

Concepts of Mathematics (21-127)

Clive Newstead

Summer II 2015

Preface

Why am I writing these notes?

It seems silly in 2015 to be writing lecture notes for an introduction to mathematics course: most universities where mathematics is taught to an advanced level have such a course, and a vast number of those courses either have lecture notes available online, or follow an easily accessible book. I spent a long time trawling the internet, looking for for a resource which...

- ...is short enough that a student could finish it in a semester;
- ...covers proof-writing methods as well as mathematical topics;
- ...avoids non-standard jargon, notation and abbreviations;
- ...defines mathematical objects by their intrinsic properties, rather than reducing them to a formal implementation.

Everything I found fell short of at least one of these criteria. This is not the fault of the authors, as their courses likely had different learning objectives than mine; but it gave me sufficient motivation to write my own set of notes. I took the liberty of adding $\text{\LaTeX}$ support.

Choices of convention

I have taken care to avoid bias in these notes, but since the mathematical community will never agree on everything, I should explain some of the choices I made. (Had I made other choices instead, those would also require explanation!)

- We take 0 to be a natural number, with the rationale that the natural numbers should be the cardinalities of the finite sets, not just the inhabited ones. Had we excluded 0, we would need to make exceptions in the section on combinatorics, and the fact that results that are true of positive natural numbers are usually also true of 0 would be obfuscated.

- I use the term *inhabited* instead of *non-empty*, because ‘ X is non-empty’ means ‘ X doesn’t not have elements’, and not using double-negatives isn’t something we want to avoid.
- We do number theory on all of $\mathbb{Z}$, rather than on $\mathbb{N}$ or on some mish-mash between $\mathbb{N}$ and $\mathbb{Z}$. This makes it easier to generalise our results to general rings, and drills in the habit of defining number theoretic concepts (like greatest common divisors) in terms of the arithmetic structure of $\mathbb{Z}$ rather than its order structure.
- Primes and irreducibles in $\mathbb{Z}$ are defined separately before being proved to coincide. This is because in a general ring they do not coincide, and thus it is a special property of $\mathbb{Z}$ that should not be glossed over.
- Concepts are not identified with their set theoretic implementation. Thus, for example, we don’t identify natural numbers with von Neumann ordinals, and a distinction is made between functions $f : X \rightarrow Y$ and their graphs $\Gamma_f \subseteq X \times Y$. This is in part to aid conceptual understanding, and in part because I don’t want to indoctrinate students into believing that any one particular axiomatic foundation is the correct one. (Those students who wish to think of a function as being a set of ordered pairs are welcome to do so!)

A note on foundations

Standard practice in non-foundational areas of mathematics is to work in terms of naïve set theory, on the understanding that their work could be formalised in ZFC if necessary. Mathematicians working at a foundational level typically formulate their work in the axiomatic framework of ZFC, or some extension of it. I have avoided discussing foundational issues at this elementary level, and as such we adopt the approach of working naïvely with sets.

Acknowledgements

I’d like to thank the following people for their useful conversations and insights which, perhaps unbeknownst to them, have helped shape the material in these notes: Jeremy Avigad, Steve Awodey, Thomas Forster, Will Gunther, Huck Gutman, Bob Harper, Brian Kell, Ben Millwood, Gareth Taylor, Andy Zucker. I would also like to thank everyone at the Eberly Center for Teaching Excellence at Carnegie Mellon University, especially Chad Hershock, whose *Evidence-Based Teaching in the Sciences* course got the ball rolling.

Contents

1	The language of mathematics	1
1.1	Symbolic logic	2
1.2	Sets and set operations	22
1.3	Induction and recursion	36
1.4	Problems	52
2	Number theory	53
2.1	Division	55
2.2	Prime numbers	63
2.3	Modular arithmetic	70
2.4	Problems	80
3	Relations and functions	81
3.1	Relations	82
3.2	Functions	93
3.3	Problems	103
4	Finite and infinite sets	104
4.1	Counting principles	105

4.2	Binomials and factorials	113
4.3	Cardinality	122
4.4	Problems	131
5	Sets with structure	132
5.1	Rings: sets with arithmetic	133
5.2	Posets: sets with order	139
5.3	Groups: sets with symmetry	139
5.4	Metric spaces: sets with distance	139
A	Proof techniques	140
A.1	Structuring a proof	140
B	Typesetting mathematics in $\text{\LaTeX}$	145
B.1	Finding the software	145
B.2	Text mode and math mode	146
B.3	Mathematical symbols	147
B.4	Organisation and formatting	148
B.5	Practice page	151
	Index	152

Chapter 1

The language of mathematics

Before we can go any distance towards achieving our goal of becoming mathematically literate, we need to introduce some basic notions that will be the underpinnings of everything we do.

