-one. We need an alternative approach. We need to know how to prove whether or not a function is one-to-one.

Definition 6.2.1. A function $f: A \to B$ is said to be **one-to-one** (or an **injection**), if distinct elements in A get mapped to distinct elements in B; written in logical form

$$(\forall x \in A)(\forall y \in A)[x \neq y \to f(x) \neq f(y)]$$

or equivalently (using the contrapositive),

$$(\forall x \in A)(\forall y \in A)[f(x) = f(y) \to x = y].$$

In many of your future mathematics courses, you will be required to prove that a given function is one-to-one. The following strategy presents a very easy method which can be used to provide such a proof.

Proof Strategy 6.2.2. To prove that a function $f: A \to B$ is one-to-one:

Prove $(\forall x \in A)(\forall y \in A)[f(x) = f(y) \to x = y]$.

That is, use the diagram:

Let $x \in A$ and $y \in A$.
  Assume $f(x) = f(y)$.
    Prove $x = y$.

Proposition 6.2.3. *Define $f: \mathbb{R} \to \mathbb{R}$ by $f(x) = 2x + 3$. Then f is one-to-one.*

Proof Analysis. First we construct the proof diagrams:

Prove $(\forall x \in \mathbb{R})(\forall y \in \mathbb{R})[f(x) = f(y) \to x = y]$

Let $x \in \mathbb{R}$ and $y \in \mathbb{R}$.
  Assume $f(x) = f(y)$.
    Prove $x = y$.

Let $x \in \mathbb{R}$ and $y \in \mathbb{R}$.
  Assume $2x + 3 = 2y + 3$.
    Prove $x = y$.

These diagrams will guide our composition of a well-structured proof. Ⓐ

Proof. We have $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x+3$. We shall prove that f is one-to-one. Let x and y be real numbers. Assume $f(x) = f(y)$. We will prove that $x = y$. Since $f(x) = f(y)$, we see that $2x+3 = 2y+3$. Subtracting 3 from both sides of this equality, we obtain $2x = 2y$. By dividing both sides by 2, we get that $x = y$. Therefore, the function f is one-to-one. □

Given a function $f: A \to B$, to show that f is not one-to-one you must find $x, y \in A$ such that $x \neq y$ and $f(x) = f(y)$, that is, you must show that f has a repeated value.

Example 1. Show that $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = (x-2)^2+1$ is not one-to-one.

Solution. Clearly, $1 \neq 3$ but $f(1) = 2 = f(3)$. Thus, f is not one-to-one. Ⓢ

Many times in mathematics one knows that a certain function is one-to-one and wants to use this fact to establish something new, for example, to show that another function is one-to-one. Our next strategy offers a simple observation that can be used in such a proof. It is a strategy that is often used in mathematical proofs.

Assumption Strategy 6.2.4. Suppose in a proof that you are *assuming* $f: A \to B$ is one-to-one. If $x, y \in A$ and you are given or can prove that $f(x) = f(y)$, then you can conclude that $x = y$.

We will apply Proof Strategy 6.2.2 and Assumption Strategy 6.2.4 in our proof of the following theorem.

Theorem 6.2.5. *Suppose $f: \mathbb{R} \to \mathbb{R}$ is one-to-one and let $c \in \mathbb{R}$ be nonzero. Define $g: \mathbb{R} \to \mathbb{R}$ by $g(x) = cf(x)$. Then g is one-to-one.*

Proof. Suppose $f: \mathbb{R} \to \mathbb{R}$ is one-to-one and $c \neq 0$. We shall prove that $g: \mathbb{R} \to \mathbb{R}$ is one-to-one, where g is defined by $g(x) = cf(x)$. Let $x, y \in \mathbb{R}$ and assume $g(x) = g(y)$. Thus, $cf(x) = cf(y)$ by the definition of g. So, $f(x) = f(y)$ because c is nonzero. Since $f(x) = f(y)$ and f is one-to-one, we have that $x = y$. Hence, g is one-to-one and this completes the proof. □

6.2.2 Onto Functions

We will now focus our attention on the co-domain of a function. Consider the function given by the arrow diagram in Fig. 6.4 on page 178. Observe that $2 \in B$ is not a value of this function. So a function may have elements in its co-domain that are not realized as a value of the function; however, when *every* element in the co-domain is a value of the function, we shall say that the function is *onto*. For example, the function in Fig. 6.5 is onto because for every element $y \in B$ there is an $x \in A$ such that $f(x) = y$, that is, every element in B has an arrow pointing to it.

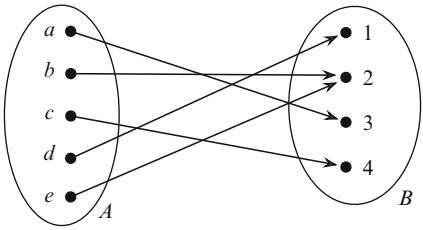


Fig. 6.5 Arrow diagram of an onto function $f: A \to B$

Definition 6.2.6. A function $f: A \to B$ is said to be **onto** (or a **surjection**), if every element $y \in B$ gets mapped-to by some $x \in A$; written in logical form

$$(\forall y \in B)(\exists x \in A)[f(x) = y].$$

In your more advanced mathematics courses, you will be asked to prove that a given function is onto. To do this, you must show that for every element y in the co-domain, there is an element x in the domain that maps to y. Thus, we have our second very important proof strategy.

Proof Strategy 6.2.7. To prove that a function $f: A \to B$ is onto:

$$\text{Prove } (\forall y \in B)(\exists x \in A)[f(x) = y].$$

In other words, use the diagram:

Let y be an element in B.
  Let $x =$ (the element in A you found).
    Prove $f(x) = y$.

Proposition 6.2.8. *Let $f: \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x + 2$. Then f is onto.*

Proof Analysis. First we construct the proof diagrams:

Prove $(\forall y \in \mathbb{R})(\exists x \in \mathbb{R})[f(x) = y]$.

Let y be a real number.
  Prove $(\exists x \in \mathbb{R})[f(x) = y]$.

Let y be a real number.
  Let $x =$ (the element in $\mathbb{R}$ you found).
    Prove $f(x) = y$.

Let y be a real number.
  Let $x = \frac{y-2}{3}$.
    Prove $f(x) = y$.

How did we find the value for x in the latter diagram? We just solved the equation $f(x) = y$ for x, that is, we solved $3x + 2 = y$ for x. This completes our analysis. This latter diagram will guide our composition of a well-structured proof. Ⓐ

Proof. Let y be a real number. Let $x = \frac{y-2}{3}$. We shall show that $f(x) = y$ as follows:

$$f(x) = f\left(\frac{y-2}{3}\right) = 3\left(\frac{y-2}{3}\right) + 2 = (y-2) + 2 = y.$$

Therefore, the function f is onto. □

Given a function $f: A \to B$, to show that f is not onto you must find a value $y \in B$ that satisfies $f(x) \neq y$ for all $x \in A$.

Proposition 6.2.9. *Let $A = \{x \in \mathbb{R} : x \neq -1\}$. Then the function $f: A \to \mathbb{R}$ defined by $f(x) = \frac{2x}{x+1}$ is not onto.*

Proof. Let $y = 2$. We shall prove that $f(x) \neq 2$ for all $x \in A$. Suppose, for a contradiction, that $f(x) = 2$ for some $x \in A$. Thus, $\frac{2x}{x+1} = 2$. Hence, $2x = 2(x+1)$. So, $2x = 2x + 2$. From this equation we derive $0 = 2$, a contradiction. □

Remark. The value $y = 2$, in the proof of Proposition 6.2.9, was found by solving the equation $\frac{2x}{x+1} = y$ for x and obtaining $x = \frac{y}{2-y}$ which is undefined when $y = 2$.

Suppose you are given that a function is onto and you need to use this fact in a mathematical proof. Our next strategy is one that will be very useful.

Assumption Strategy 6.2.10. When *assuming* $f: A \to B$ is onto, then for any $y \in B$ you can conclude that there is an $x \in A$ that satisfies $f(x) = y$.

Proof Strategy 6.2.7 and Assumption Strategy 6.2.10 will be used in our next proof of the following theorem.

Theorem 6.2.11. *Suppose $f: \mathbb{R} \to \mathbb{Q}$ is onto and let $c \in \mathbb{Q}$ be nonzero. Define $g: \mathbb{R} \to \mathbb{Q}$ by $g(x) = cf(x)$. Then g is onto.*

Proof Analysis. We are given that $f: \mathbb{R} \to \mathbb{Q}$ is onto and we want to prove that $g: \mathbb{R} \to \mathbb{Q}$ is onto. Let $y \in \mathbb{Q}$. We need to find an $x \in \mathbb{R}$ that satisfies $g(x) = y$, that is, $cf(x) = y$. First we solve this latter equation for $f(x)$ and obtain $f(x) = \frac{y}{c}$. Since f is onto and $\frac{y}{c}$ is a rational number, there is an $x \in \mathbb{R}$ such that $f(x) = \frac{y}{c}$. This is the x that we will use in our proof. Ⓐ

Proof. We are given that $f: \mathbb{R} \to \mathbb{Q}$ is onto. We shall prove that $g: \mathbb{R} \to \mathbb{Q}$ is onto, where $g(x) = cf(x)$ for all $x \in \mathbb{R}$ and $c \in \mathbb{Q}$ is nonzero. Let y be a rational number. Since f is onto and $\frac{y}{c} \in \mathbb{Q}$, there is an $x \in \mathbb{R}$ such that $f(x) = \frac{y}{c}$. We prove that $g(x) = y$ as follows: $g(x) = cf(x) = c\left(\frac{y}{c}\right) = y$. Therefore, g is onto. □

6.2.3 *Inverse Functions*

In calculus you study the inverse trigonometric functions, and you also learn that the two functions $\ln(x)$ and e^x are inverses of each other. The *inverse* of a function

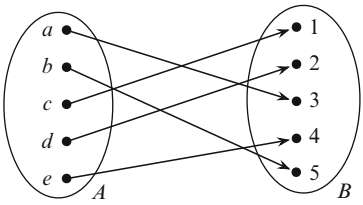


Fig. 6.6a $f: A \to B$

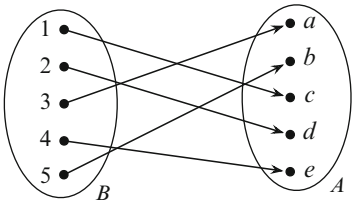


Fig. 6.6b $f^{-1}: B \to A$

is another function that "reverses the action" of the original function. Not every function has an inverse. The only functions that do have an inverse are those that are one-to-one and onto.

Theorem 6.2.12. *Suppose that $f: A \to B$ is one-to-one and onto. Then there is a function $f^{-1}: B \to A$ that satisfies*

$$f^{-1}(b) = a \text{ iff } f(a) = b \tag{6.12}$$

for all $b \in B$ and $a \in A$.

Proof. Suppose $f: A \to B$ is one-to-one and onto. We shall prove that f^{-1}, as defined by (6.12), is a function from B to A. To do this, we shall show that f^{-1} is single-valued. Let $b \in B$. Since $f: A \to B$ is onto, there is an $a \in A$ such that $f(a) = b$. Suppose that $a' \in A$ also satisfies $f(a') = b$. Thus, $f(a) = f(a')$. Because f is one-to-one, it follows that $a = a'$. Therefore, for every $b \in B$ there is exactly one $a \in A$ such that $f(a) = b$. Hence, the formula $f(a) = b$ used in (6.12) defines a function $f^{-1}: B \to A$. □

Definition 6.2.13. Suppose $f: A \to B$ is one-to-one and onto. Then the function $f^{-1}: B \to A$, defined by (6.12), is called the **inverse function** of f.

An arrow diagram of a one-to-one and onto function $f: A \to B$ is given in the Fig. 6.6a. The arrow diagram for the inverse function $f^{-1}: B \to A$ is portrayed in Fig. 6.6b. Observe that the inverse function f^{-1} reverses the action of f and that $f(x) = y$ if and only if $f^{-1}(y) = x$, for each $x \in A$ and $y \in B$.

Finding the Inverse of a "Calculus" Function

Let D and E be nonempty subsets of $\mathbb{R}$. When a function $f: D \to E$ is one-to-one and onto, then the inverse function $f^{-1}: E \to D$ exists. If a formula is given for $f(x)$, then we may be able to find a formula for $f^{-1}(y)$ by following the procedure:

Fix y in the co-domain E of f. Solve the equation $f(x) = y$ for the unique x in D, and set $f^{-1}(y)$ equal to your solution for x.

Example 2. One can show that the function $f: \mathbb{R} \to \mathbb{R}$ given by $f(x) = (x-4)^3 + 2$ is one-to-one and onto. Find a formula for the inverse function f^{-1}.

Solution. After solving the equation $(x-4)^3 + 2 = y$ for x, we obtain $x = 4 + \sqrt[3]{y-2}$. Therefore, $f^{-1}(y) = 4 + \sqrt[3]{y-2}$ is the formula for the inverse function. One can now show that $f(x) = y$ if and only if $f^{-1}(y) = x$, for all $x, y \in \mathbb{R}$. Ⓢ

Unfortunately, the above procedure for finding a formula for an inverse function can fail. For example, consider the one-to-one and onto function $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^5 + x^3$. It is impossible, using radicals, to algebraically solve the equation $x^5 + x^2 = y$ for x and thus, there is no elementary formula for f^{-1}.

Theorem 6.2.14. *Suppose $f: A \to B$ is one-to-one and onto. Let $f^{-1}: B \to A$ be the inverse of f. Then f^{-1} is also one-to-one and onto.*

Proof. Let $f: A \to B$ be one-to-one and onto. We first prove that $f^{-1}: B \to A$ is one-to-one. Let $b, b' \in B$. Assume $f^{-1}(b) = f^{-1}(b')$. Let $a \in A$ be this common value. Thus, $f^{-1}(b) = a$ and $f^{-1}(b') = a$. So $f(a) = b$ and $f(a) = b'$, by (6.12). Since f is a function, we conclude that $b = b'$. Hence, f^{-1} is one-to-one.

To prove that $f^{-1}: B \to A$ is onto, let $a \in A$. So there is a $b \in B$ be such that $f(a) = b$. By (6.12), $f^{-1}(b) = a$. Therefore, f^{-1} is onto. □

Exercises 6.2

1. Define $f: \mathbb{Z} \to \mathbb{Z}$ by $f(n) = 3n + 2$.

(a) Is f one-to-one? Prove it, or provide a counterexample.
(b) Is f onto? Prove it, or provide a counterexample.

2. Define $f: \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$.

(a) Is f one-to-one? Prove it, or provide a counterexample.
(b) Is f onto? Prove it, or provide a counterexample.

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

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

5. Let $A = \{x \in \mathbb{R} : x \neq -1\}$. Define $f: A \to \mathbb{R}$ by $f(x) = \frac{2x}{x+1}$. Prove that f is one-to-one.

6. Let $A = \{x \in \mathbb{R} : x \neq 1\}$. Define $f: A \to \mathbb{R}$ by $f(x) = \frac{3x}{2x-2}$. Prove f is not onto.

7. Define $f: \mathbb{R} \to \mathbb{R}$ by $f(x) = x - x^3$. Is f one-to-one? Is it onto?

8. Let $A = \{x \in \mathbb{R} : x \neq 2\}$ and let $B = \{y \in \mathbb{R} : y \neq 4\}$. Define $f: A \to B$ by $f(x) = \frac{4x}{x-2}$. Prove that f is onto.

9. Let $A = \{x \in \mathbb{R} : x \neq 2\}$ and let $B = \{y \in \mathbb{R} : y \neq 4\}$. Prove the function $f: A \to B$ defined by $f(x) = \frac{4x}{x-2}$ is one-to-one.

10. Let $a, b \in \mathbb{R}$ with $a \neq 0$ and define the function $f: \mathbb{R} \to \mathbb{R}$ by $f(x) = ax + b$. Given that f is one-to-one and onto, find a formula for $f^{-1}: \mathbb{R} \to \mathbb{R}$.

11. Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$, defined below, is one-to-one and onto.

$$f(x) = \begin{cases} x^2, & \text{if } x \geq 0; \\ -x^2, & \text{if } x < 0. \end{cases} \tag{6.13}$$

12. Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function defined by (6.13) in Exercise 11. Given that f is one-to-one and onto, find a formula for the inverse function $f^{-1}\colon \mathbb{R} \to \mathbb{R}$.

13. Suppose that $f\colon \mathbb{R} \to \mathbb{R}^+$ is one-to-one. Define $g\colon \mathbb{R} \to \mathbb{R}^+$ by $g(x) = (f(x))^2$. Prove that g is one-to-one.

14. Suppose that $f\colon \mathbb{R} \to \mathbb{R}^+$ is onto. Define $g\colon \mathbb{R} \to \mathbb{R}^+$ by $g(x) = (f(x))^2$. Prove that g is onto.

15. Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is one-to-one and let $a, b \in \mathbb{R}$ where $a \neq 0$. Define $g\colon \mathbb{R} \to \mathbb{R}$ by $g(x) = af(x) + b$. Prove that g is one-to-one.

16. Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is onto and let $a, b \in \mathbb{R}$ where $a \neq 0$. Define $g\colon \mathbb{R} \to \mathbb{R}$ by $g(x) = af(x) + b$. Prove that g is onto.

17. Define $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $f(m,n) = 2^m 3^n$. Prove that f is one-to-one.

18. Let $f\colon \mathbb{R} \to \mathbb{R}$ be as in Example 2. Prove that f is one-to-one and onto.

Exercise Notes: For Exercise 11, if $x^2 = y^2$ then $|x| = |y|$. For Exercise 17, let $m, n, i, j \in \mathbb{N}$. Assume $f(m,n) = f(i,j)$. Prove $m = i$ and $n = j$.

6.3 Composition of Functions

If the domain of a function equals the co-domain of another function, then we can use these two functions to construct a new function called the *composite function*. The composite function is defined by taking the output of one these functions and using that as the input for the other function. The formal mathematical definition appears below.

Definition 6.3.1. For functions $g\colon A \to B$ and $f\colon B \to C$, one forms the **composite function** $(f \circ g)\colon A \to C$ by defining $(f \circ g)(x) = f(g(x))$ for all $x \in A$.

For example, let $g\colon A \to B$ and $f\colon B \to C$ be the functions in Fig. 6.7. An arrow diagram for the composite function $(f \circ g)\colon A \to C$ appears in Fig. 6.8.

Example 1. Let $g\colon \mathbb{R} \to \mathbb{R}$ and $f\colon \mathbb{R} \to \mathbb{R}$ be the functions defined by $f(x) = \frac{1}{x^2+2}$ and $g(x) = 2x - 1$. Find formulas for $(f \circ g)(x)$ and $(g \circ f)(x)$. Is $f \circ g = g \circ f$?

Solution. Let $x \in \mathbb{R}$. We evaluate the function $(f \circ g)(x)$ as follows:

$$(f \circ g)(x) = f(g(x)) = f(2x-1) = \frac{1}{(2x-1)^2 + 2}.$$

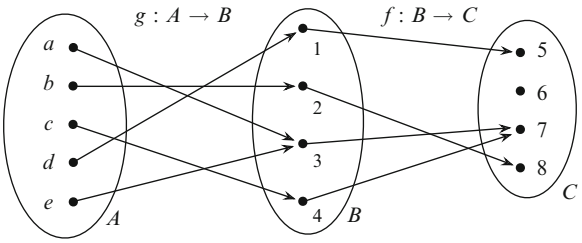


Fig. 6.7 Two functions f and g where the domain of f equals the co-domain of g

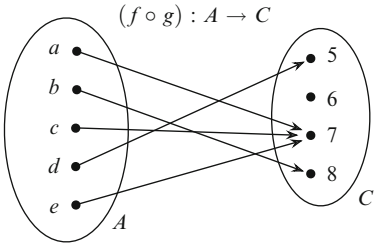


Fig. 6.8 The resulting composite function $f \circ g$ for the functions in Fig. 6.7

Thus, $(f \circ g)(x) = \frac{1}{(2x-1)^2+2}$. We evaluate $(g \circ f)(x)$ to obtain

$$(g \circ f)(x) = g(f(x)) = g\left(\frac{1}{x^2+2}\right) = 2\left(\frac{1}{x^2+2}\right) - 1 = \frac{2}{x^2+2} - 1.$$

Hence, $(g \circ f)(x) = \frac{2}{x^2+2} - 1$. Since $(f \circ g)(0) = \frac{1}{3}$ and $(g \circ f)(0) = 0$, we conclude that $f \circ g \neq g \circ f$ (review Remark 6.1.13). Ⓢ

One cannot form the composition of just any two functions. When in doubt here is a simple rule to follow: *The composition $f \circ g$ is defined when the domain of the left function f is equal to the co-domain of the right function g.*

Remark 6.3.2. Given two functions $g: A \to E$ and $f: B \to C$, if $\text{ran}(g) \subseteq B$, then one can also define the composition $(f \circ g): A \to C$. In other words, if $f(b)$ is defined for every value b of the function g, then one can define $f \circ g$.

6.3.1 Composing a Function with the Identity Function

The identity function just takes an input x and returns x as its output value. As a result, when one composes a function f with the identity function, the result will just be the function f.

Theorem 6.3.3. *Let f be any function $f: A \to B$. Let $i_A: A \to A$ be the identity function on A and let $i_B: B \to B$ be the identity function on B. Then*

(1) $(f \circ i_A) = f$,
(2) $(i_B \circ f) = f$.

Proof. Clearly, $(f \circ i_A)(x) = f(i_A(x)) = f(x)$ and $(i_B \circ f)(x) = i_B(f(x)) = f(x)$, for each $x \in A$. □

6.3.2 Composing a Function with Its Inverse

Since the inverse of a function "reverses the action" of the original function, the result of composing these two functions leads to "no action."

Theorem 6.3.4. *Suppose $f : A \to B$ is one-to-one and onto. Let $f^{-1} : B \to A$ be the inverse of f. Then*

(1) $f^{-1}(f(a)) = a$ *for all* $a \in A$,
(2) $f(f^{-1}(b)) = b$ *for all* $b \in B$.

Proof. First we prove (1). Let $a \in A$. Since $f(a) \in B$, let $b \in B$ be such that $f(a) = b$. Theorem 6.2.12 implies $(*)$ $f^{-1}(b) = a$. After substituting $b = f(a)$ into equation $(*)$, we see that $f^{-1}(f(a)) = a$. To prove (2), let $b \in B$. Since $f^{-1}(b) \in A$, let $a \in A$ be such that $f^{-1}(b) = a$. Thus, $(\dagger)$ $f(a) = b$ by Theorem 6.2.12. Upon substituting $a = f^{-1}(b)$ into equation $(\dagger)$, we obtain $f(f^{-1}(b)) = b$. □

Corollary 6.3.5. *Suppose $f : A \to B$ is one-to-one and onto. Let $f^{-1} : B \to A$ be the inverse of f. Then*

(1) $(f^{-1} \circ f) = i_A$
(2) $(f \circ f^{-1}) = i_B$

where i_A is the identity function on A and i_B is the identity function on B.

Proof. Since $i_A(a) = a$ for $a \in A$ and $i_B(b) = b$ for $b \in B$, items (1) and (2) follow from the corresponding items in Theorem 6.3.4. □

6.3.3 Composing One-to-One Functions

Our next theorem shows that the composition of two one-to-one functions is also one-to-one. Our proof employs Proof Strategy 6.2.2 and Assumption Strategy 6.2.4.

Theorem 6.3.6. *If $g : A \to B$ and $f : B \to C$ are one-to-one, then $(f \circ g) : A \to C$ is one-to-one.*

Proof Analysis. First we construct the proof diagrams:

Assume $g: A \to B$ is one-to-one.
Assume $f: B \to C$ is one-to-one.
Prove $(f \circ g): A \to C$ is one-to-one.

Assume $g: A \to B$ is one-to-one.
Assume $f: B \to C$ is one-to-one.
Let $x \in A$ and $y \in A$.
Assume $(f \circ g)(x) = (f \circ g)(y)$.
Prove $x = y$.

We shall use this last diagram as the guide for our proof. Ⓐ

Proof. Assume $g: A \to B$ and $f: B \to C$ are one-to-one. To prove that the function $(f \circ g): A \to C$ is one-to-one, let $x \in A$ and $y \in A$. Assume $(f \circ g)(x) = (f \circ g)(y)$. Thus, (i) $f(g(x)) = f(g(y))$ by the definition of composition. Since f is one-to-one, we conclude from (i) that $g(x) = g(y)$. Because g is one-to-one, we see that $x = y$. This completes the proof. □

6.3.4 Composing Onto Functions

The next theorem asserts that the composition of two onto functions yields an onto function. Our proof applies Proof Strategy 6.2.7 and Assumption Strategy 6.2.10.

Theorem 6.3.7. *If $g: A \to B$ and $f: B \to C$ are onto, then $(f \circ g): A \to C$ is onto.*

Proof Analysis. First we construct the proof diagrams:

Assume $g: A \to B$ is onto.
Assume $f: B \to C$ is onto.
Prove $(f \circ g): A \to C$ is onto.

Assume $g: A \to B$ is onto.
Assume $f: B \to C$ is onto.
Let z be an element in C.
Let $x =$ (the element in A you found).
Prove $(f \circ g)(x) = z$.

These proof diagrams will guide our proof. Ⓐ

Proof. Assume $g: A \to B$ and $f: B \to C$ are onto. We shall prove that the function $(f \circ g): A \to C$ is onto. Let $z \in C$. Since $f: B \to C$ is onto and $z \in C$, there is a $y \in B$ such that $f(y) = z$. Because $y \in B$ and $g: A \to B$ is onto, there is an $x \in A$ such that $g(x) = y$. We will show that $(f \circ g)(x) = z$ as follows:

$$\begin{aligned}(f\circ g)(x) &= f(g(x)) && \text{by definition of composition}\\ &= f(y) && \text{because } g(x)=y\\ &= z && \text{because } f(y)=z.\end{aligned}$$ □

Exercises 6.3

1. Let $f\colon \mathbb{R}\to\mathbb{R}$ be defined by $f(x)=x^2-1$ and let $g\colon \mathbb{R}\to\mathbb{R}$ be defined by $g(x)=x^3+1$. Evaluate the values:

(a) $(f\circ g)(1)$
(b) $(g\circ f)(1)$
(c) $(f\circ f)(1)$
(d) $(g\circ g)(1)$.

2. Let $f\colon \mathbb{R}\to\mathbb{R}$ be defined by $f(x)=x^2-1$ and let $g\colon \mathbb{R}\to\mathbb{R}$ be defined by $g(x)=x^3+1$. Obtain formulas for the following compositions:

(a) $(f\circ g)(x)$
(b) $(g\circ f)(x)$
(c) $(f\circ f)(x)$
(d) $(g\circ g)(x)$.

3. Let $f\colon A\to B$ and $g\colon B\to A$ be functions. Suppose that $f(g(b))=b$ for all $b\in B$ and $g(f(a))=a$ for all $a\in A$. Prove that f and g are one-to-one and onto.

4. For $a,b\in\mathbb{R}$ with $a\neq 0$, define the function $T_{a,b}\colon \mathbb{R}\to\mathbb{R}$ by $T_{a,b}(x)=ax+b$. Let G be the set of all such functions, that is, let $G=\{T_{a,b} : a,b\in\mathbb{R} \text{ and } a\neq 0\}$.

(a) Let $T_{a,b}\in G$ and $T_{c,d}\in G$. Show that $T_{a,b}\circ T_{c,d}=T_{ac,ad+b}$.
(b) Let $T_{a,b}\in G$ and $T_{c,d}\in G$. Show that $(T_{a,b}\circ T_{c,d})\in G$.
(c) Let $T_{a,b}\in G$. Prove that $T_{a,b}$ is one-to-one and onto.
(d) Let $I\colon \mathbb{R}\to\mathbb{R}$ be the identity function. Show that $I\in G$.
(e) Let $T_{a,b}\in G$. Show that $T_{a,b}^{-1}=T_{\frac{1}{a},-\frac{b}{a}}$ and thus, $T_{a,b}^{-1}\in G$.
(f) Find a $T_{a,b}\in G$ and $T_{c,d}\in G$ so that $(T_{a,b}\circ T_{c,d})\neq(T_{c,d}\circ T_{a,b})$.

5. Given $a\in\mathbb{Q}$ with $a\neq 0$ and $b\in\mathbb{R}$, define $T_{a,b}\colon \mathbb{R}\to\mathbb{R}$ by $T_{a,b}(x)=ax+b$. Let $H=\{T_{a,b} : a\in\mathbb{Q},\ b\in\mathbb{R} \text{ and } a\neq 0\}$.

(a) For any $T_{a,b}\in H$ and $T_{c,d}\in H$, show that $(T_{a,b}\circ T_{c,d})\in H$.
(b) Let $I\colon \mathbb{R}\to\mathbb{R}$ be the identity function. Show that $I\in H$.
(c) Let $T_{a,b}\in H$. Show that $T_{a,b}^{-1}\in H$.

6. Let $g\colon A\to B$ and $f\colon B\to C$. Suppose that $(f\circ g)\colon A\to C$ is one-to-one. Prove that g is one-to-one.

7. Let $g\colon A\to B$ and $f\colon B\to C$. Suppose that $(f\circ g)\colon A\to C$ is one-to-one and g is onto. Prove that f is one-to-one.

8. Let $f: B \to C$ and $g: A \to B$. Suppose that $(f \circ g): A \to C$ is onto. Prove that f is onto.

9. Let $g: A \to B$ and $f: B \to C$. Suppose that $(f \circ g): A \to C$ is onto and f is one-to-one. Prove that g is onto.

10. Let $h: A \to B$, $g: B \to C$ and $f: C \to D$. Prove that $(f \circ g) \circ h = f \circ (g \circ h)$.

Exercise Notes: For Exercise 4(a), find a formula for $T_{a,b}(T_{c,d}(x))$. For Exercise 4(e), find a formula for $T_{a,b}^{-1}$.

6.4 Functions Acting on Sets

There are times when we are more interested in what a function does to an entire subset of its domain, rather than how it affects an individual element in the domain. Understanding this behavior on sets can allow one to better understand the function itself and can reveal some properties concerning its domain and range. The concept of a function "acting on a set," is one that appears in every branch of mathematics.

Definition 6.4.1 (Image of a Set). Let $f: X \to Y$ be a function. Let $S \subseteq X$. The set $f[S]$, called the **image** of S, is defined by

$$f[S] = \{f(x) : x \in S\} = \{y \in Y : y = f(x) \text{ for some } x \in S\}.$$

Figure 6.9 illustrates Definition 6.4.1. The square S represents a subset of the domain of the function f. The image $f[S]$, represented by the rectangle, is the set of all values of the function that are obtained from the inputs that are in the set S.

Example 1. Let $f: \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = |x|$ and $S = \{-12, -3, 2, 3\}$. Then the image of S is $f[S] = \{f(x) : x \in S\} = \{|x| : x \in S\} = \{2, 3, 12\}$. Observe that $f(12) \in f[S]$ and yet $12 \notin S$.

Example 2. Let $f: \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$ and $S = \{-4, -3, 2, 3\}$. Then the image of S is $f[S] = \{f(x) : x \in S\} = \{x^2 : x \in S\} = \{16, 9, 4\}$. Let U be the interval $[-2, 3]$. Then $f[U] = \{f(x) : x \in U\} = \{x^2 : -2 \leq x \leq 3\} = [0, 9]$.

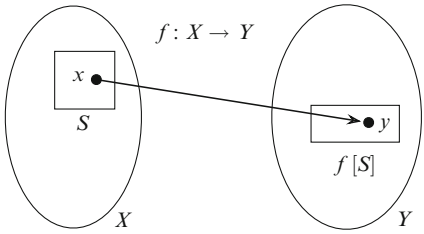


Fig. 6.9 Starting with $S \subseteq X$ we can construct the image $f[S] \subseteq Y$

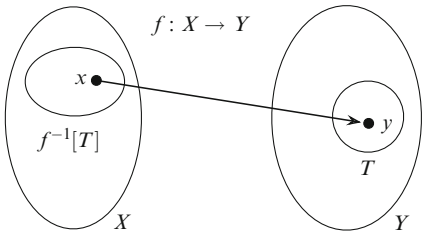


Fig. 6.10 Starting with $T \subseteq Y$ we can construct the inverse image $f^{-1}[T] \subseteq X$

Given a subset S of the domain of a function f, the set $f[S]$ is a subset of the co-domain of f which consists of all the values of the function that result from evaluating $f(x)$ for every x in the set S. We will now turn this process around. Our next definition will allow us to start with a subset T of the co-domain and then construct a subset of the domain.

Definition 6.4.2 (Inverse Image of a Set). Let $f: X \to Y$ be a function. Let $T \subseteq Y$. The set $f^{-1}[T]$ is the subset of X defined by

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

The set $f^{-1}[T]$ is called the **inverse image** of T.

Definition 6.4.2 is depicted in Fig. 6.10. The circle T represents a subset of the co-domain of the function f. The inverse image $f^{-1}[T]$ is represented by an ellipse. The set $f^{-1}[T]$ consists of those inputs in the domain whose value under f belongs to the set T.

Example 3. Let $f: \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = |x|$ and let $T = \{-8, 2, 3\}$. Then $f^{-1}[T] = \{x \in \mathbb{R} : f(x) \in T\} = \{x \in \mathbb{R} : |x| \in T\} = \{-3, -2, 2, 3\}$.

Example 4. Let $f: \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Let $T = \{-4, -3, 4, 25\}$ and $V = [-2, 9]$ where $[-2, 9] = \{x \in \mathbb{R} : -2 \leq x \leq 9\}$. Then the inverse image of T and V are given by

$$f^{-1}[T] = \{x \in \mathbb{R} : f(x) \in T\} = \{x \in \mathbb{R} : x^2 \in T\} = \{-5, -2, 2, 5\}$$

and

$$f^{-1}[V] = \{x \in \mathbb{R} : f(x) \in V\} = \{x \in \mathbb{R} : -2 \leq x^2 \leq 9\} = [-3, 3].$$

The notation f^{-1}, presented in Definition 6.4.2, should not be confused with that of an inverse function. Theorem 6.2.12 implies that the inverse function exists if and only if the original function is one-to-one and onto. Definition 6.4.2 applies to all functions, even those that are not one-to-one and onto.

Our next remark makes four observations that can be very useful when working with the image or the inverse image of a set.

Remark 6.4.3. Let $f: X \to Y$, $S \subseteq X$, $T \subseteq Y$, $a \in X$ and $b \in Y$.

1. If $a \in S$, then $f(a) \in f[S]$.
2. $b \in f[S]$ if and only if $b = f(x)$ for some $x \in S$.
3. If $a \in f^{-1}[T]$, then $f(a) \in T$.
4. If $f(a) \in T$, then $a \in f^{-1}[T]$.

Image Warning: If $f(a) \in f[S]$, then we **cannot necessarily conclude** that $a \in S$ (see Example 1); however, we can conclude that $f(a) = f(x)$ for some $x \in S$.

Theorem 6.4.4. *Let $f: X \to Y$ be a function. Let S be a subset of X, and let T be a subset of Y. Then $f[S] \subseteq T$ if and only if for all $x \in S$ we have $f(x) \in T$.*

Theorem 6.4.5. *Let $f: X \to Y$ be a function. Let C, D be subsets of X and let U, V be subsets of Y. Then*

(a) $f[C \cap D] \subseteq f[C] \cap f[D]$
(b) $f[C \cup D] = f[C] \cup f[D]$
(c) $f^{-1}[U \cap V] = f^{-1}[U] \cap f^{-1}[V]$
(d) $f^{-1}[U \cup V] = f^{-1}[U] \cup f^{-1}[V]$.

Proof. We shall prove only (a) and (d). Let $f: X \to Y$ be a function. Let C, D be subsets of X and let U, V be subsets of Y.

(a). We prove $f[C \cap D] \subseteq f[C] \cap f[D]$. Let $y \in f[C \cap D]$. We will show that $y \in f[C] \cap f[D]$. Since $y \in f[C \cap D]$, there is an $x \in C \cap D$ such that $y = f(x)$ (see Remark 6.4.3(2)). Because $x \in C \cap D$, we see that $x \in C$ and $x \in D$. Therefore, $f(x) \in f[C]$ and $f(x) \in f[D]$. Since $y = f(x)$, we conclude that $y \in f[C] \cap f[D]$.

(d). We prove $f^{-1}[U \cup V] = f^{-1}[U] \cup f^{-1}[V]$.

($\subseteq$). To show that $f^{-1}[U \cup V] \subseteq f^{-1}[U] \cup f^{-1}[V]$, let $x \in f^{-1}[U \cup V]$. We prove $x \in f^{-1}[U] \cup f^{-1}[V]$ as follows:

$$\begin{aligned}
x \in f^{-1}[U \cup V] &\Rightarrow f(x) \in U \cup V && \text{by definition of inverse image} \\
&\Rightarrow f(x) \in U \text{ or } f(x) \in V && \text{by definition of } \cup \\
&\Rightarrow x \in f^{-1}[U] \text{ or } x \in f^{-1}[V] && \text{by definition of inverse image} \\
&\Rightarrow x \in f^{-1}[U] \cup f^{-1}[V] && \text{by definition of } \cup.
\end{aligned}$$

Therefore, $f^{-1}[U \cup V] \subseteq f^{-1}[U] \cup f^{-1}[V]$.

($\supseteq$). Now we prove that $f^{-1}[U] \cup f^{-1}[V] \subseteq f^{-1}[U \cup V]$. Let $x \in f^{-1}[U] \cup f^{-1}[V]$. We prove that $x \in f^{-1}[U \cup V]$ as follows:

$$\begin{aligned}
x \in f^{-1}[U] \cup f^{-1}[V] &\Rightarrow x \in f^{-1}[U] \text{ or } x \in f^{-1}[V] && \text{by definition of } \cup \\
&\Rightarrow f(x) \in U \text{ or } f(x) \in V && \text{by def. of inverse image} \\
&\Rightarrow f(x) \in U \cup V && \text{by definition of } \cup \\
&\Rightarrow x \in f^{-1}[U \cup V] && \text{by def. of inverse image.}
\end{aligned}$$

Therefore, $f^{-1}[U] \cup f^{-1}[V] \subseteq f^{-1}[U \cup V]$. This completes the proof of (d). □

Theorem 6.4.6. *Let $f: X \to Y$ be a function. Let C, D be subsets of X. If f is one-to-one, then $f[C \cap D] = f[C] \cap f[D]$.*

Proof. Let $f: X \to Y$ be a function. Let C, D be subsets of X and assume that f is one-to-one. We shall prove that $f[C \cap D] = f[C] \cap f[D]$. Theorem 6.4.5(a) implies that $f[C \cap D] \subseteq f[C] \cap f[D]$. To show that $f[C] \cap f[D] \subseteq f[C \cap D]$, let $y \in f[C] \cap f[D]$. We will prove that $y \in f[C \cap D]$. Since $y \in f[C] \cap f[D]$, we see that $y \in f[C]$ and $y \in f[D]$. Because $y \in f[C]$, there is a $c \in C$ such that $f(c) = y$. Also, since $y \in f[D]$, there is a $d \in D$ such that $f(d) = y$. Hence, $y = f(c) = f(d)$. Since f is one-to-one, we have $c = d$. Thus, $c \in D$. So $c \in C \cap D$ and therefore, $y = f(c) \in f[C \cap D]$. We conclude that $f[C \cap D] = f[C] \cap f[D]$. □

Exercises 6.4

1. Using Definitions 6.4.1 and 6.4.2, explain why items 1–4 of Remark 6.4.3 hold.

2. Prove Theorem 6.4.4.

3. Prove item (b) of Theorem 6.4.5.

4. Prove item (c) of Theorem 6.4.5.

5. Given $a, b \in \mathbb{R}$ with $a > 0$, define the function $f: \mathbb{R} \to \mathbb{R}$ by $f(x) = ax + b$. Let $U = [2, 3]$. Using interval notation, evaluate $f[U]$ and $f^{-1}[U]$.

6. Define the function $f: \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$ and let $U = [-1, 4]$. Show the following:

(a) $f[f^{-1}[U]] \neq U$.
(b) $f^{-1}[f[U]] \neq U$.
(c) $f[f^{-1}[U]] \neq f^{-1}[f[U]]$.

7. Let $f: X \to Y$ be a function and let $A \subseteq X$ and $B \subseteq X$. Prove that if $A \subseteq B$, then $f[A] \subseteq f[B]$.

8. Let $f: \mathbb{R} \to \mathbb{R}$ be the function defined in Example 1 on page 189. Find $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$ such that $f[A] \subseteq f[B]$ and $A \not\subseteq B$.

9. Suppose $f: X \to Y$ is a one-to-one function. Let $A \subseteq X$ and $B \subseteq X$. Prove that if $f[A] \subseteq f[B]$, then $A \subseteq B$.

10. Let $f: X \to Y$ be a function and let $C \subseteq Y$ and $D \subseteq Y$. Prove that if $C \subseteq D$, then $f^{-1}[C] \subseteq f^{-1}[D]$.

11. Let $f: \mathbb{R} \to \mathbb{R}$ be the function defined in Example 3. Find $C \subseteq \mathbb{R}$ and $D \subseteq \mathbb{R}$ such that $f^{-1}[C] \subseteq f^{-1}[D]$ and $C \not\subseteq D$.

12. Suppose $f: X \to Y$ is onto and let $C \subseteq Y$ and $D \subseteq Y$. Prove if $f^{-1}[C] \subseteq f^{-1}[D]$, then $C \subseteq D$.

13. Let $f: X \to Y$ be a function. Let A be a subset of X. Prove that $A \subseteq f^{-1}[f[A]]$.

14. Suppose $f: X \to Y$ is one-to-one. Let $A \subseteq X$ and $x \in X$. Prove if $f(x) \in f[A]$, then $x \in A$.

15. Suppose that $f: X \to Y$ is one-to-one. Let $A \subseteq X$. Prove that $A = f^{-1}[f[A]]$.

16. Let $f: X \to Y$. Suppose $A = f^{-1}[f[A]]$ for all finite subsets A of X. Prove f is one-to-one.

17. Let $f: X \to Y$ be a function. Let C be a subset of Y. Prove that $f[f^{-1}[C]] \subseteq C$.

18. Assume that $f: X \to Y$ is an onto function. Let $C \subseteq Y$. Prove that $f[f^{-1}[C]] = C$.

19. Given $a, b \in \mathbb{R}$ with $a > 0$, define the function $f: \mathbb{R} \to \mathbb{R}$ by $f(x) = ax + b$. Using Exercises 15 and 18, prove that $f[f^{-1}[U]] = f^{-1}[f[U]]$ for every $U \subseteq \mathbb{R}$.

20. Let $f: X \to Y$ be a function. Let $\{C_i : i \in I\}$ be an indexed family of sets where $C_i \subseteq X$ for all $i \in I$. Prove that $f\left[\bigcup_{i \in I} C_i\right] = \bigcup_{i \in I} f[C_i]$.

Exercise Notes: Exercise 8 shows that the converse of Exercise 7 is not necessarily true for all functions. Exercise 9, however, shows that this converse is true for all one-to-one functions (review assumption strategy 6.2.4 on page 179). Similarly, Exercise 11 shows that the converse of Exercise 10 is not true for all functions; but, Exercise 12 shows that this converse is true for all functions that are onto (review Assumption Strategy 6.2.10 on page 181).

6.5 On the Size of Infinite Sets

The size of a finite set can easily be measured; for example, the size of the set $A = \{1, 2, 3, \ldots, 50\}$ is 50 because it has 50 elements, and the size of the sets $B = \{\pi, 2, 30, -2\}$ and $C = \{9, 11, -1, 5\}$ is 4. Clearly, the size of A is bigger than the size of B. In addition, the sets B and C have the same size. Can the idea of "size" be extended to infinite sets? Georg Cantor was the first mathematician to seriously address and answer this question. Cantor found a way to measure the size of any infinite set. He first observed that two sets A and B have the same size if there is a *one-to-one correspondence* between A and B; that is, there is a way of evenly matching the elements in A with the elements in B. In other words, Cantor observed that A and B have the same size, if there is a one-to-one and onto function $f: A \to B$.

For example, the arrow diagram in Fig. 6.11 presents a function that is one-to-one and onto. As a result, we can use this function to construct the following one-to-one correspondence (6.14) between the sets A and B:

$$\begin{array}{c} A: a\ b\ c\ d\ e \\ \updownarrow\updownarrow\updownarrow\updownarrow\updownarrow \\ B: 3\ 5\ 1\ 2\ 4 \end{array} \tag{6.14}$$

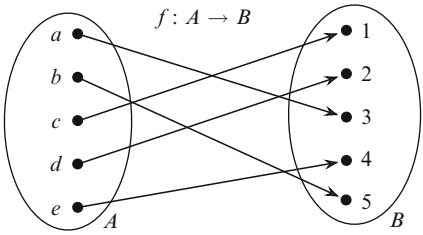


Fig. 6.11 Arrow diagram of a one-to-one and onto function

Thus, the function f allows us to set up a pairing between the elements in A and the elements in B such that each element in A is matched with exactly one element in B and each element in B is thereby matched with exactly one element in A. Cantor observed that we can now conclude that the sets A and B have the same size.

For another example, let $E = \{k \in \mathbb{N} : k \text{ is even}\}$ and let $f: \mathbb{N} \to E$ be defined by $f(n) = 2n$. Since f is one-to-one and onto, we obtain the following one-to-one correspondence between the set $\mathbb{N}$ of natural numbers and the set E of even natural numbers:

$$\begin{array}{l} \mathbb{N}: 1\ 2\ 3\ 4\ 5\ \ 6\ \cdots\ n\ \cdots \\ \quad\ \updownarrow\updownarrow\updownarrow\updownarrow\updownarrow\ \updownarrow\ \cdots\ \updownarrow\ \cdots \\ E: 2\ 4\ 6\ 8\ 10\ 12\ \cdots\ 2n\ \cdots \end{array}$$

Consequently, each natural number n corresponds to the even number $2n$, and each even natural number $2i$ is thereby matched with i. The bijection $f: \mathbb{N} \to E$ specifies a one-to-one match-up between the elements in $\mathbb{N}$ and the elements in E. Cantor concluded that the set $\mathbb{N}$ and E have the same size.

After discovering how to determine if two infinite sets have the same size, Cantor was able to prove that the set $\mathbb{Q}$ of rational numbers has the same size as the set $\mathbb{N}$ of natural numbers. As a result, Cantor conjectured that the set of real numbers also has the same size as the set $\mathbb{N}$. It came as a complete surprise to Georg Cantor in 1874 when he discovered that these two infinite sets have different sizes. In fact, Cantor showed that the set of real numbers is much larger than the set of natural numbers. This completely unexpected result would have an enormous impact on the future of mathematics.

When Cantor first presented his research on the size of infinite sets, a few of his contemporaries actually refused to accept his discoveries. Henri Poincaré referred to Cantor's work as a "disease" which would infect mathematics. Nevertheless, Cantor and his important ideas would eventually be recognized. In 1904, the Royal Society presented Cantor with its Sylvester Medal "for his brilliant" mathematical research. David Hilbert, a very influential mathematician, described Cantor's work as

> the finest product of mathematical genius and one of the supreme achievements of purely intellectual human activity.

In this section we will investigate Cantor's early work in set theory. We shall first formally define the notion of a finite set and then we will classify sets into

two categories: sets whose elements can be enumerated in a sequence (countable sets) and those sets for which it is impossible to enumerate all of their elements (uncountable sets). In mathematics, an uncountable set is an infinite set that is just too big to be countable.

6.5.1 Countable Sets

Clearly, a set is *infinite* if it is not finite. Moreover, a set is finite if it has at most n many elements for some natural number n. The following definition is just a rephrasing of this notion.

Definition 6.5.1. A set X is said to be **finite** when there is a one-to-one function $f: X \to \{1,2,3,\ldots,n\}$ for some natural number n.

It should be noted that the empty set is also considered to be finite. Now that we have a formal definition of what it means for a set to be finite, we can use the Well-Ordering Principle 4.1.1 to precisely define the notion of the "number of elements" in a finite set.

Definition 6.5.2. Suppose that X is a nonempty finite set. Let n be the least natural number such that there is a one-to-one function $f: X \to \{1,2,3,\ldots,n\}$. Then n is the **number of elements** in the set X and we write $|X| = n$.

Let X, f and n be as in Definition 6.5.2. Then one can prove that the function f must be onto (see Exercise 7). Our next theorem confirms that two nonempty finite sets have the same number of elements if and only if there is a one-to-one correspondence between the two sets.

Theorem 6.5.3. *Let A and B be nonempty finite sets and let $n,m \in \mathbb{N}$ be such that $|A| = n$ and $|B| = m$. There exists a one-to-one and onto function $h: A \to B$ if and only if $n = m$.*

Proof. See Exercise 8. □

The following theorem just shows that our definition of the number of elements in a finite set yields exactly what one would expect.

Theorem 6.5.4. *Let A be a set and let n be a natural number. If $f: A \to \{1,2,\ldots,n\}$ is one-to-one and onto, then $|A| = n$.*

Proof. Theorem 6.5.3 and Exercise 9 easily imply the result. □

Recall that $\mathbb{N} = \{1,2,3,\ldots\}$ is the set of natural numbers. A set is countable if it has the same "number of elements" as some subset of $\mathbb{N}$. In other words, a set is *countable* if there is a one-to-one correspondence between the set and a subset of $\mathbb{N}$. Our next definition captures this concept in mathematical terms.

Definition 6.5.5. A set X is **countable** if there is a one-to-one function $f: X \to \mathbb{N}$.

Hence, every finite set is countable and each subset S of $\mathbb{N}$ is countable because the identity function $\mathfrak{i}\colon S \to \mathbb{N}$ is one-to-one, where $\mathfrak{i}(x) = x$ for all $x \in S$. Thus, in particular, the set of natural numbers $\mathbb{N}$ is countable. In this section, we shall show that the set of integers and the set of rational numbers are also countable. We will then prove in Section 6.5.2 that the set of real numbers is not countable.

Definition 6.5.6. A set X is **countably infinite** if it is countable and infinite.

The set $\mathbb{N}$ is countably infinite as it is countable and it is infinite (see Exercise 10). When one can prove that a set is countable and it is clear that it is infinite, then we can conclude that the set is countably infinite. We will show in our next theorem that the set $\{0,1,2,3,\dots\}$ is countable and hence, this set countably infinite.

Theorem 6.5.7. *The set $A = \{0,1,2,3,\dots\}$ of non-negative integers is countable.*

Proof. Define $f\colon A \to \mathbb{N}$ by $f(n) = n+1$. Since f is one-to-one, A is countable. □

Theorem 6.5.8. *The set of integers $\mathbb{Z}$ is countable.*

Proof. Define the function $f\colon \mathbb{Z} \to \mathbb{N}$ be

$$f(n) = \begin{cases} 2^n, & \text{if } n \geq 0; \\ 3^{-n}, & \text{if } n < 0. \end{cases} \tag{6.15}$$

We prove that $f\colon \mathbb{Z} \to \mathbb{N}$ is one-to-one. Let i, j be integers. Assume $f(i) = f(j)$. We prove that $i = j$. Since $f(i) = f(j)$, it follows from (6.15) that we cannot have $i \geq 0$ and $j < 0$. To see why, suppose $i \geq 0$ and $j < 0$. Since $f(i) = f(j)$, definition (6.15) implies that $(*)$ $2^i = 3^{-j}$ where $i \geq 0$ and $-j > 0$. If $i > 0$, then equation $(*)$ contradicts Theorem 4.7.7, the fundamental theorem of arithmetic. If $i = 0$, then $(*)$ leads to $1 = 3^{-j}$, which is impossible because $-j \neq 0$. Similarly, we cannot have $i < 0$ and $j \geq 0$. Thus, either (a) $i \geq 0$ and $j \geq 0$, or (b) $i < 0$ and $j < 0$. Because $f(i) = f(j)$, we see that if (a) holds, then $2^i = 2^j$ and Theorem 4.7.7 implies $i = j$. If (b) holds, then $3^{-i} = 3^{-j}$. Hence $-i = -j$ and so, $i = j$. Therefore, f is one-to-one and $\mathbb{Z}$ is countable. □

Theorem 6.5.9. *Let A and B be sets where B is countable. If there is a one-to-one function $g\colon A \to B$, then A is countable.*

Proof. Let A and B be sets. Suppose B is countable and that $g\colon A \to B$ is one-to-one. We shall prove that A is countable. Since B is countable, there is a one-to-one function $f\colon B \to \mathbb{N}$. By Theorem 6.3.6, we have that $(f \circ g)\colon A \to \mathbb{N}$ is one-to-one. We conclude that A is countable. □

Theorem 6.5.10. *Suppose that B is a countable set and $A \subseteq B$. Then A is countable.*

Proof. Suppose B is countable and $A \subseteq B$. Let $\mathfrak{i}\colon A \to B$ be the identity function, that is, $\mathfrak{i}(x) = x$ for all $x \in A$. Since $\mathfrak{i}$ is one-to-one, Theorem 6.5.9 implies that A is countable. □

Theorem 6.5.11. *Suppose that A and B are countable sets. Then $A \cup B$ is countable.*

Proof. Because A and B are countable, there are one-to-one functions $f: A \to \mathbb{N}$ and $g: B \to \mathbb{N}$. Now, define the function $h: A \cup B \to \mathbb{N}$ by

$$h(x) = \begin{cases} 2^{f(x)}, & \text{if } x \in A; \\ 3^{g(x)}, & \text{if } x \in B \setminus A, \end{cases} \tag{6.16}$$

for each $x \in A \cup B$. We prove that $h: A \cup B \to \mathbb{N}$ is one-to-one. Let $x, y \in A \cup B$ and assume $h(x) = h(y)$. We shall prove that $x = y$. First, because $h(x) = h(y)$, we cannot have $x \in A$ and $y \in B \setminus A$. To see this, suppose that $x \in A$ and $y \in B \setminus A$. Since $h(x) = h(y)$, definition (6.16) implies that $2^{f(x)} = 3^{g(y)}$ and this contradicts the fundamental theorem of arithmetic. Similarly, we cannot have $y \in A$ and $x \in B \setminus A$. Thus, we must have either $x, y \in A$ or $x, y \in B \setminus A$. If $x, y \in A$, then $2^{f(x)} = 2^{f(y)}$. Theorem 4.7.7 implies that $f(x) = f(y)$ and hence $x = y$, because f is one-to-one. If $x, y \in B \setminus A$, then $2^{g(x)} = 2^{g(y)}$. We conclude that $g(x) = g(y)$. As g is one-to-one, we have $x = y$. Therefore, h is one-to-one and $A \cup B$ is countable. □

Using Theorem 6.5.11, one can prove by mathematical induction that a finite union of countable sets is countable. We shall not do this here, as we will soon prove a more general result (see Corollary 6.5.19).

Before we prove that the set of rational numbers is countable, we shall first prove that sets $\mathbb{Q}^+$ and $\mathbb{Q}^-$ are countable.

Lemma 6.5.12. *The set of positive rational numbers $\mathbb{Q}^+$ is countable.*

Proof. Define $f: \mathbb{Q}^+ \to \mathbb{N}$ by $f(\frac{m}{n}) = 2^m 3^n$ for each $\frac{m}{n} \in \mathbb{Q}^+$ in reduced form, where m and n are natural numbers.[2] We prove that f is one-to-one. Let $\frac{a}{b}, \frac{c}{d} \in \mathbb{Q}^+$ be in reduced form. Assume $f(\frac{a}{b}) = f(\frac{c}{d})$. Thus, $2^a 3^b = 2^c 3^d$. Theorem 4.7.7 implies that $a = c$ and $b = d$. Hence $\frac{a}{b} = \frac{c}{d}$. Therefore, f is one-to-one and $\mathbb{Q}^+$ is countable. □

Lemma 6.5.13. *The set of negative rational numbers $\mathbb{Q}^-$ is countable.*

Proof. Define $f: \mathbb{Q}^- \to \mathbb{Q}^+$ by $f(q) = -q$ for each $q \in \mathbb{Q}^-$. Clearly, f is one-to-one. Since $\mathbb{Q}^+$ is countable, we have that $\mathbb{Q}^-$ is countable by Theorem 6.5.9. □

Theorem 6.5.14. *The set of rational numbers $\mathbb{Q}$ is countable.*

Proof. Clearly $\mathbb{Q} = \mathbb{Q}^- \cup \{0\} \cup \mathbb{Q}^+$ and so, $\mathbb{Q}$ is the union of three sets. From Lemmas 6.5.12 and 6.5.13, we conclude that each set in this union is countable. Therefore, $\mathbb{Q}$ is countable by Theorem 6.5.11. □

Our next result shows that any countably infinite set can be put into a one-to-one correspondence with the set of natural numbers.

[2] There is only one way that such a positive rational number can be written in reduced form (see Definition 3.8.7 and Exercise 6 on page 141). Thus, f is well-defined.

Theorem 6.5.15. *If A is a countably infinite set, then there is a function $g\colon \mathbb{N} \to A$ that is one-to-one and onto.*

Proof. Assume A is a countably infinite set. Thus, there is a function $f\colon A \to \mathbb{N}$ that is one-to-one. Since A is not finite, the range of f must be an infinite subset of $\mathbb{N}$. Let $R = \text{ran}(f)$. Since R is an infinite set of natural numbers, we can list the elements of R in strictly increasing order, say $R = \{n_1, n_2, n_3, \dots\}$ where $n_i < n_{i+1}$ for all $i \in \mathbb{N}$. So, $(\star)$ $n_1 < n_2 < n_3 < \cdots$, where n_1 is the first element in R, n_2 is the second element in R, etc.[3] Define $g\colon \mathbb{N} \to A$ as follows:

$$g(i) = a \text{ if and only if } f(a) = n_i. \tag{6.17}$$

Since f is one-to-one, it follows that g is a function.[4] To prove that g is one-to-one, let i, j be natural numbers. Assume that $g(i) = g(j)$ and let $a \in A$ be this common value. So, $g(i) = a$ and $g(j) = a$. It follows from (6.17) that $f(a) = n_i$ and $f(a) = n_j$. Hence, $n_j = n_j$. Because the ordering $(\star)$ of R is without repetition, we conclude that $i = j$. Therefore, g is one-to-one. To prove that g is onto, let $a \in A$. Since $f(a) \in R$, let $n_j \in R$ be so that $f(a) = n_j$. Thus, $g(j) = a$ and so, g is onto. □

Corollary 6.5.16. *A set A is countably infinite if and only if there is a one-to-one and onto function $g\colon \mathbb{N} \to A$.*

Proof. If A is a countably infinite set, then Theorem 6.5.15 implies there is one-to-one and onto function $g\colon \mathbb{N} \to A$. Conversely, suppose that there is a one-to-one and onto function $g\colon \mathbb{N} \to A$. Thus, A is infinite (see Exercise 10). Theorem 6.2.12 asserts the existence of the inverse function $g^{-1}\colon A \to \mathbb{N}$. Furthermore, Theorem 6.2.14 implies that $g^{-1}\colon A \to \mathbb{N}$ is one-to-one. Therefore, A is countable and infinite. □

Theorem 6.5.17. *Suppose that A is a countably infinite set. Then there exists an enumeration $a_1, a_2, a_3, \dots, a_n, \dots$ of all of the elements in A such that every element in A appears in this enumeration exactly once.*

Proof. Suppose A is countably infinite. Theorem 6.5.15 implies there is a function $g\colon \mathbb{N} \to A$ that is one-to-one and onto. For each $n \geq 1$, let $a_n = g(n)$. Since g is one-to-one and onto, it follows that the enumeration $a_1, a_2, a_3, \dots, a_n, \dots$ lists every element in A exactly once. □

As a result of Theorem 6.5.17, countably infinite sets are said to be *denumerable*; that is, we can list the elements of a denumerable set in the same way that we list the natural numbers, namely, $1, 2, 3, 4, 5, \dots$. We will soon show that it is impossible to list all of the real numbers in such a manner. In other words, $\mathbb{R}$ is not denumerable.

Theorem 6.5.18. *If $\{A_i : i \in \mathbb{N}\}$ is a family of countable sets, then $\bigcup_{i \in \mathbb{N}} A_i$ is countable.*

---

[3] The sequence $n_1, n_2, n_3, \dots$ can be defined by recursion using the Well-Ordering Principle 4.1.1.

[4] $g(i)$ is the unique element $a \in A$ such that $f(a) = n_i$, the i-th element in R. See the proof of Theorem 6.2.12.

Proof. Suppose $\{A_i : i \in \mathbb{N}\}$ is a family of countable sets. Since A_i is countable, there is a one-to-one function $f_i : A_i \to \mathbb{N}$ for each $i \in \mathbb{N}$.[5] Now consider the infinite list of all the primes in strictly increasing order $p_1 < p_2 < p_3 < \cdots < p_i < \cdots$ where $p_1 = 2$, $p_2 = 3$, etc. Define the function $g : \bigcup_{i\in\mathbb{N}} A_i \to \mathbb{N}$ by

$$g(x) = p_i^{f_i(x)} \text{ where } i \text{ is the least } i \in \mathbb{N} \text{ such that } x \in A_i$$

for all $x \in \bigcup_{i\in\mathbb{N}} A_i$. We shall prove that g is one-to-one. Let $x, y \in \bigcup_{i\in\mathbb{N}} A_i$ and assume $g(x) = g(y)$. Let $i \in \mathbb{N}$ be the least natural number such that $x \in A_i$ and let $j \in \mathbb{N}$ be the least such that $y \in A_j$. As $g(x) = g(y)$, we have $p_i^{f_i(x)} = p_j^{f_j(x)}$. It follows, from Theorem 4.7.7, that $p_i = p_j$ and $f_i(x) = f_j(y)$. Since $p_i = p_j$, we must have that $i = j$. Thus, $f_i(x) = f_i(y)$. We conclude that $x = y$, because f_i is one-to-one. Therefore, g is one-to-one and $\bigcup_{i\in\mathbb{N}} A_i$ is countable. □

We can now prove that a countable union of countable sets is countable. Thus using countable sets one can construct many more countable sets.

Corollary 6.5.19. *Let J be a nonempty countable set. If $\{B_j : j \in J\}$ is a family of countable sets, then $\bigcup_{j\in J} B_j$ is countable.*

Proof. Let J be a nonempty countable set and assume $\{B_j : j \in J\}$ is a family of countable sets. Thus, B_j is countable for each $j \in J$. Since J is countable, there is a one-to-one function $h : J \to \mathbb{N}$. Let $k \in J$ be fixed. Consider the family $\{A_i : i \in \mathbb{N}\}$ defined by

$$A_i = \begin{cases} B_j, & \text{if } i \in \operatorname{ran}(h) \text{ and } h(j) = i; \\ B_k, & \text{if } i \notin \operatorname{ran}(h), \end{cases} \tag{6.18}$$

for each $i \in \mathbb{N}$. Therefore, A_i is countable for all $i \in \mathbb{N}$ and $\bigcup_{j\in J} B_j = \bigcup_{i\in\mathbb{N}} A_i$. Theorem 6.5.18 now implies that $\bigcup_{j\in J} B_j$ is countable. □

Since a finite set is countable, Corollary 6.5.19 implies that a finite union of countable sets is also countable. This completes our introduction to countable sets.

6.5.2 Uncountable Sets

In the previous section we showed that there are many countable sets. Are there any sets that are not countable? In this section we will show that there are such sets; but first, we identify a slightly easier way to say that a set is "not countable."

[5] We are tacitly using the axiom of choice to obtain the choice set $\{f_i : i \in \mathbb{N}\}$ for the family $\{F_i : i \in \mathbb{N}\}$, where F_i is the set of one-to-one functions $f_i : A_i \to \mathbb{N}$ for each $i \in \mathbb{N}$ (see page 167).

Definition 6.5.20. A set is **uncountable** if it is not countable.

An uncountable set is an infinite set that is much "bigger" than any countable set. Cantor was the first mathematician to discover and prove that uncountable sets exist. In his proof, Cantor introduced a new and very clever proof technique which is often referred to as a *diagonal argument*. This argument has had a profound influence on mathematics ever since its introduction. The proof of our next theorem illustrates Cantor's argument. When you read this proof, see if you can find the *diagonal*.

