1 Logic

1.1 Negations

A *negation* of a statement is a statement that is true exactly when the original statement is false.

Exercise 1.1. Negate the sentence "Every horse is red."

Of course, one negation would be "Not every horse is red". But try to come up with a negation that doesn't have "not" or "it's not true that" in front, and that illustrates better how you would *show* someone that not every horse is red.

To formalize what's going on above, it's convenient to adopt some mathematical shorthand. Let's write $\forall$ for 'for all', $\exists$ for 'there exists', and denote the negation of a statement P by $not\, P$. Then

$$not\,(\forall X, P(X)) = \exists X \text{ such that } not\, P(X), \tag{1}$$

$$not\,(\exists X \text{ such that } P(X)) = \forall X, not\, P(X). \tag{2}$$

Here, we write $P(X)$ to indicate that the P is a statement that accepts an input X, and the truth of P depends on what X is. For example, P could be the statement *"X is a politician"*, so then P(Kamala Harris) is true, while P(my cat) is false. Note that the statement "Every horse is red." can be rewritten as "For all horses H, H is red." Using (1) above, check that your answer in Exercise 1.1 is correct.

Exercise 1.2. Negate the sentence "In every war there is a hero who does not die." *Hint: this is a 'for all' statement that has a "there exists" statement inside. Try to write the sentence first with parentheses around the inner exists statement, and find the negation by working outside in.*

Remark (Vacuous truth)**.** Is the following statement true?

> "All the people on Jupiter are friendly."

Well, the negation would be "There is a person on Jupiter who is not friendly," which is certainly false since there are no people on Jupiter. So, the statement above should be true! This is an example of what's called *vacuous truth*, where a statement about some kind of object is automatically true because there are no such objects, so you can't find a counterexample. In some sense, it's a mathematical convention to say that the negation of "$\forall X, P(X)$" should be "$\exists X$ such that $P(X)$" even when there are no X's, but it makes everything simpler to interpret things in this way.

Let's write $P \wedge Q$ for P *and* Q, and write $P \vee Q$ for P *or* Q. So, $P \wedge Q$ is true when *both* P and Q are true, while $P \vee Q$ is true if *at least one* of the two is true. This interpretation of 'or' is typical in math; the statement P *or* Q *but not both* is a different thing, usually referred to as *exclusive or*.

Fact 1.3. $not\,(P \wedge Q) = (not\,P) \vee (not\,Q)$

Proof. To prove such a statement, we check for each pair of truth values for P and Q whether the two sides of the equation have the same truth value. For instance, if both P and Q are true, then both sides of the equation are false. One way to organize this kind of case by case analysis is with a *truth table*. For example,

P	Q	$P \wedge Q$	$(not\,P) \vee (not\,Q)$
T	T	T	F
T	F	F	T
F	T	F	T
F	F	F	T

indicates the possible truth values (T or F) of the statements $P \wedge Q$ and $(not\,P) \vee (not\,Q)$, depending on whether P, Q are true or false. Since these values are negations of each other, the Fact follows.

In general, some explanation might be necessary to convince the reader of the correct value of each table entry. For instance, if P is false and Q is true, then $not\,P$ is true and $not\,Q$ is false, so at least one of the two is true, i.e. $(not\,P) \vee (not\,Q)$ is true. The justifications for the other entries are similar in length. □

Exercise 1.4. Write a truth table showing that $not\,(P \vee Q) = not\,P \wedge not\,Q$.

Exercise 1.5. Negate the following statements, using the rules discussed above to turn all 'there exists' into 'for all's, and 'and's into 'or's, and vice versa.

(a) For every horse that jumps higher than my pet rabbit, there exists a lion that is tired and wants to eat that horse.

(b) There exists a real number x such that for all real numbers y, we have $x \geq y$.

(c) For all real numbers x, y, we have either $x \geq y$ or $x \leq y$.

1.2 Implications

We write $P \implies Q$, read P *implies* Q, if whenever P is true, then Q is true. Some other ways to describe this situation are Q*, if* P, and P *only if* Q. For example, the following English sentences all have the same meaning:

(a) The fact that she is a lawyer implies that she passed the bar exam.

(b) If she is a lawyer, then she passed the bar exam.

(c) She passed the bar exam, if she's a lawyer.

(d) She is a lawyer only if she passed the bar exam.

The fact that the last sentence has the same meaning as the others can be confusing, partly because it's a grammatical structure that isn't used so often in English. However, one can reformulate it as "The only way that she can be a lawyer is if she passed the bar exam," which makes it more clearly the same as the others.

Remark (Vacuous truth)**.** We talked about vacuous truth of 'for all' statements in the previous section. A similar phenomenon arises when you have a statement P that's always false and you ask whether the statement '$P \implies Q$' is true. Really, you can interpret this as a 'for all' statement: '$P \implies Q$' is the same as saying 'for all possible conditions in which P is true, Q is also true.' So, if there are no conditions where P is true, the statement $P \implies Q$ is called *vacuously true*.

Exercise 1.6. Write a truth table that shows that $\mathit{not}\,(P \implies Q) = P \wedge (\mathit{not}\,Q)$.

Exercise 1.7. Suppose I am the coach of our dodgeball team and you all are the players. I tell you "If we win tonight, then I will buy you pizza tomorrow." When can you rightly claim to have been lied to?

Exercise 1.8. Negate the sentence "If it's raining outside, then I will bring an umbrella and if my raincoat is back from the cleaners, I will wear it."

Exercise 1.9. Let A represent "6 is an even number" and B represent "6 is a multiple of 4." Express each of the following in ordinary English sentences and state whether the statement is true or false.

(a) $\mathit{not}\,A$

(b) $A \wedge B$

(c) $A \vee B$

(d) $(not\, A) \vee B$

(e) $A \wedge (not\, B)$

(f) $A \implies B$

(g) $B \implies A$

Remark 1.10. We used the symbols $not\,, \wedge, \vee$ partly for shorthand, and partly to shock your brains into thinking about these concepts more rigorously. However, in the rest of this text we will just write 'not' into English sentences, and write 'and' and 'or' instead of $\wedge, \vee$.

1.3 Converse and contrapositive

The **converse** of an implication $A \implies B$ is $B \implies A$, while the **contrapositive** of $A \implies B$ is $not\, B \implies not\, A$.

Exercise 1.11. Provide an example of a true statement whose converse is false.

Exercise 1.12. Find the converse and contrapositive of the following statements:

(a) If n is an even natural number, then $n + 1$ is an odd natural number.

(b) If it rains today, then I bring my umbrella.

Theorem 1.13. *Assume A and B are statements. Then $A \implies B$ if and only if $not\, B \implies not\, A$. That is, an implication is equivalent to its contrapositive.*

Proof. Here is the relevant truth table.

A	B	$A \implies B$	$not\, B \implies not\, A$
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

Alternatively, $A \implies B$ means that 'B is true whenever A is'. So, the only way that B can be false is if A is false. □

1.4 Proof methods

In a *direct* mathematical proof you start with a set of assumptions and write down a logical argument that eventually results in the desired conclusion. Here's an example.

Fact 1.14. *If integers m, n are perfect squares, so is their product mn.*

Proof. Say $m = x^2$ and $n = y^2$, where x, y are integers. Then

$$mn = x^2y^2 = (xy)^2,$$

which is a perfect square. □

However, sometimes it's convenient to write proofs with a different logical structure. For instance, Theorem 1.13 says that if you want to prove an implication, you can prove its contrapositive instead. Let's define an integer n to be *even* if $n = 2k$ for some $k \in \mathbb{Z}$, and say that n is *odd* if this is not the case. Prove the following by proving its contrapositive, being careful to use the definition above. In particular, do *not* try to use that odd numbers can be expressed as $2k + 1$ for some $k \in \mathbb{Z}$.

Exercise 1.15. Let x and y be integers. If xy is odd, then both x and y are odd.

Here is another statement that is most naturally proved via the contrapositive.

Exercise 1.16. Let $x, y \in \mathbb{R}$. If $\forall \epsilon > 0$ we have $|x - y| < \epsilon$, then $x = y$.

Another common proof method is to assume that the thing you want to prove is false, and show that something ridiculous arises from this. This is called *proof by contradiction*. For one example, recall that a real number is *rational* if it is the ratio a/b of two integers, where $b \neq 0$, and is *irrational* otherwise.

Fact 1.17. *The sum of a rational number and an irrational number is irrational.*

Proof. Suppose that x is rational and y is irrational. Hoping for a contradiction, suppose that $x + y$ is rational. Write $x = a/b$ and $x + y = c/d$. Then

$$y = x + y - x = c/d - a/b = (cb - ad)/bd,$$

which is rational, a contradiction to our assumptions. □

Here's another exercise that you can prove by contradiction.

Exercise 1.18. There do not exist integers x, y such that $3x + 6y = 1$.

Try to prove the following by contradiction, and also via the contrapositive. It will be essentially the same proof in both cases, just worded slightly differently.

Exercise 1.19. If x, y, z are real numbers and $x + y \geq z$, either $x \geq z/2$ or $y \geq z/2$.

Contradiction is sometimes the most natural method for a problem. However, a direct proof or a proof via the contrapositive is often also an option, and you should always try to use the method with the fewest logical cartwheels! Also, in longer proofs by contradiction it can be hard to trust your mathematical intuition, since in the entirety of the proof, you are working in an impossible universe.

Exercise 1.20. (Hard) Show by contradiction that there's no natural number $n \geq 2$ such that $\sum_{k=2}^{n} \frac{1}{k}$ is an integer. *Hint: suppose you have some n where the sum is an integer x, suppose that the biggest power of 2 that you see as a denominator in this sum is 2^m, multiply everything by 2^{m-1}, put all the fractions in lowest terms, and try to get a contradiction by looking at whether denominators are even or odd.*

Finally, we write $P \iff Q$, read 'P *if and only if* Q' when $P \implies Q$ and $Q \implies P$, i.e. when P is true exactly when Q is true. Sometimes we abbreviate and write P *iff* Q. We'll often find ourself wanting to prove such statements, which are known as 'bidirectional implications', or just 'if and only if' statements. When doing so, remember that there are two implications to prove, separately. First you prove $P \implies Q$, and then you prove $Q \implies P$. Here's an example.

Exercise 1.21. Suppose that x, y are positive real numbers. Show that $\frac{x+y}{2} = \sqrt{xy}$ if and only if $x = y$.

As another example, try to rewrite Exercise 1.16 as an iff statement, and prove it.

1.5 The importance of definitions

Exercise 1.22. Pair up with a friend or divide into groups. Try to write down a precise definition of one of the following geometric objects. Then have your friend or another group try to challenge your definition by coming up with either an example of some object that fits your definition literally, but isn't the object you're describing, or an example of the requested object that doesn't fit your definition.

(a) a circle,

(b) a line.

If you finish these, you could also try defining a 'polygon'!

2 Sets

A *set* is a collection of objects, called its *elements*. We write $x \in S$ if x is an element of a set S. Here, $\in$ is also read 'is in'. Sets are described by specifying their elements in some way, between curly braces $\{\quad\}$. Here are some examples.

- We can describe sets by listing all of their elements. For instance, the set

  $$S = \{1, 2, 3, 4, 5\}$$

  contains precisely the five smallest positive integers. As shorthand, we sometimes list only some of the elements, with ellipses to indicate other elements. For example, the set $T = \{3, 4, 5, \ldots, 100\}$ contains the positive integers from 3 to 100, including 6 through 99, even though these latter are not explicitly written. We'll also use ellipses lazily to describe some sets whose elements we already know and understand, like the sets of *natural numbers* and *integers*

  $$\mathbb{N} := \{1, 2, 3, \ldots\}, \quad \mathbb{Z} := \{\ldots, -2, -1, 0, 1, 2, \ldots\},$$

  and the set of *rational numbers*, i.e. the set of integer fractions

  $$\mathbb{Q} := \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0 \right\}.$$

  We'll also sometimes refer to the set $\mathbb{R}$ of all *real numbers*, which we won't rigorously define, but which is probably familiar from calculus.

  In this course, you'll often see ':=' written instead of '=' when we want to emphasize that something is a definition*, rather than an equality of previously defined objects.*

- Another way to describe a set is via a rule that picks out certain elements of a larger set. The notation looks like

  $\{$ things in the bigger set $\mid$ conditions they have to satisfy to be in the new set $\}$,

  where the vertical bar is pronounced 'such that'. As an example, we can write the set of even natural numbers as

  $$\{x \in \mathbb{N} \mid x = 2k \text{ for some } k \in \mathbb{N}\},$$

  read "the set of all x in the natural numbers such that $x = 2k$ for some k in the natural numbers". Similarly, the set of all real numbers whose squares are less than 3 is written

  $$\{x \in \mathbb{R} \mid x^2 < 3\}.$$

- Finally, we can describe a set via a rule (a 'function') that produces its elements from elements of another set. Such a description is written as

  $$\{ \text{ rule producing new elements from old elements } \mid \text{ old elements } \}$$

  As an example, the set of even integers can be written as

  $$\{2k \mid k \in \mathbb{Z}\},$$

  read "the set of all $2k$ such that k is in $\mathbb{Z}$", and the set of all perfect squares is

  $$\{n^2 \mid n \in \mathbb{Z}\}.$$

We say two sets A and B are *equal* if they contain precisely the same elements, that is, if $x \in A$ if and only if $x \in B$. Sets cannot contain multiple copies of elements: each element is either in a set S, or isn't. So for instance, the expression $\{1, 1, 2\}$ doesn't define a set that is distinct from $\{1, 2\}$.

Definition 2.1. The *empty set* is the set with no elements, and it is denoted $\emptyset$.

Exercise 2.2. Is it true that every element of the empty set is a cat?

2.1 Subsets of sets

We say that a set A is a *subset* of a set B, written $A \subset B$, if every element of A is also an element of B, that is, if $x \in A$, then $x \in B$. Note that

$$A = B \text{ if and only if } A \subset B \text{ and } B \subset A.$$

We call A a *proper subset* of B if $A \subset B$ and $A \neq B$, in which case we'll write $A \subsetneq B$.

Remark 2.3. Some authors write $\subseteq$ and $\subset$ for subset and proper subset, in analogy with $\leq$ and $<$. But in practice, one talks about subsets a lot more than proper subsets, so many mathematicians use the simpler symbol for the more common notion.

Exercise 2.4. Let $A = \{1, \{2\}\}$. Is $1 \in A$? Is $2 \in A$? Is $\{1\} \subset A$? Is $\{2\} \subset A$? Is $1 \subset A$? Is $\{1\} \in A$? Is $\{2\} \in A$? Is $\{\{2\}\} \subset A$? Explain.

The *power set* of a set A is the set $\mathcal{P}(A)$ of all subsets of A.

Exercise 2.5. Write out the elements of $\mathcal{P}(A)$ when $A = \{1, 2, 3\}$.

Exercise 2.6. What's $\mathcal{P}(\emptyset)$? Prove it.

Exercise 2.7. Show that every nonempty set A has a proper subset.

Fact 2.8. *Suppose $A \subset B$ and $B \subset C$. Show that $A \subset C$.*

This is our first (simple) example of a proof of set inclusion, so we'll demonstrate the form the proof should take, which you should follow whenever you want to prove one set is contained in another.

Proof. Let $a \in A$. Since $A \subset B$, we have $a \in B$. But then since $B \subset C$, we have $a \in C$. So, every element of A lies in C, proving $A \subset C$. □

Exercise 2.9. Is there a set that has exactly 3 subsets?

Exercise 2.10. Suppose that A, B are sets. Show that $A \subset B \iff$ every subset of A is a subset of B.

Exercise 2.11. Suppose X, A, B are sets and $A, B \subset X$. Show that

$$A \subset B \iff (X \setminus B) \subset (X \setminus A).$$

2.2 Unions, Intersections and Products

Definition 2.12. Let A and B be two sets. The *union* of A and B is the set

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

The *intersection* of A and B is the set

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

Note the similarity between the symbols $\cup, \cap$ and $\vee, \wedge$.

Theorem 2.13 (Distribution of Union and Intersection)**.** *If A, B, and C are sets,*

(a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,

(b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Each of these inequalities can be proved via a *double inclusion proof*, where we prove that the two sets are equal by showing that each is a subset of the other. Here's an outline of the proof of part (a).

Proof of (a). We first show that $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$. So, let $x \in A \cup (B \cap C)$.

...

Hence, $x \in (A \cup B) \cap (A \cup C)$ as desired.

We now show that $A \cup (B \cap C) \supset (A \cup B) \cap (A \cup C)$. Let $x \in (A \cup B) \cap (A \cup C)$.

...

Therefore, $x \in A \cup (B \cap C)$. □

Exercise 2.14. Fill in the missing parts of the proof above.

Exercise 2.15. Prove part (b).

Exercise 2.16. If A, B are sets, prove that $A \cup (A \cap B) = A$ using a double inclusion.

Two sets A and B are called *disjoint* if $A \cap B = \emptyset$.

Exercise 2.17. Can a set be disjoint from itself?

Definition 2.18. Let A and B be two sets. The *difference* of A and B is the set

$$A \setminus B = \{x \in A \mid x \notin B\}.$$

When $B \subset A$, the set $A \setminus B$ is also called the *complement* of B in A.

Theorem 2.19. *(DeMorgan's Laws) Let X be a set, and let $A, B \subset X$. Then:*

(a) $X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B)$

(b) $X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B)$

Exercise 2.20. Prove part (a) of DeMorgan's laws, using double inclusion.