The first notion is that of a **proposition**, which is quite simply a statement that is either true or false, as opposed to being nonsensical. We will look at how to build propositions from other propositions using **logical operators** like ‘and’, ‘or’, ‘not’ and ‘implies’, and how to prove statements built up from others in this way. Some statements will involve variables—for example, the statement ‘ n is even’ can be true or false depending on the value of n ; we can reason about such statements using **quantifiers**, which will be introduced soon.

We will then talk about **sets**, which allow us to reason about collections of mathematical objects. Everything we do can be phrased in terms of sets, and the language of set theory will be very useful for formulating our problems later on.

Our focus will then turn to the set of natural numbers. The ordering of the natural numbers satisfies a special property called the **well-ordering principle**; this allows us to prove statements about natural numbers using a proof technique called **induction**, and to define sequences of natural numbers using **recursion**.

With these tools under our belt, we’ll be free to take on the world! (Well, mathematically speaking at least.)

1.1 Symbolic logic

GOAL

After working through this section you should be able to identify the logical structure of a proposition and use this structure to formulate appropriate proof strategies.

Consider the following statement:

This sentence is false.

Is it true or false? If you think about this for a couple of seconds then you'll get into a bit of a pickle.

Now consider the following statement:

The happiest donkey in the world.

Is it true or false? Well it's not even a sentence; it doesn't make sense to even *ask* if it's true or false!

Clearly we'll be wasting our time trying to write proofs of statements like the two listed above. This motivates the following informal definition:

1.1.1 Definition. A **proposition** is a statement to which it is possible to assign a **truth value** ('true' or 'false').¹ If a proposition is true, a **proof** of the proposition is a logically valid argument demonstrating that it is true.

Thus the statements given above are not propositions because there is no possible way of assigning them a truth value.

1.1.2 Exercise. Think of an example of a true proposition, a false proposition, and a statement that is not a proposition. Compare with a neighbour, and discuss why each is true, false or not a proposition.

¹All that matters is that it *makes sense* to say that it is true or false, regardless of whether it actually *is* true or false. The truth value of many propositions is unknown, even very simple ones.

Symbolic approach

So that we can study propositions in an abstract setting, we will represent propositions by symbols, typically the letters p , q , r and s . (It is rare that we will speak about more than four propositions at the same time; if we need to, we'll just use more letters!) We call these **propositional variables**: they are 'propositional' because they represent propositions, and they are 'variables' because we will make no assumptions about their truth value (unless explicitly stated).

This symbolic approach will allow us to decompose complex propositions into simpler ones and investigate their logical structure, which in turn will help us work out how to structure our proofs.

For example, consider the following:

Let n be an integer. If n is prime and $n > 2$ then n is odd.

The three statements ' n is prime', ' $n > 2$ ' and ' n is odd' are all propositions in their own right, despite the fact that they all appear in a more complex proposition. We can really examine the logical structure of the proposition by replacing these simpler propositions with symbols. Referring to ' n is prime' as p , ' $n > 2$ ' as q , and ' n is odd' as r , the structure of the second proposition is:

If p and q , then r .

Thus the propositions p, q, r are tied together by language, namely the word 'and' and the construction 'if-then'. Soon we will give precise definitions of what these words mean; in the abstract setting they are called *logical operators*.

Observing logical structure in complex propositions allows us to make an educated guess about how to proceed with a proof of the statement if it is true. Indeed, it is a safe bet that in order to prove 'if p and q , then r ', you should derive a proof of r based on the assumption that p and q are both true.

The value of reducing statements to symbolic expressions is that it forces us to remove ambiguity and gives a clear-cut and precise way of knowing when we've done what we set out to do.

Logical operators

1.1.3 Definition. A **logical operator** is a rule that constructs a new proposition out of other propositions or propositional variables.

For example, from the propositional variables p, q we can construct several new propositions off the bat:

‘ p and q ’ ‘ p or q ’ ‘if p , then q ’ ‘ p is false’

Each of these will refer to different logical operator—respectively called *conjunction*, *disjunction*, *implication* and *negation*—and there are many more where they came from!

Relying on our understanding of the English language to interpret what these logical operators mean will cause us some trouble; the next few pages introduce the most commonly used logical operators, together with their precise definitions.

Conjunction ($\wedge$)

Conjunction is the logical operator which makes precise what we mean when we say ‘and’.

1.1.4 Definition. Let p and q be propositions. The **conjunction** of p and q , denoted $p \wedge q$ (read: ‘ p and q ’) (`\wedge` in $\text{\LaTeX}$) is a proposition which is true if both p and q are true, and false otherwise.

PROOF TIP

A proof of a proposition of the form $p \wedge q$ can be given by tying together two proofs, one being a proof that p is true, the other being a proof that q is true.

On the other hand, if you are *assuming* that $p \wedge q$ is true, then you can deduce that p is true and that q is true.

It is not always obvious when conjunction is being used; sometimes it sneaks in without the word ‘and’ ever being mentioned! Be on the look-out for occasions like this, such as in the following exercise.

