Math 205-01

Foundations of Mathematics

Spring 2013

Instructor: R. Weir

The tools of mathematics
PRELIMINARY VERSION[1]

Tamara J. Lakins
Allegheny College

[1] December 29, 2012. © by Tamara J. Lakins, 2012, 2013. Do not disseminate.

Contents

Preface: To students

The idea for this book was conceived as a direct result of my experience teaching the introduction to proofs course at Allegheny College. My experience in teaching this material, and the feedback from the many students in my classes over the years, led me to write a text that I hope you find clear and effective. My hope is that you find it useful not only in this course, but also in your future mathematics, or related, courses.

This semester (Spring 2013) will be the fifth "classroom test" of this textbook. I am very interested in hearing your reaction to the book. What parts of it are clear and helpful, and what parts are hard to understand? Please keep a list of your comments about the book during the semester and give it to your instructor at the end of the course. Since no amount of proofreading will catch all typographical errors, please also keep a list of typos, and inform your instructor right away of any you find.

The text is essentially complete, but it is not yet in final form and has not yet been submitted to a publisher. For this reason, ***do not disseminate this version in any way.***

CHAPTER 1

Language, logic, and proof

1.1. Language and logic

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.

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 often needed in mathematics and computer science, such as sets, functions, and relations.

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.

□

We will often represent propositions with capital letters, such as P, Q, or R. Next we consider some non-examples 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 which 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.

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

TABLE 1.1. Truth table for $\wedge$

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 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 turns out that 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:

$$\text{if } a \text{ and } b \text{ are integers with } ab = 0, \text{ then } a = 0 \text{ or } b = 0.$$

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.

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

TABLE 1.2. Truth table for $\vee$

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.

P	$\neg P$
T	F
F	T

TABLE 1.3. Truth table for $\neg$

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.

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

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

□

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. Then*

(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 a final example 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" 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 Law* 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.

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 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

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

TABLE 1.5. Truth table for $\Rightarrow$

in a truth table (see Table 1.5). 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.

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

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

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.

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

TABLE 1.7. Truth table for $\Leftrightarrow$

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.*

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

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

EXAMPLE 1.1.9. 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) $1 + 1 = 2$ is necessary for $3 < 6$.

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

$$\text{if } 3 < 6 \text{ then } 1 + 1 = 2.$$

We can now see that this is a true implication by Table 1.7, since the hypothesis $3 < 6$ is true and also the conclusion $1+1 = 2$ is true. Here we can also see that mathematical implication has nothing to do with "causality"; the fact that $1 + 1 = 2$ is not "caused" by the fact that $3 < 6$.

(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, the two statements $|x| = 1$ and $(x = 1) \vee (x = -1)$ have the same truth value (see Proposition 1.1.10(3) and Definition 2.1.10).

□

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

PROPOSITION 1.1.10. *Let P and Q be statements. Then*

(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 and Exercise 1.1.2f. Use truth tables. □

Proposition 1.1.10(2) indicates how to find a useful denial of an implication.

EXAMPLE 1.1.11. 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 1.1.10(2) and DeMorgan's Laws, a useful denial of the given statement is, in natural English,

$$n \text{ is prime, and } n \text{ is an even integer other than } 2.$$

□

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

DEFINITION 1.1.12. 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 1.1.13. 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.

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

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 1.1.14. *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 1.1.13 again. We learn in calculus that statement (1.1) is a true statement; hence, its contrapositive, statement (1.3) is also true. We also learn in calculus that the converse of (1.1), statement (1.2), is a false statement. To prove this, 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).

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 $\mathbb{N} = \{1, 2, 3 \ldots\}$, 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 proposition is to modify it with a *quantifier.*

DEFINITION 1.1.15. Let $\mathcal{U}$ be the universe under consideration, and $P(x)$ be a predicate whose only free variable is x. Then the statements

"for all x, $P(x)$" Notation: $(\forall x)P(x)$
"there exists x such that $P(x)$" Notation: $(\exists x)P(x)$

are propositions.

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 1.1.16. 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. The statement $(\forall x \in \mathbb{R})(x^2 \geq 0)$ is true because the square of a real number is never negative. (See Exercise 2.1.2.)

(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 which 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 which 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, and yet 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 which shows that $(\forall x \in \mathbb{R})(x^3+52x^2+79x+1000 \geq 0)$ is a false statement.

□

Note that Definition 1.1.15 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. Notice that 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 here the definitions and notation for the various universes $\mathcal{U}$ we will consider in this text, namely, the natural numbers, the integers, the rational numbers, and the real numbers. Note that because every rational number has infinitely many representations (for example, $\frac{1}{2}=\frac{3}{6}=\frac{-2}{-4}=\dots$), 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. 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.

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, $b \in \mathbb{Z}$ 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 are women" is "at least one member of this class is a man". 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 "is a woman" was negated to become "is a man". 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 1.1.17. *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 1.1.15. □

Proposition 1.1.17 indicates how to find a useful denial of a quantified statement.

EXAMPLE 1.1.18. 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.

$$\begin{array}{ll} \neg(\forall x \in \mathbb{R})[x > 2 \Rightarrow x^2 > 4] & \text{is equivalent to} \\ (\exists x \in \mathbb{R})[\neg(x > 2 \Rightarrow x^2 > 4)], & \text{by Proposition 1.1.17(1),} \end{array}$$

which is equivalent to

$$\begin{array}{ll} (\exists x \in \mathbb{R})[x > 2 \wedge x^2 \leq 4] & \text{by Proposition 1.1.10(2).} \end{array}$$

Thus, a useful denial of the statement

$$\text{for all real numbers } x, \text{ if } x > 2, \text{ then } x^2 > 4$$

is

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

□

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 the notation $(\forall x \in \mathcal{U})P(x)$, which explicitly indicates 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 which are modified in other ways, such as 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

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

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 to 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.

$$\text{If } n \text{ is odd, then } n + 1 \text{ is even.}$$

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 always treat an 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" should be interpreted as

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

To repeat, *an implication will always be 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

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

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 1.1.19. A *tautology* is a proposition which is true for every possible assignment of truth values to the statement letters that occur in it. A *contradiction* is a proposition which 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? 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) is false; i.e., give a specific example of a function 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 and let $\mathcal{U}$ be the intended universe.

(a) Use Definition 1.1.15 to explain why the following statements are logically equivalent.
  (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 1.1.10, Proposition 1.1.17, 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 $2n + 3 = 5n + 1$.
(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 a *useful denial* (i.e., a statement equivalent to the negation) of each statement. (Here, $\vec{v_1}, \vec{v_2}, \vec{0}$ are fixed vectors, x_1, x_2, x, y are fixed real numbers, and f is a fixed function.)
(a) $(x_1\vec{v_1} + x_2\vec{v_2} = \vec{0}) \Rightarrow (x_1 = 0 \wedge x_2 = 0)$.
(b) $(\forall x_1)(\forall x_2)[f(x_1) = f(x_2) \Rightarrow x_1 = x_2]$.
(c) $(\forall y)(\exists x)[y = f(x)]$.

(14) Find the converse and contrapositive of each statement. (Here, $\vec{v_1}, \vec{v_2}, \vec{0}$ are fixed vectors, x_1, x_2 are fixed real numbers, and f is a fixed function.)
(a) $(x_1\vec{v_1} + x_2\vec{v_2} = \vec{0}) \Rightarrow (x_1 = 0 \wedge x_2 = 0)$.
(b) $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$.

(15) Find a *useful denial* (i.e., a statement equivalent to the negation) of each statement, and express it in mathematically precise, natural English. Express all conditional statements in the form "if ... then ...". (Here, a, b, c, n are fixed integers, f is a fixed function, and x_0, L, M 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 every integer a and every integer b, if $n = ab$, then $a = 1$ or $b = 1$.

(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 $0 < |x - a| < \delta$.

(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 statement. 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 $P \Rightarrow Q$.

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

TABLE 1.10. Truth table for $\Rightarrow$

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. Thus "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, 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 1.10. 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 1.10 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.

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.

TABLE 1.11

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 1.1.15 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 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.

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 1.12

Definition 1.1.15 also tells us how to prove (directly) that a statement of the form $(\exists x)P(x)$ is true.

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.

TABLE 1.13

Let's begin by proving the following:

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

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 is to 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 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 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 odd. Although not explicitly stated in Definition 1.2.1, it is further understood that *if* we know that n is odd, *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.

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 Section 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

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

Remember 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* 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.

*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*, Cambridge University Press, 1994. It is also used in Peter J. Eccles' book *An introduction to mathematical reasoning: numbers, sets and functions*, Cambridge University Press, 1997, where Velleman is credited with the terminology.

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 1.11 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 1.13 tells us that our goal is to:

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

We are *Given* that:

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

This means:

there exists an integer k 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 communicate a mathematical proof of this statement to others.

PROPOSITION 1.2.2. *If m is an even integer and n is an odd integer, then $m + n$ is an odd integer.*

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 we can fix $i \in \mathbb{Z}$ such that $m = 2i$. Similarly, since n is odd, by definition 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.

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

by

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

Or, we could have replaced it by

> Since m is even, by definition 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 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 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 formal 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 Appendix A 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 the Integers 1.2.3[†], below, 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 mathematical statements, no matter how "simple" they may seem, as "obvious".

Basic Properties of Integers 1.2.3.
For all integers a, b, c,

(Closure under $+$, $\cdot$)	$a + b$ 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)	$a + 0 = a$, $a \cdot 1 = a$, and $a \cdot 0 = 0$
(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 $a \neq 0$.

[†] and also the *Principle of Mathematical Induction*, to be discussed in Chapter 3.

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.
  (c) Prove that if m is even and n is odd, then $m + n$ is odd.

(3) Let m and n be integers.
  (a) Prove that if m is even, then mn is even.
  (b) Prove that if m and n are odd, then mn is odd.

CHAPTER 2

Techniques of proof

2.1. More direct proofs

In this chapter, we consider several examples which 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"* 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 $12 = 5n$.

As another example, note that $12 \mid 48$ since $48 = 12 \cdot 4$. □

Note that in Example 2.1.2, we have $3 \mid 12$ and $12 \mid 48$. Also note that $3 \mid 48$, since $48 = 3 \cdot 18$. This is an example of the "transitive" property of

*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 "$\mid$", while "$\frac{a}{b}$" (and "a/b") is a noun; it is the name of a particular rational number.

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	$a \mid c$
$a \mid b$	
$b \mid c$	

Our *Goal* is to show:

$$a \mid c.$$

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

find an integer k 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	find $k \in \mathbb{Z}$
$a \mid b$	with $c = ak$
$b \mid c$	

Now we consider our *Givens.* We know:

$$a \mid b,$$

so we can

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

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

Similarly, since

$$b \mid c,$$

we can

fix $m \in \mathbb{Z}$ 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 the 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. The list of axioms you may assume about the real numbers is given below (one additional axiom, the *Completeness Axiom*, will be discussed in Chapter 9). As with the Basic Properties of the Integers 1.2.3, we could take a smaller list of statements 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)	$a + 0 = a$, $a \cdot 1 = a$, 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 $a \neq 0$, there is a unique real number $a^{-1} = \frac{1}{a}$ such that $a \cdot a^{-1} = a \cdot \frac{1}{a} = 1$
(Division)	$\frac{b}{a}$ 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 $a < b$ and $c > 0$, then $ac < bc$
(Order property 3)	if $a < b$ and $c < 0$, then $ac > bc$

We recall also 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 is also a rational number; 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.

We'll begin by proving that

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

(Recall that we denote the set of all real numbers by $\mathbb{R}$). Our Given-Goal diagram contains the usual information.

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

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 2.1.5. *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 concepts that one already knows about, and do they 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).

EXAMPLE 2.1.6. 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 reviewing what is means for an integer to be prime.

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

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

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

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

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

For all positive integers n, $n^2 + n + 41$ 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. 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 2.1.8. *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 2.1.8 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 2.1.8 (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 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, we wish to determine how the expressions $\frac{x}{x+1}$ and $\frac{x+1}{x+2}$, where x is a positive real number, are related to each 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**
$x \in \mathbb{R}$	$\frac{x}{x+1} < \frac{x+1}{x+2}$
$x > 0$	

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 is 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 backward, 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 2.1.9. *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 2.1.9 isn't particularly profound; the point here is to demonstrate the method of working backward. 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 (⇒) direction

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

and the backward (⇐) 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 2.1.10. 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**
$a,\ b \in \mathbb{R}$	$\lvert a\rvert \leq b$
$b \geq 0$	
$-b \leq a \leq b$	
(i.e.,	
$-b \leq a$ and $a \leq b$)	

PROPOSITION 2.1.11. *For all real numbers $a,\ b \in \mathbb{R}$ with $b \geq 0$, $|a| \leq b$ iff $-b \leq a \leq b$.*

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 2.1.10, 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 2.1.10. 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 2.1.10, and hence $|a| \leq b$ by our assumption that $a \leq b$.

Case II: $a < 0$.

Then $|a| = -a$ by Definition 2.1.10. Since $-b \leq a$ by our assumption, we have $b \geq -a$ by an order property. Thus $|a| \leq b$, as desired.

□

Exercises 2.1

(1) Let a, b, and c be integers. Prove that for all integers x and y, if $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$.

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

(3) Prove that for all real numbers a and b, if $0 < a < b$, then $0 < a^2 < b^2$.

(4) 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".)

(5) 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".)

(6) Prove that for all $a, b \in \mathbb{Z}^+$, if $a \mid b$, then $a \leq b$.

(7) Let a and b be positive integers. Prove that if $a \mid b$ and $b \mid a$, then $a = b$.

(8) Prove that for all integers m, if m is odd, then there exists $k \in \mathbb{Z}$ such that $m^2 = 8k + 1$.

(9) Using definitions, prove by cases that for every integer n, $n^2 + n + 3$ is odd.

(10) 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 $15m + 12n = -6$.

(11) Let $a \in \mathbb{R}$. Prove that

(a) $|a| = |-a|$ (Hint: Consider three cases.)

(b) $-|a| \leq a \leq |a|$

(12) Let $a, b \in \mathbb{R}$. Prove that

(a) If $|a| = |b|$, then $a = b$ or $a = -b$.

(b) $|ab| = |a||b|$

(c) $|a - b| = |b - a|$

(d) (Triangle Inequality) $|a + b| \leq |a| + |b|$ (Hint: Use Exercise 11b and Proposition 2.1.11, or a proof by cases.)

(13) Recall that if $a \geq 0$ is a real number, then $\sqrt{a}$ is defined to be the nonnegative real number b with the property that $b^2 = a$. Prove that for all real numbers x, $\sqrt{x^2} = |x|$.

(14) Prove that for all real numbers x, $x^2 - x - 2 > 0$ if and only if $x < -1$ or $x > 2$.

(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 real numbers $x \geq 1$, $\frac{3|x-2|}{x} \leq 4$. (**HINT:** Consider two cases, based on the definition of absolute value.)

(18) Prove that for all nonnegative real numbers x, $\frac{2|x-3|}{x+1} \leq 7$.

2.2. Indirect proofs: Proofs by contradiction and contrapositive

2.2.1. Proof by contradiction. In Sections 1.2.2 and 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

$$\boxed{\text{there do not exist integers } m,\ n \in \mathbb{Z} \text{ such that } 14m + 21n = 100.}$$

Our Given-Goal diagram is the following:

Given	**Goal**
Nothing	$\neg(\exists m, n \in \mathbb{Z})[14m + 21n = 100]$

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

$$\neg(\exists m, n \in \mathbb{Z})[14m + 21n = 100]$$

as

$$(\forall m, n \in \mathbb{Z})[14m + 21n \neq 100]$$

does not seem to help:

Given	**Goal**
$m,\ n \in \mathbb{Z}$ arbitrary	$14m + 21n \neq 100$

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

$$0 = 1 \tag{2.6}$$

or

$$Q \wedge \neg Q, \quad \text{where } Q \text{ is some (possibly different) statement,} \tag{2.7}$$

is true. A statement such as (2.6) or (2.7) 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

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.

proof technique is called *proof by contradiction*, and it is an example of an *indirect* method of proof.

