MATH 267: Introduction to Abstract Math

Angel V. Kumchev

Spring 2026

Contents

1 Theorems, Definitions, Proofs...

Mathematics is about solving problems rigorously. Since its historical development was intertwined with that of philosophy, the edifice of mathematics is heavily influenced by the philosophical struggle to build a perfect society. Thus, the solution of every problem is a formal true "statement," called a *theorem*, for which we have supplied a "rigorous proof." Here are some examples of theorems.[1]

Theorem 1.1. *$\sqrt{2}$ is irrational.*

Theorem 1.2. *If n is an integer, then $\sqrt[3]{n}$ is irrational.*

Theorem 1.3. *Every integer $n \geq 2$ can be expressed as a product of primes.*

Theorem 1.4. *There are infinitely many primes.*

Theorem 1.5. *The sum of two rational numbers is a rational number.*

Theorem 1.6. *The sum of two irrational numbers is an irrational number.*

Theorem 1.7. *If $f : (0, \infty) \to \mathbb{R}$ is increasing and $f(x) \leq 1$ for all $x > 0$, then the curve $y = f(x)$ has a horizontal asymptote as $x \to \infty$.*

Theorem 1.8. *If A and B are 3×3 matrices, then $\det(AB) = (\det A)(\det B)$.*

Theorem 1.9. $1 - \dfrac{1}{2} + \dfrac{1}{3} - \dfrac{1}{4} + \cdots = \ln 2.$

Theorem 1.10. $1 + \dfrac{1}{2^2} + \dfrac{1}{3^2} + \dfrac{1}{4^2} + \cdots = \dfrac{\pi^2}{6}.$

Theorem 1.11. $\displaystyle\int \frac{dx}{x^2 - 1} = \frac{1}{2} \ln \left| \frac{x-1}{x+1} \right| + C.$

The above list aims to make two points:

- Depending on context, a fact may or may not be considered a theorem: Theorem 1.11 is a prime example. However, at least formally, almost every mathematical fact is a theorem. Consequently, almost every mathematical fact has a proof, even though its proof may not fit our idea of what a proof looks like (again, think of the proof of Theorem 1.11).
- In order to understand the statement of a theorem, we need to have a clear understanding of all the terms in its statement. This is very important in rigorous mathematics, so we will focus heavily in this course on learning to respect and utilize formal definitions.

[1] Though two of these are, in fact, *not* theorems. Can you tell which two?

As one starts to focus on formal definitions and rigorous proofs, one needs to follow some fundamental rules:

- Every mathematical term must be *defined*. Since this process must start somewhere, there are a few agreed-upon exceptions to this rule—they are called *undefined terms* (or *primitive terms*, depending on the author).
- Once an English word has been defined as a mathematical term, its lay definition becomes irrelevant. It is therefore extremely important to know the precise definitions of mathematical terminology and not to rely on everyday English to infer mathematical meaning.
- Likewise, sometimes different branches of mathematics may use the same word to refer to different notions (just as an English word may have multiple meanings) or may have different definitions for the same term (see Exercise 1.14 below). One must be aware of the definition in one's current context, as this is the correct definition one must work with.
- Every mathematical fact must be *proved*. That means using "formal logic" to deduce the new fact from known facts. Since the process must originate with *some* fact, we must agree to accept a few basic facts as true without proof—these are the *axioms*. New mathematical facts that we prove are called *theorems*. To keep things organized, we distinguish several types of theorems: *lemmas, propositions, corollaries*; occasionally, you may encounter *examples, facts, etc.* as well as special names for yet-unproven claims such as *conjectures*.
- Once we agree on some axioms, a theorem is true if it has been deduced from those axioms by valid formal logic. Note that if two branches of mathematics are based on different sets of axioms, it is possible for them to obtain two theorems that appear at odds with each other, yet each theorem is true within its own sets of axioms. The most famous example of this phenomenon has to do with the so-called *Euclid postulate* in geometry. In particular, the rules of formal logic that we use to make deductions are themselves a type of axioms that we accept across all of mathematics. It is imperative to understand those clearly; otherwise, one finds oneself using different axioms from everyone else and giving a "proof" that is not really such.

Note that when we say that we must define every mathematical "term" we use, that includes also every piece of mathematical notation—after all, notation is merely a clever replacement for terminology. That means that whenever one encounters a piece of notation that they are unfamiliar with, they should search the text they are reading (or their prior mathematical background) for the definition of that piece of notation. In the same vein, one should make always sure that whenever a mathematical text uses a piece of terminology that one is familiar with, that term is used with the definition that one knows.

Example 1.12. In mathematics, the letter $\mathbb{N}$ is commonly used to denote the *set of natural numbers*. You probably have heard of "natural numbers"—but can you define them? Complete the definition:

The set $\mathbb{N}$ of natural numbers is…

You may be surprised to find out that depending on which mathematical textbook you consult, you can find two slightly different definitions. They both work equally well in general, and in different situations each has advantages over the other. In this class, we will agree that the first natural number is 0. □

Example 1.13. Here is the definition of the concept of divisibility of one integer by another.

Definition 1.1. Let a and b be integers. We say that b *divides* a if $a = bq$ for some integer q. When b divides a, we write $b \mid a$; otherwise, we write $b \nmid a$.

For example,

$$2 \mid 6, \quad 3 \mid 6, \quad 4 \nmid 6, \quad 0 \mid 0, \quad 0 \nmid 5, \quad 5 \mid 0.$$

Give three additional examples of pairs a, b such that $b \mid a$ and two additional examples of pairs a, b such that $b \nmid a$. Your examples should involve only integers a and b that exceed 10. □

Exercise 1.14. Here is another definition.[2]

Definition 1.2. An integer $n > 1$ is called *reducible* if we can express it in the form $n = ab$, where a and b are integers and $1 < a, b < n$; if $n > 1$ is not reducible, n is called *irreducible*.

For example, 6 is reducible because $6 = 2 \cdot 3$ and $1 < 2, 3 < 6$. On the other hand, 3 is irreducible: if $3 = ab$, where a and b are integers and $1 < a, b < 3$, then we must have $a = b = 2$; however, $2 \cdot 2 \neq 3$.

Give three additional examples of reducible integers and two additional examples of irreducible integers. Explain why your examples work. □

[2]You may know Definition 1.2 as the definition of what it means for n to be a prime or a composite number. In this course, we will give a different definitions of those terms, and we will call the above concept (ir)reducibility.

2 Logical Statements

Definition 2.1. A *statement* is a sentence that is true or false, but not both.

Exercise 2.1. Below are some examples and non-examples of statements. Classify them; if a sentence is a statement, determine whether it is true or false.

(a) $2 + 2 = 5$

(b) $2 + 3 \leq 5$

(c) 7 is a prime number.

(d) $x + 6 = 0$

(e) All prime numbers are odd.

(f) 2 is an odd prime number.

(g) This statement is false.

Many statements in mathematics are compound, so it is important to define rigorously how they are built from basic statements. To that end, we introduce the concepts of:

- a *logical variable*: a variable (e.g., P, Q, R, S) that can be replaced by any logical statement; it will then take a value of either T (for "true") or F (for "false");
- a *statement form*: an expression that involves logical variables and logical connectives that turns into a statement when all the logical variables are replaced by logical statements; the statement form then takes on one of the values T or F.

Definition 2.2. If P is a statement, its *negation*, denoted $\neg P$, is the statement that is true/false if P is false/true:

P	$\neg P$
T	F
F	T

Definition 2.3. If P and Q are two statements, their *disjunction*, denoted $P \vee Q$, is the statement that is false only if P and Q are both false:

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

Definition 2.4. If P and Q are two statements, their *conjunction*, denoted $P \wedge Q$, is the statement that is true only if P and Q are both true:

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

Definition 2.5. If P and Q are two statements, the *implication* (or *conditional statement*) $P \to Q$ is the statement that is false only if P is true and Q is false:

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

Definition 2.6. If P and Q are two statements, their *equivalence* (or *biconditional statement*), denoted $P \leftrightarrow Q$, is the statement that is true/false if P and Q have the same/opposite true values:

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

A truth table, similar to those that appear in Definitions 2.2–2.6 above, can be constructed for any statement form. Two statement forms F and G (that involve the same truth variables) are called *equivalent* if the statement form $F \leftrightarrow G$ is a *tautology*: i.e., it is true for all possible values of the logical variables that appear in the two forms. Equivalently, the two forms are equivalent if their truth tables are the same. Truth tables are often used to verify that certain logical forms are equivalent.

Example 2.2. We claim that

$$\neg(P \to Q) \leftrightarrow (P \wedge \neg Q). \qquad (*)$$

One way to verify $(*)$ is to construct a joint truth table of the negation $\neg(P \to Q)$ and of the conjunction $P \wedge \neg Q$:

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

Since the columns for $\neg(P \to Q)$ and $P \wedge \neg Q$ are identical, the equivalence $(*)$ is satisfied.

Theorem 2.3. *If P, Q are statements, then:*

(1) $\neg(\neg P) \leftrightarrow P$;

(2) $\neg(P \wedge Q) \leftrightarrow (\neg P \vee \neg Q)$;

(3) $\neg(P \vee Q) \leftrightarrow (\neg P \wedge \neg Q)$;

(4) $\neg(P \to Q) \leftrightarrow (P \wedge \neg Q)$.

Proof. The proof of part (4) was given in Example 2.2. The other parts can be proved similarly using truth tables. □

Exercise 2.4. For each given statement: determine whether the statement is true or false; state its negation.

(a) If $5 \mid 60$, then $5 \mid 6$ or $5 \mid 10$.
(b) If $5 \mid 99$, then $5 \mid 9$ or $5 \mid 11$.
(c) If $8 \mid 24$, then $8 \mid 6$ or $8 \mid 4$.
(d) If $8 \mid 48$, then $8 \mid 24$ or $8 \mid 2$.

2.1 Sets

Most of the time, mathematical statements involve variables, so they are not logical statements. However, those variables are seldom "free." It is common to restrict one's variables to the elements of some "set": a collection of objects that have some common defining characteristic; in this way, one may combine logic and variables. However, before we can discuss that idea, we need to have a brief discussion of sets.

The basic notions of a set S and an element s of a set are introduced in Chapter 4 of the textbook; the brief version of that introduction is that *set* and *element* are primary terms, explained only intuitively. We do introduce formal notation to indicate whether a given object s is an element of a given set S: namely, we write $s \in S$ when s is among the elements of S and $s \notin S$ otherwise.

When it comes to describing sets, there are two ways. First, we may list all the elements:

$$A = \{a, b, c\}, \quad B = \{1, 2, 3, \dots, 2026\}, \quad \mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}.$$

However, for more complicated sets, this approach fails, and we are forced to describe a set in the form

$$S = \{x \in X : p(x)\},$$

where X is some existing set and $p(x)$ is a "property" of x, which turns into a statement when x is replaced by an element of X. For example,

$$(0, 2] = \{x \in \mathbb{R} : 0 < x \leq 2\}$$

is such a description of the semi-open interval $(0, 2]$: the property $p(x)$ in this case is the double inequality $0 < x \leq 2$; for any real x, this inequality is either true, or false. For example, since $0 < 1.33 \leq 2$, we have $1.33 \in (0, 2]$; however, $-1 \notin (0, 2]$ and $\pi \notin (0, 2]$.

Notation

This is a good place to introduce some standard notation.

$\mathbb{N}$ is the set of *natural numbers*, $\mathbb{N} = \{0, 1, 2, 3, \dots\}$.
$\mathbb{Z}$ is the set of *integers*, $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$.
$\mathbb{Q}$ is the set of *rational numbers*, $\mathbb{Q} = \{a/b : a, b \in \mathbb{Z}, b \neq 0\}$.
$\mathbb{R}$ is the set of *real numbers*. (We won't give a more explicit description here.)
$\mathbb{C}$ is the set of *complex numbers*, $\mathbb{C} = \{a + \mathrm{i}b : a, b \in \mathbb{R}\}$; here, $\mathrm{i} = \sqrt{-1}$.
If $n \geq 2$ is an integer, $\mathbb{R}^n$ is the set of n-dimensional (row) vectors,

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) : x_1, x_2, \dots, x_n \in \mathbb{R}\}.$$

If a and b are real and $a < b$, then we define the *intervals*

$$(a, b) = \{x \in \mathbb{R} : a < x < b\}, \qquad (a, b] = \{x \in \mathbb{R} : a < x \leq b\},$$
$$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}, \qquad [a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

2.2 Variables and Quantifiers

We now return to the discussion of mathematical sentences involving variables. We assume that there is some "universe" X that we confine our discussion to: that is, there is some set X and we only consider objects $x \in X$.

Definition 2.7. Let A be a set and $p(x)$ be a property that becomes a statement when x is replaced by any element $a \in A$. If there is no element $a \in A$ such that $p(a)$ is false, we say that $p(x)$ *is true for all* $x \in A$ and write

$$\forall x \in A,\ p(x);$$

this is a *universally quantified* statement. If there is some element $a \in A$ such that $p(a)$ is true, we say that $p(x)$ *is true for some* $x \in A$ and write

$$\exists x \in A : p(x);$$

this is an *existentially quantified* statement.

Exercise 2.5. For each of the following quantified statements, determine whether it is true or false:

(a) $\forall x \in \mathbb{R},\ x^2 + 1 = 0$;
(b) $\forall x \in \mathbb{R},\ x^2 + 1 > 0$;
(c) $\exists x \in \mathbb{R} : x^2 + 1 = 0$;
(d) $\exists x \in \mathbb{R} : x^2 - 2 = 0$;
(e) $\exists x \in \mathbb{Q} : x^2 - 2 = 0$;
(f) $\forall y \in \mathbb{Z}, \exists x \in \mathbb{Z} : x + y = 3x$;
(g) $\forall y \in \mathbb{Q}, \exists x \in \mathbb{Q} : x + y = 3x$;
(h) $\exists x \in \mathbb{Q} : \forall y \in \mathbb{Z}, xy = 2x + 3y - 6$.

Exercise 2.6. The following statements are, in fact, quantified statements in disguise. Rewrite them formally using quantifiers.

(a) If n is an integer, then $\sqrt[3]{n}$ is irrational.

(b) If p is prime, then $\sqrt{p}$ is irrational.

(c) Every integer $n \geq 2$ can be expressed as a product of primes.

(d) The sum of two rational numbers is a rational number.

(e) A composite integer n is divisible by some prime $p \leq \sqrt{n}$.

It is important to learn how to negate quantified statements. The basic rule is that when we negate a quantified statement, the domain of the quantified variable remains the same, the quantifier changes type ($\forall$ becomes $\exists$ and vice versa), and the condition gets negated. In other words,

$$\neg(\forall x \in A,\ p(x)) \quad \leftrightarrow \quad \exists x \in A : \neg p(x),$$

and

$$\neg(\exists x \in A : p(x)) \quad \leftrightarrow \quad \forall x \in A,\ \neg p(x).$$

Example 2.7. Let us negate the following statement:

> If 6 divides the product of two integers, then 6 must divide at least one of them.

Note that this is really an universal statement: we implicitly assume that the statement is true for the product of *any* product of two integers. Thus, a formal restatement is

$$\forall a \in \mathbb{Z}, \forall b \in \mathbb{Z},\ 6 \mid ab \to (6 \mid a \vee 6 \mid b).$$

The negation of this statement is

$$\exists a \in \mathbb{Z} : \neg(\forall b \in \mathbb{Z}, 6 \mid ab \to (6 \mid a \vee 6 \mid b)),$$

or equivalently:

$$\exists a \in \mathbb{Z} : \exists b \in \mathbb{Z} : \neg(6 \mid ab \to (6 \mid a \vee 6 \mid b)).$$

The latter, in turn, is equivalent to

$$\exists a \in \mathbb{Z} : \exists b \in \mathbb{Z} : 6 \mid ab \wedge \neg(6 \mid a \vee 6 \mid b),$$

and further to

$$\exists a \in \mathbb{Z} : \exists b \in \mathbb{Z} : 6 \mid ab \wedge 6 \nmid a \wedge 6 \nmid b.$$

We can now restate the negation of the given statement in plain English:

> There is some pair of integers such that 6 divides their product but does not divide either of them.

□

Exercise 2.8. Write the given statement formally using quantifiers, negate the formal statement, and then write the negation in plain English.

(a) For every $\varepsilon > 0$, there is some $\delta > 0$ such that $|f(x) - L| < \varepsilon$ whenever $|x - a| < \delta$.

(b) For every $\varepsilon > 0$, there is some $\delta > 0$ such that $|f(x)| > \varepsilon^{-1}$ whenever $|x - a| < \delta$.

(c) For every $\varepsilon > 0$, there is some $\delta > 0$ such that $|f(x) - L| < \varepsilon$ whenever $|x| > \delta^{-1}$.

(d) For every $\varepsilon > 0$ and $\delta > 0$, one has $|f(x) - L| < \varepsilon$ for some x such that $0 < |x - a| < \delta$.

(e) For every $\varepsilon > 0$ and $\delta > 0$, one has $|f(x) - L| < \varepsilon$ for some x such that $|x| > \delta^{-1}$.

(f) For every $\varepsilon > 0$ and $\delta > 0$, one can find an x such that $|x| > \delta$ and $|f(x)| > \varepsilon^{-1}$.

(g) The integer n is not divisible by the square of any prime.

(h) The integer n is not divisible by any prime $p < z$.

(i) If a prime p divides n, then n is divisible also by p^2.

(j) If a prime divides n, then n is divisible also by its square.

Recall Definition 1.2. It may appear familiar: this is probably how you think of prime numbers. At least, that is how they are typically introduced in elementary and middle school math (though somewhat less formally). Here, we will give a different definition of a prime number.

Definition 2.8. An integer $p > 1$ is called *prime*, if it has the following property:

$$\forall a \in \mathbb{Z}, \forall b \in \mathbb{Z}, (p \mid ab \to p \mid a \text{ or } p \mid b). \tag{P}$$

An integer $n > 1$ that is not prime is called *composite*.

Exercise 2.9. Use the above definition to show that 6 is composite.

Example 2.10. Let us show that 2 is prime.

To show that 2 has property (P), suppose that a and b are two arbitrary integers. We must check that

$$2 \mid ab \to (2 \mid a \vee 2 \mid b).$$

This implication is vacuously true when $2 \nmid ab$, so let us suppose that a and b are integers such that $2 \mid ab$; that is, ab is even. We must then show that a or b is even. The only way for this to fail is if a and b are both odd. But the product of two odd numbers is odd.[1] Since ab is assumed even, at least one of a or b must also be even. □

Note that if $p = ab$ with $1 < a, b < p$ (that is, if p fails to be irreducible), then p fails property (P) above (the proof generalizes the solution of Exercise 2.9). Therefore, every prime p (in the sense of our definition) is also irreducible. It is not as clear whether there are some irreducible integers that are composite (that is, they fail property (P)). It turns out that there are no such integers, and so being prime is equivalent to being irreducible, but we will not discuss that here. To put it simply: for us primality and compositeness are defined by property (P).

[1] We will prove this later; for the time being, we may use this "known fact."

3 Simplest Proofs

Most theorems are either universal or existential statements. In this lecture, we will discuss how to approach the proofs of such statements, and we will see some basic examples of such proofs.

Consider the existential statement

$$\exists x \in X : P(x),$$

where $P(x)$ is a property whose truth depends on the value of $x \in X$. In its simplest form, the proof of this existential statement proceeds as follows:

> "Let $x = a$ (where $a \in X$ is specific). Then $P(a)$ is true, because … QED"

Here is an example.

Example 3.1. There is a complex number w such that $(2 + \mathrm{i})w = 1$.

Let $w = 0.4 - 0.2\mathrm{i}$. Then

$$(2 + \mathrm{i})w = (2 + \mathrm{i})\,(0.4 - 0.2\mathrm{i}) = (0.8 - (-0.2)) + (-0.4 + 0.4)\,\mathrm{i} = 1 + 0\mathrm{i} = 1. \qquad \square$$

Note that in this "proof," we had figured in advance what choice of w would make the statement $(2+\mathrm{i})w = 1$ true, and then we simply checked that this identity does hold for our chosen w. The final product is neat but counterintuitive, because it hides the analysis that led us to select this particular w. In more complex situations, one usually tries to explain the source of such choices: see Example 3.3 below.

Next, consider the universal statement

$$\forall x \in X, P(x).$$

In its simplest form, the proof of this universal statement proceeds as follows:

> "Let $x \in X$. Then … So, $P(x)$ must be true."

The derivations of many identities fit this model. Here is an example.

Example 3.2. For $x \in \mathbb{R}$, one has

$$\sin(3x) = 3\sin x - 4\sin^3 x.$$

We really claim that the following universal statement is true:

$$\forall x \in \mathbb{R},\ \sin(3x) = 3\sin x - 4\sin^3 x.$$

Thus, we will follow the above paradigm.

Let $x \in \mathbb{R}$. Then

$$\begin{aligned}\sin(3x) &= \sin(x+2x) = \sin x \cos(2x) + \cos x \sin(2x) \\ &= \sin x \left(1 - 2\sin^2 x\right) + \cos x (2 \sin x \cos x) \\ &= \sin x - 2\sin^3 x + 2 \sin x \cos^2 x \\ &= \sin x - 2\sin^3 x + 2 \sin x \left(1 - \sin^2 x\right) \\ &= 3\sin x - 4\sin^3 x.\end{aligned}$$

Note that in the last example, we assumed that we already know the trig identities for $\sin(x+y)$, $\sin(2x)$, and $\cos(2x)$. Had we assumed familiarity only with the identity

$$\sin(x+y) = \sin x \cos y + \cos x \sin y,$$

our proof would have been longer, since we would need to use this formula to derive also expressions for $\sin(2x)$ and $\cos(2x)$.

Example 3.3. Let $z \in \mathbb{C}$ be non-real. Then there is a $w \in \mathbb{C}$ such that $zw = 1$.

Here, we are interested in the following formal statement (that includes multiple quantifiers):

$$\forall z \in \mathbb{C}, z \notin \mathbb{R} \to (\exists w \in \mathbb{C} : zw = 1).$$

Thus, we will first select an arbitrary complex number z that is non-real (the hypothesis of the implication is true), and then figure a choice of w that establishes the truth of the existential statement (for that chosen z).

Let $z = a + b\mathrm{i}$, with $a, b \in \mathbb{R}$ and $b \neq 0$. We will show that

$$w = \frac{1}{a^2+b^2}(a - b\mathrm{i}) = \frac{a}{a^2+b^2} + \frac{-b}{a^2+b^2}\,\mathrm{i}$$

has the desired property. First, we note that w is well-defined, because

$$a^2 + b^2 \geq 0 + b^2 = b^2 > 0.$$

Further,

$$\begin{aligned}zw &= (a+b\mathrm{i})\left(\frac{a}{a^2+b^2} + \frac{-b}{a^2+b^2}\,\mathrm{i}\right) \\ &= \left(\frac{a^2}{a^2+b^2} - \frac{-b^2}{a^2+b^2}\right) + \left(\frac{-ab}{a^2+b^2} + \frac{ba}{a^2+b^2}\right)\mathrm{i} = 1 + 0\mathrm{i};\end{aligned}$$

that is, $zw = 1$ as desired. □

Recall that an integer n is called *even* (respectively, *odd*) if $2 \mid n$ (respectively, $2 \nmid n$).

Example 3.4. The sum of any two even integers is even.

Suppose that m and n are two even integers. We will show that $m+n$ is also even. Since m and n are even, there are integers p and q such that $m = 2p$ and $n = 2q$. Thus,

$$m + n = 2p + 2q = 2(p + q).$$

Since $p + q$ is an integer (as the sum of two integers), we conclude that $2 \mid (m + n)$: that is, $m + n$ is even. □

Remark. To prove similar statements involving odd integers is more difficult, because the oddness of n does not come with a convenient representation by an equation: it merely refutes the existence of an equation $n = 2q$ for any $q \in \mathbb{Z}$. Fortunately, there is an equivalent way to describe what an odd integer is: n is odd if and only if $n - 1$ is even. This translates into $n = 2q + 1$ for some $q \in \mathbb{Z}$: this is the "convenient representation" we wished for.

Example 3.5. The product of any two odd integers is odd.

Suppose that m and n are two odd integers. We will show that mn is also odd. Since m and n are odd, by the previous remark, there are integers p and q such that $m = 2p + 1$ and $n = 2q + 1$. Thus,

$$mn = (2p + 1)(2q + 1) = 4pq + 2p + 2q + 1 = 2(2pq + p + q) + 1.$$

Since $2pq + p + q$ is an integer (it is obtained by addition and multiplication of integers), we conclude that $2 \mid (mn - 1)$: that is, mn is odd. □

Remark. Example 3.5 ties the loose ends left at the end of Example 2.10, where we used the above fact to prove that 2 is a prime number in the sense of Definition 2.8.

Example 3.6. The sum of any two rational numbers is rational.

Suppose that x and y are two rationals. We will show that $x + y$ is also rational. Since x and y are rational, there are integers a, b, c, d such that

$$x = \frac{a}{b}, \quad y = \frac{c}{d}, \quad b, d \neq 0.$$

Thus, $bd \neq 0$, and

$$x + y = \frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{bc}{bd} = \frac{ad + bc}{bd}.$$

Since $ad + bc$ and bd are integers, with $bd \neq 0$, we conclude that $x + y$ is rational. □

Exercise 3.7. Prove the following statements.

(a) The product of any two even integers is even.

(b) The product of an even integer and another integer is even.

(c) The sum of an even integer and an odd integer is odd.

(d) The sum of any two odd integers is even.

(e) If a, b, c, x, y are integers such that $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$.

(f) The product of any two rational numbers is rational.

(g) If $a, b \in \mathbb{Q}$ and $a \neq 0$, there is a rational x such that $ax = b$.

The above proofs are examples of what is known as "direct proofs": we proceed from the assumptions towards the conclusion of the theorem. Sometimes it is easier to approach an implication differently. One such indirect approach uses the *contrapositive* of an implication $P \to Q$: that is, the implication $\neg Q \to \neg P$. These two implications are, in fact, equivalent:

Theorem 3.8. *An implication $P \to Q$ is logically equivalent to its contrapositive implication $\neg Q \to \neg P$.*

Proof. Use a truth table. □

The next lemma is an example of a *proof by contraposition*—where we prove an implication by proving its contrapositive.

Lemma 3.9. *If the product of two integers is even, at least one of the two integers must be even.*

Proof. We will prove this by contraposition. Let m and n be two integers such that the conclusion is false: that is, suppose that both m and n are odd. We need to show that then mn is also odd. That is exactly what we proved in Example 3.5. □

Another kind of an indirect proof is the *proof by contradiction*. In that kind of proof, we assume that the result is false and proceed to reach some sort of *contradiction*: we prove the truth of something that we know to be false. We then conclude that the original statement must be true. The proof of the next theorem is an example of such a proof.

Theorem 3.10. *The number $\sqrt{2}$ is irrational.*

Proof. See the proof of Theorem 5.2 in the textbook. □

Remark. Example 2.10 provides another example of a proof by contradiction, albeit a more implicit one. Thus, we now have two different indirect proofs that 2 is prime: a proof by contradiction in Example 2.10 and a proof by contraposition in Lemma 3.9 (which checks property (P) for $p = 2$). Note that both of those relied essentially on Example 3.5. This is quite common: it is often possible to frame an argument both as a proof by contraposition and a proof by contradiction.

Exercise 3.11. Use the fact that 3 is prime (recall Definition 2.8) to prove that $\sqrt{3}$ is irrational.

Exercise 3.12.

(a) Use Lemma 3.9 to prove that if n^3 is even, then n must be even.

(b) Use part (a) to prove that $\sqrt[3]{2}$ is irrational.[1]

(c) Use the fact that 5 is prime to prove that if $5 \mid n^4$, then $5 \mid n$.

(d) Use part (c) to prove that $\sqrt[4]{5}$ is irrational.

(e) Prove that any real root r of the equation $x^3 + 6x - 3 = 0$ must be irrational.

(f) Use Theorem 3.10 to prove that $3 - \sqrt{2}$ is irrational.

(g) Use part (b) to prove that $-1 + 5\sqrt[3]{2}$ is irrational.

(h) If a, b are rational, with $a \neq 0$, and x is irrational, prove that $ax + b$ is irrational.

[1] Recall that $\sqrt[3]{2}$ is a solution of the equation $x^3 = 2$.

4 Mathematical Induction

4.1 Basic Induction

Mathematical induction is a method for proving statements of the form

$$\forall n \in \mathbb{Z}^+, P(n)$$

where $P(n)$ is a property of an integer n that turns into a statement when n is replaced by a specific integer: e.g., $P(n)$ may represent an identity like $(n+1)^2 = n^2 + 2n + 1$, or an inequality like $3n^2 \leq 2^n$. In this lecture, we will follow the textbook and refer to such properties as *assertions*. In its basic form, the method of mathematical induction applies the following theorem to a specific choice of $P(n)$ to prove an universal statement of the above form.