Theorem 6.5.21. *Let $S = \{0,1,2,\ldots,8\}$. Consider the set $\mathcal{F}$ consisting of all the functions $f\colon \mathbb{N} \to S$, that is, let $\mathcal{F} = \{f \mid f\colon \mathbb{N} \to S\}$. Then the set $\mathcal{F}$ is uncountable.*

Proof. Let S and $\mathcal{F}$ be as stated in the theorem. We prove that $\mathcal{F}$ is uncountable. Suppose, for a contradiction, that $\mathcal{F}$ is countable. Since the set $\mathcal{F}$ is infinite (see Exercise 19), Theorem 6.5.17 implies that there is an enumeration

$$f_1, f_2, f_3, \ldots, f_n, \ldots \tag{6.19}$$

of all of the functions in $\mathcal{F}$; that is, every function in $\mathcal{F}$ appears in the list (6.19). Define the function $g\colon \mathbb{N} \to S$ by

$$g(i) = \begin{cases} f_i(i) + 1, & \text{if } f_i(i) < 8; \\ 0, & \text{if } f_i(i) = 8, \end{cases} \tag{6.20}$$

for each $i \in \mathbb{N}$. Since $g\colon \mathbb{N} \to S$, we see that $g \in \mathcal{F}$. Since each function in $\mathcal{F}$ appears in the list (6.19), the function g is in this list. So there is an $n \in \mathbb{N}$ such that $g = f_n$. Thus, $g(i) = f_n(i)$ for all $i \in \mathbb{N}$. Consequently, $(\star)$ $g(n) = f_n(n)$. Since $f_n\colon \mathbb{N} \to S$, either $f_n(n) < 8$ or $f_n(n) = 8$. If $f_n(n) < 8$, then $g(n) = f_n(n) + 1$ by (6.20). In addition, we have that $g(n) = f_n(n)$ by $(\star)$. We conclude that $f_n(n) + 1 = f_n(n)$. Hence $1 = 0$, which is a contradiction. If $f_n(n) = 8$, then $g(n) = 0$ by (6.20). Moreover, we have that $g(n) = f_n(n)$ by $(\star)$. Thus $0 = 8$, which is also a contradiction. Therefore, $\mathcal{F}$ is not countable and thus, $\mathcal{F}$ is uncountable. □

Where Is the Diagonal?

One may wonder why the technique used in the proof of Theorem 6.5.21 is referred to as a diagonal argument. To answer this inquiry, we shall now revisit this proof. Given a function $f\colon \mathbb{N} \to S$ there is a way of writing the values of f as an infinite sequence of terms from the set $\{0,1,2,\ldots 8\}$. For example suppose $f(i) = 3$ if i is even and $f(i) = 5$ of i is odd. So

$$f(1) = 5,\ f(2) = 3,\ f(3) = 5,\ f(4) = 3,\ \ldots$$

and we can represent f as follows:

$$f = \langle 5, 3, 5, 3, 5, 3, 5, \ldots \rangle .$$

Furthermore if you are told that $h\colon \mathbb{N} \to S$ is represented by

$$h = \langle 4, 2, 6, 3, 8, 0, 1, \ldots \rangle,$$

then you know that

$$h(1) = 4,\ h(2) = 2,\ h(3) = 6,\ h(4) = 3,\ h(5) = 8,\ h(6) = 0,\ h(7) = 1,\ \ldots.$$

Consider the list (6.19) of functions in $\mathcal{F}$. Let us represent each f_i in this list as a sequence, that is, let

$$f_i = \langle f_i(1),\ f_i(2),\ f_i(3),\ f_i(4),\ f_i(5),\ f_i(6),\ \ldots \rangle.$$

Using this notation, we shall now rewrite the list of functions (6.19) in the following vertical form:

$$\begin{aligned}
f_1 &= \langle f_1(1),\ f_1(2),\ f_1(3),\ f_1(4),\ f_1(5),\ f_1(6),\ \ldots \rangle \\
f_2 &= \langle f_2(1),\ f_2(2),\ f_2(3),\ f_2(4),\ f_2(5),\ f_2(6),\ \ldots \rangle \\
f_3 &= \langle f_3(1),\ f_3(2),\ f_3(3),\ f_3(4),\ f_3(5),\ f_3(6),\ \ldots \rangle \\
f_4 &= \langle f_4(1),\ f_4(2),\ f_4(3),\ f_4(4),\ f_4(5),\ f_4(6),\ \ldots \rangle \\
f_5 &= \langle f_5(1),\ f_5(2),\ f_5(3),\ f_5(4),\ f_5(5),\ f_5(6),\ \ldots \rangle \\
f_6 &= \langle f_6(1),\ f_6(2),\ f_6(3),\ f_6(4),\ f_6(5),\ f_6(6),\ \ldots \rangle \\
&\vdots
\end{aligned} \tag{6.21}$$

In the proof of Theorem 6.5.21, we defined a function $g\colon \mathbb{N} \to S$ that is not equal to any function in the list (6.21). This is done by going down this list and assigning a value to $g(i)$ that is different from the diagonal value $f_i(i)$ for each f_i appearing in (6.21). To illustrate this idea, let us give some specific values to the entries that can appear in the diagonal of (6.21). Suppose $f_1(1) = 6$, $f_2(2) = 4$, $f_3(3) = 8$, $f_4(4) = 7$, $f_5(5) = 1$ and $f_6(6) = 0$. Thus, (6.21) becomes:

$$\begin{aligned}
f_1 &= \langle \underline{\ 6\ },\ f_1(2),\ f_1(3),\ f_1(4),\ f_1(5),\ f_1(6),\ \ldots \rangle \\
f_2 &= \langle f_2(1),\ \underline{\ 4\ },\ f_2(3),\ f_2(4),\ f_2(5),\ f_2(6),\ \ldots \rangle \\
f_3 &= \langle f_3(1),\ f_3(2),\ \underline{\ 8\ },\ f_3(4),\ f_3(5),\ f_3(6),\ \ldots \rangle \\
f_4 &= \langle f_4(1),\ f_4(2),\ f_4(3),\ \underline{\ 7\ },\ f_4(5),\ f_4(6),\ \ldots \rangle \\
f_5 &= \langle f_5(1),\ f_5(2),\ f_5(3),\ f_5(4),\ \underline{\ 1\ },\ f_5(6),\ \ldots \rangle \\
f_6 &= \langle f_6(1),\ f_6(2),\ f_6(3),\ f_6(4),\ f_6(5),\ \underline{\ 0\ },\ \ldots \rangle \\
&\vdots \\
g &= \langle \underline{\ 7\ },\ \underline{\ 5\ },\ \underline{\ 0\ },\ \underline{\ 8\ },\ \underline{\ 2\ },\ \underline{\ 1\ },\ \ldots \rangle
\end{aligned} \tag{6.22}$$

We have put the function g below the infinite list (6.22) where the values of g are determined by applying definition (6.20) given in the proof of Theorem 6.5.21. For example, to evaluate $g(1)$ we see that $f_1(1) = 6 < 8$ and so, $g(1) = 7$ by (6.20). Thus, $g(1) \neq f_1(1)$ and we are thereby assured that $g \neq f_1$. Now we evaluate $g(2)$. Since $f_2(2) = 4 < 8$, we obtain $g(2) = 5$. So $g(2) \neq f_2(2)$ and hence, $g \neq f_2$. Again, because $f_3(3) = 8$, we obtain $g(3) = 0$ and $g \neq f_3$. Continuing in this manner we construct a function $g\colon \mathbb{N} \to S$ that is different from every function in the list (6.22). This is the clever diagonal argument that Cantor introduced to mathematics.

Theorem 6.5.22. *Let A and B be sets. Suppose that A is uncountable and $g\colon A \to B$ is a one-to-one function. Then B is uncountable.*

Proof. Assume A is uncountable and that $g\colon A \to B$ is one-to-one. We shall prove that B is uncountable. Suppose, for a contradiction, that B is countable. Since $g\colon A \to B$ is one-to-one, Theorem 6.5.9 implies that A is countable. This contradicts our assumption that A is uncountable. Therefore, B is uncountable. □

Before we prove our next theorem, we make an observation. Suppose that $f\colon \mathbb{N} \to \{0,1,2,\ldots,8\}$ is a function. Thus, $0 \leq f(n) \leq 8$ for all $n \in \mathbb{N}$. So we can use f to define a real number by means of an infinite decimal expansion. Let $f_n = f(n)$ for each $n \in \mathbb{N}$. Then we have the real number given by the infinite decimal expansion $0.f_1f_2f_3f_4\cdots f_n\cdots$. For example, suppose $f(1) = 2$, $f(2) = 4$, $f(3) = 1$, $f(4) = 8, \ldots$. Then

$$0.f_1f_2f_3f_4\cdots f_n\cdots = 0.2418\cdots f_n\cdots .$$

Theorem 6.5.23. *The set of real numbers $\mathbb{R}$ is uncountable.*

Proof. Let $S = \{0,1,2,\ldots,8\}$ and let $\mathcal{F} = \{f \mid f\colon \mathbb{N} \to S\}$. By Theorem 6.5.21, we know that $\mathcal{F}$ is uncountable. For each $f \in \mathcal{F}$ let us write $(\star)$ $f_n = f(n)$ for all $n \in \mathbb{N}$. So f_n is a natural number satisfying $0 \leq f_n \leq 8$ for every $n \in \mathbb{N}$. Define the function $G\colon \mathcal{F} \to \mathbb{R}$ by

$$G(f) = 0.f_1f_2f_3\cdots \tag{6.23}$$

for each $f \in \mathcal{F}$. We will prove that G is one-to-one. Let f and h be functions in $\mathcal{F}$. Assume $G(f) = G(h)$. We must prove that $f = h$. Since $G(f) = G(h)$, we conclude from (6.23) that

$$0.f_1f_2f_3\cdots = 0.h_1h_2h_3\cdots$$

Theorem 4.6.2 implies $f_n = h_n$ for all $n \in \mathbb{N}$. From $(\star)$, we see that $f(n) = h(n)$ for all $n \in \mathbb{N}$. Therefore $f = h$ and thus, G is one-to-one. Since $G\colon \mathcal{F} \to \mathbb{R}$ is one-to-one and $\mathcal{F}$ is uncountable, Theorem 6.5.22 implies that $\mathbb{R}$ is uncountable. □

6.5.3 Cardinality

The cardinality of a set is a measure of how many elements are in the set. In particular, the set $A = \{1, 2, 3, \ldots, 25\}$ contains 25 elements and so, the cardinality of A is 25. We let $|A|$ denote the cardinality of A and thus, $|A| = 25$. The cardinality of an infinite set X is also denoted by $|X|$ and we will present a method for measuring the size of the set X that does not rely on numbers. There are examples, as we will see, of two infinite sets where one of these sets has cardinality much larger than that of the other infinite set. In other words, it is possible for one infinite set to have "many more" elements than another infinite set.

What does it mean to say that two sets have the same cardinality, that is, the same size? Georg Cantor discovered a mathematically precise and simple answer to this question.

Definition 6.5.24. Let A and B be sets. Then A has the **same cardinality** as B, denoted by $|A| = |B|$, if there is a function $f : A \to B$ that is one-to-one and onto.

Remark 6.5.25. Unfortunately, the expression $|A| = |B|$ looks like an equation; however, the assertion $|A| = |B|$ should be viewed only as an abbreviation for the statement "A has the same cardinality as B." In other words, $|A| = |B|$ means that "there is a function $f : A \to B$ that is one-to-one and onto."

What does it mean to say that one set has smaller cardinality than another set? Cantor found a simple answer to this question, as well.

Definition 6.5.26. Let A and B be sets. We say that A has **cardinality strictly less** than that of B, denoted by $|A| < |B|$, if there is a one-to-one function $f : A \to B$ and there is no function $g : A \to B$ that is both one-to-one and onto.

Using our cardinality notation, we will summarize some of the results that were previously established about countable and uncountable sets.

Theorem 6.5.27 (Cantor). $|\mathbb{N}| = |\mathbb{Z}|$ *and* $|\mathbb{N}| = |\mathbb{Q}|$.

Proof. Theorems 6.5.8 and 6.5.14 imply that $\mathbb{Z}$ and $\mathbb{Q}$ are countably infinite. By Theorem 6.5.15, there are functions $f : \mathbb{N} \to \mathbb{Z}$ and $g : \mathbb{N} \to \mathbb{Q}$ that are one-to-one and onto. Therefore, $|\mathbb{N}| = |\mathbb{Z}|$ and $|\mathbb{N}| = |\mathbb{Q}|$. □

Theorem 6.5.28 (Cantor). $|\mathbb{N}| < |\mathbb{R}|$.

Proof. Consider the function $f : \mathbb{N} \to \mathbb{R}$ defined by $f(n) = n$. This function is one-to-one. We now show that there is no function $g : \mathbb{N} \to \mathbb{R}$ that is one-to-one and onto. Suppose, for a contradiction, there is such a function g. Then $g^{-1} : \mathbb{R} \to \mathbb{N}$ would be one-to-one by Theorem 6.2.14 and thus, $\mathbb{R}$ would be countable. This contradicts Theorem 6.5.23. Therefore, $|\mathbb{N}| < |\mathbb{R}|$. □

Recalling Definition 5.1.3, for any set A the power set $\mathcal{P}(A)$ is the set of all subsets of A, that is, $\mathcal{P}(A) = \{X : X \subseteq A\}$. For example, consider the set $A = \{a, b\}$. Then $\mathcal{P}(A) = \{\varnothing, \{a\}, \{b\}, \{a, b\}\}$ and we observe that $|A| < |\mathcal{P}(A)|$. One can prove

that if A is any finite set with n many elements, then $\mathcal{P}(A)$ has 2^n many elements. Thus, $|A| < |\mathcal{P}(A)|$ whenever A is a finite set. But what happens if A is an infinite set? Cantor answered this intriguing question as well, using his diagonalization argument.

Theorem 6.5.29 (Cantor). *Let A be any set. Then $|A| < |\mathcal{P}(A)|$.*

Proof. Let A be a set. Consider the function $f: A \to \mathcal{P}(A)$ defined by $f(a) = \{a\}$ for all $a \in A$. It is easy to show that f is one-to-one. We now show that there is no function $g: A \to \mathcal{P}(A)$ that is one-to-one and onto. Suppose, for a contradiction, that there is a bijection $g: A \to \mathcal{P}(A)$. Observe that $g(x) \subseteq A$ for all $x \in A$. Let $X = \{x \in A : x \notin g(x)\}$. Clearly, $X \subseteq A$ and thus $X \in \mathcal{P}(A)$. Since g is onto, there is an $a \in A$ such that $g(a) = X$. There are two cases to consider, namely, either $a \in X$ or $a \notin X$. For the first case, assume that $a \in X$. Then, from the definition of X, we conclude that $a \notin g(a)$. Since $g(a) = X$, we have that $a \notin X$ which contradicts our assumption. For the other case, assume that $a \notin X$. Thus, $a \in g(a)$ by the definition of X. As $g(a) = X$, we deduce that $a \in X$ which is again a contradiction. So, there is no $g: A \to \mathcal{P}(A)$ that is one-to-one and onto. Therefore, $|A| < |\mathcal{P}(A)|$. □

Definition 6.5.30. Let A and B be sets. We say that A has **cardinality less than or equal** to B, denoted by $|A| \leq |B|$, if there is a function $f: A \to B$ that is one-to-one.

Our next theorem is very useful for proving many results about cardinality.

Theorem 6.5.31 (Schröder-Bernstein). *If $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

Theorem 6.5.31 states that if there are one-to-one functions $f: A \to B$ and $g: B \to A$, then there is a one-to-one *and* onto function $h: A \to B$. The theorem sounds very reasonable; however, its proof is quite challenging and even eluded the brilliant Georg Cantor. The mathematicians Ernst Schröder and Felix Bernstein discovered a proof based on standard set theory. We will not prove Theorem 6.5.31 in this book; but, we have covered all of the set-theoretic tools needed to read and understand a proof of this deep theorem (see [12, pp. 298–300]).

Our next theorem, also due to Cantor, uses the Schröder-Bernstein theorem to show that the interior of the unit square S has the same cardinality as the interior of the unit interval I (see Fig. 6.12). Thus, there are just as many points in the unit square as there are points in the unit interval. Cantor initially believed that the set of points in the two-dimensional square S must have cardinality much larger than the set of points in the one-dimensional interval I. Then he discovered a proof showing that his initial belief was wrong. This prompted Cantor to exclaim "I see it but I do not believe it" (see [4, p. 273]).

Theorem 6.5.32 (Cantor). *Let $S = (0,1) \times (0,1)$ and let $I = (0,1)$. Then $|S| = |I|$.*

Proof. We will first define a one-to-one function $f: I \to S$ and then we shall define a one-to-one function $g: S \to I$. For each $z \in I$ define $f(z) = (z, \frac{1}{2})$. It is easy to see that f is one-to-one. By Definition 6.5.30, we have that $|I| \leq |S|$. To define the function g, let $(x,y) \in S$. So, $0 < x < 1$ and $0 < y < 1$. Now, let $x = 0.x_1x_2x_3\ldots$

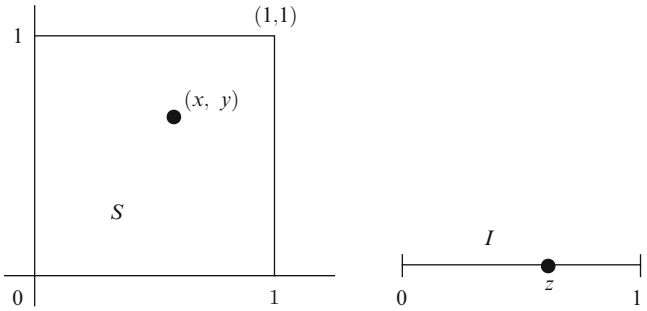


Fig. 6.12 The unit square and the unit interval

and let $y = 0.y_1y_2y_3\ldots$ be infinite decimal expansions of x and y, where $0 \leq x_i \leq 9$ and $0 \leq y_i \leq 9$ for each $i \in \mathbb{N}$. To eliminate any duplicate representations, we insist that whenever x has two decimal representations, one ending with a string 0's and the other ending with a string of 9's, we will choose the one ending with a string 0's (see Remark 4.6.3 on page 130). Similarly, we will not use a decimal expansion for y that ends in a string of 9's. Define $g(x,y) = 0.x_1y_1x_2y_2x_3y_3\cdots$. One can show that g is one-to-one (see Exercise 3). Thus, $|S| \leq |I|$. Theorem 6.5.31 now implies that $|S| = |I|$. □

Exercises 6.5

1. Let $A = \{4,8,12,16,\ldots\}$ and let $B = \{n \in \mathbb{Z} : n < -25\}$.
   (a) Define a one-to-one and onto function $f : A \to \mathbb{N}$.
   (b) Define a one-to-one and onto function $g : B \to \mathbb{N}$.

2. Let A and B be as in Exercise 1. Define a one-to-one function $h : A \cup B \to \mathbb{N}$.

3. Prove that the function $g : S \to I$ defined in the proof of Theorem 6.5.32 is one-to-one.

4. Suppose that the set B is finite and $A \subseteq B$. Prove that A is finite.

5. Suppose that A and B are finite sets. Prove that $A \cup B$ is finite.

6. Suppose that A and B are finite sets. Prove that $A \times B$ is finite.

7. Let X, f and n be as in Definition 6.5.2. Prove that the function f is onto.

8. Using Definition 6.5.2, prove Theorem 6.5.3. (For the direction ($\Leftarrow$) use Exercise 7.)

9. Using Definition 6.5.2, Exercise 7 and mathematical induction, prove that $|\{1,2,\ldots,n\}| = n$ for all natural numbers n.

10. Prove that for all $n \in \mathbb{N}$ there is no one-to-one function $f : \mathbb{N} \to \{1,2,3,\ldots,n\}$. Conclude that if $g : \mathbb{N} \to A$ is one-to-one, then A is infinite.

11. Let $A = \{x \in \mathbb{R} : 0 < x < 1\}$ and $B = \{x \in \mathbb{R} : 2 < x < 5\}$. Prove that $|A| = |B|$.

12. Prove that $\mathbb{N} \times \mathbb{N}$ is countable.

13. Let A and B be countable sets. Prove that $A \times B$ is countable.

14. Let A be uncountable. Prove that $A \times B$ is uncountable for any nonempty set B.

15. Prove that there exists a function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ that is one-to-one and onto.

16. Let A be a set. Suppose that $f : \mathbb{N} \to A$ is onto. Prove that A is countable.

17. Let A and B be sets. Suppose A is uncountable and B is countable. Prove that $A \setminus B$ is uncountable.

18. Prove that the set of irrational numbers is uncountable; that is, prove that $\mathbb{R} \setminus \mathbb{Q}$ is uncountable.

19. Let $\mathcal{F}$ be as in Theorem 6.5.21. Let $(\star)$ $f_1, f_2, \ldots, f_n$ be a finite list of functions in $\mathcal{F}$. Using the argument in the proof of Theorem 6.5.21, define a new function $g \in \mathcal{F}$ that is not in the list $(\star)$. Therefore, $\mathcal{F}$ is infinite.

20. Let $S = \{\langle a_1, a_2, \ldots, a_k \rangle : a_1, a_2, \ldots, a_k \in \mathbb{N} \text{ for some } k \in \mathbb{N}\}$, the set of all finite sequences of elements from $\mathbb{N}$. Prove that S is countable.

21. Let A and B be countably infinite sets. Prove that $|A| = |B|$.

22. Let A and B be sets. Prove that if $|A| = |B|$, then $|B| = |A|$.

23. Let A, B and C be sets. Prove that if $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.

24. Let A, B and C be sets. Prove that if $|A| < |B|$ and $|B| = |C|$, then $|A| < |C|$.

25. Let A and B be sets. Suppose B is countable. Prove that if $|A| \leq |B|$, then A is also countable.

26. Let A, B and C be sets. Prove that if $|A| < |B|$ and $|B| < |C|$, then $|A| < |C|$.

27. Prove that the set $\mathcal{P}(\mathbb{N})$ is uncountable.

28. Suppose someone asserts that the set of real numbers in the interval $(0, 1)$ is countable and that all of these real numbers can be enumerated as in (6.24), where each such real number is represented by an infinite decimal expansion:

$$\begin{aligned} x_1 &= .12345689234\cdots \\ x_2 &= .68729958219\cdots \\ x_3 &= .05050506620\cdots \\ x_4 &= .57591884622\cdots \\ &\vdots \\ x_i &= .x_{i1}x_{i2}x_{i3}x_{i4}x_{i5}\cdots \\ &\vdots \end{aligned} \tag{6.24}$$

You are to show that this assertion is false. Using Cantor's diagonal argument, define a decimal expansion for a real number b in $(0, 1)$ that is not in the list (6.24). Ensure that your decimal expansion $b = .b_1b_2b_3\ldots b_i\ldots$ contains

neither of the digits 0 or 9.[6] Then identify the first 4 decimal digits in the decimal expansion of b and prove $b \neq x_n$ for all $n \in \mathbb{N}$.

29. Let $S = \{q \in \mathbb{Q} : 0 < q < 1\}$. Theorems 6.5.10 and 6.5.14 imply S is countable. We can thus enumerate all of the elements in S in a list $(\star)$ $q_1, q_2, q_3, \ldots$, by Theorem 6.5.17. Since each of these rational numbers has an infinite decimal expansion $q_i = .q_{i1}q_{i2}q_{i3}q_{i4}q_{i5}\cdots$, one can define a real number $b \in (0,1)$ that is not in the list $(\star)$, just as in Exercise 28. Is b a rational number? Justify your answer.

30. Let $A = \{a,b,c,d,w,y,z\}$ and let $g : A \to \mathcal{P}(A)$ be the function given by

$$\begin{aligned} g(a) &= \{b,c,g\} \\ g(b) &= \{a,b,c,w,z\} \\ g(c) &= \{b,c\} \\ g(d) &= \{d\} \\ g(w) &= A \\ g(y) &= \{a,b,c,d,w\} \\ g(z) &= \varnothing. \end{aligned}$$

The function g is one-to-one. The proof of Theorem 6.5.29 shows that g is not onto because the subset of A defined by $X = \{x \in A : x \notin g(x)\}$ is not in the range of g. Evaluate the set X.

Exercise Notes: For Exercise 3, use Exercise 2 on page 136. For Exercise 7, assume that f is not onto. Let $1 \leq i \leq n$ be the largest such that $i \notin \text{ran}(f)$. If $i = n$ then get a contradiction. If $i < n$, let $a \in X$ be such that $f(a) = n$. Define a one-to-one function $g : X \to \{1,2,\ldots,n-1\}$. For Exercise 9, in the inductive step suppose that $|\{1,2,\ldots,n,n+1\}| = k < n+1$. Let $f : \{1,2,\ldots,n,n+1\} \to \{1,2,\ldots,k\}$ be one-to-one and onto. Let $1 \leq j \leq k$ be such that $f(n+1) = j$ and let $1 \leq \ell \leq n$ be such that $f(\ell) = k$. Define a one-to-one function $g : \{1,2,\ldots,n\} \to \{1,2,\ldots,k-1\}$. For Exercise 10, suppose that $n \in \mathbb{N}$ is the least such that there is a one-to-one function $f : \mathbb{N} \to \{1,2,\ldots,n\}$. Show that $1 < n$ and show that f is onto. Let $i \in \mathbb{N}$ be such that $f(n) = i$. Then $h : \mathbb{N} \to \mathbb{N} \setminus \{i\}$ defined by $h(k) = i+k$ is one-to-one. Define a one-to-one function $g : \mathbb{N} \to \{1,2,\ldots,n-1\}$. For Exercise 16, let $I_a = \{n \in \mathbb{N} : f(n) = a\}$ for each $a \in A$. Since f is onto, each I_a is nonempty and has a least element. For Exercise 20, define $h : S \to \mathbb{N}$ by $h(\langle a_1, a_2, \ldots, a_m\rangle) = 2^{a_1} \cdot 3^{a_2} \cdot 5^{a_3} \cdots p_m^{a_m}$ where p_m is the m-th prime. For Exercise 28, review the proof of Theorem 6.5.21 and note that the diagonal digit x_{ii} is in the i-th decimal place for each x_i in (6.24); for example, $x_{22} = 8$ and $x_{44} = 9$.

---

[6] So b will have a unique decimal representation. Thus, if the decimal expansions of $b, x \in (0,1)$ have different digits in at least one decimal place, then $b \neq x$. See Remark 4.6.3 on page 130.

CHAPTER 7

Relations

We are already familiar with the relations $a = b$ (equality), $a < b$ (less than), $X \subseteq Y$ (subset), and $m \mid n$ (evenly divides). Many of the fundamental concepts of mathematics can be described in terms of relations. In this chapter we shall view relations as mathematical objects and explore various properties that relations may possess.

7.1 Relations on a Set

We first recall the definition of an ordered pair and that of a Cartesian product of a set with itself.

Definition 7.1.1. An **ordered pair** has the form (a,b), where a is called the first component and b is called the second component.

Example 1. $(2,3)$ is an ordered pair and so is $(3,2)$. Note that these are different ordered pairs, that is, $(2,3) \neq (3,2)$.

Definition 7.1.2. Given a set A, the **Cartesian product** $A \times A$ is defined to be

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

In other words, $A \times A$ is the set of *all* ordered pairs with first component in A and second component also in A.

Definition 7.1.3. A **relation** R on A is a subset of $A \times A$, that is, $R \subseteq A \times A$.

We will use the symbols R and $\sim$ to denote relations. Suppose that R is a relation on the set A. For $a,b \in A$, we shall write aRb to mean that $(a,b) \in R$. When we use the notation aRb, we shall say that "a is related to b." Similarly, when $\sim$ is a relation on A, we write $a \sim b$ to mean that $(a,b) \in \sim$ and say that a is related to b. We shall also write $a \not\sim b$ to assert that "a is not related to b."

Example 2. Let $A = \{2,a,b,c,3\}$. Then $R = \{(2,a), (2,b), (3,b), (3,c), (3,3)\}$ is a relation on A. Thus $3Rc$. Furthermore, $\sim = \{(b,c), (c,c), (2,3)\}$ is another relation on A. So, $b \sim c$ and $c \not\sim b$.

An ordered pair (a,b) is often viewed as an arrow $a \longrightarrow b$, going from a to b. The double arrow $a \longleftrightarrow b$ is also used to portray the ordered pairs (a,b) and (b,a), that is, $a \longrightarrow b$ and $b \longrightarrow a$. Using arrows to represent ordered pairs, we can "draw" a picture of a relation, called a **directed graph**.

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_7, 209

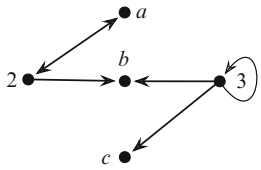


Fig. 7.1 Visualizing a relation by a directed graph

Example 3. The relation $R = \{(2,a), (a,2), (2,b), (3,b), (3,c), (3,3)\}$ defined on the set $A = \{2,a,b,c,3\}$ is represented by the directed graph in Fig. 7.1.

Given a property $P(x,y)$, there is another way to specify a relation on A; namely, let $\sim$ be the relation on A defined by $x \sim y$ if and only if $P(x,y)$, for all $x,y \in A$. Consequently, the relation $\sim$ is the set $\{(x,y) \in A \times A : P(x,y)\}$.

Example 4. Let $A = \{0,2,3,4,5,6,7,8\}$. For $x,y \in A$ define $x \sim y$ if and only if $x \,|\, y$ and $x < y$. Determine the pairs in the relation $\sim$.

Solution. Thus, $\sim = \{(x,y) \in A \times A : x \,|\, y \text{ and } x < y\}$ and we obtain

$$\sim = \{(2,4), (2,6), (2,8), (3,6), (4,8)\}. \qquad \text{Ⓢ}$$

7.1.1 Reflexive, Symmetric, and Transitive Relations

The concept of equality permeates all of mathematics. We now recognize three fundamental properties of equality. For quantities x, y, and z, we have the following:

1. $x = x$ (reflexive).
2. If $x = y$, then $y = x$ (symmetric).
3. If $x = y$ and $y = z$, then $x = z$ (transitive).

In this section we will investigate relations that share some, or all, of the above properties that hold for equality. We shall define what it means for a relation to be reflexive, symmetric, and transitive. We also present proof strategies that can be used to prove that a relation has one of these properties. These strategies will be employed throughout most of this chapter. Relations that have all three of these properties frequently appear in mathematics.

Reflexive Relations

A relation on a set A is *reflexive* if every element in the set A is related to itself.

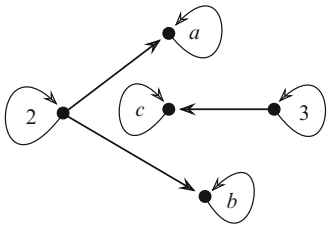


Fig. 7.2 A reflexive relation

Definition 7.1.4. A relation $\sim$ on a set A is **reflexive** if $(\forall x \in A)(x \sim x)$, that is, when $x \sim x$ for all $x \in A$.

Example 5. Let $A = \{a, b, c, 2, 3\}$. The relation

$$\sim = \{(2,a), (2,b), (3,c), (2,2), (a,a), (b,b), (3,3), (c,c)\}$$

on A is reflexive, which is represented by the directed graph in Fig. 7.2, where every element x in A has an arrow $x \longrightarrow x$ pointing to itself by means of a loop.

Proof Strategy 7.1.5. To prove that a relation $\sim$ on A is reflexive:

$$\text{Prove } (\forall x \in A)(x \sim x).$$

In other words, use the diagram

$$\begin{array}{l} \text{Let } x \in A. \\ \quad \text{Prove } x \sim x. \end{array}$$

Symmetric Relations

A relation on A is *symmetric* if whenever x is related to y, then y is also related to x.

Definition 7.1.6. A relation $\sim$ on a set A is **symmetric** when

$$(\forall x \in A)(\forall y \in A)(x \sim y \to y \sim x),$$

that is, if $x \sim y$ then $y \sim x$, for all $x, y \in A$.

Example 6. Consider the relation $\sim$ on the set $A = \{a, b, c, 2, 3\}$ defined by

$$\sim = \{(2,a), (a,2), (2,b), (b,2), (2,2), (3,b), (b,3), (3,c), (c,3), (3,3), (c,c)\}.$$

The relation $\sim$ is symmetric and is portrayed in Fig. 7.3. Observe that whenever there is an arrow $x \longrightarrow y$, then there is an arrow $y \longrightarrow x$.

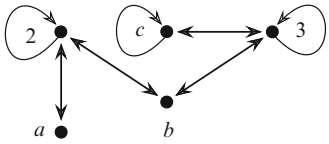


Fig. 7.3 A symmetric relation

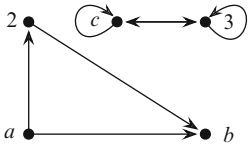


Fig. 7.4 A transitive relation

Proof Strategy 7.1.7. To prove that a relation $\sim$ on A is symmetric:

$$\text{Prove } (\forall x \in A)(\forall y \in A)(x \sim y \to y \sim x).$$

That is, use the diagram

$$\begin{array}{l} \text{Let } x, y \in A. \\ \quad \text{Assume } x \sim y. \\ \quad\quad \text{Prove } y \sim x. \end{array}$$

Transitive Relations

A relation on a set A is *transitive* if whenever x is related to y and y is related to z, then x is also related to z.

Definition 7.1.8. A relation $\sim$ on a set A is **transitive** if

$$(\forall x \in A)(\forall y \in A)(\forall z \in A)[(x \sim y \wedge y \sim z) \to x \sim z],$$

in other words, if $x \sim y$ and $y \sim z$, then $x \sim z$ for all $x, y, z \in A$.

Example 7. Draw a transitive relation on the set $A = \{a, b, c, 2, 3\}$ and then express this relation as a set of ordered pairs.

Solution. Figure 7.4 illustrates a transitive relation on A. Thus, for all $x, y, x \in A$ if there is an arrow $x \longrightarrow y$ and an arrow $y \longrightarrow z$, then there is an arrow $x \longrightarrow z$. For example, in the figure we have the arrows $a \longrightarrow 2$ and $2 \longrightarrow b$. So $a \longrightarrow b$. Since the arrows $c \longrightarrow 3$ and $3 \longrightarrow c$ appear in this graph, we also have the arrow $c \longrightarrow c$. Written as a set, the relation is $\{(a,2), (2,b), (a,b), (c,3), (3,c), (c,c), (3,3)\}$. Ⓢ

Proof Strategy 7.1.9. To prove that a relation $\sim$ on A is transitive:

$$\text{Prove } (\forall x \in A)(\forall y \in A)(\forall z \in A)[(x \sim y \wedge y \sim z) \to x \sim z].$$

In other words, use the diagram

Let $x, y, z \in A$.
  Assume $x \sim y$.
  Assume $y \sim z$.
    Prove $x \sim z$.

Example 8. Define the relation $\sim$ on the set of natural numbers $\mathbb{N}$ by

$$x \sim y \text{ if and only if } x = yk \text{ for some } k \in \mathbb{N}. \tag{7.1}$$

Determine whether or not the relation $\sim$ is reflexive, symmetric, or transitive. You must justify your answers with a proof if the relation is reflexive, symmetric, or transitive. You must provide a counterexample if the relation fails to be reflexive, symmetric, or transitive.

Solution. We have the relation $\sim$ on the set $\mathbb{N}$ defined by (7.1).

- The relation $\sim$ is reflexive.
  Proof. Let x be a natural number. Since $x = x \cdot 1$ where $1 \in \mathbb{N}$, we conclude that $x \sim x$. □
- The relation $\sim$ is not symmetric. Let $x = 6$ and $y = 2$. Since $x = y \cdot 3$, we see that $x \sim y$. We also see that $y \not\sim x$ because $2 \neq 6 \cdot k$ for any $k \in \mathbb{N}$.
- The relation $\sim$ is transitive.

  Proof. Let x, y, z be natural numbers. Assume $x \sim y$ and $y \sim z$. Thus, (1) $x = yk$ and (2) $y = zj$ for some $k, j \in \mathbb{N}$. By substituting the value for y in equation (2) into equation (1), we obtain $x = (zj)k = z(jk)$. Therefore, $x = z(jk)$ where $jk \in \mathbb{N}$. Hence, $x \sim z$. □

This completes our solution. Ⓢ

Exercises 7.1

1. Let $\sim$ be the relation on the set $A = \{0, 1, 2, 3, 4, 5\}$ defined by $a \sim b$ if and only if $a \mid (b + 1)$. Represent this relation as a set of ordered pairs.

2. Let $\sim$ be the relation on the set $\mathbb{R}$ defined by $x \sim y$ if and only if $x + y \geq 0$. Prove that this relation is symmetric. Find counterexamples showing that this relation is not reflexive and not transitive.

3. Let $\sim$ be the relation on the set $\mathbb{R}$ defined by $x \sim y$ if and only if $xy \geq 0$. Prove that this relation is reflexive and symmetric. Find a counterexample showing that this relation is not transitive.

4. The relation $\sim$ on the set $\mathbb{Z}$ is defined by $m \sim n$ if and only if $m - n$ is even. Prove that this relation is reflexive, symmetric, and transitive.

5. The relation $\sim$ on the set $\mathbb{Z}$ is defined by $m \sim n$ if and only if $m - n$ is odd. Prove that this relation is symmetric. Find counterexamples showing that this relation is not reflexive and not transitive.

6. The relation $\sim$ on the set $\mathbb{R}$ is defined by $x \sim y$ if and only if $|x| = |y|$. Prove that this relation is reflexive, symmetric, and transitive.

7. The relation $\sim$ on the set $\mathbb{Z}$ is defined by $m \sim n$ if and only if $3 \,|\, (m - n)$. Prove that this relation is reflexive, symmetric, and transitive.

8. Let the relation $\sim$ on the set $\mathbb{N}$ be defined by $m \sim n$ if and only if $m \,|\, n$. Prove that this relation is reflexive and transitive. Find a counterexample showing that this relation is not symmetric.

9. The relation $\sim$ on the set $\mathbb{R}$ is defined by $x \sim y$ if and only if $\sin(x) = \sin(y)$. Prove that this relation is reflexive, symmetric, and transitive.

7.2 Equivalence Relations and Partitions

Because the equality relation has been so useful, mathematicians have generalized this concept. A relation is called an equivalence relation if it satisfies the three key properties that are normally associated with equality.

Definition 7.2.1. A relation $\sim$ on a set A is called an **equivalence relation** if it is reflexive, symmetric, and transitive.

Thus, to prove that a relation is an equivalence relation, three distinct proofs are required; that is, one must prove that the relation is (1) reflexive, (2) symmetric, and (3) transitive. Equivalence relations are used in many areas of mathematics. An equivalence relation allows one to connect those elements of a set that have a particular property in common. Example 1, below, identifies three equivalence relations. The relation in item 1 joins the even integers and links the odd integers, the relation in item 2 associates the real numbers that have the same absolute value, and the equivalence relation in item 3 unites those integers (as we will see) that have the same remainder when divided by 3.

Example 1. One can show that each one of the relations below is an equivalence relation (see Exercises 4, 6 and 7 of Section 7.1):

1. The relation $\sim$ on the set $\mathbb{Z}$ defined by $m \sim n$ if and only if $m - n$ is even.
2. The relation $\sim$ on the set $\mathbb{R}$ defined by $x \sim y$ if and only if $|x| = |y|$.
3. The relation $\sim$ on the set $\mathbb{Z}$ defined by $m \sim n$ if and only if $3 \,|\, (m - n)$.

The main result that we will establish in the section is that an equivalence relation on a set A induces a partition of A into disjoint subsets. This will allow us to create a new mathematical object from an old one. For each $a \in A$, we must first form the set of all those elements in A that are related to a.

Definition 7.2.2. Let $\sim$ be an equivalence relation on a set A. Let a be an element in A. The **equivalence class** of a, denoted by $[a]_\sim$, is the set of all elements in A that are related to a; namely,

$$[a]_\sim = \{x \in A : x \sim a\}.$$

In Definition 7.2.2 we shall write $[a] = [a]_\sim$ when the relation $\sim$ is understood.

Example 2. Let us work with the equivalence relation $\sim$ on $\mathbb{R}$ defined by $x \sim y$ if and only if $|x| = |y|$. We evaluate the equivalence classes $[1]$, $[-1]$, and $[2]$ as follows:

$$[1] = \{x \in \mathbb{R} : x \sim 1\} = \{x \in \mathbb{R} : |x| = |1|\} = \{1, -1\}$$

$$[-1] = \{x \in \mathbb{R} : x \sim -1\} = \{x \in \mathbb{R} : |x| = |-1|\} = \{-1, 1\}$$

$$[2] = \{x \in \mathbb{R} : x \sim 2\} = \{x \in \mathbb{R} : |x| = |2|\} = \{2, -2\}.$$

Thus, $[1] = [-1]$ and $[1] \neq [2]$. In addition, observe that $[1] \cap [2] = \varnothing$.

Example 3. Let $\sim$ be the equivalence relation on $\mathbb{Z}$ defined by $m \sim n$ if and only if $m - n$ is even. Thus, $[k] = \{m \in \mathbb{Z} : m \sim k\} = \{m \in \mathbb{Z} : m - k \text{ is even}\}$ for each integer k. We evaluate the equivalence classes $[1]$, $[2]$, and $[3]$ as follows:

$$[1] = \{m \in \mathbb{Z} : m - 1 \text{ is even}\} = \{\ldots, -5, -3, -1, 1, 3, 5, 7, \ldots\}$$

$$[2] = \{m \in \mathbb{Z} : m - 2 \text{ is even}\} = \{\ldots, -6, -4, -2, 2, 4, 6, 8, \ldots\}$$

$$[3] = \{m \in \mathbb{Z} : m - 3 \text{ is even}\} = \{\ldots, -5, -3, -1, 1, 3, 5, 7, \ldots\}.$$

Thus, $[1] = [3]$ and $[3] \neq [2]$. Furthermore, $[3] \cap [2] = \varnothing$.

Remark 7.2.3. Let $\sim$ be an equivalence relation on a set A and let $a \in A$. Then $[a] \subseteq A$ and furthermore, $x \in [a]$ if and only if $x \sim a$, for each $x \in A$.

For any equivalence relation, our next theorem shows that two elements are related if and only if they have exactly the same equivalence classes.

Theorem 7.2.4. *Let $\sim$ be an equivalence relation on A. Then for all $a \in A$ and $b \in A$,*

$$a \sim b \text{ if and only if } [a] = [b].$$

Proof. Let $\sim$ be an equivalence relation on a set A and let $a, b \in A$. We shall prove that $a \sim b$ if and only if $[a] = [b]$.

($\Rightarrow$) Assume $a \sim b$. We prove that $[a] = [b]$, that is, we prove that these two sets are equal. First we prove that $[a] \subseteq [b]$. Let $x \in [a]$. We shall show that $x \in [b]$. Since $x \in [a]$, it follows that $x \sim a$. By assumption, we also have that $a \sim b$. So $x \sim a$ and $a \sim b$. Because $\sim$ is transitive, we conclude that $x \sim b$ and hence, $x \in [b]$. Therefore, $[a] \subseteq [b]$. We must prove that $[b] \subseteq [a]$. Since the argument to prove this

is very similar to the argument we just gave, this part of the proof will be left as an exercise. Therefore, $[a] = [b]$.

($\Leftarrow$) Assume $[a] = [b]$. Since $a \in [a]$, we see that $a \in [b]$. Hence, $a \sim b$. □

Corollary 7.2.5. *Let $\sim$ be an equivalence relation on a set A. Then for all $a, b \in A$, we have that $a \in [b]$ if and only if $[a] = [b]$.*

Let $\sim$ be an equivalence relation on a set A. The next theorem shows that the set of all equivalence classes forms a partition of the set A (see Definition 5.1.9). Thus, an equivalence relation on a set can be used to break up the set into nonempty subsets which do not overlap.

Theorem 7.2.6 (Fundamental Theorem of Equivalence Relations). *Let $\sim$ be an equivalence relation on a set A. The collection $P = \{[a] : a \in A\}$ is a partition of A.*

Proof. Let $\sim$ be an equivalence relation on a set A. We prove that the collection $P = \{[a] : a \in A\}$ is a partition of A. To do this, we show that

(1) for every element $x \in A$ we have that $x \in [x]$;
(2) for all $x, y \in A$, if $[x] \neq [y]$, then $[x] \cap [y] = \varnothing$.

To prove (1), let $x \in A$. Clearly, $[x] \in P$ and $x \in [x]$, as $\sim$ is reflexive. To prove (2), let $x, y \in A$ and assume $[x] \cap [y] \neq \varnothing$ (we are using proof by contraposition). Thus there is a $z \in A$ such that $z \in [x]$ and $z \in [y]$. Hence, $z \sim x$ and $z \sim y$. Since $\sim$ is symmetric, we have that $x \sim z$ and $z \sim y$. Because $\sim$ is transitive, we conclude that $x \sim y$. Theorem 7.2.4 now implies that $[x] = [y]$. Therefore, $P = \{[a] : a \in A\}$ is a partition of the set A. □

An equivalence relation $\sim$ on a set A breaks up A into disjoint subsets, as illustrated below

$$A = \left[\begin{array}{|c|c|c|c|c|} \hline \vdots & \vdots & \vdots & \vdots & \\ a & x & c & d & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ \hline \end{array}\right]$$
$$\uparrow \quad \uparrow \quad \uparrow \quad \uparrow$$
$$[a] \; [x] \; [c] \; [d] \; \cdots$$

Definition 7.2.7. Let $\sim$ be an equivalence relation on a set A. We let $A/_\sim$ denote the partition $\{[a] : a \in A\}$ of A. The partition $A/_\sim$ shall be referred to as the partition induced by $\sim$.

Mathematicians often use the partition of a set, induced by an equivalence relation, to learn something new about the set itself.

Example 4. Consider the equivalence relation $\sim$ on $\mathbb{Z}$ defined by $m \sim n$ if and only if $3 \mid (m-n)$. One can check, for any $m, n \in \mathbb{Z}$, that $m \sim n$ if and only if $m = 3k + n$ for some $k \in \mathbb{Z}$. Therefore,

$$[n] = \{m \in \mathbb{Z} : m \sim n\} = \{3k + n : k \in \mathbb{Z}\}$$

for each integer n. We can evaluate the equivalence classes $[0]$, $[1]$, and $[2]$ as follows:

$$[0] = \{3k : k \in \mathbb{Z}\} = \{\ldots, -9, -6, -3, 0, 3, 6, 9, \ldots\}$$

$$[1] = \{3k+1 : k \in \mathbb{Z}\} = \{\ldots, -8, -5, -2, 1, 4, 7, 10, \ldots\}$$

$$[2] = \{3k+2 : k \in \mathbb{Z}\} = \{\ldots, -7, -4, -1, 2, 5, 8, 11, \ldots\}.$$

The partition $\mathbb{Z}/_{\sim}$ is $\{[n] : n \in \mathbb{Z}\} = \{[0], [1], [2]\}$ (see Exercise 4 of this section) and is illustrated in our next figure:

$$\mathbb{Z} = \begin{array}{|c|c|c|} \hline \vdots & \vdots & \vdots \\ 6 & 7 & 8 \\ 3 & 4 & 5 \\ 0 & 1 & 2 \\ -3 & -2 & -1 \\ -6 & -5 & -4 \\ \vdots & \vdots & \vdots \\ \hline \uparrow & \uparrow & \uparrow \\ [0] & [1] & [2] \end{array}$$

Exercises 7.2

1. Let $\mathbb{R}^* = \{x \in \mathbb{R} : x \neq 0\}$ and define $x \sim y$ if and only if $x \cdot y > 0$, for $x, y \in \mathbb{R}^*$. Prove that $\sim$ is an equivalence relation on $\mathbb{R}^*$ and then identify the equivalence classes of $\sim$.
2. Let $\mathbb{R}^* = \{x \in \mathbb{R} : x \neq 0\}$ and let $\sim$ be the relation on $\mathbb{R}^*$ defined by $x \sim y$ if and only if $x \cdot y^{-1} \in \mathbb{Q}$. Prove that $\sim$ is an equivalence relation on $\mathbb{R}^*$.
3. Let $\mathbb{R}^* = \{x \in \mathbb{R} : x \neq 0\}$ and let $\sim$ be the equivalence relation on $\mathbb{R}^*$ defined by $x \sim y$ if and only if $x \cdot y^{-1} \in \{1, -1\}$. Identify the equivalence classes of $\sim$.
4. Let $\sim$ be the equivalence relation on $\mathbb{Z}$ defined by $m \sim n$ if and only if $3 \mid (m-n)$. Using the division algorithm (see Theorem 4.6.9), show that for every integer i we have that either $i \sim 0$, $i \sim 1$, or $i \sim 2$. Conclude that $[i] = [0]$, $[i] = [1]$, or $[i] = [2]$ for every integer i.
5. Let $\sim$ be the equivalence relation on $\mathbb{Z}$ defined by $m \sim n$ if and only if $3 \mid (m-n)$. From Exercise 4 we know that one of the equations $[i] = [0]$, $[i] = [1]$, $[i] = [2]$ is true for every integer i. Determine which of these equations is true for each of the integers $i = 4, 5, 6, -7$.
6. Let $f : A \to B$ be a function. Define a relation $\sim$ on A by $x \sim y$ if and only if $f(x) = f(y)$ for all $x, y \in A$. Prove that $\sim$ is an equivalence relation on A. Let $a \in A$ be given. Describe the equivalence class $[a]$.

7. Define the equivalence relation $\sim \mathbb{R}$ by $x \sim y$ if and only if $\sin(x) = \sin(y)$. Describe the equivalence classes $[0]$ and $[\frac{\pi}{2}]$.

8. Let $\sim$ be an equivalence relation on a set A and let $a, b, c \in A$. Prove each of the following statements directly from the definition of an equivalence relation and the definition of an equivalence class (that is, do not use the theorems presented in this section).

(a) $a \in [a]$.
(b) if $b \in [a]$, then $a \sim b$.
(c) if $b \sim c$ and $a \in [b]$, then $c \sim a$.
(d) if $b \sim a$ and $[b] = [c]$, then $[a] = [c]$.

9. Define a relation $\sim$ on $\mathbb{N} \times \mathbb{N}$ by $(a,b) \sim (c,d)$ iff $a+d = b+c$. Prove that $\sim$ is an equivalence relation on $\mathbb{N} \times \mathbb{N}$. List five elements in $[(3,1)]$.

10. Let $\sim$ be the relation on $\mathbb{R}$ defined by $x \sim y$ if and only if $x - y \in \mathbb{Q}$. Prove that $\sim$ is an equivalence relation on $\mathbb{R}$. Identify the equivalence classes $[0]$ and $[\sqrt{2}]$.

11. Define a relation $\sim$ on $\mathbb{N} \times \mathbb{N}$ by $(a,b) \sim (c,d)$ iff $ad = bc$.

(a) Prove that $\sim$ is an equivalence relation on $\mathbb{N} \times \mathbb{N}$.
(b) Describe the equivalence classes $[(1,1)]$ and $[(1,2)]$.

12. Let $\mathbb{Q}^* = \{x \in \mathbb{Q} : x \neq 0\}$. Define the relation $\sim$ on $\mathbb{Q}^*$ by $x \sim y$ if and only if $x \cdot y^{-1} \in \mathbb{Z}$. Show that $\sim$ is not an equivalence relation on $\mathbb{Q}^*$.

7.3 Congruence Modulo *m*

Karl Friedrich Gauss (1777–1855) has been called the "Prince of Mathematicians" for his many contributions to pure and applied mathematics. One of Gauss's most important contributions to number theory was the introduction of an equivalence relation on the integers called congruence modulo m, where $m \geq 1$ is an integer. We will explore Gauss's congruence relation and show that the operations of addition, subtraction, and multiplication preserve Gauss's relation (see Theorem 7.3.5).

Definition 7.3.1 (Congruence modulo *m*). Let $m \geq 1$ be an integer. For integers a and b, we define $a \equiv b \pmod{m}$ if and only if $m \mid (a-b)$.

When $a \equiv b \pmod{m}$ we say that a is **congruent** to b **modulo** m. We also write $a \not\equiv b \pmod{m}$, when we wish to say that a is **not** congruent to b modulo m. Here are some more examples of this notation:

- $10 \equiv 2 \pmod{4}$ because $4 \mid (10-2)$,
- $-5 \equiv 3 \pmod{4}$ since $4 \mid (-5-3)$,
- $24 \equiv 0 \pmod{4}$ as $4 \mid (24-0)$,
- $3 \not\equiv 1 \pmod{4}$ because $4 \nmid (3-1)$.

Remark 7.3.2. The notation $a \equiv b \pmod{m}$ is just a statement about divisibility and is used mainly to simplify reasoning about the divisibility concept. When $m \geq 1$ is an integer and a, b are integers, one can easily verify that the following assertions are all equivalent:

1. $a \equiv b \pmod{m}$,
2. $m \mid (a-b)$,
3. $a - b = km$ for some $k \in \mathbb{Z}$,
4. $a = b + km$ for some $k \in \mathbb{Z}$.

The above equivalences are very important and will be implicitly used for the remainder of this section.

Example 1. Let $m \geq 1$ be an integer. Then $m \equiv 0 \pmod{m}$ and $km \equiv 0 \pmod{m}$ for any integer k, because $m \mid (m-0)$ and $m \mid (km-0)$.

Example 2. Let n be an integer. By the division algorithm (see Theorem 4.6.9), there are integers k and r such that $n = 4k + r$ and $0 \leq r < 4$. Thus, we have either $n = 4k+0$, $n = 4k+1$, $n = 4k+2$, or $n = 4k+3$. Hence, $n - 0 = 4k$, $n - 1 = 4k$, $n - 2 = 4k$, or $n - 3 = 4k$. Therefore, either $n \equiv 0 \pmod{4}$, or $n \equiv 1 \pmod{4}$, or $n \equiv 2 \pmod{4}$, or $n \equiv 3 \pmod{4}$.

7.3.1 Fundamental Properties

Recall that the equality relation is reflexive, symmetric, and transitive. So equality is an equivalence relation. Notice that the congruence relation symbol $\equiv$ resembles the equality symbol. The main reason for this resemblance is that the congruence relation and the equality relation share many of the same properties. In particular, the congruence relation is also an equivalence relation. Throughout this section, m will be a natural number and a, b shall be integers.

Theorem 7.3.3. *The congruence modulo m relation is an equivalence relation on the set of integers.*

Proof. We prove that the congruence modulo m relation on the set of integers is an equivalence relation; that is, we prove the relation is (1) reflexive, (2) symmetric, and (3) transitive, respectively, as follows:

(1) Let x be an integer. Since $x - x = 0m$, we conclude that $x \equiv x \pmod{m}$.

(2) Let $x, y \in \mathbb{Z}$ and assume $x \equiv y \pmod{m}$. So $x - y = km$ for some $k \in \mathbb{Z}$. We prove that $y \equiv x \pmod{m}$. Since $x - y = km$, it follows that $y - x = (-k)m$ where $-k$ is an integer. Therefore, $y \equiv x \pmod{m}$.

(3) Let x, y, z be integers. Assume $x \equiv y \pmod{m}$ and $y \equiv z \pmod{m}$, that is, assume (a) $x - y = im$ and (b) $y - z = jm$ for some $i, j \in \mathbb{Z}$. By adding the corresponding sides of (a) and (b), we obtain $x - z = im + jm = (i+j)m$ where $i + j$ is an integer. Therefore, $x \equiv z \pmod{m}$. □

We recall a property of equality that we often use when solving equations. Given an equation $a = b$ and an integer c, we can add c to both sides of the equation to obtain $a + c = b + c$. We can also multiply both sides of the equation by c and conclude that $ac = bc$. The congruence relation also satisfies these properties.

Theorem 7.3.4. *If $a \equiv b \pmod{m}$, then for all integers c we have*

1. $a + c \equiv b + c \pmod{m}$,
2. $a - c \equiv b - c \pmod{m}$,
3. $ac \equiv bc \pmod{m}$.

Proof. Assume $a \equiv b \pmod{m}$ and let c be an integer, that is, assume that

$$a - b = mi \tag{7.2}$$

for some integer i.

1. To prove $(a+c) \equiv (b+c) \pmod{m}$, we first add and subtract c on the left side of (7.2) to obtain $a + c - c - b = mi$. Since $-c - b = -(b + c)$, we have

$$(a + c) - (b + c) = mi.$$

   Hence, $(a + c) \equiv (b + c) \pmod{m}$.
2. To prove $(a - c) \equiv (b - c) \pmod{m}$, we first subtract and add c on the left side of (7.2) to obtain $a - c + c - b = mi$. Since $c - b = -(b - c)$, we obtain

$$(a - c) - (b - c) = mi.$$

   Therefore, $(a - c) \equiv (b - c) \pmod{m}$.
3. To prove $ac \equiv bc \pmod{m}$, we multiply (both sides of) (7.2) by c which gives us $ac - bc = mic$. Thus, $ac - bc = m(ic)$. Therefore, $ac \equiv bc \pmod{m}$. □

When we have two equations $a = b$ and $c = d$, we can derive a new equation by adding both sides of these equations to obtain $a + c = b + d$. Similarly, we can multiply both sides of these equation and conclude that $ac = bd$. So, the equality relation is preserved under addition and multiplication. These important algebraic properties of equality are frequently used to solve equations.

In this section, we will develop a *congruence algebra*; that is, we will show that when we are given two congruence relations $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, we can also add, subtract, and multiply both sides of these congruences to derive a new congruence. The following theorem is a fundamental result which shows that the congruence modulo m relation is preserved under the operations of addition, subtraction, and multiplication.

Theorem 7.3.5. *If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then*

1. $(a + c) \equiv (b + d) \pmod{m}$,
2. $(a - c) \equiv (b - d) \pmod{m}$,
3. $ac \equiv bd \pmod{m}$.

Proof. Assume that $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$. Theorem 7.3.4 implies that $(a+c) \equiv (b+c) \pmod{m}$, because $a \equiv b \pmod{m}$. Similarly, since $c \equiv d \pmod{m}$, Theorem 7.3.4 also implies that $(b+c) \equiv (b+d) \pmod{m}$. Thus, we have the two congruences $(a+c) \equiv (b+c) \pmod{m}$ and $(b+c) \equiv (b+d) \pmod{m}$. Therefore, $(a+c) \equiv (b+d) \pmod{m}$ as the congruence relation is transitive by Theorem 7.3.3. The proofs of items 2 and 3 are similar and are left as an exercise. □

The next theorem shows that one can square both sides of a congruence relation to derive a new congruence.

Theorem 7.3.6. *If $a \equiv b \pmod{m}$, then $a^2 \equiv b^2 \pmod{m}$.*

Proof. Assume $a \equiv b \pmod{m}$. Theorem 7.3.5(3) implies that $aa \equiv bb \pmod{m}$. Therefore, $a^2 \equiv b^2 \pmod{m}$. □

Using Theorem 7.3.5(3) and mathematical induction on k, one can prove (see Exercise 16) the following theorem.

Theorem 7.3.7. *If $a \equiv b \pmod{m}$, then $a^k \equiv b^k \pmod{m}$ for every integer $k \geq 1$.*

Theorems 7.3.3–7.3.7 will allow us to derive new congruence relations by using congruence algebra and congruence substitution.

Example 3. Let $m > 1$ be an integer. Suppose $a \equiv 4 \pmod{m}$, $b \equiv 10 \pmod{m}$, and $c \equiv 3 \pmod{m}$. Show that $3a^2 - 2b - c^3 + 4m \equiv 1 \pmod{m}$.

Solution. We are given that

$$a \equiv 4 \pmod{m},\ b \equiv 10 \pmod{m},\ \text{and}\ c \equiv 3 \pmod{m}. \tag{7.3}$$

We will show that $3a^2 - 2b - c^3 + 4m \equiv 1 \pmod{m}$ as follows:

$$\begin{aligned} 3a^2 - 2b - c^3 + 4m &\equiv 3a^2 - 2b - c^3 \pmod{m} && \text{because } 4m \equiv 0 \pmod{m} \\ &\equiv 3 \cdot 4^2 - 2 \cdot 10 - 3^3 \pmod{m} && \text{by (7.3)} \\ &\equiv 1 \pmod{m} && \text{because } 3 \cdot 4^2 - 2 \cdot 10 - 3^3 = 1. \end{aligned}$$

Therefore, $3a^2 + 2b + c^3 + 4m \equiv 1 \pmod{m}$. Ⓢ

7.3.2 Congruence Classes

Let $m \geq 1$ be an integer. We shall temporarily use $\equiv$ to abbreviate the congruence relation $(\bmod m)$. Thus, $\equiv$ is an equivalence relation on the set of integers $\mathbb{Z}$, by Theorem 7.3.3. For each $a \in \mathbb{Z}$, recalling Definition 7.2.2, the set $[a]_{\equiv}$ is the equivalence class of a. Because the relation $\equiv$ is so closely connected with m, we will use $[a]_m$ to denote $[a]_{\equiv}$. Hence, $[a]_m = \{x \in \mathbb{Z} : x \equiv a \pmod{m}\}$. Since

$x \in [a]_m$ iff $x \equiv a \pmod{m}$ iff $x = a + mk$ for some $k \in \mathbb{Z}$, we conclude that $[a]_m = \{a + mk : k \in \mathbb{Z}\}$. We shall call $[a]_m$ the *congruence class of* $a \pmod{m}$. Applying Theorem 7.2.4, we see that

$$a \equiv b \pmod{m} \text{ if and only if } [a]_m = [b]_m \tag{7.4}$$

for all $a, b \in \mathbb{Z}$. It follows from Theorem 7.2.6 and Definition 7.2.7 that

$$\mathbb{Z}/_{\equiv} = \{[a]_m : a \in \mathbb{Z}\}$$

is the partition of the set of integers $\mathbb{Z}$ induced by $\equiv$. We shall use the notation $\mathbb{Z}_m$ to denote the partition $\mathbb{Z}/_{\equiv}$.

Lemma 7.3.8. *Let $m > 1$ be an integer. For all $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$, we have that*

1. *$[a]_m = [b]_m$ if and only if $a \equiv b \pmod{m}$;*
2. *$a \in [b]_m$ if and only if $a \equiv b \pmod{m}$.*

Proof. Since congruence modulo m is an equivalence relation on $\mathbb{Z}$, items 1 and 2 follow from Theorem 7.2.4 and Corollary 7.2.5, respectively. □

We now determine the number of distinct equivalence classes that are in $\mathbb{Z}_m$.

Theorem 7.3.9. *Let $m \geq 1$ be an integer. For each integer a there is exactly one integer r in the list $0, 1, \ldots, m-1$ such that $a \equiv r \pmod{m}$.*

Proof. Let $m \geq 1$ be an integer and let a be an integer. We will prove that there is a unique integer r in the list $0, 1, \ldots, m-1$ such that $a \equiv r \pmod{m}$.

Existence: By Theorem 4.6.9, there exist integers q and r such that $a = qm + r$ where $0 \leq r < m$. So, $a - r = qm$ and therefore, $a \equiv r \pmod{m}$ and $0 \leq r < m$.

Uniqueness: Let r be as in the existence part of our proof. Thus, $a \equiv r \pmod{m}$ and $a = qm + r$ where $0 \leq r < m$. Let i be any integer that also satisfies $a \equiv i \pmod{m}$ and $0 \leq i < m$. We shall prove that $r = i$. Since $a \equiv i \pmod{m}$, there is an integer j such that $a - i = jm$, that is, $a = jm + i$ where $0 \leq i < m$. Because $a = qm + r$ and $0 \leq r < m$, Theorem 4.6.9 implies that $r = i$. □

Example 4. Show that every perfect square is congruent to 0 or 1 $\pmod{4}$.