Let's use this technique to prove that there are no integers m and n such that $14m + 21n = 100$; i.e., $\neg(\exists m, n \in \mathbb{Z})[14m + 21n = 100]$. Since this will be a proof by contradiction, we will assume

$$\neg\neg(\exists m, n \in \mathbb{Z})[14m + 21n = 100],$$

i.e.,

$$(\exists m, n \in \mathbb{Z})[14m + 21n = 100],$$

and look for a contradiction. Our Given-Goal diagram now takes the following form.

Given	**Goal**
$m,\ n \in \mathbb{Z}$	Contradiction
$14m + 21n = 100$	

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 $14m + 21n = 100$; a logical thing to try is to factor to obtain

$$7(2m + 3n) = 100.$$

Recalling Definition 2.1.1, we see that this says that 100 is divisible by 7, which we know is false, since $100 = 2 \cdot 2 \cdot 5 \cdot 5$. We've found a contradiction[†]. Namely, we've proved that $7 \mid 100$ and $7 \nmid 100$; 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* $14m + 21n = 100$.

PROOF. Assume for the sake of a contradiction that we have integers m and n such that $14m + 21n = 100$. Then $7(2m + 3n) = 100$; i.e., $7 \mid 100$, by definition. But 100 is not divisible by 7, since $100 = 2 \cdot 2 \cdot 5 \cdot 5$, and thus 7 is

[†] We are using the *Fundamental Theorem of Arithmetic* here, which says that every positive integer greater than 1 can be written as a product of primes, where the primes that occur, and the number of times each prime factor occurs, is unique. We will prove the existence part of this theorem in Section 3.2 and the uniqueness part in Section 6.3.

not a divisor of 100 by the Fundamental Theorem of Arithmetic. Thus, we have a contradiction. Hence there do not exist integers m and n such that $14m + 21n = 100$. □

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

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

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

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 rational, 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 is odd
n^2 is odd	

By Definition 1.2.1, we know that we may fix $k \in \mathbb{Z}$ such that $n^2 = 2k+1$, but 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 1.1.14 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^2 is even
n 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	$a \geq \sqrt{b}$ or $a \leq -\sqrt{b}$
$b \geq 0$	
$a^2 \geq b$	

We could try a proof by contradiction here. Instead, we will once again take advantage of logic. Recall from Proposition 1.1.10 that a statement of the form $P \vee Q$ is logically equivalent to the statement $\neg P \Rightarrow Q$.

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

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: 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.

Note that in this case, either of the statements $a \geq \sqrt{b}$ or $a \leq -\sqrt{b}$ can be P or Q. 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	$a \leq -\sqrt{b}$
$b \geq 0$	
$a^2 \geq b$	
$a \ngeq \sqrt{b}$ (i.e., $a < \sqrt{b}$)	

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 we are working over the reals):

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

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 $a \ngeq \sqrt{b}$; 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 there are no integers m and n such that $m^2 = 4n + 2$. (**HINT:** Use Proposition 2.2.4.)

(3) Prove there are no integers m and n such that $m^2 = 4n + 3$.

(4) Prove that for all integers n, if $3 \mid n^2$, then $3 \mid n$. (**HINT:** You may use the fact (which we will prove in Chapter 6) that n is not divisible by 3 if and only if there exists an integer k such that $n = 3k + 1$ or $n = 3k + 2$.)

(5) 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.

(6) Using *definitions*, and the method of Section 2.2.3, give a *direct* proof of the fact that for all integers m and n, if mn is even, then m is even or n is even.

(7) Prove that for all real numbers x, $x^2 - x - 2 > 0$ if and only if $x < -1$ or $x > 2$.

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:

$$\neg(\exists p, q \in \mathbb{Z}^+)\left[\frac{p}{q} = \sqrt{2}\right]. \tag{2.8}$$

(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 (2.8) 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

$$p^2 = 2q^2. \tag{2.9}$$

Hence the integer p^2 is even, by definition. It follows by Proposition 2.2.4 that p is also even. Thus, again by definition, we may fix $k \in \mathbb{Z}$ such that $p = 2k$. Substituting into Equation (2.9), we have $(2k)^2 = 2q^2$, or $4k^2 = 2q^2$. By cancellation in $\mathbb{Z}$ (see Basic Integer Properties 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 we have already mentioned. 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 uniquely as a product of primes, in the sense that the primes that occur, and the number of times each prime occurs, in the factorization is unique.*

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. Here (and in the proof of Proposition 2.2.1), 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*, which we could stop and present now. By choosing to wait until Section 3.2 to prove 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, $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, and 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 Theorem 2.3.3, some prime must divide n, since n is a product of primes. Hence we can fix an integer j, $1 \leq j \leq k$, such that p_j divides n. Hence 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

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

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 2.2.4.)

(2) Prove that $\log_2 3$ is irrational. (**NOTE:** You may assume, without proof, the familiar "rules of exponents".)

(3) 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:

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

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.,

$$(\exists x \in \mathbb{Z}^+)(\forall y \in \mathbb{Z}^+)[x \leq y]. \tag{2.10}$$

Equation (2.10) 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 Equation (2.10) 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

$$\neg(\exists x \in \mathbb{Z})(\forall y \in \mathbb{Z})[x \leq y] \text{ is true; i.e., that}$$
$$(\forall x \in \mathbb{Z})(\exists y \in \mathbb{Z})[x > y] \text{ is true.}$$

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

Regardless of the number of quantifiers in a quantified statement, we always begin with the outermost quantifier and "work our way in." We must also be careful with mixed quantifiers: in general *order 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! Again, it is worth repeating, *order of quantifiers matters!*

EXAMPLE 2.4.1. Prove that

$$(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})[x + y = 0] \tag{2.11}$$

is a true statement and that

$$(\exists y \in \mathbb{R})(\forall x \in \mathbb{R})[x + y = 0] \tag{2.12}$$

is a false statement. □

PROOF. First note that statement (2.11) is one of our Basic Properties of the Real Numbers 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 (2.12) 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. □

A common statement involving mixed quantifiers that occurs in calculus is the notion of *limit.*

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.)

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. Thus Definition 2.4.2 says that for any distance $\varepsilon > 0$, we can find a 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 2.4.1, below.

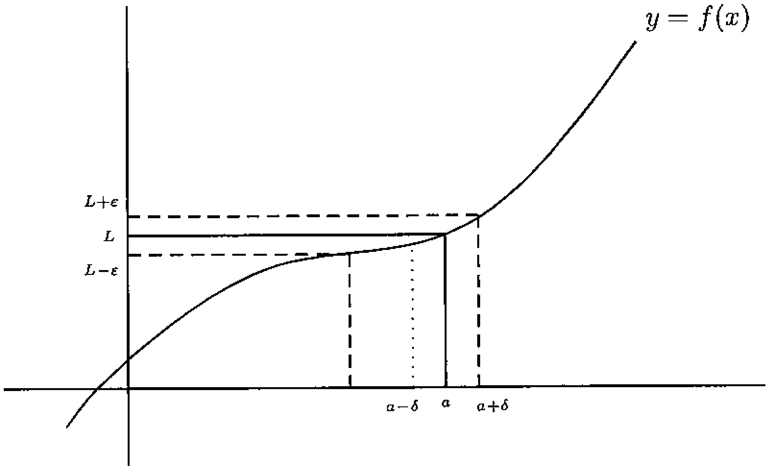


FIGURE 2.4.1. $\lim_{x \to a} f(x) = L$

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 $\lim_{x \to 2} f(x) = 5$. □

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

$$|(4x - 3) - 5| < \varepsilon, \text{i.e.,} \tag{2.13}$$

$$|4x - 8| < \varepsilon. \tag{2.14}$$

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 (2.14) is equivalent to $|4(x - 2)| < \varepsilon$, or $4|x - 2| < \varepsilon$, by Exercise 2.1.12b. 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 nonnegative, 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. □

We discuss the notion of limit further in Chapter 9.

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) $(\forall y \in \mathbb{R})(\exists x \in \mathbb{R})(\forall z \in \mathbb{R})[xy = xz]$
   (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) Assume that for every real number x, there is an integer n such that $n > x$.[‡] Prove that for every positive real number ϵ, there exists a positive integer N such that for all $n \geq N$, $\frac{1}{n} < \epsilon$.

[‡] This statement is called the *Archimedean Principle*, 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 the Basic Properties of Integers 1.2.3, such as the distributive property and the cancellation law. Another proof technique that can be useful when we want to prove a statement about all positive integers is the Principle of Mathematical Induction.

THEOREM 3.1.1 (Principle of Mathematical Induction (PMI)).
Let $P(n)$ be a statement about the positive integer n, so that n is the unique free variable in $P(n)$.

Suppose that

(PMI 1) *The statement $P(1)$ is true, and*
(PMI 2) *For all positive integers m,*

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

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

PMI essentially "says" that we can reach any positive integer n by starting at 1 and successively adding 1, which 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 $\mathbb{Z}^+$. It *is* possible to prove PMI from other statements, such as the *Well Ordering Principle*, which we discuss in Section 6.1, but then we would need to assume *those* statements as axioms.

For our first example of a proof using PMI, we will prove a statement you may have seen before, 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 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 $\sum_{k=1}^{m} k = \frac{m(m+1)}{2}$	$\sum_{k=1}^{m+1} k = \frac{(m+1)((m+1)+1)}{2}$

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

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

which we may rewrite as

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

As usual, to prove an equality we must pick one side and attempt to manipulate legally until we reach the other. We'll start with $\sum_{k=1}^{m+1} k$. The key point in the proof is that

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

i.e., that

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

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 3.1.2. *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 $\frac{1(1+1)}{2} = \frac{2}{2} = 1$, 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 $n \in \mathbb{Z}^+$,

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

□

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 3.1.2, nor to $n = 0$, as it did in the recursive definitions above. In general, the base case often corresponds to the first integer n for which the statement to be proved is true. Thus, we can state a more general version of the Principle of Mathematical Induction.

THEOREM 3.1.3 (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, and*
(PMI 2) *For all integers $m \geq n_0$,*

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

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}.

$$2^{m+1} = 2 \cdot 2^m \geq 2m^3 \qquad \text{by the Inductive Hypothesis, and}$$

$$(m+1)^3 = m^3 + 3m^2 + 3m + 1.$$

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 the 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 Equation (3.4).

We are ready to write down the formal proof.

PROPOSITION 3.1.4. *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) Prove that for all positive integers $n \geq 5$, $n^2 < 2^n$.

(7) Prove that for all positive integers $n \geq 7$, $(\frac{4}{3})^n > n$.

(8) Prove that for all positive integers $n \geq 4$, $2^n < n!$.

(9) Prove that for all positive integers n, $n^3 + 5n + 6$ is divisible by 3.

(10) Prove that for all positive integers n, $4^n - 1$ is divisible by 3.

(11) Prove that for all positive integers n, $5^{2n} - 1$ is divisible by 8.

(12) Let $a_1 = 2$, and $a_{n+1} = \frac{1}{2}(a_n + 4)$ for all $n \geq 1$.

(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, $a_n < 4$.

(c) Prove that for all positive integers n, $a_n = 4 - \frac{1}{2^{n-2}}$.

(13) The *Fibonnaci numbers* f_n, $n = 1, 2, 3, \ldots$, 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) Write out the first ten Fibonnaci numbers.

(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 of the first n Fibonnaci numbers, where $n \geq 1$ and prove it by induction.

(d) Use induction to prove that for all integers $k \geq 1$, $5 \mid f_{5k}$.

(14) Here is 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" $(\forall n)P(n)$ by induction on n. Clearly $P(1)$ is true, since any horse is the same color as itself. Now let $m \geq 1$ and assume that $P(m)$ is true; i.e., that for any set m horses, all the horses in the set are the same color. We prove $P(m+1)$ is true. Let S be a set of $m+1$ horses; say the horses in S are $h_1, h_2, \ldots, h_{m+1}$ (h for horse). Now $h_1, h_2, \ldots, h_m$ form a set of m horses, so since $P(m)$ is true, $h_1, h_2, \ldots, h_m$ are all the same color. Similarly, $h_2, h_3, \ldots, h_{m+1}$ form a set of m horses, so $h_2, h_3, \ldots, h_{m+1}$ are all the same color, again by the Inductive Hypothesis. It follows that all $m+1$ horses are the same color, since they are all the same color as horse h_2.

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 that

(PSMI 1) *The statement $P(1)$ is true, and*

(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

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

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

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

$$P(4) \text{ is true.} \tag{3.8}$$

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.

Let us now go back to the sequence that motivated our discussion:

$$\begin{aligned} a_1 &= 1, \\ a_2 &= 5, \\ \text{for all } n \geq 2,\ a_{n+1} &= a_n + 2a_{n-1}. \end{aligned}$$

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 $n \geq 1$,

$$a_n = 2^n + (-1)^n.$$

□

We now use strong induction to pay a debt from Section 2.2 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

$$(\forall m, n \in \mathbb{Z}^+)(p = mn \Rightarrow (m = 1 \text{ or } n = 1)).$$

THEOREM 3.2.2 (Fundamental Theorem of Arithmetic (Existence)). *Every positive integer $n > 1$ can be expressed as a product of primes.*[*]

PROOF. We prove the result by strong induction on n, where $n \geq 2$.

Base case: Note that 2 is prime, and hence 2 may be considered to be a product of a single 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

$$a = p_1 p_2 \dots p_i$$

and b is a product of primes

$$b = q_1 q_2 \dots q_j,$$

[*] We note that a prime number is itself considered to be a product of primes.

where $i, j \geq 1$ and $p_1, \ldots, p_i, q_1, \ldots, q_j$ are all prime integers. Hence

$$ab = p_1 p_2 \ldots p_i q_1 q_2 \ldots q_j,$$

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 a "stronger" statement than the Principle of Mathematical Induction. In fact these statements are logically equivalent. In other words, if we accept PMI as an axiom, then we can prove PSMI, and conversely, if we accept PSMI as an axiom, then we can prove PMI. (See Exercise 3.2.4.)

Exercises 3.2

(1) Let $a_1 = 2$, $a_2 = 4$, and $a_{n+1} = 5a_n - 6a_{n-1}$ for all $n \geq 2$. Conjecture a closed formula for a_n and then prove your result.

(2) Let $a_1 = 1$, $a_2 = 2$, and $a_{n+1} = \frac{1}{2}(a_n + a_{n-1})$ for all $n \geq 2$. Prove that for all positive integers n, $1 \leq a_n \leq 2$.

(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) Prove that PMI is logically equivalent to PSMI; in other words, given PMI, show that you can deduce the statement PSMI, and vice versa.

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 Section 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 $\{x \mid x \in X \text{ and } P(x)\}$.

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).

(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 elements of the set consist of all integers of the form $2k$, where k ranges through the set of positive integers.

It is important to remember that

the notation in (4.3) *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. *When in doubt, convert the definition of any set given constructively to one given conditionally.*

□

We have already made a very reasonable assumption about sets, namely, that a set is uniquely determined by its elements.* 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!)