Theorem 4.1 (Principle of mathematical induction)**.** *For an integer n, let $P(n)$ denote an assertion. Suppose that:*

(1) *$P(1)$ is true;*

(2) *for all integers $n \geq 1$, if $P(n)$ is true, then $P(n+1)$ is true.*

Then $P(n)$ holds for all positive integers n.

We will omit the proof of this theorem and simply apply it. It is equivalent to one of the basic axioms used to define the natural numbers and appears as Theorem 18.1 in the textbook, where it is presented with a proof that can be safely skipped on first (or second) reading.

Example 4.2. Prove a general formula for the sum

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

First, we need to conjecture a formula for the given sum. To that end, we will compute the sum for $n = 1, 2, \ldots, 5$. Let us denote the given sum by S_n. We find that

$$S_1 = \frac{1}{2}, \quad S_2 = \frac{2}{3}, \quad S_3 = \frac{3}{4}, \quad S_4 = \frac{4}{5}, \quad S_5 = \frac{5}{6}.$$

It appears that, at least for these five values of n, we have

$$S_n = \frac{n}{n+1}.$$

Hence, we will prove by induction that, for all $n \in \mathbb{Z}^+$,

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

When $n = 1$, we have

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

so the statement is true. Suppose now that

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

for some $n \geq 1$. We will show that

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

We have

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

Applying the inductive hypothesis to the last sum, we get

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

after some elementary algebra. This completes the induction. □

The lesson of the next example is that sometimes we restate the statement we want to prove before we use mathematical induction to prove it.

Example 4.3. Prove that for every $n \in \mathbb{Z}^+$,

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} < 2.$$

Failed attempt. When $n = 1$, the inequality claims that $\frac{3}{2} < 2$, which is true. In a standard proof by induction, the inductive step will have us show that when $n \in \mathbb{Z}^+$,

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} < 2 \quad \rightarrow \quad 1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} + \frac{1}{2^{n+1}} < 2.$$

The only sensible way to make use of the inductive hypothesis is as follows

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} + \frac{1}{2^{n+1}} < 2 + \frac{1}{2^{n+1}},$$

but this inequality is useless for the purpose of proving that

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} + \frac{1}{2^{n+1}} < 2.$$

And so, we are stuck. The workaround is to restart the process and try to prove *more* than the question asks...

Successful attempt. As we said at the end of our failed solution, we will try to prove a stronger inequality. We will use induction to show that if $n \in \mathbb{Z}^+$, then

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} \leq 2 - \frac{1}{2^n}. \qquad (*)$$

When $n = 1$, the inequality claims that $\frac{3}{2} \leq \frac{3}{2}$, which is true. Next, suppose that we know $(*)$ for some $n \in \mathbb{Z}^+$. Then

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} + \frac{1}{2^{n+1}} \leq 2 - \frac{1}{2^n} + \frac{1}{2^{n+1}} = 2 - \frac{1}{2^{n+1}},$$

which is the case $n + 1$ of $(*)$. This completes the induction. □

The assertion $P(n)$ need not be an identity or an inequality, it can be any type of property involving an integer variable. In the next theorem establishes a more general version of the principle of mathematical induction. Note that when $k = 1$, this is simply Theorem 4.1.

Theorem 4.4 (Principle of mathematical induction, II)**.** *For an integer n, let $P(n)$ denote an assertion. Suppose that k is an integer such that:*

(1) *$P(k)$ is true;*

(2) *for all integers $n \geq k$, if $P(n)$ is true, then $P(n + 1)$ is true.*

Then $P(n)$ holds for all integers $n \geq k$.

Proof. Let $Q(n)$ be the assertion $P(n + k - 1)$. We will prove that $Q(n)$ is true for all $n \geq 1$: that is, the statements $Q(1), Q(2), Q(3), \cdots$ are true. But these are the statements $P(k), P(k + 1), P(k + 2), \cdots$. Thus, if we prove $Q(n)$ for all $n \geq 1$, we will prove also that $P(n)$ is true for all $n \geq k$.

First, $Q(1)$ is the statement $P(1 + k - 1) = P(k)$, which is true by hypothesis (1).

Next, suppose that $Q(n)$ is true for some $n \geq 1$. Recall that $Q(n) = P(n + k - 1)$. Then $m = n + k - 1 \geq k$ and $P(m)$ is true; thus, hypothesis (2) implies that $P(m+1)$ is also true. However, $P(m + 1) = P(n + k) = Q(n + 1)$, so we have shown that for any $n \geq 1$, we have $Q(n) \to Q(n + 1)$. Theorem 4.1 then implies that $Q(n)$ is true for all $n \geq 1$, as desired. □

We now apply the previous theorem with $k = 2$ to prove the following result.

Theorem 4.5. *Let $p > 1$ be prime. If $a_1, a_2, \ldots, a_n$, $n \geq 2$, are integers with $p \mid a_1 a_2 \cdots a_n$, then $p \mid a_i$ for some $1 \leq i \leq n$.*

Proof. We will use induction on $n \geq 2$ to prove the following assertion:

> If $a_1, a_2, \ldots, a_n$ are any integers such that $p \mid a_1 a_2 \cdots a_n$,
> then $p \mid a_1$, or $p \mid a_2$, ..., or $p \mid a_n$.

The base case is $n = 2$. In this case, we want to prove that: if a_1 and a_2 are integers such that $p \mid a_1 a_2$, then $p \mid a_1$ or $p \mid a_2$. This is immediate from the definition of a prime number: simply take $a = a_1$ and $b = a_2$.

For the inductive step, we assume that $n \geq 2$ and we know that p has the property that for any n integers $a_1, a_2, \ldots, a_n$ such that $p \mid a_1 a_2 \cdots a_n$, we must have $p \mid a_1$, or $p \mid a_2$, ..., or $p \mid a_n$. We want to prove that p has the same property for products of $n + 1$ integers. Consider integers $a_1, a_2, \ldots, a_n, a_{n+1}$ such that $p \mid a_1 a_2 \cdots a_{n+1}$. We will show that $p \mid a_1$, or $p \mid a_2$, ..., or $p \mid a_{n+1}$. Since $a_1 a_2 \cdots a_{n+1} = (a_1 a_2 \cdots a_n) a_{n+1}$, we can use the fact that p is prime and apply Definition 2.8 to the integers $a = a_1 a_2 \cdots a_n$ and $b = a_{n+1}$ to conclude that since

$$p \mid a_1 a_2 \cdots a_{n+1} = (a_1 a_2 \cdots a_n) a_{n+1} = ab,$$

we must have $p \mid a = a_1 a_2 \cdots a_n$ or $p \mid b = a_{n+1}$. In the latter case, we are done—we have already established the p divides one of $a_1, a_2, \ldots, a_{n+1}$: namely, $p \mid a_{n+1}$. On the other hand, if $p \mid a_1 a_2 \cdots a_n$, then p must divide one of $a_1, a_2, \ldots, a_n$ by the inductive hypothesis. Thus, in all cases, p must divide one of $a_1, a_2, \ldots, a_{n+1}$; this completes the induction. □

4.2 Complete Induction

The rest of the results in this section are related to Theorem 4.6 below. When $k = 1$, this result is stated as Theorem 18.9 in the textbook (it appears within the problem section, on page 206). Problem 18.20 in the textbook gives a hint how to deduce that case from Theorem 4.1.

Theorem 4.6 (Principle of complete mathematical induction)**.** *For an integer n, let $P(n)$ denote an assertion. Suppose that k is an integer such that:*

(1) *$P(k)$ is true;*

(2) *for all integers $n \geq k$, if $P(k), P(k+1), \ldots, P(n)$ are true, then $P(n+1)$ is true.*

Then $P(n)$ holds for all integers $n \geq k$.

Theorem 4.7 (Quotient-remainder theorem)**.** *Let a, b be integers, with $b > 0$. Then there exist unique integers q, r such that*

$$a = bq + r, \quad 0 \leq r < b.$$

Remark. 1. This theorem makes two separate claims: first, that integers q, r with the stated properties exist; and second, that those integers are unique. It is a common practice to split the proofs of such results into two parts—one that establishes the existence claim and another that establishes the uniqueness claim. We will see this come into play in the proof of the quotient-remainder theorem and also in the proof of the fundamental theorem of arithmetic below.

2. It is convenient to use induction to prove the theorem for non-negative a, and then to reduce the case $a < 0$ to the case $a > 0$ by considering $|a|$. Thus, we will prove the theorem only in the case $a \geq 0$.

Proof. *Existence.* We fix $b \in \mathbb{Z}^+$ and, for $a \in \mathbb{N}$, we consider the statement

$$P(a): \quad \exists q, r \in \mathbb{Z}: \ a = bq + r,\ 0 \leq r < b.$$

We will use complete induction to prove that this statement is true for all $a \in \mathbb{N}$.

In the base case, $P(0)$ is true because we can choose $q = r = 0$. Suppose now that, for some $n \geq 0$, we know that each of the statements $P(0), P(1), \ldots, P(n)$ is true, and let $a = n + 1$. We will prove that $P(a)$ is also true. When $a < b$, we can simply take $q = 0$ and $r = a$ (so, we don't even need to use the inductive hypothesis). When $a \geq b$, we have

$$0 \leq a - b = (n + 1) - b \leq n,$$

so by the inductive hypothesis, we know that $P(a - b)$ is true. That is, there exist integers q', r such that

$$a - b = bq' + r, \quad 0 \leq r < b.$$

Hence, $a = b(q' + 1) + r$, and $P(a)$ is true because we can choose the same r and $q = q' + 1$. This completes the induction and establishes the existence claim of the theorem for $a \geq 0$.

Uniqueness. To prove the uniqueness claim, we assume that two pairs of integers q_1, r_1 and q_2, r_2 satisfy

$$a = bq_1 + r_1, \quad a = bq_2 + r_2, \quad 0 \leq r_1, r_2 < b,$$

and then show that these must in fact be the same pair: $q_1 = q_2$ and $r_1 = r_2$. Since this part of the proof does not use induction, we leave the details as an exercise. □

Exercise 4.8. Use Definition 2.8 to show that 3 and 5 are prime.

Theorem 4.9 (Fundamental Theorem of Arithmetic)**.** *Every integer $n > 1$ can be factored as a product*

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

where $p_1, p_2, \ldots, p_k$ are primes. Moreover, this factorization is unique in the sense that if

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

are two such factorizations and the primes $p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_l$ satisfy

$$p_1 \leq p_2 \leq \cdots \leq p_k, \quad q_1 \leq q_2 \leq \cdots \leq q_l,$$

then

$$k = l, \quad p_1 = q_1, \quad p_2 = q_2, \quad \ldots, \quad p_k = q_k.$$

Proof. *Uniqueness.* We argue by complete induction on $n \geq 2$. Consider first the base case $n = 2$. If $2 = p_1 p_2 \cdots p_k$ for some primes $p_1, p_2, \ldots, p_k$, $k \geq 1$, then since each $p_i \geq 2$, we have

$$2 = p_1 p_2 \cdots p_k \geq 2^k \quad \rightarrow \quad k \leq 1.$$

Hence, $k = 1$ and $p_1 = 2$. That is, any factorization of 2 as a product of primes must contain a single prime—2 itself. This proves that 2 has a unique prime factorization.

Suppose now that we know that, for some $n \geq 2$, each of the integers $2, 3, \ldots, n$ has a unique prime factorization. We will show that $n + 1$ also has a unique prime factorization. Let

$$n + 1 = p_1 p_2 \cdots p_k, \quad n + 1 = q_1 q_2 \cdots q_l$$

be two prime factorizations of $n + 1$, with the primes $p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_l$ listed in increasing order:

$$p_1 \leq p_2 \leq \cdots \leq p_k, \quad q_1 \leq q_2 \leq \cdots \leq q_l.$$

Since q_l is a prime and $q_l \mid (n + 1) = p_1 p_2 \cdots p_k$, by Theorem 4.5, we must have $q_l \mid p_i$ for some $1 \leq i \leq k$. Hence, $q_l \leq p_i$. By the same argument, we must have $p_k \leq q_j$ for some $1 \leq j \leq l$. Using the ordering of the p's and the q's, we deduce that

$$p_i \leq p_k \leq q_j \leq q_l \leq p_i \quad \rightarrow \quad p_k = q_l = p_i.$$

That is, we have $n + 1 = m p_k = m q_l$, where

$$m = p_1 p_2 \cdots p_{k-1} = q_1 q_2 \cdots q_{l-1}.$$

We now have two possibilities for m.

Case 1: $m = 1$. This means that $k = l = 1$, $n + 1$ is actually a prime, and we have $p_1 = q_1 = n + 1$. Thus, in this case, the two factorizations are identical.

Case 2: $m > 1$. Since $m = (n + 1)/p_k < n + 1$, in this case m is among the integers $2, 3, \ldots, n$ and has a unique prime factorization by the inductive hypothesis. Thus, we must have

$$k - 1 = l - 1, \quad p_1 = q_1, \quad p_2 = q_2, \quad \ldots, \quad p_{k-1} = q_{l-1} = q_{k-1}.$$

Since we already established that $p_k = q_l$, we conclude that in this case, the two factorizations are again identical. This completes the induction.

Existence. This part of the theorem can also be proved by complete induction. We leave it as an exercise. □

4.3 Recursion

The principle of mathematical undergirds the definitions of many mathematical concepts through Theorem 18.6 in the textbook, which essentially says that we can define various concepts *recursively*. For example, in (pre)calculus, we define the summation notation by

$$\sum_{k=m}^{n} a_k = a_m + a_{m+1} + \cdots + a_n \qquad (m \leq n).$$

A more rigorous defintion of the same symbol, using recursion, says that

$$\sum_{k=m}^{m} a_k = a_m, \qquad \sum_{k=m}^{n} a_k = \left(\sum_{k=m}^{n-1} a_k\right) + a_n \quad (n > m).$$

This makes sense intuitively, but to justify that intuition, we must rely on the principle of mathematical induction through Theorem 18.6. Similarly, we can define the product notation:

$$\prod_{k=m}^{m} a_k = a_m, \qquad \prod_{k=m}^{n} a_k = \left(\prod_{k=m}^{n-1} a_k\right) \cdot a_n \quad (n > m).$$

Hints to Some Exercises

Exercise 4.8. To show that 3 has property (P), we need to show that if a, b are integers, $3 \mid ab \to (3 \mid a \vee 3 \mid b)$. Prove the countrapoisitive. For any $a, b \in \mathbb{Z}$ such that $3 \nmid a$ and $3 \nmid b$, use Theorem 4.7 to write a and b as $a = 3q_1 + r_1$ and $b = 3q_2 + r_2$, $r_1, r_2 \in \{1, 2\}$. Then show that $3 \nmid ab$.

You may argue similarly to show that 5 has property (P).

5 Inequalities

We will begin by reviewing the basic properties of inequalities that you know from elementary algebra, though perhaps, you may not have seen them too often in this generality. The most basic properties of inequalities between real numbers are the following three axioms:

I. If $x > 0$ and $y > 0$, then $x + y > 0$.

II. If $x > 0$ and $y > 0$, then $xy > 0$.

III. Exactly one of the following three statements is true: $x < 0$, $x = 0$, or $x > 0$.

Using the three axioms above, we can prove other important properties of inequalities. Some of those are listed in the next theorem. Before we get to that, let us recall that the inequality $x < y$ really means that $y - x > 0$; in particular, $x < 0$ means that $-x > 0$. Also, recall the definition of $|x|$, the absolute value of x:

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

Theorem 5.1. *The following properties hold for all real a, x, y, z, with $a > 0$:*

(1) *If $x < y$ and $y < z$, then $x < z$.*

(2) *If $x < y$, then $x + z < y + z$.*

(3) *If $x < y$ and $z > 0$, then $xz < yz$.*

(4) *If $x < y$ and $z < 0$, then $yz < xz$.*

(5) *If $0 < x < y$, then $0 < y^{-1} < x^{-1}$.*

(6) *$|x| < a$ if and only if $-a < x < a$.*

Proof. 1. Since $y - x > 0$ and $z - y > 0$, axiom I gives

$$z - x = (z - y) + (y - x) > 0 \quad \Longrightarrow \quad x < z.$$

4. Since $y - x > 0$ and $-z > 0$, axiom II gives

$$xz - yz = (y - x)(-z) > 0 \quad \Longrightarrow \quad yz < xz.$$

6. First, suppose that $|x| < a$. We will show that $-a < x < a$. We consider two cases:
Case 1: $x \geq 0$. Then $|x| = x$, and we have $-a < 0 \leq x = |x| < a$.
Case 2: $x < 0$. Then $|x| = -x$, and we obtain, using (d), that

$$-x = |x| < a \quad \Longrightarrow \quad -a = a(-1) < (-x)(-1) = x < 0 < a.$$

Conversely, suppose that $-a < x < a$. We will show that $|x| < a$.
Case 1: $x \geq 0$. Then $|x| = x < a$.
Case 2: $x < 0$. Then $|x| = -x$, so $|x| = -x = x(-1) < (-a)(-1) = a$. □

Example 5.2. Prove that when $x \geq 2$,

$$\frac{1}{x} < \frac{3x}{2x^2 - 3} < \frac{3}{x}.$$

First, we will find constants A and B such that

$$Ax^2 < 2x^2 - 3 < Bx^2 \qquad \forall x \geq 2. \tag{$*$}$$

Since

$$2x^2 - 3 < 2x^2 \qquad \forall x \in \mathbb{R},$$

we can choose $B = 2$. Further, when $x \geq 2$, we have

$$2x^2 - 3 = x^2 + x^2 - 3 \geq x^2 + 4 - 3 > x^2.$$

Thus, we can choose $A = 1$ in $(*)$.

Note that when $A = 1$ and $B = 2$, all three expressions in $(*)$ are positive, so

$$\frac{1}{2x^2} < \frac{1}{2x^2 - 3} < \frac{1}{x^2}.$$

Multiplying this double inequality by $3x$ implies that

$$\frac{1}{x} < \frac{1.5}{x} = \frac{3x}{2x^2} < \frac{3x}{2x^2 - 3} < \frac{3x}{x^2} = \frac{3}{x}.$$

This establishes the claim. □

Example 5.3. Prove that 0 is the largest $m \in \mathbb{R}$ with the property that

$$\frac{3x}{2x^2 - 3} > m \qquad \text{for all } x \geq 2. \tag{$*$}$$

In the previous example, we showed that when $x \geq 2$, we have

$$\frac{3x}{2x^2 - 3} > \frac{1}{x} > 0 \qquad \text{for all } x \geq 2.$$

Thus, $m = 0$ does have the desired property. Next, we will show that no positive number has this property.

Let $m > 0$. We will show that $(*)$ fails. By Example 5.2, when $x \geq 2$, we have

$$\frac{3x}{2x^2 - 3} < \frac{3}{x}.$$

We can find an $x \geq 2$ such that $\frac{3}{x} < m$. For example, if k is an integer with $k \geq m^{-1}$ (this ensures $k \geq 1$), then $x = 4k \geq 4$ and

$$\left.\frac{3x}{2x^2 - 3}\right|_{x=4k} < \frac{3}{4k} \leq \frac{3}{4m^{-1}} = \frac{3m}{4} < m.$$

That is, $(*)$ fails. □

Theorem 5.4 (Triangle inequality)**.** *If $x, y \in \mathbb{R}$, then $|x + y| \leq |x| + |y|$.*

Proof. We leave the proof of this theorem as an optional exercise. For a hint how to proceed, see Problem 5.14 on page 55 of the textbook. □

Corollary 5.5 (Reverse triangle inequality)**.** *If $x, y \in \mathbb{R}$, then $|x - y| \geq |x| - |y|$.*

Proof. Applying the triangle inequality to y and $z = x - y$, we get

$$|x| = |z + y| \leq |z| + |y| \quad \Longrightarrow \quad |x| - |y| \leq |z| = |x - y|. \qquad \square$$

Exercise 5.6. Work through Exercise 18.3(b) on page 196 of the textbook.

Example 5.7. Prove that if $|x| \leq 0.5$, then

$$2.7 < \left|x^5 + x^2 - 3\right| < 3.3.$$

Using the triangle inequality, we find that

$$\left|x^5 + x^2\right| \leq \left|x^5\right| + \left|x^2\right| = |x|^5 + |x|^2 \leq (0.5)^5 + (0.5)^2 < 0.3.$$

Thus, using the triangle inequality one more time, we get

$$\left|x^5 + x^2 - 3\right| \leq \left|x^5 + x^2\right| + |-3| < 0.3 + |-3| = 3.3.$$

On the other hand, by the reverse triangle inequality,

$$\left|x^5 + x^2 - 3\right| = \left|3 - (x^5 + x^2)\right| \geq |3| - \left|x^5 + x^2\right| > 3 - 0.3 = 2.7. \qquad \square$$

Problems

Problem 5.1. Prove that if $x \geq 5$, then

$$0 < \frac{3x + 2}{\sqrt{x^3 - 3x^2 + 3}} < \frac{7}{\sqrt{x}}.$$

Problem 5.2. Prove that if $0 < |x - 1| < 0.1$, then

$$\left|\frac{3x - 4}{x^2 - 3} - \frac{1}{2}\right| < 2|x - 1|.$$

Hint. You may prefer to write x as $x = 1 + h$, where $|h| < 0.1$.

Problem 5.3. Use the algebraic properties of inequalities to prove that:

(a) $|x^2 + x - 1| \leq |x - 2|^2 + 5|x - 2| + 5$ for all real x;

(b) $|x^2 + x - 1| < 6.1$ for all real x with $|x - 2| < 0.2$;

(c) $|x^3 - x^2 - 3x + 2| < 8|x - 2|$ for all real x with $|x - 2| < 0.5$.

Problem 5.4. Let $a > 0$. Use part (6) of Theorem 5.1 to prove the given statements for all real x, y.
(a) $|x| \geq a$ if and only if $x \leq -a$ or $x \geq a$.
(b) $|x - y| < a$ if and only if $y - a < x < y + a$.

Problem 5.5. Prove parts (2), (3), and (5) of Theorem 5.1.

Problem 5.6. In this exercise, avoid using calculus or graphing, the basic properties of inequalities should suffice.
(a) Prove that $\frac{1}{2}$ and 4 are, respectively, the largest a and the smallest b such that

$$a \leq \frac{x - 3}{2x + 1} \leq b \qquad \text{for all } x < -1.$$

(b) Prove that there is no b such that $\dfrac{x - 3}{2x + 1} \leq b$ for all $x < -0.5$.
(c) Prove that there is no a such that $\dfrac{x - 3}{2x + 1} \geq a$ for all $x > -0.5$.

6 Sequences

Recall that a *sequence* in a set X is a function $a : \mathbb{N} \to X$. This "labels" some elements of X with the natural numbers (allowing for repeats), and we prefer to write the nth value of the sequence as a_n instead of $a(n)$. Though the set X can be arbitrary, we will focus on the cases when $X = \mathbb{R}$ or $X = \mathbb{Q}$; we will call those *real* or *rational* sequences, respectively.

Definition 6.1. A real sequence (x_n) is called *increasing* (respectively, *decreasing*) if $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$ (respectively, if $x_n \geq x_{n+1}$ for all $n \in \mathbb{N}$). A real sequence (x_n) is called *strictly increasing* (respectively, *strictly decreasing*) if $x_n < x_{n+1}$ for all $n \in \mathbb{N}$ (respectively, if $x_n > x_{n+1}$ for all $n \in \mathbb{N}$).

Example 6.1. Let $x_n = \dfrac{3n^2+1}{n^2-2n+3}$. Determine whether the sequence (x_n) is increasing, decreasing, or neither.

The sequence is increasing if $x_{n+1} - x_n \geq 0$ for all n; it is decreasing if $x_{n+1} - x_n \leq 0$ for all n; and it is neither if the difference $x_{n+1} - x_n$ can be positive for some values of n and negative for others. Hence, we will consider this difference and try to determine its sign. We have

$$\begin{aligned} x_{n+1} - x_n &= \frac{3(n+1)^2+1}{(n+1)^2-2(n+1)+3} - \frac{3n^2+1}{n^2-2n+3} \\ &= \frac{3n^2+6n+4}{n^2+2} - \frac{3n^2+1}{n^2-2n+3} = \frac{-6n^2+10n+10}{(n^2+2)(n^2-2n+3)}. \end{aligned}$$

Since

$$n^2+2 \geq 2 \quad \text{and} \quad n^2-2n+3 = (n-1)^2+2 \geq 2,$$

the sign of the difference $x_{n+1}-x_n$ is determined by the sign of the trinomial $-6n^2+10n+10$. This trinomial can be both positive and negative:

$$-6(1)^2+10(1)+10 = 14 > 0, \qquad -6(3)^2+10(3)+10 = -14 < 0.$$

Hence, the sequence is neither increasing, nor decreasing.

On the other hand, if we consider the truncated sequence $(x_n)_{n\geq 3}$, we find that this sequence is strictly decreasing. Indeed, when $n \geq 3$, we have

$$-6n^2+10n+10 \leq -18n+10n+10 = -8n+10 \leq -14 < 0. \qquad \square$$

Definition 6.2. A real sequence (x_n) is called *bounded above* (respectively, *bounded below*) if there is a real number b such that $x_n \leq b$ (respectively, $x_n \geq b$) for all $n \in \mathbb{N}$. When such a number b exists, it is called an *upper bound* (respectively, a *lower bound*) of the sequence.

A sequence (x_n) that is bounded both above and below is called *bounded*.

Example 6.2. Consider the sequence (x_n) defined for $n \geq 1$ by $x_n = 3 - \dfrac{2}{n}$. The first few terms of the sequence are $1, 2, 2\frac{1}{3}, 2\frac{1}{2}, 2\frac{3}{5}, \cdots$, so it is not difficult to guess either upper or lower bound for (x_n).

For example, the numbers 3, 4, and π are all upper bounds of (x_n), because when $n \geq 1$, we have

$$\frac{2}{n} > 0 \quad \to \quad 3 - \frac{2}{n} < 3 - 0 = 3 < \pi < 4.$$

However, 2.5 is not an upper bound because $x_5 = 2.6$ is a counterexample to the statement

"$x_n \leq 2.5$ for all $n \geq 1$."

Similarly, 2.99 is not an upper bound because $x_{1000} = 2.998$ is a counterexample to that claim.

Further, -100, 0, and 1 are lower bounds of (x_n), because when $n \geq 1$,

$$\frac{2}{n} \geq \frac{2}{1} \quad \to \quad -100 < 0 < 1 = 3 - \frac{2}{1} \leq 3 - \frac{2}{n}.$$

However, 1.000001 is not a lower bound, because $x_1 = 1$ is a counterexample to the statement

"$x_n \geq 1.000001$ for all $n \geq 1$." □

Next, we discuss the notions of a least upper and a greatest lower bounds of a sequence. The definitions below are essentially those developed in Example 19.5 and Exercise 19.7 of the textbook, except that here we do not rely on Chapter 12.

Definition 6.3. Let (x_n) be a real sequence that is bounded above. A real number U is called a *least upper bound* (or a *supremum*) of (x_n) if:

(i) $x_n \leq U$ for all $n \in \mathbb{N}$;

(ii) if b is any upper bound of (x_n) (i.e., if $x_n \leq b$ for all $n \in \mathbb{N}$), then $U \leq b$.

If U is a supremum of (x_n), we write $U = \sup(x_n)$.

Definition 6.4. Let (x_n) be a real sequence that is bounded below. A real number L is called a *greatest lower bound* (or an *infimum*) of (x_n) if:

(i) $x_n \geq L$ for all $n \in \mathbb{N}$;

(ii) if b is any lower bound of (x_n) (i.e., if $x_n \geq b$ for all $n \in \mathbb{N}$), then $L \geq b$.

If L is an infimum of (x_n), we write $L = \inf(x_n)$.

Example 6.3. Consider again the sequence (x_n) defined for $n \geq 1$ by $x_n = 3 - \dfrac{2}{n}$. We will study its sup and inf.

We already checked in Example 6.2 that 1 is a lower bound of (x_n): i.e., property (i) in Definition 6.4 holds. Moreover, if b is any lower bound, it must satisfy the inequality $x_n \geq b$ for all $n \geq 1$. In particular, when $n = 1$, we have $1 = x_1 \geq b$. Thus, property (ii) in Definition 6.4 also holds.

It is also intuitive that $\sup(x_n) = 3$. Again, we know from Example 6.2 that 3 is an upper bound of (x_n), so property (i) in Definition 6.3 holds. The check of property (ii) in that definition, however, is a little trickier. For any number b, we must show that if b is an upper bound of (x_n), then $b \geq 3$. Since 3 is not a term of the sequence, this does not follow simply from the definition of an upper bound. Instead, we will look at the contrapositive implication: we will show that if $b < 3$, then b isn't an upper bound. Suppose that $b < 3$. Then we can construct a positive integer k such that $x_k > b$, which means that b fails to be an upper bound of (x_n). For example, we can take any integer k such that $k > \dfrac{2}{3-b}$, because

$$\frac{2}{k} < \frac{2}{2/(3-b)} = 3 - b \quad \to \quad x_k = 3 - \frac{2}{k} > 3 - (3-b) = b. \qquad \square$$

Example 6.4. Consider the sequence (x_n), where $x_n = (n+1)^{(-1)^n}$. The first few terms of this sequence are $1, \frac{1}{2}, 3, \frac{1}{4}, 5, \frac{1}{6}, 7, \frac{1}{8}, \cdots$. Since all the terms are positive, 0 is a lower bound for this sequence; so is -2026. However, -2026 is definitely not the infimum of the sequence, as there are larger lower bounds, such as -6. Below, we show that $\inf(x_n) = 0$.