Exercise 2.21. Let A, B, C be sets such that $C \subset A$. By double inclusion, show that

$$A \setminus (B \setminus C) = (A \setminus B) \cup C.$$

Then give an example showing that this isn't true if we don't have $C \subset A$.

Exercise 2.22. If S, T are sets, is it true that $\mathcal{P}(S) \cup \mathcal{P}(T) = \mathcal{P}(S \cup T)$?

Definition 2.23. If A, B are sets, their *(Cartesian) product* is the set

$$A \times B := \{(a, b) \mid a \in A, b \in B\}$$

of all 'ordered pairs' of elements from A and B, respectively. We can also take a product of any number of sets, e.g.

$$A_1 \times \cdots \times A_n := \{(a_1, \ldots, a_n) \mid a_i \in A_i \text{ for all i}\}.$$

Elements of $A_1 \times \cdots \times A_n$ are called *n-tuples*. Note that $(A \times B) \times C$ is basically the same as $A \times B \times C$, although technically one should write elements of the former set as $((a, b), c)$ instead of (a, b, c), but we'll ignore this from now on. Also, we define

$$A^n = A \overset{n \text{ times}}{\times \cdots \times} A.$$

Note that if $\mathbb{R}$ is the set of real numbers, $\mathbb{R}^2$ is then the plane. If

$$M = \{Subaru, Honda, Ford, Chevrolet, \ldots\}$$

is the set of all car makes, and

$$C = \{red, blue, \ldots\}$$

is the set of all possible colors, then $M \times C$ is the set of all pairs of car makes with colors, which is useful set to choose from if you are buying a car.

Exercise 2.24. How many elements does $\{1, \ldots, m\} \times \{1, \ldots, n\}$ have?

Exercise 2.25. What is $\mathbb{N} \times \emptyset$?

Exercise 2.26. Show that for all sets A, B, C, D, we have

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

2.3 Math is broken

Let $\mathcal{S}$ be the set of all sets, and let $\mathcal{R} := \{A \in \mathcal{S} \mid A \notin A\}$. i.e. $\mathcal{R}$ is the set of all sets that don't contain themselves as elements.

Exercise 2.27. (Russel's Paradox) Find a contradiction in mathematics by studying whether $\mathcal{R}$ is an element of $\mathcal{R}$.

A colloquial restatement of this goes as follows. *In Seville, there is a barber who shaves all those men, and only those men, who do not shave themselves. So, who shaves the barber?*

Is math broken? What the exercise indicates is that it is problematic to assume that there is something like the 'set of all sets', and that we need a stricter definition of a set than 'some collection of elements'. Essentially, the way to resolve this is as follows. Starting out with some basic building blocks like the empty set, and say for simplicity $\mathbb{N}$, we only allow ourselves to look at sets constructed from these by natural set operations like unions, products, taking subsets, etc.... Writing down the rules precisely is pretty subtle, though, so we'll ignore all that and just naively assume in the future that all reasonable expressions we write down do describe sets.

2.4 The halting problem

The following is similar to Russell's paradox, although it doesn't involve sets.

Let's consider a computer program as a set of instructions that takes as input some text, and does something in response. A program *halts* if it eventually stops running. For example, consider a program **Admirer** that takes a text input X and then prints "I love X" on the screen. Compare this with a program **Stalker** that takes in an input X and then repeatedly writes "I love X" on the screen forever. The first program halts, while the second doesn't. Now each computer program can be itself encoded as text, say, so can be given as an input to another program. The *halting problem* asks if there's a single program that takes as an input a program P (encoded as text) together with another text input X, and prints "Yes" if P halts when fed the input X, and "No" if it doesn't. Here, you should imagine that X is always a string of characters, which P accepts as an input.

Exercise 2.28. Show that there is no such program. *Hint: Hoping for a contradiction, suppose there is a program that solves the halting problem, and call this program* Halt. *Write a program* Paradox *that takes as input a program P, then uses* Halt *to figure out if P halts or not when fed itself (i.e. its own source code) as its input, then prints "It doesn't halt on itself" if P doesn't halt when fed the input P, and prints "It halts!!!" repeatedly forever if P halts when fed the input P. What's the paradox?*

3 Induction

Here's a well known formula for the sum of the first n natural numbers.

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}. \tag{3}$$

Let's say we want to prove this. For $n = 1$, the formula is $1 = 1(2)/2 = 1$, which is true. For $n = 2$, the formula is $1 + 2 = 3 = 2(2 + 1)/2$, which is true. How would you prove a statement like this in general? You could keep checking individual cases like this, but you'll never be able to verify case-by-case that the formula holds for all of the infinitely many natural numbers n.

The principle of *induction* says that once you know a particular case of this result, you can prove it for all higher cases by showing that whenever a particular case is true, so is the next one. Namely, note that

$$1 + 2 + \cdots + k + (k + 1) = (1 + 2 + \cdots + k) + (k + 1),$$

so if we already know that (3) holds for the case of $n = k$, this becomes

$$\begin{aligned} &= \frac{k(k+1)}{2} + (n+1) \\ &= \frac{k(k+1) + 2(k+1)}{2} \\ &= \frac{(k+1)(k+2)}{2}, \end{aligned}$$

which is exactly the right hand side of (3) in the case that $n = k + 1$. So, if we know that (3) is true when $n = k$, it also is true when $n = k + 1$. Together with the fact that the $n = 1$ case holds, it should be intuitive that this means that (3) is true for all n. Indeed, it holds for $n = 1$, hence it holds for $n = 2$, hence for $n = 3$, etc...

To formalize this logic, set

$$P(n) := \text{“}1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}\text{ ”}.$$

That is, $P(n)$ is *the statement that the above equality is true.* Note that $P(n)$ is not equal to $1+2+3+\cdots+n$, which is a number. It is a statement. So, saying $P(n) = 6$ makes no sense, but "$P(n)$ is true" and "$P(n)$ is false" do make sense. For instance, $P(3)$ is the statement that $1 + 2 + 3 = 3(4 + 1)/2$.

So, to prove (3), we first noted that $P(1)$ is true. We then showed that whenever $P(k)$ is true, the statement $P(k+1)$ is true, and appealed to our intuition to then say

that $P(n)$ is true for all n. Here is a mathematical statement that says "We believe this approach should work!". You should regard it not as a theorem, but as an *axiom*, a statement that we think is intuitively obvious, and that we decide to assume is true in order to be able to do interesting math.

The Principle of Mathematical Induction (PMI). *Let $P(1), P(2), P(3), \ldots$ be a sequence of statements, one for each natural number. Assume the following:*

(a) $P(1)$ is true.

(b) For each $k \in \mathbb{N}$, if $P(k)$ is true, then $P(k+1)$ is true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

There is also a standard format for 'induction proofs', i.e. arguments that prove a statement about all natural numbers n using the PMI. The first step is the *base case*, in which you prove that your statement is true for $n = 1$. The second step is the *inductive case* or *inductive step*, in which you show that whenever your statement is true for $n = k$, it is also true for $n = k+1$. Here is how to format the example above.

Theorem 3.1. *For all $n \in \mathbb{N}$, we have* $1 + 2 + 3 + \cdots + n = \dfrac{n(n+1)}{2}$.

Proof by Induction. Set $P(n)$ to be the statement that $1+2+3+\cdots+n = \dfrac{n(n+1)}{2}$

For the base case, we note that $1 = \frac{1(1+1)}{2}$, so $P(1)$ is true.
For the inductive case, assume that $P(k)$ is true, so

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}.$$

Then we have

$$1 + 2 + \cdots + k + (k+1) = \ (\textit{insert calculation}) \ = \frac{(k+1)(k+2)}{2},$$

implying $P(k+1)$ is true. By the PMI, $P(n)$ is true for all n. □

You can also write induction proofs more briefly without referring explicitly to $P(n)$ or the PMI. Omitting the calculation, the general format would be as follows.

Proof by induction. For the base case, we note that $1 = \frac{1(1+1)}{2}$.

For the inductive case, assume that

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}.$$

Then we have

$$1 + 2 + \cdots + k + (k+1) = \ (\textit{insert calculation}) \ = \frac{(k+1)(k+2)}{2},$$

so the claim is true by induction. □

However, I would encourage you to write your first few induction proofs as in the first example, to make sure you are correctly identifying the statement $P(n)$.

3.1 Induction exercises

Exercise 3.2. Prove that for all $n \in \mathbb{N}$,

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

It's sometimes convenient to have the base case in the PMI not be $n = 1$, but some other integer. Indeed, suppose that $m \in \mathbb{Z}$ and that

(a) $P(m)$ is true, and

(b) for all integers $k \geq m$, if $P(k)$ is true, then $P(k+1)$ is true.

Then the same intuitive argument as above says that $P(n)$ should be true for all $n \geq m$. You can also see this directly from our earlier statement of the PMI, by applying the PMI to the new sequence of statements $Q(n) = P(n + m - 1)$.

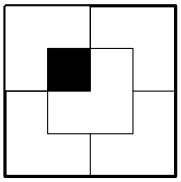
Exercise 3.3. (a) Show by induction that for all $n \geq 4$, we have $2n + 1 \leq 2^n$.

(b) Using induction again, and also part (a), show that for $n \geq 4$ we have $n^2 \leq 2^n$.

If m, n are natural numbers, we say that $m|n$, read m *divides* n, if there is some other natural number k such that $mk = n$.

Exercise 3.4. Show that if $n \in \mathbb{N}$, then $3 \mid 4^n - 1$. *Hint: for the inductive case, start with* $4^{k+1} - 1$ *and try to transform it into something involving* $4^k - 1$ *and things obviously divisible by* 3. *If you get stuck, remember that* $4 = 3 + 1$.

Exercise 3.5. A *triomino* is a 2×2 square with one square removed. Show that for any positive integer n, any $2^n \times 2^n$ checkerboard with one square removed can be tiled by triominos. *Note: in the inductive case, you* start *with a $2^{n+1} \times 2^{n+1}$ board with one square removed, and you don't get to pick which square it is. Do not start with a $2^n \times 2^n$ board and try to augment it. (Why?)*



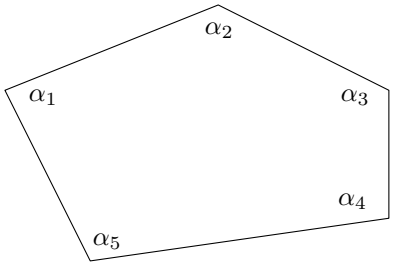
a tiling of a $2^2 \times 2^2$ checkerboard minus a square

Recall that the *power set* of a set A is the set $\mathcal{P}(A) := \{B \mid B \subset A\}$.

Exercise 3.6. Let A be a set with n elements. Show using induction that the power set $\mathcal{P}(A)$ has 2^n elements, by setting $P(n)$ to be the statement *'for every set A of n elements, $\mathcal{P}(A)$ has 2^n elements'.*

Be careful with the beginning of the induction step—you need to *start with* a set of $n + 1$ elements.

Exercise 3.7. A standard fact from Euclidean geometry is that the interior angles of a triangle (measured in radians) sum to π. Use this to prove that for $n \geq 3$, the sum of all the interior angles in a convex n-gon is $\pi(n - 2)$.



$\sum_{i=1}^{5} \alpha_i = \pi(5 - 2) = 3\pi.$

Exercise 3.8 (Monochromatic cows)**.** What is wrong with the following proof?

Theorem. *All cows are the same color.*

Proof. We will show by induction that any group of n cows is monochromatic. By showing this is true for all n, we will conclude that all cows are the same color.

Base case. If there is only one cow in a group, then clearly all cows in that group have the same color.

Inductive case. Assume that any group of n cows is monochromatic. Consider a group consisting of $n+1$ cows. First, exclude the last cow and look only at the first n cows; all these are the same color since any group of n is monochromatic. Likewise, exclude the first cow and look only at the last n cows. These too, must also be of the same color. Therefore, the first cow in the group is of the same color as the cows in the middle, who in turn are of the same color as the last cow. Hence the first cow, middle cows, and last cow are all of the same color, and we have proven that our group of $n+1$ cows is monochromatic. By induction, any group of cows is monochromatic, so all cows are the same color. □

Exercise 3.9. There are n lions on an island, all lined up in front of a piece of meat. The meat is tranquilized, and a lion that eats the meat gulps it down in one bite, then falls asleep for a whole day, with the tranquilizer coursing through its blood, so essentially it becomes the tranquilized meat itself for that day. The lions are super-intelligent and all-knowing (for instance, they know the meat is tranquilized, and they know what will happen if they eat it), they will not cooperate or share the meat, they would rather starve than be eaten, and they are ultra-disciplined, so if they do eat then they do it in the order they lined up in, and they will not cut in line or fight. What is going to happen? *Hint: you might want to try the cases of $n = 1, 2, 3$ first in order to get a feel for things. Then formulate your conjecture about what is going to happen, and prove it using induction.*

Exercise 3.10. Define a sequence of real numbers x_n by setting $x_1 = 1$ and $x_n = \sqrt{x_{n-1} + 1}$ for all $n > 1$. Prove that $x_{n+1} > x_n$ for all $n \in \mathbb{N}$.

Exercise 3.11. If $n \in \mathbb{Z}$ and $n \geq 0$, set $g_n = 2^{2^n} + 1$. Prove by induction that

$$g_0 g_1 \cdots g_{n-1} = g_n - 2$$

for all $n \geq 1$. The numbers g_n were first studied by Pierre de Fermat, who conjectured that they are always prime. While $g_0, \ldots, g_4$ are prime, unfortunately

$$g_5 = 4294967297 = 641 \cdot 6700417,$$

so Fermat's conjecture is false.

Exercise 3.12. If $n \in \mathbb{N}$, show that $\sum_{i=1}^{2n} (-1)^i i = n$.

Exercise 3.13. Show by induction that for any $n \geq 2$, any expression with n blanks and inequality signs between each consecutive pair of blanks can be filled in with the numbers $1, \ldots, n$, in some order, so that all the inequalities are true.

As an example, say I give you the expression

$$\_\_ < \_\_ > \_\_ > \_\_ < \_\_ > \_\_$$

There are 6 blank spaces here, and if we fill in the blanks like

$$3 < 4 > 2 > 1 < 6 > 5$$

then all the inequalities are actually true; for instance $3 < 4$ and $4 > 2$. You have to prove (by induction) that you can *always* do this, no matter the inequalities.

3.2 Strong Induction

Strong induction is a variation of mathematical induction in which instead of assuming that $P(k)$ is true, one gets to assume that $P(1), \ldots, P(k)$ are *all* true.

Principle of Strong Induction (PSI). *Suppose that for each $n \in \mathbb{N}$, we have a statement $P(n)$, and that the following two properties hold:*

(a) $P(1)$ is true.

(b) For every $k \in \mathbb{N}$, if $P(1), \ldots, P(k)$ are all true, then $P(k+1)$ is true as well.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Intuitively, the reason you should believe this is the same as for the PMI. Assuming $P(1)$ is true, you apply (b) with $k = 1$ and get $P(2)$ is true. Now both $P(1)$ and $P(2)$ are true, so (b) says $P(3)$ is true. Then $P(1), P(2), P(3)$ are true, so (b) says $P(4)$ is true. Continuing, you would expect that $P(n)$ is true for all n.

Here's a first example.

Definition 3.14. If $m, n \in \mathbb{N}$, we say that m *divides* n, written $m|n$, if $n = km$ for some natural number $k \in \mathbb{N}$. Here, m is called a *divisor* of n. A natural number $p \geq 2$ is *prime* if its only divisors are 1 and itself. Any natural number $n \geq 2$ that is not prime can be written as $n = km$ for some $k, m < n$, and is called *composite*.

The Prime Factorization Theorem. *Any natural number $n \geq 2$ can be written as a product $n = p_1 \cdots p_k$, where $p_1, \ldots, p_k$ are all prime.*

Proof. For the base case $n = 2$, we have that 2 is itself prime.

For the inductive case, assume that for $n = 1, \ldots, k$, we have that n can be written as a product of primes. We want to show the same for $k+1$. If $k+1$ is prime, we're done. If it's composite, write $k + 1 = ab$ where $a, b \in \mathbb{N}$ and $1 \leq a, b \leq k$. By the inductive assumption, can write a, b as products of primes:

$$a = p_1 \cdots p_k, \quad b = p_{k+1} \cdots p_m.$$

But then $k + 1 = ab = p_1 \cdots p_m$ is a product of primes. □

Try to prove the following with strong induction.

Exercise 3.15. Say that a rectangular chocolate bar is made up of n squares of chocolate. We can break the bar into two smaller bars along any line that runs between the squares, and then break those further. Show any sequence of breaks that results in individual squares at the end had $n - 1$ total breaks.

The *Fibonacci sequence* f_n is defined by letting $f_1 = f_2 = 1$ and setting $f_n = f_{n-1} + f_{n-2}$ for $n \geq 3$. The first few terms are $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$

Fact 3.16. *For all $n \in \mathbb{N}$, we have $f_n \leq 2^n$.*

There's one subtlety to the proof. To illustrate it, let's give a wrong proof first.

Proof with small mistake. For a base case, $f_1 = 1 \leq 2^1$.