1.1.5 Example. We can express the proposition ‘7 is an integer greater than 5’ in the form $p \wedge q$, by letting p be the proposition ‘7 is an integer’ and let q be the proposition ‘7 is greater than 5’. In order to prove that 7 is an integer greater than 5, we would need to give a proof that 7 is an integer, and a proof that 7 is greater than 5.

1.1.6 Exercise. Express the proposition ‘Clive is a mathematician who lives in Pittsburgh’ in the form $p \wedge q$, for propositions p and q .

The definition of $p \wedge q$ can be given by means of a **truth table**, which tells us the truth value of $p \wedge q$ for all possible assignments of truth values to p and q :

p	q	$p \wedge q$	
✓	✓	✓	← $p \wedge q$ is true when p is true and q is true
✓	×	×	← $p \wedge q$ is false when p is true and q is false
×	✓	×	← $p \wedge q$ is false when p is false and q is true
×	×	×	← $p \wedge q$ is false when p is false and q is false

Here ✓ ([L^AT_EX](#) code: `\checkmark`) denotes ‘true’ and × ([L^AT_EX](#) code: `\times`) denotes ‘false’.² There is a row for each assignment of ‘true’ (✓) or ‘false’ (×) to the propositional variables, and a column for each variable and the proposition we’re interested in.

Disjunction ($\vee$)

Disjunction is the logical operator that makes precise what we mean by ‘or’.

The word ‘or’ is especially context-dependent in English: if you say to me, ‘you can have a slice of cake or you can have a chocolate bar,’ does that mean I can have both, or not? We remove this ambiguity with the following definition; and to clarify, with this definition of ‘or’, I *can* have both the cake and the chocolate bar. Yummy.

1.1.7 Definition. Let p and q be propositions. The **disjunction** of p and q , denoted $p \vee q$ (read: ‘ p or q ’) ([L^AT_EX](#) code: `\vee`) is the proposition which is true if at least one of p or q is true, and false otherwise.

PROOF TIP

A proof of a proposition of the form $p \vee q$ can be given by simply giving a proof that p is true, or by giving a proof that q is true.

On the other hand, if you are assuming that $p \vee q$ is true for the purposes of proving something else, you should typically split into two cases, one of which proves the result on the assumption that p is true, and the other of which proves the result on the assumption that q is true.

1.1.8 Exercise. Using Definition 1.1.7, write down a truth table for $p \vee q$ (see page 5 for how it was done for $p \wedge q$).

²Instead of ✓, ×, some authors use $\top, \perp$ ([L^AT_EX](#) code: `\top, \bot`) or T, F or 1, 0.

The real power of truth tables comes when investigating how logical operators interact with each other.

1.1.9 Example. Given propositions p, q, r , when is $(p \wedge q) \vee (p \wedge r)$ true? It's not immediately obvious, but we can work it out by breaking it down into its component parts, namely the propositions $p \wedge q$ and $p \wedge r$; we'll call these **auxiliary propositions**. We can then make a column for each variable, each auxiliary proposition, and the main proposition, to find its truth values.

p	q	r	$p \wedge q$	$p \wedge r$	$(p \wedge q) \vee (p \wedge r)$
✓	✓	✓	✓	✓	✓
✓	✓	×	✓	×	✓
✓	×	✓	×	✓	✓
✓	×	×	×	×	×
×	✓	✓	×	×	×
×	✓	×	×	×	×
×	×	✓	×	×	×
×	×	×	×	×	×

variables
auxiliary propⁿs
main proposition

We can then read off the table precisely when $(p \wedge q) \vee (p \wedge r)$ is true, by comparing the entries in its column with the corresponding truth values of p, q, r .

ASIDE

If you haven't already mixed up $\wedge$ and $\vee$, you probably will soon, so here's a way of remembering which is which:

mac n cheese

If you forget whether it's $\wedge$ or $\vee$ that means 'and', just write it in place of the 'n' in 'mac n cheese':

mac $\wedge$ cheese

mac $\vee$ cheese

Clearly the first looks more correct, so $\wedge$ means 'and'. (For any Brits among you, the mnemonic 'fish n chips' works just as well.)

1.1.10 Exercise. Write a truth table for the proposition $p \wedge (q \vee r)$. Compare it with the truth table for $(p \wedge q) \vee (p \wedge r)$. What do you notice?

Hopefully, if you did the previous exercise correctly, you'll have noticed that the column for $p \wedge (q \vee r)$ is identical to the column for $(p \wedge q) \vee (p \wedge r)$. So in some sense, these two propositions are 'the same'.

1.1.11 Definition. Two propositions depending on the same propositional variables are **logically equivalent** if they have the same truth value as each other, no matter what the assignment of truth values to their propositional variables.

PROOF TIP

To prove that two propositions are logically equivalent, you can draw a truth table containing both propositions; if their columns are the same, then they are logically equivalent.