Solution. Let n be a perfect square. So $n = k^2$ for some integer k. Theorem 7.3.9 asserts that $k \equiv r \pmod{4}$ for some r in the list $0, 1, 2, 3$. Thus, either

$$k \equiv 0 \pmod{4},\ k \equiv 1 \pmod{4},\ k \equiv 2 \pmod{4}, \text{ or } k \equiv 3 \pmod{4}.$$

Theorem 7.3.6 implies that either

$$k^2 \equiv 0 \pmod{4},\ k^2 \equiv 1 \pmod{4},\ k^2 \equiv 4 \pmod{4}, \text{ or } k^2 \equiv 9 \pmod{4}. \tag{7.5}$$

Since $4 \equiv 0 \pmod 4$ and $9 \equiv 1 \pmod 4$, the last two congruences in (7.5) are redundant versions of the first two. Hence, either $k^2 \equiv 0 \pmod 4$ or $k^2 \equiv 1 \pmod 4$. As $n = k^2$, we see that $n \equiv 0 \pmod 4$ or $n \equiv 1 \pmod 4$. Ⓢ

Given an integer $m \geq 1$, Theorem 7.3.9 asserts that every integer is congruent $\pmod m$ to exactly one of the numbers in the list $0, 1, \ldots, m-1$. For this reason, we shall call this list a *complete residue system* $\pmod m$.

Corollary 7.3.10. *Let $m \geq 1$ be an integer. For each integer k there is exactly one integer r in the list $0, 1, \ldots, m-1$ such that $[k]_m = [r]_m$.*

Proof. Let k be an integer. By Theorem 7.3.9, there is exactly one integer r in the list $0, 1, \ldots, m-1$ such that $k \equiv r \pmod m$. Lemma 7.3.8(1) implies that $[k]_m = [r]_m$. It thus follows that there is exactly one such r satisfying $[k]_m = [r]_m$. □

Corollary 7.3.11. *Let $m \geq 1$ be an integer. Let k and r be distinct integers in the list $0, 1, \ldots, m-1$. Then $[k]_m \neq [r]_m$ and hence, $[k]_m$ and $[r]_m$ are disjoint.*

Proof. Let $k \neq r$ both be in the list $0, 1, \ldots, m-1$. Corollary 7.3.10 implies that $[k]_m \neq [r]_m$. Thus, by Theorem 7.2.6, we have $[k]_m \cap [r]_m = \varnothing$. □

Let $m \geq 1$ be an integer. Corollaries 7.3.10 and 7.3.11 assert that there are exactly m many distinct congruence classes $\pmod m$ for a given integer $m \geq 1$. Therefore,

$$\mathbb{Z}_m = \{[a]_m : a \in \mathbb{Z}\} = \{[0]_m, [1]_m, [2]_m, \ldots, [m-1]_m\}.$$

Example 5. Let $\mathbb{Z}$ be the set of integers. Consider the equivalence relation $\equiv$ on $\mathbb{Z}$ defined by

$$a \equiv b \text{ if and only if } a \equiv b \pmod 5,$$

and let $\mathbb{Z}_5$ be the partition of $\mathbb{Z}$ induced by $\equiv$.

1. For every $a \in \mathbb{Z}$, we have $[a]_5 = \{a + 5k : k \in \mathbb{Z}\}$.
2. Corollaries 7.3.10 and 7.3.11 imply that $\mathbb{Z} = [0]_5 \cup [1]_5 \cup [2]_5 \cup [3]_5 \cup [4]_5$ and that the sets $[0]_5, [1]_5, [2]_5, [3]_5, [4]_5$ are all mutually disjoint. Thus,

$$\mathbb{Z}_5 = \{[0]_5, [1]_5, [2]_5, [3]_5, [4]_5\}.$$

3. The set $\mathbb{Z}_5$ is a partition of $\mathbb{Z}$ as illustrated in the diagram (7.6).

$$\mathbb{Z} = \begin{array}{|c|c|c|c|c|} \hline \vdots & \vdots & \vdots & \vdots & \vdots \\ 10 & 11 & 12 & 13 & 14 \\ 5 & 6 & 7 & 8 & 9 \\ 0 & 1 & 2 & 3 & 4 \\ -5 & -4 & -3 & -2 & -1 \\ -10 & -9 & -8 & -7 & -6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \end{array} \qquad \begin{array}{ccccc} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ [0]_5 & [1]_5 & [2]_5 & [3]_5 & [4]_5 \end{array} \tag{7.6}$$

Remark 7.3.12. Let $m \geq 1$ be an integer. When the integer m is understood, we shall drop the subscripts; that is, we will just use $[a]$ instead of $[a]_m$ and shall write $\mathbb{Z}_m = \{[0],[1],[2],\ldots,[m-1]\}$. Many texts just let $\mathbb{Z}_m = \{0,1,2,\ldots,m-1\}$, which clearly simplifies the notation, and at times we will do the same.

Exercises 7.3

1. Which of the following are valid congruences? $5 \equiv 13 \pmod 4$, $4 \equiv 14 \pmod 6$, and $18 \equiv -2 \pmod{10}$.

2. By dividing 97 by 7, find a so that $a \equiv 97 \pmod 7$ and $0 \leq a < 7$. Now find b so that $b \equiv -97 \pmod 7$ and $0 \leq b < 7$.

3. Let $m \geq 1$ be an integer. Show that if $n \equiv 1 \pmod m$, then $n^2 + n \equiv 2 \pmod m$ for every integer n.

4. Show that the converse of Theorem 7.3.6 does not hold, via a counterexample.

5. Prove that for every odd integer k we have $k^2 \equiv 1 \pmod 8$.

6. Prove that $(m-1)^2 \equiv 1 \pmod m$ for every integer $m > 1$.

7. Let $m \geq 1$ be an integer and let a,b,k be integers where $k \geq 1$. Prove that if $a \equiv b \pmod m$ and $k \mid m$, then $a \equiv b \pmod k$.

8. Let $m \geq 1$ be an integer and let a,b,k be integers where $k \geq 1$. Prove that if $a \equiv b \pmod m$, then $ak \equiv bk \pmod{mk}$.

9. Let $m \geq 1$ be an integer and let a,b,k be integers. Suppose $\gcd(k,m) = 1$. Prove that if $ka \equiv kb \pmod m$, then $a \equiv b \pmod m$.

10. Show that every perfect square is congruent to 0, 1 or 4 $\pmod 8$.

11. Let p be a prime and let $1 \leq k < p$ be an integer. Prove that $\binom{p}{k} \equiv 0 \pmod p$.

12. Show that $n^3 \equiv n \pmod 3$, for all integers n.

13. For every pair of integers x and y, show that $x^2 + y^2$ is congruent to 0, 1 or 2 $\pmod 4$. Conclude that if an integer n is congruent to 3 $\pmod 4$, then n is not the sum of two perfect squares.

14. Let p be a prime and let $a,b \in \mathbb{Z}$. Prove that $(a+b)^p \equiv a^p + b^p \pmod p$.

15. Prove items 2 and 3 of Theorem 7.3.5.

16. Prove Theorem 7.3.7 using mathematical induction.

17. Let $m = 6$ and define the equivalence relation on $\mathbb{Z}$: $a \equiv b$ if and only if $a \equiv b \pmod 6$. Let $\mathbb{Z}_6 = \{[0],[1],[2],[3],[4],[5]\}$ be the partition induced by $\equiv$, where $[a] = [a]_6$ for each $a \in \mathbb{Z}$. Determine which of the following are true.

(a) $[2] = [3]$.
(b) $[2] = [-4]$.
(c) $5 \in [8]$.
(d) $5 \in [11]$.
(e) $[3+4] = [1]$.
(f) $[3 \cdot 4] = [0]$.

18. Let a and $m > 1$ be integers. Prove that $[a]_m = [0]_m$ if and only if $m \,|\, a$.

19. Let a, b and $m > 1$ be integers. Prove that $[a]_m = [b]_m$ if and only if $a - b = mi$ for some integer i.

20. Let $\mathbb{Z}_6 = \{[0],[1],[2],[3],[4],[5]\}$. Consider the purported function $f\colon \mathbb{Z}_6 \to \mathbb{Z}$ defined by $f([a]) = a$ for all $[a] \in \mathbb{Z}_6$. Show that f is not well defined.

Exercise Notes: For Exercise 3, use Theorems 7.3.6 and 7.3.5. For Exercise 5, see Exercise 4 on page 106. For Exercise 9, show that $m \,|\, k(a-b)$ and then use Theorem 4.6.14. For Exercise 11, since $1 \leq k < p$, it follows that $1 \leq p - k < p$ as well. We also know that the value $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ is a natural number by Exercise 12 on page 122. For Exercise 12, we know by Theorem 7.3.9 that there are three cases to consider: $n \equiv 0 \pmod 3$, $n \equiv 1 \pmod 3$, or $n \equiv 2 \pmod 3$. For Exercise 13, use Example 4 and congruence algebra and congruence substitution. For Exercise 14, expand $(a+b)^p$ using the Binomial Theorem (see Exercise 16 on page 122) and then use Exercise 11. For Exercise 17, review Lemma 7.3.8. For Exercise 20, review Lemma 7.3.8 and Example 4 on page 172.

7.4 Modular Arithmetic

Modular arithmetic is a system of arithmetic that is based on the congruence modulo m relation. Carl Friedrich Gauss first introduced modular arithmetic in 1801 and it has become an important tool in number theory. In the next definition, we will first recall the rules of integer arithmetic that we all learned in elementary school. These rules are referred to as the *axioms of arithmetic*. We shall soon see that modular arithmetic also satisfies these same rules. Let $\mathbb{Z}$ be the set of integers with the usual operations $+$ and $\cdot$ which we call addition and multiplication.

Integer Arithmetic 7.4.1. The number system $(\mathbb{Z}, +, \cdot)$ satisfies the following nine AXIOMS OF ARITHMETIC:

1. $a + b = b + a$ for all $a, b \in \mathbb{Z}$.
2. $(a + b) + c = a + (b + c)$ for all $a, b, c \in \mathbb{Z}$.
3. $a + 0 = a$ for all $a \in \mathbb{Z}$.
4. For all $a \in \mathbb{Z}$ there exists a $b \in \mathbb{Z}$ such that $a + b = 0$.
5. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in \mathbb{Z}$.
6. $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in \mathbb{Z}$.
7. $(b + c) \cdot a = b \cdot a + c \cdot a$ for all $a, b, c \in \mathbb{Z}$.
8. $a \cdot 1 = 1 \cdot a = a$ for all $a \in \mathbb{Z}$.
9. $a \cdot b = b \cdot a$ for all $a, b \in \mathbb{Z}$.

The above nine axioms identify the fundamental properties of integer arithmetic. For example, Axiom 1 asserts that addition is commutative and Axiom 2 states that

addition is also associative. Axiom 3 states that 0 is the *additive identity element* and Axiom 4 declares that every integer has an *additive inverse*. The distributive properties of arithmetic are given in Axioms 6–7. Axiom 8 identifies the integer 1 as the *multiplicative identity element*.

Let $m \geq 1$ be an integer. We shall now define operations of addition $\oplus$ and multiplication $\odot$ on the finite set $\mathbb{Z}_m$. For $[a], [b] \in \mathbb{Z}_m$ define $\oplus$ and $\odot$ by

$$[a] \oplus [b] = [a+b] \tag{7.7}$$

$$[a] \odot [b] = [a \cdot b]. \tag{7.8}$$

To perform the operation $[a] \oplus [b]$, one first adds the integers a and b, obtaining $a+b$. Then the value of the sum $[a] \oplus [b]$ is the congruence class $[a+b]$. Thus, the equation $[a] \oplus [b] = [a+b]$ is another way of saying "the sum of two congruence classes is the congruence class of the sum."

Similarly, to perform the operation $[a] \odot [b]$, one first multiplies the integers a and b, obtaining $a \cdot b$. Then the value of the product $[a] \odot [b]$ is the congruence class $[a \cdot b]$. Therefore, the equation $[a] \odot [b] = [a \cdot b]$ states that "the product of two congruence classes is the congruence class of the product."

Example 1. Let $m = 7$ and write $[a]$ for $[a]_7$ whenever $a \in \mathbb{Z}$. Let us evaluate and simplify the sum $[4] \oplus [5]$ in $\mathbb{Z}_7$. Thus, $[4] \oplus [5] = [4+5] = [9]$. To simplify this answer, we know from Corollary 7.3.10 there is exactly one integer r in the list $0, 1, \ldots, 6$ such that $[9] = [r]$. Clearly, $r = 2$ because $9 \equiv 2 \pmod 7$ (see Lemma 7.3.8(1)). So we have that $[4] \oplus [5] = [2]$. Let us evaluate and simplify the product $[4] \odot [5]$ in $\mathbb{Z}_7$. We obtain $[4] \odot [5] = [4 \cdot 5] = [20] = [6]$, since $20 \equiv 6 \pmod 7$.

Now, let us revisit Example 1 where $m = 7$ and $[a] = [a]_7$ for each $a \in \mathbb{Z}$. Observe that $[4] = [11]$ and $[5] = [19]$, since $4 \equiv 11 \pmod 7$ and $5 \equiv 19 \pmod 7$. Thus, $[4]$ and $[5]$ each have more than one representation. Will the sum $[11] \oplus [19]$ give us the same answer as the sum $[4] \oplus [5]$? Let us evaluate and simplify the sum $[11] \oplus [19]$ as we did in Example 1. We obtain $[11] \oplus [19] = [11 + 19] = [30] = [2]$, where the last equality holds since $30 \equiv 2 \pmod 7$. So $[4] \oplus [5] = [11] \oplus [19]$, that is, the sums are equal. It appears that the sum does not depend on the representations used to perform the operation $\oplus$. It is important to verify that this will be the case, for both of the operations $\oplus$ and $\odot$, for any $m \geq 1$.

Theorem 7.4.2 shows that the operations $\oplus$ and $\odot$ are *well-defined*, that is, the sum $\oplus$ and product $\odot$ do not depend on the representations used to perform these operations (see Section 8.1.1 on page 241).

Theorem 7.4.2. *Let $m \geq 1$ be an integer and let $[a], [b], [c], [d] \in \mathbb{Z}_m$. If $[a] = [b]$ and $[c] = [d]$, then*

(1) $[a] \oplus [c] = [b] \oplus [d]$
(2) $[a] \odot [c] = [b] \odot [d]$.

Proof. We shall prove only (1) and leave (2) as an exercise. Let $m \geq 1$ be an integer. Let $[a], [b], [c], [d] \in \mathbb{Z}_m$. Suppose that

$$[a] = [b] \text{ and } [c] = [d]. \tag{7.9}$$

We shall prove that $[a] \oplus [c] = [b] \oplus [d]$, that is, we prove that $[a + c] = [b + d]$. From (7.9) and Lemma 7.3.8(1), we conclude that

$$a \equiv b \pmod{m} \text{ and } c \equiv d \pmod{m}.$$

Theorem 7.3.5(1) implies that $(a+c) \equiv (b+d) \pmod{m}$. Therefore, $[a+c] = [b+d]$ by Lemma 7.3.8(1). □

For each integer $m \geq 1$, Theorem 7.4.2 implies that the operations of addition $\oplus$ and multiplication $\odot$ on the set $\mathbb{Z}_m$, given by (7.7) and (7.8), are well-defined. Thus, we can add and multiply any two elements in $\mathbb{Z}_m$. The number system $(\mathbb{Z}_m, \oplus, \odot)$ will be referred to as *modular arithmetic* and is used extensively in number theory, abstract algebra, and cryptography.

We know that addition is associative in integer arithmetic. The next lemma shows that addition is also associative in modular arithmetic.

Lemma 7.4.3. *Let $m \geq 1$ be an integer. The operation $\oplus$ is associative in the number system $(\mathbb{Z}_m, \oplus, \odot)$.*

Proof. Let $[a], [b], [c] \in \mathbb{Z}_m$. We prove $([a] \oplus [b]) \oplus [c] = [a] \oplus ([b] \oplus [c])$ as follows:

$$\begin{aligned}
([a] \oplus [b]) \oplus [c] &= [a+b] \oplus [c] && \text{by (7.7)} \\
&= [(a+b)+c] && \text{by (7.7)} \\
&= [a+(b+c)] && \text{because } + \text{ is associative in } \mathbb{Z} \\
&= [a] \oplus [b+c] && \text{by (7.7)} \\
&= [a] \oplus ([b] \oplus [c]) && \text{by (7.7).}
\end{aligned}$$

Therefore, $([a] \oplus [b]) \oplus [c] = [a] \oplus ([b] \oplus [c])$. □

We shall write $\oplus$ as $+$ and write $\odot$ as $\cdot$, to simplify our notation. Accordingly, we shall write (7.7) and (7.8) as

$$[a] + [b] = [a + b] \tag{7.10}$$

$$[a] \cdot [b] = [a \cdot b]. \tag{7.11}$$

In Theorem 7.4.4, below, we show that modular arithmetic satisfies the same rules of arithmetic that were given in 7.4.1.

Theorem 7.4.4 (Modular Arithmetic). *Let $m > 1$ be an integer. Let $(\mathbb{Z}_m, +, \cdot)$ be the algebraic system where the operations $+$ and $\cdot$ are defined by (7.10) and (7.11). Then $(\mathbb{Z}_m, +, \cdot)$ satisfies the following nine* AXIOMS OF ARITHMETIC*:*

1. $[a] + [b] = [b] + [a]$ *for all* $[a], [b] \in \mathbb{Z}_m$.
2. $([a] + [b]) + [c] = [a] + ([b] + [c])$ *for all* $[a], [b], [c] \in \mathbb{Z}_m$.
3. $[a] + [0] = [a]$ *for all* $[a] \in \mathbb{Z}_m$.
4. *For all* $[a] \in \mathbb{Z}_m$ *there exists a* $[b] \in \mathbb{Z}_m$ *such that* $[a] + [b] = [0]$.
5. $([a] \cdot [b]) \cdot [c] = [a] \cdot ([b] \cdot [c])$ *for all* $[a], [b], [c] \in \mathbb{Z}_m$.
6. $[a] \cdot ([b] + [c]) = [a] \cdot [b] + [a] \cdot [c]$ *for all* $[a], [b], [c] \in \mathbb{Z}_m$.
7. $([b] + [c]) \cdot [a] = [b] \cdot [a] + [c] \cdot [a]$ *for all* $[a], [b], [c] \in \mathbb{Z}_m$.
8. $[a] \cdot [1] = [1] \cdot [a] = [a]$ *for all* $[a] \in \mathbb{Z}_m$.
9. $[a] \cdot [b] = [b] \cdot [a]$ *for all* $[a], [b] \in \mathbb{Z}_m$.

Proof. Lemma 7.4.3 establishes item 2. Each of the items in 1–9 can be established by using the idea in the proof of Lemma 7.4.3. For example, to establish item 1, let $[a], [b] \in \mathbb{Z}_m$. Then $[a] + [b] = [a+b] = [b+a] = [b] + [a]$ by (7.10) and the fact that addition is commutative in $(\mathbb{Z}, +, \cdot)$. One can easily prove the other items. □

Theorem 7.4.4 states that the operations of addition and multiplication in modular arithmetic are associative, commutative, and satisfy the distributive properties given in Axioms 6–7. Axiom 3 states that $[0]$ is the *additive identity element* and Axiom 8 asserts that the congruence class $[1]$ is the *multiplicative identity element*.

From Integer Arithmetic 7.4.1 and Theorem 7.4.4, we know that integer arithmetic and modular arithmetic both satisfy the axioms of arithmetic. Thus, integer arithmetic and modular arithmetic possess the same basic rules of algebra. On the other hand, there are some differences between these two systems of arithmetic. Recall that if a and b are integers and $a \cdot b = 0$, then either $a = 0$ or $b = 0$. Also recall that if $a \cdot b = 1$, where a, b are integers, then we must have that $a = b = \pm 1$. These properties may fail for modular arithmetic.

Definition 7.4.5. Let $m > 1$ and a be integers. Let $[a] \in \mathbb{Z}_m$ be such that $[a] \neq [0]$. Then $[a]$ is a **zero divisor** in $\mathbb{Z}_m$ whenever $[a] \cdot [b] = [0]$ for some $[b] \neq [0]$ in $\mathbb{Z}_m$.

Definition 7.4.6. Let $m > 1$ and a be integers. We say $[a]$ is **invertible** in $\mathbb{Z}_m$ if $[a] \cdot [b] = [1]$ for some $[b]$ in $\mathbb{Z}_m$ called the **multiplicative inverse** of $[a]$.

Theorem 7.4.7. *Let $m > 1$ and a be integers. Then the following hold:*

(1) *The structure $(\mathbb{Z}_m, +, \cdot)$ has zero divisors if and only if m is composite.*
(2) *$[a]$ is invertible in $\mathbb{Z}_m$ if and only if* $\gcd(a, m) = 1$.
(3) *$[a]$ is a zero divisor in $\mathbb{Z}_m$ if and only if $m \nmid a$ and* $\gcd(a, m) > 1$.

Proof. Let $m > 1$ and a be integers. We shall only prove (3) and leave (1)–(2) as exercises. Let $[a] \in \mathbb{Z}_m$. We shall prove that $[a]$ is a zero divisor in $\mathbb{Z}_m$ if and only if $m \nmid a$ and $\gcd(a, m) > 1$.

($\Rightarrow$). Assume that $[a]$ is a zero divisor, that is, assume that $[a] \neq [0]$ and there is a $[b] \in \mathbb{Z}_m$ such that $[a] \cdot [b] = [0]$ where $[b] \neq [0]$. By (7.11), we have that

$$[ab] = [0],\ [a] \neq [0] \text{ and } [b] \neq [0].$$

Lemma 7.3.8(1) implies $ab \equiv 0 \pmod{m}$, $a \not\equiv 0 \pmod{m}$ and $b \not\equiv 0 \pmod{m}$. Hence, $m \mid ab$, $m \nmid a$ and $m \nmid b$. Since $m \mid ab$ and $m \nmid b$, Theorem 4.6.14 forces us to conclude that $\gcd(a,m) > 1$. Thus, $m \nmid a$ and $\gcd(a,m) > 1$.

($\Leftarrow$). Assume $m \nmid a$ and $\gcd(a,m) > 1$. Since $m \nmid a$, we infer that $[a] \neq [0]$. Let $d = \gcd(a,m)$. Thus,

$$d \mid a \text{ and } d \mid m. \tag{7.12}$$

Since $d \mid m$, we see that $d \leq m$. Furthermore, as $m \nmid a$, we also see that $d < m$ (for if $d = m$, then $m \mid a$ because $d \mid a$). Hence, $1 < d < m$. From (7.12), we obtain

$$a = di \tag{7.13}$$

$$m = dj \tag{7.14}$$

for some integers i and j. Since $1 < d < m$, (7.14) implies that $1 < j < m$ and so, $[j] \neq [0]$ by Corollary 7.3.11. Multiplying both sides of (7.13) by j and then using (7.14), we get $aj = dij = im$. Thus, $aj \equiv 0 \pmod{m}$ and so $[a] \cdot [j] = [0]$ where $[a] \neq [0]$ and $[j] \neq [0]$. Therefore, $[a]$ is a zero divisor. □

Given any integer $m \geq 1$, we have introduced $\mathbb{Z}_m = \{[0],[1],[2],\ldots,[m-1]\}$, the set of all congruence classes $\pmod{m}$. To simplify our notation, we shall write $\mathbb{Z}_m = \{0,1,2,\ldots,m-1\}$. We will now look at two different systems of modular arithmetic, namely, the system $(\mathbb{Z}_5,+,\cdot)$ and the system $(\mathbb{Z}_6,+,\cdot)$. Of course these two systems satisfy the axioms of arithmetic; however, there are some algebraic properties that they do not share.

Addition and Multiplication Tables for the System $(\mathbb{Z}_5,+,\cdot)$

Consider the set $\mathbb{Z}_5 = \{0,1,2,3,4\}$. By Theorem 7.3.9, for any two integers a and b in the list $(\star)$ $0,1,2,3,4$, there is a unique r in this list such that $a+b \equiv r \pmod{5}$. For example, $3+2 \equiv 0 \pmod{5}$ and $3+3 \equiv 1 \pmod{5}$. Similarly, for any pair of integers a and b in the list $(\star)$, there is a unique r in the list such that $a \cdot b \equiv r \pmod{5}$. For example, $3 \cdot 4 \equiv 2 \pmod{5}$ and $3 \cdot 3 \equiv 4 \pmod{5}$. The end result is the addition and multiplication tables in Table 7.1 for the system $(\mathbb{Z}_5,+,\cdot)$.

SOME OBSERVATIONS ABOUT $(\mathbb{Z}_5,+,\cdot)$:

1. Theorem 7.4.4(3) states that 0 is the additive identity element in $\mathbb{Z}_5$.
2. Theorem 7.4.4(8) states that 1 is the multiplicative identity for $\mathbb{Z}_5$.

Table 7.1 Modular arithmetic (mod 5)

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

Addition in $\mathbb{Z}_5$

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Multiplication in $\mathbb{Z}_5$

Table 7.2 Modular arithmetic (mod 6)

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

Addition in $\mathbb{Z}_6$

$\cdot$	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

Multiplication in $\mathbb{Z}_6$

3. Theorem 7.4.4(4) states that every $a \in \mathbb{Z}_5$ has an additive inverse. For example, note that $4+1=0$. So, 1 is the additive inverse of 4 and we write $-4=1$. Also, note that $3+2=0$. Thus, 2 is the additive inverse of 3 and we write $-3=2$.
4. Theorem 7.4.7(2) states that $a \in \mathbb{Z}_5$ is invertible if and only if $\gcd(a,5)=1$. Therefore $1,2,3,4$ have multiplicative inverses. Note that $2\cdot 3=1$. So, 3 is the multiplicative inverse of 2 and we write $2^{-1}=3$. Also, note that $4\cdot 4=1$. Hence, 4 is the multiplicative inverse of 4 and we write $4^{-1}=4$.
5. Since $m=5$ is a prime, $(\mathbb{Z}_5,+,\cdot)$ has no zero divisors by Theorem 7.4.7(1).

When m is a prime number, then all of the nonzero elements in the system $(\mathbb{Z}_m,+,\cdot)$ will have multiplicative inverses, as was observed in the system $(\mathbb{Z}_5,+,\cdot)$. Furthermore, if m is a prime, then the system $(\mathbb{Z}_m,+,\cdot)$ has no zero divisors. On the other hand, if m is a composite number, then $(\mathbb{Z}_m,+,\cdot)$ will have zero divisors.

Addition and Multiplication Tables for the System $(\mathbb{Z}_6,+,\cdot)$

The set of $(\text{mod } 6)$ congruence classes is $\mathbb{Z}_6=\{0,1,2,3,4,5\}$. We can perform arithmetic on the set $\mathbb{Z}_6$. Observe that $3+5\equiv 2\,(\text{mod } 6)$ and $3+3\equiv 0\,(\text{mod } 6)$. In addition, we see that $3\cdot 5\equiv 3\,(\text{mod } 6)$ and $3\cdot 3\equiv 3\,(\text{mod } 6)$. Table 7.2 presents the resulting addition table and multiplication table for the system $(\mathbb{Z}_6,+,\cdot)$.

SOME OBSERVATIONS ABOUT $(\mathbb{Z}_6,+,\cdot)$:

1. Theorem 7.4.7(2) states that $a \in \mathbb{Z}_6$ is invertible if and only if $\gcd(a,6)=1$. Therefore, 1 and 5 have multiplicative inverses. Note that $1\cdot 1=1$ and $5\cdot 5=1$.

So, 5 is the multiplicative inverse of 5 and we write $5^{-1} = 5$. Theorem 7.4.7(2) implies that 1 and 5 are the only invertible elements in $(\mathbb{Z}_6, +, \cdot)$.

2. Theorem 7.4.7(1) implies that $(\mathbb{Z}_6, +, \cdot)$ has zero divisors, as $m = 6$ is composite. Theorem 7.4.7(3) implies an $a \in \mathbb{Z}_6$ is a zero divisor if and only if $a \neq 0$ and $\gcd(a, 6) > 1$. It follows that 2, 3, and 4 are the only zero divisors in $\mathbb{Z}_6$. Observe that $2 \cdot 3 = 0$ and $4 \cdot 3 = 0$.

Exercises 7.4

1. Use Table 7.2 to evaluate $(2 \cdot 4) \cdot 5$, $2 \cdot (4 \cdot 5)$, and $2 + (4 \cdot 5)$ in $\mathbb{Z}_6$.

2. Perform the following operations in $\mathbb{Z}_8$.

(a) $[3] + [5]$
(b) $[3] \cdot [5]$
(c) $([6] \cdot [5]) + ([4] \cdot [7])$.

3. Prove item (2) of Theorem 7.4.2.

4. Find all of the zero divisors in each of the following: (a) $\mathbb{Z}_{12}$, (b) $\mathbb{Z}_{27}$, (c) $\mathbb{Z}_{13}$.

5. Find the invertible elements in each of the following: (a) $\mathbb{Z}_{12}$, (b) $\mathbb{Z}_{27}$, (c) $\mathbb{Z}_{13}$.

6. Prove Theorem 7.4.4(6) where $m > 1$ is an integer. That is, prove that

$$[a] \cdot ([b] + [c]) = [a] \cdot [b] + [a] \cdot [c]$$

for all $[a], [b], [c] \in \mathbb{Z}_m$.

7. Let $(\mathbb{Z}_m, +, \cdot)$ be the algebraic system presented in Theorem 7.4.4 where $m > 1$ is an integer. Let a, b, c, s, t be integers. Prove that $[s] \cdot [a] + [t] \cdot [b] = [c]$ if and only if $sa + tb - c = mi$ for some integer i.

8. Let a, b, m be integers where $m > 1$. Suppose $[a]_m = [b]_m$. Prove that $a = mq + r$ and $b = sm + r$ for some integers s, q, r where $0 \leq r < m$.

9. Let $\mathbb{Z}_3 = \{[0]_3, [1]_3, [2]_3\}$ and $\mathbb{Z}_6 = \{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6\}$ be the set of congruence classes (mod 3) and (mod 6). A purported function $f\colon \mathbb{Z}_6 \to \mathbb{Z}_3$ is defined by $f([a]_6) = [a]_3$ for all $[a]_6 \in \mathbb{Z}_6$. Prove that f is well-defined.

10. Given that $\mathbb{Z}_3 = \{[0]_3, [1]_3, [2]_3\}$ and $\mathbb{Z}_6 = \{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6\}$, consider the purported function $f\colon \mathbb{Z}_3 \to \mathbb{Z}_6$ defined by $f([a]_3) = [2a]_6$ for all $[a]_3 \in \mathbb{Z}_3$. Prove that f is well-defined.

11. Prove item (1) of Theorem 7.4.7.

12. Prove item (2) of Theorem 7.4.7.

13. Construct addition and multiplication tables, as in Table 7.2, for $(\mathbb{Z}_7, +, \cdot)$.

14. Construct addition and multiplication tables, as in Table 7.2, for $(\mathbb{Z}_8, +, \cdot)$. List all of the zero divisors in this modular system of arithmetic.

Exercise Notes: For Exercise 8 use Corollary 7.3.10. For Exercises 9–10, review Example 6.1.5 on page 173, Remark 6.1.6, and Lemma 7.3.8. For Exercise 11,

review Lemma 4.7.2 and note that $[x] = [0]$ if and only if $m \mid x$. For Exercise 12, review Corollary 4.6.13 and prove $[a] \cdot [b] = [1]$ if and only if $ab - 1 = mi$ for some $i \in \mathbb{Z}$.

7.5 Order Relations

We generalized the equality relation in Section 7.2 by introducing the notion of an equivalence relation. An equivalence relation satisfies three key properties that are normally associated with equality; namely, the relation is reflexive, symmetric, and transitive. In this section we will generalize the concept of 'less than or equal to.'

The relation $\leq$ on the set $\mathbb{N}$ puts an order on the natural numbers. This relation is reflexive and transitive; but, since $2 \leq 4$ and $4 \not\leq 2$, the relation $\leq$ is not symmetric. Furthermore, for $x, y \in \mathbb{N}$, if $x \leq y$ and $y \leq x$, then $x = y$. We therefore say that $\leq$ is *antisymmetric*. We now present a general definition.

Definition 7.5.1. A relation R on a set A is **antisymmetric** when

$$(\forall x \in A)(\forall y \in A)((xRy \wedge yRx) \rightarrow x = y),$$

that is, if xRy and yRx, then $x = y$ whenever $x, y \in A$.

Since many of the theorems and proofs given in a calculus book require the reader to understand inequalities, most of these books begin by reviewing the properties of the relation $\leq$ on the set of real numbers. Let us recall three fundamental properties of this relation. For all $a, b, c \in \mathbb{R}$ we have:

1. $a \leq a$ (reflexive).
2. If $a \leq b$ and $b \leq a$, then $a = b$ (antisymmetric).
3. If $a \leq b$ and $b \leq c$, then $a \leq c$ (transitive).

The above three properties are the ones that constitute the notion of 'order.' We can now generalize the concept of 'less than or equal to' and apply it to a variety of sets. To do this, we shall use the symbol $\preccurlyeq$ to denote a relation on a set.

Definition 7.5.2. A relation $\preccurlyeq$ on a set A is called a **partial order** if $\preccurlyeq$ is reflexive, antisymmetric, and transitive; that is, for all $x, y, z \in A$, the following hold:

1. $x \preccurlyeq x$.
2. If $x \preccurlyeq y$ and $y \preccurlyeq x$, then $x = y$.
3. If $x \preccurlyeq y$ and $y \preccurlyeq x$, then $x \preccurlyeq z$.

When $\preccurlyeq$ is a partial order on the set A, we shall say that the pair $(A, \preccurlyeq)$ is a **partially order set** or a **poset**. In particular, the pair $(\mathbb{R}, \leq)$ is a poset because the relation $\leq$ on the set of real numbers is reflexive, antisymmetric, and transitive. Similarly, $(\mathbb{Z}, \leq)$ and $(\mathbb{N}, \leq)$ are also posets.

Example 1. Let $\mathcal{F}$ be a family of sets and let $\subseteq$ be the subset relation. Show that $(\mathcal{F}, \subseteq)$ is a partially ordered set.

Solution. Let A, B, C be sets in $\mathcal{F}$. To show that $\subseteq$ is a partial order on $\mathcal{F}$, we prove that the following three properties hold:

1. $A \subseteq A$.
2. If $A \subseteq B$ and $B \subseteq A$, then $A = B$.
3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

Clearly, (1) holds. If $A \subseteq B$ and $B \subseteq A$, then the sets $A = B$ by the definition of set equality. Thus, (2) holds. Finally, item (3) follows from Exercise 2 on page 156. Ⓢ

Example 2. For integers a and b, the divisibility relation $a \,|\, b$ means that $b = ai$ for some $i \in \mathbb{Z}$. Let A be any subset of $\mathbb{N}$, the set of natural numbers. Show that $(A, |\,)$ is a partially ordered set.

Solution. We show that the divisibility relation on the set A is reflexive, antisymmetric, and transitive. Let a, b, c be natural numbers in A. We must verify the following three items:

1. $a \,|\, a$.
2. If $a \,|\, b$ and $b \,|\, a$, then $a = b$.
3. If $a \,|\, b$ and $b \,|\, c$, then $a \,|\, c$.

Since $a = a \cdot 1$, we have that $a \,|\, a$. Exercise 9 on page 84 shows that (2) is true. Theorem 3.5.6 establishes item (3). Ⓢ

Definition 7.5.3. Let $\preccurlyeq$ be a partial order on a set A. The relation $\preccurlyeq$ is called a **total order** if $\preccurlyeq$ satisfies the additional property:

$$(\forall x \in A)(\forall y \in A)(x \preccurlyeq y \vee y \preccurlyeq x),$$

that is, for all $x, y \in A$, either $x \preccurlyeq y$ or $y \preccurlyeq x$.

When $(A, \preccurlyeq)$ is a poset and $\preccurlyeq$ is a total order, then we shall say that $(A, \preccurlyeq)$ is a **totally ordered set**. The posets $(\mathbb{R}, \leq)$ and $(\mathbb{Q}, \leq)$ are totally ordered sets. On the other hand, let $A = \{2, 5, 4, 8, 9\}$ and let $|$ be the divisibility relation. Then $(A, |\,)$ is a poset but since $2 \nmid 5$ and $5 \nmid 2$, we see that $(A, |\,)$ is not a totally ordered set.

Definition 7.5.4. Let $\preccurlyeq$ be a partial order on a set A. For $x, y \in A$, we write $x \prec y$ if and only if $x \preccurlyeq y$ and $x \neq y$. The relation $\prec$ on A shall be called the **strict order** corresponding to $\preccurlyeq$.

Lemma 7.5.5. *Let $\preccurlyeq$ be a partial order on a set A and let $\prec$ be the strict order corresponding to $\preccurlyeq$. Then for all $x, y, z \in A$ we have the following:*

1. $x \not\prec x$.
2. *If $x \prec y$, then $y \not\prec x$.*

3. *If $x \prec y$ and $y \prec z$, then $x \prec z$.*
4. *If $\preccurlyeq$ is a total order on A, then exactly one of the following holds: $x \prec y$, $y \prec x$, or $x = y$.*

Proof. See Exercise 3. □

Let $\preccurlyeq$ be a partial order on a set A and let $x, y \in A$. If $x \prec y$, we will say that x is **smaller** than y and that y is **larger** than x.

Definition 7.5.6. Let $\preccurlyeq$ be a partial order on a set A. An element $b \in A$ is called a **maximal element** if and only if $b \not\prec x$ for all $x \in A$; that is, there is nothing in A that is larger than b.

Definition 7.5.7. Let $\preccurlyeq$ be a partial order on a set A. An element $a \in A$ is said to be a **minimal element** if and only if $x \not\prec a$ for all $x \in A$; that is, there is nothing in A that is smaller than a.

Example 3. Consider the poset $(A, |)$ where $A = \{2,3,4,5,6,9\}$ and $|$ is the divisibility relation. Then $2,3,5$ are a minimal elements and $5,6,9$ are maximal elements.

Definition 7.5.8. Suppose $\preccurlyeq$ is a partial order on a set A and $S \subseteq A$. Let $a, b \in A$.

- If b satisfies $(\forall x \in S)(x \preccurlyeq b)$, then b is called an **upper bound** for S.
- If a satisfies $(\forall x \in S)(a \preccurlyeq x)$, then a is called a **lower bound** for S.

Let $(A, |)$ be the poset in Example 3 where $A = \{2,3,4,5,6,9\}$. Thus, $S = \{2,3\}$ is a subset of A. We see that 6 is an upper bound for S. In addition, observe that there is no lower bound for S in A.

Definition 7.5.9. Suppose $\preccurlyeq$ is a partial order on a set A and $S \subseteq A$.

- If ℓ is an upper bound for S and $\ell \preccurlyeq b$ whenever b is another upper bound for S, then ℓ is called the **least upper bound** for S.
- If g is a lower bound for S and $a \preccurlyeq g$ whenever a is another lower bound for S, then g is called the **greatest lower bound** for S.

We now prove that whenever a least upper bound for a set exists, it is unique. Similarly, if a greatest lower bound exists, it too is unique.

Lemma 7.5.10. *Let $\preccurlyeq$ be a partial order on a set A and let $S \subseteq A$. If ℓ and ℓ' are least upper bounds for S, then $\ell = \ell'$. If g and g' are greatest lower bounds for S, then $g = g'$.*

Proof. Suppose that ℓ and ℓ' are least upper bounds for S. Since ℓ is a least upper bound and ℓ' is another upper bound for S, it follows from the definition of least upper bound that $\ell \preccurlyeq \ell'$. Similarly, it follows that $\ell' \preccurlyeq \ell$. By antisymmetry, we conclude that $\ell = \ell'$. An analogous argument shows that if g and g' are greatest lower bounds for S, then $g = g'$. □

Consider the poset $(\mathbb{N}, |)$ and let $S = \{6, 9, 12\}$. We see that 36 is the least upper bound for S and that 3 is the greatest lower bound for S.

Definition 7.5.11. Suppose $\preccurlyeq$ is a partial order on a set A. Let $S \subseteq A$ and let $a, b \in A$.

- If $(\forall x \in S)(x \preccurlyeq b)$ and $b \in S$, then b is called the **largest element** of S.
- If $(\forall x \in S)(a \preccurlyeq x)$ and $a \in S$, then a is called the **smallest element** of S.

Note that an upper bound, or a lower bound, for S need not be an element of the set S. A largest element of S is just an upper bound that is also an element of S (see Exercise 5). This is the only difference between an upper bound and a largest element. Similarly, a smallest element of S is a lower bound that is also an element of the set S.

Lemma 7.5.12. *Let $\preccurlyeq$ be a partial order on a set A and let $S \subseteq A$. If b and b' are largest elements of S, then $b = b'$. If a and a' are smallest elements of S, then $a = a'$.*

Proof. Let b and b' be largest elements of S. Thus, in particular, b and b' are both elements in S. Since b is a largest element of S and $b' \in S$, it follows that $b' \preccurlyeq b$. Similarly, it follows that $b \preccurlyeq b'$. By antisymmetry, we have that $b = b'$. An analogous argument shows that if a and a' are smallest elements of S, then $a = a'$. □

Let S be the interval $(2, 3]$, a subset of $\mathbb{R}$. So 3 is the largest element of S in the poset $(\mathbb{R}, \leq)$. On the other hand, 2 is the greatest lower bound for S in this poset, but the set S has no smallest element.

Let $\preccurlyeq$ be a partial order on a set A. Even though $\preccurlyeq$ may not be a total order on A, the relation $\preccurlyeq$ can be a total order on certain subsets of A.

Definition 7.5.13. Let $\preccurlyeq$ be a partial order on a set A and let $C \subseteq A$. Then C is called a **chain** in A if for all $x, y \in C$, either $x \preccurlyeq y$ or $y \preccurlyeq x$.

Example 4. Let $|$ be the divisibility relation on the set of natural numbers $\mathbb{N}$. Then $(\mathbb{N}, |)$ is a poset. The set $\{1, 2, 4, 8\}$ is a chain in $\mathbb{N}$ that has 16 as an upper bound. The set $\{2^n : n \in \mathbb{N}\}$ is also a chain in $\mathbb{N}$ and this chain has no upper bound.

We end this section with the statement of a deep theorem that has applications in linear algebra, abstract algebra, and in real analysis. The proof of this theorem requires the axiom of choice. One can find a proof of this theorem in [8].

Theorem 7.5.14 (Zorn's Lemma). *Let $(A, \preccurlyeq)$ be a nonempty partially ordered set. If every chain in A has an upper bound, then A contains a maximal element.*

Using Zorn's lemma, one can prove that every vector space has a basis. It should be noted that Zorn's lemma is actually equivalent to the axiom of choice. Thus, it is impossible to prove Zorn's lemma without using the axiom of choice.

Exercises 7.5

1. Define a relation $\preccurlyeq$ on the set of integers $\mathbb{Z}$ by

$$x \preccurlyeq y \text{ if and only if } x \leq y \text{ and } x+y \text{ is even}$$

for all $x, y \in \mathbb{Z}$. Prove that $\preccurlyeq$ is a partial order on $\mathbb{Z}$. Since $(\mathbb{Z}, \preccurlyeq)$ is a poset, answer the following questions:

(a) Is $S = \{1,2,3,4,5,6,\ldots\}$ a chain in $\mathbb{Z}$?
(b) Is $S = \{1,3,5,7,\ldots\}$ a chain in $\mathbb{Z}$?
(c) Does the set $S = \{1,2,3,4,5\}$ have a lower bound or an upper bound?
(d) Does the set $S = \{1,2,3,4,5\}$ have any maximal or minimal elements?

2. Let $\mathcal{P}(A)$ be the set of all subsets of the set $A = \{a,b,c\}$. Thus, $(\mathcal{P}(A), \subseteq)$ is a partially ordered set where

$$\mathcal{P}(A) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}.$$

Find an upper bound, the least upper bound, a lower bound, and the greatest lower bound of the following subsets of $\mathcal{P}(A)$.

(a) $S = \{\{a\}, \{a,b\}\}$
(b) $S = \{\{a\}, \{b\}\}$
(c) $S = \{\{a\}, \{a,b\}, \{a,b,c\}\}$
(d) $S = \{\{a\}, \{c\}, \{a,c\}\}$
(e) $S = \{\varnothing, \{a,b,c\}\}$
(f) $S = \{\{a\}, \{b\}, \{c\}\}$.

3. Prove Lemma 7.5.5.

4. Find the greatest lower bound of the set $S = \{15,20,30\}$ in the poset $(\mathbb{N}, |)$. Now find the least upper bound of S.

5. Let $(A, \preccurlyeq)$ be a poset and let $S \subseteq A$. Suppose that b is the largest element of S. Prove that b is also the least upper bound of S.

6. Let $\preccurlyeq$ be a partial order on a set A and let $S \subseteq A$. Suppose g and g' are both greatest lower bounds of S. Prove that $g = g'$.

7. Consider the poset $(\mathcal{P}(\mathbb{R}), \subseteq)$ where $\mathcal{P}(\mathbb{R}) = \{A : A \subseteq \mathbb{R}\}$, that is, $\mathcal{P}(\mathbb{R})$ is the set of all subsets of $\mathbb{R}$. Let C be the chain in $\mathcal{P}(\mathbb{R})$ defined by

$$C = \{\{1\}, \{1,2,3\}, \{1,2,3,4,5\}, \{1,2,3,4,5,6,7\}, \ldots\}$$

Does C have an upper bound? A least upper bound?

8. Consider the poset $(P, \subseteq)$ where $P = \{A : A \text{ is a finite subset of } \mathbb{R}\}$, that is, P is the set of all finite subsets of $\mathbb{R}$. Let C be the chain in P defined by

$$C = \{\{1\}, \{1,2,3\}, \{1,2,3,4,5\}, \{1,2,3,4,5,6,7\}, \ldots\}$$

Does C have an upper bound? A least upper bound?

9. Let $\preccurlyeq$ be a partial order on a set A and let $S \subseteq A$. Suppose a and a' are both smallest elements of S. Prove that $a = a'$.

10. Let $(A, \preccurlyeq)$ be a poset and let $S \subseteq A$. Suppose that a is the smallest element of S. Prove that a is also the greatest lower bound of S.

11. Let A be a set and let $(B, \preccurlyeq')$ be a poset. Suppose $h\colon A \to B$ is a one-to-one function. Define the relation $\preccurlyeq$ on A by $x \preccurlyeq y$ if and only if $h(x) \preccurlyeq' h(h)$ for all $x, y \in A$. Prove that $\preccurlyeq$ is a partial order on A.

12. Let $(A, \preccurlyeq)$ and $(B, \preccurlyeq')$ be posets. Suppose that a function $h\colon A \to B$ satisfies $x \preccurlyeq y$ if and only if $h(x) \preccurlyeq' h(h)$ for all $x, y \in A$. Prove that h is one-to-one.

13. Let $(A, \preccurlyeq)$ and $(B, \preccurlyeq')$ be posets. Suppose that a function $h\colon A \to B$ satisfies

$$x \preccurlyeq y \text{ if and only if } h(x) \preccurlyeq' h(h)$$

for all $x, y \in A$. Let $C \subseteq A$. Prove the following:

(a) If C is a chain in A, then the image $h[C]$ is a chain in B.
(b) If $h[C]$ is a chain in B, then C is a chain in A.
(c) If C has an upper bound, then $h[C]$ has an upper bound.
(d) If $h[C]$ has an upper bound and h is onto, then C has an upper bound.

14. Let $(A, \preccurlyeq)$ be a poset. For each $x \in A$, let $P_x = \{a \in A : a \preccurlyeq x\}$. Let $\mathcal{F}$ be the family of sets defined by $\mathcal{F} = \{P_x : x \in A\}$. Thus, $(\mathcal{F}, \subseteq)$ is a poset. Prove that $x \preccurlyeq y$ if and only if $P_x \subseteq P_y$, for all $x, y \in A$.

CHAPTER 8

Core Concepts in Abstract Algebra

Abstract algebra is typically the course where students are introduced to algebraic structures. What does 'algebraic' mean? As a means to answer this question, one of the first topics covered in an abstract algebra course is the important definition of a binary operation. This definition then leads to the study of algebraic structures and their properties. The eventual emphasis will be on two particular kinds of algebraic structures: groups and rings.

8.1 Binary Operations

A binary operation on a set A is a function that assigns to every ordered pair of elements in the set A a unique element that is also in A. Binary operations on a set A are frequently written as $a * b$, $a + b$, $a \cdot b$ or, more simply, as ab whenever $a, b \in A$. Typical examples of binary operations are the operations of addition and multiplication on the set of integers, as well as the operation of composition on a set of functions.

Definition 8.1.1. Let A be a nonempty set. We shall say that $*$ is a **binary operation** on A to mean that $*: A \times A \to A$, that is, $*$ is a function from $A \times A$ to A. Given $x, y \in A$, we shall write $x * y$ for the value $*(x, y)$.

1. The operation $*$ is **associative** on A if $(x * y) * z = x * (y * z)$ for all $x, y, z \in A$.[1]
2. The operation $*$ is **commutative** on A if $x * y = y * x$ for all $x, y \in A$.
3. The operation $*$ has an **identity element** if there is a member $e \in A$ such that $e * x = x * e = x$ for all $x \in A$.
4. Suppose $(A, *)$ has an identity element e. For $x \in A$ if there is a $y \in A$ such the $x * y = y * x = e$, then y is an **inverse** of x.

A binary operation combines two elements from a set and produces another element in the same set. In some cases a set may have several binary operations on it. Each such operation will then be represented by a distinguished name/symbol, for example, $\cdot$, $*$, $\circ$, or $+$. These symbols may, or may not, denote the usual operations of addition and multiplication, as we will see throughout this chapter.

[1] Associativity implies than one can write $x * y * z$ without ambiguity.

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_8, 239

Some Examples of Binary Operations on Sets

Example 1. The operation of subtraction $(-)$ is a binary operation on $\mathbb{Z}$. This binary operation is not associative, for example, $(7-2)-5 \neq 7-(2-5)$. It is also not commutative since $5-3 \neq 3-5$.

Example 2. The operation of subtraction $(-)$ is not a binary operation on $\mathbb{N}$, because $3-5=-2 \notin \mathbb{N}$; that is, there is a pair of natural numbers whose difference is not a natural number.

Example 3. A 2×2 **matrix** is a rectangular array of real numbers having the form

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and the number a_{ij} is called the ij-th entry of the matrix A. Let $M_2(\mathbb{R})$ be the set of all such 2×2 matrices with entries from $\mathbb{R}$. The binary operation $+$, called *matrix addition*, is defined by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

This binary operation is associative and commutative. In addition, this binary operation has the identity element $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and each matrix in $M_2(\mathbb{R})$ has an (additive) inverse (why?).

Example 4. Let $M_2(\mathbb{R})$ be the set of 2×2 matrices with entries from $\mathbb{R}$. Consider the binary operation of **matrix multiplication** $*$ defined by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}.$$

Matrix multiplication is an associative binary operation; but it is not commutative, as there are matrices A and B in $M_2(\mathbb{R})$ so that $A*B \neq B*A$. For example, let $A = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then $A*B = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B*A = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$. In addition, this binary operation has the identity element $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; however, not every matrix in $M_2(\mathbb{R})$ has a (multiplicative) inverse.

Example 5. Let $m \in \mathbb{N}$ and let $\mathbb{Z}_m = \{[0],[1],\ldots,[m-1]\}$ be the set of congruence classes (mod m). For $[a],[b] \in \mathbb{Z}_m$ define the binary operations $\oplus$ and $\odot$ on $\mathbb{Z}_m$ by

$$[a] \oplus [b] = [a+b]$$
$$[a] \odot [b] = [a \cdot b].$$

These binary operations are associative and commutative (see Theorem 7.4.4).

8.1.1 Is it a Binary Operation?

A binary operation $\divideontimes$ on a set A is a function $\divideontimes\colon A \times A \to A$. We will now restate Definition 6.1.1 on page 169 so that it directly applies to such functions.

Definition 8.1.2. Let A be a set, and let $\divideontimes \subseteq (A \times A) \times A$. Then $\divideontimes$ is said to be a **function from** $A \times A$ **to** A if the following two conditions hold:

(1) For each $(x,y) \in A \times A$ there is a $z \in A$ such that $((x,y),z) \in \divideontimes$.
(2) If $((x,y),z) \in \divideontimes$ and $((x,y),w) \in \divideontimes$, then $z = w$.

Thus, for each $x \in A$ and $y \in A$ there is exactly one $z \in A$ such that $((x,y),z) \in \divideontimes$. This unique z is called "the value of $\divideontimes$ at (x,y)" and is denoted by $\divideontimes(x,y)$ or $x \divideontimes y$. We shall refer to the function $\divideontimes$ as a **binary operation** on A.

At some point in your first abstract algebra course the topic of a "well-defined binary operation" will be discussed. Your instructor may then describe a purported binary operation $\divideontimes$ on a set A and prove that it is a function by verifying that $\divideontimes$ satisfies the above properties (1) and (2). This is usually done when the elements of the set A have multiple representations and when the proposed operation is described in terms of such representations. As a result, it is usually not at all clear that the proposed operation is in fact a function (see Section 6.1.1 starting on page 171).

Example 6. Each element in the set $\mathbb{Z}_6$ of congruence classes (mod 6) has multiple representations, for example, $[3]_6 = [9]_6 = [15]_6$ (see Lemma 7.3.8(1)). Consider the proposed binary operation $*$ on $\mathbb{Z}_6$ described by rule

$$[a]_6 * [b]_6 = [\max(a,b)]_6 \tag{8.1}$$

where $\max(a,b)$ is the maximum of the two integers a and b. Is $*$ a function from $\mathbb{Z}_6 \times \mathbb{Z}_6$ to $\mathbb{Z}_6$? It is not obvious! Show that $*$ is not a function.

Solution. Suppose, for a contradiction, that $*$ is a function. The rule in (8.1) is thus an abbreviation for the more formal description (see Example 4 on page 172)

$$x * y = z \text{ iff } (\exists a \in \mathbb{Z})(\exists b \in \mathbb{Z})\,(\,(x,y) = ([a]_6,[b]_6) \text{ and } z = [\max(a,b)]_6\,) \tag{8.2}$$

for elements x, y and z in $\mathbb{Z}_6$. Let $x = [2]_6$ and $y = [4]_6$. Since $z = [\max(2,4)]_6 = [4]_6$, the above description (8.2) would allow us to conclude that $x * y = [4]_6$. Moreover, by Lemma 7.3.8(1), we have that $[2]_6 = [8]_6$. Hence, $x = [8]_6$ and $y = [4]_6$. Upon evaluating $z = [\max(8,4)]_6 = [8]_6$, we can also conclude from (8.2) that $x * y = [8]_6$. Thus, $x * y = [4]_6$ and $x * y = [8]_6$. Since $[4]_6 \neq [8]_6$, we have derived a contradiction. Therefore, $*$ is not a function. Ⓢ

Proposition 8.1.3. *Let m be a natural number and let $\mathbb{Z}_m$ be the set of congruence classes* (mod m). *Consider the putative binary operation $\divideontimes$ on $\mathbb{Z}_m$ described by the rule*

$$[a]_m \divideontimes [b]_m = [ab - a]_m \tag{8.3}$$

where a and b are integers. Then $\divideontimes$ is a binary operation on $\mathbb{Z}_m$.

Proof Analysis. The intended subset $\divideontimes$ of $(\mathbb{Z}_m \times \mathbb{Z}_m) \times \mathbb{Z}_m$ satisfies

$$((x,y),z) \in \divideontimes \text{ iff } (\exists a \in \mathbb{Z})(\exists b \in \mathbb{Z})\,((x,y) = ([a]_m,[b]_m) \text{ and } z = [ab-a]_m)$$

(see Proposition 6.1.5 on page 173). We show how Remark 6.1.6 (on page 173) can be adapted to prove that $\divideontimes$ is a function of two inputs. We must first prove that $\divideontimes$ satisfies property (1) of Definition 8.1.2. Let $(x,y) \in \mathbb{Z}_m \times \mathbb{Z}_m$. So there are integers a and b so that $x = [a]_m$ and $y = [b]_m$. Since $z = [ab-a]_m$ is also in $\mathbb{Z}_m$, we see that $((x,y),z) \in \divideontimes$. To prove that $\divideontimes$ satisfies property (2) of Definition 8.1.2, suppose that $((x,y),z) \in \divideontimes$ and $((x,y),w) \in \divideontimes$. We must prove that $z = w$. Since $((x,y),z) \in \divideontimes$, there are integers a and b such that $x = [a]_m$, $y = [b]_m$ and $z = [ab-a]_m$. Similarly, as $((x,y),w) \in \divideontimes$, we have that $x = [c]_m$, $y = [d]_m$ and $w = [cd-c]_m$ for some integers c and d. We now have that

$$x = [a]_m = [c]_m \text{ and } y = [b]_m = [d]_m.$$

Thus, to show that $z = w$ we just need to prove that

$$([a]_m = [c]_m \text{ and } [b]_m = [d]_m) \text{ implies } [ab-a]_m = [cd-c]_m. \tag{8.4}$$

Hence, a proof of Proposition 8.1.3 can be accomplished as follows: Let a,b,c,d be integers. To prove that $\divideontimes$ is a function, we must verify that properties (1) and (2) of Definition 8.1.2 are satisfied. That property (1) holds, is obvious because $[ab-a]_m \in \mathbb{Z}_m$ when $[a]_m \in \mathbb{Z}_m$ and $[b]_m \in \mathbb{Z}_m$. For this reason, we will omit the proof of (1). To prove property (2), it is sufficient to prove (8.4) as this implication shows that the rule (8.3) is independent of the representations used for a pair of inputs $x \in \mathbb{Z}_m$ and $y \in \mathbb{Z}_m$. We will now prove that $\divideontimes$ is a function; that is, we shall prove that $\divideontimes$ is a well-defined binary operation. Ⓐ

Proof. Let a,b,c,d be integers. Assume $[a]_m = [c]_m$ and $[b]_m = [d]_m$. Lemma 7.3.8(1) implies $a \equiv c \pmod m$ and $b \equiv d \pmod m$. From $b \equiv d \pmod m$, we conclude that $(b-1) \equiv (d-1) \pmod m$ by Theorem 7.3.4(2). Since $a \equiv c \pmod m$, we obtain $(ab-a) \equiv (cd-c) \pmod m$ by Theorem 7.3.5(3). Thus, $[ab-a]_m = [cd-c]_m$. Therefore, $\divideontimes$ is a binary operation on $\mathbb{Z}_m$. □

Remark 8.1.4. As was noted in the above proof analysis, the formula (8.3) easily implies property (1) of Definition 8.1.2. Thus, in our proof of Proposition 8.1.3 we did not verify that this property holds. Most mathematicians and textbooks will do the same whenever it is obvious that property (1) holds.

Example 7. Each rational number also has multiple representations, for example, $\frac{3}{4} = \frac{6}{8} = \frac{9}{12}$. Consider the proposed binary operation $*$ on $\mathbb{Q}$ described by

$$\frac{m}{n} * \frac{a}{b} = \frac{(m-1)(a-1)}{nb} \tag{8.5}$$

where m,n,a,b are integers and n,b are nonzero. Is $*$ a function from $\mathbb{Q}\times\mathbb{Q}$ to $\mathbb{Q}$? It is not obvious! Show that $*$ is not a function.

Solution. Suppose, for a contradiction, that $*$ is a function. Thus, the rule (8.5) yields the more formal description (see Example 4 on page 172):

$$x*y=z \text{ iff } (x,y)=\left(\frac{m}{n},\frac{a}{b}\right) \text{ and } z=\frac{(m-1)(a-1)}{nb} \quad \text{for some } m,n,a,b\in\mathbb{Z} \text{ where } n\neq 0 \text{ and } b\neq 0. \tag{8.6}$$

Let $x=\frac{2}{3}$ and $y=\frac{3}{4}$. Since $z=\frac{(2-1)(3-1)}{12}=\frac{1}{6}$, the description (8.6) yields $x*y=\frac{1}{6}$. We also know that $\frac{2}{3}=\frac{4}{6}$. Thus, $x=\frac{4}{6}$ and $y=\frac{3}{4}$. Because $z=\frac{(4-1)(3-1)}{24}=\frac{1}{4}$, we can also conclude from (8.6) that $x*y=\frac{1}{4}$. We conclude that $x*y=\frac{1}{6}$ and $x*y=\frac{1}{4}$. Since $\frac{1}{6}\neq\frac{1}{4}$, we have derived a contradiction. Therefore, $*$ is not a function. Ⓢ

Proposition 8.1.5. *Let $\mathbb{Q}$ be the set of rational numbers. Consider the purported binary operation on $\mathbb{Q}$ described by the rule*

$$\frac{m}{n}*\frac{i}{j}=\frac{mj+ni}{nj} \tag{8.7}$$

where m,n,i,j are integers and n,j are nonzero. Then $*$ is a binary operation on $\mathbb{Q}$.

Proof Analysis. After adapting Remark 6.1.6, a proof of Proposition 8.1.5 can be attained as follows: Let $\frac{i}{j},\frac{a}{b},\frac{m}{n},\frac{c}{d}$ be rational numbers, where the numerators are integers and the denominators are nonzero integers. To prove that $*$ is a function, we must verify that properties (1) and (2) of Definition 8.1.2 are satisfied. To prove that (1) holds, it is enough to observe that $\frac{in+jm}{jn}\in\mathbb{Q}$ when $\frac{i}{j}\in\mathbb{Q}$ and $\frac{m}{n}\in\mathbb{Q}$. As this is clear, we shall not mention it in our proof below. To prove property (2), it is sufficient to prove that (see the proof of Proposition 6.1.5)

$$\left(\frac{m}{n}=\frac{a}{b} \text{ and } \frac{i}{j}=\frac{c}{d}\right) \text{ implies } \frac{mj+ni}{nj}=\frac{ad+bc}{bd}$$

as this implication shows that the rule (8.7) is independent of the representations used for a pair of inputs $x\in\mathbb{Q}$ and $y\in\mathbb{Q}$. Ⓐ

Proof. Let $*$ be the rule described by (8.7). Let $\frac{i}{j},\frac{a}{b},\frac{m}{n},\frac{c}{d}$ be rational numbers, where the numerators are integers and the denominators are nonzero integers. Assume that $\frac{m}{n}=\frac{a}{b}$ and $\frac{i}{j}=\frac{c}{d}$. Thus, $(\star)$ $mb=na$ and (#) $id=jc$, by Definition 2.1.3. We prove that $(mj+ni)bd=nj(ad+bc)$ as follows:

$$\begin{aligned}(mj+ni)bd&=(mj)(bd)+(ni)(bd) && \text{by distributivity}\\ &=(jd)(mb)+(nb)(id) && \text{by commutativity and associativity}\\ &=(jd)(na)+(nb)(jc) && \text{by } (\star) \text{ and (\#)}\end{aligned}$$

$$= (nj)(ad) + (nj)(bc) \quad \text{by commutativity and associativity}$$
$$= nj(ad + bc) \quad \text{by distributivity.}$$

Hence, $(mj+ni)bd = nj(ad+bc)$ and from Definition 2.1.3 we conclude that $\frac{mj+ni}{nj} = \frac{ad+bc}{bd}$. Therefore, $*$ is a binary operation on $\mathbb{Q}$. □

In elementary school we learned to use the right hand side of the rule (8.7) to add two rational numbers $\frac{m}{n}$ and $\frac{i}{j}$. School teachers rarely, if at all, explain to their students why this rule describes a well-defined binary operation.

Exercises 8.1

1. Define the binary operation $*$ on $\mathbb{R}$ by $x * y = e^{xy}$ for all $x, y \in \mathbb{R}$. Show that for all $x > 0$ there is a $y > 0$ such that $x * y = x$.

2. Define the binary operation $*$ on $\mathbb{R}$ by $x * y = \sin(xy)$ for all $x, y \in \mathbb{R}$. Show that for all $x \in \mathbb{R}$,

(a) If $x \neq 0$, then there is a y such that $x * y = 1$;
(b) If $-1 \leq x \leq 1$, then there is a y such that $x * y = x$.

3. Define the binary operation $*$ on $\mathbb{R}$ by $x * y = \sin(xy)$ for all $x, y \in \mathbb{R}$. Show that this binary operation does not have an identity element.

4. For each of the following binary operations $*$ on $\mathbb{R}$, answer the three questions: Is $*$ commutative? Is $*$ associative? Does $*$ have an identity element?

