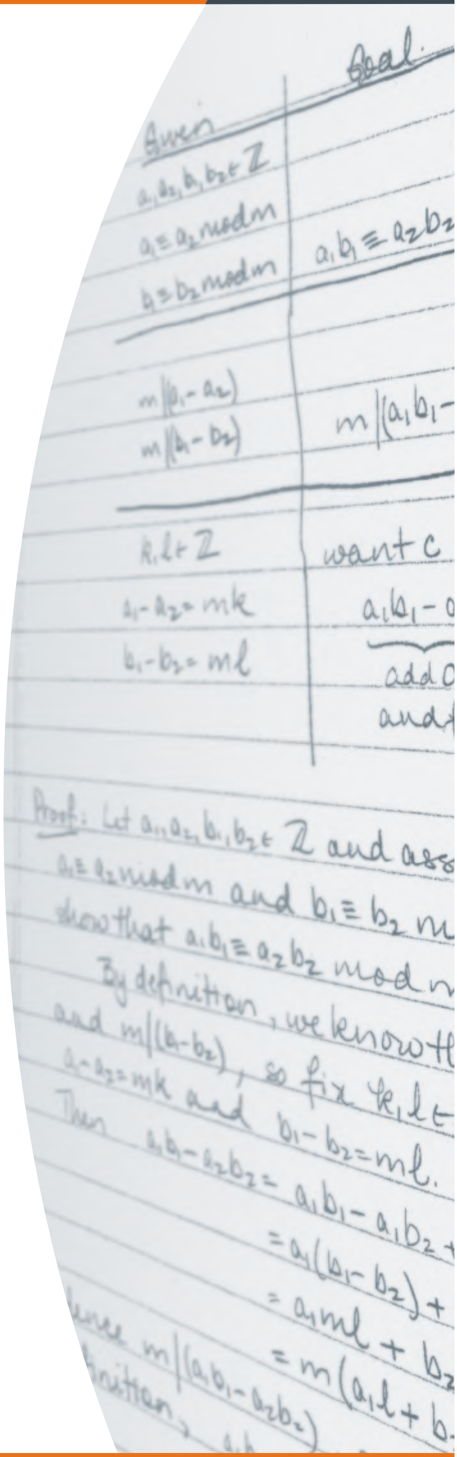
The Sally SERIES

Pure and Applied UNDERGRADUATE TEXTS 26

The Tools of Mathematical Reasoning

Tamara J. Lakins


Goal
Given
$a_1,a_2,b_1,b_2\in\mathbb{Z}$
$a_1\equiv a_2$ mod m
$b_1\equiv b_2$ mod m
$a_1b_1\equiv a_2b_2$
$m|(a_1-a_2)$
$m|(b_1-b_2)$
$m|(a_1b_1-$
$k,l\in\mathbb{Z}$
want c
$a_1-a_2=mk$
$b_1-b_2=ml$
add
and
Proof: Let $a_1,a_2,b_1,b_2\in\mathbb{Z}$ and ass
$a_1\equiv a_2$ mod m and $b_1\equiv b_2$ m
show that $a_1b_1\equiv a_2b_2$ mod m
By definition, we know th
and $m|(b_1-b_2)$, so fix $k,l\in$
$a_1-a_2=mk$ and $b_1-b_2=ml$.
Then $a_1b_1-a_2b_2=$ $a_1b_1-a_1b_2+$
$=a_1(b_1-b_2)+$
$=a_1ml+b_2$
$=m(a_1l+b$
Hence $m|(a_1b_1-a_2b_2)$
finition,

American Mathematical Society

The Tools of Mathematical Reasoning

The Sally SERIES

Pure and Applied
UNDERGRADUATE TEXTS · 26

The Tools of Mathematical Reasoning

Tamara J. Lakins

American Mathematical Society
Providence, Rhode Island

2010 *Mathematics Subject Classification*. Primary 00-01.

For additional information and updates on this book, visit
www.ams.org/bookpages/amstext-26

Library of Congress Cataloging-in-Publication Data

Names: Lakins, Tamara J., 1963-
Title: The tools of mathematical reasoning / Tamara J. Lakins.
Description: Providence, Rhode Island : American Mathematical Society, [2016] | Series: Pure and applied undergraduate texts ; volume 26 | Includes bibliographical references and index.
Identifiers: LCCN 2016021930 | ISBN 9781470428990 (alk. paper)
Subjects: LCSH: Mathematical analysis–Foundations–Textbooks. | Logic, Symbolic and mathematical–Textbooks. | AMS: General – Instructional exposition (textbooks, tutorial papers, etc.). msc
Classification: LCC QA300 .L26 2016 | DDC 511.3–dc23 LC record available at https://lccn.loc.gov/2016021930

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at http://www.ams.org/

10 9 8 7 6 5 4 3 2 1   21 20 19 18 17 16

For my husband Harald Ellers, my son Michael Ellers,
and my parents Billy and Lois Lakins,
with love

Contents

Preface

The idea for this textbook was conceived as a direct result of my experience teaching the “introduction to proofs” course at Allegheny College. When I first started teaching this course, there were only a handful of appropriate textbooks on the market. My experience teaching from various textbooks clarified in my mind what I wanted to accomplish in such a course and how to accomplish it.

One possible title (and the one used at Allegheny College) for an “introduction to proofs” course is “Foundations of Mathematics”, which can conjure up at least two possibilities for the focus of the course. The word “foundations” could be interpreted in the sense of the logician: working axiomatically and in the language of set theory, or working within a formal proof system. On the other hand, the course can be viewed as a student’s first exposure to proofs, sets, functions, etc., *as mathematicians use them*, giving students a practical collection of tools that will enable them to be successful in later mathematics courses, such as abstract algebra and real analysis. As interesting as the first interpretation is, in my opinion the second interpretation is the right one for a first exposure to these ideas and for the average mathematics major.

It is important that students begin writing proofs as early as possible in the course, hopefully by the end of the second week of classes. To achieve this, I present only enough logic for students to be able to work with the propositional connectives and the quantifiers. Although the initial treatment of this material is streamlined, the importance of this material is emphasized throughout the book. Students are frequently reminded, especially in the early chapters, of the importance of the logical structure of a mathematical statement as a framework for finding a proof of that statement. In particular, the importance of the logical structure of a mathematical definition as a framework for proving that an object has (or does not have) that property is a constant theme throughout the textbook.

Focusing on logical structure is an important first step in addressing the question, “How do I start?” that students who are learning to write proofs often ask.

To help students learn that searching for a proof is a *process*, I have adopted Velleman's approach [**15**] of using a "Given-Goal diagram" to organize what is known versus what is to be proved. Together, these methods teach students that the logical structure of the goal dictates the structure of a proof. Given-Goal diagrams can be modified by unravelling the logical structure of the statement to be proved, thereby organizing the search for a proof.

For many theorems and sample solutions to problems, this book presents several paragraphs of "scratchwork" before presenting a correct formal proof. The goal is to walk students through a complete analysis of a problem: understanding the logical structure of the statement, creating one or more Given-Goal diagrams, deploying various techniques to build a bridge from the given to the goal, and finally writing a complete concise proof. I hope that this will help demystify the process of searching for a proof.

After the introductory material on logic, my goal is to introduce students to various proof techniques. The focus is on proving simple statements about integers, rational numbers, and real numbers. It is important that students know what the "ground rules" are at the beginning: what may they assume and what requires proof? To achieve this, I provide students with the "basic properties" of integers and real numbers; all other statements require proof. My goal here is not to develop the entire theory of the integers (or the real numbers) from the axioms. Rather, my goal is to be clear about the assumptions we make. Occasionally, I feel that assuming additional axioms (such as assuming the existence part of the Fundamental Theorem of Arithmetic when proving the existence of infinitely many prime numbers) will expedite discussion and illustrate concepts. In such cases, I clearly point out when we're assuming more than we ought and emphasize that we will pay our debts later.

Quantifiers and the proper use of variables are given very careful treatment. The focus, as before, is that the logical structure of the statement determines the shape of the proof, as well as guides the search for the proof. My approach to quantifiers and variables is informed by the rigorous framework provided by first-order logic (see, for example, [**10**]). I pay particular attention to the concept of *existential instantiation*[1] as a means to stress the proper use of variables, as well as the difference between a quantified statement and a nonquantified statement; namely, if we know $(\exists x)P(x)$, then we may fix a particular element d in the universe such that $P(d)$ is true, as long as we use a new variable that doesn't already have a particular meaning in the proof. In my experience, students need time to get used to working with a single quantifier before moving on to more complicated statements. Consequently, I postpone proofs of statements *beginning* with mixed quantifiers, such as the ε-δ proofs that give students so much trouble, until after the students have more experience proving statements beginning with only one quantifier. Given-Goal diagrams are particularly effective here as a means of teaching students to unravel the logical complexity of a statement as a means of organizing and searching for the proof of that statement.

[1] I mention this terminology exactly once in the text!

One of the most difficult decisions one has to make when writing a textbook of this sort is the order in which to present the concepts of relation and function. In particular, one has to decide whether to define a function as an abstract set of ordered pairs with a particular property or as a triple consisting of two sets and a "correspondence". If a function is a triple, then one must further decide how function equality is defined. (Must the codomains agree, or not? Both definitions occur in the literature.) The cost of defining a function as a particular kind of set of ordered pairs is that other concepts, such as function composition, become more abstract and harder to work with. In my experience, this formal approach is confusing and unnecessarily abstract for students who are just beginning to learn about mathematical proofs. Functions are rarely treated as sets of ordered pairs in other undergraduate mathematics courses (unless a student takes a course in formal set theory), so doing so in this course would present an abstract view of the concept that won't be useful to students later when they need to use functions in other courses. I have chosen to define the concept of function first, as a triple consisting of two sets and a "correspondence". I include agreement of the codomains when defining function equality, to avoid the situation that can otherwise occur, in which two functions can be equal with one being onto its codomain and the other not.

In the early part of the textbook, the focus is on giving students practice with sets and the various proof techniques. Chapter 1 provides the necessary background on the logic of the propositional connectives and the quantifiers, as well as an introduction to the concept of proof. In Chapter 2, I present a variety of direct and indirect proof techniques. Chapter 3 is devoted to induction.

The goal of the later chapters is to provide students with the foundational material on sets, functions, equivalence relations, number theory, finite and infinite sets, and introductory analysis they will need in order to succeed in their later proof-based courses, particularly linear algebra, number theory, abstract algebra, and real analysis. Chapters 4 and 5 deal with sets and functions, respectively. The focus of Chapter 6 is introductory number theory: the Division and Euclidean Algorithms and elementary facts about congruences, which are also important in abstract algebra. The material in Section 6.4 and Section 6.5 on congruences and congruence classes can be delayed until the more general discussion of relations, equivalence relations, and equivalence classes in Chapter 7, which provides a further introduction to ideas important in abstract algebra. In Chapter 8, I discuss finite and infinite sets, with a focus on the material needed in real analysis, namely, the difference between countability and uncountability. Finally, Chapter 9 presents the axiomatic foundations of analysis, including the Completeness Axiom, the existence of $\sqrt{2}$, and the Archimedean Property.

One of the more difficult aspects of learning to write a proof is learning to effectively communicate that proof to others (the instructor or other students). Particularly at the beginning, I emphasize the difference between the search for a proof (including any scratchwork) and the final written proof. I have included some guidelines for writing mathematics in the appendix. These guidelines were originally inspired by the "Writing checklist" that Dr. Annalisa Crannell (Franklin & Marshall College) so generously shared with me many years ago.

I have deliberately not included “answers to selected problems” at the back of the textbook. The process of learning to find a proof, and learning to recognize when what one has written is a correct proof, is an active one, which the passive reading of solutions circumvents. The text provides plenty of examples and a great deal of commentary. Students will learn more by grappling with the problems, perhaps in consultation with the instructor or other students, without a solution easily accessible. When appropriate, I include hints for some of the more difficult problems. There are many exercises at the end of sections. Text cross-references to exercises are in the form, for example, Exercise 1.1.2a. This is a reference to exercise 2, part (a), in the section Exercises 1.1, which is at the end of Section 1.1.

Proofs are ended with the usual end-of-proof character □. I have used the symbol ◊ to mark the end of examples.

To Students

Chances are you are studying the material in this book because you are enrolled in an “introduction to proofs” course at your college or university. You will be learning how to use mathematical language and notation and how to communicate mathematical ideas clearly and precisely. And you will be learning the foundations of mathematical reasoning, the mathematician’s standard of truth. Logical reasoning skills and the ability to use mathematical language and notation properly are also essential for other scientific disciplines, such as physics and computer science.

In my experience, this type of course is usually a completely new experience for students. It is normal to feel a bit disoriented at first. It is important to persevere. It is especially important to study *actively*, by reading the textbook equipped with pencil and paper, by writing lots of proofs, and by discussing the mathematics with your instructor and fellow students. You should never expect to simply write down a proof immediately after reading the statement to be proved. As illustrated in this textbook, finding a proof is a *process* that must take you from an analysis of the statement to be proved, through the scratchwork of Given-Goal diagrams and false starts, to a final polished and correct proof.

It is important that you learn any new mathematical definition or notation right away; you cannot hope to succeed if you don’t know what the words and notation mean! This may be a new experience for you, particularly since there is not a lot of “leeway” in mathematical definitions. One needs to know the precise meaning of the words, rather than the “general idea”. Finally, you may find it strange to be writing so much in a math course. Keep in mind that our job is to *communicate* what we know, and how we know it, to others. Learning to write mathematics well requires a lot of practice and can be difficult at the beginning. One way to improve your writing is to read your mathematical statements out loud. Since notation simply abbreviates a mathematical statement and since our statements have a grammar, speaking out loud *exactly* what you’ve written (no more and no less) can help you improve your mathematical writing.

Acknowledgements

I have coded the graphics in this book using pstricks and TikZ. In particular, the code for the Venn diagrams on page 75 can be found in "Example: Set operations illustrated with Venn diagrams", authored by Uwe Ziegenhagen, published on TeXample.net on 3/18/2010. I have replaced the use of color in that example with grayscale, moved the labels, and added the rectangular box for the universe $\mathcal{U}$. The use and adaptation of this code is permitted by the Creative Commons License Deed found at creativecommons.org/licenses/by/2.5/.

I am grateful to the staff at the AMS, including Eriko Hironaka, whose support for my textbook and willingness to answer questions was gratifying, and Chris Thivierge, who provided me with all the information I needed to prepare my manuscript for the AMSTEXT series. I thank Barbara Beeton and the rest of the technical support team at the AMS for invaluable help in managing several TeX issues. I am particularly grateful to the anonymous reviewers for their careful reading of my manuscript and for their many valuable comments and suggestions for improving the exposition and exercises. Of course, any errors in the text are my own.

I am grateful to many colleagues. Michael Barry (Allegheny College), Matt Clay (University of Arkansas), Jeffry Hirst (Appalachian State University), Iraj Kalantari (Western Illinois University), and Daniel Velleman (Amherst College) provided feedback on early versions of this textbook. Craig Dodge, Harald Ellers, Rachel Weir, and Caryn Werner used early versions of this textbook in the classroom at Allegheny College and also provided valuable feedback.

My desire to write this textbook was motivated by my desire to equip students with the skills they need in order to be successful in upper-level mathematics courses. Teaching Allegheny's introduction to proofs course helped me better understand why some students find proof writing so difficult, and it helped me improve as a teacher. I am grateful to my own teachers, as well. In particular, I would like to thank Iraj Kalantari at Western Illinois University for sharing with me his enthusiasm and expertise as a mathematician, teacher, and expositor.

Finally, I am most grateful to my husband, Harald Ellers. He has given me his unwavering support, both personally and professionally. This textbook has greatly benefitted from our numerous conversations regarding mathematics, teaching, and writing.

Chapter 1

Language, Logic, and Proof

1.1. Language and logic

Like all scientific subjects, mathematics requires evidence in order to justify claims. While the lab sciences often use experimental data to justify their claims, in mathematics, logical reasoning is the standard of truth.

Mathematics is concerned with formal *statements* about mathematical objects, such as integers or functions, and whether these statements are true or false. The *language* of mathematics must therefore be precise and unambiguous—it has a vocabulary and a grammar. Logical arguments called *proofs* are used to deduce statements from basic assumptions; i.e., given a mathematical statement, we want to determine whether it is true or false and prove that our assertion is correct. For example, you may have learned in a previous course that there exist infinitely many prime numbers. Just a few definitions and proof techniques will enable us to prove this statement, which we do in Section 2.3.

The language and tools of mathematics are used by other scientists, particularly physicists and computer scientists, as well. For example, a computer scientist may wish to determine the "computational complexity" (or "hardness") of an algorithm, or even to prove that an algorithm "works" at all.

We will begin with mathematical *language*, the logical connectives and quantifiers, and then we will study the fundamental techniques of *proof*. Once armed with these tools, we are ready to study the concepts most often needed in mathematics and computer science, such as sets, functions, and relations. We then consider a variety of mathematical topics designed to prepare you for future proof-based math courses.

Definition 1.1.1. A *proposition* is a sentence (i.e., it has both a subject and a verb) which has exactly one truth value; i.e., it is either true or false, but not both.

Example 1.1.2. Consider the following examples of propositions.

(1) $2 + 3 = 6$.

Here, the verb is *equals*, which is represented notationally. (Remember we said that our mathematical language has a grammar!) Clearly this proposition is false.

(2) The 10^{46}th digit of π is 7.

At the time this book was written, the 10^{46}th digit of π was unknown. Consequently, this proposition is a bit unusual—it is certainly true or false, but not both, but which truth value it has is unknown.

(3) Every prime number is odd.

Is this proposition true or false? To answer this, you first need to know what the words *prime* and *odd* mean. (If necessary, consult Definitions 2.1.7 and 1.2.1.) ♢

We will often represent propositions with capital letters, such as P, Q, or R. Next we consider some nonexamples of propositions.

Example 1.1.3.

(1) $2 + 3$.

What is the problem here? Refer to Example 1.1.2(1).

(2) $n + 1 > 3$.

What is the problem here? It is impossible to determine a truth value. However, the situation is very different from that of Example 1.1.2(2). Here we cannot determine a truth value because the truth value depends on the value assigned to n. For example, the statement is true if $n = 4$ and false if $n = 1$. The statement "$n + 1 > 3$" *is* a sentence, though; such a statement is called a *predicate*. We can denote the predicate "$n + 1 > 3$" by the notation $P(n)$ to emphasize that n is a *free variable*, i.e., a variable that we need to "substitute for" in order to obtain a proposition.

There is another, more subtle issue, here as well. Given the predicate $P(n)$, we should really ask ourselves what we are *allowed* to substitute for n; i.e., what is the *universe*, or possible range of values, for n? So, we can see that we will either need to make the universe for a given predicate explicit or be able to deduce it from the context (here, there was no context given). ♢

So, we have two types of mathematical statements, propositions and predicates. We can build more complicated statements using *logical connectives.*

1.1.1. Basic connectives.

Suppose that P and Q are statement letters (i.e., letters that represent propositions or predicates). We define the logical connectives *conjunction*, *disjunction*, and *negation* as follows.

The *conjunction* of P and Q is the statement "P and Q", which is denoted by $P \wedge Q$; the statements P and Q are called the *conjuncts* of the statement $P \wedge Q$. The intended meaning of the statement $P \wedge Q$ is clear; the statement $P \wedge Q$ will be true when P is true and Q is true, and false otherwise. We can represent this definition by the *truth table* in Table 1.1.

Note that when we build a truth table for a compound statement involving two (or more) statement letters (say, P and Q), we must consider all the possible truth values for each statement letter. Here there are two possible truth values for

Table 1.1. Truth table for $\wedge$.

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

P (true, false), and similarly for Q, so there are $2 \cdot 2$, or 4, possible truth values for the statement $P \wedge Q$. (This method of counting is called the *multiplication principle*, which we discuss in Section 8.2.)

Next, we consider disjunction. The *disjunction* of P and Q is the statement "P or Q", which is denoted by $P \vee Q$; here, the statements P and Q are called the *disjuncts* of the statement $P \vee Q$. We must decide on the intended meaning of this connective, since it can be interpreted in one of two ways.

In the English language, the word "or" is often an *exclusive* "or". For example, at a restaurant you may be asked to choose to have soup or salad with your entree. It is understood that you should choose one or the other, but not both (unless you pay extra!).

In mathematics, however, the common usage of the word "or" is in the *inclusive* sense. For example, consider the following well-known mathematical statement:

if a and b are integers with $ab = 0$, then $a = 0$ or $b = 0$.

Here, we know that we should interpret this statement as "if a and b are integers with $ab = 0$, then $a = 0$ *or* $b = 0$ *or* possibly *both* $a = 0$ and $b = 0$".

To repeat, *in mathematics, the usage of the word "or" is always inclusive, unless explicitly stated otherwise.* The truth table for disjunction is given in Table 1.2.

Table 1.2. Truth table for $\vee$.

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

Finally, we consider negation. The *negation* of P is the statement "not P", which is denoted by $\neg P$ and interpreted just as you suspect, in Table 1.3.

Table 1.3. Truth table for $\neg$.

P	$\neg P$
T	F
F	T

Before introducing our last two basic connectives, let's consider some examples. Just as we can make more complicated statements by forming the negation of a statement, or the conjunction or disjunction of two statements, we can form other compound statements by combining connectives.

Example 1.1.4. Determine whether the following statements are true or false.

(1) $2+3=5$ and $\neg(1+1=2)$.
Since $\neg(1+1=2)$ is false, this conjunction is false.

(2) $2+3=5$ or $\neg(1+1=2)$.
This disjunction is true since $2+3=5$ is true. ◊

Example 1.1.5. Find the truth tables for the statements

$$\neg(P \wedge Q), \quad \neg P \wedge \neg Q, \quad \neg P \vee \neg Q.$$

Before we begin, notice that we have already made an assumption about the connective $\neg$, namely, that it always modifies as little as possible, unless we explicitly indicate otherwise. Thus, we should interpret the statement $\neg P \wedge \neg Q$ as $(\neg P) \wedge (\neg Q)$. In addition, the parentheses in $\neg(P \wedge Q)$ are necessary, since $\neg P \wedge Q$ would be interpreted as $(\neg P) \wedge Q$, by our previous rule. In general, therefore, just as parentheses are used in arithmetical expressions to indicate the order in which the arithmetical operations should be evaluated, parentheses are used in compound logical statements to indicate the order in which the logical connectives should be evaluated. The requested truth tables are in Table 1.4.

Table 1.4. Truth table for $\neg(P \wedge Q)$, $\neg P \wedge \neg Q$, $\neg P \vee \neg Q$.

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

◊

We see from this example that not only does the order of connectives matter, but also the sixth and eighth columns in Table 1.4 show that the statements $\neg(P \wedge Q)$ and $\neg P \vee \neg Q$ have the same logical meaning.

Definition 1.1.6. Two statements involving the same statement letters are *logically equivalent* if they have the same truth table.

In Example 1.1.5, we see that the statements $\neg(P \wedge Q)$ and $\neg P \vee \neg Q$ are logically equivalent. On the other hand, the statements $\neg(P \wedge Q)$ and $\neg P \wedge \neg Q$ are not logically equivalent because when P is true and Q is false, $\neg(P \wedge Q)$ is true, while $\neg P \wedge \neg Q$ is false. The statement that $\neg(P \wedge Q)$ is logically equivalent to $\neg P \vee \neg Q$ is one of *DeMorgan's Laws.*

Proposition 1.1.7 (DeMorgan's Laws)**.** *Let P and Q be statements.*

(1) $\neg(P \wedge Q)$ *is logically equivalent to* $\neg P \vee \neg Q$.

(2) $\neg(P \vee Q)$ *is logically equivalent to* $\neg P \wedge \neg Q$.

Proof. We proved (1) using the truth table in Table 1.4. The proof of (2) is Exercise 1.1.2d. □

We consider two more examples before introducing our final two logical connectives.

Example 1.1.8. Assume that x is a fixed real number. What is the negation of the statement $1 < x < 2$?

We must first recall that the statement $1 < x < 2$ is an abbreviation of the compound statement

$$1 < x \text{ and } x < 2.$$

The negation of this statement is

$$\text{it is not the case that } 1 < x \text{ and } x < 2,$$

or, in notation form, $\neg[(1 < x) \wedge (x < 2)]$.

While our answer is technically correct, sometimes we find that expressing a statement "negatively", as a negation, is not useful. Often, a more useful way to express the negated statement is to express it "positively", using Proposition 1.1.7 to find a logically equivalent statement. Using DeMorgan's Laws, the negation of the statement $1 < x < 2$ is logically equivalent to the statement $\neg(1 < x) \vee \neg(x < 2)$. Simplifying further, we see that the negation of the statement $1 < x < 2$ is logically equivalent to the statement

$$x \leq 1 \text{ or } x \geq 2.$$

Here we are using what is called the *Trichotomy Axiom* of real numbers: given fixed real numbers a and b, exactly one of the statements $a < b$, $a = b$, $b < a$ is true. ◊

When we use DeMorgan's Laws to express the negation of a conjunction or disjunction "positively", we shall say that we have found a *useful denial* of that statement.

Example 1.1.9. Assume that n is a fixed positive integer. Find a useful denial of the statement

$$n = 2 \text{ or } n \text{ is odd.}$$

Using DeMorgan's Laws, we see that

$$\neg[(n = 2) \vee n \text{ is odd}]$$

is logically equivalent to

$$n \neq 2 \wedge n \text{ is even,}$$

where we are using the fact that every integer is either even or odd, but not both. Thus, a useful denial of the given statement is, in natural English,

$$n \text{ is an even positive integer other than 2.}$$ ◊

We now present the final two logical connectives. As before, we let P and Q denote fixed statements.

The *implication* or *conditional* statement $P \Rightarrow Q$ is the statement "if P, then Q", or "P implies Q". The intended mathematical meaning of this connective, however, may surprise you. When should the statement $P \Rightarrow Q$ be true? Certainly the English usage of the phrase "if ..., then" conjures up the mental phrase "if P is true, then Q must also be true". However, mathematicians also consider the statement $P \Rightarrow Q$ to be true when P is false, regardless of the truth value of Q. One way to think about this is to think about when $P \Rightarrow Q$ "should" be false, namely, only when P is true and Q is false. As with the other connectives, we summarize this information in a truth table (see Table 1.5).

Table 1.5. Truth table for $\Rightarrow$.

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

In the statement $P \Rightarrow Q$, P is called the *hypothesis*, or *antecedent*, and Q is called the *conclusion* or *consequent.*

There are several other English phrases that are always interpreted to mean $P \Rightarrow Q$, or P implies Q, which are given in Table 1.6. It is important to note, therefore, that the words *if, only if, necessary, sufficient* (as well as *and* and *or*) have particular mathematical meanings, and so we must take care to use and interpret these words correctly.

Table 1.6. Alternative expressions for $P \Rightarrow Q$.

If P, then Q	Q when P
P only if Q	Q if P
P is sufficient for Q	Q is necessary for P

For our final logical connective (here, again, P and Q are fixed statements), the *biconditional* statement $P \Leftrightarrow Q$ is an abbreviation for the compound statement $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$; the truth table is given in Table 1.7.

Table 1.7. Truth table for $\Leftrightarrow$.

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$P \Leftrightarrow Q$
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

As with the conditional, there are several English phrases which can denote the biconditional (see Table 1.8). In particular, note that we interpret the $\Leftarrow$ of $P \Leftrightarrow Q$ as *if*, while we interpret the $\Rightarrow$ as *only if.*

Table 1.8. Alternative expressions for $P \Leftrightarrow Q$.

P if and only if Q
P iff Q
P is equivalent to Q
P exactly when Q
P is necessary and sufficient for Q

Example 1.1.10. Determine whether the following statements are true or false.

(1) If $7+6=14$, then $5+5=10$.

Since $7+6=14$ is false, we see that this implication is true.

(2) p is odd is necessary for p to be prime. (Here, think of p as a fixed positive integer.)

In general, we will find it easier to interpret this statement if we rephrase it as an "if . . . , then" statement using Table 1.6. In this form, the statement becomes

$$\text{if } p \text{ is prime, then } p \text{ is odd.}$$

Note again that we are relying on knowing the definitions of "prime" and "odd" (see Definitions 2.1.7 and 1.2.1). Here we can see that the truth value of the implication will depend on which positive integer p is.

If $p=2$, then the hypothesis "p is prime" is true, but the conclusion "p is odd" is false. In this case, the implication is false.

If $p=3$, then the hypothesis "p is prime" is true and also the conclusion "p is odd" is true. In this case, the implication is true. Here we can also see that mathematical implication has nothing to do with "causality". The fact that 3 is odd is not "caused" by the fact that 3 is prime.

If $p=4$ or $p=9$, then the hypothesis "p is prime" is false. Thus, in these cases, the implication is true.

We invite you to determine exactly for which positive integers p the given statement is true.

(3) $|x|=1$ iff $x=1$ or $x=-1$. (Here, think of x as a fixed real number.)

This biconditional is a true statement. When x is a fixed real number, regardless of its value, the two statements $|x|=1$ and $(x=1)\vee(x=-1)$ have the same truth value (see Proposition 1.1.11(3) and Definition 2.1.11). ◊

We will find it useful to have more than one way of thinking of certain statements.

Proposition 1.1.11. *Let P and Q be statements.*

(1) *$P \Rightarrow Q$ is logically equivalent to $(\neg P)\vee Q$.*

(2) *$\neg(P \Rightarrow Q)$ is logically equivalent to $P \wedge (\neg Q)$.*

(3) *$P \Leftrightarrow Q$ is true exactly when P and Q have the same truth value.*

Proof. See Exercise 1.1.2e, Exercise 1.1.2f, and Table 1.7. Use truth tables. □

Proposition 1.1.11(2) indicates how to find a useful denial of an implication.

Example 1.1.12. Find a useful denial of the statement

$$n \text{ is prime only if } n=2 \text{ or } n \text{ is odd.}$$

(Assume that n is a fixed positive integer.)

As before, we should first rewrite the statement as an "if . . . , then" statement; i.e.,

$$n \text{ is prime } \Rightarrow (n=2 \text{ or } n \text{ is odd}).$$

By Proposition 1.1.11(2) and Example 1.1.9, a useful denial of the given statement is, in natural English,

n is prime, and n is an even positive integer other than 2. ◊

1.1.2. Statements related to $P \Rightarrow Q$.

We consider now two statements related to the implication $P \Rightarrow Q$.

Definition 1.1.13. Let P and Q be statements.

(1) The *converse* of $P \Rightarrow Q$ is the statement $Q \Rightarrow P$.

(2) The *contrapositive* of $P \Rightarrow Q$ is the statement $(\neg Q) \Rightarrow (\neg P)$.

We illustrate these ideas using a familiar implication from calculus.

Example 1.1.14. Let f be a fixed real-valued function defined on some collection of real numbers (i.e., the type of function one considers in calculus), and let a be a fixed real number. Consider the conditional statement

(1.1) if f is differentiable at a, then f is continuous at a.

The converse of this statement is

(1.2) if f is continuous at a, then f is differentiable at a.

The contrapositive is

(1.3) if f is not continuous at a, then f is not differentiable at a. ◊

It is important to know the difference between the converse and the contrapositive of an implication, as the truth table in Table 1.9 shows.

Table 1.9. Truth table for $P \Rightarrow Q$, $Q \Rightarrow P$, $(\neg Q) \Rightarrow (\neg P)$.

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

Table 1.9 shows that the statement $P \Rightarrow Q$ and its contrapositive $(\neg Q) \Rightarrow (\neg P)$ are logically equivalent. We also see that the implication $P \Rightarrow Q$ is not logically equivalent to its converse $Q \Rightarrow P$, since when $P \Rightarrow Q$ is true, $Q \Rightarrow P$ may be true or false, depending on the truth values of P and Q. We have thus proved the following proposition.

Proposition 1.1.15. *Let P and Q be statements.*

(1) *$P \Rightarrow Q$ is logically equivalent to $(\neg Q) \Rightarrow (\neg P)$.*

(2) *$P \Rightarrow Q$ is not logically equivalent to $Q \Rightarrow P$.*

Consider Example 1.1.14 again. We learn in calculus that statement (1.1) is a true statement, regardless of which fixed function f and real number a we consider. Hence, the contrapositive statement (1.3) is also true, regardless of which fixed function f and real number a we consider. We also learn in calculus that the truth or falsity of the converse of (1.1), statement (1.2), depends on *which* fixed function f and real number a we consider. To prove that statement (1.2) does not hold *for all* functions f and *for all* real numbers a, we must demonstrate a particular function f and a particular real number a such that f is continuous at a but not differentiable at a (see Exercise 1.1.6). We discuss the quantifier "for all", and how it is negated, in Subsection 1.1.3.

1.1.3. Quantifiers. Recall that the statement "$n + 1 > 3$" on its own is not a *proposition* because it doesn't have a truth value. Instead, it is a *predicate* because it becomes a proposition when the "free variable" n is replaced by a particular value from the universe in question.

Let $P(n)$ denote the predicate "$n + 1 > 3$" (the notation makes explicit the fact that n is a free variable). As an example, we'll also assume that the universe over which n can range is $\mathbb{N} = \{1, 2, 3, \dots\}$, the set of all natural numbers (also called the set of positive integers). Then $P(2)$ is the proposition "$2+1 > 3$", which is false, and $P(7)$ is the proposition "$7 + 1 > 3$", which is true.

Another way to turn a predicate into a statement with a truth value is to modify it with a *quantifier.*

Definition 1.1.16. Let $\mathcal{U}$ be the universe under consideration and $P(x)$ be a predicate whose only free variable is x. Then the statements

$$\begin{array}{rl} \text{“for all } x,\ P(x)\text{”} & \qquad \text{notation: } (\forall x)P(x), \\ \text{“there exists } x \text{ such that } P(x)\text{”} & \qquad \text{notation: } (\exists x)P(x) \end{array}$$

are statements that have truth values.

The symbol $\forall$ is called the *universal quantifier.* The statement $(\forall x)P(x)$ is true exactly when each individual element a in the universe $\mathcal{U}$ has the property that $P(a)$ true.

The symbol $\exists$ is called the *existential quantifier.* The statement $(\exists x)P(x)$ is true exactly when the universe $\mathcal{U}$ contains at least one element a with $P(a)$ true.

We can make the universe $\mathcal{U}$ explicit by writing $(\forall x \in \mathcal{U})P(x)$ instead of $(\forall x)P(x)$, and similarly by writing $(\exists x \in \mathcal{U})P(x)$ instead of $(\exists x)P(x)$. We read the notation $(\forall x \in \mathcal{U})$ as "for all x in $\mathcal{U}$" and $(\exists x \in U)$ as "there exists x in $\mathcal{U}$". The symbol $\in$ is used in order to denote an element of a set; for example, the statement $3 \in \mathbb{N}$ says that 3 is a natural number, while the statement $\frac{1}{2} \notin \mathbb{N}$ says that $\frac{1}{2}$ is not a natural number. We discuss this concept in more depth in Chapter 4.

Example 1.1.17. Determine whether the following statements are true or false.

(1) There exists a natural number n such that $n + 1 > 3$.

The statement $(\exists n \in \mathbb{N})(n + 1 > 3)$ is true because $7 \in \mathbb{N}$ and $7 + 1 > 3$ is true.

(2) For all real numbers x, $x^2 \geq 0$.

We'll use $\mathbb{R}$ to denote the set of all real numbers. You've probably learned in a previous math course that the statement $(\forall x \in \mathbb{R})(x^2 \geq 0)$ is true; i.e., the square of a real number is never negative. (See Exercise 2.1.6, which requests a proof of this statement.)

(3) For all natural numbers n, $n+1>3$.

The statement $(\forall n \in \mathbb{N})(n+1>3)$ is false because 2 is a natural number, but $2+1>3$ is false. Here, the natural number 2 is called a *counterexample* to the statement $(\forall n \in \mathbb{N})(n+1>3)$; this means that it is an example that shows that the universal statement $(\forall n \in \mathbb{N})(n+1>3)$ is false.

(4) $(\forall x \in \mathbb{R})(x^3+52x^2+79x+1000 \geq 0)$

Here we need to think a bit before proceeding. We learn in calculus that the power function x^3 "grows faster" than the power function x^2 when x is large. Thus, when x is large negative, we expect that the polynomial $x^3+52x^2+79x+1000$ should be negative; i.e., we conjecture that we should be able to find a counterexample that shows the given statement is false. If we try $x=-2$, then we compute

$$(-2)^3+52\cdot(-2)^2+79(-2)+1000=1042,$$

which tells us nothing since $1042 \geq 0$. In other words, -2 is not the counterexample we seek. However, it is important to note that *this computation does not show that the given statement is true.* We try another value of x, such as $x=-100$. We compute

$$(-100)^3+52\cdot(-100)^2+79(-100)+1000=-486900<0.$$

Thus we have successfully found a counterexample that shows that the statement $(\forall x \in \mathbb{R})(x^3+52x^2+79x+1000 \geq 0)$ is false. $\Diamond$

Note that Definition 1.1.16 states that in order to determine the truth value of a statement $(\forall x)P(x)$ or $(\exists x)P(x)$, the universe over which x can range must be known. A quick example illustrates why. The statement $(\exists x \in \mathbb{N})(2x=3)$ is false because the equation $2x=3$ has no solution in the natural numbers. However, the statement $(\exists x \in \mathbb{R})(2x=3)$ is true because $\frac{3}{2}$ is a real number and $2\cdot\frac{3}{2}=3$.

For convenience, we collect in Table 1.10 the definitions and notation for the various number universes $\mathcal{U}$ we will consider in this text, namely, the natural numbers, the integers, the rational numbers, and the real numbers. Note that in some texts, 0 is considered to be a natural number, so it is important to check the definition in whatever source you are using. Note also that because every rational number has infinitely many representations (for example, $\frac{1}{2}=\frac{3}{6}=\frac{-2}{-4}=\cdots$), our definition of $\mathbb{Q}$ is informal. Defining $\mathbb{Q}$ more carefully, including checking that the arithmetic operations are "well-defined", is often addressed in an abstract algebra course (see also Exercise 7.2.9). Similarly, our definition of what it means to be a real number is informal only. Formally defining the concept of a real number is often addressed in courses such as real analysis or set theory; see Chapter 9.

Table 1.10. Number universes.

natural numbers $\mathbb{N}$:	$\mathbb{N}=\{1,2,3,\dots\}$
integers $\mathbb{Z}$:	$\mathbb{Z}=\{\dots,-3,-2,-1,0,1,2,3,\dots\}$
rational numbers $\mathbb{Q}$:	$x \in \mathbb{Q}$ if there exist a, $b \in \mathbb{Z}$, $b \neq 0$, such that $x=\frac{a}{b}$
real numbers $\mathbb{R}$:	*informally,* $x \in \mathbb{R}$ if x has a decimal expansion

Sometimes we will denote the set of positive integers by $\mathbb{Z}^+$. Similarly, $\mathbb{Q}^+$ denotes the set of positive rational numbers, etc.

Next, we need to determine how the quantifiers interact with negation. Here, the usual English usage of these phrases gives us exactly the right idea. The negation of the statement "all members of this class have brown hair" is "at least one member of this class doesn't have brown hair". Here we see that the universal quantifer (all members of this class) became an existential quantifier (at least one member of this class), and the statement "has brown hair" was negated to become "doesn't have brown hair". Similarly, the negation of the statement "there is a member of this class who is left-handed" is "every member of this class is right-handed".

Proposition 1.1.18. *Let $P(x)$ be a predicate and let $\mathcal{U}$ be the intended universe. Then:*

(1) $\neg(\forall x)P(x)$ *is logically equivalent to* $(\exists x)(\neg P(x))$*; i.e.,* $\neg(\forall x \in \mathcal{U})P(x)$ *is logically equivalent to* $(\exists x \in \mathcal{U})(\neg P(x))$.

(2) $\neg(\exists x)P(x)$ *is logically equivalent to* $(\forall x)(\neg P(x))$*; i.e.,* $\neg(\exists x \in \mathcal{U})P(x)$ *is logically equivalent to* $(\forall x \in \mathcal{U})(\neg P(x))$.

Proof. See Exercise 1.1.8. Use Definition 1.1.16. □

Proposition 1.1.18 tells us how to find a useful denial of a quantified statement.

Example 1.1.19. Find a useful denial of the statement

for all real numbers x, if $x > 2$, then $x^2 > 4$.

Remember that finding a useful denial of a statement means to express the negation of the statement positively. At first, you may find this type of exercise easier if you first express the statement using a mixture of English and mathematical notation, and then proceed one step at a time. The statement

$\neg(\forall x \in \mathbb{R})[x > 2 \Rightarrow x^2 > 4]$ is equivalent to

$(\exists x \in \mathbb{R})[\neg(x > 2 \Rightarrow x^2 > 4)]$, by Proposition 1.1.18(1),

which is equivalent to

$(\exists x \in \mathbb{R})[x > 2 \wedge x^2 \leq 4]$ by Proposition 1.1.11(2).

Thus, a useful denial of the statement

for all real numbers x, if $x > 2$, then $x^2 > 4$

is

there exists a real number x such that $x > 2$ and $x^2 \leq 4$. ◊

Notice in the example above that we did not allow ourselves to be distracted by the truth or falsity of the statements.

We conclude this subsection with a final comment about statements such as $(\forall x \in \mathcal{U})P(x)$, which explicitly indicate the universe $\mathcal{U}$ under consideration. The notation $(\forall x \in \mathcal{U})$ is called a *modified quantifier*, since we have modified the universal quantifier $(\forall x)$. The notation

$$(\forall x \in \mathcal{U})P(x)$$

is actually an abbreviation for the statement

$$(\forall x)(x \in \mathcal{U} \Rightarrow P(x)).$$

Similarly, the notation

$$(\exists x \in \mathcal{U})P(x)$$

is an abbreviation for

$$(\exists x)(x \in \mathcal{U} \wedge P(x)).$$

See Exercise 1.1.8b.

Quantifiers that are modified in other ways, such as the modification seen in the statement $(\forall x > 2)(x^2 > 4)$, follow the same rules. For example,

$$(\forall x > 2)(x^2 > 4)$$

is an abbreviation for

$$(\forall x)(x > 2 \Rightarrow x^2 > 4),$$

and

$$(\exists x < 2)(x^2 > 4)$$

is an abbreviation for

$$(\exists x)(x < 2 \wedge x^2 > 4).$$

1.1.4. A warning: hidden and implied quantifiers. From the beginning of this chapter, we have emphasized the language of mathematics. In order for us to understand and be understood, we must use that language (including its notation) precisely to say exactly what we mean, no more and no less. However, there are several situations in mathematical practice where the mathematical language may imply more than it says explicitly.

One important example of this is illustrated by the following statement. The universe here is the set $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ of integers.

(1.4) If n is odd, then $n + 1$ is even.

We have emphasized already that this statement is not a proposition (i.e., it doesn't have a truth value) but rather a predicate. However, the custom in mathematics is to treat an isolated implication such as this one as a *universal* statement, even though the universal quantifier $\forall$ is not explicitly mentioned. This is because n is to be treated as a fixed, but *arbitrary*, integer. Thus, the statement "if n is odd, then $n + 1$ is even" is often interpreted as

(1.5) $(\forall n \in \mathbb{Z})[\text{if } n \text{ is odd, then } n + 1 \text{ is even}]$.

Recognizing the hidden quantifiers is also important when one needs to negate a statement. If an author had intended for statement (1.4) to be interpreted as statement (1.5) (and one should be able to tell this from the context), then one needs to make the quantifiers explicit first in order to find the correct negation:

$$(\exists n \in \mathbb{Z})[n \text{ and } n + 1 \text{ are both odd}].$$

Again, do not be distracted by the truth or falsity of these statements.

So, to repeat, *an implication involving one or more free variables is often treated as a universally quantified statement.*

In addition, quantifiers may "hide" in other ways, such as in the definition of mathematical terms like "even". For example, the statement that "12 is an even integer" means that 12 is divisible by 2, which itself means that there is an integer k such that $2k = 12$ (here, $k = 6$). The phrase "12 is even" *hides* an existential quantifier, which is important to recognize.

Another example of hidden quantifiers occurs in the statement that "$\sqrt{2}$ is irrational", i.e., that "$\sqrt{2}$ is not a rational number". The statement "$\sqrt{2}$ *is* rational" means that there exist positive integers p and q such that $\sqrt{2} = \frac{p}{q}$. We can therefore express "$\sqrt{2}$ is irrational" using notation as

$$\neg(\exists p \in \mathbb{Z}^+)(\exists q \in \mathbb{Z}^+)\left[\sqrt{2} = \frac{p}{q}\right]$$

or

$$(\forall p \in \mathbb{Z}^+)(\forall q \in \mathbb{Z}^+)\left[\sqrt{2} \neq \frac{p}{q}\right].$$

Exercises 1.1

1. Let P, Q, and R be statements. Determine whether or not the two expressions in each pair are logically equivalent. In each case, demonstrate that your answer is correct.
   (a) $(P \wedge Q) \wedge R$, $\quad P \wedge (Q \wedge R)$.
   (b) $(P \vee Q) \vee R$, $\quad P \vee (Q \vee R)$.
   (c) $(P \wedge Q) \vee R$, $\quad P \wedge (Q \vee R)$.
   (d) $(P \vee Q) \wedge R$, $\quad P \vee (Q \wedge R)$.
2. Let P, Q, and R be statements. Show that the following statements are logically equivalent.
   (a) $\neg(\neg P)$ and P.
   (b) $(P \vee Q) \wedge R$ and $(P \wedge R) \vee (Q \wedge R)$.
   (c) $(P \wedge Q) \vee R$ and $(P \vee R) \wedge (Q \vee R)$.
   (d) $\neg(P \vee Q)$ and $(\neg P) \wedge (\neg Q)$.
   (e) $P \Rightarrow Q$ and $(\neg P) \vee Q$.
   (f) $\neg(P \Rightarrow Q)$ and $P \wedge (\neg Q)$.
3. Let P, Q, and R be statements. Determine whether or not the two expressions in each pair are logically equivalent. In each case, demonstrate that your answer is correct.
   (a) $(P \Rightarrow Q) \Rightarrow R$, $\quad P \Rightarrow (Q \Rightarrow R)$.
   (b) $(P \vee Q) \Rightarrow R$, $\quad (P \Rightarrow R) \vee (Q \Rightarrow R)$.
   (c) $(P \wedge Q) \Rightarrow R$, $\quad (P \Rightarrow R) \wedge (Q \Rightarrow R)$.
   (d) $P \Rightarrow (Q \vee R)$, $\quad (P \Rightarrow Q) \vee (P \Rightarrow R)$.
   (e) $P \Rightarrow (Q \wedge R)$, $\quad (P \Rightarrow Q) \wedge (P \Rightarrow R)$.
4. Propositions which are "always true" (respectively, "always false") are called tautologies (respectively, contradictions). More precisely:

Definition 1.1.20. A *tautology* is a proposition that is true for every possible assignment of truth values to the statement letters that occur in it. A *contradiction* is a proposition that is false for every possible assignment of truth values to the statement letters that occur in it.

Let P and Q be statements. Determine whether each of the following statements is a tautology, a contradiction, or neither.

(a) $P \Leftrightarrow \neg(\neg P)$.
(b) $P \wedge \neg P$.
(c) $P \vee \neg P$.
(d) $(P \wedge Q) \vee (\neg P \wedge \neg Q)$.
(e) $P \Rightarrow (Q \Rightarrow P)$.

5. Let n be a fixed positive integer. Which of the following statements are true, regardless of which positive integer n you consider? Explain *briefly*. (While you are not being asked to provide a proof, try to explain clearly.)
   (a) If n is divisible by 6, then n is divisible by 3.
   (b) If n is divisible by 3, then n is divisible by 6.
   (c) If n is divisible by 6, then n is divisible by 9.
   (d) If n is divisible by 9, then n is divisible by 6.
   (e) If n is divisible by 6, then n^2 is divisible by 6.
   (f) If n^2 is divisible by 6, then n is divisible by 6.
   (g) If n^2 is divisible by 9, then n is divisible by 9.
   (h) If n is divisible by 2 and n is divisible by 3, then n is divisible by 6.
   (i) If n is divisible by 2 and n is divisible by 6, then n is divisible by 12.

6. Give an example which shows that statement (1.2) does not hold for all functions f and for all real numbers a; i.e., give a specific example of a function f defined on the real numbers, and a specific real number a, such that f is continuous at a but not differentiable at a.

7. For each of the following statements, give an example of a mathematical universe in which the statement is true and an example of a universe in which the statement is false. Explain why your answers are correct.
   (a) $(\forall x)[0 < x^2 < 2 \Rightarrow x = 1]$.
   (b) $(\forall x)[0 < x^2 < 2 \Rightarrow (x = 1 \vee x = -1)]$.

8. Let $P(x)$ be a predicate whose only free variable is x and let $\mathcal{U}$ be the intended universe.
   (a) Use Definition 1.1.16 to explain why the following statements are logically equivalent (i.e., have the same truth value).
      (i) $\neg(\forall x)P(x)$ and $(\exists x)(\neg P(x))$.
      (ii) $\neg(\exists x)P(x)$ and $(\forall x)(\neg P(x))$.
   (b) Use Proposition 1.1.11, Proposition 1.1.18, and the definitions of the modified quantifiers on page 12 to show that the following statements are logically equivalent.
      (i) $\neg(\forall x \in \mathcal{U})P(x)$ and $(\exists x \in \mathcal{U})(\neg P(x))$.
      (ii) $\neg(\exists x \in \mathcal{U})P(x)$ and $(\forall x \in \mathcal{U})(\neg P(x))$.

9. Let $\mathcal{U}$ be the universe under consideration, and let $P(x)$ and $Q(x)$ be predicates with free variable x. Find a *useful denial* (i.e., a statement equivalent to the negation) of each statement.

(a) $(\forall x \in \mathcal{U})(P(x) \Rightarrow Q(x))$.
(b) $(\forall x \in \mathcal{U})(Q(x) \vee P(x))$.
(c) $(\exists x \in \mathcal{U})(Q(x) \wedge P(x))$.
(d) $(\exists x \in \mathcal{U})(Q(x) \wedge P(x))$. (Use an implication in your answer.)

10. Express each statement symbolically, including a quantification of all variables which makes the universe explicit. Negate the symbolic statement, and express the negation in natural language as a useful denial.
(a) The inequality $x^2 - 4x + 3 < 0$ has a real solution.
(b) The curves $y = 1 - x^2$ and $y = 3x - 2$ intersect.
(c) Every positive real number has a real square root. (Do not use the symbol $\sqrt{\ }$ in your solution.)

11. Which of the following statements are true? Explain *briefly*. (While you are not being asked to provide a proof, try to explain clearly.)
(a) There exists an integer n such that $4n + 5 = 7n + 3$.
(b) There exists a real number x such that $x^2 + 8x + 12 \geq 0$.
(c) For all real numbers x, $x^2 + 8x + 12 \geq 0$.
(d) For all real numbers x, $x^2 + 8x + 17 \geq 0$.

12. Which of the following statements are true? Explain *briefly*. (While you are not being asked to provide a proof, try to explain clearly.)
(a) For all real numbers x, $x^2 - 2x - 3 = 0$ only if $x = 3$.
(b) For all real numbers x, $x^2 - 2x - 3 = 0$ if $x = 3$.

13. Find the converse and contrapositive of each statement. Here, $\mathbf{v_1}, \mathbf{v_2}, \mathbf{0}$ are fixed vectors in a real vector space (you don't need to know what this means!), x_1, x_2 are fixed real numbers, and f is a fixed function.
(a) $(x_1\mathbf{v_1} + x_2\mathbf{v_2} = \mathbf{0}) \Rightarrow (x_1 = 0 \wedge x_2 = 0)$.
(b) $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$.

14. Find a *useful denial* (i.e., a statement logically equivalent to the negation) of each statement. Here, $\mathbf{v_1}, \mathbf{v_2}, \mathbf{0}$ are fixed vectors in a real vector space (you don't need to know what this means!) and f is a fixed function.
(a) $(\forall x_1)(\forall x_2)[(x_1\mathbf{v_1} + x_2\mathbf{v_2} = \mathbf{0}) \Rightarrow (x_1 = 0 \wedge x_2 = 0)]$.
(b) $(\forall x_1)(\forall x_2)[f(x_1) = f(x_2) \Rightarrow x_1 = x_2]$.
(c) $(\forall y)(\exists x)[y = f(x)]$.

15. Find a *useful denial* (i.e., a statement logically equivalent to the negation) of each statement, and express it in mathematically precise, natural English. Express all conditional statements in the form "if ..., then ...". Do not add implied quantifiers. (Here, a, b, c, n are fixed integers, f is a fixed function, and x_0, L are fixed real numbers.)
(a) n is not a multiple of 4 if n is even.
(b) If f has a relative maximum at x_0 and f is differentiable at x_0, then $f'(x_0) = 0$.
(c) For every integer m, m^2 is odd and $m^3 - 1$ is divisible by 4.
(d) For all integers j and k, if $n = jk$, then $j = 1$ or $k = 1$.
(e) If n is a perfect square, then there exists an integer k such that $n = 3k$ or $n = 3k + 1$.
(f) bc is divisible by a only if b is divisible by a or c is divisible by a.

(g) For all $\varepsilon > 0$ there exists $\delta > 0$ such that for all x, $|f(x) - L| < \varepsilon$ if $0 < |x - x_0| < \delta$.

(h) For every real number M there exists a real number x such that $f(x) > M$.

(i) If n is an odd integer, then there exists an integer k such that $n = 4k + 1$ or $n = 4k + 3$.

16. Write the converse and contrapositive of each conditional statement. Do not add implied quantifiers. Express all conditional statements in the form "if ..., then ...". (Here, n is a fixed integer, x is a fixed real number, S is a fixed set of real numbers, $\{a_n\}$ is a fixed sequence of real numbers, and G is a group. In this exercise, it is not necessary to know the meanings of any mathematical concepts we have not yet defined.)

(a) If $x > 1$ or $x < -1$, then $x^2 > 1$.

(b) n^2 is a multiple of 3 is sufficient for n to be a multiple of 3.

(c) S is closed and bounded is necessary for S to be compact.

(d) $\{a_n\}$ converges if $\{a_n\}$ is bounded and monotone.

(e) $\{a_n\}$ is Cauchy only if $\{a_n\}$ converges.

(f) If n is an odd integer, then there exists an integer k such that $n = 4k + 1$ or $n = 4k + 3$.

(g) If G is abelian, then every subgroup of G is normal.

(h) If n is a perfect square, then there exists an integer k such that $n = 3k$ or $n = 3k + 1$.

1.2. Proof

1.2.1. Logical arguments. So far we have concentrated on the language, notation, and grammar of mathematical statements. Now we move on to our goal of learning to construct clear and correct mathematical proofs.

What is a proof? Informally, we will define a mathematical proof to be a logical argument that establishes the truth of a mathematical statement.

What is a logical argument? We'll first consider the following familiar example from calculus.

Suppose that f is a fixed function defined on a subset of the real numbers and that a is a fixed real number. Suppose you also know:

- If f is differentiable at a, then f is continuous at a.
- f is differentiable at a.

What logical conclusion can you draw? That f is continuous at a. While you probably came to this conclusion without thinking too much about it, technically you constructed a valid logical argument using the "rule of deduction" called *modus ponens*. If we represent the statement "f is differentiable at a" by the statement letter P and the statement "f is continuous at a" by Q, then *modus ponens* says "from P and $P \Rightarrow Q$, deduce Q". We can see that it is reasonable to adopt *modus ponens* as a rule of deduction by looking again at the truth table for $P \Rightarrow Q$ in Table 1.11.

Table 1.11. Truth table for ⇒.

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

Modus ponens says that we begin by knowing that $P \Rightarrow Q$ is true and P is true. Only the first line in the truth table corresponds to this situation; in this line, Q is also true. It follows that "Q is true" is a valid conclusion, based on the hypotheses that P is true and $P \Rightarrow Q$ is true. Thus, the form of a logical argument is based on the logic of our connectives and on the logic of our quantifiers.

If we wish to prove a mathematical statement, then we must first determine the logical form of that statement.

1.2.2. Direct proofs, an introduction. We begin by considering statements of the form $P \Rightarrow Q$. To determine how we might prove that a statement with this logical form is true, we again look at the truth table for implication in Table 1.11. We see that the statement $P \Rightarrow Q$ is automatically true when P is false, so there is no need to consider this situation. In other words, we should begin by assuming that P is true. The first line in the truth table Table 1.11 then shows us how to proceed: we should demonstrate that Q is true. This type of logical argument is called a *direct* proof of the statement $P \Rightarrow Q$. It is so fundamental in mathematics that we emphasize it again in Table 1.12.

Table 1.12. Direct proof of $P \Rightarrow Q$.

To prove a statement of the form $P \Rightarrow Q$ is true (directly).
• We begin with "Assume P is true." • We must then demonstrate that Q is true.

Next, we ask how to prove a statement of the form $(\forall x)P(x)$ is true. Recall that there is an underlying universe $\mathcal{U}$ corresponding to the universal quantifier. Definition 1.1.16 tells us that the statement $(\forall x)P(x)$ is true exactly when every element a in the universe $\mathcal{U}$ has the property that $P(a)$ is true. In general, it's not possible to show $P(a)$ is true for each element a in the universe individually. Indeed, when the universe is infinite, there is no way to do this, since a proof must be finite. In a direct proof of $(\forall x)P(x)$, therefore, we demonstrate that if x is an *arbitrary, fixed* element of the universe, then $P(x)$ is true. See Table 1.13.

Definition 1.1.16 also tells us how to prove (directly) that a statement of the form $(\exists x)P(x)$ is true. See Table 1.14.

Let's begin by proving the following:

The sum of an even integer and an odd integer is odd.

Table 1.13. Direct proof of $(\forall x)P(x)$.

To prove a statement of the form $(\forall x)P(x)$ is true (directly).
• We begin with "Let x be an arbitrary (but now *fixed*), element of the universe."
• We must then demonstrate that $P(x)$ is true.

Table 1.14. Direct proof of $(\exists x)P(x)$.

To prove a statement of the form $(\exists x)P(x)$ is true (directly).
• We must *find* an element a in the universe such that $P(a)$ is true.
• In other words, we must explicitly *say* what a is *and* *demonstrate* that $P(a)$ is true.

We cannot proceed until we are sure that we know what the words mean. While you certainly know intuitively what the words "even" and "odd" mean, a mathematical proof that a particular undetermined integer is even or odd relies on knowing the precise mathematical definition of these words. Without knowing the logical structure of these definitions, we will not know what must be proved.

Definition 1.2.1. Let $n \in \mathbb{Z}$.

(1) n is *even* if there exists an integer k such that $n = 2k$; i.e.,

$$n \text{ is even if } (\exists k \in \mathbb{Z})[n = 2k].$$

(2) n is *odd* if there exists an integer k such that $n = 2k + 1$; i.e.,

$$n \text{ is odd if } (\exists k \in \mathbb{Z})[n = 2k + 1].$$

We should make two observations about this formal definition right away. First, it is almost universal in mathematics to use the word "if" in a definition when what is actually meant is "if and only if". For example, technically Definition 1.2.1 says only that, for a given integer n, *if* we know that there exists $k \in \mathbb{Z}$ such that $n = 2k$, *then* we can conclude that n is even. Although not explicitly stated in Definition 1.2.1, it is further understood that *if* we know that n is even, *then* there exists $k \in \mathbb{Z}$ such that $n = 2k$. While this completely contradicts our policy of always saying exactly what we mean, it is standard practice in mathematics that definitions have this special status, and so we may as well get used to it now.

Special status of definitions. *Mathematical definitions are always* **if and only if** *statements.*

Next, we should note that we will assume that every integer is either even or odd, but never both. You certainly believe this, but it actually requires a proof. We will make use of this assumption for now and justify it in Exercise 2.2.2 and Chapter 6.

Most importantly, we need to be sure that we recognize the logical structure of the statement to be proved, which we purposely stated first in colloquial English.

As a mathematical statement, "the sum of an even integer and an odd integer is odd" should be interpreted as follows:

> if m is an even integer and n is an odd integer, then $m + n$ is an odd integer.

Remember that we must be careful. There are assumed universal quantifiers here:

$$(\forall m \in \mathbb{Z})(\forall n \in \mathbb{Z})[(m \text{ is even and } n \text{ is odd}) \Rightarrow (m + n \text{ is odd})].$$

We should never expect to simply "write down" a proof of a statement; we will need to search for it. To organize our thoughts for this "scratchwork", we will use a "Given-Goal" diagram[1] to identify what is given and what is our goal. The universal quantifiers tell us to assume that m and n are arbitrary (and now fixed) integers.

Given	**Goal**
m, n arbitrary integers	if m is even and n is odd, then $m + n$ is odd

Right away, the logical structure of the goal, an implication, tells us what to do next.

The Goal dictates the form of the proof.

Table 1.12 tells us that we should *assume* the hypotheses that m is even and n is odd and then rewrite our goal:

Given	**Goal**
m, n arbitrary integers	
m is even	
n is odd	$m + n$ is odd

We must search for a logical way of getting to our *goal* from our *givens.* One useful way to search is with a "backward-forward" method. You should ask yourself the following questions:

What's my goal? What does it mean?

What's given? What does it mean?

Our job is to work back and forth between these ideas until we find the logical connections.

Our *Goal* is to show:

$$m + n \text{ is odd.}$$

Since the definition of *odd* is existential, Table 1.14 tells us that our goal is to

find a particular integer a such that $m + n = 2a + 1$.

[1] The notion of a "Given-Goal" diagram as a way of organizing one's thoughts regarding what is known, versus what is to be proved, was first used in Daniel J. Velleman's book *How to Prove It: A Structured Approach* [**15**]. It is also used in Peter J. Eccles's book *An Introduction to Mathematical Reasoning: Numbers, Sets and Functions* [**8**], where Velleman is credited with the terminology.

We are *Given* that

$$m \text{ is even.}$$

This means that

$$\textit{there exists} \text{ an integer } k \text{ such that } m = 2k.$$

Since such an integer *exists*, we'll *fix* one so that we can work with it; i.e., we can

$$\text{fix } i \in \mathbb{Z} \text{ such that } m = 2i.$$

Similarly, since we know

$$n \text{ is odd,}$$

we can

$$\text{fix } j \in \mathbb{Z} \text{ such that } n = 2j + 1.$$

Since we both have and want information about the integer $m + n$, it makes sense to investigate this quantity. Note that

$$m + n = 2i + (2j + 1) = (2i + 2j) + 1 = 2(i + j) + 1.$$

Since $i + j$ is an integer, we've found the integer a we are seeking.

We've convinced ourselves that the statement is true, but part of our job is to formally and effectively communicate a mathematical proof of this statement to others.

Proposition 1.2.2. *For all integers m and n, if m is even and n is odd, then $m + n$ is odd.*

Proof. Let m, $n \in \mathbb{Z}$ be arbitrary, and assume that m is even and n is odd. We show that $m + n$ is odd; i.e., we must find an integer a such that $m + n = 2a + 1$.

Since m is even, by Definition 1.2.1 we can fix $i \in \mathbb{Z}$ such that $m = 2i$. Similarly, since n is odd, by Definition 1.2.1 we can fix $j \in \mathbb{Z}$ such that $n = 2j + 1$. Then

$$\begin{aligned} m + n &= 2i + (2j + 1) \\ &= (2i + 2j) + 1 \\ &= 2(i + j) + 1, \end{aligned}$$

by the associative and distributive properties. Since $i + j$ is an integer, $m + n$ is odd, by Definition 1.2.1.

Hence, the sum of an even integer and an odd integer is odd. □

1.2.3. Some important observations. Despite the apparent simplicity of the statement and proof of Proposition 1.2.2, there are several important lessons that must be emphasized. First, in our scratchwork, we progressed from the statement

$$m \text{ is even}$$

to

$$(\exists k \in \mathbb{Z})(m = 2k)$$

to

$$\text{fix an integer } i \text{ such that } m = 2i.$$

Going from the existential statement $(\exists k \in \mathbb{Z})(m = 2k)$ (which simply asserts that something exists) to fixing a *particular* integer $i \in \mathbb{Z}$ with $m = 2i$ (which fixes a *particular example* of such an object and gives it a name) is called *existential instantiation.*

It's important here to use a new name (variable) that doesn't already have a particular meaning in your proof. To avoid wordiness in the final proof, we instantiated the existential quantifiers right away, *taking care to use a new variable each time.*

In fact, at the beginning of the final proof, the name (variable) k did not yet have a meaning. Consequently, we could have replaced the line

since m is even, by Definition 1.2.1 we can fix $i \in \mathbb{Z}$ such that $m = 2i$

by

since m is even, by Definition 1.2.1 we can fix $k \in \mathbb{Z}$ such that $m = 2k$.

Or, we could have replaced it by

since m is even, by Definition 1.2.1 we can fix $\ell \in \mathbb{Z}$ such that $m = 2\ell$.

However, once we choose a name to use to instantiate the first quantifier:

since m is even, by Definition 1.2.1 we can fix $i \in \mathbb{Z}$ such that $m = 2i$,

the meaning of the variable i becomes fixed for the rest of the proof, and *it would then be a mistake to say*,

since n is odd, by Definition 1.2.1 we can fix $i \in \mathbb{Z}$ such that $n = 2i + 1$.

Summarizing:

> If we *know* $(\exists x)P(x)$, where $P(x)$ is some predicate involving the free variable x, then we should *fix a particular* x such that $P(x)$ holds, as long as we take care *not* to use a variable whose meaning in the current proof is already fixed.

Next, note how essential the mathematical definitions of "even" and "odd" were in the proof. As we mentioned earlier, you almost certainly "knew" the definitions of these words already, at least in an intuitive sense. In mathematics, however, language must be used precisely and arguments must be rigorous. To prove that an integer is odd (or that something is a widget), we must have a precise mathematical definition of this concept, and the logical form of that definition indicates the structure of that proof.

Summarizing:

> If we wish to prove that something is a *widget* and all we know about widgets is the definition, then we must use the definition to prove it's a widget.
> The logical form of the definition of widget determines the structure of a proof that something is a widget.

It is also important to note the difference between the *search for the proof* (i.e., the scratchwork) and the mathematical proof we produced at the end. In this case, the scratchwork and the proof look pretty similar, but often it can take several approaches before you come up with the right idea for a proof. In fact, this is why reading mathematical proofs can seem difficult; you are reading the polished, finished product, and not the process by which the proof was discovered. Typically, mathematical proofs do not describe the process by which the proof was discovered (i.e., the scratchwork), although there are exceptions to this.

Furthermore, as we've emphasized from the beginning, we want to be sure to use mathematical language and notation correctly in mathematical writing. See the appendix for some suggested guidelines to help you write mathematics effectively. These guidelines review the general comments mentioned here, as well as address other issues.

Finally, it is important to note that we used some basic properties of integers in the proof of Proposition 1.2.2 (such as the associative property of addition and the distributive property). At the beginning of this section on proof, we noted that a proof is a logical argument and illustrated the role that *modus ponens* plays. Students who first learn about proof in mathematics often wonder what mathematical statements need proof, particularly when a statement "seems obvious" to them. Indeed, while our point of view in this course will be that "all" mathematical statements require proof, one cannot prove anything at all without some basic assumptions, which are often called *axioms*.

To give us a starting point, we will consider the statements found in Basic Properties of Integers 1.2.3 on page 23 (and also the *Principle of Mathematical Induction*, to be discussed in Chapter 3) to be our basic assumptions about the integers. We will accept these statements without proof, and you might take a moment to ask yourself whether you think it is reasonable for us to do so. (In fact, we could have taken a smaller list of statements as our basic assumptions about the integers and proved the rest from that smaller list.) All other statements we mention about integers, however, will require proof, unless we explicitly state otherwise. In this way, it should be very clear to you when a statement requires a proof. In general, we will never consider a mathematical statement, no matter how "simple" it may seem, as "obvious".

Basic Properties of Integers 1.2.3.

For all integers a, b, c:

(Closure under + and $\cdot$) $a + b$ and ab are also integers.

(Associative properties) $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$.

(Commutative properties) $a + b = b + a$ and $ab = ba$.

(Distributive property) $a(b + c) = ab + ac$.

(Identities) $0 \neq 1$, $a + 0 = a$, $a \cdot 1 = a$, and $a \cdot 0 = 0$.

(Additive inverses) There is a unique integer $-a = -1 \cdot a$ such that $a + (-a) = 0$.

(Subtraction) $b - a$ is defined to equal $b + (-a)$.

(No divisors of 0) If $ab = 0$, then $a = 0$ or $b = 0$.

(Cancellation) If $ab = ac$ and $a \neq 0$, then $b = c$.

(Transitive property of <) If $a < b$ and $b < c$, then $a < c$.

(Trichotomy) Exactly one of $a < b$ or $a = b$ or $a > b$ holds.

(Order property 1) If $a < b$, then $a + c < b + c$.

(Order property 2) If $c > 0$, then $a < b$ iff $ac < bc$.

(Order property 3) If $c < 0$, then $a < b$ iff $ac > bc$.

Note in particular the cancellation property of integers; there is no division operation in the integers. In order to use the cancellation property to conclude $b = c$ from $ab = ac$, where $a, b, c \in \mathbb{Z}$, we must first explain why $a \neq 0$.

Exercises 1.2

1. Let n be an integer.
   (a) Prove that if n is even, then n^2 is even.
   (b) Prove that if n is odd, then n^2 is odd.
2. Let m and n be integers.
   (a) Prove that if m and n are even, then $m + n$ is even.
   (b) Prove that if m and n are odd, then $m + n$ is even.
3. (a) Prove that for all $m \in \mathbb{Z}$, if m is even, then mn is even.
   (b) Prove that for all $m, n \in \mathbb{Z}$, if m and n are odd, then mn is odd.
4. Use the definition to prove that for all $n \in \mathbb{N}$, $4n + 7$ is odd.

Techniques of Proof

2.1. More direct proofs

In this chapter, we consider several examples that demonstrate various basic proof techniques. We begin by emphasizing again the importance of the *logical structure* of mathematical statements and definitions. The logical structure of a mathematical statement and, in particular, the logical structure of the *Goal* will dictate the form of its proof. The logical structure of a mathematical definition gives us a clear strategy for trying to establish that an object has (or doesn't have) that particular property.

In this section, we exhibit additional examples of direct proofs. For our first example, we introduce a new mathematical concept. The statement that an integer n is even is a statement about divisibility; an integer n is even if n is *divisible* by 2, or 2 *divides* n. We now define this concept more generally.

Definition 2.1.1. Let a, $b \in \mathbb{Z}$.

$$a \textit{ divides } b \text{ if there exists } n \in \mathbb{Z} \text{ such that } b = an.$$

We write $a \mid b$ for "a divides b"[1] and say that a is a *divisor* of b.

Remember the "special status" of definitions; the "if" in a *definition* (but not in other mathematical statements) is always read as "iff". We illustrate this new concept with some examples.

Example 2.1.2. Note that $3 \mid 12$ since there exists $n \in \mathbb{Z}$ such that $12 = 3n$; namely, $12 = 3 \cdot 4$.

On the other hand, $5 \nmid 12$ (i.e., 5 does not divide 12) since there does not exist an integer n such that $12 = 5n$.[2]

[1] Beware: Do not confuse this notation $a \mid b$ with the fraction notation $\frac{a}{b}$ or a/b used to denote division, which is not a legal operation in the integers. In particular, "$a \mid b$" is a complete sentence whose verb is "$|$", while "$\frac{a}{b}$" (and "a/b") is a noun; it is a name of a particular rational number.