1.1.12 Example. The propositions $p \wedge (q \wedge r)$ and $(p \wedge q) \wedge r$ are equivalent. To prove this, we'll combine the truth tables for both propositions, with auxiliary columns for the propositions $q \wedge r$ and $p \wedge q$.

p	q	r	$q \wedge r$	$p \wedge (q \wedge r)$	$p \wedge q$	$(p \wedge q) \wedge r$
✓	✓	✓	✓	✓	✓	✓
✓	✓	×	×	×	✓	×
✓	×	✓	×	×	×	×
✓	×	×	×	×	×	×
×	✓	✓	✓	×	×	×
×	✓	×	×	×	×	×
×	×	✓	×	×	×	×
×	×	×	×	×	×	×

Evidently the two propositions are equivalent since their columns are identical. Indeed, $p \wedge (q \wedge r)$ and $(p \wedge q) \wedge r$ are both true if all three of p , q and r are true, and they're both false if one or more of p , q or r is false.

Negation ($\neg$)

So far we only officially know how to prove that true propositions are true. The negation operator makes precise what we mean by 'not', which allows us to prove that false propositions are false.

1.1.13 Definition. Let p be a proposition. The **negation** of p , denoted $\neg p$ (read: 'not p ') ([L^AT_EX code: \neg](#)) is the proposition which is true if p is false, and false if p is true.

PROOF TIP

To prove a statement p is false, you can prove that its negation $\neg p$ is true.

The truth table for the negation operator is very simple:

p	$\neg p$
✓	×
×	✓

1.1.14 Example. What follows is the truth table for $p \wedge (\neg q)$ (read ‘ p and not q ’); we include a column for $\neg q$ because it appears inside the proposition.

p	q	$\neg q$	$p \wedge (\neg q)$
✓	✓	×	×
✓	×	✓	✓
×	✓	×	×
×	×	✓	×

Theoretically we could stop here: the three operators we’ve seen, $\wedge$, $\vee$ and $\neg$, can be used to give any combination of truth values to a compound proposition, in *any* number of variables!³ For example, try the following exercise:

1.1.15 Exercise. Using only two variables p, q and the operators $\wedge, \vee, \neg$, write down a proposition whose truth table column is:

p	q	???
✓	✓	×
✓	×	✓
×	✓	✓
×	×	×

We will encounter this later as the *exclusive disjunction* operator, see page 12.

The following theorem is our first big result of the course. It is a pair of dual results which relate conjunction, disjunction and negation. Informally the result says:

- Saying ‘neither p nor q is true’ is the same as saying ‘both p and q are false’;

³Proving this claim and investigating other ‘complete sets’ of operators would make a nice final project!

- Saying ‘ p and q are not both true’ is the same as saying ‘at least one of p and q is false’.

Let’s make this precise:

1.1.16 Theorem (De Morgan’s laws for logical operators). Let p and q be propositions. Then

- (a) $\neg(p \vee q)$ is logically equivalent to $(\neg p) \wedge (\neg q)$;
 (b) $\neg(p \wedge q)$ is logically equivalent to $(\neg p) \vee (\neg q)$.

Proof. (a) The following truth table demonstrates that $\neg(p \vee q)$ and $(\neg p) \wedge (\neg q)$ have the same truth value for any assignment of truth values to p and q ; hence they are logically equivalent.

p	q	$p \vee q$	$\neg(p \vee q)$	$\neg p$	$\neg q$	$(\neg p) \wedge (\neg q)$
✓	✓	✓	×	×	×	×
✓	×	✓	×	×	✓	×
×	✓	✓	×	✓	×	×
×	×	×	✓	✓	✓	✓

The proof of (b) mimics the proof of (a) and is left as an exercise. □

WRITING TIP

In mathematical texts, a proposition which is particularly important is usually called a ‘theorem’. A proposition which is proved for the purposes of being used in a proof of a later theorem is typically called a ‘lemma’. A proposition which is an easily derived consequence of a theorem is typically called a ‘corollary’.

The italicised ‘*Proof*’ signifies that the proof is beginning; since statements of theorems sometimes get very long, this helps the reader identify when the statement is finished and the proof has started. The end of the proof is signified by a square □ ([L^AT_EX](#) code: \Box).

1.1.17 Corollary.

- (a) The operator $\wedge$ can be expressed in terms of $\vee$ and $\neg$;
 (b) The operator $\vee$ can be expressed in terms of $\wedge$ and $\neg$.

Proof. (a) First note that, if p is any proposition, then p is equivalent to $\neg(\neg p)$, which we’ll write simply $\neg\neg p$. This is demonstrated by the following truth table

p	$\neg p$	$\neg\neg p$
✓	×	✓
×	✓	×

Therefore, given any propositions p and q ,

- $p \wedge q$ is equivalent to $(\neg\neg p) \wedge (\neg\neg q)$;
- ...which is equivalent to $\neg((\neg p) \vee (\neg q))$ by De Morgan's laws applied to the propositions $\neg p$ and $\neg q$.