(a) $x * y = xy + x + y$
(b) $x * y = x + xy$
(c) $x * y = |x| + |y|$.

5. Using Definition 2.1.3 (as in the proof of Proposition 8.1.5), prove that the following purported binary operations on $\mathbb{Q}^+$ are well-defined.

(a) $\frac{m}{n} * \frac{i}{j} = \frac{ni}{mj}$
(b) $\frac{m}{n} * \frac{i}{j} = \frac{(mi)^2}{(nj)^2}$

where m, n, i, j are natural numbers.

6. Let $m \geq 1$ be an integer and let $\mathbb{Z}_m$ be the set of congruence classes (mod m). Define the alleged binary operation $*$ on $\mathbb{Z}_m$ by $[a]_m * [b]_m = [a^2 + b^2]_m$ for all $[a]_m, [b]_m \in \mathbb{Z}_m$. Prove that $*$ is well-defined.

7. Let $\mathbb{Q}$ be the set of rational numbers. Show that the following purported binary operations on $\mathbb{Q}$ are not functions from $\mathbb{Q} \times \mathbb{Q}$ to $\mathbb{Q}$.

(a) $\frac{m}{n} * \frac{i}{j} = \frac{mi}{n}$
(b) $\frac{m}{n} * \frac{i}{j} = \frac{i}{n}$
(c) $\frac{m}{n} * \frac{i}{j} = \frac{i+m}{j+n}$

where m, n, i, j are integers and n, j are nonzero.

8. Let $F(\mathbb{R})$ be the set of all functions $f\colon \mathbb{R} \to \mathbb{R}$. The operation of composition $\circ$ is a binary operation on $F(\mathbb{R})$. Show that $\circ$ has an identity element.

9. Let $\mathbb{Z}_5$ be the set of congruence classes $(\text{mod } 5)$. Define the proposed binary operation $*$ on $\mathbb{Z}_5$ by $[a]_5 * [b]_5 = [|ab|]_5$ for all $[a]_5, [b]_5 \in \mathbb{Z}_5$, where $|ab|$ is the absolute value of ab. Show that $*$ is not a binary operation on $\mathbb{Z}_5$.

8.2 Algebraic Structures

Definition 8.2.1. Let A be a nonempty set with one or more binary operations $*, +, \dots$ on A. We shall call $\mathcal{A} = (A, *, +, \dots)$ an **algebraic structure**.

When it is said that $\mathcal{A} = (A, *, +, \dots)$ is an algebraic structure, one should not presume that $\mathcal{A}$ satisfies any specific 'axioms' or properties (e.g., associativity). $\mathcal{A}$ is an algebraic structure if and only if the binary operations $*, +, \dots$ are defined on all pairs of elements in the nonempty set A. Definition 8.2.1 coincides with the one given by Bourbaki [2, page xxii] followed by the statement

> for the structures defined in this way we reserve precisely the name *algebraic structures* and it is the study of these which constitutes Algebra.

Thus, abstract algebra is the study of algebraic structures and their properties.

Examples of Algebraic Structures

In elementary school and in high school we learned about the algebraic structures presented in Examples 1–3 below.

Example 1. Let $\mathbb{Z}$ be the set of integers. Let $\cdot$ be ordinary multiplication and let $+$ be ordinary addition. Then $(\mathbb{Z}, +, \cdot)$ is an algebraic structure.

Example 2. Let $\mathbb{Q}$ be the set of rational numbers. Let $\cdot$ be ordinary multiplication and let $+$ be ordinary addition. Then $(\mathbb{Q}, +, \cdot)$ is an algebraic structure.

Example 3. Let $\mathbb{R}$ be the set of real numbers and let $+$ and $\cdot$ be standard addition and multiplication, respectively. Thus $(\mathbb{R}, +, \cdot)$ is an algebraic structure.

The following algebraic structure is one that has two unusual binary operations.

Example 4. Define two binary operations on $\mathbb{R}$ by $x * y = e^{xy}$ and $x \diamondsuit y = \sin(xy)$ for all $x, y \in \mathbb{R}$. Then $(\mathbb{R}, *, \diamondsuit)$ is an algebraic structure.

Our last four examples present algebraic structures that you may see again in your future mathematics courses.

Example 5. Let $F(\mathbb{R})$ be the set of all functions of the form $f\colon \mathbb{R} \to \mathbb{R}$. For f and g in $F(\mathbb{R})$, define $(f \cdot g)\colon \mathbb{R} \to \mathbb{R}$ by $(f \cdot g)(x) = f(x)g(x)$ and define $(f + g)\colon \mathbb{R} \to \mathbb{R}$ by $(f + g)(x) = f(x) + g(x)$. Then $(F(\mathbb{R}), +, \cdot)$ is an algebraic structure.

Example 6. Let $F(\mathbb{Z})$ be the set of all functions of the form $f\colon \mathbb{Z} \to \mathbb{Z}$. Let f and g be functions in $F(\mathbb{Z})$. Define the function $(f \cdot g)\colon \mathbb{Z} \to \mathbb{Z}$ by $(f \cdot g)(x) = f(x)g(x)$ and define the function $(f+g)\colon \mathbb{Z} \to \mathbb{Z}$ by $(f+g)(x) = f(x)+g(x)$. Then $(F(\mathbb{Z}), +, \cdot)$ is an algebraic structure.

Example 7. Let $M_2(\mathbb{R})$ be the set of all such 2×2 matrices with entries from $\mathbb{R}$. Then $(M_2(\mathbb{R}), +, *)$ is an algebraic structure where $+$ is matrix addition and $*$ is matrix multiplication, as defined in Examples 3 and 4 of the previous section.

Example 8. Let $m > 1$ be an integer. By Theorem 7.4.4, we see that $(\mathbb{Z}_m, \oplus, \odot)$ is an algebraic structure.

8.2.1 Substructures

Definition 8.2.2. Let $*$ be a binary operation on a set A and let $S \subseteq A$ be nonempty. The set S is **closed under** the binary operation $*$ if and only if $x * y \in S$ for all $x, y \in S$.

Definition 8.2.3. Let $(A, *, +, \dots)$ be an algebraic structure and let $S \subseteq A$. When S is closed under all of the binary operations $*, +, \dots$, we shall say that $(S, *, +, \dots)$ is a **substructure** of $(A, *, +, \dots)$ or, more succinctly, that S is a *substructure* of A.

Example 9. Let $(\mathbb{Z}, +, \cdot)$ be the algebraic structure in Example 1. Let $E \subseteq \mathbb{Z}$ be the set of even integers. Since E is closed under addition and multiplication, we see that E is a substructure of $\mathbb{Z}$.

Example 10. Let $(\mathbb{R}, +, \cdot)$ be as in Example 3. Let $\mathbb{Q} \subseteq \mathbb{R}$ be the set of rational numbers. Because $\mathbb{Q}$ is closed under addition and multiplication, we can conclude that $\mathbb{Q}$ is a substructure of $\mathbb{R}$.

Example 11. Recall the algebraic structure $(M_2(\mathbb{R}), *, +)$ presented in Example 7. Let $S \subseteq M_2(\mathbb{R})$ be defined by

$$S = \left\{ \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} : x, y \in \mathbb{R} \right\}.$$

One can easily check that S is closed under matrix addition and multiplication. Thus, S is a substructure of $M_2(\mathbb{R})$.

Exercises 8.2

1. Let $(A, *, +, \dots)$ be an algebraic structure and let B be a substructure of A. Suppose that $*$ is commutative on A. Show that $*$ is commutative on B.

2. Let $(A,*)$ be an algebraic structure. Let $a,b \in A$ satisfy

$$a*x = x*a = x \text{ for all } x \in A$$
$$b*x = x*b = x \text{ for all } x \in A.$$

Prove that $a = b$.

3. Let $(\mathbb{R},+,\cdot)$ be the algebraic structure in Example 3. Let $S \subseteq \mathbb{R}$ be defined by $S = \{i + j\sqrt{2} : i,j \in \mathbb{Z}\}$. Show that S is a substructure of $\mathbb{R}$.

4. Let $(\mathbb{R},+,\cdot)$ be as in Example 3. Define $S \subseteq \mathbb{R}$ by $S = \{a + b\sqrt{3} : a,b \in \mathbb{Q}\}$. Show that S is a substructure of $\mathbb{R}$.

5. Let $(A,*,+)$ be an algebraic structure and let B and C be substructures of A. Suppose $B \cap C \neq \varnothing$. Prove that $B \cap C$ is a substructure of A.

6. Consider the algebraic structure $(F(\mathbb{R}),+,\cdot)$ defined in Example 5. Show that S is a substructure of $F(\mathbb{R})$ where $S = \{f \in F(\mathbb{R}) : f(1) = 0\}$.

7. Let $(F(\mathbb{Z}),+,\cdot)$ be the algebraic structure defined in Example 6. Show that S is a substructure of $F(\mathbb{Z})$ where $S = \{f \in F(\mathbb{Z}) : 5 \mid f(i) \text{ for all } i \in \mathbb{Z}\}$.

8. Let $F(\mathbb{R})$ be the set of all functions $f : \mathbb{R} \to \mathbb{R}$. For each f and g in $F(\mathbb{R})$, define the function $(f \circ g) : \mathbb{R} \to \mathbb{R}$ by $(f \circ g)(x) = f(g(x))$. Clearly we have that $(F(\mathbb{R}),\circ)$ is an algebraic structure. Let $I \subseteq F(\mathbb{R})$ be the set of those functions in $F(\mathbb{R})$ that are one-to-one. Let $S \subseteq F(\mathbb{R})$ be the set of those functions in $F(\mathbb{R})$ that are onto. Show that I and S are both substructures of $F(\mathbb{R})$.

8.3 Groups

A group is an algebraic structure $(G,*)$ with only one binary operation that satisfies three properties. Since groups are so important both within and outside of mathematics, group theory[2] has become a central subject in contemporary mathematics.

Definition 8.3.1. An algebraic structure $(G,*)$, with one binary operation $*$, is a **group** if the three GROUP AXIOMS hold:

1. $(a*b)*c = a*(b*c)$ for all $a,b,c \in G$ (associative law).
2. There is a distinguished element $e \in G$ satisfying $e*a = a*e = a$ for all $a \in G$ (e is called the **identity element**).
3. For all $a \in G$ there exists a $b \in G$ such that $a*b = b*a = e$ (b is written as a^{-1} and is called the **inverse** of a).

[2]Group Theory is the branch of mathematics that deals with groups.

We will soon show that a group has exactly one identity element and that each element of a group has exactly one inverse. We note that associativity implies that one can write $a * b * c$ without ambiguity, for any three elements a, b, c in a group.

Remark 8.3.2. Let $(G, *)$ be a group. Now let a and b be in G, and let n be a natural number. We shall write:

1. ab as shorthand for the "product" $a * b$;
2. a^n as shorthand for the "product" $\underbrace{aaa \cdots a}_{n \text{ times}}$;
3. a^{-n} as shorthand for $(a^{-1})^n$, that is, for the "product" $\underbrace{a^{-1}a^{-1}a^{-1} \cdots a^{-1}}_{n \text{ times}}$;
4. a^0 is defined to be e, the identity element in G, that is, $a^0 = e$.

One can prove, using mathematical induction, that $e^i = e$, $a^i a^j = a^{i+j}$, $(a^i)^j = a^{ij}$, and $(a^i)^{-1} = a^{-i}$ for all $i, j \in \mathbb{N}$.

Definition 8.3.3 (Abelian Group). Let $(G, *)$ be a group. We say that G is an **abelian group** if $ab = ba$ *for all* $a, b \in G$.

An abelian group is also called a *commutative group*. A group that is not abelian is called *nonabelian*.

Examples of Groups

To create groups we must first define the set G and define the binary operation $*$. Then we must verify that the structure $(G, *)$ satisfies the group axioms. We now present several examples of groups.

Example 1. Define a binary operation $*$ on the set of integers $\mathbb{Z}$ by $a * b = a + b$ whenever $a, b \in \mathbb{Z}$, where $+$ is ordinary addition. We shall verify that the $(\mathbb{Z}, +)$ satisfies the group axioms.

1. The associative law holds for $(\mathbb{Z}, +)$, because addition is associative.
2. Since $a + 0 = 0 + a = a$ for all $a \in \mathbb{Z}$, the required identity element is $e = 0$.
3. For $a \in \mathbb{Z}$, because $a + (-a) = (-a) + a = 0$, we conclude that $a^{-1} = -a$ is the required inverse of a.

Since all of the group axioms hold, we see that $(\mathbb{Z}, +)$ is a group.

Example 2. Let $+$ be the usual binary operation of addition on $\mathbb{Q}$, the set of rational numbers. As in Example 1, one can easily verify that $(\mathbb{Q}, +)$ satisfies the group axioms. Thus, $(\mathbb{Q}, +)$ is a group.

Example 3. Let $\mathbb{R}^* = \{x \in \mathbb{R} : x \neq 0\}$ be the set of nonzero real numbers and let $\cdot$ be the standard operation of multiplication on the real numbers. We shall show that algebraic structure $(\mathbb{R}^*, \cdot)$ satisfies the group axioms.

1. Since multiplication is associative, the associative law holds for $(\mathbb{R}^*, \cdot)$.
2. Clearly $e = 1$ is the identity element, because $a \cdot 1 = 1 \cdot a = a$ for all $a \in \mathbb{R}^*$.
3. For $a \in \mathbb{R}^*$, since $a \cdot (\frac{1}{a}) = (\frac{1}{a}) \cdot a = 1$, the required inverse element is $a^{-1} = \frac{1}{a}$.

Since all of the group axioms hold, we see that $(\mathbb{R}^*, \cdot)$ is a group.

Example 4. Let $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$ and $\cdot$ be the usual operation of multiplication on the real numbers. We now confirm that $(\mathbb{R}^+, \cdot)$ satisfies the group axioms.

1. The associative law holds for $(\mathbb{R}^+, \cdot)$, as multiplication is associative.
2. Let $e = 1 \in \mathbb{R}^+$. Clearly, 1 is an identity element for ordinary multiplication.
3. For each $a \in \mathbb{R}^+$, since $a \cdot (\frac{1}{a}) = (\frac{1}{a}) \cdot a = 1$, the inverse of a is $a^{-1} = \frac{1}{a} > 0$.

Since all of the group axioms hold, we see that $(\mathbb{R}^+, \cdot)$ is a group.

All of the groups in Examples 1–4 are abelian groups. Our next three examples present groups that are not abelian.

Example 5. For each $a, b \in \mathbb{R}$ with $a \neq 0$, define the function $T_{a,b}\colon \mathbb{R} \to \mathbb{R}$ by $T_{a,b}(x) = ax + b$. Let $G = \{T_{a,b} : a, b \in \mathbb{R} \text{ and } a \neq 0\}$. Now we define $T_{a,b} \circ T_{c,d}$ to be ordinary function composition. Let $T_{a,b} \in G$ and $T_{c,d} \in G$. So $a \neq 0$ and $c \neq 0$. One can show that $T_{a,b} \circ T_{c,d} = T_{ac,ad+b}$ (review Exercise 4 on page 188). Since $ac \neq 0$, we conclude that $T_{a,b} \circ T_{c,d} \in G$, that is, G is closed under $\circ$ and hence, $(G, \circ)$ is an algebraic structure.

1. The associative law holds for $(G, \circ)$, since functional composition is associative (see Exercise 10 on page 189).
2. Let $e = T_{1,0}$ be the identity function. Since $T \circ T_{1,0} = T_{1,0} \circ T = T$ for every $T \in G$, the function $T_{1,0}$ is the identity element for the binary operation $\circ$.
3. Let $T_{a,b} \in G$. Since $T_{a,b}^{-1} = T_{\frac{1}{a}, -\frac{b}{a}}$ (Exercise 4 on page 188) and $\frac{1}{a} \neq 0$, we see that $T_{a,b}^{-1}$ is also in G. Furthermore, $\left(T_{a,b} \circ T_{a,b}^{-1}\right) = \left(T_{a,b}^{-1} \circ T_{a,b}\right) = T_{1,0}$ by Corollary 6.3.5. Thus, $(T_{a,b})^{-1} = T_{a,b}^{-1}$ is the requisite inverse element.

Since all of the group axioms hold, we see that $(G, *)$ is a group. To see that this group is not abelian, review Exercise 4 on page 188.

Example 6. Consider the set $G = \{e, u, v, w, x, y\}$ and let $*$ be the binary operation defined by the following "multiplication table" where ${}_c*^r$, in the upper-left corner of the table,

${}_c*^r$	e	u	v	w	x	y
e	e	u	v	w	x	y
u	u	v	e	x	y	w
v	v	e	u	y	w	x
w	w	y	x	e	v	u
x	x	w	y	u	e	v
y	y	x	w	v	u	e

stands for "column entry" $*$ "row entry." For example, $u*x=y$ and $x*u=w$. One can check that the group axioms hold for $(G,*)$; however, when working with a multiplication table, it is often tedious to verify associativity. On the other hand, using the above multiplication table, we see that $u^{-1}=v$, $v^{-1}=u$, $w^{-1}=w$, $x^{-1}=x$, and $y^{-1}=y$. Since $u*x\neq x*u$, the group is not abelian.

We offer another example of a nonabelian group that closely resembles the group presented in Example 5.

Example 7. Let $M_2(\mathbb{R})$ be the set of all 2×2 matrices over the real numbers. Let G be the set of matrices A in $M_2(\mathbb{R})$ that have the form $A=\begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix}$ where $a\neq 0$. One can show that $A*B\in G$ for all $A\in G$ and $B\in G$, where the binary operation $*$ is matrix multiplication as defined in Example 4 on page 240. Thus, $(G,*)$ is an algebraic structure. Furthermore, one can show that $(G,*)$ is a nonabelian group with identity element $I=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (see Exercise 2).

Some Structures that are not Groups

Example 8. Let $\cdot$ be ordinary multiplication on the set of integers $\mathbb{Z}$. Then $(\mathbb{Z},\cdot)$ is an algebraic structure because $\mathbb{Z}$ is closed under multiplication. Is $(\mathbb{Z},\cdot)$ a group?

1. Since multiplication is associative, we see that the associative law holds for $(\mathbb{Z},\cdot)$.
2. Clearly, 1 is an identity element for multiplication.
3. Let $a=2$. Since there is no integer i such that $i\cdot 2=2\cdot i=1$, axiom 3 does not hold for the structure $(\mathbb{Z},\cdot)$.

Since one of the group axioms fails, we see that $(\mathbb{Z},\cdot)$ is a not a group.

Example 9. Let $G=\{x\in\mathbb{R}: x\neq 0\}$, the set of non-zero real numbers. For each $a,b\in G$, define $a*b=a^2\cdot b$ where $\cdot$ is ordinary multiplication. Since the product of non-zero real numbers is non-zero, G is closed under $*$ and so, $(G,*)$ is an algebraic structure. To show that $(G,*)$ is not a group, we must show that $(G,*)$ does not satisfy at least one of the group axioms. Does the associative law hold for $(G,*)$? Let $a=3,b=1,c=1$. Then

$$a*(b*c)=a*(b^2\cdot c)=a^2\cdot(b^2\cdot c)=3^2\cdot(1^2\cdot 1)=9$$

$$(a*b)*c=(a^2\cdot b)*c=(a^2\cdot b)^2\cdot c=(3^2\cdot 1)^2\cdot 1=81.$$

Thus $a*(b*c)\neq a*(b*c)$ and so, the associative law does not hold for $(G,*)$. Since one of the group axioms fails, we see that $(G,*)$ is a not a group.

8.3.1 Fundamental Properties of a Group

In this section we identify the important properties that are shared by all groups. For a group $(G,*)$ and $a,b \in G$, we will write ab as shorthand for the "product" $a*b$.

Lemma 8.3.4 (Uniqueness of the identity). *Let $(G,*)$ be a group. Then there is exactly one element $b \in G$ satisfying $xb = bx = x$ for all $x \in G$.*

Proof. Since G is a group, there is an identity element $e \in G$ such that for all $x \in G$

$$xe = x. \tag{8.8}$$

Suppose that $c \in G$ is also an identity element in G. We shall prove that $c = e$. Since c is an identity element, we see that for *all* $x \in G$

$$cx = x. \tag{8.9}$$

Letting $x = c$ in (8.8), we infer $ce = c$. Substituting $x = e$ into (8.9), we get $ce = e$. Therefore, $c = e$. □

Lemma 8.3.5 (Cancellation Laws). *Let $(G,*)$ be a group. Let $a,b,c \in G$. Then the following hold:*

(1) *If $ab = ac$, then $b = c$.*
(2) *If $ba = ca$, then $b = c$.*

Proof. To prove (1), assume $ab = ac$. Let a^{-1} be an inverse of a. Since $ab = ac$, we see that $a^{-1}(ab) = a^{-1}(ac)$. By associativity, we obtain $(a^{-1}a)b = (a^{-1}a)c$. Thus $eb = ec$ and so, $b = c$. A similar proof will establish (2). □

Lemma 8.3.6 (Uniqueness of an inverse). *Let $(G,*)$ be a group. For each $a \in G$ there is exactly one element $b \in G$ such that $ab = ba = e$.*

Proof. Let $a \in G$. Since G is a group, we know that there is a $b \in G$ that acts as an inverse for a. Hence,

$$ab = e. \tag{8.10}$$

Suppose that $c \in G$ is also acts as an inverse for a. We prove that $c = b$. Since c is an inverse for a, we see that

$$ac = e. \tag{8.11}$$

From (8.10) and (8.11) we obtain $ac = ab$. Therefore, $c = b$ by Lemma 8.3.5. □

Lemma 8.3.7. *Let $(G,*)$ be a group and let $x \in G$ and $y \in G$. If $xy = e$ or $yx = e$, then $x^{-1} = y$.*

Proof. Let $x \in G$ and $y \in G$. Assume $xy = e$. Thus, $x^{-1}xy = x^{-1}e$ and so, $y = x^{-1}$. Hence, $x^{-1} = y$. Similarly, if $yx = e$, then $yxx^{-1} = ex^{-1}$ which implies $x^{-1} = y$. □

Remark 8.3.8. Let G be a group and let $a, x \in G$. To prove an equation of the form $x^{-1} = a$, Lemma 8.3.7 implies that it is sufficient to prove that $ax = e$ or $xa = e$. In particular, since $ee = e$, it follows that $e^{-1} = e$.

We will apply Remark 8.3.8 in the proof of our next lemma.

Lemma 8.3.9. *Let $(G, *)$ be a group. Let $a \in G$. Then $(a^{-1})^{-1} = a$.*

Proof. Let $e \in G$ be the identity element. Let $a \in G$ and let a^{-1} be the inverse of a. Since $aa^{-1} = e$, Lemma 8.3.7 implies that $(a^{-1})^{-1} = a$. □

Lemma 8.3.10. *Let $(G, *)$ be a group. For every $a, b \in G$, we have $(ab)^{-1} = b^{-1}a^{-1}$.*

Proof. See Exercise 3. □

8.3.2 Subgroups

A common strategy in the study of a group is to investigate subsets of the group that are also groups. Let $(G, *)$ be a group and let H be a substructure of G (see Definition 8.2.3). Thus, H is a nonempty subset of G that is closed under $*$, that is, $a * b \in H$ whenever $a, b \in H$. When $(H, *)$ is also a group, we will say that H is a subgroup of G.

Definition 8.3.11. Let $(G, *)$ be a group. Then H is called a **subgroup** of G, if H is a substructure of G that also satisfies the group axioms.

Each group with two or more elements has at least two subgroups, namely, the entire group itself and the subgroup $\{e\}$ containing just the identity element.

Definition 8.3.12. Let $(G, *)$ be a group and let H be a subgroup of G. We shall call H a **proper subgroup** if $H \neq \{e\}$ and $H \neq G$.

Example 10. We introduced the group $(\mathbb{Z}, +)$ (see Example 1) with the identity element 0. Observe that $\mathbb{N}$ is a substructure of $\mathbb{Z}$; however, $(\mathbb{N}, +)$ does not satisfy the group axioms, as $\mathbb{N}$ does not have an identity element for addition. Thus, *a substructure of a group may not be a subgroup.*

Example 11. Consider the group $(G, *)$ introduce in Example 6, where $*$ is defined by the "multiplication table"

$c*^{r}$	e	u	v	w	x	y
e	e	u	v	w	x	y
u	u	v	e	x	y	w
v	v	e	u	y	w	x
w	w	y	x	e	v	u
x	x	w	y	u	e	v
y	y	x	w	v	u	e

and $G = \{e, u, v, w, x, y\}$. Let $H = \{e, w\}$. One can easily see that H is closed under $*$. So, H is a substructure of G with multiplication table given by

$c*^r$	e	w
e	e	w
w	w	e

One can now verify that $(H, *)$ satisfies the group axioms. So, H is a subgroup of G. For another example, let $K = \{e, u, v\}$. Because K is closed under $*$, we see that K is also a substructure of G with multiplication table given by

$c*^r$	e	u	v
e	e	u	v
u	u	v	e
v	v	e	u

Since $(K, *)$ satisfies the group axioms, we conclude that K is also subgroup of G.

Example 12. Let $(G, \circ)$ be the group presented in Example 5, where the binary operation $\circ$ is composition. Define $H \subseteq G$ by

$$H = \{T_{a,b} \in G : a \in \mathbb{Q} \text{ and } b \in \mathbb{R}\}.$$

Let $T_{a,b} \in H$ and $T_{c,d} \in H$. Thus, $a, c \in Q$, $a \neq 0$ and $c \neq 0$. Since $T_{a,b} \circ T_{c,d} = T_{ac,ad+b}$ and ac is a nonzero rational number, we conclude that $T_{a,b} \circ T_{c,d} \in H$. Therefore, H is closed under $\circ$ and so H is a substructure of G. We now show that the structure $(H, \circ)$ satisfies the group axioms.

1. The associative law holds for $(H, \circ)$, since functional composition is associative.
2. Recall that $T_{1,0}$ is the identity element for the group G. Since $1 \in \mathbb{Q}$, we see that $T_{1,0} \in H$. For every $T \in H$ we have $T \circ T_{1,0} = T_{1,0} \circ T = T$.
3. Let $T_{a,b} \in H$. Thus, $a \in \mathbb{Q}$ and $a \neq 0$. Hence, $\frac{1}{a} \in \mathbb{Q}$ and $\frac{1}{a} \neq 0$. In Example 5 we showed that $T_{\frac{1}{a}, -\frac{b}{a}}$ is the inverse element for $T_{a,b}$ in G. Since $T_{\frac{1}{a}, -\frac{b}{a}} \in H$, we see that $T_{a,b}$ has an inverse element in H.

Thus, all of the group axioms hold. So, H is a subgroup of G.

Our next lemma gives three (very useful) conditions that a subset of a group G must satisfy in order to be a subgroup.

Lemma 8.3.13 (Subgroup Lemma). *Let $(G, *)$ be a group and let H be a subset of G. Then H is a subgroup of G if and only if*

(1) $e \in H$,
(2) $ab \in H$ *whenever* $a, b \in H$,
(3) $a^{-1} \in H$ *whenever* $a \in H$.

Proof. Let $(G,*)$ be a group and let H be a subset of G. We shall show that H is a subgroup of G if and only if the above three properties hold.

($\Rightarrow$). Suppose H is a subgroup, that is, assume that $(H,*)$ is a group *relative* to the binary operation $*$ of G. We shall first prove that (1) and (2) hold. Let e be the identity element of G. We will show that $e \in H$. Since $(H,*)$ is a group, there is an element $e' \in H$ such that $ae' = a$ for all $a \in H$. We shall show that $e' = e$. Let $a \in H$ be an element in H. Since $ae' = a$ holds in $(H,*)$, it follows that $ae' = a$ also holds in $(G,*)$. Furthermore, because $e \in G$ is the identity element, we have $ae = a$ in G. Hence, $ae' = ae$. Lemma 8.3.5 implies that $e' = e$ and thus, $e \in H$. Item (2) holds because H is a substructure of G and is therefore closed under the binary operation.

Now, to prove that (3) holds, let $a \in H$. Let a^{-1} be the inverse of a in $(G,*)$. We shall show that $a^{-1} \in H$. Since $(H,*)$ is a group, there exists an element $a' \in H$ such that $aa' = e$. An argument similar to the one in the previous paragraph, shows that $a' = a^{-1}$ and thus, $a^{-1} \in H$.

($\Leftarrow$). Suppose H satisfies properties (1)–(3). From property (1) we see that H is nonempty. Property (2) implies that H is closed under $*$ and so, H is a substructure of G. We will show that $(H,*)$ satisfies the group axioms. Since $*$ is associative in G, we see that the associative law also holds for $(H,*)$. By (1), we have that $e \in H$ where e is an identity element for G. So, e is also the identity element for H. Let $a \in H$. By (3), we have that $a^{-1} \in H$. Since a^{-1} is the inverse for a in G, we see that a^{-1} is the inverse element for a in H as well. Therefore, H is a subgroup of G. □

Example 13. Let $(G,\circ)$ be the group presented in Example 5 on page 249, where $\circ$ is composition. Define $N \subseteq G$ to be $N = \{T_{1,b} \in G : b \in \mathbb{R}\}$. To show that N is a subgroup of G, we shall verify that items (1)–(3) of Lemma 8.3.13 hold.

(1) Clearly $T_{1,0} \in N$.
(2) Let $T_{1,b}$ and $T_{1,c}$ be in N. Since $T_{1,b} \circ T_{1,c} = T_{1,c+b}$, we see that $T_{1,b} \circ T_{1,c}$ is in N.
(3) Let $T_{1,b} \in N$. Because $T_{1,b}^{-1} = T_{1,-b}$, we conclude that $T_{1,b}^{-1} \in N$.

Since (1)–(3) hold, Lemma 8.3.13 implies that N is a subgroup of G.

Lemma 8.3.13 gives three essential properties that must be verified before one can conclude that a subset H of a group G is a subgroup. Furthermore, if one is assuming that H is a subgroup of G, then one can take advantage of these three properties. Thus, we have the following proof and assumption strategies.

Proof Strategy 8.3.14. Let $(G,*)$ be a group and let H be a subset of G. To prove that H is a subgroup of G, use the three-step proof diagram:

Step (1): Prove $e \in H$.
Step (2): Let $a,b \in H$.
Prove $ab \in H$.
Step (3): Let $a \in H$.
Prove $a^{-1} \in H$.

Assumption Strategy 8.3.15. Let $(G,*)$ be a group and suppose in a proof you are *assuming* that H is a subgroup of G. Then you know that $e \in H$, and if you are assuming or can prove that $a,b \in H$, then you can conclude that $ab \in H$ and $a^{-1} \in H$.

In the proof of our next theorem we will employ both Proof Strategy 8.3.14 and Assumption Strategy 8.3.15.

Theorem 8.3.16. *Let $(G,*)$ be a group. Suppose that H and K are subgroups of G. Then $H \cap K$ is also a subgroup of G.*

Proof. Suppose that H and K are subgroups of G. We shall prove that $H \cap K$ is a subgroup of G.

(1) Since H and K are subgroups of G, Lemma 8.3.13 implies that $e \in H$ and $e \in K$. Thus, $e \in H \cap K$.
(2) Let $a,b \in H \cap K$. So, $a,b \in H$ and $a,b \in K$. Since H and K are subgroups of G, Lemma 8.3.13 implies that $ab \in H$ and $ab \in K$. Therefore, $ab \in H \cap K$.
(3) Let $a \in H \cap K$. Thus, $a \in H$ and $a \in K$. Because H and K are subgroups of G, it follows from Lemma 8.3.13 that $a^{-1} \in H$ and $a^{-1} \in K$. Hence, $a^{-1} \in H \cap K$.

Lemma 8.3.13 now implies that $H \cap K$ is a subgroup of G. □

8.3.3 Normal Subgroups

A normal subgroup is a special kind of subgroup that can be used to construct a new group. This new group is obtained by using a relevant equivalence relation (see Section 8.6.1).

Definition 8.3.17. Let $(G,*)$ be a group and let N be a subgroup of G. For each $a \in G$, define the set $a^{-1}Na = \{a^{-1}na : n \in N\}$. In other words, the set $a^{-1}Na$ is the collection of all products of the form $a^{-1}na$ for some $n \in N$.

Clearly, for each $a \in G$, the set $a^{-1}Na$ is a subset of G.

Theorem 8.3.18. *Let $(G,*)$ be a group and let N be a subgroup of G. Then for all $a \in N$, we have that $a^{-1}Na \subseteq N$.*

Proof. Suppose that $(G,*)$ is a group and N is a subgroup of G. Let $a \in N$. So $a^{-1} \in N$, as N is a subgroup. Thus, $a^{-1}na \in N$ for all $n \in N$, because N is closed under $*$. Therefore, $a^{-1}Na \subseteq N$. □

We know, by the above theorem, that for each $a \in N$ the set $a^{-1}Na$ is a subset of N whenever N is a subgroup of G. Suppose for each $a \in G$ we also have that $a^{-1}Na$ is a subset of N. Then N will be called a normal subgroup.

Definition 8.3.19. Let $(G,*)$ be a group and let N be a subgroup of G. We shall call N a **normal subgroup** if $a^{-1}Na \subseteq N$ for *every* $a \in G$. We shall write $N \lhd G$ to mean that N is a normal subgroup of G.

In our next example, we present a normal subgroup of finite group G and then another subgroup of G that is not normal. To show that a subgroup H is not normal, one must find an element $a \in G$ and an element $h \in H$ and show that $a^{-1}ha \notin H$.

Example 14. Consider the group $(G,*)$ presented in Example 6. Recall that $*$ is defined by the "multiplication table"

$c*^{r}$	e	u	v	w	x	y
e	e	u	v	w	x	y
u	u	v	e	x	y	w
v	v	e	u	y	w	x
w	w	y	x	e	v	u
x	x	w	y	u	e	v
y	y	x	w	v	u	e

where $G = \{e,u,v,w,x,y\}$. In Example 11 we identified the subgroup $K = \{e,u,v\}$ of G. We shall show that K is a normal subgroup; that is, we show that $a^{-1}Ka \subseteq K$ for each $a \in G$. By Theorem 8.3.18, we just need to show that $a^{-1}Ka \subseteq K$ for each $a \in G \setminus K = \{w,x,y\}$. We do this as follows:

1. $w^{-1}Kw = \{w^{-1}kw : k \in K\} = \{w^{-1}ew, w^{-1}uw, w^{-1}vw\} = \{e,v,u\} \subseteq K$
2. $x^{-1}Kx = \{x^{-1}kx : k \in K\} = \{x^{-1}ex, x^{-1}ux, x^{-1}vx\} = \{e,v,u\} \subseteq K$
3. $y^{-1}Ky = \{y^{-1}ky : k \in K\} = \{y^{-1}ey, y^{-1}uy, y^{-1}vy\} = \{e,v,u\} \subseteq K$.

Therefore, K is normal subgroup of G. In Example 11, we also showed that $H = \{e,w\}$ is a subgroup of G. Observe that H is not a normal subgroup, because $x^{-1}wx = y \notin H$.

Example 15. A group can have more than one normal subgroup. Let $(G,\circ)$ be the group in Example 5 on page 249, where $\circ$ is composition.

1. Let $H = \{T_{a,b} \in G : a \in \mathbb{Q} \text{ and } b \in \mathbb{R}\}$. In Example 12 we showed that H is a subgroup of G. Furthermore, for any $T_{c,d} \in G$ and any $T_{a,b} \in H$, one can show (see Exercise 11) that $T_{c,d}^{-1}T_{a,b}T_{c,d} \in H$. Hence $T_{c,d}^{-1}HT_{c,d} \subseteq H$ and therefore, H is a normal subgroup of G.
2. In Example 13, we showed that $N = \{T_{1,b} \in G : b \in \mathbb{R}\}$ is a subgroup of G. For any $T_{c,d} \in G$ and any $T_{1,b} \in N$, one can verify that $T_{c,d}^{-1}T_{1,b}T_{c,d} = T_{1,\ell}$ for some real number ℓ. Thus $T_{c,d}^{-1}NT_{c,d} \subseteq N$. So, N is another normal subgroup of G.

Definition 8.3.19 provides a clear strategy for proving that a subgroup of a group is a normal subgroup.

Proof Strategy 8.3.20. Let $(G,*)$ be a group and let N be a subgroup of G. To prove that N is a normal subgroup of G, use the proof diagram:

Let $a \in G$.
  Prove $a^{-1}Na \subseteq N$.

In other words, use the diagram

Let $a \in G$ and $n \in N$.
  Prove $a^{-1}na \in N$.

Assumption Strategy 8.3.21. Let $(G,*)$ be a group and suppose in a proof that you are *assuming* N is a normal subgroup of G. Let $a \in G$ be any element. When you are assuming or can prove that $n \in N$, then you can conclude that $a^{-1}na \in N$ or, equivalently, that $a^{-1}na = h$ for some $h \in N$.

The proof of our next theorem uses Strategies 8.3.20 and 8.3.21.

Theorem 8.3.22. *Let $(G,*)$ be a group. If $H \lhd G$ and $K \lhd G$, then $H \cap K \lhd G$.*

Proof. Suppose $H \lhd G$ and $K \lhd G$. By Theorem 8.3.16, we know that $H \cap K$ is a subgroup of G. We shall prove that $H \cap K$ is a normal subgroup of G. Let $a \in G$ and $n \in H \cap K$. We shall prove that $a^{-1}na \in H \cap K$. Since $n \in H \cap K$, it follows that $n \in H$ and $n \in K$. Now, because H and K are normal subgroups of G, we have that $a^{-1}na \in H$ and $a^{-1}na \in K$. Hence, $a^{-1}na \in H \cap K$. □

8.3.4 The Order of an Element in a Group

In group theory the term *order* is used for two closely related concepts.

1. The *order of a group* G is simply the number of elements in G.
2. The *order of an element* a in a group is the smallest natural number m (if it exists) such that $a^m = e$, where e is the identity element of the group.

Here are the official definitions.

Definition 8.3.23. Let $(G,*)$ be a group.

1. The **order** of G, denoted by $|G|$, is the cardinality (number of elements) of G.
2. The **order** of $a \in G$, denoted by $o(a)$, is the least natural number m such that $a^m = e$. If no such m exists, then we say that the order of a *does not exist*.

Example 16. Let $(G,*)$ be the group (see Example 7) of all matrices A in $M_2(\mathbb{R})$ that have the form $A = \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix}$ where $a \neq 0$ and the operation $*$ is matrix multiplication. Since G contains an infinite number of elements, the group G has infinite order. Let $A \in G$ be given by $A = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}$. Since $A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, we conclude

that A has order 2. On the other hand, the order of the matrix $B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ does not exist, because $B^m = \begin{bmatrix} 2^m & 0 \\ 0 & 1 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ for all natural numbers m.

We shall say that G is a **finite group** if G has a finite number of elements, and we use the notation $|G|$ to denote this number.

Example 17. Let $(G,*)$ be the group in Example 6, where $G = \{e,u,v,w,x,y\}$ and $*$ is defined by the "multiplication table"

$c*^r$	e	u	v	w	x	y
e	e	u	v	w	x	y
u	u	v	e	x	y	w
v	v	e	u	y	w	x
w	w	y	x	e	v	u
x	x	w	y	u	e	v
y	y	x	w	v	u	e

We see that G is a finite group of order 6, that is, $|G| = 6$. To determine the order of the element $v \in G$, observe that $v \neq e$, $v^2 \neq e$ and $v^3 = e$. Thus, $o(v) = 3$, that is, the element v has order 3.

We now show that the order of any element in a finite group exists.

Theorem 8.3.24. *Let $(G,*)$ be a finite group. For every $a \in G$ there is a natural number k such that $a^k = e$ and thus, the order of a exists.*

Proof. Suppose that $(G,*)$ is a finite group and let $a \in G$. Consider the infinite list of elements in G

$$a, a^2, \ldots, a^i, a^{i+1}, \ldots$$

Since G is finite, there must be repetitions in the above list. So there are natural numbers $i < j$ such that $a^j = a^i$. Hence, $a^j a^{-i} = e$ and $a^{j-i} = e$. So, for $k = j - i \geq 1$, we have that $a^k = e$. Therefore, there is a natural number k such that $a^k = e$. By the well-ordering principle, there is a least such k and hence the order of a exists. □

Let $a \neq e$ be an element in a finite group G and let k be a natural number. To prove that $o(a) = k$, you must prove (1) $a^k = e$ and (2) for all $i \in \mathbb{N}$, if $i < k$ then $a^i \neq e$. We shall apply this strategy in the proof of the next lemma.

Lemma 8.3.25. *Let $(G,*)$ be a finite group and let $a \in G$. If $n = o(a)$ and $n = jk$ for natural numbers $j > 1$ and $k > 1$, then $o(a^j) = k$.*

Proof. Let $(G,*)$ be a finite group and let $a \in G$. Suppose $n = o(a)$ and $n = jk$, where $j > 1$ and $k > 1$. Since $n = o(a)$, we know that $a^n = e$ and $a^m \neq e$ whenever $m < n$ is a natural number. To show that $o(a^j) = k$, we first prove that $(a^j)^k = e$ as follows:

$(a^j)^k = a^{jk} = a^n = e$. Now, let $i < k$ be a natural number. We have $(a^j)^i = a^{ji} \neq e$ because $ji < jk = n$. Therefore, $o(a^j) = k$. □

Lemma 8.3.26. *Let $(G,*)$ be a finite group and let $a \in G$. For each $n \in \mathbb{Z}$, if $a^n = e$ then $o(a) \,|\, n$.*

Proof. Let $(G,*)$ be a finite group with $a \in G$. Let $d = o(a)$. Let n be any integer. Assume $a^n = e$. We shall prove that $d \,|\, n$. By Theorem 4.6.9 (the division algorithm), $n = qd + r$ for some $q, r \in \mathbb{Z}$ where $0 \leq r < d$. We shall prove that $r = 0$. Since $n = qd + r$, we have (see Remark 8.3.2) that

$$a^n = a^{qd+r} = a^{qd}a^r = \left(a^d\right)^q a^r.$$

So, $a^n = \left(a^d\right)^q a^r$. Since $a^n = e$ and $a^d = e$, the equation $a^n = \left(a^d\right)^q a^r$ implies that $a^r = e$. Because $0 \leq r < d$ and d is the least natural number satisfying $a^d = e$, we must have $r = 0$. Therefore, $n = qd$ and $d \,|\, n$. □

Exercises 8.3

1. Determine if the following algebraic structures are groups. If not, identify a group axiom that fails in the structure.
   (a) $(\mathbb{Z},*)$ where $a * b = a - b$ for all integers a and b.
   (b) $(\mathbb{Z},*)$ where $a * b = a + ab + b$ for all integers a and b.
   (c) $(G,+)$ where $G = \{3k : k \in \mathbb{Z}\}$ and $+$ is ordinary addition.
2. Let $(G,*)$ be the algebraic structure in Example 7 on page 250. Show that $(G,*)$ is a nonabelian group.
3. Prove Lemma 8.3.10 by showing $(ab)(b^{-1}a^{-1}) = e$.
4. Let $(G,*)$ be a group. Suppose $(ab)^2 = a^2b^2$ for all $a, b \in G$. Prove that G is abelian.
5. Let $(G,*)$ be an abelian group and let $a, b \in G$. Prove by mathematical induction that $(ab)^n = a^n b^n$ for all natural numbers n.
6. Let $(G,*)$ be a finite abelian group and let $a, b \in G$. Suppose $o(a) = m$ and $o(b) = n$. Using the result of Exercise 5, prove that $o(ab) \leq mn$.
7. Let $(G,*)$ be a finite group and let $a \in G$. Prove that if $o(a) = n$, then $o(a^{-1}) = n$.
8. Let $(G,*)$ be a group. Let $a \in G$ define $H \subseteq G$ by $H = \{a^n : n \in \mathbb{Z}\}$. Prove that H is a subgroup of G.
9. Let $(G,*)$ be a group. Let $a \in G$ and define $H = \{g \in G : ag = ga\}$. Prove that H is a subgroup of G.
10. Let $(G,*)$ be a group. Let $H = \{g \in G : gx = xg \text{ for all } x \in G\}$. Prove that H is a subgroup of G.

11. Let $(G,\circ)$ be the group presented in Example 5 on page 249. We know that $H = \{T_{a,b} \in G : a \in \mathbb{Q} \text{ and } b \in \mathbb{R}\}$ is a subgroup of G (see Example 12). Show that H is a normal subgroup of G.

12. Let $(G,*)$ be a group with subgroups H and K. Suppose $H \subseteq K$. Explain why H is also a subgroup of K.

13. Let $(G,*)$ be a group with normal subgroup N. Suppose that $N \subseteq K$ where K is also subgroup of G. Explain why N is a normal subgroup of K.

14. Let $(G,*)$ be a group and suppose $N \lhd G$. Let $n \in N$. Prove that $ana^{-1} \in N$, for all $a \in G$.

15. Let $(G,*)$ be an abelian group. Prove that every subgroup H of G is normal.

16. Let $(G,*)$ be a group and let $a \in G$. Suppose that N is a normal subgroup of G. Prove the following:

(a) For all $n \in N$ there exists a $j \in N$ such that $na = aj$.

(b) For all $n \in N$ there exists a $k \in N$ such that $an = ka$.

17. Let $(G,*)$ be a group. Suppose $H \lhd G$ and $K \lhd G$. Prove that HK is a subgroup of G, where $HK = \{hk : h \in H \text{ and } k \in K\}$.

18. Let $(G,*)$ be a group. Suppose $H \lhd G$ and $K \lhd G$. Given that HK in Exercise 17 is a subgroup of G, prove that $HK \lhd G$.

19. Let $(G,*)$ be a group and suppose that $N \lhd G$. Prove that $N \subseteq a^{-1}Na$, for all $a \in G$. Conclude that $N = a^{-1}Na$ for each $a \in G$.

20. Let $(G,*)$ be a group and let N be a normal subgroup of G. Let $a,b,c,d \in G$. Suppose $ac^{-1} \in N$ and $bd^{-1} \in N$. Prove that $(ab)(cd)^{-1} \in N$.

21. Let $(G,*)$ be a group. For $a,b \in G$, define $a \sim b$ if and only if $a = gbg^{-1}$ for some $g \in G$. Prove that $\sim$ is an equivalence relation on G.

22. Let $(G,*)$ be a group and let $\{H_i : i \in I\}$ be an indexed family of subgroups of G. Prove that $\bigcap_{i\in I} H_i$ is a subgroup of G.

23. Let $(G,*)$ be a group. Suppose that $\{N_i : i \in I\}$ is an indexed family of normal subgroups of G. Prove that $\bigcap_{i\in I} N_i$ is a normal subgroup of G.

Exercise Notes: For Exercise 2, use Example 5 as a guide. For Exercise 3, see Remark 8.3.8. For Exercise 17, use Exercise 16. For Exercise 18, the identity $xy = xaa^{-1}y$ is useful. For Exercise 19, use Exercise 14. For Exercise 20, verify that $(ab)(cd)^{-1} = (ab)(d^{-1}c^{-1}) = a(bd^{-1})c^{-1}$. Now use Exercise 16(b).

8.4 Permutation Groups

The group concept has its roots in the study of permutations. What is a permutation? If S is a nonempty set, then a **permutation** of S is a one-to-one and onto function $\sigma : S \to S$. In this section we shall be using lower case Greek letters to denote such functions.

Example 1. Let $\mathbb{Z}$ be the set of integers and define $\pi\colon \mathbb{Z} \to \mathbb{Z}$ by $\pi(x) = x+2$. Since π is one-to-one and onto, the function π is a permutation of $\mathbb{Z}$.

Example 2. Let $\mathbb{R}$ be the set of real numbers and let $a, b \in \mathbb{R}$ with $a \neq 0$. Define the function $\sigma\colon \mathbb{R} \to \mathbb{R}$ by $\sigma(x) = ax+b$. Then σ is a permutation of $\mathbb{R}$.

Definition 8.4.1. Let S be a nonempty set. Then $\text{Per}(S)$ is the set of all one-to-one and onto functions from S to S. We call $\text{Per}(S)$ the **set of all permutations** of S. For $\sigma, \tau \in \text{Per}(S)$, we let $\sigma \circ \tau$ be the composition of the functions σ and τ.

A **permutation group** is a group whose elements are permutations of a given set with composition as its binary operation. We will soon show that $(\text{Per}(S), \circ)$ is a group. One reason permutation groups are important is that every group can be represented as a group of permutations on a suitable set. Thus, every group can be thought of as set of permutations.

Remark 8.4.2. Let σ and τ be permutations in $\text{Per}(S)$ where S is a nonempty set, and let n be a natural number. Then we shall write:

1. $\sigma\tau$ as shorthand for the composition $\sigma \circ \tau$.
2. σ^{-1} for the inverse function of σ.
3. ι for the identity function from S to S (ι is the Greek letter iota).
4. σ^n for the composition $\underbrace{\sigma\sigma\sigma\cdots\sigma}_{n \text{ times}}$.
5. σ^{-n} as shorthand for $(\sigma^{-1})^n$, that is, the composition $\underbrace{\sigma^{-1}\sigma^{-1}\sigma^{-1}\cdots\sigma^{-1}}_{n \text{ times}}$.
6. $\sigma^0 = \iota$.

Note that $\sigma^{-1} \in \text{Per}(S)$ by Theorem 6.2.14. Since ι is one-to-one and onto, we see that $\iota \in \text{Per}(S)$. It follows that $\iota^i = \iota$, $\sigma^i\sigma^j = \sigma^{i+j}$, $(\sigma^i)^j = \sigma^{ij}$ and $(\sigma^i)^{-1} = \sigma^{-i}$ for all $i, j \in \mathbb{Z}$.

Lemma 8.4.3. *Let S be a nonempty set and let σ, τ, γ be elements in $Per(S)$. Then*

(a) $\sigma\tau \in Per(S)$.
(b) $(\sigma\tau)\gamma = \sigma(\tau\gamma)$.
(c) $\iota\sigma = \sigma\iota = \sigma$.
(d) $\sigma\sigma^{-1} = \sigma^{-1}\sigma = \iota$.

Thus, $(Per(S), \circ)$ is a group.

Proof. Since the composition of one-to-one and onto functions is also a one-to-one and onto function (by Theorems 6.3.6 and 6.3.7), we see that (a) holds. Part (b) follows from the fact that functional composition is associative (see Exercise 10 on page 189). Item (c) follows from Theorem 6.3.3, and Theorem 6.3.4 implies (d). This completes the proof. □

Item (a) of Lemma 8.4.3 shows that $(\text{Per}(S), \circ)$ is an algebraic structure, and the remaining items (b)–(d) show that $(\text{Per}(S), \circ)$ is a group. Since $(\text{Per}(S), \circ)$ is a

group, we know by Lemma 8.3.10 that $(\sigma\tau)^{-1} = \tau^{-1}\sigma^{-1}$ whenever σ and τ are functions in $\text{Per}(S)$. Because σ and τ are functions, one can also prove this identity directly.

8.4.1 The Symmetric Group

The quadratic formula $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ explicitly yields the roots of each second degree polynomial $ax^2 + bx + c$ when a, b, c are real numbers and $a \neq 0$. For centuries mathematicians wanted to know if it were possible to similarly solve for the roots of a polynomial of any degree. Évariste Galois solved this very difficult problem after he discovered an important connection between the ability to solve for the roots of a polynomial and a certain permutation group. This connection allowed Galois to give precise conditions that must be satisfied before one can successfully solve for the roots of a polynomial using radicals. Galois then showed that there is no formula, like the quadratic formula, that can be used to solve for the roots of each polynomial of degree n, when $n \geq 5$. It was this major discovery that gave birth to group theory. In this section we will investigate the permutation groups that Galois used to solve this famous problem in mathematics.

Definition 8.4.4. Let n be a natural number and let $S = \{1, 2, 3, \ldots, n\}$. We let S_n denote the set of all one-to-one and onto functions from S to S. In other words, $S_n = \text{Per}(S)$.

In this section we will focus on the permutation group S_n. Each $\sigma \in S_n$ is called a *permutation* on n. Define the binary operation $\circ$ on S_n to be composition, that is, $\sigma \circ \tau$ for $\sigma, \tau \in S_n$. Then $(S_n, \circ)$ is called the **symmetric group** of degree n. We note that S_n is a finite group with $n!$ many elements.

Matrix Notation

Let n be a natural number and let $S = \{1, 2, 3, \ldots, n\}$. It will be useful to have more than one way to represent a permutation in S_n. We can identify a permutation $\sigma \in S_n$ by simply listing how σ acts on each element in S as follows:

$$\sigma(1) = i_1, \sigma(2) = i_2, \ldots, \sigma(n) = i_n.$$

Another way to capture all of the values of σ is to use the matrix notation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ i_1 & i_2 & i_3 & \cdots & i_n \end{pmatrix}$$

where we list all the elements of S in the first row of the matrix and below each of these elements, we put their image under σ.

Example 3. Let $S = \{1,2,3,4,5\}$ and let $\sigma \in S_5$ be defined by $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 4 & 2 & 3 \end{pmatrix}$. So, $\sigma(1) = 5$, $\sigma(2) = 1$, $\sigma(3) = 4$, $\sigma(4) = 2$, and $\sigma(5) = 3$.

Example 4. Let $S = \{1,2,3,4\}$ and let $\sigma \in S_4$ be defined by $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$. Thus, $\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$.

Example 5. Let $S = \{1,2,3,4,5\}$. If $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 1 & 4 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 5 & 1 & 2 \end{pmatrix}$, then $\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 4 & 2 & 3 \end{pmatrix}$.

Cycle Notation

We offer an alternative method, called cycle notation, for representing permutations in S_n. Cycle notation has several advantages over matrix notation.

Definition 8.4.5. Let $i_1, i_2, \ldots, i_k$ be distinct elements in $S = \{1,2,3,\ldots,n\}$ where n is a natural number and $1 \leq k \leq n$. The expression $(i_1, i_2, i_3, \ldots, i_k)$ is called a **k-cycle** and denotes the permutation $\sigma \in S_n$ where $\sigma(i_1) = i_2$, $\sigma(i_2) = i_3$, $\ldots$, $\sigma(i_k) = i_1$, and $\sigma(j) = j$ for all $j \in S$ not listed in $(i_1, i_2, i_3, \ldots, i_k)$. We will also call $(i_1, i_2, i_3, \ldots, i_k)$ a **cycle** and refer to $i_1, i_2, i_3, \ldots, i_k$ as the **components** of the cycle.

Let $S = \{1,2,3,4,5,6\}$ and consider the permutation $\sigma \in S_6$ represented by the 4-cycle $\sigma = (1,3,4,2)$. So,

$$\sigma(1) = 3,\ \sigma(3) = 4,\ \sigma(4) = 2,\ \sigma(2) = 1 \tag{8.12}$$

where $\sigma(5) = 5$ and $\sigma(6) = 6$, because 5 and 6 are not listed in the cycle $(1,3,4,2)$. Using the notation $i \mapsto j$ to denote the fact that $\sigma(i) = j$, we can express (8.12) as

$$1 \mapsto 3 \mapsto 4 \mapsto 2 \mapsto 1.$$

We see the last item in the cycle $(1,3,4,2)$, namely 2, gets mapped to the first item in this cycle. Using (8.12), we can express σ in matrix notation and obtain $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 1 & 4 & 2 & 5 & 6 \end{pmatrix}$. One simple advantage of the cycle notation $\sigma = (1,3,4,2)$ is that it is easy to represent σ^{-1}. All one has to do is reverse the order of the cycle $\sigma = (1,3,4,2)$ to obtain $\sigma^{-1} = (2,4,3,1)$.

We note that cycle notation is not unique. Again, let $\sigma = (1,3,4,2)$ represent a permutation in S_6. Let us move the last component 2 of the cycle σ around to the first position, obtaining $\tau = (2,1,3,4)$. Thus,

$$\tau(2) = 1,\ \tau(1) = 3,\ \tau(3) = 4,\ \tau(4) = 2 \tag{8.13}$$

where $\tau(5) = 5$ and $\tau(6) = 6$, because 5 and 6 are not listed in the cycle $(2,1,3,4)$. From (8.13), we see that $\sigma = \tau$. Similarly, $\sigma = (2,1,3,4) = (4,2,1,3) = (3,4,2,1)$.

Another thing to note is that any 1-cycle is just the identity permutation. To see this, let us work in S_6 and consider the 1-cycle $\sigma = (4)$. Since 4 is the first and last item in this cycle, we see that $\sigma(4) = 4$ and $\sigma(i) = i$ for every i not listed in the cycle (4). So, σ is the identity permutation.

Now let $\sigma = (i_1, i_2, i_3, \ldots, i_k)$ be a k-cycle in S_n. Since

$$\sigma(i_1) = i_2,\ \sigma(i_2) = i_3,\ \ldots,\ \sigma(i_k) = i_1,$$

we see that σ can also be viewed as

$$\sigma = (i_1, \sigma(i_1), \sigma^2(i_1), \ldots, \sigma^{k-1}(i_1))$$

where $\sigma^k(i_1) = i_1$. For example, let $\sigma \in S_7$ be the cycle defined by $\sigma = (2,3,5,6)$. Then $\sigma = (2, \sigma(2), \sigma^2(2), \sigma^3(2))$ and $\sigma^4(2) = 2$.

8.4.2 *Cycle Products*

Suppose we have two permutations $\sigma = (i_1, \ldots, i_k)$ and $\tau = (j_1, \ldots, j_\ell)$ that are represented by cycles. Then the *product of the two cycles*

$$(i_1, \ldots, i_k)(j_1, \ldots, j_\ell)$$

denotes the composition $\sigma\tau$. We now present some examples in which we evaluate the composition of two permutations that are given in cycle notation.

Example 6. Let $\sigma \in S_6$ be defined by the 4-cycle $\sigma = (3,2,4,1)$ and let τ be the 5-cycle $\tau = (3,5,6,2,1)$. Evaluate the composition $\sigma\tau$ using the cycle product $(3,2,4,1)(3,5,6,2,1)$.

Solution. We are to evaluate the cycle product $\sigma\tau = (3,2,4,1)(3,5,6,2,1)$; that is, we need to compute $(\sigma\tau)(1)$, $(\sigma\tau)(2)$, $\ldots$, and $(\sigma\tau)(6)$. We will just determine the values for $(\sigma\tau)(1)$ and $(\sigma\tau)(5)$. One can then easily evaluate the remaining values. Of course, $(\sigma\tau)(1) = \sigma(\tau(1))$ and $(\sigma\tau)(5) = \sigma(\tau(5))$ by the definition of composition. We will use the notation $\tau : i \mapsto j$ to denote the fact that $\tau(i) = j$, and similarly for σ. Thus, $\sigma : 2 \mapsto 4$ and $\tau : 6 \mapsto 2$. To determine $\sigma(\tau(1))$ we shall use the given cycle product rewritten below as (8.14):

$$(3,2,4,1)(3,5,6,2,1). \tag{8.14}$$

We first locate 1 in the right cycle representing τ. We see that $\tau : 1 \mapsto 3$. Now we find 3 in the left cycle representing σ and we see that $\sigma : 3 \mapsto 2$. Therefore, $(\sigma\tau)(1) = 2$.

Similarly, to evaluate $(\sigma\tau)(5)$ we first find 5 in the right cycle in (8.14) and we see that $\tau : 5 \mapsto 6$ and since 6 does not appear in the left cycle, we have $\sigma : 6 \mapsto 6$. We conclude that $(\sigma\tau)(5) = 6$. Ⓢ

Example 7. Let $\sigma \in S_6$ be the 4-cycle $\sigma = (3,2,4,1)$ and let $\tau \in S_6$ be the 3-cycle $\tau = (5,3,6)$. Show that $\sigma\tau \neq \tau\sigma$.

Solution. We have that $\sigma\tau = (3,2,4,1)(5,3,6)$ and $\tau\sigma = (5,3,6)(3,2,4,1)$. Observe that 3 appears in both of the cycles σ and τ. Note that $(\sigma\tau)(3) = 6$ and $(\tau\sigma)(3) = 2$. Hence, $\sigma\tau \neq \tau\sigma$. Ⓢ

Definition 8.4.6. Two cycles are said to be **disjoint** if the cycles have no components in common.

Example 8. Consider the disjoint cycles $\sigma = (2,5,3)$ and $\tau = (8,6,4,7)$ in S_8. Show that $\sigma\tau = \tau\sigma$.

Solution. Since $\sigma\tau = (2,5,3)(8,6,4,7)$ and $\tau\sigma = (8,6,4,7)(2,5,3)$ we see that

$$\begin{aligned}
(\sigma\tau)(1) &= 1 = (\tau\sigma)(1)\\
(\sigma\tau)(2) &= 5 = (\tau\sigma)(2)\\
(\sigma\tau)(3) &= 2 = (\tau\sigma)(3)\\
(\sigma\tau)(4) &= 7 = (\tau\sigma)(4)\\
(\sigma\tau)(5) &= 3 = (\tau\sigma)(5)\\
(\sigma\tau)(6) &= 4 = (\tau\sigma)(6)\\
(\sigma\tau)(7) &= 8 = (\tau\sigma)(7)\\
(\sigma\tau)(8) &= 6 = (\tau\sigma)(8).
\end{aligned}$$

Therefore, $\sigma\tau = \tau\sigma$. Ⓢ

Our next result shows any two disjoint cycles commute.

Theorem 8.4.7. *Let σ and τ be disjoint cycles in S_n. Then $\sigma\tau = \tau\sigma$.*

Proof. Let σ and τ be disjoint cycles in S_n. Let $1 \leq i \leq n$. We will show that $(\sigma\tau)(i) = (\tau\sigma)(i)$. There are three cases to consider.

CASE 1: i does not appear in either of the cycles σ or τ. In this case we see that $\tau(i) = i$ and $\sigma(i) = i$. Consequently, $\sigma(\tau(i)) = i = \tau(\sigma(i))$.

CASE 2: i appears in τ. Since σ and τ are disjoint, we conclude that i and $\tau(i)$ do not appear in σ. Therefore, $\sigma(\tau(i)) = \tau(i)$ and $\sigma(i) = i$. Thus $\sigma(\tau(i)) = \tau(i) = \tau(\sigma(i))$.

CASE 3: i appears in σ. Because σ and τ are disjoint, we have that i and $\sigma(i)$ do not appear in τ. Thus, $\tau(i) = i$ and $\tau(\sigma(i)) = \sigma(i)$. Hence, $\sigma(\tau(i)) = \sigma(i) = \tau(\sigma(i))$.

In each case we conclude that $(\sigma\tau)(i) = (\tau\sigma)(i)$. Therefore, $\sigma\tau = \tau\sigma$. □

8.4.3 Cycle Decomposition

In this section we will present a *cycle decomposition algorithm*[3] that will allow us to express each permutation in S_n as a product (composition) of disjoint cycles. Observe that the identity permutation $\iota \in S_n$ can be written as $\iota = (1)$ and so, it is easy to express ι as a 'product' of one cycle. Before we present our algorithm, we must first explain what it means for an element to be moved by a permutation and then establish a method for generating cycles.

Definition 8.4.8. Let $S = \{1,2,3,\ldots,n\}$ where $n > 1$ is a natural number. Let $\pi \in S_n$. The permutation π is said to **move** $i \in S$, if $\pi(i) \neq i$. We let $M_\pi = \{i \in S : \pi(i) \neq i\}$ be the set of elements in S that are moved by π. We shall let $|M_\pi|$ denote the number of elements in M_π.

Example 9. Let $\tau \in S_6$ be the permutation $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 4 & 6 & 5 & 1 \end{pmatrix}$. Then τ moves $1,3,4,6$. Thus $M_\tau = \{1,3,4,6\}$ and $|M_\tau| = 4$.

We now identify a method for constructing cycles. Suppose $\sigma \in S_n$ and $i \in M_\sigma$. Let k be the least natural number satisfying $\sigma^k(i) = i$ (see Lemma 8.4.9, below) and let τ be the cycle defined by

$$\tau = (i, \sigma(i), \sigma^2(i), \ldots, \sigma^{k-1}(i)). \tag{8.15}$$

Since $\sigma^k(i) = i$, we shall say that τ is the cycle obtained by "starting at i and cycling through σ until returning to i."