[2] The fact that $5 \nmid 12$ may seem clear, but technically it requires proof. For now, let's simply note that $5 \nmid 12$ because $12 = 5 \cdot 2 + 2$; i.e., 5 "goes into" 12 twice with a "remainder" of 2. See page 40.

As another example, note that $12 \mid 72$ since $72 = 12 \cdot 6$. ♢

Note that in Example 2.1.2, we have $3 \mid 12$ and $12 \mid 72$. Also note that $3 \mid 72$, since $72 = 3 \cdot 24$. This is an example of the "transitive" property of the divisibility relation: for all integers a, b, c,

$$\boxed{\text{if } a \mid b \text{ and } b \mid c\text{, then } a \mid c.}$$

Our example is not a proof, however, so let's find a proof of this fact. As we did in Section 1.2, we organize our thoughts using a Given-Goal diagram.

Given	**Goal**
a, b, c arbitrary integers	
$a \mid b$	
$b \mid c$	$a \mid c$

Our *Goal* is to show

$$a \mid c.$$

Since the definition of *divides* is existential, our goal is to

$$\text{find an integer } k \text{ such that } c = ak.$$

(Note in particular our use of a *new* letter k here.) We can now replace the Given-Goal diagram above with the following.

Given	**Goal**
a, b, c arbitrary integers	
$a \mid b$	find $k \in \mathbb{Z}$
$b \mid c$	with $c = ak$

Now we consider our *Givens*. We know

$$a \mid b,$$

so we can

$$\text{fix } n \in \mathbb{Z} \text{ such that } b = an.$$

(Note our use of a *new* letter n here.)

Similarly, since

$$b \mid c,$$

we can

$$\text{fix } m \in \mathbb{Z} \text{ such that } c = bm$$

(and a new letter here).

We'll attempt to combine our given information into a single statement about c:

$$c = bm = (an)m = a(nm).$$

So nm is the integer k we seek. We have found a proof.

Proposition 2.1.3. *For all integers a, b, c, if $a \mid b$ and $b \mid c$, then $a \mid c$.*

Proof. Let a, b, $c \in \mathbb{Z}$ be arbitrary and assume that $a \mid b$ and $b \mid c$. We must show that $a \mid c$; i.e., we must find an integer k such that $c = ak$.

Since $a \mid b$, by Definition 2.1.1 we may fix $n \in \mathbb{Z}$ such that $b = an$. Similarly, since $b \mid c$, we may fix $m \in \mathbb{Z}$ such that $c = bm$, again by Definition 2.1.1. Then

$$\begin{aligned} c &= bm \\ &= (an)m \\ &= a(nm), \end{aligned}$$

since multiplication of integers is associative (Basic Properties of Integers 1.2.3). Since $nm \in \mathbb{Z}$, we have proved that $a \mid c$, by Definition 2.1.1, as desired. □

It is important to note that Proposition 2.1.3 describes a property of integers. The divisibility relation in Definition 2.1.1 is phrased in terms of integers. Despite the terminology "divides" in Definition 2.1.1, there is no division operation in $\mathbb{Z}$, as we noted in Section 1.2. Consequently, the proof of Proposition 2.1.3, or any statement about integer divisibility, mentions only *integers*, and never *fractions*.

So far we've been proving statements about integers. We next consider a statement about the real numbers; recall that we denote the set of all real numbers by $\mathbb{R}$. The list of axioms you may assume about the real numbers is given below. As with the Basic Properties of Integers 1.2.3, we could take a smaller list of statements from those given below as our basic assumptions about the real numbers and prove the rest from that smaller list.

Basic Properties of Real Numbers 2.1.4.

For all real numbers a, b, c:

(Closure under $+$, $\cdot$) $a + b$ and ab are also real numbers.

(Associative properties) $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$.

(Commutative properties) $a + b = b + a$ and $ab = ba$.

(Distributive property) $a(b + c) = ab + ac$.

(Identities) $0 \neq 1$, $a + 0 = a$, $a \cdot 1 = a$, and $a \cdot 0 = 0$.

(Additive inverses) There is a unique real number $-a = -1 \cdot a$ such that $a + (-a) = 0$.

(Subtraction) $b - a$ is defined to equal $b + (-a)$.

(Multiplicative inverses) If $a \neq 0$, then there is a unique real number $a^{-1} = \frac{1}{a}$ such that $a \cdot a^{-1} = a \cdot \frac{1}{a} = 1$.

(Division) When $a \neq 0$, $\frac{b}{a}$ is defined to equal $b \cdot \frac{1}{a}$.

(Transitive property of $<$) If $a < b$ and $b < c$, then $a < c$.

(Trichotomy) Exactly one of $a < b$ or $a = b$ or $a > b$ holds.

(Order property 1) If $a < b$, then $a + c < b + c$.

(Order property 2) If $c > 0$, then $a < b$ iff $ac < bc$.

(Order property 3) If $c < 0$, then $a < b$ iff $ac > bc$.

There is one additional axiom for $\mathbb{R}$, called the *Completeness Axiom.* Except for the following theorem about the existence of nth roots, which we will accept and use for now without proof, the Completeness Axiom will not be needed or discussed until Chapter 9.

Theorem 2.1.5. *Let $n \in \mathbb{Z}^+$.*

(1) *Assume n is even. Then every $x \in \mathbb{R}$ with $x \geq 0$ has a real "nth root"; i.e., when $x \geq 0$, there is a unique nonnegative real number denoted by $x^{1/n} = \sqrt[n]{x}$ which satisfies $(x^{1/n})^n = x$. Furthermore, for any $x \in \mathbb{R}$, $(x^n)^{1/n} = |x|$.*

(2) *Assume n is odd. Then every $x \in \mathbb{R}$ has a real "nth root"; i.e., for any $x \in \mathbb{R}$, there is a unique real number denoted by $x^{1/n} = \sqrt[n]{x}$ which satisfies $(x^{1/n})^n = x$. Furthermore, for any $x \in \mathbb{R}$, $(x^n)^{1/n} = x$.*

As examples, $\sqrt[3]{7}$ is the unique real number z such that $z^3 = 7$, and, for all $x \in \mathbb{R}$, $\sqrt{x^2} = |x|$. Theorem 2.1.5 will be an important tool in Chapter 5.

We also take a moment to recall that a real number z is *rational* if there exist integers a, b with $b \neq 0$ such that $z = \frac{a}{b}$. The sum and product of two rational numbers are also rational; the familiar formulas

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + cb}{bd} \quad \text{and} \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd},$$

where $a, b, c, d \in \mathbb{Z}$ and $b, d \neq 0$, establish this (note that we need the "No divisors of 0" axiom in Basic Properties of Integers 1.2.3 to conclude that $bd \neq 0$).

For our first proof about real numbers, we'll prove that

$$\boxed{\text{for all } a,\ b \in \mathbb{R} \text{ with } a < b < 0,\ a^2 > b^2.}$$

Our Given-Goal diagram contains the usual information.

Given	**Goal**
$a,\ b \in \mathbb{R}$	
$a < 0$	
$b < 0$	
$a < b$	$a^2 > b^2$

We *know* $a < b$. Since we *want* an inequality involving a^2 and b^2, we could try to multiply both sides of the inequality $a < b$ by a (respectively b) to obtain a statement about a^2 (respectively b^2). We'll need to be careful to use the order axioms in Basic Properties of Real Numbers 2.1.4 correctly.

Proposition 2.1.6. *For all $a,\ b \in \mathbb{R}$ with $a < b < 0$, $a^2 > b^2$.*

Proof. Let $a,\ b \in \mathbb{R}$ with $a < b < 0$. We prove that $a^2 > b^2$.

Multiplying both sides of the inequality $a < b$ by a gives $a^2 > ab$ by an order axiom (see Basic Properties of Real Numbers 2.1.4), since $a < 0$. Similarly, multiplying both sides of the inequality $a < b$ by b gives $ab > b^2$, since $b < 0$. Since $a^2 > ab$ and $ab > b^2$, we have $a^2 > b^2$ by the transitive property of the order relation, as desired. □

2.1.1. Counterexamples. The statement of Proposition 2.1.3 was motivated by a single example (Example 2.1.2), which wasn't much evidence! Typically, when faced with a mathematical statement, one wishes to determine whether the statement is true or false. How can one approach such a situation? One usually tries a variety of approaches. Does the statement involve familiar concepts, and do previously proved theorems about these concepts apply? Can one find examples of the statement, which can provide intuition about whether the statement is true and how one might go about finding a proof of it? If one really isn't sure whether the statement is true or false, one can alternately try to prove it and to *disprove* it (which means to show that its negation is true).

Our next example deals with prime integers, so we begin with the definition.

Definition 2.1.7. A positive integer p is *prime* if $p > 1$ and the only positive integer factors of p are 1 and p; i.e., p is *prime* if $p > 1$ and

$$(\forall a, b \in \mathbb{Z}^+)[p = ab \Rightarrow (a = 1 \text{ or } b = 1)]. \tag{2.1}$$

Example 2.1.8. Is the following statement true or false? If it is true, then prove it. If it is false, then disprove it.

For all positive integers n, $n^2 + n + 41$ is prime.

We begin by computing $n^2 + n + 41$ for several values of n:

$$1^2 + 1 + 41 = 43 \text{ is prime,}$$
$$2^2 + 2 + 41 = 47 \text{ is prime,}$$
$$3^2 + 3 + 41 = 53 \text{ is prime,}$$
$$4^2 + 4 + 41 = 61 \text{ is prime.}$$

So far it looks like we are building evidence that the statement is true. On the other hand, having many examples does not constitute a proof, and these examples give no indication for why the statement might be true (if it is). In fact, it seems unlikely that there is a formula (such as n^2+n+41) that always generates primes. Consequently, while all our examples seem to be in favor of the statement being true, we will try to disprove it.

The negation of

$$(\forall n \in \mathbb{Z}^+)(n^2 + n + 41 \text{ is prime}) \tag{2.2}$$

is

$$(\exists n \in \mathbb{Z}^+)(n^2 + n + 41 \text{ is not prime}), \tag{2.3}$$

and this is what we wish to prove. Statement (2.3) is an existential statement, so we know that we must find an example of a *particular* positive integer n such that $n^2 + n + 41$ is not prime. So far, all the values of n we have tried have resulted in a prime integer. If we seek a value of n so that $n^2 + n + 41$ is not prime, then we need to remember that a positive integer is not prime if we can factor it as a product of two positive integers, neither of which is 1, since this is the negation of statement (2.1). Since $n^2 + n + 41$ has 41 as a term, we will generate a common factor if we let $n = 41$. This example, which will demonstrate that statement (2.2) is a false statement, is called a *counterexample* to that statement.

Proof. We show that the statement

$$\text{for all positive integers } n,\ n^2+n+41 \text{ is prime}$$

is false by providing a counterexample. Note that when $n = 41$,

$$n^2+n+41 = (41)^2+41+41 = (41)(41+1+1) = (41)(43).$$

Hence $(41)^2+41+41$ is not prime, by Definition 2.1.7. □ ◊

2.1.2. Proof by cases. Sometimes one is unable to find a single argument that works in general to prove a statement. Consider the statement

$$\boxed{\text{for all integers } a,\ a(a+1) \text{ is even.}}$$

Is this statement true or false? If it is true, then we wish to prove it. If it is false, then we'll disprove it.

Again, we'll try to get a sense for the statement by computing $a(a+1)$ for various values of a.

$$\begin{aligned} a=4: &\qquad a(a+1) = 4(5) = 2(2)(5) \text{ is even,}\\ a=17: &\qquad a(a+1) = 17(18) = 2(17)(9) \text{ is even,}\\ a=-5: &\qquad a(a+1) = -5(-4) = 2(-5)(-2) \text{ is even.}\end{aligned}$$

Here, not only do our examples seem to indicate that the statement is true, but they also appear to show us why: if a is even, then $a(a+1)$ is automatically even, and if a is odd, then $a+1$ is even, again ensuring that $a(a+1)$ is even. (Note that we are continuing to assume that every integer is either even or odd, but never both.) So, we will try to prove the statement, and our scratchwork (we have no need for a Given-Goal diagram here) indicates that we should use a technique called *proof by cases*, since the argument depends on whether or not a is even.

Proposition 2.1.9. *For all integers a, $a(a+1)$ is even.*

Proof. Let $a \in \mathbb{Z}$. We show that $a(a+1)$ is even by considering two cases.

Case I: a is even.

Then $2 \mid a$, by Definition 1.2.1. Since $a \mid a(a+1)$ by Definition 2.1.1, we have that $2 \mid a(a+1)$ since the divisibility relation is transitive (Proposition 2.1.3). Hence $a(a+1)$ is even.

Case II: a is not even.

Since a is not even, we know that a is odd. Then $a+1$ is even by Exercise 1.2.2b. Then, using an argument similar to that of Case I, we have that $2 \mid (a+1)$ and $(a+1) \mid a(a+1)$, and hence $2 \mid a(a+1)$ by Proposition 2.1.3. Thus $a(a+1)$ is even.

Hence, since we have considered all possible cases for the integer a, we have proved that for all integers a, $a(a+1)$ is even. □

The most important thing about a proof by cases is that the cases need to consider all possibilities for the object in question (here, for the arbitrary integer a, that a is either even or not). Note that a proof by cases may have more than two cases.

Also note that our proof of Proposition 2.1.9 made reference to two previously proved results, namely Exercise 1.2.2b and Proposition 2.1.3. This is standard practice in mathematics, since it allows one to focus on the issues at hand and thus shortens the proof in question. While we could have included a proof of the two relevant statements (that when a is odd, then $a+1$ is even, and that the divisibility relation is transitive) within our proof of Proposition 2.1.9 (and we would have needed to, had we not previously proved those two statements), the proof we gave is more efficient. Thus, we can amend our earlier statement in Section 1.2 about how to prove that something is a widget (here, the property "even integer" is the widget in question).

> If we wish to prove that something is a *widget*, then we must either use the definition of widget or we must use a previously proved result that implies that something is a widget.
>
> If we use the definition, then its logical form determines the structure of a proof that something is a widget.
>
> If we use a previously proved result that implies that something is a widget, then we must verify the hypotheses of that result.

2.1.3. Working backwards. In this example, given a positive real number x, we wish to determine whether one of the expressions $\frac{x}{x+1}$ and $\frac{x+1}{x+2}$ is always larger than the other. We can try some simple examples, with x a positive integer, to see what we think.

x	$\frac{x}{x+1}$	$\frac{x+1}{x+2}$
1	$\frac{1}{2}$	$\frac{2}{3}$
2	$\frac{2}{3}$	$\frac{3}{4}$
3	$\frac{3}{4}$	$\frac{4}{5}$
4	$\frac{4}{5}$	$\frac{5}{6}$

It appears that, at least for positive *integers* x,

$$\boxed{\frac{x}{x+1} < \frac{x+1}{x+2}.}$$

We'll try to prove this for all positive *real numbers* x.

Given	**Goal**
$x \in \mathbb{R}$ $x > 0$	$\frac{x}{x+1} < \frac{x+1}{x+2}$

It's not clear how we should proceed, since we have hardly any information in our "Given" column. In situations such as this, it sometimes helps to "work backwards" from the Goal; i.e., we'll try to simplify or rewrite the goal to help us see how to proceed.

Warning: Working backwards is a *strategy for finding* a proof, not a proof in itself. The reason for this is that when we work backwards, we will *assume* what we are trying to prove, *which is never allowed.* When we work backwards, we are hoping to eventually find a statement that we "already" know is true, such as an instance of an axiom or a statement we have already proved. We do not know ahead of time what that statement will be. More importantly, we are also hoping that our reasoning is *reversible*, which we must also check. We've already seen that the converse of a true implication need not also be true, so that working backwards *may not work.* Thus, when working backwards, we must verify that we can construct a valid proof.

To begin, let's assume that we know we have $x > 0$ *and* $\frac{x}{x+1} < \frac{x+1}{x+2}$. We'd like to multiply both sides of the inequality by $(x+1)(x+2)$, in order to clear the fractions, but we also know that we must be careful about the sign of this expression. Since $x > 0$, we know that $x+1 > 1 > 0$ and, similarly, $x+2 > 0$. Hence $(x+1)(x+2) > 0$ by an order property. Multiplying both sides of

$$\frac{x}{x+1} < \frac{x+1}{x+2}$$

by $(x+1)(x+2)$ gives, after cancelling,

$$x(x+2) < (x+1)^2,$$

since $(x+1)(x+2) > 0$. If we rewrite this inequality as

$$x^2 + 2x < (x^2 + 2x) + 1,$$

here we see a statement that we know, from our basic properties, to be true; i.e., $x^2 + 2x < x^2 + 2x + 1$ is always true, regardless of x. *If all our steps are reversible,* then we've found a proof. We give the final proof of the desired statement below; take special note of how the proof differs from the scratchwork, where we worked backwards.

Proposition 2.1.10. *For all positive real numbers x, $\frac{x}{x+1} < \frac{x+1}{x+2}$.*

Proof. Let $x \in \mathbb{R}$ be arbitrary with $x > 0$. We must show that

$$\frac{x}{x+1} < \frac{x+1}{x+2}.$$

First note that by an order property,

$$x^2 + 2x < (x^2 + 2x) + 1 \tag{2.4}$$

and hence, by factoring,

$$x(x+2) < (x+1)^2. \tag{2.5}$$

Since $x > 0$, we know that $(x+1)(x+2) > 0$, and hence we may divide both sides of statement (2.5) by $(x+1)(x+2)$ to obtain

$$\frac{x(x+2)}{(x+1)(x+2)} < \frac{(x+1)^2}{(x+1)(x+2)},$$

by another order property. Thus we may remove a factor of 1 on each side of the inequality to obtain

$$\frac{x}{x+1} < \frac{x+1}{x+2},$$

as desired. □

The statement we just proved in Proposition 2.1.10 isn't particularly profound; the point here is to demonstrate the method of working backwards. Two things are important to note, however. First, as we mentioned above, it is important when using this method not to confuse the scratchwork with the proof. Phrased even more strongly, *the scratchwork is* **not** *the proof!*

Second, note that the proof makes no mention of how we got our inspiration to start with statement (2.4). This makes the proof seem fairly mysterious. It is important to remember this anytime you are reading a proof that makes you wonder "How did the author know to do this?" In such instances, you might try to do the scratchwork yourself.

2.1.4. Proving biconditional statements.

So far we have not proved any biconditional, or "iff", statements. Let's try to prove the familiar statement that for all real numbers a, b with $b \geq 0$,

$$\boxed{|a| \leq b \text{ iff } -b \leq a \leq b.}$$

First, we must remember that there are two statements to be proved, the forward ($\Rightarrow$) direction

$$\text{if } |a| \leq b, \text{ then } -b \leq a \leq b,$$

and the backward ($\Leftarrow$) direction

$$\text{if } -b \leq a \leq b, \text{ then } |a| \leq b.$$

Also recall that the compound inequality $-b \leq a \leq b$ is an abbreviation for the statement "$-b \leq a$ and $a \leq b$". Finally, we need to recall the definition of the absolute value function.

Definition 2.1.11. Given $x \in \mathbb{R}$, the *absolute value of* x, denoted by $|x|$, is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

As we begin our scratchwork, we note that it is likely (although not automatic) that this proof will involve cases, since the definition of $|x|$ is a definition by cases involving x. We provide the Given-Goal diagrams but leave any additional scratchwork up to you.

For ($\Rightarrow$):

Given	**Goal**
a, $b \in \mathbb{R}$	$-b \leq a \leq b$
$b \geq 0$	(i.e.,
$\lvert a\rvert \leq b$	$-b \leq a$ and $a \leq b$)

For ($\Leftarrow$):

Given	**Goal**
$a,\, b \in \mathbb{R}$	
$b \geq 0$	
$-b \leq a \leq b$	
(i.e.,	
$-b \leq a$ and $a \leq b$)	$\lvert a\rvert \leq b$

Proposition 2.1.12. *For all real numbers $a,\, b \in \mathbb{R}$ with $b \geq 0$, $|a| \leq b$ if and only if $-b \leq a \leq b$.*

Proof. Let $a,\, b \in \mathbb{R}$ be arbitrary with $b \geq 0$. We prove that $|a| \leq b$ iff $-b \leq a \leq b$.

($\Rightarrow$) Assume that $|a| \leq b$. We show that $-b \leq a \leq b$. We consider two cases.

Case I: $a \geq 0$.

Then $|a| = a$ by Definition 2.1.11, and so $a \leq b$ by our assumption that $|a| \leq b$. Also, since $b \geq 0$, we have $-b \leq 0 \leq a$; i.e., $-b \leq a$ by the transitive property. Since $-b \leq a$ and $a \leq b$, we have $-b \leq a \leq b$, as desired.

Case II: $\neg(a \geq 0)$.

Then $a < 0$ and hence $|a| = -a$, by Definition 2.1.11. So $-a \leq b$ by our assumption that $|a| \leq b$. But then $a \geq -b$ by an order property, and so

$$-b \leq a < 0 \leq b;$$

i.e., $-b \leq a \leq b$, as desired.

($\Leftarrow$) Assume that $-b \leq a \leq b$. We show that $|a| \leq b$ by again considering two cases.

Case I: $a \geq 0$.

Then $|a| = a$ by Definition 2.1.11, and hence $|a| \leq b$ by our assumption that $a \leq b$.

Case II: $a < 0$.

Then $|a| = -a$ by Definition 2.1.11. Since $-b \leq a$ by our assumption, we have $b \geq -a$ by an order property. Thus $|a| \leq b$, as desired. □

2.1.5. Uniqueness proofs. Sometimes we find that not only do we wish to prove that an object exists, but also we wish to prove that the object that exists is *unique*, i.e., that there is exactly one such object. We can express this in our mathematical language in one of two ways. Suppose that $P(x)$ is a predicate with free variable x, and think of $P(x)$ as expressing "x satisfies property P". The statement

there exists a unique x such that $P(x)$

denotes

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

or

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

Remember here that $\wedge$ denotes the logical connective "and".

Statement (2.6) states that

(1) an object x satisfying property P exists and, furthermore,

(2) if y is an arbitrary object satisfying property P, then y must be the same object as this x.

Statement (2.7) states that

(1) an object x satisfying property P exists, and

(2) if y and z are arbitrary objects both satisfying property P, then y and z are equal.

Both statement (2.6) and statement (2.7) can be represented by the notation

$$(\exists! x)P(x).$$

Regardless of whether one thinks of $(\exists! x)P(x)$ in terms of statement (2.6) or statement (2.7), note that a proof of $(\exists! x)P(x)$ always has two parts: a proof of *existence* and a proof of *uniqueness*.

Example 2.1.13. In Basic Properties of Real Numbers 2.1.4 and Basic Properties of Integers 1.2.3, we have assumed (both implicitly and explicitly) a bit more than we needed in order to not belabor these assumptions too much. For example, Basic Properties of Real Numbers 2.1.4 states that for all $a \in \mathbb{R}$, $a + 0 = a = 0 + a$ and, implicitly, that 0 is the only element of $\mathbb{R}$ with this property. Let's keep this assumption and the assumption that addition of real numbers is commutative.

Basic Properties of Real Numbers 2.1.4 also states that for all $a \in \mathbb{R}$, there exists a unique real number $-a$ such that $a + (-a) = 0 = (-a) + a$.

It's possible to replace this axiom by the weaker assumption

$$(\forall a \in \mathbb{R})(\exists b \in \mathbb{R})[a + b = 0 = b + a] \tag{2.8}$$

and then *prove*

$$(\forall a \in \mathbb{R})(\exists! b \in \mathbb{R})[a + b = 0 = b + a]. \tag{2.9}$$

Then, given $a \in \mathbb{R}$, we can *define* $-a$ to be the unique $b \in \mathbb{R}$ with $a + b = 0$.

Scratchwork: Let $a \in \mathbb{R}$. By statement (2.7), a Given-Goal diagram for the uniqueness part of statement (2.9) is

Given	**Goal**
all axioms in Basic Properties of Real Numbers, with (Additive inverses) replaced by statement (2.8)	
$a \in \mathbb{R}$ arbitrary	
$b_1, b_2 \in \mathbb{R}$ arbitrary	
$a + b_1 = 0 = b_1 + a$	
$a + b_2 = 0 = b_2 + a$	$b_1 = b_2$

Proof of statement (2.9). Let $a \in \mathbb{R}$.

(Existence): Assumption (2.8) establishes the fact that there exists $b \in \mathbb{R}$ such that $a + b = 0$.

(Uniqueness): Let $b_1, b_2 \in \mathbb{R}$ and assume that $a + b_1 = 0 = b_1 + a$ and $a + b_2 = 0 = b_2 + a$. We prove that $b_1 = b_2$. We have

$$a + b_1 = 0,$$

so

$$b_2 + (a + b_1) = b_2 + 0.$$

By associativity of addition and the fact that 0 is the additive identity, we have

$$(b_2 + a) + b_1 = b_2.$$

Since $b_2 + a = 0$, this gives

$$0 + b_1 = b_2,$$

or

$$b_1 = b_2,$$

as desired.

Hence, there is a unique $b \in \mathbb{R}$ such that $a + b = 0$. □

The point of this example was to illustrate a standard uniqueness proof, and it is necessary to prove statements of this sort in an abstract algebra course. We'll see more uniqueness statements later in this textbook. ♢

Exercises 2.1

1. Let a, b, and c be integers. Prove that for all integers m and n, if $a \mid b$ and $a \mid c$, then $a \mid (bm + cn)$.
2. Prove that for all real numbers a and b, if $0 < a < b$, then $0 < a^2 < b^2$.
3. Prove that for all integers m, if m is odd, then there exists $k \in \mathbb{Z}$ such that $m^2 = 8k + 1$.
4. Using definitions, prove by cases that for every integer n, $n^2 + n + 5$ is odd.
5. Determine whether each statement is true or false. If true, then prove it. If false, then provide a counterexample.
   (a) For all positive integers n, n is divisible by 3 is necessary for n to be divisible by 6.
   (b) For all positive integers n, n is divisible by 3 is sufficient for n to be divisible by 6.
   (c) For all real numbers x, $x^2 - 2x - 3 = 0$ only if $x = 3$.
   (d) For all real numbers x, $x^2 - 2x - 3 = 0$ if $x = 3$.
   (e) For all integers a, b, c, if $a \mid bc$, then $a \mid b$ or $a \mid c$.
   (f) For all integers a, b, c, if $a \mid (b + c)$, then $a \mid b$ or $a \mid c$.
   (g) For all even integers m and n, $4 \mid mn$.
   (h) For all integers n, if n^2 is a multiple of 4, then n is a multiple of 4.
   (i) There exist integers m and n such that $15m + 12n = -6$.

6. Prove that for all real numbers x, $x^2 \geq 0$.

7. Prove that for all real numbers x and y, if $xy = 0$, then $x = 0$ or $y = 0$.

8. Prove that for all real numbers x and y, $x^2 + xy + y^2 \geq 0$. (**HINT:** Complete the square in x.)

9. Prove that for all real numbers x and y, if $x^2 = y^2$, then $x = y$ or $x = -y$; i.e., $x = \pm y$. (Your proof should not mention anything called a "square root".)

10. Prove that for all $a, b \in \mathbb{R}^+$, $\sqrt{ab} = \sqrt{a}\sqrt{b}$ and $\sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}}$.

11. Prove that for all real numbers x and y, if $x < y$, then $x < \frac{x+y}{2} < y$.

12. Let $a \in \mathbb{R}$. Prove that:
    (a) $|a| = |-a|$. (**HINT**: Consider three cases.)
    (b) $-|a| \leq a \leq |a|$.
    (c) $|a| = 0$ if and only if $a = 0$.

13. Let $a, b \in \mathbb{R}$. Prove that:
    (a) If $|a| = |b|$, then $a = b$ or $a = -b$.
    (b) $|ab| = |a||b|$.
    (c) $|a - b| = |b - a|$.

14. Let $a, b \in \mathbb{R}$. Prove that:
    (a) (Triangle Inequality) $|a + b| \leq |a| + |b|$. (**HINT**: Use Exercise 2.1.12b and Proposition 2.1.12, or a proof by cases.)
    (b) $|a + b| = |a| + |b|$ if and only if a and b have the same sign.
    (c) $||a| - |b|| \leq |a - b|$.
    (d) $|a - b| \leq |a| + |b|$.

15. Prove that for all positive real numbers x, the sum of x and its reciprocal is greater than or equal to 2.

16. Prove that for all negative real numbers x, the sum of x and its reciprocal is less than or equal to -2.

17. Prove that for all $x, y \in \mathbb{R}^+$, $\sqrt{xy} \leq \frac{x+y}{2}$, with equality occurring if and only if $x = y$. (This result is a special case of the *Arithmetic-Geometric Mean Inequality.*)

18. Prove that for all nonnegative real numbers x, $\dfrac{2|x-3|}{x+1} \leq 7$. (**HINT:** Consider two cases, based on the definition of absolute value.)

19. Prove that for all real numbers a, b, c with $a \neq 0$, the equation $ax^2 + bx + c = 0$ has a solution $u \in \mathbb{R}$ if and only if $b^2 - 4ac \geq 0$ and $u = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. (Recall that a real number u is a *solution* of the equation $ax^2 + bx + c = 0$ if and only if $au^2 + bu + c = 0$.)

20. A 2×2 (real) *matrix* is a table of the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where $a, b, c, d \in \mathbb{R}$. Given two 2×2 matrices (the plural of "matrix" is "matrices") $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and

$B = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$, the matrix product AB is defined by

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} ax+bz & ay+bw \\ cx+dz & cy+dw \end{bmatrix}.$$

(a) Compute the matrix product $\begin{bmatrix} 2 & -3 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 5 & -1 \\ 2 & -2 \end{bmatrix}$.

(b) Prove that for all 2×2 matrices A, B, and C, $A(BC) = (AB)C$ (i.e., matrix multiplication is associative).

(c) Prove that there exist 2×2 matrices A and B with $AB \neq BA$ (i.e., in general, matrix multiplication is not commutative).

21. Let $a, b, c \in \mathbb{R}$ with $a \neq 0$. Prove that the equation $ax + b = c$ has a unique solution in $\mathbb{R}$; i.e., prove that $(\exists! x \in \mathbb{R})[ax + b = c]$.

2.2. Indirect proofs: Proofs by contradiction and contrapositive

2.2.1. Proof by contradiction. In Subsection 1.2.2 and Section 2.1, we have been concentrating on *direct* proofs of statements such as $P \Rightarrow Q$, $(\forall x)P(x)$, and $(\exists x)P(x)$. Sometimes, however, it can be difficult to determine how to proceed using these methods. For example, suppose that we want to prove that

$$\boxed{\text{there do not exist integers } m,\ n \in \mathbb{Z} \text{ such that } 12m + 9n = 100.}$$

Our initial Given-Goal diagram is the following:

Given	**Goal**
Nothing	$\neg(\exists m, n \in \mathbb{Z})[12m + 9n = 100]$

How can we proceed? We're given nothing, and we are trying to show that something does not exist. Even rewriting

$$\neg(\exists m, n \in \mathbb{Z})[12m + 9n = 100]$$

as

$$(\forall m, n \in \mathbb{Z})[12m + 9n \neq 100]$$

does not seem to help:

Given	**Goal**
m, $n \in \mathbb{Z}$ arbitrary	$12m + 9n \neq 100$

It seems difficult to show that arbitrary integers m and n, about which we know nothing, have the desired property.

So we're in a situation where we're trying to prove that a statement P is true, but we don't know how to begin. One thing to try in such a situation is to assume that $\neg P$ is true, and then try to deduce a statement R such as

$$0 = 1 \tag{2.10}$$

or

$$Q \wedge \neg Q, \quad \text{where } Q \text{ is some (possibly different) statement.} \tag{2.11}$$

A statement such as (2.10) or (2.11) is called a *contradiction.* More precisely, a statement R involving statement variables, such as $Q \wedge \neg Q$, is a *contradiction* if every assignment of truth values to the statement variables in R makes R false; see Exercise 1.1.4.

If we can find such a contradiction R, then we will have a valid proof that the implication $(\neg P) \Rightarrow R$ is true. The truth table in Table 1.5 then tells us that $(\neg P)$ must be false, since $(\neg P) \Rightarrow R$ is true and R is false. Hence, we will have proved that P is true, which is what we wanted all along. This proof technique is called *proof by contradiction*, and it is an example of an *indirect* method of proof.

Table 2.1. Proof by contradiction.

To prove a statement P is true by contradiction.
• We begin with "Assume $\neg P$ is true."
• We deduce a contradiction.
• We then conclude that P is true.

Let's use this technique to prove that there are no integers m and n such that $12m + 9n = 100$; i.e., $\neg(\exists m, n \in \mathbb{Z})[12m + 9n = 100]$. Since this will be a proof by contradiction, we will assume

$$\neg\neg(\exists m, n \in \mathbb{Z})[12m + 9n = 100].$$

By Exercise 1.1.2a, this is logically equivalent to

$$(\exists m, n \in \mathbb{Z})[12m + 9n = 100].$$

We search for a contradiction. Our Given-Goal diagram now takes the following form:

Given	**Goal**
m, $n \in \mathbb{Z}$	
$12m + 9n = 100$	Contradiction

The tricky part to a proof by contradiction is that, while we know that we are looking for a contradiction, we usually do not know ahead of time what form that contradiction will take. So we just follow where our Given column takes us logically and stay on the lookout for a contradiction.

We know $12m + 9n = 100$; a logical thing to do is to factor the left-hand side to obtain

$$3(4m + 3n) = 100.$$

Recalling Definition 2.1.1, we see that this says that 100 is divisible by 3. But $3 \nmid 100$ by the order properties of the integers, since $3 \cdot 33 = 99$ and $3 \cdot 34 = 102$. We've found a contradiction; namely, we've proved that $3 \mid 100$ and $3 \nmid 100$, i.e., a statement of the form $Q \wedge \neg Q$.

Proposition 2.2.1. *There do not exist integers m and n such that $12m+9n = 100$.*

Proof. Assume for the sake of a contradiction that we have integers m and n such that $12m + 9n = 100$. Then $3(4m + 3n) = 100$; i.e., $3 \mid 100$, by definition. But 100 is not divisible by 3 because every integer k either satisfies $k \leq 33$ or $k \geq 34$, and hence every multiple $3k$ of 3 satisfies $3k \leq 99$ or $3k \geq 102$, by the order properties of the integers. Thus, we have a contradiction. Hence there do not exist integers m and n such that $12m + 9n = 100$. □

Note that our proof by cases, above, i.e., that every multiple of 3 is either less than or equal to 99 or greater than or equal to 102, is exactly the type of proof needed in Example 2.1.2 to show that $5 \nmid 12$.

The next statement we'll consider deals with rational and irrational numbers. Recall that a real number z is *rational* if there exist integers a, b with $b \neq 0$ such that $z = \frac{a}{b}$. A real number is *irrational* if it is not rational; thus, to prove (directly) that a real number z is irrational, one must show that integers with a particular property do *not* exist. We'll now prove that

$$\boxed{\text{the sum of a rational number and an irrational number is irrational.}}$$

We begin with our initial Given-Goal diagram.

Given	**Goal**
x, $y \in \mathbb{R}$ arbitrary	
$x \in \mathbb{Q}$	
y is irrational	$x + y$ is irrational

Since trying (directly) to show that $x + y$ is irrational amounts to showing that something *does not exist*, it again seems reasonable to try an indirect proof by contradiction. We rewrite our Given-Goal diagram for this approach.

Given	**Goal**
x, $y \in \mathbb{R}$ arbitrary	
x is rational	
y is irrational	
$x + y$ is rational	Contradiction

Since we know that the sum and product of two rational numbers is rational (see page 28), we should try to exploit these facts using x and $x + y$ in some way. Since we are looking for a contradiction, it makes sense to work with y, which is easy to express in terms of x and $x + y$:

$$y = (x + y) - x.$$

We have the idea of the proof; now we need to express the details carefully.

Proposition 2.2.2. *For all real numbers x and y, if x is rational and y is irrational, then $x + y$ is irrational.*

Proof. Let x, $y \in \mathbb{R}$ be arbitrary, and assume that x is rational and y is irrational. For the sake of a contradiction, also assume that $x+y$ is rational. Note that $(-1)x$ is rational, since the product of rational numbers is another rational number. Also, $(x + y) + (-1)x$ is rational, since the sum of rational numbers is another rational number. Since $y = (x+y)+(-1)x$, this implies that y is rational, a contradiction, since we are given that y is irrational. Thus $x + y$ is irrational, as desired. □

2.2.2. Proving the contrapositive. To motivate our second method of indirect proof, we will show that

$$\boxed{\text{for all } n \in \mathbb{Z}, \text{ if } n^2 \text{ is odd, then } n \text{ is odd.}}$$

It's easy to see why we should try to use an indirect proof.

Given	**Goal**
$n \in \mathbb{Z}$ arbitrary	
n^2 is odd	n is odd

If we assume that $n \in \mathbb{Z}$ with n^2 odd, then by Definition 1.2.1, we know that we may fix $k \in \mathbb{Z}$ such that $n^2 = 2k + 1$. However, this does not give us any information about n. We could use a proof by contradiction here (try it yourself), but we will instead take advantage of what we know about implications.

Recall from Proposition 1.1.15 that an implication $P \Rightarrow Q$ is logically equivalent to its contrapositive $\neg Q \Rightarrow \neg P$.

$$\boxed{\text{Thus, to prove } P \Rightarrow Q\text{, we may choose instead to prove } \neg Q \Rightarrow \neg P.}$$

The contrapositive of

$$n^2 \text{ is odd } \Rightarrow n \text{ is odd}$$

is

$$n \text{ is even } \Rightarrow n^2 \text{ is even.}$$

We certainly know how to prove this (see Exercise 1.2.1a), and proving this statement serves as a proof of our original statement, since the two statements are logically equivalent. For practice, however, rather than simply quoting Exercise 1.2.1a, we'll provide all details in this proof.

Given	**Goal**
$n \in \mathbb{Z}$ arbitrary	
n is even	n^2 is even

Proposition 2.2.3. *For all $n \in \mathbb{Z}$, if n^2 is odd, then n is odd.*

Proof. Let $n \in \mathbb{Z}$ be arbitrary. We prove the contrapositive. Assume that n is even; we show that n^2 is also even.

Since n is even, by definition we may fix $k \in \mathbb{Z}$ such that $n = 2k$. Then

$$n^2 = (2k)^2 = 4k^2 = 2(2k^2),$$

and hence n^2 is even, also by definition.

Hence, if n^2 is odd, then n is odd. □

A similar result is also true.

Proposition 2.2.4. *For all $n \in \mathbb{Z}$, if n^2 is even, then n is even.*

Proof. This is Exercise 2.2.1. □

2.2.3. Proving or statements.

In this section we consider one useful method of proving a statement which is a disjunction. Consider the following proposition.

Proposition 2.2.5. *For all real numbers a and b with $b \geq 0$, if $a^2 \geq b$, then $a \geq \sqrt{b}$ or $a \leq -\sqrt{b}$.*

As usual, we begin with a Given-Goal diagram.

Given	**Goal**
a, $b \in \mathbb{R}$ arbitrary	
$b \geq 0$	
$a^2 \geq b$	$a \geq \sqrt{b}$ or $a \leq -\sqrt{b}$

We could try a proof by contradiction here. Instead, we will once again take advantage of logic. Recall from Proposition 1.1.11 and Exercise 1.1.2a that a statement of the form $P \vee Q$ is logically equivalent to the statement $\neg P \Rightarrow Q$. Also note that $P \vee Q$ is logically equivalent to $Q \vee P$, which is logically equivalent to the statement $\neg Q \Rightarrow P$.

Thus, to prove $P \vee Q$, we may assume $\neg P$ and prove Q, or we may assume $\neg Q$ and prove P.

Since we are not proving the statement in its original form, we view this method of proving $P \vee Q$ as an indirect proof. We can also think of this as a proof by cases, where one of the cases is automatic; for example, if P is true, then $P \vee Q$ is automatically true; thus, we need only consider the case when P is false, i.e., when $\neg P$ is true. The analogous approach with the roles of P and Q interchanged is also permitted.

Note that in this case either of the statements $a \geq \sqrt{b}$ or $a \leq -\sqrt{b}$ can be P. In general, however, your choice of P and Q will depend on the statements themselves and whatever is easier to work with. The new Given-Goal diagram follows.

Given	**Goal**
$a,\ b \in \mathbb{R}$ arbitrary	
$b \geq 0$	
$a^2 \geq b$	
$a \not\geq \sqrt{b}$ (i.e., $a < \sqrt{b}$)	$a \leq -\sqrt{b}$

To deal with these inequalities, it often helps to express them by moving all terms to one side. Thus we have $a^2 - b \geq 0$, so we can manipulate by factoring (remember that $b \in \mathbb{R}$ with $b \geq 0$):

$$(a - \sqrt{b})(a + \sqrt{b}) \geq 0.$$

We are also given that $a - \sqrt{b} < 0$, so it's now easy to see how this proof should go.

Proof of Proposition 2.2.5. Let $a,\ b \in \mathbb{R}$ be arbitrary with $b \geq 0$, and assume that $a^2 \geq b$. We prove that $a \geq \sqrt{b}$ or $a \leq -\sqrt{b}$ (note that $\sqrt{b}$ makes sense since $b \geq 0$). If $a \geq \sqrt{b}$, then we're done, so assume that $a \not\geq \sqrt{b}$; i.e., $a < \sqrt{b}$.

Since $a^2 \geq b$, we know that $a^2 - b \geq 0$, and hence $(a - \sqrt{b})(a + \sqrt{b}) \geq 0$. Since $a < \sqrt{b}$, we know $a - \sqrt{b} < 0$. Since $(a - \sqrt{b})(a + \sqrt{b}) \geq 0$ and $a - \sqrt{b} < 0$, it follows that $a + \sqrt{b} \leq 0$. To see this, note that if $a + \sqrt{b} > 0$, then $(a - \sqrt{b})(a + \sqrt{b}) < 0$ by an order property, a contradiction. Hence $a \leq -\sqrt{b}$, as desired. □

Exercises 2.2

1. Prove Proposition 2.2.4, that for all integers n, if n^2 is even, then n is even, using *definitions*.
2. Prove that there are no integers n such that n is both even and odd.
3. Prove there are no integers m and n such that $8m + 26n = 1$.
4. Prove there are no integers m and n such that $m^2 = 4n + 2$. (**HINT:** Use Proposition 2.2.4.)
5. Prove there are no integers m and n such that $m^2 = 4n + 3$.
6. In this problem, you may use the fact (which we will prove in Chapter 6) that an integer n is not divisible by 3 if and only if there exists an integer k such that $n = 3k + 1$ or $n = 3k + 2$.
   (a) Prove that for all integers n, if $3 \mid n^2$, then $3 \mid n$.
   (b) Prove that for all integers i and j, if $3 \mid (i^2 + j^2)$, then $3 \mid i$ and $3 \mid j$.
7. Prove that for all $a, b \in \mathbb{Z}^+$, if $a \mid b$, then $a \leq b$.
8. Let a and b be positive integers. Prove that if $a \mid b$ and $b \mid a$, then $a = b$.
9. Determine whether each statement is true or false. If true, then prove it. If false, then provide a counterexample.
   (a) The sum of two irrational numbers is irrational.
   (b) The product of two irrational numbers is irrational.

(c) The product of a nonzero rational number and an irrational number is irrational.

10. Using *definitions* and the method of Subsection 2.2.3, prove that for all integers m and n, if mn is even, then m is even or n is even.
11. Prove that for all real numbers x, $x^2 + 3x - 4 > 0$ if and only if $x < -4$ or $x > 1$.
12. Prove that for all real numbers x and y, if $x^3 = y^3$, then $x = y$. (Your proof should not mention anything called a "cube root". Use Exercise 2.1.8.)
13. Let $x \in \mathbb{R}$ and assume that for all $\varepsilon > 0$, $|x| < \varepsilon$. Prove that $x = 0$. (**HINT:** Use a proof by contradiction, and find a specific $\varepsilon > 0$ contradicting the hypothesis.)

2.3. Two important theorems

So far we have considered examples of direct and indirect proofs, taking advantage of the logical equivalence of statements, when necessary. The examples we have considered have been of fairly straightforward mathematical statements, since our primary goal has been to illustrate different techniques of proof.

In this section, we prove two very important facts, that $\sqrt{2}$ is irrational and that there are infinitely many prime numbers. These proofs are more sophisticated than the others we have considered up to now, but the basic setup of these indirect proofs follows the same patterns as the techniques we have discussed.

For the first result, note that the statement

$$\boxed{\sqrt{2}\text{ is irrational}}$$

is a negative one:

$$\neg(\exists p, q \in \mathbb{Z}^+)\left[\frac{p}{q} = \sqrt{2}\right]. \tag{2.12}$$

(Note that since $\sqrt{2} > 0$, we may take the universe for p and q to be $\mathbb{Z}^+$, rather than $\mathbb{Z}$.) Because the statement we wish to prove is negative, it makes sense to try a proof by contradiction. Assuming the negation of statement (2.12) yields the following Given-Goal diagram:

Given	**Goal**
$p,\ q \in \mathbb{Z}^+$ $\frac{p}{q} = \sqrt{2}$	Contradiction

Theorem 2.3.1. *$\sqrt{2}$ is irrational.*

Proof. Assume for the sake of a contradiction that $\sqrt{2}$ is rational. By definition, we may fix $p,\ q \in \mathbb{Z}^+$ such that $\frac{p}{q} = \sqrt{2}$, and by removing common factors if necessary, we may also assume that p and q do not have any common positive integer factors other than 1. (Here we are assuming that every fraction can be put in "least terms".) Then $\frac{p^2}{q^2} = 2$, and hence

$$p^2 = 2q^2. \tag{2.13}$$

Hence the integer p^2 is even, by Definition 1.2.1. It follows by Proposition 2.2.4 that p is also even. Thus, again by Definition 1.2.1, we may fix $k \in \mathbb{Z}$ such that $p = 2k$. Substituting into (2.13), we have $(2k)^2 = 2q^2$, or $4k^2 = 2q^2$. By cancellation in $\mathbb{Z}$ (see Basic Properties of Integers 1.2.3), we have $2k^2 = q^2$.

Hence q^2 is even, by definition, and again by Proposition 2.2.4, q is also even. But then p and q are both even, and so p and q have 2 as a common factor. This is a contradiction, since we assumed that p and q had no common positive integer factors other than 1.

Hence $\sqrt{2}$ is irrational. □

The argument that, for example, $\sqrt{3}$ is irrational is similar. See Exercise 2.3.1.

Our next example of an indirect proof, that there are infinitely many prime numbers, is due to Euclid. In order to prove this result, we will need three pieces of information: the definition of "prime number", the definition of "infinitely many", and the existence part of the *Fundamental Theorem of Arithmetic*, a result with which you may already be familiar about factoring positive integers as products of primes. We review these statements below.

Definition 2.3.2. The positive integer p is *prime* if

$$p > 1 \text{ and } (\forall m, n \in \mathbb{Z}^+)[p = mn \Rightarrow (m = 1 \vee n = 1)].$$

Theorem 2.3.3 (Fundamental Theorem of Arithmetic)**.** *Every positive integer greater than* 1 *can be written as a product of primes. Furthermore, this product of primes is unique, except for the order in which the factors appear.*

Note that we consider a prime number to itself be a "product of primes" with a single factor. As another example, note that the unique prime factorization of 1119250 is

$$1119250 = 2 \cdot 5 \cdot 5 \cdot 5 \cdot 11 \cdot 11 \cdot 37.$$

The only prime factors of 1119250 are 2, 5, 11, and 37, and the only factorization of 1119250 in terms of these primes, written in nondecreasing order, is the one given above.

Recall from our discussion in Subsection 2.1.5 that a proof of the Fundamental Theorem of Arithmetic will have two parts: a proof of existence and a proof of uniqueness. We will prove the existence part of Theorem 2.3.3 (i.e., the fact that every positive integer greater than 1 *can* be written as a product of primes) in Section 3.2. We prove uniqueness (which is not needed here) in Section 6.3.

We should note why we may postpone this proof. We are in essence making a promise that the proof of Theorem 2.3.3 does not depend on the statement we are currently proving, in this case, Theorem 2.3.4. If it did, then we would have what is known as a "circular argument", which is not logically valid and hence is not a proof at all. In fact, the proof of the existence part of Theorem 2.3.3 requires a proof technique called *strong induction*, and we could stop and present this proof technique now. By choosing to wait until Section 3.2 to prove the existence part of the Fundamental Theorem of Arithmetic, we are focusing on the result we are currently interested in, namely, Euclid's theorem that there are infinitely many primes.

Our proof will be an indirect proof by contradiction and will emphasize the relationship between primality and factorization.

Theorem 2.3.4 (Euclid)**.** *There are infinitely many prime numbers.*

Proof. Suppose for the sake of a contradiction that there do not exist infinitely many prime numbers, i.e., that there exist finitely many prime numbers. This means that we can form a complete list of the prime numbers:

$$p_1, p_2, \ldots, p_k,$$

where $k \in \mathbb{Z}^+$ is the number of primes and $1 < p_1 < p_2 < \cdots < p_k$. (Note that, for example, $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc.).

Consider the positive integer

$$n = p_1 \cdot p_2 \cdot \cdots \cdot p_k + 1.$$

Note that $n > p_k > 1$ by the order axioms. Thus n is not prime, by our assumption that $p_1 < p_2 < \cdots < p_k$ yields a complete list of all prime numbers. By the Fundamental Theorem of Arithmetic (Theorem 2.3.3), some prime must divide n, since n is a product of primes. Hence we can fix an integer j, $1 \le j \le k$, such that p_j divides n, and it then follows by definition that we may fix $m \in \mathbb{N}$ such that $n = p_j m$.

Thus we have $p_j m = p_1 \cdot p_2 \cdot \cdots \cdot p_k + 1$. Rewriting this gives

$$p_j m - p_1 \cdot p_2 \cdot \cdots \cdot p_k = 1, \quad \text{or}$$
$$p_j m - p_j \ell = 1,$$

where ℓ is the product of all the primes in the list $p_1, p_2, \ldots, p_k$ except for p_j (note that if $k = 1$, then $\ell = 1$). Thus

$$p_j(m - \ell) = 1,$$

and hence $p_j \mid 1$, which is a contradiction, since $p_j > 1$.

Hence there must exist infinitely many prime numbers. □

Exercises 2.3

1. Prove that $\sqrt{3}$ is irrational. (**HINT:** You will need the result of Exercise 2.2.6a.)
2. Prove that for all $x \in \mathbb{R}$, at least one of $\sqrt{3} - x$ and $\sqrt{3} + x$ is irrational.
3. Prove that $\log_2 3$ is irrational. (**NOTE:** Recall that $\log_2 3$ is the real number y such that $2^y = 3$. You may assume, without proof, the familiar "rules of exponents".)
4. In Theorem 2.3.1, what is the purpose of assuming that p and q do not have any common positive integer factors other than 1? How does the proof change if we do not make this assumption?

2.4. Proofs of statements involving mixed quantifiers

The statement of Euclid's theorem, that there exist infinitely many prime numbers, is another example of "hidden quantifiers". As we have stated it, the theorem may sound existential to you, i.e., like it has the form $(\exists n)P(n)$. However, the logical form of the statement is actually more complicated; it is a two-quantifier statement, where the type of quantifier *alternates* (here, from $\forall$ to $\exists$):

$$(\forall n \in \mathbb{Z}^+)(\exists m \in \mathbb{Z}^+)[m \geq n \text{ and } m \text{ is prime}].$$

We also say that this statement has *mixed* quantifiers.

Another example of a statement hiding a second quantifier is "There is a smallest positive integer." If we unravel this statement, it says "there is a positive integer x such that x is less than or equal to any positive integer"; i.e.,

$$(\exists x \in \mathbb{Z}^+)(\forall y \in \mathbb{Z}^+)[x \leq y]. \tag{2.14}$$

Equation (2.14) is a true statement. To prove it, we begin by noting that the outermost quantifier is $\exists$; this tells us that we must give a particular example of a positive integer x which satisfies

$$(\forall y \in \mathbb{Z}^+)[x \leq y].$$

This is easy: note that $1 \in \mathbb{Z}^+$ satisfies $(\forall y \in \mathbb{Z}^+)[1 \leq y]$.

Note also that, as in our previous discussion of quantifiers, the universe matters. If we change (2.14) to

$$(\exists x \in \mathbb{Z})(\forall y \in \mathbb{Z})[x \leq y],$$

then this statement is false. To prove this, we must prove that

$$\neg(\exists x \in \mathbb{Z})(\forall y \in \mathbb{Z})[x \leq y]$$

is true, i.e., that

$$(\forall x \in \mathbb{Z})(\exists y \in \mathbb{Z})[x > y]$$

is true.

Again, the way we begin the proof is dictated by the outermost quantifier.

Given	**Goal**
$x \in \mathbb{Z}$ arbitrary	$(\exists y \in \mathbb{Z})[x > y]$

So, given $x \in \mathbb{Z}$ arbitrary, we must demonstrate a particular integer y which is strictly less than x. Note that $y = x - 1$ is an integer and $x > y$ since $x > x - 1$. We have proved that $(\forall x \in \mathbb{Z})(\exists y \in \mathbb{Z})[x > y]$ is true, and hence the statement $(\exists x \in \mathbb{Z})(\forall y \in \mathbb{Z})[x \leq y]$ is false.

Regardless of the number of quantifiers in a quantified statement

$$(Q_1 x_1)(Q_2 x_2) \cdots (Q_k x_k)[\cdots],$$

where $Q_1, Q_2, \ldots, Q_k$ are quantifiers, we always begin with the outermost quantifier and "work our way in". When we see a string of identical quantifiers, such as in the statements

$$(\forall x)(\forall y)P(x,y) \text{ or } (\exists x)(\exists y)(\exists z)Q(x,y,z),$$

then the order of these quantifiers doesn't matter. For example,

$$(\forall x)(\forall y)P(x,y) \text{ and } (\forall y)(\forall x)P(x,y)$$

have the same logical meaning, and both can be thought of as $(\forall x, y)P(x,y)$. Similarly (considering one possible reordering),

$$(\exists x)(\exists y)(\exists z)Q(x,y,z) \text{ and } (\exists z)(\exists x)(\exists y)Q(x,y,z)$$

have the same logical meaning, and both can be thought of as $(\exists x, y, z)Q(x,y,z)$.

However, we must be careful with mixed quantifiers: in general

$$\boxed{\textit{order of mixed quantifiers matters!}}$$

We can easily see that this must be the case by comparing the following two "common sense" statements. Let S be the set of all students at your college or university. The statement

$$(\forall x \in S)(\exists y \in S)[x \text{ and } y \text{ are friends}]$$

says that every student at your college or university is friends with some student at your college or university. This is not the same statement as

$$(\exists y \in S)(\forall x \in S)[x \text{ and } y \text{ are friends}],$$

which says that some student at your college or university is friends with every student at your college or university!

In addition, moving quantifiers within a statement can change the truth value of that statement. For example, consider the statement

$$(\forall x)[(\forall y)(y > 0) \Rightarrow x > 0], \tag{2.15}$$

where the quantifiers range over $\mathbb{R}$. Statement (2.15) is true, since given an arbitrary $x \in \mathbb{R}$, the statement $(\forall y)(y > 0)$ is false in $\mathbb{R}$, so $(\forall y)(y > 0) \Rightarrow x > 0$ is true in $\mathbb{R}$.

However, the statement

$$(\forall x)(\forall y)[y > 0 \Rightarrow x > 0]$$

is false in $\mathbb{R}$; i.e.,

$$(\exists x)(\exists y)[y > 0 \wedge x \leq 0]$$

is true in $\mathbb{R}$, since $x = -1$ and $y = 3$ satisfy $y > 0 \wedge x \leq 0$.

While there do exist rules for moving quantifiers around (see [**10**]), we will not be employing them.

$$\boxed{\textit{Don't move quantifiers around within mathematical statements!}}$$

Example 2.4.1. Prove that

$$(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})[x + y = 0] \tag{2.16}$$

is a true statement and that

$$(\exists y \in \mathbb{R})(\forall x \in \mathbb{R})[x + y = 0] \tag{2.17}$$

is a false statement.

Proof. First note that statement (2.16) is one of our Basic Properties of Real Numbers 2.1.4 and hence a true statement. More formally, when $x \in \mathbb{R}$ is arbitrary, $y = -x = (-1)x$ satisfies $x + y = x + (-x) = 0$. Hence $(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})[x + y = 0]$ is a true statement.

Next, we prove (2.17) is false; i.e., we prove $(\forall y \in \mathbb{R})(\exists x \in \mathbb{R})[x + y \neq 0]$ is true. Let $y \in \mathbb{R}$ be arbitrary. We must prove that $(\exists x \in \mathbb{R})[x + y \neq 0]$. Note that if we let $x = -y + 1$, which is a real number, then

$$x + y = (-y + 1) + y = 1 \neq 0,$$

as desired. Hence $(\exists y \in \mathbb{R})(\forall x \in \mathbb{R})[x + y = 0]$ is false. □ ◊

Gven a formula $(Q_1x_1)(Q_2x_2)\cdots(Q_kx_k)P(x_1, x_2, \ldots, x_k)$, where the predicate $P(x_1, x_2, \ldots, x_k)$ is quantifier-free, an *alternation* of quantifiers means a change from $\forall$ to $\exists$ or vice versa within $(Q_1x_1)(Q_2x_2)\cdots(Q_kx_k)$. In a way that can be made mathematically precise (see [**16**]), the more alternations of quantifiers such a statement has, the more complicated the statement is. Mathematical concepts whose definitions involve mixed quantifiers are very common in calculus and real analysis. For example, the statement that there is a rational number between any two distinct real numbers is expressed as follows:

$$(\forall x, y \in \mathbb{R})(\exists z \in \mathbb{Q})[x < y \Rightarrow x < z < y].$$

An important concept involving mixed quantifiers that is studied in calculus is the notion of *limit*. Working informally for the moment, beginning students in calculus often think of the statement $\lim_{x\to a} f(x) = L$ as meaning that the values of the function f get "closer and closer" to the number L when the value x gets "closer and closer" to the number a. Unfortunately, this way of thinking about a limit is not only imprecise, it is incorrect. The precise definition of limit, which has a daunting two alternations of quantifiers, is given below.

Definition 2.4.2. Let f be a function of a single variable defined for all real numbers in an open interval containing the real number a, except possibly at a itself, and let L be a real number. Then the *limit of f at a is L*, in notation, $\lim\limits_{x\to a} f(x) = L$ if

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x)[0 < |x - a| < \delta \Rightarrow |f(x) - L| < \varepsilon].$$

(Here, all quantifiers range over the real numbers.)

To make sense of this definition, recall that the inequality $|x - a| < \delta$ says that the distance between the numbers x and a is less than δ, i.e., that x is within a distance of δ from a. Peeling off the quantifiers one at a time, Definition 2.4.2 says that for any (think: tiny) distance $\varepsilon > 0$, we can find a (think: small enough) distance $\delta > 0$ such that if $x \neq a$ is any real number within a distance of δ from a, then the value of the function $f(x)$ is within a distance of ε of L. See Figure 2.1.

Example 2.4.3. Let

$$f(x) = \begin{cases} 4x - 3 & \text{if } x \neq 2, \\ 1 & \text{if } x = 2. \end{cases}$$

Prove that $\lim\limits_{x\to 2} f(x) = 5$.

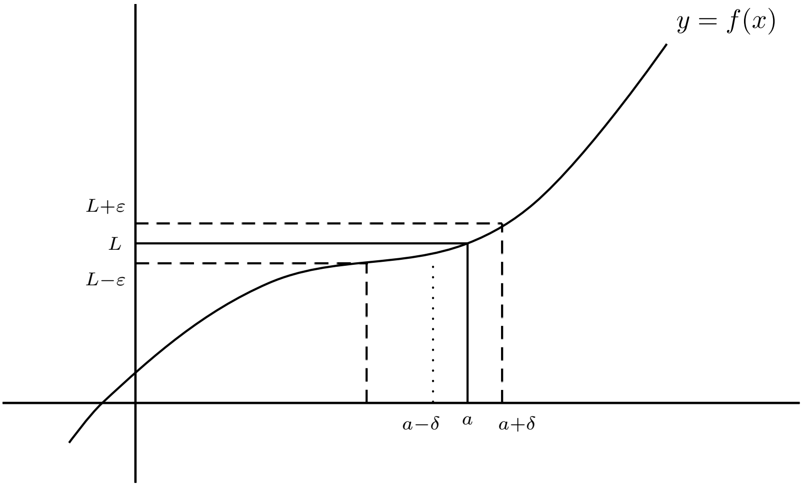


Figure 2.1. $\lim_{x\to a} f(x) = L$.

Definition 2.4.2 tells us what the Given-Goal diagram should look like.

Given	**Goal**
$\varepsilon > 0$ arbitrary	$(\exists\delta > 0)(\forall x)[0 < \lvert x-2\rvert < \delta \Rightarrow \lvert f(x)-5\rvert < \varepsilon]$

So, given $\varepsilon > 0$, we must find a real number $\delta > 0$ such that

$$(\forall x)[0 < |x-2| < \delta \Rightarrow |f(x)-5| < \varepsilon].$$

We'll work backwards. Suppose $x \in \mathbb{R}$ is arbitrary. We want $|f(x)-5| < \varepsilon$. We may assume that $x \neq 2$, since we will ultimately be assuming that $0 < |x-2|$. Hence, we may assume that $f(x) = 4x-3$. This means that we want

$$|(4x-3)-5| < \varepsilon; \quad \text{i.e.,}$$

$$|4x-8| < \varepsilon. \tag{2.18}$$

We want the δ we are looking for to satisfy $|x-2| < \delta \Rightarrow |f(x)-5| < \varepsilon$. This tells us to look for an expression involving $|x-2|$. Equation (2.18) is equivalent to $|4(x-2)| < \varepsilon$, or $4|x-2| < \varepsilon$, by Exercise 2.1.13b. Thus, we need $|x-2| < \frac{\varepsilon}{4}$, and we should try $\delta = \frac{\varepsilon}{4}$.

Proof. We prove that $\lim_{x\to 2} f(x) = 5$. Let $\varepsilon > 0$ be arbitrary. We show

$$(\exists\delta > 0)(\forall x)[0 < |x-2| < \delta \Rightarrow |f(x)-5| < \varepsilon].$$

Consider $\delta = \frac{\varepsilon}{4}$, which is positive, since ε is. We must show

$$(\forall x)[0 < |x-2| < \delta \Rightarrow |f(x)-5| < \varepsilon].$$

Let $x \in \mathbb{R}$ be arbitrary, and assume that $0 < |x-2| < \delta$; i.e., $0 < |x-2| < \frac{\varepsilon}{4}$. We must show $|f(x)-5| < \varepsilon$. To see this, note that since $x \neq 2$,

$$\begin{aligned} |f(x)-5| &= |(4x-3)-5| \\ &= |4x-8| \\ &= |4(x-2)| \\ &= 4|x-2|, \end{aligned}$$

by properties of absolute value. Thus, since $|x-2| < \frac{\varepsilon}{4}$,

$$\begin{aligned} |f(x)-5| &= 4|x-2| \\ &< 4\left(\frac{\varepsilon}{4}\right), \end{aligned}$$

by an order property. Hence $|f(x)-5| < \varepsilon$, as desired. It follows that $\lim_{x\to 2} f(x) = 5$, by Definition 2.4.2. □ ◊

The formal study of limits is part of a course in real analysis.

Exercises 2.4

1. Determine whether each statement is true or false, and prove or disprove, as appropriate.
   (a) $(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})[xy = 1]$.
   (b) $(\exists x \in \mathbb{R})(\forall y \in \mathbb{R})[xy = 1]$.
   (c) $(\exists x \in \mathbb{R})(\forall y \in \mathbb{R})[xy > 0]$.
   (d) $(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})[xy > 0]$.
   (e) $(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})(\forall z \in \mathbb{R})[xy = xz]$.
   (f) $(\exists y \in \mathbb{R})(\forall x \in \mathbb{R})(\exists z \in \mathbb{R})[xy = xz]$.
   (g) $(\forall x \in \mathbb{Q})(\exists y \in \mathbb{Z})[xy \in \mathbb{Z}]$.
   (h) $(\exists x \in \mathbb{Z}^+)(\forall y \in \mathbb{Z}^+)[y \leq x]$.
   (i) $(\forall y \in \mathbb{Z}^+)(\exists x \in \mathbb{Z}^+)[y \leq x]$.
   (j) $(\forall x, y \in \mathbb{Z})[x < y \implies (\exists z \in \mathbb{Z})[x < z < y]]$.
   (k) $(\forall x, y \in \mathbb{Q})[x < y \implies (\exists z \in \mathbb{Q})[x < z < y]]$.
2. Prove that $\lim_{x\to -\frac{1}{2}} (4x-1) = -3$.
3. Let
$$f(x) = \begin{cases} 5-2x & \text{if } x \neq 4, \\ 23 & \text{if } x = 4. \end{cases}$$
   Prove that $\lim_{x\to 4} f(x) = -3$.
4. Prove that there exists a real number M such that for all $x \in \mathbb{R}$ with $1 < x < 3$, $\left|\frac{5x^2-2x-4}{5(x^2+1)}\right| \leq M$.

5. Assume that for every real number x, there is an integer N such that $N > x$.[3] Prove that for every positive real number ε, there exists a positive integer N such that for all $n \geq N$, $\frac{1}{n} < \varepsilon$.

[3] This statement follows from a fact about real numbers called the *Archimedean Property*, which we prove in Chapter 9.

Chapter 3

Induction

3.1. Principle of Mathematical Induction

Thus far we have been proving results about the integers using only the assumptions in the Basic Properties of Integers 1.2.3. Up to now, any direct proof of a statement of the form $(\forall n \in \mathbb{Z}^+)P(n)$ has begun with the assumption "Let $n \in \mathbb{Z}^+$ be arbitrary." We consider now a new proof technique, called *proof by induction*, for proving a statement about *all* positive integers. The statement of the *Principle of Mathematical Induction* is logically complicated, so we have typeset in **bold** the words showing the logical structure of this statement.

Theorem 3.1.1 (Principle of Mathematical Induction (PMI)).

Let $P(n)$ be a statement about the positive integer n, so that n is a free variable in $P(n)$.

Suppose the following:

(PMI 1) *The statement $P(1)$ is true.*

(PMI 2) *For all positive integers m,*

$$\textit{if } P(m) \textit{ is true}, \textit{ then } P(m+1) \textit{ is true.}$$

Then, *for all positive integers n, $P(n)$ is true.*

Notice that, given the statement $P(n)$, a proof of $(\forall n \in \mathbb{Z}^+)P(n)$ consists of *first* proving the two statements (PMI 1) and (PMI 2) and then *concluding* (by induction) the statement $(\forall n \in \mathbb{Z}^+)P(n)$. Why should this conclusion be logically valid?

Well, (PMI 1) establishes that $P(1)$ is true, and (PMI 2) establishes that

$$P(1) \Rightarrow P(2).$$

Thus by *modus ponens*, $P(2)$ is true. But (PMI 2) also establishes that

$$P(2) \Rightarrow P(3).$$

Thus, from $P(2)$ and $P(2) \Rightarrow P(3)$, we conclude $P(3)$ is true, etc.

PMI says that the "etc." is valid; i.e., it says that we can reach any positive integer n by starting at 1 and successively adding 1. This is quite a reasonable statement about the positive integers. However, one can show that it is impossible to prove PMI from the Basic Properties of Integers 1.2.3. Thus, we will accept PMI as an additional axiom for $\mathbb{Z}^+$ (in fact, we are adding infinitely many axioms, one for each statement $P(n)$). It *is* possible to prove PMI from other statements about the positive integers, such as the *Well-Ordering Principle*, which we discuss in Section 6.1, but then we would need to assume *those* statements as axioms.

Several of the mathematical statements we will prove by induction are familiar summation formulas, which we will want to express using "sigma notation".

Notation 3.1.2. Suppose that $i \leq n$ are integers and $a_i, a_{i+1}, \ldots, a_n$ are real numbers. Then

$$\sum_{k=i}^{n} a_k = a_i + a_{i+1} + \cdots + a_n.$$

For our first example of a proof using PMI, we will prove a summation formula you may have seen before, perhaps when discussing Riemann sums in calculus. We will prove by induction that for all $n \in \mathbb{Z}^+$,

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2};$$

$$\text{i.e., } 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

For example, when $n = 4$, note that

$$\sum_{k=1}^{4} k = 1 + 2 + 3 + 4 = 10 \quad \text{and} \quad \frac{n(n+1)}{2} = \frac{4(5)}{2} = 10,$$

so the statement is true when $n = 4$. To prove the statement is true for *all* positive integers n, however, we will use PMI. In this first example, we will show our scratchwork in great detail.

Scratchwork. Given $n \in \mathbb{Z}^+$, we'll let $P(n)$ be the statement

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Note that $P(n)$ is a *statement* about n; it asserts that n has a particular property. The Principle of Mathematical Induction says that if we can prove that the *Base Case* (PMI 1) is true and that the *Inductive Step* (PMI 2) is true, then we may *conclude* that $(\forall n \in \mathbb{Z}^+)P(n)$ is true, which is what we want.

Base Case: Show $P(1)$ is true.

$P(1)$ is the statement

$$\sum_{k=1}^{1} k = \frac{1(1+1)}{2},$$

and our task is to prove that $P(1)$ is true. Thus, we must compute both quantities

$$\sum_{k=1}^{1} k \quad \text{and} \quad \frac{1(1+1)}{2},$$

individually, and then we must verify that these quantities are equal. Note that

$$\sum_{k=1}^{1} k = 1$$

by definition. Also note that

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

So we see that

$$\sum_{k=1}^{1} k = \frac{1(1+1)}{2};$$

i.e., $P(1)$ is true.

Inductive Step: Prove statement (PMI 2). Let's begin with a Given-Goal diagram

Given	**Goal**
$m \in \mathbb{Z}^+$ arbitrary	
$P(m)$ is true	$P(m+1)$ is true

which we rewrite as

Given	**Goal**
$m \in \mathbb{Z}^+$ arbitrary	
$\displaystyle\sum_{k=1}^{m} k = \frac{m(m+1)}{2}$	$\displaystyle\sum_{k=1}^{m+1} k = \frac{(m+1)((m+1)+1)}{2}$

We will *use the Given* (called the *Inductive Hypothesis*), which says that

$$\sum_{k=1}^{m} k = \frac{m(m+1)}{2}, \tag{3.1}$$

to help us *prove our Goal*, which says that

$$\sum_{k=1}^{m+1} k = \frac{(m+1)((m+1)+1)}{2}.$$

We may rewrite our Goal as

$$\sum_{k=1}^{m+1} k = \frac{(m+1)(m+2)}{2}.$$

To prove an equality, we must pick one side and attempt to manipulate legally until we reach the other side.