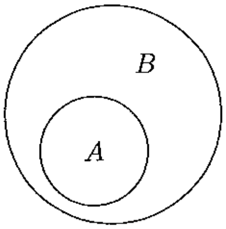
The assumption that a set is uniquely determined by its elements provides the definition of what it means for two sets to be equal (Definition 4.1.6); namely, two sets are equal if they have the same elements. Before working more with this idea, we first define what it means for one set to be "contained" 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 $A \nsubseteq B$ 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 $A \nsubseteq B$ symbolically, by forming a useful denial of Equation (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 lefthand set is also an element of the righthand set.
(3) $\{1\} \nsubseteq \{\{1\}, 2, 3\}$, since $1 \in \{1\}$, but $1 \notin \{\{1\}, 2, 3\}$.
(4) Let

$$A = \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[n = 4k + 1]\}$$
$$B = \{n \in \mathbb{Z} \mid n \text{ is odd}\}.$$

Prove that $A \subseteq B$ and $B \nsubseteq A$.

*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 $B \not\subseteq A$, we must again use Definition 4.1.4. Negating statement (4.4), 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 $B \not\subseteq A$. 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

$$A = \{n \in \mathbb{Z} \mid n + 3 \text{ is odd}\}$$
$$B = \{n \in \mathbb{Z} \mid n \text{ is even}\}.$$

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. The situation for $B \subseteq A$ is analogous.

Given	**Goal**
$n \in \mathbb{Z}$ arbitrary $n \in A$	$n \in B$

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 $n + 3$ is odd; hence, we may fix $k \in \mathbb{Z}$ such that $n + 3 = 2k + 1$. To show that $n \in B$, we must show that n is even. Note that

$$n = 2k + 1 - 3 = 2k - 2 = 2(k - 1),$$

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

$$n + 3 = 2k + 3 = 2(k + 1) + 1,$$

and hence $n + 3$ 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. The beginning Given-Goal diagram is given below.

Given	**Goal**
A, B, C arbitrary sets $A \subseteq B$ $B \subseteq C$	$A \subseteq C$

In order to start this proof correctly, we must focus on the Goal.

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 $x \in \emptyset$.

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) $\{1,2\} \in \{\{1,2,3\},\{1,3\},1,2\}$
   (c) $\{1,3\} \subseteq \{\{1,2,3\},\{1,3\},1,2\}$
   (d) $[5,6) \subseteq (4,6]$
   (e) $(7,9] \subseteq [6,9)$
   (f) $(5,9] \subseteq [6,10]$
   (g) $\{\emptyset\} \in \{\emptyset,\{\emptyset\}\}$
   (h) $\{\emptyset\} \subseteq \{\emptyset,\{\emptyset\}\}$
   (i) $\{\{\emptyset\}\} \in \{\emptyset,\{\emptyset\}\}$
   (j) $\{\{\emptyset\}\} \subseteq \{\emptyset,\{\emptyset\}\}$
   (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

$$A = \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})(n = 8k+7)\}$$
$$B = \{n \in \mathbb{Z} \mid (\exists j \in \mathbb{Z})(n = 4j+3)\}.$$

   (a) Is $A \subseteq B$? Prove or disprove.
   (b) Is $B \subseteq A$? Prove or disprove.

(5) Consider the sets

$$A = \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})(n = 4k+1)\}$$
$$B = \{n \in \mathbb{Z} \mid (\exists j \in \mathbb{Z})(n = 4j+9)\}.$$

Prove that $A = B$.

(6) Consider the sets

$$A = \{n \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[n = 3k]\}$$
$$B = \{n \in \mathbb{Z} \mid (\exists i,j \in \mathbb{Z})[n = 15i+12j]\}.$$

Prove that $A = B$.

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, B, and C 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

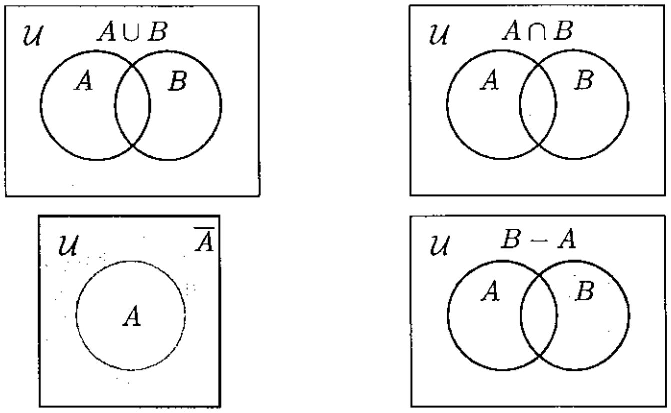
$$\begin{aligned} B - A &= \{x \in B \mid x \notin A\} \\ &= \{x \mid x \in B \text{ and } x \notin A\}. \end{aligned}$$

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 Venn diagrams which 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

$$A = \{1, 2, 4, 5, 7\}, \quad B = \{1, 3, 5, 9\}, \quad C = \{3, 6\},$$

and let the universal set be $\mathcal{U} = \mathbb{Z}^+$. Then

$$\begin{aligned}
A \cup B &= \{x \mid x \in A \text{ or } x \in B\} = \{1,2,3,4,5,7,9\} \\
A \cap B &= \{x \mid x \in A \text{ and } x \in B\} = \{1,5\} \\
A - B &= \{x \in A \mid x \notin B\} = \{2,4,7\} \\
B - A &= \{x \in B \mid x \notin A\} = \{3,9\} \\
A \cap C &= \{x \mid x \in A \text{ and } x \in C\} = \emptyset. \\
\overline{C} &= \{n \in \mathbb{Z}^+ \mid n \notin C\} = \{n \in \mathbb{Z}^+ \mid n \neq 3 \text{ and } n \neq 6\}
\end{aligned}$$

□

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. Recall that, if $a < b$ are real numbers, then

$$\begin{aligned}
(a,b) &= \{x \in \mathbb{R} \mid a < x < b\} \\
[a,b] &= \{x \in \mathbb{R} \mid a \leq x \leq b\} \\
(a,b] &= \{x \in \mathbb{R} \mid a < x \leq b\} \\
[a,b) &= \{x \in \mathbb{R} \mid a \leq x < b\} \\
(a,\infty) &= \{x \in \mathbb{R} \mid a < x\} \\
[a,\infty) &= \{x \in \mathbb{R} \mid a \leq x\} \\
(-\infty,b) &= \{x \in \mathbb{R} \mid x < b\} \\
(-\infty,b] &= \{x \in \mathbb{R} \mid x \leq b\}.
\end{aligned}$$

The next example shows that the union or intersection of two intervals need not be another interval.

EXAMPLE 4.2.4.

$$\begin{aligned}
(2,4] \cap (3,5) &= (3,4] \\
[2,4] \cap (4,5) &= \emptyset \\
(2,4) \cup (3,5) &= (2,5) \\
\overline{(2,4]} &= (-\infty,2] \cup (4,\infty)
\end{aligned}$$

□

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 $x \in B$, then $x \notin A - C$.*

PROOF. Let A, B, and C be arbitrary sets and assume that $A \cap B \subseteq C$. Assume also that $x \in B$; we show that $x \notin A - C$.

For the sake of a contraction, assume that $x \in A - C$. Then, by Definition 4.2.1(3), $x \in A$ and $x \notin C$. 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.

Hence, $x \notin A - C$, 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 $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. Let $x \in (A \cap B) \cup (A \cap C)$; 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*[†] containing x and y in that order. More generally, if $n \in \mathbb{Z}^+$ and $A_1, A_2, \ldots, A_n, A$ 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)

$$\begin{aligned} \mathbb{R} \times \mathbb{R} = \mathbb{R}^2 &= \{(x, y) \mid x, y \in \mathbb{R}\} \\ \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3 &= \{(x, y, z) \mid x, y, z \in \mathbb{R}\} \\ \mathbb{R}^n &= \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in \mathbb{R}\} \end{aligned}$$

$\mathbb{R}^2$ is the familiar *Cartesian* or *Euclidean* plane, $\mathbb{R}^3$ is Euclidean space, and $\mathbb{R}^n$ is n-dimensional Euclidean space.

□

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 that $A \times B$ (and $B \times A$) has 6 elements. In general it can be proved (see Chapter 8) 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.*

[†] Technically, an ordered pair is a set. See Exercise 4.2.20.

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 $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.

To see 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). 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,

$$\begin{aligned} 1 &\notin \mathcal{P}(X), &&\text{since} && 1 \not\subseteq X,\\ \{1\} &\in \mathcal{P}(X), &&\text{since} && \{1\} \subseteq X,\\ \emptyset &\in \mathcal{P}(X), &&\text{since} && \emptyset \subseteq X. \end{aligned}$$

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 4.2.21). This fact can help you 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 as an Exercise 4.2.18. 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 4.2.18.

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, $A \subseteq B$. □

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 if $A \subseteq B$, then $\overline{B} \subseteq \overline{A}$.

(3) Let A, B, and C be sets.
 (a) Prove that if $A \subseteq B \cup C$ and $A \cap B = \emptyset$, then $A \subseteq C$.
 (b) Prove that if $A \subseteq B \cap C$ and $A \cap B = \emptyset$, then $A = \emptyset$.

(4) Let A and B be sets.
 (a) Prove that $A = (A \cap B) \cup (A - B)$.
 (b) Prove that $A \cup B = A \cup (B - A)$.

(5) Let A, B, C, and D be sets with $C \subseteq A$ and $D \subseteq B$. Prove that $D - A \subseteq B - C$.

(6) Let A, B, and C be sets. Prove that
 (a) $(A \cup C) - B \subseteq (A - B) \cup C$.
 (b) $(A \cup C) - B = (A - B) \cup C$ iff $B \cap C = \emptyset$.

(7) 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.

(8) 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.

(9) 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.

(10) Let $A = \{a, b\}$, $B = \{1, 2\}$, and $C = \{c, d, e\}$. Find $A \times B$ and $C \times A$. Explain why $A \times (B \times C) \neq (A \times B) \times C$.

(11) 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) $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.

(12) 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.

(13) 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.

(14) Which of the following statements are true for every set A? Explain.

$$\emptyset \subseteq \mathcal{P}(A), \qquad \emptyset \in \mathcal{P}(A),$$
$$A \subseteq \mathcal{P}(A), \qquad A \in \mathcal{P}(A).$$

(15) 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) $X = \{a, b, \{a, b\}\}$.

(16) Let $A = \{1, 2\}$. Find $\mathcal{P}(\mathcal{P}(A))$.

(17) 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) $\mathcal{P}(A) - \mathcal{P}(B)$

(18) Let A and B be sets such that $A \subseteq B$. Prove that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

(19) 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: $\mathcal{P}(A - B) = \mathcal{P}(A) - \mathcal{P}(B)$.

(20) 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

$$(x, y) = (z, w) \text{ iff } x = z \text{ and } y = w.$$

(21) Use induction on $n \geq 0$ to prove that if the set A has n elements, then $\mathcal{P}(A)$ has 2^n elements.

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:

$$A \cup B \cup C = (A \cup B) \cup C, \quad \text{or}$$
$$A \cup B \cup C = A \cup (B \cup C).$$

and by Exercise 4.2.1d, both lead to the same set. This means that the notation $A \cup B \cup C$ is "unambigious", and we can therefore generalize this notion to n sets, where $n \in \mathbb{Z}^+$.

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, our theorems from Section 4.2 generalize to this setting.

Theorem 4.3.2. *Let $n \in \mathbb{Z}^+$ and $A, B_1, B_2, \ldots, B_n$ be sets. Then*

(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 others for Exercise 4.3.5.

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 $A, B_1, \ldots, B_m$,

$$A \cap (B_1 \cup B_2 \cup \cdots \cup B_m) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m).$$

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

$$(A \cap (B_1 \cup B_2 \cup \cdots \cup B_m)) \cup (A \cap B_{m+1}) = \\ ((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m)) \cup (A \cap B_{m+1})$$

by the Induction Hypothesis, and

$$((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m)) \cup (A \cap B_{m+1}) = \\ (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_m) \cup (A \cap B_{m+1})$$

as desired.

Hence by induction we have that for all $n \in \mathbb{Z}^+$, and for all sets A, $B_1, B_2, \ldots, B_n$,

$$A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

□

Note the statement of Theorem 4.3.2 has implied universal quantifiers. For example, the actual statement of Theorem 4.3.2(2) is that "for all $n \in \mathbb{Z}^+$ and *for all* sets A, $B_1, B_2, \ldots, B_n$,

$$A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)."$$

Notice in particular how these implied universal quantifiers affect the statement of the Induction Hypothesis.

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 infinite union $\bigcup_{i=1}^{\infty} A_i$ and infinite 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.,

$$A_1 = \{1,2\}, \quad A_2 = \{2,3\}, \quad A_3 = \{3,4\}, \quad \ldots.$$

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 a 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 Given-Goal diagram for $\bigcup_{i=1}^{\infty} A_i \subseteq \mathbb{Z}^+$ is

Given	**Goal**
$n \in \bigcup_{i=1}^{\infty} A_i$ arbitrary	$n \in \mathbb{Z}^+$

Definition 4.3.3 tells us that we may fix a particular $i \in \mathbb{Z}^+$ with $n \in A_i$, so that we then left with the following Given-Goal diagram:

Given	**Goal**
$n \in A_i$, where $i \in \mathbb{Z}^+$	$n \in \mathbb{Z}^+$

But this is automatic, since elements of A_i are integers by definition.

Next, the Given-Goal diagram for $\mathbb{Z}^+ \subseteq \bigcup_{i=1}^{\infty} A_i$ is

Given	**Goal**
$n \in \mathbb{Z}^+$	$(\exists i \in \mathbb{Z}^+)[n \in A_i]$

which can be replaced by

Given	**Goal**
$n \in \mathbb{Z}^+$	find a particular $i \in \mathbb{Z}^+$ with $n \in A_i$

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 $\bigcup_{i=1}^{\infty} A_i \subseteq \emptyset$ and $\emptyset \subseteq \bigcup_{i=1}^{\infty} A_i$. We know the second statement is true by Proposition 4.1.9.

The Given-Goal diagram for $\bigcap_{i=1}^{\infty} A_i \subseteq \emptyset$ is

Given	**Goal**
$n \in \bigcap_{i=1}^{\infty} A_i$ arbitrary	$n \in \emptyset$

Since $n \in \emptyset$ is always false, we must show that the Given (i.e., the hypothesis) $n \in \bigcap_{i=1}^{\infty} A_i$ is also false, which we express by:

Given	**Goal**
$n \in \mathbb{Z}^+$ arbitrary	$n \notin \bigcap_{i=1}^{\infty} A_i$

Phrased another way, rather than prove $\bigcap_{i=1}^{\infty} A_i \subseteq \emptyset$, we instead prove $\overline{\emptyset} \subseteq \overline{\bigcap_{i=1}^{\infty} A_i}$ (see Exercise 4.2.2); here, the universal set is $\mathbb{Z}^+$.

Definition 4.3.3 allows us to replace this Given-Goal diagram by

Given	**Goal**
$n \in \mathbb{Z}^+$ arbitrary	$(\exists i \in \mathbb{Z}^+)[n \notin A_i]$

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 $A_i = [0, \frac{1}{i})$. We claim that $\bigcup_{i \in \mathbb{Z}^+} A_i = [0, 1)$ and $\bigcap_{i \in \mathbb{Z}^+} A_i = \{0\}$.

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 Exercise 2.4.4, 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 $\bigcup_{i \in \mathbb{Z}^+} A_i = [0, 1)$. □

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 I and $\{A_i \mid i \in I\}$ 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 $j \in I$,* $\quad \bigcap_{i \in I} A_i \subseteq A_j$.

(2) *for each* $j \in I$, $\quad A_j \subseteq \bigcup_{i \in I} A_i$.

(3) $\overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A_i}$.

(4) $\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A_i}$.

(5) $B \cap \bigcup_{i \in I} A_i = \bigcup_{i \in I}(B \cap A_i)$.

(6) $B \cup \bigcap_{i \in I} A_i = \bigcap_{i \in I}(B \cup A_i)$.