Given a specific permutation $\sigma \in S_n$ and $i \in M_\sigma$, it is quite easy to construct the cycle τ given in (8.15).

Example 10. Let $\sigma \in S_6$ be the permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 6 & 3 & 4 & 1 \end{pmatrix}$. Since $\sigma(1) \neq 1$, we have that $1 \in M_\sigma$. Starting at 1, we cycle through σ until returning to 1 to obtain the cycle $\tau = (1, \sigma(1), \sigma^2(1), \sigma^3(1), \sigma^4(1)) = (1,3,4,5,6)$ where $\sigma^5(1) = 1$.

$$\left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \downarrow & & \downarrow & \downarrow & \downarrow & \downarrow \\ 3 & 2 & 4 & 5 & 6 & 1 \end{array} \right)$$

For a permutation σ on S and $X \subseteq S$, recall that $\sigma[X]$ is the image of X (see Definition 6.4.1).

Lemma 8.4.9. *Let σ be a permutation in S_n. Then $\sigma[M_\sigma] = M_\sigma$ and for each $i \in M_\sigma$ there is a natural number $k \leq |M_\sigma|$ such that $\sigma^k(i) = i$.*

[3] An algorithm is a step-by-step procedure for solving a problem.

Proof. Let σ be a permutation in S_n where $S = \{1, 2, \ldots, n\}$. We first prove that $\sigma[M_\sigma] \subseteq M_\sigma$. Let $x \in M_\sigma$ and thus, $\sigma(x) \neq x$. We show that $\sigma(x) \in M_\sigma$. Suppose, for a contradiction, that $\sigma(\sigma(x)) = \sigma(x)$. Since σ is one-to-one, we infer that $\sigma(x) = x$. This contradiction forces us to conclude the $\sigma(x) \in M_\sigma$ and so, $\sigma[M_\sigma] \subseteq M_\sigma$.

To prove that $M_\sigma \subseteq \sigma[M_\sigma]$, let $x \in M_\sigma$. We shall show that $x \in \sigma[M_\sigma]$. Since σ is onto, there is an $i \in S$ such that (1) $\sigma(i) = x$. We show that $i \in M_\sigma$. Suppose, for a contradiction, that $i \notin M_\sigma$. So, (2) $\sigma(i) = i$. Equations (1) and (2) yield $x = i$ and thus $x \notin M_\sigma$, which is a contradiction. Hence, must have that $i \in M_\sigma$. Now, because $\sigma(i) = x$, we see that $x \in \sigma[M_\sigma]$. So, $M_\sigma \subseteq \sigma[M_\sigma]$ and therefore, $\sigma[M_\sigma] = M_\sigma$.

Now, let $i \in M_\sigma$ and $m = |M_\sigma|$. We shall prove that there is a natural number $k \leq m$ such that $\sigma^k(i) = i$. Consider the list

$$i, \sigma(i), \sigma^2(i), \ldots, \sigma^m(i). \tag{8.16}$$

Since $i \in M_\sigma$ and $\sigma[M_\sigma] \subseteq M_\sigma$, it follows that each item in the list (8.16) is in M_σ. Furthermore, because $m = |M_\sigma|$ and the list has $m+1$ many terms, two of these terms must be equal. If the first term i in (8.16) is equal to another term $\sigma^k(i)$, then $\sigma^k(i) = i$ where $1 \leq k \leq m$ and we have our desired conclusion. Suppose that ($\star$) $\sigma^b(i) = \sigma^a(i)$ for some natural numbers $1 \leq a < b \leq m$. It follows, by applying the function σ^{-a} to both sides of the equation ($\star$), that $\sigma^{b-a}(i) = i$. Thus, $\sigma^k(i) = i$ where $k = b - a \leq m$ and k is a natural number. □

Lemma 8.4.10. *Let σ be a permutation in S_n and $i \in M_\sigma$. Let ℓ be the least natural number satisfying $\sigma^\ell(i) = i$. Then $2 \leq \ell \leq |M_\sigma|$ and $i, \sigma(i), \sigma^2(i), \ldots, \sigma^{\ell-1}(i)$ are distinct elements in M_σ.*

Proof. See Exercise 18. □

We now present our procedure for expressing a permutation on n, as a product of disjoint cycles.

Cycle Decomposition Algorithm. Given $\sigma \in S_n$ and $\sigma \neq \iota$, let L be a list of all the elements in M_σ.

1. Pick the first item, say i, in the list L.
2. Construct the cycle τ by starting at i and cycling through σ until returning to i.
3. Record the cycle τ and delete all items from the list L that appear in τ.
4. If the list L is empty, then **stop**; otherwise, go to Step 1.

After completing this algorithm, σ is the product of all the recorded disjoint cycles.

Example 11. Let $\sigma \in S_9$ be defined by $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 9 & 3 & 6 & 2 & 1 & 5 & 8 & 7 \end{pmatrix}$. Using the cycle decomposition algorithm, express σ as a product of disjoint cycles.

Solution. Because $M_\sigma = \{1,2,4,5,6,7,9\}$, we let $L = 1,2,4,5,6,7,9$. Since 1 is the first item in the list L, we start at 1 and cycle through σ until we return to 1 as follows:

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

We obtain our first cycle $\tau_1 = (1,4,6)$. Removing $1,4,6$ from the list L we obtain the new list $L = 2,5,7,9$. Since 2 is the first item in the list L, we start at 2 and cycle through σ until we return to 2:

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

We obtain our second cycle $\tau_2 = (2,9,7,5)$. Removing $2,9,7,5$ from the list L we obtain the empty list. Hence, $\sigma = \tau_2\tau_2 = (1,4,6)(2,9,7,5)$ is the product of the resulting disjoint cycles. Ⓢ

The proof of our next theorem verifies that the cycle decomposition algorithm is *correct*; that is, the proof shows that the algorithm will "always work."

Theorem 8.4.11. *Let n be a natural number. Every permutation $\sigma \in S_n$ can be expressed as a product of pairwise disjoint cycles.*

Proof. Let $S = \{1,2,\ldots,n\}$ where n is a natural number. Since $\iota = (1)$ is a 'product' of one cycle, we shall only consider permutations $\sigma \in S_n$ where $\sigma \neq \iota$. We will prove, by strong induction on $|M_\sigma|$, the following statement:

> *Whenever $|M_\sigma| \geq 2$, the permutation σ can be written as a product of disjoint cycles whose components are in M_σ.*

Base step: Suppose $|M_\sigma| = 2$. Let $i \in M_\sigma$. Lemma 8.4.10 implies that $M_\sigma = \{i, \sigma(i)\}$ and $\sigma = (i, \sigma(i))$. Thus, σ is a product of one cycle whose components are in M_σ.

Inductive step: Suppose $|M_\sigma| > 2$ and assume the strong induction hypothesis:

> If $\pi \in S_n$ satisfies $2 \leq |M_\pi| < |M_\sigma|$, then π can be expressed as a product of disjoint cycles whose components are in M_π. (IH)

Let $i \in M_\sigma$ and ℓ be the natural number given by Lemma 8.4.10. Thus, $2 \leq \ell \leq |M_\sigma|$, $\sigma^\ell(i) = i$, and $\tau = (i, \sigma(i), \sigma^2(i), \ldots, \sigma^{\ell-1}(i))$ is a cycle whose components are in M_σ. There are two cases to consider: $\ell = |M_\sigma|$ and $\ell < |M_\sigma|$.

CASE (1): $\ell = |M_\sigma|$. Since $\ell = |M_\sigma|$, it follows that $\sigma = \tau$ and σ is a product of one cycle whose components are in M_σ.

CASE (2): $\ell < |M_\sigma|$. Let $X = M_\sigma \setminus \{i, \sigma(i), \sigma^2(i), \ldots, \sigma^{\ell-1}(i)\}$. Since $2 \leq \ell < |M_\sigma|$, we see that X is nonempty and $|X| < |M_\sigma|$. One can now show that $\sigma[X] = X$ and $|X| \geq 2$ (see Exercise 19). Let $\pi\colon S \to S$ be defined by

$$\pi(a) = \begin{cases} a, & \text{if } a \notin X; \\ \sigma(a), & \text{if } a \in X \end{cases}$$

for all $a \in S$. Because $\sigma[X] = X$, one can show that π is one-to-one and onto. Thus, π is a permutation in S_n and $M_\pi = X$. Furthermore, we see that $\sigma = \tau\pi$. Since $2 \leq |M_\pi| < |M_\sigma|$, our induction hypothesis (IH) implies that π is a product of disjoint cycles whose components are in M_π. Because $\sigma = \tau\pi$, we conclude that σ can be written as product of disjoint cycles with components in M_σ. □

Let $\sigma \in S_n$ for some natural number $n > 1$. By Definition 8.3.23, the **order** of σ is the least natural number k such that $\sigma^k = \iota$, where ι is the identity permutation. Our next result shows that if σ can be written as a cycle, then the order of σ is equal to the length of the cycle.

Theorem 8.4.12. *Let $\sigma = (a_1, a_2, a_3, \ldots, a_m)$ be an m-cycle in S_n, where $1 \leq m \leq n$ are natural numbers. Then the order of σ is m.*

Proof. See Exercise 11. □

Given natural numbers m and n, we know that both m and n evenly divide the natural number mn. Thus, by the well-ordering principle, there is a smallest natural number k that is divisible by both m and n. This number k is called the least common multiple of m and n.

Definition 8.4.13. Let m and n be natural numbers. The **least common multiple** of m and n is the natural number k satisfying:

(1) $m \,|\, k$ and $n \,|\, k$.
(2) For all $i \in \mathbb{N}$, if $m \,|\, i$ and $n \,|\, i$, then $k \leq i$.

The least common multiple of m and n is denoted by $\text{lcm}(m, n)$.

For example, $\text{lcm}(12, 16) = 192$, that is, the least common multiple of 12 and 16 is 192. More generally, given a finite number of natural numbers $m_1, m_2, \ldots, m_k$ we let $\text{lcm}(m_1, m_2, \ldots, m_k)$ denote the smallest natural number ℓ such that $m_i \,|\, \ell$ for all $i = 1, \ldots, k$. Thus, $\text{lcm}(4, 3, 8) = 24$.

Theorem 8.4.14. *Let $\sigma = (a_1 a_2, a_3, \ldots, a_m)$ and $\tau = (b_1, b_2, b_3, \ldots, b_\ell)$ be disjoint cycles in S_n, where $n > 1$. Then the order of $\sigma\tau$ is* $\text{lcm}(m, \ell)$.

Proof. See Exercises 12–14. □

Let $\sigma_1\sigma_2 \cdots \sigma_k$ be a product of pairwise disjoint cycles in S_n, where $n > 1$. One can show that $o(\sigma_1\sigma_2 \cdots \sigma_k) = \text{lcm}(o(\sigma_1), o(\sigma_2), \ldots, o(\sigma_k))$ by extending the proof of Theorem 8.4.14. Thus, Theorems 8.4.11 and 8.4.12 yield a method for evaluating the order of any element in S_n.

Example 12. Find the order of the permutation $\gamma \in S_9$ given by

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

Solution. By applying the cycle decomposition algorithm, we express γ as the product of three disjoint cycles, namely, $\gamma = (1,3,5,6)(2,9,7)(4,8)$. We conclude, from Theorem 8.4.12, that $o(1,3,5,6) = 4$, $o(2,9,7) = 3$ and $o(4,8) = 2$. Since $\text{lcm}(4,3,2) = 12$, we see that $o(\gamma) = 12$ by an extension of Theorem 8.4.14. Ⓢ

A 2-cycle (i, j) is a permutation that interchanges the two elements i and j while leaving all the other elements fixed. Thus, a 2-cycle is also called a *transposition*. In the solution of our next example we will show how to write the cycle $(5,4,3,2,1)$ as a product of 2-cycles.

Example 13. Express the cycle $(5,4,3,2,1)$ in S_6 as a product of transpositions.

Solution. This can be done in several ways: $(5,4,3,2,1) = (1,2)(1,3)(1,4)(1,5)$ and $(5,4,3,2,1) = (5,4)(3,4)(3,1)(3,2)$. There is yet another way that works in every case: Start with the left most component and then "pair it" with every other component in the following order: $(5,4,3,2,1) = (5,1)(5,2)(5,3)(5,4)$. Ⓢ

Our solution in Example 13 shows there is more than one way to express a cycle as a product of transpositions. We ended this solution by applying a method that "works in every case." In our next lemma we will formally identify this method and show that each cycle can be expressed as a product of 2-cycles.

Lemma 8.4.15. *Let $n > 1$ be a natural number. Every cycle in S_n can be expressed as a product of transpositions.*

Proof. Let $\sigma = (i_1, i_2, i_3, \ldots, i_k)$ be a cycle in S_n. Then

$$(i_1, i_2, i_3, \ldots, i_k) = (i_1, i_k)(i_1, i_{k-1})(i_1, i_{k-2}) \cdots (i_1, i_3)(i_1, i_2).$$

The product on the right gives $\sigma(i_1) = i_2, \sigma(i_2) = i_3, \ldots, \sigma(i_k) = i_1$. □

We cannot expect to be able to write a cycle as a product of pairwise *disjoint* 2-cycles.

Theorem 8.4.16. *Let $n > 1$. Every permutation in S_n can be expressed as a product of transpositions.*

Proof. By Theorem 8.4.11, every permutation in S_n can be expressed as a product of disjoint cycles. Since Lemma 8.4.15 implies that every cycle can be written as a product of transpositions, we conclude that every permutation can also be expressed as a product of transpositions. □

We now know that every permutation in S_n can be written as a product of transpositions, that is, 2-cycles. We also saw that there can be more than one way to express a permutation as a product of transpositions. Even though there may be multiple ways of decomposing a particular permutation into a product of 2-cycles, these different products have something in common. If one decomposition has an even (odd) number of transpositions, then any other such decomposition will also have an even (odd) number of transpositions. In other words, it is impossible to write

a permutation π as a product of an even number of transpositions and then express π as another product which consists of an odd number of transpositions.

Theorem 8.4.17. *Let n be a natural number. Let π be a permutation in S_n. Suppose that $\sigma_1, \sigma_2, \ldots, \sigma_s$ are transpositions and that $\tau_1, \tau_2, \ldots, \tau_t$ are also transpositions. If $\pi = \sigma_1\sigma_2\cdots\sigma_s = \tau_1\tau_2\cdots\tau_t$, then s and t are either both even or both odd.*

We will not prove Theorem 8.4.17 in this book; but you will see a proof of this theorem in your abstract algebra courses (see [10, Theorem 3.3.1]). As a result of Theorem 8.4.17, we can define the *parity* of a permutation π to be *even* when it decomposes into a product of an even number of transpositions, and the *parity* of π to be *odd* if it can be decomposed into a product of an odd number of transpositions. Theorem 8.4.17 implies that a permutation cannot be both even and odd.

Exercises 8.4

1. Let $a \in S$. Define the relation $\sim$ on $\mathrm{Per}(S)$ by $\sigma \sim \tau$ if and only if $\sigma\tau^{-1}(a) = a$. Prove that $\sim$ is an equivalence relation on $\mathrm{Per}(S)$.

2. Let $a \in S$ and let $H = \{\alpha \in \mathrm{Per}(S) : \alpha(a) = a\}$. Show the following:

(a) $\iota \in H$.
(b) If $\sigma, \tau \in H$, then $\sigma\tau \in H$.
(c) If $\sigma \in H$, then $\sigma^{-1} \in H$.

Conclude that H is a subgroup of $\mathrm{Per}(S)$. Suppose that $a, b, c \in S$ are all distinct. Show that H is not a normal subgroup of $\mathrm{Per}(S)$.

3. Let $a, b \in S$ be such that $a \neq b$. Now, let $H = \{\alpha \in \mathrm{Per}(S) : \alpha(a) = a\}$ and let $K = \{\beta \in \mathrm{Per}(S) : \beta(b) = b\}$. Suppose that $\gamma \in \mathrm{Per}(S)$ satisfies $\gamma(a) = b$. Show the following:

(a) If $\beta \in K$, then $\gamma^{-1}\beta\gamma \in H$.
(b) If $\alpha \in H$, then $\gamma\alpha\gamma^{-1} \in K$.

4. Let $\alpha \in S_8$ be the following product of disjoint cycles $\alpha = (2,3,1,5,6)(4)(7,8)$. Evaluate $\alpha(i)$ for $i = 1, 2, \ldots, 8$.

5. Let $(4,5,3), (5,4) \in S_6$. Show that

(a) $(4,5,3)(5,4) = (4,3)$
(b) $(5,4)(4,5,3) = (5,3)$.

6. Let $n > 1$ be a natural number and let $S = \{1, 2, 3, \ldots, n\}$. Consider the 2-cycle $\sigma = (i, j)$ in S_n. Show that $\sigma^{-1} = \sigma$.

7. One can easily verify that $S_3 = \{\iota, (1,2,3), (1,3,2), (1,2), (1,3), (2,3)\}$. Let K be the subset of S_3 given by $K = \{\iota, (1,2,3), (1,3,2)\}$. First show that K is a subgroup of S_3, and then show that K is a normal subgroup of S_3.

8. Let m, n be a natural numbers where $m \leq n$. Let $\sigma = (a_1, a_2, a_3, \ldots, a_m)$ be an m-cycle in S_n. Show that $\sigma^{-1} = (a_m, a_{m-1}, \ldots, a_1)$.

9. Let $k \leq n$ be natural numbers and let $\pi \in S_n$. Suppose that $\pi = \sigma_1\sigma_2\cdots\sigma_{k-1}\sigma_k$ where each σ_i is a transposition. Prove that $\pi^{-1} = \sigma_k\sigma_{k-1}\cdots\sigma_2\sigma_1$.

10. Let m and n be natural numbers and let $\ell = \text{lcm}(m,n)$. Suppose that k is an integer satisfying $m \,|\, k$ and $n \,|\, k$. Prove that $\ell \,|\, k$.

11. Let $n > 1$ be a natural number. The m-cycle $\sigma = (a_1, a_2, a_3, \ldots, a_m)$ denotes a permutation in S_n where $\sigma(a_1) = a_2$, $\sigma(a_2) = a_3$, …, and $\sigma(a_m) = a_1$.

(a) Prove by induction on $k \geq 1$ that if $k + i \equiv j \,(\text{mod } m)$, then $\sigma^k(a_i) = a_j$, whenever $1 \leq i \leq m$ and $1 \leq j \leq m$.

(b) Prove that $\sigma^m = \iota$.

(c) Prove by induction on $k \geq 1$ that if $\sigma^k(a_i) = a_j$, then $k + i \equiv j \,(\text{mod } m)$, whenever $1 \leq i \leq m$ and $1 \leq j \leq m$.

(d) Prove that the order of σ is m.

12. Let $\sigma = (a_1a_2, a_3, \ldots, a_m)$ and $\tau = (b_1b_2, b_3, \ldots, b_\ell)$ be two disjoint cycles in S_n, where $n > 1$ is a natural number.

(a) Prove by induction on k that for all $k \geq 1$, we have $(\sigma\tau)^k = \sigma^k\tau^k = \tau^k\sigma^k$.

(b) Let $k \in \mathbb{N}$. Prove that $(\sigma\tau)^k(a_i) = \sigma^k(a_i)$ for each a_i in the cycle σ.

(c) Let $k \in \mathbb{N}$. Prove that $(\sigma\tau)^k(b_i) = \tau^k(b_i)$ for each b_i in the cycle τ.

(d) Let $k \in \mathbb{N}$. Prove that if $(\sigma\tau)^k = \iota$, then $\sigma^k = \iota$ and $\tau^k = \iota$.

13. Let $\sigma = (a_1a_2, a_3, \ldots, a_m)$ and $\tau = (b_1b_2, b_3, \ldots, b_\ell)$ be disjoint cycles in S_n, where $n > 1$ is a natural number. By Theorem 8.4.12, $m = o(\sigma)$ and $\ell = o(\tau)$. Suppose $(\sigma\tau)^k = \iota$ where $k \in \mathbb{N}$. Using Exercise 12(d), prove that $m \,|\, k$ and $\ell \,|\, k$.

14. Let $\sigma = (a_1a_2, a_3, \ldots, a_m)$ and $\tau = (b_1b_2, b_3, \ldots, b_\ell)$ be disjoint cycles in S_n, where $n > 1$ is a natural number. Prove that if $k \in \mathbb{N}$ satisfies $m \,|\, k$ and $\ell \,|\, k$, then $(\sigma\tau)^k = \iota$. Conclude from Exercise 13 that $o(\sigma\tau) = \text{lcm}(m, \ell)$.

15. Write each of the given permutations in S_7 as a product of disjoint cycles.

(a) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 1 & 4 & 6 & 7 & 2 \end{pmatrix}$ (b) $(1,3,4,5)(5,4,7,6)$.

16. Using Theorems 8.4.12 and 8.4.14, evaluate the order of the two permutations in S_7:

(a) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 1 & 4 & 6 & 7 & 2 \end{pmatrix}$ (b) $(1,3,4,5)(5,4,7,6)$.

17. Let $n \geq 3$ and let $\sigma \in S_n$. Prove that if $\alpha\sigma = \sigma\alpha$ for all $\alpha \in S_n$, then $\sigma = \iota$.

18. Prove Lemma 8.4.10.

19. Let X and σ be as in Case (2) of the inductive step in the proof on page 268 of Theorem 8.4.11. Prove that $\sigma[X] = X$ and $|X| \geq 2$.

Exercise Notes: For Exercise 11(a), note that when $(k+1)+i \equiv j \,(\text{mod } m)$, we have $k + i \equiv j - 1 \,(\text{mod } m)$. If $j = 1$, then $j \equiv m + 1 \,(\text{mod } m)$. For Exercise 11(c): When $\sigma^{k+1}(a_i) = a_j$, we have $\sigma^k(\sigma(a_i)) = a_j$. If $i < m$ then $\sigma(a_i) = a_{i+1}$. If $i = m$ then $\sigma(a_i) = a_1$. For Exercise 18, review the proof of Lemma 8.4.9. For Exercise 19, to prove $\sigma[X] = X$, first review the proof of Lemma 8.4.9. Using the fact that $\sigma[X] = X$ and X is nonempty, one can show that $|X| \geq 2$.

8.5 Rings

Any book on abstract algebra will define a ring to be an algebraic structure which is endowed with two binary operations (usually called addition and multiplication) and satisfies certain axioms. These axioms ensure that a ring will possess many of the familiar properties of addition and multiplication that hold in the number system $(\mathbb{Z}, +, \cdot)$ (see Axioms of Arithmetic 7.4.1). To qualify as a ring, an algebraic structure of the form $(R, +, \cdot)$ must satisfy the ring axioms.

Definition 8.5.1. Let $(R, +, \cdot)$ be an algebraic structure where $+$ and $\cdot$ are two binary operations. Then $(R, +, \cdot)$ is called a **ring** if the seven RING AXIOMS hold:

1. $a + b = b + a$ for all $a, b \in R$.
2. $(a + b) + c = a + (b + c)$ for all $a, b, c \in R$.
3. There is a $0 \in R$ such that $a + 0 = a$ for all $a \in R$ (0 is called the **zero element**).
4. For every $a \in R$ there is a $b \in R$ such that $a + b = 0$ (b is written as $-a$ and is called the **additive inverse** of a).
5. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$.
6. $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in R$.
7. $(b + c) \cdot a = b \cdot a + c \cdot a$ for all $a, b, c \in R$.

We shall say that $(R, +, \cdot)$ is a **ring with unity** if it satisfies the additional axiom:

8. There is a **unity element** $1 \in R$, where $1 \neq 0$, such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in R$ (sometimes 1 is called the **multiplicative identity element**).

We will say that $(R, +, \cdot)$ is a **commutative ring** if it also satisfies the axiom:

9. $a \cdot b = b \cdot a$ for all $a, b \in R$.

Our next two results show, respectively, that a ring has exactly one zero element and each element of the ring has exactly one additive inverse. One can also show that a ring with unity has only one unity element.

Lemma 8.5.2. *Let $(R, +, \cdot)$ be a ring. There is exactly one element $b \in R$ satisfying $a + b = a$ for all $a \in R$.*

Proof. See Exercise 4. □

Lemma 8.5.3. *Let $(R, +, \cdot)$ be a ring. For each $a \in R$ there is exactly one element $b \in R$ such that $a + b = 0$.*

Proof. See Exercise 5. □

Remark 8.5.4. Axioms 1–4 of the ring axioms assert that $(R, +)$ is an abelian group with identity element 0. Lemma 8.5.2 shows that identity element 0 is unique, and Lemma 8.5.3 shows that the additive inverse $-a$ is unique for each element a in the ring. Furthermore, Axiom 5 states that multiplication is associative. Axioms 6 and 7 affirm that $(R, +, \cdot)$ satisfies the distributive laws.

Whenever a and b are in a ring, Lemma 8.5.3 implies that if $a+b=0$, then $b=-a$. This fact will be used several times in our proof of Lemma 8.5.8 below.

Definition 8.5.5. Let $(R,+,\cdot)$ be a ring. We say that an element $a\neq 0$ in R is a **zero divisor** if $a\cdot b=0$ for some $b\neq 0$ in R.

Definition 8.5.6. Let $(R,+,\cdot)$ be a commutative ring with unity. R shall be called an **integral domain** if $a\cdot b=0$ implies that $a=0$ or $b=0$, for all $a,b\in R$.

Thus, a commutative ring with unity is an integral domain whenever the ring has no zero divisors; that is, the only way that a product can be zero is when one of the factors is zero. Below, we present some examples of rings with zero divisors and rings that are integral domains.

Definition 8.5.7. Let $(R,+,\cdot)$ be a ring with unity element 1. We say that an element $a\in R$ is a **unit** if a has a multiplicative inverse, that is, $a\cdot b=b\cdot a=1$ for some $b\in R$.

Let $(R,+,\cdot)$ be a ring. For $a,b\in R$ we shall often write $a\cdot b$ as ab. In addition, we may write $a+(-b)$ more succinctly as $a-b$.

Examples of Rings

Example 1. Let $R=\mathbb{Z}$ be the set of integers. Define $+$ as ordinary addition and $\cdot$ as ordinary multiplication of integers. Then $(\mathbb{Z},+,\cdot)$ is a commutative ring with unity. In addition, $(\mathbb{Z},+,\cdot)$ is an integral domain.

Example 2. Let $R=\mathbb{Q}$ with $+$ and $\cdot$ defined, respectively, to be the usual addition and multiplication of rational numbers. Then $(\mathbb{Q},+,\cdot)$ is a commutative ring with unity. The ring $(\mathbb{Q},+,\cdot)$ is also an integral domain.

Example 3. Let $R=\mathbb{R}$ with $+$ and $\cdot$ defined to be the standard addition and multiplication of real numbers. Then $(\mathbb{R},+,\cdot)$ is a commutative ring with unity. The ring $(\mathbb{R},+,\cdot)$ is also an integral domain.

Example 4. It follows from Theorem 7.4.4 that $(\mathbb{Z}_6,+,\cdot)$ is a commutative ring with unity element $[1]$ and with zero element $[0]$. Furthermore, $(\mathbb{Z}_6,+,\cdot)$ is not an integral domain because $[2]\cdot[3]=[6]=[0]$ with $[2]\neq[0]$ and $[3]\neq[0]$.

Example 5. Let $C(\mathbb{R})$ be the set of all continuous functions $f\colon\mathbb{R}\to\mathbb{R}$. We now define how to add and multiply two functions in $C(\mathbb{R})$. For each $f\in C(\mathbb{R})$ and $g\in C(\mathbb{R})$, define the function $(f+g)\colon\mathbb{R}\to\mathbb{R}$ by $(f+g)(x)=f(x)+g(x)$ and define the function $(f\cdot g)\colon\mathbb{R}\to\mathbb{R}$ by $(f\cdot g)(x)=f(x)g(x)$ for all $x\in\mathbb{R}$. Because f and g are continuous, we have that $f+g$ and $f\cdot g$ are also continuous functions in $C(\mathbb{R})$. The zero element of $C(\mathbb{R})$ is the continuous, constant function $\overline{0}\colon\mathbb{R}\to\mathbb{R}$ defined by $\overline{0}(x)=0$ for all $x\in\mathbb{R}$. The additive inverse of each $f\in C(\mathbb{R})$ is the continuous function $(-f)\colon\mathbb{R}\to\mathbb{R}$ defined by $(-f)(x)=-f(x)$. Thus, $(C(\mathbb{R}),+,\cdot)$ is a commutative ring with unity element $\overline{1}$, where $\overline{1}$ is the function in $C(\mathbb{R})$ defined

by $\overline{1}(x) = 1$ for all $x \in \mathbb{R}$. Furthermore, $(C(\mathbb{R}), +, \cdot)$ is not an integral domain. Let $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ be the continuous nonzero functions defined by

$$f(x) = \begin{cases} 0, & \text{if } x \leq 1; \\ 2x-2, & \text{if } x > 1 \end{cases} \qquad g(x) = \begin{cases} 2x-2, & \text{if } x \leq 1; \\ 0, & \text{if } x > 1. \end{cases}$$

Thus $f \cdot g = \overline{0}$ where $f \neq \overline{0}$ and $g \neq \overline{0}$.

Example 6. Let p be a prime. By Theorem 7.4.4, $(\mathbb{Z}_p, +, \cdot)$ is a commutative ring with unity element $[1]$ and zero element $[0]$. Moreover, the ring $(\mathbb{Z}_p, +, \cdot)$ is also an integral domain. To prove this, let $[a], [b] \in \mathbb{Z}_p$. Assume $[a] \cdot [b] = [0]$. Thus, $[ab] = [0]$. Therefore, $ab \equiv 0 \pmod{p}$ and so, $p \,|\, (ab)$. Lemma 4.7.2 implies that $p \,|\, a$ or $p \,|\, b$, that is, $[a] = [0]$ or $[b] = [0]$. Hence, $(\mathbb{Z}_p, +, \cdot)$ is an integral domain.

Example 7. Let $F(\mathbb{Z})$ be the set of all functions of the form $f\colon \mathbb{Z} \to \mathbb{Z}$. Let f and g be in $F(\mathbb{Z})$. Define the function $(f+g)\colon \mathbb{Z} \to \mathbb{Z}$ by $(f+g)(i) = f(i) + g(i)$ and define the function $(f \cdot g)\colon \mathbb{Z} \to \mathbb{Z}$ by $(f \cdot g)(i) = f(i)g(i)$ for all $i \in \mathbb{Z}$. The constant function $\overline{0}\colon \mathbb{Z} \to \mathbb{Z}$, defined by $\overline{0}(i) = 0$ for all $i \in \mathbb{Z}$, is the zero element of $F(\mathbb{Z})$. For each $f \in F(\mathbb{Z})$, the additive inverse of f is the function $(-f)\colon \mathbb{Z} \to \mathbb{Z}$ defined by $(-f)(i) = -f(i)$ for all $i \in \mathbb{Z}$. We see that $(F(\mathbb{Z}), +, \cdot)$ is a commutative ring with unity element $\overline{1}$, where $\overline{1}$ is the function in $F(\mathbb{Z})$ defined by $\overline{1}(i) = 1$ for all $i \in \mathbb{Z}$.

Example 8. Let $M_2(\mathbb{R})$ be the set of all 2×2 matrices over the real numbers. Define $+$ to be matrix addition and $*$ to be matrix multiplication. Thus, $(M_2(\mathbb{R}), +, *)$ is a ring with unity element $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and zero element $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. In addition, $(M_2(\mathbb{R}), +, *)$ is not a commutative ring, because there are matrices A and B such that $A * B \neq B * A$ (see Example 4 on page 240). Furthermore, $(M_2(\mathbb{R}), +, *)$ has zero divisors. Let A and B in $M_2(\mathbb{R})$ be given by $A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Since $A * B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, we see that A and B are zero divisors.

8.5.1 Fundamental Properties of Rings

In the previous section, our first example of a ring was the ring of integers $(\mathbb{Z}, +, \cdot)$. We also presented some examples of rings that do not look anything like the ring of integers; however, because these rings all satisfy the ring axioms, there are some important properties that they all share. The next lemma establishes some algebraic identities that are true for all rings.

Lemma 8.5.8. *Let $(R, +, \cdot)$ be a ring and let $a, b \in R$. Then*

(1) $a0 = 0a = 0$.
(2) $a(-b) = (-a)b = -(ab)$.

(3) $-(-a) = a$
(4) $-(a+b) = (-a)+(-b)$
(5) $(-a)(-b) = ab$.
(6) *If R has unity element* 1, *then* $(-1)a = -a$.
(7) *If* $a+b = a+c$, *then* $b = c$.

Proof. We shall prove items (1)–(7). Throughout our proof we shall be referring to Axioms 1–9 of the ring axioms given in Definition 8.5.1.

(1) We show that $a0 = 0$. Since $0+0 = 0$, we have $a(0+0) = a0$. By applying Axiom 6, we derive $a0+a0 = a0$. Upon adding $-a0$ to both sides, we obtain the equation $(a0+a0)+(-a0) = a0+(-a0)$. So, $(a0+a0)+(-a0) = 0$ by Axiom 3. Hence, $a0+(a0+(-a0)) = 0$ by Axiom 2. From Axiom 4, we have $a0+0 = 0$ and then $a0 = 0$ by Axiom 3. A similar argument shows that $0a = 0$.
(2) We first establish that $a(-b) = -(ab)$, by showing that $ab+a(-b) = 0$, as follows: $ab+a(-b) = a(b+(-b)) = a0 = 0$ where the last equality holds by item (1) above. A similar argument shows that $(-a)b = -(ab)$.
(3) Since $a+(-a) = 0$, the additive inverse of $(-a)$ is a. Therefore, $-(-a) = a$.
(4) We establish the equation $-(a+b) = (-a)+(-b)$ by showing that

$$(a+b)+((-a)+(-b)) = 0$$

which follows from Axioms 1, 2, 3, and 4.
(5) We shall prove $(-a)(-b) = ab$ by using item (2) twice:

$$(-a)(-b) = -(a(-b)) = -(-(ab)) = ab$$

where the last equality follows from (3).
(6) Suppose that R has unity element 1. We prove that $(-1)a = -a$, by showing that $a+(-1)a = 0$, as follows: $a+(-1)a = 1a+(-1)a = (1+(-1))a = 0a = 0$.
(7) Suppose $a+b = a+c$. Then $(-a)+(a+b) = (-a)+(a+c)$. Using Axioms 1, 2, 3 and 4, one can show that $b = c$. □

Lemma 8.5.9 (Cancellation Law). *Let $(R,+,\cdot)$ be an integral domain and let $a,b,c \in R$ where $c \neq 0$. If $ac = bc$, then $a = b$.*

Proof. Let $a,b,c \in R$ where $c \neq 0$. Assume $ac = bc$. Thus, $ac-bc = (a-b)c = 0$. Because $c \neq 0$ and $(R,+,\cdot)$ is an integral domain, we conclude that $a-b = 0$ and hence, $a = b$. □

8.5.2 Subrings

Let $(R,+,\cdot)$ be a ring and let $S \subseteq R$ be a substructure of R (see Definition 8.2.3). So, S is nonempty and is closed under the binary operations $+$ and $\cdot$ (addition and multiplication). If $(S,+,\cdot)$ is also a ring, then we say that S is a subring of R.

Definition 8.5.10. Let $(R,+,\cdot)$ be a ring. Then S is called a **subring** of R if S is a substructure of R that also satisfies the ring axioms. In other words, the structure $(S,+,\cdot)$ is a ring.

Lemma 8.5.11. *Let $(R,+,\cdot)$ be a ring and let $S \subseteq R$. Then S is a subring of R if and only if $0 \in S$, $a \cdot b \in S$, $-b \in S$, and $a+b \in S$ for all $a,b \in S$.*

Proof. The ideas used in the proof of Lemma 8.3.13 (page 253) apply here. □

Example 9. Let $(\mathbb{R},+,\cdot)$ be the ring of real numbers and define $\mathbb{Z}[\sqrt{3}] \subseteq \mathbb{R}$ by $\mathbb{Z}[\sqrt{3}] = \{i+j\sqrt{3} : i,j \in \mathbb{Z}\}$. Show that $\mathbb{Z}[\sqrt{3}]$ is a subring of $\mathbb{R}$.

Solution. We shall apply Lemma 8.5.11. Since $0 = 0+0\sqrt{3}$, we see that $0 \in \mathbb{Z}[\sqrt{3}]$. Let $i+j\sqrt{3}$ and $m+n\sqrt{3}$ be elements in $\mathbb{Z}[\sqrt{3}]$, where $i,j,m,n \in \mathbb{Z}$. We shall show that $(i+j\sqrt{3})(m+n\sqrt{3})$ is in $\mathbb{Z}[\sqrt{3}]$ as follows:

$$\begin{aligned}
(i+j\sqrt{3})(m+n\sqrt{3}) &= im+in\sqrt{3}+jm\sqrt{3}+jn\sqrt{3}\sqrt{3} && \text{by algebra} \\
&= im+in\sqrt{3}+jm\sqrt{3}+3jn && \text{because } \sqrt{3}\sqrt{3}=3 \\
&= im+(in+jm)\sqrt{3}+3jn && \text{by distributivity} \\
&= (im+3jn)+(in+jm)\sqrt{3} && \text{by algebra.}
\end{aligned}$$

Since $im+3jn$ and $in+jm$ are in $\mathbb{Z}$, we see that $(i+j\sqrt{3})(m+n\sqrt{3})$ is in $\mathbb{Z}[\sqrt{3}]$. Because $-(i+j\sqrt{3}) = -i+(-j)\sqrt{3}$, we have that $-(i+j\sqrt{3}) \in \mathbb{Z}[\sqrt{3}]$. Finally, observe that

$$(i+j\sqrt{3})+(m+n\sqrt{3}) = (i+m)+(j+n)\sqrt{3}$$

and thus $(i+j\sqrt{3})+(m+n\sqrt{3}) \in \mathbb{Z}[\sqrt{3}]$, because $i+m$ and $j+n$ are in $\mathbb{Z}$. Lemma 8.5.11 implies that $\mathbb{Z}[\sqrt{3}]$ is a subring of $\mathbb{R}$. Ⓢ

Lemma 8.5.12. *Let $(R,+,\cdot)$ be a commutative ring with unity element 1. Suppose that S is a subring of R with $1 \in S$. If R is an integral domain, then S is also an integral domain.*

Proof. Since R is commutative and $1 \in S$, it follows that S is a commutative ring with unity element. Let $a,b \in S$ be such that $a \cdot b = 0$. Because R is an integral domain, we have either $a = 0$ or $b = 0$. Therefore, S is an integral domain. □

8.5.3 Ideals

An ideal is a special subset of a ring that allows one to generalize some important properties of the integers like "even number" and "multiple of 5." For example, the set of even integers contains 0 and is closed under addition. Moreover, any integer multiple of an even integer is also even.

Definition 8.5.13. Let $(R,+,\cdot)$ be a ring. A subset $I \subseteq R$ is called an **ideal** of R if the following three properties hold:

(1) $0 \in I$.
(2) For all $i \in I$ and all $j \in I$, we have $i+j \in I$ and $-i \in I$.
(3) For all $r \in R$ and all $i \in I$, we have $ir \in I$ and $ri \in I$.

Item (2) of Definition 8.5.13 states that I is closed under addition and additive inverses. Item (3) asserts that for any element in $i \in I$ and any element in $r \in R$, the products $i \cdot r$ and $r \cdot i$ are in I. Since $I \subseteq R$, we conclude that I is closed under the multiplication of elements in I, as well. It follows that an ideal I of a ring R is a subring of R. Before proceeding further, we shall present some examples of ideals.

Example 10. Let $(R,+,\cdot)$ be a ring with zero element 0. Then the sets $\{0\}$ and R are ideals of the ring R.

Example 11. Let $(\mathbb{Z},+,\cdot)$ be the ring of integers and let m be a natural number. Then the set $I = \{mk : k \in \mathbb{Z}\}$ is an ideal of the ring of integers. To verify this, we shall show that I satisfies items (1)–(3) of Definition 8.5.13. (1) Clearly, $0 = m \cdot 0$ and thus, $0 \in I$. (2) Let $i \in I$ and $j \in I$. So $i = mk$ and $j = m\ell$ for some $k, \ell \in Z$. We see that $i + j = mk + m\ell = m(k+\ell)$. Since $k + \ell \in \mathbb{Z}$, we conclude that $i + j \in I$. Since $-i = -mk = m(-k)$, where $-k \in \mathbb{Z}$, we have that $-i \in I$. (3) Let $r \in \mathbb{Z}$ and let $i \in I$. So, $i = mk$ for some $k \in \mathbb{Z}$. Since $ri = r(mk) = m(rk)$ where $rk \in \mathbb{Z}$, we see that $ri \in I$. Because the ring $\mathbb{Z}$ is commutative, we also have that $ir \in I$. Therefore, I is an ideal of the ring of integers.

Example 12. From Theorem 7.4.4 we conclude that $(\mathbb{Z}_{12},+,\cdot)$ is a commutative ring with zero element $[0]$. One can show that $I = \{[0],[4],[8]\}$ is an ideal of $\mathbb{Z}_{12}$.

Example 13. Let $(C(\mathbb{R}),+,\cdot)$ be the ring of all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ given in Example 5. Recall that zero element of $C(\mathbb{R})$ is the constant function $\overline{0}\colon \mathbb{R} \to \mathbb{R}$ defined by $\overline{0}(x) = 0$ for all $x \in \mathbb{R}$. Let a be a fixed real number. We will show that the set $I = \{f \in C(\mathbb{R}) : f(a) = 0\}$ is an ideal of the ring $C(\mathbb{R})$. To do this, we shall confirm that I satisfies items (1)–(3) of Definition 8.5.13. (1) Clearly, $\overline{0}(a) = 0$ and thus, $\overline{0} \in I$. (2) Let $f \in I$ and $g \in I$. So $f(a) = 0$ and $g(a) = 0$. We see that $(f+g)(a) = f(a) + g(a) = 0$ and conclude that $f + g \in I$. Since $(-f)(a) = -f(a) = 0$, we have that $(-f) \in I$. (3) Let $h \in C(\mathbb{R})$ and let $f \in I$. Thus, $(h \cdot f)(a) = h(a) \cdot f(a) = h(a) \cdot 0 = 0$. We see that $hf \in I$. Because the ring $C(\mathbb{R})$ is commutative, we also have that $fh \in I$. Therefore, I is an ideal.

One of the motivations for the concept of an ideal is that it can be used to build a new ring called the quotient ring (see Section 8.6.2). To prove that a subset I of a ring R is an ideal, you must show that I satisfies properties (1)–(3) of Definition 8.5.13. On the other hand, when assuming that I is an ideal, then you know that $0 \in I$, and if you are assuming or can prove that $i, j \in I$, then you can conclude that $i + j \in I$ and $-i \in I$. In addition, if you have that $i \in I$ and $r \in R$, then you can infer that $ir \in I$ and $ri \in I$.

Theorem 8.5.14. *Let $(R,+,\cdot)$ be a ring. Suppose that I and J are ideals of R. Then the set $K=\{i+j : i\in I \text{ and } j\in J\}$ is also an ideal of R.*

Proof. Let $(R,+,\cdot)$ be a ring and let I and J be ideals of R. We will prove that

$$K=\{i+j : i\in I \text{ and } j\in J\}$$

is an ideal of R by establishing items (1)–(3) of Definition 8.5.13.

(1) We first prove that $0\in K$. Since I and J are ideals, it follows that $0\in I$ and $0\in J$. Thus, $0+0\in K$ and, because $0+0=0$, we see that $0\in K$.

(2) Let $x,y\in K$. So $x=i+j$ and $y=m+n$ for some $i,m\in I$ and $j,n\in J$. Because I and J are ideals, we have that $i+m\in I$ and $j+n\in J$. Since

$$x+y=(i+j)+(m+n)=(i+m)+(j+n),$$

we conclude that $x+y\in K$. Furthermore, $-x=-i+-j$ and thus, $-x\in K$ because $-i\in I$ and $-j\in J$.

(3) Let $r\in R$ and let $x\in K$. Given that $x\in K$, we have $x=i+j$ for some $i\in I$ and $j\in J$. Since I and J are ideals, we know that $ri,ir\in I$ and $rj,jr\in J$. We evaluate rx to be

$$rx=r(i+j)=ri+rj$$

and therefore, $rx\in K$. A similar argument shows that $xr\in K$.

Therefore, K is an ideal of R. □

Exercises 8.5

1. Let R be the set of matrices of the form $\begin{bmatrix}0 & a\\ 0 & b\end{bmatrix}$ where $a,b\in\mathbb{R}$. Show that $(R,+,*)$ is a ring, where $+$ and $*$ are matrix addition and multiplication. Is this ring R commutative? Does R have a unity element? Does R have any zero divisors?

2. Let $(F(\mathbb{Z}),+,\cdot)$ be the commutative ring defined in Example 7. Show that $(F(\mathbb{Z}),+,\cdot)$ is not an integral domain.

3. Let $(R,+,\cdot)$ be a ring and let $a,b\in R$. Prove that $-(a-b)=b-a$.

4. Prove Lemma 8.5.2.

5. Prove Lemma 8.5.3.

6. Prove Lemma 8.5.11.

7. Find all of the zero divisors (if any) of the following rings: $\mathbb{Z}_8$, $\mathbb{Z}_{25}$, and $\mathbb{Z}_{17}$. Then find all of the elements in each of these rings that are units.

8. Let $(\mathbb{R},+,\cdot)$ be the ring of real numbers. Define the subset $\mathbb{Z}[\sqrt{2}]$ of $\mathbb{R}$ by $\mathbb{Z}[\sqrt{2}]=\{a+b\sqrt{2} : a,b\in\mathbb{Z}\}$.

(a) Prove that $\mathbb{Z}[\sqrt{2}]$ is a subring of $\mathbb{R}$ (see Example 9).
(b) Prove that $\mathbb{Z}[\sqrt{2}]$ is an integral domain.
(c) Let $a, b \in \mathbb{Z}$ be such that $a + b\sqrt{2} = 0$. Prove that $a = b = 0$.
(d) Show that $1 + \sqrt{2}$ is a unit in $\mathbb{Z}[\sqrt{2}]$.
(e) Show that $2 + 2\sqrt{2}$ is not a unit in $\mathbb{Z}[\sqrt{2}]$.
(f) Prove that for $a, b \in \mathbb{Z}$, if $a^2 - 2b^2 = \pm 1$, then $a + b\sqrt{2}$ is a unit in $\mathbb{Z}[\sqrt{2}]$.

9. Consider the ring $(\mathbb{Z}[\sqrt{2}], +, \cdot)$ given in Exercise 8. Let $S \subseteq \mathbb{Z}[\sqrt{2}]$ be defined by $S = \{a + b\sqrt{2} : a, b \in \mathbb{Z} \text{ and } a \text{ is even}\}$. Prove that S is a subring of $\mathbb{Z}[\sqrt{2}]$.

10. Let $(R, +, \cdot)$ be a ring and suppose that $\{S_\ell : \ell \in L\}$ is an indexed family of subrings of R. Prove that $\bigcap_{\ell \in L} S_\ell$ is a subring of R.

11. We know from Theorem 7.4.4 that $(\mathbb{Z}_{12}, +, \cdot)$ is a commutative ring with zero element $[0]$. Let $I = \{[0], [4], [8]\}$. Show that I is an ideal of $\mathbb{Z}_{12}$.

12. Let $(\mathbb{Z}[\sqrt{2}], +, \cdot)$ be the ring given in Exercise 8. Let $I \subseteq \mathbb{Z}[\sqrt{2}]$ be defined by $I = \{a + b\sqrt{2} \in \mathbb{Z}[\sqrt{2}] : 3 \mid a \text{ and } 3 \mid b\}$. Prove that I is an ideal of the ring $\mathbb{Z}[\sqrt{2}]$.

13. By Example 9, we know that $(\mathbb{Z}[\sqrt{3}], +, \cdot)$ is a ring. Let $I \subseteq \mathbb{Z}[\sqrt{3}]$ be defined by $I = \{a + b\sqrt{3} \in \mathbb{Z}[\sqrt{3}] : 3 \mid a\}$. Prove that I is an ideal of $\mathbb{Z}[\sqrt{3}]$.

14. Let $(M_2(\mathbb{R}), +, *)$ be the ring of 2×2 matrices given in Example 8. Define $S_2(\mathbb{R}) \subseteq M_2(\mathbb{R})$ by

$$S_2(\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a, b \in \mathbb{R} \right\}.$$

Show that $S_2(\mathbb{R})$ is a subring of the ring $M_2(\mathbb{R})$.

15. Let $(S_2(\mathbb{R}), +, *)$ be the ring of 2×2 matrices given in Exercise 14, where $+$ and $*$ are matrix addition and multiplication, respectively. Define $I \subseteq S_2(\mathbb{R})$ by

$$I = \left\{ \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} : b \in \mathbb{R} \right\}.$$

Show that I is an ideal of the ring $S_2(\mathbb{R})$.

16. Let $(\mathbb{R}, +, \cdot)$ be the ring of real numbers. Let $\mathbb{Q}$ be the set of rational numbers. Define the subset $\mathbb{Q}(\sqrt{3})$ of $\mathbb{R}$ by $\mathbb{Q}(\sqrt{3}) = \{a + b\sqrt{3} : a, b \in \mathbb{Q}\}$.
(a) Prove that $\mathbb{Q}(\sqrt{3})$ is a subring of $\mathbb{R}$.
(b) Prove that for all $x \in \mathbb{Q}(\sqrt{3})$, if $x \neq 0$, then x is a unit in $\mathbb{Q}(\sqrt{3})$.

17. Let $(R, +, \cdot)$ be a commutative ring and let $a \in R$. Define $I = \{x \in R : xa = 0\}$. Prove that I is an ideal of R.

18. Let $(R, +, \cdot)$ be a ring. Suppose that $I \subseteq R$ and $J \subseteq R$ are ideals of R. Prove that $I \cap J$ is also an ideal of R.

19. Let $(R, +, \cdot)$ be a commutative ring. Let I be an ideal of R and let $a \in R$. Define $H \subseteq R$ by $H = \{ax : x \in I\}$. Prove that H is an ideal of R.

20. Let $(F(\mathbb{Z}), +, \cdot)$ be the ring in Example 7 on page 275, where $F(\mathbb{Z})$ is the set of all functions of the form $f : \mathbb{Z} \to \mathbb{Z}$. Let $I = \{f \in F(\mathbb{Z}) : 5 \mid f(1)\}$. Prove that I is an ideal of $F(\mathbb{Z})$.

21. Let $(R,+,\cdot)$ be a ring. Suppose that $\{I_\ell : \ell \in L\}$ is an indexed family of ideals of R. Prove that $\bigcap_{\ell \in L} I_\ell$ is an ideal of R.

Exercise Notes: For Exercise 4, show that the equation $a+b=a$ implies that $b=0$. For Exercise 5, show that the equation $a+b=0$ implies $b=-a$. For Exercise 8(b), use Lemma 8.5.12 and the fact that $(\mathbb{R},+,\cdot)$ is an integral domain. For Exercise 8(c), assume $b \neq 0$ and derive a contradiction. For Exercise 16(b), rationalize the denominator in $\frac{1}{a+b\sqrt{3}}$.

8.6 Quotient Algebras

In Section 7.4 we constructed the modular number system $(\mathbb{Z}_m,+,\cdot)$ from the integer number system $(\mathbb{Z},+,\cdot)$. The key ingredient that allowed us to construct the new system $(\mathbb{Z}_m,+,\cdot)$ was the congruence modulo m relation. This equivalence relation on $\mathbb{Z}$ is called a congruence relation because addition and multiplication preserve the relation, namely: For all a,b,c,d in $\mathbb{Z}$, if $a \equiv b \pmod m$ and $c \equiv d \pmod m$, then $(a+c) \equiv (b+d) \pmod m$ and $ac \equiv bd \pmod m$ (see Theorem 7.3.5).

In this section we will see how this method of construction can be generalized. As a result we will be able to build new algebraic structures from old ones. We first must define what it means for a binary operation to preserve an equivalence relation.

Definition 8.6.1. Let $\sim$ be an equivalence relation on a set A and let $*$ be a binary operation on A. Then $*$ **preserves the relation** $\sim$ when for all $a,b,c,d \in A$

$$\text{if } a \sim c \text{ and } b \sim d, \text{ then } (a*b) \sim (c*d).$$

Definition 8.6.2. Let $\mathcal{A} = (A,*,+,\dots)$ be an algebraic structure and let $\sim$ be an equivalence relation on the set A. We say that $\sim$ is a **congruence relation** on $\mathcal{A}$ if all of the binary operations $*,+,\dots$ preserve the relation $\sim$.

Examples of Congruence Relations

The definition of a congruence relation depends on the type of algebraic structure under consideration. We first give an example of a congruence relation defined on a group and then give two examples of a congruence relation defined on a ring.

Example 1. Let $\mathbb{R}^* = \{x \in \mathbb{R} : x \neq 0\}$, the set of non-zero real numbers. Thus, $(\mathbb{R}^*,\cdot)$ is an abelian group where $\cdot$ is ordinary multiplication (see Example 3 on page 248). Let $\sim$ be the equivalence relation on $\mathbb{R}^*$ defined by $x \sim y$ if and only if $x \cdot y^{-1} \in \mathbb{Q}$. Show that $\sim$ is a congruence relation on $(\mathbb{R}^*,\cdot)$.

Solution. Let $x, y, z, w \in \mathbb{R}^*$ and assume that $x \sim y$ and $z \sim w$, that is,

$$xy^{-1} \in \mathbb{Q} \text{ and } zw^{-1} \in \mathbb{Q}.$$

We shall prove that $(x \cdot z) \sim (y \cdot w)$. Since $xy^{-1} \in \mathbb{Q}$ and $zw^{-1} \in \mathbb{Q}$, we have $(xy^{-1})(zw^{-1}) \in \mathbb{Q}$. By algebra, we conclude that $(xy^{-1})(zw^{-1}) = (xz)(yw)^{-1}$ and hence, $(xz)(yw)^{-1} \in \mathbb{Q}$. Therefore, $(x \cdot z) \sim (y \cdot w)$. Ⓢ

Example 2. Let $(\mathbb{Z}, +, \cdot)$ be the ring of integers, where $+$ and $\cdot$ are the standard operations of addition and multiplication. Let $m > 1$ be a natural number. Define the equivalence relation $\sim$ on $\mathbb{Z}$ by

$$x \sim y \text{ if and only if } x \equiv y \pmod{m} \tag{8.17}$$

for all $x, y \in \mathbb{Z}$. Then $\sim$ is a congruence relation on $(\mathbb{Z}, +, \cdot)$ by Theorem 7.3.5.

Example 3. Let $(F(\mathbb{Z}), +, \cdot)$ be the ring in Example 7 on page 275, where $F(\mathbb{Z})$ is the set of all functions of the form $f\colon \mathbb{Z} \to \mathbb{Z}$. Recall that $(f+g)\colon \mathbb{Z} \to \mathbb{Z}$ is defined by $(f+g)(i) = f(i) + g(i)$ and the function $(f \cdot g)\colon \mathbb{Z} \to \mathbb{Z}$ is defined by $(f \cdot g)(i) = f(i)g(i)$. Let $\sim$ be the equivalence relation on $F(\mathbb{Z})$ defined by $f \sim g$ if and only if $5 \,|\, (f(1) - g(1))$, for all $f, g \in F(\mathbb{Z})$. Show that $\sim$ is a congruence relation on $(F(\mathbb{Z}), +, \cdot)$.

Solution. For $f, g \in F(\mathbb{Z})$, observe that

$$f \sim g \text{ if and only if } f(1) \equiv g(1) \pmod{5}. \tag{8.18}$$

Let $f, g, h, k \in F(\mathbb{Z})$. Assume $f \sim g$ and $h \sim k$. We shall show that

$$(f+h) \sim (g+k) \text{ and } (f \cdot h) \sim (g \cdot k).$$

Since $f \sim g$ and $h \sim k$, we see from (8.18) that

$$f(1) \equiv g(1) \pmod{5} \text{ and } h(1) \equiv k(1) \pmod{5}.$$

Thus, by Theorem 7.3.5, we have that

$$(f(1) + h(1)) \equiv (g(1) + k(1)) \pmod{5} \text{ and } (f(1) \cdot h(1)) \equiv (g(1) \cdot k(1)) \pmod{5}.$$

Hence, $(f+h) \sim (g+k)$ and $(f \cdot h) \sim (g \cdot k)$ and therefore, $\sim$ is a congruence relation on $(F(\mathbb{Z}), +, \cdot)$. Ⓢ

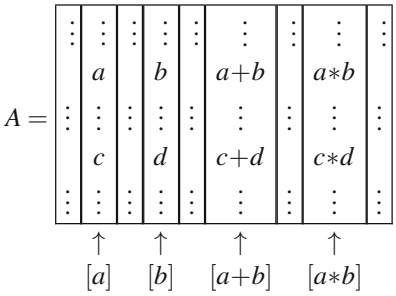
The next theorem introduces an important concept that allows one to construct a new algebraic structure from an old algebraic structure. The elements of the new structure are the equivalence classes of a congruence relation on the old structure. Recall that when $\sim$ is an equivalence relation on a set A, we let $A/\!\sim$ denote the partition $\{[a] : a \in A\}$ of A induced by $\sim$ (see Definition 7.2.7).

Theorem 8.6.3 (Quotient Algebra). *Suppose that $\mathcal{A} = (A, *, +, \dots)$ is an algebraic structure. Let $\sim$ be a congruence relation on $\mathcal{A}$. Then there exist well-defined binary operations $\circledast, \oplus, \dots$ on $A/_\sim$ such that*

$$[a] \circledast [b] = [a * b]$$
$$[a] \oplus [b] = [a + b]$$
$$\vdots$$

for all $a, b \in A$. Therefore, $\mathcal{B} = (A/_\sim, \circledast, \oplus, \dots)$ is also an algebraic structure, called the quotient algebra.

The conclusion of Theorem 8.6.3 is illustrated in the following figure where $A/_\sim = \{\dots, [a], \dots, [b], \dots, [a+b], \dots, [a*b], \dots\}$.



Before reading the proof below, one should review Section 8.1.1 which begins on page 241 and ends on page 244.

Proof (of Theorem 8.6.3). We only prove that if $*$ preserves the relation $\sim$, then there exists a well-defined binary operation $\circledast$ on $A/_\sim$ such that $[a] \circledast [b] = [a * b]$ for all $a, b \in A$. (The argument is the same for the other binary operations.) So assume that $*$ preserves the relation $\sim$. Thus, for all $a, b, c, d \in A$

$$\text{if } a \sim c \text{ and } b \sim d, \text{ then } (a * b) \sim (c * d). \tag{8.19}$$

We must show that the definition $[a] \circledast [b] = [a * b]$ for any $a, b \in A$, produces a well-defined binary operation. Assume $[a] = [c]$ and $[b] = [d]$. We will show that $[a * b] = [c * d]$. Since $[a] = [c]$ and $[b] = [d]$, we have that $a \sim c$ and $b \sim d$, by Theorem 7.2.4. So, by (8.19), we see that $(a * b) \sim (c * d)$. Theorem 7.2.4 implies that $[a * b] = [c * d]$. Therefore, $\circledast$ is a well-defined binary operation on $A/_\sim$. □

In the next two sections, we will use Theorem 8.6.3 to construct quotient groups and quotient rings.

8.6.1 Quotient Groups

We shall apply the method given in Theorem 8.6.3 to construct a new group from an old group. The new group can then yield some important information about the original group.

Theorem 8.6.4 (Quotient Group). *Let $(G,*)$ be a group and let $\sim$ be a congruence relation on $(G,*)$. Then $(G/_\sim,\circledast)$ is a group, where the binary operation $\circledast$ on $G/_\sim$ is well-defined by*

$$[a]\circledast[b]=[a*b]$$

for all $a,b\in G$. The group $(G/_\sim,\circledast)$ is called a quotient group.

Proof. Theorem 8.6.3 asserts that the binary operation $\circledast$ on $G/_\sim$ is well-defined by

$$[x]\circledast[y]=[x*y] \tag{8.20}$$

for all $x,y\in G$. We shall show that the algebraic structure $(G/_\sim,\circledast)$ satisfies the group axioms. Recall that $G/_\sim\ =\{[a]:a\in G\}$.

1. ASSOCIATIVITY: For $[a],[b],[c]\in G/_\sim$, we have the following:

$$\begin{aligned}
[a]\circledast([b]\circledast[c]) &= [a]\circledast[b*c] && \text{by (8.20)}\\
&= [a*(b*c)] && \text{by (8.20)}\\
&= [(a*b)*c)] && \text{by associativity in } (G,*)\\
&= [a*b]\circledast[c] && \text{by (8.20)}\\
&= ([a]\circledast[b])\circledast[c] && \text{by (8.20).}
\end{aligned}$$

   Therefore, $[a]\circledast([b]\circledast[c])=([a]\circledast[b])\circledast[c]$.
2. IDENTITY ELEMENT EXISTS: Let $e\in G$ be the identity element for the group $(G,*)$. Thus, $[e]\in G/_\sim$. We show that $[e]$ is the identity element for $(G/_\sim,\circledast)$. Let $[a]\in G/_\sim$. Then $[a]\circledast[e]=[a*e]=[a]$ and $[e]\circledast[a]=[e*a]=[a]$, by (8.20) and the fact that e is the identity element in G. So $[a]\circledast[e]=[e]\circledast[a]=[a]$ and therefore, $[e]$ is the identity element for $(G/_\sim,\circledast)$.
3. INVERSE ELEMENTS EXIST: Let $[a]\in G/_\sim$. Let $[a]^{-1}=[a^{-1}]$. One can check that $[a]\circledast[a]^{-1}=[a]^{-1}\circledast[a]=[e]$, using (8.20) and the fact that a^{-1} is the inverse for a in G.

Since all of the group axioms hold, we see that $(G/_\sim,\circledast)$ is a group. □

Definition 8.6.5. Let $(G,*)$ be a group and let H be a subgroup of G. Let $\sim_H$ be the relation on G defined by $a\sim_H b$ if and only if $ab^{-1}\in H$, for all $a,b\in G$.

Theorem 8.6.6. *Let $(G,*)$ be a group and let H be a subgroup of G. The relation $\sim_H$ is an equivalence relation on G.*

Proof. See Exercise 5. □

Definition 8.6.7. Let $(G,*)$ be a group and let H be a subgroup of G. For each $a \in G$, the equivalence class of a is defined by $[a] = \{b \in G : b \sim_H a\}$. We write G/H as the set of equivalence classes of the relation $\sim_H$, that is, $G/H = G/_{\sim_H} = \{[a] : a \in G\}$.

The notion of a right coset is very useful for 'computing' each equivalence class in G/H, whenever H is a subgroup of a group G.

Definition 8.6.8 (Right Cosets). Let H be a subgroup of a group G. A **right coset** of H is a subset of G having the form $Ha = \{ha : h \in H\}$ for some $a \in G$.

Theorem 8.6.9. *Let $(G,*)$ be a group and let H be a subgroup of G. Consider the equivalence relation $\sim_H$. For each $a \in G$ we have that $[a] = Ha$. Consequently, $G/H = \{[a] : a \in G\} = \{Ha : a \in G\}$.*

Proof. Let G and H be as stated in the theorem. Recall that the equivalence relation $\sim_H$ on G is defined by

$$a \sim_H b \text{ if and only if } ab^{-1} \in H \tag{8.21}$$

for all $a, b \in G$. Now, let $a \in G$. We shall prove that $[a] = Ha$, that is, we shall prove that these two sets are equal. For each $x \in G$ we see that

$$\begin{aligned} x \in [a] \text{ iff } & x \sim_H a && \text{by the definition of } [a] \\ \text{iff } & xa^{-1} \in H && \text{by (8.21)} \\ \text{iff } & (xa^{-1})a \in Ha && \text{by the definition of } Ha \\ \text{iff } & x \in Ha && \text{since } (xa^{-1})a = x. \end{aligned}$$

Therefore, $[a] = Ha$. Consequently, $G/H = \{[a] : a \in G\} = \{Ha : a \in G\}$. □

Normal subgroups are important in group theory because they can be used to construct quotient groups.