We'll start with $\sum_{k=1}^{m+1} k$. The key point in the proof is that, since $m \in \mathbb{Z}^+$,

$$\begin{aligned}\sum_{k=1}^{m+1} k &= 1 + 2 + \cdots + (m+1)\\ &= (1 + 2 + \cdots + m) + (m+1);\end{aligned}$$

i.e.,

$$\sum_{k=1}^{m+1} k = \left(\sum_{k=1}^{m} k\right) + (m+1), \tag{3.2}$$

by the associative property of addition. Equation (3.2) shows us clearly where the Inductive Hypothesis (3.1) will be useful; namely, we may replace $\sum_{k=1}^{m} k$ by $\frac{m(m+1)}{2}$. From here, the computation should be routine.

In our formal writeup of this proof by induction, pay close attention to the *format* of the proof, which we'll use in all proofs by induction.

Proposition 3.1.3. *For all $n \in \mathbb{Z}^+$,*

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Proof. Given $n \in \mathbb{Z}^+$, let $P(n)$ denote the statement

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

We prove $(\forall n \in \mathbb{Z}^+)P(n)$ by induction on n.

Base Case: We must show that $P(1)$ is true.

Note that $P(1)$ is the statement

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

Since $\sum_{k=1}^{1} k = 1$ by definition and since $\frac{1(1+1)}{2} = \frac{2}{2} = 1$, we see that $\sum_{k=1}^{1} k = \frac{1(1+1)}{2}$. Hence, $P(1)$ is true.

Inductive Step: Let $m \in \mathbb{Z}^+$ and assume that $P(m)$ is true; i.e., assume the Inductive Hypothesis

$$\sum_{k=1}^{m} k = \frac{m(m+1)}{2}.$$

We must show that $P(m+1)$ is true; i.e., we must show that

$$\sum_{k=1}^{m+1} k = \frac{(m+1)((m+1)+1)}{2} = \frac{(m+1)(m+2)}{2}.$$

Note that

$$\begin{aligned}\sum_{k=1}^{m+1} k &= (1+2+\cdots+m)+(m+1)\\ &= \left(\sum_{k=1}^{m} k\right)+(m+1)\\ &= \frac{m(m+1)}{2}+(m+1) \quad \text{by the Inductive Hypothesis.}\end{aligned}$$

Thus, by adding fractions,

$$\begin{aligned}\sum_{k=1}^{m+1} k &= \frac{m(m+1)+2(m+1)}{2}\\ &= \frac{(m+1)(m+2)}{2} \quad \text{(factoring } (m+1) \text{ in the numerator),}\end{aligned}$$

as desired. Hence $P(m+1)$ is true.

Thus, by the Principle of Mathematical Induction, we have proved that for all $n \in \mathbb{Z}^+$,

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}. \qquad \square$$

Note that the idea of proof by induction is the same idea used when making an "inductive" or "recursive" definition. For example, the familiar function 2^n, where $n \geq 0$ is an integer, is defined by recursion as follows:

$$2^0 = 1$$

and

$$\text{for all } m \geq 0, \quad 2^{m+1} = 2 \cdot 2^m.$$

Note that, just as with a proof by induction, a recursive definition has a *base case* (sometimes more than one) and an *inductive step*, which in a recursive definition is often called the *recursion step.* Not surprisingly, one often uses proof by induction to prove statements about concepts that are defined recursively.

The factorial function $n!$ is also defined by recursion on $n \geq 0$:

$$0! = 1$$

and

$$\text{for all } m \geq 0, \ (m+1)! = (m+1)m!.$$

For example (showing all steps in the recursion),

$$\begin{aligned}
5! &= 5 \cdot 4! \\
&= 5 \cdot 4 \cdot 3! \\
&= 5 \cdot 4 \cdot 3 \cdot 2! \\
&= 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1! \\
&= 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 0! \\
&= 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1, \qquad \text{since } 0! = 1; \text{ i.e.,} \\
5! &= 120.
\end{aligned}$$

Notice also that in a recursive definition or proof by induction, the base case need not correspond to $n = 1$, as it did in Proposition 3.1.3, nor to $n = 0$, as it did in the recursive definitions above. Thus, we can state a more general version of the Principle of Mathematical Induction.

Theorem 3.1.4 (Principle of Mathematical Induction (modified))**.** *Let $P(n)$ be a statement about the integer n, where n is free in $P(n)$.*

Suppose that there is an integer n_0 such that:

(PMI 1) *The statement $P(n_0)$ is true.*

(PMI 2) *For all integers $m \geq n_0$,*

if $P(m)$ is true, then $P(m+1)$ is true.

Then, *for all integers $n \geq n_0$, $P(n)$ is true.*

For our next example, we will prove that

$$\boxed{\text{for all integers } n \geq 10,\ n^3 \leq 2^n.}$$

The fact that 2^n, $n \geq 0$, is defined by recursion on n tells us that it is reasonable to try induction on $n \geq 10$. We do the scratchwork for the inductive step; you should verify for yourself why we chose $n = 10$ as the base case.

Scratchwork: The Given-Goal diagram for the Inductive Step is below; we have identified the Inductive Hypothesis by (IH).

Given	**Goal**
$m \in \mathbb{Z}^+$	
$m \geq 10$	
$m^3 \leq 2^m$ (IH)	$(m+1)^3 \leq 2^{m+1}$

We begin by examining $(m+1)^3$ and 2^{m+1}.

$$\begin{aligned}
2^{m+1} &= 2 \cdot 2^m \geq 2m^3 \qquad \text{by the Inductive Hypothesis, and} \\
(m+1)^3 &= m^3 + 3m^2 + 3m + 1.
\end{aligned}$$

Working backwards, we want to argue that

$$2m^3 \geq m^3 + 3m^2 + 3m + 1, \tag{3.3}$$

so it will suffice to argue that

$$m^3 \geq 3m^2 + 3m + 1. \tag{3.4}$$

Throughout we'll make use of the order properties in Basic Properties of Integers 1.2.3. Note that since $1 \leq m$, we have $1 \leq m \leq m^2$, and hence

$$3m^2 + 3m + 1 \leq 3m^2 + 3m^2 + m^2 = 7m^2.$$

Also, $7 \leq m$ and $m^2 \geq 0$, so $7m^2 \leq m^3$. Thus we have

$$3m^2 + 3m + 1 \leq 3m^2 + 3m^2 + m^2 = 7m^2 \leq m^3,$$

establishing (3.4).

We are ready to write down the formal proof.

Proposition 3.1.5. *For all integers $n \geq 10$, $n^3 \leq 2^n$.*

Proof. Let $n \in \mathbb{Z}$ with $n \geq 10$, and let $P(n)$ denote the statement

$$n^3 \leq 2^n.$$

We prove by induction on n that for all integers $n \geq 10$, $P(n)$ is true.

Base Case: We must show that $10^3 \leq 2^{10}$.

Note that $10^3 = 1000$ and $2^{10} = 1024$, so the Base Case holds; i.e., $10^3 \leq 2^{10}$.

Inductive Step: Let $m \in \mathbb{Z}$ with $m \geq 10$ and assume that $m^3 \leq 2^m$. We must prove that $(m+1)^3 \leq 2^{m+1}$.

To see this, first note that since $1 \leq m$, $1 \leq m \leq m^2$. In addition, $7m^2 \leq m^3$, since $7 \leq m$ and $m^2 \geq 0$. Thus,

$$\begin{aligned}
(m+1)^3 &= m^3 + 3m^2 + 3m + 1 \\
&\leq m^3 + 3m^2 + 3m^2 + m^2 \\
&= m^3 + 7m^2 \\
&\leq m^3 + m^3 \\
&= 2m^3 \\
&\leq 2 \cdot 2^m, \qquad \text{by the Inductive Hypothesis.}
\end{aligned}$$

Hence, $(m+1)^3 \leq 2^{m+1}$, as desired.

Thus, by PMI, we have that for all integers $n \geq 10$, $n^3 \leq 2^n$. □

Exercises 3.1

All exercises should be proved using induction.

1. Prove that for all positive integers n,

$$\sum_{k=1}^{n} k^2 = 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

2. Prove that for all integers $n \geq 0$,
$$\sum_{k=0}^{n} 2^k = 1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1} - 1.$$
3. Prove that for all positive integers n,
$$\sum_{k=1}^{n} (2k-1) = 1 + 3 + 5 + \cdots + (2n-1) = n^2.$$
4. Prove that for all positive integers n,
$$\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{1}{1(2)} + \frac{1}{2(3)} + \frac{1}{3(4)} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}.$$
5. Prove that for all positive integers n,
$$\sum_{k=1}^{n} (2k-1)^2 = (1)^2 + (3)^2 + (5)^2 + \cdots + (2n-1)^2 = \frac{4n^3 - n}{3}.$$
6. Conjecture a formula for $\sum_{k=1}^{n} (-1)^k k^2$, and then prove the formula is correct using induction.
7. Prove that for all positive integers n, $n < 10^n$.
8. Prove that for all positive integers $n \geq 5$, $n^2 < 2^n$.
9. Prove that for all positive integers $n \geq 7$, $(\frac{4}{3})^n > n$.
10. Prove that for all positive integers $n \geq 4$, $2^n < n!$.
11. Prove that for all positive integers n, $n^3 + 8n + 9$ is divisible by 3.
12. Prove that for all positive integers n, $4^n - 1$ is divisible by 3.
13. Prove that for all positive integers n, $3^{2n} - 1$ is divisible by 8.
14. Let $a_1 = 2$, and let $a_{n+1} = \frac{1}{2}(a_n + 3)$ for all $n \geq 1$. (We are essentially defining a function $a(n) = a_n$ for $n \in \mathbb{Z}^+$ by recursion.)
    (a) Prove that for all positive integers n, $a_n < a_{n+1}$.
    (b) Without using part (c) below, prove that for all positive integers n, $a_n < 3$.
    (c) Prove that for all positive integers n, $a_n = 3 - \frac{1}{2^{n-1}}$.
15. Let $r \in \mathbb{R}$ with $r \neq 1$. Prove that
$$\sum_{k=0}^{n-1} r^k = 1 + r + \cdots + r^{n-1} = \frac{1 - r^n}{1 - r}.$$
    (For this notation to make sense, we "bend the rules" and declare $0^0 = 1$.)
16. Prove Bernoulli's Inequalty: Let $x > -1$. Then for all $n \in \mathbb{Z}^+$, $(1+x)^n \geq 1+nx$.
17. The *Fibonacci numbers* f_n, $n \in \mathbb{Z}^+$, are defined recursively by the formulas $f_1 = 1$, $f_2 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for all $n \geq 3$.
    (a) Find the first ten Fibonacci numbers $f_1, \ldots, f_{10}$.
    (b) Compute $f_1 + f_2$, $f_1 + f_2 + f_3$, $f_1 + f_2 + f_3 + f_4$, $f_1 + f_2 + f_3 + f_4 + f_5$.
    (c) Conjecture a formula for the sum $f_1 + \cdots + f_n$ of the first n Fibonacci numbers, where $n \geq 1$, and then prove the formula is correct using induction.
    (d) Use induction to prove that for all integers $k \geq 1$, $5 \mid f_{5k}$.

18. Given integers $k, n \geq 0$ with $k \leq n$, we define the *binomial coefficient* $\binom{n}{k}$ by
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$
    (a) Prove that for all integers n, k with $1 \leq k \leq n$, $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$. (This part does not require induction.)
    (b) Use part (a) to prove by induction on n that for all integers $n \geq 0$, for all integers k with $0 \leq k \leq n$, $\binom{n}{k}$ is an integer.
19. Let $x, y \in \mathbb{R}$. Prove the *binomial theorem*: for all integers $n \geq 1$,
$$\begin{aligned}(x+y)^n &= \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\ &= \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 \\ &\quad + \cdots + \binom{n}{n-2} x^2 y^{n-2} + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n.\end{aligned}$$
20. The following is a well-known example of a "proof" that all horses are the same color.

    Let $P(n)$ denote the statement "For every set of n horses, all the horses in the set are the same color." We "prove" $(\forall n \in \mathbb{Z}^+) P(n)$ by induction on n. Note that $P(1)$ is true, since for any set containing a single horse, all the horses in that set have the same color, namely the color of that single horse. Next, let $m \geq 1$ and assume that $P(m)$ is true, i.e., that for any set of m horses, all the horses in the set are the same color. We prove that $P(m+1)$ is true. Let S be a set of $m+1$ horses named $x_1, x_2, \ldots, x_{m+1}$. Now the horses $x_1, x_2, \ldots, x_m$ are a set of m horses, so by the Inductive Hypothesis, all the horses $x_1, x_2, \ldots, x_m$ in this set have the same color. Furthermore, the horses $x_1, x_3, \ldots, x_{m+1}$ (i.e., all the horses except for horse x_2) are a set of m horses, so by the Inductive Hypothesis, all the horses $x_1, x_3, \ldots, x_{m+1}$ in this set have the same color. It follows that all of the $m+1$ horses $x_1, x_2, \ldots, x_{m+1}$ are the same color, since they are all the same color as horse x_1. Thus, $P(m+1)$ is true. It follows by PMI that $P(n)$ is true for all $n \in \mathbb{Z}^+$; i.e., all horses have the same color.

    What's wrong with this proof?

3.2. Strong induction

Recall that, informally, a sequence is a list of real numbers. More precisely, a sequence is given by a function defined on the positive integers. For example, the function $a(n) = \frac{1}{n}$, for $n \geq 1$, defines the sequence
$$1, \frac{1}{2}, \frac{1}{3}, \ldots.$$
In general we write a_n, rather than $a(n)$, and a sequence
$$a_1, a_2, a_3, \ldots$$
is denoted by $\{a_n\}_{n=1}^{\infty}$.

Consider the following recursively defined sequence $\{a_n\}_{n=1}^{\infty}$.

$$\begin{aligned} a_1 &= 1, \\ a_2 &= 5, \\ \text{for all } n \geq 2,\ a_{n+1} &= a_n + 2a_{n-1}. \end{aligned}$$

Let's try to guess a formula for a_n. Computing the first few integers in the sequence shows that

$$\begin{aligned} a_3 &= a_2 + 2a_1 = 5 + 2(1) = 7, \\ a_4 &= a_3 + 2a_2 = 7 + 2(5) = 17, \\ a_5 &= a_4 + 2a_3 = 17 + 2(7) = 31, \\ a_6 &= a_5 + 2a_4 = 31 + 2(17) = 65. \end{aligned}$$

Note that these numbers are almost, but not quite, powers of 2:

$$\begin{aligned} a_1 &= 1 = 2^1 - 1, \\ a_2 &= 5 = 2^2 + 1, \\ a_3 &= 7 = 2^3 - 1, \\ a_4 &= 17 = 2^4 + 1, \\ a_5 &= 31 = 2^5 - 1, \\ a_6 &= 65 = 2^6 + 1. \end{aligned}$$

Thus, a reasonable guess is that

$$a_n = 2^n + (-1)^n \quad \text{for } n \geq 1.$$

This formula is called a *closed formula* for the recursively defined sequence; i.e., it describes a_n as a function of n.

How can we prove that our closed formula is correct? It seems clear that induction is needed; the recursive definition has a base case (in fact, two of them) and an inductive step. However, the inductive step seems problematic. If we set up the inductive step according to PMI, then we have the following Given-Goal diagram.

Given	**Goal**
$m \in \mathbb{Z}^+$ $(m \geq 2)$ $a_{m+1} = a_m + 2a_{m-1}$ $a_m = 2^m + (-1)^m$ (IH)	$a_{m+1} = 2^{m+1} + (-1)^{m+1}$

However, the inductive step in the recursive definition defines a_{m+1} in terms of *two* predecessors (a_m and a_{m-1}), rather than the usual *one* predecessor a_m. Our usual inductive hypothesis will give us information about a_m, but no information about a_{m-1}. Thus, we appear to need a new form of induction.

Theorem 3.2.1 (Principle of Strong Mathematical Induction (PSMI))**.** *Let $P(n)$ be a statement about the positive integer n.*

Suppose the following:

(PSMI 1) *The statement $P(1)$ is true.*

(PSMI 2) *For all positive integers m,*

(3.5) *if for all integers k with $1 \leq k \leq m, P(k)$ is true, then $P(m+1)$ is true.*

Then*, for all positive integers n, $P(n)$ is true.*

Statement (3.5) is complicated, so let's see what it says for various values of m. When $m = 1$, statement (3.5) says that

if for all integers k with $1 \leq k \leq 1, P(k)$ is true, then $P(2)$ is true;

i.e.,

if $P(1)$ is true, then $P(2)$ is true.

Given the base case, it follows that

(3.6) $P(2)$ is true.

When $m = 2$, statement (3.5) says that

if for all integers k with $1 \leq k \leq 2, P(k)$ is true, then $P(3)$ is true;

i.e.,

if $P(1)$ and $P(2)$ are true, then $P(3)$ is true.

Given the base case and statement (3.6), it follows that

(3.7) $P(3)$ is true.

When $m = 3$, statement (3.5) says that

if for all integers k with $1 \leq k \leq 3, P(k)$ is true, then $P(4)$ is true;

i.e.,

if $P(1), P(2),$ and $P(3)$ are true, then $P(4)$ is true.

Given the base case and statements (3.6) and (3.7), it follows that

$P(4)$ is true.

We can see from these examples that it is reasonable to accept PSMI as an axiom. If we can prove statements (PSMI 1) and (PSMI 2), then it is reasonable to conclude that for all positive integers n, $P(n)$ is true. Furthermore, just as with PMI, the "starting integer" can be any integer n_0, rather than 1.

Example 3.2.2. Let us now go back to the sequence that motivated our discussion:

$$a_1 = 1,$$
$$a_2 = 5,$$
$$\text{for all } n \geq 2,\ a_{n+1} = a_n + 2a_{n-1}.$$

We will prove by strong induction on n that for all $n \in \mathbb{Z}^+$, $a_n = 2^n + (-1)^n$. We will not provide any scratchwork, since we already know how to set up induction proofs, but rather we'll simply indicate how the framework will change. As before, we will have a base case and an inductive step. However, because this *particular* recursive definition has two base cases, our induction proof will also have two base cases. For the inductive step, (PSMI 2) tells us to begin with an arbitrary integer $m \geq 2$ (2 rather than 1 because of the two base cases), and assume the strong induction hypothesis that is indicated in statement (3.5).

Proof. For $n \in \mathbb{Z}^+$, let $P(n)$ denote the statement

$$a_n = 2^n + (-1)^n.$$

We prove $(\forall n \in \mathbb{Z}^+)P(n)$ by strong induction on n.

Base Case: We show $P(1)$ and $P(2)$.

Since $2^1 + (-1)^1 = 2 - 1 = 1$ and $a_1 = 1$ by definition of the sequence, $P(1)$ is true.

Since $2^2 + (-1)^2 = 4 + 1 = 5$ and $a_2 = 5$ by definition of the sequence, $P(2)$ is true.

Inductive Step: Let $m \in \mathbb{Z}$ with $m \geq 2$, and assume that for all integers k with $1 \leq k \leq m$, $P(k)$ is true; i.e., we assume that for all integers k with $1 \leq k \leq m$, $a_k = 2^k + (-1)^k$. We must prove that $P(m+1)$ is true, i.e., that $a_{m+1} = 2^{m+1} + (-1)^{m+1}$.

Note that $a_{m+1} = a_m + 2a_{m-1}$ by definition, since $m \geq 2$. Thus,

$$\begin{aligned} a_{m+1} &= a_m + 2a_{m-1} \\ &= 2^m + (-1)^m + 2(2^{m-1} + (-1)^{m-1}) \end{aligned}$$

by the inductive hypothesis for $k = m-1, m$. Hence,

$$\begin{aligned} a_{m+1} &= 2^m + (-1)^m + 2(2^{m-1} + (-1)^{m-1}) \\ &= 2^m + (-1)^m + 2^m + 2(-1)^{m-1} \\ &= 2 \cdot 2^m + (-1)^{m-1}(-1+2) \\ &= 2^{m+1} + (-1)^{m-1} \\ &= 2^{m+1} + (-1)^{m-1}(-1)^2 \\ &= 2^{m+1} + (-1)^{m+1}, \end{aligned}$$

as desired.

Hence, by PSMI, we have that for all integers $n \geq 1$,

$$a_n = 2^n + (-1)^n. \qquad \square \ \lozenge$$

We now use strong induction to pay a debt from Section 2.3 and prove the existence part of the Fundamental Theorem of Arithmetic (Theorem 2.3.3). Recall first from Definition 2.1.7 that a positive integer p is *prime* if $p > 1$ and

$$(\forall a, b \in \mathbb{Z}^+)(p = ab \Rightarrow (a = 1 \text{ or } b = 1)).$$

Theorem 3.2.3 (Fundamental Theorem of Arithmetic (Existence))**.** *Every positive integer $n > 1$ can be written as a product of primes; i.e., for every positive integer $n > 1$, there exist $s \in \mathbb{Z}^+$ and primes $p_1, p_2, \ldots, p_s$ such that $n = p_1 p_2 \cdots p_s$.*

Proof. We prove the result by strong induction on n, where $n \geq 2$.

Base Case: Note that 2 is prime, and hence $2 = p_1$, where p_1 is prime.

Inductive Step: Let $m \in \mathbb{Z}$ with $m \geq 2$ and assume that for all integers k with $2 \leq k \leq m$, k is a product of primes. We must prove that $m+1$ is a product of primes.

Case I: $m+1$ is prime.

Then $m+1$ is a product of primes, as in the Base Case.

Case II: $m+1$ is not prime.

Then, by Definition 2.1.7, we may fix a, $b \in \mathbb{Z}^+$ such that

$$m+1 = ab, \text{ where neither } a \text{ nor } b \text{ is } 1.$$

Note then that $1 < a < m+1$, so $2 \leq a \leq m$, and also $2 \leq b \leq m$. Thus, by the Inductive Hypothesis applied to each of a and b, a is a product of primes

$$a = p_1 p_2 \cdots p_i$$

and b is a product of primes

$$b = q_1 q_2 \cdots q_j,$$

where $i, j \geq 1$ and $p_1, \ldots, p_i, q_1, \ldots, q_j$ are all primes. Hence

$$ab = p_1 p_2 \cdots p_i q_1 q_2 \cdots q_j,$$

and so ab is a product of primes.

Hence by PSMI, every positive integer $n > 1$ is a product of primes. □

We end this section by commenting that one might wonder whether the Principle of *Strong* Mathematical Induction is in fact a “stronger” statement than the Principle of Mathematical Induction. In fact these statements are logically equivalent. In other words, if we accept PSMI as an axiom, then we can prove PMI (this should be easy to prove), and conversely, if we accept PMI as an axiom, then we can prove PSMI. (See Exercise 3.2.5.)

Exercises 3.2

1. Let $a_1 = 2$, $a_2 = 4$, and $a_{n+1} = 7a_n - 10a_{n-1}$ for all $n \geq 2$. Conjecture a closed formula for a_n and then prove your result.
2. Let $a_1 = 3$, $a_2 = 4$, and $a_{n+1} = \frac{1}{3}(2a_n + a_{n-1})$ for all $n \geq 2$. Prove that for all positive integers n, $3 \leq a_n \leq 4$.
3. Without using the Fundamental Theorem of Arithmetic, use strong induction to prove that for all positive integers n with $n \geq 2$, n has a prime factor.
4. For $i \in \mathbb{Z}^+$, let p_i denote the ith prime number, so that
$$p_1 = 2, \qquad p_2 = 3, \qquad p_3 = 5, \ldots.$$
Prove that for all $n \in \mathbb{Z}^+$, $p_n \leq 2^{2^{n-1}}$. (**HINT:** For the induction step, given $m \in \mathbb{Z}^+$, show that $p_{m+1} \leq p_1 p_2 \cdots p_m + 1$.)
5. In this exercise, you will prove that PMI is logically equivalent to PSMI; in other words, given PMI, show that you can deduce the statement PSMI, and vice versa.

(a) Assume that PSMI is true. To prove that PMI is true, let $P(n)$ be a statement about the positive integer n. *Assume that*

(i) $P(1)$ is true and

(ii) for all $m \in \mathbb{Z}^+$, if $P(m)$ is true, then $P(m+1)$ is true.

Your goal is to prove that for all $n \in \mathbb{Z}^+$, $P(n)$ is true. You should do this using PSMI. This means that you should *prove*

(iii) $P(1)$ is true and

(iv) for all $m \in \mathbb{Z}^+$, if for all integers k with $1 \leq k \leq m$, $P(k)$ is true, then $P(m+1)$ is true.

You can then conclude from PSMI that for all $n \in \mathbb{Z}^+$, $P(n)$ is true.

(b) Assume that PMI is true. To prove that PSMI is true, let $P(n)$ be a statement about the positive integer n. *Assume that*

(i) $P(1)$ is true and

(ii) for all $m \in \mathbb{Z}^+$, if for all integers k with $1 \leq k \leq m$, $P(k)$ is true, then $P(m+1)$ is true.

Your goal is to prove that for all $n \in \mathbb{Z}^+$, $P(n)$ is true. You should do this using PMI applied to a slightly different statement. Let $Q(n)$ be the statement $(\forall k \leq n)P(k)$. Use PMI to prove that for all $n \in \mathbb{Z}^+$, $Q(n)$ is true. This means that you should *prove*

(iii) $Q(1)$ is true and

(iv) for all $m \in \mathbb{Z}^+$, if $Q(m)$ is true, then $Q(m+1)$ is true.

You can then conclude from PMI that for all $n \in \mathbb{Z}^+$, $Q(n)$ is true. Finally, explain why for all $n \in \mathbb{Z}^+$, $P(n)$ is true.

Chapter 4

Sets

4.1. The language of sets

Sets occur everywhere in mathematics and other subjects whose foundations are mathematical. We have been using the word "set" since Chapter 1, but we have never defined what this word means. In many ways, this situation cannot be avoided. Take a moment now to try to define what you mean by a "set" .... Do you find yourself saying something like "a collection of objects"? Can you define *mathematically* what this means? Of course not; "collection" is just as mathematically vague as "set"!

Just as Euclid took concepts such as "point" and "line" in geometry as *basic* or *undefined*, we will (for now) take the concept of "set" to be undefined. We will insist that for any particular set A, membership in A (i.e., whether an object is in the set A or not) must be well-defined (i.e., for any fixed object x in the underlying universe under discussion, the answer to the question, "Is x in A?" must be either always "yes" or always "no", regardless of when the question is asked). As we have done since Chapter 1, we will use the notation $\in$ to denote membership in a set.

Notation 4.1.1. Given a set A, we write $x \in A$ for "x is an element of the set A", and we write $x \notin A$ for "x is not an element of the set A".

There are several ways that one can specify or describe a set. One way is to explicitly list its elements inside a pair of curly braces, such as

$$\begin{aligned} A &= \{1, 2, 3\}, \\ B &= \{1, 2, 3, \dots, 10\}, \\ C &= \{2, 4, 6, \dots\}. \end{aligned}$$

Here, B appears to be the set of positive integers between 1 and 10, inclusive, and C appears to be the set of even positive integers. The notation "$\dots$", however, is imprecise. We are inferring from the information given that the underlying *universal* set is the set $\mathbb{Z}^+$ of positive integers (recall our discussion of universal sets in Subsection 1.1.3) and that the patterns we see exhibited in the sets B and C continue.

A more precise way of describing a set is to use "set-builder" notation to specify the precise property that the elements of the set should satisfy. This description is sometimes called a *conditional definition* of the set.

Notation 4.1.2. Let $P(x)$ be a statement with a single free variable x. Then the notation

$$\{x \mid P(x)\} \qquad \text{or} \qquad \{x : P(x)\}$$

denotes the set of all objects x in the underlying universal set such that $P(x)$ is true.

When X is a set, the notation $\{x \in X \mid P(x)\}$ is an abbreviation for the set $\{x \mid x \in X \text{ and } P(x)\}$.

Example 4.1.3.

(1) A conditional definition of the set $A = \{1, 2, 3, \dots, 10\}$ above is

$$A = \{n \in \mathbb{Z}^+ \mid 1 \le n \le 10\}.$$

Note here that we have made the underlying universal set explicit.

(2) As with our discussion of quantifiers, the underlying universal set matters. The set

$$D = \{x \in \mathbb{R} \mid 1 \le x \le 10\}$$

is certainly different from the set A above, because $\pi \in D$, while $\pi \notin A$.

Recall that the set D is also denoted in interval notation by $[1, 10]$. (With interval notation, the underlying universal set is always assumed to be $\mathbb{R}$, unless explicitly indicated otherwise. See Notation 4.1.10.)

(3) The set $C = \{2, 4, 6, \dots\}$ of even positive integers can be defined conditionally by

$$\{n \in \mathbb{Z}^+ \mid n \text{ is even}\} \tag{4.1}$$

or, making the quantifier complexity explicit,

$$\{n \in \mathbb{Z}^+ \mid (\exists k \in \mathbb{Z}^+)[n = 2k]\}. \tag{4.2}$$

We can also denote this set using what is sometimes called a *constructive definition*:

$$\{2k \mid k \in \mathbb{Z}^+\}. \tag{4.3}$$

Here, the notation indicates that the elements of the set are all integers of the form $2k$, where k ranges through the set of positive integers.

It is important to remember that

$$\boxed{\textit{the notation in } (4.3) \textit{ is an abbreviation for the set given in } (4.2).}$$

While the notation in (4.3) is more concise, it is dangerous in the sense that it hides the existential quantifier explicitly given in (4.2). Since we have already seen that knowing the logical form of a statement is essential when writing proofs, we must be sure to be aware of such hidden quantifiers. *Students who are learning about sets for the first time may wish, if given a constructive* *definition of a set, to convert the definition into a conditional definition of that set.* ◊

We have already made a very reasonable assumption about sets, namely, that a set is uniquely determined by its elements.[1] The unique set with no elements is denoted by $\emptyset$ and is called the *empty set* or *null set*. It is important to note right away that $\emptyset$ is different from $\{\emptyset\}$. (Be sure you can explain why!)

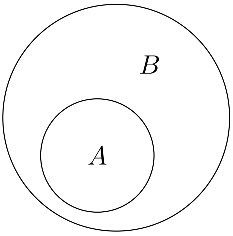
The assumption that a set is uniquely determined by its elements provides the definition of what it means for two sets to be equal (Definition 4.1.6); namely, two sets are equal if they have the same elements. Before working more with this idea, we first define what it means for one set to be "contained" in another set.

Definition 4.1.4. Let A and B be sets. Then A is a *subset* of B (in notation, $A \subseteq B$) if every element of A is also an element of B. Symbolically, $A \subseteq B$ if

$$(\forall x)[x \in A \Rightarrow x \in B]. \tag{4.4}$$

We'll write $A \not\subseteq B$ if A is not a subset of B.

We can denote an arbitrary set as the interior of a circle, so that the following diagram illustrates that $A \subseteq B$.



Take a moment to write the definition of $A \not\subseteq B$ symbolically, by forming a useful denial of (4.4). You'll find this symbolic representation useful when writing proofs.

Example 4.1.5.

(1) $\mathbb{N} \subseteq \mathbb{Z}$, $\mathbb{Z} \subseteq \mathbb{Q}$, $\mathbb{Q} \subseteq \mathbb{R}$.

(2) $\{\{1\}, 2\} \subseteq \{\{1\}, 2, 3\}$, since every element (how many are there?) of the left-hand set is also an element of the right-hand set.

(3) $\{1\} \not\subseteq \{\{1\}, 2, 3\}$, since $1 \in \{1\}$, but $1 \notin \{\{1\}, 2, 3\}$.

(4) Let

$$\begin{aligned} A &= \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[n = 4k + 1]\}, \\ B &= \{n \in \mathbb{Z} \mid n \text{ is odd}\}. \end{aligned}$$

Prove that $A \subseteq B$ and $B \not\subseteq A$.

[1] This assumption is actually the *Axiom of Extensionality* in formal, axiomatic set theory.

Scratchwork. First, we'll give several examples of elements of A, to illustrate the conditional definition of this set.

$$\begin{aligned} -3 \in A, &\text{ since } -3 = 4 \cdot -1 + 1, \\ 1 \in A, &\text{ since } \quad 1 = 4 \cdot 0 + 1, \\ 5 \in A, &\text{ since } \quad 5 = 4 \cdot 1 + 1, \\ 9 \in A, &\text{ since } \quad 9 = 4 \cdot 2 + 1. \end{aligned}$$

Remember that these computations *are not a proof.* These computations simply help build our intuition by giving us a better sense for which integers live in the set and which integers may not.

We must use Definition 4.1.4 to prove that $A \subseteq B$; the Given-Goal diagram is below.

Given	**Goal**
$n \in \mathbb{Z}$ arbitrary	
$n \in A$	$n \in B$

Since we'll be given $n \in A$, the definition of A tells us we may fix $k \in \mathbb{Z}$ such that $n = 4k + 1$. The definition of B tells us that our goal is to show that n is odd, which we certainly know how to do.

To show that $B \nsubseteq A$, we must again use Definition 4.1.4. Negating the statement $(\forall x)[x \in B \Rightarrow x \in A]$, we see that we must show

$$(\exists x)[x \in B \text{ and } x \notin A].$$

Our computations above seem to imply that the odd integer 3 is not an element of A, although we will prove this carefully using a proof by contradiction.

We're ready to write down the formal proofs.

Proof. Let

$$\begin{aligned} A &= \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[n = 4k + 1]\}, \\ B &= \{n \in \mathbb{Z} \mid n \text{ is odd}\}. \end{aligned}$$

We first show that $A \subseteq B$. Let $n \in A$ be arbitrary. We must show that $n \in B$.

Since $n \in A$, we may fix $k \in \mathbb{Z}$ such that $n = 4k + 1$. Then

$$n = 4k + 1 = 2(2k) + 1,$$

and hence n is odd, by definition. Since n is odd, $n \in B$ by definition of B. Thus, $A \subseteq B$.

Next we show that $B \nsubseteq A$. To see this, note that 3 is odd, so $3 \in B$, by definition of B. We claim $3 \notin A$. Assume for the sake of a contradiction that $3 \in A$. Then we may fix $k \in \mathbb{Z}$ such that $3 = 4k + 1$, by definition of A. But then $2 = 4k$, where $k \in \mathbb{Z}$; i.e., $4 \mid 2$, which is a contradiction. Hence $3 \notin A$, and so $B \nsubseteq A$, as desired. □ ◊

We've already noted that the definition of set equality states that a set is uniquely determined by its elements; by Definition 4.1.4, we can also phrase this in terms of the subset relation $\subseteq$.

Definition 4.1.6. Let A and B be sets. Then $A = B$ if

$$(\forall x)[x \in A \Leftrightarrow x \in B].$$

Equivalently, $A = B$ iff $A \subseteq B$ and $B \subseteq A$.

For the particular sets A and B in Example 4.1.5(4), we had $A \subseteq B$ but $A \neq B$ (since $B \nsubseteq A$). In this case, we say that A is a *proper subset* of B and write $A \subsetneq B$.

Example 4.1.7. Let

$$\begin{aligned} A &= \{n \in \mathbb{Z} \mid n+5 \text{ is odd}\}, \\ B &= \{n \in \mathbb{Z} \mid n \text{ is even}\}. \end{aligned}$$

Prove that $A = B$.

Scratchwork. Definition 4.1.6 tells us that we must prove that $A \subseteq B$ and $B \subseteq A$. Definition 4.1.4 tells us that to prove $A \subseteq B$, we use the following Given-Goal diagram.

Given	**Goal**
$n \in \mathbb{Z}$ arbitrary	
$n \in A$	$n \in B$

The situation for $B \subseteq A$ is analogous.

Proof. We first show that $A \subseteq B$. Let $n \in A$; we must show that $n \in B$.

Since $n \in A$, we know that $n + 5$ is odd; hence, we may fix $k \in \mathbb{Z}$ such that $n + 5 = 2k + 1$. To show that $n \in B$, we must show that n is even. Note that

$$n = 2k + 1 - 5 = 2k - 4 = 2(k - 2),$$

and hence n is even, as desired. Thus $n \in B$, and we may conclude that $A \subseteq B$.

Next we show that $B \subseteq A$. Let $n \in B$; we must show that $n \in A$.

Since $n \in B$, we know that n is even. Hence, we may fix $k \in \mathbb{Z}$ such that $n = 2k$. Then

$$n + 5 = 2k + 5 = 2(k + 2) + 1,$$

and hence $n + 5$ is odd. Thus $n \in A$, and we may conclude that $B \subseteq A$.

Since $A \subseteq B$ and $B \subseteq A$, we have that $A = B$, by Definition 4.1.6. □ ◊

So far we have seen some examples of how to show that one specified set is a subset of, or is equal to, another set. We now establish two general results regarding subsets.

Theorem 4.1.8. *Let A, B, and C be sets. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.*

Scratchwork. We begin with an important reminder.

$$\boxed{\textit{In order to start this proof correctly, we must focus on the Goal.}}$$

The beginning Given-Goal diagram is given below.

Given	**Goal**
A, B, C arbitrary sets	
$A \subseteq B$	
$B \subseteq C$	$A \subseteq C$

Note that Definition 4.1.4 tells us how to prove that $A \subseteq C$.

Given	**Goal**
A, B, C arbitrary sets	
$A \subseteq B$	
$B \subseteq C$	
$n \in A$ arbitrary	$n \in C$

Proof. Let A, B, and C be sets, and assume that $A \subseteq B$ and $B \subseteq C$. We show that $A \subseteq C$.

Let $n \in A$ be arbitrary; we must show that $n \in C$. Since $n \in A$ and $A \subseteq B$, we have that $n \in B$, by Definition 4.1.4. Similarly, since $n \in B$ and $B \subseteq C$, we have that $n \in C$, by Definition 4.1.4, as desired.

Hence, by Definition 4.1.4, $A \subseteq C$. □

Recall that the unique set with no elements is $\emptyset$, the empty set.

Proposition 4.1.9. *For all sets A, $\emptyset \subseteq A$ and $A \subseteq A$.*

Scratchwork. The only possible delicate issue here is that the definition of $\subseteq$ is given in terms of $\in$, and $\emptyset$ has no elements! Thus, we take a moment to examine more closely the symbolic form of the statement $\emptyset \subseteq A$, according to Definition 4.1.4:

$$\emptyset \subseteq A \Leftrightarrow (\forall x)[x \in \emptyset \Rightarrow x \in A].$$

Since $\emptyset$ has no elements, any statement of the form $x \in \emptyset$ is false. If we recall the truth table for $\Rightarrow$ in Table 1.5, then it is clear how the proof of $\emptyset \subseteq A$ should go.

Proof. Let A be an arbitrary set. The proof of $A \subseteq A$ is Exercise 4.1.2. We prove that $\emptyset \subseteq A$. We must show

$$(\forall x)[x \in \emptyset \Rightarrow x \in A].$$

Given any arbitrary x, $x \in \emptyset$ is false, and hence the implication

$$x \in \emptyset \Rightarrow x \in A$$

is true. Thus by definition $\emptyset \subseteq A$, as desired. □

We say that a statement such as

$$(\forall x)[x \in \emptyset \Rightarrow x \in A]$$

is *vacuously true*, since there are no x such that $x \in \emptyset$.

We conclude this section by recalling interval notation.

Notation 4.1.10. If $a < b$ are real numbers, then

$$\begin{aligned}
(a, b) &= \{x \in \mathbb{R} \mid a < x < b\}, \\
[a, b] &= \{x \in \mathbb{R} \mid a \leq x \leq b\}, \\
(a, b] &= \{x \in \mathbb{R} \mid a < x \leq b\}, \\
[a, b) &= \{x \in \mathbb{R} \mid a \leq x < b\}, \\
(a, \infty) &= \{x \in \mathbb{R} \mid a < x\}, \\
[a, \infty) &= \{x \in \mathbb{R} \mid a \leq x\}, \\
(-\infty, b) &= \{x \in \mathbb{R} \mid x < b\}, \\
(-\infty, b] &= \{x \in \mathbb{R} \mid x \leq b\}.
\end{aligned}$$

Exercises 4.1

1. State whether the following are true or false. Briefly explain your answers.
   (a) $\{1, 2, 3\} \in \{\{1, 2, 3\}, \{1, 3\}, 1, 2, 3\}$.
   (b) $\{1, 3\} \in \{\{1, 2, 3\}, \{1, 2\}, 1, 3\}$.
   (c) $\{1, 2\} \subseteq \{\{1, 2, 3\}, \{1, 2\}, 1, 3\}$.
   (d) $[5, 6) \subseteq (4, 6]$.
   (e) $(7, 9] \subseteq [6, 9)$.
   (f) $(5, 9] \subseteq [6, 10]$.
   (g) $\{0\} \in \{0, \{0\}\}$.
   (h) $\{0\} \subseteq \{0, \{0\}\}$.
   (i) $\{\{0\}\} \in \{0, \{0\}\}$.
   (j) $\{\{0\}\} \subseteq \{0, \{0\}\}$.
   (k) For every set A, $\{\emptyset\} \subseteq A$.
2. Let A be a set. Prove that $A \subseteq A$.
3. Let A and B be sets. Prove that if $x \notin B$ and $A \subseteq B$, then $x \notin A$.
4. Consider the sets

$$\begin{aligned}
A &= \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})(n = 12k + 11)\}, \\
B &= \{n \in \mathbb{Z} \mid (\exists j \in \mathbb{Z})(n = 4j + 3)\}.
\end{aligned}$$

   (a) Is $A \subseteq B$? Prove or disprove.
   (b) Is $B \subseteq A$? Prove or disprove.
5. Consider the sets

$$\begin{aligned}
A &= \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})(n = 4k + 1)\}, \\
B &= \{n \in \mathbb{Z} \mid (\exists j \in \mathbb{Z})(n = 4j - 7)\}.
\end{aligned}$$

   Prove that $A = B$.

6. Consider the sets

$$\begin{aligned} A &= \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[n = 3k]\}, \\ B &= \{n \in \mathbb{Z} \mid (\exists i, j \in \mathbb{Z})[n = 15i + 12j]\}. \end{aligned}$$

   Prove that $A = B$.

7. Let $A = \{x + y\sqrt{2} \mid x, y \in \mathbb{Q}\} \subseteq \mathbb{R}$.
   (a) Prove that for all $x, y \in \mathbb{Q}$, $x + y\sqrt{2} = 0$ if and only if $x = y = 0$.
   (b) Prove that for all $z_1, z_2 \in A$, $z_1 + z_2, z_1 z_2 \in A$ and, for $z_2 \neq 0$, $\dfrac{z_1}{z_2} \in A$.

8. Let $M = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \,\middle|\, a, b, c, d \in \mathbb{R} \right\}$. Let $X = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M \,\middle|\, ad - bc \neq 0 \right\}$. Prove that for all $A, B \in X$, $AB \in X$ (see Exercise 2.1.20).

4.2. Operations on sets

This section deals with operations we may perform on sets to build "new" sets from "old" ones. We begin with several commonly used set operations.

Definition 4.2.1. Let A and B be sets.

(1) The *union* of A and B is the set

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

(2) The *intersection* of A and B is the set

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

(3) The *complement of A in B*, also called the *difference of B and A*, is the set

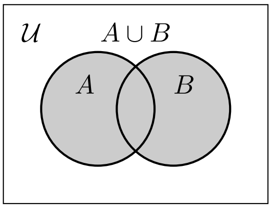
$$\begin{aligned} B - A &= \{x \in B \mid x \notin A\} \\ &= \{x \mid x \in B \text{ and } x \notin A\}. \end{aligned}$$

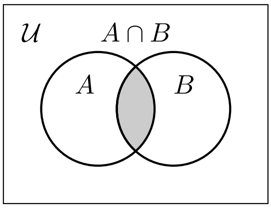
The set difference $B - A$ is sometimes denoted by $B \backslash A$.

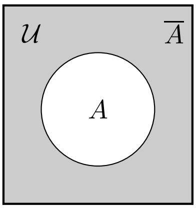
(4) If $\mathcal{U}$ is the universal set under discussion (so that $A \subseteq \mathcal{U}$), then $\mathcal{U} - A$ is denoted by $\overline{A}$ and is called the *complement* of A; i.e.,

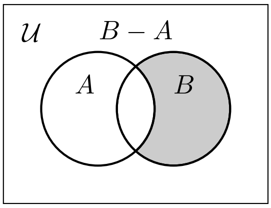
$$\overline{A} = \{x \in \mathcal{U} \mid x \notin A\}.$$

Below we give diagrams, called Venn diagrams, that illustrate the various set operations. In a Venn diagram, the universal set $\mathcal{U}$ is denoted by a rectangle. As before, we denote an arbitrary set as the interior of a circle. In each case, the shaded region represents the set operation being illustrated.









Example 4.2.2. Let

$$A = \{1, 3, 4, 5, 7, 9\}, \quad B = \{3, 6, 7, 10\}, \quad C = \{2, 6\},$$

and let the universal set be $\mathcal{U} = \mathbb{Z}^+$. Then

$$\begin{aligned} A \cup B &= \{x \mid x \in A \text{ or } x \in B\} = \{1, 3, 4, 5, 6, 7, 9, 10\}, \\ A \cap B &= \{x \mid x \in A \text{ and } x \in B\} = \{3, 7\}, \\ A - B &= \{x \in A \mid x \notin B\} = \{1, 4, 5, 9\}, \\ B - A &= \{x \in B \mid x \notin A\} = \{6, 10\}, \\ A \cap C &= \{x \mid x \in A \text{ and } x \in C\} = \emptyset, \\ \overline{C} &= \{n \in \mathbb{Z}^+ \mid n \notin C\} = \{n \in \mathbb{Z}^+ \mid n \neq 2 \text{ and } n \neq 6\}. \end{aligned} \qquad \Diamond$$

Definition 4.2.3. Two sets A and B are *disjoint* if $A \cap B = \emptyset$.

In Example 4.2.2, sets A and C are disjoint. In the next example, the sets (which are subsets of $\mathbb{R}$) are given in interval notation.

Example 4.2.4. This example shows that the union or intersection of two intervals need not be another interval.

$$\begin{aligned} (2, 4] \cap (3, 5) &= (3, 4], \\ [2, 4] \cap (4, 5) &= \emptyset, \\ (2, 4) \cup (3, 5) &= (2, 5), \\ \overline{(2, 4]} &= (-\infty, 2] \cup (4, \infty). \end{aligned} \qquad \Diamond$$

Before stating some useful properties that the set operations $\cup$, $\cap$, and set complement possess, we first include a proof illustrating some of these operations.

Proposition 4.2.5. *Let A, B, and C be sets. If $A \cap B \subseteq C$ and $x \in A - C$, then $x \notin B$.*

Proof. Let A, B, and C be arbitrary sets and assume that $A \cap B \subseteq C$. Assume also that $x \in A - C$; we show that $x \notin B$.

For the sake of a contradiction, assume that $x \in B$. Since $x \in A - C$, $x \in A$ and $x \notin C$, by Definition 4.2.1(3). Since $x \in A$ and $x \in B$, by Definition 4.2.1(2) we know that $x \in A \cap B$. Since $A \cap B \subseteq C$, we may conclude that $x \in C$, a contradiction, since $x \notin C$.

Hence, $x \notin B$, as desired. □

The next theorem enumerates some properties of our new set operations.

Theorem 4.2.6. *Let A, B, and C be subsets of some universal set $\mathcal{U}$. Then:*

(1) $A \cup A = A$.

(2) $A \cap A = A$.

(3) $A \cup \emptyset = A$.

(4) $A \cap \emptyset = \emptyset$.

(5) $A \cap B \subseteq A$.

(6) $A \subseteq A \cup B$.

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

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

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

(10) $A \cap B = B \cap A$.

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

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

(13) $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$.

(14) $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$.

(15) $A \cup \overline{A} = \mathcal{U}$.

(16) $A \cap \overline{A} = \emptyset$.

(17) $\overline{\overline{A}} = A$.

Proof. Let A, B, and C be subsets of some universal set $\mathcal{U}$. We will prove properties (12) and (14), showing all details, and leave the rest for the exercises at the end of the section.

Proof of (12): We show that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

First we show that $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. Let $x \in A \cap (B \cup C)$; we must show that $x \in (A \cap B) \cup (A \cap C)$. Since $x \in A \cap (B \cup C)$, by Definition 4.2.1(2) we know that $x \in A$ and $x \in B \cup C$. Since $x \in B \cup C$, by Definition 4.2.1(1) we know that $x \in B$ or $x \in C$.

Case I: $x \in B$.

Then we have that $x \in A$ and $x \in B$, and hence $x \in A \cap B$ by Definition 4.2.1(2). It follows that $x \in (A \cap B) \cup (A \cap C)$ by Definition 4.2.1(1).

Case II: $x \notin B$.

Then we have that $x \in C$, and since $x \in A$ also, we know that $x \in A \cap C$ by Definition 4.2.1(2). Hence $x \in (A \cap B) \cup (A \cap C)$ by Definition 4.2.1(1).

Thus, in any case, we have that $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

Next, we must show that $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. We begin by letting $x \in (A \cap B) \cup (A \cap C)$; we must show that $x \in A \cap (B \cup C)$.

Since $x \in (A \cap B) \cup (A \cap C)$, we know that $x \in A \cap B$ or $x \in A \cap C$ by Definition 4.2.1(1). As before, we consider two cases.

Case I: $x \in A \cap B$.

Then we have that $x \in A$ and $x \in B$ by Definition 4.2.1(2). Since $x \in B$, we know that $x \in B \cup C$ by Definition 4.2.1(1). Hence $x \in A \cap (B \cup C)$ by Definition 4.2.1(2).

Case II: $x \notin A \cap B$.

Then we have that $x \in A \cap C$. Hence $x \in A$ and $x \in C$ by Definition 4.2.1(2), and it follows that $x \in A$ and $x \in B \cup C$ by Definition 4.2.1(1). Thus we have $x \in A \cap (B \cup C)$ by Definition 4.2.1(2).

Hence $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$, and so $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof of (14): We show that $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$.

First we show that $\overline{(A \cap B)} \subseteq \overline{A} \cup \overline{B}$. Let $x \in \overline{(A \cap B)}$; we must show that $x \in \overline{A} \cup \overline{B}$. Since $x \in \overline{(A \cap B)}$, we know that $x \in \mathcal{U}$ and $x \notin A \cap B$ by Definition 4.2.1(4). It follows by the negation of Definition 4.2.1(2) that $x \notin A$ or $x \notin B$. Hence by Definition 4.2.1(4), $x \in \overline{A}$ or $x \in \overline{B}$; i.e., $x \in \overline{A} \cup \overline{B}$, by Definition 4.2.1(1), as desired. Thus $\overline{(A \cap B)} \subseteq \overline{A} \cup \overline{B}$.

Next we show that $\overline{A} \cup \overline{B} \subseteq \overline{(A \cap B)}$. Let $x \in \overline{A} \cup \overline{B}$; we must show that $x \in \overline{(A \cap B)}$. Since $x \in \overline{A} \cup \overline{B}$, we know by Definition 4.2.1(1) that $x \in \overline{A}$ or $x \in \overline{B}$. If $x \in \overline{A}$, then $x \in \mathcal{U}$ and $x \notin A$ by Definition 4.2.1(4). Since $x \notin A$, $x \notin A \cap B$ by Definition 4.2.1(2). Hence $x \in \overline{A \cap B}$ by Definition 4.2.1(4). Similarly, if $x \notin \overline{A}$, then $x \in \overline{B}$. It follows by Definition 4.2.1(4) that $x \in \mathcal{U}$ and $x \notin B$. Since $x \notin B$, $x \notin A \cap B$ by Definition 4.2.1(2). Hence $x \in \overline{A \cap B}$ by Definition 4.2.1(4). Thus, in any case, $x \in \overline{(A \cap B)}$ and hence $\overline{A} \cup \overline{B} \subseteq \overline{(A \cap B)}$.

It follows that $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$. □

Just as in Proposition 1.1.7, properties (13) and (14) from Theorem 4.2.6 are often called *DeMorgan's Laws.*

You are probably familiar with the next set operation.

Definition 4.2.7. Let A and B be sets. The *Cartesian product* of A and B is the set

$$A \times B = \{(x, y) \mid x \in A \text{ and } y \in B\},$$

where (x, y) denotes the *ordered pair*[2] containing x and y in that order. More generally, if $n \in \mathbb{Z}^+$ and A_1, A_2, …, A_n, A are sets, then

$$A_1 \times A_2 \times \cdots \times A_n = \{(x_1, x_2, \ldots, x_n) \mid \text{ for all } i, 1 \leq i \leq n, x_i \in A_i\}$$

is a set of ordered n-tuples and

$$A^n = \{(x_1, x_2, \ldots, x_n) \mid \text{ for all } i, 1 \leq i \leq n, x_i \in A\}.$$

Example 4.2.8.

(1) Let $A = \{1, 2\}$ and $B = \{\pi, e, \{0\}\}$.

$$A \times B = \{(1, \pi), (1, e), (1, \{0\}), (2, \pi), (2, e), (2, \{0\})\},$$
$$B \times A = \{(\pi, 1), (\pi, 2), (e, 1), (e, 2), (\{0\}, 1), (\{0\}, 2)\}.$$

[2] Technically, an ordered pair is a set. See Exercise 4.2.25.

(2) $\mathbb{R}^2$ is the familiar *Cartesian*, or *Euclidean*, plane, $\mathbb{R}^3$ is Euclidean space, and $\mathbb{R}^n$ is n-dimensional Euclidean space.

$$\begin{aligned} \mathbb{R} \times \mathbb{R} = \mathbb{R}^2 &= \{(x, y) \mid x, y \in \mathbb{R}\}, \\ \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3 &= \{(x, y, z) \mid x, y, z \in \mathbb{R}\}, \\ \mathbb{R}^n &= \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in \mathbb{R}\}. \end{aligned} \qquad \Diamond$$

We define two ordered pairs (x, y) and (a, b) to be *equal* if and only if $x = a$ and $y = b$. Thus we see from Example 4.2.8(1) that, in general, $A \times B \neq B \times A$. More generally,

$$(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n) \text{ iff for all } i, 1 \leq i \leq n, x_i = y_i.$$

Note also in Example 4.2.8(1) that the set A has 2 elements, B has 3 elements, and $A \times B$ (and $B \times A$) has 6 elements. In general it can be proved (see Theorem 8.2.7) that when A has m elements and B has n elements, where $m, n \in \mathbb{Z}$ with $m, n \geq 0$, then $A \times B$ has mn elements.

The following proposition indicates some of the relationships between the Cartesian product and the other set operations.

Proposition 4.2.9. *Let A, B, C, and D be sets. Then:*

(1) $A \times \emptyset = \emptyset = \emptyset \times A$.

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

(3) $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

(4) $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.

(5) $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$. *In general, equality need not hold.*

Proof. Let A, B, C, and D be sets. We prove parts (2) and (5); the rest of the proofs are left for the exercises.

Proof of (2): We show $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

Let $(x, y) \in A \times (B \cup C)$; we must show that $(x, y) \in (A \times B) \cup (A \times C)$. Since $(x, y) \in A \times (B \cup C)$, we know that $x \in A$ and $y \in B \cup C$; i.e., $y \in B$ or $y \in C$. If $y \in B$, then $(x, y) \in A \times B$, and hence $(x, y) \in (A \times B) \cup (A \times C)$. If $y \notin B$, then $y \in C$. Thus $(x, y) \in (A \times B) \cup (A \times C)$, since $(x, y) \in A \times C$. Hence $A \times (B \cup C) \subseteq (A \times B) \cup (A \times C)$.

Next, let $(x, y) \in (A \times B) \cup (A \times C)$; we show that $(x, y) \in A \times (B \cup C)$. Since $(x, y) \in (A \times B) \cup (A \times C)$, we know that $(x, y) \in A \times B$ or $(x, y) \in A \times C$. If $(x, y) \in A \times B$, then $x \in A$ and $y \in B$, so $y \in B \cup C$. Otherwise, $(x, y) \in A \times C$, so $x \in A$ and $y \in C$, so $y \in B \cup C$. Thus, in any case, $(x, y) \in A \times (B \cup C)$. Hence $(A \times B) \cup (A \times C) \subseteq A \times (B \cup C)$.

It follows that $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

Proof of (5): We show that $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.

Let $(x, y) \in (A \times B) \cup (C \times D)$; we show that $(x, y) \in (A \cup C) \times (B \cup D)$. Since $(x, y) \in (A \times B) \cup (C \times D)$, we know that $(x, y) \in A \times B$ or $(x, y) \in C \times D$. If $(x, y) \in A \times B$, then $x \in A$ and $y \in B$, so $x \in A \cup C$ and $y \in B \cup D$. If $(x, y) \notin A \times B$, then $(x, y) \in C \times D$. So, $x \in C$ and $y \in D$, and hence again

$x \in A \cup C$ and $y \in B \cup D$. Hence in any case, $(x, y) \in (A \cup C) \times (B \cup D)$. Thus $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.

We next show that equality need not hold in general, i.e., that, in general, $(A \cup C) \times (B \cup D) \subseteq (A \times B) \cup (C \times D)$ is false. We must provide a counterexample. Let $A = \mathbb{Z} = D$ and $B = \emptyset = C$. Then $A \times B = \emptyset = C \times D$, by part (1) of this proposition. Thus $(A \times B) \cup (C \times D) = \emptyset$ and

$$(A \cup C) \times (B \cup D) = \mathbb{Z} \times \mathbb{Z} \neq \emptyset = (A \times B) \cup (C \times D).$$ □

Our last operation defines the collection of all subsets of a set.

Definition 4.2.10. Let X be a set. The *power set* of X is the set

$$\mathcal{P}(X) = \{A \mid A \subseteq X\},$$

the set of all subsets of X.

Example 4.2.11. Let $X = \{1, 2, 3\}$. Note that

$$\{1, 2\} \subseteq X, \text{ so } \{1, 2\} \in \mathcal{P}(X).$$

Also,

$$\begin{array}{rcr} 1 \notin \mathcal{P}(X), & \text{since} & 1 \nsubseteq X, \\ \{1\} \in \mathcal{P}(X), & \text{since} & \{1\} \subseteq X, \\ \emptyset \in \mathcal{P}(X), & \text{since} & \emptyset \subseteq X. \end{array}$$

We can list all the elements of $\mathcal{P}(X)$:

$$\mathcal{P}(X) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$ ◊

Note that in Example 4.2.11, the set X has 3 elements and $\mathcal{P}(X)$ has $2^3 = 8$ elements. In general, if a set X has n elements, where $n \in \mathbb{Z}$, $n \geq 0$, then $\mathcal{P}(X)$ has 2^n elements (see Exercise 4.2.26). This fact can help you check that you have the right number of subsets when you are computing the power set of a finite set.

The next proposition shows how the power set operation interacts with subsethood.

Proposition 4.2.12. *Let A and B be sets. Then*

$$A \subseteq B \Leftrightarrow \mathcal{P}(A) \subseteq \mathcal{P}(B).$$

Scratchwork. We will prove only the backward direction and leave the forward direction for Exercise 4.2.19. Our Given-Goal diagram is below.

Given	**Goal**
A, B arbitrary sets	
$\mathcal{P}(A) \subseteq \mathcal{P}(B)$	$A \subseteq B$

As always,

the structure of the proof is determined by the Goal,

so we rewrite the Given-Goal diagram.

Given	**Goal**
A, B arbitrary sets	
$\mathcal{P}(A) \subseteq \mathcal{P}(B)$	
$x \in A$ arbitrary	$x \in B$

In order to make use of the hypothesis that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$, we need to turn information about $x \in A$ into information about $\mathcal{P}(A)$. The important thing to remember is that *elements* of $\mathcal{P}(A)$ correspond to *subsets* of A, by Definition 4.2.10. We have an element x in A; in order to use the hypothesis that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$, we must use x to find a *subset* of A.

Proof. Let A and B be sets. The proof that

$$A \subseteq B \Rightarrow \mathcal{P}(A) \subseteq \mathcal{P}(B)$$

is Exercise 4.2.19.

We prove

$$\mathcal{P}(A) \subseteq \mathcal{P}(B) \Rightarrow A \subseteq B.$$

Assume that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$; we must show that $A \subseteq B$. Let $x \in A$ be arbitrary. Since $x \in A$, by definition, $\{x\} \subseteq A$. Thus by Definition 4.2.10, $\{x\} \in \mathcal{P}(A)$. Since $\mathcal{P}(A) \subseteq \mathcal{P}(B)$, we know that $\{x\} \in \mathcal{P}(B)$. Then, again by Definition 4.2.10, $\{x\} \subseteq B$. Hence, $x \in B$, as desired.

Thus, $A \subseteq B$. □

In Exercise 4.2.20, you are asked to use the fact that $A \subseteq A$ to provide a different proof of

$$\mathcal{P}(A) \subseteq \mathcal{P}(B) \Rightarrow A \subseteq B.$$

Exercises 4.2

1. Let A, B, and C be subsets of some universal set $\mathcal{U}$. Prove the following statements from Theorem 4.2.6.
   (a) $A \cup A = A$ and $A \cap A = A$.
   (b) $A \cup \emptyset = A$ and $A \cap \emptyset = \emptyset$.
   (c) $A \cap B \subseteq A$ and $A \subseteq A \cup B$.
   (d) $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$.
   (e) $A \cup B = B \cup A$ and $A \cap B = B \cap A$.
   (f) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
   (g) $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$.
   (h) $A \cup \overline{A} = \mathcal{U}$ and $A \cap \overline{A} = \emptyset$.
   (i) $\overline{\overline{A}} = A$.
2. Let A and B be sets, where $\mathcal{U}$ is the underlying universal set. Prove that $A \subseteq B \Leftrightarrow \overline{B} \subseteq \overline{A}$.

3. Let A and B be sets.
   (a) Prove that $A \subseteq B \Leftrightarrow A \cap B = A$.
   (b) Prove that $A \subseteq B \Leftrightarrow A \cup B = B$.
4. Let A, B, and C be sets. Prove that if $A \subseteq B \cup C$ and $A \cap B = \emptyset$, then $A \subseteq C$.
5. Let A and B be sets.
   (a) Prove that $A = (A \cap B) \cup (A - B)$.
   (b) Prove that $A \cup B = A \cup (B - A)$.
   (c) Prove that $A - B = A \cap \overline{B}$.
   (d) Prove that $(A \cup B) \cap \overline{A} = B - A$.
   (e) Prove that $(A - B) \cup (B - A) = (A \cup B) - (A \cap B)$.
6. Let A, B, C, and D be sets with $C \subseteq A$ and $D \subseteq B$. Prove that $D - A \subseteq B - C$.
7. Let A, B, and C be sets. Prove:
   (a) $(A \cup B) - C \subseteq (A - C) \cup B$.
   (b) $(A \cup B) - C = (A - C) \cup B$ iff $B \cap C = \emptyset$.
8. Let A, B, and C be sets.
   (a) Prove or disprove: if $A \subseteq B \cup C$, then $A \subseteq B$ or $A \subseteq C$.
   (b) State the converse of part (a) and prove or disprove.
9. Let A, B, and C be sets.
   (a) Prove or disprove: if $A \subseteq B \cap C$, then $A \subseteq B$ and $A \subseteq C$.
   (b) State the converse of part (a) and prove or disprove.
10. Let A, B, and C be sets.
    (a) Prove or disprove: if $A - C \subseteq B - C$, then $A \subseteq B$.
    (b) State the converse of part (a) and prove or disprove.
11. Let $A = \{a, b\}$, $B = \{1, 2\}$, and $C = \{c, d, e\}$. Find $A \times B$ and $C \times A$. Explain why $A \times (B \times C) \neq (A \times B) \times C$.
12. Let A, B, C, and D be sets. Prove the following statements from Proposition 4.2.9.
    (a) $A \times \emptyset = \emptyset = \emptyset \times A$.
    (b) $A \times (B \cap C) = (A \times B) \cap (A \times C)$.
    (c) $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.
13. Let A, B, C be sets with $A \neq \emptyset$. Prove that if $A \times B = A \times C$, then $B = C$. Is the statement still true if $A = \emptyset$? Prove your answer.
14. Let A and B be nonempty sets. Prove that $A \times B = B \times A$ iff $A = B$. Is this statement true if one of A or B is empty? Prove your answer.
15. Which of the following statements are true for every set A? Explain.
$$\begin{aligned} &\emptyset \subseteq \mathcal{P}(A), \qquad && \emptyset \in \mathcal{P}(A), \\ &A \subseteq \mathcal{P}(A), \qquad && A \in \mathcal{P}(A). \end{aligned}$$
16. Find the power set $\mathcal{P}(X)$ for the following sets.
    (a) $X = \{1, 2\}$.
    (b) $X = \{0, \triangle, \square\}$.
    (c) $X = \{1, \{2, \{3\}\}\}$.
    (d) $X = \{a, b, \{a, b\}\}$.
17. Let $A = \{1, 2\}$. Find $\mathcal{P}(\mathcal{P}(A))$.

18. Let $\mathcal{U} = \{1, 2, 3\}$ be the universal set for $A = \{1, 2\}$ and $B = \{2, 3\}$. Find
    (a) $\mathcal{P}(A) \cap \mathcal{P}(B)$,
    (b) $\mathcal{P}(\overline{A}) \cup \mathcal{P}(\overline{B})$,
    (c) $\mathcal{P}(A) - \mathcal{P}(B)$.

19. Let A and B be sets such that $A \subseteq B$. Prove that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

20. Let A and B be sets such that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$. Use the fact that $A \subseteq A$ to show that $A \subseteq B$.

21. Let A and B be sets.
    (a) Prove or disprove: $\mathcal{P}(A \cup B) = \mathcal{P}(A) \cup \mathcal{P}(B)$.
    (b) Prove or disprove: $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cap \mathcal{P}(B)$.
    (c) Prove or disprove: $\mathcal{P}(A - B) = \mathcal{P}(A) - \mathcal{P}(B)$.

22. Prove that for all $a, b \in \mathbb{R}$ with $a \neq b$ there exist $\varepsilon, \delta \in \mathbb{R}^+$ such that

$$\{x \in \mathbb{R} \mid |x - a| < \varepsilon\} \cap \{x \in \mathbb{R} \mid |x - b| < \delta\} = \emptyset.$$

23. Let X be a set. For sets $A, B \in \mathcal{P}(X)$, define the *symmetric difference* $A \vartriangle B$ of A and B by

$$A \vartriangle B = (A - B) \cup (B - A) = (A \cup B) - (A \cap B)$$

(see Exercise 4.2.5e).

(a) Let $X = \mathbb{Z}$ and

$$\begin{aligned} A &= \{n \in \mathbb{Z} \mid n \geq 5\}, & B &= \{n \in \mathbb{Z} \mid n \leq 7\}, \\ C &= \{-7, -2, 1, 5, 9, 11\}, & D &= \{-2, 0, 3, 5, 10, 11, 12\}. \end{aligned}$$

Find $A \vartriangle B$ and $C \vartriangle D$.

(b) Let

$$X = \mathbb{R}, \qquad A = [-3, 1] \cup (6, 9), \qquad B = [-2, 10).$$

Find $A \vartriangle B$.

(c) Prove that for all sets $A, B, C \in \mathcal{P}(X)$,
    (i) $A \vartriangle B = B \vartriangle A$ ($\vartriangle$ is commutative),
    (ii) $A \vartriangle (B \vartriangle C) = (A \vartriangle B) \vartriangle C$ ($\vartriangle$ is associative).

(d) Prove that there exists a unique set $I \in \mathcal{P}(X)$ such that for all sets $A \in \mathcal{P}(X)$, $A \vartriangle I = A$.

(e) Prove that for the set I in part (d) above, for all sets $A \in \mathcal{P}(X)$, there exists a unique set $B \in \mathcal{P}(X)$ such that $A \vartriangle B = I$.

(f) Prove that for all sets $A, B \in \mathcal{P}(X)$, there exists a unique set $C \in \mathcal{P}(X)$ such that $A \vartriangle C = B$.

24. We say that a set $S \subseteq \mathbb{R}$ of real numbers is *open* if for all $x \in S$, there exists a real number $r > 0$ such that $(x - r, x + r) \subseteq S$. The set S is *closed* if $\mathbb{R} - S$ is open.
    (a) Let $a, b \in \mathbb{R}$ with $a < b$. Prove that the intervals (a, b), (a, ∞), $(-\infty, b)$, and $(-\infty, \infty)$ are open.
    (b) Prove that $\mathbb{R}$ and $\emptyset$ are both open and closed.
    (c) Let $U, V \subseteq \mathbb{R}$ be open sets. Prove that $U \cup V$ and $U \cap V$ are open.
    (d) Let $a, b \in \mathbb{R}$ with $a < b$. Prove that the intervals $[a, b]$, $[a, \infty)$, $(-\infty, b]$ are closed.

25. In a course in formal set theory, the ordered pair (x, y) is defined to be the set $\{\{x\}, \{x, y\}\}$. Use this definition to prove that
$$(x, y) = (z, w) \text{ iff } x = z \text{ and } y = w.$$
26. Use induction on $n \geq 0$ to prove that if the set A has n elements, then $\mathcal{P}(A)$ has 2^n elements. (This exercise is also Exercise 8.2.9a, where a hint is given.)

4.3. Arbitrary unions and intersections

4.3.1. Arbitrary finite union and intersection. In the previous sections, we have examined several types of operations on sets, including the union $A \cup B$ and intersection $A \cap B$ of two sets A and B. If we are given three sets, say $A = \{1, 3, 5\}$, $B = \{2, 5, 7\}$, and $C = \{3, 5\}$, then only a moment's thought tells you that $A \cup B \cup C = \{1, 2, 3, 5, 7\}$ and $A \cap B \cap C = \{5\}$. However, since union and intersection are operations defined on two sets, not three, one must be more careful to define what one means by, say, $A \cup B \cup C$. There are two obvious ways to define it:
$$\begin{aligned} A \cup B \cup C &= (A \cup B) \cup C \quad \text{or} \\ A \cup B \cup C &= A \cup (B \cup C), \end{aligned}$$
and by Exercise 4.2.1d, both lead to the same set. This means that the notation $A \cup B \cup C$ is unambiguous, and we can therefore generalize this notion to n sets, where $n \in \mathbb{Z}^+$, using induction.

For the remainder of this section, we fix a universal set $\mathcal{U}$.

Definition 4.3.1. Let $n \in \mathbb{Z}^+$ and $A_1, A_2, \ldots, A_n$ be sets. Then
$$\begin{aligned} \bigcup_{i=1}^{n} A_i &= A_1 \cup A_2 \cup \cdots \cup A_n \\ &= \{x \in \mathcal{U} \mid \text{there exists } i \in \mathbb{Z}^+ \text{ with } 1 \leq i \leq n \text{ such that } x \in A_i\}, \\ \bigcap_{i=1}^{n} A_i &= A_1 \cap A_2 \cap \cdots \cap A_n \\ &= \{x \in \mathcal{U} \mid \text{for all } i \in \mathbb{Z}^+ \text{ with } 1 \leq i \leq n,\ x \in A_i\}. \end{aligned}$$

Not surprisingly, parts (11)–(14) of Theorem 4.2.6 generalize to this setting.

Theorem 4.3.2. *Let $n \in \mathbb{Z}^+$. Then for all sets $A, B_1, B_2, \ldots, B_n$,*

(1) $A \cup (B_1 \cap B_2 \cap \cdots \cap B_n) = (A \cup B_1) \cap (A \cup B_2) \cap \cdots \cap (A \cup B_n)$.

(2) $A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)$.

(3) $\overline{B_1 \cup B_2 \cup \cdots \cup B_n} = \overline{B_1} \cap \overline{B_2} \cap \cdots \cap \overline{B_n}$.