Since $p \wedge q$ is equivalent to $\neg((\neg p) \vee (\neg q))$, which contains only the operators $\neg$ and $\vee$, the result has been shown.

The proof of (b) mimics the proof of (a) and is left as an exercise. □

This means that just *two* operators, say $\wedge$ and $\neg$, suffice for expressing all other possible operators! However, there is no real virtue in being stingy with our operators; after all, the whole point of everything we're doing is to communicate mathematical ideas. The expression

$$\neg((\neg p) \vee (\neg q))$$

is a lot harder to read and *much* harder to understand than the expression

$$p \wedge q$$

So we'll keep $\wedge$ for now, and we'll go one step further: there is one especially crucial operator that we have not yet defined, namely *implication*.

Implication ($\Rightarrow$)

The implication operator makes precise what we mean when we say 'if p , then q ' or ' p implies q '. The definition of the implication operator might seem unnatural at first, but this will be discussed as an aside after the definition has been given.

1.1.18 Definition. Let p and q be propositions. The proposition $p \Rightarrow q$ (read: 'if p then q ', or ' p implies q ') ([L^AT_EX](#) code: `\Rightarrow`) is *false* if p is true and q is false, and true otherwise.

PROOF TIP

To prove that $p \Rightarrow q$ is true, you can give a proof of q , reasoning from the assumption that p is true. Alternatively, if you know that p is false, a proof that p is false suffices to prove that $p \Rightarrow q$ is true; in this latter case, we say $p \Rightarrow q$ is ‘trivially true’.

On the other hand, to prove a proposition q , it suffices to prove $p \Rightarrow q$ and p are both true. This is especially useful when *applying* theorems. Theorems often take the form ‘assuming some hypotheses about X , something is true of X ’; we then obtain ‘something is true of X ’ by proving that the hypotheses hold.

The truth table for the implication operator is as follows:

p	q	$p \Rightarrow q$
✓	✓	✓
✓	×	×
×	✓	✓
×	×	✓

1.1.19 Exercise. Use a truth table to show that $p \Rightarrow q$ is equivalent to $(\neg p) \vee q$.

1.1.20 Exercise. Let p and q be propositions. Find an expression equivalent to $\neg(p \Rightarrow q)$ which uses only the operators $\wedge$, $\vee$ and $\neg$. How could you use this equivalence to prove that an implication $p \Rightarrow q$ is false?

ASIDE

The biggest source of confusion for most people about the implication operator is why $p \Rightarrow q$ is true whenever p is false, even if q is also false.

The reason behind this confusion is that people tend to think of implication in terms of *causation*, i.e. that $p \Rightarrow q$ is a statement asserting ‘ q is true because of p ’. This is *not* what ‘implies’ means here! The statement $p \Rightarrow q$ says nothing about the truth value of q *unless* we know that p is true.

Think of it this way: $p \Rightarrow q$ means that I can give you a proof of q so long as you can give me a proof of p . If p has no proofs, my job is done before I even started! The only way I can fail is if you have a proof of p but I have no proof of q .

Other operators ($\Leftrightarrow$, $\oplus$...)

There are many other operators we can define, but we will focus on just two more.

1.1.21 Definition. Let p and q be propositions. The proposition $p \Leftrightarrow q$ (read ‘ p if and only if q ’) ([L^AT_EX code: `\Leftrightarrow`](#)) is true when p and q have the same truth value, and false otherwise. The operator $\Leftrightarrow$ is called the **biconditional operator**.

1.1.22 Exercise. Show that $p \Leftrightarrow q$ is logically equivalent to $(p \Rightarrow q) \wedge (q \Rightarrow p)$.

PROOF TIP

To prove $p \Leftrightarrow q$ is true, you can give two proofs: a proof that $p \Rightarrow q$ is true, and a proof that $q \Rightarrow p$ is true.

1.1.23 Definition. Let p and q be propositions. The proposition $p \oplus q$ (read ‘ p or q but not both’) ([L^AT_EX code: `\oplus`](#)) is true when p and q have different truth values, and false otherwise. The operator $\oplus$ is called the **exclusive disjunction operator**.

Computer scientists and logicians often refer to $\oplus$ as ‘xor’ or ‘exclusive or’.

Proof principles

The foregoing section gives rise to some very useful proof techniques.

The first technique, known as the **law of the excluded middle**, allows you to prove a statement is true by showing that

1.1.24 Lemma. Let p, q, r be propositions. Then $(p \vee q) \Rightarrow r$ is logically equivalent to $(p \Rightarrow r) \wedge (q \Rightarrow r)$.