For the inductive step, assume that $f_n \leq 2^n$ for $n = 1, \ldots, k$. Then we have

$$f_{k+1} = f_k + f_{k-1} \leq 2^k + 2^{k-1} \leq 2^k + 2^k = 2^{k+1}. \qquad \square$$

So what's the subtlety? If we're using $n = 1$ as our base case, we need the inductive step to start applying immediately with $k = 1$. However, when $k = 1$ we're writing $f_2 = f_1 + f_0$, and there's no 0^{th} Fibonacci number in our definition. To fix this, we can just verify both the claim for *both* $n = 1$ and $n = 2$, and then start the induction at 2 instead of 1. We format this as follows.

Correct proof. For bases cases, $f_1 = 1 \leq 2^1$ and $f_2 = 1 \leq 2^2$. For the inductive step, assume that $k \geq 2$ and that $f_n \leq 2^n$ for $n = 1, \ldots, k$. Then we have

$$f_{k+1} = f_k + f_{k-1} \leq 2^k + 2^{k-1} \leq 2^k + 2^k = 2^{k+1}. \qquad \square$$

Here's a similar exercise, where you'll again need two base cases in order for the recurrence relation defining the Fibonacci numbers to kick in in the inductive step.

Exercise 3.17. Show that for every $n \in \mathbb{N}$, f_n is even if $3|n$, and odd otherwise.

Exercise 3.18. Let $\alpha = \frac{1+\sqrt{5}}{2}, \beta = \frac{1-\sqrt{5}}{2}$. Show that for all $n \in \mathbb{N}$, we have

$$f_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}.$$

Hint: first prove that α, β both satisfy the equation $x^2 = 1 + x$.

A *polynomial* is a function of the form $f(x) = a_n x^n + \cdots + a_1 x + a_0$, where $a_n, \ldots, a_0 \in \mathbb{R}$. By discarding some of the terms, we can assume that the *leading coefficient* a_n of f is nonzero, and then the number n is called the *degree* of f. A real number x is called a *root* of f if $f(x) = 0$.

Theorem 3.19. *Any nonzero degree n polynomial has at most n roots.*

For example, the polynomials $f(x) = x^2 - 2$, $g(x) = x^2$, $h(x) = x^2 + 1$ all have degree 2. The roots of f are $x = \pm\sqrt{2}$, the polynomial g has only $x = 0$ as a root, and h has no roots. In all cases, the number of roots is at most 2, the degree. The only exception is that the zero polynomial $z(x) = 0$ has degree 0, which is at most 2, but *every* real number is a root of z, so z has infinitely many roots.

Exercise 3.20. Prove Theorem 3.19 by filling in the following outline.

Proof Outline. We'll do a strong induction proof. For the base case, consider a nonzero, degree zero polynomial $f(x) = a_0$. Then $f(x)$ has zero roots, since ($\cdots$ *insert solution (a) here* $\cdots$)

For the inductive case, suppose that for $n = 1, \ldots, k$, any nonzero degree n polynomial has at most n roots. Now let $f(x) = a_{k+1}x^{k+1} + \cdots + a_1x + a_0$ be a nonzero degree $k+1$ polynomial, and hoping for a contradiction, assume that $f(x)$ has $k+2$ distinct roots, which we'll call $x_1, \ldots, x_{k+2}$. Consider the polynomial

$$g(x) = f(x) - a_{k+1}(x - x_1)\cdots(x - x_{k+1}).$$

The degree of g is at most k, since ($\cdots$ *insert solution (b) here* $\cdots$). Moreover, g is not the zero polynomial, since ($\cdots$ *insert solution (c) here* $\cdots$).

So, by the inductive hypothesis, g has at most k roots. But ($\cdots$ *insert solution (d) here* $\cdots$). So this is a contradiction. Hence our initial assumption that f had $k+2$ roots had to be wrong, so f has at most $k+1$ roots as desired. This finishes the inductive case. □

Exercise 3.21. Use strong induction to prove that any non-negative integer n can be written as a sum of distinct powers of 2. In other words, show that for each n,

$$n = 2^{a_l} + 2^{a_{l-1}} + \cdots + 2^{a_1}$$

for distinct integers $a_l > a_{l-1} > \cdots > a_1 \geq 0$. *Hint: for the inductive step, when you are trying to prove that $k+1$ can be written like this, start by letting 2^a be the largest power of 2 that is less than or equal to $k+1$. And remember, you need to prove that all the powers of 2 are distinct, i.e. different!*

Note that if you then write a string of 1's and 0's where the 1's are in the $a_l, \ldots, a_1$ places, counting from the right, you get the binary expansion of n. For instance, $43 = 2^5 + 2^3 + 2^1 + 2^0$ is 101011 in binary. So, this exercise is showing that any positive integer can be written in binary. Note that $n = 0$ can be written as an 'empty sum' with no terms.

The game of Nim involves two players and two piles of pennies. On each turn, a player removes some non-zero number of pennies from one of the piles. A player loses if on their turn, there are no pennies left in either pile.

Exercise 3.22. Show that if the two piles initially have the same number of pennies, there is a strategy in which the second player can always win!

You should prove using strong induction that for each $n \in \mathbb{N}$, the second player always has a winning strategy in a Nim game in which both players start with n pennies. To get some intuition, try doing the $n = 0, n = 1, n = 2$ cases first.

3.3 The Well Ordering Principle

Here is another reasonable sounding statement about natural numbers that can be used in place of induction in many proofs.

The Well Ordering Principle (WOP). *Every nonempty subset of* $\mathbb{N}$ *has a least element.*

Here a *least element* is an element x such that $x \leq y$ for any other element of the set. Note that the WOP does not hold if we replace $\mathbb{N}$ with $\mathbb{Z}$: for instance, $\mathbb{Z}$ itself is not empty, and it does not have a least element, since for any $x \in \mathbb{Z}$, we have $x > x - 1$, and $x - 1 \in \mathbb{Z}$ too.

Exercise 3.23. Does every nonempty subset of the closed interval $[0, 1] \subset \mathbb{R}$ have a least element?

Here's how to use the WOP in place of induction. Say you have a statement $P(n)$ about natural numbers that you want to show is true for all n. If this is *not* the case, then the set $S = \{n \mid P(n) \text{ is false}\}$ is not empty, so the WOP says that it has a least element. One then takes this least element n such that $P(n)$ is false, and tries to come up with some sort of contradiction.

For example, here's a proof of the Prime Factorization Theorem using the WOP.

The Prime Factorization Theorem. *Any natural number* $n \geq 2$ *can be written as a product* $n = p_1 \cdots p_k$*, where* $p_1, \ldots, p_k$ *are all prime.*

Proof. Suppose the theorem is not true. Then the WOP implies that there is a *least* natural number x that can't be written as products of primes. If x is prime, it's a product of 1 prime, which is a contradiction. So, x is composite. Write $x = ab$ where

$a, b \in \mathbb{N}$ and $1 \leq a, b < x$. Since both $a, b < x$, they are not in S, and hence we can write them as products of primes:

$$a = p_1 \cdots p_k, \quad b = p_{k+1} \cdots p_m.$$

But then $x = ab = p_1 \cdots p_m$ is a product of primes, a contradiction. □

In fact, any induction theorem can be worded to use the WOP instead.

Exercise 3.24. Use the WOP to prove that for all $n \in \mathbb{N}$, we have

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

Hint: assuming the claim isn't true, take the smallest n where the equality fails.

The WOP is also useful when one wants to prove there is an example of something that is in some sense simpler than other such somethings. For example, in the following problem, we want to show the existence of a cycle of length 3, which is the shortest that a cycle can be. So, we use the WOP to say there *is* a shortest cycle, and then prove that it must have length 3.

Exercise 3.25. A *round robin* is a tournament in which each player p plays each other player q exactly once. We write $p > q$ if player p wins, and $q > p$ otherwise. A *cycle* is a sequence of players $p_1, \ldots, p_k$ such that $p_1 < p_2 < \cdots < p_k < p_1$. The number k is called the *length* of the cycle. Show that in every round robin tournament, if there is a cycle, then there is a cycle of length 3. *Hint: Suppose we are given a tournament that has a cycle. The WOP says that there is a* shortest *cycle $p_1 < \cdots < p_k < p_1$.*

Here's an example with a similar flavor.

Exercise 3.26. If m, n are natural numbers, the fraction $\frac{m}{n}$ is said to be *in lowest terms* if there is no natural number $d \geq 2$ that divides both m and n. Given any fraction $\frac{a}{b}$, show that there is another fraction $\frac{m}{n}$ in lowest terms such that $\frac{a}{b} = \frac{m}{n}$. *Hint: given a fraction $\frac{a}{b}$, the WOP says that there is a fraction $\frac{m}{n}$ with $\frac{a}{b} = \frac{m}{n}$ where the numerator m is as small as possible. Show that this $\frac{m}{n}$ is in lowest terms.*

3.4 Equivalence of the axioms

Above, we introduced the PMI, the PSI and the WOP as *axioms*, statements that we decide are intuitively true and that we want to be able to assume in order to do interesting math. However, it turns out that all three axioms are *equivalent*, in the sense that you can use either one to prove the other.

Exercise 3.27. Show that the WOP implies the PMI. *Hint: start with some property $P(n)$ as given in Theorem 1, and assume the PMI fails for P. Your goal is to define some nonempty set $S \subset \mathbb{N}$, depending on P, and use the fact that S has a least element to get a contradiction.*

Exercise 3.28. Use the PMI to prove the PSI.

Exercise 3.29. Use the PSI to prove the WOP.

4 Number Theory

We will start in now with a bit of number theory. As always, $\mathbb{Z}$ will denote the set of integers. You are allowed to assume that all usual properties of addition, subtraction and multiplication of integers work in the usual ways, and interact as expected with the usual order $<$ on $\mathbb{Z}$. For instance, addition is commutative and multiplication distributes over addition, and if $a \leq b$ then $a + c \leq b + c$ for all c.

4.1 Division with remainder

In high school or earlier, you probably considered statements like *"if you divide 8 by 3, you get 2, but with a remainder of 2."* Here's a theorem that makes sense of this.

Theorem 4.1 (Division with remainder)**.** *If $a, b \in \mathbb{Z}$ and $b > 0$, then there exist unique integers q and r such that $a = bq + r$ and $0 \leq r < b$.*

Intuitively, in the setting above, dividing a by b gives q with a remainder of r.

Exercise 4.2. If $a = 7$ and $b = 314$, find q and r. What about if $a = -11$ and $b = 30$?

The following two exercises complete the proof of Theorem 1.

Exercise 4.3. Fix $b > 0$. Via strong induction, show that every integer $a \geq 0$ can be written as $a = qb + r$ for some integer r with $0 \leq r < b$. Then explain why the same result follows for negative a.

Proof. First, if $a < b$, we can take $a = 0 \cdot b + r$, with $r = a$. We can take all these cases as the base cases. For the inductive case, suppose that $k \geq b$ and the claim is true for all $0 \leq a < k$. Then $0 \leq a = k - b < k$, so we have $k - b = qb + r$ with $0 \leq r < b$. But then $k = (q + 1)b + r$, so we're done. □

Exercise 4.4. Suppose q_1, r_1 and q_2, r_2 are integers with $0 \leq r_1, r_2 < b$ and

$$a = bq_1 + r_1 = bq_2 + r_2.$$

Show that $q_1 = q_2$ and $r_1 = r_2$. *Hint: show that $b(q_1 - q_2) = r_2 - r_1$ and then figure out what range of numbers the right side lies in.*

4.2 Greatest common divisors and $\mathbb{Z}$-linear combinations

Recall that if $a, b \in \mathbb{Z}$, then a *divides* b if there is some integer k with $b = ak$, and we call a a *divisor* of b, while b is a *multiple* of a. We write $a|b$ when this is the case.

Definition 4.5. Let $a, b \in \mathbb{Z}$, not both zero. A *common divisor* of a and b is defined to be any integer c such that $c|a$ and $c|b$. The largest common divisor is usually called the *greatest common divisor* of a and b, and is denoted $gcd(a, b)$.

Exercise 4.6. List all the common divisors of 18 and 24, and find $gcd(18, 24)$.

Given $a, b \in \mathbb{Z}$, an expression of the form $xa + yb$, where $x, y \in \mathbb{Z}$, is called a $\mathbb{Z}$-*linear combination* of a, b. The numbers x, y are called the *coefficients* of the linear combination. For instance, the following are $\mathbb{Z}$-linear combinations of 3 and 7:

$$-2 \cdot 3 + 5 \cdot 7 = 36, \qquad 0 \cdot 3 - 1 \cdot 7 = -7, \qquad 100 \cdot 3 + 2 \cdot 7 = 314.$$

Exercise 4.7. Suppose that $a, b \in \mathbb{Z}$ and suppose that n is a $\mathbb{Z}$-linear combination of a, b. Show that any common divisor of a, b is a divisor of n.

In Exercise 4.6, you computed the gcd of 18 and 24 by just listing all the common divisors and taking the biggest one. This is not an effective method of computing gcd's of very large numbers though, since it's computationally intensive to find all divisors of a large number. In fact, this computational difficulty is what makes many common encryption schemes work, e.g. those used to encrypt web traffic.

Fortunately, there's a better way to compute gcds. The key is:

Exercise 4.8. If $a, b, q, r \in \mathbb{Z}$ and $a = qb + r$, then $gcd(a, b) = gcd(b, r)$. *Hint: show every common divisor of a, b is a common divisor of b, r, and vice versa.*

How do we use this to compute gcds? Well, if you start with two (say, positive) numbers a, b, you can assume a is the bigger one, and then write $a = qb + r$ using division with remainder. The exercise says $gcd(a, b) = gcd(b, r)$, and this is easier to compute, since r is smaller than a. Moreover, if you then use division with remainder and the exercise *again*, you can perhaps reduce the computation of $gcd(b, r)$ to an even simpler computation, etc...

For example, say we want to compute $gcd(93, 36)$. We write:

$$\begin{aligned}
93 &= 2 \cdot 36 + 21 && \leftarrow \text{ reduces it to computing } gcd(36, 21)\\
36 &= 1 \cdot 21 + 15 && \leftarrow \text{ reduces it to computing } gcd(21, 15)\\
21 &= 1 \cdot 15 + 6 && \leftarrow \text{ reduces it to computing } gcd(15, 6)\\
15 &= 2 \cdot 6 + 3 && \leftarrow \text{ reduces it to computing } gcd(6, 3)\\
6 &= 2 \cdot 3 + 0 && \leftarrow \text{ says that 3 divides 6, so actually } gcd(6, 3) = 3!
\end{aligned}$$

Tracing through all these steps and applying the exercise each time then shows that $gcd(93, 36) = 3$. This procedure is called the *Euclidean algorithm*, after the Greek mathematician Euclid, who lived around 300 B.C.E. Here's a formal statement.

Theorem 4.9 (The Euclidean Algorithm)**.** *Let $a, b \in \mathbb{Z}$ be positive integers, not both zero. Then we can apply division with remainder repeatedly to find q_i, r_i as follows:*

$$\begin{array}{rcll} a & = & bq_1 + r_1 & 0 < r_1 < b \\ b & = & r_1q_2 + r_2 & 0 < r_2 < r_1 \\ r_1 & = & r_2q_3 + r_3 & 0 < r_3 < r_2 \\ & \vdots & & \\ r_{k-2} & = & r_{k-1}q_k + r_k & 0 < r_k < r_{k-1} \\ r_{k-1} & = & r_kq_{k+1}, & \end{array}$$

and where $r_k = gcd(a, b)$.

Exercise 4.10. Use the Euclidean algorithm to find $gcd(18, 24)$, $gcd(75, -21)$ and $gcd(145, 690)$.

Exercise 4.11. Using Exercise 4.8, write a formal proof that the Euclidean Algorithm works. There's almost a proof above, but why does the algorithm always end in finitely many steps with a remainder of zero?

We say a, b are *relatively prime* or *co-prime* if $gcd(a, b) = 1$.

Exercise 4.12. Let f_n be the Fibonacci sequence. Show that f_n and f_{n+1} are relatively prime for all positive integers n.

Exercise 4.13. Show that any integer $n \geq 8$ can be written as $n = x \cdot 3 + y \cdot 5$ for some nonnegative $x, y \in \mathbb{Z}$. *Hint: use strong induction on n.*

Note that in Exercise 4.13, we're only using nonnegative coefficients. So, the exercise is showing, for instance, that any possible amount n of postage that is at least 8 cents can be made using some combination of 3 cent and 5 cent stamps. In general, though, $\mathbb{Z}$-linear combinations can have negative coefficients.

4.3 Bézout's Identity

The following is one of the most important properties of gcds. We'll use it in the next section to prove certain fundamental facts about prime numbers.

Theorem 4.14 (Bézout's Identity)**.** *If $a, b \in \mathbb{Z}$, not both zero, then $gcd(a, b)$ is a $\mathbb{Z}$-linear combination of a, b.*

Exercise 4.15. Find x, y such that $gcd(21, 27) = x21 + y27$.

You can prove Theorem 4.14 using the following exercise, in combination with the Euclidean algorithm.

Exercise 4.16. Suppose that $a, b \in \mathbb{Z}$. If m, n are both $\mathbb{Z}$-linear combinations of a, b and $m = qn + r$, where $q, r \in \mathbb{Z}$, then r is also a $\mathbb{Z}$-linear combination of a, b.