(4) $\overline{B_1 \cap B_2 \cap \cdots \cap B_n} = \overline{B_1} \cup \overline{B_2} \cup \cdots \cup \overline{B_n}$.

Proof. We prove only part (2) by induction on $n \geq 1$, leaving the other parts for Exercise 4.3.6.

Base Case: Let A, B_1 be sets.

The statement of the Base Case is that

$$A \cap B_1 = A \cap B_1,$$

which is certainly true.

Inductive Step: Let $m \geq 1$ and assume that for all sets $C, D_1, \ldots, D_m$,

$$C \cap (D_1 \cup D_2 \cup \cdots \cup D_m) = (C \cap D_1) \cup (C \cap D_2) \cup \cdots \cup (C \cap D_m).$$

Next, let $A, B_1, \ldots, B_{m+1}$ be arbitrary sets. We must prove that

$$A \cap (B_1 \cup B_2 \cup \cdots \cup B_{m+1}) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_{m+1}).$$

First note that

$$\begin{aligned} A \cap (B_1 \cup B_2 \cup \cdots \cup B_{m+1}) &= A \cap ((B_1 \cup B_2 \cup \cdots \cup B_m) \cup B_{m+1}) \\ &= (A \cap (B_1 \cup B_2 \cup \cdots \cup B_m)) \cup (A \cap B_{m+1}) \end{aligned}$$

by Theorem 4.2.6(12). Then

$$\begin{aligned} &(A \cap (B_1 \cup B_2 \cup \cdots \cup B_m)) \cup (A \cap B_{m+1}) \\ &\qquad = ((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m)) \cup (A \cap B_{m+1}) \end{aligned}$$

by the Induction Hypothesis for $C = A$ and $D_i = B_i$, $1 \leq i \leq m$. Hence

$$\begin{aligned} &((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m)) \cup (A \cap B_{m+1}) \\ &\qquad = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m) \cup (A \cap B_{m+1}) \end{aligned}$$

as desired.

Hence, by mathematical induction we have that for all $n \in \mathbb{Z}^+$ and for all sets A, $B_1, B_2, \ldots, B_n$,

$$A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n). \qquad \square$$

Notice in particular the universal quantifiers on the sets in Theorem 4.3.2 and how these universal quantifiers affected the Inductive Step of the preceding proof.

4.3.2. Index sets. Another way to visualize the finite union $\bigcup_{i=1}^{n} A_i$ (where $i \in \mathbb{Z}^+$ and for all $1 \leq i \leq n$, A_i is a set) is to note that the subscripts $1, 2, \ldots, n$ on the sets $A_1, A_2, \ldots, A_n$ form an *index set* $I = \{1, 2, \ldots, n\}$. Each element $i \in I$ corresponds to a set A_i, and $\{A_i \mid i \in I\}$ is called an *indexed family of sets.* Then

$$\bigcup_{i=1}^{n} A_i = \{x \in \mathcal{U} \mid (\exists i \in I)[x \in A_i]\}$$

and, analogously,

$$\bigcap_{i=1}^{n} A_i = \{x \in \mathcal{U} \mid (\forall i \in I)[x \in A_i]\}.$$

If we have an infinite list $A_1, A_2, A_3, \ldots$ of sets, then the index set is $\mathbb{Z}^+$.

Definition 4.3.3. Given sets A_i, $i \in \mathbb{Z}^+$, with underlying universal set $\mathcal{U}$, the union $\bigcup_{i=1}^\infty A_i$ and intersection $\bigcap_{i=1}^\infty A_i$ are defined by

$$\bigcup_{i=1}^\infty A_i = \{x \in \mathcal{U} \mid (\exists i \in \mathbb{Z}^+)[x \in A_i]\},$$

$$\bigcap_{i=1}^\infty A_i = \{x \in \mathcal{U} \mid (\forall i \in \mathbb{Z}^+)[x \in A_i]\}.$$

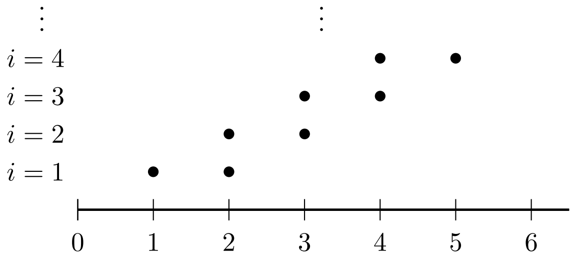
We may also denote $\bigcup_{i=1}^\infty A_i$ by $\bigcup_{i\in\mathbb{Z}^+} A_i$ and $\bigcap_{i=1}^\infty A_i$ by $\bigcap_{i\in\mathbb{Z}^+} A_i$.

We begin with an easy example to illustrate these concepts.

Example 4.3.4. Given $i \in \mathbb{Z}^+$, let $A_i = \{i, i+1\}$; i.e.,

$$A_1 = \{1,2\}, \quad A_2 = \{2,3\}, \quad A_3 = \{3,4\}, \quad \ldots .$$

We'll find $\bigcup_{i=1}^\infty A_i$ and $\bigcap_{i=1}^\infty A_i$. A picture is useful for visualizing these sets. We draw the sets $A_1, A_2, \ldots$ above a number line:



For $\bigcup_{i=1}^\infty A_i$, we want the collection of all numbers that show up in at least one A_i, so it appears that $\bigcup_{i=1}^\infty A_i = \mathbb{Z}^+$. For $\bigcap_{i=1}^\infty A_i$, we want the collection of all numbers that show up in all the A_i's, so it appears that $\bigcap_{i=1}^\infty A_i = \emptyset$.

Proving that our claims are true is not difficult, but it does require that we completely understand Definition 4.3.3. Students who are encountering these ideas for the first time may wish to omit the proofs on their first reading. For this first example, we'll show all details in order to fully illustrate the concepts.

First, to show that $\bigcup_{i=1}^\infty A_i = \mathbb{Z}^+$, we must show that $\bigcup_{i=1}^\infty A_i \subseteq \mathbb{Z}^+$ and $\mathbb{Z}^+ \subseteq \bigcup_{i=1}^\infty A_i$. The containment $\bigcup_{i=1}^\infty A_i \subseteq \mathbb{Z}^+$ is automatic, since the elements of A_i are positive integers by definition.

Next, the Given-Goal diagram for $\mathbb{Z}^+ \subseteq \bigcup_{i=1}^\infty A_i$ is

Given	**Goal**
$n \in \mathbb{Z}^+$	$(\exists i \in \mathbb{Z}^+)[n \in A_i]$

Here we will use the definition of the A_i's.

Similarly, to show that $\bigcap_{i=1}^\infty A_i = \emptyset$, we must show that $\emptyset \subseteq \bigcap_{i=1}^\infty A_i$ and $\bigcap_{i=1}^\infty A_i \subseteq \emptyset$. We know the first statement is true by Proposition 4.1.9. For the second statement, Exercise 4.2.2 tells us we can prove $\mathbb{Z}^+ \subseteq \overline{\bigcap_{i=1}^\infty A_i}$ instead, since $\overline{\emptyset} = \mathbb{Z}^+$ in this context. Our Given-Goal diagram is

Given	**Goal**
$n \in \mathbb{Z}^+$ arbitrary	$n \notin \bigcap\limits_{i=1}^{\infty} A_i$ i.e., $(\exists i \in \mathbb{Z}^+)[n \notin A_i]$

So, our goal is to demonstrate a particular $i \in \mathbb{Z}^+$ with $n \notin A_i$.

We are now ready to prove that $\bigcup_{i=1}^{\infty} A_i = \mathbb{Z}^+$ and $\bigcap_{i=1}^{\infty} A_i = \emptyset$.

Proof. First we show that $\bigcup_{i=1}^{\infty} A_i = \mathbb{Z}^+$. Let $n \in \bigcup_{i=1}^{\infty} A_i$. We must show that $n \in \mathbb{Z}^+$. Since $n \in \bigcup_{i=1}^{\infty} A_i$, by Definition 4.3.3 we know we may fix $i \in \mathbb{Z}^+$ such that $n \in A_i$. Since $A_i \subseteq \mathbb{Z}^+$, we immediately have that $n \in \mathbb{Z}^+$. Hence $\bigcup_{i=1}^{\infty} A_i \subseteq \mathbb{Z}^+$.

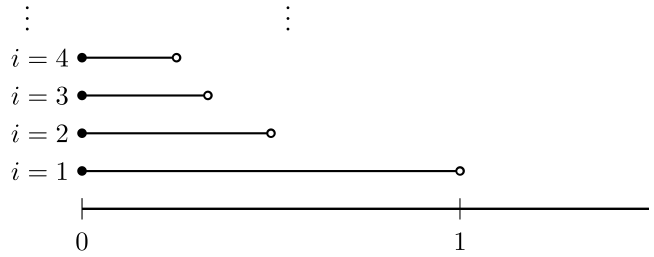
Next let $n \in \mathbb{Z}^+$. We must show that $n \in \bigcup_{i=1}^{\infty} A_i$. Note that $A_n = \{n, n+1\}$, so that $n \in A_n$, and hence $n \in \bigcup_{i=1}^{\infty} A_i$, as desired. Thus $\mathbb{Z}^+ \subseteq \bigcup_{i=1}^{\infty} A_i$ and hence $\mathbb{Z}^+ = \bigcup_{i=1}^{\infty} A_i$.

Next we show that $\bigcap_{i=1}^{\infty} A_i = \emptyset$. Note first that $\emptyset \subseteq \bigcap_{i=1}^{\infty} A_i$ by Proposition 4.1.9. To show that $\bigcap_{i=1}^{\infty} A_i \subseteq \emptyset$, we let $n \in \mathbb{Z}^+$ be arbitrary. We must show that $n \notin \bigcap_{i=1}^{\infty} A_i$. Note that $n \notin A_{n+1} = \{n+1, n+2\}$, and hence by Definition 4.3.3, $n \notin \bigcap_{i=1}^{\infty} A_i$. Thus $\bigcap_{i=1}^{\infty} A_i = \emptyset$, as desired. □ ◊

A typical example in an analysis course is the following.

Example 4.3.5. Given $i \in \mathbb{Z}^+$, define $A_i = [0, \frac{1}{i})$. Find $\bigcup_{i \in \mathbb{Z}^+} A_i$ and $\bigcap_{i \in \mathbb{Z}+} A_i$.

As before, we first make a picture to illustrate the family of sets.



We claim that $\bigcup_{i \in \mathbb{Z}^+} A_i = [0, 1)$ and $\bigcap_{i \in \mathbb{Z}^+} A_i = \{0\}$.

Once again, students who are encountering these ideas for the first time may wish to omit the proof on their first reading. The proof relies on the result of Exercise 2.4.5, which follows from the Completeness Axiom for $\mathbb{R}$ and is proved in Corollary 9.3.2(3).

Proof. First we show that $\bigcup_{i \in \mathbb{Z}^+} A_i = [0, 1)$. Let $x \in \bigcup_{i \in \mathbb{Z}^+} A_i$, and by Definition 4.3.3, fix $i \in \mathbb{Z}^+$ such that $x \in A_i$. We must show that $x \in [0, 1)$. Since $x \in A_i$, we know that $x \in [0, \frac{1}{i})$. Thus $0 \leq x < \frac{1}{i} \leq 1$, since $i \geq 1$. Hence $x \in [0, 1)$ as desired and $\bigcup_{i \in \mathbb{Z}^+} A_i \subseteq [0, 1)$.

Next let $x \in [0, 1)$. Then $x \in A_1$ by definition, and hence $x \in \bigcup_{i \in \mathbb{Z}^+} A_i$ by definition. Thus $[0, 1) \subseteq \bigcup_{i \in \mathbb{Z}^+} A_i$, and hence $\bigcup_{i \in \mathbb{Z}^+} A_i = [0, 1)$.

Next we show that $\bigcap_{i\in\mathbb{Z}^+} A_i = \{0\}$. Let $x \in \bigcap_{i\in\mathbb{Z}^+} A_i$. We must show that $x = 0$. Assume for a contradiction that $x \neq 0$; then $x > 0$, since all elements of $\bigcap_{i\in\mathbb{Z}^+} A_i$ are nonnegative. By the result of Exercise 2.4.5, we may fix $n \in \mathbb{Z}^+$ such that $\frac{1}{n} < x$. Thus $x \notin A_n = [0, \frac{1}{n})$, and hence $x \notin \bigcap_{i\in\mathbb{Z}^+} A_i$, by Definition 4.3.3. This is a contradiction, and hence $x = 0$. Thus $\bigcap_{i\in\mathbb{Z}^+} A_i \subseteq \{0\}$.

Next let $x \in \{0\}$; i.e., let $x = 0$. We must show that $x \in \bigcap_{i\in\mathbb{Z}^+} A_i$. Let $i \in \mathbb{Z}^+$ be arbitrary. Since $A_i = [0, \frac{1}{i})$, $0 \in A_i$. Thus $x \in \bigcap_{i\in\mathbb{Z}^+} A_i$ by Definition 4.3.3, and hence $\{0\} \subseteq \bigcap_{i\in\mathbb{Z}^+} A_i$. It follows that $\bigcap_{i\in\mathbb{Z}^+} A_i = \{0\}$. □ ◊

In the previous examples, our index sets have been subsets of the integers. Note that we may take any nonempty set I to be an index set.

Definition 4.3.6. Let I be a nonempty set and let $\{A_i \mid i \in I\}$ be a family of sets indexed by I, with underlying universal set $\mathcal{U}$. Then

$$\bigcup_{i\in I} A_i = \{x \in \mathcal{U} \mid (\exists i \in I)[x \in A_i]\},$$

$$\bigcap_{i\in I} A_i = \{x \in \mathcal{U} \mid (\forall i \in I)[x \in A_i]\}.$$

We use this definition to prove that the generalization of Theorem 4.2.6 holds in this context.

Theorem 4.3.7. *Let I be a nonempty set and let $\{A_i \mid i \in I\}$ be an indexed family of sets, relative to some universal set $\mathcal{U}$. Let B be a set. Then:*

(1) *For each $j \in I$, $\displaystyle\bigcap_{i\in I} A_i \subseteq A_j$.*

(2) *For each $j \in I$, $\displaystyle A_j \subseteq \bigcup_{i\in I} A_i$.*

(3) $\displaystyle B \cup \bigcap_{i\in I} A_i = \bigcap_{i\in I}(B \cup A_i)$.

(4) $\displaystyle B \cap \bigcup_{i\in I} A_i = \bigcup_{i\in I}(B \cap A_i)$.

(5) $\displaystyle \overline{\bigcup_{i\in I} A_i} = \bigcap_{i\in I} \overline{A_i}$.

(6) $\displaystyle \overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A_i}$.

Proof. Let I be a nonempty set and let $\{A_i \mid i \in I\}$ be an indexed family of sets, relative to some universal set $\mathcal{U}$. Let B be a set. We prove parts (3) and (5), leaving the rest of the results for Exercise 4.3.7.

Proof of (3): We first show that $B \cup \bigcap_{i\in I} A_i \subseteq \bigcap_{i\in I}(B \cup A_i)$.

Let $x \in B \cup \bigcap_{i\in I} A_i$. Then $x \in B$ or $x \in \bigcap_{i\in I} A_i$.

Case I: $x \in B$.

Then for all $i \in I$, $x \in B \cup A_i$, and hence $x \in \bigcap_{i\in I}(B \cup A_i)$.

Case II: $x \notin B$.

Then $x \in \bigcap_{i\in I} A_i$. Hence, for all $i \in I$, $x \in A_i$. This implies that for all $i \in I$, $x \in B \cup A_i$. Thus $x \in \bigcap_{i\in I}(B \cup A_i)$.

Hence, in any case, $B \cup \bigcap_{i\in I} A_i \subseteq \bigcap_{i\in I}(B \cup A_i)$.

Next we show that $\bigcap_{i\in I}(B \cup A_i) \subseteq B \cup \bigcap_{i\in I} A_i$.

Let $x \in \bigcap_{i\in I}(B \cup A_i)$. Then for all $i \in I$, $x \in B \cup A_i$.

Case I: $x \in B$.

Then $x \in B \cup \bigcap_{i\in I} A_i$.

Case II: $x \notin B$.

Then, since for all $i \in I$, $x \in B \cup A_i$, it follows that for all $i \in I$, $x \in A_i$. Hence $x \in \bigcap_{i\in I} A_i$, and so $x \in B \cup \bigcap_{i\in I} A_i$.

Thus $B \cup \bigcap_{i\in I} A_i \subseteq \bigcap_{i\in I}(B \cup A_i)$, and hence the two sets are equal.

Proof of (5): We first show that $\overline{\bigcup_{i\in I} A_i} \subseteq \bigcap_{i\in I} \overline{A_i}$.

Let $x \in \overline{\bigcup_{i\in I} A_i}$. Then $x \in \mathcal{U}$ and $x \notin \bigcup_{i\in I} A_i$. By Definition 4.3.6, it follows that for all $i \in I$, $x \notin A_i$, and hence for all $i \in I$, $x \in \overline{A_i}$. Thus $x \in \bigcap_{i\in I} \overline{A_i}$ and so $\overline{\bigcup_{i\in I} A_i} \subseteq \bigcap_{i\in I} \overline{A_i}$.

Next, we show that $\bigcap_{i\in I} \overline{A_i} \subseteq \overline{\bigcup_{i\in I} A_i}$.

Let $x \in \bigcap_{i\in I} \overline{A_i}$. Then for all $i \in I$, $x \in \overline{A_i}$; i.e., $x \in \mathcal{U}$ and for all $i \in I$, $x \notin A_i$. Hence, by Definition 4.3.6, $x \notin \bigcup_{i\in I} A_i$, and so $x \in \overline{\bigcup_{i\in I} A_i}$. Thus $\bigcap_{i\in I} \overline{A_i} \subseteq \overline{\bigcup_{i\in I} A_i}$, and so the two sets are equal. □

Exercises 4.3

1. For $i \in \mathbb{Z}^+$, let $A_i = (-i, i)$.
   (a) Find $\displaystyle\bigcup_{i=1}^{\infty} A_i$ and $\displaystyle\bigcap_{i=1}^{\infty} A_i$.
   (b) Prove your answers to part (a) are correct.
2. For $i \in \mathbb{Z}^+$ with $i \geq 2$, let $A_i = \left[\frac{1}{i}, i\right)$.
   (a) Find $\displaystyle\bigcup_{i=2}^{\infty} A_i$ and $\displaystyle\bigcap_{i=2}^{\infty} A_i$.
   (b) Prove your answers to part (a) are correct.
3. For $i \in \mathbb{Z}^+$ with $i \geq 2$, let $A_i = \left(\frac{1}{i}, i\right]$.
   (a) Find $\displaystyle\bigcup_{i=2}^{\infty} A_i$ and $\displaystyle\bigcap_{i=2}^{\infty} A_i$.
   (b) Prove your answers to part (a) are correct.
4. For $i \in \mathbb{Z}^+$ with $i \geq 1$, let $A_i = \left[0, 1 - \frac{1}{i}\right]$.
   (a) Find $\displaystyle\bigcup_{i\in\mathbb{Z}^+} A_i$ and $\displaystyle\bigcap_{i\in\mathbb{Z}^+} A_i$.
   (b) Prove your answers to part (a) are correct.
5. For $i \in \mathbb{Z}^+$ let $A_i = \left[1 - \frac{1}{i}, 3 - \frac{1}{i}\right)$.
   (a) Find $\displaystyle\bigcup_{i\in\mathbb{Z}^+} A_i$ and $\displaystyle\bigcap_{i\in\mathbb{Z}^+} A_i$.

   (b) Prove your answers to part (a) are correct.

6. Finish the proof of Theorem 4.3.2. Let $n \in \mathbb{Z}^+$. Prove by induction on n that for all sets $A, B_1, B_2, \ldots, B_n$,
   (a) $A \cup (B_1 \cap B_2 \cap \cdots \cap B_n) = (A \cup B_1) \cap (A \cup B_2) \cap \cdots \cap (A \cup B_n)$.
   (b) $\overline{B_1 \cup B_2 \cup \cdots \cup B_n} = \overline{B_1} \cap \overline{B_2} \cap \cdots \cap \overline{B_n}$.
   (c) $\overline{B_1 \cap B_2 \cap \cdots \cap B_n} = \overline{B_1} \cup \overline{B_2} \cup \cdots \cup \overline{B_n}$.

7. Complete the proof of Theorem 4.3.7.

8. Let I be a nonempty set and let $\{A_i \mid i \in I\}$ be an indexed family of sets. Prove that $\bigcap_{i \in I} A_i \subseteq \bigcup_{i \in I} A_i$.

9. Let I be a nonempty set and let $\{A_i \mid i \in I\}$ be an indexed family of sets. Let X and Y be sets.
   (a) Suppose that for all $i \in I$, $X \subseteq A_i$. Prove that $X \subseteq \bigcap_{i \in I} A_i$.
   (b) Suppose that for all $i \in I$, $A_i \subseteq X$. Prove that $\bigcup_{i \in I} A_i \subseteq X$.

10. Let $\{A_i \mid i \in \mathbb{Z}^+\}$ be an indexed family of sets.
   (a) Assume that for all $i \in \mathbb{Z}^+$, $A_i \subseteq A_{i+1}$. Prove that $\bigcap_{i \in \mathbb{Z}^+} A_i = A_1$.
   (b) Assume that for all $i \in \mathbb{Z}^+$, $A_{i+1} \subseteq A_i$. Prove that $\bigcup_{i \in \mathbb{Z}^+} A_i = A_1$.

11. (See Exercise 4.2.24.)
   (a) Prove that for all $n \in \mathbb{Z}^+$, for all sets $A_1, A_2, \ldots, A_n$, if A_i is open for all i, $1 \leq i \leq n$, then $\bigcup_{i=1}^{n} A_i$ is open.
   (b) Prove that for all $n \in \mathbb{Z}^+$, for all sets $A_1, A_2, \ldots, A_n$, if A_i is closed for all i, $1 \leq i \leq n$, then $\bigcup_{i=1}^{n} A_i$ is closed.

12. (See Exercise 4.2.24.) Let $I \neq \emptyset$ and let $\{A_i \mid i \in I\}$ be an indexed family of sets.
   (a) Prove that if A_i is open for all $i \in I$, then $\bigcup_{i \in I} A_i$ is also open.
   (b) Prove that if A_i is closed for all $i \in I$, then $\bigcap_{i \in I} A_i$ is also closed.

4.4. Axiomatic set theory

We have been dealing with sets on a very informal basis, as many mathematicians do. We have been willing to write down any object of the form $\{x \mid P(x)\}$, for any "reasonable" conditional definition $P(x)$, and call it a "set". Based on our experience so far, defining a set is a routine matter of writing down a reasonable conditional definition $P(x)$, and our approach has been consistent with the early history of set theory.

We present now a very brief history of set theory, which is based on the engaging longer account presented in Robert S. Wolf's *A Tour through Mathematical Logic* [**16**]. See also [**9**] and [**12**].

From about 1870 to 1900, German mathematicians Richard Dedekind and Georg Cantor, and others, worked to develop a theory of sets. Their motivation was their desire to convince mathematicians of the time to adopt the use of infinite sets as mathematical objects. Their theory, which is now called "naive set theory",[3] had only two basic assumptions.

Extensionality: Two sets are equal exactly when they have the same elements.

Comprehension: Any collection $\{x \mid P(x)\}$ is a set as long as the defining condition $P(x)$ is well-defined: given x, it must be clear that either x has this property or it does not.

This has been the approach we have taken, and while so far we've encountered no problems, potential problems do exist.

Consider now the following collection of sets that are not elements of themselves:

$$A = \{x \mid x \notin x\}.$$

It's reasonable to ask whether A is an element of A. Note that if $A \in A$, then A must satisfy its defining condition, and hence $A \notin A$. But if $A \notin A$, then A satisfies its defining condition, and hence $A \in A$.

Hence we've proved

$$A \in A \Leftrightarrow A \notin A. \tag{4.5}$$

However, it certainly must be the case that $A \in A$ or $A \notin A$, but not both, so statement (4.5) is false. Thus we have the contradiction that statement (4.5) is simultaneously true and false. This contradiction is called *Russell's paradox* in honor of English mathematician Bertrand Russell, who announced the contradiction in 1902 (German mathematician Ernst Zermelo also discovered this contradiction earlier and independently).

The ultimate effect of Russell's paradox was the development of *axiomatic set theory* as a "formal system". In a "first-order" formal system for set theory, the only mathematical statements that can be written down are equalities involving the variables (such as $x = y$), statements that say that one object is an element of another (such as $x \in y$), and statements built up from these "atomic" formulas using the logical connectives ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$) and quantifiers ($\forall$, $\exists$) applied to variables (only). In the intended interpretation of the formal system, variables denote sets (only). Finally, a list of axioms is given, which provides a set of rules that indicate what types of sets exist. For example, the statement

$$(\forall x)(\forall y)(\exists z)(\forall w)[w \in z \leftrightarrow (w = x \vee w = y)],$$

called the *Pairing Axiom*, states (in the intended interpretation) that for any sets x and y, the set $\{x, y\}$ exists.

Our axioms for the integers give a list of properties that the integers (and possibly other sets of "numbers"?) possess. Any first-order formal system of axioms

[3] "Naive set theory" just means a nonaxiomatic approach to set theory.

for set theory provides a list of rules that indicates what types of sets exist. In fact, there are several ways that one can axiomatize set theory; the earliest axiomatization was given by Zermelo in 1908. A common first-order axiomatization of set theory is denoted ZF, for Zermelo-Fraenkel set theory. It contains axioms that assert, among other things, the existence of $\emptyset$, finite sets, an infinite set, and any set built up from other sets using $\cap$, $\cup$, $\subseteq$, and the power set operation (handled carefully). The "subset axioms" (informally) assert that "definable" subsets of any given set exist; for example, when $P(x)$ is a formula in the language of set theory whose only free variable is x, then $\{x \in z \mid P(x)\}$ is a subset axiom that defines the particular subset of the set z whose elements satisfy $P(x)$. The subset axioms prevent sets from getting "too big", as is the problem with the Russell set A above. In ZF set theory, Russell's paradox becomes a proof that there is no "set of all sets".

Proof. Suppose for the sake of a contradiction that there exists a set X such that for all sets y, $y \in X$. By a subset axiom, $Y = \{z \in X \mid z \notin z\}$ is also a set. Since $Y \in X$, by definition we have $Y \in Y$ if and only if $Y \notin Y$. This is a contradiction, since $Y \in Y$ and $Y \notin Y$ cannot be both simultaneously true or both simultaneously false. Hence no such set X of all sets exists. □

Although the language of ZF may seem limited to you (again, the only allowed "atomic" formulas in ZF are formulas of the form $x = y$ or $x \in y$), it turns out that a great deal of mathematics can be derived in ZF. Informally, what we mean by this statement is that many mathematical concepts (such as the natural numbers, integers, rational numbers, and real numbers) can be defined in, and the properties about these concepts can be proved in, ZF. Furthermore, while it is not possible to formally prove that ZF is "free from contradictions" (this statement can be made precise), so far no contradictory statement has been found to be provable from ZF, and most mathematicians believe that ZF is free from contradictions.

Adding the *Axiom of Choice* (first used by Zermelo in 1904) to ZF creates the formal system ZFC. Informally, (one form of) the Axiom of Choice states that for every family of nonempty sets, there exists a "choice function" f such that $f(X) \in X$ for every set X in the family (i.e., f chooses an element of X for all X in the family, "all at once"). As described in [**16**], the Axiom of Choice was controversial when it was first introduced. First, it asserts the existence of something (a choice function) without indicating how to define it. Furthermore, it implies both "believable" mathematical statements (such as the statement that the Cartesian product of any family of nonempty sets is nonempty[4]) and "unbelievable" mathematical statements (such as the *Banach-Tarski paradox*, which states that "a solid sphere can be decomposed into a finite number of pieces that can be reassembled (using only 'rigid motions', translation and rotation) into two spheres of the same radius as the original!") [**16**, p. 226]. At this point in time, the Axiom of Choice (AC) is accepted by most mathematicians as just another tool of mathematical reasoning. In fact, contemporary mathematicians rarely point out when they use the Axiom of Choice.

Studying set theory from an axiomatic point of view, including the effect of including or not including the Axiom of Choice, and the fact that a great deal of

[4] Another "believeable" statement implied by the Axiom of Choice can be found in Exercise 5.4.10b.

mathematics can be derived in axiomatic set theory, is the topic of another course. Interested readers should consult [**9**] or [**12**]. What's important for us is that, with care, issues like Russell's paradox do not come up in day-to-day mathematics.

Chapter 5

Functions

5.1. Definitions

Just as with the notion of set, you have been working with functions in several previous mathematics courses. You already have an intuitive feeling about functions and how to work with them, most likely from a calculus course. In such a course, a function (defined on some subset of $\mathbb{R}$) is defined to be a correspondence which sends each real number in its domain to a unique real number. However, using the word "correspondence" to define "function" has the same problem as using the word "collection" to define "set"; it is mathematically imprecise and essentially leaves the term "function" undefined. In addition, we will want to consider functions defined on sets X other than the real numbers, for which the outputs of the function reside in some set Y.

Given arbitrary nonempty sets X and Y, one can formally define a function from X to Y to be a subset $F \subseteq X \times Y$ of ordered pairs (x, y) such that for each $x \in X$, there exists a unique $y \in Y$ with $(x, y) \in F$. Given $x \in X$, the unique $y \in Y$ such that $(x, y) \in F$ is denoted notationally by $F(x)$. Some authors choose to define the phrase "function from X to Y" in this way, which enables them to work with a formal set-theoretical definition from which to prove theorems. However, while it is easy to avoid adding another undefined term by formalizing the definition of function as a particular kind of set, doing so adds a layer of abstraction. In practice, most mathematicians work with functions informally, as they do with sets, and an informal definition of function suffices to prove the theorems of interest in many areas of mathematics. When a more formal approach is needed, one can return to the set-theoretical definition.

In this book, an informal definition of function is adequate for our purpose. All further definitions regarding functions will be precise, as will be the theorems we state and prove about functions. Thus, we begin with the definition you are accustomed to, with the exception that we will define a function to be a *triple* consisting of two sets, corresponding to the domain and “target set” of outputs of the function, and a “correspondence” between these sets.[1]

Definition 5.1.1. Let X and Y be nonempty sets. A *function f from the set X to the set Y* is a correspondence that assigns to each element $x \in X$ a unique element $y \in Y$, which is denoted by $f(x)$.

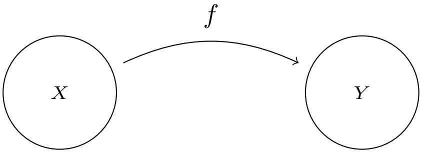
The set X is called the *domain* of f. We often denote the domain of a function f by $\operatorname{dom} f$. The set Y is called the *codomain* of f.

If $x \in X$ and $y \in Y$ are such that $y = f(x)$, then y is called the *value of f at x*, or *the image of x under f*, and x is called a *preimage* of y under f. We may also say that *f maps x to y*.

Note, then, that a function is specified by giving a domain, a codomain, and a correspondence. However, it is important to emphasize that a correspondence does not have to be specified by a “rule”, a formula, or an algorithm. Although each x in the domain of a function f is assigned to a unique element $f(x)$ of the codomain, we may not have any information regarding “how” that correspondence takes place.

Notation 5.1.2. We indicate that f is a function from domain X to codomain Y by writing $f : X \to Y$. (Note that this notation then implies that X and Y are nonempty.)

We often use the following picture to denote a function $f : X \to Y$.



Example 5.1.3. Let $X = \{1, 2, 3, 4\}$ and $Y = \{a, b, c, d, e\}$. Since X is finite, we may define a function $f : X \to Y$ by simply stating what the correspondence is for each $x \in X$:

$$f(1) = c, \qquad f(2) = e,$$
$$f(3) = e, \qquad f(4) = a.$$

Note that

- f is defined on each element of the set X; i.e., $\operatorname{dom} f = X$;
- the image of 4 under f is a, since $f(4) = a$;
- both 2 and 3 are preimages of e under f since $f(2) = f(3) = e$;
- $b \in Y$ is not the image of any element of X under f. ♢

Right away we see that there is no requirement that every element of the codomain of a function must be the image of some element of the domain. There is sometimes a difference between the codomain of a function, which you can think of

[1] Note that not all authors define a function as a triple, so always check and adhere to the definition for whatever source you are using.

as the "target set" in which the values of the function live, and the *range*, or *image* of a function, which is the set of actual values attained by the function.

Definition 5.1.4. Let X and Y be sets, and let $f : X \to Y$. The *range of* f (also called the *image of* f) is the set

$$\{y \in Y \mid (\exists x \in X)[y = f(x)]\} = \{f(x) \mid x \in X\}.$$

We denote the range (or image) of the function f by $\operatorname{ran} f$ (or $\operatorname{im} f$).

In Example 5.1.3 above, we see from our computations that $\operatorname{ran} f = \{a, c, e\}$, so that $\operatorname{ran} f$ is not equal to the codomain of f, which is the set $Y = \{a, b, c, d, e\}$.

Recall that we can define a function as a set of ordered pairs. We are used to thinking of this set as the "graph" of the function.

Definition 5.1.5. Let X and Y be sets, and let $f : X \to Y$. The *graph of* f is the set

$$\begin{aligned} G_f &= \{(x, y) \in X \times Y \mid y = f(x)\} \\ &= \{(x, f(x)) \mid x \in X\}. \end{aligned}$$

Note that we can determine a function from its domain, codomain, and graph.

In Example 5.1.3 above, $G_f = \{(1, c), (2, e), (3, e), (4, a)\}$. In that example, we gave the correspondence that defines the function f by explicitly indicating, for each element of the domain, the corresponding element in the codomain. Often functions are defined by formulas.

Example 5.1.6. Let $f : \mathbb{Z} \to \mathbb{R}$ and $g : \mathbb{Z} \to \mathbb{R}$ by, for all $n \in \mathbb{Z}$,

$$\begin{aligned} f(n) &= \cos(n\pi), \\ g(n) &= (-1)^n. \end{aligned}$$

We'll find the graphs G_f and G_g of these two functions. Note that when n is an even integer,

$$\begin{aligned} f(n) &= \cos(n\pi) = 1 \text{ and} \\ g(n) &= (-1)^n = 1, \end{aligned}$$

and when n is odd,

$$\begin{aligned} f(n) &= \cos(n\pi) = -1 \text{ and} \\ g(n) &= (-1)^n = -1. \end{aligned}$$

Thus, we see that $\operatorname{ran} f = \{-1, 1\} = \operatorname{ran} g$. Furthermore,

$$\begin{aligned} G_f &= \{(n, 1) \mid n \in \mathbb{Z} \text{ is even}\} \cup \{(n, -1) \mid n \in \mathbb{Z} \text{ is odd}\} \\ &= \{(2m, 1) \mid m \in \mathbb{Z}\} \cup \{(2m + 1, -1) \mid m \in \mathbb{Z}\} \\ &= G_g. \end{aligned}$$

Note that while the correspondences of the functions are given by different formulas, the graphs of the functions are the same, which means that the *correspondences* between the domain $\mathbb{Z}$ and the codomain $\mathbb{R}$ are the same. In other words, while $f, g : \mathbb{Z} \to \mathbb{R}$ were defined by different formulas, they are the *same function.* ◊

Definition 5.1.7 (Function equality)**.** Let A, B, C, D be sets. Let $f : A \to B$ and $g : C \to D$. Then $f = g$ if

(1) $A = C$ and $B = D$ and

(2) for all $x \in A$, $f(x) = g(x)$.

Definition 5.1.7 says that a function $f : X \to Y$ is determined by its graph, not by its rule or formula. In Example 5.1.6, $\operatorname{dom} f = \operatorname{dom} g = \mathbb{Z}$, the codomains of f and g are both $\mathbb{R}$, and for all $n \in \mathbb{Z}$, $f(n) = g(n)$. Hence $f = g$, by Definition 5.1.7.

Example 5.1.8. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$,

$$\begin{aligned} f(x) &= \sqrt{x^2}, \\ g(x) &= x. \end{aligned}$$

Note that $f \neq g$, since Definition 5.1.7(2) does not hold: we can find $x \in \mathbb{R}$ such that $f(x) \neq g(x)$.

$$\begin{aligned} f(-5) &= \sqrt{(-5)^2} = \sqrt{25} = 5, \text{ and} \\ g(-5) &= -5. \end{aligned}$$

Hence, by Definition 5.1.7, $f \neq g$. ◊

The function g in Example 5.1.8 is called the *identity function* on $\mathbb{R}$. We can define this notion more generally.

Definition 5.1.9. Let X be a set. The *identity function on X* is the function $I_X : X \to X$ defined by, for all $x \in X$, $I_X(x) = x$.

We consider several more examples below. Our next example considers a polynomial function.

Definition 5.1.10. Let $n \in \mathbb{Z}$ with $n \geq 0$, and let $a_0, a_1, \ldots, a_n \in \mathbb{R}$ such that $a_n \neq 0$. The function $p : \mathbb{R} \to \mathbb{R}$ is a *polynomial of degree n with real coefficients* $a_0, a_1, \ldots, a_n$ if for all $x \in \mathbb{R}$,

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

The function $q : \mathbb{R} \to \mathbb{R}$ such that $q(x) = 0$ for all $x \in \mathbb{R}$ is the *zero polynomial*, whose degree is undefined.

Example 5.1.11. Let $f : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$, $f(x) = x^2 + 1$.

It is important to again emphasize the definitions and proper use of notation and terminology. First note that the graph of f is

$$\begin{aligned} G_f &= \{(x, y) \in \mathbb{R}^2 \mid y = x^2 + 1\} \\ &= \{(x, x^2 + 1) \mid x \in \mathbb{R}\}. \end{aligned}$$

Here the graph of f can be illustrated by the usual graph in $\mathbb{R}^2$. (See Figure 5.1.)

In addition,

- f is the *function*;
- $f(x)$ is the *image of x under f* (note that f and $f(x)$ are not the same!);
- 3 and -3 are *preimages of* 10 *under f*, since $f(3) = 10 = f(-3)$;

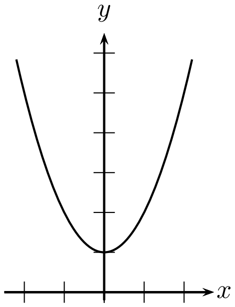


Figure 5.1. Graph of $y = x^2 + 1$.

- 0 is not the image of any real number under f, since for no $x \in \mathbb{R}$ can we have $x^2 + 1 = 0$, or $x^2 = -1$. Phrased another way, note that for all $x \in \mathbb{R}$, $x^2 + 1 \geq 0 + 1 = 1$. Hence $\operatorname{ran} f \subseteq [1, \infty)$, $0 \notin \operatorname{ran} f$, and $\operatorname{ran} f \neq \mathbb{R}$.

In fact, $\operatorname{ran} f = [1, \infty)$. To show that $[1, \infty) \subseteq \operatorname{ran} f$, we need Theorem 2.1.5 on the existence of nth roots.

Scratchwork: We know that we must begin with an arbitrary element of $[1, \infty)$. Definition 5.1.4 tells us exactly how to prove that a real number is an element of $\operatorname{ran} f$.

We know that

$$\begin{aligned}\operatorname{ran} f &= \{y \in \mathbb{R} \mid (\exists x \in \operatorname{dom} f)[y = f(x)]\} \\ &= \{y \in \mathbb{R} \mid (\exists x \in \mathbb{R})[y = x^2 + 1]\}.\end{aligned}$$

We thus have the following Given-Goal diagram.

Given	**Goal**
$y \in \mathbb{R}$ with $y \geq 1$ arbitrary	find $x \in \operatorname{dom} f$ with $y = x^2 + 1$

We work backwards. We want $y = x^2 + 1$, so we need $x^2 = y - 1$. Since $y \geq 1$, we know $y - 1 \geq 0$. Thus by Theorem 2.1.5, $\sqrt{y-1}$ exists (i.e., it is a real number) and $(\sqrt{y-1})^2 = y - 1$; i.e., $\sqrt{y-1}$ is the x we seek. We are ready to prove our claim.

Claim. $\operatorname{ran} f = [1, \infty)$.

Proof. We first show that $\operatorname{ran} f \subseteq [1, \infty)$. Let $y \in \operatorname{ran} f$. We must show $y \in [1, \infty)$; i.e., $y \geq 1$. By Definition 5.1.4, we may fix $x \in \operatorname{dom} f = \mathbb{R}$ such that $y = x^2 + 1$. Since $x^2 \geq 0$, $y = x^2 + 1 \geq 1$. Hence $\operatorname{ran} f \subseteq [1, \infty)$.

Next, we show that $[1, \infty) \subseteq \operatorname{ran} f$. Let $y \in [1, \infty)$. We must find $x \in \operatorname{dom} f = \mathbb{R}$ such that $y = f(x)$.

Consider $x = \sqrt{y-1}$, which exists by Theorem 2.1.5 since $y - 1 \geq 0$; i.e., $x \in \operatorname{dom} f = \mathbb{R}$. Then

$$\begin{aligned} f(x) &= f(\sqrt{y-1}) \\ &= (\sqrt{y-1})^2 + 1 \\ &= y - 1 + 1 = y, \end{aligned}$$

as desired. Hence $[1, \infty) \subseteq \operatorname{ran} f$. □ ◊

As we have previously emphasized, we note that you should be sure not to confuse the scratchwork, where we worked backwards to find the desired $x \in \operatorname{dom} f$ such that $y = f(x)$, and the actual proof that $[1, \infty) \subseteq \operatorname{ran} f$. The definition of $\operatorname{ran} f$ is existential, and hence we followed our usual procedure of explicitly stating the object we sought ($x = \sqrt{y-1}$) and verifying that it worked ($x \in \operatorname{dom} f$ and $y = f(x)$). The proof looks very different from the scratchwork, and in particular, the proof generally does not show how the object we sought was obtained.

The next example emphasizes again that a function is specified by giving the domain, the codomain, and the correspondence.

Example 5.1.12. Let $g : \mathbb{R} \to [1, \infty)$ by, for all $x \in \mathbb{R}$, $g(x) = x^2 + 1$. Note that by Definition 5.1.7, g is not the same function as the function f defined in Example 5.1.11. This is because while f and g have the same domain and are defined by the same formula, f and g have *different codomains.* ◊

In courses like calculus, functions are typically specified by a formula only, and the domain and codomain are taken from context.

Example 5.1.13. Let $f(x) = \frac{2x+1}{x-4}$.

We will adopt the following convention.

Convention. When the domain and codomain of a function are not given, we take the domain of the function to be the *implicit*, or *natural*, domain. The universe under consideration is taken from context, and the implicit domain is the largest subset of that universe on which the function is defined. Similarly, the codomain is taken from context.

This function is a typical function from calculus, and the codomain of any such function is $\mathbb{R}$ (which says that f is a *real-valued* function), unless we explicitly specify otherwise. The domain is a subset of $\mathbb{R}$. Here, the implicit domain is

$$\operatorname{dom} f = \{x \in \mathbb{R} \mid x \neq 4\} = (-\infty, 4) \cup (4, \infty).$$

Thus $f : (-\infty, 4) \cup (4, \infty) \to \mathbb{R}$.

We next find $\operatorname{ran} f$ and verify that our answer is correct.

Scratchwork for $\operatorname{ran} f$*:* We must find which numbers y are of the form $f(x)$, for some $x \in \operatorname{dom} f$. One way of doing this is to take advantage of the fact that f is a rational function (i.e., a function of the form $\frac{P(x)}{Q(x)}$, where $P(x)$ and $Q(x)$ are polynomials) in which the degrees of the numerator and denominator are both 1. We could use long division, but instead we'll add 0 in a clever way to the numerator of $f(x)$; our goal here is to achieve a term in the numerator which is a factor of $x-4$:

$$\frac{2x+1}{x-4}=\frac{(2x+1)-8+8}{x-4}=\frac{2x-8+9}{x-4}=\frac{2(x-4)+9}{x-4}=2+\frac{9}{x-4}.$$

Since $\frac{9}{x-4}$ is never 0, we see that $\frac{2x+1}{x-4}=2+\frac{9}{x-4}$ can never equal 2. Thus we conjecture that $\operatorname{ran} f=\{y\in\mathbb{R}\mid y\neq 2\}=(-\infty,2)\cup(2,\infty)$.

The Given-Goal diagram for showing $\{y\in\mathbb{R}\mid y\neq 2\}\subseteq\operatorname{ran} f$ is similar to the previous example.

Given	**Goal**
$y\in\mathbb{R}$ with $y\neq 2$ arbitrary	find $x\in\operatorname{dom} f$ with $y=\frac{2x+1}{x-4}$

As before, we should work backwards to find the desired x such that $y=f(x)$, and we must not forget to verify that $x\in\operatorname{dom} f$.

Claim. $\operatorname{ran} f=\{y\in\mathbb{R}\mid y\neq 2\}$.

Proof. We first show that $\operatorname{ran} f\subseteq\{y\in\mathbb{R}\mid y\neq 2\}$. Let $y\in\operatorname{ran} f$. Then we may fix $x\in\operatorname{dom} f$, i.e., $x\in\mathbb{R}$ with $x\neq 4$, such that $y=f(x)=\frac{2x+1}{x-4}$. Since

$$\frac{2x+1}{x-4}=\frac{2(x-4)+9}{x-4}=2+\frac{9}{x-4}$$

and $\frac{9}{x-4}\neq 0$, $y=2+\frac{9}{x-4}\neq 2$. Hence $\operatorname{ran} f\subseteq\{y\in\mathbb{R}\mid y\neq 2\}$.

Next we show that $\{y\in\mathbb{R}\mid y\neq 2\}\subseteq\operatorname{ran} f$. Let $y\in\mathbb{R}$ with $y\neq 2$. We must find $x\in\operatorname{dom} f$, i.e., $x\in\mathbb{R}$ with $x\neq 4$, such that $y=f(x)$. Consider $x=\frac{4y+1}{y-2}$ (found by working backwards), which is defined since $y\neq 2$. Note that $x\in\operatorname{dom} f$, i.e., $x\neq 4$, since

$$x=\frac{4y+1}{y-2}=\frac{(4y+1)-8+8}{y-2}=\frac{4(y-2)+9}{y-2}=4+\frac{9}{y-2}$$

and $\frac{9}{y-2}\neq 0$.

Next, note that $y=f(x)$, since

$$\begin{aligned}
f(x)&=f\left(\frac{4y+1}{y-2}\right)\\
&=\frac{2\left(\frac{4y+1}{y-2}\right)+1}{\left(\frac{4y+1}{y-2}\right)-4}\\
&=\frac{2\left(\frac{4y+1}{y-2}\right)+1}{\left(\frac{4y+1}{y-2}\right)-4}\cdot\frac{y-2}{y-2}\\
&=\frac{2(4y+1)+y-2}{4y+1-4(y-2)}\\
&=\frac{9y}{9}=y,
\end{aligned}$$

as desired.

Hence, $\{y \in \mathbb{R} \mid y \neq 2\} \subseteq \operatorname{ran} f$, and so

$$\operatorname{ran} f = \{y \in \mathbb{R} \mid y \neq 2\}. \qquad \square$$

Two additional remarks are in order. First, another way we could have conjectured the fact that $\operatorname{ran} f = \{y \in \mathbb{R} \mid y \neq 2\}$ is to begin with $y = f(x) = \frac{2x+1}{x-4}$ and note that solving for x requires that $y \neq 2$. Second, we could have used a proof by contradiction to show that when $x = \frac{4y+1}{y-2}$, $x \neq 4$. ♢

It is important to note that in this last example, we have been able to verify $\operatorname{ran} f$ algebraically. This is not always possible, and indeed, finding the range of an arbitrary real-valued function can be very difficult. Sometimes analytic (i.e., calculus) methods are necessary, such as the use of a theorem from calculus called the Intermediate Value Theorem (see [**11**] or any calculus textbook).

Of course, not all functions map subsets of the real numbers to the real numbers. We consider several further examples.

Example 5.1.14. Let $f : \mathbb{Z} \to \mathbb{Z}$ by, for all $n \in \mathbb{Z}$,

$$f(n) = \begin{cases} n-1 & \text{if } n \text{ is even,} \\ n+5 & \text{if } n \text{ is odd.} \end{cases}$$

This piecewise-defined function gives the value of $f(n)$ according to whether $n \in \mathbb{Z}$ is even or odd. For example, $f(-2) = -2-1 = -3$, and $f(11) = 11+5 = 16$.

What about $\operatorname{ran} f$? First, as an example, note that $-54 \in \operatorname{ran} f$. To see this, note that $n-1$ is odd when $n \in \mathbb{Z}$ is even, and $n+5$ is even when $n \in \mathbb{Z}$ is odd. Working backwards, we therefore see that, since -54 is even, we need an odd integer n such that $f(n) = n+5 = -54$. Thus, $-54 \in \operatorname{ran} f$ since $f(-59) = -59+5 = -54$. In Exercise 5.1.5, you will prove that $\operatorname{ran} f = \mathbb{Z}$. ♢

Example 5.1.15. Recall that $\mathcal{P}(\mathbb{Z})$ is the set of all subsets of $\mathbb{Z}$. Consider the function $f : \mathcal{P}(\mathbb{Z}) \to \mathcal{P}(\mathbb{Z})$ by, for all $A \in \mathcal{P}(\mathbb{Z})$, $f(A) = \overline{A}$, where $\overline{A}$ is the complement of A in $\mathbb{Z}$. Note therefore that each element of the domain of f, i.e., each input of f, is a *set* of integers and that each element of the codomain of f, and hence each output of f, is a *set* of integers. Note that

$$\begin{aligned} f(\{-3,2,17\}) &= \{n \in \mathbb{Z} \mid n \neq -3 \text{ and } n \neq 2 \text{ and } n \neq 17\}, \\ f(\{99,100,101,\ldots\}) &= f(\{n \in \mathbb{Z} \mid n \geq 99\}) \\ &= \{n \in \mathbb{Z} \mid n \leq 98\} = \{\ldots,96,97,98\}, \\ f(\{n \in \mathbb{Z} \mid n \neq 0\}) &= \{0\}, \quad \text{and} \\ f(E) &= O, \end{aligned}$$

where $E = \{n \in \mathbb{Z} \mid n \text{ is even}\}$ and $O = \{n \in \mathbb{Z} \mid n \text{ is odd}\}$.

Since $\overline{\overline{A}} = A$ for all sets $A \subseteq \mathbb{Z}$ by Theorem 4.2.6(17), we suspect that $\operatorname{ran} f = \mathcal{P}(\mathbb{Z})$. Note that $\operatorname{ran} f \subseteq \mathcal{P}(\mathbb{Z})$ by the definition of f. Each element of $\operatorname{ran} f$ must be an element of the codomain of f, which is specified here to be $\mathcal{P}(\mathbb{Z})$. Phrased another way, any output of f is a set of integers, by definition of f.

To show that $\mathcal{P}(\mathbb{Z}) \subseteq \operatorname{ran} f$, we know we must begin with an arbitrary element of $\mathcal{P}(\mathbb{Z})$; i.e., we must begin with an arbitrary set B of integers. As before, the definition of $\operatorname{ran} f$ tells us exactly how to proceed.

Given	**Goal**
$B \in \mathcal{P}(\mathbb{Z})$ arbitrary	find $X \in \mathcal{P}(\mathbb{Z})$ with $B = f(X) = \overline{X}$

Claim. $\operatorname{ran} f = \mathcal{P}(\mathbb{Z})$.

Proof. First note that since $f : \mathcal{P}(\mathbb{Z}) \to \mathcal{P}(\mathbb{Z})$, we know that $\operatorname{ran} f \subseteq \mathcal{P}(\mathbb{Z})$ by definition. Thus, we must show that $\mathcal{P}(\mathbb{Z}) \subseteq \operatorname{ran} f$.

Let $B \in \mathcal{P}(\mathbb{Z})$; i.e., $B \subseteq \mathbb{Z}$. We must find $X \subseteq \mathbb{Z}$ with $B = f(X)$. By Theorem 4.2.6(17), $f(\overline{B}) = \overline{\overline{B}} = B$. Thus $B \in \operatorname{ran} f$, and so $\mathcal{P}(\mathbb{Z}) \subseteq \operatorname{ran} f$, as desired. Hence $\operatorname{ran} f = \mathcal{P}(\mathbb{Z})$. □ ◊

Example 5.1.16. Let $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ by, for all $m, n \in \mathbb{Z}$, $f((m, n)) = m + n$. Here, the function f maps each *ordered pair* (m, n) of integers to a *single* integer. For example,

$$f((15, 7)) = 22 \qquad \text{and} \qquad f((4, -9)) = -5.$$

To simplify the notation, the value $f((m, n))$ of f at (m, n) is often denoted by $f(m, n)$ instead. In other words, we might write $f(15, 7) = 22$, rather than $f((15, 7)) = 22$.

We know that $\operatorname{ran} f \subseteq \mathbb{Z}$, since $f : \mathbb{Z}^2 \to \mathbb{Z}$ (here, we are using the notation $\mathbb{Z}^2$ as usual for the set $\mathbb{Z} \times \mathbb{Z}$). We show $\operatorname{ran} f = \mathbb{Z}$ by showing that $\mathbb{Z} \subseteq \operatorname{ran} f$. The Given-Goal diagram emphasizes what we must show:

Given	**Goal**
$k \in \mathbb{Z}$ arbitrary	find $(m, n) \in \mathbb{Z}^2$ with $k = f(m, n)$

i.e.,

Given	**Goal**
$k \in \mathbb{Z}$ arbitrary	find $m, n \in \mathbb{Z}$ with $k = m + n$

Claim. $\operatorname{ran} f = \mathbb{Z}$.

Proof. First note that since $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, $\operatorname{ran} f \subseteq \mathbb{Z}$, by definition. Thus, we must show that $\mathbb{Z} \subseteq \operatorname{ran} f$.

Let $k \in \mathbb{Z}$. We must find $m, n \in \mathbb{Z}$ with $f(m, n) = k$. Note that $f(k, 0) = k + 0 = k$. Hence $k \in \operatorname{ran} f$, and so $\mathbb{Z} \subseteq \operatorname{ran} f$, as desired. □ ◊

A function $f : X \times X \to X$ which maps each ordered pair of elements of a set X to X is often called a *binary operation* on X. Thus, the function f in Example 5.1.16 is a binary operation on $\mathbb{Z}$. Phrased more naturally, addition (+) is a binary operation on the integers.

Example 5.1.17. Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ by, for all $x, y \in \mathbb{R}$, $g(x,y) = (-y,x)$. Here, the function maps ordered pairs of real numbers to ordered pairs of real numbers. For example,

$$g(\sqrt{2}, 3.97) = (-3.97, \sqrt{2}) \qquad \text{and} \qquad g\left(-\pi, \frac{\sqrt[3]{4}}{7}\right) = \left(\frac{-\sqrt[3]{4}}{7}, -\pi\right).$$

Let's show that $\operatorname{ran} g = \mathbb{R}^2$. The Given-Goal diagram for $\mathbb{R}^2 \subseteq \operatorname{ran} g$ is given below; note our careful use of variables here, which follows our usual policy of taking care not to use a variable whose meaning in the current proof is already fixed.

Given	**Goal**
$(z,w) \in \mathbb{R}^2$ arbitrary	find $(x,y) \in \mathbb{R}^2$ with $g(x,y) = (z,w)$

Before you read the proof below, work backwards to determine the candidates for x and y.

Claim. $\operatorname{ran} g = \mathbb{R}^2$.

Proof. First note that since $g : \mathbb{R}^2 \to \mathbb{R}^2$, $\operatorname{ran} g \subseteq \mathbb{R}^2$ by definition. Thus, we must show that $\mathbb{R}^2 \subseteq \operatorname{ran} g$.

Let $(z,w) \in \mathbb{R}^2$. We must find $(x,y) \in \mathbb{R}^2$ such that $g(x,y) = (z,w)$. Consider $(x,y) = (w,-z)$; i.e., $x = w$ and $y = -z$. Then

$$\begin{aligned} g(x,y) &= g(w,-z) \\ &= (-(-z), w) \\ &= (z,w), \end{aligned}$$

as desired. Hence $\mathbb{R}^2 \subseteq \operatorname{ran} g$. □ ◊

Exercises 5.1

1. For each of the following functions determine the domain and range of f. Prove that your answer for $\operatorname{ran} f$ is correct.
   (a) $f(x) = 7 - 2x$.
   (b) $f(x) = \dfrac{3x-2}{2x+1}$.
   (c) $f(x) = \dfrac{4x-2}{3x+1}$.
   (d) $f(x) = x^2 + 4x + 1$. (**HINT:** Complete the square to help you find a conjecture for $\operatorname{ran} f$.)

(e) $f(x) = \dfrac{1}{1+x^2}$. (**HINT:** Note that $f(x) > 0$ for all x (why?), and work backwards or analyze the form of $f(x)$ to find an additional restriction on the values of $f(x)$.)
(f) $f(x) = 4 - \sqrt{1-x}$.

2. Let $f : \mathbb{R}^2 \to \mathbb{R}$ by, for all $x, y \in \mathbb{R}$, $f(x,y) = y$. (Note that f is a *projection* function; it projects all inputs $(x,y) \in \mathbb{R}^2$ onto their second coordinate.) Prove that $\operatorname{ran} f = \mathbb{R}$.
3. Let $f(x,y) = (2y, \frac{1}{x})$. What is the implied domain of f; i.e., what is the largest subset of $\mathbb{R}^2$ on which f is defined? What is the most natural codomain of f? Find $\operatorname{ran} f$ and prove that your answer is correct.
4. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ by, for all $(a_1, \ldots, a_n) \in \mathbb{R}^n$, $f(a_1, \ldots, a_n) = (-a_n, \ldots, -a_1)$. Prove that $\operatorname{ran} f = \mathbb{R}^n$.
5. Let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by, for all $n \in \mathbb{Z}$,
$$f(n) = \begin{cases} n-1 & \text{if } n \text{ is even,} \\ n+5 & \text{if } n \text{ is odd.} \end{cases}$$
Prove that $\operatorname{ran} f = \mathbb{Z}$.
6. This problem assumes the Fundamental Theorem of Arithmetic; see Theorem 2.3.3.

   Let $f : \mathbb{Z}^+ \to \mathbb{Z} \times \mathbb{Z}$ by, for all $n \in \mathbb{Z}^+$, $f(n) = (a,b)$, where a and b are the unique integers such that $n = 2^a \cdot b$, with b odd.
   (a) Find $f(1)$, $f(32)$, $f(100)$, and $f(112)$.
   (b) Find $n, m \in \mathbb{Z}^+$ such that $f(n) = (5,3)$ and $f(m) = (1,1)$.
7. This problem requires calculus. Let
$$\mathsf{P} = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \mid n \geq 0, a_0, \ldots, a_n \in \mathbb{R}\}$$
be the set of all polynomials with real coefficients. Let $F : \mathsf{P} \to \mathsf{P}$ by, for all $p \in \mathsf{P}$, $F(p) = p'$, where p' is the derivative of p.
   (a) Find $F(3x^5 - \frac{7}{3}x^2 + x)$.
   (b) Find $p \in \mathsf{P}$ such that $F(p) = \frac{1}{2}x^3 + 5x^2 - 4$. Is p unique?
   (c) Find $\operatorname{ran} F$ and prove your answer is correct.
8. Are the functions $f(x) = \dfrac{16 - x^2}{x+4}$ and $g(x) = 4 - x$ equal? Why or why not?
9. Let $f : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}$. Say that f is *increasing on D* if for all $x, y \in D$,
$$x < y \implies f(x) < f(y).$$
Similarly, f is *decreasing on D* if for all $x, y \in D$,
$$x < y \implies f(x) > f(y).$$
   (a) Show that $f(x) = 5 - 2x$ is decreasing on $\mathbb{R}$.
   (b) Show that $f(x) = x^2$ is increasing on $[0, \infty)$.
   (c) Show that $f(x) = x^2$ is decreasing on $(-\infty, 0]$.
   (d) Show that $f(x) = x^3$ is increasing on $\mathbb{R}$. (**HINT:** See Exercise 2.1.8.)
10. A function $F : \mathbb{R} \to \mathbb{R}$ is a
   (a) *rigid motion* if for all $x, y \in \mathbb{R}$ with $x \neq y$, $|x - y| = |F(x) - F(y)|$;

(b) *translation* if there exists $b \in \mathbb{R}$ such that for all $x \in \mathbb{R}$, $F(x) = x + b$ (in this case, we say that F is the "translation by b" and denote F by the notation T_b);

(c) *reflection* if there exists $a \in \mathbb{R}$ such that for all $x \in \mathbb{R}$, $F(x) = 2a - x$ (in this case, we say that F is the "reflection through the point a" and denote F by R_a).

Prove that F is a rigid motion if and only if F is a translation or a reflection.

11. The function $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a *metric* if for all $x, y, z \in \mathbb{R}$,
    - $d(x, y) \geq 0$;
    - $d(x, y) = 0$ if and only if $x = y$;
    - $d(x, y) = d(y, x)$;
    - $d(x, y) \leq d(x, z) + d(z, y)$.

    Prove that the following functions are metrics.

    (a) $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by, for all $x, y \in \mathbb{R}$, $d(x, y) = |x - y|$.

    (b) $d_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by, for all $x, y \in \mathbb{R}$, $d_1(x, y) = \min\{1, d(x, y)\}$, where d is the metric in part (a) and $\min\{a, b\}$ denotes the least element in $\{a, b\}$, when $a, b \in \mathbb{R}$.

    (c) $d_2 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by, for all $x, y \in \mathbb{R}$, $d_2(x, y) = \frac{d(x,y)}{1+d(x,y)}$, where d is the metric in part (a). Also prove that d_2 is a *bounded* metric; i.e., there exists $M \in \mathbb{R}$ such that for all $x, y \in \mathbb{R}$, $d_2(x, y) \leq M$.

12. Let $*$ be a binary operation on $\mathbb{Z}$ (i.e., $* : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$) defined by, for all $a, b \in \mathbb{Z}$, $*(a, b) = a + b - 31$. For convenience, we write $a * b = a + b - 31$ instead of $*(a, b) = a + b - 31$.

    (a) Prove that for all $a, b, c \in \mathbb{Z}$, $a * (b * c) = (a * b) * c$ (i.e., $*$ is associative).

    (b) Prove that there exists a unique $e \in \mathbb{Z}$ such that for all $a \in \mathbb{Z}$, $e * a = a = a * e$.

    (c) For the integer e in part (b), prove that for all $a \in \mathbb{Z}$, there exists $b \in \mathbb{Z}$ such that $a * b = e = b * a$.

    This problem proves that $\mathbb{Z}$, together with the binary operation $*$, forms a *group* (see Table 6.2).

5.2. Function composition

Function composition is a way of constructing new functions from "old" ones.

Definition 5.2.1. Let A, B, C, and D be sets. Let $f : A \to B$ and $g : C \to D$, with $\operatorname{ran} f \subseteq C$. The *composite* (*composition*) *of* f *and* g is the function $g \circ f : A \to D$ defined by, for all $x \in A$, $(g \circ f)(x) = g(f(x))$.

When $f : A \to B$ and $g : B \to C$, then we have the following picture for the function $g \circ f : A \to C$.

$$\begin{tikzcd} A \arrow[r, bend left, "f"] \arrow[rr, bend right, "g \circ f"'] & B \arrow[r, bend left, "g"] & C \end{tikzcd}$$

Example 5.2.2. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$,

$$\begin{aligned} f(x) &= x^2, \\ g(x) &= x + 1. \end{aligned}$$

Then $f \circ g : \mathbb{R} \to \mathbb{R}$ is defined by, for all $x \in \mathbb{R}$,

$$(f \circ g)(x) = f(g(x)) = f(x+1) = (x+1)^2,$$

and $g \circ f : \mathbb{R} \to \mathbb{R}$ is defined by, for all $x \in \mathbb{R}$,

$$(g \circ f)(x) = g(f(x)) = g(x^2) = x^2 + 1.$$

We suspect that $f \circ g \neq g \circ f$, but recall that functions defined by different formulas might nevertheless be equal, i.e., the same function. Thus, we must use Definition 5.1.7 to prove $f \circ g \neq g \circ f$. Since $f \circ g$ and $g \circ f$ have the same domains and codomains, we must show that

$$(\exists x \in \mathbb{R})[(f \circ g)(x) \neq (g \circ f)(x)].$$

Thus, we see that $f \circ g \neq g \circ f$ since

$$\begin{aligned} &(f \circ g)(1) = 4 \quad \text{and} \\ &(g \circ f)(1) = 2. \end{aligned}$$

♢

Example 5.2.2 illustrates the following important fact.

$$\boxed{\text{Usually } f \circ g \neq g \circ f \text{, even when both compositions are defined.}}$$

Example 5.2.3. Consider the sets $X = \{1, 2, 3, 4\}$, $Y = \{a, b, c, d, e\}$, and $Z = \{0, 5, 10, 15, 20\}$. Define functions $f : X \to Y$ and $g : Y \to Z$ by

$$f(1) = c, \qquad f(2) = e, \qquad f(3) = e, \qquad f(4) = a,$$
$$g(a) = 10, \quad g(b) = 0, \quad g(c) = 5, \quad g(d) = 20, \quad g(e) = 15.$$

Then $g \circ f : X \to Z$, and $(g \circ f)(1) = g(f(1)) = g(c) = 5$. Similarly,

$$(g \circ f)(2) = 15, \qquad (g \circ f)(3) = 15, \qquad (g \circ f)(4) = 10.$$

On the other hand, the composite function $f \circ g$ is not defined, since the range of g is not a subset of the domain of f. ♢

Example 5.2.4. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$,

$$f(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ x - 2 & \text{if } x < 0; \end{cases}$$

$$g(x) = \begin{cases} x + 3 & \text{if } x \geq 4, \\ 2x & \text{if } x < 4. \end{cases}$$

We find $g \circ f$ and leave $f \circ g$ for Exercise 5.2.2. Given $x \in \mathbb{R}$,

$$(g \circ f)(x) = \begin{cases} g(x^2) & \text{if } x \geq 0, \\ g(x-2) & \text{if } x < 0. \end{cases}$$

To compute $g(x^2)$ when $x \geq 0$, the definition of g tells us that we must consider whether $x^2 \geq 4$ or $x^2 < 4$. Since $x \geq 0$, $x^2 \geq 4$ when $x \geq 2$, and $x^2 < 4$ when $0 \leq x < 2$. Thus, $g(x^2) = x^2 + 3$ when $x \geq 2$, and $g(x^2) = 2x^2$ when $0 \leq x < 2$. Similarly, to compute $g(x-2)$ when $x < 0$, we must consider whether $x - 2 \geq 4$ or $x - 2 < 4$. When $x < 0$, $x - 2 < -2 < 4$, so $g(x-1) = 2(x-2)$. Thus, we have

$$(g \circ f)(x) = \begin{cases} x^2 + 3 & \text{if } x \geq 2, \\ 2x^2 & \text{if } 0 \leq x < 2, \\ 2x - 4 & \text{if } x < 0. \end{cases} \qquad \Diamond$$

We end this section by proving two important facts about function composition. Recall that I_X denotes the identity function on the set X.

Proposition 5.2.5. *Let X, Y, Z, and W be sets. Let $f : X \to Y$, $g : Y \to Z$, and $h : Z \to W$. Then*

(1) *$(h \circ g) \circ f = h \circ (g \circ f)$; i.e., function composition is* associative*, and*

(2) *$f \circ I_X = f = I_Y \circ f$.*

A picture illustrating the functions is given below.

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[rr, bend left, "g \circ f"] & Y \arrow[r, "g"] \arrow[rr, bend right, "h \circ g"'] & Z \arrow[r, "h"] & W
\end{tikzcd}$$

Proof. Let $f : X \to Y$, $g : Y \to Z$, and $h : Z \to W$.

(1) We use Definition 5.1.7 to show that $(h \circ g) \circ f = h \circ (g \circ f)$. First note that $(h \circ g) \circ f, h \circ (g \circ f) : X \to W$. Next, we must show that for all $x \in X$, $((h \circ g) \circ f)(x) = (h \circ (g \circ f))(x)$.

Let $x \in X$. Then

$$\begin{aligned} ((h \circ g) \circ f)(x) &= (h \circ g)(f(x)) = h(g(f(x))), \quad \text{and} \\ (h \circ (g \circ f))(x) &= h((g \circ f)(x)) = h(g(f(x))). \end{aligned}$$

Hence $(h \circ g) \circ f = h \circ (g \circ f)$ by Definition 5.1.7.

(2) We show that $f \circ I_X = f$ and leave the proof that $I_Y \circ f = f$ as Exercise 5.2.3. First note that $I_X : X \to X$ and $f : X \to Y$, so $f \circ I_X : X \to Y$. Next let $x \in X$. Then

$$(f \circ I_X)(x) = f(I_X(x)) = f(x),$$

by definition of the identity function I_X. Hence $f \circ I_X = f$, as desired, by Definition 5.1.7. $\square$

Exercises 5.2

1. Find $f \circ g$ and $g \circ f$ for each pair of functions f and g.
   (a) $f, g : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2 - 3x$ and $g(x) = 5x - 2$.
   (b) $f, g : \mathbb{Z} \to \mathbb{Z}$ by $f(n) = 2n + 3$ and

$$g(n) = \begin{cases} 2n - 1 & \text{if } n \text{ is even,} \\ n + 1 & \text{if } n \text{ is odd.} \end{cases}$$

2. Let $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$,

$$f(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ x - 2 & \text{if } x < 0, \end{cases}$$

$$g(x) = \begin{cases} x + 3 & \text{if } x \geq 4, \\ 2x & \text{if } x < 4. \end{cases}$$

   Find $f \circ g$.

3. Complete the proof of Proposition 5.2.5(2). Let X and Y be sets, and let $f : X \to Y$. Prove that $I_Y \circ f = f$.

4. See Exercise 5.1.10.
   (a) Prove that for all $a, b \in \mathbb{R}$, $R_a \circ R_b$ is either the identity function $I_{\mathbb{R}}$ or a translation T_c for some $c \in \mathbb{R}$. When $R_a \circ R_b \neq I_{\mathbb{R}}$, find a formula for c in terms of a and b, and indicate the relationship between $R_a \circ R_b$ and $R_b \circ R_a$.
   (b) Prove that for all $a, b \in \mathbb{R}$, $T_a \circ T_b$ is also a translation T_c for some $c \in \mathbb{R}$. Find a formula for c in terms of a and b, and indicate the relationship between $T_a \circ T_b$ and $T_b \circ T_a$.
   (c) Prove that for all $a, b \in \mathbb{R}$, $R_a \circ T_b$ (respectively, $T_b \circ R_a$) is a reflection R_c (respectively, R_d) for some $c \in \mathbb{R}$ (respectively, $d \in \mathbb{R}$). Find formulas for c and d in terms of a and b.

5.3. One-to-one and onto functions

Recall that the definition of $f : X \to Y$ states that each $x \in X$ is mapped via f to a unique output $f(x)$ in Y. Note that the definition of the word function does not imply that every element of the codomain has a unique preimage, or indeed any preimage at all. We have already seen in Example 5.1.11 an example of a function $f : \mathbb{R} \to \mathbb{R}$, namely $f(x) = x^2 + 1$, where it's possible for an element of the codomain to have more than one preimage (here, $f(-2) = 5 = f(2)$) and for an element of the codomain to have no preimages at all (here, $0 \notin \operatorname{ran} f$). Functions that *do* possess the properties that every element of the codomain has a unique preimage are important in mathematics.