Exercises 4.3

(1) For $i \in \mathbb{Z}^+$, let $A_i = (-i, i)$.

(a) Find $\bigcup_{i=1}^{\infty} A_i$ and $\bigcap_{i=1}^{\infty} A_i$.

(b) Prove your answers to part (a) are correct.

(2) For $i \in \mathbb{Z}^+$ with $i \geq 2$, let $A_i = [\frac{1}{i}, i)$.

(a) Find $\bigcup_{i=2}^{\infty} A_i$ and $\bigcap_{i=2}^{\infty} A_i$.

(b) Prove your answers to part (a) are correct.

(3) For $i \in \mathbb{Z}^+$ with $i \geq 2$, let $A_i = (\frac{1}{i}, i]$.

(a) Find $\bigcup_{i=2}^{\infty} A_i$ and $\bigcap_{i=2}^{\infty} A_i$.

(b) Prove your answers to part (a) are correct.

(4) For $i \in \mathbb{Z}^+$ let $A_i = [1 - \frac{1}{i}, 3 - \frac{1}{i})$.

(a) Find $\bigcup_{i \in \mathbb{Z}^+} A_i$ and $\bigcap_{i \in \mathbb{Z}^+} A_i$.

(b) Prove your answers to part (a) are correct.

(5) Finish the proof of Theorem 4.3.2. Let $n \in \mathbb{Z}^+$ and $A, B_1, B_2, \ldots, B_n$ be sets. Prove by induction on n that

(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) $\overline{B_1 \cap B_2 \cap \cdots \cap B_n} = \overline{B_1} \cup \overline{B_2} \cup \cdots \cup \overline{B_n}$.

(6) Let I be a nonempty set and let $\{A_i \mid i \in I\}$ be an indexed family of sets. Prove that $\bigcap_{i \in I} A_i \subseteq \bigcup_{i \in I} A_i$.

(7) 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 $\bigcup_{i \in I} A_i \subseteq X$.

4.4. Axiomatic set theory*

We have been dealing, as many mathematicians do, with sets on a very informal basis. 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 $P(x)$.

Now, consider the following collection of sets which are not elements of themselves:

$$A = \{x \mid x \notin x\}.$$

It's reasonable to ask whether A an element of A.

Note that if $A \in A$, then by the definition of A, $A \notin A$.

Hence $A \notin A$.

But if $A \notin A$, then by definition of A, $A \in A$.

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*, and it lead to the development of *axiomatic set theory*.

There are several ways that one can axiomatize set theory. Our axioms for the integers give a list of properties that the integers (and possibly other sets of "numbers"?) possess. Any system of axioms for set theory provides a set of "rules" that indicate what types of sets exist. The most common axiomatizations provide axioms that assert, among other things, the existence of $\emptyset$, finite sets, $\mathbb{N}$, and any set built up from other sets using $\cap$, $\cup$, $\subseteq$, and the power set operation (handled carefully). The "subset axiom", which (informally) asserts that "definable" subsets of any given set exist, prevents sets from getting "out of control", or "too big", as is the problem with the Russell set A above.

Studying set theory from an axiomatic point of view is the topic of another course. Interested readers should consult [**6**]. In day–to–day mathematics, issues like Russell's Paradox rarely come up.

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 most likely 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 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, the notion of function is studied within the context of axiomatic set theory.

In this book, an informal definition of function is adequate for our purpose. All further definitions regarding functions will be precise, as will 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.*

*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.

DEFINITION 5.1.1. Let X and Y be sets. A *function f from the set X to the set Y* is a correspondence which assigns to each element $x \in X$ a unique element $y \in Y$, which is denoted by $f(x)$.

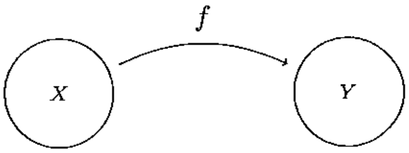
The set X is called the *domain* of f. We often denote the domain of a function f by $\operatorname{dom} f$. The set Y is called the *codomain* of f.

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 $f : X \to Y$.

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 $x \in X$:

$$\begin{aligned} f(1) &= c & \qquad f(2) &= e \\ f(3) &= e & \qquad f(4) &= a. \end{aligned}$$

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; i.e., there is sometimes a difference between the codomain of a function, which you can think of as the "target set" in which 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 $x \in X$, $I_X(x) = x$.

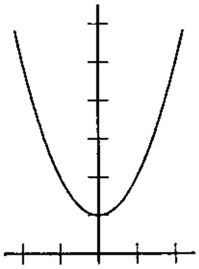
We consider several more examples below.

EXAMPLE 5.1.10. 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 $\mathbb{R}^2$.



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 $f(3) = 10 = f(-3)$;
- 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, $\operatorname{ran} f = [1, \infty)$, although one needs ideas from analysis to prove this, since one needs to prove that every nonnegative real number has a nonnegative square root.

We state here the general theorem which asserts when nth roots exist. As already noted, the proof requires ideas from analysis and is omitted here.

THEOREM 5.1.11. *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* $x \in \mathbb{R}$, $(x^n)^{1/n} = x$.

We return to the claim that $\operatorname{ran} f = [1, \infty)$. This statement says that two sets are equal. We've already done the scratchwork for $\operatorname{ran} f \subseteq [1, \infty)$, so we need to find a proof that $[1, \infty) \subseteq \operatorname{ran} f$.

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

$$\operatorname{ran} f = \{y \in \mathbb{R} \mid (\exists x \in \operatorname{dom} f)[y = f(x)]\} \tag{5.1}$$

$$= \{y \in \mathbb{R} \mid (\exists x \in \mathbb{R})[y = x^2 + 1]\}. \tag{5.2}$$

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 5.1.11, $\sqrt{y-1}$ 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 5.1.11 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 $[1, \infty) \subseteq \operatorname{ran} f$. □

We note, as we have done in the past, that you should be sure not to confuse the scratchwork, where we worked backwords 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 5.1.10. This is because while f and g have the same domain and are defined by the same formula, f and g have *different codomains.* □

We have emphasized that a function is alway specified by giving the domain, codomain, and correspondence. In courses like calculus, often just a formula is given.

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 $f(x)$ (note that 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 be 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

$$\begin{aligned}
f(x) &= f\left(\frac{4y+1}{y-2}\right) \\
&= \frac{2\left(\frac{4y+1}{y-2}\right)+1}{\left(\frac{4y+1}{y-2}\right)-4} \\
&= \frac{2\left(\frac{4y+1}{y-2}\right)+1}{\left(\frac{4y+1}{y-2}\right)-4} \cdot \frac{y-2}{y-2} \\
&= \frac{2(4y+1)+y-2}{4y+1-4(y-2)} \\
&= \frac{8y+2+y-2}{4y+1-4y+8} \\
&= \frac{9y}{9} = y,
\end{aligned}$$

as desired.

Hence, $\{y \in \mathbb{R} \mid y \neq 2\} \subseteq \operatorname{ran} f$, and so

$$\operatorname{ran} f = \{y \in \mathbb{R} \mid y \neq 2\}.$$

□

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 the Intermediate Value Theorem.

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 $n \in \mathbb{Z}$,

$$f(n) = \begin{cases} n-1 & \text{if } n \text{ is even;} \\ n+3 & \text{if } n \text{ is odd.} \end{cases}$$

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 $f(11) = 11 + 3 = 14$.

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 $n + 3$ 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 $f(n) = n + 3 = -54$. Thus, $-54 \in \operatorname{ran} f$ since $f(-57) = -57 + 3 = -54$. 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 $\mathbb{Z}$. Let $f : \mathcal{P}(\mathbb{Z}) \to \mathcal{P}(\mathbb{Z})$ 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 each element of the codomain of f, and hence each output of f, is a *set* of integers. Note that

$$\begin{aligned} f(\{-3, 2, 17\}) &= \{n \in \mathbb{Z} \mid n \neq -3 \text{ and } n \neq 2 \text{ and } n \neq 17\}, \\ f(\{99, 100, 101, \dots\}) &= f(\{n \in \mathbb{Z} \mid n \geq 99\}), \\ &= \{n \in \mathbb{Z} \mid n \leq 98\} = \{\dots, 96, 97, 98\}, \\ f(\{n \in \mathbb{Z} \mid n \neq 0\}) &= \{0\}, \quad \text{and} \\ f(E) &= O, \end{aligned}$$

where $E = \{n \in \mathbb{Z} \mid n \text{ is even}\}$ and $O = \{n \in \mathbb{Z} \mid n \text{ is odd}\}$.

Next, we show 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.

Next, we 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 $B = f(X) = \overline{X}$

Theorem 4.2.6(17) tells us the set X of integers we seek.

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), $f(\overline{B}) = \overline{\overline{B}} = B$. Hence $\mathcal{P}(\mathbb{Z}) \subseteq \operatorname{ran} f$, as desired. □

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 an *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,

$$g(\sqrt{2}, 3.97) = (-3.97, \sqrt{2}) \qquad \text{and} \qquad g\left(-\pi, \tfrac{\sqrt[3]{4}}{7}\right) = \left(\tfrac{-\sqrt[3]{4}}{7}, -\pi\right).$$

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

$$\begin{aligned} g(x,y) &= f(w,-z) \\ &= (-(-z), w) \\ &= (z, w), \end{aligned}$$

as desired. Hence $\mathbb{R}^2 \subseteq \operatorname{ran} g$. □

Exercises 5.1

(1) For each of the following functions determine the domain of f and $\operatorname{ran} f$. Prove 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

$$f(n) = \begin{cases} n - 1 & \text{if } n \text{ is even;} \\ n + 3 & \text{if } n \text{ is odd.} \end{cases}$$

Prove that $\operatorname{ran} f = \mathbb{Z}$.

(6) (This problem assumes the Fundamental Theorem of Arithmetic 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) In this problem

$$\mathbb{P} = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \mid n \geq 0, a_0, \ldots, a_n \in \mathbb{R}\}$$

denotes the set of all polynomials with real coefficients. Let $F : \mathbb{P} \to \mathbb{P}$ by, for all $p \in \mathbb{P}$, $F(p) = p'$, where p' is the (symbolic) derivative of p.

(a) Find $F(3x^5 - \frac{7}{3}x^2 + x)$.

(b) Find $p \in \mathbb{P}$ such that $F(p) = \frac{1}{2}x^3 + 5x^2 - 4$.

(8) Are the functions $f(x) = \dfrac{9 - x^2}{x + 3}$ and $g(x) = 3 - x$ 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 $f(x) = x^2$ is increasing on $[0, \infty)$.

(b) Show that $f(x) = x^2$ is decreasing on $(-\infty, 0]$.

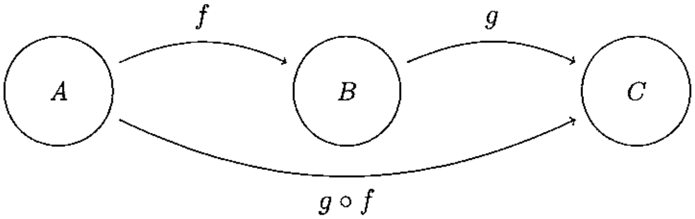
(c) Show that $f(x) = x^3$ is increasing on $\mathbb{R}$. (**HINT:** Argue by cases.)

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 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 $g : B \to C$, we have the following picture for $g \circ f : A \to C$.



EXAMPLE 5.2.2. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ by, for all $x \in \mathbb{R}$,

$$f(x) = x^2$$
$$g(x) = x + 1.$$

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

$$(f \circ g)(1) = 4, \quad \text{and}$$
$$(g \circ f)(1) = 2.$$

□

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 $x \in \mathbb{R}$,

$$f(x) = \begin{cases} x^2, & \text{if } x \geq 0 \\ x - 1, & \text{if } x < 0, \end{cases}$$

$$g(x) = \begin{cases} x + 1, & \text{if } x \geq 1 \\ 2x, & \text{if } x < 1. \end{cases}$$

We find $g \circ f$ and leave $f \circ g$ as an exercise. Given $x \in \mathbb{R}$,

$$(g \circ f)(x) = \begin{cases} g(x^2), & \text{if } x \geq 0 \\ g(x - 1), & \text{if } x < 0. \end{cases}$$

To compute $g(x^2)$ when $x \geq 0$, the definition of g tells us that we must consider whether $x^2 \geq 1$ or $x^2 < 1$. Since $x \geq 0$, we know that $x^2 \geq 0$. Thus $x^2 \geq 1$ when $x \geq 1$, and $x^2 < 1$ when $0 \leq x < 1$. Similarly, to compute $g(x - 1)$ when $x < 0$, we must consider whether $x - 1 \geq 1$ or $x - 1 < 1$. When $x < 0$, $x - 1 < -1 < 1$, so $g(x - 1) = 2(x - 1)$. Thus, we have

$$(g \circ f)(x) = \begin{cases} x^2 + 1, & \text{if } x \geq 1 \\ 2x^2, & \text{if } 0 \leq x < 1 \\ 2x - 2, & \text{if } x < 0. \end{cases}$$

□

We next prove some 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 composite 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

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)), \quad \text{and}$$
$$((h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x)).$$

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.

□

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 $f(x) = x^2 + 2x$ and $g(x) = 2x + 1$

(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 $x \in \mathbb{R}$,

$$f(x) = \begin{cases} x^2 & \text{if } x \geq 0 \\ x - 1 & \text{if } x < 0, \end{cases}$$

$$g(x) = \begin{cases} x + 1 & \text{if } x \geq 1 \\ 2x & \text{if } x < 1. \end{cases}$$

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 $I_Y \circ f = f$.

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 5.1.10 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

$$(\forall x_1, x_2 \in X)[x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)] \tag{5.3}$$

or, equivalently,

$$(\forall x_1, x_2 \in X)[f(x_1) = f(x_2) \Rightarrow x_1 = x_2]. \tag{5.4}$$

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

$$(\forall y \in Y)(\exists x \in X)[y = f(x)]. \tag{5.5}$$

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 $f : X \underset{\text{onto}}{\overset{\text{1-1}}{\to}} Y$.

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 (5.3) and (5.5), 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 (5.3) and (5.4) give us two ways to show that a function is 1-1 (in fact, there are other ways, as well; see Exercise 5.3.6). Often (although not always) it is the definition provided by (5.4) 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 (5.4) 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**
$x_1, x_2 \in \mathbb{R}$ arbitrary $f(x_1) = f(x_2)$ $a \neq 0$	$x_1 = x_2$

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 (5.5). Again, we begin with the general Given-Goal diagram.

Given	**Goal**
$y \in \mathbb{R}$ arbitrary	$(\exists x \in \mathbb{R})[y = f(x)]$

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

$$f(1) = c, \qquad f(2) = e, \qquad f(3) = e, \qquad f(4) = a,$$
$$g(a) = 10, \quad g(b) = 0, \quad g(c) = 5, \quad g(d) = 20, \quad g(e) = 15.$$

Then f is not 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 8.2.7. □

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.

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 is needed in order to determine whether or not the function is "onto". Recall also that we showed in Example 5.1.10 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 as an exercise.

THEOREM 5.3.6. *Let $f : X \to Y$ and $\operatorname{ran} f$ 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.

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

$$\begin{aligned} (x_1)^2 + 1 &= (x_2)^2 + 1, \quad \text{so} \\ (x_1)^2 &= (x_2)^2. \end{aligned}$$

By Exercise 2.1.4, we obtain

$$x_1 = x_2 \qquad \text{or} \qquad x_1 = -x_2,$$

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 as an exercise the verification that we can restrict the domain of f 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. Since the proof of this theorem involves formalizing the notion of function in axiomatic set theory and the Axiom of Choice, we omit it here.