We will check that $L = 0$ satisfies both conditions of Definition 6.4. We already observed that $x_n > 0$ for all $n \in \mathbb{N}$; hence, $L = 0$ is a lower bound for (x_n). Now, let b be any lower bound for (x_n). We will show that $b \leq 0$ by showing that no positive b can be a lower bound for (x_n).

Let $b > 0$. To show that b is not a lower bound for the sequence, we need to find a term x_k such that $x_k < b$. If k is odd, we have $x_k = (k+1)^{-1}$, so it suffices to choose an odd integer k so that $k > b^{-1}$—for example, we may choose $k = 2m + 1$, where m is an integer with $m \geq b^{-1} > 0$. Then, k is a positive odd integer such that

$$x_k = \frac{1}{k+1} = \frac{1}{2m+2} \leq \frac{1}{2b^{-1}+2} < \frac{1}{2b^{-1}} = \frac{b}{2} < b.$$

Therefore, b is not a lower bound for (x_n). This establishes that $\inf(x_n) = 0$. $\square$

Exercise 6.5. Work through Exercise 19.6 on page 211 of the textbook.

Exercise 6.6. Work through Exercise 19.10 on pages 212–213 of the textbook.

Exercise 6.7. Let (x_n) be a real sequence that is bounded above. Prove that $\sup(x_n)$ is unique.

Exercise 6.8. Let (x_n) be a real sequence that is bounded below. Prove that $\inf(x_n)$ is unique.

Note that while these exercises establish the uniqueness of the supremum and the infimum of a bounded sequence when those exist, it is not obvious that they do exist. The following is a fundamental property of the real numbers.

Axiom 6.9 (Completeness axiom of $\mathbb{R}$)**.** *Every bounded above real sequence (x_n) has a least upper bound in $\mathbb{R}$.*

Exercise 6.10. Use the above axiom to prove that every bounded below real sequence (x_n) has a greatest lower bound in $\mathbb{R}$. This is essentially Exercise 19.11 on page 213 of the textbook.

Example 6.11. Let (x_n) be a bounded real sequence, with $\alpha = \inf(x_n)$ and $\beta = \sup(x_n)$. Consider the sequence (y_n) defined by $y_n = 2x_n + 5$. We will show that $\inf(y_n) = 2\alpha + 5$ and $\sup(y_n) = 2\beta + 5$.

First, we check that $2\alpha + 5$ is a lower bound of (y_n). Since α is a lower bound of (x_n), for all $n \geq 1$, we have

$$\alpha \leq x_n \quad \rightarrow \quad 2\alpha + 5 \leq 2x_n + 5 = y_n.$$

This proves that $2\alpha + 5 \leq y_n$ for all $n \in \mathbb{N}$: that is, $2\alpha + 5$ is a lower bound of (y_n).

Next, we will show that if b is a lower bound of (y_n), then $b \leq 2\alpha + 5$. Suppose that b is a lower bound of (y_n):

$$b \leq y_n = 2x_n + 5 \quad \forall n \in \mathbb{N}.$$

We can rewrite the last statement as

$$\frac{b-5}{2} \leq x_n \quad \forall n \in \mathbb{N},$$

which says that $\frac{b-5}{2}$ is a lower bound of (x_n). Thus, by the definition of $\inf(x_n)$,

$$\frac{b-5}{2} \leq \alpha \quad \rightarrow \quad b \leq 2\alpha + 5,$$

as desired.

The proof that $\sup(y_n) = 2\beta + 5$ is similar, so we leave it as an exercise. □

Exercise 6.12. Let (x_n) be a bounded real sequence, with $\alpha = \inf(x_n)$ and $\beta = \sup(x_n)$. Consider the sequence (y_n) defined by $y_n = -3x_n + 4$. Show that $\inf(y_n) = -3\beta + 4$ and $\sup(y_n) = -3\alpha + 4$.

Hints to Some Exercises

Exercise 6.7. Let (x_n) be a real sequence and suppose that U_1 and U_2 are suprema of (x_n). Since U_1 is an upper bound for (x_n) and U_2 is a supremum, $U_2 \leq U_1$. Similarly (supply a detailed explanation), $U_1 \leq U_2$. Hence, $U_1 = U_2$, which proves the uniqueness of the supremum.

7 Convergence of Real Sequences

Start by reading pages 223–228 in the textbook.

7.1 Basic Definitions

Definition 7.1. We say that a real sequence (x_n) *converges* if there is a real number L such that for all $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $|x_n - L| < \varepsilon$ for all $n \geq N$. When such an L exists, we call L the *limit* of the sequence and write $x_n \to L$ or $\lim x_n = L$; we also say that (x_n) *converges to* L. When such an L does not exist, we say that the sequence is *divergent*.

Remark. It is often helpful to write the above definition formally, using quantifiers and inequalities. To that end, the statement $\lim x_n = L$ is a compressed version of:

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} : \forall n \geq N, |x_n - L| < \varepsilon.$$

To establish formally even a simple statement involving limits, such as $\lim\limits_{n\to\infty} \dfrac{n}{n+2} = 1$, we need to write a brief proof that follows a standard blueprint based on this formal version. The next couple of examples illustrate that; they complement Examples 20.2 and 20.3 in the textbook.

Example 7.1. Prove that $\lim\limits_{n\to\infty} \dfrac{4}{n^2} = 0$.

We verify the definition of $\lim\limits_{n\to\infty} x_n = L$ in the special case when $x_n = \dfrac{4}{n^2}$ and $L = 0$. Let $\varepsilon > 0$. We choose $N > 2/\sqrt{\varepsilon}$. Then for $n \geq N$, we have

$$|x_n - 0| = \left|\frac{4}{n^2} - 0\right| = \frac{4}{n^2} \leq \frac{4}{N^2} < \frac{4}{(2/\sqrt{\varepsilon})^2} = \frac{4}{4/\varepsilon} = \varepsilon. \qquad \square$$

Example 7.2. Prove that $\lim\limits_{n\to\infty} \dfrac{3n^2+1}{n^2-2n+3} = 3$.

We verify the definition of $\lim x_n = L$ in the special case when $x_n = \dfrac{3n^2+1}{n^2-2n+3}$ and $L = 3$. Let $\varepsilon > 0$. We choose $N \geq 1 + 40/\varepsilon$, which is at least 2. Then for $n \geq N$, we have

$$\begin{aligned} |x_n - 3| &= \left|\frac{3n^2+1}{n^2-2n+3} - 3\right| = \left|\frac{6n-8}{n^2-2n+3}\right| = \frac{|6n-8|}{|(n-1)^2+2|} = \frac{|6n-8|}{(n-1)^2+2} \\ &\leq \frac{10n}{(n-1)^2+2} < \frac{10n}{(n/2)^2} = \frac{40}{n} \leq \frac{40}{N} < \frac{40}{40/\varepsilon} = \varepsilon. \end{aligned}$$

Here, we have used the inequalities

$$|6n-8| \leq |6n| + |-8| = 6n + 8 \leq 10n$$

and

$$(n-1)^2 + 2 > (n-1)^2 \geq (n/2)^2,$$

both of which hold for all $n \geq 2$. □

Example 7.3. Prove that $\lim_{n\to\infty} \dfrac{3n^2+1}{n^2-2n+3} \neq 2$.

To prove this, we verify the negation of the definition of $\lim x_n = L$ (recall Exercise 20.5 in the textbook) in the special case when $x_n = \dfrac{3n^2+1}{n^2-2n+3}$ and $L = 2$. That is, we must verify the following quantified statement:

$$\exists \varepsilon > 0 : \forall N \in \mathbb{N}, \exists n \geq N : \left| \frac{3n^2+1}{n^2-2n+3} - 2 \right| \geq \varepsilon.$$

Choose $\varepsilon = \frac{1}{4}$ (which is > 0). Let $N \in \mathbb{N}$ and choose $n = \max(N, 2)$ (so that, $n \geq N$ and $n \geq 2$). In Example 7.2, we showed that $n^2 - 2n + 3 > 0$. Using this inequality and the fact that $n \geq 2$, we get

$$\begin{aligned} |x_n - 2| &= \left| \frac{3n^2+1}{n^2-2n+3} - 2 \right| = \left| \frac{n^2+4n-5}{n^2-2n+3} \right| = \frac{|(n+5)(n-1)|}{|n^2-2n+3|} \\ &= \frac{(n+5)(n-1)}{n^2-2n+3} \geq \frac{n(n-1)}{n^2+3} \geq \frac{n(n/2)}{n^2+3} > \frac{n^2/2}{2n^2} = \frac{1}{4} = \varepsilon. \end{aligned}$$

Here, we have used the inequality $n^2 + 3 < 2n^2$, which holds for all $n > \sqrt{3}$. □

Example 7.4. Let (x_n) be a convergent real sequence with $\lim x_n = 3$, and let (y_n) be defined by $y_n = 5x_n + 1$. Prove that $\lim y_n = 16$.

We verify the definition of $\lim x_n = L$ with y_n in place of x_n and $L = 16$. Let $\varepsilon > 0$. Before we specify our choice of N, we will stop to record our hypothesis: if $\varepsilon_1 > 0$, then we can find an $N_1 \in \mathbb{N}$ with the property that

$$|x_n - 3| < \varepsilon_1 \quad \text{for all } n \geq N_1.$$

We wish to make use of this when $\varepsilon_1 = \frac{1}{5}\varepsilon$, so we will choose N to be the integer N_1 that corresponds to $\varepsilon_1 = \varepsilon/5$. Then for $n \geq N$, we have

$$|y_n - 16| = |(5x_n + 1) - 16| = |5(x_n - 3)| = 5|x_n - 3| < 5\varepsilon_1 = \varepsilon. \qquad \square$$

7.2 Important Properties

The following two theorems are proved in the textbook.

Theorem 7.5. *If a sequence converges, then its limit is unique.*

Theorem 7.6. *If a sequence converges, then it is bounded.*

Theorem 7.7. *Let (x_n) and (y_n) be convergent real sequences such that $x_n \to L$ and $y_n \to M$. Then*

(1) *$x_n + y_n \to L + M$;*

(2) *$\alpha x_n \to \alpha L$ for any real α;*

(3) *$x_n y_n \to LM$;*

(4) *when $L \neq 0$, $1/x_n \to 1/L$.*

Proof. 3. Let $\varepsilon > 0$. First, by Theorem 7.6, the sequence (x_n) is bounded; hence, there is a real $K > 0$ such that $|x_n| \leq K$ for all $n \in \mathbb{N}$. Since $\lim x_n = L$ and $\lim y_n = M$, there exist integers N_1 and N_2 such that:

$$\forall n \geq N_1, \;\; |x_n - L| < \varepsilon_1; \qquad \forall n \geq N_2, \;\; |y_n - M| < \varepsilon_2; \tag{7.1}$$

where

$$\varepsilon_1 = \frac{\varepsilon}{2(|M|+1)}, \qquad \varepsilon_2 = \frac{\varepsilon}{2K}. \tag{7.2}$$

Choose $N = \max(N_1, N_2)$. Then for $n \geq N$, both inequalities in (7.1) hold, and we have

$$\begin{aligned}
|x_n y_n - LM| &= |x_n y_n - x_n M + x_n M - LM| \\
&\leq |x_n y_n - x_n M| + |x_n M - LM| && \text{(Theorem 5.4)} \\
&= |x_n|\,|y_n - M| + |M|\,|x_n - L| \\
&\leq K|y_n - M| + |M|\,|x_n - L| && \text{(by choice of } K\text{)} \\
&< K\varepsilon_2 + |M|\varepsilon_1 && \text{(by (7.1))} \\
&= \frac{\varepsilon}{2}\left(1 + \frac{|M|}{|M|+1}\right) < \varepsilon. && \text{(by (7.2))}
\end{aligned}$$

(Take note how the last two lines in this chain explain the choices for ε_1 and ε_2 in (7.2).) This proves that $\lim x_n y_n = LM$. □

Exercise 7.8. Let (x_n) be a convergent sequence, and define (y_n) by $y_n = x_{n+1}$. Prove that (y_n) is also convergent and $\lim y_n = \lim x_n$.

Exercise 7.9. Let (x_n) be a real sequence defined recursively by $x_0 = 2$, $x_{n+1} = \dfrac{x_n}{2} + \dfrac{1}{x_n}$.

(a) Prove that $x_n > \sqrt{2}$ for all $n \in \mathbb{N}$.

(b) Prove that (x_n) is decreasing.

(c) Use parts (a) and (b) and Theorem 20.12 in the textbook to prove that (x_n) is convergent.

(d) Use the result of Exercise 7.8 and Theorem 7.7 to find $\lim x_n$.

8 Cauchy Sequences and Completeness

8.1 Cauchy Sequences

Definition 8.1. A real sequence (x_n) is called a *Cauchy sequence* (or a *fundamental sequence*) if it has the following property:

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} : \forall m \geq N,\ \forall n \geq N,\ \ |x_m - x_n| < \varepsilon.$$

Example 8.1. We will show that the sequence (x_n), where $x_n = \dfrac{4n-3}{3n+5}$, is Cauchy.

Let $\varepsilon > 0$ and let $N \in \mathbb{N}$ be chosen so that $N \geq 4/\varepsilon$. We have

$$|x_m - x_n| = \left|\frac{4m-3}{3m+5} - \frac{4n-3}{3n+5}\right| = \left|\frac{29(m-n)}{(3m+5)(3n+5)}\right| < \frac{29|m-n|}{(3m)(3n)}.$$

If $m \geq n \geq N$, we have $|m-n| = m-n$, so we deduce

$$|x_m - x_n| < \frac{29(m-n)}{9mn} < \frac{4m}{mn} = \frac{4}{n} \leq \frac{4}{N} \leq \frac{4}{4/\varepsilon} = \varepsilon;$$

and if $n > m \geq N$, we have $|m-n| = n-m$, so we deduce

$$|x_m - x_n| < \frac{29(n-m)}{9mn} < \frac{4n}{mn} = \frac{4}{m} \leq \frac{4}{N} \leq \varepsilon.$$

Therefore, whenever $m \geq N$ and $n \geq N$, we have $|x_m - x_n| < \varepsilon$. This proves that (x_n) is Cauchy. □

Theorem 8.2. *If (x_n) is a convergent sequence, then it is Cauchy.*

Proof. Let $L = \lim x_n$. Fix $\varepsilon > 0$. Then there is an integer N such that, for any integer n,

$$n \geq N \quad \rightarrow \quad |x_n - L| < \frac{\varepsilon}{2}.$$

Therefore, when $m, n \geq N$, we have

$$|x_m - x_n| = |(x_m - L) + (L - x_n)| \leq |x_m - L| + |L - x_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

where the first inequality follows from the triangle inequality. □

While the above theorem establishes that every convergent sequence must be Cauchy, Exercise 7.9 gives an example of a sequence in $\mathbb{Q}$, which is Cauchy but does not converge in $\mathbb{Q}$ (because it converges to $\sqrt{2}$—which is irrational). The next theorem states that this cannot occur if we ask whether a Cauchy sequence converges in $\mathbb{R}$.

Theorem 8.3 (Completeness of $\mathbb{R}$)**.** *Every Cauchy sequence (x_n) in $\mathbb{R}$ is convergent to some real number.*

We will not prove this theorem here, but we stop to point out that it is, in fact, equivalent to the completeness axiom of $\mathbb{R}$. On the one hand, the proof of Theorem 8.3 draws on the completeness axiom (since it uses Theorem 20.11 in the textbook); on the other hand, if we assume that Theorem 8.3 is true, we can turn the completeness axiom into a theorem, whose proof will depend on the above theorem.

This begs a question: Since we are focusing on real sequences here, why bother with Cauchy sequences at all? After all, they are merely convergent sequences in disguise... One context where Cauchy sequences are useful is for showing that a sequence is *divergent*. To prove that a sequence (x_n) diverges using the definition, we have to prove that $\lim x_n \neq L$ *for all* L. This often requires multiple cases. On the other hand, one can often prove that (x_n) is not Cauchy is relatively easily.

Exercise 8.4. State without using negatives what it means for a sequence (x_n) to not be Cauchy.

Example 8.5. Prove that (x_n) is divergent for the following choices of x_n:

$$x_n = (-1)^{n+1}, \quad x_n = \sqrt{2n + \sin n}.$$

We show that the sequences are not Cauchy, and therefore, not convergent. We need to find a choice of ε such that, for all $N \in \mathbb{N}$, we can find an $m > N$ and an $n > N$ for which

$$|x_m - x_n| \geq \varepsilon.$$

First, we consider the case $x_n = (-1)^{n+1}$. We have

$$|x_m - x_n| = \left|(-1)^{m+1} - (-1)^{n+1}\right| = \left|(-1)^{m-n} - 1\right| = \begin{cases} 2 & \text{if } m-n \text{ is odd,} \\ 0 & \text{if } m-n \text{ is even.} \end{cases}$$

In particular, when $m = n + 1$, we find that

$$|x_{n+1} - x_n| = 2.$$

Thus, if we choose $\varepsilon = 1$ and if $N \in \mathbb{N}$, we can choose $m = N + 2$ and $n = N + 1$ so that

$$|x_m - x_n| = 2 > 1 = \varepsilon.$$

This proves that the first sequence is not Cauchy.

Next, we consider the case $x_n = \sqrt{2n + \sin n}$. We have

$$\begin{aligned} |x_m - x_n| &= \left|\sqrt{2m + \sin m} - \sqrt{2n + \sin n}\right| = \left|\frac{\left(\sqrt{2m + \sin m}\right)^2 - \left(\sqrt{2n + \sin n}\right)^2}{\sqrt{2m + \sin m} + \sqrt{2n + \sin n}}\right| \\ &= \frac{|2m - 2n + \sin m - \sin n|}{\sqrt{2m + \sin m} + \sqrt{2n + \sin n}} \geq \frac{|2m - 2n + \sin m - \sin n|}{\sqrt{2m + 1} + \sqrt{2n + 1}}, \end{aligned}$$

after using that $\sin x \leq 1$ for all $x \in \mathbb{R}$. Moreover, if $x \geq 1$, we have $2x + 1 \leq 3x$, while the reverse triangle inequality yields

$$\begin{aligned} |2m - 2n + \sin m - \sin n| &\geq |2m - 2m + \sin m| - |\sin n| \geq |2m - 2n + \sin m| - 1 \\ &\geq |2m - 2m| - |\sin m| - 1 \geq 2|m - n| - 2. \end{aligned}$$

When $m \geq n + 1$, the last expression is $2(m - n) - 2 \geq 0$, and we find that

$$|x_m - x_n| \geq \frac{2(m-n) - 2}{\sqrt{3m} + \sqrt{3n}}.$$

At this point, we specify $m = 2n + 1$ (which is $> n + 1$) and get

$$|x_{2n+1} - x_n| \geq \frac{2n}{\sqrt{6n + 2} + \sqrt{3n}} > \frac{2n}{\sqrt{8n} + \sqrt{4n}} = \frac{\sqrt{n}}{\sqrt{2} + 1}.$$

Thus, if we choose $\varepsilon = \frac{1}{3}$ and if $N \in \mathbb{N}$, we can choose $n = N + 1$ and $m = 2n + 1 (> N)$ so that

$$|x_m - x_n| > \frac{\sqrt{n}}{\sqrt{2} + 1} \geq \frac{1}{\sqrt{2} + 1} > \frac{1}{3} = \varepsilon.$$

This proves that the second sequence is not Cauchy either. □

8.2 More on Completeness of the Reals*

As we saw over the last few lectures, most basic properties of the reals and the rationals are alike, but occasionally, a question about sequences may have an answer in $\mathbb{R}$ but not in $\mathbb{Q}$. It turns out that the distinguishing property between the reals and the rationals is the notion of "completeness." We saw one version of it in Axiom 6.9 and another in Theorem 8.3. There are other versions of the competeness property of $\mathbb{R}$ that apply not only to sequences but also to more general sets of real numbers. We will now discuss one of these versions, but first we need to introduce some notation.

Definition 8.2. A non-empty set A of real numbers is called *bounded above* (respectively, *bounded below*) if there is a real number b such that $x \leq b$ for all $x \in A$ (respectively, $x \geq b$ for all $x \in A$). When such a number b exists, it is called an *upper bound* (respectively, a *lower bound*) of A.

A set A that is bounded both above and below is called *bounded*.

Example 8.6. For example, the numbers 2, 3, and π are all upper bounds of the interval $(0, 2]$, because $x \leq 2 < 3 < \pi$ for all $x \in (0, 2]$. However, 1.999 is not an upper bound because the number $a = 2 \in (0, 2]$ is a counterexample to the statement

"$x \leq 1.999$ for all $x \in (0, 2]$."

Similarly, -100, -1, and 0 are lower bounds of $(0, 2]$, because $-100 < -1 < 0 < x$ for all $x \in (0, 2]$. However, 10^{-50} is not a lower bound, because the number $a = 10^{-60} \in (0, 2]$ is a counterexample to the statement

"$x \geq 10^{-50}$ for all $x \in (0, 2]$." □

Exercise 8.7. Is the set $\{x^2 - 3 : x \leq 2\}$ bounded above?

As is evident from the last example, if a set has one upper bound, it has infinitely many upper bounds, and similarly for lower bounds. However, some of those upper and lower bounds are more meaningful than others. This observation motivates the next two notions.

Definition 8.3. Let A be a bounded above set A of real numbers. A real number u is called a *least upper bound*, or a *supremum*, of A if:

(i) $a \leq u$ for all $a \in A$;

(ii) $u \leq b$ for every upper bound b of A.

When it exists, the least upper bound of a set A is denoted $\sup A$.

Definition 8.4. Let A be a bounded below set A of real numbers. A real number l is called a *greatest lower bound*, or an *infimum*, of A if:

(i) $a \geq l$ for all $a \in A$;

(ii) $l \geq b$ for every lower bound b of A.

When it exists, the greatest lower bound of a set A is denoted $\inf A$.

Example 8.8. Let us consider again the interval $(0, 2]$ it is a bounded set, so it makes sense to ask about its sup and inf.

It is not difficult to see that $\sup(0, 2] = 2$. We already checked that 2 is an upper bound of $(0, 2]$: i.e., property (i) in Definition 8.3 holds. Moreover, if b is any upper bound, it must satisfy the inequality $x \leq b$ for all $x \in (0, 2]$. In particular, when $x = 2$, we have $2 \leq b$. Thus, property (ii) in Definition 8.3 also holds.

It is also intuitive that $\inf(0, 2] = 0$. Again, we know from Example 8.6 that 0 is a lower bound of $(0, 2]$, so property (i) in Definition 8.4 holds. The check of property (ii) in that definition, however, is a little trickier. For any number b, we must show that if b is a lower bound of $(0, 2]$, then $b \leq 0$. Since $0 \notin (0, 2]$, this does not follow simply from the definition of a lower bound. Instead, we will look at the contrapositive implication: we will show that if $b > 0$, then b isn't a lower bound. Suppose that $b > 0$. Then we can construct[1] a number $c \in (0, 2]$ with $0 < c < b$, which means that b fails to be a lower bound of $(0, 2]$. □

[1] A simple idea is to try $c = b/2$, but one needs to be careful: b may be large, and that may make this simple choice too big to fit in $(0, 2]$. Try to fix this issue!

Exercise 8.9. Express the results of Problem 5.6 using the terminology Definitions 8.2–8.4.

Exercise 8.10. Let a, b be real numbers with $a < b$, and define

$$A = \{a\}, \quad B = \{a, b\}.$$

Find $\inf A$, $\sup A$, $\inf B$, and $\sup B$.

Exercise 8.11. Prove that $\inf \{x^2 - 3 : x \geq 1\} = -2$.

There are two facts about suprema and infima that intuitively seem obvious, but that is not the case if formal proofs are required. The first is their uniqueness.

Theorem 8.12. *Let A be a non-empty set of real numbers. If A has a supremum/infimum, it is unique.*

Proof. The case of $\sup A$ is Lemma 12.3 in the textbook. □

Then there is the question when a bounded above/below set A has a supremum/infimum. At first, it may seem that the answer to such a question should be "always," but that is only partially true. It turns out that this is true for the real numbers, because they are defined so: this is one of the axioms of $\mathbb{R}$. Sometimes it is stated as the axiom below, and sometimes in the form of our Axiom 6.9. In the latter case, Axiom 8.13 will be a theorem.

Axiom 8.13 (Completeness axiom of $\mathbb{R}$, II)**.** *Every bounded above set of real numbers has a least upper bound in $\mathbb{R}$.*

Once one has the completeness axiom at one's disposal, one can prove the respective fact about existence of infimum of a bounded below set: see Exercise 12.9 in the textbook.

Theorem 8.14. *Every bounded below set of real numbers has a greatest lower bound in $\mathbb{R}$.*

We mentioned already that the completeness axiom fails for $\mathbb{Q}$. Example 8.16 below demonstrates that. However, before we engage with that example to illustrate that, we will state the following result, which is proved in Theorem 12.12 in the textbook.

Theorem 8.15. *If a and b are real numbers such that $a < b$, then there is a rational number r such that $a < r < b$.*

Example 8.16. Consider the set

$$S = \left\{x \in \mathbb{Q}^+ : x^2 < 2\right\}.$$

We will prove that $\sup S = \sqrt{2}$.

First, we check that $\sqrt{2}$ is an upper bound of S. Let $x \in S$. Then $x > 0$ and $x^2 < 2$. Thus,

$$x^2 - 2 < 0 \quad \to \quad \left(x - \sqrt{2}\right)\left(x + \sqrt{2}\right) < 0.$$

Since the sum of positive numbers is positive, we have $x + \sqrt{2} > 0$, so we deduce that

$$x - \sqrt{2} < 0 \quad \to \quad x < \sqrt{2}.$$

This proves that $\sqrt{2}$ is an upper bound of S.

Next, we will show that no real number $y < \sqrt{2}$ is an upper bound of S. This will establish that $\sup S = \sqrt{2}$. Suppose that $y < \sqrt{2}$. If $y < 1$, then 1 is an element of S with $y < 1$, so y is not an upper bound of S. On the other hand, if $1 < y < \sqrt{2}$, then by Theorem 8.15, there is a rational number r such that

$$y < r < \sqrt{2} \quad \to \quad y^2 < r^2 < 2.$$

Therefore, r is an element of S with $y < r$, and so y is again not an upper bound of S. □

Remark. Note that in the above example we showed that S is a set of rational numbers whose unique supremum in $\mathbb{R}$ is $\sqrt{2}$. However, we proved earlier (in Theorem 3.10) that the real number $\sqrt{2}$ is irrational. Therefore, S is a set of rationals that has a supremum in $\mathbb{R}$ but not in $\mathbb{Q}$.

8.3 Additional Examples*

Example 8.17. Let A be a bounded set of real numbers, and define $\alpha = \inf A$ and $\beta = \sup A$. Consider the sets

$$B = \{2a + 5 : a \in A\}, \quad C = \{-3a + 4 : a \in A\}.$$

(a) We will show that $\inf B = 2\alpha + 5$. First, we check that $2\alpha + 5$ is a lower bound of B. Let $x \in B$. Then $x = 2a + 5$ for some $a \in A$. Since α is a lower bound of A, we have

$$\alpha \leq a \quad \to \quad 2\alpha + 5 \leq 2a + 5 = x.$$

This proves that $2\alpha + 5 \leq x$ for all $x \in B$: that is, $2\alpha + 5$ is a lower bound of B.

Next, we will show that if b is a lower bound of B, then $b \leq 2\alpha + 5$. Suppose that b is a lower bound of B:

$$b \leq 2a + 5 \quad \forall a \in A.$$

We can rewrite the last statement as

$$\frac{b - 5}{2} \leq a \quad \forall a \in A,$$

which says that $\frac{b-5}{2}$ is a lower bound of A. Thus, by the definition of $\inf A$,

$$\frac{b - 5}{2} \leq \alpha \quad \to \quad b \leq 2\alpha + 5,$$

as desired.

(b) We will show that $\sup C = -3\alpha + 4$. First, we check that $-3\alpha + 4$ is an upper bound of C. Let $x \in C$. Then $x = -3a + 4$ for some $a \in A$. Since α is a lower bound of A, we have

$$\alpha \leq a \quad \rightarrow \quad -3\alpha + 4 \geq -3a + 4 = x.$$

This proves that $-3\alpha + 4 \geq x$ for all $x \in C$: that is, $-3\alpha + 4$ is an upper bound of C.