Definition 5.3.1. Let X, Y be sets, and let $f : X \to Y$.

(1) The function f is *one-to-one* (1-1) if

$$(\forall x_1, x_2 \in X)[x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)] \tag{5.1}$$

or, equivalently,

$$(\forall x_1, x_2 \in X)[f(x_1) = f(x_2) \Rightarrow x_1 = x_2]. \tag{5.2}$$

We may also say that f is *injective*, or is *an injection*, and write $f : X \overset{\text{1-1}}{\to} Y$.

(2) The function f is *onto* if

$$(\forall y \in Y)(\exists x \in X)[y = f(x)]. \tag{5.3}$$

We may also say that f is *surjective*, or is *a surjection*, and write $f : X \underset{\text{onto}}{\to} Y$. (Note that a function $f : X \to Y$ is onto iff $\operatorname{ran} f = Y$.)

(3) The function f is *bijective*, or is a *bijection* (or a 1-1 *correspondence*), if f is both an injection and a surjection, i.e., f is both 1-1 and onto, and we write $f : X \overset{\text{1-1}}{\underset{\text{onto}}{\to}} Y$.

We rephrase the example at the beginning of this section in terms of this new language. Before you read the example below, first find useful denials of statements (5.1) and (5.3), in order to find the definitions of the statements "f is not 1-1" and "f is not onto".

Example 5.3.2. Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2 + 1$ for all $x \in \mathbb{R}$. Then f is *not* 1-1 because $f(2) = 5 = f(-2)$. Since $f(x) \geq 1$ for all $x \in \mathbb{R}$, we know that $0 \notin \operatorname{ran} f$, and hence f is *not onto.* ♢

Note that statements (5.1) and (5.2) give us two ways to show that a function is 1-1 (in fact, there are other ways, as well; see Exercise 5.3.8). Often (although not always) it is the definition provided by (5.2) that is the most useful.

Example 5.3.3. Let $a, b \in \mathbb{R}$ with $a \neq 0$. Let $f : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$, $f(x) = ax + b$. We show that f is a bijection.

As remarked above, we use (5.2) to show that f is 1-1. It is worthwhile writing down the general Given-Goal diagram. We will also need to keep track of and use the additional hypothesis that $a \neq 0$.

Given	**Goal**
$a \neq 0$	
$x_1, x_2 \in \mathbb{R}$ arbitrary	
$f(x_1) = f(x_2)$	$x_1 = x_2$

Proof that f is 1-1. Let $x_1, x_2 \in \mathbb{R}$ and assume that $f(x_1) = f(x_2)$. We must show that $x_1 = x_2$.

Since $f(x_1) = f(x_2)$, we know that $ax_1 + b = ax_2 + b$. Then $ax_1 = ax_2$, and since $a \neq 0$, we may divide both sides by a to obtain $x_1 = x_2$, as desired. Thus, f is 1-1, by definition. □

To show that f is onto, we use (5.3). Again, we begin with the general Given-Goal diagram.

Given	**Goal**
$y \in \mathbb{R}$ arbitrary	$(\exists x \in \mathbb{R})[y = f(x)]$

Proof that f is onto. Let $y \in \mathbb{R}$ be given. We must find $x \in \mathbb{R}$ such that $y = f(x)$. Consider $x = \frac{y-b}{a}$, which is a real number since $a \neq 0$ (and which we found in the usual way by working backwards). Then

$$\begin{aligned} f(x) &= f\left(\frac{y-b}{a}\right) \\ &= a\left(\frac{y-b}{a}\right) + b \\ &= (y-b) + b \\ &= y, \end{aligned}$$

as desired. Hence f is onto, by definition. □ ◊

The functions given in Example 5.2.3, which mapped finite sets to finite sets, were given explicitly, so that one can determine whether the functions are 1-1 or onto by observation.

Example 5.3.4. Consider the sets $X = \{1, 2, 3, 4\}$, $Y = \{a, b, c, d, e\}$, and $Z = \{0, 5, 10, 15, 20\}$. Define functions $f : X \to Y$ and $g : Y \to Z$ by

$$\begin{gathered} f(1) = c, \qquad f(2) = e, \qquad f(3) = e, \qquad f(4) = a, \\ g(a) = 10, \quad g(b) = 0, \quad g(c) = 5, \quad g(d) = 20, \quad g(e) = 15. \end{gathered}$$

Then f is not 1-1, since $f(2) = e = f(3)$. Similarly, f is not onto, since we can see that, for example, $b \notin \operatorname{ran} f$.

On the other hand, g is a bijection; we can see from its definition that distinct elements of Y are mapped to distinct elements of Z, and $\operatorname{ran} g = Z$.

We will see in Chapter 8 that, since X and Y are finite sets of different sizes, no function from one of these sets to the other can be a bijection. On the other hand, since Y and Z are finite sets of the same size, any function from one of these sets to the other is 1-1 if and only if it is onto. See Exercise 8.2.13. ◊

Example 5.3.5. We showed above that the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2+1$ for all $x \in \mathbb{R}$ is not onto, since $f(x) \geq 1$ for all $x \in \mathbb{R}$. We now show that $g : \mathbb{R} \to [1, \infty)$ by $g(x) = x^2 + 1$ is an onto function.

Proof. Let $y \in [1, \infty)$ be arbitrary. We must find $x \in \mathbb{R}$ such that $y = g(x)$; i.e., $y = x^2 + 1$. Consider $x = \sqrt{y-1}$, which makes sense since $y - 1 \geq 0$ (and which we found in the usual way by working backwards). Then

$$\begin{aligned} g(x) &= g(\sqrt{y-1}) \\ &= (\sqrt{y-1})^2+1 \\ &= (y-1)+1 \\ &= y, \end{aligned}$$

as desired. Hence g is onto. □ ◊

Example 5.3.5 illustrates that the codomain of a function must be specified in order to determine whether or not the function is "onto". Recall also that we showed in Example 5.1.11 that $\operatorname{ran} f = [1,\infty)$; the same steps that established $[1,\infty) \subseteq \operatorname{ran} f$ also show that the function g in Example 5.3.5 is onto. The function g shows that (in an abuse of language) a function always maps its domain onto its range. More precisely we have the following theorem, whose proof we leave for Exercise 5.3.5.

Theorem 5.3.6. *Let $f : X \to Y$ and let $\operatorname{ran} f$ be the range of f. Then the function $g : X \to \operatorname{ran} f$ by $g(x) = f(x)$ for all $x \in X$ is onto.*

Note also in Example 5.3.5 that the function g is not 1-1 for the same reason that the function f isn't (for example, $g(2) = 5 = g(-2)$). However, by "restricting the domain" of g, we can obtain a 1-1 function.

Example 5.3.7. Let $h : [0,\infty) \to [1,\infty)$ by $h(x) = x^2+1$, for all $x \in [0,\infty)$. Then h is a bijection.

Proof. To see that h is 1-1, we let $x_1, x_2 \in \mathbb{R}$ with $x_1, x_2 \geq 0$ and assume that $h(x_1) = h(x_2)$. We must show that $x_1 = x_2$.

Since $h(x_1) = h(x_2)$, we have

$$\begin{aligned} (x_1)^2+1 &= (x_2)^2+1, && \text{so} \\ (x_1)^2 &= (x_2)^2, && \text{and hence} \\ \sqrt{(x_1)^2} &= \sqrt{(x_2)^2}. \end{aligned}$$

By Theorem 2.1.5, we obtain $|x_1| = |x_2|$, and since $x_1, x_2 \geq 0$, we have $x_1 = x_2$, as desired. Hence, h is 1-1.

The proof that h is onto is exactly the proof given above that g is onto, except that *we must verify that the candidate* $x = \sqrt{y-1}$ from Example 5.3.5 *is in the domain of* h. Since $\sqrt{y-1} \geq 0$, $x \in \operatorname{dom} h = [0,\infty)$, and hence h is onto.

Since h is both 1-1 and onto, h is a bijection. □ ◊

We leave for Exercise 5.3.6 the verification that we can restrict the domain of the function f above in a different way to yield a different 1-1 function with the same formula. Namely, the function $k : (-\infty,0] \to [1,\infty)$ defined by $k(x) = x^2+1$ for all $x \in (-\infty,0]$ is also a bijection. These examples demonstrate a kind of analogue to Theorem 5.3.6 for the notion of 1-1-ness.

Theorem 5.3.8. *Let $f : X \to Y$ be onto. Then there exists a subset $X_0 \subseteq X$ such that restricting the domain of f to X_0 yields a* 1-1 *and onto function; i.e., the function $g : X_0 \to Y$ defined by $g(x) = f(x)$ for all $x \in X_0$ is a bijection.*

The proof of Theorem 5.3.8 involves the Axiom of Choice (see Section 4.4), since for each element of the codomain Y we must choose an element $x \in X$ such that $f(x) = y$. We therefore omit this proof.

As we noted in Section 5.1, it is not always easy, or even possible, to show that a function $f : X \to Y$ is onto using algebraic methods. For example, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3 - x$ for all $x \in \mathbb{R}$. Showing f is onto algebraically amounts to solving the equation $x^3 - x = y$ for x, regardless of $y \in \mathbb{R}$, which is not easy. The easiest way to show f is onto is by using the Intermediate Value Theorem from calculus.

We consider one final example.

Example 5.3.9. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ by, for all $x, y \in \mathbb{R}$, $f(x, y) = (-y, x)$. We show that f is a bijection.

Scratchwork: As before, we will use statements (5.2) and (5.3) to prove that the function f is 1-1 and onto. However, since an arbitrary element of $\mathbb{R}^2$ is an ordered pair, our Given-Goal diagrams must reflect this.

For showing f is 1-1:

Given	**Goal**
$(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ arbitrary	$(x_1, y_1) = (x_2, y_2)$
$f(x_1, y_1) = f(x_2, y_2)$	i.e., $x_1 = x_2$ and $y_1 = y_2$

For showing f is onto:

Given	**Goal**
$(z, w) \in \mathbb{R}^2$ arbitrary	$(\exists (x, y) \in \mathbb{R}^2)[(z, w) = f(x, y)]$

Note that this is the same Given-Goal diagram as the one we used in Example 5.1.17 to show that the range of this function is all of $\mathbb{R}^2$; i.e., we've already done the work that shows that f is onto.

Proof. We showed that f is onto in Example 5.1.17. We now show that f is 1-1. Let $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ and assume that $f(x_1, y_1) = f(x_2, y_2)$. We must show that $(x_1, y_1) = (x_2, y_2)$; i.e., we must show that $x_1 = x_2$ and $y_1 = y_2$.

Since $f(x_1, y_1) = f(x_2, y_2)$, we know that $(-y_1, x_1) = (-y_2, x_2)$. By definition of equality of ordered pairs (see page 78), we have that $-y_1 = -y_2$, and hence $y_1 = y_2$, and also $x_1 = x_2$. Thus f is 1-1.

Since f is 1-1 and onto, f is a bijection. □ ♢

We end this section by considering various important results about function composition and 1-1 or onto functions.

Theorem 5.3.10. *Let X, Y, Z be sets. Let $f : X \to Y$ and $g : Y \to Z$.*

(1) *If f and g are both 1-1, then $g \circ f$ is 1-1.*

(2) *If f and g are both onto, then $g \circ f$ is onto.*

(3) *If f and g are both bijections, then $g \circ f$ is a bijection.*

(4) *If $g \circ f$ is 1-1, then f is 1-1, but g need not be.*

(5) *If $g \circ f$ is onto, then g is onto, but f need not be.*

Proof. Let $f : X \to Y$ and $g : Y \to Z$. Recall that $g \circ f : X \to Z$.

Note that (3) follows immediately from (1) and (2). We prove (1) and (5) and leave (2) and (4) for Exercise 5.3.7. For each of these statements, remember that

$$\boxed{\textit{it is the goal that determines how the proof should proceed.}}$$

Proof of (1): Assume that f and g are both 1-1. We show that $g \circ f$ is 1-1.

Let $x_1, x_2 \in X$ and assume that $(g \circ f)(x_1) = (g \circ f)(x_2)$. We must show that $x_1 = x_2$. Since $(g \circ f)(x_1) = (g \circ f)(x_2)$, we know that

$$g(f(x_1)) = g(f(x_2)).$$

Since g is 1-1, it follows that

$$f(x_1) = f(x_2).$$

Finally, since f is 1-1, it follows that $x_1 = x_2$ as desired. Hence $g \circ f$ is 1-1.

Proof of (5): Assume that $g \circ f$ is onto. We must show that g is onto.

Since $g : Y \to Z$, statement (5.3) tells us to begin by letting $z \in Z$ be arbitrary. We must find $y \in Y$ such that $z = g(y)$.

We know that $g \circ f : X \to Z$ is onto and $z \in Z$, so we may fix $x \in X$ such that $z = (g \circ f)(x)$. But then $z = g(f(x))$, and so $y = f(x)$ has the property that $z = g(y)$. Note that $y = f(x) \in Y$ since $f : X \to Y$. Hence g is onto, as desired.

Next, we must provide a counterexample which shows that when $g \circ f$ is onto, f need not be onto. Since we are constructing a counterexample, we must provide *specific* functions $f : X \to Y$ and $g : Y \to Z$ with this property. Rather than try to work with formulas of familiar functions, the easiest thing to do is to work with functions defined on finite sets.

Let $X = Y = \{1, 2\}$, and let $Z = \{1\}$. Define $f : X \to Y$ by

$$f(1) = f(2) = 1,$$

and define $g : Y \to Z$ by

$$g(1) = g(2) = 1.$$

Then $(g \circ f) : X \to Z$ is onto, since $\operatorname{ran}(g \circ f) = Z = \{1\}$. However, f is not onto, since $2 \in Y$, but $2 \notin \operatorname{ran} f$. □

Exercises 5.3

1. For each function f,
   (i) determine whether f is 1-1;
   (ii) determine whether f is onto.
   Prove your answers.
   (a) $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x + |x|$.
   (b) $f : \{x \in \mathbb{R} \mid x \neq \frac{3}{5}\} \to \{y \in \mathbb{R} \mid y \neq \frac{2}{5}\}$ by $f(x) = \frac{2x+1}{5x-3}$.
   (c) $f : \{x \in \mathbb{R} \mid x \neq -\frac{d}{c}\} \to \{y \in \mathbb{R} \mid y \neq \frac{a}{c}\}$ by $f(x) = \frac{ax+b}{cx+d}$, where $a, b, c, d \in \mathbb{R}$ have the property that $ad - bc \neq 0$ and $c \neq 0$.
   (d) $f : (-\infty, 3] \to [2, \infty)$ by $f(x) = (x-3)^2 + 2$.
   (e) $f : (-\infty, 1] \to (-\infty, 4]$ by $f(x) = 4 - \sqrt{1 - x^3}$.
   (f) $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x + y$.
   (g) $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = (x - y)^3$.
   (h) $f : \mathbb{R} \to \mathbb{R}^2$ by $f(x) = (x, x)$.
   (i) $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x, y) = (x + y, x - y)$.
   (j) $f : \mathbb{R}^2 \to \mathbb{R}^3$ by $f(x, y) = (x + y, x - y, xy)$.
   (k) $f : \mathbb{R}^3 \to \mathbb{R}^3$ by $f(x, y, z) = (x + y, y + z, x + z)$.
   (l) $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f(x, y, z) = (x + y, y + z)$.
   (m) $f : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+$ by $f(m, n) = 2^{m-1}(2n - 1)$. (**HINT:** You will need the Fundamental Theorem of Arithmetic (Theorem 2.3.3).)
   (n) $F : \mathsf{P} \to \mathsf{P}$ by $F(p) = p'$. (See Exercise 5.1.7.)
   (o) $f : \mathcal{C} \to \mathbb{Z}$, where $\mathcal{C} = \{A \in \mathcal{P}(\mathbb{Z}) \mid A \text{ is finite}\}$ and $f(A)$ is the sum of all elements of A.

2. Let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by

$$f(n) = \begin{cases} n - 1 & \text{if } n \text{ is even,} \\ n + 3 & \text{if } n \text{ is odd.} \end{cases}$$

   Prove that f is a bijection. (**HINT:** To prove that f is 1-1, let $n_1, n_2 \in \mathbb{Z}$ and assume that $f(n_1) = f(n_2)$, as usual. Then consider cases for n_1 and n_2. How many cases are there?)

3. For each of the piecewise-defined functions f,
   (i) determine whether f is 1-1;
   (ii) determine whether f is onto.
   Prove your answers.
   (a) $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ 2x & \text{if } x < 0. \end{cases}$$

   (b) $f : \mathbb{Z} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} n + 1 & \text{if } n \text{ is even,} \\ 2n & \text{if } n \text{ is odd.} \end{cases}$$

(c) $f : \mathbb{Z} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} 2n+1 & \text{if } n \text{ is even,} \\ n+3 & \text{if } n \text{ is odd.} \end{cases}$$

4. Let X be a set. Prove that the identity function $I_X : X \to X$ is a bijection.
5. Prove Theorem 5.3.6.
6. Let $k : (-\infty, 0] \to [1, \infty)$ by $k(x) = x^2 + 1$ for all $x \in (-\infty, 0]$. Prove that k is a bijection.
7. Let X, Y, Z be sets, and let $f : X \to Y$ and $g : Y \to Z$. Prove the following statements from Theorem 5.3.10.
   (a) If f and g are both onto, then $g \circ f$ is onto.
   (b) If $g \circ f$ is 1-1, then f is 1-1.
   (c) Give an example of particular functions $f : X \to Y$ and $g : Y \to Z$ with the property that $g \circ f$ is 1-1 but g is not 1-1.
8. Let $X, Y \subseteq \mathbb{R}$, and assume that $f : X \to Y$ is increasing on X (see Exercise 5.1.9). Prove that f is 1-1. Similarly, prove that $f : X \to Y$ is 1-1 when f is decreasing on X. (**HINT:** Use statement (5.1), rather than statement (5.2).)
9. Prove that every rigid motion $F : \mathbb{R} \to \mathbb{R}$ is 1-1, using the *definition* of rigid motion. (See Exercise 5.1.10.)
10. Let $C : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by, for all $(x, y) \in \mathbb{N} \times \mathbb{N}$,

$$C(x,y) = \sum_{i=1}^{x+y-1} (i-1) + y = \frac{(x+y)^2 - 3x - y + 2}{2}.$$

Show that C is a bijection. (**HINT:** Use statement (5.1), rather than statement (5.2). For $(x_1, y_1), (x_2, y_2) \in \mathbb{N}^2$, consider the cases $x_1 + y_1 = x_2 + y_2$ and $x_1 + y_1 \neq x_2 + y_2$.)

5.4. Invertible functions

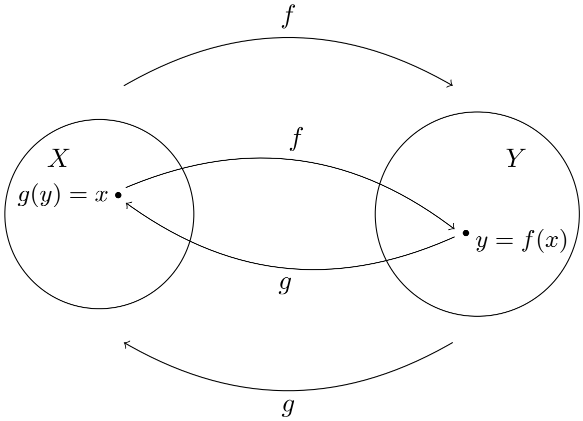
In this section, we answer the question of when it is possible to "reverse", or "undo", the action of a function. As it turns out, the answer is connected to the property of being a bijection.

Definition 5.4.1. Let X, Y be sets, and let $f : X \to Y$. We say that f is *invertible* if there exists a function $g : Y \to X$ such that for all $x \in X$ and for all $y \in Y$,

$$y = f(x) \quad \Leftrightarrow \quad x = g(y).$$

We say that such a function g is an *inverse function* of f.

The following picture illustrates Definition 5.4.1.



Note that if a function $f : X \to Y$ is invertible and $g : Y \to X$ is an inverse function of f, then Definition 5.4.1 implies that g is also invertible and that f is an inverse of g; see Exercise 5.4.3.

Example 5.4.2. Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 3x - 1$. Then f is invertible because the function $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = \frac{x+1}{3}$ satisfies $y = f(x) \Leftrightarrow x = g(y)$ for all $x, y \in \mathbb{R}$. To see this, note that given $x, y \in \mathbb{R}$,

$$\begin{aligned} y = f(x) &\Leftrightarrow y = 3x - 1 \\ &\Leftrightarrow y + 1 = 3x \\ &\Leftrightarrow x = \frac{y+1}{3} \\ &\Leftrightarrow x = g(y). \end{aligned}$$

Thus, we see that g is an inverse of f. ♢

Our first result characterizes inverse functions in terms of their composition.

Proposition 5.4.3. *Let X and Y be sets, and let $f : X \to Y$ and $g : Y \to X$. Then f is invertible and g is an inverse function of f iff $g \circ f = I_X$ and $f \circ g = I_Y$.*

Proof. Let $f : X \to Y$ and $g : Y \to X$.

($\Rightarrow$) Assume that f is invertible and that g is an inverse function of f. Then by Definition 5.4.1 we have that for all $x \in X$ and for all $y \in Y$,

$$y = f(x) \Leftrightarrow x = g(y). \tag{5.4}$$

We must show that $g \circ f = I_X$ and $f \circ g = I_Y$.

First note that $g \circ f : X \to X$. Let $x \in X$ be arbitrary, and define $y = f(x)$. Then

$$\begin{aligned} (g \circ f)(x) &= g(f(x)) \\ &= g(y) \\ &= x, && \text{by (5.4),} \\ &= I_X(x), && \text{by definition.} \end{aligned}$$

Hence $g \circ f = I_X$.

Similarly, $f \circ g = I_Y$, which we leave as an exercise (Exercise 5.4.4).

($\Leftarrow$) Assume that f and g satisfy $g \circ f = I_X$ and $f \circ g = I_Y$. To show that f is invertible and g is an inverse function of f, we must let $x \in X$ and $y \in Y$ be arbitrary and show that (5.4) is true.

First we assume that $y = f(x)$ and show that $x = g(y)$.

$$\begin{aligned} g(y) &= g(f(x)) \\ &= (g \circ f)(x) \\ &= I_X(x), && \text{by hypothesis,} \\ &= x, \end{aligned}$$

as desired. Similarly, $x = g(y) \Rightarrow y = f(x)$, which we leave as an exercise (Exercise 5.4.4). □

Note that Proposition 5.4.3 gives an alternate way to prove that two functions are inverse functions. The following corollary is immediate.

Corollary 5.4.4. *Let X, Y be sets and let $f : X \to Y$. Then f is invertible iff there exists a function $g : Y \to X$ such that $g \circ f = I_X$ and $f \circ g = I_Y$.*

Example 5.4.5. We show that the functions in Example 5.4.2 are inverse functions, this time using Proposition 5.4.3 instead of Definition 5.4.1. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$, $f(x) = 3x - 1$ and $g(x) = \frac{x+1}{3}$. Then $g \circ f : \mathbb{R} \to \mathbb{R}$ satisfies $g \circ f = I_{\mathbb{R}}$. To see this, let $x \in \mathbb{R}$. Then

$$\begin{aligned} (g \circ f)(x) &= g(f(x)) \\ &= g(3x - 1) \\ &= \frac{(3x-1)+1}{3} \\ &= \frac{3x}{3} \\ &= x = I_{\mathbb{R}}(x). \end{aligned}$$

Similarly, $f \circ g : \mathbb{R} \to \mathbb{R}$ satisfies $f \circ g = I_{\mathbb{R}}$, which we leave for Exercise 5.4.5. Hence f and g are inverse functions, by Proposition 5.4.3. ◊

Note that not all functions are invertible.

Example 5.4.6.

(1) The function $f : \mathbb{R} \to [1, \infty)$ by, for all $x \in \mathbb{R}$, $f(x) = x^2 + 1$ is not 1-1. For example, $f(1) = 2 = f(-1)$. It follows that f is not invertible; it is not possible to find a function $g : [1, \infty) \to \mathbb{R}$ such that, for all $x \in \mathbb{R}$ and for all $y \in [1, \infty)$, $y = f(x) \Leftrightarrow x = g(y)$. If an inverse function g existed, then we would need to have $g(2) = 1$ and $g(2) = -1$, which is not possible since g is a function.

(2) The function $f : \mathbb{Z} \to \mathbb{Z}$ by, for all $n \in \mathbb{Z}$, $f(n) = 2n + 1$ is not invertible; it is not possible to find a function $g : \mathbb{Z} \to \mathbb{Z}$ such that, for all $n, m \in \mathbb{Z}$, $m = f(n) \Leftrightarrow n = g(m)$. If an inverse function g for f existed, then the integer $n = g(2)$ would have to satisfy $f(n) = 2$. However, one can prove that $\operatorname{ran} f$ is

the set of odd integers, so no such integer n exists. The problem here is that *f is not onto.* $\Diamond$

The examples above motivate the next theorem, which characterizes invertible functions as exactly those which are bijections.

Theorem 5.4.7. *Let X, Y be sets and let $f : X \to Y$.*

(1) *Then f is invertible iff f is a bijection.*

(2) *If f is invertible, then its inverse function is unique.*

Notation 5.4.8. When $f : X \to Y$ is invertible, the unique inverse function is denoted by f^{-1}, and $f^{-1} : Y \to X$.

So, in Example 5.4.6(1), the function f is not invertible because f is not a bijection; f is not 1-1. In Example 5.4.6(2), the function f is not invertible because f is not a bijection; f is not onto.

Proof. Let $f : X \to Y$. We first prove (1).

($\Rightarrow$) Assume that f is invertible. We must show that f is a bijection.

Since f is invertible, by Proposition 5.4.3 we may fix a function $g : Y \to X$ such that $g \circ f = I_X$ and $f \circ g = I_Y$. Note that by Exercise 5.3.4, the identity functions I_X and I_Y are bijections. Hence $g \circ f$ is a bijection. Thus, $g \circ f$ is 1-1, and so it follows by Theorem 5.3.10(4) that f is also 1-1. Similarly $f \circ g$ is a bijection. Thus, $f \circ g$ is onto, and so it follows by Theorem 5.3.10(5) that f is also onto. Thus f is 1-1 and onto; i.e., f is a bijection, as desired.

($\Leftarrow$) Assume that f is a bijection. We must show that f is invertible. By Definition 5.4.1, we must define a function $g : Y \to X$ such that for all $x \in X$ and for all $y \in Y$, $y = f(x) \Leftrightarrow x = g(y)$.

To define g, let $y \in Y$ be given. Since f is onto, we can fix $x \in X$ such that $y = f(x)$. Since f is 1-1, this x is unique. Hence we define $g(y)$ to be this unique $x \in X$ such that $f(x) = y$. It follows by definition of g that for all $x \in X$ and for all $y \in Y$, $y = f(x) \Leftrightarrow x = g(y)$. Hence, by Definition 5.4.1, f is invertible.

To prove (2), we must assume that f is invertible and prove that the inverse function of f is unique. We use the standard method for proving that an object is unique, given that it exists, found in statement (2.7) in Subsection 2.1.5. Assume that we have functions $g_1 : Y \to X$ and $g_2 : Y \to X$ such that g_1 and g_2 are both inverses of f. We must prove that $g_1 = g_2$. Using Definition 5.1.7, we let $y \in Y$ be arbitrary and prove that $g_1(y) = g_2(y)$.

Let $x_1, x_2 \in X$ be such that $x_1 = g_1(y)$ and $x_2 = g_2(y)$. Then $f(x_1) = y$, since g_1 is an inverse of f, and similarly $f(x_2) = y$, since g_2 is an inverse of f. Since f is invertible, we know that f is a bijection, as already proved above. Thus f is 1-1. Since we have $f(x_1) = f(x_2)$, it follows that $x_1 = x_2$; i.e., $g_1(y) = g_2(y)$, as desired. Thus $g_1 = g_2$. $\square$

Corollary 5.4.9. *Let X, Y be sets and let $f : X \to Y$. If f is a bijection, then $f^{-1} : Y \to X$ is a bijection.*

Proof. Let $f : X \to Y$ be a bijection. By Theorem 5.4.7, f is invertible, and $f^{-1} : Y \to X$ is the inverse function of f. But then f is the inverse function of f^{-1} by Exercise 5.4.3, so f^{-1} is invertible by definition. Thus f^{-1} is a bijection, again by Theorem 5.4.7. □

Since every function maps its domain onto its range (see Theorem 5.3.6), we have the following (more precisely stated) corollary.

Corollary 5.4.10. *Let X, Y be sets, and assume $f : X \to Y$ is* 1-1*. Then the function $g : X \to \operatorname{ran} f$ defined by, for all $x \in X$, $g(x) = f(x)$ is invertible.*

Finally, the uniqueness statement in Theorem 5.4.7, together with Proposition 5.4.3, says that if you have a "candidate" function $g : Y \to X$ which you think is the inverse of a given function $f : X \to Y$, then it's easy to verify this using function composition.

Corollary 5.4.11. *Let X and Y be sets, and let $f : X \to Y$ and $g : Y \to X$. If $g \circ f = I_X$ and $f \circ g = I_Y$, then $g = f^{-1}$ and $f = g^{-1}$.*

We next present several examples illustrating these ideas.

Example 5.4.12. Consider the sets $X = \{1, 2, 3, 4\}$, $Y = \{a, b, c, d, e\}$, and $Z = \{0, 5, 10, 15, 20\}$. Define functions $f : X \to Y$ and $g : Y \to Z$ by

$$f(1) = c, \qquad f(2) = e, \qquad f(3) = e, \qquad f(4) = a,$$
$$g(a) = 10, \quad g(b) = 0, \quad g(c) = 5, \quad g(d) = 20, \quad g(e) = 15.$$

Then f is not invertible, since f isn't 1-1. On the other hand, we noted in Example 5.3.4 that g is a bijection, and hence g is invertible. The inverse function $g^{-1} : Z \to Y$ is defined by

$$g^{-1}(0) = b, \quad g^{-1}(5) = c, \quad g^{-1}(10) = a, \quad g^{-1}(15) = e, \quad g^{-1}(20) = d. \qquad \diamondsuit$$

Example 5.4.13. Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3 + 1$, for all $x \in \mathbb{R}$. Show that f is a bijection and find f^{-1}.

First we show that f is 1-1. Let $x_1, x_2 \in \mathbb{R}$ and assume that $f(x_1) = f(x_2)$. Then

$$\begin{aligned} (x_1)^3 + 1 &= (x_2)^3 + 1, && \text{so} \\ (x_1)^3 &= (x_2)^3, && \text{and so} \\ \sqrt[3]{(x_1)^3} &= \sqrt[3]{(x_2)^3}. \end{aligned}$$

Thus $x_1 = x_2$ by Theorem 2.1.5(2). (Note that it is important here that we are dealing with an *odd* root, rather than an *even* root.) Hence, f is 1-1.

Next we show that f is onto. Let $y \in \mathbb{R}$ be arbitrary. Consider $x = \sqrt[3]{y-1}$, which is a real number, and note that

$$f(x) = f(\sqrt[3]{y-1}) = (\sqrt[3]{y-1})^3 + 1 = (y-1) + 1 = y.$$

Hence f is onto.

Thus f is a bijection, and hence f is invertible, by Theorem 5.4.7.

Note that what we have actually shown is that for all $x \in \mathbb{R}$,

$$y = x^3 + 1 \Leftrightarrow x = \sqrt[3]{y-1};$$

i.e.,

$$y = f(x) \Leftrightarrow x = g(y),$$

where $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = \sqrt[3]{x-1}$. Thus, since the inverse of f is unique, $f^{-1} : \mathbb{R} \to \mathbb{R}$ by $f^{-1}(x) = \sqrt[3]{x-1}$ for all $x \in \mathbb{R}$. ♢

We can see from Example 5.4.13 that if we can show *algebraically* that a function $f : X \to Y$ is a bijection, and in particular that it is onto, then we will automatically find a formula for the inverse function $f^{-1} : Y \to X$.

Example 5.4.14. Let $a, b \in \mathbb{R}$ with $a \neq 0$ and let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = ax + b$ for all $x \in \mathbb{R}$. We showed in Example 5.3.3 that f is a bijection, and hence f is invertible. The work done in that example also shows that $f^{-1} : \mathbb{R} \to \mathbb{R}$ by $f^{-1}(x) = \frac{x-b}{a}$ for all $x \in \mathbb{R}$. ♢

Example 5.4.15. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ by, for all $x, y \in \mathbb{R}$, $f(x, y) = (-y, x)$. We showed in Example 5.3.9 that f is a bijection, and hence f is invertible. The work done in that example also shows that $f^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ by $f^{-1}(x, y) = (y, -x)$ for all $(x, y) \in \mathbb{R}^2$. ♢

It is not always possible to use algebraic methods to find a formula for the inverse function of a bijection.

Example 5.4.16. In an analysis or calculus course, the *natural logarithm* function $\ln : (0, \infty) \to \mathbb{R}$ is defined by $\ln x = \int_1^x \frac{1}{t}\,dt$, for all $x \in (0, \infty)$. Using the derivative of ln, one can prove that ln is increasing and hence 1-1, and using the Intermediate Value Theorem, one can prove that ln is onto. Thus ln is a bijection, and so ln is invertible. The inverse function $\ln^{-1} : \mathbb{R} \to (0, \infty)$ is defined by, for all $x \in \mathbb{R}$ and for all $y \in \mathbb{R}^+$,

$$\ln^{-1}(x) = y \Leftrightarrow x = \ln y.$$

The inverse function $\ln^{-1}$ is usually called exp, the *natural exponential function*, since it satisfies the usual rules of exponents and its derivative is itself. Thus we have that for all $x \in \mathbb{R}$ and for all $y \in \mathbb{R}^+$,

$$\exp(x) = y \Leftrightarrow x = \ln y,$$

or, in more familiar notation,

$$e^x = y \Leftrightarrow x = \ln y. \qquad ♢$$

Exercises 5.4

1. For each function f,
   (i) determine whether f is 1-1;
   (ii) determine whether f is onto;
   (iii) use your answers to (i) and (ii) to determine whether f is invertible, and if f *is* invertible, then find the inverse function f^{-1}.

Prove your answers.

(a) $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x + |x|$.

(b) $f : \{x \in \mathbb{R} \mid x \neq \frac{3}{5}\} \to \{y \in \mathbb{R} \mid y \neq \frac{2}{5}\}$ by $f(x) = \frac{2x+1}{5x-3}$.

(c) $f : \{x \in \mathbb{R} \mid x \neq -\frac{d}{c}\} \to \{y \in \mathbb{R} \mid y \neq \frac{a}{c}\}$ by $f(x) = \frac{ax+b}{cx+d}$, where $a, b, c, d \in \mathbb{R}$ have the property that $ad - bc \neq 0$ and $c \neq 0$.

(d) $f : (-\infty, 3] \to [2, \infty)$ by $f(x) = (x-3)^2 + 2$.

(e) $f : (-\infty, 1] \to (-\infty, 4]$ by $f(x) = 4 - \sqrt{1 - x^3}$.

(f) $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x + y$.

(g) $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = (x - y)^3$.

(h) $f : \mathbb{R} \to \mathbb{R}^2$ by $f(x) = (x, x)$.

(i) $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x, y) = (x + y, x - y)$.

(j) $f : \mathbb{R}^2 \to \mathbb{R}^3$ by $f(x, y) = (x + y, x - y, xy)$.

(k) $f : \mathbb{R}^3 \to \mathbb{R}^3$ by $f(x, y, z) = (x + y, y + z, x + z)$.

(l) $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f(x, y, z) = (x + y, y + z)$.

(m) $f : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+$ by $f(m, n) = 2^{m-1}(2n - 1)$. (**HINT:** You will need the Fundamental Theorem of Arithmetic (Theorem 2.3.3).)

(n) $F : \mathsf{P} \to \mathsf{P}$ by $F(p) = p'$. (See Exercise 5.1.7.)

(o) $f : \mathcal{C} \to \mathbb{Z}$, where $\mathcal{C} = \{A \in \mathcal{P}(\mathbb{Z}) \mid A \text{ is finite}\}$ and $f(A)$ is the sum of all elements of A.

2. For each of the piecewise-defined functions f,

(i) determine whether f is 1-1;

(ii) determine whether f is onto;

(iii) use your answers to (i) and (ii) to determine whether f is invertible, and if f *is* invertible, then find the inverse function f^{-1}.

Prove your answers.

(a) $f : \mathbb{Z} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} n - 1 & \text{if } n \text{ is even,} \\ n + 3 & \text{if } n \text{ is odd.} \end{cases}$$

(b) $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ 2x & \text{if } x < 0. \end{cases}$$

(c) $f : \mathbb{Z} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} n + 1 & \text{if } n \text{ is even,} \\ 2n & \text{if } n \text{ is odd.} \end{cases}$$

(d) $f : \mathbb{Z} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} 2n + 1 & \text{if } n \text{ is even,} \\ n + 3 & \text{if } n \text{ is odd.} \end{cases}$$

3. Let $f : X \to Y$ be invertible and let $g : Y \to X$ be an inverse function of f. Use Definition 5.4.1 to prove that g is also invertible and that f is an inverse function of g.

4. Provide the missing details in the proof of Proposition 5.4.3.

5. Provide the missing details in Example 5.4.5.

6. Use Definition 5.4.1 to provide an alternate proof of the fact in Theorem 5.4.7 that if $f : X \to Y$ is invertible, then f is a bijection.
7. Give a direct proof (i.e., using definitions) of Corollary 5.4.9.
8. Let $f : X \to Y$ and $g : Y \to Z$ be invertible functions. Prove that $g \circ f : X \to Z$ is invertible and that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. (**HINT:** Use Corollary 5.4.11.)
9. Let $X, Y \subseteq \mathbb{R}$, and assume that $f : X \to Y$ is invertible and increasing on X (see Exercise 5.1.9). Prove that $f^{-1} : Y \to X$ is increasing on Y.
10. Let X, Y be sets and let $f : X \to Y$. Say that f is *left-invertible* if there exists a function $g : Y \to X$ such that $g \circ f = I_X$; in this case, g is called a *left inverse* of f. Say that f is *right-invertible* if there exists a function $h : Y \to X$ such that $f \circ h = I_Y$; in this case, h is called a *right inverse* of f.
    (a) Prove that $f : X \to Y$ is 1-1 if and only if f is left-invertible.
    (b) Prove that $f : X \to Y$ is onto if and only if f is right-invertible. (**NOTE:** The proof that f is right-invertible when f is onto requires the Axiom of Choice (see Section 4.4). Do you see why?)
    (c) Give an example of a function that has a left inverse but no right inverse and an example of a function that has a right inverse but no left inverse.
11. Let X, Y be sets and let $f : X \to Y$.
    (a) Prove that f is 1-1 if and only if for all sets Z, for all functions $h : Z \to X$ and $k : Z \to X$, if $f \circ h = f \circ k$, then $h = k$.
    (b) Prove that f is onto if and only if for all sets Z, for all functions $h : Y \to Z$ and $k : Y \to Z$, if $h \circ f = k \circ f$, then $h = k$.

5.5. Functions and sets

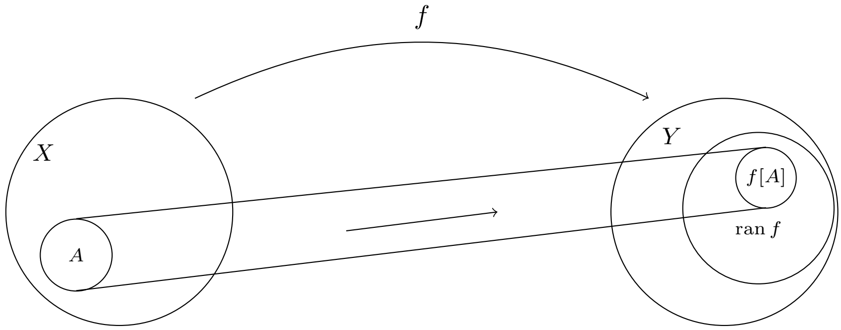
Given a function $f : X \to Y$, f is defined on *elements* of X. However, sometimes we also want to consider the image of an entire *subset* of X under f.

Definition 5.5.1. Let $f : X \to Y$ and $A \subseteq X$. The *image of A under f* is the set

$$\{y \in Y \mid (\exists x \in A)[y = f(x)]\} = \{f(x) \mid x \in A\},$$

which is denoted by the notation $f[A]$.

By definition, $f[A]$ is the set of all images of elements of A under f. A picture can help us visualize this concept.



Warning: Note the potential for confusion with this notation. Given a function $f : X \to Y$, an element $x \in X$, and a subset $A \subseteq X$,

$$f(x) \text{ is an } \textit{element} \text{ of } Y,$$
$$f[A] \text{ is a } \textit{subset} \text{ of } Y.$$

Our use of square brackets, rather than parentheses, is meant to help you remember the difference between the image of an *element* of the domain under f and the image of a *subset* of the domain under f. See Subsection 5.5.1 for further discussion regarding notation.

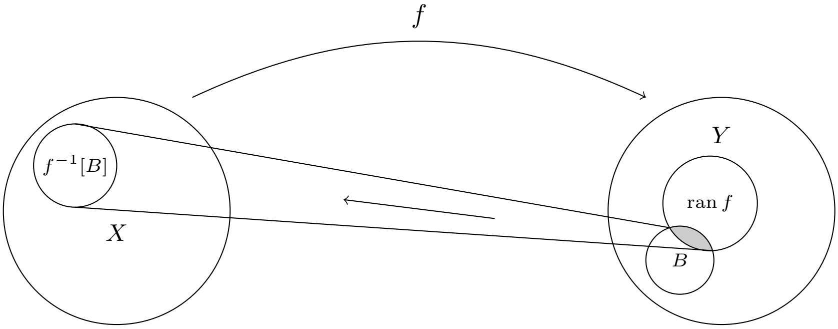
The next definition is of a related concept.

Definition 5.5.2. Let $f : X \to Y$ and $B \subseteq Y$. The *inverse image of B under f* (sometimes called the *preimage of B under f*) is the set

$$\{x \in X \mid f(x) \in B\},$$

which is denoted by the notation $f^{-1}[B]$.

By definition, $f^{-1}[B]$ is the set of all elements in X whose image under f is in B, i.e., the set of all preimages of elements of B. Again, a picture can help clarify this concept.



Warning: Once again, note the potential for confusion with this notation. In particular, it is important to note in Definition 5.5.2 that the inverse image $f^{-1}[B]$ under f has nothing to do with inverse functions.

We are not claiming in Definition 5.5.2 *that f is invertible, and indeed f need not be invertible.* The "exponent" -1 in the notation $f^{-1}[B]$ is merely notation. See Subsection 5.5.1 for further discussion regarding notation.

Example 5.5.3. Let $X = \{1, 2, 3, 4, 5, 6\}$ and $Y = \{5, 10, 15, 20, 25\}$. We let $f : X \to Y$ by

$$\begin{array}{lll} f(1) = 15, & f(2) = 10, & f(3) = 5, \\ f(4) = 10, & f(5) = 15, & f(6) = 10. \end{array}$$

We find $f[\{2, 3\}]$, $f[\{2, 6\}]$, $f^{-1}[\{5, 10\}]$, and $f^{-1}[\{20, 25\}]$.

Note that

$$\begin{aligned} f[\{2,3\}] &= \{f(x) \mid x \in \{2,3\}\} \\ &= \{f(2), f(3)\} \\ &= \{5, 10\}, \\ f[\{2,6\}] &= \{f(2), f(6)\} \\ &= \{10\} \end{aligned}$$

and

$$\begin{aligned} f^{-1}[\{5,10\}] &= \{x \in X \mid f(x) \in \{5,10\}\} \\ &= \{2,3,4,6\}, \\ f^{-1}[\{20,25\}] &= \{x \in X \mid f(x) \in \{20,25\}\} \\ &= \emptyset. \end{aligned}$$

♢

Example 5.5.4. Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 2x + 5$ for all $x \in \mathbb{R}$. Let $A = [1,2]$, $B = (-\infty, -2)$, and $C = (4, \infty)$. We find $f[A]$, $f[B]$, and $f^{-1}[C]$.

$$\begin{aligned} f[A] &= \{y \in \mathbb{R} \mid (\exists x \in A)[y = f(x)]\} \\ &= \{y \in \mathbb{R} \mid (\exists x \in \mathbb{R})[1 \le x \le 2 \text{ and } y = 2x + 5]\} \\ &= \{2x + 5 \mid 1 \le x \le 2\}. \end{aligned}$$

Note that by properties of real numbers,

$$\begin{aligned} 1 \le x \le 2 &\Leftrightarrow 2 \le 2x \le 4 \\ &\Leftrightarrow 7 \le 2x + 5 \le 9. \end{aligned}$$

Hence $f[A] = [7, 9]$. See Figure 5.2.

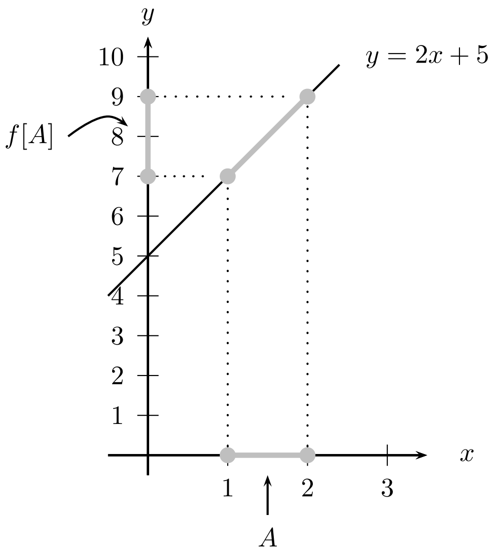


Figure 5.2. $f[A] = [7, 9]$.

Also, $f[B] = \{2x + 5 \mid x < -2\}$. Since $x < -2$ iff $2x + 5 < 1$, $f[B] = (-\infty, 1)$. See Figure 5.3.

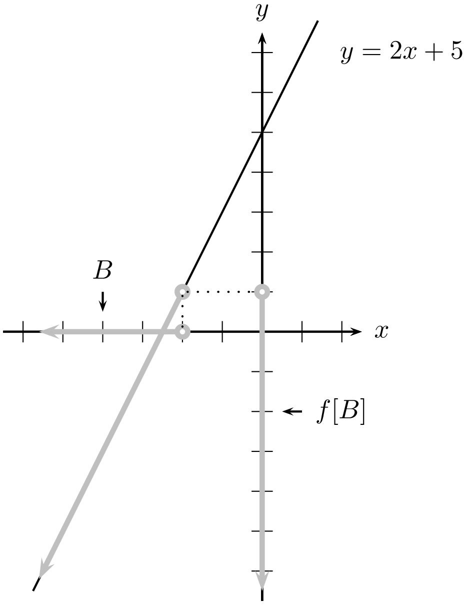


Figure 5.3. $f[B] = (-\infty, 1)$.

Finally,

$$\begin{aligned} f^{-1}[C] &= \{x \in \mathbb{R} \mid f(x) \in C\} \\ &= \{x \in \mathbb{R} \mid 2x + 5 > 4\} \\ &= \left\{x \in \mathbb{R} \mid x > -\frac{1}{2}\right\} \\ &= \left(-\frac{1}{2}, \infty\right). \end{aligned}$$

See Figure 5.4.

Note that in the previous example our task is much easier because the function $f(x) = 2x + 5$ is *increasing*. When the function is not 1-1, we must be much more careful.

Example 5.5.5. Let $f : [0, 2\pi] \to [-1, 1]$ by $f(x) = \cos x$. Let $A = \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ and $B = [0, 1]$. We find $f[A]$ and $f^{-1}[B]$.

From the graph of $y = \cos x$ we see that

$$\begin{aligned} f[A] &= \{f(x) \mid x \in A\} \\ &= \left\{\cos x \;\middle|\; \frac{\pi}{2} \le x \le \frac{3\pi}{2}\right\} \\ &= [-1, 0]. \end{aligned}$$

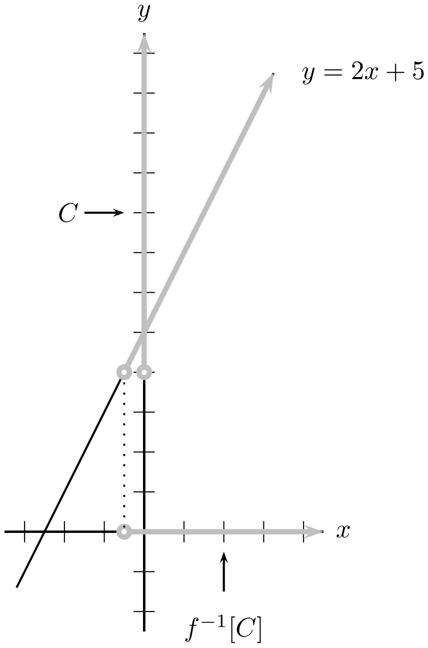


Figure 5.4. $f^{-1}[C] = (-\frac{1}{2}, \infty)$.

◊

It's important to note here that $f\left[[\frac{\pi}{2}, \frac{3\pi}{2}]\right]$ is not simply equal to $[f(\frac{\pi}{2}), f(\frac{3\pi}{2})]$; in fact, $[f(\frac{\pi}{2}), f(\frac{3\pi}{2})]$ isn't an interval. See Figure 5.5.

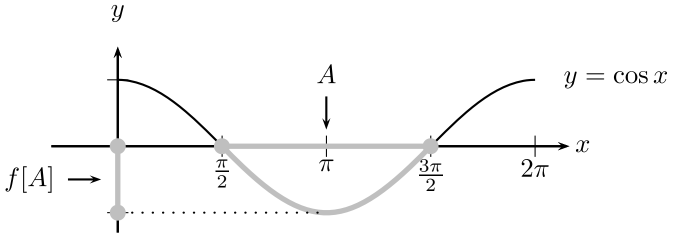


Figure 5.5. $f[A] = [-1, 0]$.

Again using the graph of the function and paying particular attention to the fact that the function is not 1-1 (i.e., some elements of B may have more than one preimage), we also see that

$$\begin{aligned} f^{-1}[B] &= \{x \in [0, 2\pi] \mid f(x) \in B\} \\ &= \{x \in [0, 2\pi] \mid 0 \leq \cos x \leq 1\} \\ &= \left[0, \frac{\pi}{2}\right] \cup \left[\frac{3\pi}{2}, 2\pi\right]. \end{aligned}$$

See Figure 5.6. ◊

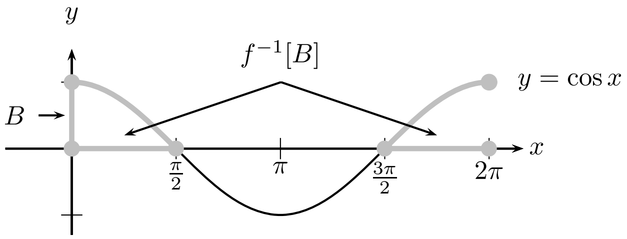


Figure 5.6. $f^{-1}[B] = \left[0, \frac{\pi}{2}\right] \cup \left[\frac{3\pi}{2}, 2\pi\right]$.

Now that we've seen several examples, we prove a theorem about how these new concepts interact with the set operations $\cup$ and $\cap$.

Theorem 5.5.6. *Let $f : X \to Y$, and let $A, B \subseteq X$ and $C, D \subseteq Y$. Then:*

(1) $f[A \cup B] = f[A] \cup f[B]$.

(2) $f[A \cap B] \subseteq f[A] \cap f[B]$, *but in general, equality need not hold.*

(3) $f^{-1}[C \cup D] = f^{-1}[C] \cup f^{-1}[D]$.

(4) $f^{-1}[C \cap D] = f^{-1}[C] \cap f^{-1}[D]$.

Proof. We prove (1) and (3) and leave (2) and (4) for Exercises 5.5.3 and 5.5.4.

Let $f : X \to Y$, $A, B \subseteq X$, and $C, D \subseteq Y$.

(1) We show $f[A \cup B] = f[A] \cup f[B]$. First let $y \in f[A \cup B]$. Then by Definition 5.5.1, we may fix $x \in A \cup B$ such that $y = f(x)$. Since $x \in A \cup B$, $x \in A$ or $x \in B$. We argue by cases. If $x \in A$, then $y \in f[A]$ by Definition 5.5.1 since $y = f(x)$. It follows that $y \in f[A] \cup f[B]$. Similarly, if $x \notin A$, then $x \in B$, so $y \in f[B]$ by Definition 5.5.1, and hence $y \in f[A] \cup f[B]$. Thus $f[A \cup B] \subseteq f[A] \cup f[B]$.

Next let $y \in f[A] \cup f[B]$. Then $y \in f[A]$ or $y \in f[B]$. Without loss of generality[2], assume that $y \in f[A]$, since the argument for $y \in f[B]$ is analogous. Since $y \in f[A]$, by Definition 5.5.1 we may fix $x \in A$ such that $y = f(x)$. Then $x \in A \cup B$ and $y = f(x)$, so again by Definition 5.5.1, $y \in f[A \cup B]$. Hence $f[A] \cup f[B] \subseteq f[A \cup B]$, and so $f[A \cup B] = f[A] \cup f[B]$.

(3) We show $f^{-1}[C \cup D] = f^{-1}[C] \cup f^{-1}[D]$. First let $x \in f^{-1}[C \cup D]$. Then by Definition 5.5.2, $f(x) \in C \cup D$, so $f(x) \in C$ or $f(x) \in D$. We argue by cases. If $f(x) \in C$, then we know by Definition 5.5.2 that $x \in f^{-1}[C]$, and hence $x \in f^{-1}[C] \cup f^{-1}[D]$. Similarly, if $f(x) \notin C$, then $f(x) \in D$. Once again by Definition 5.5.2 we have that $x \in f^{-1}[D]$, and hence $x \in f^{-1}[C] \cup f^{-1}[D]$. Thus $f^{-1}[C \cup D] \subseteq f^{-1}[C] \cup f^{-1}[D]$.

Next, let $x \in f^{-1}[C] \cup f^{-1}[D]$. Without loss of generality, we assume that $x \in f^{-1}[D]$, since the argument for $x \in f^{-1}[C]$ is analogous. By

[2] Or we could again argue by cases.

Definition 5.5.2, $f(x) \in D$, and hence $f(x) \in C \cup D$. Again by Definition 5.5.2, $x \in f^{-1}[C \cup D]$. Thus $f^{-1}[C] \cup f^{-1}[D] \subseteq f^{-1}[C \cup D]$, and so $f^{-1}[C \cup D] = f^{-1}[C] \cup f^{-1}[D]$. □

Note in our proof of Theorem 5.5.6(1) and (3), we have introduced the phrase "without loss of generality" (sometimes abbreviated by "WLOG") as a way of handling a proof in which the argument for one case is *completely analogous* to the argument for the other cases. Specifically, our use of the phrase "without loss of generality, assume that $y \in f[A]$, since the argument for $y \in f[B]$ is analogous" in the proof of part (1) means that the proof for the case when $y \in f[B]$ can be obtained from the proof of the case when $y \in f[A]$ by simultaneously replacing every occurrence of A with B and every occurrence of B with A. If it was not the case that the proof of the second case was entirely analogous to the proof of the first case, then it would be incorrect to use "without loss of generality" to argue only one case.

5.5.1. More on notation. We emphasize once again that, given a function $f : X \to Y$, f is defined on *elements* of X, not on *subsets* of X. Given a subset $A \subseteq X$, we defined the image of A under f to be the set

$$f[A] = \{y \in Y \mid (\exists x \in A)[y = f(x)]\},$$

the set of all images of elements of A. What we are actually doing here is defining a *new* function from $\mathcal{P}(X)$ to $\mathcal{P}(Y)$. As we noted at the beginning of this section, we are using square brackets [ ], rather than parentheses ( ), around the input as a signal to help us remember that $f(x)$ is an element of Y and $f[A]$ is a subset of Y. The notation $f[A]$ is one of the standard notations for this concept.

It's important to note that many authors use the notation $f(A)$ for the set

$$\{y \in Y \mid (\exists x \in A)[y = f(x)]\}.$$

Here there is no signal from the notation, so the reader is responsible for determining from the context whether $f(A)$ denotes the image of an *element* of the domain or the image of a *subset* of the domain.

The situation with the inverse image (preimage) of a set $B \subseteq Y$ under f,

$$f^{-1}[B] = \{x \in X \mid f(x) \in B\},$$

is similar. As before, what we are actually doing is defining a *new* function from $\mathcal{P}(Y)$ to $\mathcal{P}(X)$. The potential for confusion is even greater: the function f *may not be invertible*, i.e., f^{-1} may not exist, and the "exponent" is nothing more than notation. Again, the notation $f^{-1}[B]$ is one of the standard notations for this concept.

It's important to note that many authors use the notation $f^{-1}(B)$ for the set

$$\{x \in X \mid f(x) \in B\}.$$

As before, one must determine the meaning of the notation $f^{-1}(B)$ from the context in which it appears.

In the future, always check and adhere to the notation for these sets given by whatever source you are using.

Exercises 5.5

1. Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2 + 1$. Find each of the following. You do not need to prove your answers, but you might want to draw a picture. Be careful!

$$f[\{-3, 2, 7\}], \quad f[[-1, 3]], \quad f[(-\infty, -2)],$$
$$f^{-1}[\{-2, 5, 16\}], \quad f^{-1}[[-2, 4]], \quad f^{-1}[[8, \infty)]$$

2. Let $f : X \to Y$ where $X = \{1, 2, 3, 4, 5, 6\}$, $Y = \{p, q, r, s, t, z\}$, and $f(1) = p$, $f(2) = p$, $f(3) = s$, $f(4) = t$, $f(5) = z$, and $f(6) = t$. Find each of the following:

$$f[\{1, 3, 4, 6\}], \quad f^{-1}[\{p, q, s\}], \quad f^{-1}[\{r\}], \quad f^{-1}[f[\{1, 4, 5\}]].$$

3. Prove Theorem 5.5.6(2). Let X and Y be sets, $A, B \subseteq X$, and $f : X \to Y$.
   (a) Prove that $f[A \cap B] \subseteq f[A] \cap f[B]$.
   (b) Give an example of sets X and Y, subsets $A, B \subseteq X$, and a function $f : X \to Y$ such that $f[A \cap B] \neq f[A] \cap f[B]$.

4. Prove Theorem 5.5.6(4). Let X, Y be sets, $C, D \subseteq Y$, and $f : X \to Y$. Prove that $f^{-1}[C \cap D] = f^{-1}[C] \cap f^{-1}[D]$.

5. Let X and Y be sets, $A, B \subseteq X$, and $f : X \to Y$.
   (a) Prove that if $A \subseteq B$, then $f[A] \subseteq f[B]$.
   (b) Give an example of sets X and Y, subsets $A, B \subseteq X$, and $f : X \to Y$ such that $f[A] \subseteq f[B]$, but $A \nsubseteq B$.

6. Let X and Y be sets, $A \subseteq X$, and $f : X \to Y$.
   (a) Prove that $A \subseteq f^{-1}[f[A]]$.
   (b) Give an example of sets X and Y, a subset $A \subseteq X$, and $f : X \to Y$ such that $A \neq f^{-1}[f[A]]$.

7. Let X and Y be sets, $B \subseteq Y$, and $f : X \to Y$.
   (a) Prove that $f[f^{-1}[B]] \subseteq B$.
   (b) Give an example of sets X and Y, a subset $B \subseteq Y$, and $f : X \to Y$ such that $f[f^{-1}[B]] \neq B$.

8. Let X and Y be sets, $A, B \subseteq X$, and $f : X \to Y$ be 1-1. Prove that $f[A \cap B] = f[A] \cap f[B]$.
   Warning: If you do not use the hypothesis that f is 1-1 at some point, then you do not have a proof.

9. Let X and Y be sets, $A, B \subseteq X$, and $f : X \to Y$ be 1-1. Prove that if $f[A] \subseteq f[B]$, then $A \subseteq B$.
   Warning: If you do not use the hypothesis that f is 1-1, then you do not have a proof.

10. Let X and Y be sets, $A \subseteq X$, and $f : X \to Y$ be 1-1. Prove that $f^{-1}[f[A]] = A$.
    Warning: If you do not use the hypothesis that f is 1-1 at some point, then you do not have a proof.

11. Let X and Y be sets, $B \subseteq Y$, and $f : X \to Y$ be onto. Prove that $f[f^{-1}[B]] = B$.

    Warning: If you do not use the hypothesis that f is onto at some point, then you do not have a proof.

Chapter 6

An Introduction to Number Theory

The goal of the first five chapters of this book is to help you master the concepts essential for success in most mathematics courses: logic, proof techniques, sets, and functions. We turn now to material chosen to help prepare you for future math courses, such as abstract algebra and analysis.

In this chapter, we return to studying properties of the integers. By now, you should be feeling confident in your ability to write proofs, so we will be including far fewer Given-Goal diagrams than we have seen previously. However, it is good practice for you to construct them yourself.

6.1. The Division Algorithm and the Well-Ordering Principle

So far throughout this text, we have assumed the fact that every integer is either even or odd, but not both, and we have used this fact several times to divide our proofs into two cases. While this assumption seems quite harmless, it in fact requires proof. In this section we will prove a theorem commonly known as the "Division Algorithm", which is one of the most important theorems about integers, since it implies that proofs involving integers may be divided into finitely many cases (such as two, for "even" and "odd").

Theorem 6.1.1 (Division Algorithm)**.** *Let $a, b \in \mathbb{Z}$ with $b > 0$. Then there exist unique $q, r \in \mathbb{Z}$ such that $a = bq + r$, where $0 \leq r < b$.*

Note that when $b = 2$, the Division Algorithm says that every integer a can be uniquely expressed in one of two forms:

$$a = 2q + 0 \quad \text{or} \quad a = 2q + 1, \quad \text{where } q \in \mathbb{Z};$$

i.e., every integer a is either even or odd. Next, we consider some specific examples that demonstrate the conclusion of the Division Algorithm.

Example 6.1.2.

(1) $a = 75$ and $b = 12$.

Here, the conclusion of the Division Algorithm is essentially what you remember about long division: "how many times does 12 go into 75, and what's the remainder?" In other words, $q = 6$, since 72 is the largest multiple of 12 that is less than or equal to 75, and the remainder r is 3; i.e., $75 = 12{\cdot}6{+}3$.

(2) $a = -4$ and $b = 3$.

When $a < 0$, we must be more careful; the r we're looking for must satisfy $0 \leq r < 3 = b$. Again we look for the largest multiple of 3 that is less than or equal to -4, which is -6, so $q = -2$ and $r = 2$; i.e., $-4 = 3 \cdot -2 + 2$. $\diamondsuit$

These examples, while seemingly trivial, give us the idea behind how to prove the "existence" part of the Division Algorithm. Given $a, b \in \mathbb{Z}$ with $a, b > 0$, our intuition tells us that we can find q by adding b to itself consecutively until we reach a multiple of b strictly greater than a; "backing up" one copy of b gives us a multiple of b from which we can compute q. Once we have q, it is easy to compute r.

Another way of thinking about this is that we can keep subtracting b from a until we first reach a remainder strictly smaller than b. This gives us both q and r. Our job is to formalize this idea (in fact, you might be convinced already). To write it down in general, we need to know that the process of subtracting will stop, regardless of which numbers a and b we begin with. It turns out that we have two options for formalizing this argument; we can use a proof by induction or we can use another important fact about the nonnegative integers, called the Well-Ordering Principle. In what follows, we denote the set of nonnegative integers by $\mathbb{Z}^{\geq 0}$; i.e., $\mathbb{Z}^{\geq 0} = \{n \in \mathbb{Z} \mid n \geq 0\}$.

Well-Ordering Principle 6.1.3. Every nonempty set of nonnegative integers has a least element. In notation,

$$\text{if } S \subseteq \mathbb{Z}^{\geq 0} \text{ and } S \neq \emptyset, \text{ then } (\exists m \in S)(\forall x \in S)[m \leq x].$$

Phrased another way, the Well-Ordering Principle states that there does not exist an infinite "descending chain" of positive integers

$$0 < \cdots < n_3 < n_2 < n_1.$$

Like the Principle of Mathematical Induction, the Well-Ordering Principle is an *axiom* about the nonnegative integers; it is not possible for us to prove the Well-Ordering Principle solely from the Basic Properties of Integers 1.2.3. In fact, the Well-Ordering Principle (WOP) is *equivalent* to PMI; i.e., one can prove WOP by assuming PMI, and conversely one can prove PMI by assuming WOP. See Exercise 6.1.8.

We now use the Well-Ordering Principle to formalize our intuition about the proof of the Division Algorithm; note that we no longer need the assumption that $a > 0$. As we noted in Subsection 2.1.5, the proof of the Division Algorithm must prove "existence" as well as "uniqueness". We'll be thinking about uniqueness in terms of statement (2.6) in that subsection.

Proof of Theorem 6.1.1. Let $a, b \in \mathbb{Z}$ with $b > 0$.

Existence: We prove there exist $q, r \in \mathbb{Z}$ such that $a = bq + r$ and $0 \leq r < b$. Let

$$S = \{n \in \mathbb{Z} \mid n \geq 0 \text{ and } (\exists x \in \mathbb{Z})[n = a - bx]\}.$$

(Note that when $a > 0$, the elements of this set that are less than or equal to a will be the numbers you get by successively subtracting copies of b.)

To use the Well-Ordering Principle, we must show that $S \neq \emptyset$. We consider cases. Note that if $a \geq 0$, then $a - b \cdot 0 = a \geq 0$, so $a \in S$ since it has the right form. If $a < 0$, then $a - b \cdot a = a(1 - b) \geq 0$ since $a < 0$ and $b \geq 1$. In this case $a - b \cdot a \in S$. Hence in any case, $S \neq \emptyset$.

Thus, by the Well-Ordering Principle 6.1.3, S has a least element r. Fix $q \in \mathbb{Z}$ such that $r = a - bq$, which is possible by definition of S. Then $a = bq + r$, as desired.

We know that $r \geq 0$ by definition of S, so we must show that $r < b$. Suppose for the sake of a contradiction that $r \geq b$. Then

$$\begin{aligned} 0 \leq r - b &= a - bq - b \\ &= a - (b + 1)q, \end{aligned}$$

so $r - b \in S$ by definition. But $r - b < r$ since $b > 0$, contradicting the fact that r is the least element of S. Hence $r < b$, as desired.

Uniqueness: Suppose that we also have $q_1, r_1 \in \mathbb{Z}$ with $a = bq_1 + r_1$, where $0 \leq r_1 < b$. We must show that $q = q_1$ and $r = r_1$.

So, we have

$$a = bq + r = bq_1 + r_1,$$

so

$$r - r_1 = bq_1 - bq = b(q_1 - q).$$

In other words, $b \mid (r - r_1)$. However, $0 \leq r < b$ and $0 \leq r_1 < b$, so $-b < r - r_1 < b$. Thus, since $b \mid (r - r_1)$, $r - r_1 = 0$. Hence $r = r_1$, as desired. Since $bq + r = bq_1 + r_1$, this implies that $bq = bq_1$. Hence $q = q_1$ by cancellation in $\mathbb{Z}$, since $b \neq 0$. □

As mentioned earlier, the strength of the Division Algorithm is that it can be used to reduce proofs about integers to finitely many cases. For example, when $b = 3$, the Division Algorithm says that every integer a can be expressed in exactly one of three forms:

$$a = 3q + 0, \quad a = 3q + 1, \quad \text{or } a = 3q + 2, \quad \text{where } q \in \mathbb{Z}.$$

This was the fact provided in Exercise 2.2.6a.

The Division Algorithm also says that every integer a can be expressed in exactly one of four forms:

$$a = 4q + 0, \quad a = 4q + 1, \quad a = 4q + 2, \quad \text{or } a = 4q + 3, \quad \text{where } q \in \mathbb{Z};$$

here, we are taking $b = 4$. Which version of the Division Algorithm we use (or try to use) depends on the question being asked.

Example 6.1.4. Prove that the square of any integer has one of the forms $3k$ or $3k+1$, where $k \in \mathbb{Z}$.

Given the form of this statement, it makes sense to apply the Division Algorithm with $b = 3$.

Proof. Let $n \in \mathbb{Z}$. By the Division Algorithm (Theorem 6.1.1), n can be written uniquely in exactly one of the forms $3q$, $3q+1$, or $3q+2$, where $q \in \mathbb{Z}$.

Case I: $n = 3q$.
Then
$$n^2 = (3q)^2 = 9q^2 = 3(3q^2),$$
so $n^2 = 3k$, where $k = 3q^2 \in \mathbb{Z}$.

Case II: $n = 3q+1$.
Then
$$n^2 = (3q+1)^2 = 9q^2 + 6q + 1 = 3(3q^2 + 2q) + 1,$$
so $n^2 = 3k+1$, where $k = 3q^2 + 2q \in \mathbb{Z}$.

Case III: $n = 3q+2$.
Then
$$n^2 = (3q+2)^2 = 9q^2 + 12q + 4 = 3(3q^2 + 4q + 1) + 1,$$
so $n^2 = 3k+1$, where $k = 3q^2 + 4q + 1 \in \mathbb{Z}$.

Hence, the square of any integer has one of the forms $3k$ or $3k+1$, where $k \in \mathbb{Z}$.
□ ◊

Exercises 6.1

1. Prove that the fourth power of any integer has one of the forms $5k$ or $5k+1$, where k is an integer.
2. Prove that the cube of any integer has one of the forms $9k$, $9k+1$, or $9k+8$, where $k \in \mathbb{Z}$.
3. Use the Division Algorithm (Theorem 6.1.1) to prove that for all $n \in \mathbb{Z}^+$,
$$6 \mid n(n+1)(2n+1).$$
4. Assume that the set $I \subseteq \mathbb{Z}$ satisfies the following properties:
   (a) There exists $n \in I$ such that $n \neq 0$.
   (b) If $m, n \in I$, then $m + n \in I$.
   (c) If $m \in I$ and $a \in \mathbb{Z}$, then $am \in I$.
   Prove that there exists $n_0 \in \mathbb{Z}$ such that $I = \{kn_0 \mid k \in \mathbb{Z}\}$.
5. (Division Algorithm for polynomials with real coefficients) Let $n, m \in \mathbb{Z}^{\geq 0}$. Let f and g be polynomials with real coefficients of degrees n and m, respectively, so that neither polynomial is the zero polynomial. Prove that there exist polynomials q and r such that $f(x) = q(x)g(x) + r(x)$ for all $x \in \mathbb{R}$, where either $r = 0$, or $r \neq 0$ and $\deg r < \deg g$. (**HINT:** Use strong induction on the degree of f. Write f and g in polynomial form as in Definition 5.1.10 and consider the

polynomial $f(x) - \frac{a_n}{b_m} x^{n-m} g(x)$, where a_n is the coefficient of x^n in f and b_m is the coefficient of x^m in g. Note that the result also holds when $f = 0$.)

6. Let $n \in \mathbb{Z}^{\geq 0}$. Let f be a polynomial with real coefficients of degree n.
   (a) Let $a \in \mathbb{R}$. Using Exercise 6.1.5, prove that $f(a) = 0$ if and only if there exists a polynomial q with real coefficients such that $f(x) = q(x)(x - a)$ for all $x \in \mathbb{R}$.
   (b) Prove by induction on the degree n of f that there are at most n real numbers a such that $f(a) = 0$.