Theorem 8.6.10. *Let $(G,*)$ be a group. If N is a normal subgroup of G, then the binary operation $*$ preserves the equivalence relation $\sim_N$.*

Proof. Let $(G,*)$ be a group and let N be a normal subgroup of G. Let $a,b,c,d \in G$ and assume $a \sim_N c$ and $b \sim_N d$. Thus, by Definition 8.6.5, $ac^{-1} \in N$ and $bd^{-1} \in N$. By Exercise 20 on page 260, we have $(ab)(cd)^{-1} \in N$. Hence $(ab) \sim_N (cd)$, that is, $(a*b) \sim_N (c*d)$ □

The next theorem gives "possibly the single most important construction in group theory" (see [10, page 79]).

Theorem 8.6.11. *Let $(G,*)$ be a group and let N be a normal subgroup of G. Then the algebraic structure $(G/N, \circledast)$ is a group where the binary operation $\circledast$ on G/N is well-defined by $[a] \circledast [b] = [a*b]$ for all $a,b \in G$.*

Proof. Theorem 8.6.10 asserts that $\sim_N$ is a congruence relation on $(G,*)$. Thus, Theorem 8.6.3 implies that the binary operation $\circledast$ on G/N is well-defined. Hence, $(G/N,\circledast)$ is a group by Theorem 8.6.4. □

Examples of Quotient Groups

Example 4. Let $(\mathbb{R}^*,\cdot)$ be the group of nonzero real numbers under multiplication. Because $(\mathbb{R}^*,\cdot)$ is an abelian group, it follows that $N=\{-1,1\}$ is a normal subgroup of $\mathbb{R}^*$ (see Exercise 15 on page 260). We will construct the quotient group $(\mathbb{R}^*/N,\odot)$. The relation $\sim_N$ is defined by $x\sim_N y$ if and only if $xy^{-1}\in N$, for all $x,y\in\mathbb{R}^*$. Since $xy^{-1}\in N$ means that $\frac{x}{y}=\pm 1$, we have that $x\sim_N y$ if and only if $x=\pm y$. Because $x=\pm|x|$, we see that $x\sim|x|$ and thus, by Theorem 7.2.4, we have that $[x]=[\,|x|\,]$ for each $x\in\mathbb{R}^*$. Hence, $\mathbb{R}^*/N=\{[x]:x\in\mathbb{R}^+\}$ and

$$[x]\odot[y]=[x\cdot y] \text{ and } [x]^{-1}=[x^{-1}]=\left[\frac{1}{x}\right]$$

for all $x,y\in\mathbb{R}^+$. So the quotient group $(\mathbb{R}^*/N,\odot)$ is very similar[4] to the group $(\mathbb{R}^+,\cdot)$ of positive real numbers under multiplication (see Example 4 on page 249).

Example 5. Define the function $T_{a,b}\colon\mathbb{R}\to\mathbb{R}$ by $T_{a,b}(x)=ax+b$, whenever $a,b\in\mathbb{R}$ with $a\neq 0$. Let $G=\{T_{a,b}:a,b\in\mathbb{R}\text{ and }a\neq 0\}$, the set of all such functions. We know that $(G,\circ)$ is a group where $\circ$ is functional composition (see Example 5 on page 249). In Example 15 on page 256 we showed that $N=\{T_{1,b}\in G:b\in\mathbb{R}\}$ is a normal subgroup of G. Let us construct and investigate the quotient group $(G/N,\circledcirc)$. Recall that $\sim_N$ is defined by

$$T_{a,b}\sim_N T_{c,d} \text{ if and only if } T_{a,b}\circ T_{c,d}^{-1}\in N$$

for all $T_{a,b},T_{c,d}\in G$. One can easily show that $T_{a,b}\circ T_{c,d}^{-1}=T_{\frac{a}{c},\ell}$ for some real number ℓ. Since $T_{\frac{a}{c},\ell}\in N$ means that $\frac{a}{c}=1$, we conclude that

$$T_{a,b}\sim_N T_{c,d} \text{ if and only if } a=c. \tag{8.22}$$

Therefore, (8.22) implies that $T_{a,b}\sim_N T_{a,d}$ for any value of d. So, we can use $d=0$. Consequently, for every $a,b\in\mathbb{R}$ we have that $T_{a,b}\sim_N T_{a,0}$. Thus, $[T_{a,b}]=[T_{a,0}]$. We conclude that $G/N=\{[T_{a,0}]:a\in\mathbb{R}\}$. Since $T_{a,0}\circ T_{c,0}=T_{ac,0}$ and $T_{a,0}^{-1}=T_{\frac{1}{a},0}$, we also have that

$$[T_{a,0}]\circledcirc[T_{c,0}]=[T_{a,0}\circ T_{c,0}]=[T_{ac,0}] \text{ and } [T_{a,0}]^{-1}=[T_{a,0}^{-1}]=[T_{\frac{1}{a},0}].$$

[4] In fact, the two groups are *isomorphic* (see page 68 of [10]).

The above operations in $(G/N, \circledcirc)$ behave very much like those in the group $(\mathbb{R}^*, \cdot)$ of nonzero real numbers (see Example 3 on page 248). In particular, the group $(G/N, \circledcirc)$ is abelian. Thus, we have constructed the abelian group $(G/N, \circledcirc)$ from the nonabelian group $(G, \circ)$.

8.6.2 Quotient Rings

In Section 7.4 we constructed the ring $(\mathbb{Z}_m, \oplus, \odot)$ from the ring of integers $(\mathbb{Z}, +, \cdot)$. We will now investigate how to generalize this construction. The first step in the construction of $(\mathbb{Z}_m, \oplus, \odot)$ was the equivalence relation $a \equiv b \pmod{m}$ on $\mathbb{Z}$, given in Definition 7.3.1. Similarly, the first step in our generalization will be to define an appropriate equivalence relation on the elements of a ring.

Theorem 8.6.12. *Let $(R, +, \cdot)$ be a ring and let $I \subseteq R$ be an ideal. Define the relation $\sim$ on R by*

$$a \sim b \text{ if and only if } (a - b) \in I \tag{8.23}$$

for all $a, b \in R$. Then $\sim$ is an equivalence relation on R.

Proof. See Exercise 13. □

Definition 8.6.13. Let $(R, +, \cdot)$ be a ring. Whenever $I \subseteq R$ is an ideal, let $\sim_I$ be the equivalence relation on R defined by $a \sim_I b$ if and only if $(a - b) \in I$, for all $a, b \in R$. Let R/I denote the set of equivalence classes of the relation $\sim_I$, that is, $R/I = R/_{\sim_I}$.

Theorem 8.6.14. *Let $(R, +, \cdot)$ be a ring and let $I \subseteq R$ be an ideal. Then $\sim_I$ is a congruence relation on $(R, +, \cdot)$.*

Proof. We must show that $+$ and $\cdot$ preserve the equivalence relation $\sim_I$. Let $a, b, c, d \in R$ and assume $a \sim_I c$ and $b \sim_I d$, that is, assume that

$$(a - c) \in I \tag{8.24}$$

$$(b - d) \in I. \tag{8.25}$$

We first prove $(a + b) \sim_I (c + d)$. Since I is closed under $+$, we conclude from (8.24) and (8.25) that $(a - c) + (b - d) \in I$. In addition, by Lemma 8.5.8 and the ring axioms, we have $(a - c) + (b - d) = (a + b) - (c + d)$. Thus, $(a + b) - (c + d) \in I$ and hence, $(a + b) \sim_I (c + d)$.

To prove $ab \sim_I cd$, we need to show that $(ab - cd) \in I$. Since I is an ideal, we see from (8.24) that $(a - c)b = (ab - cb) \in I$ (see Definition 8.5.13(2)). In addition, from (8.25), we conclude that $c(b - d) = (cb - cd) \in I$, again because I is an ideal. Thus, $(ab - cb) + (cb - cd) \in I$, as I is a closed under $+$. Hence, $(ab - cd) \in I$. Therefore, $\sim_I$ is a congruence relation. □

Abstract algebra evolved from the desire to find the roots of polynomials. Our next theorem can be used (as you will see in your abstract algebra courses) to construct a new ring in which a particular polynomial has a root.

Theorem 8.6.15 (Quotient Ring). *Let $(R,+,\cdot)$ be a ring and let $I \subseteq R$ be an ideal. Then $(R/I,\oplus,\odot)$ is also a ring where the binary operations $\oplus$ and $\odot$ on R/I are well-defined by*

$$[a]\oplus[b] = [a+b] \tag{8.26}$$

$$[a]\odot[b] = [a\cdot b] \tag{8.27}$$

for all $a,b \in R$. The ring $(R/I,\oplus,\odot)$ is called a quotient ring.

Proof. Theorem 8.6.14 states that $\sim_I$ is a congruence relation on $(R,+,\cdot)$. Thus, the binary operations $\odot$ and $\oplus$ on R/I are well-defined by Theorem 8.6.3. The proof that $(R/I,\oplus,\odot)$ is a ring is analogous to the proof of Theorem 8.6.4. □

Examples of Quotient Rings

Example 6. Let $(\mathbb{Z},+,\cdot)$ be the ring of integers and let m be a natural number. Then the set $I = \{mk : k \in \mathbb{Z}\}$ is an ideal of the ring of integers. We will construct and evaluate the quotient ring $(\mathbb{Z}/I,\oplus,\odot)$. The equivalence relation $\sim_I$ is defined by $x \sim_I y$ if and only if $x-y \in I$, for all $x,y \in \mathbb{Z}$. Since $x-y \in I$ means that $m \,|\, (x-y)$, we see that

$$x \sim_I y \text{ if and only if } x \equiv y \pmod{m} \tag{8.28}$$

for all $x,y \in \mathbb{Z}$. Thus, the relation $\sim_I$ is the congruence relation $(\text{mod } m)$ and hence, $\mathbb{Z}/I = \mathbb{Z}_m$. Moreover, it follows that the definitions of addition and multiplication for the ring $\mathbb{Z}/I$ are exactly the same as that for $(\text{mod } m)$ modular arithmetic. Therefore, $(\mathbb{Z}/I,\oplus,\odot) = (\mathbb{Z}_m,\oplus,\odot)$.

Example 7. Let $(C(\mathbb{R}),+,\cdot)$ be the ring of all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ given in Example 5 on page 274. Let a be a fixed real number. We know by Example 13 on page 278 that $I = \{f \in C(\mathbb{R}) : f(a) = 0\}$ is an ideal of $C(\mathbb{R})$. We will construct and evaluate the quotient ring $(C(\mathbb{R})/I,\oplus,\odot)$. The relation $\sim_I$ is defined by

$$f \sim_I g \text{ if and only if } f-g \in I$$

for all $f,g \in C(\mathbb{R})$. Since $f-g \in I$ means that $(f-g)(a) = 0$, we see that

$$f \sim_I g \text{ if and only if } f(a) = g(a)$$

for all $f,g \in C(\mathbb{R})$. For each $v \in \mathbb{R}$ let $\overline{v} \in C(\mathbb{R})$ be the constant function $\overline{v}\colon \mathbb{R} \to \mathbb{R}$ defined by $\overline{v}(x) = v$ for all $x \in \mathbb{R}$. For every $f \in C(\mathbb{R})$ if $f(a) = v$, then $f \sim_I \overline{v}$

and thus, $[f] = [\overline{v}]$. Therefore, $C(\mathbb{R})/I = \{[\overline{v}] : v \in \mathbb{R}\}$. Because $\overline{v} + \overline{w} = \overline{v+w}$ and $\overline{v} \cdot \overline{w} = \overline{v \cdot w}$, we see that

$$[\overline{v}] \oplus [\overline{w}] = [\overline{v} + \overline{w}] = [\overline{v+w}] \text{ and } [\overline{v}] \odot [\overline{w}] = [\overline{v} \cdot \overline{w}] = [\overline{v \cdot w}]$$

for all $v, w \in \mathbb{R}$. We observe that the operations of addition and multiplication in $(C(\mathbb{R})/I, \oplus, \odot)$ act very much like[5] those in the ring $(\mathbb{R}, +, \cdot)$ of real numbers.

Exercises 8.6

1. Consider the algebraic structure $(\mathbb{R}, +, \cdot)$ where $\mathbb{R}$ is the set of real numbers, $+$ is addition, and $\cdot$ is multiplication. Let $\sim$ be the equivalence relation on the set of real numbers defined by $x \sim y$ if and only if $|x| = |y|$.

(a) Prove that $\cdot$ preserves the relation $\sim$.

(b) Show that $+$ does not preserve the relation $\sim$.

2. Let $A = \mathbb{Z} \times \mathbb{N} = \{(x, y) : x \in \mathbb{Z} \text{ and } y \in \mathbb{N}\}$. Define the equivalence relation $\sim$ on A by

$$(x, y) \sim (s, t) \text{ iff } xt = ys$$

and define the binary operation $*$ on A by $(x, y) * (s, t) = (xs, yt)$. Prove that $*$ preserves the relation $\sim$.

3. Let $A = \mathbb{Z} \times \mathbb{N} = \{(x, y) : x \in \mathbb{Z} \text{ and } y \in \mathbb{N}\}$. The relation $\sim$ on A define by

$$(x, y) \sim (s, t) \text{ iff } xt = ys$$

is an equivalence relation. Consider the binary operation $+$ defined on A by $(x, y) + (s, t) = (xt + ys, yt)$. Prove that $+$ preserves the relation $\sim$.

4. Let $\mathcal{F} = (F(\mathbb{R}), +, \cdot)$ be the algebraic structure defined in Example 5 on page 245. Let $\sim$ be the equivalence relation on the set $F(\mathbb{R})$ defined by $f \sim g$ if and only if $f(1) = g(1)$. Prove that $\sim$ is a congruence relation on $\mathcal{F}$.

5. Prove Theorem 8.6.6.

6. Let $(G, *)$ be a group and let N be a subgroup of G. Define the equivalence relation $\sim$ on G by $a \sim b$ if and only if $ab^{-1} \in N$, for all $a, b \in G$. Prove that

(a) $[e] = N$, and

(b) For all $a \in N$ we have that $[a] = [e] = N$.

7. Let $(G, *)$ be a group and let $\sim$ be an equivalence relation on G. Suppose that $*$ preserves the relation $\sim$. Let $a, b, c \in G$. Prove the following:

(a) If $a \sim b$, then $a * c \sim b * c$.

[5] Actually, the two rings are *isomorphic* (see page 141 of [10]).

(b) If $a \sim b$, then $c * a \sim c * b$.
(c) If $a \sim b$, then $a^{-1} \sim b^{-1}$.

8. Let $(\mathbb{R}^*, \cdot)$ be the group of nonzero real numbers under multiplication. Consider the normal subgroup $\mathbb{R}^+$ of $\mathbb{R}^*$, where $\mathbb{R}^+$ is the set of positive real numbers. Construct and evaluate the quotient group $(\mathbb{R}^*/\mathbb{R}^+, \odot)$.

9. Consider the group $(G, \circ)$ where $G = \{T_{a,b} : a, b \in \mathbb{R} \text{ and } a \neq 0\}$ and $\circ$ is functional composition (see Example 5 on page 249). One can verify that $N = \{T_{a,b} : a = \pm 1 \text{ and } b \in \mathbb{R}\}$ is a normal subgroup of G. Construct and evaluate the quotient group $(G/N, \circledcirc)$.

10. Let $(G, *)$ be a group with normal subgroup N. Suppose that $N \subseteq K$ where K is also subgroup of G. First show that N is a normal subgroup of K. Now show that $K/N \subseteq G/N$ and that $(K/N, \circledast)$ is a subgroup of $(G/N, \circledast)$.

11. Let $(G, *)$ be a group and let $\sim$ be an equivalence relation on G. Assume that $*$ preserves the relation $\sim$. Let $H = \{a \in G : a \sim e\}$ where $e \in G$ is the identity element. Using the results of Exercise 7 prove the following:

(a) H is a subgroup of G.
(b) H is a normal subgroup of G.
(c) For all $x, y \in G$, we have $x \sim_H y$ if and only if $x \sim y$.

12. Let $(G, *)$ be a group and let N be a *normal subgroup* of G. Thus, $(G/N, \circledast)$ is a group by Theorem 8.6.11. Suppose that $\overline{M}$ is a subgroup of G/N. Define $M \subseteq G$ by $M = \{a \in G : [a] \in \overline{M}\}$.

(a) Prove that M is a subgroup of G.
(b) Prove that $N \subseteq M$ (see Exercise 6).
(c) Prove that if $\overline{M} \lhd G/N$, then $M \lhd G$.

13. Prove Theorem 8.6.12.

14. Let $\sim$ be a congruence relation on a ring $(R, +, \cdot)$. Let $a, b, c \in R$. Prove the following:

(a) If $a \sim b$, then $a \cdot c \sim b \cdot c$.
(b) If $a \sim b$, then $c \cdot a \sim c \cdot b$.

15. Let $\sim$ be a congruence relation on a ring $(R, +, \cdot)$. Let 0 be the zero element of R and define $I = \{a \in R : a \sim 0\}$. Using the results of Exercise 14 prove that I is an ideal of R.

16. Let I be an ideal of the ring $(R, +, \cdot)$ and let $a \in R$. Then $I + a = \{i + a : i \in I\}$ is called a *right coset* of the ideal I. Now, let $\sim_I$ be as defined in Definition 8.6.13. Prove that $[a] = I + a$, where $[a] = \{x \in R : x \sim_I a\}$.

17. Let $(R, +, \cdot)$ be a ring and let I be an ideal of R. Prove that if R is a commutative ring, then R/I is also commutative.

18. Let $(\mathbb{Z}[\sqrt{3}], +, \cdot)$ be the ring introduced in Example 9 on page 277. Consider the ideal of $\mathbb{Z}[\sqrt{3}]$ given by $I = \{a + b\sqrt{3} \in \mathbb{Z}[\sqrt{3}] : 3 \,|\, a\}$. Using the ideal I, construct and evaluate the quotient ring $(\mathbb{Z}[\sqrt{3}]/I, \oplus, \odot)$. Conclude that the operations of addition and multiplication in the quotient ring act very much like those of the ring $(\mathbb{Z}_3, \oplus, \odot)$.

19. Let $(S_2(\mathbb{R}),+,\cdot)$ be the ring presented in Exercise 14 on page 280 where $S_2(\mathbb{R})$ is defined by $S_2(\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a,b \in \mathbb{R} \right\}$. Let J be the ideal of $S_2(\mathbb{R})$ given by $J = \left\{ \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} : b \in \mathbb{R} \right\}$. Construct and evaluate the quotient ring $(S_2(\mathbb{R})/J,\oplus,\odot)$. Conclude that the operations of addition and multiplication in the quotient ring act very much like the ring $(\mathbb{R},+,\cdot)$ of real numbers.

20. The ring $(\mathbb{Z}[\sqrt{2}],+,\cdot)$ is defined in Exercise 8 on page 279. Let $I \subseteq \mathbb{Z}[\sqrt{2}]$ be the ideal of $\mathbb{Z}[\sqrt{2}]$ defined by $I = \{a+b\sqrt{2} \in \mathbb{Z}[\sqrt{2}] : 3 \,|\, a \text{ and } 3 \,|\, b\}$. Construct and evaluate the quotient ring $(\mathbb{Z}[\sqrt{2}]/I,\oplus,\odot)$.

21. Let $(F(\mathbb{Z}),+,\cdot)$ be the ring in Example 7 on page 275. Consider the ideal $I = \{f \in F(\mathbb{Z}) : 5 \,|\, f(1)\}$ of $F(\mathbb{Z})$. Construct and evaluate the quotient ring $(F(\mathbb{Z})/I,\oplus,\odot)$. Conclude that the operations of addition and multiplication in the quotient ring act very much like the ring $(\mathbb{Z}_5,\oplus,\odot)$.

Exercise Notes: For Exercise 5, review Lemma 8.3.13. For Exercise 10, look over Exercise 13 on page 260 and Theorem 8.6.9. For Exercise 16, review the proof of Theorem 8.6.9.

CHAPTER 9

Core Concepts in Real Analysis

Real analysis is a branch of mathematics that studies the set $\mathbb{R}$ of real numbers and provides a theoretical foundation for the fundamental principles of the calculus. The main concepts studied in a first real analysis course are bounded sets of real numbers, functions, limits, sequences, continuity, differentiation, integration, and sequences of functions. Among the first topics covered in such a course are the field axioms and the definition of an ordered field. The concept of an ordered field forms a basis for the algebraic operations and properties of order that are essential in calculus and in real analysis.

9.1 Fields

We present eight fundamental properties that involve addition and multiplication. These properties are called the *field axioms*. The real number system $(\mathbb{R},+,\cdot)$ is said to be a field because it satisfies the field axioms, and from these axioms one can derive all of the other algebraic properties that hold for the real numbers.

Definition 9.1.1. Let $\mathcal{F}=(F,+,\cdot)$ be an algebraic structure where $+$ and $\cdot$ are two binary operations called addition and multiplication, respectively. Then $\mathcal{F}$ is called a field if the following axioms are satisfied:

A1. $x+y=y+x$ for all $x,y\in F$.
A2. $x+(y+z)=(x+y)+z$ for all $x,y,z\in F$.
A3. There is an element $0\in F$ such that $x+0=x$ for all $x\in F$ (0 is called the **zero element**).
A4. For all $x\in F$ there exists a $y\in F$ such that $x+y=0$ (y is written as $-x$ and is called the **additive inverse** of x).

M1. $x\cdot y=y\cdot x$ for all $x,y\in F$.
M2. $x\cdot(y\cdot z)=(x\cdot y)\cdot z$ for all $x,y,z\in F$.
M3. There is an element $1\in F$ such that $1\neq 0$ and $x\cdot 1=x$ for all $x\in F$ (1 is called the **multiplicative identity element**).
M4. For all $x\in F$ if $x\neq 0$, then there exists a $y\in F$ such that $x\cdot y=1$ (y is written as x^{-1}, or $\frac{1}{x}$, and it is called the **multiplicative inverse** of x).

D1. $x\cdot(y+z)=(x\cdot y)+(x\cdot z)$ for all $x,y,z\in F$.

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7_9,

Let $\mathcal{F} = (F, +, \cdot)$ be a field and let $x \in F$. Axiom A4 states that $x + (-x) = 0$. From axiom A1 we conclude that $(-x) + x = 0$. Similarly, if $x \neq 0$, then axiom M4 yields the equation $x \cdot x^{-1} = 1$ and from axiom M1 we have that $x^{-1} \cdot x = 1$.

In our next result, items (2) and (3) respectively show that the zero element in a field is unique and that each element in the field has a unique additive inverse.

Proposition 9.1.1. *Let $(F, +, \cdot)$ be a field. Then for all $x, y, z \in F$ the following hold:*

(1) *If $x + y = x + z$, then $y = z$.*
(2) *If $x + y = x$, then $y = 0$.*
(3) *If $x + y = 0$, then $y = -x$.*
(4) $-(-x) = x$.

Proof. Let $x, y, z \in F$. To prove (1) assume $x + y = x + z$. We prove that $y = z$ as follows:

$$\begin{aligned}
x + y &= x + z && \text{by assumption} \\
y + x &= z + x && \text{by axiom A1} \\
(y + x) + (-x) &= (z + x) + (-x) && \text{add } (-x) \text{ to both sides} \\
y + (x + (-x)) &= z + (x + (-x)) && \text{by axiom A2} \\
y + 0 &= z + 0 && \text{by axiom A4} \\
y &= z && \text{by axiom A3.}
\end{aligned}$$

To prove (2) assume $x + y = x$. By axiom A3 we have that $x + 0 = x$. So, we see that $x + y = x + 0$. Item (1) implies that $y = 0$. We now prove (3). Suppose $x + y = 0$. Because $x + (-x) = 0$ by axiom A4, we have that $x + y = x + (-x)$. Item (1) yields $y = -x$. Finally, to prove (4) note that $(-x) + x = 0$ by axioms A4 and A1. Item (3) allows us to infer that $x = -(-x)$. □

In a similar manner one can prove our next proposition which asserts that the multiplicative identity element in a field is unique and each nonzero element has a unique multiplicative inverse.

Proposition 9.1.2. *Let $(F, +, \cdot)$ be a field. For all $x, y, z \in F$ the following statements hold:*

(1) *If $x \neq 0$ and $x \cdot y = x \cdot z$, then $y = z$.*
(2) *If $x \neq 0$ and $x \cdot y = x$, then $y = 1$.*
(3) *If $x \neq 0$ and $x \cdot y = 1$, then $y = x^{-1}$, that is, $y = \frac{1}{x}$.*
(4) $(x^{-1})^{-1} = x$, *that is,* $\frac{1}{\frac{1}{x}} = x$.

Proposition 9.1.3. *Let $(F, +, \cdot)$ be a field. Then for all $x, y \in F$ we have the next four properties:*

(1) $x \cdot 0 = 0$.
(2) *If $x \neq 0$ and $y \neq 0$, then $x \cdot y \neq 0$.*

(3) $(-x)\cdot y = -(x\cdot y) = x\cdot(-y)$.
(4) $(-x)\cdot(-y) = x\cdot y$.

Proof. Let $x, y \in F$. To prove (1) observe that

$$\begin{aligned} x\cdot 0 &= x\cdot(0+0) && \text{because } 0+0=0 \text{ by axiom A3} \\ &= x\cdot 0 + x\cdot 0 && \text{by axiom D1.} \end{aligned}$$

Thus, $x\cdot 0 + x\cdot 0 = x\cdot 0$. Proposition 9.1.2(2) implies that $x\cdot 0 = 0$. To prove (2), assume that $x \neq 0$ and $y \neq 0$. Suppose, for a contradiction, that $x\cdot y = 0$. Since $x\cdot 0 = 0$ by (1), we conclude that $x\cdot y = x\cdot 0$. As $x \neq 0$, item (1) of Proposition 9.1.2 implies that $y = 0$ which contradicts our assumption that $y \neq 0$. Therefore, $x\cdot y \neq 0$. To establish (3), we see that

$$x\cdot y + x\cdot(-y) = x\cdot(y+(-y)) = x\cdot 0 = 0$$

by axiom D1, axiom A4, and item (1). Hence, $x\cdot y + x\cdot(-y) = 0$. Proposition 9.1.1(2) implies that $x\cdot(-y) = -(x\cdot y)$. Similarly, one can show that $(-x)\cdot y = -(x\cdot y)$. Finally,

$$(-x)\cdot(-y) = -(x\cdot(-y)) = -(-(x\cdot y)) = x\cdot y$$

by item (3) and Proposition 9.1.1(4). □

When x and y are elements in a field, one can write $x\cdot y$ as xy, denote $x+(-y)$ by $x-y$, and write $x\cdot x$ as x^2. One can also express x^{-1} as $\frac{1}{x}$ and write $\frac{y}{x}$ for $y\cdot\frac{1}{x}$.

As noted at the beginning of this section, the algebraic system $(\mathbb{R}, +, \cdot)$ is a field. In addition, the system $(\mathbb{Q}, +, \cdot)$ is also a field. Thus, these systems satisfy all of the properties presented in Propositions 9.1.1–9.1.3.

9.1.1 Ordered Fields

In Section 7.5 we said that a partial order $\leq$ on a set F is a relation that is reflexive, antisymmetric, and transitive. Furthermore, if for every $x, y \in F$ we have either $x \leq y$ or $y \leq x$, then $\leq$ is a total order. We also defined the strict order $<$ on F by $x < y$ if and only if $x \leq y$ and $x \neq y$, for all $x, y \in F$ (see Definitions 7.5.2–7.5.4). We shall write $x > y$ to mean $y < x$ whenever $x, y \in F$.

The familiar relation $\leq$ on the set of real numbers is a total order. Moreover, this relation satisfies many additional properties that involve addition and multiplication; two of which are presented in our next definition.

Definition 9.1.4. A structure $(F,+,\cdot,\leq)$ is called an **ordered field** if $(F,+,\cdot)$ is a field and $\leq$ is a total order on F such that for all $x,y,z \in F$ we have the following:

1. If $x < y$, then $x+z < y+z$.
2. If $x > 0$ and $y > 0$, then $x \cdot y > 0$.

When $x > 0$ we shall say that x is *positive* and if $x < 0$, then x is *negative*.

The structures $(\mathbb{Q},+,\cdot,\leq)$ and $(\mathbb{R},+,\cdot,\leq)$ are ordered fields where $\leq$ is the usual 'less than or equal' relation on the rational and real numbers. In fact, using the axioms of set theory, one can actually construct these two number systems and prove that they are ordered fields. The standard properties of inequality that hold in these two fields also hold in every ordered field. In particular, when F is an ordered field and $x,y,z \in F$, Lemma 7.5.5 implies the following:

1. If $x < y$ and $y < z$, then $x < z$. (transitivity)
2. Exactly one of the following holds: $x < y$, $y < x$, or $x = y$. (trichotomy)

Furthermore, in an ordered field one can prove that the result of multiplying both sides of an inequality by a positive value preserves the inequality and the result of multiplying both sides by a negative value reverses the inequality.

Proposition 9.1.5. *Let $(F,+,\cdot,\leq)$ be an ordered field. For all $x,y,z \in F$ the following properties of inequality hold:*

(1) *If $x > 0$, then $-x < 0$. If $x < 0$, then $-x > 0$.*
(2) *If $x > 0$ and $y < z$, then $xy < xz$.*
(3) *If $x < 0$ and $y < z$, then $xy > xz$.*
(4) *If $x \neq 0$, then $x^2 > 0$. In particular, $1 > 0$.*
(5) *If $x > 0$, then $\frac{1}{x} > 0$.*
(6) *If $0 < x < y$, then $0 < \frac{1}{y} < \frac{1}{x}$.*

Proof. We will prove (1), (2), (4), (5) and leave items (3) and (6) as exercises. To prove (1), assume that $x > 0$. From Definition 9.1.4(1) we obtain

$$x + (-x) > 0 + (-x).$$

Thus, by axioms A4 and A3, we have $0 > -x$; that is, $-x < 0$. Now assume that $x < 0$. Then $x + (-x) < 0 + (-x)$. Hence, $0 < -x$ and so, $-x > 0$.

To prove item (2), assume $x > 0$ and $y < z$. Since $y < z$, we conclude that

$$y + (-y) < z + (-y)$$

by Definition 9.1.4(1). Thus, $0 < z - y$. Because $x > 0$ and $z - y > 0$, we conclude that $x(z-y) > 0$ from Definition 9.1.4(2). Axiom D1 and Proposition 9.1.3(3) imply that $x(z-y) = xz - xy$ and hence, $0 < xz - xy$. Therefore, $0 + xy < (xz - xy) + xy$ by (1) of Definition 9.1.4. Thus, $xy < (xz - xy) + xy$. From axioms A1, A2, and A4 we have that $(xz - xy) + xy = xz$. Therefore, $xy < xz$.

To establish (4), assume $x \neq 0$. So, by trichotomy, we must have either $x > 0$ or $x < 0$. If $x > 0$, then $xx > 0$ by (2) of Definition 9.1.4. Thus, $x^2 > 0$. If $x < 0$, then $-x > 0$ by item (1) of the proposition. Again by Definition 9.1.4(2), we obtain $(-x)(-x) > 0$. Proposition 9.1.3(4) implies that $(-x)(-x) = xx$. Hence, $x^2 > 0$. Since $1^2 = 1$, we infer that $1 > 0$.

Finally, to prove item (5), assume that $x > 0$. Suppose, for a contradiction, that $\frac{1}{x} \leq 0$. If $\frac{1}{x} = 0$, then $x \cdot \frac{1}{x} = x \cdot 0 = 0$. Since $x \cdot \frac{1}{x} = 1$, we conclude that $1 = 0$ which contradicts axiom M3. If $\frac{1}{x} < 0$, then $x \cdot \frac{1}{x} < x \cdot 0 = 0$, as $x > 0$. Thus, $1 < 0$. This is impossible, because $1 > 0$ by item (4) of the proposition. Therefore, $\frac{1}{x} > 0$. □

Exercises 9.1

1. Let $(F, +, \cdot)$ be a field with $x, y \in F$. Prove the following:
   (a) $-x - y = -(x + y)$.
   (b) $-(y - x) = x - y$.

2. Let $(F, +, \cdot, \leq)$ be an ordered field with $x, y \in F$. Prove the following:
   (a) $1 < 1 + 1$ and $\frac{1}{1+1} < 1$.
   (b) $x < y$ if and only if $-y < -x$.
   (c) $x \cdot y = 0$ if and only if $x = 0$ or $y = 0$.

3. Let $(F, +, \cdot, \leq)$ be an ordered field where x and y are elements in F. Prove if $x < y$, then $x(y + 1) < y(x + 1)$.

4. Prove items (3) and (6) of Proposition 9.1.5.

5. Let $(F, +, \cdot, \leq)$ be an ordered field where x is an element in F. Prove that if $x < 0$, then $\frac{1}{x} < 0$.

6. Let $(F, +, \cdot, \leq)$ be an ordered field with $x, y \in F$. Prove that if $x < 0$ and $y < 0$, the $xy > 0$.

7. Let $(F, +, \cdot, \leq)$ be an ordered field where a, b, c, d are elements in F. Prove that if $a < b$ and $c < d$, then $a + c < b + d$.

8. Let $(F, +, \cdot, \leq)$ be an ordered field where a, b, c, d are positive elements in F. Prove that if $a < b$ and $c < d$, then $ac < bd$. Conclude that $a^2 < b^2$.

9. Let $(F, +, \cdot, \leq)$ be an ordered field where a, b, c, d are positive elements in F. Prove the following:
   (a) If $a < c$, then $\frac{a}{b} < \frac{c}{b}$.
   (b) If $d < b$, then $\frac{a}{b} < \frac{a}{d}$.
   (c) If $a \leq c$ and $d \leq b$, then $\frac{a}{b} \leq \frac{c}{d}$.

10. Let $(F, +, \cdot, \leq)$ be an ordered field where x and y are positive elements in F. Prove that if $x^2 < y^2$, then $x < y$.

11. Let $(F, +, \cdot)$ be a field and let $x, y \in F$ where $x \neq -y$. Prove that $x^2 + y^2 > xy$ and $x^2 + y^2 > -xy$.

9.2 The Real Field

For the remainder of this chapter, we will be investigating the ordered field of real numbers $(\mathbb{R},+,\cdot,\leq)$ which is sometimes referred to as the *real field*. Since the real field is an ordered field, we can apply all of the propositions and definitions that we presented in the previous section. In particular, using the properties of an ordered field we can prove our next theorem.

Theorem 9.2.1. *Let $x,y\in\mathbb{R}$. Suppose $x\leq y+\varepsilon$ for all $\varepsilon>0$. Then $x\leq y$.*

Proof. Let $x,y\in\mathbb{R}$ and assume that $x\leq y+\varepsilon$ for all $\varepsilon>0$. We will prove that $x\leq y$. Suppose, for a contradiction, that $x>y$. So $x-y>0$ and $\varepsilon=\frac{x-y}{2}>0$. Thus, $x\leq y+\varepsilon$ by our assumption. Since $\varepsilon=\frac{x-y}{2}<x-y$, we obtain

$$x\leq y+\varepsilon=y+\left(\frac{x-y}{2}\right)<y+(x-y)=x$$

which implies that $x<x$. This contradiction forces us to conclude that $x\leq y$. □

The *absolute value function* is a very important function in real analysis because it provides us with a notion of distance between two real numbers. Many of the proofs in real analysis establish inequalities that involve the absolute value of a real number or the absolute value of the difference of two real numbers.

Definition 9.2.2 (Absolute Value). Given a real number x, the **absolute value** of x, denoted by $|x|$, is defined by

$$|x|=\begin{cases}x, & \text{if } x\geq 0;\\ -x, & \text{if } x<0.\end{cases}$$

Theorem 9.2.3 (Basic Properties of Absolute Value). *For all $a,x\in\mathbb{R}$, with $a\geq 0$, the following hold:*

(a) $0\leq|x|$ *and* $|-x|=|x|$.
(b) $x\leq|x|$ *and* $-x\leq|x|$.
(c) *If* $|x|=0$, *then* $x=0$.
(d) $|x|\leq a$ *if and only if* $-a\leq x\leq a$.
(e) $|xy|=|x|\,|y|$.
(f) $\left|\frac{x}{y}\right|=\frac{|x|}{|y|}$ *when* $y\neq 0$.
(g) $|x|^2=x^2$.
(h) $|x+y|\leq|x|+|y|$. (triangle inequality)

Proof. Items (a)–(g) follow directly from Definition 9.2.2. We shall prove (h). Since $|x+y|^2=(x+y)^2$, we have that

$$|x+y|^2=x^2+2xy+y^2\leq x^2+|2xy|+y^2=|x|^2+2\,|x|\,|y|+|y|^2=(|x|+|y|)^2.$$

So, $|x+y|^2 \leq (|x|+|y|)^2$. Exercise 10 (page 297) implies that $|x+y| \leq |x|+|y|$. □

The triangle inequality is used frequently in real analysis. For example, suppose $|x-\ell| < 1$. Using the triangle inequality we can show that $|x| < |\ell|+1$ as follows:

$$|x| = |x-\ell+\ell| \leq |x-\ell|+|\ell| < 1+|\ell|\,.$$

Thus, $|x| < |\ell|+1$. In the proof of our next theorem, we will be using the triangle inequality to derive other useful properties of the absolute value function.

Theorem 9.2.4 (More Properties of Absolute Value). *For all $x, y, k \in \mathbb{R}$, where $k > 0$, we have*

(1) $|x| < k$ *if and only if* $-k < x < k$.
(2) $|x| > k$ *if and only if* $x < -k$ *or* $x > k$.
(3) $|x| - |y| \leq |x-y|$.
(4) $|y| - |x| \leq |x-y|$.
(5) $\big|\,|x| - |y|\,\big| \leq |x-y|$. (backward triangle inequality)[1]

Proof. Items (1)–(2) follow from Definition 9.2.2. We shall prove (3), (4), and (5). To prove (3), observe that $|x| = |x-y+y|$. So the triangle inequality implies that

$$|x| = |x-y+y| \leq |x-y|+|y|\,.$$

Hence $|x| \leq |x-y|+|y|$ and thus, $|x|-|y| \leq |x-y|$.

Now we prove (4). By interchanging x and y in (3), we obtain $|y|-|x| \leq |y-x|$. Since $|y-x| = |x-y|$, we also have that $|y|-|x| \leq |x-y|$.

Finally, we prove (5). From (4) we conclude that $-|x-y| \leq -(|y|-|x|)$ and thus,

$$-|x-y| \leq |x|-|y|\,.$$

From (3) we see that $|x|-|y| \leq |x-y|$. Hence, $-|x-y| \leq |x|-|y| \leq |x-y|$. Theorem 9.2.3(d) implies that $\big|\,|x|-|y|\,\big| \leq |x-y|$. Our proof is complete. □

Given a finite nonempty set of real numbers A, we let $\max A$, or $\max(A)$, denote the maximum number in A. We also define $\min A$, or $\min(A)$, to be the minimum number in A. For example, $\max\{-1,2,\pi,3\} = \pi$ and $\min\{-1,2,\pi,3\} = -1$.

Lemma 9.2.5. *Let x, a, and b be real numbers. If $a \leq x \leq b$, then $|x| \leq \max\{|a|,|b|\}$.*

Proof. Assume $a \leq x \leq b$. We shall prove that $|x| \leq \max\{|a|,|b|\}$. First suppose that $x \geq 0$. Because $x \leq b$, we see that $b \geq 0$. So $|x| = x$ and $|b| = b$. Since $x \leq b$, we deduce that $|x| \leq |b|$ and therefore, $|x| \leq \max\{|a|,|b|\}$. Now suppose that $x < 0$. Then, because $a \leq x$, we have that $a < 0$. Thus, $|x| = -x$ and $|a| = -a$. Since $a \leq x$, we see that $-x \leq -a$ and hence, $|x| \leq |a|$. Therefore, $|x| \leq \max\{|a|,|b|\}$. □

[1] This inequality is also called the *reverse triangle inequality*.

Exercises 9.2

1. Prove (4), (5), (6), and (7) of Theorem 9.2.3.
2. Prove (1) and (2) of Theorem 9.2.4.
3. Let x, a, and b be real numbers. Prove that if $a \leq x \leq b$, then $\min\{|a|, |b|\} \leq |x|$.
4. Let a and b be real numbers. Using Theorem 9.2.4, prove that $|a+b| \geq |a| - |b|$.
5. Prove by mathematical induction that for all $n \geq 1$ if $x_1, x_2, \ldots, x_n$ are real numbers, then $|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|$.
6. Let x, y be real numbers. Prove that $x^2 + y^2 = 0$ if and only if $x = 0$ and $y = 0$.
7. Let x and y be real numbers. Prove that $|x+y| = |x| + |y|$ if and only if $xy \geq 0$.
8. Let x and y be real numbers. Prove that $\max\{x, y\} = \frac{x+y+|x-y|}{2}$. Conclude that $|x| = 2\max\{x, 0\} - x$.
9. Let x and y be real numbers. Prove that $\min\{x, y\} = \frac{x+y-|x-y|}{2}$. Conclude that $|x| = x - 2\min\{x, 0\}$.

9.3 The Completeness Axiom

Since the real field is ordered, we often use the real line to give a geometric picture of this field where every real number appears as a point on the line. The field of rational numbers is another ordered field which we shall call the *rational field*. Since the rational field is also ordered, we can imagine a rational number line that consists of just the rational numbers. Because $\sqrt{2}$ is not a rational number, it does not appear on the rational number line; however, it does appear on the real line. So, in a sense, the real line fills in the holes that exist in the rational line. Does the real line possess any holes? The real line is without holes, because the real field satisfies a key property called *completeness*. We shall express this completeness property in terms of a formal mathematical axiom. To state this axiom, we need some preliminary definitions.

Definition 9.3.1 (Upper and Lower Bounds). Let $S \subseteq \mathbb{R}$ be nonempty.

- The set S is **bounded above** if there is a real number b such that $x \leq b$ for all $x \in S$. The number b is called an **upper bound** for S.
- The set S is **bounded below** if there is a real number a such that $a \leq x$ for all $x \in S$. The number a is called a **lower bound** for S.
- If S has both a lower bound and an upper bound, then we say that S is **bounded**.

Theorem 9.3.2. *Let $S \subseteq \mathbb{R}$ be nonempty. Then S is bounded if and only if there is an $M > 0$ so that $|x| \leq M$ for all $x \in S$.*

Proof. Let $S \subseteq \mathbb{R}$ be nonempty. Assume that S is bounded. So there are nonzero real numbers a and b such that $a \leq x \leq b$ for all $x \in S$. Let $M = \max\{|a|, |b|\} > 0$.

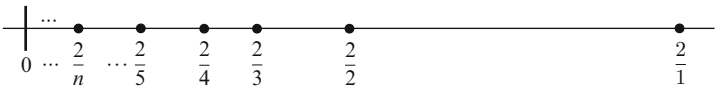


Fig. 9.1 The set $\{\frac{2}{n} : n \in \mathbb{N}\}$ is bounded below by 0 and bounded above by 3

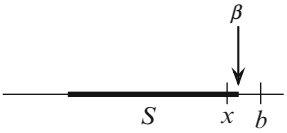


Fig. 9.2a $\beta = \sup(S)$

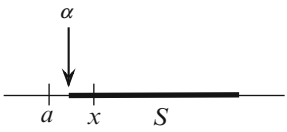


Fig. 9.2b $\alpha = \inf(S)$

Lemma 9.2.5 implies that $|x| \leq M$ for all $x \in S$. For the converse, assume $|x| \leq M$ for all $x \in S$ where $M > 0$. From Theorem 9.2.3(b), we conclude that $x \leq |x| \leq M$ and $-x \leq |x| \leq M$. Thus, $-M \leq x \leq M$ for all $x \in S$ and so, S is bounded. □

Definition 9.3.3 (Supremum and Infimum). Let $S \subseteq \mathbb{R}$ be nonempty.

- Suppose that β is an upper bound for S. If β is the **least upper bound** for S, then β is called the **supremum** of S and we write $\beta = \sup(S)$.
- Suppose that α is a lower bound for S. If α is the **greatest lower bound** for S, then α is called the **infimum** of S and we write $\alpha = \inf(S)$.

The set $S = \{\frac{2}{n} : n \in \mathbb{N}\}$ (see Fig. 9.1) is bounded. We note that 0 is the greatest lower bound for S and that 2 is the least upper bound for S. Thus, $0 = \inf(S)$ and $2 = \sup(S)$.

Let $S \subseteq \mathbb{R}$ be nonempty. The equation $\beta = \sup(S)$ means that (i) β is an upper bound for S and (ii) β is the smallest upper bound for S. On the other hand, the equation $\alpha = \inf(S)$ means that (i) α is a lower bound for S and (ii) α is the largest lower bound for S. The next remark repeats and clarifies these observations.

Remark 9.3.4. Let $S \subseteq \mathbb{R}$ be nonempty and let $\alpha, \beta \in \mathbb{R}$ (see Fig. 9.2a, b). Then we have that

1. $\beta = \sup(S)$ if and only if the following two conditions hold:
   (i) $x \leq \beta$ for all $x \in S$,
   (ii) for all real numbers b, if b is an upper bound for S, then $\beta \leq b$.
2. $\alpha = \inf(S)$ if and only if the following two conditions hold:
   (i) $\alpha \leq x$ for all $x \in S$,
   (ii) for all real numbers a, if a is a lower bound for S, then $a \leq \alpha$.

The properties of an ordered field that we investigated in Section 9.1 identify many of the important properties possessed by the real numbers. In real analysis, however, almost every significant result relies on our next axiom which proclaims that every nonempty set of real numbers that is bounded above has a supremum.

Completeness Axiom. Every nonempty $S \subseteq \mathbb{R}$ that is bounded above has a least upper bound.

The completeness axiom just asserts that if a set $S \subseteq \mathbb{R}$ is nonempty and bounded above, then there is a real number β that satisfies the equation $\beta = \sup(S)$. Even though it is stated as an axiom, Richard Dedekind in 1858 discovered a proof of the completeness axiom using concepts from set theory (see [11, Theorem 1.19]). Moreover, one can prove that the ordered field of rational numbers does not satisfy the completeness axiom; that is, one can show that there are bounded subsets of $\mathbb{Q}$ which do not have a least upper bound in $\mathbb{Q}$. Thus, the real field is a very special ordered field.

9.3.1 Proofs on the Supremum of a Set

Remark 9.3.4(1) inspires the following very useful strategy for proving equations of the form $\beta = \sup(S)$.

Proof Strategy 9.3.5. Given a real number β and a nonempty $S \subseteq \mathbb{R}$, to prove that $\beta = \sup(S)$ use the two-step proof diagram:

Step (1): Prove $x \leq \beta$ for all $x \in S$.
Step (2): Assume b is an upper bound for S.
Prove $\beta \leq b$.

In other words, to prove that $\beta = \sup(S)$ you must first prove that β is an upper bound for S and then prove that β is the smallest upper bound for S. Similarly, Remark 9.3.4(1) yields a strategy that allows one to take advantage of an assumption having the form $\beta = \sup(S)$.

Assumption Strategy 9.3.6. Let $S \subseteq \mathbb{R}$ be nonempty and let $\beta \in \mathbb{R}$. Suppose you are *assuming* that $\beta = \sup(S)$. Then you can infer (1) $x \leq \beta$ for all $x \in S$, and (2) whenever b is an upper bound for S, you can deduce that $\beta \leq b$.

Our proof of the next theorem employs both Proof Strategy 9.3.5 and Assumption Strategy 9.3.6. Given a set $S \subseteq \mathbb{R}$ and a real number k we can form the new set of real numbers defined by $A = \{kx : x \in S\}$. The set A is sometimes denoted by kS. We will now show that there is a connection between the supremum of S (if it exists) and the supremum of A.

Theorem 9.3.7. *Let $S \subseteq \mathbb{R}$ be nonempty and bounded above, and let $k > 0$. Suppose $A = \{kx : x \in S\}$. Then the set A is bounded above and* $\sup(A) = k \sup(S)$.

Proof Analysis. Since the set S is bounded above, there is a real number β satisfying $\beta = \sup(S)$ by the completeness axiom. Let $k > 0$ and $A = \{kx : x \in S\}$. We need to prove that $k\beta = \sup(A)$. First note that every element in A has the form kx for an $x \in S$. So to prove that $k\beta$ is an upper bound for A, we just need to show that

$kx \leq k\beta$ for all $x \in S$. Appealing to Proof Strategy 9.3.5, we construct the following proof diagram:

Assume $\beta = \sup(S)$.
  Prove $kx \leq k\beta$ for all $x \in S$.
  Assume c is an upper bound for A.
    Prove $k\beta \leq c$.

The first line of the above proof diagram indicates that, in our proof, we will be assuming $\beta = \sup(S)$. Thus, we can use Assumption Strategy 9.3.6. The second line in this proof diagram asserts that we must prove that $k\beta$ is an upper bound for the set A. The third line states that we will assume c is an upper bound for A. We will then have to prove that $k\beta \leq c$. Since every element in A has the form kx for an $x \in S$, to say that c is an upper bound for A just means that $kx \leq c$ for all $x \in S$. We now have all of the necessary ingredients that we need to compose a correct proof of the theorem. Ⓐ

Proof (of Theorem 9.3.7). Suppose that $S \subseteq \mathbb{R}$ is nonempty and bounded above. The completeness axiom asserts that S has a least upper bound β and so, $\beta = \sup(S)$. Hence, $x \leq \beta$ for all $x \in S$. Let $k > 0$ and $A = \{kx : x \in S\}$. We shall prove that $k\beta = \sup(A)$. We first prove that $k\beta$ is an upper bound for A. Since $x \leq \beta$ for all $x \in S$ and $k > 0$, we have $kx \leq k\beta$ for all $x \in S$. Thus, $k\beta$ is an upper bound for A.

Let c be an upper bound for A. We shall prove that $k\beta \leq c$. Since c is an upper bound for A, we have that $kx \leq c$ for all $x \in S$. Because $k > 0$, we see that $x \leq \frac{c}{k}$ for all $x \in S$. Thus, $\frac{c}{k}$ is an upper bound for S. Since β is the smallest upper bound for S, we conclude that $\beta \leq \frac{c}{k}$ and so $k\beta \leq c$. Therefore, $k\beta = \sup(A)$. □

9.3.2 *Proofs on the Infimum of a Set*

Our next proof strategy, motivated by Remark 9.3.4(2), can be used to prove an equation of the form $\alpha = \inf(S)$.

Proof Strategy 9.3.8. Given a real number α and a nonempty $S \subseteq \mathbb{R}$, to prove that $\alpha = \inf(S)$ use the two-step proof diagram:

Step (1): Prove $\alpha \leq x$ for all $x \in S$.
Step (2): Assume a is a lower bound for S.
  Prove $a \leq \alpha$.

The completeness axiom asserts that a nonempty set of real numbers, which is *bounded above*, has a supremum. In the proof of our next theorem, we will show that the completeness axiom implies that if a set S is nonempty and *bounded below*, then S has an infimum; that is, there is a real number α satisfying the equation $\alpha = \inf(S)$. Before reading the proof of Theorem 9.3.9, one should read Remark 9.3.4 and Proof Strategy 9.3.8.

Theorem 9.3.9. *Let $S \subseteq \mathbb{R}$ be nonempty and bound below. Then S has a greatest lower bound.*

Proof. Let $S \subseteq \mathbb{R}$ be a nonempty set with lower bound a. Let $S^* = \{-x : x \in S\}$. So every element in S^* has the form $-x$ for an $x \in S$. Since a is a lower bound for S, we have $a \leq x$ for all $x \in S$. Thus, $-x \leq -a$ for all $x \in S$ and so, $-a$ is an upper bound for S^*. By the completeness axiom, S^* has a least upper bound β. Hence, $\beta = \sup(S^*)$; that is, $-x \leq \beta$ for all $x \in S$ and β is the smallest such upper bound.

We shall prove that $-\beta = \inf(S)$. Since $-x \leq \beta$ for each $x \in S$, it follows that $-\beta \leq x$ for all $x \in S$. Hence, $-\beta$ is a lower bound for S. To prove that $-\beta$ is the largest lower bound for S, let a be a lower bound for S. By the argument in the first paragraph of this proof, we see that $-a$ is an upper bound for the set S^*. Since β is the least such upper bound for S^*, we have that $\beta \leq -a$. So, $a \leq -\beta$. Therefore, $-\beta$ is the greatest lower bound for S, that is, $-\beta = \inf(S)$. □

Assumption Strategy 9.3.10. Let $S \subseteq \mathbb{R}$ be nonempty and let $\alpha \in \mathbb{R}$. Suppose you are *assuming* that $\alpha = \inf(S)$. Then you can conclude (1) $\alpha \leq x$ for all $x \in S$, and (2) whenever a is a lower bound for S, you can deduce that $a \leq \alpha$.

Theorem 9.3.11. *Let $S \subseteq \mathbb{R}$ be nonempty and bounded below, and let $k > 0$. Suppose $A = \{kx : x \in S\}$. Then the set A is bounded below and* $\inf(A) = k\inf(S)$.

Proof Analysis. Since the set S is bounded below, there is a real number α satisfying $\alpha = \inf(S)$ by the completeness axiom (that is, by Theorem 9.3.9). Let $k > 0$ and $A = \{kx : x \in S\}$. We must prove that $k\alpha = \inf(A)$. First note that every element in A has the form kx for $x \in S$. So to prove that $k\alpha$ is a lower bound for A, we just need to show that $k\alpha \leq kx$ for all $x \in S$. Appealing to Proof Strategy 9.3.8, we construct the following proof diagram:

Assume $\alpha = \inf(S)$.
  Prove $k\alpha \leq kx$ for all $x \in S$.
  Assume c is a lower bound for A.
    Prove $c \leq k\alpha$.

The first line of the above proof diagram indicates that, in our proof, we will be assuming $\alpha = \inf(S)$. Thus, we can use Assumption Strategy 9.3.10. The second line asserts that we must prove that $k\alpha$ is a lower bound for the set A. The third line states that we will assume that c is a lower bound for A. We will then have to prove that $c \leq k\alpha$. Since every element in A has the form kx for $x \in S$, to say that c is a lower bound for A just means that $c \leq kx$ for all $x \in S$. We have now identified all of the essential components that we need to prove the theorem. Ⓐ

Proof (of Theorem 9.3.11). Let $S \subseteq \mathbb{R}$ be nonempty and bounded below. Let $k > 0$ and $A = \{kx : x \in S\}$. Theorem 9.3.9 implies that S has a greatest lower bound α and so, $\alpha = \inf(S)$. Thus, $\alpha \leq x$ for all $x \in S$. First we prove that $k\alpha$ is a lower bound for A. Since $\alpha \leq x$ for all $x \in S$ and $k > 0$, we have that $k\alpha \leq kx$ for all $x \in S$. Hence, $k\alpha$ is a lower bound for A.

Let c be a lower bound for A. We will prove that $c \leq k\alpha$. Since c be a lower bound for A, we see that $c \leq kx$ for all $x \in S$. Because $k > 0$, we have that $\frac{c}{k} \leq x$ for all $x \in S$. Thus, $\frac{c}{k}$ is a lower bound for S. Since α is the largest lower bound for S, we conclude that $\frac{c}{k} \leq \alpha$. Therefore, $c \leq k\alpha$ and $k\alpha = \inf(A)$. □

Theorem 9.3.12. *Let $S \subseteq \mathbb{R}$ be nonempty and bounded above, and let $k < 0$. Suppose $A = \{kx : x \in S\}$. Then the set A is bounded above and $\sup(A) = k\inf(S)$.*

Proof. See Exercise 7. □

Theorem 9.3.13. *Let $S \subseteq \mathbb{R}$ be nonempty and bounded above, and let $k < 0$. Suppose $A = \{kx : x \in S\}$. Then the set A is bounded below and $\inf(A) = k\sup(S)$.*

Proof. See Exercise 8. □

Theorem 9.3.14. *Let $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$ be non-empty. Suppose $x \leq y$ for all $x \in A$ and all $y \in B$. Then A is bounded above, B is bounded below, and $\sup(A) \leq \inf(B)$.*

Proof. Suppose that

$$x \leq y \text{ for all } x \in A \text{ and all } y \in B. \tag{9.1}$$

Let $y \in B$ be arbitrary. From (9.1) we have that $x \leq y$ for all $x \in A$. Thus, y is an upper bound for A (see Fig. 9.3). The completeness axiom implies that $\gamma = \sup(A)$ exists. Since y is an upper bound for A and γ is the least such upper bound, it follows that $\gamma \leq y$. As y is an arbitrary element in B, it follows that $\gamma \leq y$ for all $y \in B$. Thus, γ is a lower bound for B. Theorem 9.3.9 implies that $\delta = \inf(B)$ exists. Because γ is a lower bound for B and δ is the greatest such lower bound, we see that $\gamma \leq \delta$. Therefore, $\sup(A) \leq \inf(B)$. □

Definition 9.3.15 (Maximum and Minimum Elements). Let $S \subseteq \mathbb{R}$.

- If $b = \sup(S)$ and $b \in S$, then b is called the **maximum element** of S and we write $b = \max(S)$.
- If $a = \inf(S)$ and $a \in S$, then a is called the **minimum element** of S and we write $a = \min(S)$.

For example, let $S = [2,5]$. Because $5 = \sup(S)$ and $5 \in S$, we see that $5 = \max(S)$. Since $2 = \inf(S)$ and $2 \in S$, we conclude that $2 = \min(S)$. For another example, let $T = \{\frac{1}{n} : n \in \mathbb{N}\}$. Because $1 = \sup(T)$ and $1 \in T$, we have that $1 = \max(T)$. Since $0 = \inf(T)$ and $0 \notin T$, we infer that $\min(T)$ is undefined.

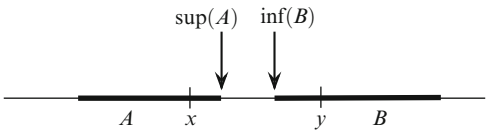


Fig. 9.3 An illustration for the proof of Theorem 9.3.14

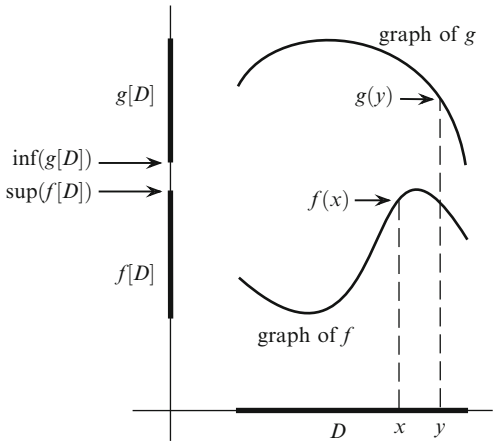


Fig. 9.4 Representation of Theorem 9.3.18

9.3.3 Bounded Functions

Let $f: D \to \mathbb{R}$ be a function and let $S \subseteq D$. Then $f[S] = \{f(x) : x \in S\}$ is the image of S and $\text{ran}(f) = \{f(x) : x \in D\}$ is the range of f. Observe that $\text{ran}(f) = f[D]$. Thus, our next definition states that a function is bounded if the range of the function is a bounded set.

Definition 9.3.16. A function $f: D \to \mathbb{R}$ is **bounded** if the set $f[D]$ is bounded.

Suppose $f: D \to \mathbb{R}$ is bounded. Thus, the set $f[D] = \{f(x) : x \in D\}$ is bounded. Let $\beta = \sup(f[D])$ and $\alpha = \inf(f[D])$. Then $\alpha \leq f(x) \leq \beta$ for all $x \in D$.

Remark 9.3.17. Let $f: D \to \mathbb{R}$. Theorem 9.3.2 implies that there are two equivalent ways of saying that the set $f[D] = \{f(x) : x \in D\}$ is bounded, namely:

(1) There are real numbers a, b such that $a \leq f(x) \leq b$ for all $x \in D$,
(2) There is a real number $M > 0$ such that $|f(x)| \leq M$ for all $x \in D$.

Theorem 9.3.18. *Let $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ be functions. If $f(x) \leq g(y)$ for all $x, y \in D$, then the set $f[D]$ is bounded above, the set $g[D]$ is bounded below, and* $\sup(f[D]) \leq \inf(g[D])$.

Proof. Assume $(\star)$ $f(x) \leq g(y)$ for all $x, y \in D$ (see Fig. 9.4). We will show that $a \leq b$ for all $a \in f[D]$ and all $b \in g[D]$. Let $a \in f[D]$ and $b \in g[D]$. So $a = f(x)$ and $b = g(y)$ for some $x, y \in D$. Hence, $a = f(x) \leq g(y) = b$ by $(\star)$. Thus, $a \leq b$ for all $a \in f[D]$ and all $b \in g[D]$. Theorem 9.3.14 implies that $f[D]$ is bounded above, $g[D]$ is bounded below, and that $\sup(f[D]) \leq \inf(g[D])$. □

Our next theorem shows that the sum of two bounded functions is also bounded.

Theorem 9.3.19. *If $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ are bounded, then $(f+g): D \to \mathbb{R}$ is bounded and*

(a) $\sup((f+g)[D]) \leq \sup(f[D]) + \sup(g[D])$,
(b) $\inf(f[D]) + \inf(g[D]) \leq \inf((f+g)[D])$.

Proof. Assume $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ are bounded. We shall prove only (b) and leave (a) as an exercise. Since $f[D]$ and $g[D]$ are bounded below, let $\varepsilon = \inf(f[D])$ and $\delta = \inf(g[D])$. Consequently,

$$\varepsilon \leq f(x) \text{ for all } x \in D \tag{9.2}$$

$$\delta \leq g(x) \text{ for all } x \in D. \tag{9.3}$$

Inequalities (9.2) and (9.3) imply that $\varepsilon + \delta \leq f(x) + g(x)$ for all $x \in D$ and, since $(f+g)(x) = f(x) + g(x)$, we conclude that $\varepsilon + \delta \leq (f+g)(x)$ for all $x \in D$. Thus, $\varepsilon + \delta$ is a lower bound for $(f+g)[D]$. Hence, $\gamma = \inf((f+g)[D])$ exists. Because γ is the greatest lower bound for $(f+g)[D]$, we obtain $\varepsilon + \delta \leq \gamma$ and therefore, $\inf(f[D]) + \inf(g[D]) \leq \inf((f+g)[D])$. □

9.3.4 *Alternative Proof Strategies*

Our earlier proofs on the supremum and infimum relied on Remark 9.3.4 in which we noted that the equation $\beta = \sup(S)$ means that (i) β is an upper bound for S and (ii) β is the smallest upper bound for S. Given (i), another way to state (ii) is to say that "every number $r < \beta$ is not an upper bound for S." A similar observation holds for the infimum of a set. We record these observations in Remark 9.3.20, below, where items 1 and 2 are illustrated in Fig. 9.5a, b, respectively.

Remark 9.3.20. Let $S \subseteq \mathbb{R}$ be nonempty and let $\alpha, \beta \in \mathbb{R}$. Then

1. $\beta = \sup(S)$ if and only if the following two conditions hold:
   (i) $x \leq \beta$ for all $x \in S$,
   (ii) for all real numbers r if $r < \beta$, then there is an $x \in S$ so that $r < x$.
2. $\alpha = \inf(S)$ if and only if the following two conditions hold:
   (i) $\alpha \leq x$ for all $x \in S$,
   (ii) for all real numbers q if $\alpha < q$, then there is an $x \in S$ so that $x < q$.

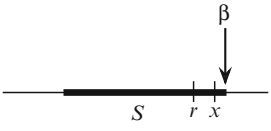


Fig. 9.5a $\beta = \sup(S)$

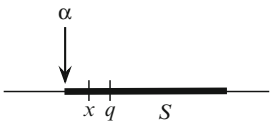


Fig. 9.5b $\alpha = \inf(S)$

Remark 9.3.20(1) thus leads to the following alternative proof and assumption strategies for dealing with equations of the form $\beta = \sup(S)$.