Next, we will show that if b is an upper bound of C, then $b \geq -3\alpha + 4$. Suppose that b is an upper bound of C:

$$b \geq -3a + 4 \quad \forall a \in A.$$

We can rewrite the last statement as

$$\frac{b-4}{-3} \leq a \quad \forall a \in A,$$

which says that $\frac{b-4}{-3}$ is a lower bound of A. Thus, by the definition of $\inf A$,

$$\frac{b-4}{-3} \leq \alpha \quad \rightarrow \quad b \geq -3\alpha + 4,$$

as desired. □

Example 8.18. Let A and B be bounded sets of real numbers, and define $\alpha = \sup A$ and $\beta = \sup B$. We will prove that

$$\sup \{a + b : a \in A, b \in B\} = \alpha + \beta.$$

First, we check that $\alpha + \beta$ is an upper bound of the target set

$$S = \{a + b : a \in A, b \in B\}.$$

Let $x \in S$. Then $x = a + b$ for some $a \in A$ and $b \in B$. Since α and β are upper bounds of A and B, we have

$$a \leq \alpha,\ b \leq \beta \quad \rightarrow \quad x = a + b \leq \alpha + \beta.$$

This proves that $x \leq \alpha + \beta$ for all $x \in S$: that is, $\alpha + \beta$ is an upper bound of S.

Next, we will show that if $y < \alpha + \beta$, then y is not an upper bound of S. Suppose that $y < \alpha + \beta$ and define $\epsilon = (\alpha + \beta - y)/2$. Then $\epsilon > 0$ and $\alpha - \epsilon < \alpha$. Since α is the least upper bound of A, we know that $\alpha - \epsilon$ is not an upper bound of A. Thus, there must exist some $a \in A$ such that $\alpha - \epsilon < a$. Similarly, because $\beta - \epsilon < \beta = \sup B$, there must exist some $b \in B$ such that $\beta - \epsilon < b$. Hence,

$$a + b > (\alpha - \epsilon) + (\beta - \epsilon) = \alpha + \beta - 2\epsilon = y.$$

As $a + b \in S$, this proves that y is not an upper bound of S. □

Exercise 8.19. Let A be a bounded set of real numbers, and define $\alpha = \inf A$ and $\beta = \sup A$. Prove the given statements.

(a) $\sup\{3a - 2 : a \in A\} = 2\beta - 3$

(b) $\inf\{-4a : a \in A\} = -4\beta$

(c) $\sup\{a^3 + 1 : a \in A\} = \beta^3 + 1$

(d) $\inf\{a^3 + 1 : a \in A\} = \alpha^3 + 1$

Exercise 8.20. Let A and B be bounded sets of real numbers. Prove that

$$\sup\{a - b : a \in A, b \in B\} = \sup A - \inf B, \quad \inf\{a + b : a \in A, b \in B\} = \inf A + \inf B.$$

Exercise 8.21. Give an example of a bounded set A of real numbers such that:

(a) $\sup\{a^2 + 1 : a \in A\} \neq \beta^2 + 1$, where $\beta = \sup A$;

(b) $\inf\{a^2 + 1 : a \in A\} \neq \alpha^2 + 1$, where $\alpha = \inf A$;

(c) $\sup\{a^2 + 1 : a \in A\} = \beta^2 + 1$, where $\beta = \sup A$;

(d) $\inf\{a^2 + 1 : a \in A\} = \alpha^2 + 1$, where $\alpha = \inf A$;

(e) $\sup\{a^2 + 1 : a \in A\} = \alpha^2 + 1$, where $\alpha = \inf A$;

(f) $\inf\{a^2 + 1 : a \in A\}$ is neither $\alpha^2 + 1$ nor $\beta^2 + 1$, where $\alpha = \inf A$ and $\beta = \sup A$.

Hints on Some Exercises

Exercise 8.4. The answer is embedded in the solution of Example 8.5.

Exercise 8.7. No. If b is any number, show that there is a number y in the given set such that $y > b$. You probably want to reduce to the case $b > 0$ first.

Exercise 8.9. For example, part (b) of Problem 5.6 says that the set $\left\{\dfrac{x-3}{2x+1} : x < -0.5\right\}$ is not bounded above. Part (c) is similar, and part (a) says something about the inf and sup of another set.

Exercise 8.21. The sets in your examples can be chosen fairly simple. For example, in part (a), you may choose $A = \{-3, 1\}$. Then $\beta = 1$ but $\sup\{a^2 + 1 : a \in A\} = \sup\{2, 10\} = 10$. (Any other set $\{a, b\}$ with $a < 0 < b < |a|$ works as well, as do many more complicated examples.)

9 Sets and Set Operations

9.1 Sets: The Basics

The basic notions of a set S and an element s of a set were introduced in passing in §2.1 (and in Chapter 4 of the textbook). Chapter 6 of the textbook recalls those two concepts and explains that these two notions are among the "undefined terms" in mathematics. Next, we will define some of the basic notions related to sets.

First, we will introduce the *empty set*, denoted $\varnothing$, which is a set with no elements: $\varnothing = \{\}$. Further, if A and B are two sets, we say that A is a *subset* of B, and write $A \subseteq B$, if

$$\forall x \in A, x \in B.$$

We say that A is a *proper subset* of B if $A \subseteq B$ but $A \neq B$; in this case, we write $A \subset B$.

Remark. A comment regarding notation. The textbook uses $A \subseteq B$ to denote that A is a subset of B and $A \subset B$ to denote that A is a proper subset of B. We follow the same convention here, but you should be aware for future reference that these are not the only widely adopted mathematical notations for these notions. It is also common to use both $A \subseteq B$ and $A \subset B$ to denote that A is a subset of B and to write $A \subsetneq B$ when A is a proper subset of B.

Exercise 9.1. Which of the following symbolic statements are true?

(a) $x \in \{x, y, z\}$
(b) $x \subseteq \{x, y, z\}$
(c) $\{x\} \in \{x, y, z\}$
(d) $\{x\} \subseteq \{x, y, z\}$
(e) $\{x\} \in \{\{x\}, \{y\}, \{z\}\}$
(f) $\{x\} \subseteq \{\{x\}, \{y\}, \{z\}\}$
(g) $\{x\} \subseteq \{x, \{x\}, \{y\}\}$
(h) $\{x\} \in \{\{x\}, \{y\}, \{y\}\}$

A major point in Chapter 6 of the textbook involves *equality* of sets. The idea of set equality is simple: two sets are equal, if they have the same elements. Unfortunately, except to point quickly why $A \neq B$ for sets like $A = \{1, 2\}$ and $B = \{2, 3\}$, this definition is not very helpful. When the sets involved are more complicated, the following definition of set equality is much more useful.

Definition 9.1. We say that two sets A and B are *equal*, and write $A = B$, if $A \subseteq B$ and $B \subseteq A$.

Since proofs of set equality reduce to proving that one set is a subset of another, Chapter 6 goes over several examples of proving that $A \subseteq B$ for specific sets A and B. Some of those examples may seem unnecessarily complicated to you, but that is not the case: the point of those examples is to introduce you to the general method for proving set inclusion in a familiar setting. This method has several informal names, including the "element method." Here, we include some additional examples.

Example 9.2. Prove that $A = B$, where A and B are the sets

$$A = \big\{n \in \mathbb{Z} : n = 4q + 2 \text{ for some } q \in \mathbb{Z}\big\}, \quad B = \big\{n \in \mathbb{Z} : n = 2m \text{ for some odd } m \in \mathbb{Z}\big\}.$$

We have to prove that two sets are equal, so the proof breaks naturally into two parts.

"$A \subseteq B$". Suppose that $n \in A$. Then

$$n = 4q + 2 = 2(2q + 1) \quad \text{for some } q \in \mathbb{Z}.$$

Hence, there is an odd integer m (namely, $m = 2q+1$) such that $n = 2m$. Therefore, $n \in B$. This establishes that $A \subseteq B$.

"$B \subseteq A$". Suppose that $n \in B$, that is, $n = 2m$ for some odd integer m. By the definition of odd, $m - 1 = 2q$ for some $q \in \mathbb{Z}$. Hence,

$$n = 2m = 2(2q + 1) = 4q + 2,$$

and we have $n \in A$. This establishes that $B \subseteq A$. □

Example 9.3. If A and B are any two sets of real numbers, we define the set

$$A + B = \{a + b : a \in A,\ b \in B\}.$$

If $A = \{1, 2\}$, $B = \{3\}$, $C = (0, 1]$, and $D = [2, 4)$, prove that:

$$A + B = \{4, 5\}, \quad B + C = (3, 4], \quad C + D = (2, 5).$$

We do not need the element method to prove the first two identities:

$$\begin{aligned} A + B &= \{a + b : a \in \{1, 2\},\ b = 3\} = \{a + 3 : a \in \{1, 2\}\} \\ &= \{1 + 3, 2 + 3\} = \{4, 5\}; \\ B + C &= \{b + c : b = 3,\ c \in (0, 1]\} = \{3 + c : 0 < c \leq 1\} \\ &= \{x : 0 + 3 < x \leq 1 + 3\} = (3, 4]. \end{aligned}$$

We will use the element to prove that $C + D = (2, 5)$. First, let $x \in C + D$. Then $x = c + d$ for some real c, d such that $0 < c \leq 1$ and $2 \leq d < 4$. Hence,

$$2 = 0 + 2 \leq 0 + d < c + d \leq 1 + d < 1 + 4 = 5,$$

that is, $x \in (2, 5)$. This proves that $C + D \subseteq (2, 5)$.

Next, suppose that $x \in (2, 5)$. We will show that we can find real c, d such that $0 < c \leq 1$, $2 \leq d < 4$, and $x = c + d$. We will consider two cases:

Case 1: $3 \leq x < 5$. Then $(x - 1) \in [2, 4)$, and we can choose $c = 1$ and $d = x - 1$.

Case 2: $2 < x < 3$. Then $(x - 2) \in (0, 1)$, and we can choose $c = x - 2$ and $d = 2$.

Thus, in all cases $x \in C + D$, which proves that $(2, 5) \subseteq C + D$. □

Example 9.4. Prove that $A \subset B$, where

$$A = \left\{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 \leq 1\right\}, \quad B = \left\{x \in \mathbb{R}^3 : \max(|x_1|, |x_2|, |x_3|) \leq 1\right\}.$$

First, we prove that $A \subseteq B$. Let $x \in A$. Then $x = (x_1, x_2, x_3)$, where $x_1^2 + x_2^2 + x_3^2 \leq 1$. Let $i \in \{1, 2, 3\}$ be chosen so that $\max(|x_1|, |x_2|, |x_3|) = |x_i|$. Then

$$x_i^2 \leq x_1^2 + x_2^2 + x_3^2 \leq 1 \quad \rightarrow \quad -1 \leq x_i \leq 1 \quad \rightarrow \quad |x_i| \leq 1.$$

That is, $\max(|x_1|, |x_2|, |x_3|) \leq 1$, which means that $x \in B$.

To prove that the inclusion is proper, it suffices to find an $x \in B$ such that $x \notin A$. There are many possible examples, but an easy way to construct one is to pick an $x = (1, a, b)$, where $0 < |a| \leq 1$ and $|b| \leq 1$: for example, $x = (1, 0.5, 0.25)$. □

9.2 Sets: Basic Operations

Chapters 6 and 9 in the textbook introduce also several "operations" that we can use to define new sets using existing ones. In all of these definitions, we assume that there is a "universal set" X such that all the sets we consider are subsets of this universal set. In practice, we seldom need worry about the universal set[1].

Definition 9.2. Let A and B be two sets (subsets of a universe X). The *union* of A and B, denoted $A \cup B$, is the set

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

the *intersection* of A and B, denoted $A \cap B$, is the set

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

The *set difference* of A and B, denoted $A \setminus B$, is the set

$$A \setminus B = \{x \in X : x \in A \text{ and } x \notin B\};$$

and the *complement* of A, denoted A^c, is the set

$$A^c = X \setminus A = \{x \in X : x \notin A\}.$$

Chapter 9 of the textbook introduces some further ways to construct new sets from an existing sets in a given universe. These concepts expand the universe as well.

Definition 9.3. Let A and be a set. The *power set* of A, denoted $\mathcal{P}(A)$, is the set of all subsets of A.

[1] The notion of complement (see Definition 9.2) is a notable exception to this remark.

The next definition assumes that you are familiar with the concept of an *ordered pair* (x, y), where x and y are known objects. Recall that (x, y) is essentially the set $\{x, y\}$, where we have designated x as the first component of the pair and y as its second component. In particular, while $\{a, a\} = \{a\}$ (because sets ignore repetitions), we have $(a, a) \neq a$.

Definition 9.4. Let A and B be two sets. Their *Cartesian product*, denoted $A \times B$, is the set

$$A \times B = \{(x, y) : x \in A,\ y \in B\}.$$

Exercise 9.5. **(a)** Let $X = \{\varnothing, \{\varnothing\}, \{1\}\}$. List four sets in $\mathcal{P}(X)$.
(b) Let $X = \mathbb{R}$, $A = [1, 2)$, $B = (0, 3)$, $C = (-\infty, 1]$, and $D = \{2, 3, 5, 7\}$. Express the following in interval notation:

$$A \cup B, \quad A \cup C, \quad B \cap C, \quad A \setminus B, \quad B \setminus A, \quad D \setminus C, \quad A^c \cup (B \cap C), \quad (B \cup D) \cap (A \cup C).$$

Example 9.6. Prove that the set $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ is not a Cartesian product.

Since $D \subset \mathbb{R}^2$, if $D = A \times B$, then A and B must be subsets of $\mathbb{R}$. Thus, we need to show that for any $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$, then $D \neq A \times B$. Suppose that $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$ are such that $D \subseteq A \times B$. Since $(0.99, 0) \in D$, we must have $0.99 \in A$. Also, since $(0, 0.99) \in D$, we must have $0.99 \in B$. It follows that $(0.99, 0.99) \in A \times B$; however,

$$0.99^2 + 0.99^2 = 1.9602 > 1 \quad \to \quad (0.99, 0.99) \notin D.$$

Therefore, $D \neq A \times B$. □

9.3 Simple Proofs Involving Sets

Next, we will practice proving "set identities" between abstract set expressions. Our main focus here is on the method of proof: We rely on Definition 9.1, and so every proof breaks naturally in two (similarly to the solution of Example 9.2). Since we know nothing about the sets in the statements, we must rely only on the definitions of the various set concepts.

In the following proposition, we state some of the properties of sets listed in Theorem 7.4 in the textbook.

Proposition 9.7. *Let A, B, C be subsets of a universal set X. Then:*

(1) $\varnothing \cap A = \varnothing$ *and* $\varnothing \cup A = A$;
(2) $A \cap B \subseteq A$;
(3) $A \subseteq A \cup B$;
(4) $A \cap B = B \cap A$ *and* $A \cup B = B \cup A$;
(5) $(A \cap B) \cap C = A \cap (B \cap C)$ *and* $(A \cup B) \cup C = A \cup (B \cup C)$;
(6) $A \cap A = A = A \cup A$;
(7) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ *and* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;

(8) *If $B \subseteq C$, then $A \cap B \subseteq A \cap C$ and $A \cup B \subseteq A \cup C$.*

Remark. If we make the substitutions $0 \to \varnothing$, $\times \to \cap$, $+ \to \cup$, and $\leq \to \subseteq$ into the above properties, we obtain, respectively:

(1*) $0 \times A = 0$ and $0 + A = A$;
(2*) $A \times B \leq A$;
(3*) $A \leq A + B$;
(4*) $A \times B = B \times A$ and $A + B = B + A$;
(5*) $(A \times B) \times C = A \times (B \times C)$ and $(A + B) + C = A + (B + C)$;
(6*) $A \times A = A = A + A$;
(7*) $A + B \times C = (A + B) \times (A + C)$ and $A \times (B + C) = (A \times B) + (A \times C)$;
(8*) If $B \leq C$, then $A \times B \leq A \times C$ and $A + B \leq A + C$.

If A, B, C are real numbers instead of sets, then (1*), (4*), (5*), the second identity in (7*), and the part of (8*) about addition are known properties of the reals. Furthermore, if $0 \leq A, B \leq 1$, properties (2*), (3*) and the part of (8*) about multiplication are also well-known for numbers. The only "properties" that are unfamiliar to us are (6*) and the first part of (7*), and it is easy to show that those identities are simply not true if A is a number.

Proof of Proposition 9.7. (2) Let $x \in A \cap B$. Then $x \in A$ and $x \in B$; in particular, $x \in A$. Since x was an arbitrary element of $A \cap B$, this proves that $A \cap B \subseteq A$.

(5) We will prove the first identity. First, let $x \in (A \cap B) \cap C$. Then

$$x \in (A \cap B) \text{ and } x \in C \quad \to \quad (x \in A,\ x \in B) \text{ and } x \in C.$$

Thus, x in an element of all three sets A, B, C. We can rearrange this as

$$x \in A \text{ and } (x \in B,\ x \in C) \quad \to \quad x \in A \text{ and } (x \in B \cap C);$$

that is, $x \in A \cap (B \cap C)$. This proves that $(A \cap B) \cap C \subseteq A \cap (B \cap C)$.

Conversely, let $x \in A \cap (B \cap C)$. Then

$$x \in A \text{ and } x \in (B \cap C) \quad \to \quad x \in A \text{ and } (x \in B,\ x \in C).$$

Thus, x in an element of all three sets A, B, C. We can rearrange this as

$$(x \in A,\ x \in B) \text{ and } x \in C \quad \to \quad x \in (A \cap B) \text{ and } x \in C;$$

that is, $x \in (A \cap B) \cap C$. This proves that $A \cap (B \cap C) \subseteq (A \cap B) \cap C$.

(7) This is the first part is Theorem 7.1 in the textbook. □

Hints on Some Exercises

Exercise 9.1. Statements (a), (d), (e), (g), and (h) are true; the other three are false.

Exercise 9.5. **(a)** We have

$$\mathcal{P}(X) = \Big\{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{1\}\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{1\}\}, \{\{\varnothing\}, \{1\}\}, X\Big\}.$$

A list of any four of the eight sets listed above is an acceptable answer.
(b) We have

$$\begin{aligned} A \cup B &= (0,3), & A \cup C &= (-\infty, 2), \\ B \cap C &= (0,1], & A \setminus B &= \varnothing, \\ B \setminus A &= (0,1) \cup [2,3), & D \setminus C &= \{2,3,5,7\}, \\ A^c \cup (B \cap C) &= (-\infty, 1] \cup [2, \infty), & (B \cup D) \cap (A \cup C) &= (0,2). \end{aligned}$$

10 Functions

The notion of a function from one set to another is introduced in Chapter 14 of the textbook as a special case of the notion of a relation, which appears briefly at the end of Chapter 9. The definition in the textbook can be rephrased without recourse to relations as follows.

10.1 Definition and Basic Examples

Definition 10.1. If X and Y are non-empty sets, then a *function f from X to Y*, is a rule f, which can be applied to every element $x \in X$ to select a unique element $y \in Y$, called the *function value at x* and denoted $f(x)$: $y = f(x)$. The set X is called the *domain* of the function and Y—its *codomain*.

Remark. One way to visualize a function is by its *graph*: that is, the set

$$\{(x, y) \in X \times Y : y = f(x)\}$$

in the Cartesian product $X \times Y$. This way of thinking of the graph as a visualization of the abstract rule defining the function is common in calculus and below. In calculus, we then distinguish graphs (of equations) which can be used to define a function from those which cannot be used that way. In more abstract settings, where we cannot really see the graph anyway, we sometimes identify the graph and the function. This is the approach the authors promote in the textbook: when they write $(x, y) \in f$, they mean that the function *is* the graph and the correspondence rule is that for a pair $(x, y) \in f$, the value $f(x)$ of the function is set equal to y, the second component of the pair (x, y).

Woven into the examples of Chapter 14 of the textbook is an important observation: that a "rule" or a formula may appear to define a function, while that is not really true. You are likely unfamiliar with such ambiguities, since they are usually resolved quietly in lower-level math courses. However, in advanced mathematics, you will face such issues regularly, and when you do, you need to know how to recognize and resolve them. When we want to distinguish whether a rule defines a function or fails to, we will often phrase the question as "Is this a well-defined function?" The two possible answers to this question are: "Yes, the function is well-defined." (meaning that the rule does define a function); or "No, this function is not well-defined." (meaning that it is not a function at all).

But what does it mean for a function to be (or not to be) "well-defined"? There are three basic ways in which an attempt to define a function $f : X \to Y$ can fail:

- The rule may fail for some $x \in X$. An example is $f(x) = 1/x$ as a function $f : \mathbb{R} \to \mathbb{R}$. This "function" is undefined when $x = 0$.
- The rule may sometimes select a value $y = f(x)$ that is not in Y. An example is $f(x) = x - 1$ as a function $f : \mathbb{N} \to \mathbb{N}$. This function would well-defined if the chosen codomain was $\mathbb{Z}$ or $\mathbb{R}$, but when the codomain is set to $\mathbb{N}$, $f(0)$ fails: $f(0) = -1 \notin \mathbb{N}$.

- The rule may be ambiguous: for some values of $x \in X$, it may be possible to apply the rule in different ways to arrive at different values for $f(x)$.

The issues in the first two bullets above are usually easy to spot and, if necessary, resolve. The last of these three issues is harder to deal with. When the rule is indeed ambiguous, we can often find an example that demonstrates that the function is not well-defined. The hard part is when a function *is* well-defined even though there is a reason to question whether it is. When that occurs, we need to prove that the rule is unambiguous—that is essentially a uniqueness proof. Below, we include some examples of such proofs.

Example 10.1. Recall that a rational number x can be expressed as a fraction $x = a/b$, where $a, b \in \mathbb{Z}$ and $b \neq 0$. Do the formulas below define functions from $\mathbb{Q}$ to $\mathbb{Q}$?

$$f(a/b) = \frac{a^2 + b^2}{2ab}, \quad g(a/b) = a, \quad h(a/b) = \frac{2ab}{a^2 + b^2}.$$

The first of the three formulas does not define a function, because when $x = 0$, we have $x = 0/1$, but the expression for $f(0/1)$ in undefined.

The second formula is ambiguous: we have $0.4 = 2/5 = 4/10 = 40/100$; however,

$$g(2/5) = 2, \quad g(4/10) = 4, \quad g(40/100) = 40.$$

Thus, this formula cannot be used to define a function.

The final formula does define a function. To prove that, we must prove that:

(i) $h(x)$ is defined for every $x \in \mathbb{Q}$;

(ii) the expression $2ab/(a^2 + b^2)$ is always a rational number;

(iii) if $x = a/b$ and $x = c/d$ are two representations of the same $x \in \mathbb{Q}$, then the expressions $h(a/b)$ and $h(c/d)$ result in the same value for $h(x)$.

It is easy to justify (i) and (ii): since $a, b \in \mathbb{Z}$ and $b \neq 0$, both $2ab$ and $a^2 + b^2$ are also integers and $a^2 + b^2 > 0$; in particular, $a^2 + b^2 \neq 0$. To justify (iii), let a, b, c, d be integers with $b \neq 0$, $d \neq 0$, and $a/b = c/d = x$. Then

$$h(a/b) = \frac{2ab}{a^2 + b^2} = \frac{(2ab)/b^2}{(a^2 + b^2)/b^2} = \frac{2(a/b)}{(a/b)^2 + 1} = \frac{2x}{x^2 + 1}.$$

We see that the value of $h(a/b)$ depends only on the value of the fraction $x = a/b$ and not on the individual values of a and b. In particular, if we repeat the calculation for $h(c/d)$, we get

$$h(c/d) = \frac{2cd}{c^2 + d^2} = \frac{(2cd)/d^2}{(c^2 + d^2)/d^2} = \frac{2(c/d)}{(c/d)^2 + 1} = \frac{2x}{x^2 + 1} = h(a/b),$$

which completes the proof of (iii). □

10.2 Injective, Surjective and Bijective Functions

After introducing the notion of a function in Chapter 14, in Chapter 15 the authors of the textbook introduce three important classes of functions. Recall that the book uses the notation $\text{ran}(f)$ to denote the *range* of the function f: these are the elements $y \in Y$ that are values of f (i.e., $y = f(x)$ for some $x \in X$).

Definition 10.2. A function $f : X \to Y$ is *one-to-one* (or *injective*) if for every $y \in \text{ran}(f)$, there is a unique $x \in X$ with $f(x) = y$: that is,

$$\forall x_1, x_2 \in X, (f(x_1) = f(x_2) \quad \to \quad x_1 = x_2).$$

A function $f : X \to Y$ is *onto* (or *surjective*) if $\text{ran}(f) = Y$: that is,

$$\forall y \in Y, \exists x \in X : y = f(x).$$

A function $f : X \to Y$ that is both one-to-one and onto is called a *bijection* between X and Y.

The textbook develops these notions painstakingly through a series of examples and exercises on pages 158–161: you should read carefully through the examples and work through the exercises. Below, we include a couple of additional examples.

Example 10.2. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f(x, y) = (2x + y, 3x - y + 1).$$

Check whether f is one-to-one and onto.

To check that f is one-to-one, we must show that for all pairs $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$,

$$f(x_1, y_1) = f(x_2, y_2) \quad \to \quad (x_1, y_1) = (x_2, y_2).$$

Let $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ and suppose that $f(x_1, y_1) = f(x_2, y_2)$. This means that

$$(2x_1 + y_1, 3x_1 - y_1 + 1) = (2x_2 + y_2, 3x_2 - y_2 + 1),$$

or equivalently:

$$\begin{cases} 2x_1 + y_1 = 2x_2 + y_2, \\ 3x_1 - y_1 + 1 = 3x_2 - y_2 + 1; \end{cases} \quad \leftrightarrow \quad \begin{cases} 2(x_1 - x_2) + (y_1 - y_2) = 0, \\ 3(x_1 - x_2) - (y_1 - y_2) = 0. \end{cases}$$

Thus,

$$5(x_1 - x_2) = [2(x_1 - x_2) + (y_1 - y_2)] + [3(x_1 - x_2) - (y_1 - y_2)] = 0 + 0 = 0,$$

and

$$5(y_1 - y_2) = 3[2(x_1 - x_2) + (y_1 - y_2)] - 2[3(x_1 - x_2) - (y_1 - y_2)] = 0 - 0 = 0,$$

so we conclude that

$$x_1 - x_2 = y_1 - y_2 = 0 \quad \to \quad x_1 = x_2,\ y_1 = y_2 \quad \to \quad (x_1, y_1) = (x_2, y_2).$$

To check that f is onto, we must check whether given an arbitrary $(a, b) \in \mathbb{R}^2$, we can find a pair $(x, y) \in \mathbb{R}^2$ with $f(x, y) = (a, b)$. Let $(a, b) \in \mathbb{R}^2$. We want to solve

$$(2x + y, 3x - y + 1) = (a, b)$$

for x and y. It is not difficult to see that the solution is

$$x = \frac{a + b - 1}{5}, \quad y = \frac{3a - 2b + 2}{5}.$$

Once we have these values, it is easy to check that they are correct by checking that $f(x, y) = (a, b)$:

$$\begin{aligned} f(x, y) &= \left(2\left(\frac{a + b - 1}{5}\right) + \left(\frac{3a - 2b + 2}{5}\right), 3\left(\frac{a + b - 1}{5}\right) - \left(\frac{3a - 2b + 2}{5}\right) + 1\right) \\ &= \left(\frac{2a + 2b - 2 + 3a - 2b + 2}{5}, \frac{3a + 3b - 3 - 3a + 2b - 2 + 5}{5}\right) = (a, b). \end{aligned}$$

□

Example 10.3. Let $f : [0, \infty)^2 \to [0, \infty)^2$ be given by

$$f(x, y) = (x^2 + y^2, 2xy).$$

Check whether f is one-to-one and onto. If f is not onto, find $\text{ran}(f)$.

Because of the symmetry in the definition of f, it is easy to see that it is not one-to-one: for example,

$$(1, 2) \neq (2, 1), \quad f(1, 2) = (1^2 + 2^2, 2(1)(2)) = (5, 4) = (2^2 + 1^2, 2(2)(1)) = f(2, 1).$$

Let us check next whether f is onto. Let $(a, b) \in [0, \infty)^2$, and consider the vector equation

$$(x^2 + y^2, 2xy) = (a, b).$$

which we want to solve for x and y. This is equivalent to the (non-linear) system

$$\begin{cases} x^2 + y^2 = a, \\ 2xy = b. \end{cases} \tag{$*$}$$

If $b = 0$, the second equation leads to two possibilities: $x = 0$ or $y = 0$. When $x = 0$, the first equation gives

$$y^2 = a \quad \to \quad y = \sqrt{a};$$

and when $y = 0$, the first equation gives

$$x^2 = a \quad \to \quad x = \sqrt{a}.$$

Hence, when (a, b) is of the form $(a, 0)$, we have two possible solutions to $(*)$: $(\sqrt{a}, 0)$ and $(0, \sqrt{a})$ (which collapse to a single solution when $a = 0$).