THEOREM 5.3.8. *Let $f : X \to Y$. Then there exists a subset $X_0 \subseteq X$ such that restricting the domain of f to X_0 yields a 1-1 function; i.e., the function $g : X_0 \to Y$ defined by, for all $x \in X_0$, $g(x) = f(x)$ is 1-1.*

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 (which is just a statement about the range of the function) using algebraic methods. For example, the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3 - x$, for all $x \in \mathbb{R}$, is onto, but the easiest way to show this is by using the Intermediate Value Theorem from calculus. Furthermore, methods other than (5.3) and (5.4) from Definition 5.3.1 can be used to show that a function is 1-1.

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 (5.4) and (5.5) 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 :

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.

Given	**Goal**
$(z, w) \in \mathbb{R}^2$ arbitrary	$(\exists\ (x, y) \in \mathbb{R}^2)[(z, w) = f(x, y)]$

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 75), 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 properties that 1-1 and onto functions possess.

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) as exercises.

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. Remember that

> *it is the goal that determines how the proof should proceed.*

Since $g : Y \to Z$, statement (5.5) 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 $y = f(x) \in Y$ 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) $f : \mathbb{R} - \{\frac{3}{5}\} \to \mathbb{R} - \{\frac{2}{5}\}$ by $f(x) = \frac{2x+1}{5x-3}$
(c) $f : \mathbb{R} - \{-\frac{d}{c}\} \to \mathbb{R} - \{\frac{a}{c}\}$ 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 $f(x, y, z) = (x, x + y, x + z)$
(l) $f : \mathbb{R}^3 \to \mathbb{R}^3$ by $f(x, y, z) = (x + y, y + z, x + z)$
(m) $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f(x, y, z) = (x + y, y + z)$
(n) $F : \mathbb{P} \to \mathbb{P}$ 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

$$f(n) = \begin{cases} 2n+1 & \text{if } n \text{ is even;} \\ 4n+1 & \text{if } n \text{ is odd.} \end{cases}$$

(4) Let X be a set. Prove that the identity function $I_X : X \to X$ is a bijection.

(5) 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.

(6) 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.

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.

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 $y \in Y$,

$$y = f(x) \Leftrightarrow x = g(y). \tag{5.6}$$

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

$$\begin{aligned} (g \circ f)(x) &= g(f(x)) \\ &= g(y) \\ &= x, \quad \text{by (5.6)}, \\ &= I_X(x), \quad \text{by definition.} \end{aligned}$$

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 (5.6) is true.

First we assume that $y = f(x)$ and show that $x = g(y)$.

$$\begin{aligned} g(y) &= g(f(x)) \\ &= (g \circ f)(x) \\ &= I_X(x), \quad \text{by hypothesis,} \\ &= x, \end{aligned}$$

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 as an exercise. Hence f and g are inverse functions, by Proposition 5.4.3. □

Note that not all functions are invertible.

EXAMPLE 5.4.6.

(1) We showed in Example 5.3.5 that 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.*

□

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 a standard method for proving that an object is unique, given that it exists. 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 $g_1 = g_2$. □

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

$$g^{-1}(0) = b, \quad g^{-1}(5) = c, \quad g^{-1}(10) = a, \quad g^{-1}(15) = e, \quad g^{-1}(20) = d.$$

□

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

$$\begin{aligned}(x_1)^3 + 1 &= (x_2)^3 + 1, \quad \text{so}\\ (x_1)^3 &= (x_2)^3, \quad \text{and so}\\ \sqrt[3]{(x_1)^3} &= \sqrt[3]{(x_2)^3}.\end{aligned}$$

Thus $x_1 = x_2$ by Theorem 5.1.11(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 $(x, y) \in \mathbb{R}^2$. □

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 facts about the derivative, one can prove that ln is 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 rule 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,

$$e^x = y \Leftrightarrow x = \ln y.$$

□

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) $f : \mathbb{R} - \{\frac{3}{5}\} \to \mathbb{R} - \{\frac{2}{5}\}$ by $f(x) = \frac{2x+1}{5x-3}$

(c) $f : \mathbb{R} - \{-\frac{d}{c}\} \to \mathbb{R} - \{\frac{a}{c}\}$ 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 $f(x, y, z) = (x, x + y, x + z)$

(l) $f : \mathbb{R}^3 \to \mathbb{R}^3$ by $f(x, y, z) = (x + y, y + z, x + z)$

(m) $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f(x, y, z) = (x + y, y + z)$

(n) $F : \mathbb{P} \to \mathbb{P}$ 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

$$f(n) = \begin{cases} 2n+1 & \text{if } n \text{ is even;} \\ 4n+1 & \text{if } n \text{ is odd.} \end{cases}$$

(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) 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.

(6) Give a direct proof (i.e., using definitions) of Corollary 5.4.9.

(7) 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.)

(8) 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.

5.5. Functions and sets

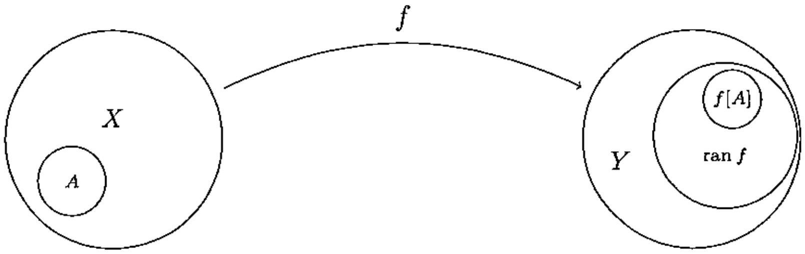
Given a function $f : X \to Y$, f is defined on *elements* of X. However, sometimes we also want to consider the image of an entire *subset* of X under f.

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. When $f : X \to Y$, $x \in X$, and $A \subseteq X$,

$f(x)$ is an *element* of Y

$f[A]$ is a *subset* of Y.

Our use of square brackets, rather than parentheses, is meant to help you remember the difference.

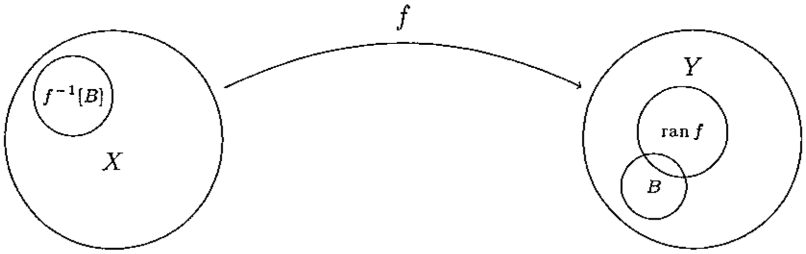
The next definition is of a related concept.

Definition 5.5.2. Let $f : X \to Y$ and $B \subseteq Y$. The *inverse image of B under f* 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.

EXAMPLE 5.5.3. Let $X = \{0,1,2,3,4,5\}$ and $Y = \{7,9,11,12,13\}$. Let $f : X \to Y$ by

$$\begin{array}{lll} f(0) = 11 & f(1) = 9 & f(2) = 7 \\ f(3) = 9 & f(4) = 11 & f(5) = 9. \end{array}$$

Then

$$\begin{aligned} f[\{2,3\}] &= \{f(x) \mid x \in \{2,3\}\} \\ &= \{f(2), f(3)\} \\ &= \{7,9\}, \\ f[\{1,5\}] &= \{f(1), f(5)\} \\ &= \{9\}, \\ f^{-1}[\{7,9\}] &= \{x \in X \mid f(x) \in \{7,9\}\} \\ &= \{1,2,3,5\}, \\ f^{-1}[\{12,13\}] &= \{x \in X \mid f(x) \in \{12,13\}\} \\ &= \emptyset. \end{aligned}$$

□

EXAMPLE 5.5.4. Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 3x + 7$ for all $x \in \mathbb{R}$. Let $A = [0,1]$, $B = (-\infty, -2)$, and $C = (0, \infty)$. A graph of $y = f(x)$ is given below in Figure 5.5.1.

Then

$$\begin{aligned} f[A] &= \{y \in \mathbb{R} \mid (\exists x \in A)[y = f(x)]\} \\ &= \{y \in \mathbb{R} \mid (\exists x \in \mathbb{R})[0 \le x \le 1 \text{ and } y = 3x + 7]\} \\ &= \{3x + 7 \mid 0 \le x \le 1\}. \end{aligned}$$

Note that by properties of real numbers,

$$\begin{aligned} 0 \le x \le 1 &\Leftrightarrow 0 \le 3x \le 3 \\ &\Leftrightarrow 7 \le 3x + 7 \le 10. \end{aligned}$$

Hence $f[A] = [7,10]$.

Also, $f[B] = \{3x + 7 \mid x < -2\}$. Since $x < -2$ iff $3x + 7 < 1$, $f[B] = (-\infty, 1)$.

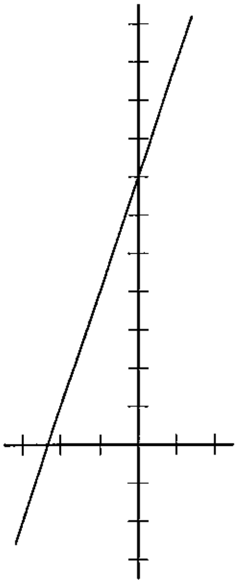


FIGURE 5.5.1. Graph of $f(x) = 3x + 7$

Finally,

$$\begin{aligned} f^{-1}[C] &= \{x \in \mathbb{R} \mid f(x) \in C\} \\ &= \{x \in \mathbb{R} \mid 3x + 7 > 0\} \\ &= \left\{x \in \mathbb{R} \mid x > -\frac{7}{3}\right\} \\ &= \left(-\frac{7}{3}, \infty\right). \end{aligned}$$

□

Note that in the previous example, our task is much easier because the function $f(x) = 3x + 7$ 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 $A = [\frac{\pi}{2}, \frac{3\pi}{2}]$ and $B = [0, 1]$. A graph of $y = f(x)$ is given below in Figure 5.5.2.

From the graph of $f(x) = \cos x$ we see that

$$\begin{aligned} f[A] &= \{f(x) \mid x \in A\} \\ &= \left\{\cos x \;\middle|\; \frac{\pi}{2} \le x \le \frac{3\pi}{2}\right\} \\ &= [-1, 0]. \end{aligned}$$

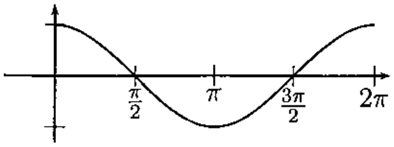


FIGURE 5.5.2. Graph of $f(x) = \cos x$

It's important to note here that $f\left[\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]\right]$ 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.

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

$$\begin{aligned} f^{-1}[B] &= \{x \in [0, 2\pi] \mid f(x) \in B\} \\ &= \{x \in [0, 2\pi] \mid 0 \leq \cos x \leq 1\} \\ &= \left[0, \frac{\pi}{2}\right] \cup \left[\frac{3\pi}{2}, 2\pi\right]. \end{aligned}$$

□

Now that we've seen several examples, we prove a theorem about how induced set functions 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) as exercises.

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[†], 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 $f[A \cup B] = f[A] \cup f[B]$.

[†] Or, we could again argue by cases.

(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, assume that $x \in f^{-1}[D]$, since the argument for $x \in f^{-1}[C]$ is analogous. By 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 $f^{-1}[C \cup D] = f^{-1}[C] \cup f^{-1}[D]$.

□

5.5.1. Functions and sets, more carefully. We emphasize 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. As we noted at the beginning of this section, there is the potential for the notation $f[A]$ to be confusing. We are using [ ], rather than ( ), around the "input" as a signal to help us remember the difference: while $f(x)$ is an element of Y, $f[A]$ is a subset of Y. Although this notation is one of the standard notations for this concept (as is, for example, the notation $f(A)$ for the same concept), in some sense we are using the notation for the function f in two different ways, and we must tell from context what is meant by that notation. Technically, we are actually defining a *new* function $\overrightarrow{f} : \mathcal{P}(X) \to \mathcal{P}(Y)$ defined on subsets of X; given $A \in \mathcal{P}(X)$ (i.e., $A \subseteq X$), $\overrightarrow{f}(A) = \{y \in Y \mid (\exists x \in A)[y = f(x)]\}$.

The situation with the inverse image of a set $B \subseteq Y$ under f,

$$f^{-1}[B] = \{x \in X \mid f(x) \in B\},$$

is similar, although as we mentioned earlier, the potential for confusion is even higher. Not only do we need to tell the difference between elements and subsets of Y, but also the function f *may not be invertible*, i.e., f^{-1} may not exist. The "exponent" is nothing more than notation, and again, this notation is standard in mathematics (as is, for example, the notation $f^{-1}(B)$ for the same concept); mathematicians can tell from context what the notation means. Technically, however, we are actually defining a *new* function $\overleftarrow{f} : \mathcal{P}(Y) \to \mathcal{P}(X)$ defined on subsets of Y; given $B \in \mathcal{P}(Y)$ (i.e., $B \subseteq Y$), $\overleftarrow{f}(B) = \{x \in X \mid f(x) \in B\}$.

The notation $\overrightarrow{f}$ and $\overleftarrow{f}$ is due to Eccles [**5**], who uses it in his textbook at this same level.

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!

$$f[\{-3, 2, 7\}], \quad f[[-1, 3]], \quad f[(-\infty, -2)],$$
$$f^{-1}[\{-2, 5, 16\}], \quad f^{-1}[[-2, 3]], \quad f^{-1}[[8, \infty)]$$

(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) Complete the proof of 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) Complete the proof of 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 a function $f : X \to Y$ such that $f[A] \subseteq f[B]$, but $A \not\subseteq B$.

(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 a function $f : X \to Y$ 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

In this chapter, we return to studying properties of the integers.

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, 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.

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 which 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 which is less than or equal to -4, which is -6, so $q = -2$ and $r = 2$; i.e., $-4 = 3 \cdot -2 + 2$

□

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 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:

$$\text{If } S \subseteq \mathbb{Z}^{\geq 0} \text{ and } S \neq \emptyset, \text{ then } (\exists m \in S)(\forall x \in S)[m \leq x].$$

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 the Integers 1.2.3.* 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 $a > 0$.

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 which are less than or equal to a will be the numbers you get by successively subtracting off 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

$$\begin{aligned} 0 \leq r - b &= a - bq - b \\ &= a - (b + 1)q, \end{aligned}$$

*In fact, WOP is *equivalent* to PMI; i.e., one can prove WOP by assuming PMI, and conversely one can prove PMI by assuming WOP.

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 $b \mid (r - r_1)$, $r - r_1 = 0$, and hence $r = r_1$, as desired. Since $bq + r = bq_1 + r_1$, this implies that $bq = bq_1$, and 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:

$$a = 3q + 0, \quad a = 3q + 1, \quad \text{or } a = 3q + 2, \quad \text{where } q \in \mathbb{Z}.$$

In fact, this was the hint in Exercise 2.2.4.

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 $b = 4$. What 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 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 $9k + 8$.

(3) Use the Division Algorithm 6.1.1 to prove that for all $n \in \mathbb{Z}^+$, $6 \mid n(n + 1)(2n + 1)$.

(4) Prove Theorem 6.1.1 using a proof by induction, rather than the Well Ordering Principle.

(5) Prove that PMI is logically equivalent to WOP. (**HINT:** For WOP $\Rightarrow$ PMI, proceed by contradiction and consider $\{n \in \mathbb{N} \mid P(n) \text{ is false}\}$. For PMI $\Rightarrow$ WOP, proceed by contradiction and let $P(n)$ be the statement "$n \notin S$".)

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, and an algorithm for computing it. We begin, as usual, with a definition.

DEFINITION 6.2.1. Let $a, b \in \mathbb{Z}$ with at least one of a, 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 $\gcd(a, b)$.