Proof of Theorem 4.14. You can use induction and Exercise 4.16 to say that all the remainders that appear in the Euclidean algorithm are all $\mathbb{Z}$-linear combinations of the original a, b. So, this is the case for the final nonzero remainder, which is $gcd(a, b)$. □

One can use the proof above to *explicitly* find the coefficients x, y such that $xa+yb = gcd(a, b)$. It's almost more confusing to write an explicit procedure for this than to do it: the point is just to go through the Euclidean algorithm, and write each of the remainders that comes up as a $\mathbb{Z}$-linear combination of a, b, using the $\mathbb{Z}$-linear combinations for the previous remainders.

Here's a simple example. To compute $gcd(36, 26)$, we proceed as follows:

$$\begin{aligned} 36 &= 1 \cdot 26 + 10 \\ 26 &= 2 \cdot 10 + 6 \\ 10 &= 1 \cdot 6 + 4 \\ 6 &= 1 \cdot 4 + 2 \\ 4 &= 2 \cdot 2, \end{aligned}$$

so $gcd(36, 26) = 2$. To write 2 as a $\mathbb{Z}$-linear combination, you

$$\begin{aligned} 10 &= 36 - 26 \\ 6 &= 26 - 2 \cdot 10 = 26 - 2(36 - 26) = -2 \cdot 36 + 3 \cdot 26 \\ 4 &= 10 - 6 = (36 - 26) - (-2 \cdot 36 + 3 \cdot 26) = 3 \cdot 36 - 4 \cdot 26 \\ 2 &= 6 - 4 = (-2 \cdot 36 + 3 \cdot 26) - (3 \cdot 36 - 4 \cdot 26) = -5 \cdot 36 + 7 \cdot 26. \end{aligned}$$

Exercise 4.17. Use the Euclidean algorithm to find $gcd(105, 135)$ and x, y such that $x105 + y135 = gcd(105, 135)$.

Here's one corollary of Bézout's Identity. To state it, let $a, b \in \mathbb{Z}$ and let

$$S(a,b) := \{xa + yb \mid x, y \in \mathbb{Z}\}$$

be the set of all $\mathbb{Z}$-linear combinations of a and b.

Exercise 4.18. Using Bézout's Identity, show that

$$S(a,b) = \{n \cdot gcd(a,b) \mid n \in \mathbb{Z}\},$$

that is $S(a,b)$ is exactly the set of all integer multiples of $gcd(a,b)$.

Above, we used Bézout to understand $S(a,b)$. You can also work in the opposite direction, proving Bézout's Identity by understanding from scratch the properties of $S(a,b)$. So, let's consider $S(a,b)$ just as the set of all $\mathbb{Z}$-linear combinations of a, b, and pretend we don't know about Exercise 4.18.

First, note that as long as a, b aren't both zero, the set $S(a,b)$ contains positive integers, e.g. it contains $\pm a$ and $\pm b$, and one of these four is positive. So, $S(a,b)$ contains a *least* positive integer $d(a,b)$, by the well ordering principle.

Exercise 4.19. Show that $d(a,b)$ is a common divisor of a, b.

Exercise 4.20. Suppose that c is a common divisor of a, b. Show that $c|d(a,b)$, and therefore $c \leq d(a,b)$.

Exercise 4.21. Use Exercises 4.19 and 4.20 to prove Bézout's Identity.

Here's a similar exercise, for extra practice.

Exercise 4.22. Suppose $S \subset \mathbb{N} \cup \{0\}$ is a nonempty subset such that whenever $x, y \in S$ and $y \geq x$, then $y - x \in S$ as well.

(a) Show with a quick induction proof that for all $q \in \mathbb{N}$, we have that if $\ell, x \in S$ and $q\ell \leq x$, then $x - q\ell \in S$.

(b) Show that the least nonzero element $\ell \in S$ divides every element of S.

4.4 More primes

Recall that a natural number $n \geq 2$ is *prime* if its only positive divisors are 1 and itself, and otherwise it is *composite*. Note that if n is composite, then $n = ab$ for two integers a, b with $2 \leq a, b \leq n$. Previously, we showed:

Theorem 4.23. *Any natural number n can be written as a product*

$$n = p_1 \cdots p_k,$$

where $p_1, \ldots, p_k$ are all prime.

Often, when we write n as such a product, we say that we have *factored* n. The terms in the product are the *factors*.

Exercise 4.24. If p is prime and p does not divide $a \in \mathbb{N}$, show that $gcd(p, a) = 1$.

Exercise 4.25. Let $a, b, n \in \mathbb{N}$ and assume $gcd(a, n) = 1$ and $n|ab$. Show that $n|b$. Then conclude that if p is prime and $p|ab$, then $p|a$ or $p|b$.

Exercise 4.26. Let p be prime and $a_1, \ldots, a_k$ be natural numbers. If $p|a_1 \cdots a_k$, show that $p|a_i$ for some i. *Hint: do induction on k.*

In particular, if the a_i in Exercise 4.26 are all prime, then we have that $p = a_i$ for some i. Using this, you should now prove:

Theorem 4.27 (The Fundamental Theorem of Arithmetic)**.** *Every integer $n \geq 1$ may be factored into a product of primes in a unique way up to the order of the factors. In other words, if $n = p_1 \cdots p_k = q_1 \cdots q_l$ are two prime factorizations of n, then $k = l$ and we can reorder the q_i's so that $p_1 = q_1, p_2 = q_2, \ldots, p_k = q_k$.*

Hint: it may be useful to do induction on n. The base case $n = 1$ is trivial.

Exercise 4.28. Suppose that $n = p_1^{e_1} \cdots p_k^{e_k}$, where $p_1, \ldots, p_k$ are distinct primes and the exponents are integers $e_i \geq 0$, and let $d \in \mathbb{N}$. Show that $d|n$ if and only if

$$d = p_1^{j_1} \cdots p_k^{j_k}$$

for some $j_i \in \mathbb{Z}$ with $0 \leq j_i \leq e_i$ for all $i = 1, \ldots, k$. *Hint: you should do this in two directions. First, show that if $d = p_1^{j_1} \cdots p_k^{j_k}$, then $d|n$. Then show that any divisor d of n has this form, using the uniqueness of prime factorizations (Theorem 4.27).*

Exercise 4.29. Suppose that $n = p_1^{e_1} \cdots p_k^{e_k}$ and $m = p_1^{f_1} \cdots p_k^{f_k}$, where $p_1, \ldots, p_k$ are distinct primes and $e_i, f_i \geq 0$. If

$$g := p_1^{\min\{e_1, f_1\}} \cdots p_k^{\min\{e_k, f_k\}},$$

show that $g = gcd(m, n)$. *Hint: use the previous problem.*

Here, $\min\{x, y\}$ is just the minimum of x, y, i.e. whichever one is smaller. Note that in the exercise, we can allow e_i or f_i to be zero. This allows us to apply the exercise to *any* pair of natural numbers n, m. For instance, we can write

$$60 = 2^2 \cdot 3^1 \cdot 5^1 \cdot 7^0, \quad 35 = 2^0 \cdot 3^0 \cdot 5^1 \cdot 7^1,$$

and then $gcd(60, 35) = 2^0 \cdot 3^0 \cdot 5^1 \cdot 7^0 = 5$.

A natural number n is a *perfect square* if there is some $a \in \mathbb{N}$ with $a^2 = n$. Here are some examples of perfect squares and their prime factorizations:

$$9 = 3^2, \quad 64 = 2^6 \quad 36 = 2^2 \cdot 3^2, \quad 400 = 2^4 \cdot 5^2.$$

Exercise 4.30. Show that n is a perfect square if and only if every prime factor occurs an even number of times in the (essentially unique) prime factorization of n.

A real number x is defined to be *rational* if there exist integers p and q such that $x = p/q$ and *irrational* otherwise.

Exercise 4.31. Show that if $n \in \mathbb{N}$ and $\sqrt{n}$ is rational, then n is a perfect square.

So in particular, $\sqrt{2}$ is irrational.

Exercise 4.32. Show that there are infinitely many primes. *Hint: hoping for a contradiction, suppose the only primes are $p_1, \ldots, p_k$ and consider $n = p_1 \cdots p_k + 1$.*

Exercise 4.33. Show that there are infinitely many primes of the form $4n + 3$. *Hint: you might find it useful to show that the product of two numbers of the form $4n + 1$ are also of that form. For example, $5, 9$ and $5 \cdot 9 = 45$ are all of the form $4n + 1$, for $n = 1, 2, 11$, respectively. Note that all primes except 2 are odd, and every odd prime is either of the form $4n + 1$ or $4n + 3$ for some n.*

Finally, try to digest the statement of the following theorem, for inspiration. I don't expect you to prove it, although if you do some sort of medal is in order.

Theorem 4.34 (The prime number theorem)**.** *Given a positive number x, let $\pi(x)$ be the number of primes that are less than or equal to x. Then*

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x} = 1.$$

What does this mean? Well, if x is really large, the theorem implies that

$$\frac{\pi(x)}{x} \approx \frac{1}{\ln(x)}.$$

Here, the left hand side is the *proportion* of the numbers $1, \ldots, x$ that are prime. For instance, it turns out that there are 50,847,534 primes that are less than a billion. The proportion of numbers less than a billion that are prime is then

$$50,847,534/1,000,000,000 = .05847534,$$

while $1/\ln(1,000,000,000) \approx 0.04825494243$, which is pretty close. Note that as $x \to \infty$, the percentage of numbers less than or equal to x that are prime gets smaller and smaller. This is because primes are becoming sparser as you look at bigger numbers, since there are more numbers available to be factors.

5 Complex numbers

The real numbers $\mathbb{R}$ are great, but they could be better in some ways! For instance, some polynomial equations can't be solved, e.g. $x^2 + 1 = 0$ has no real solutions. To fix this, we introduce a new number called i, with the property that $i^2 = -1$, and look at what we get when we throw this in with all the other real numbers.

We'd like be able to add and multiply, so we're forced to also consider expressions of the form $x + iy$, where $x, y \in \mathbb{R}$, which we call *complex numbers*. Define

$$(x + iy) + (x' + iy') := (x + x') + i(y + y'), \tag{4}$$
$$(x + iy)(x' + iy') := (xx' - yy') + i(yx' + xy'). \tag{5}$$

Here, to justify the second equality, expand out the left hand side and use the fact that $i^2 = -1$. We let $\mathbb{C}$ denote the set of all complex numbers.

Fact 5.1. *Defined as in* (4) *and* (8)*, addition and multiplication on* $\mathbb{C}$ *satisfy:*

(a) for $z, w \in \mathbb{C}$*, we have* $z \cdot w = w \cdot z$ *and* $z + w = w + z$*, (commutativity)*

(b) for $z, u, w \in \mathbb{C}$*, we have* $z \cdot (u \cdot w) = (z \cdot u) \cdot w$*, (associativity)*

(c) for $z, u, w \in \mathbb{C}$*, we have* $z \cdot (u + w) = z \cdot u + z \cdot w$*. (distributivity)*

You can try to prove these properties if you like, just using the analogous properties of addition and multiplication of real numbers.

Definition 5.2. Suppose that $z = x + iy \in \mathbb{C}$. Then

- the *real part* of z is $Re(z) := x$,
- the *imaginary part* of z is $Im(z) := y$,
- the *(complex) conjugate* of z is $\bar{z} := x - iy$,
- the *absolute value* or *modulus* of z is $|z| := \sqrt{x^2 + y^2}$.

Geometrically, we imagine $\mathbb{C}$ as the plane. The real numbers form the horizontal axis and the *imaginary numbers* iy, where $y \in \mathbb{R}$, form the vertical axis. If $z \in \mathbb{C}$, then $Re(z)$ and $Im(z)$ are the horizontal/vertical coordinates of z. The absolute value $|z|$ is the distance from the origin to z, and the complex conjugate $\bar{z}$ is obtained by reflecting z through the real axis.

Exercise 5.3. Suppose that $z, w \in \mathbb{C}$. Show that

(a) $|z|^2 = z\bar{z}$,

(b) $\overline{z+w} = \bar{z} + \bar{w}$, and $\overline{z \cdot w} = \bar{z} \cdot \bar{w}$,

(c) $|zw| = |z||w|$. *Hint: use (a) and (b) instead of just writing it out.*

(d) Show that if $z \in \mathbb{C}$ then $|Re(z)| \leq |z|$ and $|Im(z)| \leq |z|$.

(e) Show that if $z \in \mathbb{C}$ then $z + \bar{z} = 2Re(z)$, while $z - \bar{z} = 2iIm(z)$.

Theorem 5.4 (The Triangle Inequality)**.** *If $z, w \in \mathbb{C}$ then $|z+w| \leq |z| + |w|$.*

Try to visualize this geometrically. If you form a triangle in $\mathbb{C}$ with vertices $0, z, z+w$, then the side lengths are $|z|, |w|$, and $|z+w|$, respectively. Any one side length is at most the sum of the other two, which is the statement of the triangle inequality.

Exercise 5.5. Let's prove the triangle inequality algebraically!

(a) Given $v, w \in \mathbb{C}$, show that $|z+w|^2 = |z|^2 + |w|^2 + 2Re(z\bar{w})$. *Hint: it'll be useful to use Exercise 5.3.*

(b) Prove the triangle inequality.

We'd also like to be able to divide complex numbers, so suppose that $z, w \in \mathbb{C}$ and $w \neq 0$. What should z/w be? Well, if division works as expected, we should get

$$\frac{z}{w} = \frac{z}{w} \cdot \frac{\bar{w}}{\bar{w}} = \frac{z \cdot \bar{w}}{w\bar{w}} = \frac{z \cdot \bar{w}}{w\bar{w}} = z \cdot \frac{\bar{w}}{|w|^2}, \tag{6}$$

and here $\bar{w}/|w|^2$ is a complex number divided by a real number, which makes sense as you can just divide the real and imaginary parts of $\bar{w}$ by $|w|^2$ separately. For example,

$$\frac{2+3i}{3+5i} = \frac{2+3i}{3+5i} \cdot \frac{3-5i}{3-5i} = \frac{(2+3i)(3-5i)}{34} = \frac{6+9i-10i-15i^2}{34} = \frac{21}{34} - \frac{1}{34}i.$$

You can then check that if you define division by

$$\frac{z}{w} := z \cdot \frac{\bar{w}}{|w|^2},$$

it satisfies all the usual properties you'd like, for instance

$$\frac{z}{w} + \frac{u}{v} = \frac{zv + uw}{wv}, \quad \frac{z}{w} \cdot \frac{u}{v} = \frac{zu}{wv} \tag{7}$$

Exercise 5.6. Compute $(1+2i)/(4-2i)$.

Exercise 5.7. Prove one of the two properties in (7).

Finally, we mention the following important theorem, which illustrates why complex numbers are sometimes better than real numbers.

Theorem 5.8 (The Fundamental Theorem of Algebra)**.** *Every nonconstant polynomial $p(z) = a_n z^n + a_{n-1}x^{n-1} + \cdots + a_0$, where $a_i \in \mathbb{C}$, has a root $z \in \mathbb{C}$.*

The nonconstant assumption rules out polynomials like $p(z) = 5$, which certainly has no roots. The proof is a bit too difficult for this class, but if you're interested, then take a complex analysis class! Note that the theorem isn't true with $\mathbb{R}$ instead of $\mathbb{C}$, since $p(x) = x^2 + 1$ has no roots. Indeed, we started the section by saying that this was the case, so we should throw in another number i that is a solution to this equation. But we could just as well have started with $x^4 - 6x^3 + 15x^2 - 18x + 10 = 0$, which it turns out also doesn't have any real solutions, and tried to add in a solution to this. The FTA says that actually, as soon as we add in i, we have solutions to *all* nonconstant polynomial equations.

5.1 The exponential function and polar coordinates

Definition 5.9. If $z \in \mathbb{C}$, we define $e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \in \mathbb{C}$.

Why does this series converge? It's essentially the same as proving convergence when z is a real number, which you probably did in a Calculus class at some point. One just has to develop convergence of sequences and series for complex numbers instead of real numbers, and then one can use the ratio test, for instance, which for complex series says that $\sum_{n=0}^{\infty} z_n$ converges if $\lim_{n\to\infty} |z_{n+1}|/|z_n| < 1$.

Fact 5.10. *If $z, w \in \mathbb{C}$, we have $e^z e^w = e^{z+w}$.*

We'll accept the above on faith for the moment. Its proof is not so hard, but it would take us a bit far afield.

Theorem 5.11 (Euler's Theorem)**.** *If $\theta \in \mathbb{R}$, we have $e^{i\theta} = \cos(\theta) + i\sin(\theta)$.*

Proof. We just manipulate the series definition of $e^{i\theta}$ into a form involving the Taylor series expansions of $\cos(\theta)$ and $\sin(\theta)$, as follows:

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) + \\ &= \cos(\theta) + i\sin(\theta). \quad \square \end{aligned}$$

Corollary 5.12. $e^{i\pi} = -1$

Proof. Just plug in $\theta = \pi$ to Euler's Theorem. $\square$

More generally, if $r > 0$ and $\theta \in \mathbb{R}$, consider the complex number

$$z = re^{i\theta} = r\cos(\theta) + ir\sin(\theta).$$

Here, $|z| = r$, and θ measures the angle from the positive real axis to the the line segment from 0 to z. Since we can pick r and θ arbitrarily, we have:

Fact 5.13. *Every $z \in \mathbb{C}$ can be written as $z = r \cdot e^{i\theta}$ for some $r \geq 0$ and $\theta \in \mathbb{R}$.*

Here, $z = r \cdot e^{i\theta}$ is called writing z in *polar coordinates*.