Suppose now that $b > 0$. Then the second equation in $(*)$ gives

$$x \neq 0, \qquad y = \frac{b}{2x}.$$

Substitution of the latter expression for y into the first equation in $(*)$ gives

$$x^2 + \frac{b^2}{4x^2} = a \quad \to \quad 4x^4 - 4ax^2 + b^2 = 0. \tag{$**$}$$

This is a quadratic equation for x^2, which has no real solutions unless

$$(-4a)^2 - 4(4)b^2 = 16(a^2 - b^2) \geq 0.$$

We can use this observation to construct an example that shows that f is not onto either: we simply need to choose a pair (a, b) with $0 < a < b$. For example, when $(a, b) = (1, 3)$, equation $(**)$ is $4x^4 - 4x^2 + 3 = 0$. This has no real solutions, because

$$4x^4 - 4x^2 + 3 = (2x^2 - 1)^2 + 2 \geq 2 > 0.$$

In order to determine the range of f, we return to equation $(**)$. We must determine for which pairs $(a, b) \in [0, \infty)^2$, $b > 0$, that equation has solutions $x > 0$. Earlier, we saw that we must have

$$0 \leq 16(a^2 - b^2) = 16(a + b)(a - b) \quad \to \quad a - b \geq 0 \quad \to \quad a \geq b.$$

Under this condition, we can solve $(**)$ for x^2 to get

$$x^2 = \frac{4a \pm \sqrt{16(a^2 - b^2)}}{8} = \frac{a \pm \sqrt{a^2 - b^2}}{2}.$$

Since $b^2 > 0$, we have

$$\sqrt{a^2 - b^2} < \sqrt{a^2} = a \quad \to \quad \frac{a - \sqrt{a^2 - b^2}}{2} > \frac{a - a}{2} = 0.$$

So, each of the solutions for x^2 produces two solutions for x:

$$x = \pm\sqrt{\frac{a + \sqrt{a^2 - b^2}}{2}}, \quad x = \pm\sqrt{\frac{a - \sqrt{a^2 - b^2}}{2}};$$

however, only two of these four solutions yield solutions to $(*)$:

$$x_1 = \sqrt{\frac{a + \sqrt{a^2 - b^2}}{2}}, \quad x_2 = \sqrt{\frac{a - \sqrt{a^2 - b^2}}{2}}.$$

We conclude that when $a \geq 0,\ b = 0$ or $a \geq b > 0$, $(*)$ has solutions $(x, y) \in [0, \infty)^2$; and when $b > a \geq 0$, $(*)$ does not have any real solutions. Therefore,

$$\text{ran}(f) = \big\{(a, b) \in \mathbb{R}^2 : 0 \leq b \leq a\big\}.$$

□

Exercise 10.4. In the last example, we showed that when $0 \leq b \leq a$, the equation $f(x, y) = (a, b)$ has two solutions (x_1, y_1) and (x_2, y_2) in the first quadrant. Show that, in fact, $y_1 = x_2$ and $y_2 = x_1$.

Exercise 10.5. Let f be the function from the last example, and let

$$A = \big\{(a, b) \in \mathbb{R}^2 : 0 \leq b \leq a\big\}.$$

Prove that $f|_A$, the function f with its domain restricted to A, is a bijection from A to A.

You may recall the notion of composition of functions from high-school mathematics. It generalizes naturally to functions between arbitrary sets.

Definition 10.3. Let X, Y, Z be sets and let $f : X \to Y$ and $g : Y \to Z$ be functions. The *composition* of g and f, denoted $g \circ f$, is the function $g \circ f : X \to Z$, defined by

$$(g \circ f)(x) = g(f(x)) \quad \forall x \in X.$$

Theorem 10.6. *Let X, Y, Z be sets and let $f : X \to Y$ and $g : Y \to Z$ be functions.*

(1) *If both f and g are one-to-one, then so is $g \circ f$.*

(2) *If both f and g are onto, then so is $g \circ f$.*

(3) *If both f and g are bijections, then so is $g \circ f$. That is, the composition of two bijections is also a bijection.*

Proof. (1) Suppose that f and g are one-to-one and $x_1, x_2 \in X$ are such that

$$(g \circ f)(x_1) = (g \circ f)(x_2) \quad \to \quad g(f(x_1)) = g(f(x_2)).$$

Since g is one-to-one, we have that ($y_1 = f(x_1)$ and $y_2 = f(x_2)$ are elements of Y such that $g(y_1) = g(y_2)$)

$$g(f(x_1)) = g(f(x_2)) \quad \to \quad f(x_1) = f(x_2).$$

Further, since f is one-to-one, we have

$$f(x_1) = f(x_2) \quad \rightarrow \quad x_1 = x_2.$$

Therefore, $g \circ f$ is one-to-one.

(2) Suppose that f and g are onto and $z \in Z$. Since g is onto, there is some $y \in Y$ such that $g(y) = z$, and since f is onto, there is some $x \in X$ such that $f(x) = y$. Thus,

$$(g \circ f)(x) = g(f(x)) = g(y) = z.$$

Therefore, $g \circ f$ is onto. □

11 Equivalent Sets

The notion of equivalence of sets is a generalization of the notion of comparing the numbers of elements of finite sets. Here is the formal definition.

11.1 Basic Examples

Definition 11.1. If A and B are nonempty sets, we say that they are *equivalent* (or *equipotent*), and write $A \approx B$, if there exists a bijection $f : A \to B$.

This concept is introduced in Chapter 21 of the textbook. The central fact in that chapter is Theorem 21.1, which establishes that set equivalence is what is called an "equivalence relation". Since we have not yet discussed equivalence relations in this course, we provide here an alternative statement of Theorem 21.1 that avoids that terminology.

Theorem 11.1. *Let $\mathcal{A}$ be a nonempty collection of sets. Equivalence between elements of $\mathcal{A}$, as defined above, has the following properties:*

(1) *$A \approx A$ for any set $A \in \mathcal{A}$.*

(2) *If $A, B \in \mathcal{A}$ and $A \approx B$, then also $B \approx A$.*

(3) *If $A, B, C \in \mathcal{A}$, with $A \approx B$ and $B \approx C$, then also $A \approx C$.*

Proof. (1) Let $A \in \mathcal{A}$. The function $i_A : A \to A$ defined by $i_A(x) = x$ for all $x \in A$ (this is called the *identity function on A*) is a bijection. So, $A \approx A$.

(2) Let $A, B \in \mathcal{A}$ and $A \approx B$. Then there is a function $f : A \to B$ that is both one-to-one and onto. We will define a function $g : B \to A$ by the following rule:

> For every $y \in B$: if $y = f(x)$ for some $x \in A$, we define $g(y) = x$.

We will show that g is a bijection, and hence, $B \approx A$. That requires that we prove three things: that g is well-defined; that g is one-to-one; and that g is onto.

To show that g is well-defined we need to make three observations. Most importantly, we must check that the above defining rule is unambiguous—that is, that we cannot apply it to the same y in two different ways to produce different values for $g(y)$. For such ambiguity to occur, we would have $x_1, x_2 \in A$ with $x_1 \neq x_2$ and $f(x_1) = f(x_2) = y$. However, this is impossible due to the injectivity of f. Further, we need to verify that g is defined for all $y \in B$. This follows from the surjectivity of f: for every $y \in B$, there is a value $x \in A$ such that $f(x) = y$; hence, $g(y) = x$. Finally, the defining rule of g only selects $x \in A$ as potential values of $g(y)$, so it is impossible for $g(y) \notin A$ to occur. Together, these observations establish that $g : B \to A$ is a well-defined function.

Next, we will show that g is onto. Let $x \in A$. Then $y = f(x)$ is an element of B, and $g(y) = x$, by the defining property of g (and its well-definedness). Hence, $A = \text{ran}(g)$.

Finally, we will show that g is one-to-one. Suppose that $y_1, y_2 \in B$ satisfy $g(y_1) = g(y_2) = x$, say. Then we must have $f(x) = y_1$ and $f(x) = y_2$. Since f is a function, this may only occur if $y_1 = y_2$—which establishes the injectivity of g.

(3) This follows from part (3) of Theorem 10.6. □

Chapter 21 in the textbook contains several examples of proofs that different sets are equivalent, some of them in the form of theorems (e.g., Theorem 21.5). You should study carefully all the material between Example 21.2 and Theorem 21.9. Here, we include a couple more basic examples.

Example 11.2. We have $[0, 1) \approx (-\infty, 3]$.

We need to construct a bijection $f : [0, 1) \to (-\infty, 3]$. We will find a decreasing function f such that

$$f(0) = 3, \quad \lim_{x \to 1^-} f(x) = -\infty.$$

Any function of the form

$$g(x; c) = c + \frac{1}{x - 1}$$

satisfies the limit condition for $x \to 1^-$. If we choose $c = 4$, the function $f(x) = g(x; 4)$ will satisfy also the condition $f(0) = 3$. Thus, we choose

$$f(x) = 4 + \frac{1}{x - 1} = \frac{4x - 3}{x - 1}.$$

Next, we will prove that this function is a bijection from $[0, 1)$ to $(-\infty, 3]$.

The function f is one-to-one, because if $x_1, x_2 \in [0, 1)$ satisfy $f(x_1) = f(x_2)$, we find

$$4 + \frac{1}{x_1 - 1} = 4 + \frac{1}{x_2 - 1} \quad \Longrightarrow \quad \frac{1}{x_1 - 1} = \frac{1}{x_2 - 1} \quad \Longrightarrow \quad x_1 - 1 = x_2 - 1;$$

hence, $x_1 = x_2$.

To show that f is onto, let $y \in (-\infty, 3]$ and consider the real number[1]

$$x = 1 + \frac{1}{y - 4} = \frac{y - 3}{y - 4}.$$

Note that when $y \leq 3$, we have $y - 3 \leq 0$ and $y - 4 < 0$. Thus, $0 \leq x < 1$ and

$$f(x) = 4 + \frac{1}{\dfrac{y - 3}{y - 4} - 1} = 4 + \frac{1}{1/(y - 4)} = 4 + (y - 4) = y.$$

This proves that f is onto. □

The next example is a little trickier and is presented as an exercise.

[1] This is the solution of the equation $f(x) = y$.

Exercise 11.3. Prove that $(0,1] \approx (0,1)$ by showing that the following function is a bijection between these intervals:

$$f(x) = \begin{cases} \frac{1}{n+1} & \text{if } x = \frac{1}{n} \in A; \\ x & \text{if } x \notin A. \end{cases}$$

Here, A is the set $A = \left\{\frac{1}{n} : n \in \mathbb{Z}^+\right\} = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots\right\}$.

Exercise 11.4. Prove that $[0,1] \approx (0,1]$ and $[0,1] \approx (0,1)$.

The last few examples and exercises demonstrate that sometimes proving set equivalence using the definition can be cumbersome. The next theorem can sometimes be used to simplify such proofs. The theorem (in a different form) and its proof appear in Chapter 24 of the textbook, but in this course we will be concerned only with its applications. Thus, we will not discuss the proof of the Cantor–Schröder–Bernstein theorem.

Theorem 11.5 (Cantor–Schröder–Bernstein)**.** *Let A and B be two non-empty sets, and suppose that there exist injections $f : A \to B$ and $g : B \to A$. Then $A \approx B$.*

Example 11.6. We can use the Cantor–Schröder–Bernstein theorem to prove more quickly that $[0,1] \approx (0,1)$ by constructing two injections $f : [0,1] \to (0,1)$ and $g : (0,1) \to [0,1]$. Note that since we do not insist that these function are to be onto, it is easy to come up with many such examples.

We can choose f to be any increasing function such that $0 < f(0) < f(1) < 1$; for example, $f(x) = \frac{1}{3}x + \frac{1}{3}$. It is not difficult to prove that this function is one-to-one. We can define g using the same formula[2], or even simpler, we may take $g(x) = x$, which is also one-to-one. Hence, by the Cantor–Schröder–Bernstein theorem, $[0,1] \approx (0,1)$. □

Example 11.7. We can use the Cantor–Schröder–Bernstein theorem to prove that $\mathbb{N} \approx \mathbb{N}^2$.

It is easy to come up with injections $f : \mathbb{N} \to \mathbb{N}^2$; for example, $f(x) = (x, 0)$ is injective (prove this!). It is trickier to construct a one-to-one function $g : \mathbb{N}^2 \to \mathbb{N}$. If $m, n \in \mathbb{N}$, we define

$$g(m,n) = 2^m 3^n.$$

We will use the fundamental theorem of arithmetic to prove that this function is one-to-one.

Suppose that $(m_1, n_1) \in \mathbb{N}^2$ and $(m_2, n_2) \in \mathbb{N}^2$ are such that $g(m_1, n_1) = g(m_2, n_2)$: that is,

$$2^{m_1} 3^{n_1} = 2^{m_2} 3^{n_2} = k, \quad \text{say.}$$

If $k > 1$, these are two factorizations of k into primes: one factorization contains m_1 2s and n_1 3s, and the other m_2 2s and n_2 3s. Since the fundamental theorem of arithmetic (Theorem 4.9) states that such factorizations are unique, we must have $m_1 = m_2$ and $n_1 = n_2$: i.e., $(m_1, n_1) = (m_2, n_2)$. If $k = 1$, since m_1, m_2, n_1, n_2 are all non-negative, we must have $m_1 = n_1 = m_2 = n_2 = 0$ and we again get that $(m_1, n_1) = (m_2, n_2)$. This proves that g is one-to-one.

Thus, it follows from the Cantor–Schröder–Bernstein theorem that $\mathbb{N} \approx \mathbb{N}^2$. □

[2]However, since the domains of f and g are different, these functions are different, even when they are given by the same formula. Thus, we cannot use f in place of g.

11.2 Cardinality

The idea of equivalent sets lies at the core of the answer to the question "How many elements does a set have?" This is a very basic question for a finite set, but for infinite sets the answer is quite difficult, since a mere "infinite" is often insufficient. The concept of cardinality allows mathematicians to compare infinite sets and declare that two set have "the same number of elements" or that one has "more elements than the other" in a similar way to how we claim without much thought that the sets $\{1, 2, 3\}$ and $\{a, b, c\}$ have the same number of elements, whereas the set $\{\spadesuit, \heartsuit, \diamondsuit, \clubsuit\}$ has more elements than either of those two sets.

For a finite set A, the *cardinality* of A is just a synonym for the number of elements of A. Chapters 21 and 22 of the textbook spend a considerable amount of time going over that. In fact, Chapter 21 defines a finite set as a set A such that $A \approx \{1, 2, \ldots, n\}$ for some $n \in \mathbb{Z}^+$. Then in Theorem 22.6, the textbook proves that there is a unique such n and that unique n is the cardinality (i.e., the number of elements) of A. The main thrust of that narrative is that there is some universal model set—namely, $\{1, 2, \ldots, n\}$ that all other finite sets of n elements are compared with, and a set A has n elements if and only if it is equivalent to this model set.

Chapters 22 and 23 extend the above model for comparing finite sets to infinite sets. The basic message is that we will come up with some infinite model sets, and then we will identify that other infinite sets have the same cardinality as one of those model sets. The first model set is the set of natural numbers $\mathbb{N}$. In Theorem 22.3, the textbook proves that $\mathbb{N}$ is infinite; Chapter 23 calls this set and any set $A \approx \mathbb{N}$ *countably infinite*. Theorems 21.5, 23.8 and 23.11 establish that several other "standard sets" that are also countably infinite, whereas Theorem 23.12 establishes that $\mathbb{R}$ is not countable.[3] We state this last result also here.

Theorem 11.8. *The set of all real numbers is uncountable:* $\mathbb{R} \not\approx \mathbb{N}$.

Remark. Before we engage with the proof of this theorem and the subsequent example, we need to review some facts about decimal representations of real numbers. Recall that a real number $x \in (0, 1)$ has a decimal representation $x = 0.x_1x_2 \cdots x_n \cdots$, where $0 \leq x_j \leq 9$ for all $j \geq 1$. Moreover, if we impose the restriction that we do not allow representations of the form $x = 0.x_1x_2 \cdots x_n000 \cdots$, the decimal representation is unique. Indeed, the only real numbers with multiple decimal representations are the finite decimals, which have two very closely related decimal representations: if $x_n \geq 1$, then

$$x = 0.x_1x_2 \cdots x_{n-1}x_n000 \cdots = 0.x_1x_2 \cdots x_{n-1}x'_n999 \cdots , \quad x'_n = x_n - 1;$$

for example, $0.5000 \cdots = 0.4999 \cdots$. In other words, if we agree to use only *strictly infinite* decimal representations—that is, if we always write a finite decimal in the form $x = 0.x_1x_2 \cdots x_n999 \cdots$, then every real number $x \in (0, 1)$ will have a unique strictly infinite decimal representation:

$$0.x_1x_2 \cdots x_n \cdots = 0.y_1y_2 \cdots y_n \cdots \quad \to \quad x_j = y_j \text{ for all } j \geq 1,$$

[3] You should study all those proofs in the textbook, and we will cover some of them in lecture.

provided that the two representations on the left are strictly infinite.

Proof of Theorem 11.8. We argue by contradiction. Suppose that $\mathbb{R}$ is countable: $\mathbb{R} \approx \mathbb{N}$. As part of the homework, you will show that $\mathbb{R} \approx (0,1)$, and it is not difficult to show that $\mathbb{N} \approx \mathbb{Z}^+$. Hence, by Theorem 11.1, we have also $\mathbb{Z}^+ \approx (0,1)$, so there is a bijection $f : \mathbb{Z}^+ \to (0,1)$.

Since the values $f(n)$ lie in the interval $(0,1)$, we can express each of them as a strictly infinite decimal:

$$\begin{aligned} f(1) &= 0.a_{11}a_{12}a_{13}\cdots a_{1n}\cdots \\ f(2) &= 0.a_{21}a_{22}a_{23}\cdots a_{2n}\cdots \\ &\vdots \\ f(n) &= 0.a_{n1}a_{n2}a_{n3}\cdots a_{nn}\cdots \\ &\vdots \end{aligned}$$

where $0 \le a_{nk} \le 9$ are the decimal digits of $f(n)$. We will use these representations to construct a number $\xi \in (0,1)$ that is *not* a value of f: this will contradict the assumption that f is onto.

We define

$$\xi = 0.c_1c_2c_3\cdots c_n\cdots ; \qquad c_n = \begin{cases} 7 & \text{if } 0 \le a_{nn} \le 4, \\ 3 & \text{if } 5 \le a_{nn} \le 9. \end{cases}$$

Note that this choice[4] ensures that c_n, the nth decimal digit of ξ, is different from the corresponding digit of $f(n)$. Since $\xi \in (0,1)$ and f is onto, there must be some $m \in \mathbb{Z}^+$ such that $f(m) = \xi$: that is,

$$0.a_{m1}a_{m2}a_{m3}\cdots a_{mn}\cdots = 0.c_1c_2c_3\cdots c_n\cdots ,$$

and since these are strictly infinite decimal representations, using our remark, we deduce

$$a_{mj} = c_j \quad \text{for all } j \ge 1.$$

However, when $j = m$, this contradicts the choice of c_m. □

We now see that there are at least two different sizes of infinite sets: the countably infinite sets, equivalent to $\mathbb{N}$, and sets equivalent to $\mathbb{R}$—such as any interval, finite or infinite. (We have seen already that several different intervals are equivalent, and it is not difficult to generalize those examples to show that any two intervals are equivalent to one another and to the whole real line.) The cardinality of $\mathbb{R}$ (or of any interval) is called *cardinality of the continuum*. In our last example, we show that a solid square has the same cardinality as the interval $(0,1)$; from that it will follow that $\mathbb{R}^2$, and indeed any finite-dimensional space $\mathbb{R}^k$, has cardinality of the continuum.

[4] Which is different from the one we made in lecture but works equally well...

Example 11.9. We have $(0,1)^2 \approx (0,1)$.

We can use the Cantor–Schröder–Bernstein theorem to show this. First, it is an easy exercise to show that the function $f : (0,1) \to (0,1)^2$ given by $f(x) = (x,x)$ is one-to-one (check this!). The bulk of the work will go into defining a one-to-one function $g : (0,1)^2 \to (0,1)$. For every $x, y \in (0,1)$, write x and y in terms of their strictly infinite decimal representations:

$$x = 0.x_1x_2\cdots x_n\cdots \qquad \text{and} \qquad y = 0.y_1y_2\cdots y_n\cdots ;$$

we define $g(x,y)$ to have the following decimal representation:

$$g(x,y) = 0.x_1y_1x_2y_2\cdots x_ny_n\cdots .$$

For example, since $\frac{1}{2} = 0.4999\cdots$ and $\frac{1}{3} = 0.3333\cdots$, we have $g(\frac{1}{2}, \frac{1}{3}) = 0.43939393\cdots$. We will prove that g is a one-to-one function.

Suppose that (x,y) and (u,v) are two pairs in $(0,1)^2$ such that $g(x,y) = g(u,v)$. Writing u, v in terms of their strictly infinite decimal representations (similarly to x and y), we find that

$$0.x_1y_1x_2y_2\cdots x_ny_n\cdots = 0.u_1v_1u_2v_2\cdots u_nv_n\cdots . \qquad (*)$$

Note that these two decimal representations are strictly infinite: for example, if the sequence $x_1, y_1, x_2, y_2, \ldots$ contained only zeros starting from some x_j or y_j, then the expansions of both x and y would contain only zeros after x_j and y_j; this would contradict the assumption that x and y are written as strictly infinite decimals. Therefore, $0.x_1y_1x_2y_2\cdots x_ny_n\cdots$ is a strictly infinite decimal expansion, and similarly for $0.u_1v_1u_2v_2\cdots u_nv_n\cdots$. But if the expansions in $(*)$ are strictly infinite, then that equality implies

$$x_1 = u_1, \quad y_1 = v_1, \quad x_2 = u_2, \quad \ldots;$$

thus, x and u have the same decimal expansion, and so do y and v. We conclude that

$$x = u, \quad y = v \qquad \to \qquad (x,y) = (u,v).$$

This proves that g is one-to-one. □

Hints on Some Exercises

Exercise 11.3. You need to show that f is both one-to-one and onto. For each of those subproofs, you will need to consider several cases depending on whether chosen numbers belong to A or not. For example, to prove that f is onto, you must show that for each $y \in (0,1)$, you can find an $x \in [0,1)$ with $f(x) = y$. The argument is different for $y \in A$ and $y \notin A$:

Case 1: $y \notin A$. Then $f(y) = y$, so we can choose $x = y$.

Case 2: $y \in A$, say $y = \frac{1}{n}$. Since $y \neq 1$, we must have $n \geq 2$. Hence, $x = \frac{1}{n-1} \in A$ and

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

To prove that f is one-to-one, we must prove that if $0 < x_1, x_2 \leq 1$ are such that $f(x_1) = f(x_2)$, then $x_1 = x_2$. This proof will break into three cases: both x_1 and x_2 belong to A; one of x_1 or x_2 belongs to A and the other does not; and neither x_1 nor x_2 belongs to A. In the first and third cases, you can use the definition of f to deduce that $x_1 = x_2$; in the remaining case (when only one of x_1 and x_2 is an element of A), you can use the definition of f to show that $f(x_1) \neq f(x_2)$, and so this case is actually impossible.

Exercise 11.4. One possible bijection $g : [0, 1] \to (0, 1]$ is the following. Let $f : (0, 1] \to (0, 1)$ be the function from Exercise 11.3, and define

$$g(x) = \begin{cases} f(x) & \text{if } 0 < x \leq 1, \\ 1 & \text{if } x = 0. \end{cases}$$

Show that this is a bijection.[5] Once we know that $[0, 1] \approx (0, 1]$ and $(0, 1] \approx (0, 1)$, the equivalence $[0, 1] \approx (0, 1)$ follows from Theorem 10.6, part (3).

[5] You don't need to repeat all the work from the last exercise.

12 Relations

At the end of Chapter 9, the textbook introduces the concept of a relation: if A and B are two non-empty sets, then a *relation* from A to B is a subset R of the Cartesian product $A \times B$. When R is a relation from A to B and two elements $a \in A$ and $b \in B$ satisfy $(a, b) \in R$, we say that they are *in relation* and write $a\,R\,b$. Also, when R is a relation from a set A to itself, we usually say that R is a *relation on* A.

In Chapter 10, the authors of the textbook cover several examples of relations. Note that in most of those examples, the relations express relationships between the elements of A and B (or between two elements of A). Many familiar relationships (for which we have special symbols) fall under the abstract concept of a relation: for example, the relationships "is greater than", "divides", "is equal to", "is equivalent to", and "belongs to" ($>$, $|$, $=$, $\approx$, and $\in$, respectively) are all examples of relations. This observation motivates the notation aRb: it generalizes $x < y$ and $x \in \mathbb{R}$.

The utility of the abstract concept of relation (as a subset of $A \times B$) comes from the realization that it allows us to draw analogies between different relations that may at first appear unrelated to one another. For example, the relations $\leq$ on $\mathbb{R}$ and $\subseteq$ on the sets in a given universe turn out to have a lot in common. In light of that analogy, our approach of proving that two sets A and B are equal by showing that $A \subseteq B$ and $B \subseteq A$ is similar to how we sometimes prove that $x = y$ for two numbers by showing that $x \leq y$ and $y \leq x$.

Two important types of relations are known as "equivalence relations" and "order relations," respectively. We discuss those next.

12.1 Equivalence Relations

Definition 12.1. Let R be a relation on a set X. We say that R is an *equivalence relation* on X, if it has the following three properties:

- R is *reflexive*: for all $x \in X$, we have xRx;
- R is *symmetric*: for all $x, y \in X$, if xRy, then also yRx;
- R is *transitive*: for all $x, y, z \in X$, if xRy and yRz, then also xRz.

We saw one equivalence relation in the previous lecture: the equivalence of sets is an equivalence relation. This is the meaning of Theorem 11.1; indeed, this is how that theorem is stated in the textbook. Chapter 10 of the textbook presents several other examples of equivalence relations, some of them in the form of exercises. Below we include a few more.

Example 12.1. Let X be the set of pairs $(x, y) \in \mathbb{Z}^2$, where $y \neq 0$. Define a relation $\sim$ on X by: For $(x, y), (u, v) \in X$, $(x, y) \sim (u, v)$ if $xv = yu$. Prove that $\sim$ is an equivalence relation.

First, let $(x, y) \in X$. Then $xy = yx$, which proves that $(x, y) \sim (x, y)$. Thus, $\sim$ is reflexive.

Next, let $(x,y),(u,v) \in X$, with $(x,y) \sim (u,v)$: that is, we have $xv = yu$. But then, we have also $uy = vx$, which proves that $(u,v) \sim (x,y)$. Thus, $\sim$ is symmetric.

Finally, let $(x,y),(u,v),(p,q) \in X$, with $(x,y) \sim (u,v)$ and $(u,v) \sim (p,q)$. Then $xv = yu$ and $uq = vp$. To prove that $(x,y) \sim (p,q)$, we must show that $xq = yp$. Since $v \neq 0$, we find that

$$xv = yu \quad \Longrightarrow \quad xvq = yuq = yvp \quad \Longrightarrow \quad xq = yp.$$

This proves that $(x,y) \sim (p,q)$. Thus, $\sim$ is transitive. □

Example 12.2. Let X be the set of all Cauchy sequences (x_n) in $\mathbb{Q}$. We define a relation $\sim$ on X by: For $(x_n),(y_n) \in X$, $(x_n) \sim (y_n)$ if $\lim\limits_{n\to\infty}(x_n - y_n) = 0$. We will prove that $\sim$ is an equivalence relation.

First, let (x_n) be any sequence in $\mathbb{Q}$ (Cauchy or not). Then

$$\lim_{n\to\infty}(x_n - x_n) = \lim_{n\to\infty} 0 = 0.$$

In particular, $(x_n) \sim (x_n)$ for any sequence $(x_n) \in X$. Thus, $\sim$ is reflexive.

Next, let $(x_n),(y_n) \in X$, with $(x_n) \sim (y_n)$. Then, by the properties of limits,

$$\lim_{n\to\infty}(x_n - y_n) = 0 \quad \to \quad \lim_{n\to\infty}(y_n - x_n) = 0,$$

which proves that $(y_n) \sim (x_n)$. Thus, $\sim$ is symmetric.

Finally, let $(x_n),(y_n),(z_n) \in X$, with $(x_n) \sim (y_n)$ and $(y_n) \sim (z_n)$: that is,

$$\lim_{n\to\infty}(x_n - y_n) = \lim_{n\to\infty}(y_n - z_n) = 0.$$

By the properties of limits,

$$\begin{aligned}\lim_{n\to\infty}(x_n - z_n) &= \lim_{n\to\infty}[(x_n - y_n) + (y_n - z_n)] \\ &= \lim_{n\to\infty}(x_n - y_n) + \lim_{n\to\infty}(y_n - z_n) = 0,\end{aligned}$$

which proves that $(x_n) \sim (z_n)$. Thus, $\sim$ is transitive. □

Example 12.3. We say that two $n \times n$ matrices A and B are *similar*, and write $A \sim B$, if there is an invertible $n \times n$ matrix P such that $A = PBP^{-1}$. Prove that $\sim$ is an equivalence relation on the set of $n \times n$ matrices.

First, $\sim$ is reflexive, because for any matrix A, we have $IAI^{-1} = IAI = A$, where I is the $n \times n$ identity matrix.