1.1.25 Theorem (Principle of the excluded middle). Let p be a proposition. Then $p \vee \neg p$ is true.

1.1.26 Exercise. Prove Lemma 1.1.24 and Theorem 1.1.25.

1.1.27 Corollary. Let p and q be propositions. If $p \Rightarrow q$ and $\neg p \Rightarrow q$ are true, then q is true.

Proof. By Lemma 1.1.24, it suffices to show that if $(p \vee \neg p) \Rightarrow q$ is true, then q is true. By the principle of the excluded middle (Theorem 1.1.25), $p \vee \neg p$ is true, and hence $(p \vee \neg p) \Rightarrow q$ is true if and only if q is true. But this is precisely what we wanted to prove. $\square$

PROOF TIP

To prove a proposition q is true, you can split into cases on another proposition p . In particular, you can give two proofs:

- A proof that $p \Rightarrow q$ is true;
- A proof that $\neg p \Rightarrow q$ is true.

1.1.28 Example. To do: PEM example

The second technique, known as *proof by contradiction*, allows you to prove a statement is true by showing that the assumption that it is false leads to a nonsensical claim.

1.1.29 Theorem (Principle of contradiction). Let p be a proposition, and suppose the proposition q is false. If $(\neg p) \Rightarrow q$ is true, then p is true.

Proof. Consider the following truth table:

p	q	$\neg p$	$(\neg p) \Rightarrow q$
✓	✓	×	✓
✓	×	×	✓
×	✓	✓	✓
×	×	✓	×

The only row in which q is false and $(\neg p) \Rightarrow q$ is true is the second row, in which p is true. Hence, from these assumptions, p must be true. $\square$

PROOF TIP

To prove a statement p is true, it suffices to show that its negation $\neg p$ implies something that is known to be false. We then say p is true ‘by contradiction’.

1.1.30 Example. We will prove by contradiction that chickens are not fish.

Let p denote the proposition ‘chickens are not fish’. Its negation $\neg p$ is the proposition ‘chickens are fish’. We show that $\neg p$ implies something known to be false.

Well, all fish have gills, so $\neg p$ implies that chicken have gills, which is false. This contradicts the assumption that chicken are fish. Hence, chicken are not fish. $\square$

We will put proof by contradiction to heavy use in later sections, when we have some mathematical objects to talk about.

The third proof technique is known as *proof by contraposition*, which is handy when proofs of implications are giving you trouble.

1.1.31 Theorem (Principle of contraposition). Let p and q be propositions. Then $p \Rightarrow q$ is logically equivalent to $(\neg q) \Rightarrow (\neg p)$.

Proof. Consider the following truth table:

p	q	$p \Rightarrow q$	$\neg q$	$\neg p$	$(\neg q) \Rightarrow (\neg p)$
✓	✓	✓	×	×	✓
✓	×	×	✓	×	×
×	✓	✓	×	✓	✓
×	×	✓	✓	✓	✓

Since the third and sixth columns are identical, the two propositions are logically equivalent. $\square$

PROOF TIP

To prove an implication $p \Rightarrow q$, you can instead prove the implication $\neg q \Rightarrow \neg p$; that is, assuming that q is false, show that p must be false. We then say $p \Rightarrow q$ is true ‘by contraposition’.

1.1.32 Example. Fix two natural numbers m and n . We will prove that if $mn > 64$, then either $m > 8$ or $n > 8$. Letting p be the proposition ‘ $mn > 64$ ’, q be the proposition ‘ $m > 8$ ’ and r be the proposition ‘ $n > 8$ ’, the statement ‘if $mn > 64$, then either $m > 8$ or $n > 8$ ’ becomes

$$p \Rightarrow (q \vee r)$$

By contraposition, this is equivalent to

$$\neg(q \vee r) \Rightarrow \neg p$$

By de Morgan’s laws, this is equivalent to

$$((\neg q) \wedge (\neg r)) \Rightarrow \neg p$$

Let’s spell this out. The proposition $\neg p$ means $mn \leq 64$, and the proposition $(\neg q) \wedge (\neg r)$ means that $m \leq 8$ and $n \leq 8$. So what we need to prove is:

If $m \leq 8$ and $n \leq 8$ then $mn \leq 64$.

Well this is certainly true! If you multiply two natural numbers which are less than or equal to 8, then their product must be less than or equal to 8^2 , which is equal to 64.

1.1.33 Corollary. Let p and q be propositions. Then $p \Leftrightarrow q$ is equivalent to

$$(p \Rightarrow q) \wedge ((\neg p) \Rightarrow (\neg q))$$

Proof. Left as an exercise. You can prove it directly, or apply reasoning you’ve already acquired to the result of Theorem 1.1.31. $\square$

Free and bound variables

If all we have to work with is propositions then our ability to do mathematical reasoning will be halted pretty quickly. For example, consider the following statement:

x is divisible by 7

This statement seems like the kind of thing we should probably be able to work with if we’re doing mathematics. It makes sense if x is a whole number, such as 28 or 41; but it doesn’t make sense at all if x is a parrot called Alex.⁴ In any case, even when it does make sense, its truth depends on the value of x ; indeed, ‘28 is divisible by 7’ is a true proposition, but ‘41 is divisible by 7’ is a false proposition.