One obvious way of computing the gcd of two integers, such as -12 and 30, is to list the positive divisors of each and pick out the greatest one that is common to both:

positive divisors of -12:	$1, 2, 3, 4, 6, 12$
positive divisors of 30:	$1, 2, 3, 5, 6, 10, 15, 30.$

So, we see that $(-12, 30) = 6$.

This method of computing the greatest common divisor of two integers is tedious when the numbers are large; in this section, we use the Division Algorithm to give an algorithm (called the *Euclidean Algorithm*) for computing gcd's.

We first state an informal version of the Euclidean Algorithm, as well as prove the result used in establishing its correctness.

EUCLIDEAN ALGORITHM (INFORMAL) 6.2.2.

(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.

We need to verify the claim in step (3).

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. □

Before giving a precise statement of the Euclidean Algorithm (our informal version above will suffice for now) and proving it, we give an example of how it is used to find the gcd of any two positive integers.

EXAMPLE 6.2.4. Find $(1414, 666)$.

We use the Division Algorithm 6.1.1 with $a = 1414$ and $b = 666$ to obtain $1414 = 666 \cdot 2 + 82$; i.e., $q = q_1 = 2$ and $r = r_1 = 82$.

We then repeat, this time taking $b = 666$ in place of a and $r = 82$ in place of b to obtain $666 = 82 \cdot 8 + 10$; i.e., we now have $q_2 = 8$ and $r_2 = 10$. We continue until we reach a remainder of 0; this is most easily followed as a sequence of Division Algorithm computations.

$$1414 = 666 \cdot 2 + 82 \tag{6.1}$$

$$666 = 82 \cdot 8 + 10 \tag{6.2}$$

$$82 = 10 \cdot 8 + 2 \tag{6.3}$$

$$10 = 2 \cdot 5 + 0. \tag{6.4}$$

By repeated application of Lemma 6.2.3, note that

$$(1414, 666) = (666, 82) = (82, 10) = (10, 2) = (2, 0) = 2.$$

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 $(1414, 666) = 2$.

We can also use this process to show that 2, the gcd of 1414 and 666, can be expressed explicitly in terms of 1414 and 666. We will show that we can find $x, y \in \mathbb{Z}$ such that $2 = 1414x + 666y$. This fact, that the gcd 2 is an integer "linear combination" of 1414 and 666, 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:

$$\begin{aligned}
2 &= 82 - 10 \cdot 8 = 82 + 10(-8) \\
10 &= 666 - 82 \cdot 8 = 666 + 82(-8) \\
82 &= 1414 - 666 \cdot 2 = 1414 + 666(-2).
\end{aligned}$$

Then substitute:

$$\begin{aligned}
2 &= 82 + 10(-8) \\
&= 82 + (666 + 82(-8))(-8) \\
&= 82 + 666(-8) + 82(64) \\
&= 82(65) + 666(-8) \\
&= (1414 + 666(-2))(65) + 666(-8) \\
&= 1414(65) + 666(-130) + 666(-8) \\
&= 1414(65) + 666(-138).
\end{aligned}$$

So, $2 = 1414x + 666y$, where $x = 65$ and $y = -138$ (you should confirm this statement using your calculator). □

The precise statement of the Euclidean Algorithm is notationally complicated. To see how to formalize the Euclidean Algorithm, note that we can view the computations (6.1)–(6.4) as follows:

$$\begin{aligned} a &= bq_1 + r_1 \\ b &= r_1q_2 + r_2 \\ r_1 &= r_2q_3 + r_3 \\ r_2 &= r_3q_4 + r_4, \end{aligned}$$

where $a = 1414$, $b = 666$ and

$$\begin{aligned} q_1 &= 2 & \qquad r_1 &= 82 \\ q_2 &= 8 & r_2 &= 10 \\ q_3 &= 8 & r_3 &= 2 \\ q_4 &= 6 & r_4 &= 0. \end{aligned}$$

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 decrease to $r_4 = 0$, where the desired gcd is $d = r_3$.

THEOREM 6.2.5 (Euclidean Algorithm). *Let $a, b \in \mathbb{Z}^+$ and let $d = (a, b)$. Then there are two finite lists of integers $q_1, q_2, \ldots, q_{k+1}$ and $r_0, r_1, r_2, \ldots, r_{k+1}$, where $k \geq 0$, such that $b = r_0 > r_1 > r_2 > \cdots > r_k > r_{k+1} = 0$,*

$$\begin{aligned} a &= bq_1 + r_1 \\ b &= r_1q_2 + r_2 \\ r_1 &= r_2q_3 + r_3 \\ &\vdots \\ r_{k-2} &= r_{k-1}q_k + r_k \\ r_{k-1} &= r_kq_{k+1} + r_{k+1}, \end{aligned}$$

and $d = r_k$.

PROOF. Informally, repeated applications of the Division Algorithm 6.1.1 and Lemma 6.2.3. Formally, a proof by induction or the Well Ordering Principle (see Exercise 6.2.5). □

We next formally define the terminology used in Example 6.2.4.

DEFINITION 6.2.6. Let $a, b, n \in \mathbb{Z}$. The integer n is a *linear combination* of a and b if there exists $x, y \in \mathbb{Z}$ such that $n = ax + by$.

COROLLARY 6.2.7. *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. Informally (for $a, b > 0$), repeated back substitution using the computations that arise when the Euclidean Algorithm 6.2.5 is used to compute (a, b). Formally, a proof by induction. □

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 6.2.8. *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 $(-1414, 666) = (1414, 666) = 2$ by our previous computation. We can rewrite the linear combination found in Example 6.2.4 as follows:

$$\begin{aligned} 2 &= 1414(65) + 666(-138) \\ &= -1414(-65) + 666(-138) \end{aligned}$$

to write $(-1414, 666)$ as a linear combination of -1414 and 666.

Exercises 6.2

(1) (a) Use the Euclidean Algorithm to find $(12378, 3054)$ and write $(12378, 3054)$ as an integer linear combination of 12378 and 3054.
(b) Repeat part (a) for $(227, 143)$.
(c) Repeat part (a) for $(272, 1479)$.

(2) Prove Proposition 6.2.8.

(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) 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 6.2.9. 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 6.2.1. Use Corollary 6.2.7 to show that d is a gcd of a and b as defined in Definition 6.2.9. 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 6.2.9 exists.

(b) Using the definition in Definition 6.2.9, 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 6.2.9, and prove that $d_1 = d_2$.)

(c) Show that if the greatest common divisor (as defined in Definition 6.2.9) of a and b is d , then $d = (a, b)$ (as defined in Definition 6.2.1). In other words, show that for all $c \in \mathbb{Z}^+$, if $c \mid a$ and $c \mid b$, then $c \leq d$.

(5) Use induction to prove the Euclidean Algorithm 6.2.5.

6.3. Relatively prime integers and the Fundamental Theorem of Arithmetic

Corollary 6.2.7 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 6.2.7, 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 6.2.7.

($\Leftarrow$) 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 numbers. (See Definition 2.1.7.)

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. □

COROLLARY 6.3.4. *Let $p, a, b \in \mathbb{Z}^+$, where p is prime. If $p \mid ab$, then $p \mid a$ or $p \mid b$.*

PROOF. Exercise 6.3.1. □

Corollary 6.3.4 is the missing tool that we need in order to prove the uniqueness part of the Fundamental Theorem of Arithmetic Theorem 2.3.3.

THEOREM 6.3.5 (Fundamental Theorem of Arithmetic (Uniqueness)). *Every positive integer greater than* 1 *can be written uniquely as a product of primes, in the sense that the primes that occur, and the number of times each prime occurs, in the factorization is unique.*

Exercises 6.3

(1) Prove Corollary 6.3.4. Let a, b, p be positive integers. Assume that p is prime and $p \mid ab$. Prove that $p \mid a$ or $p \mid b$. Prove that this result does not hold in general, when p is not prime.

(2) 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 $\frac{a}{d}, \frac{b}{d} \in \mathbb{Z}$.)

(3) Prove (without using the Fundamental Theorem of Arithmetic) 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 $(a, b) \neq 1$.

6.4. 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.

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. 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.

THEOREM 6.4.4.

(1) *For all* $a \in \mathbb{Z}$, $a \equiv a \mod m$ *(i.e.,* $\equiv$ *is "reflexive").*
(2) *For all* $a, b \in \mathbb{Z}$, *if* $a \equiv b \mod m$, *then* $b \equiv a \mod m$ *(i.e.,* $\equiv$ *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.,* $\equiv$ *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

$$(a - b) + (b - c) = mk + m\ell, \quad \text{so}$$
$$a - c = m(k + \ell).$$

Thus $m \mid (a - c)$ and hence $a \equiv c \mod m$, by Definition 6.4.1.

□

6.4.1. Remainders mod m.

We can use the Division Algorithm 6.1.1 to show that every integer is congruent modulo m to a unique remainder upon division by m.

DEFINITION 6.4.5. The set of (least, nonnegative) *remainders modulo m* is the set

$$R_m = \{r \in \mathbb{Z} \mid 0 \leq r < m\}.$$

For example, $R_4 = \{0, 1, 2, 3\}$, the set of possible remainders upon division by 4 obtained from the Division Algorithm.

THEOREM 6.4.6. *Let $a, b \in \mathbb{Z}$.*

(1) *There is a unique $r \in R_m$ such that $a \equiv r \mod m$.*

(2) *The congruence $a \equiv b \mod m$ holds if and only if there exists $r \in R_m$ such that $a \equiv r \mod m$ and $b \equiv r \mod m$ (i.e., a and b have the same remainder $r \in R_m$ when divided by m).*

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

(1) Let $r \in R_m$, so that $0 \leq r < m$. By Proposition 6.4.3,

$$a \equiv r \mod m \Leftrightarrow (\exists q \in \mathbb{Z})[a = mq + r].$$

The Division Algorithm 6.1.1 says that there exist unique $q, r \in \mathbb{Z}$ such that $a = mq+r$ and $0 \leq r < m$. Thus a is congruent modulo m to a unique element of R_m.

(2)

($\Rightarrow$) Assume $a \equiv b \mod m$. By the Division Algorithm 6.1.1, fix $q_1, q_2, r_1, r_2 \in \mathbb{Z}$ such that $a = mq_1 + r_1$ and $b = mq_2 + r_2$ and $r_1, r_2 \in R_m$. Then

$$\begin{aligned} r_1 &\equiv a \mod m && \text{by Proposition 6.4.3 and symmetry of } \equiv \mod m \\ &\equiv b \mod m && \text{by transitivity, since } a \equiv b \mod m \\ &\equiv r_2 \mod m && \text{by transitivity, since } b \equiv r_2 \mod m. \end{aligned}$$

Since $r_1, r_2 \in R_m$ and $r_1 \equiv r_1 \mod m$, we have $r_1 = r_2$ by (1).

($\Leftarrow$) Assume we have $r \in R_m$ 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 of $\equiv \mod m$.

□

Theorem 6.4.6 says, for example, that each integer a is congruent to exactly one of 0, 1, 2, or 3 modulo 4, which 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 mod m, we can greatly simply the proofs we saw in Section 6.1. We consider this idea in Section 6.4.2.

6.4.2. Arithmetic mod m.

Not surprisingly, perhaps, we can perform arithmetic mod m. The next theorem shows that all but one of the arithmetic operations behave as expected.

THEOREM 6.4.7. *Let $a_1, a_2, b_1, b_2, c \in \mathbb{Z}$ such 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) $a_1 b_1 \equiv a_2 b_2 \mod m$, *and*
(5) $a_1 c \equiv a_2 c \mod m$.

PROOF. We prove (4) and leave the rest to the exercises.

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

$$\begin{aligned} a_1 b_1 - a_2 b_2 &= a_1 b_1 - a_1 b_2 + a_1 b_2 - a_2 b_2 \\ &= a_1(b_1 - b_2) + b_2(a_1 - a_2) \\ &= a_1 m\ell + b_1 mk \\ &= m(a_1 \ell + b_1 k). \end{aligned}$$

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 6.4.8. Prove that the square of any integer has one of the forms $3k$, $3k + 1$, where $k \in \mathbb{Z}$.

PROOF. By Theorem 6.4.6, 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 6.4.7 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 $a \equiv b \mod n$ and $m \mid n$, then $a \equiv b \mod m$.

(3) Let $a, b \in \mathbb{Z}$ and $n \in \mathbb{Z}^+$. Prove that if $a \equiv b \mod n$, then $(a, n) = (b, n)$. (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 $ab \equiv ac \mod n$ need not imply that $b \equiv c \mod n$. Give an example to show that $a^2 \equiv b^2 \mod n$ need not imply that $a \equiv b \mod n$.

(5) Use congruences to prove that for all integers a, $a^3 \equiv 0, 1,$ or $6 \mod 7$.

6.5. Congruence classes

Theorem 6.4.6 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 to the same element of R_m modulo 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.

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

$$\begin{aligned}
[0]_4 &= \{x \in \mathbb{Z} \mid x \equiv 0 \mod 4\} && (6.5)\\
&= \{x \in \mathbb{Z} \mid 4 \mid x\} && (6.6)\\
&= \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k]\} && (6.7)\\
&= \{\dots, -8, -4, 0, 4, 8, \dots\}. && (6.8)
\end{aligned}$$

In fact (this requires proof), $[0]_4 = [4]_4 = [-8]_4 = \cdots$; 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.

$$\begin{aligned}
[1]_4 &= \{x \in \mathbb{Z} \mid x \equiv 1 \mod 4\} && (6.9)\\
&= \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k + 1]\} && (6.10)\\
&= \{\dots, -7, -3, 1, 5, 9, \dots\} && (6.11)\\
&= [9]_4 = [-7]_4 = \cdots . && (6.12)
\end{aligned}$$

$$\begin{aligned}
[2]_4 &= \{x \in \mathbb{Z} \mid x \equiv 2 \mod 4\} && (6.13)\\
&= \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k + 2]\} && (6.14)\\
&= \{\dots, -6, -2, 2, 6, 10, \dots\} && (6.15)\\
&= [10]_4 = [-6]_4 = \cdots . && (6.16)
\end{aligned}$$

$$\begin{aligned}
[3]_4 &= \{x \in \mathbb{Z} \mid x \equiv 3 \mod 4\} && (6.17)\\
&= \{x \in \mathbb{Z} \mid (\exists k \in \mathbb{Z})[x = 4k + 3]\} && (6.18)\\
&= \{\dots, -5, -1, 3, 7, 11, \dots\} && (6.19)\\
&= [7]_4 = [-5]_4 = \cdots . && (6.20)
\end{aligned}$$

Equations (6.7), (6.10), (6.14), and (6.18), in conjunction with the Division Algorithm 6.1.1 tell us that $[0]_4$, $[1]_4$, $[2]_4$, $[3]_4$ are the only congruence classes modulo 4.

DEFINITION 6.5.4. The set $\mathbb{Z}_m = \{[0]_m, [1]_m, \ldots, [m-1]_m\}$ is a *complete set of congruence classes modulo m.*

Note that while there is a similarity[†] between the sets $R_4 = \{0, 1, 2, 3\}$ and $\mathbb{Z}_4 = \{[0]_4, [1]_4, [2]_4, [3]_4\}$, they are not equal as sets. While R_4 is a set of integers, $\mathbb{Z}_4$ is a set of sets of integers.

Furthermore, Theorem 6.4.6 tells us (since every integer is congruent to exactly one element of R_4 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. □

These facts are summarized in general in the following theorem.

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.[‡] □

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 $[2+3]_4 = [5]_4$. 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

$$[10]_4 +_4 [-1]_4 = [9]_4 = [5]_4 = [2]_4 +_4 [3]_4.$$

Theorem 6.4.7 and Theorem 6.5.5 can be used to show that, in general, this arithmetic of congruence classes is "well-defined", in the following sense.