Next, suppose that $A \sim B$. Then $A = PBP^{-1}$ for some invertible matrix P. Recall that if P is invertible, then so is the matrix $Q = P^{-1}$ and $Q^{-1} = P$. Hence,

$$QAQ^{-1} = P^{-1}AP = P^{-1}\big(PBP^{-1}\big)P = \big(P^{-1}P\big)B\big(PP^{-1}\big) = IBI = B.$$

That is, $B \sim A$, which proves that $\sim$ is symmetric.

Finally, suppose that $A \sim B$ and $B \sim C$. Then $A = PBP^{-1}$ and $B = QCQ^{-1}$ for some invertible matrices P, Q. Recall that if P and Q are invertible, then so is the matrix $R = PQ$ and $R^{-1} = Q^{-1}P^{-1}$. Hence,

$$A = PBP^{-1} = P\big(QCQ^{-1}\big)P^{-1} = (PQ)C\big(Q^{-1}P^{-1}\big) = RCR^{-1}.$$

That is, $A \sim C$, which proves that $\sim$ is transitive. □

An important concept related to equivalence relations is that of equivalence classes. For an equivalence relation $\sim$ on a set X, the textbook defines the *equivalence class* of $a \in X$ with respect to $\sim$ (sometimes *modulo* $\sim$) as the set

$$E_a = \big\{x \in X : x \sim a\big\}.$$

Another, much more commonly used notation for equivalence classes is

$$[a]_\sim = \big\{x \in X : x \sim a\big\}.$$

The following theorem appears as Lemma 11.3 in the textbook. It captures a very important property of equivalence classes

Theorem 12.4. *Let $\sim$ be an equivalence relation on a non-empty set X. The equivalence classes $[a]_\sim$, $a \in X$, are mutually disjoint: that is, if $[a]_\sim \cap [b]_\sim \neq \varnothing$, then $[a]_\sim = [b]_\sim$.*

Proof. Suppose that $[a]_\sim \cap [b]_\sim \neq \varnothing$. First, we show that $[a]_\sim \subseteq [b]_\sim$. Let $x \in [a]_\sim$; then $x \sim a$. Since $[a]_\sim \cap [b]_\sim \neq \varnothing$, there is some $y \in X$ such that $y \in [a]_\sim$ and $y \in [b]_\sim$. Hence, $y \sim a$ and $y \sim b$. By the symmetry of $\sim$, we have also $a \sim y$. We can now use the transitivity of $\sim$ to conclude that

$$x \sim a,\ a \sim y,\ y \sim b \quad \Longrightarrow \quad x \sim y,\ y \sim b \quad \Longrightarrow \quad x \sim b.$$

Thus, $x \in [b]_\sim$. This proves that $[a]_\sim \subseteq [b]_\sim$.

Next, note that in the above proof of $[a]_\sim \subseteq [b]_\sim$ we can easily switch the places of a and b to prove that $[b]_\sim \subseteq [a]_\sim$. Hence, $[a]_\sim = [b]_\sim$. □

Example 12.5. Consider the equivalence classes $[(0,1)]_\sim$, $[(2,4)]_\sim$, and $[(-3,5)]_\sim$ for the equivalence relation $\sim$ from Example 12.1. We have

$$[(0,1)]_\sim = \big\{(x,y) \in X : x = 0\big\} = \big\{(0,y) : y \in \mathbb{Z}, y \neq 0\big\}.$$

Similarly, the equivalence class $[(2,4)]_\sim$ contains the pairs $(x,y) \in X$ satisfying the equations

$$4x = 2y \quad \leftrightarrow \quad 2x = y,$$

and the class $[(-3,5)]_\sim$ contains the pairs with $5x = -3y$. For example, five pairs from the equivalence class $[(2,4)]_\sim$ are: $(2,4)$, $(1,2)$, $(-3,-6)$, $(19,38)$, and $(-98,-196)$.

In general, we have

$$\begin{aligned}[(a,b)]_\sim &= \big\{(x,y) \in X : (x,y) \sim (a,b)\big\} \\ &= \big\{(x,y) \in X : xb = ya\big\} \\ &= \big\{(x,y) \in \mathbb{Z}^2 : x/y = a/b\big\},\end{aligned}$$

so the equivalence class $[(a,b)]_\sim$ contains all the pairs ("numerator", "denominator") of fractions x/y equal to the fraction a/b whose numerator and denominator are the components of the pair (a,b) used to generate the equivalence class $[(a,b)]_\sim$. □

12.2 Order Relations*

Order relations are another common type of relation, which not mentioned explicitly in the textbook but appears in several examples. Here is the definitions of two order relations.

Definition 12.2. Let R be a relation on a set X. We say that R is a *partial order relation* (or simply a *partial order*) on X, if it is reflexive, transitive, and *anti-symmetric*: that is, if R has the following property.

For all $x, y \in X$, if xRy and yRx, then $x = y$.

A relation R on a set X is a *strict order* on X, if it is transitive and *irreflexive*: that is, if $(x,x) \notin R$ for any $x \in X$.

Exercise 12.6. The relation $\leq$ on $\mathbb{R}$ and the relation $\subseteq$ on any family of sets are partial orders. The relations $<$ on $\mathbb{R}$ and $\subset$ on any family of sets are strict orders.

Example 12.7. The relation $\mid$ on $\mathbb{Z}^+$ is a partial order. First, for every integer n, we have $n \mid n$ (because $n = 1 \cdot n$); hence, $\mid$ is reflexive.

Next, suppose that $m, n \in \mathbb{Z}^+$ with $m \mid n$ and $n \mid m$. Then $m = np$ and $n = mq$ for some $p, q \in \mathbb{Z}^+$. We get

$$n = mq = (np)q = n(pq) \quad \Longrightarrow \quad pq = 1.$$

Since $p, q \in \mathbb{Z}^+$, this is possible only if $p = q = 1$. Hence, $m = n$ and $\mid$ is anti-symmetric.

Finally, let $k, m, n \in \mathbb{Z}^+$ with $k \mid m$ and $m \mid n$. Then $m = kp$ and $n = mq$ for some $p, q \in \mathbb{Z}^+$. Hence,

$$n = mq = (kp)q = k\underbrace{(pq)}_{\in\mathbb{Z}} \quad \Longrightarrow \quad k \mid n,$$

which proves that $\mid$ is transitive. □

Exercise 12.8. Explain why $\mid$ is not a partial order on $\mathbb{Z}$.

Example 12.9. Define a relation $\prec$ on $\mathbb{R}^n$ by

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \prec \mathbf{y} = (y_1, y_2, \ldots, y_n) \quad \leftrightarrow \quad x_1 \leq y_1,\ x_2 \leq y_2,\ \ldots,\ x_n \leq y_n.$$

It is easy to prove that this is a partial order on $\mathbb{R}^n$. For example, to prove the anti-symmetry of this relation, suppose that $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, with $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{y} \prec \mathbf{x}$. Then $x_i \leq y_i$ and $y_i \leq x_i$ for all $i = 1, \ldots, n$. Since the usual inequlity $\leq$ is anti-symmetric, we deduce that $x_i = y_i$ for all $i = 1, \ldots, n$. Therefore, $\mathbf{x} = \mathbf{y}$, which proves that $\prec$ is anti-symmetric. The proofs that $\prec$ is reflexive and transitive are equally straightforward, since $\leq$ on $\mathbb{R}$ is reflexive and transitive. □

Note that the usual order $\leq$ on $\mathbb{R}$ can be applied to any pair of real numbers: if $x, y \in \mathbb{R}$, then at least one of $x \leq y$ or $y \leq x$ is true. Such order relations are called *linear orders*. On the other hand, when $n \geq 2$, the order $\prec$ on $\mathbb{R}^n$ from the last example fails to be linear: if $\mathbf{x} = (1, 0)$ and $\mathbf{y} = (0, 2)$, both $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{y} \prec \mathbf{x}$ fail (because $1 > 0$ and $2 > 0$, respectively).

Exercise 12.10. Define the relation $\prec$ on $\mathbb{R}^3$ by: $(x_1, x_2, x_3) \prec (y_1, y_2, y_3)$ if:

$$x_1 < y_1; \quad \text{or} \quad x_1 = y_1,\ x_2 < y_2; \quad \text{or} \quad x_1 = y_1,\ x_2 = y_2,\ x_3 < y_3. \tag{L}$$

Prove that $\prec$ is a strict order on $\mathbb{R}^3$. This is called the "lexicographic order" on $\mathbb{R}^3$.

Hints on Some Exercises

Exercise 12.6. You need to show that all four relations are transitive, that $\leq$ and $\subseteq$ are reflexive and anti-symmetric, and that $<$ and $\subset$ are irreflexive. For the two inequalities, $\leq$ and $<$, these are basic properties of inequalities between numbers: for example, if $x \leq y$ and $y \leq z$, then also $x \leq z$, which establishes the transitivity of $\leq$.

To check that $\subseteq$ is reflexive, we observe that $A \subseteq A$ for every set A. The anti-symmetry of $\subseteq$ is the very definition of equality of sets: if $A \subseteq B$ and $B \subseteq A$, then $A = B$ because this is how we defined $A = B$. To prove the transitive property, let A, B, C be three sets such that $A \subseteq B$ and $B \subseteq C$. We want to show that $A \subseteq C$. Let $x \in A$. Since $A \subseteq B$, we have $x \in B$. But then, since $B \subseteq C$, we have $x \in C$. Since $x \in C$ for all x with $x \in A$, we find that $A \subseteq C$. That is, $\subseteq$ is transitive.

To check that $\subset$ is irreflexive is simple: since $A = A$, the proper inclusion $A \subset A$ always fails. To prove the transitive property, let A, B, C be three sets such that $A \subset B$ and $B \subset C$. We want to show that $A \subset C$. This requires proving that $A \subseteq C$ and $A \neq C$. We know that

$$A \subseteq B, \quad B \subseteq C, \quad A \neq B, \quad \text{and} \quad B \neq C.$$

In the previous paragraph, we proved that the first two of these conditions imply that $A \subseteq C$. Since $A \neq B$, there is a $b \in B$ such that $b \notin A$. Moreover, since $B \subseteq C$, we have $b \in C$. That is, $b \in C$ and $b \notin A$, proving that $A \neq C$. Thus, $\subset$ is transitive.

Exercise 12.8. When we consider $|$ on $\mathbb{Z}$, the anti-symmetric property fails. For example

$2 \mid (-2)$ and $(-2) \mid 2$, but $2 \neq -2$.

Exercise 12.10. Since $x_i < x_i$ fails for $i = 1, 2, 3$, condition $(\star)$ with $\mathbf{y} = \mathbf{x}$ fails for all $\vec{x}$. This establishes irreflexivity. To prove that $\prec$ is transitive, let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3$ satisfy $\mathbf{x} \prec \mathbf{y}$ and $\mathbf{y} \prec \mathbf{z}$. We want to show that $\mathbf{x} \prec \mathbf{z}$. We will consider several cases and subcases.

Case 1: $x_1 < y_1$. The condition $\mathbf{y} \prec \mathbf{z}$ means that $y_1 \leq z_1$, so we get $x_1 < z_1$. Hence, $\mathbf{x} \prec \mathbf{z}$.

Case 2: $x_1 = y_1$, $x_2 < y_2$. We split this case in two subcases.

Case 2.1: $y_1 < z_1$. Then $x_1 = y_1 < z_1$, so we get $\mathbf{x} \prec \mathbf{z}$.

Case 2.2: $y_1 = z_1$. Then the condition $\mathbf{y} \prec \mathbf{z}$ means that $y_2 \leq z_2$, so we get $x_1 = y_1 = z_1$ and $x_2 < z_2$. Hence, $\mathbf{x} \prec \mathbf{z}$.

Case 3: $x_1 = y_1$, $x_2 = y_2$, $x_3 < y_3$. We split this case in three subcases.

Case 3.1: $y_1 < z_1$. Then $x_1 = y_1 < z_1$, so we get $\mathbf{x} \prec \mathbf{z}$.

Case 3.2: $y_1 = z_1, y_2 < z_2$. Then $x_1 = y_1 = z_1$ and $x_2 = y_2 < z_2$. Hence, $\mathbf{x} \prec \mathbf{z}$.

Case 3.3: $y_1 = z_1, y_2 = z_2, y_3 < z_3$. Then $x_1 = z_1$, $x_2 = z_2$, and $x_3 < z_3$. Hence, $\mathbf{x} \prec \mathbf{z}$.

13 Congruence Modulo m

The relation "congruence modulo m," introduced in Chapter 27 of the textbook is a very important example of an equivalence relation.

13.1 The Basic Concept

Definition 13.1. Let $m > 1$ be an integer. We say that two integers a and b are *congruent modulo* m, and write $a \equiv b \pmod{m}$, if $m \mid (a - b)$.

Since the textbook leaves the proof of that this relation (Theorem 27.2 on page 301) as a practice problem, we include it here. Note that the proof of this theorem is really another example akin to Examples 12.1–12.3.

Theorem 13.1. *Let $m \in \mathbb{Z}$, with $m > 1$. Then congruence modulo m is an equivalence relation on $\mathbb{Z}$.*

The proof of the theorem, and many other proofs involving congruences rely on the following basic lemma.

Lemma 13.2. *Let $a, b, m \in \mathbb{Z}$, with $m > 1$. Then $a \equiv b \pmod{m}$ if and only if there exists an integer k such that $a = b + mk$.*

Proof. By the definitions of congruence and divisibility,

$$\begin{aligned} a \equiv b \pmod{m} \quad &\leftrightarrow \quad m \mid (a - b) \\ &\leftrightarrow \quad a - b = mk \\ &\leftrightarrow \quad a = b + mk, \end{aligned}$$

for some $k \in \mathbb{Z}$. □

Proof of Theorem 13.1. For every integer x, we have $m \mid (x - x) = 0$. Hence, $x \equiv x \pmod{m}$, which proves that congruence modulo m is reflexive.

Next, suppose that $x, y \in \mathbb{Z}$, with $x \equiv y \pmod{m}$. By the lemma, there is an integer k such that $x = y + mk$. But then, we have also $y = x + ml$, where $l = -k$ is also an integer. Hence, by the lemma, $y \equiv x \pmod{m}$, proving that congruence modulo m is symmetric.

Finally, suppose that $x, y, z \in \mathbb{Z}$, with $x \equiv y \pmod{m}$ and $y \equiv z \pmod{m}$. By the lemma, there are integers k, l such that $x = y + mk$ and $y = z + ml$. It follows that

$$x = y + mk = (z + ml) + mk = z + m\underbrace{(k + l)}_{\in \mathbb{Z}}.$$

Thus, another appeal to the lemma gives that $x \equiv z \pmod{m}$, and this proves that congruence modulo m is transitive. □

Having shown that congruence modulo m is an equivalence relation, we now take a look at its equivalence classes. The equivalence class $[a]$ of an integer a under congruence modulo m is denoted $[a]_m$, that is,

$$[a]_m = \{x \in \mathbb{Z} \mid x \equiv a \pmod{m}\}.$$

The equivalence class $[a]_m$ is called the *congruence class* of a modulo m or the *residue class* of a modulo m. According to Theorem 12.4, the different congruence classes modulo m are disjoint. It would be nice to have a complete list of those congruence classes. The next theorem provides such a list.

Theorem 13.3. *Let $m \in \mathbb{Z}$, with $m > 1$. Then $[0]_m, [1]_m, \ldots, [m-1]_m$ is a complete list of congruence classes modulo m. Moreover, the classes in this list are pairwise distinct.*

We first prove a lemma.

Lemma 13.4. *Let $a, b, m \in \mathbb{Z}$, with $m > 1$. Then $[a]_m = [b]_m$ if and only if $a \equiv b \pmod{m}$.*

Proof. First, suppose that $[a]_m = [b]_m$. Since $a \equiv a \pmod{m}$, we have $a \in [a]_m$, and hence, $a \in [b]_m$. Thus, $a \equiv b \pmod{m}$.

Conversely, suppose that $a \equiv b \pmod{m}$. Then $a \in [b]_m$; of course, we have also $a \in [a]_m$, so $a \in [a]_m \cap [b]_m$. Since $[a]_m \cap [b]_m \neq \varnothing$, it follows from Theorem 12.4 that $[a]_m = [b]_m$. □

Proof of Theorem 13.3. The first claim of the theorem is that the congruence class $[x]_m$ of every integer x appears in the above list. Let $x \in \mathbb{Z}$. By the quotient-remainder theorem, there exist integers q and r such that

$$x = mq + r, \quad 0 \leq r < m.$$

By Lemma 13.2, we have $x \equiv r \pmod{m}$, so $[x]_m = [r]_m$ by the last lemma. Since r is one of the numbers $0, 1, \ldots, m-1$, this proves that $[x]_m$ is one of the congruence classes in the above list.

To prove the second claim of the theorem, we assume that $[a]_m = [b]_m$ for some integers a and b in the list $0, 1, \ldots, m-1$. Then $b \in [a]_m$, so $b \equiv a \pmod{m}$. Thus, by the definition of congruence, $m \mid (a - b)$. However,

$$a - b \leq (m-1) - b \leq m - 1 \quad \text{and} \quad a - b \geq -b \geq -(m-1),$$

that is, $a - b$ is one of the integers $0, \pm 1, \pm 2, \ldots, \pm(m-1)$. Since 0 is the only integer in this list that is divisible by m, we conclude that $a = b$. This proves that all the congruence classes in the list $[0]_m, [1]_m, \ldots, [m-1]_m$ are distinct. □

The set of all congruence classes modulo m is denoted $\mathbb{Z}_m$ and called the *ring of integers modulo m*. (We won't get into a discussion why call it a "ring.") Theorem 13.3 tells us the cardinality of $\mathbb{Z}_m$ is m and

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

Note that we have also

$$\mathbb{Z}_m = \big\{[1]_m, [2]_m, \ldots, [m]_m\big\} \quad \text{and} \quad \mathbb{Z}_m = \big\{[-1]_m, [-2]_m, \ldots, [-m]_m\big\}.$$

Indeed, because $[m]_m = [0]_m$, the first of these sets is the same as $\{[0]_m, [1]_m, \ldots, [m-1]_m\}$, except that $[0]_m$ has been relabeled and moved from first to last place in the list. Similarly, since we have $[-j]_m = [m-j]_m$ for all $j \in \mathbb{Z}$, the set $\{[-1]_m, [-2]_m, \ldots, [-m]_m\}$ simply lists the elements of $\{[0]_m, [1]_m, \ldots, [m-1]_m\}$ in reverse order (and with negative labels). There are a myriad of other representations of $\mathbb{Z}_m$. For example, if m is odd—say, $m = 2k+1$, we have

$$\mathbb{Z}_m = \{[2]_m, [4]_m, \ldots, [2m]_m\},$$

because the congruence classes in the latter set are

$$[2]_m,\ [4]_m,\ \ldots,\ [2k]_m = [m-1]_m,\ [2(k+1)]_m = [m+1]_m = [1]_m,$$
$$[2(k+2)] = [m+3]_m = [3]_m,\ \ldots,\ [2(2k)]_m = [2m-2]_m = [m-2]_m,\ [2m]_m = [m]_m.$$

Note that these are exactly the classes $[1]_m, [2]_m, \ldots, [m]_m$ rearranged so that those with even labels (namely, $[2]_m, [4]_m, \ldots, [m-1]_m$) appear before those with odd labels (namely, $[1]_m, [3]_m, \ldots, [m]_m$).

13.2 Additional Examples

Congruence classes are an excellent source of examples for some of the concepts that we have encountered before. Some such examples appear in Chapter 27 of the textbook—study carefully Example 27.9 and Exercise 27.10 on pages 305–306. Here, we present some additional examples.

Example 13.5. Determine whether the rules below define functions from $\mathbb{Z}_8$ to $\mathbb{Z}_{12}$: for every $x \in \mathbb{Z}$,

$$f([x]_8) = [5x+2]_{12}; \qquad g([x]_8) = [3x-4]_{12}.$$

It is easy to see that f is not a well-defined function, because this rule is ambiguous. For example, $[0]_8 = [8]_8$, but

$$f([0]_8) = [5(0)+2]_{12} = [2]_{12} \neq [42]_{12} = [5(8)+2]_{12} = f([8]_8).$$

This is only one example that shows that it is possible to produce different outputs by applying f to the same congruence class modulo 8. There are many others.

The second rule is very different, because g *is* a well-defined function. Thus, we aren't able to find counterexamples like the above. However, failure to produce a counterexample is not a proof that none exists. Thus, we need to write a formal proof that g is well-defined. To that end...

Let $x, y \in \mathbb{Z}$ be such that $[x]_8 = [y]_8$. To prove that g is well-defined, we must show that the values of $g([x]_8)$ and $g([y]_8)$ are the same. That is, we need to show that

$$[x]_8 = [y]_8 \quad \to \quad [3x-4]_{12} = [3y-4]_{12}.$$

By Lemma 13.2, the assumption $[x]_8 = [y]_8$ yields $x = y + 8k$ for some $k \in \mathbb{Z}$. Hence,

$$3x - 4 = 3(y + 8k) - 4 = 3y + 24k - 4 = (3y - 4) + 12(2k).$$

Since $2k$ is an integer, this equation and Lemma 13.2 give $[3x - 4]_{12} = [3y - 4]_{12}$, and this proves that g is well-defined. □

Since congruence classes modulo m are equivalence classes for an equivalence relation, Theorem 12.4 tells us that the intersection $[a]_m \cap [b]_m$ of two of those is either empty—if they are distinct, or equal to either of them—if they are the same set. What happens if we take the intersection of two congruence classes with different moduli though? It turns out that the intersection $[a]_m \cap [b]_n$ is either the empty set, or a congruence class to a modulus that depends on m and n. The next example shows some of the possibilities.

Example 13.6. Show that

$$[2]_5 \cap [2]_{10} = [2]_{10}, \quad [2]_8 \cap [3]_6 = \varnothing, \quad [1]_8 \cap [3]_6 = [9]_{24}.$$

We will use the "element method" to prove all three set identities. If $x \in [2]_5 \cap [2]_{10}$, then $x \in [2]_{10}$ by the very definition of intersection of sets; hence, $[2]_5 \cap [2]_{10} \subseteq [2]_{10}$. On the other hand, if $x \in [2]_{10}$, then Lemma 13.2 gives, for some $k \in \mathbb{Z}$,

$$x = 2 + 10k = 2 + 5\underbrace{(2k)}_{\in\mathbb{Z}} \quad \to \quad x \in [2]_5.$$

Hence, $x \in [2]_5 \cap [2]_{10}$, proving that $[2]_{10} \subseteq [2]_5 \cap [2]_{10}$.

Next, let $x \in [2]_8 \cap [3]_6$. Since $x \in [2]_8$, by Lemma 13.2, we have $x = 2 + 8k$ for some $k \in \mathbb{Z}$. Similarly, since $x \in [3]_6$, we have $x = 3 + 6l$ for some $l \in \mathbb{Z}$. We conclude that

$$2(1 + 4k) = x = 3 + 6l = 1 + 2(1 + 3l).$$

However, since $4k + 1$ and $1 + 3l$ are integers, the left side of this identity is even, while its right side is odd. This is a contradiction, which proves that $[2]_8 \cap [3]_6 = \varnothing$.

The third identity is slightly more involved. First, let $x \in [9]_{24}$. Then $x = 9 + 24k$ for some $k \in \mathbb{Z}$, and we obtain that

$$x = 9 + 24k = 1 + 8\underbrace{(1 + 3k)}_{\in\mathbb{Z}} \quad \to \quad x \in [1]_8,$$

and also that

$$x = 9 + 24k = 3 + 6\underbrace{(1 + 4k)}_{\in\mathbb{Z}} \quad \to \quad x \in [3]_6.$$

Hence, $x \in [1]_8 \cap [3]_6$, which proves that $[9]_{24} \subseteq [1]_8 \cap [3]_6$.

Conversely, let $x \in [1]_8 \cap [3]_6$. Then there exist integers k, l such that $x = 1 + 8k$ and $x = 3 + 6l$. We want to show that x can then be expressed in the form $x = 9 + 24n$ for

some integer n. To get there, we apply the quotient-remainder theorem to write $k = 3q + r$ with some integers q, r and the remainder $r \in \{0, 1, 2\}$. Substitution into the existing representations of x then gives

$$3 + 6l = x = 1 + 8k = 1 + 24q + 8r \quad \rightarrow \quad 6\underbrace{(l - 4q - r)}_{\in \mathbb{Z}} = 2r - 2.$$

The left side of the last equation is an integer multiple of 6, while its right side can take on only three values: $0, \pm 2$. These can only equal each other if $r = 1$ (and $l - 4q - r = 0$). Hence,

$$x = 1 + 8k = 1 + 8(3q + 1) = 9 + 24q \quad \rightarrow \quad x \in [9]_{24},$$

and this proves that $[1]_8 \cap [3]_6 \subseteq [9]_{24}$. □

13.3 Modular Arithmetic

The major point of Example 27.9 in the textbook (and of a related homework problem) is that given two congruence classes modulo m, we can "add" or "multiply" them as follows:

$$[a]_m \oplus [b]_m = [a + b]_m, \quad [a]_m \odot [b]_m = [ab]_m.$$

Moreover, we can "raise congruence classes to a power:"

$$[a]_m^k = \underbrace{[a]_m \odot [a]_m \odot \cdots \odot [a]_m}_{k \text{ copies}}.$$

We can use this to evaluate *efficiently* remainders modulo m.

Example 13.7.

Evaluate the remainders of $17^2, 17^5, 17^8, 17^{20}$ after division by 100.

The remainder r of x after division by 100 satisfies $[x]_{100} = [r]_{100}$ and also $0 \leq r < 100$, so we can find r by identifying $[x]_{100}$ among the classes $[0]_{100}, [1]_{100}, \ldots, [99]_{100}$.

We have $[17^2]_{100} = [289]_{100} = [89]_{100}$, so the first remainder is 89.

Since $[17^2]_{100} = [89]_{100} = [-11]_{100}$, we further have

$$[17^4]_{100} = [17^2]_{100} \odot [17^2]_{100} = [-11]_{100} \odot [-11]_{100} = [121]_{100} = [21]_{100},$$

and

$$[17^5]_{100} = [17^4]_{100} \odot [17]_{100} = [21]_{100} \odot [17]_{100} = [357]_{100} = [57]_{100}.$$

Hence, the second remainder is 57.

Further, we have

$$[17^8]_{100} = [17^4]_{100} \odot [17^4]_{100} = [21]_{100} \odot [21]_{100} = [441]_{100} = [41]_{100},$$

and

$$\begin{aligned}[17^{20}]_{100} &= [17^8]_{100} \odot [17^8]_{100} \odot [17^4]_{100} = [41]_{100} \odot [41]_{100} \odot [21]_{100} \\ &= [1681]_{100} \odot [21]_{100} = [-19]_{100} \odot [21]_{100} = [-399]_{100} = [1]_{100}.\end{aligned}$$

Hence, the last two remainders are 41 and 1, respectively. □

Exercise 13.8. Find the last two decimal digits of 17^{2026}.

Remark. We can combine modular arithmetic with the decimal representation of integers to explain why the divisibility tests you know from grade school work and also to obtain new ones. The next example illustrates the basic idea behind such divisibility tests. For its purposes, recall that the decimal representation $(d_k \cdots d_2 d_1 d_0)_{10}$, where $d_0, d_1, \ldots, d_k$ are decimal digits, of an integer n means that

$$n = d_0 + d_1 \times 10 + d_2 \times 10^2 + \cdots + d_k \times 10^k.$$

For example, $n = (2026)_{10}$ (or as we write more commonly $n = 2026$) means that

$$n = 6 + 2 \times 10 + 0 \times 100 + 2 \times 1000.$$

Example 13.9. Prove that an integer n with a decimal representation $n = (d_k \cdots d_1 d_0)_{10}$ is divisible by 9 if and only if its sum of digits $d_0 + d_1 + \cdots + d_k$ is.

Since the question whether 9 divides n or not is equivalent to the question whether $[n]_9 = [0]_9$, we can use the decimal representation of n to rephrase this question. We have

$$\begin{aligned}[n]_9 &= \left[d_0 + d_1 \times 10 + d_2 \times 10^2 + \cdots + d_k \times 10^k\right]_9 \\ &= [d_0]_9 \oplus \left([d_1]_9 \odot [10]_9\right) \oplus \left([d_2]_9 \odot [10^2]_9\right) \oplus \cdots \oplus \left([d_k]_9 \odot [10^k]_9\right).\end{aligned}$$

Further, using that $10 \equiv 1 \pmod 9$, we have $[10]_9 = [1]_9$. More generally, if $j \geq 2$,

$$[10^j]_9 = \underbrace{[10]_9 \odot [10]_9 \odot \cdots \odot [10]_9}_{j \text{ copies}} = \underbrace{[1]_9 \odot [1]_9 \odot \cdots \odot [1]_9}_{j \text{ copies}} = [1^j]_9 = [1]_9.$$

Hence,

$$\begin{aligned}[n]_9 &= [d_0]_9 \oplus \left([d_1]_9 \odot [1]_9\right) \oplus \left([d_2]_9 \odot [1]_9\right) \oplus \cdots \oplus \left([d_k]_9 \odot [1]_9\right) \\ &= [d_0 + d_1 \cdot 1 + d_2 \cdot 1 + \cdots + d_k \cdot 1]_9 = [d_0 + d_1 + \cdots + d_k]_9.\end{aligned}$$

Therefore,

$$9 \mid n \quad \leftrightarrow \quad [0]_9 = [n]_9 = [d_0 + d_1 + \cdots + d_k]_9 \quad \leftrightarrow \quad 9 \mid (d_0 + d_1 + \cdots + d_k).$$

□

Exercise 13.10. Use the decimal representation of n in a similar manner to Example 13.9 to prove that:

(a) $n = (d_k \cdots d_2 d_1 d_0)_{10}$ is divisible by 5 if and only if its units digit d_0 is 0 or 5;
(b) $n = (d_k \cdots d_2 d_1 d_0)_{10}$ is divisible by 4 if and only if the two-digit number $(d_1 d_0)_{10}$ is;
(c) $n = (d_k \cdots d_2 d_1 d_0)_{10}$ is divisible by 25 if and only if $(d_1 d_0)_{10}$ is one of $00, 25, 50$, or 75;
(d) $n = (d_k \cdots d_2 d_1 d_0)_{10}$ is divisible by 8 if and only if $(d_2 d_1 d_0)_{10}$ is;
(e) $n = (d_k \cdots d_2 d_1 d_0)_{10}$ is divisible by 11 if and only if the sum $d_0 - d_1 + d_2 - d_3 + \cdots$ is.