Exercise 5.14. (a) If $re^{i\theta} = 1$, show that $r = 1$ and $\theta = 2\pi n$ for some $n \in \mathbb{Z}$.

(b) Using (a), show that if $re^{i\theta} = r'e^{i\theta'}$, then $r = r'$, and if $r \neq 0$ then we have $\theta' - \theta = 2\pi n$ for some $n \in \mathbb{Z}$.

One advantage of polar coordinates is that the formula for complex multiplication is simpler than that given in (8): from Fact 5.10 we get

$$(re^{i\theta}) \cdot (r'e^{i\theta'}) = (rr')e^{i(\theta+\theta')}. \tag{8}$$

In particular, using the multiplication formula (8) and induction one can prove that

$$z = re^{i\theta}, n \in \mathbb{N} \quad \Longrightarrow \quad z^n = r^n e^{in\theta},$$

which is sometimes called DeMoivre's Theorem.

Exercise 5.15. If $z = re^{i\theta}$, write $1/z$ in polar coordinates.

Exercise 5.16. Since $e^{i\theta}e^{i\theta'} = e^{i(\theta+\theta')}$, Euler's Theorem says that

$$(\cos(\theta) + i\sin(\theta))(\cos(\theta') + i\sin(\theta')) = \cos(\theta + \theta') + i\sin(\theta + \theta') \tag{9}$$

Show how to derive from this the angle sum formulas for cosine and sine.

5.2 Roots of unity

An n^{th} *root of unity* is a complex number $z \in \mathbb{C}$ such that $z^n = 1$. We let

$$U_n := \{z \in \mathbb{C} \mid z^n = 1\}$$

be the set of all n^{th} roots of unity. Note that if $z \in U_n$ then $|z|^n = |z^n| = |1| = 1$, so $|z| = 1$. So, U_n is a subset of the *unit circle*

$$S^1 := \{z \in \mathbb{C} \mid |z| = 1\} \subset \mathbb{C}$$

Exercise 5.17. If $z, w \in U_n$, show that $zw \in U_n$ and $1/z \in U_n$.

Exercise 5.18. If $d|n$, show that $U_d \subset U_n$.

In the language of abstract algebra, this shows U_n is a *group*.

Exercise 5.19. If $m, n \in \mathbb{N}$, show that $U_n \cap U_m = U_g$, where $g = gcd(m, n)$. *Hint: use Bézout's Theorem to show LHS $\subset$ RHS.*

Exercise 5.20. In polar coordinates, show that $U_n = \{e^{2\pi i \frac{k}{n}} \mid k = 1, \ldots, n\}$. *Hint: to show every element $z \in U_n$ is of the form $e^{2\pi i \frac{k}{n}}$, first write $z = e^{i\theta}$ where $\theta \in [0, 2\pi)$ (why can you do this?) and try to show θ is of the desired form.*

Exercise 5.21. For each $n \geq 2$, show that the sum of all elements of U_n is zero. *Hint: there are multiple ways to do this. Set $c = e^{2\pi i/n}$. One method is to note that multiplying by c just rearranges the elements of U_n. Another method is to observe that you're trying to sum $1 + c + c^2 + \cdots + c^{n-1}$. It's a good idea here to multiply that whole expression by $1 - c$. You probably did something like this in Calculus II..*

Exercise 5.22. Show that the product of all elements of U_n is 1 if n is odd, and -1 if n is even. *Hint: one way to do it is to note that if $z \in U_n$, so is $1/z$. Another method is to just multiply out the product of the elements using polar coordinates.*

The *order* of $z \in U_n$ is the smallest $d \in \mathbb{N}$ such that $z^d = 1$. For example, we have $-1, i \in U_4$, since $(-1)^4 = i^4 = 1$, but $-1 \in \mathbb{C}$ has order 2, while i has order 4.

Exercise 5.23. If $z \in U_n$ has order d, show that $d|n$. *Hint: division with remainder.*

Exercise 5.24. Show that $e^{2\pi i \frac{k}{n}} \in U_n$ has order $n/\gcd(k, n)$. *Hint: show that the order is the minimal d such that $n|k \cdot d$, and then argue that $k \cdot d = lcm(k, n)$.*

An element $z \in U_n$ is called *primitive* if it has order n, rather than something smaller. Note that for each n, the element $e^{2\pi i \frac{1}{n}} \in U_n$ is primitive.

Exercise 5.25. Suppose that $z \in U_n$ is primitive. Show that $U_n = \{z, z^2, \ldots, z^n\}$.

That is, if $z \in U_n$ is primitive then *all* elements of U_n are powers of z. As a hint, try to show that all the above powers of z are distinct.

Exercise 5.26. Set $z = \frac{1}{2} - i\frac{\sqrt{3}}{2}$. Show that z is a root of unity, and find its order. Then write z^{100} in the form $x + iy$.

6 Functions

A *function* is a rule that assigns to every element of a set A, an element of another set B. We write functions in the form $f : A \longrightarrow B$. Here, the element of B that the function assigns to an element $a \in A$ is written $f(a)$. The set A is called the *domain* of A, and the set B is the *codomain*.

Example 6.1. Here are some examples of functions! In each example, we specify the domain and codomain, then write down the rule that tells us how to transform elements of the domain into elements of the codomain.

(a) $f : \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = x^2,$

(b) $g : \mathbb{R} \setminus \{0\} \longrightarrow \mathbb{R}, \quad g(x) = \frac{1}{x},$

(c) $g : \mathbb{R} \longrightarrow \mathbb{R}, \quad g(x) = \begin{cases} \frac{1}{x} & \text{if } x \neq 0 \\ 37 & \text{otherwise} \end{cases}$

(d) $h : \mathbb{C} \longrightarrow \mathbb{R}, \quad h(z) = |z|,$

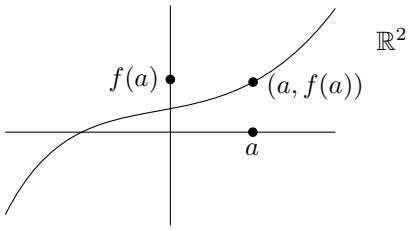
(e) $j : \{\heartsuit, \diamond, 2\} \longrightarrow \{3, 4, \emptyset\}, \quad j(\heartsuit) = j(\diamond) = 4, \; j(2) = \emptyset.$

(f) If A is *any* set, the *identity function* on A is $id_A : A \longrightarrow A$, $id_A(a) = a$.

How do we visualize functions? If $f : A \longrightarrow B$ is a function, then we define $graph(f)$ to be the subset of the product $A \times B$ given by

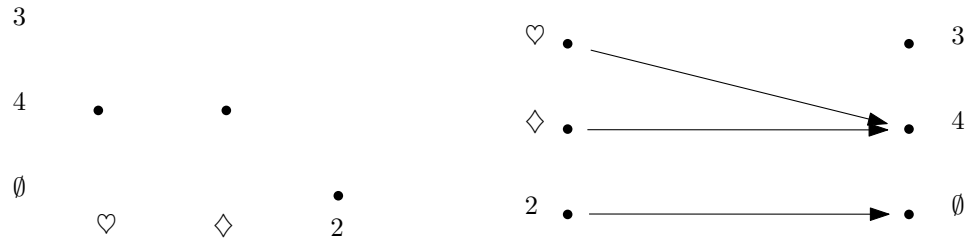
$$graph(f) := \{(a, f(a)) \mid a \in A\} \subset A \times B.$$

Graphs are especially important for functions from $\mathbb{R}$ to $\mathbb{R}$, since the product $\mathbb{R} \times \mathbb{R}$ can be visualized as a plane, with $graph(f)$ a curve.



For functions like the j in example (e), you could arrange the elements of the domain horizontally and the elements of the codomain vertically and illustrate $graph(j)$ by

placing a dot at each point $(a, j(a))$, as on the left below. However, it's sometimes more useful to draw both the domain and codomain vertically and depict the function by drawing an arrow from each a to $j(a)$, as on the right below. The latter is sometimes referred to as a dot and arrow diagram.



Exercise 6.2. What is a function, really? That is, say that you're comfortable with all the stuff from our set theory sheet. Can you give a definition of a function using only the language of set theory, without saying vague things like 'rule that assigns'? *Hint: instead of thinking about a function as a rule, think about it as its graph.*

6.1 Injectivity and surjectivity

Let $f : A \longrightarrow B$ be a function. Then f is *injective* (or *one-to-one*) if for all $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$, we have $a_1 = a_2$. And f is *surjective* (or *onto*) if for every $b \in B$, there is some $a \in A$ with $f(a) = b$. We say f is *bijective* if it is both injective and surjective, or in other words, if for every $b \in B$, there is exactly one $a \in A$ such that $f(a) = b$. A function that satisfies one of these properties is called an *injection, surjection,* or *bijection* respectively.

Exercise 6.3. Determine whether the following are injective, surjective or bijective.

(a) $id_A : A \longrightarrow A$, where A is a set,

(b) $f : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z},\ f(n) = (n, n)$,

(c) $f : \mathbb{Z} \to \mathbb{Z},\ f(n) = \begin{cases} n & n \geq 0 \\ n + 5 & n < 0 \end{cases}$,

(d) $f : \mathbb{C} \longrightarrow \mathbb{C},\ \ f(z) = z^2$,

(e) $g : \mathbb{C} \longrightarrow \mathbb{R},\ \ g(z) = |z|$,

(f) $h : \mathbb{C} \longrightarrow \mathbb{C}, \quad h(z) = e^z$,

(g) $g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}, \ g(m, n) = m + n$,

(h) $h : \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}, \ h(x, y) = \frac{x}{y}$,

(i) $\phi : \mathbb{N} \longrightarrow \mathbb{N}$, where $\phi(x)$ is the number of primes less than or equal to x,

(j) $d : \mathbb{N} \longrightarrow \mathbb{N}$, where $d(n)$ is the number of divisors (in $\mathbb{N}$, say) of n.

Exercise 6.4. Let A, B be nonempty sets and define the *projection onto* A to be

$$\pi_A : A \times B \longrightarrow A, \quad \pi_A(a, b) = a.$$

(a) Show that π_A is surjective.

(b) Show that if B has more than one element, then π_A is not injective.

Exercise 6.5. If $f, g : \mathbb{Z} \longrightarrow \mathbb{Z}$ are functions, their *sum* is the function

$$f + g : \mathbb{Z} \longrightarrow \mathbb{Z}, \quad (f + g)(n) = f(n) + g(n).$$

For example, note that if $f(n) = n + 5$ and $g(n) = -5$, then we have $(f + g)(n) = n + 5 - 5 = n$, so in other words $f + g = id_{\mathbb{Z}}$.

(a) (Easier) Show there are *injective* $f, g : \mathbb{Z} \longrightarrow \mathbb{Z}$ such that $id_{\mathbb{Z}} = f + g$, where $id_{\mathbb{Z}} : \mathbb{Z} \longrightarrow \mathbb{Z}$ is the identity function $id_{\mathbb{Z}}(n) = n$.

(b) (Medium) Show there are *surjective* $f, g : \mathbb{Z} \longrightarrow \mathbb{Z}$ such that $id_{\mathbb{Z}} = f + g$.

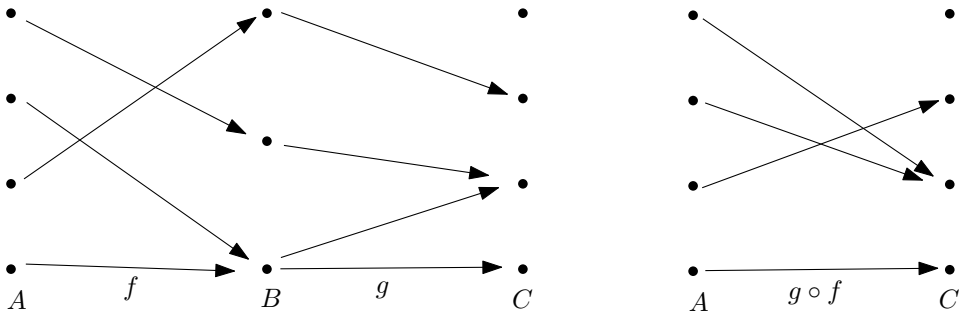
(c) (*Very* hard) Show there are *bijective* $f, g : \mathbb{Z} \longrightarrow \mathbb{Z}$ such that $id_{\mathbb{Z}} = f + g$. *Note: I'm not sure if there's a good way to write down actual formulas for f, g here. However, you can prove that they exist using a recursive construction.*

6.2 Compositions

The *composition* of two functions $f : A \longrightarrow B$ and $g : B \longrightarrow C$ is the function

$$g \circ f : A \longrightarrow C, \ g \circ f(a) = g(f(a)).$$

It's instructive to visualize compositions with dot diagrams: to find where $g \circ f$ sends $a \in A$, you follow the f-arrow into B, and then follow the g arrow into C. Note that



in order to form the composition $g \circ f$ of two functions, you need the codomain of f to be equal to the domain of g.

It is legitimately confusing that when we form a composition in which we first apply f, then apply g, we write $g \circ f$ rather than $f \circ g$. The reason is that we first apply f to $a \in A$, giving $f(a)$, and then we apply g to get $g(f(a))$, so we call the composition $g \circ f$ to have the orders of the functions agree in the two expressions. At its root, this awkwardness is due to the convention in math that inputs are written to the *right* of the function, so in other words we write $f(a)$ instead of $(a)f$.

Example 6.6. If $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ and $f(x) = x^2 + 1$, $g(x) = \sin(x)\cos(x)$, then

$$g \circ f(x) = \sin(x^2 + 1)\cos(x^2 + 1), \quad f \circ g(x) = (\sin(x)\cos(x))^2 + 1.$$

Exercise 6.7. Write nice-looking formulas for the compositions $f \circ g$ and $g \circ f$, where

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}, \ f(x, y) = \frac{x}{y^2 + 1}, \quad g : \mathbb{R} \longrightarrow \mathbb{R}^2, \ g(x) = \left(x^2, x\right).$$

Exercise 6.8. Solve the following problems about compositions.

(a) Show that the composition of two injective functions is injective.

(b) Show that the composition of two surjective functions is surjective.

(c) Show that if $g \circ f$ is injective, so is f.

(d) Show that if $g \circ f$ is surjective, so is g.

(e) If $g \circ f$ is injective, does g have to be injective? Either prove it or give a counterexample.

(f) If $g \circ f$ is surjective, does f have to be surjective? Either prove it or give a counterexample.

(g) Find nonbijective functions f, g such that $g \circ f$ is a bijection.

Exercise 6.9. For every $n \in \mathbb{N}$, construct a function $f_n : \mathbb{C} \longrightarrow \mathbb{C}$ such that

$$f_n \circ \overset{n \text{ times}}{\cdots} \circ f_n = id_{\mathbb{C}},$$

but where any composition of fewer than n copies of f_n is not the identity.

6.3 Images and Preimages

Definition 6.10 (Image)**.** If $f : A \longrightarrow B$ and $X \subset A$, the *image of X under f* is

$$f(X) := \{f(x) \mid x \in X\}.$$

The image $f(A)$ of the entire domain A is called the *image* or *range* of f. Note that $f : A \longrightarrow B$ is surjective if and only if $f(A) = B$.

Exercise 6.11. Identify the following images.

(a) $f([-1, 6])$, where $f : \mathbb{R} \longrightarrow \mathbb{R},\ f(x) = x^2$,

(b) the image of $f : \mathbb{C} \longrightarrow \mathbb{C},\ f(z) = e^z + 1$,

(c) $id_A(C)$, where $id_A : A \longrightarrow A$ is the identity function and $C \subset A$,

(d) the image of $f : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z},\quad f(x, y) = 6x + 10y$.

Definition 6.12 (Preimage)**.** If $f : A \longrightarrow B$ and $Y \subset B$, the *preimage of Y* is

$$f^{-1}(Y) := \{a \in A \mid f(a) \in Y\}.$$

As an example, if $f : \mathbb{R} \longrightarrow \mathbb{R},\ f(x) = x^2$, then $f^{-1}(\{-3, 1, 2\}) = \{-1, 1, -\sqrt{2}, \sqrt{2}\}$, because these are the 4 real numbers whose squares are in the set $\{-3, 1, 2\}$. Note that the definition does *not* say to take the square roots of the numbers $-3, 1, 2$.

Exercise 6.13. Identify the following preimages.

(a) $f^{-1}([4, 9])$, where $f : \mathbb{R} \longrightarrow \mathbb{R},\ f(x) = x^2$,

(b) $g^{-1}([1, 2] \times [3, 5])$, where $g : \mathbb{R} \longrightarrow \mathbb{R}^2,\ g(x) = (x, x^2)$.

(c) $h^{-1}(S^1)$, where $S^1 \subset \mathbb{C}$ is the unit circle and $h : \mathbb{C} \longrightarrow \mathbb{C},\ h(z) = e^z$.

(d) $j^{-1}(E)$, where $j : \mathbb{Z} \longrightarrow \mathbb{Z},\ j(x) = 3x$ and $E \subset \mathbb{Z}$ is the set of even numbers.

Exercise 6.14. Let $d, n \in \mathbb{N}$. Let U_n be the set of n^{th} roots of unity, define

$$f : U_n \longrightarrow U_n,\ f(z) = z^d.$$

(a) Show that $f^{-1}(\{1\}) = U_g$, where $g := gcd(d, n)$.