THEOREM 6.5.6. *Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ and assume that $a_1 \equiv a_2 \mod m$ and $b_1 \equiv b_2 \mod m$. Then*

(1) $[a_1 + b_1]_m = [a_2 + b_2]_m$,
(2) $[a_1 - b_1]_m = [a_2 - b_2]_m$,
(3) $[a_1 b_1]_m = [a_2 b_2]_m$.

PROOF. Exercise 6.5.2. □

[†] They are actually the "same" in a mathematical sense that one makes precise in an abstract algebra course.

[‡] We are relegating the proof of this important theorem to the exercises, since we will prove a more general result Theorem 7.2.9 (which implies this one) in Section 7.2.

We may thus define the arithmetic operations $+_m$, $-_m$, and $\cdot_m$ on $\mathbb{Z}_m$ as follows.

DEFINITION 6.5.7. Given $a, b \in \mathbb{Z}$,

(1) $[a]_m +_m [b]_m = [a+b]_m$,
(2) $[a]_m -_m [b]_m = [a-b]_m$,
(3) $[a]_m \cdot_m [b]_m = [ab]_m$.

The algebraic structure $\langle \mathbb{Z}_4, +_4 \rangle$ has the following addition table.

$+_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[0]_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[1]_4$	$[1]_4$	$[2]_4$	$[3]_4$	$[0]_4$
$[2]_4$	$[2]_4$	$[3]_4$	$[0]_4$	$[1]_4$
$[3]_4$	$[3]_4$	$[0]_4$	$[1]_4$	$[2]_4$

While we have emphasized that the sets R_m and $\mathbb{Z}_m$ are not the same, we can "rename" elements of $\mathbb{Z}_m$ by the corresponding elements of R_m (the technical concept here is that of a "group isomorphism", which is studied in a course on algebraic structures) and write simply $+$ for $+_m$ to obtain a simpler-looking table.

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

Exercises 6.5

(1) Prove Theorem 6.5.5.
(2) Prove Theorem 6.5.6.

CHAPTER 7

Relations and partitions

Congruence modulo m is an example of a "relation" on $\mathbb{Z}$ (in fact, it is an "equivalence relation"), and we noted in the previous chapter that the congruences classes modulo m of the integers form a "partition" of $\mathbb{Z}$. In this chapter, we formally define and investigate these notions.

7.1. Introduction

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 $(3, b) \notin R$. □

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

$$\begin{aligned} &a\, R\, b \quad \text{if } (a, b) \in R, \text{ and} \\ &a \not R\, b \quad \text{if } (a, b) \notin R. \end{aligned}$$

Using the notation in Definition 7.1.3, we see in Example 7.1.2 that $3\, R\, a$, while $3 \not R\, b$.

While the purpose of the first example is to emphasize that a relation is nothing more than a set of ordered pairs, the second 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 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

$$\begin{aligned} (1,2) &\sim (-4,-8) && \text{because } (1)(-8) = (2)(-4), \text{ and} \\ (-1,3) &\not\sim (2,-5) && \text{because } (-1)(-5) \neq (3)(2). \end{aligned}$$

□

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.,

$$(\forall a \in \operatorname{dom} R)[(\exists b \in Y)[(a,b) \in R] \text{ and}$$
$$[(\forall b, c \in Y)[((a,b) \in R \text{ and } (a,c) \in R) \implies b = c]].$$

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 $R \subseteq X \times Y$.

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) The congruence relation $\equiv \mod m$ ($m \in \mathbb{Z}^+$) 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, 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 certainly 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 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 $2 \not R\, 4$.

□

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 which 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 which are equivalent under this relation into "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 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

$$\mathbb{Z}/{\equiv_m} = \mathbb{Z}_m = \{[0]_m, [1]_m, \ldots, [m-1]_m\}.$$

□

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:

$$(a,b) \sim (c,d) \text{ iff } ad = bc, \quad \text{for all } a, c \in \mathbb{Z},\ b, d \in \mathbb{Z}^*,$$

where $\mathbb{Z}^* = \mathbb{Z} - \{0\}$.

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 $(-9,-18)$.

This equivalence relation is used to rigorously define the rational numbers from the integers. □

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 7.2.3, you are asked to show that $\sim$ is an equivalence relation on $\mathbb{R} \times \mathbb{R}$. Here,

$$\begin{aligned}[(1,2)] &= \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid (x,y) \sim (1,2)\} \\ &= \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = 5\},\end{aligned}$$

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". □

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 7.2.1.

(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 (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 (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) Determine whether the following relations on the given sets are reflexive, symmetric, or transitive.
   (a) The divisibility relation $|$ on $\mathbb{Z}$.
   (b) The relation $\subseteq$ on $\mathcal{P}(\mathbb{Z})$.
   (c) The relation $\sim$ on $\mathbb{R}$ defined by $x \sim y$ iff $xy \geq 0$.

(2) 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 $xy = 1$. Explicitly find the equivalence classes $[3]$, $[-\frac{2}{3}]$, 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 $[(4, 5)]$ and $[(0, 0)]$. Describe

$[(a, b)]$ for any fixed values of a and b in $\mathbb{R}$, both as a set, and geometrically.

(d) 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 3 elements of the equivalence class $[(2, 0)]$ and 3 elements of the equivalence class $[(0, 3)]$.

Note: Since subtraction is not always defined on $\mathbb{Z}^{\geq 0}$ (for example, $2 - 3$ is 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 $\mathbb{Z}^{\geq 0}$.

(3) 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 $\mathbb{R} \times \mathbb{R}$.

(4) 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.

(5) Let $A = \{1, 2, 3\}$. Find a nontrivial (i.e., nonempty) relation on A (i.e., a set of ordered pairs which is a subset of $A \times A$) which is symmetric and transitive, but not reflexive. Why does your relation not contradict the statement in Exercise 7.2.4?

7.3. Partitions

Having used the term "partition" informally, we now give a precise definition.

DEFINITION 7.3.1. Let X be a set, and let Π be a collection of subsets of X (i.e., $\Pi \subseteq \mathcal{P}(X)$). The collection Π is a *partition* of X if

(1) For all $A \in \Pi$, $A \neq \emptyset$.
(2) For all $A, B \in \Pi$, $A = B$ or $A \cap B = \emptyset$.
(3) For all $x \in X$ there exists $A \in \Pi$ such that $x \in A$. (This says that the sets in Π *cover* X).

EXAMPLE 7.3.2.

(1) Let $X = \{1,2,3,4,5\}$. Then $\Pi = \{\{1,4\},\{3\},\{2,5\}\}$ is a partition of X; i.e., every integer in X is in exactly one of the sets $\{1,4\}$, $\{3\}$, or $\{2,5\}$.
(2) The collection $\Pi = \{\mathbb{Z}^+, \mathbb{Z}^-, \{0\}\}$ is a partition of $\mathbb{Z}$; in other words, every integer falls into exactly one category: it is either positive, negative, or zero.

□

As we have already noted informally above, every equivalence relation on a set 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) $\Pi = \big\{\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\} \mid r \in \mathbb{R}^{\geq 0}\big\}$ is a partition of $\mathbb{R} \times \mathbb{R}$.

□

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 Π be a partition of the nonempty set X. Define a relation $\sim$ on X by, for all $a, b \in X$,*

$$a \sim b \iff (\exists A \in \Pi)[a \in A \text{ and } b \in A].$$

Then $\sim$ is an equivalence relation on X. Furthermore, the equivalence classes of $\sim$ are exactly the elements of the partition Π; i.e., $X/{\sim} = \Pi$.

PROOF. Let Π be a partition of the set X and let $\sim$ be defined by, for all $a, b, \in X$,

$$a \sim b \Leftrightarrow (\exists A \in \Pi)[a \in A \text{ and } b \in A].$$

$\sim$ is reflexive: Let $a \in X$. By Definition 7.3.1 (3), since Π covers X, we can fix $A \in \Pi$ 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 $A \in \Pi$ such that $a \in \Pi$ and $b \in \Pi$, 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 $A \in \Pi$ such that $a \in A$ and $b \in A$. Similarly, since $b \sim c$, we can fix $B \in \Pi$ such that $b \in B$ and $c \in B$. Then $b \in A \cap B$, so $A \cap B \neq \emptyset$. Since Π 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 $X/{\sim} = \Pi$.

Let $[a] \in X/{\sim}$. Then $a \in X$, so we can fix $A \in \Pi$ 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 $[a] \in \Pi$ and so $X/{\sim} \subseteq \Pi$.

Next let $A \in \Pi$. 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 $\Pi \subseteq X/{\sim}$. Hence $X/{\sim} = \Pi$ 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) $\{\{1, 2\}, \{3, 4, 5\}\}$
  (b) $\{\{2, 3, 4, 5\}, \{1\}\}$

(2) Let X and Y be sets and $f : X \to Y$ be onto. For all $b \in Y$, let $A_b = f^{-1}[\{b\}]$. Prove that $\{A_b \mid b \in Y\}$ is a partition of X.

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 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.

NOTATION 8.1.1. For $n \in \mathbb{N}$, we denote by $\mathbb{N}_n$ the set

$$\mathbb{N}_n = \{1, 2, \dots, n\} = \{i \in \mathbb{N} \mid i \leq n\}.$$

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 $f : X \underset{\text{onto}}{\overset{1\text{-}1}{\to}} Y$.

(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 exists $f : \{1, 2, \ldots, n\} \underset{\text{onto}}{\overset{1\text{-}1}{\to}} X$.
If $\mathbb{N}_n \approx X$, then we say that the *cardinality* of X is n and write $|X| = n$. We define $|\emptyset| = 0$.

(3) We say X is *infinite* if X is not finite.

Note that $\approx$ is an "equivalence relation"* on the collection of all sets. See Exercise 8.1.5.

While these definitions seem quite sensible and straightforward, it turns out that several "common sense" facts regarding these notions are fairly complicated to prove, 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 $\mathbb{N}_n \approx X$. 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 proof. One additional "common sense" result is given below.

THEOREM 8.1.3. *Let A and B be sets with $A \subseteq B$.*

(1) *If B is finite, then so is A.*
(2) *If A is infinite, then so is B.*

We now consider some examples that illustrate the definitions in Definition 8.1.2. First, let's make the bijections in our original example explicit.

EXAMPLE 8.1.4. Let $X = \{3, -6, \pi, 4.7, \sqrt{2}\}$ and $Y = \{\blacksquare, \clubsuit, \spadesuit, \bigstar, \blacklozenge\}$. The function $f : X \to Y$ defined by

$$f(3) = \blacksquare, \quad f(-6) = \clubsuit, \quad f(\pi) = \spadesuit, \quad f(4.7) = \bigstar, \quad f(\sqrt{2}) = \blacklozenge$$

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 which are equinumerous.

*We put the term equivalence relation in quotation marks since the collection of all sets is not a set.

EXAMPLE 8.1.5. $\mathbb{N} \approx \mathbb{Z}$.

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 $\mathbb{Z}$.

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \dots \\ 0 & 1 & -1 & 2 & -2 & 3 & \dots \end{array}$$

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 the codomain of f is indeed $\mathbb{Z}$, 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 $f(n) \leq 0$ 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 $n \geq 1$, then $f(2n) = \frac{2n}{2} = n$. If $n \leq 0$, then $-2n + 1 \geq 1$ (you should check this) and $f(-2n+1) = \frac{-((-2n+1)-1)}{2} = n$. Hence f is onto. □

Exercises 8.1

(1) The bijection g given in Example 8.1.4 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 8.1.5 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 for all 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 $X \approx Z$.

(6) Let X and Y be sets and assume that $X \approx Y$. Prove directly (from the definitions) that
  (a) If X is finite, then Y is finite and $|X| = |Y|$.
  (b) If X is infinite, then Y is infinite.

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. In particular, we will for now continue to assume that when a nonempty set A is finite, then its cardinality $|A|$ is a unique natural number n.

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 $f : \mathbb{N}_n \xrightarrow[\text{onto}]{1\text{-}1} A$; 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

$$A = \{a_1, a_2, a_3, \ldots, a_n\}.$$

This idea will be useful in the next result, which is one of several useful facts about cardinalities of finite sets.

THEOREM 8.2.1. *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**
$m, n \neq 0$	
$\|A\| = n$, $\|B\| = m$	
$A \cap B = \emptyset$	$\|A \cup B\| = n + m$

We can use Definition 8.1.2(2) to immediately rewrite this as follows.

Given	**Goal**
$m, n \neq 0$	
$A \cap B = \emptyset$	
$f : \mathbb{N}_n \xrightarrow[\text{onto}]{1\text{-}1} A$	
$g : \mathbb{N}_m \xrightarrow[\text{onto}]{1\text{-}1} B$	$h : \mathbb{N}_{n+m} \xrightarrow[\text{onto}]{1\text{-}1} A \cup B$

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.

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 $|B| = m$. Note that if $A = \emptyset$, then $n = 0$ and $A \cup B = B$ is finite, and $|A \cup B| = |B| = m = n + m$. Similarly, if $B = \emptyset$, then $m = 0$ and $A \cup B = A$ is finite, and $|A \cup B| = |A| = n = n + m$.

Next, assume that $A \neq \emptyset$ and $B \neq \emptyset$. By Definition 8.1.2(2), fix $m, n \in \mathbb{N}$ and bijections $f : \mathbb{N}_n \overset{\text{1-1}}{\underset{\text{onto}}{\to}} A$ and $g : \mathbb{N}_m \overset{\text{1-1}}{\underset{\text{onto}}{\to}} B$. We show $|A \cup B| = n + m$ by constructing a bijection $h : \mathbb{N}_{n+m} \overset{\text{1-1}}{\underset{\text{onto}}{\to}} A \cup B$. 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 must show that h is a bijection. First note that h is onto. Let $y \in A \cup B$, so that $y \in A$ or $y \in B$. If $y \in A$, then since f is onto, we can fix $i \in \mathbb{N}_n$ (i.e., $1 \leq i \leq n$) such that $f(i) = y$. Then by definition, $h(i) = f(i) = y$. If $y \in B$, then since g is onto, we can fix $i \in \mathbb{N}_m$ (i.e., $1 \leq i \leq m$) such that $g(i) = y$. Then $n + 1 \leq n + i \leq n + m$, and by definition $h(n + i) = g(n + i - n) = g(i) = y$. Hence h is onto.

Next we show that h is 1-1. Let $i, j \in \mathbb{N}_{n+m}$ and assume that $h(i) = h(j)$. We consider several cases in order to show that $i = j$.

Case I: $1 \leq i \leq n$ and $1 \leq j \leq n$.

Then $h(i) = f(i)$ and $h(j) = f(j)$, and hence $f(i) = f(j)$. Since f is 1-1, we have $i = j$.

Case II: $1 \leq i \leq n$ and $n + 1 \leq j \leq n + m$.

Then $h(i) = f(i)$ and $h(j) = g(j - n)$. Since $f(i) \in A$ and $g(j - n) \in B$, the fact that $h(i) = h(j)$ implies that both $f(i)$ and $g(j - n)$ are in $A \cap B$. But $A \cap B = \emptyset$, which is a contradiction, and hence this case cannot occur.

We leave proofs of the remaining two cases, which are analogous, for Exercise 8.2.1.

□

COROLLARY 8.2.2. *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 $\cup_{i=1}^{k} A_i = A_1 \cup A_2 \cup \cdots \cup A_k$ 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 8.2.2. □

THEOREM 8.2.3. *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. Exercise 8.2.4. □

COROLLARY 8.2.4. *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 8.2.5. □

The next result first appeared as Exercise 4.2.21.