Proof Strategy 9.3.21. Given a real number β and a nonempty $S \subseteq \mathbb{R}$, to prove that $\beta = \sup(S)$ use the two-step proof diagram:

Step (1): Prove $x \leq \beta$ for all $x \in S$.
Step (2): Assume $r < \beta$.
Prove $r < x$ for some $x \in S$.

Assumption Strategy 9.3.22. Let $S \subseteq \mathbb{R}$ be nonempty and let $\beta \in \mathbb{R}$. Suppose you are *assuming* that $\beta = \sup(S)$. Then you can conclude that (1) $x \leq \beta$ for all $x \in S$, and (2) whenever $r < \beta$ there is an $x \in S$ such that $r < x$.

The proof of our next theorem again illustrates the ideas that are used to prove theorems on the supremum of a set of real numbers. Actually, we will present two proofs of this theorem. Our first proof applies the alternative Strategies 9.3.21 and 9.3.22. In our second proof we use Strategies 9.3.5 and 9.3.6.

Theorem 9.3.23. *Let $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$ be nonempty and bounded above. Then the set $C = \{x+y : x \in A \text{ and } y \in B\}$ is bounded above and* $\sup(C) = \sup(A) + \sup(B)$.

Proof Analysis. Since the sets A and B are bounded above, there are real numbers α and β satisfying $\alpha = \sup(A)$ and $\beta = \sup(B)$ by the completeness axiom. Thus, we must prove that $\alpha + \beta = \sup(C)$. First note that every element in $z \in C$ has the form $z = x + y$ for some $x \in A$ and some $y \in B$. So to prove that $\alpha + \beta$ is an upper bound for C, we need to show that $x + y \leq \alpha + \beta$ for each $x \in A$ and each $y \in B$. Appealing to Proof Strategy 9.3.21, we construct the following proof diagram:

Assume $\alpha = \sup(A)$ and $\beta = \sup(B)$.
Prove $z \leq \alpha + \beta$ for all $z \in C$.
Assume $s < \alpha + \beta$.
Prove $s < z$ for some $z \in C$.

The first line of this proof diagram indicates that we will be assuming $\alpha = \sup(A)$ and $\beta = \sup(B)$. Thus, we can use Assumption Strategy 9.3.22. The second line asserts that we must prove that $\alpha + \beta$ is an upper bound for the set C. The third line states that we will assume that s is a value satisfying $s < \alpha + \beta$. We will then have to prove that there is an element in C that is larger than s. Since every element in C has the form $x + y$ for $x \in A$ and $y \in B$, we just need to find an $x \in A$ and a $y \in B$ such that $s < x + y$. We now present our first proof. Ⓐ

Proof (of Theorem 9.3.23). Suppose that $\alpha = \sup(A)$ and $\beta = \sup(B)$. We shall prove that $\alpha + \beta = \sup(C)$, where $C = \{x + y : x \in A \text{ and } y \in B\}$. Let $z \in C$. So $z = x + y$ for some $x \in A$ and some $y \in B$. Since $x \leq \alpha$ and $y \leq \beta$, it follows that $z = (x + y) \leq (\alpha + \beta)$. We conclude that $\alpha + \beta$ is an upper bound for C.

To prove that $\alpha+\beta$ is the least upper bound for C, assume that $s<\alpha+\beta$. So, $s-\beta<\alpha$. Since $\alpha=\sup(A)$, there is an $x\in A$ such that $s-\beta<x$. Thus, $s-x<\beta$. Since $\beta=\sup(B)$, there is a $y\in B$ such that $s-x<y$. Hence, $s<x+y$ and $x+y\in C$. Therefore, $\alpha+\beta=\sup(C)$. □

We will give another proof of Theorem 9.3.23 using Strategies 9.3.5 and 9.3.6. Thus, the logical structure of our second proof will be as follows:

Assume $\alpha=\sup(A)$ and $\beta=\sup(B)$.
Prove $z\le\alpha+\beta$ for all $z\in C$.
Assume c is an upper bound for C.
Prove $\alpha+\beta\le c$.

Of course, the initial part of our next proof will be the same as in the first proof. The only difference between these two proofs lies in the second part of both proofs. In the second part of the proof below, we will be assuming that c is an upper bound for C and will prove that $\alpha+\beta\le c$.

Proof (of Theorem 9.3.23). Let $\alpha=\sup(A)$ and $\beta=\sup(B)$. We shall prove that $\alpha+\beta=\sup(C)$, where $C=\{x+y:x\in A \text{ and } y\in B\}$. Let $z\in C$. So $z=x+y$ for some $x\in A$ and some $y\in B$. Since $x\le\alpha$ and $y\le\beta$, we have that $z=(x+y)\le(\alpha+\beta)$. So $\alpha+\beta$ is an upper bound for C.

To prove that $\alpha+\beta$ is the least upper bound for C, suppose that c is an upper bound for C. Because c is an upper bound for C, we have that

$$x+y\le c \text{ for all } x\in A \text{ and all } y\in B. \tag{9.4}$$

Let $y\in B$ be arbitrary. From (9.4) we have that $x+y\le c$ for all $x\in A$. Hence, $x\le c-y$ for all $x\in A$. Thus $c-y$ is an upper bound for A. Since $\alpha=\sup(A)$, we infer that $\alpha\le c-y$. Since y was arbitrary, we conclude that $y\le c-\alpha$ for all $y\in B$. So $c-\alpha$ is an upper bound for B. Since $\beta=\sup(B)$, we have that $\beta\le c-\alpha$. Hence, $\alpha+\beta\le c$, and therefore $\alpha+\beta=\sup(C)$. □

Remark 9.3.20(2) (see Fig. 9.5b) provides us with the following alternative strategies for working with an equation of the form $\alpha=\inf(S)$.

Proof Strategy 9.3.24. Given a real number α and a nonempty $S\subseteq\mathbb{R}$, to prove that $\alpha=\inf(S)$ use the two-step proof diagram:

Step (1): Prove $\alpha\le x$ for all $x\in S$.
Step (2): Assume $\alpha<q$.
Prove $x<q$ for some $x\in S$.

Assumption Strategy 9.3.25. Let $S\subseteq\mathbb{R}$ be nonempty and let $\alpha\in\mathbb{R}$. Suppose you are *assuming* that $\alpha=\inf(S)$. Then you can infer (1) $\alpha\le x$ for all $x\in S$, and (2) whenever $\alpha<q$ there is an $x\in S$ such that $x<q$.

Exercises 9.3

1. For each of the following subsets S of $\mathbb{R}$, answer the following two questions: Is the set S bounded above? Is the set S bounded below?

(a) $S = [2,5]$ (b) $S = [2,5)$
(c) $S = (2,\infty)$ (d) $S = \mathbb{N}$
(e) $S = \{x \in \mathbb{R} : (x^2+1)^{-1} > \frac{1}{2}\}$ (f) $S = \{\frac{1}{n} : n \in \mathbb{N}\}$
(g) $S = \{q \in \mathbb{Q} : 0 \le q \le \sqrt{2}\}$ (h) $S = \{x \in \mathbb{R} : |2x+1| < 5\}$.

2. For each subset S of $\mathbb{R}$ identify the $\sup(S)$ and $\inf(S)$, if they exist.

(a) $S = [2,5]$ (b) $S = [2,5)$
(c) $S = (2,\infty)$ (d) $S = \mathbb{N}$
(e) $S = \{x \in \mathbb{R} : (x^2+1)^{-1} > \frac{1}{2}\}$ (f) $S = \{\frac{1}{n} : n \in \mathbb{N}\}$
(g) $S = \{q \in \mathbb{Q} : 0 \le q \le \sqrt{2}\}$ (h) $S = \{x \in \mathbb{R} : |2x+1| < 5\}$.

3. For each subset S of $\mathbb{R}$ identify the $\max(S)$ and $\min(S)$, if they exist.

(a) $S = \{2,5,6\}$ (b) $S = \mathbb{N}$
(c) $S = (2,\infty)$ (d) $S = \{q \in \mathbb{Q} : 0 \le q \le \sqrt{2}\}$.

4. Suppose $S \subseteq \mathbb{R}$ is nonempty and bounded. Let $A \subseteq S$ be nonempty. Prove that A is bounded. Then prove that $\sup(A) \le \sup(S)$ and $\inf(S) \le \inf(A)$.

5. Suppose $S \subseteq \mathbb{R}$ is nonempty and bounded. Let $\beta = \sup(S)$. Prove that for every $\varepsilon > 0$ there exists an $x \in S$ such that $\beta - \varepsilon < x$.

6. Suppose $S \subseteq \mathbb{R}$ is nonempty and bounded. Let $\alpha = \inf(S)$. Prove that for every $\varepsilon > 0$ there exists an $x \in S$ such that $x < \alpha + \varepsilon$.

7. Prove Theorem 9.3.12.

8. Prove Theorem 9.3.13.

9. Let $S \subseteq \mathbb{R}$ be nonempty and bounded. Let $k \in \mathbb{R}$ and define $A = \{k + x : x \in S\}$. Prove (a) $\sup(A) = k + \sup(S)$ and (b) $\inf(A) = k + \inf(S)$.

10. Let S be a bounded (nonempty) subset of $\mathbb{R}$. Let $k > 0$ and $c \in \mathbb{R}$. Prove that the set $A = \{kx + c : x \in S\}$ is bounded. Now prove (a) $\sup(A) = k\sup(S) + c$ and (b) $\inf(A) = k\inf(S) + c$.

11. Using the alternative Proof and Assumption Strategies 9.3.21 and 9.3.22, give a proof of Theorem 9.3.7 that is different than the one presented in the text.

12. Suppose $S \subseteq \mathbb{R}$ and $T \subseteq \mathbb{R}$ are nonempty and bounded. Prove that $S \cup T$ is bounded, and then prove the following:

(a) $\sup(S \cup T) = \max\{\sup(S), \sup(T)\}$
(b) $\inf(S \cup T) = \min\{\inf(S), \inf(T)\}$.

13. Prove part (a) of Theorem 9.3.19.

14. Let $S \subseteq \mathbb{R}^+$ and $T \subseteq \mathbb{R}^+$ be nonempty and bounded above. Prove that the set $P = \{xy : x \in S \text{ and } y \in T\}$ is bounded above and that $\sup(P) = \sup(S) \cdot \sup(T)$.

15. Let $f : D \to \mathbb{R}$ be bounded. Let $E \subseteq D$ be nonempty. Prove that the set $f[E]$ is bounded. Then prove that $\sup(f[E]) \le \sup(f[D])$ and $\inf(f[D]) \le \inf(f[E])$.

16. Let $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ be bounded. Assume $f(x) \leq g(x)$ for all $x \in D$. Prove that $\sup(f[D]) \leq \sup(g[D])$.

17. (The Archimedean Property) Prove that for each $x \in \mathbb{R}$ there is an $n \in \mathbb{N}$ such that $x < n$.

18. Suppose $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$ are nonempty and bounded below. Prove that the set $C = \{x+y : x \in A \text{ and } y \in B\}$ is bounded below and $\inf(C) = \inf(A) + \inf(B)$.

19. Let $S \subseteq \mathbb{R}^+$ and $T \subseteq \mathbb{R}^+$ be nonempty and bounded below. Prove that the set $P = \{xy : x \in S \text{ and } y \in T\}$ is bounded below and $\inf(P) = \inf(S) \cdot \inf(T)$.

Exercise Notes: For Exercise 1, $|2x+1| < 5$ is equivalent to $-5 < 2x+1 < 5$. For Exercises 5 and 6, see Remark 9.3.20. For Exercises 9 and 10, review the analysis and proofs for Theorems 9.3.7 and 9.3.11. For Exercise 14, review the analysis and proof of Theorem 9.3.23. For Exercise 17, let $x \in \mathbb{R}$ and suppose, for a contradiction, that $n \leq x$ for all $n \in \mathbb{N}$. Thus, $\mathbb{N} \subseteq \mathbb{R}$ is bounded above. Hence, by the completeness axiom, $\beta = \sup(\mathbb{N})$ exists. Since $\beta - 1 < \beta$, there is an $m \in \mathbb{N}$ such that $\beta - 1 < m$ (see Remark 9.3.20(1-ii)). Therefore, $\beta < m+1 = n \in \mathbb{N}$. For Exercise 18, review strategies 9.3.24 and 9.3.25. For Exercise 19, show that $\alpha = \inf(S) \geq 0$ and that $\beta = \inf(T) \geq 0$; and review strategies 9.3.8 and 9.3.10. Let c be a lower bound for P. If $c \leq 0$, then clearly $c \leq \alpha\beta$. Suppose $c > 0$ and let $y \in T$. Prove that $\frac{c}{y}$ is a lower bound for S.

9.4 Convergence of Sequences

Sequences are fundamental in real analysis and, while you may already be familiar with sequences, it is important to have a formal definition. A sequence, as depicted in Fig. 9.6, will be defined as a function from the set of natural numbers to the set of real numbers $\mathbb{R}$.

Definition 9.4.1. A **sequence** is a function $s\colon \mathbb{N} \to \mathbb{R}$. We shall denote the value $s(n)$ by s_n. We will write s as $\langle s_n \rangle$ or as $\langle s_1, s_2, s_3, \dots \rangle$.

Figure 9.6 illustrates the functional view of a sequence where the value $s(n)$, or s_n, is the length of the n-th arrow. An arrow that is pointing up has positive length and an arrow pointing down has negative length. Furthermore, a sequence $\langle s_n \rangle$ is often viewed as an infinite list $\langle s_1, s_2, s_3, \dots \rangle$ of real numbers. The numbers $s_1, s_2, \dots$ are called the *terms* of the sequence and s_n is called the n-th term. For example, we can view the sequence $\left\langle \frac{1}{n} \right\rangle$ as the infinite list $\left\langle 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \right\rangle$ where $\frac{1}{5}$ is the fifth term of this sequence.

The limit of a sequence is one of the oldest and most important concepts in real analysis. A sequence converges to a limit ℓ if the terms of the sequence get closer and closer to the real number ℓ. We now give a precise definition of this concept.

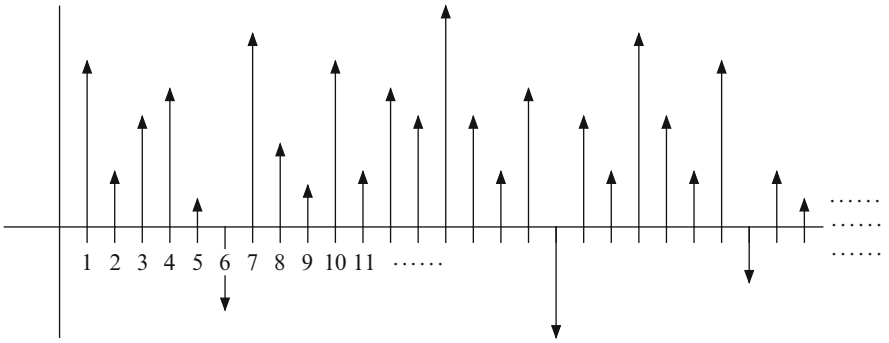


Fig. 9.6 Functional representation of a sequence: $s(4) = s_4 > 0$ and $s(6) = s_6 < 0$

Definition 9.4.2. A sequence $\langle s_n \rangle$ **converges** to the real number ℓ provided that for every $\varepsilon > 0$ there exists a natural number N such that for all $n \in \mathbb{N}$, if $n > N$ then $|s_n - \ell| < \varepsilon$. If a sequence does not converge, then we shall say that it **diverges**.

Figure 9.7 illustrates the convergence concept given in Definition 9.4.2 and, in this figure, when $n > N$, the length of the n-th arrow is within ε of ℓ; in other words, if $n > N$ then $|s_n - \ell| < \varepsilon$. When a sequence $\langle s_n \rangle$ converges to ℓ, then ℓ is called the **limit** of the sequence $\langle s_n \rangle$ and we write $\lim_{n\to\infty} s_n = \ell$.

The logical form of the convergence concept can be expressed as

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(n > N \to |s_n - \ell| < \varepsilon) \tag{9.5}$$

and this logical form motivates our next proof strategy.

Proof Strategy 9.4.3. To prove that $\lim_{n\to\infty} s_n = \ell$ use the proof diagram:

Let $\varepsilon > 0$ be a real number.
Let $N =$ (the natural number you found).
Let $n > N$ be a natural number.
Prove $|s_n - \ell| < \varepsilon$.

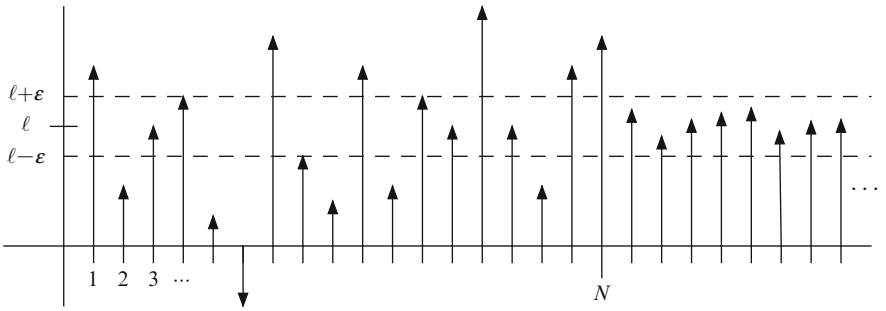


Fig. 9.7 For all $n > N$ we have $|s_n - \ell| < \varepsilon$

To apply Proof Strategy 9.4.3 on a specific sequence, first let $\varepsilon > 0$. We must find a natural number N such that when $n > N$, we can prove that $|s_n - \ell| < \varepsilon$. To find the desired N, we will first attempt the following:

Using algebra and properties of inequality on the expression $|s_n - \ell|$, 'extract out' a larger value that resembles $\frac{1}{n}$.

We shall then use this larger value to find N so that when $n > N$ we will have that $|s_n - \ell| < \varepsilon$. We shall illustrate this idea in our proof analysis of Theorems 9.4.5–9.4.8. Before we discuss these theorems, we identify three substitution properties of inequality that follow from Exercise 9 on page 297. These properties are very useful for proving theorems about convergence and extend the substitution properties 3.3.3 given on page 69.

Substitution Properties of Inequality 9.4.4. *Let a, b, c, d be positive real numbers. Then the following hold:*

(1) *Given the ratio $\frac{a}{b}$, if $a < c$, then you can conclude that $\frac{a}{b} < \frac{c}{b}$.*
(Replacing a numerator with a larger value yields a larger ratio.)

(2) *Given the ratio $\frac{a}{b}$, if $d < b$, then you can conclude that $\frac{a}{b} < \frac{a}{d}$.*
(Replacing a denominator with a smaller value yields a larger ratio.)

(3) *Given the ratio $\frac{a}{b}$, if $a \leq c$ and $d \leq b$, then you can conclude that $\frac{a}{b} \leq \frac{c}{d}$.*
(Replacing a numerator with a larger value and denominator with a smaller value yields a larger ratio.)

The Substitution Properties 9.4.4 and 3.3.3 will be implicitly used in nearly all of the remaining proofs of this chapter. To apply the above 1–3, you must first know that a, b, c, d are positive. To illustrate how we will be using these properties, suppose $x > 0$ and that we working with the ratio $\frac{x}{4}$. It follows that $\frac{x}{4} < \frac{x+1}{4}$ by item (1) of 9.4.4 because $x < x + 1$. We can also conclude from item (2) of 9.4.4 that $\frac{x}{4} < \frac{x}{2}$ since $2 < 4$.

Example 1. Property (2) in 9.4.4 implies the following assertions:

1. $\frac{1}{n} < \frac{1}{N}$ when $n > N > 0$.
2. $\frac{1}{\sqrt{n}} < \frac{1}{\sqrt{N}}$ when $n > N > 0$, by Theorem 3.8.6 on page 95.
3. $\frac{1}{2^n} < \frac{1}{n}$ when $2^n > n > 0$.

We shall now apply Proof Strategy 9.4.3 in the proofs of our next four theorems. In each such proof we will first let $\varepsilon > 0$ and then we will tell the reader the value for N that we will use to complete the proof. Prior to each of these proofs, we shall also present a proof analysis in which we discuss how we actually found N. In these proofs we will also be using some of the inequalities identified in Example 1.

Theorem 9.4.5. $\lim_{n\to\infty} \frac{1}{n} = 0.$

Proof Analysis. Given $\varepsilon > 0$, we must find an $N \in \mathbb{N}$ such that if $n > N$, then $\left|\frac{1}{n} - 0\right| < \varepsilon$. Since $\left|\frac{1}{n} - 0\right| = \frac{1}{n}$, we need to find N so that if $n > N$, then $\frac{1}{n} < \varepsilon$. Solving the inequality $\frac{1}{n} < \varepsilon$ for n, we see that we must have $n > \frac{1}{\varepsilon}$. So if we take a natural number $N \geq \frac{1}{\varepsilon}$, then we will be able to prove the desired result. We shall now present a logically correct proof using Proof Strategy 9.4.3 as a guide. Ⓐ

Proof. Let $\varepsilon > 0$ and let $N \geq \frac{1}{\varepsilon}$ be a natural number. For each $n > N$ we prove that $\left|\frac{1}{n} - 0\right| < \varepsilon$ as follows:

$$\begin{aligned}
\left|\frac{1}{n} - 0\right| &= \left|\frac{1}{n}\right| && \text{by algebra} \\
&= \frac{1}{n} && \text{because } \frac{1}{n} > 0 \\
&< \frac{1}{N} && \text{because } n > N \\
&\leq \frac{1}{\frac{1}{\varepsilon}} && \text{because } N \geq \frac{1}{\varepsilon} \\
&= \varepsilon && \text{by algebra.}
\end{aligned}$$

Therefore, $\left|\frac{1}{n} - 0\right| < \varepsilon$. □

In the proof of our next theorem we will be using item (2) of Example 1. Item (3) of the same example will be used in the proof of the subsequent Theorem 9.4.7.

Theorem 9.4.6. $\lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0.$

Proof Analysis. Given $\varepsilon > 0$, we must find an $N \in \mathbb{N}$ so that when $n > N$, we have $\left|\frac{1}{\sqrt{n}} - 0\right| < \varepsilon$. Since $\left|\frac{1}{\sqrt{n}} - 0\right| = \frac{1}{\sqrt{n}}$, we need to find N such that if $n > N$, then $\frac{1}{\sqrt{n}} < \varepsilon$. Solving the inequality $\frac{1}{\sqrt{n}} < \varepsilon$ for n, we obtain $n > \frac{1}{\varepsilon^2}$. So if we take a natural number $N \geq \frac{1}{\varepsilon^2}$, then we will be able to prove the desired result. Ⓐ

Proof. Let $\varepsilon > 0$. Let $N \geq \frac{1}{\varepsilon^2}$ be a natural number. For $n > N$, we have that

$$\begin{aligned}
\left|\frac{1}{\sqrt{n}} - 0\right| &= \left|\frac{1}{\sqrt{n}}\right| && \text{by algebra} \\
&= \frac{1}{\sqrt{n}} && \text{because } \frac{1}{\sqrt{n}} > 0 \\
&< \frac{1}{\sqrt{N}} && \text{because } n > N
\end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{\sqrt{\frac{1}{\varepsilon^2}}} && \text{because } N \geq \frac{1}{\varepsilon^2} \\ &= \varepsilon && \text{by algebra.} \end{aligned}$$

Therefore, $\left|\frac{1}{\sqrt{n}} - 0\right| < \varepsilon$. □

Theorem 9.4.7. $\lim_{n\to\infty} 1 + \frac{1}{2^n} = 1$.

Proof Analysis. Given $\varepsilon > 0$, we need to find an $N \in \mathbb{N}$ so that if $n > N$, then $\left|1 + \frac{1}{2^n} - 1\right| < \varepsilon$. Since $\left|1 + \frac{1}{2^n} - 1\right| = \frac{1}{2^n}$, we want an $N \in \mathbb{N}$ so that if $n > N$, then $\frac{1}{2^n} < \varepsilon$. Solving the inequality $\frac{1}{2^n} < \varepsilon$ for n is difficult and so, we use a different approach. It is easy to show by induction that $n < 2^n$ and thus, $\frac{1}{2^n} < \frac{1}{n}$ when $n \geq 1$. We solve the inequality $\frac{1}{n} < \varepsilon$ for n and obtain $n > \frac{1}{\varepsilon}$. So if we take $N \geq \frac{1}{\varepsilon}$, then we will be able to prove the desired result using Proof Strategy 9.4.3. Ⓐ

Proof. Let $\varepsilon > 0$ and let $N \geq \frac{1}{\varepsilon}$ be a natural number. For each $n > N$ we have that

$$\begin{aligned} \left|\left(1 + \frac{1}{2^n}\right) - 1\right| &= \left|\frac{1}{2^n}\right| && \text{by algebra} \\ &= \frac{1}{2^n} && \text{because } \frac{1}{2^n} > 0 \\ &< \frac{1}{n} && \text{because } n < 2^n \\ &< \frac{1}{N} && \text{because } n > N \\ &\leq \frac{1}{\frac{1}{\varepsilon}} && \text{because } N \geq \frac{1}{\varepsilon} \\ &= \varepsilon && \text{by algebra.} \end{aligned}$$

Therefore, $\left|1 + \frac{1}{2^n} - 1\right| < \varepsilon$. □

Theorem 9.4.8. $\lim_{n\to\infty} \frac{2n+3}{3n+5} = \frac{2}{3}$.

Proof Analysis. For $\varepsilon > 0$, we want an $N \in \mathbb{N}$ so that if $n > N$, then $\left|\frac{2n+3}{3n+5} - \frac{2}{3}\right| < \varepsilon$. We see that $\left|\frac{2n+3}{3n+5} - \frac{2}{3}\right| = \left|\frac{-1}{9n+15}\right| = \frac{1}{9n+15}$, as $9n + 15 > 0$ for $n \in \mathbb{N}$. Thus we need to find N so that if $n > N$ then $\frac{1}{9n+15} < \varepsilon$. One could now solve the inequality $\frac{1}{9n+15} < \varepsilon$ for n, but we take an easier approach. Since $\frac{1}{9n+15} < \frac{1}{9n}$ (see 9.4.4(2)), we shall find an N so that if $n > N$ then $\frac{1}{9n} < \varepsilon$. Solving the inequality $\frac{1}{9n} < \varepsilon$ for n, we obtain $n > \frac{1}{9\varepsilon}$. If we take $N \geq \frac{1}{9\varepsilon}$, then we will be able to prove the desired result. Ⓐ

Proof. For $\varepsilon > 0$, let $N \geq \frac{1}{9\varepsilon}$ be a natural number. For $n > N$ we have

$$\begin{aligned}
\left|\frac{2n+3}{3n+5} - \frac{2}{3}\right| &= \left|\frac{(2n+3)3 - 2(3n+5)}{3(3n+5)}\right| && \text{common denominator} \\
&= \left|\frac{6n+9-6n+10)}{3(3n+5)}\right| && \text{by distributivity} \\
&= \left|\frac{-1}{9n+15}\right| && \text{by algebra} \\
&= \frac{1}{9n+15} && \text{because } 9n+15 > 0 \\
&< \frac{1}{9n} && \text{because } 9n < 9n+15 \\
&< \frac{1}{9N} && \text{because } n > N \\
&\leq \frac{1}{9\frac{1}{9\varepsilon}} && \text{because } N \geq \frac{1}{9\varepsilon} \\
&= \varepsilon && \text{by algebra.}
\end{aligned}$$

Therefore, $\left|\frac{2n+3}{3n+5} - \frac{2}{3}\right| < \varepsilon$ and this completes the proof. □

Remark 9.4.9. In each of the proofs of Theorems 9.4.5–9.4.8, we identified a real number x and then selected a natural number $N \geq x$. The Archimedean Property (see Exercise 17 on page 311) asserts that whenever x is a real number, there is such a natural number $N \geq x$.

Suppose in a proof you are assuming that a given sequence converges and you want to use this assumption to prove that another sequence converges. The next strategy will then be essential.

Assumption Strategy 9.4.10. Suppose you are *assuming* that $\lim_{n\to\infty} s_n = \ell$. Then for any $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $|s_n - \ell| < \varepsilon$ for all $n > N$.

Given that $\lim_{n\to\infty} s_n = \ell$, using Assumption Strategy 9.4.10, you can conclude that for any positive value $v > 0$ there is a natural number N such that for all $n > N$ we have that $|s_n - \ell| < v$. We shall express this observation as "we can make $|s_n - \ell|$ as small as we want."

We apply this idea in the proof of our next theorem where we will be assuming that $\lim_{n\to\infty} s_n = \ell$ with $\ell > 0$. In the proof, we will use the positive value $v = \varepsilon\sqrt{l}$. As $\lim_{n\to\infty} s_n = \ell$, we will conclude there is an N such that $|s_n - \ell| < \varepsilon\sqrt{l}$ for all $n > N$.

Theorem 9.4.11. *Suppose that* $\lim_{n\to\infty} s_n = \ell$ *where* $\ell > 0$ *and* $s_n > 0$ *for all* $n \geq 1$. *Then* $\lim_{n\to\infty} \sqrt{s_n} = \sqrt{\ell}$.

Proof Analysis. In a proof of the above theorem, we assume that $\lim_{n\to\infty} s_n = \ell$ and we must prove that $\lim_{n\to\infty} \sqrt{s_n} = \sqrt{\ell}$. How can one apply Proof Strategy 9.4.3 and Assumption Strategy 9.4.10 to find such a proof? First of all, our proof will need to have the following logical structure:

Assume $\lim_{n\to\infty} s_n = \ell$.
Let $\varepsilon > 0$ be a real number.
  Let $N =$ (the natural number you found).
    Let $n > N$ be a natural number.
      Prove $\left|\sqrt{s_n} - \sqrt{\ell}\right| < \varepsilon$.

For any given $\varepsilon > 0$ we must find a natural number N such that if $n > N$, then $|\sqrt{s_n} - \sqrt{\ell}| < \varepsilon$. We shall use the assumption $\lim_{n\to\infty} s_n = \ell$ to find the desired N. Here is the basic plan that we will apply to get N.

> *Using algebra and properties of inequality on the expression* $|\sqrt{s_n} - \sqrt{\ell}|$, *"extract out" a larger value containing* $|s_n - \ell|$ *and no other occurrences of* s_n.

Since $\lim_{n\to\infty} s_n = \ell$, we can make $|s_n - \ell|$ "as small as we want." We should then be able to make $|\sqrt{s_n} - \sqrt{\ell}| < \varepsilon$ and find the N that we need. Let us execute this plan! First we start with $|\sqrt{s_n} - \sqrt{\ell}|$ and extract out $|s_n - \ell|$ as follows:

$$\left|\sqrt{s_n} - \sqrt{\ell}\right| = \left|\frac{(\sqrt{s_n} - \sqrt{\ell})}{1}\frac{(\sqrt{s_n} + \sqrt{\ell})}{(\sqrt{s_n} + \sqrt{\ell})}\right| \quad \text{rationalizing the numerator}$$

$$= \left|\frac{s_n - \ell}{\sqrt{s_n} + \sqrt{\ell}}\right| \quad \text{by algebra}$$

$$= \frac{|s_n - \ell|}{\sqrt{s_n} + \sqrt{\ell}} \quad \text{because } \sqrt{s_n} + \sqrt{\ell} > 0$$

$$< \frac{|s_n - \ell|}{\sqrt{\ell}} \quad \text{because } \sqrt{\ell} < \sqrt{s_n} + \sqrt{\ell}.$$

We started with $|\sqrt{s_n} - \sqrt{\ell}|$ and extracted out the larger value $\frac{|s_n-\ell|}{\sqrt{\ell}}$ which contains $|s_n - \ell|$ and no other occurrences of s_n. Consequently, $\left|\sqrt{s_n} - \sqrt{\ell}\right| < \frac{|s_n-\ell|}{\sqrt{\ell}}$. Thus, if $\frac{|s_n-\ell|}{\sqrt{\ell}} < \varepsilon$, then we will have that $\left|\sqrt{s_n} - \sqrt{\ell}\right| < \varepsilon$. How small must $|s_n - \ell|$ be to ensure that $\frac{|s_n-\ell|}{\sqrt{\ell}} < \varepsilon$? To answer this question, we just solve this latter inequality for $|s_n - \ell|$ to obtain $|s_n - \ell| < \varepsilon\sqrt{\ell}$. Hence, we need an N so that $|s_n - \ell| < \varepsilon\sqrt{\ell}$

when $n > N$. Since $\lim_{n\to\infty} s_n = \ell$, there is such an N. This is the value for N that we will use in our proof. Ⓐ

Proof (of Theorem 9.4.11). Assume that $\lim_{n\to\infty} s_n = \ell$ where $\ell > 0$ and $s_n > 0$ for all $n \geq 1$. We shall prove that $\lim_{n\to\infty} \sqrt{s_n} = \sqrt{\ell}$. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} s_n = \ell$, there is a natural number N such that $(\star)$ $|s_n - \ell| < \varepsilon\sqrt{\ell}$ for all $n > N$. Let $n > N$ be a natural number. We prove that $\left|\sqrt{s_n} - \sqrt{\ell}\right| < \varepsilon$ as follows:

$$\begin{aligned}
\left|\sqrt{s_n} - \sqrt{\ell}\right| &= \left|\frac{(\sqrt{s_n} - \sqrt{\ell})}{1}\frac{(\sqrt{s_n} + \sqrt{\ell})}{(\sqrt{s_n} + \sqrt{\ell})}\right| && \text{rationalizing the numerator} \\
&= \left|\frac{s_n - \ell}{\sqrt{s_n} + \sqrt{\ell}}\right| && \text{as } (\sqrt{s_n} - \sqrt{\ell})(\sqrt{s_n} + \sqrt{\ell}) = s_n - \ell \\
&= \frac{|s_n - \ell|}{\sqrt{s_n} + \sqrt{\ell}} && \text{because } \sqrt{s_n} + \sqrt{\ell} > 0 \\
&< \frac{|s_n - \ell|}{\sqrt{\ell}} && \text{because } \sqrt{\ell} < \sqrt{s_n} + \sqrt{\ell} \\
&< \frac{\varepsilon\sqrt{\ell}}{\sqrt{\ell}} && \text{by } (\star) \text{ because } n > N \\
&= \varepsilon && \text{by algebra.}
\end{aligned}$$

Therefore, $\left|\sqrt{s_n} - \sqrt{\ell}\right| < \varepsilon$ and this completes the proof. □

Theorem 9.4.12 (Uniqueness of the Limit). *If a sequence converges, then there is only one limit of the sequence.*

Proof. Suppose the sequence $\langle s_n \rangle$ converges. To prove that there is only one limit of this sequence, suppose that ℓ and ℓ' are both limits of the sequence $\langle s_n \rangle$. We shall prove that $\ell = \ell'$. For a contradiction, assume $\ell \neq \ell'$. Let $\varepsilon = |\ell - \ell'|$. Since $\ell \neq \ell'$, we have that $\varepsilon > 0$. Because $\langle s_n \rangle$ converges to ℓ, there is an $N \in \mathbb{N}$ such that $|s_n - \ell| < \frac{\varepsilon}{2}$ for all $n > N$. Since $\langle s_n \rangle$ also converges to ℓ', there is an $N' \in \mathbb{N}$ such that $|s_n - \ell'| < \frac{\varepsilon}{2}$ for all $n > N'$. Therefore, for all $n > \max\{N, N'\}$ we have that

$$\left|\ell - \ell'\right| = \left|(\ell - s_n) + (s_n - \ell')\right| \leq |\ell - s_n| + \left|s_n - \ell'\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, $|\ell - \ell'| < \varepsilon$. As $\varepsilon = |\ell - \ell'|$, we can conclude that $\varepsilon < \varepsilon$. This contradiction completes the proof of the theorem. □

If you are assuming that a given sequence converges and you need to prove that another sequence converges, then Assumption Strategy 9.4.10 will be useful.

For instance, suppose you are assuming that $\lim_{n\to\infty} a_n = a$ and you are also working with $\varepsilon > 0$. Using Assumption Strategy 9.4.10 you can conclude for any positive $v < \varepsilon$ (for example, $v = \frac{\varepsilon}{2}$), there is an N' such that for all $n > N'$ we have $|a_n - a| < v$.

Theorem 9.4.13. *Let $\langle s_n \rangle$, $\langle a_n \rangle$ be sequences and let $\ell \in \mathbb{R}$. If*

(1) $|s_n - \ell| \leq k\,|a_n|$ *for all $n \geq m$ where $k > 0$ and $m \in \mathbb{N}$,*
(2) $\lim_{n\to\infty} a_n = 0$,

then $\lim_{n\to\infty} s_n = \ell$.

Proof Analysis. Let us assume that $|s_n - \ell| \leq k\,|a_n|$ for all $n \geq m$, where $k > 0$ and $m \in \mathbb{N}$. We shall also assume that $\lim_{n\to\infty} a_n = 0$. Because $\lim_{n\to\infty} a_n = 0$, we can make $|a_n| = |a_n - 0|$ "as small as we want." Given $\varepsilon > 0$, we must find an $N \in \mathbb{N}$ so that if $n > N$, then $|s_n - \ell| < \varepsilon$. Since $|s_n - \ell| \leq k\,|a_n|$ whenever $n \geq m$, we need to find an $N \geq m$ so that $k\,|a_n| < \varepsilon$ for $n > N$. Solving the inequality $k\,|a_n| < \varepsilon$ for $|a_n|$, we obtain $|a_n| < \frac{\varepsilon}{k}$. Because $\lim_{n\to\infty} a_n = 0$, there is an $N' \in \mathbb{N}$ such that if $n > N'$, then $|a_n - 0| = |a_n| < \frac{\varepsilon}{k}$. Thus, we will use $N = \max\{N', m\}$. This value for N will ensure that when $n > N$, we will have $n > N'$ and $n > m$. Ⓐ

Proof. Let $\langle s_n \rangle$, $\langle a_n \rangle$ be sequences and let $\ell \in \mathbb{R}$. Assuming (1) and (2), in the statement of the theorem, we prove that $\lim_{n\to\infty} s_n = \ell$. Let $\varepsilon > 0$. By (1), we have

$$|s_n - \ell| \leq k\,|a_n| \text{ for all } n \geq m \tag{9.6}$$

where $k > 0$ and $m \in \mathbb{N}$. By (2) there is an $N' \in \mathbb{N}$ such that

$$|a_n - 0| = |a_n| < \frac{\varepsilon}{k} \text{ for all } n > N'. \tag{9.7}$$

Let $N = \max\{m, N'\}$. If $n > N$, then

$$\begin{aligned} |s_n - \ell| &\leq k\,|a_n| && \text{by (9.6) because } n > N \geq m \\ &< k\left(\frac{\varepsilon}{k}\right) && \text{by (9.7) because } n > N \geq N' \\ &= \varepsilon && \text{by algebra.} \end{aligned}$$

Therefore $|s_n - \ell| < \varepsilon$. This completes the proof of the theorem. □

Corollary 9.4.14. *Let x be a real number satisfying $0 < x < 1$. Then* $\lim_{n\to\infty} x^n = 0$.

Proof. Let $x \in \mathbb{R}$ be such that $0 < x < 1$. Since $\frac{1}{x} > 1$, there is a $c > 0$ such that $1 + c = \frac{1}{x}$. Let n be a natural number. By Bernoulli's inequality (see Exercise 7 on page 126) we have that $(1+c)^n \geq 1 + nc$. Hence, $\frac{1}{x^n} = (1+c)^n \geq 1 + nc > nc$ and

so, $x^n < \frac{1}{nc}$. Thus, $|x^n - 0| = x^n < \frac{1}{cn}$ for all $n \geq 1$. Because $\lim_{n\to\infty} \frac{1}{n} = 0$ and $c > 0$, Theorem 9.4.13 implies that $\lim_{n\to\infty} x^n = 0$. □

The next corollary can be helpful for proving that a particular sequence $\langle u_n \rangle$ converges to r if you know that another related sequence converges.

Lemma 9.4.15. *If* $\lim_{n\to\infty} s_n = \ell$, *then* $\lim_{n\to\infty} (s_n - \ell) = 0$.

Proof. Assume $\lim_{n\to\infty} s_n = \ell$. We shall prove that $\lim_{n\to\infty} (s_n - \ell) = 0$. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} s_n = \ell$, there is an $N \in \mathbb{N}$ where $|s_n - \ell| < \varepsilon$ for all $n > N$. So if $n > N$, then $|(s_n - \ell) - 0| = |s_n - \ell| < \varepsilon$ and this completes the proof. □

Corollary 9.4.16. *Let* $\langle s_n \rangle$ *be a sequence such that* $\lim_{n\to\infty} s_n = \ell$ *for a real number* ℓ. *Suppose that* $\langle u_n \rangle$ *is a sequence satisfying* $|u_n - r| \leq k\,|s_n - \ell|$ *for all* $n \geq m$, *where* $k > 0$ *and* r *are real numbers, and* $m \in \mathbb{N}$. *Then* $\lim_{n\to\infty} u_n = r$.

Proof. Suppose that $\lim_{n\to\infty} s_n = \ell$ and $(\star)$ $|u_n - r| \leq k\,|s_n - \ell|$ for all $n \geq m$, where $k > 0$ and $m \in \mathbb{N}$. Since $\lim_{n\to\infty} s_n = \ell$, Lemma 9.4.15 implies that $\lim_{n\to\infty} (s_n - \ell) = 0$. Theorem 9.4.13, together with $(\star)$, implies that $\lim_{n\to\infty} u_n = r$. □

Recall that the logical form of the assertion $\lim_{n\to\infty} s_n = \ell$ can be expressed as

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(n > N \to |s_n - \ell| < \varepsilon). \tag{9.8}$$

Upon taking the negation of (9.8) and applying the appropriate logic laws, we obtain our next remark which expresses what it means for a sequence not to converge to ℓ.

Remark 9.4.17. The sequence $\langle s_n \rangle$ does **not** converge to ℓ if and only if there exists an $\varepsilon > 0$ such that for all $N \in \mathbb{N}$ there is an $n > N$ such that $|s_n - \ell| \geq \varepsilon$.

Definition 9.4.18. Let $x \in \mathbb{R}$ and let $\varepsilon > 0$. The open interval $(x - \varepsilon, x + \varepsilon)$, centered at x, is called a **neighborhood** of x.

Let x be a real number and let U_ε be the neighborhood $(x - \varepsilon, x + \varepsilon)$ of x where $\varepsilon > 0$. Then a real number s is in U_ε if and only if $|s - x| < \varepsilon$. Thus, U_ε consists of all the points whose distance from x is less than ε. Our next theorem shows that the notion of convergence can be expressed in terms of neighborhoods (see Fig. 9.8).

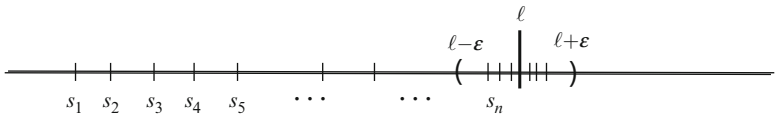


Fig. 9.8 An illustration for Theorem 9.4.19

Theorem 9.4.19. *Let $\langle s_n \rangle$ be a sequence and let ℓ be a real number. The following statements are equivalent:*

1. $\lim_{n\to\infty} s_n = \ell$.
2. *For every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ so that for all $n \in \mathbb{N}$, if $n > N$ then $|s_n - \ell| < \varepsilon$.*
3. *For every neighborhood U_ε of ℓ there is an $N \in \mathbb{N}$ so that for all $n \in \mathbb{N}$, if $n > N$ then $s_n \in U_\varepsilon$.*

Corollary 9.4.20. *Let $\langle s_n \rangle$ be a sequence of distinct points and suppose $\langle s_n \rangle$ converges to ℓ. Then every neighborhood of ℓ contains an infinite number of points from the sequence $\langle s_n \rangle$.*

Proof. Suppose U is a neighborhood of ℓ and that the sequence $\langle s_n \rangle$ converges to ℓ. Theorem 9.4.19 states that there is an $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n > N$ then $s_n \in U$. Therefore, an infinite number of points from the sequence $\langle s_n \rangle$ are in U. □

9.4.1 Bounded Sequences

Definition 9.4.21. A sequence $\langle s_n \rangle$ is **bounded** if there are real numbers a and b such that $a \le s_n \le b$ for all $n \ge 1$.

Remark 9.4.22. A sequence $\langle s_n \rangle$ is bounded if and only if there is an $M > 0$ such that $|s_n| \le M$ for all $n \in \mathbb{N}$ (see Theorem 9.3.2).

Theorem 9.4.23. *Let $\langle s_n \rangle$ be a convergent sequence. Then $\langle s_n \rangle$ is bounded.*

Proof. Assume $\lim_{n\to\infty} s_n = \ell$. Thus, for any $\varepsilon > 0$ there is an $N \in \mathbb{N}$ so that $|s_n - \ell| < \varepsilon$ for all $n > N$. Let us take $\varepsilon = 1$ and let $N \in \mathbb{N}$ be such that $|s_n - \ell| < 1$ for all $n > N$. By the backward triangle inequality, we conclude that $|s_n| - |\ell| \le |s_n - \ell| < 1$ for all $n > N$. Hence $|s_n| - |\ell| < 1$, that is, $|s_n| < |\ell| + 1$ for all $n > N$. Let

$$M = \max\{|s_1|, \ldots, |s_N|, |\ell| + 1\}.$$

We see that $|s_n| \le M$ for all $n \in \mathbb{N}$. Therefore, $\langle s_n \rangle$ is a bounded sequence. □

Remark 9.4.24. If a sequence $\langle s_n \rangle$ is unbounded, then the sequence $\langle s_n \rangle$ diverges.

In our next example, we show that the sequence of harmonic sums diverges.

Example 2. Consider the sequence $\langle s_n \rangle$ where $s_n = \sum_{k=1}^{n} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$. Now imagine n to be very large and write

$$\begin{aligned} s_n &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots + \frac{1}{n} \\ &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \cdots + \frac{1}{16}\right) + \cdots + \frac{1}{n} \end{aligned}$$

$$> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \left(\frac{1}{16} + \cdots + \frac{1}{16}\right) + \cdots + \frac{1}{n}$$
$$= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{n}.$$

It is clear that by taking n sufficiently large we can introduce as many $\frac{1}{2}$'s in the sum as we wish. Therefore, the sequence $\langle s_n \rangle$ is unbounded and so, it diverges.

Exercises 9.4

1. Let $a \in \mathbb{R}$. Prove that the sequence $\left\langle a + (-1)^n \frac{2n+1}{n} \right\rangle$ is bounded.
2. Let $k \neq 0$. Use Definition 9.4.2 to prove that $\lim_{n\to\infty} \frac{k}{n} = 0$.
3. Use Definition 9.4.2 to prove that $\lim_{n\to\infty} \frac{n+1}{n+2} = 1$.
4. Use Definition 9.4.2 to prove that $\lim_{n\to\infty} \frac{3n}{n+2} = 3$.
5. Use Definition 9.4.2 to prove that $\lim_{n\to\infty} \frac{6n-7}{3n-2} = 2$.
6. Use Definition 9.4.2 to prove that $\lim_{n\to\infty} \frac{6n-7}{2n-7} = 3$.
7. Prove that the limits given in Exercises 3–6 hold by applying Theorems 9.4.13 and 9.4.5.
8. Let $\ell \geq 0$. Prove that $\lim_{n\to\infty} (-1)^n \neq \ell$.
9. Prove Theorem 9.4.11 using Corollary 9.4.16.
10. Let $\langle s_n \rangle$ be a convergent sequence and let $c \in \mathbb{R}$. Suppose $\lim_{n\to\infty} s_n = \ell$. Prove that $\lim_{n\to\infty} (c + s_n) = c + \ell$.
11. Let $\langle s_n \rangle$ be a convergent sequence and let $c \in \mathbb{R}$ be nonzero. Suppose $\lim_{n\to\infty} s_n = \ell$. Prove that $\lim_{n\to\infty} (cs_n) = c\ell$.
12. Let $\langle s_n \rangle$ be a convergent sequence. Prove that if $\lim_{n\to\infty} s_n = \ell$, then $\lim_{n\to\infty} |s_n| = |\ell|$.
13. Let $\langle s_n \rangle$ be a convergent sequence and let $M > 0$. Prove that if $\lim_{n\to\infty} s_n = \ell$ and $|s_n| \leq M$ for all $n \geq 1$, then $\lim_{n\to\infty} {s_n}^2 = \ell^2$.
14. Let $c \in \mathbb{R}$ be constant. Prove that $\lim_{n\to\infty} c = c$.
15. Let $\langle x_n \rangle$ and $\langle y_n \rangle$ be two convergent sequences. Prove that there exists an $M > 0$ such that $|x_n| \leq M$ and $|y_n| \leq M$ for all $n \geq 1$.
16. Let $\langle s_n \rangle$ a convergent sequence. Suppose that $\lim_{n\to\infty} s_n = \ell$. Prove that there exists an $M > 0$ such that $|s_n + \ell| \leq M$ for all $n \geq 1$.
17. Use Theorems 9.4.13 and 9.4.5 to prove that $\lim_{n\to\infty} \frac{\sin(n)}{n} = 0$.
18. Use Theorems 9.4.13 and 9.4.6 to prove that $\lim_{n\to\infty} (\sqrt{n+1} - \sqrt{n}) = 0$.

19. Let $\langle s_n \rangle$ a convergent sequence. Prove that if $\lim_{n\to\infty} s_n = \ell$, then $\lim_{n\to\infty} s_n^2 = \ell^2$.

Exercise Notes: For Exercise 5, observe that $n \leq 3n-2$ when $n \geq 1$. For Exercise 6, note that $n < |2n-7|$ if $n > 7$. In this case, we would need N to be at least 7 and so, any $N \geq \max\{7, \frac{14}{\varepsilon}\}$ could be used in the proof. For Exercise 8, apply Remark 9.4.17. For Exercise 18, show that $\left|\sqrt{n+1} - \sqrt{n}\right| = \frac{1}{\sqrt{n+1}+\sqrt{n}} \leq \frac{1}{2\sqrt{n}}$.

9.5 Limit Theorems for Sequences

What are limit theorems? A limit theorem states that if you know the limits of some given sequences, then you can determine the limit of another sequence which is related to the given sequences. Limit theorems have the form:

Theorem. *Suppose that* $\lim_{n\to\infty} s_n = s$ *and* $\lim_{n\to\infty} t_n = t$. *Then one can evaluate the limit, say* $\lim_{n\to\infty} u_n = u$, *of a sequence* $\langle u_n \rangle$ *that is constructed from* $\langle s_n \rangle$ *and* $\langle t_n \rangle$.

How does one prove theorems of this form? The sequence $\langle u_n \rangle$ is defined from the sequences $\langle s_n \rangle$ and $\langle t_n \rangle$. We are also assuming that $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$, and we must prove that $\lim_{n\to\infty} u_n = u$. The following strategy will guide us.

Proof Strategy 9.5.1. To prove that $\lim_{n\to\infty} u_n = u$, our proof must contain the structure

Assume $\lim_{n\to\infty} s_n = s$.
Assume $\lim_{n\to\infty} t_n = t$.
Let $\varepsilon > 0$ be a real number.
  Let $N =$ (the natural number you found).
    Let $n > N$ be a natural number.
      Prove $|u_n - u| < \varepsilon$.

To apply Proof Strategy 9.5.1, let $\varepsilon > 0$. We must find a natural number N such that if $n > N$, then $|u_n - u| < \varepsilon$. To find the desired N, we shall use the assumptions $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$. Here is the basic idea that we will apply to get N.

> *Using algebra and properties of inequality on the expression* $|u_n - u|$, *"extract out" a larger value containing* $|s_n - s|$ *and* $|t_n - t|$, *and no other occurrences of* s_n *or* t_n.

Since $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$, we can make $|s_n - s|$ and $|t_n - t|$ "as small as we want." We should then be able to make $|u_n - u| < \varepsilon$ and find the N that we need. We will apply these ideas in our proof analysis of the next two theorems.

Theorem 9.5.2. *Suppose that* $\lim_{n\to\infty} s_n = s$ *and* $\lim_{n\to\infty} t_n = t$. *Then* $\lim_{n\to\infty}(s_n + t_n) = s + t$.

Proof Analysis. Let $\varepsilon > 0$. We are assuming that $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$. So we can make $|s_n - s|$ and $|t_n - t|$ as small as we want. We need to make

$$|(s_n + t_n) - (s + t)| < \varepsilon.$$

Using algebra and properties of inequality on the expression $|(s_n + t_n) - (s + t)|$ we extract out $|s_n - s|$ and $|t_n - t|$ as follows:

$$\begin{aligned} |(s_n + t_n) - (s + t)| &= |(s_n - s) + (t_n - t)| && \text{by algebra} \\ &\leq |s_n - s| + |t_n - t| && \text{by the triangle inequality.} \end{aligned}$$

Thus, if we have that $|s_n - s| < \frac{\varepsilon}{2}$ and $|t_n - t| < \frac{\varepsilon}{2}$, then we can conclude that

$$|(s_n + t_n) - (s + t)| < \varepsilon.$$

Since $\lim_{n\to\infty} s_n = s$, there is an N_s such that $|s_n - s| < \frac{\varepsilon}{2}$ when $n > N_s$. Similarly, there is an N_t such that $|t_n - t| < \frac{\varepsilon}{2}$ when $n > N_t$. Let $N = \max\{N_s, N_t\}$. When $n > N$, we can be assured that $n > N_s$ and $n > N_t$ both hold. We now offer a logically correct proof using Proof Strategy 9.5.1. Ⓐ

Proof. Suppose $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$. We shall prove that $\lim_{n\to\infty}(s_n + t_n) = s + t$. Let $\varepsilon > 0$. Since $\lim_{n\to\infty} s_n = s$, there is an $N_s \in \mathbb{N}$ such that

$$|s_n - s| < \frac{\varepsilon}{2} \text{ for all } n > N_s. \tag{9.9}$$

Also, because $\lim_{n\to\infty} t_n = t$, there is an $N_t \in \mathbb{N}$ such that

$$|t_n - t| < \frac{\varepsilon}{2} \text{ for all } n > N_t. \tag{9.10}$$

Let $N = \max\{N_s, N_t\}$. Thus for $n > N$ we have that

$$\begin{aligned} |(s_n + t_n) - (s + t)| &= |(s_n - s) + (t_n - t)| && \text{by distributivity and commutativity} \\ &\leq |s_n - s| + |t_n - t| && \text{by the triangle inequality} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{by (9.9) and (9.10)} \\ &= \varepsilon && \text{by algebra.} \end{aligned}$$

Hence, $|(s_n + t_n) - (s + t)| < \varepsilon$. Therefore, $\lim_{n\to\infty}(s_n + t_n) = s + t$. □

In the proof of Theorem 9.5.2 we were able to 'cleanly' extract out $|s_n - s|$ and $|t_n - t|$, that is, there were no additional factors. This may not be the case when applying Strategy 9.5.1 in the proof of other such theorems. After extracting out $|s_n - s|$ or $|t_n - t|$, there may be times when 'unwanted' factors appear that involve n, s_n or t_n. We will then have to find an upper bound for these factors. This will be done in the proofs of Theorems 9.5.3 and 9.5.5, below. In our proof of Theorem 9.5.3, we obtain the unwanted factor $|t_n|$. Since the sequence $\langle t_n \rangle$ converges, Theorem 9.4.23 implies that there is a $K > 0$ such that $|t_n| \leq K$ for all $n \geq 1$. We will thus have our desired upper bound. Actually, we shall use $M = \max\{K, |s|\}$ for the upper bound to avoid the possibility that $s = 0$.

Theorem 9.5.3. *Suppose* $\lim_{n\to\infty} s_n = s$ *and* $\lim_{n\to\infty} t_n = t$. *Then* $\lim_{n\to\infty} (s_n t_n) = st$.

Proof Analysis. We are assuming that $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$. So, we can make $|s_n - s|$ and $|t_n - t|$ as small as we want. We need to make $|s_n t_n - st| < \varepsilon$. Since the sequence $\langle t_n \rangle$ converges, we know by Theorem 9.4.23 that this sequence is bounded. Thus, there is a $K > 0$ such that

$$|t_n| \leq K \text{ for all } n \geq 1. \tag{9.11}$$

Let $M = \max\{K, |s|\}$. Using algebra and properties of inequality on the expression $|s_n t_n - st|$, we extract out $|s_n - s|$ and $|t_n - t|$ as follows[2]:

$$\begin{aligned}
|s_n t_n - st| &= |s_n t_n - st_n + st_n - st| && \text{because } -st_n + st_n = 0 \\
&= |t_n(s_n - s) + s(t_n - t)| && \text{by distributivity and commutativity} \\
&\leq |t_n(s_n - s)| + |s(t_n - t)| && \text{by the triangle inequality} \\
&= |t_n|\,|s_n - s| + |s|\,|t_n - t| && \text{by Theorem 9.2.3(e)} \\
&\leq M\,|s_n - s| + M\,|t_n - t| && \text{by (9.11) and } K, |s| \leq M.
\end{aligned}$$

So if $|s_n - s| < \frac{\varepsilon}{2M}$ and $|t_n - t| < \frac{\varepsilon}{2M}$, then we can conclude that $|s_n t_n - st| < \varepsilon$. One can now compose a logically correct proof guided by Proof Strategy 9.5.1. Ⓐ

Proof. See Exercise 7. □

Our next lemma states that if the limit of a convergent sequence is nonzero and all the terms in the sequence are also nonzero, then there is a single positive number γ smaller than the absolute values of the limit and of all the terms in the sequence.

Lemma 9.5.4. *Suppose that* $\lim_{n\to\infty} t_n = t$ *where* $t \neq 0$ *and* $t_n \neq 0$ *for all* $n \geq 1$. *Then there is a real number* $\gamma > 0$ *such that* $\gamma \leq |t|$ *and* $\gamma \leq |t_n|$ *for all* $n \geq 1$.

[2]The algebraic "trick," of adding and subtracting the same value, is used often in analysis.

Proof Analysis. To better understand how to prove this lemma, we first construct the following proof diagram:

Assume $\lim_{n\to\infty} t_n = t$.
Assume $t \neq 0$ and $t_n \neq 0$ for all $n \geq 1$.
Let $\gamma =$ (the positive real number you found).
Prove $\gamma \leq |t_n|$ for all $n \geq 1$.
Prove $\gamma \leq |t|$.

Since $\lim_{n\to\infty} t_n = t$, we can make $|t_n - t|$ as small as we would like. To find a positive γ such that $\gamma \leq |t_n|$ for all $n \geq 1$, we shall first start with $|t_n|$ and extract out a smaller value involving $|t_n - t|$. This can be done using the backward triangle inequality (see Exercise 4 on page 300) as follows: $|t_n| = |t + (t_n - t)| \geq |t| - |t_n - t|$. Thus,

$$|t| - |t_n - t| \leq |t_n|\,. \qquad (\star)$$

We know that $|t|$ is positive. So to ensure that $|t| - |t_n - t|$ is positive, we will need $|t_n - t|$ to be smaller than $|t|$, say $\frac{|t|}{2}$. Since $\lim_{n\to\infty} t_n = t$, there is a natural number N such that $|t - t_n| < \frac{|t|}{2}$ for all $n > N$. Using properties of inequality, we can conclude from $(\star)$ that $\frac{|t|}{2} < |t_n|$ for all $n > N$. In addition, we have that $\frac{|t|}{2} < |t|$. Thus our positive value for γ should satisfy $\gamma \leq \frac{|t|}{2}$. To make sure that $\gamma \leq |t_n|$ for *all* $n \geq 1$, we will use $\gamma = \min\{|t_1|, |t_2|, \ldots, |t_N|, \frac{|t|}{2}\}$. Ⓐ

Proof. Suppose that $\lim_{n\to\infty} t_n = t$ where $t \neq 0$ and $t_n \neq 0$ for all $n \geq 1$. We prove that there is a real number $\gamma > 0$ such that $\gamma \leq |t|$ and $\gamma \leq |t_n|$ for all $n \geq 1$. Because $\lim_{n\to\infty} t_n = t$ and $\frac{|t|}{2} > 0$, there is an $N \in \mathbb{N}$ such that

$$|t - t_n| < \frac{|t|}{2} \text{ for all } n > N. \qquad (9.12)$$

Let $\gamma = \min\{|t_1|, |t_2|, \ldots, |t_N|, \frac{|t|}{2}\}$. Hence, $\gamma > 0$ and $\gamma \leq \frac{|t|}{2} < |t|$. If $n \leq N$, then clearly $\gamma \leq |t_n|$. Suppose $n > N$. We establish the inequality $|t_n| > \frac{|t|}{2}$ as follows:

$$\begin{aligned}
|t_n| &= |t + (t_n - t)| && \text{because } t - t = 0 \\
&\geq |t| - |t_n - t| && \text{by the backward triangle inequality} \\
&> |t| - \frac{|t|}{2} && \text{by (9.12)} \\
&= \frac{|t|}{2} && \text{by algebra.}
\end{aligned}$$

Thus, $|t_n| > \frac{|t|}{2}$. Since $\gamma \leq \frac{|t|}{2}$ and $\frac{|t|}{2} < |t_n|$, we have $\gamma \leq |t_n|$. □

Our next theorem shows that whenever a convergent sequence $\langle t_n \rangle$ satisfies the conditions of Lemma 9.5.4, then the reciprocal sequence $\left\langle \frac{1}{t_n} \right\rangle$ also converges.

Theorem 9.5.5. *Suppose that $t \neq 0$ and $t_n \neq 0$ for all $n \geq 1$. If $\lim_{n\to\infty} t_n = t$, then $\lim_{n\to\infty} \frac{1}{t_n} = \frac{1}{t}$.*

Proof Analysis. We are assuming that $\lim_{n\to\infty} t_n = t$. So we can make $|t_n - t|$ as small as we want. Let $\varepsilon > 0$. We need to make $\left|\frac{1}{t_n} - \frac{1}{t}\right| < \varepsilon$. Since $t \neq 0$ and $t_n \neq 0$ for all $n \geq 1$, we know by Lemma 9.5.4 that there is a real number $\gamma > 0$ such that

$$\gamma \leq |t| \text{ and } \gamma \leq |t_n| \text{ for all } n \geq 1. \qquad (\star)$$

We use algebra and properties of inequality on the expression $\left|\frac{1}{t_n} - \frac{1}{t}\right|$ to extract out a larger value that contains $|t_n - t|$, and no other factor involving $|t_n|$ as follows:

$$\left|\frac{1}{t_n} - \frac{1}{t}\right| = \left|\frac{t - t_n}{t_n t}\right| = \frac{|t - t_n|}{|t_n|\,|t|} \leq \frac{|t - t_n|}{\gamma^2}$$

where the inequality follows from $(\star)$. So if $|t_n - t| < \gamma^2\varepsilon$, then we can conclude that $\left|\frac{1}{t_n} - \frac{1}{t}\right| < \varepsilon$. We can now present a logically correct proof directed by Proof Strategy 9.5.1. Ⓐ

Proof. Assume that $\lim_{n\to\infty} t_n = t$ where $t \neq 0$ and $t_n \neq 0$ for all $n \geq 1$. By Lemma 9.5.4, there is a real number $\gamma > 0$ such that $\gamma \leq |t|$ and $\gamma \leq |t_n|$ for all $n \geq 1$. So,

$$\gamma^2 \leq |t_n t| \text{ for all } n \geq 1. \qquad (9.13)$$

To prove that $\lim_{n\to\infty} \frac{1}{t_n} = \frac{1}{t}$, let $\varepsilon > 0$. Since $\lim_{n\to\infty} t_n = t$, there is an $N \in \mathbb{N}$ such that

$$|t_n - t| < \varepsilon\gamma^2 \text{ for all } n > N. \qquad (9.14)$$

For $n > N$ we have

$$\begin{aligned}
\left|\frac{1}{t_n} - \frac{1}{t}\right| = \left|\frac{t - t_n}{t_n t}\right| &= \frac{|t - t_n|}{|t_n|\,|t|} && \text{common denominator and Theorem 9.2.3(e, f)} \\
&\leq \frac{|t - t_n|}{\gamma^2} && \text{by (9.13)} \\
&< \frac{\varepsilon\gamma^2}{\gamma^2} && \text{by (9.14)} \\
&= \varepsilon && \text{by algebra.}
\end{aligned}$$

Hence, $\left|\frac{1}{t_n} - \frac{1}{t}\right| < \varepsilon$. Therefore, $\lim_{n\to\infty} \frac{1}{t_n} = \frac{1}{t}$. □

Theorems 9.5.3 and 9.5.5 imply our next theorem.