(b) Show that $f(U_n) = U_m$, where if $g = gcd(d, n)$ then $m = n/g$. *Hint: first, try showing that $f(U_n) \subset U_m$. The converse is a bit harder. Write $z \in U_m$ in polar coordinates and use Bézout's theorem to transform the g into a linear combination of d, n, then simplify.*

Exercise 6.15. Let $f : A \longrightarrow B$ be a function.

(a) If $X_1 \subset X_2 \subset A$, show that $f(X_1) \subset f(X_2)$.

(b) If $Y_1 \subset Y_2 \subset B$, show that $f^{-1}(Y_1) \subset f^{-1}(Y_2)$.

Exercise 6.16. Let $f : A \to B$ be a function. In each of the following, decide if the statement is true. If so, prove it. If not, give an explicit counterexample and then try to prove it under an additional assumption that f is either surjective or injective.

(a) $f^{-1}(f(X)) = X$ for all $X \subset A$.

(b) $f(f^{-1}(Y)) = Y$ for all $Y \subset B$.

(c) $f(X_1 \cap X_2) = f(X_1) \cap f(X_2)$ for all $X_1, X_2 \subset A$.

(d) $f^{-1}(Y_1 \cap Y_2) = f^{-1}(Y_1) \cap f^{-1}(Y_2)$ for all $Y_1, Y_2 \subset B$.

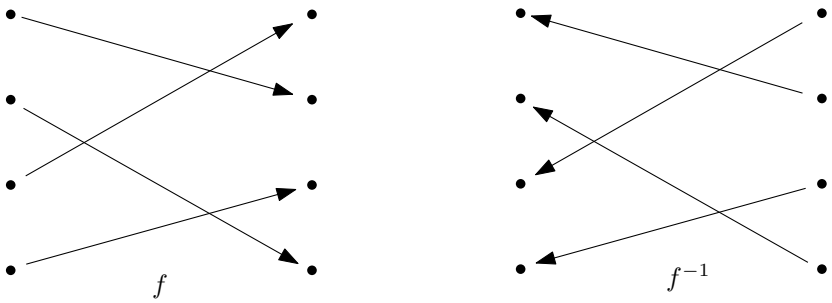
(e) $f(X_1) \setminus f(X_2) = f(X_1 \setminus X_2)$ for all $X_1, X_2 \subset A$.

(f) $f^{-1}(Y_1) \setminus f^{-1}(Y_2) = f^{-1}(Y_1 \setminus Y_2)$ for all $Y_1, Y_2 \subset B$.

6.4 Inverses

Let $f : A \longrightarrow B$ be a function. Then f is a bijection if and only if for every $b \in B$, there is exactly one $a \in A$ such that $f(a) = b$.

Definition 6.17 (Inverse)**.** The *inverse* of a bijection $f : A \longrightarrow B$ is the function $f^{-1} : B \longrightarrow A$, where $f^{-1}(b)$ is the unique element $a \in A$ such that $f(a) = b$.



If we draw a dot and arrow diagram of f, then f^{-1} is the function corresponding to the diagram with all the arrows reversed, as pictured below.

Here's an explicit example. Say we have the function

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = 2x - 1.$$

I know that if $y \in \mathbb{R}$, where here $\mathbb{R}$ is regarded as the codomain, then

$$f(x) = y \iff 2x - 1 = y \iff x = \frac{1}{2}(y + 1).$$

This shows that $x = \frac{1}{2}(y + 1)$ is the unique $x \in \mathbb{R}$ such that $f(x) = y$, so f is a bijection and its inverse is the function

$$f^{-1} : \mathbb{R} \longrightarrow \mathbb{R}, \quad f^{-1}(y) = \frac{1}{2}(y + 1).$$

Remark 6.18. The notation f^{-1} in Definition 6.17 is abusive, since we previously used the same notation to denote the preimage. However, it should always be clear from context whether we are referring to the preimage or to the inverse function. As a confusing example, note that if $f : A \longrightarrow B$ is a bijection and $b \in B$ then

$$f^{-1}(\{b\}) = \{f^{-1}(b)\}.$$

Make sure you understand what all the curly braces are doing here! On the left, f^{-1} means preimage. On the right, f^{-1} is the inverse function. Also, remember:

> *If f is not a bijection, there is *no* function called f^{-1}, and this notation always means preimage.*

Exercise 6.19. Show the following are bijections, and find the inverse.

(a) $f : \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) = (5x - 2)/12$.

(b) $g : \mathbb{N} \cup \{0\} \longrightarrow \mathbb{Z},\ g(n) = \begin{cases} \frac{n}{2} & n \text{ is even} \\ -\frac{n+1}{2} & n \text{ is odd.} \end{cases}$

(c) $h : \mathbb{R}^2 \longrightarrow \mathbb{R}^2,\ \ h(x, y) = (2x, x + y).$

(d) $j : \mathbb{Z} \longrightarrow \mathbb{Z},\ \ j(n) = \begin{cases} -n + 5 & -10 \leq n \leq 10 \\ n + 5 & \text{otherwise.} \end{cases}$

Exercise 6.20. Write down a bijection $g : \mathbb{N} \longrightarrow \{-1, -3, 2, 3, 4, 5, \ldots\}$.

If A is a set, recall that the *identity function* on A is $id_A : A \longrightarrow A,\ id_A(a) = a$.

Exercise 6.21. If $f : A \longrightarrow B$ is a bijection, show that

$$f \circ f^{-1} = id_B, \ \text{ and } \ f^{-1} \circ f = id_A.$$

Exercise 6.22. Suppose $f : A \longrightarrow B$ and there is a function $g : B \longrightarrow A$ such that both $g \circ f = id_A$ and $f \circ g = id_B$.

(a) Prove that f is a bijection, by proving injectivity and surjectivity separately.

(b) Prove that $g = f^{-1}$.

Say you are given a function $f : A \longrightarrow B$ and you want to prove it's a bijection and find its inverse. Exercise 6.22 says that all you have to do is construct some function $g : B \longrightarrow A$ and show that $g \circ f = id_A$ and $f \circ g = id_B$, and then you're guaranteed that f is a bijection with g its inverse. Here are some examples.

Exercise 6.23. Let $U_7 \subset \mathbb{C}$ be the set of 7^{th} roots of unity. Show that the functions

$$f : U_7 \longrightarrow U_7,\ f(z) = z^3, \quad g : U_7 \longrightarrow U_7,\ g(z) = z^5$$

are inverses.

Exercise 6.24. For complex numbers $a, b, c, d \in \mathbb{C}$, $c \neq 0$, let

$$f : \mathbb{C} \setminus \left\{ -\frac{d}{c} \right\} \longrightarrow \mathbb{C} \setminus \left\{ \frac{a}{c} \right\}, \ \ f(z) = \frac{az + b}{cz + d}.$$

Show that f is a bijection whose inverse is given by $g(z) = \frac{dz+b}{cz+a}$.

Exercise 6.25. If $f : A \longrightarrow B$ is a bijection, show that f^{-1} is a bijection too.

Exercise 6.26. If $f : A \longrightarrow B$ and $g : B \longrightarrow C$ are bijections, we know from Exercise 6.8 that $g \circ f$ is a bijection too. Show that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

It may seem confusing at first that the ordering of f, g switches when we take the inverse. But if you put a shirt on and then a sweater, to reverse this operation you have to first take the sweater off, and then the shirt.

Exercise 6.27. (Harder) Find a bijection $f : [0, 1] \longrightarrow [0, 1)$. *Hint: a definition in cases will be useful. You want to find a way to absorb* 1 *into a subset of* $[0, 1)$.

7 Equivalence Relations

Let A be a set. Formally, a *relation* on A is just a subset of $A \times A$. However, we view relations as follows. We pick a symbol (usually $\sim$) and we write $a \sim b$ if (a, b) is in our subset, and $a \not\sim b$ if (a, b) isn't in our subset. Here, $a \sim b$ is usually read as *'a is related to b'*, while $a \not\sim b$ is read as *a is not related to b.* As examples:

(a) If $A = \{1, 2, 3\}$, the subset $\{(1, 1), (1, 2)\} \subset A \times A$ corresponds to the relation $\sim$ on A where $1 \sim 1$ and $1 \sim 2$.

(b) We can define a relation $\heartsuit$ on $\mathbb{N}$ by declaring $a \heartsuit b$ if $a \cdot b < 5$. So for instance, $2 \heartsuit 2$, but it's not true that $3 \heartsuit 4$. Here, the associated subset of $\mathbb{N} \times \mathbb{N}$ is

$$\{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (3, 1), (4, 1), (2, 2)\}$$

(c) $\leq$ is a relation on $\mathbb{R}$, so here a real number x is related to y if $x \leq y$.

(d) Let A be a set and $f : A \longrightarrow A$ be a function. Then $a \sim f(a)$ defines a relation on A. As a subset of $A \times A$, this relation is the same as $graph(f)$.

Definition 7.1. A relation $\sim$ is an *equivalence relation* if the following hold:

(1) for all $a \in A$, we have $a \sim a$ (reflexivity),

(2) for all $a, b \in A$, if $a \sim b$ then $b \sim a$ (symmetry),

(3) for all $a, b, c \in A$, if $a \sim b$ and $b \sim c$ then $a \sim c$ (transitivity).

When $\sim$ is an equivalence relation, we sometimes read $a \sim b$ as 'a is equivalent to b', rather than saying 'a is related to b', but either works. Note that the relation on $\mathbb{Z}$ defined by $a \sim b$ if $a \leq b$ is not an equivalence relation: we have $2 \leq 3$ but $3 \not\leq 2$, so the relation isn't symmetric.

Example 7.2 (Equality)**.** If A is any set, define $a \sim b$ if $a = b$. This is an equivalence relation on A. Namely, $a = a$ for all $a \in A$, so it's reflexive. If $a = b$, then $b = a$, so it's symmetric. If $a = b$ and $b = c$, then $a = c$, so it's transitive.

Example 7.3 (The trivial equivalence relation)**.** If A is any set, define $a \sim b$ for all $a, b \in A$. Then $\sim$ is an equivalence relation on A, since as everything is related to everything, all conditions are automatically satisfied.

The point of an equivalence relation is to introduce a notion of 'similarity' between elements of A that behaves kind of like equality.

Exercise 7.4. Are the following equivalence relations?

(a) (Informal) Let L be the set of lines on the plane. For $a, b \in L$ write $a||b$ if a and b are parallel. Is $||$ an equivalence relation? *Here, lines are parallel if they are disjoint or identical. It's ok to write an informal argument for this one. Just say why you think it is or isn't an equivalence relation.*

(b) (Informal) Let $\mathcal{P}$ be the set of polygons in the plane. Set $P \sim Q$ if P is congruent to Q. *As in (a), just say whether you think this is an equivalence relation or not, and briefly explain how you came to that conclusion.*

(c) The relation on $A = \{a, b, c\}$ corresponding to the subset

$$\{(a, a), (b, b), (c, c), (a, b), (a, c), (c, a), (b, c), (c, b)\} \subset A \times A.$$

(d) Fix $n \in \mathbb{Z}$, and for $a, b \in \mathbb{Z}$ let $a \sim_n b$ if $a - b$ is a multiple of n.

(e) The relation $\sim$ on $\mathbb{Z}$ where $a \sim b$ if $a - b$ is odd.

(f) Let $\sim$ be the relation on $\mathbb{Z}$ where $a \sim b$ if $gcd(a, b) = 1$.

(g) Let $\sim$ be the relation on $\mathbb{Z}$ where $a \sim b$ if $a + b$ is a multiple of 2.

(h) Let $\sim$ be the relation on $\mathbb{Z}$ where $a \sim b$ if $a + b$ is a multiple of 3.

(i) The relation $\approx$ on $\mathbb{R}$ where $x \approx y$ if $|x - y| < .001$.

(j) The relation $\sim$ on $\mathbb{R}$ where $x \sim y$ if $x - y \in \mathbb{Z}$.

(k) The relation $\sim$ on $\mathbb{R}$ where $x \sim y$ if $xy > 0$. What about with $\geq$?

(l) The relation on $\mathbb{R}^2$ given by $(x, y) \sim (a, b)$ if $x = a$.

(m) The relation $\sim$ on $\emptyset$ given by $a \sim b$ if I am President of the United States.

(n) If X is a set, the relation $\sim$ on the power set $\mathcal{P}(X)$, where $A \sim B$ if $A \cap B \neq \emptyset$.

(o) If X is a set, the relation $\subset$ on $\mathcal{P}(X)$, i.e. A is related to B if $A \subset B$.

(p) Let $\mathcal{F}$ be the set of all functions $f : \mathbb{R} \longrightarrow \mathbb{R}$, and define $f \sim g$ if there is some finite subset $S \subset \mathbb{R}$ such that $f(x) = g(x)$ for all $x \notin S$.

(q) Let $\mathcal{F}$ be the set of functions $f : \mathbb{R} \longrightarrow \mathbb{R}$, and define $f \sim g$ if $f(0) = g(0)$.

(r) *Let $d, n \in \mathbb{N}$ with $d|n$. Let $U_n \subset \mathbb{C}$ be the set of n^{th} roots of unity, and for $z, w \in U_n$ define $z \sim w$ if $z = uw$, where $u \in U_d$.*

(s) *On $\mathbb{R}^2$, define $(x, y) \sim (a, b)$ if there's a real number $r > 0$ such that $(x, y) = (ra, rb)$.*

Here is an especially important example. Suppose that $f : A \longrightarrow B$ is a function. Let's define a relation $\sim_f$ on A by declaring $a \sim_f b$ when $f(a) = f(b)$.

Exercise 7.5. Show that $\sim_f$ is an equivalence relation.

For example, if we take $f : \mathbb{C} \longrightarrow \mathbb{R}$, $f(z) = |z|$ then $z, w \in \mathbb{R}^2$ are related exactly when they have the same modulus. In general, this exercise says that we get an equivalence relation whenever we define a relation by saying '*$a \sim b$ if a and b have the same* ____', since we can take the blank as our function in Exercise 7.5.

Exercise 7.6. Explain why the equivalence relations in parts (l) and (q) of Exercise 7.4 arise from the construction in Exercise 7.5.

Exercise 7.7. Say that A is a set, $f : A \longrightarrow A$ is a function and we define $\sim$ by setting $a \sim f(a)$. Given one example of a set A and a function f where $\sim$ *is* an equivalence relation, and another example where it is not.

Exercise 7.8. Is the following theorem and proof correct?

Theorem 7.9. *Suppose $\sim$ is a relation on a set A that's symmetric and transitive. Then $\sim$ is reflexive.*

Proof. Let $a \in A$. Pick some $b \in A$ with $a \sim b$. Then $b \sim a$, by symmetry. Since $a \sim b$ and $b \sim a$, by transitivity we have $a \sim a$. □

Exercise 7.10. Suppose that $\sim_1, \sim_2$ are two equivalence relations on A. Define a new relation $\sim$ by setting $a \sim b$ when $a \sim_1 b$ and $a \sim_2 b$.

(a) Show that $\sim$ is an equivalence relation.

(b) Show via an explicit example that if we use 'or' instead of 'and' in the construction then $\sim$ is not always an equivalence relation.

(c) (Symmetrizing) Using part (a), show that if $\sim$ is a reflexive, transitive relation on A, then the relation $\sim'$ defined by $a \sim' b$ if $a \sim b$ and $b \sim a$ is an equivalence relation on A. We call $\sim'$ the *symmetrization* of $\sim$.

Exercise 7.11. Say that $\sim$ is a relation on A. The *equivalence relation* $\sim'$ *generated by* $\sim$ is defined by $a \sim' b$ if there is a sequence $a = a_0, \ldots, a_n = b$ such that $a_i \sim a_{i+1}$ or $a_{i+1} \sim a_i$ for each $i = 0, \ldots, n-1$.

(a) Say that $\sim$ on $A = \{1, 2, 3, 4\}$ is defined by $1 \sim 2$ and $4 \sim 2$. Describe the equivalence relation $\sim'$ generated by $\sim$ by writing out *all* equivalent pairs.

(b) Show in general that $\sim'$ is actually an equivalence relation. *Hint: when proving reflexivity you can interpret* $a \sim' a$ *as vacuously true. Why?*

Exercise 7.12. Suppose that $\sim_A$ and $\sim_B$ are equivalence relations on sets A and B, respectively. Define a new relation $\sim$ on $A \times B$ by letting

$$(a, b) \sim (a', b') \text{ if } a \sim_A a' \text{ and } b \sim_B b'.$$

Is $\sim$ an equivalence relation? What if we use 'or' instead of 'and'?

7.1 Equivalence classes

Let A be a set and let $\sim$ be an equivalence relation on A. Then for $a \in A$, the $\sim$*-equivalence class of* a is defined as

$$[a]_\sim = \{x \in A \mid a \sim x\}$$

When the equivalence relation is understood, we will usually just write $[a]$ instead of $[a]_\sim$. For example, if ℓ is a line in $\mathbb{R}^2$ and we use the equivalence relation $||$ from Exercise 7.4 (a), then $[\ell]$ is the set of lines parallel to a.

Exercise 7.13. Determine the following equivalence classes, by writing out a set theoretic description of all the elements in them.

(a) $[5]$, where $\sim$ is the equivalence relation on $\mathbb{Z}$ defined by $a \sim b$ if $a - b$ is even.

(b) The $\sim_f$-equivalence class of the point $1 \in \mathbb{C}$, where $f : \mathbb{C} \longrightarrow \mathbb{R}$, $f(z) = |z|$, and $\sim_f$ is the associated equivalence relation as in Exercise 7.5.