7. Prove Theorem 6.1.1 using a proof by induction, rather than the Well-Ordering Principle.

8. In this exercise, you will prove that PMI is logically equivalent to WOP; in other words, given PMI, show that you can deduce the statement WOP, and vice versa. We will be taking 0 as the base case in PMI.
   (a) Assume that WOP is true. To prove that PMI is true, let $P(n)$ be a statement about $n \in \mathbb{Z}^{\geq 0}$. *Assume that*
      (i) $P(0)$ is true, and
      (ii) for all $m \in \mathbb{Z}^{\geq 0}$, if $P(m)$ is true, then $P(m+1)$ is true.

      Your goal is to prove that for all $n \in \mathbb{Z}^{\geq 0}$, $P(n)$ is true. You should do this using a proof by contradiction and by applying WOP to the set $S = \{n \in \mathbb{Z}^{\geq 0} \mid P(n) \text{ is false}\}$. This means that you should *prove* that S is nonempty, so that you can apply WOP, to conclude that S has a least element n_0. What do statements (i) and (ii) imply about n_0?
   (b) Assume that PMI is true. To prove that WOP is true, let $S \subseteq \mathbb{Z}^{\geq 0}$ be nonempty. Your goal is to prove that S has a least element, which you should do by contradiction. Given $n \in \mathbb{Z}^{\geq 0}$, let $P(n)$ be the statement "for all $k \in \mathbb{Z}^{\geq 0}$ with $1 \leq k \leq n$, $k \notin S$". Use PMI to prove that the statement $(\forall n \in \mathbb{Z}^{\geq 0}) P(n)$ is true. What conclusion can you then draw about S?

6.2. Greatest common divisors and the Euclidean Algorithm

In this section, we discuss another concept you're probably familiar with, that of the greatest common divisor of two given nonzero integers. We'll also discuss a very useful algorithm not only for computing the greatest common divisor of two positive integers, but also for expressing the greatest common divisor as a "linear combination" of those integers, which is an important tool in number theory and abstract algebra. We begin with an example that relies on your intuition.

Example 6.2.1. The greatest common divisor of the integers -24 and 30 is exactly what it sounds like: the greatest positive integer that divides both -24 and 30.

Positive divisors of -24:	$1, 2, 3, 4, 6, 8, 12, 24.$
Positive divisors of 30:	$1, 2, 3, 5, 6, 10, 15, 30.$
Common positive divisors of -12 and 30:	$1, 2, 3, 6.$

So, the greatest common divisor of -24 and 30 is 6. ◊

We can see from this example why the greatest common divisor of two integers, not both of which are 0, exists and is unique. The set of positive integer divisors of a nonzero integer is finite, and the maximum element of a finite set is unique (see Exercise 8.2.14). We are ready for the definition.

Definition 6.2.2. Let $a, b \in \mathbb{Z}$ with at least one of a and b nonzero. The *greatest common divisor* (*gcd*) of a and b is the unique positive integer d such that

(1) $d \mid a$ and $d \mid b$ and

(2) for all $c \in \mathbb{Z}^+$, if $c \mid a$ and $c \mid b$, then $c \leq d$.

We denote the gcd of a and b by (a, b) or $\gcd(a, b)$.

Computing the greatest common divisor of two integers using the definition is tedious when the numbers are large. In this section, we use the Division Algorithm to derive an algorithm, called the *Euclidean Algorithm*, for computing gcd's. The algorithm relies on the following lemma.

Lemma 6.2.3. *Let $a, b \in \mathbb{Z}$ with $a \neq 0$ and $b \neq 0$. Assume we have $q, r \in \mathbb{Z}$ such that $a = bq + r$. Then $(a, b) = (b, r)$.*

Proof. Let $a, b, q, r \in \mathbb{Z}$ with $a \neq 0, b \neq 0$, and $a = bq + r$. To show that $(a, b) = (b, r)$, we show that the pair a, b and the pair b, r have exactly the same common divisors; it follows immediately that the pair a, b and the pair b, r have exactly the same greatest common positive divisor.

Let $D_1 = \{d \in \mathbb{Z} \mid d \mid a \text{ and } d \mid b\}$ and $D_2 = \{d \in \mathbb{Z} \mid d \mid b \text{ and } d \mid r\}$. We show that $D_1 = D_2$. First let $d \in D_1$ and show that $d \in D_2$. Since $d \in D_1$, we know $d \mid a$ and $d \mid b$. Thus, we may fix $m, n \in \mathbb{Z}$ such that $a = dn$ and $b = dm$. Then $r = a - bq = dn - dmq = d(n - mq)$, so $d \mid r$. Hence $d \in D_2$ and $D_1 \subseteq D_2$. The argument that $D_2 \subseteq D_1$ is similar, and we leave it to you to finish. □

Given $a, b \in \mathbb{Z}^+$, the Division Algorithm produces a remainder r with $0 \leq r < b$. If we repeatedly apply the Division Algorithm, starting with a and b to produce q and r, then Lemma 6.2.3 says that we can replace the question of computing (a, b) with computing (b, r), an "easier" gcd to compute since it involves smaller integers. We just need to argue that this process of repeatedly applying the Division Algorithm stops. For this, we will again need the Well-Ordering Principle 6.1.3.

Before giving a precise statement and proof of the Euclidean Algorithm, which we do at the end of this section, we will state the algorithm informally. We'll also give an example of how it is used to find the gcd of any two positive integers.

Euclidean Algorithm (Informal) 6.2.4.

(1) Given $a, b \in \mathbb{Z}^+$.

(2) If $b \mid a$, then $(a, b) = b$, and STOP.

(3) If $b \nmid a$, then use the Division Algorithm to find $q, r \in \mathbb{Z}$ such that $a = bq + r$, where $0 \leq r < b$. Note that $(a, b) = (b, r)$.

(4) Repeat from step (2), replacing a by b and b by r.

Intuitively we see that, given $a, b \in \mathbb{Z}^+$, the algorithm must halt. Otherwise, the sequence of remainders we obtain will form an infinite descending chain in $\mathbb{Z}^+$, contradicting the Well-Ordering Principle.

Example 6.2.5. Find $(1962, 924)$.

We use the Division Algorithm (Theorem 6.1.1) with $a = 1962$ and $b = 924$ to obtain $1962 = 924 \cdot 2 + 114$; i.e., $q = q_1 = 2$ and $r = r_1 = 114$.

We then repeat, this time taking $b = 924$ in place of a and $r = 114$ in place of b to obtain $924 = 114 \cdot 8 + 12$; i.e., we now have $q_2 = 8$ and $r_2 = 12$. We continue until we reach a remainder of 0; this is most easily followed as a sequence of Division Algorithm computations:

$$1962 = 924 \cdot 2 + 114, \tag{6.1}$$
$$924 = 114 \cdot 8 + 12, \tag{6.2}$$
$$114 = 12 \cdot 9 + 6, \tag{6.3}$$
$$12 = 6 \cdot 2 + 0. \tag{6.4}$$

By repeated application of Lemma 6.2.3, note that

$$(1962, 924) = (924, 114) = (114, 12) = (12, 6) = (6, 0) = 6.$$

This process shows that the last nonzero remainder we obtain in the repeated computations above is the gcd of the two numbers we began with; here $(1962, 924) = 6$.

We can also use this process to show that 6, the gcd of 1962 and 924, can be expressed explicitly in terms of 1962 and 924 in a particularly useful way. We will show that we can find $x, y \in \mathbb{Z}$ such that $6 = 1962x + 924y$. This fact, that the gcd 6 is an integer "linear combination" of 1962 and 924, is extremely important in number theory. We obtain x and y by running the Euclidean Algorithm backwards. First rewrite the sequence of computations (6.3), (6.2), and (6.1) as follows:

$$\begin{aligned} 6 &= 114 - 12 \cdot 9 = 114 + 12(-9), \\ 12 &= 924 - 114 \cdot 8 = 924 + 114(-8), \\ 114 &= 1962 - 924 \cdot 2 = 1962 + 924(-2). \end{aligned}$$

Then substitute:

$$\begin{aligned} 6 &= 114 + 12(-9) \\ &= 114 + (924 + 114(-8))(-9) \\ &= 114 + 924(-9) + 114(72) \\ &= 114(73) + 924(-9) \\ &= (1962 + 924(-2))(73) + 924(-9) \\ &= 1962(73) + 924(-146) + 924(-9) \\ &= 1962(73) + 924(-155). \end{aligned}$$

So, $6 = 1962x + 924y$, where $x = 73$ and $y = -155$ (you should confirm this statement using your calculator). ♢

We next formally define the terminology used in Example 6.2.5.

Definition 6.2.6. Let $a, b, n \in \mathbb{Z}$. The integer n is a *linear combination* of a and b if there exist $x, y \in \mathbb{Z}$ such that $n = ax + by$.

Although the Euclidean Algorithm can be used to find the greatest common divisors of positive integers only, it is easy to adapt when one or both of the integers are negative, by the following fact.

Proposition 6.2.7. *Let $a, b \in \mathbb{Z}$ such that not both a and b equal 0. Then $(a, b) = (|a|, |b|)$.*

Proof. Exercise 6.2.2. □

Thus $(-1962, 924) = (1962, 924) = 6$ by our previous computation. We can rewrite the linear combination found in Example 6.2.5 as

$$\begin{aligned} 6 &= 1962(73) + 924(-155) \\ &= -1962(-73) + 924(-155) \end{aligned}$$

to write $(-1962, 924)$ as a linear combination of -1962 and 924.

Corollary 6.2.8. *Let $a, b \in \mathbb{Z}$ such that not both a and b equal 0, and let $d = (a, b)$. Then d is a linear combination of a and b.*

Proof (Informal Sketch). Informally (for $a \geq b > 0$), repeated back substitution using the computations that arise when the Euclidean Algorithm (Theorem 6.2.9) is used to compute (a, b). Formally, one proves by strong induction on k that all the remainders r_k generated by the proof of the Euclidean Algorithm are linear combinations of a and b. See Exercise 6.2.7. □

6.2.1. The Euclidean Algorithm, more formally.

The precise statement of the Euclidean Algorithm is notationally complicated. Students who are encountering these ideas for the first time may wish to omit the statement and proof of the Euclidean Algorithm.

To see how to formalize the Euclidean Algorithm, note that we can view the computations (6.1)–(6.4) as follows:

$$\begin{aligned} a &= bq_1 + r_1, \\ b &= r_1 q_2 + r_2, \\ r_1 &= r_2 q_3 + r_3, \\ r_2 &= r_3 q_4 + r_4, \end{aligned}$$

where $a = 1962$, $b = 924$, and

$$\begin{aligned} q_1 &= 2, & r_1 &= 114, \\ q_2 &= 8, & r_2 &= 12, \\ q_3 &= 9, & r_3 &= 6, \\ q_4 &= 2, & r_4 &= 0. \end{aligned}$$

So, we are really asserting the existence of a "list of quotients" q_1, q_2, q_3, q_4 and a "list of remainders" r_1, r_2, r_3, r_4 which strictly decrease to $r_4 = 0$, where the desired gcd is the last nonzero remainder r_3. In fact, the list of remainders starts earlier: defining $r_0 = b$ and $r_{-1} = a$ is consistent with the pattern we see here since

$$\begin{aligned} r_0 &= b = r_1 q_2 + r_2, \\ r_{-1} &= a = r_0 q_1 + r_1. \end{aligned}$$

Theorem 6.2.9 (Euclidean Algorithm)**.** *Let $a, b \in \mathbb{Z}^+$ with $a \geq b$ and let $d = (a, b)$. Then there exist finite lists of integers $q_1, q_2, \dots, q_{n+1}$ and $r_{-1}, r_0, r_1, r_2, \dots, r_{n+1}$, where $n \geq 0$, such that $a = r_{-1} \geq b = r_0 > r_1 > r_2 > \cdots > r_n > r_{n+1} = 0$,*

$$\begin{aligned} a &= bq_1 + r_1, \\ b &= r_1 q_2 + r_2, \\ r_1 &= r_2 q_3 + r_3, \\ &\vdots \\ r_{n-2} &= r_{n-1} q_n + r_n, \\ r_{n-1} &= r_n q_{n+1} + r_{n+1}, \end{aligned}$$

and $d = r_n$.

Proof. Informally, the Euclidean Algorithm follows from repeated applications of the Division Algorithm (Theorem 6.1.1) and Lemma 6.2.3. Formally, the proof requires induction or the Well-Ordering Principle.

Let $a, b \in \mathbb{Z}^+$ with $a \geq b$, and let $r_{-1} = a$ and $r_0 = b$. We define the two lists of integers by recursion, using the Division Algorithm to obtain:

$$\begin{array}{lll} r_1 \text{ and } q_1 \text{ with} & a = r_{-1} = r_0 q_1 + r_1 = bq_1 + r_1, & \text{where } 0 \leq r_1 < r_0, \\ r_2 \text{ and } q_2 \text{ with} & b = r_0 = r_1 q_2 + r_2, & \text{where } 0 \leq r_2 < r_1, \\ r_3 \text{ and } q_3 \text{ with} & r_1 = r_2 q_3 + r_3, & \text{where } 0 \leq r_3 < r_2, \end{array}$$

and in general, for $k \geq 1$,

$$r_k \text{ and } q_k \text{ with} \qquad r_{k-2} = r_{k-1} q_k + r_k, \qquad \text{where } 0 \leq r_k < r_{k-1},$$

until a remainder of 0 is obtained, i.e., until $n \geq 0$ is obtained with

$$r_{n-1} = r_n q_{n+1} + r_{n+1}, \qquad \text{where } r_{n+1} = 0.$$

This $n \geq 0$ exists, since if a remainder of 0 is never obtained, then $\{r_n \mid n \geq 0\}$ is a nonempty set of positive integers

$$\cdots < r_3 < r_2 < r_1 < r_0$$

with no least element, contradicting the Well-Ordering Principle 6.1.3.

Then by Lemma 6.2.3 we have

$$(a, b) = (b, r_1) = (r_1, r_2) = \cdots = (r_{n-1}, r_n) = (r_n, r_{n+1}) = (r_n, 0) = r_n.$$

To prove this rigorously, prove by induction on k that if $0 \leq k \leq n$, then $(r_{k-1}, r_k) = (r_k, r_{k+1})$. See Exercise 6.2.6. Hence $(a, b) = r_n$, the last nonzero remainder, as desired. □

Exercises 6.2

1. (a) Use the Euclidean Algorithm to find $(14670, 4257)$ and write $(14670, 4257)$ as an integer linear combination of 14670 and 4257.
   (b) Repeat part (a) for $(1207, 569)$.
   (c) Repeat part (a) for $(7776, 16650)$.
2. Prove Proposition 6.2.7.
3. Let $a, b \in \mathbb{Z}$ with a and b not both zero, and let $d = (a, b)$. Let

$$S = \{n \in \mathbb{Z} \mid (\exists x, y \in \mathbb{Z})[n = ax + by]\}$$

   (i.e., S is the set of all integer linear combinations of a and b), and let

$$T = \{n \in \mathbb{Z} \mid (\exists m \in \mathbb{Z})[n = dm]\}$$

   (i.e., T is the set of all integer multiples of d). Prove that $S = T$.
4. Let $a, b \in \mathbb{Z}$ with a and b not both zero, and let $d = (a, b)$. Prove that d is the least positive integer linear combination of a and b; i.e., for all $c \in \mathbb{Z}^+$, if c is an integer linear combination of a and b, then $d \leq c$.
5. One can define the gcd of two positive integers a and b strictly in terms of the divisibility relation $|$, which is useful since it generalizes to other "algebraic structures" without a linear order relation $\leq$.

   Definition 6.2.10. Let $a, b, d \in \mathbb{Z}^+$. We say that d is a *greatest common divisor* of a and b if
   (a) $d \mid a$ and $d \mid b$ and
   (b) for all $c \in \mathbb{Z}^+$, if $c \mid a$ and $c \mid b$, then $c \mid d$.

   Let $a, b, d \in \mathbb{Z}^+$.
   (a) Let $d = (a, b)$ be the gcd of a and b as defined in Definition 6.2.2. Use Corollary 6.2.8 to show that d is a gcd of a and b as defined in Definition 6.2.10. In other words, show that for all $c \in \mathbb{Z}^+$, if $c \mid a$ and $c \mid b$, then $c \mid d$. This proves that for all $a, b \in \mathbb{Z}^+$, an integer d satisfying Definition 6.2.10 exists.
   (b) Using Definition 6.2.10, prove that the greatest common divisor of a and b is unique. (**Hint:** Assume that for $a, b \in \mathbb{Z}^+$, we have $d_1, d_2 \in \mathbb{Z}^+$ both satisfying Definition 6.2.10, and prove that $d_1 = d_2$.)
   (c) Show that if the greatest common divisor (as defined in Definition 6.2.10) of a and b is d , then $d = (a, b)$ (as defined in Definition 6.2.2). In other words, show that for all $c \in \mathbb{Z}^+$, if $c \mid a$ and $c \mid b$, then $c \leq d$.
6. Fill in the missing induction argument in the proof of the Euclidean Algorithm (Theorem 6.2.9).
7. Prove Corollary 6.2.8.

6.3. Relatively prime integers and the Fundamental Theorem of Arithmetic

The result of Corollary 6.2.8 is particularly useful when the greatest common divisor of two integers is 1.

Definition 6.3.1. Let $a, b \in \mathbb{Z}$ such that not both a and b equal 0. If $(a, b) = 1$, then the integers a and b are called *relatively prime* (or *coprime*).

For example, 12 and 15 are not relatively prime, since $(12, 15) = 3$, while 24 and 35 are relatively prime, since $(24, 35) = 1$. By Corollary 6.2.8, we know that we can write 1 as a linear combination of 24 and 35: you can check that $1 = 24(-16) + 35(11)$. It turns out that this property characterizes when two integers are relatively prime.

Theorem 6.3.2. *Let $a, b \in \mathbb{Z}$ such that not both of a and b equal 0. Then a and b are relatively prime if and only if there exist $x, y \in \mathbb{Z}$ such that $ax + by = 1$.*

Proof. Let $a, b \in \mathbb{Z}$ such that not both of a and b equal 0.

($\Rightarrow$) Corollary 6.2.8.

($\Leftarrow$) Assume that we are given $x, y \in \mathbb{Z}$ such that $ax + by = 1$. We show that $(a, b) = 1$. To do this, assume that we have $d \in \mathbb{Z}^+$ such that $d \mid a$ and $d \mid b$. We must show that $d = 1$.

Since $d \mid a$ and $d \mid b$, we may fix $m, n \in \mathbb{Z}$ such that $a = dm$ and $b = dn$. Then we have $ax + by = 1$, so $dmx + dny = 1$, or $d(mx + ny) = 1$. Hence $d \mid 1$. Since $d > 0$, this implies that $d = 1$, as desired. Hence a and b are relatively prime. □

As already mentioned, Theorem 6.3.2 is a very useful tool when one knows that two integers are relatively prime. We'll use it to prove a result known as Euclid's Lemma, which implies an important fact about prime divisors of products. (See Definition 2.1.7, if necessary, for the definition of *prime* integer.)

Theorem 6.3.3 (Euclid's Lemma)**.** *Let $a, b, c \in \mathbb{Z}$. If $(a, b) = 1$ and $a \mid bc$, then $a \mid c$.*

Proof. Let $a, b, c \in \mathbb{Z}$. Assume that $(a, b) = 1$ and $a \mid bc$. We show that $a \mid c$.

Since $(a, b) = 1$, by Theorem 6.3.2 we may fix $x, y \in \mathbb{Z}$ with the property that $ax + by = 1$. Since $a \mid bc$, we may fix $m \in \mathbb{Z}$ such that $bc = am$. Then

$$(ax + by)c = axc + byc = c.$$

Substituting for bc gives $axc + amy = c$, or $a(xc + my) = c$. Thus $a \mid c$, as desired. □

Several important corollaries follow from Theorem 6.3.3, and they will lead us back to the Fundamental Theorem of Arithmetic.

Corollary 6.3.4. *Let $a, b \in \mathbb{Z}$ and $p \in \mathbb{Z}^+$ be prime. If $p \mid ab$, then $p \mid a$ or $p \mid b$.*

Proof. Exercise 6.3.1. □

Corollary 6.3.5. *Let $p \in \mathbb{Z}^+$ be prime, $r \in \mathbb{Z}^+$, and $b_1, b_2, \ldots, b_r \in \mathbb{Z}$ be such that $p \mid b_1 b_2 \cdots b_r$. Then there exists $i \in \mathbb{Z}^+$, $1 \leq i \leq r$, such that $p \mid b_i$.*

Proof. We prove for all primes $p \in \mathbb{Z}^+$ and all $b_1, b_2, \ldots, b_r \in \mathbb{Z}$,

$$p \mid b_1 b_2 \cdots b_r \implies \text{there exists } i \in \mathbb{Z}^+, 1 \leq i \leq r, \text{such that } p \mid b_i \tag{6.5}$$

by induction on r.

Base Case: Given $p \in \mathbb{Z}^+$ prime and $b_1 \in \mathbb{Z}$, statement (6.5) for $r = 1$ is immediate.

Inductive Step: Let $r \in \mathbb{Z}^+$. We assume the inductive hypothesis that for all primes $q \in \mathbb{Z}^+$ and all $c_1, c_2, \ldots, c_r \in \mathbb{Z}$,

$$q \mid c_1 c_2 \cdots c_r \implies \text{there exists } i \in \mathbb{Z}^+, 1 \leq i \leq r, \text{such that } q \mid c_i. \tag{6.6}$$

Next, let $p \in \mathbb{Z}^+$ be prime and $b_1, b_2, \ldots, b_r, b_{r+1} \in \mathbb{Z}$, and assume that

$$p \mid b_1 b_2 \cdots b_r b_{r+1}. \tag{6.7}$$

We prove that there exists $i \in \mathbb{Z}^+$, $1 \leq i \leq r+1$, such that $p \mid b_i$.

Note that $p \mid ab_{r+1}$, where $a = b_1 b_2 \cdots b_r$. By Corollary 6.3.4, $p \mid b_{r+1}$ or $b \mid a$. If $p \mid b_{r+1}$, then we're done, so assume $p \mid a$. Then $p \mid b_1 b_2 \cdots b_r$, so by the inductive hypothesis for $q = p$ and $c_j = b_j$, $1 \leq j \leq r$, there exists $i \in \mathbb{Z}^+$, $1 \leq i \leq r$, such that $p \mid b_i$, completing the Inductive Step.

Hence, by PMI, for all $r \in \mathbb{Z}^+$, if $p \in \mathbb{Z}^+$ is prime and $b_1, b_2, \ldots, b_r \in \mathbb{Z}$ are such that $p \mid b_1 b_2 \cdots b_r$, then there exists $i \in \mathbb{Z}^+$, $1 \leq i \leq r$, such that $p \mid b_i$. □

Note our careful use of variables in the previous induction proof. The statement to be proved by induction is a universal statement, and so the inductive hypothesis is also universal. We use different variables in the universal quantifiers in the inductive hypothesis to remind ourselves that statement (6.6) holds *for all* $q, c_1, c_2, \ldots, c_r$ and not just for the *specific* $p, b_1, b_2, \ldots, b_r, b_{r+1}$ in statement (6.7). In this case, we didn't need the full generality of the inductive hypothesis, but in other proofs we might (see the discussion regarding the proof of Theorem 8.2.9).

Corollary 6.3.6. *Let $r \in \mathbb{Z}^+$ and $p, q_1, q_2, \ldots, q_r \in \mathbb{Z}^+$ be prime with $p \mid q_1 q_2 \cdots q_r$. Then there exists $i \in \mathbb{Z}^+$, $1 \leq i \leq r$, such that $p = q_i$.*

Proof. Exercise 6.3.2. □

Corollary 6.3.6 is the missing tool that we need in order to prove the uniqueness part of the Fundamental Theorem of Arithmetic (Theorem 2.3.3). Note that this proof is a much more sophisticated proof by induction than we have considered so far. The proof of Corollary 6.3.5 is a warm-up induction proof.

Theorem 6.3.7 (Fundamental Theorem of Arithmetic (Uniqueness))**.** *Every positive integer greater than* 1 *can be written as a product of primes. Furthermore, this product of primes is unique, except for the order in which the factors appear.*

Proof (Uniqueness). We first restate the uniqueness statement by taking advantage of notation, which will conveniently set up the induction argument:

Fundamental Theorem of Arithmetic (Uniqueness). For all $r \in \mathbb{Z}^+$, for all $s \in \mathbb{Z}^+$ with $r \leq s$,

(6.8) for all primes $p_1 \leq p_2 \leq \cdots \leq p_r, q_1 \leq q_2 \leq \cdots \leq q_s$,

if $p_1p_2 \cdots p_r = q_1q_2 \cdots q_s$, then $r = s$ and for all $i \in \mathbb{Z}^+, 1 \leq i \leq r, p_i = q_i$.

Note that this phrasing of the Fundamental Theorem of Arithmetic says that the factorization of any fixed integer $n > 1$ as a product of primes written in nondecreasing order is unique.

We prove this by induction on r (i.e., by induction on the length of the factorization).

Base Case: Assume we have $s \in \mathbb{Z}^+$ and primes $p_1, q_1, q_2, \ldots, q_s \in \mathbb{Z}^+$ such that $q_1 \leq q_2 \leq \cdots \leq q_s$ and

$$n = p_1 = q_1q_2 \cdots q_s.$$

We show that $s = 1$ and $q_1 = p_1$.

Let $i \in \mathbb{Z}^+$ with $1 \leq i \leq s$. Note that $q_i \mid p_1$. Hence, since q_i and p are prime, $q_i = p_1$ by Corollary 6.3.6. Thus, since i is arbitrary, $p_1 = p_1^s$. Since $s \geq 2$ gives a contradiction, $s = 1$. Hence $n = p_1 = q_1$.

Inductive Step: Let $r \in \mathbb{Z}^+$. We assume the result for r (i.e., we assume that for all $s \in \mathbb{Z}^+$ with $r \leq s$, statement (6.8) is true), and we prove it for $r + 1$.

Let $t \in \mathbb{Z}$ with $r + 1 \leq t + 1$ (i.e., $r \leq t$), and assume we have primes

$$P_1 \leq P_2 \leq \cdots \leq P_r \leq P_{r+1} \text{ and } Q_1 \leq Q_2 \leq \cdots \leq Q_t \leq Q_{t+1}$$

such that

$$n = P_1P_2 \cdots P_rP_{r+1} = Q_1Q_2 \cdots Q_tQ_{t+1}.$$

Then $P_{r+1} \mid Q_1Q_2 \cdots Q_tQ_{t+1}$, so by Corollary 6.3.6, we may fix $i \in \mathbb{Z}^+$, $1 \leq i \leq t + 1$, such that $P_{r+1} = Q_i$. Note then that $P_{r+1} \leq Q_{t+1}$. Reasoning analogously, $Q_{t+1} \mid P_1P_2 \cdots P_rP_{r+1}$, so $Q_{t+1} \leq P_{r+1}$. Thus $P_{r+1} = Q_{t+1}$.

By cancellation in $\mathbb{Z}$, we have

$$P_1P_2 \cdots P_r = Q_1Q_2 \cdots Q_t.$$

Hence by the induction hypothesis for t and $P_1, \ldots, P_r$, $Q_1, \ldots Q_t$, we conclude that $r = t$, and for all $i \in \mathbb{Z}^+$, $1 \leq i \leq r$, $P_i = Q_i$; i.e., $r + 1 = t + 1$ and for all $i \in \mathbb{Z}^+$, $1 \leq i \leq r + 1$, $P_i = Q_i$, completing the Inductive Step.

Hence, by PMI, statement (6.8) is true for all $r, s \in \mathbb{Z}^+$ with $r \leq s$; i.e., the factorization of any fixed integer $n > 1$ as a product of primes written in nondecreasing order is unique. □

It is also possible to prove Theorem 6.3.7 by strong induction on the positive integer being factored, rather than on the length of the factorization; see Exercise 6.3.12.

Exercises 6.3

1. Prove Corollary 6.3.4: for all $a, b \in \mathbb{Z}$, for all $p \in \mathbb{Z}^+$, if p is prime and $p \mid ab$, then $p \mid a$ or $p \mid b$. Prove that this result does not hold in general, when p is not prime.
2. Prove Corollary 6.3.6.
3. Let $a, b \in \mathbb{Z}$ with a and b not both zero. Prove that if $d = (a, b)$, then $(\frac{a}{d}, \frac{b}{d}) = 1$. (Note that if $d = (a, b)$, then $\frac{a}{d}, \frac{b}{d} \in \mathbb{Z}$.)
4. *Without using the Fundamental Theorem of Arithmetic*, prove that for all $a, b, c \in \mathbb{Z}$, if $a \mid c$ and $b \mid c$ and $(a, b) = 1$, then $ab \mid c$. Prove that this result does not hold in general when $(a, b) \neq 1$.
5. *Using the Fundamental Theorem of Arithmetic*, prove that for all $a, b, c \in \mathbb{Z}$, if $a \mid c$ and $b \mid c$ and $(a, b) = 1$, then $ab \mid c$. Prove that this result does not hold in general when $(a, b) \neq 1$.
6. Let $p \in \mathbb{Z}^+$ be prime and let $i \in \mathbb{Z}^+$ be such that $0 < i < p$. Prove that p divides the binomial coefficient $\binom{p}{i}$. (See Exercise 3.1.18.)
7. Let $a, b \in \mathbb{Z}$ such that $(a, b) = 1$. Prove that $(a + b, a - b) = 1$ or 2.
8. Let $a, b, c \in \mathbb{Z}$. Prove that $(a, c) = (b, c) = 1$ if and only if $(ab, c) = 1$.
9. Let $a, b \in \mathbb{Z}$. Prove that $(a, b) = 1$ if and only if $(ab, a + b) = 1$.
10. (Rational Root Theorem) Let $r, s \in \mathbb{Z}$ with $s \neq 0$ and $(r, s) = 1$, and let p be a polynomial with integer coefficients; i.e., $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, where $a_0, a_1, \ldots, a_n \in \mathbb{Z}$. Prove that if $\frac{r}{s}$ is a rational root of p, i.e., $p(\frac{r}{s}) = 0$, then $r \mid a_0$ and $s \mid a_n$.
11. Note that by the Fundamental Theorem of Arithmetic (Theorem 6.3.7), every positive integer n greater than 1 has a unique *prime-power factorization*; i.e., every integer $n > 1$ can be written uniquely in the form
$$n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r},$$
where $p_1 < p_2 < \cdots < p_r$ are prime and $k_i \in \mathbb{Z}^+$ for all $1 \leq i \leq r$.
   (a) Let $a, b \in \mathbb{Z}^+$ with $a < b$ and let $p_1, p_2, \ldots, p_r$ be the list of all distinct primes that are factors of a *or* b.
      (i) Explain why there exist $i_1, i_2, \ldots, i_r, j_1, j_2, \ldots, j_r \in \mathbb{Z}^{\geq 0}$ such that
$$a = p_1^{i_1} p_2^{i_2} \cdots p_r^{i_r} \quad \text{and} \quad b = p_1^{j_1} p_2^{j_2} \cdots p_r^{j_r}.$$
      (ii) Prove that
$$(a, b) = p_1^{\min\{i_1, j_1\}} p_2^{\min\{i_2, j_2\}} \cdots\ p_r^{\min\{i_r, j_r\}},$$
where $\min\{m, n\}$ denotes the minimum of the integers m and n.
   (b) Prove that all the exponents in the prime-power factorization of the integer $n > 1$ are even if and only if n is a perfect square (i.e., iff there exists $m \in \mathbb{Z}^+$ such that $m^2 = n$).
12. Provide another proof of the uniqueness part of the Fundamental Theorem of Arithmetic; i.e., prove using strong induction on n that for all $n \in \mathbb{Z}^+$ with $n \geq 2$, n can be written uniquely as a product of primes.

6.4. Congruences

Another important topic in number theory is the notion of "congruence modulo m". In fact, this is a notion with which you are already familiar, since we "tell time" modulo 12 (or 24).

Definition 6.4.1. Let $a, b \in \mathbb{Z}$, and let $m \in \mathbb{Z}^+$. The integers a and b are *congruent modulo* m, written $a \equiv b \mod m$, if $m \mid (a - b)$.

For the remainder of this chapter, m will always denote a fixed positive integer greater than 1 (since congruence modulo 1 isn't interesting). We first consider some examples of this concept and then consider its basic properties.

Example 6.4.2.

(1) $41 \equiv 5 \mod 12$, since $41 - 5 = 36$ and $12 \mid 36$.

(2) $-15 \equiv 13 \mod 4$, since $-15 - 13 = -28$ and $4 \mid -28$.

(3) $25 \not\equiv 12 \mod 7$, since $25 - 12 = 13$ and $7 \nmid 13$. ♢

The next proposition, which gives another way to think of congruence mod m, is easy to prove.

Proposition 6.4.3. *Let $a, b \in \mathbb{Z}$. Then*

$$a \equiv b \mod m \Leftrightarrow (\exists q \in \mathbb{Z})[a = mq + b].$$

Proof. Exercise 6.4.1. □

Note that congruence mod m has "equality-like" properties; namely, it is "reflexive", "symmetric", and "transitive".

Theorem 6.4.4.

(1) *For all $a \in \mathbb{Z}$, $a \equiv a \mod m$ (i.e., congruence modulo m is "reflexive").*

(2) *For all $a, b \in \mathbb{Z}$, if $a \equiv b \mod m$, then $b \equiv a \mod m$ (i.e., congruence modulo m is "symmetric").*

(3) *For all $a, b, c \in \mathbb{Z}$, if $a \equiv b \mod m$ and $b \equiv c \mod m$, then $a \equiv c \mod m$ (i.e., congruence modulo m is "transitive").*

Proof. Let $a, b, c \in \mathbb{Z}$.

(1) To show that $a \equiv a \mod m$, note that $a - a = 0$, and hence $m \mid a - a$. Thus $a \equiv a \mod m$ by Definition 6.4.1.

(2) Assume that $a \equiv b \mod m$. We must show that $b \equiv a \mod m$. Since $a \equiv b \mod m$, by Definition 6.4.1 we know that $m \mid (a - b)$. Thus we may fix $\ell \in \mathbb{Z}$ such that $a - b = m\ell$. But then $b - a = m(-\ell)$, so $m \mid (b - a)$ and $b \equiv a \mod m$ by Definition 6.4.1.

(3) Assume that $a \equiv b \mod m$ and $b \equiv c \mod m$. We must show that $a \equiv c \mod m$. Since $a \equiv b \mod m$ and $b \equiv c \mod m$, by Definition 6.4.1 we know that $m \mid (a-b)$ and $m \mid (b-c)$. Thus we may fix $k, \ell \in \mathbb{Z}$ such that $a-b = mk$ and $b-c = m\ell$. Then

$$\begin{aligned} (a-b)+(b-c) &= mk + m\ell, \quad \text{so} \\ a - c &= m(k+\ell). \end{aligned}$$

Thus $m \mid (a-c)$ and hence $a \equiv c \mod m$, by Definition 6.4.1. □

The Division Algorithm (Theorem 6.1.1) implies that every integer is congruent modulo m to a unique nonnegative remainder $r < m$ upon division by m.

Theorem 6.4.5. *Let $a, b \in \mathbb{Z}$.*

(1) *There is a unique $r \in \mathbb{Z}$ such that $0 \le r < m$ and $a \equiv r \mod m$.*

(2) *The congruence $a \equiv b \mod m$ holds if and only if there exists $r \in \mathbb{Z}$ with $0 \le r < m$ such that $a \equiv r \mod m$ and $b \equiv r \mod m$ (i.e., a and b have the same nonnegative remainder $r < m$ when divided by m).*

Proof. Let $a, b \in \mathbb{Z}$.

(1) By the Division Algorithm (Theorem 6.1.1), there exists a unique remainder $r \in \mathbb{Z}$, $0 \le r < m$, such that

$$(\exists q \in \mathbb{Z})[a = mq + r].$$

By Proposition 6.4.3, it follows that there exists a unique remainder $r \in \mathbb{Z}$, $0 \le r < m$, such that $a \equiv r \mod m$.

(2) We must prove the biconditional.

($\Rightarrow$) Assume $a \equiv b \mod m$. By the Division Algorithm (Theorem 6.1.1), we may fix integers q_1, q_2, r_1, r_2 such that $a = mq_1 + r_1$, $b = mq_2 + r_2$, and also $0 \le r_1, r_2 < m$. Then $r_1 \equiv a \mod m$ by Proposition 6.4.3 and the symmetric property Theorem 6.4.4(2). Hence $r_1 \equiv b \mod m$, since $r_1 \equiv a \mod m$ and $a \equiv b \mod m$, by the transitive property Theorem 6.4.4(3). Thus $r_1 \equiv r_2 \mod m$, again by transitivity, since $b \equiv r_2 \mod m$. Since $0 \le r_1, r_2 < m$ and $r_1 \equiv r_1 \mod m$, we have $r_1 = r_2$ by part (1) of this theorem.

($\Leftarrow$) Assume we have $r \in \mathbb{Z}$, $0 \le r < m$, such that $a \equiv r \mod m$ and also $b \equiv r \mod m$. Then $a \equiv b \mod m$ by the symmetric and transitive properties in Theorem 6.4.4. □

Theorem 6.4.5 says, for example, that each integer a is congruent to exactly one of 0, 1, 2, or 3 modulo 4. This is simply a restatement of the fact that, by the Division Algorithm, each integer a takes exactly one of the following four forms: $a = 4k$, $a = 4k+1$, $a = 4k+2$, or $a = 4k+3$, where $k \in \mathbb{Z}$. By computing modulo m, we can greatly simplify the proofs we saw in Section 6.1.

Thus, we wish to define "arithmetic modulo m". The next theorem shows that all but one of the arithmetic operations behave as expected.

Theorem 6.4.6. *Let $a_1, a_2, b_1, b_2, c \in \mathbb{Z}$ and assume that $a_1 \equiv a_2 \mod m$ and $b_1 \equiv b_2 \mod m$. Then:*

(1) $a_1 + b_1 \equiv a_2 + b_2 \mod m$.

(2) $a_1 - b_1 \equiv a_2 - b_2 \mod m$.

(3) $a_1 + c \equiv a_2 + c \mod m$.

(4) $a_1 b_1 \equiv a_2 b_2 \mod m$.

(5) $a_1 c \equiv a_2 c \mod m$.

Proof. We prove (4) and leave the rest to Exercise 6.4.7. See also Exercises 6.4.4 and 6.4.8.

Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$. Assume that $a_1 \equiv a_2 \mod m$ and $b_1 \equiv b_2 \mod m$. We prove that $a_1 b_1 \equiv a_2 b_2 \mod m$. By Definition 6.4.1, we know that $m \mid (a_1 - a_2)$ and $m \mid (b_1 - b_2)$, so we fix $k, \ell \in \mathbb{Z}$ such that $a_1 - a_2 = mk$ and $b_1 - b_2 = m\ell$. We must show that $m \mid (a_1 b_1 - a_2 b_2)$. Note that

$$\begin{aligned} a_1 b_1 - a_2 b_2 &= a_1 b_1 - a_1 b_2 + a_1 b_2 - a_2 b_2 \\ &= a_1(b_1 - b_2) + b_2(a_1 - a_2) \\ &= a_1 m\ell + b_2 mk \\ &= m(a_1 \ell + b_2 k). \end{aligned}$$

Hence, $m \mid (a_1 b_1 - a_2 b_2)$ and so $a_1 \equiv a_2 \mod m$, as desired. □

The notion of congruences can greatly simplify "divisibility" proofs, as well as proofs using the Division Algorithm. As an example, we'll provide another solution to the problem from Example 6.1.4.

Example 6.4.7. Prove that the square of any integer has one of the forms $3k$ or $3k + 1$, where $k \in \mathbb{Z}$.

Proof. By Proposition 6.4.3, it suffices to prove that the square of any integer must be congruent to 0 or 1 modulo 3. Let $x \in \mathbb{Z}$. As before, we consider 3 cases, and we use Theorem 6.4.6 to compute.

Case I: $x \equiv 0 \mod 3$.
Then $x^2 \equiv 0^2 \equiv 0 \mod 3$.

Case II: $x \equiv 1 \mod 3$.
Then $x^2 \equiv 1^2 \equiv 1 \mod 3$.

Case III: $x \equiv 2 \mod 3$.
Then $x^2 \equiv 2^2 \equiv 4 \equiv 1 \mod 3$.

Hence for any $x \in \mathbb{Z}$, x^2 is congruent to 0 or 1 modulo 3. □ ◊

Exercises 6.4

1. Prove Proposition 6.4.3.
2. Let $a, b \in \mathbb{Z}$ and $m, n \in \mathbb{Z}^+$. Prove that if $a \equiv b \mod m$ and $n \mid m$, then $a \equiv b \mod n$.

3. Let $a, b \in \mathbb{Z}$ and $m \in \mathbb{Z}^+$. Prove that if $a \equiv b \mod m$, then $(a, m) = (b, m)$. (Hint: Use the same method that was used to prove Lemma 6.2.3.)
4. Give an example to show that $a \neq 0$ and $ab \equiv ac \mod m$ need not imply that $b \equiv c \mod m$. Give an example to show that $a^2 \equiv b^2 \mod m$ need not imply that $a \equiv b \mod m$.
5. Use congruences to prove that for all integers a, $a^3 \equiv 0, 1,$ or $6 \mod 7$.
6. Redo Exercises 6.1.1, 6.1.2, and 6.1.3 using congruences in place of the Division Algorithm.
7. Complete the proof of Theorem 6.4.6.
8. Let $a, b, c \in \mathbb{Z}$ and $m \in \mathbb{Z}^+$. Prove that if $(c, m) = 1$ and $ac \equiv bc \mod m$, then $a \equiv b \mod m$.
9. Let $p \in \mathbb{Z}^+$ be prime and let $a, b \in \mathbb{Z}$. Prove that $(a + b)^p \equiv a^p + b^p \mod p$. (**HINT:** See Exercise 3.1.19.)
10. Let $a, b \in \mathbb{Z}$ and $m \in \mathbb{Z}^+$.
    (a) Prove that the congruence $ax \equiv b \mod m$ has a solution if and only if $d \mid b$, where $d = (a, m)$.
    (b) Solve the following congruences.
        (i) $3x \equiv 5 \mod 7$.
        (ii) $12x \equiv 5 \mod 9$.
11. Let $n \in \mathbb{Z}^+$ be such that $n \geq 2$, and assume that $(n - 1)! \equiv -1 \mod n$. Prove that n is prime.
12. (a) Let $p \in \mathbb{Z}^+$ be prime. Prove that for all $x \in \mathbb{Z}$, $x^2 \equiv 1 \mod p$ if and only if $x \equiv 1 \mod p$ or $x \equiv -1 \mod p$.
    (b) Let $p > 3$ be prime and let $U = \{2, 3, \ldots, p-2\}$. Prove that if $a \in U$, then there exists a unique $x \in U$ such that $ax \equiv 1 \mod p$. (See Exercise 6.4.10.)
    (c) (Wilson's Theorem) Prove that for all primes p, $(p - 1)! \equiv -1 \mod p$. (**HINT:** First try an example where $p = 13$, to see how parts (a) and (b) are relevant.)
13. In this exercise, we consider the following theorem of Dirichlet.

    Theorem 6.4.8 (Dirichlet's Theorem on Primes in Arithmetic Progressions (1837))**.** *Let $a, m \in \mathbb{Z}^+$ with $(a, m) = 1$. Then there exist infinitely many primes p such that $p \equiv a \mod m$; i.e., there are infinitely many primes p of the form $mk + a$, where $k \in \mathbb{Z}^{\geq 0}$.*

    A proof of Dirichlet's Theorem in its full generality uses methods beyond the scope of this textbook, but it is possible to prove some special cases of this theorem using the methods we have discussed.

    (a) Prove that there exist infinitely many primes congruent to 3 modulo 4; i.e., there exist infinitely many primes of the form $4k + 3$, where $k \in \mathbb{Z}^{\geq 0}$. (**HINT:** Mimic the proof of Euclid's Theorem that there exist infinitely many primes (Theorem 2.3.4). Assume that $p_1 < p_2 < \cdots < p_n$, where $n \in \mathbb{Z}^+$, is a complete and increasing list of the primes congruent to 3 modulo 4, and consider the positive integer $n = 4p_1 \cdot p_2 \cdot \cdots \cdot p_n - 1$.)
    (b) Prove that there exist infinitely many primes congruent to 5 modulo 6; i.e., there exist infinitely many primes of the form $6k + 5$, where $k \in \mathbb{Z}^{\geq 0}$.

(c) Why does adapting your proofs of the first two parts of this problem fail when trying to prove that there are infinitely many primes congruent to 7 modulo 8, i.e., infinitely many primes of the form $8k+7$, where $k \in \mathbb{Z}^{\geq 0}$?

6.5. Congruence classes

Theorem 6.4.5 essentially says that when $a, b \in \mathbb{Z}$ with $a \equiv b \mod m$, then a and b are the "same", since they are congruent modulo m to the same nonnegative remainder $r < m$. It will be convenient, therefore, to "lump together" into a single set integers which are the "same" modulo m and then treat that set as a single mathematical object. This is a common mathematical technique in abstract algebra, and we consider the general idea in Chapter 7.

Definition 6.5.1. Let $a \in \mathbb{Z}$. The *congruence class of a modulo m* is the set

$$[a]_m = \{x \in \mathbb{Z} \mid x \equiv a \mod m\}.$$

Example 6.5.2. Note that $27 \equiv -13 \mod 4$ (since $27 - (-13) = 40$), so $27 \in [-13]_4$. By symmetry, $-13 \equiv 27 \mod 4$, so $-13 \in [27]_4$.

Since $27 \not\equiv 2 \mod 4$ (i.e., $27 - 2 = 25$ is not divisible by 4), $27 \notin [2]_4$. ♢

Example 6.5.3. Continuing to work modulo 4, note that

$$\begin{aligned}[0]_4 &= \{x \in \mathbb{Z} \mid x \equiv 0 \mod 4\} = \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k]\} \\ &= \{\dots, -8, -4, 0, 4, 8, \dots\}.\end{aligned} \tag{6.9}$$

In fact (this requires proof; see Theorem 6.5.5(2)), $[0]_4 = [4]_4 = [-8]_4 = \cdots$; i.e., the congruence class $[0]_4$ has many different "names".

We can continue in this way to find all of the congruence classes mod 4.

$$\begin{aligned}[1]_4 &= \{x \in \mathbb{Z} \mid x \equiv 1 \mod 4\} = \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k+1]\} \\ &= \{\dots, -7, -3, 1, 5, 9, \dots\} \\ &= [9]_4 = [-7]_4 = \cdots.\end{aligned} \tag{6.10}$$

$$\begin{aligned}[2]_4 &= \{x \in \mathbb{Z} \mid x \equiv 2 \mod 4\} = \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k+2]\} \\ &= \{\dots, -6, -2, 2, 6, 10, \dots\} \\ &= [10]_4 = [-6]_4 = \cdots.\end{aligned} \tag{6.11}$$

$$\begin{aligned}[3]_4 &= \{x \in \mathbb{Z} \mid x \equiv 3 \mod 4\} = \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k+3]\} \\ &= \{\dots, -5, -1, 3, 7, 11, \dots\} \\ &= [7]_4 = [-5]_4 = \cdots.\end{aligned} \tag{6.12}$$

♢

Equations (6.9), (6.10), (6.11), and (6.12), in conjunction with the Division Algorithm (Theorem 6.1.1), tell us that $[0]_4$, $[1]_4$, $[2]_4$, $[3]_4$ are the only congruence classes modulo 4. In general, if $m \in \mathbb{Z}^+$, then $[0]_m$, $[1]_m$, …, $[m-1]_m$ are the only congruence classes modulo m.

Definition 6.5.4. The set of *integers modulo m*, denoted $\mathbb{Z}_m$, is the set

$$\mathbb{Z}_m = \{[0]_m, [1]_m, \dots, [m-1]_m\}.$$

Note that while there is a similarity between the set $\{0, 1, 2, 3\}$ and the set $\mathbb{Z}_4 = \{[0]_4, [1]_4, [2]_4, [3]_4\}$, they are not equal as sets. While $\{0, 1, 2, 3\}$ is a set of integers, $\mathbb{Z}_4$ is a set of *sets* of integers.

Furthermore, Theorem 6.4.5 tells us (since every integer is congruent to exactly one element of the set $\{0, 1, 2, 3\}$ modulo 4) that the congruence classes modulo 4 "partition" $\mathbb{Z}$ into four pairwise disjoint sets, i.e., that each integer is in *exactly one* of these congruence classes (see Definition 7.3.1). These facts are summarized in Theorem 6.5.5. See Figure 6.1.

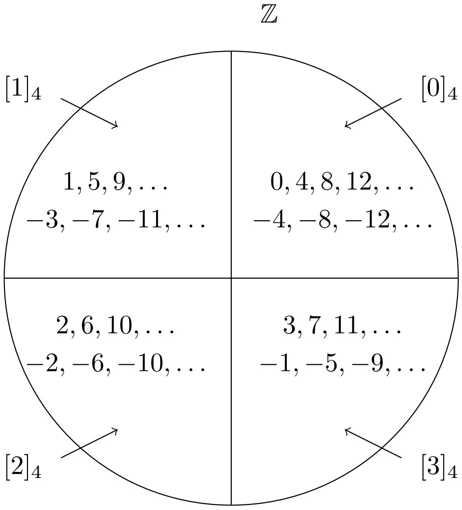


Figure 6.1. $\mathbb{Z}_4$.

Theorem 6.5.5 (Congruence classes modulo m form a "partition" of $\mathbb{Z}$)**.**

(1) *For all* $a \in \mathbb{Z}$, $a \in [a]_m$.

(2) *For all* $a, b \in \mathbb{Z}$, $a \equiv b \mod m$ *iff* $[a]_m = [b]_m$.

(3) *For all* $a, b \in \mathbb{Z}$, $a \not\equiv b \mod m$ *iff* $[a]_m \cap [b]_m = \emptyset$.

Proof. Exercise 6.5.1. (We are relegating the proof of this important theorem to the exercises, since we will prove the more general result Theorem 7.2.9, which implies this one.) □

Just as $\mathbb{Z}$ is an "algebraic structure" (i.e., a set equipped with an algebraic operation $+$), we can make $\mathbb{Z}_m$ into an algebraic structure by defining an addition $+_m$ of congruence classes modulo m, using arithmetic modulo m. For example, it is reasonable to expect that $[2]_4 +_4 [3]_4$ should equal $[2+3]_4 = [5]_4 = [1]_4$. However, there is a potential problem here, since every congruence class has "infinitely many names". For example, since $[10]_4 = [2]_4$ and $[-1]_4 = [3]_4$, we need to be sure that

$$[10]_4 +_4 [-1]_4 = [9]_4 = [1]_4 = [5]_4 = [2]_4 +_4 [3]_4.$$

Theorem 6.4.6 and Theorem 6.5.5 can be used to show that this arithmetic of congruence classes is "well-defined" (see Subsection 6.5.1), in the following sense.

Theorem 6.5.6. *Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ and assume that $[a_1]_m = [a_2]_m$ and $[b_1]_m = [b_2]_m$. Then:*

(1) $[a_1 + b_1]_m = [a_2 + b_2]_m$.

(2) $[a_1 - b_1]_m = [a_2 - b_2]_m$.

(3) $[a_1 b_1]_m = [a_2 b_2]_m$.

Scratchwork. Let's examine the Given-Goal diagram for part (1).

Given	**Goal**
$a_1, a_2, b_1, b_2 \in \mathbb{Z}$	
$[a_1]_m = [a_2]_m$	
$[b_1]_m = [b_2]_m$	$[a_1 + b_1]_m = [a_2 + b_2]_m$

By Theorem 6.5.5(2), we may rewrite the Given-Goal diagram as follows.

Given	**Goal**
$a_1, a_2, b_1, b_2 \in \mathbb{Z}$	
$a_1 \equiv a_2 \mod m$	
$b_1 \equiv b_2 \mod m$	$a_1 + b_1 \equiv a_2 + b_2 \mod m$

But this is just the Given-Goal diagram for the proof of Theorem 6.4.6(1). We'll include the proof of this statement, for completeness.

Proof. We prove (1) and leave the rest to Exercise 6.5.2.

Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ and assume that $[a_1]_m = [a_2]_m$ and $[b_1]_m = [b_2]_m$. We prove that $[a_1 + b_1]_m = [a_2 + b_2]_m$.

Since $[a_1]_m = [a_2]_m$, by Theorem 6.5.5(2), we know that $a_1 \equiv a_2 \mod m$. Thus, by Proposition 6.4.3, we may fix $q_1 \in \mathbb{Z}$ such that $a_1 = mq_1 + a_2$. Similarly, since $[b_1]_m = [b_2]_m$, we know that $b_1 \equiv b_2 \mod m$. Hence, we may fix $q_2 \in \mathbb{Z}$ such that $b_1 = mq_2 + b_2$. Then

$$a_1 + b_1 = (mq_1 + a_2) + (mq_2 + b_2) = m(q_1 + q_2) + (a_2 + b_2).$$

Thus, again by Proposition 6.4.3, $a_1 + b_1 \equiv a_2 + b_2 \mod m$, and so by Theorem 6.5.5(2), $[a_1 + b_1]_m = [a_2 + b_2]_m$, as desired. □

We may thus define the arithmetic binary operations $+_m$, $-_m$, and $\cdot_m$ on $\mathbb{Z}_m$ as follows.

Definition 6.5.7. Given $[a]_m, [b]_m \in \mathbb{Z}_m$,

(1) $[a]_m +_m [b]_m = [a + b]_m$,

(2) $[a]_m -_m [b]_m = [a - b]_m$,

(3) $[a]_m \cdot_m [b]_m = [ab]_m$.

Table 6.1. Addition table for $\mathbb{Z}_4$

$+_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[0]_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[1]_4$	$[1]_4$	$[2]_4$	$[3]_4$	$[0]_4$
$[2]_4$	$[2]_4$	$[3]_4$	$[0]_4$	$[1]_4$
$[3]_4$	$[3]_4$	$[0]_4$	$[1]_4$	$[2]_4$

For example, the addition table for the algebraic structure $\langle \mathbb{Z}_4, +_4 \rangle$ can be found in Table 6.1.

In general, $\langle \mathbb{Z}_m, +_m \rangle$ is an example of an algebraic structure called a *group*, a concept you will likely study in a future course called abstract algebra (see [**14**]). A group consists of a set G and a binary operation $*$ on G (see page 102) satisfying the group axioms given in Table 6.2.

Table 6.2. Group axioms

(G1) $*$ is associative
$(\forall a, b, c \in G)[a * (b * c) = (a * b) * c]$.

(G2) existence of identity
$(\exists e \in G)(\forall a \in G)[a * e = a = e * a]$.

(G3) existence of inverses
$(\forall a \in G)(\exists b \in G)[a * b = e = b * a]$.

For example, the set $\mathbb{Z}$ of integers is a group under addition $+$ by the Basic Properties of Integers 1.2.3. Similarly, $\mathbb{Z}_m$ is a group under $+_m$ (see Exercise 6.5.3). Considering again $\langle \mathbb{Z}_4, +_4 \rangle$, note that Table 6.1 shows that $[0]_4$ is the identity element e in $\mathbb{Z}_4$ and that the inverse of $[0]_4$ is $[0]_4$, the inverse of $[1]_4$ is $[3]_4$ (and vice versa), and the inverse of $[2]_4$ is $[2]_4$.

6.5.1. Well-definedness. What, exactly, do we mean when we say that a function $f : X \to Y$ is "well-defined"? In some sense, this traditional terminology is misleading; a "well-defined" function $f : X \to Y$ is just a function: each input in X maps to a unique output in Y. Well-definedness simply hasn't arisen before now. For example, the function $f : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$, $f(x) = x^2$ is clearly a well-defined function. Technically, what we mean is that if $x_1, x_2 \in \mathbb{R}$ with $x_1 = x_2$, then $f(x_1) = f(x_2)$; i.e., $(x_1)^2 = (x_2)^2$. For this function, well-definedness of f is just a property of equality.

The situation is more complicated when trying to define a function on a set of congruence classes, such as a function $f : \mathbb{Z}_4 \to \mathbb{Z}_4$. As we've already noted, elements in $\mathbb{Z}_4$ have more than one name; for example, $[1]_4 = [5]_4 = [149]_4 = [-35]_4$. Given $[n]_4 \in \mathbb{Z}_4$, if we try to define $f([n]_4)$ in terms of n (which is likely), then we have a potential problem because it is possible to have $[n_1]_4 = [n_2]_4$ without having $n_1 = n_2$. In the example above, if using the name $[5]_4$ leads to a different answer than using the name $[1]_4$ when computing $f([1]_4)$, then we would be trying to assign two different values for the output of $f([1]_4)$. If this happens, then f isn't a function at all (f is "ill-defined")!

Theorem 6.5.6 states that the binary function $+_4 : \mathbb{Z}_4 \times \mathbb{Z}_4 \to \mathbb{Z}_4$, which is defined by $f([a]_4, [b]_4) = [a+b]_4$, i.e., in terms of a and b, is well-defined. Similarly, $-_4 : \mathbb{Z}_4 \times \mathbb{Z}_4 \to \mathbb{Z}_4$ and $\cdot_4 : \mathbb{Z}_4 \times \mathbb{Z}_4 \to \mathbb{Z}_4$ are well-defined.

Exercises 6.5

1. Prove Theorem 6.5.5.
2. Complete the proof of Theorem 6.5.6.
3. Let $m \in \mathbb{Z}^+$, $m > 1$.
   (a) Show that $\langle \mathbb{Z}_m, +_m \rangle$ is a group by verifying the group axioms in Table 6.2:
      (i) Show that for all $[a]_m, [b]_m, [c]_m \in \mathbb{Z}_m$,
      $$[a]_m +_m ([b]_m +_m [c]_m) = ([a]_m +_m [b]_m) +_m [c]_m.$$
      (ii) Show that for all $[a]_m \in \mathbb{Z}_m$,
      $$[a]_m +_m [0]_m = [a]_m = [0]_m +_m [a]_m.$$
      (iii) Show that for all $[a]_m \in \mathbb{Z}_m$ there exists $[b]_m \in \mathbb{Z}_m$ such that
      $$[a]_m +_m [b]_m = [0]_m = [b]_m +_m [a]_m.$$
   (b) In fact, $+_m$ is commutative, making $\langle \mathbb{Z}_m, +_m \rangle$ an *abelian*[1] group: show that for all $[a]_m, [b]_m \in \mathbb{Z}_m$,
   $$[a]_m +_m [b]_m = [b]_m +_m [a]_m.$$
4. Prove that if $n \in \mathbb{Z}^+$ is odd, then the sum of all the elements in $\mathbb{Z}_n$ is $[0]_n$.
5. Given $m \in \mathbb{Z}^+$, $m > 1$, let $\mathbb{Z}_m^* = \mathbb{Z}_m - \{[0]_m\}$.
   (a) Find the multiplication tables for $\langle \mathbb{Z}_4^*, \cdot_4 \rangle$ and $\langle \mathbb{Z}_5^*, \cdot_5 \rangle$.
   (b) Is $\langle \mathbb{Z}_4^*, \cdot_4 \rangle$ a group? Is $\langle \mathbb{Z}_5^*, \cdot_5 \rangle$ a group? Why are we considering $\mathbb{Z}_m^*$, rather than $\mathbb{Z}_m$, under multiplication modulo m?
6. Let $m \in \mathbb{Z}^+$, $m > 1$. Prove that $\langle \mathbb{Z}_m^*, \cdot_m \rangle$ is a group if and only if m is prime. (**HINT:** For ($\Rightarrow$), prove the contrapositive. What element of $\mathbb{Z}_m^*$ plays the role of the (multiplicative) identity e? The definition of "m isn't prime" will be useful for showing that axiom **(G3)** in Table 6.2 isn't satisfied. For ($\Leftarrow$), you need to verify that the three group axioms in Table 6.2 are satisfied, where the binary operation $*$ is $\cdot_m$. What element of $\mathbb{Z}_m^*$ plays the role of the (multiplicative) identity e? For axiom **(G3)**, you will find Theorem 6.3.2 useful.)
7. Let $n \in \mathbb{Z}^+$, $n > 1$, and let $f : \mathbb{Z}_n \to \mathbb{Z}$ by, for all $[a]_n \in \mathbb{Z}_n$, $f([a]_n) = (a, n)$. Prove that f is well-defined. (**HINT:** See Exercise 6.4.3.)
8. Let $f : \mathbb{Z}_5 \to \mathbb{Z}$ by, for all $[a]_5 \in \mathbb{Z}_5$, $f([a]_5) = (2, a)$. Prove that f is not well-defined.

[1] Abelian groups are named after the Norwegian mathematician Niels Abel, who lived during 1802–1829.

9. Let $m, n \in \mathbb{Z}^+$ with $(m, n) = 1$. Let $f : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$ by, for all $a \in \mathbb{Z}$, $f([a]_{mn}) = ([a]_m, [a]_n)$.
   (a) Prove that f is well-defined.
   (b) Let $m = 4$ and $n = 7$. Find $a \in \mathbb{Z}$ such that $f([a]_{28}) = ([3]_4, [5]_7)$.
   (c) Prove that f is a bijection.[2] (**HINT:** To prove that f is onto, given $([b]_m, [c]_n) \in \mathbb{Z}_m \times \mathbb{Z}_n$, consider $x = cmr + bns$, where $1 = mr + ns$.)

[2] The fact that f is onto is known as the *Chinese Remainder Theorem*, as problems using this fact appeared in Chinese mathematical writing as early as the first century C.E.

Chapter 7

Equivalence Relations and Partitions

Congruence modulo m is an example of an "equivalence relation" on $\mathbb{Z}$, and we noted in the previous chapter that the congruence classes modulo m of the integers form a "partition" of $\mathbb{Z}$. In this chapter, we formally define and investigate these notions. Our focus is on equivalence relations and partitions, as these concepts can be found in many different mathematical areas of study. In this chapter, we have omitted Given-Goal diagrams, not because they aren't important any more, but because you can construct them yourself.

7.1. Introduction

We must define the more general notion of "relation" before we can study equivalence relations. Our introduction here is very brief and focuses on binary relations. A more thorough study of relations can be found in [**15**].

Definition 7.1.1. Let A and B be sets. A (*binary*) *relation R from A to B* is a subset of $A \times B$. A (*binary*) *relation on A* is a subset of $A \times A$.

Example 7.1.2. Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{a, b, c\}$. Then

$$R = \{(1, b), (1, c), (3, a), (4, a), (4, c), (5, b)\} \subseteq A \times B,$$

so R is a relation from A to B. Note that 3 is "R-related" to a since $(3, a) \in R$, while 3 is not "R-related" to b since $(3, b) \notin R$. $\lozenge$

When $R \subseteq A \times B$ is a relation from A to B, elements $a \in A$ and $b \in B$ are "R-related" when $(a, b) \in R$. It will often be more transparent to use the following notation.

Definition 7.1.3. Let A and B be sets, and let $R \subseteq A \times B$ be a relation from A to B. For $a \in A$ and $b \in B$, we write

$$\begin{aligned} a\, R\, b &\quad \text{iff } (a,b) \in R \text{ and} \\ a \not R\, b &\quad \text{iff } (a,b) \notin R. \end{aligned}$$

Using the notation in Definition 7.1.3, we see in Example 7.1.2 that $3\ R\ a$, while $3 \not R\ b$.

The purpose of the first example is to emphasize that a relation is nothing more than a set of ordered pairs. The second example emphasizes a relation you are already familiar with.

Example 7.1.4. Let $R = \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid x < y\}$. Then $(e,\pi) \in R$ and $(\frac{3}{2}, -3) \notin R$. In the notation of Definition 7.1.3, $e\ R\ \pi$ and $\frac{3}{2} \not R -3$. However, it is standard to identify the relation R with the symbol $<$ used to define it and to write $e < \pi$ and $\frac{3}{2} \not< -3$, as usual.

Since $<$ is a relation on $\mathbb{R}$, it is a set of ordered pairs in $\mathbb{R} \times \mathbb{R}$, and hence it has a graph in $\mathbb{R} \times \mathbb{R}$. ♢

Another common notation used to represent a relation is the symbol "$\sim$".

Example 7.1.5. Let $\mathbb{Z}^* = \mathbb{Z} - \{0\}$. Define the relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}^*$ by, for all $a, c \in \mathbb{Z}$ and all $b, d \in \mathbb{Z}^*$,

$$(a,b) \sim (c,d) \text{ iff } ad = bc.$$

(Technically, $\sim$ is a subset of $(\mathbb{Z} \times \mathbb{Z}^*) \times (\mathbb{Z} \times \mathbb{Z}^*)$; i.e., $\sim$ is a set of ordered pairs of ordered pairs!) Note that

$$\begin{aligned} (1,2) &\sim (-4,-8) &\quad& \text{because } (1)(-8) = (2)(-4) \text{ and} \\ (-1,3) &\not\sim (2,-5) &\quad& \text{because } (-1)(-5) \neq (3)(2). \end{aligned}$$ ♢

Example 7.1.6. In Definition 5.1.1, we defined a function from a set X to Y informally as a *correspondence*. While this is a familiar "definition", it is not a precise definition. What, after all, is a "correspondence"? How do we define it mathematically? It turns out that it does no harm to think of a function in this way, since in fact the notion can be defined rigorously, in terms of relations.

We first define the notion of the domain of a relation.

Definition 7.1.7. Let R be a relation from a set X to a set Y; i.e., $R \subseteq X \times Y$. The *domain* of R is the set

$$\operatorname{dom} R = \{a \in X \mid (\exists b \in Y)[(a,b) \in R]\}.$$

Definition 7.1.8. A relation R from a set X to a set Y is a *function* if for all $a \in \operatorname{dom} R$ there exists a unique $b \in Y$ such that $(a,b) \in R$; i.e.,

$$(\forall a \in \operatorname{dom} R)(\exists! b \in Y)[(a,b) \in R].$$

Furthermore, $R : X \to Y$ if $\operatorname{dom} R = X$. ♢

Exercises 7.1

1. Suppose that the relation R is a function $R : X \to Y$, as defined by Definition 7.1.8.
   (a) Write down the definition of "R is 1-1", keeping in mind that $R \subseteq X \times Y$.
   (b) Write down the definition of "R is onto", keeping in mind that $R \subseteq X \times Y$.
   (c) Under the assumption that R is a bijection (i.e., R is 1-1 and onto), write down the definition of the inverse function R^{-1}, keeping in mind that $R \subseteq X \times Y$.
2. Suppose that the relation f is a function $f : X \to Y$ and the relation g is a function $g : Y \to Z$, as defined by Definition 7.1.8. Write down the definition of the composite function $g \circ f$, keeping in mind that $f \subseteq X \times Y$ and $g \subseteq Y \times Z$.

7.2. Equivalence relations

Relations such as congruence modulo m, which are reflexive, symmetric, and transitive, as described in Theorem 6.4.4, are called "equivalence relations". We formalize these notions in the next definition.

Definition 7.2.1. Let $\sim$ be a relation on a set A.

(1) $\sim$ is *reflexive* if $(\forall a \in A)[a \sim a]$.

(2) $\sim$ is *symmetric* if $(\forall a, b \in A)[a \sim b \Rightarrow b \sim a]$.

(3) $\sim$ is *transitive* if $(\forall a, b, c \in A)[(a \sim b \text{ and } b \sim c) \Rightarrow a \sim c]$.

(4) $\sim$ is an *equivalence relation* if $\sim$ is reflexive, symmetric, and transitive.

We consider several examples.

Example 7.2.2.

(1) The equality relation $=$ is an equivalence relation on any set X, since
- for all $x \in X$, $x = x$,
- for all $x, y \in X$, if $x = y$, then $y = x$, and
- for all $x, y, z \in X$, if $x = y$ and $y = z$, then $x = z$.

(2) For each $m \in \mathbb{Z}^+$, the congruence relation "$\equiv \mod m$" is an equivalence relation on $\mathbb{Z}$, by Theorem 6.4.4. ◊

Not all relations are equivalence relations. Note that Definition 7.2.1 implies that a relation $\sim$ on a set A is *not symmetric* if

$$(\exists a, b \in A)[a \sim b \text{ and } b \not\sim a].$$

As an exercise (see Exercise 7.2.1), you should write down the definitions of "$\sim$ is *not reflexive*" and "$\sim$ is *not transitive*".

Example 7.2.3. Determine whether the following relations are reflexive, symmetric, or transitive.

(1) $\leq$ on $\mathbb{R}$.
- $\leq$ is reflexive since, given $x \in \mathbb{R}$, $x \leq x$.
- $\leq$ is not symmetric because $1 \leq 2$, but $2 \not\leq 1$.
- $\leq$ is transitive since, given $x, y, z \in \mathbb{R}$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

(2) Let $A = \{1, 2, 3, 4\}$ and let the relation R on A be defined by

$$R = \{(1,1), (2,3), (3,2), (2,2), (3,4), (4,3)\}.$$

- R is not reflexive because $3 \not R\, 3$; i.e., $(3,3) \notin R$.
- R is symmetric. Here we must consider all possibilities for all ordered pairs $(a,b) \in R$, to be sure that (b,a) is also in R when (a,b) is.
- R is not transitive because $2\, R\, 3$ and $3\, R\, 4$, but $2 \not R\, 4$. ◊

In the next example, we show that the relation $\sim$ defined in Example 7.1.5 is an equivalence relation.

Example 7.2.4. Recall that $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ and define the relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}^*$ by, for all $a, c \in \mathbb{Z}$, $b, d \in \mathbb{Z}^*$,

$$(a,b) \sim (c,d) \text{ iff } ad = bc.$$

We show that $\sim$ is reflexive, symmetric, and transitive.

$\sim$ **is reflexive:** Let $(a,b) \in \mathbb{Z} \times \mathbb{Z}^*$. Then $(a,b) \sim (a,b)$ because $ab = ba$, and hence $\sim$ is reflexive.

$\sim$ **is symmetric:** Let $(a,b), (c,d) \in \mathbb{Z} \times \mathbb{Z}^*$ and assume that $(a,b) \sim (c,d)$. We must show that $(c,d) \sim (a,b)$. Since $(a,b) \sim (c,d)$, we know that $ad = bc$. By properties of equality and multiplication of integers, we know that $cb = da$. It follows by definition that $(c,d) \sim (a,b)$, and hence $\sim$ is symmetric.

$\sim$ **is transitive:** Let $(a,b), (c,d), (m,n) \in \mathbb{Z} \times \mathbb{Z}^*$ and assume that $(a,b) \sim (c,d)$ and $(c,d) \sim (m,n)$. We must show that $(a,b) \sim (m,n)$. Since $(a,b) \sim (c,d)$, we know that $ad = bc$, and since $(c,d) \sim (m,n)$, we know that $cn = dm$. To show that $(a,b) \sim (m,n)$, we must show that $an = bm$.

Since $ad = bc$, we know that $adn = bcn$, and since $cn = dm$, we know that $bcn = bdm$. Thus $adn = bdm$. Since $d \neq 0$ (remember that $d \in \mathbb{Z}^* = \mathbb{Z} - \{0\}$), cancellation implies that $an = bm$ as desired. Thus $(a,b) \sim (m,n)$, and hence $\sim$ is transitive.