Theorem 9.5.6. *Suppose that* $\lim_{n\to\infty} s_n = s$ *and* $\lim_{n\to\infty} t_n = t$. *If* $t \neq 0$ *and* $t_n \neq 0$ *for all* $n \geq 1$, *then* $\lim_{n\to\infty} \frac{s_n}{t_n} = \frac{s}{t}$.

Proof. Suppose that $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$, where $t \neq 0$ and $t_n \neq 0$ for all $n \geq 1$. Since $\frac{s_n}{t_n} = s_n\left(\frac{1}{t_n}\right)$, Theorems 9.5.5 and 9.5.3 imply that $\lim_{n\to\infty} \frac{s_n}{t_n} = \frac{s}{t}$. □

Theorem 9.5.7 (Squeeze Theorem). *Let* $\langle s_n\rangle$, $\langle y_n\rangle$ *and* $\langle t_n\rangle$ *be sequences such that* $\lim_{n\to\infty} s_n = \ell$ *and* $\lim_{n\to\infty} t_n = \ell$. *If* $s_n \leq y_n \leq t_n$ *for all* $n \geq 1$, *then* $\lim_{n\to\infty} y_n = \ell$.

Proof. We have that $\lim_{n\to\infty} s_n = \ell$ and $\lim_{n\to\infty} t_n = \ell$. Assume $s_n \leq y_n \leq t_n$ for all $n \geq 1$. To prove that $\lim_{n\to\infty} y_n = \ell$, let $\varepsilon > 0$. Since $\lim_{n\to\infty} s_n = \ell$, there is an N_s such that

$$|s_n - \ell| < \varepsilon \text{ for all } n > N_s. \tag{9.15}$$

Similarly, since $\lim_{n\to\infty} t_n = \ell$, there is an N_t such that

$$|t_n - \ell| < \varepsilon \text{ for all } n > N_t. \tag{9.16}$$

Let $N = \max\{N_s, N_t\}$ and let $n > N$. We shall prove that $|y_n - \ell| < \varepsilon$. By assumption, we have $s_n \leq y_n \leq t_n$ and so, $s_n - \ell \leq y_n - \ell \leq t_n - \ell$. Lemma 9.2.5 implies that $|y_n - \ell| \leq \max\{|s_n - \ell|, |t_n - \ell|\}$. Since $n > N_s$ and $n > N_t$, it follows from (9.15) and (9.16) that $\max\{|s_n - \ell|, |t_n - \ell|\} < \varepsilon$. Therefore, $|y_n - \ell| < \varepsilon$. □

Exercises 9.5

1. Suppose $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$. Using Proof Strategy 9.5.1 as a guide, prove that $\lim_{n\to\infty}(s_n - t_n) = s - t$.

2. Suppose $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$. Let $a, b \in \mathbb{R}$ be nonzero. Using Proof Strategy 9.5.1, prove that $\lim_{n\to\infty}(as_n + bt_n) = as + bt$.

3. Assume $\lim_{n\to\infty} x_n = c$. Suppose that a sequence $\langle y_k\rangle$ satisfies $|x_k - y_k| < \frac{1}{k}$ for all $k \geq 1$. Prove that $\lim_{n\to\infty} y_n = c$.

4. Let $\langle a_n\rangle$ and $\langle b_n\rangle$ be sequences, and let $\ell, m \in \mathbb{R}$. Suppose that $\lim_{n\to\infty}(a_n + b_n) = \ell$ and $\lim_{n\to\infty}(a_n - b_n) = m$. Use Theorem 9.5.2 to prove that $\langle a_n\rangle$ converges and to evaluate the limit $\lim_{n\to\infty} a_n$.

5. Let $\langle a_n \rangle$ and $\langle b_n \rangle$ be sequences, and let $\ell, m \in \mathbb{R}$. Assume $\lim_{n\to\infty}(a_n + b_n) = \ell$ and $\lim_{n\to\infty}(a_n - b_n) = m$. Using Exercise 1, prove that $\langle b_n \rangle$ converges and evaluate the limit $\lim_{n\to\infty} b_n$.

6. Suppose that $\langle a_n \rangle$ is bounded and that $\lim_{n\to\infty} b_n = 0$. Prove that $\lim_{n\to\infty} a_n b_n = 0$.

7. Prove Theorem 9.5.3.

8. Suppose $\lim_{n\to\infty} s_n = c$ and let $\sigma : \mathbb{N} \to \mathbb{N}$ be one-to-one. Prove that $\lim_{n\to\infty} s_{\sigma(n)} = c$.

Exercise Notes: For Exercise 7, use Proof Strategy 9.5.1 and the proof analysis on page 325 as a guide. For Exercise 8, observe that the set $\{n \in \mathbb{N} : \sigma(n) \leq N\}$ is finite for any $N \in \mathbb{N}$, because σ is one-to-one.

9.6 Continuous Functions

What does it mean to say that a function $f : D \to \mathbb{R}$ is continuous? Intuitively, a continuous function is one for which a small change in the input results in a small change in the output. In this section we will be presuming that $D \subseteq \mathbb{R}$. Here is the precise definition which is illustrated in Fig. 9.9.

Definition 9.6.1. Let $f : D \to \mathbb{R}$ and let $c \in D$. We say that f is **continuous** at c when the following holds: For every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x \in D$ if $|x - c| < \delta$, then $|f(x) - f(c)| < \varepsilon$.

The logical form of Definition 9.6.1 can be expressed as

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in D)(|x - c| < \delta \to |f(x) - f(c)| < \varepsilon) \tag{9.17}$$

and it is this logical form that drives our next proof strategy.

Proof Strategy 9.6.2. Let $f : D \to \mathbb{R}$ and $c \in D$. To prove that f is continuous at c use the proof diagram:

Let $\varepsilon > 0$ be a real number.
Let $\delta =$ (the positive real number found).
Let x be a real number in D.
Assume $|x - c| < \delta$.
Prove $|f(x) - f(c)| < \varepsilon$.

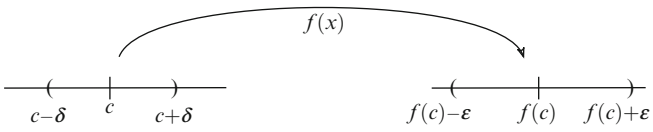


Fig. 9.9 Illustration for Definition 9.6.1

To apply Proof Strategy 9.6.2 to a specific function f, first let $\varepsilon > 0$. We must find a $\delta > 0$ such that when $|x-c| < \delta$, we can then prove that $|f(x)-f(c)| < \varepsilon$. To find the desired δ, we use algebra and properties of inequality on the expression $|f(x)-f(c)|$ to "extract out" a larger value containing $|x-c|$ and no other occurrences of x. We should then be able to find δ so that when $|x-c| < \delta$, we will have that $|f(x)-f(c)| < \varepsilon$.

In the proof analysis of our next proposition we shall apply the above procedure to find δ; moreover, in the proof of the proposition we will not show the reader how we found δ. At the beginning of the proof we shall let $\varepsilon > 0$ and then just identify the value for δ that we will use to complete the proof.

Proposition 9.6.3. *Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x+5$ and let $c \in \mathbb{R}$. Then f is continuous at c.*

Proof Analysis. Let $c \in \mathbb{R}$ and let $\varepsilon > 0$. We need to find a $\delta > 0$ so that when $|x-c| < \delta$, we can conclude that $|f(x)-f(c)| < \varepsilon$. To find this δ, we use algebra and properties of inequality on the expression $|f(x)-f(c)|$ to extract out $|x-c|$ as follows:

$$\begin{aligned} |f(x)-f(c)| &= |(3x+5)-(3c+5)| && \text{by definition of } f \\ &= |3(x-c)| && \text{by algebra} \\ &= 3\,|x-c| && \text{by Theorem 9.2.3(e).} \end{aligned}$$

Thus, if we have $3\,|x-c| < \varepsilon$, then we will be able to deduce that $|f(x)-f(c)| < \varepsilon$. Solving the inequality $3\,|x-c| < \varepsilon$ for $|x-c|$ we obtain $|x-c| < \frac{\varepsilon}{3}$. So, we will let $\delta = \frac{\varepsilon}{3}$. We now give a correct proof, guided by Proof Strategy 9.6.2. Ⓐ

Proof. Let $f(x) = 3x+5$ and let $c \in \mathbb{R}$. We shall prove that f is continuous at c. Let $\varepsilon > 0$ be given. Now, let $\delta = \frac{\varepsilon}{3}$ and let $x \in \mathbb{R}$ satisfy $|x-c| < \delta$. We prove $|f(x)-f(c)| < \varepsilon$ as follows:

$$\begin{aligned} |f(x)-f(c)| &= |(3x+5)-(3c+5)| && \text{by definition of } f \\ &= |3(x-c)| && \text{by algebra} \\ &= 3\,|x-c| && \text{by Theorem 9.2.3(e)} \\ &< 3\delta && \text{because } |x-c| < \delta \\ &= 3\frac{\varepsilon}{3} && \text{because } \delta = \frac{\varepsilon}{3} \\ &= \varepsilon && \text{by algebra.} \end{aligned}$$

Hence, $|f(x)-f(c)| < \varepsilon$ and therefore f is continuous at c. □

In the proof of Proposition 9.6.3 we were able to 'cleanly' extract out $|x-c|$; however, this may not be the case when applying Strategy 9.6.2 in the proof of other such results. Sometimes, after extracting out $|x-c|$, there will be some 'unwanted' factors involving x. If such factors appear, then we will have to find an appropriate upper bound for these factors. We will execute this idea in our next proof analysis.

Proposition 9.6.4. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$ and let $c \in \mathbb{R}$. Then f is continuous at c.*

Proof Analysis. Let $c \in \mathbb{R}$ and $\varepsilon > 0$. We need a $\delta > 0$ so that when $|x-c| < \delta$, we can conclude that $|f(x) - f(c)| < \varepsilon$. To find this δ, we use algebra and properties of inequality on the expression $|f(x) - f(c)|$ to extract out $|x-c|$ as follows:

$$\begin{aligned} |f(x) - f(c)| &= \left|x^2 - c^2\right| && \text{by definition of } f \\ &= |(x+c)(x-c)| && \text{by factoring} \\ &= |x+c|\,|x-c| && \text{by Theorem 9.2.3(e).} \end{aligned} \tag{$\star$}$$

In this case, we end up with the unwanted factor $|x+c|$, which involves x. We need to find an upper bound for this factor. How can we do this? We know that we can make $|x-c|$ as small as we would like. Thus, to obtain an upper bound for $|x+c|$, we start with $|x+c|$ and extract out a larger value containing $|x-c|$ as follows:

$$\begin{aligned} |x+c| &= |x-c+2c| && \text{because } c = -c+2c \\ &\leq |x-c| + 2\,|c| && \text{by the triangle inequality.} \end{aligned}$$

So if we have $|x-c| < 1$, then we will be able to derive $|x+c| \leq 1 + 2\,|c|$ and thus, $M = 1 + 2\,|c|$ will be our desired upper bound. Hence, if we let $\delta \leq 1$ and $|x-c| < \delta$, then we can continue ($\star$) above to obtain $|f(x) - f(c)| = |x+c|\,|x-c| \leq M\,|x-c|$ and thus, we have cleanly extracted out $|x-c|$ from $|f(x) - f(c)|$. Furthermore, if we have $M\,|x-c| < \varepsilon$, then we can conclude that $|f(x) - f(c)| < \varepsilon$. Solving the inequality $M\,|x-c| < \varepsilon$ for $|x-c|$, we obtain $|x-c| < \frac{\varepsilon}{M}$. So we will also need $\delta \leq \frac{\varepsilon}{M}$. Since we will require that $\delta \leq 1$ and $\delta \leq \frac{\varepsilon}{M}$ in our proof, we shall use $\delta = \min\{1, \frac{\varepsilon}{M}\}$. We now have all of the ingredients that are needed to prove the theorem. Ⓐ

Proof. Let $f(x) = x^2$ and let $c \in \mathbb{R}$. Let $\varepsilon > 0$ be given and let $\delta = \min\{1, \frac{\varepsilon}{M}\}$ where $M = 1 + 2\,|c|$. Let $x \in \mathbb{R}$ satisfy $|x-c| < \delta$. Since $|x-c| < \delta \leq 1$, it follows that

$$\begin{aligned} |x+c| &= |x-c+2c| && \text{because } c = -c+2c \\ &\leq |x-c| + 2\,|c| && \text{by the triangle inequality} \\ &< 1 + 2\,|c| = M && \text{because } |x-c| < 1. \end{aligned}$$

Therefore, $(\star)$ $|x+c| \leq M$. We prove that $|f(x)-f(c)| < \varepsilon$ as follows:

$$\begin{aligned}
|f(x)-f(c)| &= \left|x^2-c^2\right| && \text{by definition of } f \\
&= |(x+c)(x-c)| && \text{by factoring} \\
&= |x+c|\,|x-c| && \text{by Theorem 9.2.3(e)} \\
&\leq M\,|x-c| && \text{by } (\star) \\
&< M\delta && \text{because } |x-c| < \delta \\
&\leq M\frac{\varepsilon}{M} && \text{because } \delta \leq \frac{\varepsilon}{M} \\
&= \varepsilon && \text{by algebra.}
\end{aligned}$$

Thus, $|f(x)-f(c)| < \varepsilon$. Therefore, f is continuous at c. □

The logical form of the definition of continuous at a point is given in (9.17). By taking the negation of this logical form and using logic laws, we obtain our next remark which expresses what it means for a function not to be continuous at a point.

Remark 9.6.5. Let $f\colon D \to \mathbb{R}$ and let $c \in D$. Then f is **not** continuous at c if and only if *there exists an* $\varepsilon > 0$ *such that for all* $\delta > 0$*, there is an* $x \in D$ *such that* $|x-c| < \delta$ *and* $|f(x)-f(c)| \geq \varepsilon$.

9.6.1 Algebraic Operations on Functions

Given two functions $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$, we can form new functions by adding, subtracting, multiplying, and dividing the values of f and g.

Definition 9.6.6. Let $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ be functions. We can define five new functions as follows:

(1) The **sum** $(f+g)\colon D \to \mathbb{R}$ is defined by $(f+g)(x) = f(x)+g(x)$ for all $x \in D$.
(2) The **difference** $(f-g)\colon D \to \mathbb{R}$ is defined by $(f-g)(x) = f(x)-g(x)$ for all $x \in D$.
(3) The **product** $(fg)\colon D \to \mathbb{R}$ is defined by $(fg)(x) = f(x)g(x)$ for all $x \in D$.
(4) For $k \in \mathbb{R}$, the **constant multiple** $(kf)\colon D \to \mathbb{R}$ is defined by $(kf)(x) = kf(x)$ for all $x \in D$.
(5) Suppose $g(x) \neq 0$ for all $x \in D$. The **quotient** $\left(\frac{f}{g}\right)\colon D \to \mathbb{R}$ is defined by $\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$ for all $x \in D$.

We will show, assuming f and g are continuous at a point c, that the operations defined in Definition 9.6.6 will each produce a function that is also continuous at c. Thus, the operations (1)–(5) preserve continuity. How can we take advantage of an assumption stating that a function is continuous? The next strategy will be used when working with such an assumption.

Assumption Strategy 9.6.7. Let $f: D \to \mathbb{R}$ and let $c \in D$. Suppose you are *assuming* that f is continuous at c. Then for any $\varepsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ whenever $x \in D$ satisfies $|x - c| < \delta$.

Given that f is continuous at c, using Assumption Strategy 9.6.7, we can conclude that for any positive value $v > 0$, there is a $d > 0$ such that whenever $x \in D$ and $|x - c| < d$, we will have that $|f(x) - f(c)| < v$. We shall express this observation as "we can make $|f(x) - f(c)|$ as small as we want."

9.6.2 Preservation-of-Continuity Theorems

How can one prove theorems that have the following form?

Theorem. *Let $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ be functions and let $c \in D$. Suppose that f is continuous at c and g is continuous at c. Then the function $h: D \to \mathbb{R}$, constructed from f and g, is also continuous at c.*

In a proof of a theorem having the above form, we would assume that f and g are continuous at c and then we would have to prove that h is continuous at c. To prove such a theorem, we shall apply the following strategy.

Proof Strategy 9.6.8. To prove that h is continuous at c, use the proof diagram:

Assume f and g are continuous at c.
Let $\varepsilon > 0$ be a real number.
  Let $\delta =$ (the positive real number found).
    Let x be a real number in D.
      Assume $|x - c| < \delta$.
        Prove $|h(x) - h(c)| < \varepsilon$.

To apply Proof Strategy 9.6.8, let $\varepsilon > 0$. We must find a $\delta > 0$ such that $|h(x) - h(c)| < \varepsilon$ when $|x - c| < \delta$. We will use the assumption that f and g are continuous at c, to find the desired δ. Since f and g are continuous at c, we can make $|f(x) - f(c)|$ and $|g(x) - g(c)|$ "as small as we want." Here is the basic idea that we will apply to get δ.

> *Using algebra and properties of inequality on the expression $|h(x) - h(c)|$, extract out a larger value containing $|f(x) - f(c)|$ and $|g(x) - g(c)|$, and no other occurrences of x, $f(x)$ or $g(x)$.*

Since we can make $|f(x) - f(c)|$ and $|g(x) - g(c)|$ as small as we want, we should then be able to find δ. We will use these ideas in our proof analysis of the next two theorems. Our first theorem shows that sum of two continuous functions is also continuous. The second theorem shows that the reciprocal of a nonzero continuous function is continuous as well.

Theorem 9.6.9. *Let $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ be functions and let $c \in D$. Suppose that f and g are continuous at c. Then $(f+g)$ is continuous at c.*

Proof Analysis. Let $\varepsilon > 0$. We are given that f and g are continuous at c. So we can make $|f(x)-f(c)|$ and $|g(x)-g(c)|$ as small as we want. We need to make $|(f+g)(x)-(f+g)(c)| < \varepsilon$. Using algebra and properties of inequality, we extract out $|f(x)-f(c)|$ and $|g(x)-g(c)|$ from the expression $|(f+g)(x)-(f+g)(c)|$ as follows:

$$\begin{aligned} |(f+g)(x)-(f+g)(c)| &= |(f(x)+g(x))-(f(c)+g(c))| && \text{by definition of } f+g \\ &= |(f(x)-f(c))+(g(x)-g(c))| && \text{by algebra} \\ &\leq |f(x)-f(c)|+|g(x)-g(c)| && \text{by triangle inequality.} \end{aligned}$$

So if we have that $|f(x)-f(c)| < \frac{\varepsilon}{2}$ and $|g(x)-g(c)| < \frac{\varepsilon}{2}$, then we can deduce

$$|(f+g)(x)-(f+g)(c)| < \varepsilon.$$

We now present a logically correct proof directed by Proof Strategy 9.6.8. Ⓐ

Proof. Let $f: D \to \mathbb{R}$ and $g: D \to \mathbb{R}$ be functions. Let $c \in D$. Assume that f and g are continuous at c. We shall prove that $f+g$ is continuous at c. Let $\varepsilon > 0$. Since f is continuous at c, there is a $\delta_1 > 0$ such that for all $x \in D$ we have

$$|f(x)-f(c)| < \frac{\varepsilon}{2} \text{ when } |x-c| < \delta_1. \tag{9.18}$$

Because g is continuous at c, there is a $\delta_2 > 0$ such that for all $x \in D$ we have

$$|g(x)-g(c)| < \frac{\varepsilon}{2} \text{ when } |x-c| < \delta_2. \tag{9.19}$$

Let $\delta = \min\{\delta_1, \delta_2\}$ and $x \in D$ satisfy $|x-c| < \delta$. So $|x-c| < \delta_1$ and $|x-c| < \delta_2$. So, (9.18) and (9.19) apply. We prove that $|(f+g)(x)-(f+g)(c)| < \varepsilon$ as follows:

$$\begin{aligned} |(f+g)(x)-(f+g)(c)| &= |(f(x)+g(x))-(f(c)+g(c))| && \text{by definition of } f+g \\ &= |(f(x)-f(c))+(g(x)-g(c))| && \text{by algebra} \\ &\leq |f(x)-f(c)|+|g(x)-g(c)| && \text{by triangle inequality} \\ &< \frac{\varepsilon}{2}+\frac{\varepsilon}{2} && \text{by (9.18) and (9.19)} \\ &= \varepsilon && \text{by algebra} \end{aligned}$$

and thus, $|(f+g)(x)-(f+g)(c)| < \varepsilon$. Therefore, $(f+g)$ is continuous at c. □

In the proof of Theorem 9.6.9 we were able to 'cleanly' extract out $|f(x)-f(c)|$ and $|g(x)-g(c)|$, that is, there were no additional factors. When applying Proof

Strategy 9.6.8, after extracting out $|f(x)-f(c)|$ or $|g(x)-g(c)|$, there may be some 'unwanted' factors involving x, $f(x)$ or $g(x)$. One must then find an appropriate upper bound for these factors. This will be done in our proofs of Theorems 9.6.11 and 9.6.13. In the proof of Theorem 9.6.11, we will obtain the unwanted factor $\frac{1}{|g(x)|}$. Our next lemma yields an upper bound for this factor.

Lemma 9.6.10. *Let $g\colon D \to \mathbb{R}$ be continuous at $c \in D$ where $g(c) \neq 0$. Then there exist an $m > 0$ and a $\delta > 0$ such that for all $x \in D$ if $|x-c| < \delta$, then $|g(x)| > m$.*

Proof Analysis. To better understand how to prove this lemma, we first construct the proof diagram:

Assume g is continuous at c.
Assume $g(c) \neq 0$.
  Let $\delta =$ (the positive real number found).
  Let $m =$ (the positive real number you found).
    Let x be a real number in D.
      Assume $|x-c| < \delta$.
        Prove $|g(x)| > m$.

Since g is continuous at c, we can make $|g(x)-g(c)|$ as small as we would like. To find a positive δ and a positive m such that $|g(x)| > m$ when $|x-c| < \delta$, we shall first start with $|g(x)|$ and extract out a smaller value involving $|g(x)-g(c)|$. This can be done using the backward triangle inequality (see Exercise 4 on page 300) as follows: $|g(x)| = |g(c)+(g(x)-g(c))| \geq |g(c)| - |g(x)-g(c)|$. Thus,

$$|g(x)| \geq |g(c)| - |g(x)-g(c)| \qquad (\star)$$

Since $|g(c)| > 0$, to ensure that $|g(c)| - |g(x)-g(c)|$ is positive, we would need $|g(x)-g(c)|$ to be smaller than $|g(c)|$, say $\frac{|g(c)|}{2}$. Because g is continuous at c, there is a $\delta > 0$ such that $|g(x)-g(c)| < \frac{|g(c)|}{2}$ when $|x-c| < \delta$. We can conclude from $(\star)$ that $|g(x)| > \frac{|g(c)|}{2}$ when $|x-c| < \delta$. So we shall use $m = \frac{|g(c)|}{2}$ in our proof. Ⓐ

Proof. Let $g\colon D \to \mathbb{R}$ be continuous at $c \in D$ where $g(c) \neq 0$. Consider the positive value $\frac{|g(c)|}{2}$. Because g is continuous at c, there exists a $\delta > 0$ such that

$$|g(x)-g(c)| < \frac{|g(c)|}{2} \text{ when } x \in D \text{ and } |x-c| < \delta. \qquad (9.20)$$

Let $x \in D$ satisfy $|x-c| < \delta$. We shall show that $|g(x)| > \frac{|g(c)|}{2}$ as follows:

$$\begin{aligned} |g(x)| &= |g(c)+(g(x)-g(c))| && \text{because } g(c)-g(c)=0 \\ &\geq |g(c)| - |g(x)-g(c)| && \text{by the backward triangle inequality} \end{aligned}$$

$$> |g(c)| - \frac{|g(c)|}{2} \qquad \text{by (9.20)}$$

$$= \frac{|g(c)|}{2} \qquad \text{by algebra.}$$

Therefore, for $m = \frac{|g(c)|}{2}$ we have $|g(x)| > m$ whenever $x \in D$ and $|x-c| < \delta$. □

Theorem 9.6.11. *Let $g\colon D \to \mathbb{R}$ be such that $g(x) \neq 0$ for all $x \in D$. Suppose that g is continuous at $c \in D$. Then $\left(\frac{1}{g}\right)$ is continuous at c.*

Proof Analysis. Let $\varepsilon > 0$. We are given that g is continuous at c. So we can make $|g(x) - g(c)|$ as small as we want. We need to make $\left|\frac{1}{g(x)} - \frac{1}{g(c)}\right| < \varepsilon$. Since g is continuous at c, Lemma 9.6.10 implies that there is a $\delta_1 > 0$ and an $m > 0$ such that

$$|g(x)| > m \text{ when } |x-c| < \delta_1 \text{ and } x \in D. \qquad (\star)$$

Note that when $|g(x)| > m$, we have that $\frac{1}{|g(x)|} < \frac{1}{m}$. Also observe that $(\star)$ implies that $|g(c)| > m$, because $|c-c| = 0 < \delta_1$. Thus, $\frac{1}{|g(c)|} < \frac{1}{m}$. Let $x \in D$ and assume $|x-c| < \delta_1$. We extract out $|g(x) - g(c)|$ from the expression $\left|\frac{1}{g(x)} - \frac{1}{g(c)}\right|$ as follows:

$$\left|\frac{1}{g(x)} - \frac{1}{g(c)}\right| = \left|\frac{g(c) - g(x)}{g(x)g(c)}\right| = \frac{|g(x) - g(c)|}{|g(x)|\,|g(c)|} \leq \frac{|g(x) - g(c)|}{m^2}.$$

So, if $|g(x) - g(c)| < m^2\varepsilon$, then we will be able to deduce $\left|\frac{1}{g(x)} - \frac{1}{g(c)}\right| < \varepsilon$. We can now give a logically correct proof using Proof Strategy 9.6.8 as a guide. Ⓐ

Proof. Let $g\colon D \to \mathbb{R}$ and let $c \in D$. Suppose g is continuous at c and $g(c) \neq 0$. Lemma 9.6.10 implies that there is a $\delta_1 > 0$ and an $m > 0$ such that

$$|g(x)| > m \text{ when } |x-c| < \delta_1 \text{ and } x \in D. \qquad (9.21)$$

To prove that $\left(\frac{1}{g}\right)$ is continuous at c, let $\varepsilon > 0$. Since g is continuous at c, there is a δ_2 such that

$$|g(x) - g(c)| < \varepsilon m^2 \text{ when } |x-c| < \delta_2 \text{ and } x \in D. \qquad (9.22)$$

Let $\delta = \min\{\delta_1, \delta_2\}$ and $x \in D$ be such that $|x-c| < \delta$. Thus $|x-c| < \delta_1$ and $|x-c| < \delta_2$. Hence, (9.21) and (9.22) apply. We prove $\left|\frac{1}{g(x)} - \frac{1}{g(c)}\right| < \varepsilon$ as follows:

$$\begin{aligned}
\left|\frac{1}{g(x)}-\frac{1}{g(c)}\right| &= \left|\frac{g(c)-g(x)}{g(x)g(c)}\right| && \text{common denominator}\\
&= \frac{|g(x)-g(c)|}{|g(x)|\,|g(c)|} && \text{by Theorem 9.2.3(e, f)}\\
&\leq \frac{|g(x)-g(c)|}{m^2} && \text{by (9.21)}\\
&< \frac{\varepsilon m^2}{m^2} && \text{by (9.22)}\\
&= \varepsilon && \text{by algebra}
\end{aligned}$$

and so $\left|\frac{1}{g(x)}-\frac{1}{g(c)}\right| < \varepsilon$. Therefore, $\left(\frac{1}{g}\right)$ is continuous at c. □

Suppose that f and g are continuous at a point c. To prove that fg is continuous at c, one would start with $|(fg)(x)-(fg)(c)|$ and then extract out $|f(x)-f(c)|$ and $|g(x)-g(c)|$. To do this, one can get $|f(x)|$ as an unwanted factor. Our next lemma can be used to obtain an upper bound for this factor.

Lemma 9.6.12. *Let $f\colon D \to \mathbb{R}$ and let $c \in D$. If f is continuous at c, then there is a $\delta > 0$ and an $M > 0$ such that $|f(x)| \leq M$ for all $x \in D$ satisfying $|x-c| < \delta$.*

Proof. Let $f\colon D \to \mathbb{R}$ be continuous at $c \in D$. So there exists a $\delta > 0$ such that

$$|f(x)-f(c)| < 1 \text{ whenever } x \in D \text{ and } |x-c| < \delta. \tag{9.23}$$

Let $x \in D$ satisfy $|x-c| < \delta$. Thus,

$$\begin{aligned}
|f(x)| &= |(f(x)-f(c))+f(c)| && \text{because } -f(c))+f(c)=0\\
&\leq |f(x)-f(c)|+|f(c)| && \text{by the triangle inequality}\\
&< 1+|f(c)| && \text{by (9.23).}
\end{aligned}$$

Let $M = 1+|f(c)|$. Therefore, $|f(x)| \leq M$ for all $x \in D$ satisfying $|x-c| < \delta$. □

Theorem 9.6.13. *Suppose that $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ are continuous at $c \in D$. Then fg is continuous at c.*

Proof. See Exercise 11. □

Theorem 9.6.14. *Suppose that $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ are continuous at $c \in D$. If $g(c) \neq 0$, then $\left(\frac{f}{g}\right)$ is continuous at c.*

Proof. Suppose $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ are continuous at $c \in D$. Assume $g(c) \neq 0$. Theorem 9.6.11 implies that $\left(\frac{1}{g}\right)$ is continuous at c. Since $\left(\frac{f}{g}\right) = f \cdot \left(\frac{1}{g}\right)$, we see that $\left(\frac{f}{g}\right)$ is continuous at c by Theorem 9.6.13. □

Our next theorem shows that the operation of composition preserves continuity.

Theorem 9.6.15. *Let $f: D \to \mathbb{R}$ and $g: E \to \mathbb{R}$ be functions such that $f[D] \subseteq E$. Suppose f is continuous at $c \in D$ and g is continuous at $f(c)$. Then $(g \circ f): D \to \mathbb{R}$ is continuous at c.*

Proof. Let $f: D \to \mathbb{R}$ and $g: E \to \mathbb{R}$ be functions such that $f[D] \subseteq E$. Suppose that f is continuous at $c \in D$ and g is continuous at $f(c)$. We shall prove that $g \circ f$ is continuous at c. Let $\varepsilon > 0$. Since g is continuous at $f(c)$, there is a $\delta_1 > 0$ such that

$$|g(y) - g(f(c))| < \varepsilon \text{ for all } y \in E \text{ satisfying } |y - f(c)| < \delta_1. \tag{9.24}$$

Because f is continuous at c and $\delta_1 > 0$, there is a $\delta > 0$ such that

$$|f(x) - f(c)| < \delta_1 \text{ for all } x \in D \text{ satisfying } |x - c| < \delta. \tag{9.25}$$

Now, let $x \in D$ be such that $|x - c| < \delta$. Then

$$\begin{aligned} |(g \circ f)(x) - (g \circ f)(c)| &= |g(f(x)) - g(f(c))| && \text{by definition of } g \circ f \\ &< \varepsilon && \text{by (9.25) and (9.24).} \end{aligned}$$

Hence, $|(g \circ f)(x) - (g \circ f)(c)| < \varepsilon$. Therefore, $g \circ f$ is continuous at c. □

Our next lemma shows that when a continuous function is nonzero at a point, then the function is also nonzero in a neighborhood of the point.

Lemma 9.6.16. *Let $f: D \to \mathbb{R}$ be continuous at $c \in D$ and suppose $f(c) \neq 0$.*

(a) *If $f(c) > 0$, there is a $\delta > 0$ where $f(x) > 0$ for all $x \in D$ satisfying $|x - c| < \delta$.*
(b) *If $f(c) < 0$, there is a $\delta > 0$ where $f(x) < 0$ for all $x \in D$ satisfying $|x - c| < \delta$.*

Proof. Suppose $f: D \to \mathbb{R}$ is continuous at $c \in D$ and $f(c) \neq 0$. Let $\varepsilon = |f(c)| > 0$. Since f is continuous at c, there exists a $\delta > 0$ such that

$$|f(x) - f(c)| < \varepsilon \text{ when } x \in D \text{ and } |x - c| < \delta. \tag{9.26}$$

To prove (a), assume $f(c) > 0$. Thus $\varepsilon = f(c)$. Let $x \in D$ satisfy $|x - c| < \delta$. We will prove that $f(x) > 0$. Since $|x - c| < \delta$ and $\varepsilon = f(c)$, we conclude from (9.26) that $|f(x) - f(c)| < f(c)$. Because $f(c) - f(x) \leq |f(x) - f(c)|$, we obtain the inequality $f(c) - f(x) < f(c)$. Therefore, $f(c) - f(c) < f(x)$ and so, $f(x) > 0$. The proof of (a) is now complete. The proof of (b) is very similar (see Exercise 12). □

Definition 9.6.17. We say $f: D \to \mathbb{R}$ is **continuous** when f is continuous at every point $c \in D$.

Many of the important theorems in real analysis assume that a function is continuous at every point in its domain. For example, the fundamental theorem of calculus assumes that a function is continuous on a closed interval.

Proposition 9.6.18. *Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is continuous. Then the function $g\colon \mathbb{R} \to \mathbb{R}$ defined by $g(x) = 3f(x) - 2$ is also continuous.*

Proof. Assume that $f\colon \mathbb{R} \to \mathbb{R}$ is continuous. We will prove that g is continuous, where $g\colon \mathbb{R} \to \mathbb{R}$ is defined by $g(x) = 3f(x) - 2$. Let c be any real number and let $\varepsilon > 0$. Since f is continuous, there is a $\delta > 0$ such that

$$|f(x) - f(c)| < \frac{\varepsilon}{3} \text{ when } x \in \mathbb{R} \text{ and } |x - c| < \delta. \tag{9.27}$$

Assume $|x - c| < \delta$. We prove that $|g(x) - g(c)| < \varepsilon$ as follows:

$$\begin{aligned} |g(x) - g(c)| &= |(3f(x) - 2) + (3f(c) - 2)| && \text{by definition of } g \\ &= 3\,|f(x) - f(c)| && \text{by algebra and Theorem 9.2.3(e)} \\ &< 3\frac{\varepsilon}{3} = \varepsilon && \text{by (9.27) and algebra} \end{aligned}$$

Thus, $|g(x) - g(c)| < \varepsilon$. Therefore, g is continuous. □

Exercises 9.6

For Exercises 1–5 apply Proof Strategy 9.6.2. Apply Proof Strategy 9.6.8 in Exercises 7 to 11.

1. Let m and d be real numbers where $m \neq 0$. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = mx + d$. Prove that f is continuous.
2. Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = |x|$ for all $x \in \mathbb{R}$. Prove that f is continuous.
3. Let $f\colon [2,4] \to \mathbb{R}$ be defined by $f(x) = 3x^2 + 1$ for all $x \in [2,4]$. Prove that f is continuous.
4. Let $f\colon \mathbb{R}^+ \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$ for all $x \in \mathbb{R}^+$. Prove that f is continuous.
5. Let $f\colon [1,3] \to \mathbb{R}$ be defined by $f(x) = \frac{1}{x+2}$ for all $x \in [1,3]$. Prove that f is continuous.
6. Suppose that $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ are continuous at $c \in D$. Prove that there exist an $M > 0$ and a $\delta > 0$ such that $|f(x)| \leq M$ and $|g(x)| \leq M$ when $|x - c| < \delta$ and $x \in D$.
7. Let $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ be functions. Suppose that f and g are continuous at $c \in D$. Prove that $f - g$ is continuous at c.
8. Suppose that $f\colon D \to \mathbb{R}$ is continuous at $c \in D$ and $k \in \mathbb{R}$ is nonzero. Prove that kf is continuous at c.
9. Let $f\colon D \to \mathbb{R}$ be continuous at $c \in D$. Define $g\colon D \to \mathbb{R}$ by $g(x) = |f(x)|$ for all $x \in D$. Prove that g is continuous at c.

10. Let $f\colon D \to \mathbb{R}$ and $g\colon D \to \mathbb{R}$ be functions. Suppose g is continuous at $c \in D$ and $|f(x) - f(c)| \leq |g(x)|\,|x - c|$ for all $x \in D$. Prove that f is continuous at c.

11. Prove Theorem 9.6.13.

12. Prove Lemma 9.6.16(b).

13. Let $f\colon D \to \mathbb{R}$ be continuous at $c \in D$ and let $h \in \mathbb{R}$. Prove the following:

(a) If $f(c) > h$, then there is a $\delta > 0$ such that $f(x) > h$ whenever $x \in D$ and $|x - c| < \delta$.

(b) If $f(c) < h$, then there is a $\delta > 0$ such that $f(x) < h$ whenever $x \in D$ and $|x - c| < \delta$.

Exercise Notes: For Exercise 3, note that $|x + c| = x + c \leq 8$ when $x, c \in [2, 4]$. For Exercise 4, evaluate $(\sqrt{x} - \sqrt{c})\frac{\sqrt{x}+\sqrt{c}}{\sqrt{x}+\sqrt{c}}$. For Exercises 6 and 10, use Lemma 9.6.12. For Exercise 11, to prove that the function fg is continuous at c, use the identity $|f(x)g(x) - f(c)g(c)| = |f(x)g(x) - f(x)g(c) + f(x)g(c) - f(c)g(c)|$. Now extract out $|g(x) - g(c)|$ and $|f(x) - f(c)|$ and use Lemma 9.6.12. For the proof of Exercise 13, read the proof of Lemma 9.6.16.

APPENDIX

Summary of Strategies

Proof Strategies

1. To prove an algebraic equation, try one of the following:

(a) Transform one side of the equation into the other side of the equation.
(b) Derive the equation from any previously given, or assumed, equations.

2. To prove $P \to Q$, try one of the following:

(a) Assume P
Prove Q.

(b) Assume $\neg Q$
Prove $\neg P$.

3. To prove $P \wedge Q$, try the following:

Prove P
Prove Q.

4. To prove $P \vee Q$, try one of the following:

(a) Assume $\neg P$
Prove Q.

(b) Assume $\neg Q$
Prove P.

(c) Try using a division by cases. In each case, prove P or prove Q.

5. To prove $P \leftrightarrow Q$, try the following:

Prove $P \to Q$
Prove $Q \to P$.

6. To prove P by contradiction:

Assume $\neg P$
Derive "a contradiction.'

7. To prove $\forall x P(x)$, or $(\forall x \in A)P(x)$:

Let x, or respectively let $x \in A$, be arbitrary. Now prove $P(x)$.

8. To prove $\exists x P(x)$, or $(\exists x \in A)P(x)$, try one of the following:

(a) Find an x, or respectively an $x \in A$, such that $P(x)$ is true. Now prove $P(x)$.
(b) Prove $P(x)$ for some x, or respectively some $x \in A$.

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7,

9. To prove $\exists! x P(x)$:

First prove $\exists x P(x)$. Then, assuming $P(x)$ and $P(y)$, prove $x = y$.

Assumption Strategies

1. When assuming $P \to Q$:

(a) If you are assuming or can prove P, then you can conclude Q.
(b) If you are assuming or can prove $\neg Q$, then you can conclude $\neg P$.

2. When assuming $P \vee Q$:

(a) If required to prove R, try the following division by cases,

Case 1: Assume P.
Prove R.
Case 2: Assume Q.
Prove R.

(b) If you are assuming or can prove $\neg P$, then you can conclude Q.
(c) If you are assuming or can prove $\neg Q$, then you can conclude P.

3. When assuming $P \wedge Q$:

You can assume P and assume Q.

4. When assuming $P \leftrightarrow Q$:

(a) If you are assuming or can prove P, then you can conclude Q.
(b) If you are assuming or can prove Q, then you can conclude P.

5. When assuming $\neg P$:

(a) In a proof by contradiction try to prove P and thereby derive a contradiction.
(b) Reexpress $\neg P$ as a positive statement, and try to *use* the positive statement.

6. When assuming $\forall x P(x)$, or $(\forall x \in A) P(x)$:

You may plug in *any useful* value for x, or respectively $x \in A$; say a, and assume $P(a)$.

7. When assuming $\exists x P(x)$, or $(\exists x \in A) P(x)$:

Introduce a **new** variable x_0, or respectively $x_0 \in A$, representing an object which makes $P(x_0)$ true and then assume $P(x_0)$.

8. Any established theorem can be assumed and used in a proof.

Well-Ordering and Induction Proof Strategies

Let b be a fixed integer and let n be an integer variable.

1. To prove $(\forall n \geq b)\, P(n)$ by the well-ordering principle, use:

Assume that $\neg P(n)$ holds for some integer $n \geq b$.
Let $N \geq b$ be the smallest such integer satisfying $\neg P(N)$.
Derive "a contradiction."

2. To prove $(\forall n \geq b)\, P(n)$ by mathematical induction, use:

Base step: Prove $P(b)$.
Inductive step: Let $n \geq b$ be an integer.
Assume $P(n)$.
Prove $P(n+1)$.

3. To prove $(\forall n \geq b)\, P(n)$ by strong induction with one base step, use:

Base step: Prove $P(b)$.
Inductive step: Let $n > b$ be an integer.
Assume $P(k)$ whenever $b \leq k < n$.
Prove $P(n)$.

4. To prove $(\forall n \geq b)\, P(n)$ by strong induction with multiple base steps, identify the integer $c > b$ and use:

Base step: Prove $P(b)$.
Base step: Prove $P(b+1)$.
$\vdots$
Base step: Prove $P(c)$.
Inductive step: Let $n > c$ be an integer.
Assume $P(k)$ whenever $b \leq k < n$.
Prove $P(n)$.

Proof and Assumption Strategies for Set Theory

1. To prove $A \subseteq B$, use the form:

Let $x \in A$.
Prove $x \in B$.

2. To prove $A = B$, try one of the following:

(a) Prove $A \subseteq B$
Prove $B \subseteq A$.

(b)
Let x be arbitrary.
    Prove $x \in A \leftrightarrow x \in B$.

3. When *assuming* $A \subseteq B$, if you know or can prove $x \in A$, then you can conclude $x \in B$. If you know or can prove $x \notin B$, then you can conclude $x \notin A$.
4. When *assuming* $A = B$, if you know or can prove $x \in A$, then you can conclude $x \in B$. If you know or can prove $x \notin A$, then you can conclude $x \notin B$.

Proof and Assumption Strategies for Functions

1. To prove $f = g$ where $f : A \to B$ and $g : A \to B$, use:

Let $x \in A$.
    Prove $f(x) = g(x)$.

2. To prove that a function $f : A \to B$ is one-to-one, use:

Let $x \in A$ and $y \in A$.
    Assume $f(x) = f(y)$.
        Prove $x = y$.

3. To prove that a function $f : A \to B$ is onto, use:

Let $y \in B$.
    Let $x =$ (the element in A you found).
        Prove $f(x) = y$.

4. When *assuming* $f : A \to B$ is one-to-one. If you are also assuming or can prove that $f(x) = f(y)$, then you can conclude that $x = y$ whenever $x, y \in A$.
5. When *assuming* $f : A \to B$ is onto, then for any $y \in B$ you can conclude that $f(x) = y$ for some $x \in A$.

Proof Strategies for Relations

1. To prove a relation $\sim$ on A is reflexive, use:

Let $x \in A$.
    Prove $x \sim x$.

2. To prove a relation $\sim$ on A is symmetric, use:

Let $x, y \in A$.
    Assume $x \sim y$.
        Prove $y \sim x$.

3. To prove a relation $\sim$ on A is transitive, use:

Let $x, y, z \in A$.
  Assume $x \sim y$.
  Assume $y \sim z$.
    Prove $x \sim z$.

Proof and Assumption Strategies for Abstract Algebra

1. Let $(G, *)$ be a group and let H be a subset of G. To prove that H is a subgroup of G, use:

Prove $e \in H$.
Let $a, b \in H$.
  Prove $ab \in H$.
Let $a \in H$.
  Prove $a^{-1} \in H$.

2. Let $(G, *)$ be a group and let N be a subgroup of G. To prove that N is a normal subgroup of G, use:

Let $a \in G$ and $n \in N$.
  Prove $a^{-1}na \in N$.

3. Let $(G, *)$ be a group and suppose in a proof you are *assuming* that H is a subgroup of G. Then you know that $e \in H$, and if you are assuming or can prove that $a, b \in H$, then you can conclude that $ab \in H$ and $a^{-1} \in H$.

4. Let $(G, *)$ be a group and suppose in a proof that you are *assuming* N is a normal subgroup of G. Let $a \in G$ be any element. If you are assuming or can prove that $n \in N$, then you can conclude that $a^{-1}na \in N$ or, equivalently, that $a^{-1}na = h$ for some $h \in N$.

Proof and Assumption Strategies for Real Analysis

1. Given a real number β and a nonempty $S \subseteq \mathbb{R}$, to prove that $\beta = \sup(S)$, use one of the following:

(a)
Prove $x \leq \beta$ for all $x \in S$.
Assume b is an upper bound for S.
  Prove $\beta \leq b$.

(b)
Prove $x \leq \beta$ for all $x \in S$.
Assume $r < \beta$.
  Prove $r < x$ for some $x \in S$.

2. Given a real number α and a nonempty $S \subseteq \mathbb{R}$, to prove that $\alpha = \inf(S)$, use one of the following:

(a)
Prove $\alpha \leq x$ for all $x \in S$.
Assume a is a lower bound for S.
Prove $a \leq \alpha$.

(b)
Prove $\alpha \leq x$ for all $x \in S$.
Assume $\alpha < q$.
Prove $x < q$ for some $x \in S$.

3. Let $\langle s_n \rangle$ be a sequence and $\ell \in \mathbb{R}$. To prove that $\lim_{n\to\infty} s_n = \ell$, use:

Let $\varepsilon > 0$ be a real number.
Let $N =$ (the natural number you found).
Let $n > N$ be a natural number.
Prove $|s_n - \ell| < \varepsilon$.

4. Let $f\colon D \to \mathbb{R}$ and $c \in D$. To prove that f is continuous at c, use:

Let $\varepsilon > 0$ be a real number.
Let $\delta =$ (the positive value you found).
Let $x \in D$.
Assume $|x - c| < \delta$.
Prove $|f(x) - f(c)| < \varepsilon$.

5. Let $S \subseteq \mathbb{R}$ be nonempty and $\beta \in \mathbb{R}$. When you are *assuming* that $\beta = \sup(S)$. Then (1) $x \leq \beta$ for all $x \in S$; (2) whenever b is an upper bound for S, you can deduce that $\beta \leq b$; and (3) whenever $r < \beta$ there is an $x \in S$ such that $r < x$.

6. Let $S \subseteq \mathbb{R}$ be nonempty and $\alpha \in \mathbb{R}$. When you are *assuming* that $\alpha = \inf(S)$. Then (1) $\alpha \leq x$ for all $x \in S$; (2) whenever a is a lower bound for S, you can deduce that $a \leq \alpha$; and (3) whenever $\alpha < q$ there is an $x \in S$ such that $x < q$.

7. Suppose you are *assuming* that $\lim_{n\to\infty} s_n = \ell$. Then for any $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $|s_n - \ell| < \varepsilon$ for all $n > N$.

8. Let $f\colon D \to \mathbb{R}$ and $c \in D$. Suppose you are *assuming* that f is continuous at c. Then for any $\varepsilon > 0$ there is an $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ whenever $x \in D$ satisfies $|x - c| < \delta$.

References

1. Barwise, J., Etchemendy, J.: The Language of First-Order Logic. CSLI Lecture Notes, vol. 23, 3rd edn. Stanford University Center for the Study of Language and Information, Stanford (1993). Including the program Tarski's World
2. Bourbaki, N.: Elements of Mathematics. Algebra, Part I: Chapters 1–3. Hermann, Paris (1974). Translated from the French
3. Courant, R., Robbins, H.: What Is Mathematics? Oxford University Press, New York (1941)
4. Dunham, W.: Journey Through Genius. Penguin Books, New York (1991)
5. Enderton, H.B.: Elements of Set Theory. Academic (Harcourt), New York (1977)
6. Enderton, H.B.: A Mathematical Introduction to Logic, 2nd edn. Harcourt/Academic, Burlington (2001)
7. Epp, S.: Discrete Mathematics with Applications. Thompson, Belmont (2004)
8. Halmos, P.R.: Naive Set Theory. Springer, New York (1974). Reprint of the 1960 edition, Undergraduate Texts in Mathematics
9. Hawking, S.W.: A Brief History of Time. Bantam Books, Toronto/New York (1988)
10. Herstein, I.N.: Abstract Algebra, 3rd edn. Prentice Hall, Upper Saddle River (1996)
11. Rudin, W.: Principles of Mathematical Analysis. International Series in Pure and Applied Mathematics, 3rd edn. McGraw-Hill, New York (1976)
12. Velleman, D.J.: How to Prove It. Cambridge University Press, Cambridge, UK (1994)

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7,

Index of Special Symbols

$\wedge, \vee, \neg$, 2
Ⓢ, 4
T, F, 4
$\Leftrightarrow$, 7
$\rightarrow$, 12
$\leftrightarrow$, 16
$\therefore$, 19
$a \notin A$, 30
$x \in A$, 30
$\mathbb{N}$, 31
$\mathbb{Q}$, 31
$\mathbb{R}$, 31
$\mathbb{Z}$, 31
$\{x \in U : P(x)\}$, 31
$\varnothing$, 32
$\mathbb{Q}^+$, 32
$\mathbb{Q}^-$, 32
$\mathbb{R}^+$, 32
$\mathbb{R}^-$, 32
$\mathbb{Z}^-$, 32
$\subseteq$, 32
∞, 33
$\forall, \exists$, 34
$(\forall x \in A), (\exists x \in A)$, 39
$(\forall x < a), (\exists x < a)$, 40
$\exists!$, 54
$\square$, 64
Ⓐ, 68
$m \mid n$, 81
$m \nmid n$, 81
$|x|$, 88
$\Rightarrow, \Leftarrow$, 90
$\sum_{k=1}^{n} a_k$, 108
$\binom{n}{k}$, 113
$n!$, 113
$\gcd(m,n)$, 135
$\varnothing$, 143
$\subseteq$, 143
$\mathcal{P}$, 144
$\not\subseteq$, 144
$\cap, \cup, \setminus$, 145
$A \times B$, 147
$\{C_i : i \in I\}$, 157
$\bigcap_{i \in I} C_i$, 159
$\bigcup_{i \in I} C_i$, 159
$\bigcap \mathcal{F}$, 162
$\bigcup \mathcal{F}$, 162
$f : A \rightarrow B$, 170
$f(x)$, 170
$\mathrm{ran}(f)$, 174
$i_A : A \rightarrow A$, 174
$f^{-1} : B \rightarrow A$, 182
$f \circ g$, 184
$f[S]$, 189
$f^{-1}[T]$, 190
$|A|$, 203
$|A| < |B|$, 203
$|A| = |B|$, 203
$|A| \leq |B|$, 204
(a,b), 209
aRb, 209
$a \sim b$, 209
$[a]_\sim$, 215
$[a]$, 215
$a \equiv b \pmod{m}$, 218
$[a]_m$, 222
$\mathbb{Z}_m$, 222
$\mathbb{Z}_m$, 223
$\odot$, 226
$\oplus$, 226
$\preccurlyeq$, 232

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7,

$F(\mathbb{R})$, 245
$F(\mathbb{Z})$, 246
$\lhd$, 256
$\mathrm{Per}(S)$, 261
$\mathrm{lcm}(m,n)$, 269
$\circledast$, 283
$\oplus$, 283
$a \sim_H b$, 284
$\odot$, 288
$|x|$, 298
max, 299
min, 299
$\inf(S)$, 301
$\sup(S)$, 301
max, 305
min, 305
$\langle s_n \rangle$, 311
$\lim_{n\to\infty} s_n = \ell$, 312

Index

A
abelian group, 248
absolute value, 88, 298–299
abstract algebra, vii, viii, 227, 245
algebraic structure, 245
Archimedean property, 311
arrow diagram, 170
ascending prime factorization, 139
associative laws
  for sets, 154
  logical, 9
assumption strategies
  biconditional, 90
  conditional, 70
  conjunction, 82
  disjunction, 86
  division by cases, 87
  for all, 72
  one-to-one function, 179
  onto function, 181
  proof by cases, 86
  set equality, 154
  subset relation, 151
  there exists, 74
axiom of choice, 167
axiomatic method, 166
axioms of arithmetic, 225

B
backward triangle inequality, 299
base step, 104
base value, 104
Bernoulli's inequality, 127, 319
biconditional
  assumption strategy, 90
  connective, 16
  law, 16
  proof strategy, 89
binary operation, 239
  associative, 239
  closed under, 246
  commutative, 239
  identity element, 239
  inverse, 239
  relation preserving, 281
  well-defined formula, 226
binomial coefficient, 113
binomial theorem, 122
bound variable, 36
bounded
  function, 306
  sequence, 321
  set, 300
bounded above, 300
bounded below, 300
bounded number quantifiers, 40
bounded quantifier negation laws, 44, 45
bounded set quantifiers, 39

C
cancellation law
  for groups, 251
  for integral domains, 276
Cantor's diagonal argument, 200, 206
Cantor, Georg, 164, 193
cardinality, 203
Cartesian product
  of 2 sets, 147
  of a set with itself, 209
chain, 235
choice set, 167

D.W. Cunningham, *A Logical Introduction to Proof*, DOI 10.1007/978-1-4614-3631-7,

CL (conditional law), 14
closed under, 31, 246
closed-form solution, 110
Cohen, Paul, 168
commutative laws
  logical, 9
complete residue system, 223
completeness axiom, 302
components of a cycle, 263
composite function, 184
composite number, 100
comprehension principle, 164
conditional
  assumption strategies, 70
  conclusion, 12
  connective, 12
  hypothesis, 12
  laws, 14
  proof strategy, 68
congruence algebra, 220
congruence relation, 281
  modulo *m*, 218
  on an algebraic structure, 281
conjecture, 61, 62
conjunction
  assumption strategy, 82
  connective, 2
  proof strategy, 80
conjunctive addition, *see* inference rules, *see* inference rules
conjunctive simplification, *see* inference rules
constant function, 174
constructive proof, 73
continuous function
  at a point, 329
    assumption strategy, 333, 346
    proof strategy, 329
  on its domain, 338
contradiction
  law, 10
  logical, 7
  proof strategy, 93
contrapositive
  logical, 14
  proof strategy, 91
contrapositive law, 17
converse, 15
converse error, 22
corollary, 97
countable set, 195
countably infinite, 196
counterexample
  refuting a conjecture, 62
  refuting a universal statement, 35
cycle decomposition algorithm, 267
cycle notation, *see* permutation

D
De Morgan's laws
  for sets, 161, 162
  logical, 8
decimal expansions, 129, 130, 136, 205, 206
deduction, 26
denumerable, 198
diagonal argument, 200, 206
difference, *see* set operations
direct proof, 68, 91
directed graph, 209
disjoint cycles, 265
disjoint sets, 143, 146
disjunction
  assumption strategies, 86
  connective, 2
  proof strategies, 85
disjunctive addition, *see* inference rules
disjunctive syllogism, *see* inference rules, *see* inference rules
distributive laws
  for sets, 154
  logical, 9
divisibility, 81
division algorithm, 134
division by cases, 85
DML (De Morgan's law), 8
DNL (double negation law), 10

E
either-or, 3
element, 30, 143
element of, 30, 143
empty set, 32, 143
enumeration, 198
equations, how to prove, 63
equivalence class, 215
equivalence relation, 214
Euclid's lemma, 139
even integer, 75
existence proofs, *see* there exists
existential statement, 35

F
factorial, 113
fallacy, 22, 59
family of sets, 157
Fibonacci sequence, 137

field, 293
  additive inverse, 293
  algebraic properties, 293–295
  axioms, 293
  multiplicative identity, 293
  multiplicative inverse, 293
  ordered, 295–297
  zero element, 293
finite set, 195
  number of elements, 195
for all
  assumption strategy, 72
  proof strategy, 71
free variable, 36
function
  co-domain, 169
  domain, 169
  continuous at a point, 329
  equality, 175
    proof strategy, 175
  image of a set, 189
  inverse image of a set, 190
  one-to-one
    assumption strategy, 179
    definition, 178
    proof strategy, 178
  onto
    assumption strategy, 181
    definition, 180
    proof strategy, 180
  range, 174
  single-valued, 170
  well-defined, 171–173
fundamental theorem
  of arithmetic, 140
  of equivalence relations, 216

G
Gödel, Kurt, 168
Galois, Évariste, 262
Gauss, Carl Friedrich, 218, 225
geometric sequence, 118
geometric sum, 118
greatest common divisor, 135
greatest lower bound, 234, *see* infimum
group, 247
  abelian, 248
  axioms, 247
  identity element, 247
  order of an element, 257
  order of the group, 257
  subgroup, *see* subgroup
group theory, 247

I
ideal, 278
idempotent laws, 9
identity element, *see* binary operation, group
identity function, 174
if and only if, *see* biconditional
if-then statement, *see* conditional
iff, 16
image, *see* function
index set, 157
indexed family of sets, 157
indexed set, 157
indirect proof, 91
induction, *see* mathematical induction
induction conclusion (IC), 104
induction hypothesis (IH), 104
inductive step, 104
inductively defined sequence, 124
inequality
  how to prove, 64
  in ordered fields, 295–297
  laws of, 64
  substitution properties, 69, 313
inference rules, 24
  conjunctive addition, 24, 58
  conjunctive simplification, 24
  disjunctive addition, 24
  disjunctive syllogism, 24, 58
  modus ponens, 20
  modus tollens, 20
  substitution, 25
infimum
  assumption strategy, 304, 309
  definition, 301
  proof strategy, 303, 309
infinite set, 195
injection, *see* one-to-one
integer arithmetic, 225
  additive identity, 226
  additive inverse, 226
  multiplicative identity, 226
integers, 31
intersection, *see* set operations
interval notation, 33
invalid argument
  in propositional logic, 21
inverse error, 23
inverse function, 182
inverse image, *see* function
irrational numbers, 31, 97, 206

L
largest element, 235
least common multiple, 269

least upper bound, 234, *see* supremum
lemma, 101
limit of a sequence, *see* sequence
logic law, 8
logic laws
  substitution, 10
logical connectives, 2
logical equivalence, 7
  not equivalent, 8
lower bound, 234, 300

M
mathematical induction, 102
  base step, 104
  base value, 104
  induction conclusion (IC), 104
  induction hypothesis (IH), 104
  inductive step, 104
  principle of, 103
  proof strategy, 103
  strong induction I, 128
  strong induction II, 131
  strong induction proof strategies
    multiple base steps, 131
    one base step, 128
mathematical proof, 66
matrix, 240
maximal element, 234
maximum, 299, 305
member, 30
member of, 30
minimal element, 234
minimum, 299, 305
mod, 218
modular arithmetic, 225, 227
  additive identity, 228
  additive inverse, 230
  invertible, 228
  multiplicative identity, 228
  multiplicative inverse, 228
  zero divisor, 228
modus ponens, *see* inference rules
modus tollens, *see* inference rules
moved element, by a permutation, 266
multiplicative inverse, 31

N
natural numbers, 31
necessary and sufficient, 14
negation
  connective, 2
  double negation law, 10
neighborhood, 320
neither-nor, 3
nonconstructive proof, 73
normal subgroup, 256
  assumption strategy, 257
  proof strategy, 256

O
odd integer, 75
one-to-one, *see* function
one-to-one correspondence, 193
only if, 14
onto, *see* function
open sum, 110
open-form solution, 110
order of group element, *see* group
ordered field, 296
ordered pair, 147, 209
  equality, 147
  first component, 147
  second component, 147

P
pairwise disjoint, 148
partial order, 232
partially order set, 232
partition, 148, 214
perfect cube, 41
permutation
  cycle notation, 263
    2-cycle, 270
    k-cycle, 263
  matrix notation, 262
  parity, 271
permutation group, 261
permutation of a set, 260
poset, 232
power set, 144, 203
  cardinality of, 204
predicate, 29
  domain, 29
prime number, 100
proof strategies
  algebraic equations, 63
  biconditional, 89
  conditional, 68
  conjunction, 80
  contradiction, 93
  contrapositive, 91
  disjunction, 85
  for all, 71

function equality, 175
induction, 103
one-to-one function, 178
onto function, 180
reflexive relation, 211
set equality, 151
strong induction, 128, 131
subset relation, 150
symmetric relation, 212
there exits, 73
transitive relation, 212
uniqueness, 78
well-ordering, 100
proper subgroup, 252
proper subset, 143
proposition, 1
propositional components, 1, 4
propositional logic laws, 9
propositional sentence, 2, 17

Q
QNL (quantifier negation laws), 43
quantifier negation laws, 43–45
quantifiers, 34
adjacent, 46
existential, 34
mixed, 49, 77
multiple, 46
non-adjacent, 49
number bounded, 40
set bounded, 39
uniqueness, 54
universal, 34
quotient algebra, 283
quotient group, 284
quotient ring, 288

R
rational field, 300
rational numbers, 31
equality, 32
reduced form, 95
real analysis, vii, viii, 42, 88, 293, 311
real field, 298
real numbers, 31
recursively defined sequence, 124
relation
antisymmetric, 232
equivalence, 214
on a set, 209
reflexive, 211
proof strategy, 211
symmetric, 211
proof strategy, 212
transitive, 212
proof strategy, 212
relatively prime, 135
reverse triangle inequality, *see* backward triangle inequality
right coset of a subgroup, 285
right coset of an ideal, 290
ring, 273
additive inverse, 273
axioms, 273
commutative, 273
integral domain, 274
subring, 277
unit, 274
with unity, 273
zero divisor, 274
zero element, 273
round-robin proof, 90
Russel's paradox, 164, 166

S
sequence, 106, 311
bounded, 321
convergent, 312
limit of, 312
assumption strategy, 316
proof strategy, 312
limit theorems, 323
term, 311
set, 30, 143
set equality
assumption strategy, 154
definition, 143, 151
proof strategy, 151
double-subset strategy, 152
iff strategy, 152
set operations
complement, 145
difference, 145
intersection
of indexed family, 159
of two sets, 145
union
of indexed family, 158
of two sets, 145
single-valued, 170
smallest element, 235
square root, 95
squeeze theorem, 328
strict order, 233
string of 0's, 130

string of 9's, 130
subgroup, 252
  assumption strategy, 254
  proof strategy, 254
  proper, 252
subset relation
  assumption strategy, 151
  definition, 32, 143
  proof strategy, 150
substructure, 246
summand, 69
summation notation, 108
  index, 108
  lower limit, 108
  shift rule, 112
  summand, 108
  upper limit, 108
supremum
  assumption strategy, 302, 308
  definition, 301
  proof strategy, 302, 308
surjection, *see* onto
symmetric group, 262

T
Tarski's World, 37
Tarskian predicates, 38, 49
tautology
  law, 10
  logical, 7
theorem, 62
there exists
  assumption strategy, 74
  proof strategies, 73
total order, 233
totally ordered set, 233
transitivity law, *see* inequality
transposition, 270
triangle inequality, 98, 298
trichotomy law, 64, 296
truth assignment, 4
truth set, 31, 144
truth table, 4, 6
truth value, 4

U
uncountable set, 200
union, *see* set operations
uniqueness
  proof strategies, 78
  quantifier, 54
unity, *see* ring
universal converse error, 59
universal inverse error, 59
universal proofs, *see* for all
universal statement, 35
universe of discourse, 30, 31
upper bound, 234, 300

V
valid argument
  in predicate logic, 58
  in propositional logic, 20
  substitution, 23
Venn diagram, 143

W
well-defined, 171
well-ordering principle, 99
  general form, 99
  proof strategy, 100
without loss of generality, 87

Z
Zermelo-Fraenkel axioms, 164
ZF, 167
ZFC, 167
Zorn's Lemma, 235