13.4 The GCD and Modular Inverses

Part of Chapter 27 in the textbook deals with the following question: If $m > 1$ and $[a]_m \in \mathbb{Z}_m$ is a congruence class modulo m, then is there another class $[x]_m$ such that $[a]_m \odot [x]_m = [1]_m$? If such a class $[x]_m$ exists, then it is called a *modular inverse* of $[a]_m$. It turns out that the answer to that question, given in Corollary 27.11 on page 306, depends on the greatest common divisor of a and the modulus m, denoted $\gcd(a, m)$. For that reason, we must get to know greatest common divisors.

There are two common definitions of the greatest common divisor (which you may have seen in the past as the "greatest common factor"). The following is the definition that you are more likely to have seen in the past.

Definition 13.2. Let a, b be integers, not both zero. Then the *greatest common divisor* of a and b, denoted $\gcd(a, b)$, is any integer $d \in \mathbb{Z}^+$ such that:

(i) $d \mid a$ and $d \mid b$;

(ii) if c is an integer such that $c \mid a$ and $c \mid b$, then $c \leq d$.

The textbook, however, defines $\gcd(a, b)$ slighly differently.

Definition 13.3. Let a, b be integers, not both zero. Then the *greatest common divisor* of a and b, denoted $\gcd(a, b)$, is any integer $d \in \mathbb{Z}^+$ such that:

(i) $d \mid a$ and $d \mid b$;

(ii) if c is an integer such that $c \mid a$ and $c \mid b$, then $c \mid d$.

Quick examination reveals that the two definitions differ ever so slightly in their choice of condition (ii). The version of (ii) in Definition 13.2 says that if c is a common divisor of a and b, then $c \leq d$. In other words, $\gcd(a, b)$ is greater than any other common divisor of a and b. The version of (ii) in Definition 13.3, on the other hand, says that any other common divisor of a and b must divide $\gcd(a, b)$. It turns out that the two definitions are equivalent, but we won't dwell on that here. Instead, we will simply acknowledge this fact and will accept Definition 13.3 as our working definition of the gcd. You may wonder why though. The answer is that while the other definition may feel more natural, it turns out to be less helpful when we try to prove mathematical facts about gcd's.

We need to figure out two things about gcd's. First, we need to figure out how to calculate $\gcd(a, b)$ for given integers a and b. The second issue we want to address is how to make use of $\gcd(a, b)$ in proofs involving a and b. Let us turn to the first of these issues. Its resolution depends on the following lemma.

Lemma 13.11. *If a, b, q, r are integers, with $b > 0$ and $a = bq + r$, then $\gcd(a, b) = \gcd(r, b)$.*

Proof. Let $d = \gcd(a, b)$. By condition (i) of the definition, we have that $d \mid a$ and $d \mid b$; hence, $a = da_1$ and $b = db_1$ for some $a_1, b_1 \in \mathbb{Z}$. We deduce that

$$r = a - bq = da_1 - (db_1)q = d\underbrace{(a_1 - b_1 q)}_{\in\mathbb{Z}} \quad \Longrightarrow \quad d \mid r.$$

Since we already knew that $d \mid b$, this proves that d is a common divisor of r and b.

Next, suppose that $c \mid r$ and $c \mid b$; hence, $r = cr_1$ and $b = cb_1$ for some $r_1, b_1 \in \mathbb{Z}$. We deduce that

$$a = bq + r = (cb_1)q + cr_1 = c\underbrace{(b_1 q + r_1)}_{\in\mathbb{Z}} \quad \Longrightarrow \quad c \mid a.$$

Since we already knew that $c \mid b$, this proves that c is a common divisor of a and b. Since $d = \gcd(a, b)$, it follows that $c \mid d$.

We have shown that d is a common divisor of r and b such that every common divisor c of r and b satisfies $c \mid d$. Definition 13.3 then gives $d = \gcd(r, b)$. □

The above lemma is at the core of the Euclidean algorithm described in Problem 27.20 in the textbook.

Example 13.12. Let us apply Lemma 13.11 to calculate $\gcd(1547, 527)$.

In this soluition, we write simply (a, b) instead of $\gcd(a, b)$. Using repeatedly the lemma, we find

$$\begin{aligned}(1547, 527) &= (2 \cdot 527 + 493, 527) = (493, 527)\\ &= (493, 1 \cdot 493 + 34) = (493, 34)\\ &= (14 \cdot 34 + 17, 34) = (17, 34) = 17.\end{aligned}$$

□

In proofs involving gcd's we often want to connect $\gcd(a, b)$ directly to a and b. The following theorem (this is Theorem 27.6 in the textbook) tends to be very helpful to that end.

Theorem 13.13. *If a, b are integers, not both zero, then $\gcd(a, b)$ equals the least positive integer d which can be represented as $d = ax + by$ for some integers x, y.*

Corollary 13.14. *If a, b are integers, then $\gcd(a, b) = 1$ if and only if $ax + by = 1$ for some integers x, y.*

Proof. If $\gcd(a, b) = 1$, then such integers x, y exist by Theorem 13.13. Conversely, if such x and y exist, then 1 is a positive integer of the form $ax + by$. Since there is no smaller positive integer, 1 must be the least integer of this form—and that equals $\gcd(a, b)$ by the theorem. □

Corollary 13.15. *If* $\gcd(a,b) = 1$ *and* $a \mid bc$*, then* $a \mid c$*.*

Proof. By Theorem 13.13, there exist $x, y \in \mathbb{Z}$ such that $ax + by = 1$. We deduce that

$$c = c(1) = c(ax + by) = a(cx) + (bc)y = a(cx) + (ak)y = a\underbrace{(cx + ky)}_{\in\mathbb{Z}} \quad \Longrightarrow \quad a \mid c,$$

for some integer k. (Such a k exists, because $a \mid bc$ by assumption.) □

Corollary 13.16. *If* $\gcd(a,b) = 1$ *and* $\gcd(a,c) = 1$*, then* $\gcd(a,bc) = 1$*.*

Proof. Using the hypotheses $\gcd(a,b) = \gcd(a,c) = 1$ and Theorem 13.13, we can find integers x, y, u, v such that

$$ax + by = 1, \quad au + cv = 1.$$

Hence,

$$1 = (ax + by)(au + cv) = a(aux + cvx + buy) + (bc)(vy).$$

Since $m = aux + cvx + buy$ and $n = vy$ are integers, it follows that 1 is an integer of the form $am + (bc)n$. It must therefore be the least integer of this form: that is, $1 = \gcd(a, bc)$. □

Corollary 13.17. *If* $\gcd(a,b) = 1$*,* $a \mid c$ *and* $b \mid c$*, then* $ab \mid c$*.*

Recall the definitions of a prime number and an irreducible integer from the beginning of the course. Back then, we showed that all primes are irreducible and stated that each irreducible is also prime, but we did not prove the latter. We can prove that fact easily using Theorem 13.13 and its corollaries.

Corollary 13.18. *If* $p > 1$ *is irreducible, then* p *is prime.*

Proof. Suppose that $p > 1$ is irreducible and a, b are integers such that $p \mid ab$. Since p is irreducible, its only positive divisors are 1 and p, so $\gcd(p,a) = p$ or 1. We will consider those two cases separately.

Case 1: $\gcd(p,a) = p$. Then, we must have $p \mid a$, by the definition of gcd.

Case 2: $\gcd(p,a) = 1$. Then by Corollary 13.15 with $a = p$, $b = a$ and $c = b$, we have $p \mid b$.

Therefore, in all cases $p \mid a$ or $p \mid b$: that is, p has property (P) from Definition 2.8. □

The above corollaries represented a slight detour to illustrate how we use Theorem 13.13 to prove some basic facts about divisibility. Its main application to modular arithmetic, however, comes through its applications to finding "modular inverses," also called "modular reciprocals." Recall that the textbook introduces the *reciprocal modulo* m of a as any solution x of the congruence $ax \equiv 1 \pmod{m}$. Together Corollary 27.11 and Lemma 27.12 in the textbook give the following result.

Theorem 13.19. *Let $m > 1$ be an integer. If $\gcd(a, m) = 1$, then there exists a unique congruence class $[b]_m$ such that $[a]_m \odot [b]_m = [1]_m$. Moreover, $[b]_m = [u]_m$ where u, v are any pair of integers with $au + mv = 1$.*

Example 13.20. Find the reciprocal of 7 modulo 11.

Since $7(-3) = -21 \equiv 1 \pmod{11}$, we have $[7]_{11} \odot [-3]_{11} = [1]_{11}$, and the reciprocal class of $[7]_{11}$ is $[-3]_{11} = [8]_{11}$. □

The above solution was quick, but unfortunately amounted to a "lucky guess." When the modulus is larger than 20 or so, guesswork becomes inefficient. The next couple of examples show you how to apply the Euclidean algorithm, explained in Problem 27.20 in the textbook, to answer such questions in a more methodical fashion.

Example 13.21. Find the reciprocal of 119 modulo 239 and 127 modulo 256.

We begin with the reciprocal modulo 239. We want to find integers u, v such that $119u + 239v = 1$. If we do, then $[u]_{239}$ will be the modular reciprocal we seek. Let us apply the quotient-remainder theorem to 239 and 119: we get

$$239 = 119 \cdot 2 + 1 \quad \leftrightarrow \quad 1 = 119(-2) + 239(1).$$

Thus, we can use $u = -2$ and $v = 1$, and we find that the reciprocal of 119 modulo 239 is $[-2]_{239} = [237]_{239}$.

Next, we consider the reciprocal of 127 modulo 256. We want to find integers u, v such that $127u + 256v = 1$. If we do, then $[u]_{256}$ will be the modular reciprocal we seek. Again, we apply the quotient-remainder theorem to 256 and 127: we get $256 = 2(127) + 2$, which we can use to rewrite the original equation for u, v as

$$1 = 127u + [2(127) + 2]v = 127(u + 2v) + 2v = 127w + 2v, \quad \text{say.}$$

Note that if we find v, w that solve the last equation, we immediately get also a solution u, v of the original equation $119u + 256v = 1$: v stays the same and $u = w - 2v$.

We can approach the equation $127w + 2v = 1$ in the same way: the quotient-remainder theorem for 127 and 2 yields

$$127 = 2 \cdot 63 + 1 \quad \leftrightarrow \quad 1 = 127(1) + 2(-63).$$

Thus,

$$(v, w) = (-63, 1) \quad \rightarrow \quad (u, v) = (1 - 2(-63), -63) = (127, -63),$$

and we find that the reciprocal of 127 modulo 256 is $[127]_{256}$. □

In the previous example, we used the Euclidean algorithm "in reverse" in two very fortunate situations. We can do this in general, but we usually need to go through multiple iterations. In such situations, doing repeatedly what we did in the second part of the last example is too time consuming. The next example demonstrates an alternative bookkeeping.

Example 13.22. Find the reciprocal of 119 modulo 256.

We first compute $\gcd(256, 119)$ using the Euclidean algorithm:

$$\begin{aligned}(256, 119) &= (119 \cdot 2 + 18, 119) = (18, 119)\\ &= (18, 18 \cdot 6 + 11) = (18, 11)\\ &= (11 \cdot 1 + 7, 11) = (7, 11)\\ &= (7, 7 \cdot 1 + 4) = (7, 4)\\ &= (4 \cdot 1 + 3, 4) = (3, 4) = 1.\end{aligned}$$

Take note of our quotient-remainder results, but solved for the remainder:

$$\mathbf{18} = \mathbf{256} - \mathbf{119}(2), \quad \mathbf{11} = \mathbf{119} - \mathbf{18}(6), \quad \mathbf{7} = \mathbf{18} - \mathbf{11}, \quad \mathbf{4} = \mathbf{11} - \mathbf{7}, \quad \mathbf{3} = \mathbf{7} - \mathbf{4}.$$

Using these in reverse order, we now have

$$\begin{aligned}1 &= \mathbf{4} - \mathbf{3} = \mathbf{4} - (\mathbf{7} - \mathbf{4})\\ &= \mathbf{4}(2) - \mathbf{7} = (\mathbf{11} - \mathbf{7})(2) - \mathbf{7}\\ &= \mathbf{11}(2) - \mathbf{7}(3) = \mathbf{11}(2) - (\mathbf{18} - \mathbf{11})(3)\\ &= \mathbf{11}(5) - \mathbf{18}(3) = (\mathbf{119} - \mathbf{18}(6))(5) - \mathbf{18}(3)\\ &= \mathbf{119}(5) - \mathbf{18}(33) = \mathbf{119}(5) - (\mathbf{256} - \mathbf{119}(2))(33)\\ &= \mathbf{119}(71) - \mathbf{256}(33) = \mathbf{119}(71) + \mathbf{256}(-33).\end{aligned}$$

Hence, $[71]_{256}$ is the reciprocal of $[119]_{256}$: $71 \cdot 119 \equiv 1 \pmod{256}$. □

Hints on Some Exercises

Exercise 13.8. The last two decimal digits of x are the remainder r of x when divided by 100: that is, the number $r \in \{0, 1, \ldots, 99\}$ such that $[x]_{100} = [r]_{100}$. We know from Example 13.7 that $[17^{20}]_{100} = [1]_{100}$, $[17^4]_{100} = [21]_{100}$, and $[17]^2_{100} = [-11]_{100}$. Hence,

$$\begin{aligned}[17^{2026}]_{100} &= [17^{20 \cdot 101 + 6}]_{100} = [17^{20}]^{101}_{100} \odot [17]^4_{100} \odot [17]^2_{100}\\ &= [1]^{101}_{100} \odot [21]_{100} \odot [-11]_{100} = [-231]_{100} = [69]_{100}.\end{aligned}$$

Thus, the last two decimal digits of 17^{2026} are 69.

Exercise 13.10. **(a)** Arguing similarly to Example 13.9, we see that

$$5 \mid n \quad \leftrightarrow \quad [0]_5 = [n]_5 = \left[d_0 + d_1 \times 10 + \cdots + d_k \times 10^k\right]_5 = [d_0]_5,$$

since $d_1 \times 10 + \cdots + d_k \times 10^k$ is divisible by 5. So,

$$5 \mid n \quad \leftrightarrow \quad [0]_5 = [d_0]_5 \quad \leftrightarrow \quad 5 \mid d_0,$$

and the only decimal digits divisible by 5 are 0 and 5.

(d) Arguing similarly to Example 13.9, we see that

$$8 \mid n \quad \leftrightarrow \quad [0]_8 = [n]_8 = \left[d_0 + d_1 \times 10 + \cdots + d_k \times 10^k\right]_8 = [d_0 + d_1 \times 10 + d_2 \times 10^2]_8,$$

since $d_3 \times 10^3 + \cdots + d_k \times 10^k$ is divisible by 8. So,

$$8 \mid n \quad \leftrightarrow \quad [0]_8 = [d_0 + d_1 \times 10 + d_2 \times 10^2]_8 \quad \leftrightarrow \quad 8 \mid (d_2 d_1 d_0)_{10}.$$

(e) Since $10 \equiv -1 \pmod{11}$, we get $[10^j]_{11} = [(-1)^j]_{11}$ for all $j \geq 1$. Hence, arguing similarly to Example 13.9, we see that

$$[n]_{11} = \left[d_0 + d_1 \times 10 + \cdots + d_k \times 10^k\right]_{11} = [d_0 - d_1 + d_2 - d_3 + \cdots + (-1)^k d_k]_{11}.$$

Thus,

$$\begin{aligned} 11 \mid n \quad &\leftrightarrow \quad [0]_{11} = [n]_{11} = [d_0 - d_1 + d_2 - d_3 + \cdots]_{11} \\ &\leftrightarrow \quad 11 \mid (d_0 - d_1 + d_2 - d_3 + \cdots). \end{aligned}$$

14 Families of Sets

Sometimes, we need to work with sets whose elements are themselves sets, grouped by some common property. It is common to refer to such sets as *families* or *collections* of sets. For example, the power set $\mathcal{P}(A)$ of a set A is a family—the family of all subsets of A. Similarly, the set of all finite open intervals in $\mathbb{R}$,

$$\mathcal{I} = \{(a, b) : -\infty < a < b < \infty\},$$

is a family of sets.

When we work with collections of sets, the elements of the collection are often indexed by the elements of another set: say, the collection $\mathcal{A}$ may contain a set A_i for every element i of some known set I (the "index set"). Then we may describe the collection $\mathcal{A}$ as $\mathcal{A} = \{A_i : i \in I\}$ and call it an *indexed family*. For example, the family $\mathcal{I}$ of all open intervals above is indexed by pairs of real numbers. Another example of an indexed collection of sets, indexed by the positive integers is

$$\mathcal{A} = \left\{(0, 1/n] : n \in \mathbb{Z}^+\right\}.$$

14.1 Unions and Intersections of Families of Sets

Definition 14.1. Let $\mathcal{A} = \{A_i : i \in I\}$ be an indexed family of subsets of an universal set X. The *union of (the sets in)* $\mathcal{A}$, denoted $\cup\mathcal{A}$ or $\cup_{i\in I} A_i$, is given by

$$\bigcup_{i\in I} A_i = \left\{x \in X : x \in A_i \text{ for some } i \in I\right\}.$$

When $I \neq \varnothing$, the *intersection of (the sets in)* $\mathcal{A}$, denoted $\cap\mathcal{A}$ or $\cap_{i\in I} A_i$, is given by

$$\bigcap_{i\in I} A_i = \left\{x \in X : x \in A_i \text{ for all } i \in I\right\}.$$

Example 14.1. Consider the family of the intervals $A_n = (0, 1/n]$, $n \in \mathbb{Z}^+$. Then

$$\bigcup_{n\in\mathbb{Z}^+} A_n = (0, 1] \quad \text{and} \quad \bigcap_{n\in\mathbb{Z}^+} A_n = \varnothing.$$

Suppose that $x \in (0, 1] = A_1$. Then by the definition of union, $x \in \cup_n A_n$. This proves that $(0, 1] \subseteq \cup_n A_n$. Conversely, suppose that $x \in \cup_n A_n$. Then, by the definition of union, $x \in A_n$ for some $n \in \mathbb{Z}^+$. Hence,

$$0 \leq x < 1/n \quad \rightarrow \quad x \in (0, 1].$$

This proves that $\cup_n A_n \subseteq (0, 1]$.

Now, suppose that $\cap_n A_n \neq \varnothing$. Then there is a real number x with $x \in \cap_n A_n$: that is, $0 < x \leq 1/n$ for all $n \in \mathbb{Z}^+$. In particular, we have $x > 0$. Next, we choose an integer m such that $m > 1/x$. Then $x > 1/m$, so $x \notin A_m$, and by the definition of intersection, $x \notin \cap_n A_n$; a contradiction. Therefore, $\cap_n A_n = \varnothing$. □

The properties of arbitrary unions and intersections listed in the next theorem are generalizations of some of those listed in Proposition 9.7.

Theorem 14.2. *Let $\mathcal{A} = \{A_i : i \in I\}$ be an indexed family of sets. Let $j \in I$ and let B be a set. Then:*

(1) $\bigcap_{i\in I} A_i \subseteq A_j \subseteq \bigcup_{i\in I} A_i$;

(2) $B \cap (\bigcup_{i\in I} A_i) = \bigcup_{i\in I}(B \cap A_i)$;

(3) $B \cup (\bigcap_{i\in I} A_i) = \bigcap_{i\in I}(B \cup A_i)$;

(4) $(\bigcup_{i\in I} A_i)^c = \bigcap_{i\in I} A_i^c$;

(5) $(\bigcap_{i\in I} A_i)^c = \bigcup_{i\in I} A_i^c$;

(6) $(\bigcup_{i\in I} A_i) \times B = \bigcup_{i\in I}(A_i \times B)$;

(7) $(\bigcap_{i\in I} A_i) \times B = \bigcap_{i\in I}(A_i \times B)$.

Proof. (3) First, suppose that $x \in B \cup (\bigcap_{i\in I} A_i)$. Then $x \in B$ or $x \in \cap_i A_i$. We consider two cases:

Case 1: $x \in B$. Then $x \in B \cup A_i$ for all $i \in I$. Therefore, by the definition of intersection, $x \in \cap_i(B \cup A_i)$.

Case 2: $x \notin B$. Then $x \in \cap_i A_i$, so by the definition of intersection, $x \in A_i$ for all $i \in I$. Therefore, $x \in B \cup A_i$ for all $i \in I$, and by the definition of intersection, $x \in \cap_i(B \cup A_i)$.

We conclude that $x \in \bigcap_{i\in I}(B \cup A_i)$. Thus, $B \cup (\bigcap_{i\in I} A_i) \subseteq \bigcap_{i\in I}(B \cup A_i)$.

Next, suppose that $x \in \bigcap_{i\in I}(B \cup A_i)$. We consider two cases:

Case 1: $x \in B$. Then, by the definition of union, $x \in B \cup (\cap_i A_i)$.

Case 2: $x \notin B$. By the definition of intersection,

$$x \in \bigcap_{i\in I}(B \cup A_i) \quad \rightarrow \quad x \in (B \cup A_i) \quad \text{for all } i \in I.$$

Since $x \notin B$, we deduce that $x \in A_i$ for all $i \in I$. Thus, $x \in \cap_i A_i \subseteq B \cup (\cap_i A_i)$.

Therefore, $x \in B \cup (\bigcap_{i\in I} A_i)$, and $\bigcap_{i\in I}(B \cup A_i) \subseteq B \cup (\bigcap_{i\in I} A_i)$.

(7) First, let $x \in (\bigcap_{i\in I} A_i) \times B$. This means that $x = (a, b)$, with $a \in \cap_i A_i$ and $b \in B$. By the definition of intersection, $a \in A_i$ for all $i \in I$. Hence, $(a, b) \in A_i \times B$ for all $i \in I$; by the definition of intersection, this leads to $x = (a, b) \in \bigcap_{i\in I}(A_i \times B)$. Thus, $(\bigcap_{i\in I} A_i) \times B \subseteq \bigcap_{i\in I}(A_i \times B)$.

Conversely, let $x \in \bigcap_{i\in I}(A_i \times B)$; then $x \in A_i \times B$ for all $i \in I$. Thus, x is an ordered pair, $x = (a, b)$, where $b \in B$ and $a \in A_i$ for all $i \in I$. The latter means that $a \in \cap_i A_i$; together with $b \in B$, this gives $x = (a, b) \in (\cap_i A_i) \times B$, proving that $\bigcap_{i\in I}(A_i \times B) \subseteq (\bigcap_{i\in I} A_i) \times B$. □

Corollary 14.3. *Let $\mathcal{A} = \{A_i : i \in I\}$ and $\mathcal{B} = \{B_j : j \in J\}$ be two indexed families of sets. Then*

$$\Big(\bigcup_{i\in I} A_i\Big) \cap \Big(\bigcup_{j\in J} B_j\Big) = \bigcup_{(i,j)\in I\times J} (A_i \cap B_j).$$

Proof. Let $B = \cup_j B_j$. By Proposition 9.7(4) and Theorem 14.2(2), we get

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

Next, for a fixed $i \in I$, we apply Theorem 14.2(2) with $B = A_i$ and $\mathcal{A} = \mathcal{B}$ to get

$$A_i \cap B = A_i \cap \Big(\bigcup_{j\in J} B_j\Big) = \bigcup_{j\in J}(A_i \cap B_j).$$

Hence,

$$\Big(\bigcup_{i\in I} A_i\Big) \cap \Big(\bigcup_{j\in J} B_j\Big) = \bigcup_{i\in I}\Big(\bigcup_{j\in J}(A_i \cap B_j)\Big) = \bigcup_{(i,j)\in I\times J} (A_i \cap B_j).$$

□

14.2 Open and Closed Subsets of the Reals

Example 14.4. A set $U \subseteq \mathbb{R}$ is called *open* if it can be expressed as a union of a family of open intervals: that is, U is open if $U = \bigcup_{i\in I}(a_i, b_i)$ for some index set I and pairs of real numbers $a_i < b_i$, $i \in I$. In this example, we will look at some examples and non-examples of open sets.

(a) Any (finite) open interval (a, b) is an open set, because it is the union of the one-set family $\{(a, b)\}$. However, we have also

$$(a, b) = \bigcup_{a<x<b} (a, x),$$

so (a, b) is open also because it is the union of the infinite family of intervals $\{(a, x) : a < x < b\}$. The point we are making here is that the family of intervals that establish the openness of a set need not be unique.[1]

(b) An infinite open interval of the form (a, ∞), $a \in \mathbb{R}$, is open. For example, it is the union of all open intervals (a, b), $b > a$:

$$(a, \infty) = \bigcup_{b>a}(a, b). \qquad (*)$$

[1] In fact, if you think briefly about the example of the open interval, you will realize that this family is almost never unique. What is the one situation, when it *is* unique?

Let us prove this.

First, let $x \in (a, \infty)$. Then $a < x < x + 1$, so $x \in (a, x + 1)$. Since $(a, x + 1)$ is an open interval in our family, $x \in \bigcup_{b>a}(a, b)$. This proves that $(a, \infty) \subseteq \bigcup_{b>a}(a, b)$.

Conversely, if x belongs to the union, then $x \in (a, b)$ for some $b > a$. Hence, $a < x < b$; in particular, the condition $x > a$ means that $x \in (a, \infty)$. This completes the proof of $(*)$.

(c) Similarly to part (b), we can show that an infinite open interval of the form $(-\infty, b)$, $b \in \mathbb{R}$, is also open—this is Exercise 14.5 below. Later, we will show that the union of any two open sets is open (see Exercise 14.7 below in the case of a family of two sets). Taking that fact for granted, we can now show that $\mathbb{R}$ is open:

$$\mathbb{R} = (-\infty, 1) \cup (0, \infty);$$

we already proved that $(0, \infty)$ is open, and $(-\infty, 1)$ is open by Exercise 14.5.

(d) The semi-open interval $(0, 1]$ is not an open set. Indeed, assume that it was: i.e., assume that

$$(0, 1] = \bigcup_{i \in I} (a_i, b_i)$$

for some family of open intervals. Since $1 \in (0, 1]$, by the definition of union, $1 \in (a_j, b_j)$ for some $j \in I$; hence, $a_j < 1 < b_j$. Let $1 < c < b_j$. Then $c \in (a_j, b_j)$, and so c belongs to the union of the whole family. Thus, $c \in (0, 1]$. This contradicts our choice that $c > 1$. Therefore, $(0, 1]$ is not open.

(e) The empty set is open, because it is the union of the empty family of intervals. □

Exercise 14.5. Let $b \in \mathbb{R}$. Prove that the interval $(-\infty, b)$ is an open set.

Exercise 14.6. Generalize part (d) of Example 14.4 to show that the closed interval $[a, b]$ is not open for any $a, b \in \mathbb{R}$, $a < b$.

Exercise 14.7. Show that if $\mathcal{U} = \{U_i : i \in I\}$ is a family of open subsets of $\mathbb{R}$, then their union $\bigcup_{i \in I} U_i$ is also an open subset of $\mathbb{R}$.

Example 14.8. A set $F \subseteq \mathbb{R}$ is called *closed* if its complement is open: that is, F is closed if $\mathbb{R} \setminus F$ is open (in the sense of Example 14.4).

(a) An infinite interval of the form $(-\infty, b]$, $b \in \mathbb{R}$, is a closed set, because its complement is the interval (b, ∞), which is open by Example 14.4(b).

(b) Any (finite) closed interval $[a, b]$ is a closed set, because its complement is the set $-\infty, a) \cup (b, \infty)$, which is an open set by Example 14.4(b) and Exercises 14.5 and 14.7.

(c) The empty set $\varnothing$ is closed, because its complement is $\mathbb{R}$, which is an open set by Example 14.4(c). The set $\mathbb{R}$ of all real numbers is also closed, because its complement is the empty set, which is open by Example 14.4(e).