(c) $[9]$, where $\sim$ is the equivalence relation on $\mathbb{R}$ defined by $x \sim y$ if $\lfloor x \rfloor = \lfloor y \rfloor$, where $\lfloor x \rfloor$ (read as *the floor of* x) is the largest integer that is less than or equal to x. *Express your answer here in interval notation!*

Exercise 7.14. Let $\mathcal{F}$ be the set of all functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ and let $Z : \mathbb{R} \longrightarrow \mathbb{R}$ be the zero function, i.e. $Z(x) = 0$ for all $x \in \mathbb{R}$. For each of the following equivalence relations on $\mathcal{F}$, find one other element of $\mathcal{F}$ that is *in* $[Z]$, and one that is *not* in $[Z]$.

(a) Set $f \sim g$ if there is a finite set $S \subset \mathbb{R}$ with $f(x) = g(x)$ for all $x \in \mathbb{R} \setminus S$.

(b) Set $f \sim g$ if $f(5) = g(5)$.

(c) Set $f \sim g$ if either of the following properties hold:

- there are $x, y \in \mathbb{R}$ such that $f(x) = 0$ and $g(y) = 0$, or
- $f(x)g(x) > 0$ for all $x \in \mathbb{R}$.

The following exercise is vital!

Exercise 7.15. Suppose that $\sim$ is an equivalence relation on A. Show that

(a) if $a \sim b$ then $[a] = [b]$,

(b) if $a \not\sim b$ then $[a] \cap [b] = \emptyset$.

Definition 7.16. If $\sim$ is an equivalence relation on A, the set

$$A/\sim := \{[a] \mid a \in X\},$$

of all $\sim$-equivalence classes is called *the quotient of A by $\sim$*, or just the *quotient set*.

For example, suppose that $\sim$ is the equivalence relation on $A = \{1, \ldots, 6\}$ where each element is related to itself, $2 \sim 3$, and $4 \sim 5 \sim 6$. Then

$$A/\sim = \{\{1\}, \{2, 3\}, \{4, 5, 6\}\}.$$

Here, it is called a 'quotient' because we are dividing A into pieces.

Exercise 7.17. Say that $A = \{1, 2\}$ and $\sim$ is the 'equality' equivalence relation on A, as described in Example 7.2. Find $A/\sim$.

Exercise 7.18. Say that $\sim$ is the equivalence relation on $\mathbb{R}$ given by $x \sim y$ if either $x = y = 0$ or $xy > 0$. Find $\mathbb{R}/\sim$.

Really, the point of a quotient set is to transform 'equivalence' into 'equality'. Namely, say $\sim$ is an equivalence relation on A. Each element $a \in A$ determines an element $[a] \in A/\sim$, and by Exercise 7.15 we have

$$a \sim b \quad \iff \quad [a] = [b].$$

So equivalent elements of A become *equal* elements of $A/\sim$.

Example 7.19. Suppose that $\sim_n$ is the equivalence relation on $\mathbb{Z}$ where $a \sim_n b$ if $a - b$ is a multiple of n. We define

$$\mathbb{Z}/n\mathbb{Z} := \mathbb{Z}/\sim_n = \{[x] \mid x \in \mathbb{Z}\}.$$

Here, the notation '$\mathbb{Z}/n\mathbb{Z}$' comes from abstract algebra, wherein $n\mathbb{Z}$ is notation for the set of multiples of n, and 'dividing by $n\mathbb{Z}$' corresponds to dividing up the integers into classes differing by multiples of n.

By division with remainder, every $m \in \mathbb{Z}$ has the form

$$m = qn + r, \quad q \in \mathbb{Z}, r \in \{0, \ldots, n-1\},$$

in which case $m \sim r$, so $[m] = [r]$. Moreover, if $r, s \in \{0, \ldots, n-1\}$ and $r \neq s$, then $s - r$ cannot be divisible by n, so $r \not\sim s$, implying $[r] \neq [s]$. In other words, the equivalence relation $\sim_n$ has exactly n equivalence classes, and

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], [2], \ldots, [n-1]\}.$$

Exercise 7.20. Define a map $\pi : A \longrightarrow A/\sim$, where $\pi(a) = [a]$. Show that the equivalence relation $\sim_\pi$ of Exercise 7.5 is just the original $\sim$. This shows that in fact, every equivalence relation comes from the construction in Exercise 7.5.

7.2 The rational numbers and well-definedness

Many natural mathematical objects are constructed as quotient sets of equivalence relations. For instance, let's say we have only defined integers and their arithmetic and we want to rigorously define 'rational numbers', i.e. ratios of integers. There are a number of properties we want them to satisfy, like

$$\frac{2}{3} = \frac{4}{6} = \frac{-6}{-9}, \quad \frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ab}{cd}.$$

However, it's not clear what an expression like $\frac{a}{b}$ should even mean, other than as a pair of integers (a, b), written vertically with a horizontal line between them. And if we just define a rational number to be a pair of integers written in this way, we won't have desired equalities like $\frac{2}{3} = \frac{4}{6}$.

Equivalence relations now come to the rescue. Since

$$\frac{a}{b} = \frac{c}{d} \iff ad = bc,$$

we can express when $\frac{a}{b}, \frac{c}{d}$ are 'equivalent', i.e. represent the same rational number, just in terms of integer arithmetic. We encode this in an equivalence relation, and then pass to the quotient set to transform equivalence into equality.

To formalize this, we'll consider integer pairs (a, b). You should think of a and b as the numerator and denominator of a fraction, but as we're trying to rigorously *define* rational numbers, we won't use the notation $\frac{a}{b}$ yet. Set

$$S := \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid b \neq 0\}$$

and define a relation $\sim$ on S via $(a, b) \sim (c, d)$ if $ad = bc$.

Exercise 7.21. Show that $\sim$ is an equivalence relation.

Definition 7.22. The *set* $\mathbb{Q}$ *of rational numbers* is defined to be the quotient set

$$\mathbb{Q} := S \,/\, \sim .$$

So, a 'rational number' is a $\sim$-equivalence class in S. For example,

$$\begin{aligned}[(1, 2)] &= \{(a, b) \in S \mid 1 \cdot b = 2 \cdot a\} \\ &= \{\ldots, (-3, -6), (-2, -4), (-1, -2), (1, 2), (2, 4), (3, 6), \ldots\}.\end{aligned}$$

which intuitively is the set of all pairs that represent the rational number $\frac{1}{2}$. One way to think about this is that there are many different ways to represent a given rational number as an integer fraction, so instead of picking any one representation, we look at the set of *all* possible representations. Note that

$$(2, 3) \sim (4, 6) \sim (-6, -9) \quad \Longrightarrow \quad [(2, 3)] = [(4, 6)] = [(-6, -9)],$$

so in $\mathbb{Q}$, we have equalities corresponding to $\frac{2}{3} = \frac{4}{6} = \frac{-6}{-9}$, as desired.

With the definition above, how do you define arithmetic operations on $\mathbb{Q}$? For instance, suppose we have $x, y \in \mathbb{Q}$ and want to define $x + y$. Well, by definition we have $x = [(a, b)]$ and $y = [(c, d)]$ for some $a, b, c, d \in \mathbb{Z}$, where $b, d \neq 0$. We'd like to make our definition so that it satisfies the usual law

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd},$$

and the way to do this using our current notation is to write:

$$\text{if } x = [(a, b)] \text{ and } y = [(c, d)], \text{ then } x + y := [(ad + bc, bd)]. \tag{10}$$

There's a subtlety here. Suppose for a moment that we try to define a similar operation called $\oplus$ on $\mathbb{Q}$, where $x \oplus y := [(a+c, b+d)]$. (This is the naive way that children without our sophistication might try to add fractions.) For example,

$$\text{if } x = [(0,1)] \text{ and } y = [(1,2)], \text{ then } x \oplus y := [(1,3)],$$
$$\text{if } x = [(0,2)] \text{ and } y = [(3,6)], \text{ then } x \oplus y := [(3,8)].$$

However, the x's in the two examples are actually *equal*, as are the two y's, since $(0,1) \sim (0,2)$ and $(1,2) \sim (3,6)$. However, $(1,3) \not\sim (3,8)$, so the resulting $x \oplus y$ doesn't just depend on the two inputs x, y. Rather, it depends on some additional choice involving how to represent x, y as equivalence classes. That's like saying 'Whenever you want to calculate $2+3$, flip a coin. If it turns up heads, write 5. Otherwise, write 6.' In this situation we say that $\oplus$ is not *well-defined*, which just means that our definition of $\oplus$ didn't make sense.

In contrast, we have the following.

Theorem 7.23. *The operation* $+ : \mathbb{Q} \times \mathbb{Q} \longrightarrow \mathbb{Q}$ *defined by*

$$\text{if } x = [(a,b)] \text{ and } y = [(c,d)], \text{ then } x + y := [(ad+bc, bd)]$$

is well-defined.

Again, all this means is that the formula above actually defines a function $+$. To prove it, you have to deal with this ambiguity wherein the definition of $x+y$ at least a priori depends on how you're representing an equivalence class. By Exercise 7.15,

$$[(a,b)] = [(c,d)] \iff (a,b) \sim (c,d).$$

So to show $+$ is well defined, we write $x = [(a,b)] = [(a',b')]$ and $y = [(c,d)] = [(c',d')]$ and show that we get the same equivalence class out from $+$ whether we use the pairs $(a,b), (c,d)$ or $(a',b'), (c',d')$ in our computation. So in other words, we assume $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$ and want to show

$$(ad+bc, bd) \sim (a'd' + b'c', b'd'). \tag{11}$$

Proof of Theorem 7.23. Assume $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$. Then $ab' = ba'$ and $cd' = dc'$, so it follows that

$$(ad+bc) \cdot b'd' = adb'd' + bcb'd' = ba'dd' + bb'dc' = (a'd' + b'c')bd.$$

This proves (11), which shows $+$ is well defined. □

Exercise 7.24. Which of the following are well-defined?

(a) $f : \mathbb{Q} \longrightarrow \mathbb{Z}$, $f([(a,b)]) = b$.

(b) $g : \mathbb{Q} \longrightarrow \mathbb{Q}$, $g([(a,b)]) = [(b,a)]$.

(c) $h : \mathbb{Z}/n\mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z}$, $h([n]) = [n+1]$.

(d) $f : \mathbb{R}/\sim \longrightarrow \mathbb{R}/\sim$, $f([x]) = [\sin(x)]$, where $\sim$ is defined by $x \sim \pm x$.

(e) $g : \mathbb{R}/\sim \longrightarrow \mathbb{R}/\sim$, $g([x]) = [x+1]$, where $\sim$ is defined by $x \sim \pm x$.

Exercise 7.25. Consider the equivalence relation $\sim$ on the set $\mathcal{F}$ of all functions $f : \mathbb{R} \longrightarrow \mathbb{R}$, where $f \sim g$ if there's a finite set $S \subset \mathbb{R}$ such that $f(x) = g(x)$ for all $x \in \mathbb{R} \setminus S$. Is the function $Z : \mathcal{F}/\sim \longrightarrow \mathbb{R}$, $Z([f]) = f(0)$ well defined?

Exercise 7.26. Define multiplication on $\mathbb{Q}$ and show it is well-defined.

Exercise 7.27. In this problem we define $<$ on $\mathbb{Q}$.

(a) If $x \in \mathbb{Q}$, show that $x = [(a,b)]$, where $a, b \in \mathbb{Z}$ and $b > 0$. *Hint: The whole point here is to make b positive rather than nonzero.*

Using part (a), if $x, y \in \mathbb{Q}$ we can write $x = [(a,b)]$ and $y = [(c,d)]$ with $b, d > 0$. In this case we define $x < y$ if $ad < bc$.

(b) Show that $<$ is well-defined on $\mathbb{Q}$.

(c) Show that if $x < y$ and $y < z$ then $x < z$.

Exercise 7.28. Prove the distributive property: $x(y+z) = xy+xz$ for all $x, y, z \in \mathbb{Q}$.

Exercise 7.29. Say that $\sim, \sim'$ are two equivalence relations on a set A. We say that $\sim'$ is *finer than*[1] $\sim$ if $a \sim' b \implies a \sim b$.

(a) On $\mathbb{Z}$, show that $\sim_n$ is finer than $\sim_d$ if and only if $d|n$. (Here, $a \sim_n b$ if $n|a-b$, and $\sim_d$ is defined similarly.)

(b) If $\sim'$ is finer than $\sim$, show that the map

$$\pi : A/\sim' \longrightarrow A/\sim, \quad \pi([a]') = [a]$$

is a well-defined surjection, where $[a]', [a]$ denote $\sim', \sim$ equivalence classes.

[1]It's called this since $\sim'$ then remembers more details about $a \in A$ than $\sim$ does. Think about performing a fine analysis of something vs a coarse analysis.

(c) If $\sim'$ is not finer than $\sim$, show that π above is not well defined.

Exercise 7.30. Suppose that $f : A \longrightarrow B$ is a function and $\sim_f$ is the equivalence relation on A defined by $a \sim_f a'$ if $f(a) = f(a')$. Show that the map

$$F : A/\sim \longrightarrow B, \quad F([a]) = f(a)$$

is a well-defined injection.

8 Modular Arithmetic

Fix $n \in \mathbb{N}$, and let $\sim_n$ be the equivalence relation on $\mathbb{Z}$ given by $a \sim_n b$ if $n|a-b$. As described in Example 7.19, we define

$$\mathbb{Z}/n\mathbb{Z} := \mathbb{Z}/\sim_n = \{[0], [1], \ldots, [n-1]\}.$$

Define operations $+, \cdot$ on $\mathbb{Z}/n\mathbb{Z}$ as follows.

$$[a] + [b] = [a + b].$$
$$[a] \cdot [b] = [a \cdot b].$$

Theorem 8.1. *The operations $+$ and $\cdot$ above are well-defined.*

One can show that $+, \cdot$ on $\mathbb{Z}/n\mathbb{Z}$ obey lots of the same properties that you know from integer arithmetic, like commutativity, associativity and distributivity. The elements [0] and [1] also function as 'additive and multiplicative identities', since

$$[0] + [a] = [0 + a] = [a], \quad [1] \cdot [a] = [1 \cdot a] = [a].$$

A set where you can add and multiply, satisfying the assumptions above, is called a *commutative ring*. You may see these later in an algebra class! The name actually is inspired by the additive structure of $\mathbb{Z}/n\mathbb{Z}$. For instance, in $\mathbb{Z}/12\mathbb{Z}$ we have $[7]+[6] = [13] = [1]$, and if you imagine $[0], \ldots, [11]$ as corresponding to hours on a clock, then you compute the sum by starting at [7] and counting off 6 more hours, arriving at [1]. This picture suggests arranging the elements of $\mathbb{Z}/n\mathbb{Z}$ in a circle, i.e. in a 'ring'.

Exercise 8.2. Prove the distributive property in $\mathbb{Z}/n\mathbb{Z}$, namely that if $a, b, c \in \mathbb{Z}$ then $[a]([b] + [c]) = [a][b] + [a][c]$. *(This should be rather easy, and just uses the distributive property for integer arithmetic.)*

Exercise 8.3. Suppose that $m|n$. Writing the $\sim_n$-equivalence class of $a \in \mathbb{Z}$ as $[a]_n$, and the $\sim_m$-equivalence class as $[a]_m$, define a function

$$\pi : \mathbb{Z}/n\mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z}, \quad \pi([a]_n) = [a]_m.$$

Here, we write the $\sim_n$ and $\sim_m$ equivalence classes of a as $[a]_n$ and $[a]_m$, to avoid confusing the two. Show that this function is well-defined, and explain how you could introduce someone who uses 24 hr clocks to 12 hr clocks.

Exercise 8.4. Is the function π in Exercise 8.3 well-defined if $n = 3, m = 2$, say?

In practice, when we work at a higher level with $\mathbb{Z}/n\mathbb{Z}$, we usually drop the brackets from our notation and represents its elements as $0, \ldots, n-1$, bearing in mind that really these numbers represent their associated equivalence classes. Using this notation here's a multiplication table for $\mathbb{Z}/3\mathbb{Z}$:

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

For instance, the bottom right entry reflects that $2 \cdot 2 = 1 \in \mathbb{Z}/3\mathbb{Z}$, which is code for the fact that $[2] \cdot [2] = [4] = [1]$, since $4 \sim_3 1$.

Exercise 8.5. Write out completely a 'multiplication table' for $\mathbb{Z}/4\mathbb{Z}$.

Exercise 8.6. In $\mathbb{Z}/7\mathbb{Z}$, what is $3 \cdot 4 + 3 \cdot 6 \cdot 5 \cdot (2+6) - 1$? You should be able to do this in your head, without a calculator and without ever even thinking about a number bigger than 30.

Exercise 8.7. What's the remainder when you divide 2^{140} by 7? *A word of caution: the fact that* 140 *is divisible by* 7 *is not relevant.*

Note that if $[a] \in \mathbb{Z}/n\mathbb{Z}$, then $[a] + [-a] = [a-a] = [0]$. We say here that $[-a]$ is an *additive inverse* for $[a]$, since it's an element we can add to $[a]$ to get back to the additive identity. Similarly, a *multiplicative inverse* for $[a]$ is an element $[b] \in \mathbb{Z}/n\mathbb{Z}$ that we can multiply $[a]$ by to get back to the multiplicative identity:

$$[a] \cdot [b] = [1] \in \mathbb{Z}/n\mathbb{Z}.$$

Note that $[a] \cdot [b] = [1]$ means $[ab] = [1]$ which means $n | ab - 1$.