Since $\sim$ is reflexive, symmetric, and transitive, $\sim$ is an equivalence relation, by definition. ◊

7.2.1. Equivalence classes.

In Section 6.5, we used the fact that congruence modulo m is an equivalence relation in order to identify integers that are congruent modulo m; this identification yielded the congruence classes modulo m. Similarly, if we have an arbitrary equivalence relation on a set X, we may identify elements of X that are equivalent under this relation to form "equivalence classes".

Definition 7.2.5. Let $\sim$ be an equivalence relation on a nonempty set X, and let $a \in X$. The *equivalence class of* a is the set

$$[a] = \{x \in X \mid x \sim a\}.$$

The set of all equivalence classes of $\sim$ is denoted by

$$X/\!\sim\, = \{[a] \mid a \in X\}.$$

Note that $X/{\sim} \subseteq \mathcal{P}(X)$.

As usual, we consider several examples.

Example 7.2.6. For convenience of notation, let $\equiv_m$ denote congruence modulo $m \in \mathbb{Z}^+$; i.e., for all $a, b \in \mathbb{Z}$, let $a \equiv_m b$ iff $a \equiv b \mod m$. Then, as we have already noted, $\equiv_m$ is an equivalence relation. Given $a \in \mathbb{Z}$, the equivalence class $[a] = \{n \in \mathbb{Z} \mid n \equiv_m a\}$ is just the congruence class of a modulo m; i.e.,

$$[a] = [a]_m = \{n \in \mathbb{Z} \mid n \equiv a \mod m\}.$$

The set of all equivalence classes of $\equiv_m$ is

$$\mathbb{Z}/{\equiv_m} = \mathbb{Z}_m = \{[0]_m, [1]_m, \ldots, [m-1]_m\}. \qquad \diamondsuit$$

Example 7.2.7. Once again, recall the definition of the equivalence relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}^*$ from Example 7.1.5:

$$(a, b) \sim (c, d) \text{ iff } ad = bc, \quad \text{for all } a, c \in \mathbb{Z},\ b, d \in \mathbb{Z}^*, \tag{7.1}$$

where $\mathbb{Z}^* = \mathbb{Z} - \{0\}$.

We consider the equivalence class $[(1,2)]$. Since $(-4)(2) = (-8)(1)$, $(-4,-8) \sim (1,2)$, and hence $(-4,-8) \in [(1,2)]$. Since $(-1)(2) \neq (2)(1)$, $(-1,2) \not\sim (1,2)$, and hence $(-1,2) \notin [(1,2)]$. We can use its definition to compute the equivalence class $[(1,2)]$. Given an arbitrary $(a,b) \in \mathbb{Z} \times \mathbb{Z}^*$,

$$(a,b) \in [(1,2)] \Leftrightarrow (a,b) \sim (1,2) \Leftrightarrow 2a = b.$$

Thus $[(1,2)] = \{(a,b) \in \mathbb{Z} \times \mathbb{Z}^* \mid b = 2a\}$. Other elements of $[(1,2)]$ are $(1,2)$, $(7,14)$, and $(-9,-18)$.

Recall that our definition of $\mathbb{Q}$ in Table 1.10 was informal. The equivalence relation in this example is used to rigorously define the rational numbers from the integers; i.e., formally, $\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z}^+)/{\sim}$. Informally, you should think of $(a,b) \in \mathbb{Z} \times \mathbb{Z}^+$ as the fraction $\frac{a}{b}$. Informally, statement (7.1) says that $\frac{a}{b} = \frac{c}{d}$ iff $ad = bc$. The equivalence class $[(1,2)]$ collects together in a single set all the "names" for the fraction $\frac{1}{2}$:

$$\frac{1}{2} = \frac{-4}{-8} = \frac{7}{14} = \frac{-9}{-18} = \cdots,$$

since $(1,2), (-4,-8), (7,14), (-9,-18)$ are all elements of $[(1,2)]$. In Exercise 7.2.9, you will show that the formal definitions of $+_{\mathbb{Q}}$ and $\cdot_{\mathbb{Q}}$ for $\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z}^+)/{\sim}$ are well-defined. $\diamondsuit$

Example 7.2.8. Consider the relation $\sim$ on $\mathbb{R} \times \mathbb{R}$ defined by, for all $a_1, a_2, b_1, b_2 \in \mathbb{R}$,

$$(a_1, a_2) \sim (b_1, b_2) \text{ iff } (a_1)^2 + (a_2)^2 = (b_1)^2 + (b_2)^2.$$

In Exercise 7.2.5, you are asked to show that $\sim$ is an equivalence relation on $\mathbb{R} \times \mathbb{R}$. As an example,

$$\begin{aligned}[(1,2)] &= \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid (x,y) \sim (1,2)\} \\ &= \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = 5\};\end{aligned}$$

i.e., the equivalence class $[(1,2)]$ is the set of all points in $\mathbb{R} \times \mathbb{R}$ on the circle $x^2 + y^2 = 5$.

Thus $(\mathbb{R} \times \mathbb{R})/{\sim}$ is the set of all graphs of circles in the plane centered at the origin. The equivalence class

$$[(0,0)] = \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = 0\} = \{(0,0)\}$$

is called a "degenerate circle". See Figure 7.1 on page 164 for a picture of the equivalence classes of this equivalence relation. ♢

In Section 6.5, we spoke of the set of congruence classes of the integers modulo m forming a "partition" of the integers. Similarly, the set of equivalence classes of an equivalence relation on a set X "partitions" X into pairwise disjoint subsets, as follows.

Theorem 7.2.9. *Let $\sim$ be an equivalence relation on a nonempty set X.*

(1) *For all $a \in X$, $a \in [a]$.*

(2) *For all $a, b \in X$, $a \sim b$ iff $[a] = [b]$.*

(3) *For all $a, b \in X$, $a \not\sim b$ iff $[a] \cap [b] = \emptyset$.*

Proof. Let $\sim$ be an equivalence relation on $X \neq \emptyset$.

(1) Let $a \in X$. Then $a \sim a$, since $\sim$ is reflexive, and hence $a \in [a]$ by Definition 7.2.5.

(2) Let $a, b \in X$.

($\Rightarrow$) Assume $a \sim b$; we must show $[a] = [b]$. Since $[a]$ and $[b]$ are sets, we prove the usual set containments. Let $x \in [a]$. Then $x \sim a$ by definition. Hence we have $x \sim a$ and $a \sim b$, and so $x \sim b$ by transitivity of $\sim$. Thus $x \in [b]$ and $[a] \subseteq [b]$. Next let $x \in [b]$. Then $x \sim b$ by definition. Since $a \sim b$, we know that $b \sim a$ since $\sim$ is symmetric. Thus $x \sim b$ and $b \sim a$, and so $x \sim a$ by transitivity. Thus $x \in [a]$ and $[b] \subseteq [a]$, and so $[a] = [b]$.

($\Leftarrow$) Assume that $[a] = [b]$; we must show that $a \sim b$. Note that $a \in [a]$, by part (1). Since $[a] = [b]$, $a \in [b]$, and hence $a \sim b$ by definition, as desired.

(3) Let $a, b \in X$.

($\Rightarrow$) We prove the contrapositive. Assume that $[a] \cap [b] \neq \emptyset$. We must show that $a \sim b$. Since $[a] \cap [b] \neq \emptyset$, we may fix $x \in X$ such that $x \in [a] \cap [b]$. Then $x \sim a$, since $x \in [a]$, so $a \sim x$ since $\sim$ is symmetric. Similarly, $x \sim b$, since $x \in [b]$. Thus $a \sim x$ and $x \sim b$, and hence $a \sim b$ by transitivity of $\sim$.

($\Leftarrow$) We prove the contrapositive. Assume that $a \sim b$. We must show that $[a] \cap [b] \neq \emptyset$. Note that $a \in [a]$ by part (1), and since $a \sim b$, $a \in [b]$ by definition. Thus $a \in [a] \cap [b]$ and so $[a] \cap [b] \neq \emptyset$. □

Exercises 7.2

1. Let $\sim$ be a relation on a set A.
   (a) Give the definition of "$\sim$ is not reflexive".
   (b) Give the definition of "$\sim$ is not transitive".

2. Determine whether the following relations on the given sets are reflexive, symmetric, or transitive.
   (a) The relation
   $$R = \{(1,1),(1,3),(1,4),(1,5),(2,5),(3,1),(3,4),(4,1),(4,3),(5,1),(5,2)\}$$
   on the set $A = \{1,2,3,4,5\}$.
   (b) The divisibility relation $\mid$ on $\mathbb{Z}$.
   (c) The relation $\subseteq$ on $\mathcal{P}(\mathbb{Z})$.
   (d) The relation $\sim$ on $\mathbb{R}$ defined by $x \sim y$ iff $xy \geq 0$.
   (e) The relation R on $\mathbb{Z}$ defined by $m\ R\ n$ iff $2 \mid (m+n)$.
   (f) The relation R on $\mathbb{Z}$ defined by $m\ R\ n$ iff $3 \mid (m+n)$.
3. Let $A = \{1,2,3,4\}$. Give an example of a relation R on A (as a set of ordered pairs) such that:
   (a) R is reflexive, symmetric, and transitive.
   (b) R is reflexive, symmetric, and not transitive.
   (c) R is reflexive, not symmetric, and transitive.
   (d) R is reflexive, not symmetric, and not transitive.
   (e) R is not reflexive, symmetric, and transitive.
   (f) R is not reflexive, symmetric, and not transitive.
   (g) R is not reflexive, not symmetric, and transitive.
   (h) R is not reflexive, not symmetric, and not transitive.
4. For each of the following, prove that the relation is an equivalence relation. Then give the information about the equivalence classes, as specified.
   (a) The relation $\sim$ on $\mathbb{Z}$ defined by $x \sim y$ iff $x^2 = y^2$. Explicitly find the equivalence classes $[0]$, $[4]$, and $[-72]$.
   (b) The relation $\sim$ on $\mathbb{R}$ defined by $x \sim y$ iff $x = y$ or $xy = 2$. Explicitly find the equivalence classes $[2]$, $[3]$, $[-\frac{4}{5}]$, and $[0]$.
   (c) The relation $\sim$ on $\mathbb{R} \times \mathbb{R}$ defined by $(x,y) \sim (u,v)$ iff $4x - y = 4u - v$. Explicitly find the equivalence classes $[(5,2)]$ and $[(0,0)]$. For fixed values $a, b \in \mathbb{R}$, describe $[(a,b)]$, both as a set and geometrically.
   (d) The relation $\sim$ on $\mathbb{R} \times \mathbb{R}$ defined by $(x,y) \sim (u,v)$ iff $x^2 - y = u^2 - v$. Explicitly find the equivalence classes $[(5,2)]$ and $[(0,0)]$. For fixed values $a, b \in \mathbb{R}$, describe $[(a,b)]$, both as a set and geometrically.
   (e) The relation $\sim$ on $\mathbb{R}^+ \times \mathbb{R}^+$ defined by $(x,y) \sim (u,v)$ iff $x^2 v = u^2 y$. Explicitly find the equivalence classes $[(5,2)]$ and $[(1,4)]$. For fixed values $a, b \in \mathbb{R}^+$, describe $[(a,b)]$, both as a set and geometrically.
   (f) The relation $\sim$ on $\mathbb{Z}^{\geq 0} \times \mathbb{Z}^{\geq 0}$ defined by $(x,y) \sim (z,w)$ iff $x + w = z + y$. Give three elements of the equivalence class $[(2,0)]$ and three elements of the equivalence class $[(0,3)]$.
   Note: Since subtraction and division are not always defined in $\mathbb{Z}^{\geq 0}$ (for example, $2 - 3$ and $1 \div 2$ are not defined in $\mathbb{Z}^{\geq 0}$), your proof should be expressed in terms of the operation of addition and use the cancellation property in $\mathbb{Z}^{\geq 0}$. This equivalence relation is used to rigorously define the integers $\mathbb{Z}$ from the nonnegative integers $\mathbb{Z}^{\geq 0}$.
5. Define $\sim$ on $\mathbb{R} \times \mathbb{R}$ by, for all $a_1, a_2, b_1, b_2 \in \mathbb{R}$, $(a_1, a_2) \sim (b_1, b_2)$ iff $(a_1)^2 + (a_2)^2 = (b_1)^2 + (b_2)^2$. Show that $\sim$ is an equivalence relation on $\mathbb{R} \times \mathbb{R}$.

6. Let R be a relation on a set A such that $(\forall a \in A)(\exists b \in A)[aRb]$. Show that if R is symmetric and transitive, then R is reflexive.
   Warning: If you do not use all of the hypotheses in this statement, then you do not have a proof.
7. Let $A = \{1, 2, 3\}$. Find a nontrivial (i.e., nonempty) relation on A (i.e., a set of ordered pairs that is a subset of $A \times A$) that is symmetric and transitive, but not reflexive. Why does your relation not contradict the statement in Exercise 7.2.6?
8. Let $n \geq 1$, and let $S_n = \{f \mid f : \{1, 2, \ldots, n\} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} \{1, 2, \ldots, n\}\}$. Define a relation $\sim$ on S_n by, for all $f, g \in S_n$, $f \sim g$ iff there exists $h \in S_n$ such that $g = h^{-1} \circ f \circ h$.
   (a) Show that $\sim$ is an equivalence relation on S_n.
   (b) Find S_3.
   (c) Find the equivalence classes of $\sim$ for S_3.
9. Let $\sim$ be the equivalence relation defined in Example 7.2.7. In that example, we noted that the formal definition of the set of rational numbers is $\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z}^+)/{\sim}$. In this exercise, you will show that the usual formulas for addition and multiplication of rational numbers, now expressed formally in $(\mathbb{Z} \times \mathbb{Z}^+)/{\sim}$, are well-defined binary operations on $\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z}^+)/{\sim}$.

   Let $(a_1, b_1), (a_2, b_2), (c_1, d_1), (c_2, d_2) \in \mathbb{Z} \times \mathbb{Z}^+$. Assume that
   $$[(a_1, b_1)] = [(c_1, d_1)] \text{ and } [(a_2, b_2)] = [(c_2, d_2)].$$
   Prove:
   (a) $[(a_1b_2 + a_2b_1, b_1b_2)] = [(c_1d_2 + d_1c_2, d_1d_2)]$.
   (b) $[(a_1a_2, b_1b_2)] = [(c_1c_2, d_1d_2)]$.
   It then follows that $+_{\mathbb{Q}} : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ and $\cdot_{\mathbb{Q}} : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ by, for all $[(a_1, b_1)], [(a_2, b_2)] \in (\mathbb{Z} \times \mathbb{Z}^+)/{\sim}$,
   $$[(a_1, b_1)] +_{\mathbb{Q}} [(a_2, b_2)] = [(a_1b_2 + a_2b_1, b_1b_2)] \text{ and}$$
   $$[(a_1, b_1)] \cdot_{\mathbb{Q}} [(a_2, b_2)] = [(a_1a_2, b_1b_2)]$$
   are well-defined.
10. Suppose that $\sim$ is an equivalence relation on a set A and that $f : A \to A$ such that for all $a, b \in A$,
    $$a\,R\,b \implies f(a)\;R\;f(b).$$
    Prove that the function $F : (A/{\sim}) \to (A/{\sim})$ defined by $F([a]) = [f(a)]$ is well-defined.

7.3. Partitions

Having used the term “partition” informally several times, we now give a precise definition.

Definition 7.3.1. Let X be a nonempty set, and let $\mathbf{P}$ be a collection of subsets of X (i.e., $\mathbf{P} \subseteq \mathcal{P}(X)$). The collection $\mathbf{P}$ is a *partition* of X if

(1) For all $A \in \mathbf{P}$, $A \neq \emptyset$.

(2) For all $A, B \in \mathbf{P}$, $A = B$ or $A \cap B = \emptyset$.

(3) For all $x \in X$ there exists $A \in \mathbf{P}$ such that $x \in A$. (This says that the sets in $\mathbf{P}$ *cover* X.)

Example 7.3.2.

(1) Let $X = \{1, 2, 3, 4, 5, 6\}$. Then $\mathbf{P} = \{\{1, 3, 4\}, \{2, 5\}, \{6\}\}$ is a partition of X; i.e., every integer in X is in exactly one of the sets $\{1, 3, 4\}$, $\{2, 5\}$, or $\{6\}$.

(2) The collection $\mathbf{P} = \{\mathbb{Q}^+, \mathbb{Q}^-, \{0\}\}$ is a partition of $\mathbb{Q}$; in other words, every rational number falls into exactly one category: positive, negative, or zero. ♢

As we have already noted informally above, every equivalence relation on a set gives rise to, or *induces*, a partition of that set.

Corollary 7.3.3 (Corollary to Theorem 7.2.9)**.** *Let $\sim$ be an equivalence relation on a nonempty set X. Then $X/{\sim} = \{[a] \mid a \in X\}$, the set of all equivalence classes of $\sim$, is a partition of X.*

Proof. Given an equivalence relation $\sim$ on a nonempty set X, $X/{\sim}$ satisfies Definition 7.3.1(1) and (3) by Theorem 7.2.9(1) and $X/{\sim}$ satisfies Definition 7.3.1(2) by Theorem 7.2.9(2) and (3), since for any $a, b \in X$, either $a \sim b$ or $a \not\sim b$. □

The following examples are restatements of Examples 6.5.3 and 7.2.8.

Example 7.3.4.

(1) $\mathbb{Z}_4 = \{[0]_4, [1]_4, [2]_4, [3]_4\}$ is a partition of $\mathbb{Z}$.

(2) $\mathbf{P} = \big\{\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\} \mid r \in \mathbb{R}^{\geq 0}\big\}$ is a partition of $\mathbb{R} \times \mathbb{R}$.

Recall from Example 7.2.8 that

$$\begin{aligned} [(1, 2)] &= \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = 5\}, \\ [(0, 0)] &= \{(0, 0)\}, \text{ and in general,} \\ [(a, b)] &= \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = a^2 + b^2\} \end{aligned}$$

when $(a, b) \in \mathbb{R} \times \mathbb{R}$. This says that the Cartesian plane $\mathbb{R} \times \mathbb{R}$ can be partitioned into pairwise disjoint circles (where $\{(0, 0)\}$ is a degenerate circle) that cover $\mathbb{R} \times \mathbb{R}$. See Figure 7.1. ♢

The converse of Corollary 7.3.3 is also true; i.e., not only does every equivalence relation on a nonempty set give rise to a partition of that set, but also every partition of a nonempty set gives rise to an equivalence relation on that set. In fact, we can prove something slightly stronger.

Theorem 7.3.5. *Let $\mathbf{P}$ be a partition of the nonempty set X. Define a relation $\sim$ on X by, for all $a, b \in X$,*

$$a \sim b \iff (\exists A \in \mathbf{P})[a \in A \text{ and } b \in A].$$

Then $\sim$ is an equivalence relation on X. Furthermore, the equivalence classes of $\sim$ are exactly the elements of the partition $\mathbf{P}$; i.e., $X/{\sim} = \mathbf{P}$.

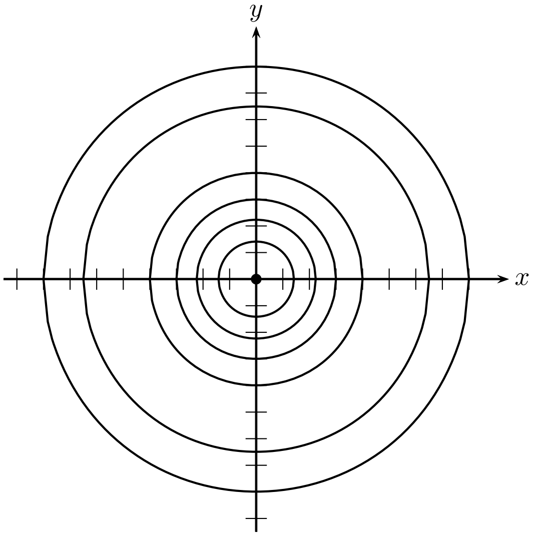


Figure 7.1. $\mathbb{R} \times \mathbb{R}$ partitioned into nested circles.

Proof. Let $\mathbf{P}$ be a partition of the nonempty set X and let $\sim$ be defined by, for all $a, b \in X$,

$$a \sim b \Leftrightarrow (\exists A \in \mathbf{P})[a \in A \text{ and } b \in A].$$

- $\sim$ **is reflexive:** Let $a \in X$. By Definition 7.3.1(3), since $\mathbf{P}$ covers X, we can fix $A \in \mathbf{P}$ such that $a \in A$. Thus $a \sim a$.
- $\sim$ **is symmetric:** Let $a, b \in X$, and assume that $a \sim b$. Then we can fix $A \in \mathbf{P}$ such that $a \in \mathbf{P}$ and $b \in \mathbf{P}$, by definition of $\sim$. But then $b \sim a$ is immediate.
- $\sim$ **is transitive:** Let $a, b, c \in X$, and assume that $a \sim b$ and $b \sim c$. We prove that $a \sim c$.

  Since $a \sim b$, we can fix $A \in \mathbf{P}$ such that $a \in A$ and $b \in A$. Similarly, since $b \sim c$, we can fix $B \in \mathbf{P}$ such that $b \in B$ and $c \in B$. Then $b \in A \cap B$, so $A \cap B \neq \emptyset$. Since $\mathbf{P}$ is a partition, by Definition 7.3.1(2), we know that $A = B$ or $A \cap B = \emptyset$. Hence $A = B$ and $a, c \in A$. Thus $a \sim c$ by definition of $\sim$.

Thus, $\sim$ is an equivalence relation on X. We next show that $X/{\sim} = \mathbf{P}$.

Let $[a] \in X/{\sim}$. Then $a \in X$, so we can fix $A \in \mathbf{P}$ such that $a \in A$, by Definition 7.3.1(3). Note that $[a] = A$, since for all $x \in X$,

$$\begin{aligned} x \in [a] &\Leftrightarrow x \sim a \\ &\Leftrightarrow x \in A \end{aligned}$$

by definition of $\sim$. Hence $[a] \in \mathbf{P}$ and so $X/{\sim} \subseteq \mathbf{P}$.

Next let $A \in \mathbf{P}$. Then $A \neq \emptyset$ by Definition 7.3.1(1), so we can fix an element $a \in A$. Once again we can show that $[a] = A$, so that $A \in X/{\sim}$ and $\mathbf{P} \subseteq X/{\sim}$. Hence $X/{\sim} = \mathbf{P}$ as desired. □

Exercises 7.3

1. Let $A = \{1, 2, 3, 4, 5\}$. Give the equivalence relation (as a set of ordered pairs in $A \times A$) associated with these partitions of A:
   (a) $\{\{2, 4\}, \{1, 3, 5\}\}$,
   (b) $\{\{2, 3, 4, 5\}, \{1\}\}$,
   (c) $\{\{2\}, \{5\}, \{1, 3, 4\}\}$,
   (d) $\{\{1, 5\}, \{2, 3\}, \{4\}\}$.
2. Let X and Y be sets and let $f : X \to Y$ be onto. For all $b \in Y$, let $A_b = f^{-1}[\{b\}]$.
   (a) Prove that $\{A_b \mid b \in Y\}$ is a partition of X.
   (b) Write down the equivalence relation $\sim$ on X associated with this partition of X.
   (c) For $f : \mathbb{Z} \to \{0, 1\}$ by, for all $n \in \mathbb{Z}$,
   $$f(n) = \begin{cases} 0 & \text{if } n \text{ is even,} \\ 1 & \text{if } n \text{ is odd,} \end{cases}$$
   find $A_0 = f^{-1}[\{0\}]$ and $A_1 = f^{-1}[\{1\}]$. Is $\{A_0, A_1\}$ a partition of $\mathbb{Z}$?
   (d) For $f : \mathbb{R} \to \mathbb{R}^{\geq 0}$ by, for all $x \in \mathbb{R}$, $f(x) = x^2$, find A_0, A_4, A_5, and A_π. Is $\{A_b \mid b \in \mathbb{R}^{\geq 0}\}$ a partition of $\mathbb{R}$?

Chapter 8

Finite and Infinite Sets

Finite and infinite sets are familiar notions, but we have yet to define these terms rigorously. In this chapter, we define mathematically what it means for two sets to have the "same size". We discuss several facts regarding finite sets, and we address the question of whether all infinite sets have the same size.

8.1. Introduction

Until now, we have not questioned what we mean by a finite or infinite set. If we wish to prove statements about these concepts, however, we must have mathematical definitions to work with. It seems quite clear to us, and we can see "at a glance", that the set

$$X = \left\{3, -6, \pi, 4.7, \sqrt{2}\right\}$$

is finite and that X is the "same size" as the set

$$Y = \{\blacksquare, \clubsuit, \spadesuit, \bigstar, \blacklozenge\}.$$

By "counting", we see that each of these sets has five elements. Mathematically, this says that we can put each of X and Y into 1-1 correspondence with the set $\{1, 2, 3, 4, 5\}$. But we don't even need to work out how many elements these sets have in order to see that they are the "same size", since we can see directly that they have the same size by matching their elements, i.e., by constructing a bijection between them.

$$\begin{array}{ccccc}
3 & -6 & \pi & 4.7 & \sqrt{2} \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\blacksquare & \clubsuit & \spadesuit & \bigstar & \blacklozenge
\end{array}$$

We make these notions precise below. For convenience, we first fix notation. Recall that the set $\mathbb{N}$ of natural numbers is the set $\mathbb{N} = \mathbb{Z}^+ = \{1, 2, 3, \dots\}$.

Notation 8.1.1. For $n \in \mathbb{N}$, we denote by $\mathbb{N}_n$ the set

$$\mathbb{N}_n = \{1, 2, \ldots, n\} = \{i \in \mathbb{N} \mid i \leq n\}.$$

We define $N_0 = \emptyset$.

Definition 8.1.2. Let X and Y be sets.

(1) We say X is *equinumerous with* Y, denoted by $X \approx Y$, if there exists a bijection $f : X \underset{\text{onto}}{\overset{\text{1-1}}{\to}} Y$.

(2) We say X is *finite* if $X = \emptyset$ or there exists $n \in \mathbb{N}$ such that $\mathbb{N}_n \approx X$; i.e., there is a bijection $f : \{1, 2, \ldots, n\} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} X$.

(3) If $\mathbb{N}_n \approx X$, then we say that the *cardinality* of X is n and write $|X| = n$. We define the cardinality $|\emptyset|$ of the empty set to be $|\emptyset| = 0$.

(4) We say X is *infinite* if X is not finite.

Note that $\approx$ is an "equivalence relation"[1] on the collection of all nonempty sets. See Exercise 8.1.9.

While these definitions seem quite sensible and straightforward, it turns out that several "common sense" facts regarding these notions are fairly complicated to prove *from the definitions*, and we delay these proofs until Section 8.2. For example, while it is hard to imagine how it could possibly be otherwise, we need to prove that when X is finite, the cardinality of X is well-defined; i.e., if X is a nonempty finite set, then there exists a *unique* natural number n such that $\mathbb{N}_n \approx X$ (see Corollary 8.2.2). For now, we will assume this fact. Furthermore, while it seems obvious that the set $\mathbb{N}$ of natural numbers is infinite, this fact also requires a proof from the definition (see Corollary 8.2.4).

We now consider some examples that illustrate the concepts in Definition 8.1.2. First, let's make the bijections in our original example explicit.

Example 8.1.3. Let $X = \{3, -6, \pi, 4.7, \sqrt{2}\}$ and $Y = \{■, ♣, ♠, ★, ♦\}$. The function $f : X \to Y$ defined by

$$f(3) = ■, \quad f(-6) = ♣, \quad f(\pi) = ♠, \quad f(4.7) = ★, \quad f(\sqrt{2}) = ♦$$

is a bijection that shows that $X \approx Y$.

The function $g : \mathbb{N}_5 \to X$ defined by

$$g(1) = \pi, \quad g(2) = 4.7, \quad g(3) = -6, \quad g(4) = \sqrt{2}, \quad g(5) = 3$$

is a bijection that shows that X is finite and $|X| = 5$. ♢

The bijection g in the previous example may not have been the one you expected to see. In fact, there are many bijections from $\mathbb{N}_5$ to the set X in that example that illustrate that $|X| = 5$ (see Exercise 8.1.1).

The next example demonstrates two infinite sets that are equinumerous.

[1] We put the term *equivalence relation* in quotation marks since the collection of all nonempty sets is not a set.

Example 8.1.4. $\mathbb{N} \approx \mathbb{Z}$.

It is often helpful to view the desired bijection using a picture. We need to define the function illustrated below and show that it is a bijection between $\mathbb{N}$ and $\mathbb{Z}$.

$$\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & \ldots \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \ldots \\
0 & 1 & -1 & 2 & -2 & 3 & \ldots
\end{array}$$

Formally, define $f : \mathbb{N} \to \mathbb{Z}$ by, for all $n \in \mathbb{N}$,

$$f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even,} \\ \frac{-(n-1)}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Note that $\operatorname{ran} f \subseteq \mathbb{Z}$ (so that $\mathbb{Z}$ is a reasonable codomain), since if n is even, then $\frac{n}{2} \in \mathbb{Z}$, and if n is odd, then $\frac{-(n-1)}{2} \in \mathbb{Z}$.

To see that f is 1-1, let $n_1, n_2 \in \mathbb{N}$ and assume that $f(n_1) = f(n_2)$. If n_1 and n_2 are both even, then we have that $\frac{n_1}{2} = \frac{n_2}{2}$, and hence $n_1 = n_2$, as desired. Similarly, if n_1 and n_2 are both odd, then we have that $\frac{-(n_1-1)}{2} = \frac{-(n_2-1)}{2}$, and hence $n_1 = n_2$. Note that we cannot have n_1 even and n_2 odd (or vice versa), since $f(n) > 0$ when n is even and $f(n) \le 0$ when n is odd (which you can easily check). Hence f is 1-1.

To see that f is onto, assume that $n \in \mathbb{Z}$. If $n \ge 1$, then $f(2n) = \frac{2n}{2} = n$. If $n \le 0$, then $-2n+1 \ge 1$ and $f(-2n+1) = \frac{-((-2n+1)-1)}{2} = n$. Hence f is onto. $\Diamond$

Exercises 8.1

1. The bijection g given in Example 8.1.3 is not the only one that shows that the set X defined there has cardinality 5. Find another one. How many such bijections are there?
2. The bijection given in Example 8.1.4 is not the only one that shows that $\mathbb{N}$ and $\mathbb{Z}$ are equinumerous. Find another one, and prove that it is a bijection.
3. Prove that $\mathbb{N} \approx O^*$, where O^* is the set of positive odd integers.
4. Prove that $\mathbb{Z} \approx E^*$, where E^* is the set of positive even integers.
5. Prove that the interval $(0, 1) \subseteq \mathbb{R}$ is equinumerous with the interval $(3, 7)$. (**HINT:** Graph $(0, 1)$ on the x-axis and $(3, 7)$ on the y-axis, and think about lines.)
6. Let $a, b, c, d \in \mathbb{R}$ with $a < b$ and $c < d$. Prove that $(a, b) \approx (c, d)$. Which other pairs of bounded intervals with these endpoints are equinumerous?

7. Prove that $[0,1] \approx (0,1)$. (**HINT:** The following picture gives the idea of the proof.)

$$\begin{array}{ccccccc} 0 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \cdots \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \cdots \end{array}$$

8. Prove that $(0,1) \approx \mathbb{R}$. (**HINT:** You will find the inverse tangent function, whose properties you may assume, useful.)

9. Prove the following for all nonempty sets X, Y, Z:
   (a) $X \approx X$.
   (b) If $X \approx Y$, then $Y \approx X$.
   (c) If $X \approx Y$ and $Y \approx Z$, then $X \approx Z$.

10. Let X and Y be sets and assume that $X \approx Y$. Prove:
    (a) If X is finite and nonempty, then Y is finite and $|X| = |Y|$.
    (b) If X is infinite, then Y is infinite.

11. Let A be a set and let x be any element in the underlying universe $\mathcal{U}$. Prove that $\{x\} \times A \approx A$ and $A \times \{x\} \approx A$.

8.2. Finite sets

In this section, we prove several useful facts about finite sets, as well as prove the statements from Section 8.1 that were given there without proof.

Definition 8.1.2(2) states that for $n \neq 0$, $|A| = n$ exactly when there is a bijection $f : \mathbb{N}_n \to A$, which is the natural mathematical notion that corresponds to our informal idea of "counting" the elements of A. Another way to think of this is to note that when $|A| = n$, $n \neq 0$, the elements of A can be enumerated as a list of n elements (and so the list ends). To see this, let $f : \mathbb{N}_n \underset{\text{onto}}{\overset{\text{1-1}}{\to}} A$; we can list the elements of A as

$$f(1), f(2), f(3), \ldots, f(n).$$

If we define $a_i = f(i)$ for all $i \in \mathbb{N}_n$, then we can write A as

$$A = \{a_1, a_2, a_3, \ldots, a_n\}.$$

It is easy to believe that the cardinality of a nonempty finite set is a unique natural number, but proving it from our definitions is required. In essence, while it is difficult to imagine it happening, we need to be sure that there do not exist a set X and functions $f : \mathbb{N}_n \underset{\text{onto}}{\overset{\text{1-1}}{\to}} X$ (which says $|X| = n > 0$) and $g : \mathbb{N}_m \underset{\text{onto}}{\overset{\text{1-1}}{\to}} X$ (which says $|Y| = m > 0$) with $n \neq m$.

Proving that the cardinality of a nonempty finite set A is a unique natural number relies on the following "common sense" fact. We'll leave the proof of this fact for Subsection 8.2.1.

Theorem 8.2.1. *For all $n \in \mathbb{N}$, there does not exist a bijection $f : \mathbb{N}_n \to X$, where X is a proper subset of $\mathbb{N}_n$.*

Taking $X = \mathbb{N}_m$, where $m < n$, Theorem 8.2.1 implies that there does not exist a 1-1 function $f : \mathbb{N}_n \to \mathbb{N}_m$. This fact is known as the "Pigeonhole Principle". In more colorful language:

> If $n > m$ and n pigeons are put into m pigeonholes, then at least one pigeonhole contains more than one pigeon.

Along with the Addition and Multiplication Principles, the Pigeonhole Principle is an important mathematical tool in combinatorics. The Pigeonhole Principle seems like another "obvious" statement, and it is often treated as such in mathematics courses that make use of it.

Accepting Theorem 8.2.1 for now without proof, we see that this result is the tool we need to show that our definition of a finite set agrees with our intuition about finite sets.

Corollary 8.2.2. *The cardinality of a finite set is well-defined; i.e., for all sets X, if X is finite, then there exists a unique $n \in \mathbb{Z}$, $n \geq 0$, such that $|X| = n$.*

Proof. Assume that the cardinality of the finite set X is not well-defined. We cannot have $|X| = 0$ and $|X| = n$ for some $n \in \mathbb{N}$, since $|X| = 0$ implies that $X = \emptyset$ and $|X| = n$, $n \in \mathbb{N}$, implies that $X \neq \emptyset$.

Hence we may fix $n, m \in \mathbb{N}$ with $n > m$ and bijections $f : \mathbb{N}_n \underset{\text{onto}}{\overset{\text{1-1}}{\to}} X$ and $g : \mathbb{N}_m \underset{\text{onto}}{\overset{\text{1-1}}{\to}} X$. Then $g^{-1} : X \underset{\text{onto}}{\overset{\text{1-1}}{\to}} \mathbb{N}_m$ by Corollary 5.4.9, and so $g^{-1} \circ f : \mathbb{N}_n \overset{\text{1-1}}{\to} \mathbb{N}_m$ by Theorem 5.3.10(1). Since $n > m$, this contradicts the Pigeonhole Principle. Hence there is a unique $n \in \mathbb{Z}$ with $n \geq 0$ such that $|X| = n$. □

Corollary 8.2.3. *Let A be a set.*

(1) *If A is finite, then A is not equinumerous with a proper subset of itself.*

(2) *If A is equinumerous with a proper subset of itself, then A is infinite.*

Proof. Exercise 8.2.2. □

Corollary 8.2.4. *The set $\mathbb{N}$ of natural numbers is infinite.*

Proof 1. Suppose for the sake of a contradiction that $\mathbb{N}$ is finite. Since $\mathbb{N} \neq \emptyset$, we may fix $n \in \mathbb{N}$ such that $\mathbb{N} \approx \mathbb{N}_n$. So, we fix a function $f : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} \mathbb{N}_n$. Define the function $g : \mathbb{N}_{n+1} \to \mathbb{N}_n$ by, for all $k \in \mathbb{N}_{n+1}$, $g(k) = f(k)$. (The function g is called the *restriction* of f to $\mathbb{N}_{n+1}$ and is sometimes denoted by $f \restriction \mathbb{N}_{n+1}$.) Then g is 1-1, since f is, contradicting the Pigeonhole Principle. Hence $\mathbb{N}$ is infinite. □

Proof 2. By Exercise 8.1.3, $\mathbb{N}$ is equinumerous with a proper subset of itself, namely, the set of odd positive integers. Hence by Corollary 8.2.3(2), $\mathbb{N}$ is infinite. □

We're ready now to prove some theorems about cardinalities of finite sets.

Theorem 8.2.5. *Let A and B be finite sets with $A \cap B = \emptyset$, and let $n, m \geq 0$ be natural numbers such that $|A| = n$ and $|B| = m$. Then $A \cup B$ is finite and $|A \cup B| = n + m$.*

Scratchwork. For now, let's assume that $n, m \neq 0$. It's worth writing down a Given-Goal diagram.

Given	**Goal**
$m, n \neq 0$	
$\lvert A\rvert = n$, $\lvert B\rvert = m$	
$A \cap B = \emptyset$	$\lvert A \cup B\rvert = n + m$

We can use Definition 8.1.2(2) to immediately rewrite this as follows.

Given	**Goal**
$m, n \neq 0$	
$A \cap B = \emptyset$	
$f : \mathbb{N}_n \underset{\text{onto}}{\overset{\text{1-1}}{\to}} A$	
$g : \mathbb{N}_m \underset{\text{onto}}{\overset{\text{1-1}}{\to}} B$	$h : \mathbb{N}_{n+m} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} A \cup B$

Informally, we can see exactly why $|A \cup B| = n + m$. As discussed above, the bijection f can be used to list A (here $f(i) = a_i$ for all $i \in \mathbb{N}_n$):

$$a_1, a_2, \ldots, a_n,$$

and the bijection g can be used to list B (here $g(i) = b_i$ for all $i \in \mathbb{N}_m$):

$$b_1, b_2, \ldots, b_m.$$

Thus we must construct the bijection h which lists $A \cup B$ as

$$a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m.$$

In otherwords, we need to construct the bijection

$$\begin{array}{cccccccc} 1 & 2 & \ldots & n & n+1 & n+2 & \ldots & n+m \\ \downarrow & \downarrow & \ldots & \downarrow & \downarrow & \downarrow & \ldots & \downarrow \\ a_1 & a_2 & \ldots & a_n & b_1 & b_2 & \ldots & b_m. \end{array}$$

The fact that $A \cap B = \emptyset$ will be needed to show that there are no repetitions in this listing of $A \cup B$, i.e., that the function h is 1-1, so that there are actually $n + m$ elements in this list.

The proof of the theorem simply expresses these ideas formally and, of course, checks the details.

Proof. Let A and B be finite sets with $A \cap B = \emptyset$, and let $n, m \geq 0$ be such that $|A| = n$ and $|B| = m$. We assume that $A \neq \emptyset$ and $B \neq \emptyset$, leaving the case when A or B is empty for Exercise 8.2.1.

By Definition 8.1.2(2), we may fix $m, n \in \mathbb{N}$ and bijections $f : \mathbb{N}_n \underset{\text{onto}}{\overset{1\text{-}1}{\to}} A$ and $g : \mathbb{N}_m \underset{\text{onto}}{\overset{1\text{-}1}{\to}} B$. We show $|A \cup B| = n + m$ by constructing an explicit bijection $h : \mathbb{N}_{n+m} \underset{\text{onto}}{\overset{1\text{-}1}{\to}} A \cup B$. Define h by, for $i \in \mathbb{N}_{n+m}$,

$$h(i) = \begin{cases} f(i) & \text{if } 1 \leq i \leq n, \\ g(i-n) & \text{if } n+1 \leq i \leq n+m. \end{cases}$$

We leave the verification that h is a bijection for Exercise 8.2.1. □

Corollary 8.2.6. *Let $A_1, A_2, \ldots, A_k$, where $k \in \mathbb{N}$, be a family of pairwise disjoint finite sets (i.e., $A_i \cap A_j = \emptyset$ when $i \neq j$). Then the set $\bigcup_{i=1}^{k} A_i = A_1 \cup A_2 \cup \cdots \cup A_k$ is finite and $|A_1 \cup A_2 \cup \cdots \cup A_k| = |A_1| + |A_2| + \cdots + |A_k|$.*

Proof. The proof is by induction on k, the number of sets. See Exercise 8.2.4. □

Theorem 8.2.7. *Let A and B be finite sets, and let $n, m \geq 0$ be such that $|A| = n$ and $|B| = m$. Then $A \times B$ is finite, and $|A \times B| = nm$.*

Proof (Informal Sketch). Let A and B be finite sets, and let $n, m \geq 0$ be such that $|A| = n$ and $|B| = m$. We assume that $A \neq \emptyset$ and $B \neq \emptyset$ (i.e., $n, m > 0$), leaving the case when A or B is empty for Exercise 8.2.7.

We have a bijection establishing

$$A = \{a_1, a_2, \ldots, a_n\}$$

and, similarly,

$$B = \{b_1, b_2, \ldots, b_m\}.$$

We must construct a bijection which lists all the elements of $A \times B$ without repetition. We can visualize counting $A \times B$ using n rows, each containing m ordered pairs:

$$\begin{gathered} (a_1, b_1), (a_1, b_2), \ldots, (a_1, b_m) \\ (a_2, b_1), (a_2, b_2), \ldots, (a_2, b_m) \\ \vdots \\ (a_n, b_1), (a_n, b_2), \ldots, (a_n, b_m). \end{gathered} \tag{8.1}$$

The fact that $(x, y) = (u, v)$ if and only if $x = u$ and $y = v$ is used to show that there are no repetitions in this listing of $A \times B$, so that there are actually nm elements in this list. See Exercise 8.2.7. □

As before, the proof of the theorem simply expresses these ideas formally and, of course, checks the details.

Corollary 8.2.8. *Let $A_1, A_2, \ldots, A_k$, where $k \in \mathbb{N}$, be a family of finite sets. Then $A_1 \times A_2 \times \cdots \times A_k$ is finite and*

$$|A_1 \times A_2 \times \cdots \times A_k| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_k|.$$

Proof. The proof is by induction on k, the number of sets. See Exercise 8.2.8. □

Theorems 8.2.5 and 8.2.7 are important tools in the mathematical subject called *combinatorics*, which studies methods of *counting* (see [**5**]). For example, a typical exercise in combinatorics is to count the number of five-card poker hands that are "full houses" (i.e., three of one face value, or kind, and two of a different kind). In combinatorics, Theorem 8.2.5 is called the *Addition Principle*, and it says that the number of ways of making a single selection from categories A or B is equal to $|A|+|B|$ when the categories are finite and disjoint. Theorem 8.2.7 is typically called the *Multiplication Principle*; the Multiplication Principle states that the number of ways to first select one object from a finite category A and then, independently of the object chosen, select one object from a finite category B is $|A| \cdot |B|$. Both of these statements generalize as in Corollaries 8.2.6 and 8.2.8.

The next counting result first appeared as an exercise in the chapter on induction: Exercise 4.2.26. Recall that if A is a set, then $\mathcal{P}(A)$ denotes the power set of A, i.e., the set of all subsets of A.

Theorem 8.2.9. *Let A be a finite set, and let $n \geq 0$ be such that $|A| = n$. Then $\mathcal{P}(A)$ is finite and $|\mathcal{P}(A)| = 2^n$.*

Scratchwork. This result can be proved by induction on the cardinality of A. More precisely, one proves the following statement by induction on $n \geq 0$:

$$\boxed{\begin{array}{l}\text{for all } n \geq 0, \text{ for all sets } A,\\ \text{if } |A| = n, \text{ then } |\mathcal{P}(A)| = 2^n.\end{array}}$$

Thinking of the statement of the theorem in this way will ensure that the inductive hypothesis is general enough to be useful to us, since it is a *universally quantified* statement about sets. The inductive step assumes the result for $n \geq 0$; namely,

$$(\dagger) \qquad \boxed{\text{for any set } X, \text{ if } |X| = n, \text{ then } |\mathcal{P}(X)| = 2^n.}$$

We then assume we have a set $A = \{a_1, a_2, \ldots, a_{n+1}\}$ and apply the inductive hypothesis (†) to a *different* set X in order to show $|\mathcal{P}(A)| = 2^{n+1}$.

A small example helps to illustrate how the inductive hypothesis (†) is applied using a *new* set X. Let $A = \{a_1, a_2, a_3\}$ have cardinality 3. Note that

$$\mathcal{P}(A) = \{\emptyset, \{a_1\}, \{a_2\}, \{a_3\}, \{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_3\}, \{a_1, a_2, a_3\}\}.$$

We will think of $\mathcal{P}(A)$ differently, by choosing a designated element of A, say a_3, and dividing the elements of $\mathcal{P}(A)$ into two categories: those subsets of A not containing a_3 and those subsets of A containing a_3.

Subsets of A not containing a_3	**Subsets of A containing a_3**
$\emptyset$	$\{a_3\}$
$\{a_1\}$	$\{a_1, a_3\}$
$\{a_2\}$	$\{a_2, a_3\}$
$\{a_1, a_2\}$	$\{a_1, a_2, a_3\}$

Note that the subsets of A not containing a_3 are just the subsets of the set $X = A - \{a_3\} = \{a_1, a_2\}$, to which the inductive hypothesis applies. Thus, there are $2^2 = 4$ sets in this category. Each subset of A containing $\{a_3\}$ in the second category arises from adding a_3 to a set in the first category. This means that there are $2^2 = 4$ sets in the second category, and hence $4 + 4 = 8$ subsets of A overall.

The proof of the theorem expresses these ideas formally and checks the details.

Proof. Exercise 8.2.9. □

We end with another "common sense" result about subsets of finite sets. We include this proof, as it gives another good example of an inductive proof of a universally quantified statement.

Theorem 8.2.10. *Let A and B be sets with $A \subseteq B$. If B is finite, then A is also finite and $|A| \leq |B|$. Furthermore, $A \subsetneq B$ iff $|A| < |B|$.*

Scratchwork. We will prove the first statement by induction on the cardinality of B. More precisely, we will be proving the following statement by induction on n:

$$\text{(8.2)} \qquad \boxed{\begin{array}{l}\text{for all } n \geq 0, \text{ for all sets } A \text{ and } B,\\ \text{if } A \subseteq B \text{ and } |B| = n, \text{ then } A \text{ is finite and } |A| \leq |B|.\end{array}}$$

Thinking of the statement of the theorem in this way will ensure that the inductive hypothesis is general enough to be useful to us. As in the induction proof of Theorem 8.2.9, the inductive step involves choosing a designated element $b_{n+1} \in B$. Our cases depend on whether or not $b_{n+1} \in A$, and we apply the induction hypothesis to a *different* set.

Proof. We prove statement (8.2) by induction on n.

Base Case: Let A and B be sets with $A \subseteq B$ and $|B| = 0$. We must show that A is finite.

Since $|B| = 0$, it must be the case that $B = \emptyset$. Since $A \subseteq B$, we must also have $A = \emptyset$, and hence A is finite by Definition 8.1.2(2) and $|A| = |B| = 0$.

Inductive Step: Let $m \geq 0$ and assume that for all sets X and Y, if $X \subseteq Y$ and $|Y| = m$, then X is finite and $|X| \leq |Y|$ (this is our inductive hypothesis).

Next, let A and B be sets with $A \subseteq B$ and $|B| = m + 1$. We must prove that A is finite and $|A| \leq m + 1$.

Since $|B| = m + 1$, we may fix a bijection $f : \mathbb{N}_{m+1} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} B$; i.e., $B = \{b_1, b_2, \ldots, b_m, b_{m+1}\}$, where $f(i) = b_i$ for all $1 \leq i \leq m + 1$. We consider two cases.

Case I: $A \subseteq \{b_1, b_2, \ldots, b_m\}$.

Then $Y = \{b_1, b_2, \ldots, b_m\}$ is a finite set with $|Y| = m$. Hence by the inductive hypothesis (with $X = A$), A is finite and

$$|A| \leq |Y| = m < m + 1 = |B|.$$

Case II: Otherwise.

Then $b_{m+1} \in A$. Let $X = A - \{b_{m+1}\}$. Note that $A = X \cup \{b_{m+1}\}$ and $X \cap \{b_{m+1}\} = \emptyset$. Since $A \subseteq B$, we have $X \subseteq Y = \{b_1, b_2, \ldots, b_m\}$, where

Y has cardinality m. Thus, again by the inductive hypothesis, X is a finite set and $|X| \leq |Y| = m$. Hence $A = X \cup \{b_{m+1}\}$ is finite and

$$|A| = |X| + 1 \leq m + 1 = |B|,$$

by Theorem 8.2.5.

It follows by induction that statement (8.2) is true, and hence any subset of a finite set is finite. The fact that $A \subsetneq B$ iff $|A| < |B|$ is left for Exercise 8.2.11. □

The following corollary is immediate.

Corollary 8.2.11. *Let A and B be sets with $A \subseteq B$. If A is infinite, then B is infinite.*

8.2.1. Debts paid. We must now pay our debts and prove Theorem 8.2.1. Our goal is to prove the result from our definitions, rather than relying on our intuition about finite sets. Students who are encountering these ideas for the first time may wish to omit this subsection on their first reading.

Theorem 8.2.12. *For all $n \in \mathbb{N}$, there does not exist a bijection $f : \mathbb{N}_n \to X$, where X is a proper subset of $\mathbb{N}_n$.*

Scratchwork: We will prove the result by induction on n. Let's discuss the strategy of the inductive step.

We let $m \geq 1$ and assume that for all sets $X \subsetneq \mathbb{N}_m$, there does not exist a bijection from $\mathbb{N}_m \to X$ (this is our inductive hypothesis). Let $A \subsetneq \mathbb{N}_{m+1}$, and assume for the sake of a contradiction that we have $f : \mathbb{N}_{m+1} \xrightarrow[\text{onto}]{\text{1-1}} A$. We need to convert this f into a form to which the inductive hypothesis applies.

Given	**Goal**
$m \geq 1$	
$A \subsetneq \mathbb{N}_{m+1}$	find $X \subsetneq \mathbb{N}_m$ and g such that
$f : \mathbb{N}_{m+1} \xrightarrow[\text{onto}]{\text{1-1}} A$	$g : \mathbb{N}_m \xrightarrow[\text{onto}]{\text{1-1}} X$

To move from $\mathbb{N}_{m+1}$ to $\mathbb{N}_m$, we must consider $m+1$, and the question is whether or not $m + 1 \in A$.

If $m + 1 \notin A$, then $A \subseteq \mathbb{N}_m$, so to apply the inductive hypothesis, we consider the effect of f on $\mathbb{N}_m$, which will map $\mathbb{N}_m$ to the *proper* subset $A - \{f(m+1)\}$ of $\mathbb{N}_m$ (i.e., we will *restrict* f to $\mathbb{N}_m$).

If $m + 1 \in A$, then we can fix $k \in \mathbb{N}_{m+1}$ such that $f(k) = m + 1$. If $k = m + 1$, then we can restrict f to $\mathbb{N}_m$ and consider the proper subset $A - \{m+1\}$. Otherwise, we have the picture in Figure 8.1.

In this case, we should swap the values of $f(k)$ and $f(m+1)$ and then restrict f to $\mathbb{N}_m$.

Proof. We prove Theorem 8.2.12 by induction on n.

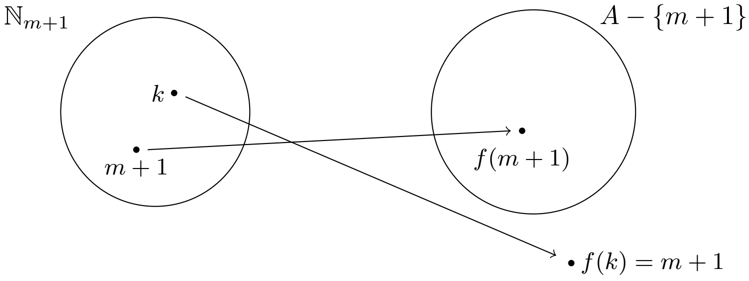


Figure 8.1. The case when $f(k) = m+1 \in A$, $k \neq m+1$.

Base Case: When $n = 1$, the only proper subset of $\mathbb{N}_1 = \{1\}$ is $\emptyset$. There is no bijection $f : \{1\} \to \emptyset$, as desired.

Inductive Step: Let $m \in \mathbb{N}$ and assume the result for m; i.e., we assume that for all proper subsets $X \subsetneq \mathbb{N}_m$, there does not exist a bijection from $\mathbb{N}_m$ to X.

We prove the result for $m+1$. Let $A \subsetneq \mathbb{N}_{m+1}$, and assume for the sake of a contradiction that we have a bijection $f : \mathbb{N}_{m+1} \to A$.

Case I: $m+1 \notin A$.

In this case, note that $A \subseteq \mathbb{N}_m$ and the function $g = f \restriction \mathbb{N}_m$ (i.e., $g : \mathbb{N}_m \to A - \{f(m+1)\}$ by, for all $i \leq m$, $g(i) = f(i)$) is a bijection from $\mathbb{N}_m$ to $A - \{f(m+1)\} \subsetneq A \subseteq \mathbb{N}_m$. This contradicts the induction hypothesis for the proper subset $X = A - \{f(m+1)\} \subsetneq \mathbb{N}_m$.

Case II: $m+1 \in A$.

Then, since f is onto, we may fix $k \in \mathbb{N}_{m+1}$ such that $f(k) = m+1$. Note that if $k \neq m+1$, then $f(m+1) \in \mathbb{N}_m$ since f is 1-1. Consider $g : \mathbb{N}_m \to A - \{m+1\}$ by, for all $i \in \mathbb{N}_m$,

$$g(i) = \begin{cases} f(i) & \text{if } i \neq k, \\ f(m+1) & \text{if } i = k. \end{cases}$$

Then g is a bijection from $\mathbb{N}_m$ to the proper subset $A - \{m+1\} \subsetneq \mathbb{N}_m$, contradicting the induction hypothesis.

Hence, it follows by induction that for all $n \in \mathbb{N}$, there does not exist a bijection $f : \mathbb{N}_n \to X$, where X is a proper subset of $\mathbb{N}_n$. □

Exercises 8.2

1. Complete the proof of Theorem 8.2.5:
   (a) Prove the result in the case that $A = \emptyset$ or $B = \emptyset$.
   (b) Prove that the function h defined in the proof is a bijection. (**HINT:** The proof that h is 1-1 is a proof by cases, similar to the proof in Example 8.1.4. Cases are also needed to prove that h is onto.)

2. Prove Corollary 8.2.3.

3. Let A and B be finite sets with $|A| = n$ and $|B| = m$, where $m, n \in \mathbb{Z}$ with $n > m \geq 0$. Prove that there is no 1-1 function $f : A \overset{\text{1-1}}{\to} B$.

4. Prove Corollary 8.2.6.

5. Let A and B be finite sets.
   (a) Prove that $A \cap B$ is finite.
   (b) Prove that $A - B$ is finite and that $|A - B| = |A| - |A \cap B|$.
   (c) Prove that $A \cup B$ is finite and that $|A \cup B| = |A| + |B| - |A \cap B|$.

6. Let A be a finite set and B be an infinite set. Prove that $B - A$ is infinite.

7. Prove Theorem 8.2.7. (**HINT:** Consider the lines of (8.1) as listing the elements of the sets $\{a_1\} \times B$, $\{a_2\} \times B$, $\ldots$, $\{a_n\} \times B$.)

8. Prove Corollary 8.2.8.

9. Prove Theorem 8.2.9
   (a) Use a proof by induction on $n = |A|$ (see the scratchwork in Section 8.2).
   (b) Use a less formal argument invoking the Multiplication Principle.

10. Let A and B be nonempty finite sets. Use the Multiplication Principle to show that $|\{f \mid f : A \to B\}| = |B|^{|A|}$.

11. Complete the proof of Theorem 8.2.10. Let A and B be sets with $A \subseteq B$. Prove that $A \subsetneq B$ iff $|A| < |B|$. (**HINT:** Use Corollary 8.2.3(1).)

12. Let A be a nonempty finite set, $B \neq \emptyset$, and $f : A \to B$. Prove that the image $f[A]$ is also finite and $|f[A]| \leq |A|$. Furthermore f is 1-1 iff $|f[A]| = |A|$.

13. Let A and B be nonempty finite sets with $|A| = |B|$, and let $f : A \to B$. Prove that f is 1-1 iff f is onto.

14. Let $n \in \mathbb{Z}^+$ and $A = \{a_1, a_2, \ldots, a_n\} \subseteq \mathbb{R}$ be a finite set of real numbers. We say that $M \in \mathbb{R}$ is a *maximum element of* A if $M \in A$ and for all $x \in A$, $x \leq M$. Similarly, we say that $m \in \mathbb{R}$ is a *minimum element of* A if $m \in A$ and for all $x \in A$, $m \leq x$.
    (a) Using induction on the cardinality of the set, prove that every nonempty finite set has both a maximum and a minimum element. That is, prove the following by induction on n:

    for all $n > 0$, for all sets $A \subseteq \mathbb{R}$ with $|A| = n$,
    there exist $m, M \in A$ such that
    for all $x \in A$, $m \leq x \leq M$.

    (b) Prove that the maximum and minimum elements of a finite set are unique. That is, prove that if $M_1, M_2 \in \mathbb{R}$ are both maximums of the nonempty finite set A, then $M_1 = M_2$, and analogously for minimums.

8.3. Infinite sets

In this section, we specifically study infinite sets. Our focus is on material most useful for a future course in real analysis, and our ultimate goal is to prove that $\mathbb{R}$ is not equinumerous with $\mathbb{Q}$.

Recall that in Example 8.1.4 we showed that $\mathbb{N} \approx \mathbb{Z}$. Sets that are equinumerous with $\mathbb{N}$ are called "denumerable" or "countably infinite", since they can be "listed" or "enumerated" without repetition. Recall the picture from Example 8.1.4, which shows this enumeration of $\mathbb{Z}$:

$$0, 1, -1, 2, -2, 3, \dots .$$

Definition 8.3.1.

(1) The set X is *denumerable* (*enumerable*) if there exists a bijection $f : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} X$, i.e., if $\mathbb{N} \approx X$.

(2) The set X is *countable* if X is finite or denumerable.

(3) The set X is *uncountable* if X is not countable, i.e., if X is infinite and not denumerable.

It's worth reiterating that, informally, a set X is denumerable if X can be listed without repetition: if $f : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} X$, then we can list the elements of X as

$$f(1), f(2), f(3), \dots .$$

If we define $x_i = f(i)$ for all $i \in \mathbb{N}$, then we can write X as

$$X = \{x_1, x_2, x_3, \dots\}.$$

Note that because $\mathbb{N}$ is infinite (see Corollary 8.2.4), every denumerable set is infinite (by Exercise 8.1.10b).

The use of the word "countable" to describe the sets that are either finite or denumerable is deliberate. Informally, a countable set is one whose elements can be "counted", or listed, although the list may or may not "end". Note that some authors define the terms "denumerable" and "countable" to have the opposite meanings, so be sure to check the definitions when using a different source.

The goal of this section is to prove that the set $\mathbb{Q}$ of rational numbers is denumerable, while the set $\mathbb{R}$ of real numbers is uncountable. This will show that $\mathbb{Q}$ and $\mathbb{R}$ are not equinumerous, i.e., the (possibly) surprising fact that infinite sets do not all have the same "size". We will prove the results necessary to establish these facts, leaving a more detailed investigation of the "size" (cardinality) of an infinite set for a course in set theory (see [**9**] or [**12**]). In some cases, we will provide an informal sketch of the proof, leaving a formal proof for the exercises or for Subsection 8.3.1. Students who are new to these ideas may wish to omit Subsection 8.3.1 on their first reading.

We first note that, while an infinite set A is countable if there is a bijection from $\mathbb{N}$ onto A, in fact, showing there is a surjection from $\mathbb{N}$ onto A is enough (and often easier).

Theorem 8.3.2. *Let A be a nonempty set. If there exists a surjective function $g : \mathbb{N} \underset{onto}{\to} A$, then A is countable; i.e., A is finite or denumerable.*

Proof (Informal Sketch). Assume that $g : \mathbb{N} \xrightarrow[\text{onto}]{} A$. If A is finite, then A is countable, and we're done. So we assume that A is infinite and show that A is denumerable by finding a bijection $f : \mathbb{N} \xrightarrow[\text{onto}]{\text{1-1}} A$.

The idea here is that g may not be 1-1, so g may list elements of A more than once. To define f, we must "skip over" elements of A that have already been listed by g and construct a new list with no "repeats" that contains all the elements of A.

We can visualize the proof as follows, with the enumerations given vertically.

Enumeration of A with repeats	Repeats	New enumeration of A
$g(1)$		$f(1)$
$g(2)$	$= g(1)$	
$g(3)$	$= g(1)$	
$g(4)$		$f(2)$
$g(5)$		$f(3)$
$g(6)$	$= g(4)$	
$g(7)$		$f(4)$
$g(8)$	$= g(5)$	
⋮	⋮	⋮

□

Proof. See Subsection 8.3.1. □

Theorem 8.3.3. *Let A and B be denumerable sets. Then $A \cup B$ is also denumerable.*

Scratchwork. Let A and B be denumerable, and fix bijections $f : \mathbb{N} \xrightarrow[\text{onto}]{\text{1-1}} A$ and $g : \mathbb{N} \xrightarrow[\text{onto}]{\text{1-1}} B$. Thus, we can list the elements of A:

$$f(1), f(2), f(3), f(4), \ldots$$

or more simply:

$$a_1, a_2, a_3, a_4, \ldots,$$

and we can list the elements of B:

$$g(1), g(2), g(3), g(4), \ldots$$

or more simply:

$$b_1, b_2, b_3, b_4, \ldots.$$

Unlike the proof of the analogous result for finite sets (Theorem 8.2.5), we cannot show that $A \cup B$ is denumerable by first listing A and then listing B, since the list for A doesn't end.

So, we will instead construct a list of the elements of $A \cup B$ by interleaving or "dovetailing"[2] the enumerations of A and B:

$$a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4, \ldots.$$

[2] The terminology here is deliberate. A dovetail is a type of carpenter's joint, such as the joint between the sides and front of a drawer, that interlocks the objects being joined.

This shows that $A \cup B$ is denumerable by Theorem 8.3.2, since every element of $A \cup B$ (which is infinite) appears (at least once) on this list.

Proof. Let A and B be denumerable, and fix bijective functions $f : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} A$ and $g : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} B$. Define $h : \mathbb{N} \to A \cup B$ by, for all $n \in \mathbb{N}$,

$$h(n) = \begin{cases} f(\frac{n}{2}) & \text{if } n \text{ is even,} \\ g(\frac{n+1}{2}) & \text{if } n \text{ is odd.} \end{cases}$$

We show that h is onto. Assume $y \in A \cup B$.

Case I: $y \in A$.

Since f is onto, we may fix $n \in \mathbb{N}$ such that $f(n) = y$. Then $h(2n) = f(\frac{2n}{2}) = f(n) = y$.

Case II: $y \notin A$.

Then $y \in B$, since $y \in A \cup B$. Since g is onto, we may fix $n \in \mathbb{N}$ such that $g(n) = y$. Note that $2n - 1 \in \mathbb{N}$ and $h(2n-1) = g\left(\frac{(2n-1)+1}{2}\right) = y$.

Hence h is onto. It follows by Theorem 8.3.2 that $A \cup B$ is countable.

Since $A \subseteq A \cup B$ and A is infinite, $A \cup B$ is also infinite by Corollary 8.2.11. Hence $A \cup B$ is denumerable. □

Note that in the proof above, the listing of the elements of $A \cup B$ constructed by the function h will have repeats when $A \cap B \neq \emptyset$, but what matters is that the list contains all elements of $A \cup B$.

Corollary 8.3.4. *The union of finitely many denumerable sets is denumerable.*

Proof. The proof is by induction on k, the number of sets. See Exercise 8.3.7. □

A slight modification gives the following.

Corollary 8.3.5. *The union of finitely many countable sets is countable.*

Just these few tools are enough to sketch a proof that $\mathbb{Q}$ is denumerable.

Theorem 8.3.6 (Cantor, 1874)**.** *The set $\mathbb{Q}$ of rational numbers is denumerable.*

Proof (Informal Sketch). Note that

$$\begin{aligned} \mathbb{Q} &= \left\{ \frac{a}{b} \;\middle|\; a, b \in \mathbb{Z} \text{ and } b > 0 \right\} \\ &= \mathbb{Q}^+ \cup \mathbb{Q}^- \cup \{0\}, \end{aligned}$$

where $\mathbb{Q}^+ = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}^+ \right\}$ and $\mathbb{Q}^- = \left\{ -\frac{a}{b} \mid a, b \in \mathbb{Z}^+ \right\}$.

We first show that $\mathbb{Q}^+$ is denumerable. Begin by writing the elements of $\mathbb{Q}^+$, with repeats, in a rectangular array; in the first row we list all rational numbers with 1 as a denominator (this set is denumerable, since we've listed it), in the second row we list all rational numbers with 2 as a denominator (this set is denumerable, since we've listed it), etc. We now dovetail the enumerations. See Figure 8.2.

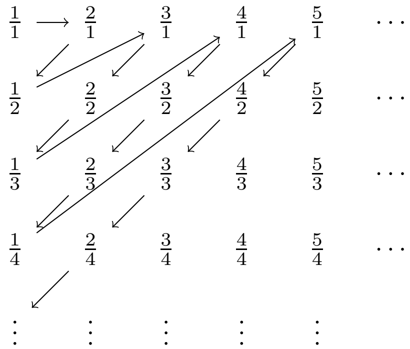


Figure 8.2. Dovetailing the enumeration of $\mathbb{Q}^+$.

We claim that every positive rational number is on this list (which has repeats); i.e., we can write down a function $f : \mathbb{N} \underset{\text{onto}}{\rightarrow} \mathbb{Q}^+$, as shown. So $\mathbb{Q}^+$ is countable, by Theorem 8.3.2, and hence denumerable, since $\mathbb{Q}^+$ is infinite.

We next note that $\mathbb{Q}^+ \approx \mathbb{Q}^-$, since the function $g : \mathbb{Q}^+ \to \mathbb{Q}^-$ by, for all $x \in \mathbb{Q}^+$, $g(x) = -x$ is a bijection. Since $\approx$ is transitive, $\mathbb{Q}^-$ is also denumerable.

Since $\mathbb{Q}^+$ is infinite, it follows that $\{0\} \cup \mathbb{Q}^+ \cup \mathbb{Q}^-$ is denumerable, by Corollary 8.3.5. □

Proof. Two rigorous proofs that $\mathbb{Q}^+$ is denumerable are outlined in Exercise 8.3.12. □

The method used to show that $\mathbb{Q}^+$ is denumerable is called "Cantor's first diagonalization method".

A similar argument will show that the Cartesian product $\mathbb{N} \times \mathbb{N}$ is denumerable; see Exercise 8.3.9. The following theorem is then easy to prove.

Theorem 8.3.7. *Let A and B be denumerable sets. Then $A \times B$ is denumerable.*

Proof. Exercise 8.3.10. □

The first diagonalization method can also be used to prove that a denumerable union of denumerable sets is denumerable, as long as one is willing to use the Axiom of Choice (see Section 4.4). So as to avoid the Axiom of Choice, we state a special case of this theorem, which you are asked to prove pictorially in Exercise 8.3.11.

Theorem 8.3.8. *Assume that the enumerations*

$$
\begin{aligned}
A_1 &= \{a_1^1, a_1^2, a_1^3, \dots\},\\
A_2 &= \{a_2^1, a_2^2, a_2^3, \dots\},\\
A_3 &= \{a_3^1, a_3^2, a_3^3, \dots\},\\
&\vdots
\end{aligned}
$$

of denumerable sets $A_1, A_2, A_3, \ldots$ are given. Then the set $A = \bigcup_{i \in \mathbb{N}} A_i$ is denumerable.

Proof. Exercise 8.3.11. □

For our proof that the set of real numbers is uncountable, we need one last theorem. Informally, it says that if we can list the elements of a set, then we can list the elements of any subset of that set. The proof is similar to the proof of Theorem 8.3.2.

Theorem 8.3.9. *Let A be denumerable and let $B \subseteq A$ be infinite. Then B is denumerable.*

Proof (Informal Sketch). Assume that A is denumerable and fix $f : \mathbb{N} \overset{\text{1-1}}{\underset{\text{onto}}{\to}} A$. Let $B \subseteq A$ be infinite. We must define a function $g : \mathbb{N} \overset{\text{1-1}}{\underset{\text{onto}}{\to}} B$ to show that B is denumerable. The idea is to use f to enumerate A but "skip over" any elements of A that are not also in B. The example below shows the general idea, with the enumerations of A and B given vertically.

enumeration of A	in B?	enumeration of B
$f(1)$	No	
$f(2)$	Yes	$g(1)$
$f(3)$	Yes	$g(2)$
$f(4)$	No	
$f(5)$	Yes	$g(3)$
$f(6)$	No	
$f(7)$	No	
$f(8)$	Yes	$g(4)$
$\vdots$	$\vdots$	$\vdots$

□

Proof. Exercise 8.3.15. □

Corollary 8.3.10. *Let A and B be sets with $B \subseteq A$.*

(1) *If A is countable, then B is countable.*

(2) *If B is uncountable, then A is uncountable.*

Proof. Exercise 8.3.16. □

We are now ready to show that the set $\mathbb{R}$ of real numbers is uncountable.

Theorem 8.3.11 (Cantor, 1874). *$\mathbb{R}$ is uncountable.*

Proof. Recall that to show that a set is uncountable we must show that it is neither finite nor denumerable. Since $\mathbb{R}$ is infinite, we must show that it is not denumerable. We will do this using a proof by contradiction. First, we must give some necessary background information about real numbers, which we present here without proof and discuss further in Chapter 9.

Claim. Every real number has a decimal expansion

$$d.d_1d_2d_3d_4\ldots,$$

where $d \in \mathbb{Z}$ and for all $i \in \mathbb{Z}^+$, $d_i \in \mathbb{Z}$ with $0 \leq d_i \leq 9$. The integer d is called the *integer part* of the real number, and $.d_1d_2d_3d_4\ldots$ is called the *decimal part*.

In most cases, the decimal expansion of a real number is unique, except for instances such as

$$0.4\overline{9} = 0.49999\ldots = 0.5 = 0.50000\ldots.$$

The interval $[0,1]$ consists of all real numbers that can expressed with an integer part of 0. We will show that $[0,1]$ is uncountable; it follows that $\mathbb{R}$ is uncountable by Corollary 8.3.10.