THEOREM 8.2.5. *Let A be a finite set, and let $n \geq 0$ be such that $|A| = n$. Then $\mathcal{P}(A)$ is finite and $|\mathcal{P}(A)| = 2^n$.*

As already mentioned, throughout we have been assuming that the cardinality of a finite set is a well-defined concept. Before moving on to a proof of this fact (and its corollary that $\mathbb{N}$ is an infinite set), we prove Theorem 8.1.3(1), which was stated without proof in Section 8.1, that when $A \subseteq B$ are sets and B is finite, then A is finite.

Scratchwork. We will prove this result by induction on the cardinality of B. More precisely, we will be proving the following statement:

$$(*) \qquad \begin{array}{l} \text{For all } n \geq 0 \text{, for all sets } A \text{ and } B, \\ \text{if } A \subseteq B \text{ and } |B| = n \text{, then } A \text{ is finite.} \end{array}$$

Thinking of the statement of theorem in this way will ensure that the inductive hypothesis is general enough to be useful to us.

PROOF OF THEOREM 8.1.3(1). We prove statement $(*)$ 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).

Inductive Step: Let $n \geq 0$ and assume that for all sets X and Y, if $X \subseteq Y$ and $|Y| = n$, then X is finite (this is our inductive hypothesis).

Next, let A and B be sets with $A \subseteq B$ and $|B| = n + 1$. We must prove that A is finite.

Since $|B| = n + 1$, we may fix a bijection $f : \mathbb{N}_{n+1} \underset{\text{onto}}{\overset{\text{1-1}}{\to}} B$; i.e., $B = \{b_1, b_2, \ldots, b_n, b_{n+1}\}$, where $f(i) = b_i$ for all $1 \leq i \leq n+1$. We consider two cases.

Case I: $A \subseteq \{b_1, b_2, \ldots, b_n\}$.

Then $Y = \{b_1, b_2, \ldots, b_n\}$ is a finite set with $|Y| = n$ (be sure you see why this is true), and hence by the inductive hypothesis (with $X = A$), A is finite.

Case II: Otherwise.

Then $b_{n+1} \in A$. Note that $A = (A - \{b_{n+1}\}) \cup \{b_{n+1}\}$ and $(A - \{b_{n+1}\}) \cap \{b_{n+1}\} = \emptyset$. The set $\{b_{n+1}\}$ is finite, since it has one element. Also note that, since $A \subseteq B$, the set $X = A - \{b_{n+1}\}$ is a subset of $Y = \{b_1, b_2, \ldots, b_n\}$, which has cardinality n. Thus, again by the inductive hypothesis, $X = A - \{b_{n+1}\}$ is a finite set. Thus A is finite by Theorem 8.2.1.

It follows by induction that statement $(*)$ is true, and hence any subset of a finite set is finite.

□

8.2.1. Debts paid: The Pigeonhole Principle*. There is no reason to believe (without proof) that the cardinality of a nonempty finite set is a unique natural number. 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 $n \neq m$. If we can prove the following special case of this statement, then our desired result will follow.

THEOREM 8.2.6. *For all $n, m \in \mathbb{N}$, if $n > m$, then there does not exist a 1-1 function $f : \mathbb{N}_n \to \mathbb{N}_m$.*

We can rephrase Theorem 8.2.6 as

for all $n, m \in \mathbb{N}$, if $n > m$ then

for any function $f : \mathbb{N}_n \to \mathbb{N}_m$, f is not 1-1,

which is known as the "Pigeonhole Principle". In more colorful language:

> If $n > m$ and n pigeons are put into m pigeon holes, then at least one pigeonhole contains more than one pigeon.

COROLLARY 8.2.7. *The set $\mathbb{N}$ of natural numbers is infinite.*

Exercises 8.2

(1) Complete the proof of Theorem 8.2.1.

(2) Prove Corollary 8.2.2.

(3) Let A be a finite set and B be an infinite set. Prove that $B - A$ is infinite.

(4) Prove Theorem 8.2.3.

(5) Prove Corollary 8.2.4.

(6) Prove that if A is a finite set and $f : A \to B$, then the image $f[A]$ is also finite.

(7) Let A and B be finite sets with $|A| = |B|$, and let $f : A \to B$. Prove that f is 1-1 iff f is onto.

8.3. Infinite sets

In this section, we specifically study infinite sets. In Example 8.1.5, we showed that $\mathbb{N} \approx \mathbb{Z}$, or that $\mathbb{Z}$ is equinumerous with a proper subset of itself. This is true in general for infinite sets and in fact is a characterization of infinite sets (i.e., the notions are equivalent).

THEOREM 8.3.1 (Dedekind). *A set X is infinite if and only if it is equinumerous with a proper subset of itself.*

The fact that $\mathbb{N} \approx \mathbb{Z}$ says (informally) that the elements of $\mathbb{Z}$ can be "listed", or "enumerated" without repetition. Recall the picture from Example 8.1.5, which shows this enumeration of $\mathbb{Z}$:

$$0, 1, -1, 2, -2, 3, \dots .$$

DEFINITION 8.3.2.

(1) The set X is *denumerable (enumerable)* if there is is 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 8.2.7), every denumerable set is infinite (by Exercise 8.1.6b).

The use of the word "countable" to describe a set as 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.

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 **[6]**). In some cases, we will provide an informal sketch of the proof as well as the formal proof. Students who are new to these ideas may wish to omit the formal proofs at the end of the section 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.

Theorem 8.3.3. *Let A be a 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 (informal sketch). Assume that $g : \mathbb{N} \underset{\text{onto}}{\rightarrow} A$. 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 $f : \mathbb{N} \overset{\text{1-1}}{\underset{\text{onto}}{\rightarrow}} A$.

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" containing all the elements of A.

We can visualize the proof as follows:

Enumeration of A	Repeats	New enumeration of A
with repeats		
$g(1)$		$f(1)$
$g(2)$	$= g(1)$	
$g(3)$	$= g(1)$	
$g(4)$		$f(2)$
$g(5)$		$f(3)$
$g(6)$	$= g(4)$	
$g(7)$		$f(4)$
$g(8)$	$= g(5)$	
$\vdots$	$\vdots$	$\vdots$

□

Proof. See the end of this section. □

Theorem 8.3.4. *Let A and B be denumerable sets. Then $A \cup B$ is also denumerable.*

Proof (informal sketch). Let A and B be denumerable, and fix bijections $f : \mathbb{N} \overset{\text{1-1}}{\underset{\text{onto}}{\rightarrow}} A$ and $g : \mathbb{N} \overset{\text{1-1}}{\underset{\text{onto}}{\rightarrow}} B$. 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:

$$b_1, b_2, b_3, b_4, \ldots.$$

Informally, we construct a list of the elements of $A \cup B$ by "dovetailing" the enumerations of A and B:

$$a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4, \ldots.$$

This shows that $A \cup B$ is denumerable by Theorem 8.3.3, since every element of $A \cup B$ (which is infinite) appears (at least once) on this list.

Formally, we 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 must 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 8.3.3 that $A \cup B$ is countable. Since $A \subseteq A \cup B$ and A is infinite, $A \cup B$ is also infinite by Theorem 8.1.3(2). 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 8.3.5. *The union of finitely many denumerable sets is denumerable.*

PROOF. By induction on the number of sets. □

COROLLARY 8.3.6. *The union of finitely many countable sets is countable.*

Just these few tools are enough to prove that $\mathbb{Q}$ is denumerable.

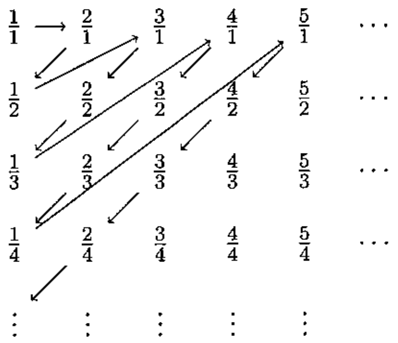
THEOREM 8.3.7 (Cantor, 1874). *The set $\mathbb{Q}$ of rational numbers is denumerable.*

PROOF (INFORMAL SKETCH). Note that

$$\begin{aligned} \mathbb{Q} &= \left\{\frac{a}{b} \mid a, b \in \mathbb{Z} \text{ and } b > 0\right\} \\ &= \mathbb{Q}^+ \cup \mathbb{Q}^- \cup \{0\}, \end{aligned}$$

where $\mathbb{Q}^+ = \{\frac{a}{b} \mid a, b \in \mathbb{Z}^+\}$ and $\mathbb{Q}^- = \{-\frac{a}{b} \mid a, b \in \mathbb{Z}^+\}$.

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), and etc. We now dovetail the enumerations.



$$\begin{array}{cccccc}
\frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \frac{4}{1} & \frac{5}{1} & \cdots \\
\frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \frac{4}{2} & \frac{5}{2} & \cdots \\
\frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{4}{3} & \frac{5}{3} & \cdots \\
\frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \frac{4}{4} & \frac{5}{4} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots &
\end{array}$$

We claim that every positive rational number is on this list (which has repeats); i.e., we can write down a function $f : \mathbb{N} \underset{\text{onto}}{\to} \mathbb{Q}^+$, as shown. So $\mathbb{Q}^+$ is countable, by Theorem 8.3.3, and hence denumerable, since $\mathbb{Q}^+$ is infinite.

We next note that $\mathbb{Q}^+$ is equinumerous with $\mathbb{Q}^-$; i.e., there is a bijection $g : \mathbb{Q}^+ \overset{\text{1-1}}{\underset{\text{onto}}{\to}} \mathbb{Q}^-$. Since $\approx$ is transitive, $\mathbb{Q}^-$ is also denumerable.

It follows that $\{0\} \cup \mathbb{Q}^+ \cup \mathbb{Q}^-$ is denumerable, by Corollary 8.3.5. □

A similar argument will show that the cartesian product $A \times B$ is denumerable when A and B are denumerable sets; for example, $\mathbb{N} \times \mathbb{N}$ is denumerable. The method used to show that $\mathbb{Q}^+$ is denumerable is called "Cantor's first diagonalization argument".

Certainly any infinite subset of a denumerable set is denumerable; i.e., if we can list a the elements of a set, then we can list the elements of any subset of that set. This is the final result we need in order to prove that the set of real numbers is uncountable.

THEOREM 8.3.8. *Let A be denumerable and $B \subseteq A$ 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 enumerable. 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. Similar to the proof of Theorem 8.3.4. □

COROLLARY 8.3.9. *Let A be countable and $B \subseteq A$. Then B is countable.*

We are now ready to show that the set $\mathbb{R}$ of real numbers is uncountable.

THEOREM 8.3.10 (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; we must first give some necessary background information about real numbers, which we present 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 $0.d_1d_2d_3d_4\ldots$ 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 the contrapositive of Theorem 8.3.8.

Assume for the sake of a contradiction that $[0, 1]$ is countable. Since $[0, 1]$ is infinite (can you explain why?), it is denumerable. Hence we may fix a function $f : \mathbb{N} \xrightarrow[\text{onto}]{\text{1-1}} [0, 1]$ which 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:

$$\begin{aligned} f(1) &= 0.d_1^1 d_2^1 d_3^1 d_4^1 d_5^1 \dots \\ f(2) &= 0.d_1^2 d_2^2 d_3^2 d_4^2 d_5^2 \dots \\ f(3) &= 0.d_1^3 d_2^3 d_3^3 d_4^3 d_5^3 \dots \\ f(4) &= 0.d_1^4 d_2^4 d_3^4 d_4^4 d_5^4 \dots \\ f(5) &= 0.d_1^5 d_2^5 d_3^5 d_4^5 d_5^5 \dots \\ &\vdots \end{aligned}$$

We obtain a contradiction by constructing a real number

$$r = 0.r_1 r_2 r_3 r_4 r_5 \dots$$

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 $f : \mathbb{N} \underset{\text{onto}}{\overset{1\text{-}1}{\rightarrow}} [0,1]$ and $r \in [0,1]$, $r \in \operatorname{ran} f$. Thus we can fix $k \in \mathbb{Z}^+$ such that

$$r = f(k) = 0.d_1^k d_2^k d_3^k d_4^k d_5^k \dots d_k^k \dots .$$

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, $r \neq f(k)$, a contradiction.

Thus $[0,1]$, and hence $\mathbb{R}$, is uncountable. □

The method used to show that $[0,1]$ is uncountable is called "Cantor's second diagonalization argument".

One important consequence of the fact that $\mathbb{R}$ is uncountable is that any *list* of real numbers

$$x_1, x_2, x_3, \dots$$

cannot include *all* real numbers. In other words, no proof of a statement of the form

For all $x \in \mathbb{R} \dots$

can begin with the statement

Let $x_1, x_2, x_3, \dots$ be a list of all real numbers.

We end this section with the proof of Theorem 8.3.3, which was sketched informally above.

PROOF OF THEOREM 8.3.3. 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 $f : \mathbb{N} \underset{\text{onto}}{\overset{1\text{-}1}{\to}} A$.

We define $f : \mathbb{N} \underset{\text{onto}}{\overset{1\text{-}1}{\to}} A$ by recursion; the idea here is to make f 1-1 "as we go" and to 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). 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 $i \leq n$.

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 (make sure you can explain why!).

To see that f is onto, let $y \in A$ and show that $y \in \operatorname{ran} f$. Since g is onto, we may fix $m \in \mathbb{Z}^+$ such that $g(m) = y$. We claim that $y \in \{f(1), \ldots, f(m+1)\}$, and hence $y \in \operatorname{ran} f$. □

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 Theorem 8.3.1 to prove that $\mathbb{N}$ is infinite.

(3) Prove that the union of a finite set and a denumerable set is denumerable.

(4) Prove that the union of a finite set and a countable set is countable.

(5) Prove that the union of finitely many denumerable sets is denumerable.

(6) Prove that the union of finitely many countable sets is countable.

(7) Prove that the union of denumerably many denumerable sets is denumerable.

(8) Prove that a set A is countable iff there exists a 1-1 function $f : A \to \mathbb{N}$.

(9) Assume that A is uncountable and B is a countable subset of A. Prove that $A - B$ is uncountable.

(10) 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).

(11) Prove that $\mathcal{P}(\mathbb{Z}^+)$ is uncountable.

APPENDIX A

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 for students when they are first beginning. The following guidelines for writing mathematics point out some of the issues you'll want to keep in mind.

Guidelines for writing mathematics*

(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.

*These guidelines were originally inspired by a "Writing checklist" by Dr. Annalisa Crannell of Franklin and Marshall College. I have adapted them over time in the various "proof courses" I have taught in my career at Allegheny College.

(4) "There is no room in mathematics for ambiguity."[†] 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 scratch work 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.

[†] From Ethan D. Block's *Proofs and Fundamentals: A First Course in Abstract Mathematics*, Birkhäuser, Boston, 2000, p. 94.

Bibliography

[1] Robert G. Bartle and Donald R. Sherbert, *Introduction to real analysis,* 3rd edition, John Wiley & Sons, Inc., 2000.

[2] Robert J. Bond and William J. Keane, *An introduction to abstract mathematics,* Brooks/Cole, 1999.

[3] David M. Burton *Elementary number theory,* 5th edition, McGraw-Hill, 2002.

[4] Keith Devlin, *Sets, functions, and logic: an introduction to abstract mathematics,* 2nd edition, Chapman & Hall Mathematics, 1992.

[5] Peter J. Eccles, *An introduction to mathematical reasoning: numbers, sets and functions,* Cambridge University Press, 1997.

[6] Herbert B. Enderton, *Elements of set theory,* Academic Press, Inc., 1977.

[7] Russell A. Gordon, *Real analysis: a first course,* 2nd edition, Pearson Education, Inc., 2002.

[8] Daniel J. Velleman, *How to prove it: a structured approach,* Cambridge University Press, 1994.

Index