While every element of $\mathbb{Z}/n\mathbb{Z}$ has an additive inverse, not every element has a multiplicative inverse!

Exercise 8.8. Using the table from Exercise 8.5, find out which elements of $\mathbb{Z}/4\mathbb{Z}$ have multiplicative inverses.

Theorem 8.9. *An element $a \in \mathbb{Z}/n\mathbb{Z}$ has a multiplicative inverse in $\mathbb{Z}/n\mathbb{Z}$ if and only if the greatest common divisor $gcd(a, n) = 1$.*

As a hint for the proof, remember our work on $\mathbb{Z}$-linear combinations.

Exercise 8.10. Let $U_n = \{z \in \mathbb{C} \mid z^n = 1\}$ be the set of n^{th} roots of unity, and let

$$f : \mathbb{Z}/n\mathbb{Z} \longrightarrow U_n, \quad f([k]) = e^{i2\pi k/n}.$$

(a) Show that f is a well-defined bijection.

(b) Show that $f([k] + [l]) = f([k]) \cdot f([l])$, so in some sense the additive structure of $\mathbb{Z}/n\mathbb{Z}$ is the same as the multiplicative structure of U_n.

Exercise 8.11. Say that $x \in \mathbb{N}$. Using modular arithmetic, show that x is divisible by 3 if and only if its decimal digits sum to a multiple of 3.

For instance, we have that 2352 is divisible by 3 since $2+3+5+2 = 12$ is divisible by 3, while 2452 is not divisible by 3 since $2 + 4 + 5 + 2 = 13$, which is not divisible by 3. To do the exercise, recall that in decimal notation we have

$$x = a_n a_{n-1} \cdots a_1 a_0 \iff x = a_n 10^n + a_{n-1} 10^{n-1} + \cdots a_0 10^0$$

So as an example, $2352 = 2 \cdot 10^3 + 3 \cdot 10^2 + 5 \cdot 10^1 + 2 \cdot 10^0$.

Exercise 8.12. Say that $x \in \mathbb{N}$. Using modular arithmetic, show that x is divisible by 11 if and only if the *alternating sum* of its decimal digits is divisible by 11.

For example, the alternating sum of digits of 123653 is $-1+2-3+6-5+3$. This isn't divisible by 11, so neither is 123653.

Exercise 8.13. Show that the equation $3x^2 - 5y^2 = 1$ has no integer solutions. *Hint: what does this turn into in $\mathbb{Z}/5\mathbb{Z}$?*

9 Cardinality

Definition 9.1. Given two sets A, B we say that A, B have the *same size*, or alternatively *same cardinality*, written $A \approx B$, if there is a bijection $f : A \longrightarrow B$.

This $\approx$ satisfies the following three properties.

(a) For any set A, we have $A \approx A$, via the bijection $id_A : A \longrightarrow A$, $id_A(a) = a$.

(b) If $A \approx B$ then $B \approx A$, since any bijection $f : A \longrightarrow B$ has an inverse $f^{-1} : B \longrightarrow A$, which is also a bijection.

(c) If $A \approx B$ and $B \approx C$, then $A \approx C$. Indeed, bijections $A \longrightarrow B$ and $B \longrightarrow C$ compose to give a bijection $A \longrightarrow C$.

So if the 'set of all sets' were a set, $\approx$ would define an equivalence relation on it.

Example 9.2. When A, B are finite, we have $A \approx B$ exactly when A, B have the same number of elements, since then those elements can be matched up to give the desired bijection.

Remark 9.3. What do 'finite' and 'number of elements' really mean? In fact, there's a whole theory here that we should really develop before moving forward, but it's less interesting on a first pass than a discussion of infinite sets, so we'll skip it. Briefly, though, if $n \in \mathbb{N}$ then a set A 'has n-elements' if there's a bijection $\{1, \ldots, n\} \longrightarrow A$, and A is called 'finite' if it is either empty or has n-elements for some $n \in \mathbb{N}$.

Exercise 9.4. Show that $\mathbb{N} \approx \mathbb{N} \cup \{0\}$.

Exercise 9.5. Show that $\mathbb{Z} \approx \mathbb{N}$.

Exercise 9.6. If $A \approx \emptyset$, show that $A = \emptyset$.

9.1 Countability

You may be used to thinking that all infinite sets have the same cardinality, but there are many different 'sizes' of infinity.

Definition 9.7. If $A \approx \mathbb{N}$, we say that A is *countably infinite*. We say A is *countable* if it is finite or countably infinite.

For example, $\mathbb{Z}$ is countably infinite, and $\{1, 8, \heartsuit\}$ and $\mathbb{Z}$ are both countable.

Exercise 9.8. If A is countable and $A \approx B$, then B is countable.

Remark 9.9. A set A is countably infinite exactly when its elements can be arranged into an infinite list, i.e. when A can be written in the form

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

Indeed, if A is countable, then there's a bijection $f : \mathbb{N} \longrightarrow A$, and if we set $a_i := f(i)$, then $A = \{a_1, a_2, \ldots ...\}$ as above. Conversely, if $A = \{a_1, a_2, \ldots ...\}$, then we can define a bijection $f : \mathbb{N} \longrightarrow A$ by setting $f(i) = a_i$.

Here's a first example of an uncountable set.

Theorem 9.10. *There's no surjection $f : \mathbb{N} \longrightarrow (0, 1)$. In particular, there's no bijection, so $(0, 1)$ is uncountable.*

The proof is Cantor's famous diagonal argument.

Proof. Let $f : \mathbb{N} \longrightarrow (0, 1)$ be a function. We'll show there's some $x \in (0, 1)$ such that $x \neq f(i)$ for all $i \in \mathbb{N}$. This will show f isn't surjective.

Let's write out decimal expansions of all the numbers $f(i) \in (0, 1)$ as follows.

$$\begin{aligned} f(1) &= .a_{11}\, a_{12}\, a_{13}\, a_{14} \ldots \\ f(2) &= .a_{21}\, a_{22}\, a_{23}\, a_{24} \ldots \\ f(3) &= .a_{31}\, a_{32}\, a_{33}\, a_{34} \ldots \\ f(4) &= .a_{41}\, a_{42}\, a_{43}\, a_{44} \ldots \\ &\vdots \end{aligned}$$

We want to construct some $x \in (0, 1)$ that's not equal to any of these. So, set

$$x = .x_1\, x_2\, x_3 \ldots, \quad x_i = \begin{cases} 3 & a_{ii} = 4 \\ 4 & a_{ii} \neq 4. \end{cases}$$

For example, suppose we have

$$\begin{aligned} f(1) &= .3869 \ldots \\ f(2) &= .0482 \ldots \\ f(3) &= .4490 \ldots \\ f(4) &= .2224 \ldots \\ &\vdots \end{aligned}$$

Then $x = .4343\ldots$. By construction $x_i \neq a_{ii}$, so x and $f(i)$ differ in the i^{th} decimal place, and hence aren't equal. (Note that since x doesn't have 0's and 9's in its decimal expansion, it has a *unique* decimal example, so to check it's not equal to any of the $f(i)$, it suffices to check that the decimal expansions above are different. Contrast this with the two decimal expansions for $.10000\ldots = .099999\ldots$.) □

Exercise 9.11. Draw the graph of a bijection $f : (0, 1) \longrightarrow \mathbb{R}$, and conclude that $\mathbb{R}$ is also uncountable.

Exercise 9.12. (a) If $B \subset \mathbb{N}$, show that B is countable.

(b) Using part (a), show very quickly that if A is countable and $f : B \longrightarrow A$ is injective, then B is countable. In particular, a subset of a countable set is countable.

Exercise 9.13. (a) If $f : \mathbb{N} \longrightarrow B$ is surjective, show that B is countable. *Hint: if $b \in B$, let $g(b) = \min\{x \in A \mid f(x) = b\}$, i.e. the least element that f takes to b. Show that g is an injection.*

(b) Using part (a), show that if A is countable and $f : A \longrightarrow B$ is surjective, then B is countable.

In particular, we have the following useful criteria for countability:

Theorem 9.14. *Let A be a set. Then the following are equivalent.*

(a) A is countable,

(b) there is an injection $f : A \longrightarrow \mathbb{N}$,

(c) there is a surjection $g : \mathbb{N} \longrightarrow A$.

Proof. We know that (b) and (c) imply that A is countable from above, so it only remains to show that if A is countable, then (b) and (c) hold.

If A is countably infinite, then there's a bijection $f : A \longrightarrow \mathbb{N}$. So we can use f as the injection in (b) and f^{-1} as the surjection in (c).

If A is finite, then there's a bijection $f : A \longrightarrow \{1, \ldots, n\}$ for some n, which gives an injection as in (a), and we can define a surjection $g : \mathbb{N} \longrightarrow A$ by

$$g(x) = \begin{cases} f^{-1}(x) & x \leq n \\ 1 & \text{otherwise.} \end{cases}$$

□

Exercise 9.15. Show that if A, B are both countably infinite, so is $A \cup B$.

Exercise 9.16. (a) Show that $\mathbb{N} \times \mathbb{N}$ is countably infinite. *Hint: here are two possible approaches. The easiest involves constructing an injection $\mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ by using uniqueness of prime factorizations. Alternatively, think about snakes! Make an $\mathbb{N} \times \mathbb{N}$ grid, and starting at $(1, 1)$, try to wind through it.*

(b) Using (a), show that if A, B are both countable, so is $A \times B$.

A version of the following was suggested by F. Dong on his Intro to Abstract Math Final Exam, in Fall 2024! It gives another possible solution to Exercise 9.16 (a).

Exercise 9.17. Let $f : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N},\ f(x, y) = (x+y)^2 + x$. Show that f is injective.

Exercise 9.18. Using the previous exercises and our definition of $\mathbb{Q}$ as a quotient set, show that $\mathbb{Q}$ is countably infinite.

Exercise 9.19. Show that the set $\mathbb{R} \setminus \mathbb{Q}$ of all irrational numbers is uncountable.

9.2 Ordering sets by cardinality

If there is an injection $f : A \longrightarrow B$, we write $A \preceq B$. If $A \preceq B$ but $A \not\approx B$, i.e. there's an injection $A \longrightarrow B$ but there is no such bijection, we write $A \prec B$. For example, $\mathbb{N} \prec \mathbb{R}$ because the inclusion $i : \mathbb{N} \longrightarrow \mathbb{R},\ i(n) = n$ is an injection, but we proved in Exercise 9.11 that there's no bijection $\mathbb{N} \longrightarrow \mathbb{R}$.

Note that to prove that $A \prec B$, it does *not* suffice to produce an injection that is not a bijection. For example, $f : \mathbb{N} \longrightarrow \mathbb{N},\ f(x) = x + 1$ is an injection that is not a bijection, but it's not true that $\mathbb{N} \prec \mathbb{N}$. The point is to construct an injection, and then show separately there is no possible (unrelated) bijection.

Exercise 9.20. In this exercise, we show that if A is a set, then $A \prec \mathcal{P}(A)$.

(a) Construct an injective function $i : A \longrightarrow \mathcal{P}(A)$, showing that $A \preceq \mathcal{P}(A)$.

(b) Suppose that $f : A \longrightarrow \mathcal{P}(A)$ is a function. Show that the subset

$$X = \{a \in A \mid a \notin f(a)\} \subset A$$

has the property that $X \neq f(b)$ for all $b \in A$. Conclude that $A \not\approx \mathcal{P}(A)$, which together with (a) implies that $A \prec \mathcal{P}(A)$.

(c) Suppose now that $A = \mathbb{N}$. To each subset $B \subset \mathbb{N}$, associate an infinite string $b_1 b_2 b_3 \ldots$ of 0's and 1's, where $b_i = 1$ if $i \in B$, and $b_i = 0$ otherwise. Explain how from this perspective, the proof in (b) above is exactly the same as Cantor's diagonal argument, from Theorem 9.10.

The following theorem may look obvious at first, but it's really not!

Theorem 9.21. *(Schroeder–Bernstein) If $A \preceq B$ and $B \preceq A$, then $A \approx B$.*

Proof. We can assume A, B are disjoint, take injections

$$f : A \longrightarrow B,\ g : B \longrightarrow A,$$

and for convenience we combine them to give an injection

$$h : A \cup B \longrightarrow A \cup B, \quad h(x) = \begin{cases} f(x) & x \in A \\ g(x) & x \in B. \end{cases}$$

If $x \in A \cup B$ lies in the image of h, we define $h^{-1}(x) \in A \cup B$ to be the unique element of $A \cup B$ that h takes to x. Then given $x \in A \cup B$, we can repeatedly apply h^{-1}, giving elements $h^{-1}(x), h^{-2}(x), \ldots$. Let's call these elements the *ancestors* of x. If all these ancestors are always in the image of h, we can keep going forever. However, if there's some ancestor $h^{-n}(x)$ that doesn't lie in the image of h, it has no further ancestors, and we call $h^{-n}(x)$ the *original ancestor*. We now define a bijection

$$H : A \longrightarrow B, \quad H(a) = \begin{cases} g^{-1}(a) & \text{if } a \text{ has an original ancestor, which lies in } B \\ f(a) & \text{otherwise.} \end{cases}$$

We first claim that H is injective. So, assume $a \neq a'$ and $H(a) = H(a')$. Since f, g are injective, we can assume a, a' are in the two separate cases in the definition of H. So, a has an original ancestor that lies in B, and a' doesn't, but $g^{-1}(a) = f(a')$. However, here $a' = f^{-1}(g^{-1}(a)) = h^{-2}(a)$, so if a has an original ancestor in B, then a' will too, and we're done.

Next, we want to show that H is surjective. Pick some $b \in B$. Suppose first that b has an original ancestor, which lies in B. Let $a = g(b)$. Then a has the same original ancestor, which lies in B, so $H(a) = g^{-1}(a) = b$ and we're done. On the other hand, say that b either doesn't have an original ancestor, or it has one that lies in A. Then in particular b has ancestors, i.e. it lies in the image of h, so we can set $a := h^{-1}(b) = f^{-1}(b)$. This a also either doesn't have an original ancestor or has one that lies in A, so $H(a) = f(a) = b$. ∎

Alternative proof. The following is the same proof as above, but phrased more visually. As before, suppose we have injective functions $f : A \longrightarrow B$ and $g : B \longrightarrow A$. Draw elements of A and B as dots: $\bullet$ for A and $\odot$ for B. Draw an arrow from each $a \in A$ to $f(a)$, and from each $b \in B$ to $f(b)$. If you pick a point x in $A \cup B$, you can look at all the points you get from that point by tracking forwards and backwards along the arrows. Let's call such a set of points a *lineage*. Since every dot has exactly one arrow pointing out of it, and at most one arrow pointing into it, lineages look like

(a) a biinfinite line $\cdots \to \bullet \to \odot \to \bullet \to \odot \to \cdots$,

(b) a circle $\bullet \to \odot \to \cdots \to \bullet \to \odot \to \bullet$, where the first and last $\bullet$ are the same,

(c) a ray $\bullet \to \odot \to \bullet \to \odot \to \cdots$, or

(d) a ray $\odot \to \bullet \to \odot \to \bullet \to \cdots$

Think of $A \cup B$ as broken up into all the possible lineages above. Then we can make a bijection $H : A \longrightarrow B$ by defining it separately on each lineage, and matching up the $\bullet$'s to $\odot$'s in that lineage. In cases (a)-(c), just define $H(\bullet)$ to be the $\odot$ to the immediate right. In case (d), just define $H(\bullet)$ to be the $\odot$ to the immediate left. □

Exercise 9.22. Show that $\mathcal{P}(\mathbb{N}) \approx \mathbb{R}$. *Hint: for convenience, produce injections in both directions. You will probably find using binary or decimal expansions useful.*

Exercise 9.23. If $A \preceq B$ and $B \prec C$, show that $A \prec C$. *Hint: Amazingly, this isn't obvious. Use the Schroeder-Bernstein theorem.*

Exercise 9.24. Let $\mathcal{A}$ be a set whose elements are sets. Show that there is some set B such that $A \preceq B$ for all $A \in \mathcal{A}$.

This exercise suggests an unimaginable number of different infinite cardinalities. Namely, given a set A, let $\mathcal{P}^n(A)$ be the n^{th} *iterated power set* of a set A, defined by

$$\mathcal{P}^n(A) := \mathcal{P}(\cdots \mathcal{P}(\mathcal{P}(A)) \cdots).$$

Starting with $\mathbb{N}$, we can construct a sequence of sets as follows:

$$\mathbb{N} \prec \mathcal{P}(\mathbb{N}) \prec \mathcal{P}^2(\mathbb{N}) \prec \ \cdots \preceq B \prec \mathcal{P}(B) \prec \mathcal{P}^2(B) \prec \cdots \preceq C \prec \mathcal{P}(C) \prec \cdots,$$

where here, B is a set that is at least as big in size as all $\mathcal{P}^n(\mathbb{N})$, and then C is defined similarly. Of course, this goes on forever, even after we run out of letters in the alphabet, and the exercise above shows that even after you repeat this procedure forever, there's STILL a bigger set than everything so far constructed. Then you can repeat the process with that bigger set, and continue...

Exercise 9.25. Write out all the elements of $\mathcal{P}^3(\emptyset)$. Then try to come up with a formula for the number of elements in the set $\mathcal{P}^n(\emptyset)$.