Assume for the sake of a contradiction that $[0,1]$ is countable. Since $[0,1]$ is infinite, it is denumerable. Hence we may fix a function $f : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} [0,1]$ that lists the real numbers in $[0,1]$ as $f(0), f(1), f(2), \ldots$. Given $i, j \in \mathbb{N}$, let d_i^j denote the ith decimal digit of the real number $f(j)$. Thus, the real numbers in $[0,1]$ form the following (now vertical) list:

$$\begin{aligned}
f(1) &= 0.d_1^1d_2^1d_3^1d_4^1d_5^1\ldots,\\
f(2) &= 0.d_1^2d_2^2d_3^2d_4^2d_5^2\ldots,\\
f(3) &= 0.d_1^3d_2^3d_3^3d_4^3d_5^3\ldots,\\
f(4) &= 0.d_1^4d_2^4d_3^4d_4^4d_5^4\ldots,\\
f(5) &= 0.d_1^5d_2^5d_3^5d_4^5d_5^5\ldots,\\
&\vdots
\end{aligned}$$

We obtain a contradiction by constructing a real number

$$r = 0.r_1r_2r_3r_4r_5\ldots$$

in the interval $[0,1]$ such that $r \notin \operatorname{ran} f$. Given $i \in \mathbb{N}$, the ith decimal digit r_i of r is defined as follows:

$$r_i = \begin{cases} 7 & \text{if } d_i^i \neq 7,\\ 2 & \text{if } d_i^i = 7. \end{cases}$$

Since $f : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} [0,1]$ and $r \in [0,1]$, $r \in \operatorname{ran} f$. Thus we can fix $k \in \mathbb{Z}^+$ such that

$$r = f(k) = 0.d_1^kd_2^kd_3^kd_4^kd_5^k\ldots d_k^k\ldots.$$

The kth decimal digit of r is r_k, by definition. Also, the kth decimal digit of r is d_k^k, since $r = f(k)$. However, $r_k \neq d_k^k$ by definition of r_k. Since r is not a real number with two different decimal expansions (we have avoided the repeating 9's problem), $r \neq f(k)$, a contradiction.

Thus $[0,1]$, and hence $\mathbb{R}$, is uncountable. □

To help us understand the proof of Theorem 8.3.11, let's look at the definition of r in that proof from a hypothetical listing $f(0), f(1), f(2), \ldots$ of the real numbers in $[0, 1]$:

$$\begin{aligned}
f(1) &= 0.257132601\ldots,\\
f(2) &= 0.372694237\ldots,\\
f(3) &= 0.999352196\ldots,\\
f(4) &= 0.50000000\ldots,\\
f(5) &= 0.62347182\ldots,\\
&\vdots
\end{aligned}$$

We can see the diagonalization if we circle the digit in each $f(k)$, $k \geq 1$, that is changed to form the decimal digit r_k.

$$\begin{aligned}
f(1) &= 0.\textcircled{2}57132601\ldots,\\
f(2) &= 0.3\textcircled{7}2694237\ldots,\\
f(3) &= 0.99\textcircled{9}352196\ldots,\\
f(4) &= 0.500\textcircled{0}0000\ldots,\\
f(5) &= 0.6234\textcircled{7}182\ldots,\\
&\vdots
\end{aligned}$$

Then r is defined to be $r = 0.72772\ldots$, and we can see that, for all $k \in \mathbb{Z}^+$, $r \neq f(k)$ because $r_k \neq d_k^k$.

The method used to show that $[0, 1]$ is uncountable is called "Cantor's second diagonalization method". One important consequence of the fact that $\mathbb{R}$ is uncountable is that any *list* of real numbers

$$x_1, x_2, x_3, \ldots$$

cannot include *all* real numbers. In other words, no proof of a statement of the form

$$\text{for all } x \in \mathbb{R} \ldots$$

can begin with

$$\text{let } x_1, x_2, x_3, \ldots \text{ be a list of all real numbers.}$$

8.3.1. Debts paid.

In this subsection, which is optional, we provide a rigorous proof of Theorem 8.3.2.

Theorem 8.3.2. *Let A be a nonempty set. If there exists a surjective function $g : \mathbb{N} \underset{onto}{\rightarrow} A$, then A is countable; i.e., A is finite or denumerable.*

Proof. Let A be a set and assume that there is a surjection $g : \mathbb{N} \underset{\text{onto}}{\rightarrow} A$. We show that A is countable. If A is finite, then A is countable, and we're done. So we assume that A is infinite, and we show that A is denumerable by finding a bijection $f : \mathbb{N} \underset{\text{onto}}{\overset{\text{1-1}}{\rightarrow}} A$.

Remember that the idea is to define $f(n)$ to be the nth new element in the sequence $g(1), g(2), g(3), \ldots$. Formally, we define $f : \mathbb{N} \overset{\text{1-1}}{\underset{\text{onto}}{\to}} A$ by recursion; we make f 1-1 "as we go" and ensure that each element of the range of g is in the range of f.

First let $f(1) = g(1)$. Next, for the recursion, let $n \in \mathbb{N}$ and assume that $f(1), \ldots, f(n)$ have been defined (so each is an element of A) in such a way that for all i and j with $1 \leq i, j \leq n$,

$$i \neq j \implies f(i) \neq f(j)$$

(i.e., the part of f defined so far is 1-1). We also inductively assume that

$$g(1), \ldots, g(n) \in \{f(1), \ldots, f(n)\}.$$

We will define $f(n+1)$.

The set $\{f(1), \ldots, f(n)\}$ is finite, and $\operatorname{ran} g = A$ is infinite, so the set of positive integers

$$X = \{m \in \mathbb{N} \mid g(m) \notin \{f(1), \ldots, f(n)\}\}$$

is nonempty. Hence X has a least element m_0 by the Well-Ordering Principle 6.1.3. Define $f(n+1)$ to be $g(m_0)$, so that $g(m_0) \notin \{f(1), \ldots, f(n)\}$. Note that by definition, $f(n+1) \neq f(i)$ for all $i \leq n$. Furthermore, the definition of m_0 ensures that $m_0 > n$ and

$$g(1), \ldots, g(m_0) \in \{f(1), \ldots, f(n+1)\}.$$

In particular, since $m_0 > n$, we have that

$$g(1), \ldots, g(n+1) \in \{f(1), \ldots, f(n+1)\}.$$

We can then prove by induction on n that for all positive integers i and n, if $i < n$, then $f(i) \neq f(n)$. It follows that f is 1-1. Induction also proves that for all positive integers n,

$$g(1), \ldots, g(n) \in \{f(1), \ldots, f(n)\}.$$

Hence f is onto A, since g is. □

Exercises 8.3

1. Let A be a denumerable set and let x be any element of the underlying universe. Prove that $A \cup \{x\}$ is denumerable. (**HINT:** Consider two cases, depending on whether $x \in A$.)
2. Use the definition to prove that any denumerable set is equinumerous with a proper subset of itself.
3. Use the definition to prove that the set $X = \{n \in \mathbb{Z} \mid 5 \mid n\}$ is denumerable.
4. Given $n \in \mathbb{N}$, define $Q_n = \left\{\frac{a}{n} \mid a \in \mathbb{N}\right\}$. Use the definition to prove that for all $n \in \mathbb{N}$, Q_n is denumerable.
5. Prove that the union of a finite set and a denumerable set is denumerable.
6. Prove that the union of a finite set and a countable set is countable.
7. Prove Corollary 8.3.4.

8. Prove Corollary 8.3.5.
9. Prove that $\mathbb{N} \times \mathbb{N}$ is denumerable in the following ways:
   (a) Pictorially and formally, using Cantor's first diagonalization method (see Exercise 5.3.10).
   (b) Formally, by showing that the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by
   $$f(m, n) = 2^{m-1}(2n - 1)$$
   for all $(m, n) \in \mathbb{N} \times \mathbb{N}$ is a bijection. (**HINT:** You will need the Fundamental Theorem of Arithmetic (Theorem 3.2.3).)
10. Prove Theorem 8.3.7.
11. Prove Theorem 8.3.8 pictorially.
12. Prove that $\mathbb{Q}^+$ is denumerable in the following ways:
   (a) By proving the existence of a surjection $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}^+$ and explaining why the result follows from this fact.
   (b) Using Exercise 8.3.4 and Theorem 8.3.8.
13. Prove that the set of all finite subsets of $\mathbb{Z}^+$ is denumerable.
14. Prove that a set A is countable iff there exists an injection $f : A \overset{1\text{-}1}{\to} \mathbb{N}$.
15. Prove Theorem 8.3.9. (**HINT:** The proof is similar to the proof of Theorem 8.3.2.)
16. Prove Corollary 8.3.10.
17. Assume that A is uncountable and B is a countable subset of A. Prove that $A - B$ is uncountable.
18. A *sequence* $(x_1, x_2, x_3, \dots)$ of real numbers is a function $f : \mathbb{N} \to \mathbb{R}$ defined by $f(n) = x_n$ for all $n \in \mathbb{N}$. Prove that the set X of infinite binary sequences (i.e., infinite sequences of 0's and 1's) is uncountable (**HINT:** Use Cantor's second diagonalization method.)
19. A real number α is defined to be *algebraic* if there exists a polynomial p of degree at least 1 with *integer* coefficients such that $p(\alpha) = 0$. A real number is *transcendental* if it is not algebraic.
   (a) Prove that all rational numbers are algebraic.
   (b) Prove that $\sqrt{2}$, $\sqrt{3}$, and $\sqrt{2} + \sqrt{3}$ are algebraic.
   (c) Prove that if $\alpha \in \mathbb{R}$ is algebraic and $\alpha \neq 0$, then $\alpha^{-1} = \frac{1}{\alpha}$ is algebraic.
   (d) Let A be the set of real algebraic numbers. Prove that A is denumerable. (**HINT:** You will need to use Exercise 6.1.6b.)[3]
   (e) Prove that the set of transcendental real numbers is uncountable. (Note: This problem is not asking you to exhibit a transcendental number. It turns out that the familiar numbers e and π are transcendental, but it is quite difficult to prove this.)

[3] It turns out that the set of all real algebraic numbers is an ordered field (see Definition 9.1.1), but this is beyond the scope of this book. Furthermore, the set of all *complex* algebraic numbers (i.e., the set of all complex numbers $\alpha = a + bi$ $(a, b \in \mathbb{R},\ i^2 = -1)$ such that $p(\alpha) = 0$ for some polynomial with integer coefficients) forms a denumerable *algebraically closed field*, i.e., a denumerable field $\mathcal{F}$ with the property that for every polynomial p of degree at least 1 with coefficients in $\mathcal{F}$, there exists $\alpha \in \mathcal{F}$ such that $p(\alpha) = 0$.

8.4. What next?

Our approach to finite and infinite sets in this chapter has been a "blinders on" approach. We have focused only on the most basic facts about finite sets, and our discussion of infinite sets had as its goal the results about the sizes of $\mathbb{Q}$ and $\mathbb{R}$.

In the previous section, we have shown that there are at least two different "sizes" of infinite sets: the size of $\mathbb{N}$ (i.e., the size of a denumerable set) and the size of $\mathbb{R}$. At this point, one can ask several questions. What do we mean by the "size" of an infinite set? Are there more than two "sizes of infinity"? Does there exist an infinite set whose size is strictly between the size of $\mathbb{N}$ and the size of $\mathbb{R}$?

Recall from our earlier discussion of the history of set theory in Section 4.4 (which is drawn from [**16**]) that Georg Cantor contributed to the development of "naive" set theory in the late nineteenth and early twentieth centuries. As part of his theory of infinite sets, he defined a theory of *sizes* (or *cardinalities*) of infinite sets. While a rigorous definition of the cardinality of an arbitrary set is beyond the scope of this book, the idea of equinumerosity, which we've already considered, and the idea of "comparing" sizes of sets can be safely expressed inside our informal set theory.

Definition 8.4.1. Let A and B be sets.

(1) If there is a 1-1 function $f : A \overset{\text{1-1}}{\to} B$, then we write $A \preceq B$.

(2) If $A \preceq B$ and $A \not\approx B$, then we write $A \prec B$.

What you should be thinking here is that, however the size or cardinality $|X|$ of the set X is defined, $A \approx B$ should mean that A and B have the same number of elements:

$$A \approx B \Leftrightarrow |A| = |B|.$$

Furthermore, $A \preceq B$ should mean that A has at most as many elements as B (or B has at least as many elements as A). In other words, cardinality should be defined so that there is an ordering $\leq$ of cardinalities that satisfies

$$A \preceq B \Leftrightarrow |A| \leq |B|.$$

Finally, $A \prec B$ should mean that A has strictly fewer elements than B:

$$A \prec B \Leftrightarrow |A| < |B|.$$

Thus, we hope that for all sets A and B,

$$(A \preceq B \text{ and } B \preceq A) \Rightarrow A \approx B.$$

In fact, the proof of this statement, which is called the Cantor-Schröder-Bernstein Theorem, is nontrivial (see [**9**] or [**16**]).

One can prove in ZFC that for every set X, $|X|$ exists (i.e., $|X|$ is a set). In addition, the statement that for all sets A and B, $A \preceq B$ or $B \preceq A$ is equivalent to the Axiom of Choice.

In the previous section we showed that $\mathbb{N} \not\approx \mathbb{R}$. Notice that $\mathbb{N} \preceq \mathbb{R}$ since the inclusion function $I : \mathbb{N} \to \mathbb{R}$ by $I(x) = x$ for all $x \in \mathbb{N}$ is 1-1. Thus $\mathbb{N} \prec \mathbb{R}$; i.e., $|\mathbb{N}| < |\mathbb{R}|$.

The diagonalization argument that establishes $\mathbb{N} \prec \mathbb{R}$ can be generalized to show that for any size of an infinite set, there is always a set whose size is strictly larger. In fact, the proof of this result is the earliest example of a proof using a diagonalization argument.

Theorem 8.4.2 (Cantor, 1873)**.** *No set is equinumerous with its power set.*

Proof. Let A be a set, and assume for a contradiction that $A \approx \mathcal{P}(A)$. Fix a bijection $f : A \underset{\text{onto}}{\overset{\text{1-1}}{\to}} \mathcal{P}(A)$. Note that, given $x \in A$, $f(x) \in \mathcal{P}(A)$, so that $f(x) \subseteq A$. We diagonalize to construct a set $B \subseteq A$ such that $B \notin \operatorname{ran} f$. We define

$$B = \{x \in A \mid x \notin f(x)\}.$$

Since $B \subseteq A$ and f is onto, we can fix $a \in A$ such that $B = f(a)$. We now ask whether $a \in B$. Notice that

$$a \in B \Rightarrow a \notin f(a) \Rightarrow a \notin B$$

and

$$a \notin B \Rightarrow a \in f(a) \Rightarrow a \in B.$$

Thus, $a \in B \Leftrightarrow a \notin B$, which is a contradiction.

Hence, A is not equinumerous with $\mathcal{P}(A)$. □

Note the similarity of this proof to the argument in Russell's paradox, which was also used to show in ZF that there does not exist a set of all sets.

Corollary 8.4.3. *For any set A, $A \prec \mathcal{P}(A)$; i.e., $|A| < |\mathcal{P}(A)|$.*

Proof. Let A be a set. We have already shown that $A \not\approx \mathcal{P}(A)$. As an exercise, you should show that the function $f : A \to \mathcal{P}(A)$ by, for all $x \in A$, $f(x) = \{x\}$ is 1-1, which implies that $A \preceq B$. □

Note that Corollary 8.4.3 tells us that there is no largest size of infinite set, since

$$|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < |\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))| < \cdots .$$

A natural question to ask, then, is whether there exists a set A such that $|\mathbb{N}| < A < |\mathcal{P}(\mathbb{N})|$. Cantor conjectured that no such set A existed but could not prove it.

Continuum Hypothesis 8.4.4 (Cantor, 1878)**.** There does not exist a set A such that $\mathbb{N} \prec A \prec \mathcal{P}(\mathbb{N})$.

The Continuum Hypothesis can be stated another way. As we have already noted, the set $\mathbb{Q}$ of rational numbers is countable and $|\mathbb{N}| = |\mathbb{Q}|$. We know that $\mathbb{R}$ is uncountable; in fact, it is possible to prove that $\mathbb{R} \approx \mathcal{P}(\mathbb{N})$. Thus, the Continuum Hypothesis can be rephrased as

> there is no set $X \subseteq \mathbb{R}$ such that $|\mathbb{Q}| < |X| < |\mathbb{R}|$; i.e., every uncountable subset of real numbers has the same cardinality as $\mathbb{R}$.

Perhaps surprisingly, the Continuum Hypothesis can be neither proved nor disproved from the usual axioms of ZF set theory; set theorists say that the formal statement of the Continuum Hypothesis (which we denote by CH) is *independent* of ZF. To prove that the Continuum Hypothesis cannot be disproved in ZF, one must show that, assuming ZF is consistent, there is a universe (i.e., *model*) satisfying the axioms of ZF where CH is true. This was proved by Kurt Gödel in 1938; in fact, the universe constructed by Gödel also satisfies the Axiom of Choice, so that AC also cannot be disproved in ZF. To prove that the Continuum Hypothesis cannot be proved in ZF, one must show that, assuming ZF is consistent, there is a universe satisfying the axioms of ZF where CH is false. This was proved by Paul Cohen in 1963; in fact, the Axiom of Choice is also false in the universe constructed by Cohen, so that AC also cannot be proved in ZF.

For more information, see [**12**] and [**16**].

Chapter 9

Foundations of Analysis

9.1. Introduction

We have spent considerable time studying properties of integers; in this chapter, we return to a discussion of the real numbers. In Chapter 2, we noted that our definition of $\mathbb{R}$ was informal only. We also quite confidently proved that $\sqrt{2}$ is irrational, i.e., that $\sqrt{2}$ is a real number that is not rational. However, we never proved that $\sqrt{2}$ *is a real number*! You might wonder what there is to prove. Basically, we are asserting that

$$(\exists x \in \mathbb{R})[x^2 = 2].$$

How do we know that such a real number x exists? What, in fact, is a real number?

Perhaps a different question one could ask is *why* the definition we gave in Chapter 2 is only an informal one. There, we said that x is a real number exactly when it has a decimal expansion. So, for example, we are very comfortable in saying that

$$0.35210476\ldots$$

is a real number (although we're not sure what happens at the ...). But what does the decimal notation really mean?

$$0.35210476\ldots = \frac{3}{10} + \frac{5}{10^2} + \frac{2}{10^3} + \frac{1}{10^4} + \frac{0}{10^5} + \frac{4}{10^6} + \frac{7}{10^7} + \frac{6}{10^8} + \cdots .$$

This infinite sum should make you nervous (especially if you have taken calculus); what does it mean to say that the sum of infinitely many numbers is a number (in the language of calculus, how do you know that this infinite series *converges*)?

In fact, there are several ways to rigorously define what is meant by a real number; here, we will take an axiomatic point of view, which is the point of view often taken in a beginning undergraduate course in analysis. This means that we will write down a short list of properties that $\mathbb{R}$ "clearly" satisfies and from which we can prove that all other properties follow. The properties we have been using since Chapter 2 are those of an ordered field.

Definition 9.1.1. An *ordered field* is a nonempty set F of objects equipped with two operations called *addition* and *multiplication* and a relation $<$ on F such that the following properties hold.

(Closure under +) For all $a, b \in F$, there is a unique element $a + b \in F$.
(Closure under ·) For all $a, b \in F$, there is a unique element $ab \in F$.
(A1) For all $a, b, c \in F$, $a + (b + c) = (a + b) + c$.
(A2) For all $a, b \in F$, $a + b = b + a$.
(A3) F contains an element 0 such that for all $a \in F$, $a + 0 = a$.
(A4) For all $a \in F$ there exists $b \in F$ such that $a + b = 0$.
(M1) For all $a, b, c \in F$, $a(bc) = (ab)c$.
(M2) For all $a, b \in F$, $ab = ba$.
(M3) F contains an element $1 \neq 0$ such that for all $a \in F$, $a \cdot 1 = a$.
(M4) For all $a \in F$ with $a \neq 0$, there exists $b \in F$ such that $ab = 1$.
(D) For all $a, b, c \in F$, $a(b + c) = ab + ac$.
(O1) For all $a, b \in F$, exactly one of $a < b$ or $a = b$ or $b < a$ holds.
(O2) For all $a, b, c \in F$, if $a < b$ and $b < c$, then $a < c$.
(O3) For all $a, b, c \in F$, if $a > 0$ and $b > 0$, then $a + b > 0$.
(O4) For all $a, b, c \in F$, if $a > 0$ and $b > 0$, then $ab > 0$.
(O5) For all $a, b \in F$, $a < b$ iff $b - a > 0$.

It certainly seems reasonable to assume that $\mathbb{R}$ is an ordered field. However, $\mathbb{Q}$ is also an ordered field, and yet $\mathbb{Q}$ and $\mathbb{R}$ seem very different. Certainly, as we saw in the previous chapter, $\mathbb{Q}$ is denumerable while $\mathbb{R}$ is uncountable. Right away this tells us that $\mathbb{R}$ satisfies some property not on this list that $\mathbb{Q}$ does not satisfy. To help isolate what this property could be, notice further that, informally, $\mathbb{Q}$ has a "hole" where $\sqrt{2}$ "ought to be", and $\mathbb{R}$ does not. This can be expressed in several ways.

In $\mathbb{Q}$:

- the equation $x^2 = 2$ has no solution;
- the sequence of rational numbers

  $$1, 1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots$$

  has no limit;
- the set $\{x \in \mathbb{Q}^+ \mid x^2 < 2\}$ is "bounded above" but has no "least upper bound".

In $\mathbb{R}$:

- the equation $x^2 = 2$ has a solution;
- the sequence of rational (and hence real) numbers

  $$1, 1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots$$

  has a limit;
- the set $\{x \in \mathbb{R}^+ \mid x^2 < 2\}$ has a "least upper bound".

In turns out that we need only assume one additional axiom (the *Completeness Axiom*) in order to capture what makes $\mathbb{R}$ different from $\mathbb{Q}$. That axiom is related to the notion of a *least upper bound* of a set, mentioned above, which we define in the next section.

9.2. The Completeness Axiom

First, we review the notion of a maximum (or minimum) of a set.

Definition 9.2.1. Let $S \subseteq \mathbb{R}$ be nonempty. Let $M, m \in \mathbb{R}$.

(1) We say that M is the *maximum element of* S and write $\max S = M$ if $M \in S$ and for all $x \in S$, $x \leq M$.

(2) We say that m is the *minimum element of* S and write $\min S = m$ if $m \in S$ and for all $x \in S$, $m \leq x$.

Note that one should prove that the maximum and minimum elements of a nonempty finite set $S \subseteq \mathbb{R}$ are well-defined. See Exercise 8.2.14.

Example 9.2.2.

(1) Let $S = \{0, -\frac{16}{3}, -27, 2798.999, \pi\} \subseteq \mathbb{R}$. Then $\max S = 2798.999$ and $\min S = -27$.

(2) Let $S = [-1, 2) = \{x \in \mathbb{R} \mid -1 \leq x < 2\} \subseteq \mathbb{R}$. From the definition of S, we see that $\min S = -1$. However, S has no maximum element. While this may seem obvious, we'll prove it. Suppose for the sake of a contradiction that we have $M \in \mathbb{R}$ such that $\max S = M$. Then $M \in S$, so $M < 2$ by definition of S. By Exercise 2.1.11, we know that $M < \frac{M+2}{2} < 2$, so $\frac{M+2}{2} \in S$ is greater than $\max S$, a contradiction. ♢

Example 9.2.2(1) illustrates an important property of finite sets.

Proposition 9.2.3. *Every finite, nonempty set $S \subseteq \mathbb{R}$ has both a maximum element and a minimum element.*

Proof. This was Exercise 8.2.14. □

While the set $[-1, 2)$ has no maximum element, it is bounded above. The difference is that an upper bound of a nonempty set is not required to be a member of the set (and may or may not be an element of the set).

Definition 9.2.4. Let $S \subseteq \mathbb{R}$ be nonempty.

(1) The real number M is an *upper bound of* S if for all $x \in S$, $x \leq M$.

(2) The set S is *bounded above* if there exists $M \in \mathbb{R}$ such that M is an upper bound of S (i.e., if $(\exists M \in \mathbb{R})(\forall x \in S)[x \leq M]$).

(3) The real number m is a *lower bound of* S if for all $x \in S$, $m \leq x$.

(4) The set S is *bounded below* if there exists $m \in \mathbb{R}$ such that m is a lower bound of S (i.e., if $(\exists m \in \mathbb{R})(\forall x \in S)[m \leq x]$).

(5) The set S is *bounded* if S is both bounded above and bounded below.

Let $S \subseteq \mathbb{R}$ be nonempty and let $m, M \in \mathbb{R}$. As an exercise, you should write down the definitions of "M is not an upper bound of S", "m is not a lower bound of S", "S is not bounded above", and "S is not bounded below". It's also worth taking a moment to write down the general Given-Goal diagrams for showing that "$M \in \mathbb{R}$ is an upper bound of S", "M is not an upper bound of S", and "S is not bounded above". We leave writing down the analogous Given-Goal diagrams for proving that "$m \in \mathbb{R}$ is a lower bound of S", "$m \in \mathbb{R}$ is not a lower bound of S", and "S is not bounded below" as an exercise.

Table 9.1. M is an upper bound of S.

Given	**Goal**
$S \neq \emptyset$	
$x \in S$ arbitrary	
definition of M	$x \leq M$

Table 9.2. M is not an upper bound of S.

Given	**Goal**
$S \neq \emptyset$	
definition of M	find $x \in S$ such that $x > M$

Table 9.3. S is not bounded above.

Given	**Goal**
$S \neq \emptyset$	
$M \in \mathbb{R}$ arbitrary	find $x \in S$ such that $x > M$

Example 9.2.5.

(1) Let $S = [-1, 2) = \{x \in \mathbb{R} \mid -1 \leq x < 2\} \subseteq \mathbb{R}$. Note that S is bounded below; each of the numbers $-25, -472, -1$ is a lower bound of S (in fact, S has infinitely many lower bounds). In addition, S is bounded above; each of the numbers $49, e, 2.0000001, 2$ is an upper bound of S (in fact, S has infinitely many upper bounds). Thus, S is a bounded set.

(2) Let $S = \{\frac{1}{n} \mid n \in \mathbb{Z}^+\} = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\} \subseteq \mathbb{R}$. Then S is bounded, since (for example) 12 is an upper bound of S and $-\frac{8}{9}$ is a lower bound of S.

(3) Let $S = (0, \infty) = \{x \in \mathbb{R} \mid x > 0\} \subseteq \mathbb{R}$. Note that S is bounded below by any real number $r \leq 0$, since if $r \leq 0$ and $x \in S$, then $r \leq 0 < x$. However, S is not bounded above. Again, while this may seem obvious to you, we will prove it. By Definition 9.2.4, we must prove

$$(\forall M \in \mathbb{R})(\exists x \in S)[x > M].$$

Let $M \in \mathbb{R}$. If $M \leq 0$, then $1 \in S$ and $1 > M$. If $M > 0$, then $M + 1 > 0$. Thus $M + 1 \in S$ with $M + 1 > M$. Thus $S = (0, \infty)$ is not bounded above.

(4) Let $S = \mathbb{Z}^+ = \{1, 2, 3, \ldots\} \subseteq \mathbb{R}$. Again, S is bounded below, say by any real number $r \leq 1$. It seems obvious that S is not bounded above. However, unlike the previous example, we cannot yet prove this, for the proof requires the Completeness Axiom! (Try it: one needs to prove

$$(\forall M \in \mathbb{R})(\exists n \in \mathbb{Z}^+)[n > M].$$

If you find yourself talking about decimal expansions, then you are using the Completeness Axiom.) ♢

While the set $[-1, 2)$ has many upper bounds (and lower bounds), the number 2 appears to be its "least upper bound".

Definition 9.2.6. Let $S \subseteq \mathbb{R}$ be nonempty.

(1) Suppose that S is bounded above. The real number β is the *supremum* (*least upper bound*) *of S* if
   (a) β is an upper bound of S and
   (b) for all $\gamma \in \mathbb{R}$, if γ is an upper bound of S, then $\beta \leq \gamma$.
   If β is the supremum of S, then we write $\sup S = \beta$.

(2) Suppose that S is bounded below. The real number α is the *infimum* (*greatest lower bound*) *of S* if
   (a) α is a lower bound of S and
   (b) for all $\gamma \in \mathbb{R}$, if γ is a lower bound of S, then $\gamma \leq \alpha$.
   If α is the infimum of S, then we write $\inf S = \alpha$.

Note that one should prove that if they exist, then the supremum and infimum of a nonempty set $S \subseteq \mathbb{R}$ are unique. See Exercise 9.2.3.

Let $S \subseteq \mathbb{R}$ be nonempty and let $\alpha, \beta \in \mathbb{R}$. Note that in order to prove that $\sup S = \beta$ (respectively, $\inf S = \alpha$), one must prove *both* statements (1a) and (1b) (respectively, (2a) and (2b)) in Definition 9.2.6. It's worth taking a moment to write down the general Given-Goal diagrams for proving that β is the supremum of S; we leave writing down the analogous Given-Goal diagrams for proving that α is the infimum of S as an exercise. The Given-Goal diagram for showing that β is an upper bound of S is given in Table 9.1. Two possible Given-Goal diagrams for showing that an upper bound β of S is the least of all upper bounds of S are given in Table 9.4. The left Given-Goal diagram corresponds to a direct proof, while the right Given-Goal diagram corresponds to proving the contrapositive.

Table 9.4. The upper bound β is the least of all upper bounds of S.

Given	**Goal**
$S \neq \emptyset$ definition of β $(\forall x \in S)[x \leq \beta]$ $\gamma \in \mathbb{R}$ $(\forall x \in S)[x \leq \gamma]$	$\beta \leq \gamma$

Given	**Goal**
$S \neq \emptyset$ definition of β $(\forall x \in S)[x \leq \beta]$ $\gamma \in \mathbb{R}$ arbitrary $\gamma < \beta$	γ is not an upper bound of S

Also useful is the general Given-Goal diagram for showing that an upper bound β is not the least upper bound of S. We again leave the analogous Given-Goal diagram for infimums as an exercise.

Table 9.5. Upper bound β is *not* $\sup S$.

Given	**Goal**
$S \neq \emptyset$	
definition of β	find $\gamma \in \mathbb{R}$ such that
$(\forall x \in S)[x \leq \beta]$	γ is an upper bound of S and $\gamma < \beta$

Even though we don't have all the necessary tools available to us right now, it is still useful to consider some examples. One of these relies, for now, on our intuition.

Example 9.2.7.

(1) Let $S = [-1, 2) = \{x \in \mathbb{R} \mid -1 \leq x < 2\}$. We claim that $\inf S = -1$ and $\sup S = 2$.

That $\inf S = -1$ follows from the fact that $\min S = -1$. See Proposition 9.2.8. We'll show that $\sup S = 2$. By Definition 9.2.6, we must show that

(a) 2 is an upper bound of S, which we can see from the definition of S, and

(b) if $\gamma \in \mathbb{R}$ is an upper bound of S, then $2 \leq \gamma$.

We prove the contrapositive of (b). Following the right Given-Goal diagram in Table 9.4, let $\gamma \in \mathbb{R}$ and assume that $\gamma < 2$. We show that γ is not an upper bound of S. Note that if $\gamma < -1$, then γ is not an upper bound of S because $-1 \in S$. So we assume that $-1 \leq \gamma < 2$. But then by Exercise 2.1.11, $\gamma < \frac{\gamma+2}{2} < 2$; i.e., $\gamma < \frac{\gamma+2}{2}$ and $\frac{\gamma+2}{2} \in S$, so that γ is not an upper bound of S. Hence statement (b) is true and $2 = \sup S$.

(2) Let $S = \{\frac{1}{n} \mid n \in \mathbb{Z}^+\} = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\} \subseteq \mathbb{R}$. Then $\inf S = 0$ and $\sup S = 1$. That $\sup S = 1$ follows since $\max S = 1$. We defer proving $\inf S = 0$ until Example 9.3.3, after we have discussed the Completeness Axiom.

(3) Let $S = (0, \infty) = \{x \in \mathbb{R} \mid x > 0\}$. We claim that $\inf S = 0$ and that $\sup S$ doesn't exist.

The definition of S shows that 0 is a lower bound of S. To show that 0 is the greatest lower bound of S, we assume that $\gamma \in \mathbb{R}$ satisfies $0 < \gamma$, and we show that γ is not a lower bound of S. Let $x = \frac{\gamma}{2}$. Then $0 < x < \gamma$, and by definition $x \in S$. Hence γ is not a lower bound of S, and so $\inf S = 0$ by Definition 9.2.6. We showed in Example 9.2.5(3) that S is not bounded above, and hence $\sup S$ does not exist, again by Definition 9.2.6. ◊

We mentioned some straightforward properties about $\sup S$ and $\inf S$ in the example above, which we now state for convenience.

Proposition 9.2.8. *Let $S \subseteq \mathbb{R}$ be nonempty.*

(1) *If* $\max S$ *exists, then* $\sup S = \max S$*. Conversely, if* $\sup S \in S$*, then* $\sup S = \max S$*.*

(2) *If* $\min S$ *exists, then* $\inf S = \min S$. *Conversely, if* $\inf S \in S$, *then* $\inf S = \min S$.

Proof. Exercise 9.2.5. □

Example 9.2.7 shows that an arbitrary nonempty set of real numbers may or may not have a supremum or infimum. One set that demonstrates why $\mathbb{Q}$ has "holes" is $S = \{x \in \mathbb{Q} \mid x > 0 \text{ and } x^2 < 2\}$. This set is bounded above in $\mathbb{Q}$ by, say, 2. This set has no supremum (i.e., no least upper bound) in $\mathbb{Q}$ (hence the "hole"), while the analogous subset of $\mathbb{R}$ does have a supremum in $\mathbb{R}$, namely, $\sqrt{2}$ (all of these statements require proof). Thus, one of the problems with $\mathbb{Q}$ is that there exist nonempty subsets of $\mathbb{Q}$ that are bounded above in $\mathbb{Q}$ but with no least upper bound in $\mathbb{Q}$. This is the property that distinguishes $\mathbb{Q}$ from $\mathbb{R}$.

Completeness Axiom 9.2.9. Every nonempty set of real numbers that has an upper bound has a supremum.

The analogous property regarding sets that are bounded below follows immediately.

Corollary 9.2.10. *Every nonempty set of real numbers that has a lower bound has an infimum.*

Proof. Exercise 9.2.6. □

The Completeness Axiom is the last axiom that we need to assume about the real numbers, and it will guarantee the existence of $\sqrt{2}$ and all the other real roots. We say that $\mathbb{R}$ is a *complete ordered field*, while $\mathbb{Q}$ is an incomplete ordered field. In fact, in a sense that can be made mathematically precise, $\mathbb{R}$ is the *only* "completion" of $\mathbb{Q}$ with respect to the usual order $<$ on $\mathbb{Q}$.

The Completeness Axiom also guarantees that every real number has a (usually) unique decimal expansion and that every decimal expansion represents a real number. This is the use of the Completeness Axiom in the proof in Section 8.3 that $\mathbb{R}$ is uncountable, and this is what makes that proof fail in $\mathbb{Q}$. We consider decimal expansions in Section 9.3.

9.2.1. The existence of $\sqrt{2}$.

In this subsection, which is optional, we prove using the Completeness Axiom that $\sqrt{2}$ is a real number. We also show that $\mathbb{Q}$ does not satisfy a Completeness Axiom. Students who are encountering these ideas for the first time may wish to omit this subsection on their first reading.

Theorem 9.2.11. *There exists a real number* x *such that* $x^2 = 2$.

Scratchwork: We consider the set $S = \{x \in \mathbb{R} \mid x > 0 \text{ and } x^2 < 2\}$. We must use the Completeness Theorem to argue that $\beta = \sup S$ exists and that $1 < \beta < 2$.

Our proof will repeatedly make use of the following fact about real numbers (see Exercise 2.1.2 and Proposition 2.2.5):

$$(\forall w, z > 0)[w < z \Leftrightarrow w^2 < z^2].$$

It then follows that

(9.1) if $z > 0$ and $z^2 > 2$, then z is an upper bound of S,

since for all $x \in S$, $x > 0$ and $x^2 < 2$.

To show that $\beta = \sqrt{2}$, we must show that $\beta^2 = 2$. The Trichotomy Axiom asserts that either

$$\beta^2 < 2 \qquad \text{or} \qquad \beta^2 = 2 \qquad \text{or} \qquad \beta^2 > 2.$$

Thus, we must show that both of the assumptions $\beta^2 < 2$ and $\beta^2 > 2$ lead to a contradiction.

Our intuition should be that if $\beta^2 < 2$, then β is too small to be an upper bound of S, which contradicts the fact that $\beta = \sup S$. The specific version of the Given-Goal diagram in Table 9.2 that we need is below.

Given	**Goal**
$\beta = \sup S$	
$1 < \beta < 2$	find $x \in \mathbb{R}$ with $x > 0$ and $x^2 < 2$ (i.e., $x \in S$)
$\beta^2 < 2$	satisfying $\beta < x$

It will be helpful to make our Given-Goal diagram more explicit.

Given	**Goal**
$\beta = \sup S$	
$1 < \beta < 2$	find $h \in \mathbb{R}$ with $0 < h < 1$ and $(\beta + h)^2 < 2$
$\beta^2 < 2$	(i.e., $\beta + h \in S$, where $\beta < \beta + h$)

We work backwards to find h, under the assumption that $0 < h < 1$ (this will have to be verified in the proof), which implies that $h^2 < h$.

$$\begin{aligned}
(\beta + h)^2 &= \beta^2 + 2\beta h + h^2 \\
&< \beta^2 + 2\beta h + h \\
&= \beta^2 + h(2\beta + 1) \\
&< \beta^2 + 5h, \quad \text{since } \beta < 2 \text{ and } h > 0.
\end{aligned}$$

Thus, to make $(\beta + h)^2 < 2$, we can make $\beta^2 + 5h \leq 2$, so we'll try $h = \frac{2-\beta^2}{5}$. Again, we will have to demonstrate in the proof that this definition of h satisfies $0 < h < 1$.

Moving to the next case, our intuition should be that if $\beta^2 > 2$, then β is too big to be the *least* upper bound of S, which again contradicts the fact that $\beta = \sup S$. The specific version of the Given-Goal diagram in Table 9.5 that we need is below.

Given	**Goal**
$\beta = \sup S$	
$1 < \beta < 2$	find $z \in \mathbb{R}$ with $z < \beta$ such that
$\beta^2 > 2$	z is an upper bound of S

As before, it will be helpful to make our Given-Goal diagram more explicit. Note that, by statement (9.1), we can ensure that a real number $z > 0$ is an upper bound of S by ensuring that $z^2 > 2$.

Given	**Goal**
$\beta = \sup S$	
$1 < \beta < 2$	find $h \in \mathbb{R}$ with $0 < h < 1$ and $(\beta - h)^2 > 2$
$\beta^2 > 2$	(i.e., $\beta - h$ is an upper bound of S, where $\beta - h < \beta$)

We work backwards to find h, under the assumption that $0 < h < 1$ (this will have to be verified in the proof).

$$\begin{aligned}(\beta - h)^2 &= \beta^2 - 2\beta h + h^2\\ &> \beta^2 - 2\beta h\\ &> \beta^2 - 4h,\end{aligned}$$

since $0 < \beta < 2$ and $h > 0$. Thus, to make $(\beta - h)^2 > 2$, we can make $\beta^2 - 4h \geq 2$, so we'll try $h = \frac{\beta^2 - 2}{4}$. Again, we will have to demonstrate in the proof that this definition of h satisfies $0 < h < 1$.

We are now ready for the proof that $\sqrt{2}$ is a real number.

Proof. Let $S = \{x \in \mathbb{R} \mid x > 0 \text{ and } x^2 < 2\} \subseteq \mathbb{R}$. Then $S \neq \emptyset$, since $\left(\frac{4}{3}\right)^2 = \frac{16}{9} < 2$, and so $\frac{4}{3} \in S$. In addition, S is bounded above by $\frac{3}{2}$ by statement (9.1), since $\left(\frac{3}{2}\right)^2 = \frac{9}{4} > 2$. Hence by the Completeness Axiom, $\beta = \sup S$ exists.

Note that $1 < \frac{4}{3} \leq \beta$ since $\frac{4}{3} \in S$. Furthermore $\beta \leq \frac{3}{2} < 2$ since $\frac{3}{2}$ is an upper bound of S. We show that $\beta^2 = 2$ by showing that $\beta^2 < 2$ and $\beta^2 > 2$ lead to contradictions (note the use of the Trichotomy Axiom here).

Assume that $\beta^2 < 2$. We will find $h \in \mathbb{R}$ with $0 < h < 1$ and $(\beta + h)^2 < 2$. This implies that $\beta + h \in S$, but $\beta < \beta + h$, so then β can't be an upper bound of S, a contradiction.

Consider $h = \frac{2-\beta^2}{5}$. Then $h > 0$ since $\beta^2 < 2$. In addition, $h = \frac{2-\beta^2}{5} = \frac{2}{5} - \frac{\beta^2}{5} < \frac{2}{5} < 1$, and hence $h^2 < h$. Thus, since $\beta < 2$ and $0 < h < 1$, we have

$$\begin{aligned}(\beta+h)^2 &= \beta^2 + 2\beta h + h^2\\ &< \beta^2 + 2\beta h + h\\ &= \beta^2 + h(2\beta+1)\\ &< \beta^2 + 5h\\ &= \beta^2 + 5\left(\frac{2-\beta^2}{5}\right) = 2,\end{aligned}$$

as desired, our contradiction.

Next assume that $\beta^2 > 2$. We'll find $h \in \mathbb{R}$ such that $0 < h < 1$ and $(\beta - h)^2 > 2$. Then $\beta - h > 0$, since $1 < \beta$ and $h < 1$. Furthermore, $\beta - h$ is an upper bound of S by statement (9.1). If we can find such an h, then $\beta - h < \beta$, and therefore β isn't the least upper bound of S, a contradiction.

Consider $h = \frac{\beta^2-2}{4}$. Note that $h > 0$ and $h = \frac{\beta^2}{4} - \frac{1}{2} < \frac{4}{4} - \frac{1}{2} < 1$, since $0 < \beta < 2$. Also, since $0 < \beta < 2$ and $h > 0$, we have

$$\begin{aligned}(\beta-h)^2 &= \beta^2 - 2\beta h + h^2\\ &> \beta^2 - 2\beta h\\ &> \beta^2 - 4h\\ &= \beta^2 - 4\left(\frac{\beta^2-2}{4}\right) = 2,\end{aligned}$$

as desired, our contradiction.

Thus, $\beta^2 < 2$ and $\beta^2 > 2$ both lead to contradictions. Hence by the Trichotomy Axiom, $\beta^2 = 2$; i.e., $\beta = \sqrt{2}$. □

Corollary 9.2.12 (Corollary to the proof of Theorem 9.2.11)**.** $\mathbb{Q}$ *does not satisfy a Completeness Axiom.*

Proof. Let $T = \{x \in \mathbb{Q} \mid x > 0 \text{ and } x^2 < 2\}$, and note, as in the proof of Theorem 9.2.11, that $\frac{3}{2} \in \mathbb{Q}$ is an upper bound of T. We show that T has no supremum in $\mathbb{Q}$. Assume for the sake of a contradiction that there exists a rational number β such that $\beta = \sup T$. By the proof of Theorem 9.2.11, $\beta^2 < 2$ leads to a contradiction since $h = \frac{2-\beta^2}{5} \in \mathbb{Q}$, and $\beta + h \in T$ with $\beta < \beta + h$, so that β cannot be an upper bound of T. Similarly, by the proof of Theorem 9.2.11, $\beta^2 > 2$ leads to a contradiction since $h = \frac{\beta^2-2}{4} \in \mathbb{Q}$, and $\beta - h$ is an upper bound of T with $\beta - h < \beta$, so that β cannot be the least upper bound of T. Since $\mathbb{Q}$ is an ordered field, it follows by the Trichotomy Axiom that $\beta^2 = 2$, which contradicts Theorem 2.3.1, since $\beta \in \mathbb{Q}$. Hence T has no supremum in $\mathbb{Q}$. □

Exercises 9.2

1. Let $a < b$ be real numbers and let $S = \{x \in \mathbb{R} \mid a < x < b\}$. Show that $\inf S = a$ and $\sup S = b$.
2. Find $\sup\{x \in \mathbb{R} \mid x^2 + 4 < 5x\}$.
3. Let $S \subseteq \mathbb{R}$ be nonempty.
   (a) Prove that if $\sup S$ exists, then it is unique; i.e., if S is bounded above and both $\beta_1, \beta_2 \in \mathbb{R}$ satisfy the definition of $\sup S$ in Definition 9.2.6, then $\beta_1 = \beta_2$.
   (b) Prove that if $\inf S$ exists, then it is unique; i.e., if S is bounded below and both $\alpha_1, \alpha_2 \in \mathbb{R}$ satisfy the definition of $\inf S$ in Definition 9.2.6, then $\alpha_1 = \alpha_2$.
4. Assume $S \subseteq \mathbb{R}$ is nonempty and bounded, and let $\alpha, \beta \in \mathbb{R}$ such that α is a lower bound of S and β is an upper bound of S.
   (a) Prove that $\beta = \sup S$ iff $(\forall \varepsilon > 0)(\exists x \in S)[\beta - \varepsilon < x]$.
   (b) Prove that $\alpha = \inf S$ iff $(\forall \varepsilon > 0)(\exists x \in S)[x < \alpha + \varepsilon]$.
5. Prove Proposition 9.2.8.
6. Prove Corollary 9.2.10. (**HINT:** Use "mirrors". Let $T \subseteq \mathbb{R}$ be bounded below, and let $\alpha \in \mathbb{R}$ be a lower bound of T. Apply the Completeness Axiom to the set $S = \{-x \mid x \in T\}$.)
7. Let $S \subseteq \mathbb{R}$ be bounded and nonempty, and let $T \subseteq S$ be nonempty. Prove that T is bounded and
$$\inf S \leq \inf T \leq \sup T \leq \sup S.$$
8. Let $S \subseteq \mathbb{R}$ be nonempty and let $c \in \mathbb{R}$ be nonzero. Let $T = \{cx \in x \in S\}$.
   (a) Assume that S is bounded above and $c > 0$. Prove that T is bounded above and $\sup T = c \cdot \sup S$.
   (b) Assume that S is bounded above and $c < 0$. Prove that T is bounded below and $\inf T = c \cdot \sup S$.
   (c) Assume that S is bounded below and $c > 0$. Prove that T is bounded below and $\inf T = c \cdot \inf S$.
   (d) Assume that S is bounded below and $c < 0$. Prove that T is bounded above and $\sup T = c \cdot \inf S$.
9. Let $S \subseteq \mathbb{R}$ be bounded and nonempty, and let $c \in \mathbb{R}$. Let $T = \{x + c \mid x \in S\}$. Prove that T is bounded and that $\sup T = c + \sup S$ and $\inf T = c + \inf S$.
10. Use the method of Theorem 9.2.11 to prove the following:
    (a) $\sqrt{3}$ is a real number; i.e., there exists a real number x such that $x^2 = 3$.
    (b) $\sqrt[3]{2}$ is a real number; i.e., there exists a real number x such that $x^3 = 2$.
11. In this problem, we (finally) formally define the term "interval".

    Definition 9.2.13. The set $S \subseteq \mathbb{R}$ is an *interval* if S contains at least two distinct real numbers and, for all $x, y \in S$, if $x < y$, then for all $z \in \mathbb{R}$ with $x < z < y$, $z \in S$.

Let $S \subseteq \mathbb{R}$. Prove that S is an interval if and only if S takes one of the forms in Notation 4.1.10.

9.3. The Archimedean Property and its consequences

Probably one of the most surprising claims made in Section 9.2 was that we did not have the tools necessary to prove that the set of positive integers (as a subset of $\mathbb{R}$) is not bounded above in $\mathbb{R}$. What is needed is a consequence of the Completeness Axiom known as the Archimedean Property, which is a very useful tool in real analysis.

Theorem 9.3.1 (The Archimedean Property)**.** *Let $a, b \in \mathbb{R}$ with $0 < a < b$. Then there exists $n \in \mathbb{Z}^+$ such that $na > b$.*

A picture of the Archimedean Property is given in Figure 9.1. Imagine having line segments of length a and b. The Archimedean Property essentially says that, regardless of how large b is and how small a is, it is possible to find a large enough positive integer n such that laying down n line segments of length a end to end on a line will produce a line segment that is longer than b.

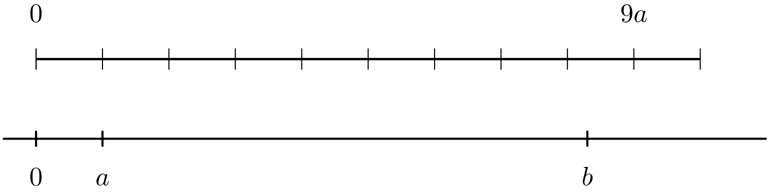


Figure 9.1. The segment of length $9a$ is longer than the segment of length b.

Scratchwork: Given real numbers $0 < a < b$, we will proceed by contradiction. Our intuition is that the contradiction will involve the Completeness Axiom, so we will be looking for (in this instance) a nonempty set of real numbers that is bounded above but has no supremum. Writing down the Given part of the Given-Goal diagram gives us the right idea for what set to look at.

Given	**Goal**
$0 < a < b$ in $\mathbb{R}$ $(\forall n \in \mathbb{Z}^+)[na \leq b]$	$\{na \mid n \in \mathbb{Z}^+\}$ is bounded above but has no supremum

Proof. Let $a, b \in \mathbb{R}$ and assume that $0 < a < b$. Assume also for the sake of a contradiction that for all $n \in \mathbb{Z}^+$, $na \leq b$.

Consider $S = \{na \mid n \in \mathbb{Z}^+\}$. Then $S \neq \emptyset$, and S is bounded above by b, by assumption. By the Completeness Axiom, S has a supremum; let $\beta = \sup S$.

Note that $\beta - a < \beta$, since $a > 0$. Hence $\beta - a$ is not an upper bound of S since $\beta = \sup S$. Thus we can fix $x \in S$ such that $\beta - a < x$. Since $x \in S$, we can fix $m \in \mathbb{Z}^+$ such that $x = ma$. Then

$$\begin{aligned} \beta &< x + a \\ &= ma + a \\ &= (m+1)a. \end{aligned}$$

But then β is not an upper bound of S, since $(m+1)a \in S$ by definition, contradicting the fact that $\sup S = \beta$.

Hence there exists $n \in \mathbb{Z}^+$ such that $na > b$, as desired. □

The facts below, which all follow from the Archimedean Property, are frequently used in proofs in analysis. In part (3) below, you should imagine that, regardless of how small $x > 0$ is, you can always find a large enough positive integer n where $\frac{1}{n}$ is even smaller.

Corollary 9.3.2.

(1) *$\mathbb{Z}^+$ is not bounded above in $\mathbb{R}$; i.e.,*

$$(\forall x \in \mathbb{R})(\exists n \in \mathbb{Z}^+)[n > x].$$

(2) *For all $x \in \mathbb{R}$, there exists $n \in \mathbb{Z}$ with $n \leq x < n + 1$.*

(3) *For all $x \in \mathbb{R}^+$, there exists $n \in \mathbb{Z}^+$ with $0 < \frac{1}{n} < x$.*

Scratchwork for part (2): Given $x \in \mathbb{R}$, by part (1) there is a positive integer greater than x. The least such positive integer should be the integer $n + 1$ we seek, and it exists by the Well-Ordering Principle.

Proof. We leave parts (1) and (3) for Exercise 9.3.1 and prove (2).

Let $x \in \mathbb{R}$; we must find $n \in \mathbb{Z}$ such that $n \leq x < n + 1$. We consider three cases.

Case I: $x \in \mathbb{Z}$.

Then $x \leq x < x + 1$, and we're done (i.e., x is the integer n we seek).

Case II: $x \notin \mathbb{Z}$ and $x > 0$.

Let $S = \{m \in \mathbb{Z}^+ \mid x < m\}$. Note that $S \neq \emptyset$, since otherwise $m \leq x$ for all positive integers m; i.e., x is an upper bound of $\mathbb{Z}^+$, contradicting part (1). Thus by the Well-Ordering Principle 6.1.3, S has a least element. Let m_x be the least element of S, and note that $x < m_x$. (We are using a subscript notation here to emphasize the fact that this integer depends on x.)

Notice that $m_x - 1 \leq x$ by definition of m_x. Taking $n = m_x - 1 \in \mathbb{Z}$, we have $n \leq x < n + 1$, as desired.

Case III: $x \notin \mathbb{Z}$ and $x < 0$.

We use "mirrors". Note that $-x > 0$, so by Case II we can fix $m \in \mathbb{Z}$ such that $m \leq -x < m + 1$. Then $-m - 1 < x \leq -m$. Since $x \notin \mathbb{Z}$, we have $-m - 1 < x < m$. Thus $n = -m - 1$ is the integer we seek. □

We are now in a position to complete our discussion of Example 9.2.7(2).

Example 9.3.3. Let $S = \{\frac{1}{n} \mid\in \mathbb{Z}^+\}$. We show that $\inf S = 0$.

Certainly 0 is a lower bound of S; i.e., for all $n \in \mathbb{Z}^+$, $0 \leq \frac{1}{n}$. We must show that 0 is the greatest lower bound of S. Let $\alpha \in \mathbb{R}$ be such that $0 < \alpha$. We show that α is not a lower bound of S. By Corollary 9.3.2(3), we can fix $n \in \mathbb{Z}^+$ such that $\frac{1}{n} < \alpha$. Since $\frac{1}{n} \in S$, α is not a lower bound of S, as desired. Hence $\inf S = 0$. ♢

The results here on the Archimedean Property form the beginnings of a course in real analysis, which you may take in future.

9.3.1. Decimal expansions. This subsection is optional. Students who are encountering these ideas for the first time may wish to omit the details of this subsection on their first reading.

So what, after all, is a real number? That's easy: a real number is an element of the set $\mathbb{R}$.

But then, what's $\mathbb{R}$? Well, the approach that's often taken in a real analysis course is that we simply make the *assumption* that $\mathbb{R}$ is a complete ordered field; i.e., $\mathbb{R}$ is a set F that satisfies all the axioms of an ordered field in Definition 9.1.1 as well as the Completeness Axiom 9.2.9. Thus, the definition of $\mathbb{R}$ is quite abstract. If you're starting to think that you no longer know what a real number is, then that's normal.

So, let's connect this back to what you do know; let's go back to decimal expansions. This is how we originally *informally* defined $\mathbb{R}$, and we needed decimal expansions in our proof that $\mathbb{R}$ is uncountable (Theorem 8.3.11).

First, we need to make sense of the infinite sum

$$0.35210476\ldots = \frac{3}{10} + \frac{5}{10^2} + \frac{2}{10^3} + \frac{1}{10^4} + \frac{0}{10^5} + \frac{4}{10^6} + \frac{7}{10^7} + \frac{6}{10^8} + \cdots$$

at the beginning of this chapter. We avoid the sum altogether and go back to thinking in terms of "terminating" decimal expansions, which we do understand. In other words, we want to think of $0.35210476\ldots$ as the "limit" of the sequence

$$0.3, 0.35, 0.352, 0.3521, 0.35210, 0.352104, 0.3521047, 0.35210476, \ldots$$

of rational numbers. Thus, we want to think of $0.35210476\ldots$ as the *supremum* of the set

$$\{0.3, 0.35, 0.352, 0.3521, 0.35210, 0.352104, 0.3521047, 0.35210476, \ldots\}.$$

Definition 9.3.4. Let $d_0 \in \mathbb{Z}^{\geq 0}$ and, for all $i \in \mathbb{Z}^+$, $d_i \in \mathbb{Z}$ with $0 \leq d_i \leq 9$. We define the *decimal expansion* $d_0.d_1d_2d_3\ldots$ to be notation that represents the supremum of the set

$$\{d_0, d_0.d_1, d_0.d_1d_2, \ldots\} = \{d_0\} \cup \left\{ d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n}{10^n} \;\middle|\; n \in \mathbb{Z}^+ \right\}$$

of real numbers. The integers $d_1, d_2, d_3, \ldots$ are called the *decimal digits* of the decimal expansion $d_0.d_1d_2d_3\ldots$. We say that the decimal expansion $d_0.d_1d_2d_3\ldots$ is *terminating* if there exists $N \in \mathbb{Z}^+$ such that for all $n \geq N$, $d_n = 0$. Otherwise, we say that the decimal expansion is *nonterminating*.

Notice that the supremum of the set mentioned in Definition 9.3.4 exists because $d_0 + 1 \in \mathbb{R}$ is an upper bound for that set (this requires explanation). Furthermore, every decimal expansion is a nonnegative real number. As an example, for the decimal expansion of $\sqrt{2}$,

$$1.414213562\ldots = \sup\{1, 1.4, 1.414, 1.4142, 1.41421, 1.414213, \ldots\}.$$

We will prove that every nonnegative real number has (is) a decimal expansion. It then follows that every real number has (is) a decimal expansion (see Exercise 9.3.6).

Theorem 9.3.5. *Every nonnegative real number is a decimal expansion.*

Proof. Let $x \in \mathbb{R}^+$. We will define by recursion a sequence $\{d_n\}_{n=0}^{\infty}$ of decimal digits of x; i.e., the sequence $\{d_n\}_{n=0}^{\infty}$ will have the property that $x = \sup S$, where

$$S = \{d_0, d_0.d_1, d_0.d_1d_2, \ldots\} = \{d_0\} \cup \left\{ d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n}{10^n} \,\middle|\, n \in \mathbb{Z}^+ \right\}.$$

To define d_0, by Corollary 9.3.2(2), fix $d_0 \in \mathbb{Z}$ such that $d_0 \leq x < d_0 + 1$, and note that $d_0 \geq 0$.

Next, define d_1 to be the largest integer such that $0 \leq d_1 \leq 9$ and

$$d_0 + \frac{d_1}{10} \leq x.$$

Note that d_1 exists because it is the maximum element of the nonempty finite set

$$A = \left\{ m \in \mathbb{Z}^{\geq 0} \,\middle|\, d_0 + \frac{m}{10} \leq x \right\}.$$

This set A is nonempty because $0 \in A$ and finite because $A \subseteq \{0, 1, \ldots, 9\}$: if $m \in \mathbb{Z}$ with $m \geq 10$, then

$$d_0 + \frac{m}{10} \geq d_0 + \frac{10}{10} = d_0 + 1 > x,$$

a contradiction. Hence, for the recursion step, let $n \in \mathbb{Z}^+$ and assume that $d_0, d_1, \ldots, d_n$ have been defined and inductively satisfy the property that for all i, $1 \leq i \leq n$, d_i is the largest integer such that $0 \leq d_i \leq 9$ and

$$d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_i}{10^i} \leq x.$$

We define d_{n+1} to be the largest integer such that $0 \leq d_{n+1} \leq 9$ and

$$d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n}{10^n} + \frac{d_{n+1}}{10^{n+1}} \leq x.$$

An argument similar to that for d_1 proves that d_{n+1} exists: if $m \in \mathbb{Z}$ with $m \geq 10$, then

$$\begin{aligned} d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n}{10^n} + \frac{m}{10^{n+1}} &\geq d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n}{10^n} + \frac{10}{10^{n+1}} \\ &= d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n + 1}{10^n} \\ &> x, \end{aligned}$$

by the inductive hypothesis on d_n.

Given the recursively defined sequence $\{d_n\}_{n=0}^{\infty}$, we then define

$$S = \{d_0\} \cup \left\{ d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_n}{10^n} \;\middle|\; n \in \mathbb{Z}^+ \right\},$$

which is bounded above by x, by construction. By the Completeness Axiom, let $\delta \in \mathbb{R}$ be the supremum of S, and note that δ is a decimal expansion, by definition.

We claim that $\delta = x$. We know that $\delta \leq x$ by definition of $\sup S$, since x is an upper bound of S.

Assume for a contradiction that $\delta < x$; i.e., $x - \delta > 0$. By Corollary 9.3.2(3), we can fix $k \in \mathbb{Z}^+$ such that $\frac{1}{k} < x - \delta$. Then $\frac{1}{10^k} < \frac{1}{k} < x - \delta$ by Exercise 3.1.7, so $\delta + \frac{1}{10^k} < x$. Also note that, by maximality of d_k and the fact that $\delta = \sup S$,

$$\begin{aligned} x &< d_0 + \frac{d_1}{10} + \frac{d_2}{10^2} + \cdots + \frac{d_k + 1}{10^k} \\ &\leq \delta + \frac{1}{10^k} \\ &< x. \end{aligned}$$

This is our contradiction. Hence $\delta = x$, so x is a decimal expansion, as desired. □

We end this subsection by noting that we need slightly more than Theorem 9.3.5 for our proof of Theorem 8.3.11 that $\mathbb{R}$ is uncountable. That proof also relies on the fact that the decimal expansion of a real number is unique, except for instances such as

$$0.4\overline{9} = 0.49999\ldots = 0.5 = 0.50000\ldots.$$

What is needed are the following two additional results about decimal expansions, whose proofs we omit. See [**13**], where hints are provided.

Theorem 9.3.6. *If $x \in \mathbb{R}^+$ has more than one decimal expansion, then there exist $m, n \in \mathbb{Z}^+$ such that $x = \frac{m}{10^n}$, and in this case, the rational number x has exactly two decimal expansions.*

Theorem 9.3.7. *If $\delta_1, \delta_2 \in \mathbb{R}^+$ are nonterminating decimal expansions that disagree in at least one decimal digit, then $\delta_1 \neq \delta_2$.*

We've made the connection from our abstract definition of $\mathbb{R}$ back to decimal expansions, but in fact we don't really need it to study real analysis. A real number is just an element of $\mathbb{R}$, and $\mathbb{R}$ is a set that satisfies the axioms of a complete ordered field. This is enough. With these axioms, and the Completeness Axiom in particular, it is possible to define the usual notions from calculus, most importantly, the notion of a *limit*. And, we can prove all the important theorems of calculus, such as the Intermediate Value Theorem and the Mean Value Theorem. This is what we want, and it is the subject of a course in real analysis.

Exercises 9.3

1. Prove parts (1) and (3) of Corollary 9.3.2.
2. Let $S = \{1 - \frac{1}{n} \mid n \in \mathbb{Z}^+\}$. Prove that $\inf S = 0$ and $\sup S = 1$.

3. Let $a \in \mathbb{R}^+$. Prove that for all $x \in \mathbb{R}$, there exists $n \in \mathbb{Z}$ such that $na \le x < (n+1)a$.
4. (Nested Interval Theorem) For all $n \in \mathbb{Z}^+$, let $I_n = [a_n, b_n] \subseteq \mathbb{R}$ (so $a_n < b_n$). Assume that for all $n \in \mathbb{Z}^+$, $I_{n+1} \subseteq I_n$. Prove that $\bigcap_{n \in \mathbb{Z}^+} I_n \neq \emptyset$.
5. Does the result of Exercise 9.3.4 still hold if we replace $I_n = [a_n, b_n]$ by $I_n = (a_n, b_n)$ for all $n \in \mathbb{Z}^+$? Prove that your answer is correct.
6. Given a decimal expansion $d_0.d_1d_2d_3\ldots$, where $d_0 \in \mathbb{Z}^{\ge 0}$ and, for all $i \in \mathbb{Z}^+$, $d_i \in \mathbb{Z}$ with $0 \le d_i \le 9$, show that the set $X =$
$$\begin{aligned}&\{-d_0, -d_0.d_1, -d_0.d_1d_2, \ldots\}\\ &= \{-d_0\} \cup \left\{-d_0 - \frac{d_1}{10} - \frac{d_2}{10^2} - \cdots - \frac{d_n}{10^n} \;\middle|\; n \in \mathbb{Z}^+\right\}\end{aligned}$$
is bounded below. Thus, we define $-d_0.d_1d_2d_3\ldots$ to be the infimum of the set X and say that $-d_0.d_1d_2d_3\ldots$ is also a decimal expansion. Prove that every negative real number is a decimal expansion. (**HINT:** You will need Exercise 9.2.8b.)
7. Prove that $0.\overline{9} = 0.999\ldots = 1$. In other words, prove that
$$1 = \sup\left\{0, \frac{9}{10}, \frac{99}{100}, \frac{999}{1000}, \ldots\right\}.$$
8. (a) Prove that for all real numbers $x < y$, there exists $z \in \mathbb{Q}$ such that $x < z < y$ (i.e., prove that $\mathbb{Q}$ is *dense* in $\mathbb{R}$). (**HINT:** Consider cases $0 < x$, $x < y < 0$, and $x < 0 < y$. You will need Corollary 9.3.2(2) and (3).)

   (b) Prove that for all $x, y \in \mathbb{R}$ with $x < y$, there exists an irrational number z such that $x < z < y$. (**HINT:** Apply part (a) to $\frac{x}{\sqrt{2}}$ and $\frac{y}{\sqrt{2}}$.)

9.4. What next?

Let's go back to the last paragraph of Subsection 9.3.1. We said there that

> a real number is just an element of $\mathbb{R}$, and $\mathbb{R}$ is a set that satisfies the axioms of a complete ordered field.

How do we know that there is *any* set that satisfies the axioms of a complete ordered field?

Let's briefly return to set theory. We said in Section 4.4 that "a great deal of mathematics can be derived in ZF." What we mean is that, from the axioms in ZF, we can give set-theoretic definitions of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$, including the relevant arithmetic operations and order relations, *and prove that these definitions satisfy the usual properties.* See Exercise 7.2.4f for $\mathbb{Z}$, as well as Example 7.2.7 and Exercise 7.2.9 for $\mathbb{Q}$.

There are two ways to give a definition of $\mathbb{R}$ in ZF. One method defines a real number to be an equivalence class of *Cauchy sequences* (a particular type of sequence), and the other defines a real number to be a *Dedekind cut* (a particular type of set of rational numbers). In either case, one must define in ZF the arithmetic operations and order relation on $\mathbb{R}$ and prove in ZF that the axioms of a complete ordered field are satisfied. Saying more than this is beyond the scope of this book, so interested readers should consult [**9**] or [**12**].

Writing Mathematics

Chances are the mathematics course you are currently taking will require far more "writing in English" than any other math course you have previously taken, which may be surprising to you. Whereas in previous courses you may have written solutions to problems by simply writing line after line of formulas, with no English words at all, now you are required to write complete English sentences and paragraphs. This does not mean that no formulas will appear, but rather, formulas should be incorporated *grammatically* into sentences.

In any subject, whether it be history, biology, economics, or mathematics, our job is to *communicate* what we know, and how we know it, to others. Learning to write mathematics well requires a lot of practice and can be difficult at the beginning. The following guidelines[1] for writing mathematics point out some of the issues you'll want to keep in mind.

(1) All proofs (or solutions involving some sort of explanation) should be written in grammatically correct, complete English sentences.

(2) Begin a proof by assuming the relevant hypotheses. End each proof with a sentence that reiterates what has been proved.

- For example, if you are trying to prove that the product of two odd numbers is odd, then you should *begin* by saying, "Let m and n be odd integers." The last line of the proof might be something like, "Hence, mn is odd, as desired."

(3) Proofs (or solutions involving some sort of explanation) should include enough detail for the reader to understand your reasoning. Do not assume that the reader knows what you are talking about. Assume that your reader has the same mathematical background as you but does not know the proof you are writing.

[1] These guidelines were originally inspired by a "Writing checklist" created by Dr. Annalisa Crannell of Franklin & Marshall College. I have adapted them over time in the various "proof courses" I have taught in my career at Allegheny College.

(4) Be sure that what you have written is mathematically precise. Mean what you write, and write what you mean.

(5) Use proper mathematical notation and terminology.
- All variables must be *explicitly* defined.
  - For example, if you are trying to prove that the product of two odd numbers is odd, then you should *begin* by saying, "Let m and n be odd integers." If you are then tempted to write "$m = 2k+1$", then you should explicitly identify what k is and explain why it exists.
- Mathematical symbols should not be confused with English words. For example, the symbol "=" should be used only in mathematical formulas and computations, not as the verb "is" in a sentence.

(6) Proofs should explicitly make reference to any definitions or theorems used.

(7) Proofs should not contain any scratchwork or work done in the margins, nor any large sections of "crossed out" work.

(8) Proofread all solutions for correctness and clarity. Recopying your solutions is one useful way to accomplish this.

Bibliography

[1] R. B. J. T. Allenby, *Rings, Fields, and Groups: An Introduction to Abstract Algebra*, Edward Arnold (Publishers) Ltd., 1983.

[2] Robert G. Bartle and Donald R. Sherbert, *Introduction to Real Analysis,* 3rd edition, John Wiley & Sons, Inc., 2000.

[3] John A. Beachy and William D. Blair, *Abstract Algebra*, 2nd edition, Waveland Press, Inc., 1990.

[4] Robert J. Bond and William J. Keane, *An Introduction to Abstract Mathematics*, Brooks/Cole, 1999.

[5] Richard A. Brualdi, *Introductory Combinatorics*, 5th edition, Pearson Education, Inc., 2009.

[6] David M. Burton, *Elementary Number Theory*, 5th edition, McGraw-Hill, 2002.

[7] Keith Devlin, *Sets, Functions, and Logic: An Introduction to Abstract Mathematics,* 2nd edition, Chapman & Hall Mathematics, 1992.

[8] Peter J. Eccles, *An Introduction to Mathematical Reasoning: Numbers, Sets and Functions*, Cambridge University Press, 1997.

[9] Herbert B. Enderton, *Elements of Set Theory*, Academic Press, Inc., 1977.

[10] Herbert B. Enderton, *A Mathematical Introduction to Logic*, 2nd edition, Harcourt/Academic Press, 2001.

[11] Russell A. Gordon, *Real Analysis: A First Course*, 2nd edition, Pearson Education, Inc., 2002.

[12] Karel Hrbacek and Thomas Jech, *Introduction to Set Theory*, 2nd edition, revised and expanded, Marcel Dekker, Inc., 1984.

[13] Kenneth Rogers, *Advanced Calculus*, Charles E. Merrill Publishing Co., 1976.

[14] Dan Saracino, *Abstract Algebra: A First Course*, 2nd edition, Waveland Press, 2008.

[15] Daniel J. Velleman, *How to Prove It: A Structured Approach*, Cambridge University Press, 1994.

[16] Robert S. Wolf, *A Tour through Mathematical Logic*, The Mathematical Association of America (Incorporated), 2005.

Index

In the first part of this index, mathematical symbols and expressions are listed in their order of appearance in the book.

This accessible textbook gives beginning undergraduate mathematics students a first exposure to introductory logic, proofs, sets, functions, number theory, relations, finite and infinite sets, and the foundations of analysis. The book provides students with a quick path to writing proofs and a practical collection of tools that they can use in later mathematics courses such as abstract algebra and analysis. The importance of the logical structure of a mathematical statement as a framework for finding a proof of that statement, and the proper use of variables, is an early and consistent theme used throughout the book.

ISBN 978-1-4704-2899-0

9 781470 428990

AMSTEXT/26

For additional information and updates on this book, visit
www.ams.org/bookpages/amstext-26

AMS on the Web
www.ams.org

This series was founded by the highly respected mathematician and educator, Paul J. Sally, Jr.