This means that the statement ‘ x is divisible by 7’ isn’t a proposition—*quel horreur!* But it *almost* is a proposition: if we know that x refers somehow to a whole number, then it becomes a proposition as soon as a particular numerical value of x is specified. Such a symbol x is called a **free variable** or **parameter**. To indicate that a statement p contains x as a free variable, we will write $p(x)$. When we replace x by a specific value, say 28, we write $p(28)$; this is called **substitution** of a value for a variable.

Some statements might have several free variables. For example, the statement ‘ $y = x + 3$ ’ is a true proposition when $x = 3$ and $y = 6$, but it’s a false proposition when $x = 1$ and $y = 2$. What really matters is that we have a notion of what it is appropriate to use as values of x and y —namely, they should be numbers—and that whenever we use such values, what comes out is a proposition. To indicate that a statement p contains x and y as free variables, we will write $p(x, y)$.

⁴Alex the parrot is the only non-human animal to have ever been observed to ask an existential question; he died in September 2007. It is unlikely that Alex was divisible by 7, even when he was alive.

1.1.34 Definition. A **logical formula** is a statement containing some number of **variables**, each with a **range**, such that the statement becomes a proposition when values for all the variables are substituted from their respective ranges.

1.1.35 Example. As mentioned before, the statement $p(x)$ defined by ‘ x is divisible by 7’ is a formula if the range of x is the integers. For example, $p(28)$ is true and $p(41)$ is false.

1.1.36 Exercise. Write down a logical formula $p(x, y)$ with two free variables x, y ranging over the integers. Is the proposition $p(3, 7)$ true or false? For what values of y is $p(0, y)$ true?

We can obtain propositions from logical formulae in ways other than simply substituting for a variable. For example, the assertion that *every* substitution for a variable makes the formula true, is in itself a proposition. This can be done using *quantifiers*.

Universal quantifier ($\forall$)

The universal quantifier makes precise what we mean when we say ‘for all’, or ‘ $p(x)$ is always true no matter what value x takes’.

1.1.37 Definition. Let $p(x)$ be a logical formula. The proposition ‘ $\forall x, p(x)$ ’ (read ‘for all x , $p(x)$ ’) (**L^AT_EX** code: `\forall`) is true if $p(x)$ is true no matter what value of x is substituted, and false otherwise.

PROOF TIP

To prove a proposition of the form $\forall x, p(x)$, you can *introduce* x , with no assumptions about the value of x other than that it comes from its specified range.

A typical way of doing introducing x is saying ‘let x be arbitrary’, provided its range is understood. You may need to specify the range; for example, if it were the integers, you could introduce x by saying ‘let x be an arbitrary integer’. When you’ve done this, you can proceed to a proof that $p(x)$ is true.

The word ‘arbitrary’ is usually dropped in practice.

1.1.38 Exercise. Let $p(x)$ be the formula ‘ x is divisible by 7’, where x ranges over the integers. Write out the propositions $\forall x, p(x)$ and $\forall x, \neg p(x)$ in English.

1.1.39 Example. Consider the proposition

For all integers n , if n is even then $n + 1$ is odd.

This proposition takes the form $\forall n, (p(n) \Rightarrow q(n))$, where $p(n)$ is the statement ‘ n is even’ and $q(n)$ is the statement ‘ $n + 1$ is odd’. The range of n is implicitly specified to be the integers.

A proof would proceed as follows:

- (i) Let n be an (arbitrary) integer.
- (ii) Assume that n is even.
- (iii) From the above two assumptions, derive the fact that $n + 1$ is odd.

Step (i) is introduction of the variable n . For the rest of the proof we may treat n as if it’s any old integer, but whatever we say about n must be true no matter what value n takes. Having introduced n , we now need to prove $p(n) \Rightarrow q(n)$.

Step (ii) uses our proof strategy for proving implications: prove the proposition to the right of the $\Rightarrow$ symbol from the assumption that what is to the left of the $\Rightarrow$ symbol is true. This means that for the remainder of the proof, we may assume that n is even.

Step (iii) finishes off the proof.

COMMON ERROR

Consider the following (non-)proof of the proposition $\forall n, n^2 \geq 0$, where n ranges over the integers.

Let n be an arbitrary integer, say $n = 17$. Then $17^2 = 289 \geq 0$, so the statement is true. $\square$

The error made here is that the *writer* has picked an arbitrary value of n , not the *reader*. (In fact, the above argument actually proves $\exists n, n^2 \geq 0$; see below.)