(d) The semi-open interval $(0, 1]$ is not a closed set. Indeed, assume that it was. Then its complement, $(-\infty, 0] \cup (1, \infty)$, must be open: i.e.,

$$(-\infty, 0] \cup (1, \infty) = \bigcup_{i \in I} (a_i, b_i)$$

for some family of open intervals. Since $0 \in (-\infty, 0] \cup (1, \infty)$, by the definition of union, $0 \in (a_j, b_j)$ for some $j \in I$; hence, $a_j < 0 < b_j$. Let $0 < c < \min(b_j, 1)$. Then $c \in (a_j, b_j)$, and so c belongs to the union of the whole family. Thus, $c \in (-\infty, 0] \cup (1, \infty)$. This contradicts our choice that $0 < c < 1$. Therefore, $(0, 1]$ is not closed. □

14.3 Topologies on a Set*

The family of open sets of real numbers is a special case of a mathematical notion called "topology."

Definition 14.2. Let X be a non-empty set. A family $\mathcal{T}$ of subsets of X is called a *topology* on X if it has the following properties:

O_1. $X, \varnothing \in \mathcal{T}$;

O_2. if $\mathcal{U} = \{U_i : i \in I\} \subseteq \mathcal{T}$ is any family of sets from $\mathcal{T}$, then $\bigcup_{i \in I} U_i \in \mathcal{T}$;

O_3. if $\{U_1, U_2, \ldots, U_n\} \subseteq \mathcal{T}$, $n \in \mathbb{Z}^+$, is any finite family of sets from $\mathcal{T}$, then $\bigcap_{i=1}^{n} U_i \in \mathcal{T}$.

If $\mathcal{T}$ is a topology on X, the sets $U \in \mathcal{T}$ are called $\mathcal{T}$-*open*.

Example 14.4(e) and Exercise 14.7 verify that the collection of open subsets of $\mathbb{R}$ satisfies axioms O_1 and O_2, and the next example verifies that the collection of open subsets of $\mathbb{R}$ satisfies axiom O_3. Thus, the collection of open subsets of $\mathbb{R}$ (as defined in Example 14.4) is a topology on $\mathbb{R}$. This topology is called the "usual" or "standard topology" on $\mathbb{R}$.

Example 14.9. Let $U_1, U_2, \ldots, U_n$, $n \geq 2$, be open sets. We will prove that their intersection $\bigcap_{i=1}^{n} U_i$ is also open.

Consider first the case $n = 2$. We have

$$U_1 = \bigcup_{i \in I} (a_i, b_i), \quad U_2 = \bigcup_{j \in J} (c_j, d_j)$$

for some index sets I, J and families of open intervals $\{(a_i, b_i)\}_{i \in I}$, $\{(c_j, d_j)\}_{j \in J}$. By Corollary 14.3, we have

$$U_1 \cap U_2 = \bigcup_{(i,j) \in I \times J} \big[(a_i, b_i) \cap (c_j, d_j)\big].$$

Let us consider the intersection of two open intervals (a, b) and (c, d). Let $p = \max(a, c)$ and $q = \min(b, d)$. We have

$$x \in (a, b) \cap (c, d) \quad \leftrightarrow \quad a < x < b,\ c < x < d \quad \leftrightarrow \quad p < x < q,$$

so

$$(a,b)\cap(c,d)=\begin{cases}(p,q) & \text{if } p<q,\\ \varnothing & \text{if } p\geq q.\end{cases}$$

If we write $p_{ij}=\max(a_i,c_j)$ and $q_{ij}=\min(b_i,d_j)$, we see that

$$(a_i,b_i)\cap(c_j,d_j)=(p_{ij},q_{ij}) \text{ or } \varnothing.$$

Hence, after dropping the empty sets from the union, we get

$$U_1\cap U_2=\bigcup_{(i,j)\in K}(p_{ij},q_{ij}),\quad K=\{(i,j)\in I\times J : p_{ij}<q_{ij}\}.$$

That is, $U_1\cap U_2$ is the union of a family of open intervals and must be open.

We can establish the general case by induction on $n\geq 2$. We just proved the base case $n=2$. Suppose that $n\geq 2$ and we know that the intersection of every n open sets is open. Consider open sets $U_1,U_2,\ldots,U_{n+1}$. By the inductive hypothesis, $V=\bigcap_{i=1}^{n}U_i$ is open. Then, by the base case $n=2$, $V\cap U_{n+1}$ must be open. Since,

$$\bigcap_{i=1}^{n+1}U_i=\Big(\bigcap_{i=1}^{n}U_i\Big)\cap U_{n+1}=V\cap U_{n+1},$$

this completes the induction. □

It turns out that a set can have many topologies. The next example, presents another topology on $\mathbb{R}$, different from the standard topology.

Example 14.10. The family

$$\mathcal{T}=\{\varnothing\}\cup\{U\subseteq\mathbb{R} : \mathbb{R}\setminus U \text{ is finite}\}.$$

is a topology on $\mathbb{R}$.

O_1. The empty set is included explicitly in $\mathcal{T}$ and $\mathbb{R}\in\mathcal{T}$, because $\mathbb{R}\setminus\mathbb{R}$ is empty, and hence, finite.

O_2. Let $\mathcal{U}=\{U_i : i\in I\}$ be a family of sets from $\mathcal{T}$. Suppose first that none of the sets in $\mathcal{U}$ is empty. Then for all $i\in I$, $F_i=\mathbb{R}\setminus U_i$ is a finite set. By Theorem 14.2(4) and Theorem 14.2(1),

$$\Big(\bigcup_{i\in I}U_i\Big)^c=\bigcap_{i\in I}U_i^c=\bigcap_{i\in I}F_i\subseteq F_j,$$

for any fixed index $j\in I$. Since F_j is finite, it follows that $\cup_i U_i$ has a finite complement. Therefore, $\bigcup_{i\in I}U_i\in\mathcal{T}$.

If $\mathcal{U}$ does contain the empty set, let $j\in I$ be its index in the family: $U_j=\varnothing$; let $I'=I\setminus\{j\}$ and $\mathcal{U}'=\mathcal{U}\setminus\{\varnothing\}=\{U_i : i\in I'\}$. Then the union of $\mathcal{U}$ and that of $\mathcal{U}'$ are the same:

$$\bigcup_{i\in I}U_i=\Big(\bigcup_{i\in I'}U_i\Big)\cup U_j=\Big(\bigcup_{i\in I'}U_i\Big)\cup\varnothing=\bigcup_{i\in I'}U_i.$$

Since none of the sets in $\mathcal{U}'$ is empty, we can apply the already established case to the latter union. Therefore, $\bigcup_{i\in I} U_i$ is also in $\mathcal{T}$ (since the two unions are in fact the same).

O_3. Let $\{U_1, U_2, \ldots, U_n\}$, $n \geq 1$, be a finite family of sets from $\mathcal{T}$. If any of the sets U_i is empty, then $\cap_i U_i = \varnothing$ is an element of $\mathcal{T}$. Next, suppose that none of the sets $U_1, \ldots, U_n$ is empty. Then for all $i = 1, 2, \ldots, n$, $F_i = \mathbb{R} \setminus U_i$ is a finite set. By Theorem 14.2(5),

$$\left(\bigcap_{i=1}^{n} U_i\right)^c = \bigcup_{i=1}^{n} U_i^c = \bigcup_{i=1}^{n} F_i.$$

Since a finite union of finite sets is finite, it follows that $\cap_i U_i$ has a finite complement. Therefore, $\bigcap_{i=1}^{n} U_i \in \mathcal{T}$. □

Hints on Some Exercises

Exercise 14.5. Show that $(-\infty, b) = \bigcup_{a<b}(a, b)$.

Exercise 14.6. Assume that $[a, b]$ is open and use a in a similar way to how 0 is used in the example to "find" a c such that $c < a$ but c also belongs to $[a, b]$.

Exercise 14.7. Each open set U_i in the family is the union of some family of open intervals: say,

$$U_i = \bigcup_{j\in J_i} (a_{i,j}, b_{i,j}).$$

(Since the intervals $(a_{i,j}, b_{i,j})$ depend on U_i, their endpoints depend on two indices, and the index set J_i of a particular family of intervals will in general depend on i.) Let $K = \{(i, j) : i \in I,\ j \in J_i\}$. It is easy to show that

$$\bigcup_{i\in I} U_i = \bigcup_{(i,j)\in K} (a_{i,j}, b_{i,j}).$$

This proves that the union of the given family is an open subset of $\mathbb{R}$.

15 Images and Inverse Images

In this lecture, we introduce two important examples of families of sets related to a given function.

15.1 Definition and Basic Examples

If $f : X \to Y$ is any function, then for any set $A \subseteq X$, we can define its *image (under f)* by

$$f(A) = \big\{f(x) : x \in A\big\} = \big\{y \in Y : y = f(x) \text{ for some } x \in A\big\}.$$

Further, for any set $B \subseteq Y$, we can define its *inverse image (under f)*, called also the *pre-image of B*, by

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

Remark. Note that when we apply the definition of image to the set $A = X$, we get

$$f(X) = \big\{f(x) : x \in X\big\} = \big\{y \in Y : y = f(x) \text{ for some } x \in X\big\}.$$

This is the range of f. For this reason, the range of a function $f : X \to Y$ is often denoted by $f(X)$ (and not $\text{ran}(f)$).

Chapter 17 in the textbook contains several examples of computations of images and inverse images. Here, we add a couple more.

Example 15.1. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2 - 4x = (x-2)^2 - 4$. Then

$$f(\{-1, 0, 1, 3\}) = \{f(-1), f(0), f(1), f(3)\} = \{5, 0, -3\},$$

and the range of f is

$$\begin{aligned} f(\mathbb{R}) &= \big\{y \in \mathbb{R} : y = (x-2)^2 - 4 \text{ for some } x \in \mathbb{R}\big\} \\ &= \big\{y \in \mathbb{R} : y + 4 = (x-2)^2 \text{ for some } x \in \mathbb{R}\big\} = [-4, \infty), \end{aligned}$$

because the equation $y + 4 = (x-2)^2$ has real solutions exactly when $y + 4 \geq 0$.

Similarly, we have

$$\begin{aligned} f([2,4]) &= \big\{y \in \mathbb{R} : y = (x-2)^2 - 4 \text{ for some } x \in [2,4]\big\} \\ &= \big\{y \in \mathbb{R} : y + 4 = (x-2)^2 \text{ for some } x \in [2,4]\big\} \\ &= \big\{y \geq -4 : \pm\sqrt{y+4} = x - 2 \text{ for some } x \in [2,4]\big\} \\ &= \big\{y \geq -4 : \text{ at least one of } 2 \pm \sqrt{y+4} \text{ lies in } [2,4]\big\}. \end{aligned}$$

We have $2 - \sqrt{y+4} \geq 2$ only when $y = -4$. On the other hand,

$$2 \leq 2 + \sqrt{y+4} \leq 4 \quad \leftrightarrow \quad 0 \leq \sqrt{y+4} \leq 2 \quad \leftrightarrow \quad y + 4 \leq 4.$$

Hence,

$$f([2,4]) = \big\{y \geq -4 : y = -4 \text{ or } y \leq 0\big\} = [-4,0].$$

Further, we have

$$f^{-1}(\{0\}) = \big\{x \in \mathbb{R} \mid x^2 - 4x = 0\big\} = \{0,4\}, \quad f^{-1}(\{-4\}) = \{2\}, \quad f^{-1}(\{-5\}) = \varnothing.$$

Also,

$$\begin{aligned} f^{-1}([-3,5]) &= \big\{x \in \mathbb{R} : f(x) \in [-3,5]\big\} = \big\{x \in \mathbb{R} : -3 \leq (x-2)^2 - 4 \leq 5\big\} \\ &= \big\{x \in \mathbb{R} : 1 \leq (x-2)^2 \leq 9\big\} \\ &= \big\{x \in \mathbb{R} : 1 \leq x-2 \leq 3 \text{ or } -3 \leq x-2 \leq -1\big\} \\ &= \big\{x \in \mathbb{R} : 3 \leq x \leq 5 \text{ or } -1 \leq x \leq 1\big\} = [-1,1] \cup [3,5]. \end{aligned}$$

Exercise 15.2. Let $f : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

(a) Find $f(\{0\})$, $f(\{1,2,3\})$, $f((1,2))$, and $f((-1,3))$.
(b) Find the inverse image $f^{-1}((a,b))$ for an arbitrary open interval (a,b). You will need to consider several cases for this.

Exercise 15.3. Let $X = \big\{a, b, \{a\}, \{b\}, \{a,b\}\big\}$ and define $f : X \to \mathbb{N}$ by

$$f(a) = f(b) = 0, \quad f(\{a\}) = f(\{b\}) = 1, \quad f(\{a,b\}) = 2.$$

(a) Find $f(\{a,b\})$ in two different ways.
(b) Comment on what is wrong with the way part (a) is phrased and propose an alternative wording.

Theorems 17.5 and 17.7 in the textbook establish several properties of images and pre-images of unions and intersections of sets. The next theorem contains some generalizations of those properties to arbitrary unions and intersections.

Theorem 15.4. *Let $f : X \to Y$ be a function, let $\mathcal{A} = \{A_i : i \in I\}$ be a family of subsets of X, and let $\mathcal{B} = \{B_j : j \in J\}$ be a family of subsets of Y. Then:*

(1) $f\left(\cup_{i\in I} A_i\right) = \cup_{i\in I} f(A_i)$;

(2) $f\left(\cap_{i\in I} A_i\right) \subseteq \cap_{i\in I} f(A_i)$;

(3) $f^{-1}\left(\cup_{j\in J} B_j\right) = \cup_{j\in J} f^{-1}(B_j)$;

(4) $f^{-1}\left(\cap_{j\in J} B_j\right) = \cap_{j\in J} f^{-1}(B_j)$.

Proof. (1) Let $y \in f\left(\cup_{i\in I} A_i\right)$. Then $y = f(x)$ for some $x \in \cup_{i\in I} A_i$. Then by the definition of union, $x \in A_i$ for some $i \in I$. Hence, $f(x) \in f(A_i)$, that is, $y \in f(A_i)$. Thus, by the definition of union, $y \in \cup_{i\in I} f(A_i)$. This proves that $f\left(\cup_{i\in I} A_i\right) \subseteq \cup_{i\in I} f(A_i)$.

Now, let $y \in \cup_{i\in I} f(A_i)$. By the definition of union, $y \in f(A_i)$ for some $i \in I$. That is, $y = f(x)$ for some $x \in A_i$. Since by the definition of union, $x \in \cup_{i\in I} A_i$, we get that $y = f(x) \in f\left(\cup_{i\in I} A_i\right)$. This proves that $\cup_{i\in I} f(A_i) \subseteq f\left(\cup_{i\in I} A_i\right)$. □

15.2 Continuous Functions*

Advanced mathematics often uses images and inverse images of families of sets with some special property to define new concepts. Here, we give one such example. The next definition looks very different from the definition of a continuous function in calculus, but is in fact a generalization of that definition. Indeed, if we make the calculus definition of continuity rigorous, it will become a more cumbersome version of the definition below.

Definition 15.1. Let $D \subseteq \mathbb{R}$ be a non-empty set of real numbers. A function $f : D \to \mathbb{R}$ is called *continuous*, if for every open interval $(p, q) \subset \mathbb{R}$, its inverse image $f^{-1}((p, q))$ satisfies

$$f^{-1}((p, q)) = D \cap U \quad \text{for some an open set } U \subseteq \mathbb{R}.$$

Example 15.5. Let $f : [-1, 1] \to \mathbb{R}$ be given by $f(x) = x^2 + 1$. We will prove that f is continuous.

We need to compute the inverse image $f^{-1}((p, q))$ of an arbitrary open interval (p, q). We have

$$f^{-1}((p, q)) = \left\{x \in [-1, 1] : p < x^2 + 1 < q\right\} = \left\{x \in [-1, 1] : p - 1 < x^2 < q - 1\right\}.$$

Recall that

$$x^2 < a \quad \leftrightarrow \quad \begin{cases} -\sqrt{a} < x < \sqrt{a} & \text{if } a > 0, \\ x \in \varnothing & \text{if } a \leq 0; \end{cases}$$

and

$$x^2 > a \quad \leftrightarrow \quad \begin{cases} x < -\sqrt{a} \text{ or } x > \sqrt{a} & \text{if } a > 0, \\ x \neq 0 & \text{if } a = 0, \\ x \in \mathbb{R} & \text{if } a < 0. \end{cases}$$

We will use these two facts to guide our choice of several cases for p, q, and in each case we will calculate $f^{-1}((p, q))$.

Case 1: $q \leq 1$. Then the inequality $x^2 < q - 1$ has no solution, so $f^{-1}((p, q)) = \varnothing$—which is an open subset of $\mathbb{R}$.

Case 2: $1 < q \leq 2$. Then the inequality $x^2 < q - 1$ has solutions $-\sqrt{q-1} < x < \sqrt{q-1}$. Since in this case $\sqrt{q-1} \leq 1$, we get

$$f^{-1}((p,q)) = \big\{x \in (-\sqrt{q-1}, \sqrt{q-1}) : x^2 > p - 1\big\}.$$

Case 2.1: $p < 1$. Then the inequality $x^2 > p - 1$ holds for all x, so

$$f^{-1}((p,q)) = (-\sqrt{q-1}, \sqrt{q-1}),$$

an open subset of $\mathbb{R}$.

Case 2.2: $p = 1$. Then the inequality $x^2 > p - 1$ holds for all $x \neq 0$, so

$$f^{-1}((p,q)) = \big(-\sqrt{q-1}, 0\big) \cup \big(0, \sqrt{q-1}\big),$$

which is also an open subset of $\mathbb{R}$.

Case 2.3: $1 < p < q$. Then the inequality $x^2 > p - 1$ holds when $x > \sqrt{p-1}$ or $x < -\sqrt{p-1}$. Combining this with the restriction that $-\sqrt{q-1} < x < \sqrt{q-1}$, we get

$$f^{-1}((p,q)) = \big(-\sqrt{q-1}, -\sqrt{p-1}\big) \cup \big(\sqrt{p-1}, \sqrt{q-1}\big),$$

which is again an open subset of $\mathbb{R}$.

Case 3: $q > 2$. Then the inequality $x^2 < q - 1$ has solutions $-\sqrt{q-1} < x < \sqrt{q-1}$. Since in this case $\sqrt{q-1} > 1$, this condition follows from the restriction that $x \in [-1, 1]$. Thus,

$$f^{-1}((p,q)) = \big\{x \in [-1,1] : x^2 > p - 1\big\}.$$

Case 3.1: $p < 1$. Then the inequality $x^2 > p - 1$ holds for all x, so $f^{-1}((p,q)) = [-1,1]$. This is not an open set itself, but it is the intersection of the domain $[-1,1]$ and an open set: for example, $(-2,2)$ is an open set and

$$[-1,1] = [-1,1] \cap (-2,2).$$

Case 3.2: $p = 1$. Then the inequality $x^2 > p - 1$ holds for all $x \neq 0$, so

$$f^{-1}((p,q)) = [-1,0) \cup (0,1] = [-1,1] \cap \big((-2,0) \cup (0,2)\big),$$

and $(-2,0) \cup (0,2)$ is an open set.

Case 3.3: $1 < p < 2$. Then the inequality $x^2 > p - 1$ holds when $x > \sqrt{p-1}$ or $x < -\sqrt{p-1}$. Combining this with the restriction that $-1 \leq x \leq 1$, we get

$$\begin{aligned} f^{-1}((p,q)) &= \big[-1, -\sqrt{p-1}\big) \cup \big(\sqrt{p-1}, 1\big] \\ &= [-1,1] \cap \Big(\big(-2, -\sqrt{p-1}\big) \cup \big(\sqrt{p-1}, 2\big)\Big), \end{aligned}$$

which is again an intersection of $[-1,1]$ and an open set.

Case 3.4: $2 \leq p < q$. Again, the inequality $x^2 > p - 1$ holds for $x > \sqrt{p-1}$ or $x < -\sqrt{p-1}$. However, in this case $\sqrt{p-1} \geq 1$, so these inequalities are inconsistent with the restriction that $-1 \leq x \leq 1$. Hence, $f^{-1}((p,q)) = \varnothing$, an open set.

We see that in all cases, the inverse image $f^{-1}((p,q))$ is either an open subset of $\mathbb{R}$, or the intersection of such a set and the domain $[-1,1]$. Therefore, f is continuous. $\square$

Example 15.6. Let $f : \mathbb{N} \to \mathbb{R}$ be a function. Then it is continuous.

This may seem a little counterintuitive at first—we have no information what function we are talking about. How is it possible to tell that it is continuous? The key point is that the domain of f is $\mathbb{N}$ and the elements of $\mathbb{N}$ do not get close to one another (if $m, n \in \mathbb{N}$ and $m \neq n$, then $|m - n| \geq 1$).

Let (p, q) be an interval. Then $f^{-1}((p, q))$ is a subset of $\mathbb{N}$: either $f^{-1}((p, q)) = \varnothing$ (and so, it is open), or

$$f^{-1}((p, q)) = \{n_1, n_2, \dots, n_i, \dots\}$$

for some integers $n_1 < n_2 < \cdots < n_i < \cdots$. Let us write this set as $\{n_i : i \in I\}$, where the set of indices I is either a finite set $\{1, 2, \dots, m\}$, or $I = \mathbb{Z}^+$. Either way, we have

$$f^{-1}((p, q)) = \bigcup_{i \in I} \{n_i\} = \bigcup_{i \in I} \big(\mathbb{N} \cap (n_i - 0.5, n_i + 0.5)\big) = \mathbb{N} \cap \bigcup_{i \in I} (n_i - 0.5, n_i + 0.5).$$

Since the last union is an open set (it is the union of open intervals), this proves that $f^{-1}((p, q))$ is the intersection of an open subset of $\mathbb{R}$ and the domain $\mathbb{N}$. Thus, f is continuous. □

Example 15.7. Let $f : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0. \end{cases}$$

We will show that f is not continuous.

Let (p, q) be an interval. Then

$$\begin{aligned} f^{-1}((p, q)) &= \big\{x \in \mathbb{R} : p < f(x) < q\big\} \\ &= \big\{x \leq 0 : p < 0 < q\big\} \cup \big\{x > 0 : p < 1 < q\big\}. \end{aligned}$$

Let us consider this in the case when $p < 0 < q \leq 1$. In this case, the condition $p < 0 < q$ is true, and the condition $p < 1 < q$ is false. Thus, the first set in the above union is $(-\infty, 0]$, and the second set is $\varnothing$; hence, $f^{-1}((p, q)) = (-\infty, 0]$. Since the domain of f is $\mathbb{R}$, the condition that f is continuous reduces to requiring that $f^{-1}((p, q))$ be open. However, similarly to Example 14.4(d), we can show that $(-\infty, 0]$ is not open. Therefore, f is not continuous. □

Exercise 15.8. In the previous example, we did not need to compute $f^{-1}((p, q))$ in all possible cases for p, q, since we needed to find just one counterexample. Find $f^{-1}((p, q))$ for all possible choices of p, q.

Example 15.9. Let $f : \mathbb{R} \to \mathbb{R}^+$ be a continuous function, and define $g : \mathbb{R} \to \mathbb{R}^+$ by[1]

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

[1]This is well-defined, because f takes only positive values; in particular, $f(x)$ is never 0.

Then g is also continuous.

Let (p, q) be an interval. Then

$$g^{-1}((p,q)) = \{x \in \mathbb{R} : p < g(x) < q\} = \{x \in \mathbb{R} : p < \frac{1}{f(x)} < q\}.$$

We consider three cases depending on the signs of p, q.

Case 1: $q \le 0$. Then $g^{-1}((p,q)) = \varnothing$, because $q \le 0 < 1/f(x)$.

Case 2: $0 < p < q$. Then

$$g^{-1}((p,q)) = \{x \in \mathbb{R} : 1/q < f(x) < 1/p\} = f^{-1}((1/q, 1/p)),$$

which is an open subset of $\mathbb{R}$ by the continuity of f.

Case 3: $p \le 0 < q$. Then the inequality $p < 1/f(x)$ is superfluous, and we get

$$g^{-1}((p,q)) = \{x \in \mathbb{R} : f(x) > 1/q\} = \bigcup_{r > 1/q} \{x \in \mathbb{R} : 1/q < f(x) < r\},$$

by an argument similar to the solution of Example 14.4(b). The continuity of f means that the sets

$$\{x \in \mathbb{R} : 1/q < f(x) < r\} = f^{-1}((1/q, r))$$

are open subsets of $\mathbb{R}$. Therefore, in this case, $g^{-1}((p,q))$ is the union of open sets and must also be open by Exercise 14.7. □

Hints on Some Exercises

Exercise 15.2. **(a)** We have $f(\{0\}) = \{0\}$, $f(\{1,2,3\}) = \{1, \frac{1}{2}, \frac{1}{3}\}$, and $f((1,2)) = (\frac{1}{2}, 1)$. Also,

$$\begin{aligned} f((-1,3)) &= \{y \in \mathbb{R} : y = f(x) \text{ for some } x \in (-1,3)\} \\ &= \{y \in \mathbb{R} : y = \tfrac{1}{x} \text{ for some } x \in (0,3)\} \cup \{f(0)\} \\ &\quad \cup \{y \in \mathbb{R} : y = \tfrac{1}{x} \text{ for some } x \in (-1,0)\} \\ &= \{y \in \mathbb{R} : \tfrac{1}{y} \in (0,3)\} \cup \{0\} \cup \{y \in \mathbb{R} : \tfrac{1}{y} \in (-1,0)\} \\ &= \{y \in \mathbb{R} : y > \tfrac{1}{3}\} \cup \{0\} \cup \{y \in \mathbb{R} : y < -1\} = (-\infty, -1) \cup \{0\} \cup (\tfrac{1}{3}, \infty). \end{aligned}$$

(b) We consider several cases depending on the signs of a and b.

Case 1: $0 < a < b$. Note that $f(x) > 0$ only if $x > 0$. Thus, in this case,

$$\begin{aligned} f^{-1}((a,b)) &= \{x \in \mathbb{R} : a < f(x) < b\} = \{x > 0 : a < \tfrac{1}{x} < b\} \\ &= \{x > 0 : \tfrac{1}{b} < x < \tfrac{1}{a}\} = (\tfrac{1}{b}, \tfrac{1}{a}). \end{aligned}$$

Case 2: $0 = a < b$. Then,

$$\begin{aligned} f^{-1}((0,b)) &= \{x \in \mathbb{R} : 0 < f(x) < b\} = \{x > 0 : 0 < \tfrac{1}{x} < b\} \\ &= \{x > 0 : x > \tfrac{1}{b}\} = (\tfrac{1}{b}, \infty). \end{aligned}$$

Case 3: $a < b < 0$. Similarly to Case 1, we get $f^{-1}((a,b)) = \left(\frac{1}{b}, \frac{1}{a}\right)$.
Case 4: $a < b = 0$. Similarly to Case 2, we get $f^{-1}((a,0)) = \left(-\infty, \frac{1}{a}\right)$.
Case 5: $a < 0 < b$. Then,

$$\begin{aligned} f^{-1}((a,b)) &= \big\{x \in \mathbb{R} : a < f(x) < b\big\} \\ &= \big\{x \in \mathbb{R} : a < f(x) < 0\big\} \cup \big\{x \in \mathbb{R} : f(x) = 0\big\} \cup \big\{x \in \mathbb{R} : 0 < f(x) < b\big\} \\ &= \left(-\infty, \tfrac{1}{a}\right) \cup \{0\} \cup \left(\tfrac{1}{b}, \infty\right), \end{aligned}$$

where we used the results of Cases 2 and 4 and noted that $f(x) = 0$ if and only if $x = 0$.

Exercise 15.3. (a) If $f(\{a,b\})$ is meant to represent the value $f(x)$ when $x = \{a,b\} \in X$, then $f(\{a,b\}) = 2$. Note that this is a number.

If $f(\{a,b\})$ is meant to represent the image of $f(A)$ when $A = \{a,b\} \subseteq X$, then

$$f(\{a,b\}) = \{f(a), f(b)\} = \{0,0\} = \{0\}.$$

Note that this is a *set* of numbers.

(b) Normally, this wording would be just fine. In the present example, however, the set $\{a,b\}$ is both an element and a subset of X, which leads to a possible ambiguity in the meaning of $f(\{a,b\})$. Note that there is no question what we mean by $f(a)$ or by $f(\{a,\{a\}\})$.

The solution of part (a) provides a path to eliminating this ambiguity. If we want to ask for the first answer, we could say something along the lines of

> "Find the function value $f(\{a,b\})$."

And if we want to ask for the second answer, we could say something along the lines of

> "Find the image $f(\{a,b\})$ of the set $\{a,b\}$."

Exercise 15.8. When $p \geq 1$, or $0 \leq p < q \leq 1$, or $q \leq 0$, we have $f^{-1}((p,q)) = \varnothing$.
When $0 \leq p < 1 < q$, we have $f^{-1}((p,q)) = (0,\infty)$.
When $p < 0 < 1 < q$, we have $f^{-1}((p,q)) = \mathbb{R}$.