Your proof should make no assumptions about the value of n other than its range. Here is a correct proof:

Let n be an arbitrary integer. Either $n \geq 0$ or $n < 0$. If $n \geq 0$ then $n^2 \geq 0$, since the product of two nonnegative numbers is nonnegative; if $n < 0$ then $n^2 \geq 0$, since the product of two negative numbers is positive. $\square$

Existential quantifier ($\exists$)

The existential quantifier makes precise what we mean when we say ‘there exists’, or ‘ $p(x)$ is true for some value of x in its range’.

1.1.40 Definition. Let $p(x)$ be a logical formula. The proposition ‘ $\exists x, p(x)$ ’ (read ‘there exists x such that $p(x)$ ’) ([L^AT_EX code: `\exists`](#)) is true if $p(x)$ is true for at least one substitution of the variable x .

PROOF TIP

To prove a proposition of the form $\exists x, p(x)$, you can find a particular value of x for which $p(x)$ is true, say when $x = a$, and prove $p(a)$.

1.1.41 Exercise. Let $p(x)$ be the formula ‘ x is divisible by 7’, where x ranges over the integers. Write out the propositions $\exists x, p(x)$ and $\exists x, \neg p(x)$ in English. For each, either prove that it is true, or prove that it is false.

1.1.42 Example. Consider the proposition

There exists a whole number which is odd and greater than 3.

This proposition takes the form $\exists n, (p(n) \wedge q(n))$, where $p(n)$ is the statement ‘ n is odd’ and $q(n)$ is the statement ‘ n is greater than 3’.

A proof would proceed by finding a particular value of n such that $p(n)$ and $q(n)$ are both true. Well, we know that 5 is odd, and 5 is certainly greater than 3! This means that $p(5) \wedge q(5)$ is true. Since we’ve proved the proposition for a value of n , we now know that $\exists n, (p(n) \wedge q(n))$ is true.

Quantifiers behave in an interesting way with the negation operator. Intuitively this makes sense: for example, to show ‘ x is even’ isn’t true for *all* x , it suffices to find a single x for which ‘ x is even’ is false. Thus, we can disprove $\forall x, (x \text{ is even})$ by proving $\exists x, (x \text{ is not even})$. This will be useful when cooking up proof strategies.

1.1.43 Theorem (De Morgan’s laws for quantifiers). Let $p(x)$ be a logical formula. Then

- (a) $\neg(\exists x, p(x))$ is logically equivalent to $\forall x, (\neg p(x))$;
- (b) $\neg(\forall x, p(x))$ is logically equivalent to $\exists x, (\neg p(x))$.

Proof. (a) We need to show that $\forall x, (\neg p(x))$ is true when $\neg(\exists x, p(x))$ is true, and false when it is false.

Suppose $\neg(\exists x, p(x))$ is true. Then $\exists x, p(x)$ is false, which means it is not the case that at least one value of x makes $p(x)$ true. Since no values of x make $p(x)$ true, this must mean that *all* values of x make $\neg p(x)$ true. So from the assumption that x takes any value whatsoever, we know that $\neg p(x)$ is true. Hence $\forall x, (\neg p(x))$ is true.

Conversely, suppose $\neg(\exists x, p(x))$ is false. Then $\exists x, p(x)$ is true, so there is some fixed value of x making $p(x)$ true. Therefore it is not the case that $\neg p(x)$ is true for all values of x : if x takes this special value then $p(x)$ is true, so $\neg p(x)$ is false! Hence $\forall x, (\neg p(x))$ is false.

The proof of (b) mimics the proof of (a) and is left to the reader. $\square$

Bound variables

When a variable is quantified, we say it is **bound**. Bound variables behave differently from free variables in a number of ways, for example

- Propositions cannot have free variables, but they can have bound variables.
- It is possible to substitute a value for a free variable, but not for a bound variable.

1.1.44 Example. Consider the formula, where x, y, z range over the natural numbers:

$$\forall x, \exists y, x^2 + y^2 + z^2 = 1$$

In this formula, the variables x and y are bound, but the variable z is free. To see this, note that we *can* substitute for z ; substituting 2 for z yields:

$$\forall x, \exists y, x^2 + y^2 + 2^2 = 1$$

which is a false proposition. However we cannot substitute for x or y ; trying to substitute 2 for x yields:

$$\forall 2, \exists y, 2^2 + y^2 + z^2 = 1$$

which *must* be nonsense: the phrase ‘for all 2, ...’ doesn’t even make sense!

1.1.45 Exercise. For each of the following formulae, where all variables range over the integers, write down the formula using quantifiers and specify which variables are free and which are bound:

- If n is prime and $n > 2$ then n is odd.
- There exist x and y such that $ax + by = 1$.
- No integer value of x satisfies $0x = 1$.

Summary of logical operators

Operator	Notation	True when...	False when...
conjunction	$p \wedge q$	p, q both true	p false or q false or both
disjunction	$p \vee q$	p true or q true or both	p, q both false
negation	$\neg p$	p false	p true
implication	$p \Rightarrow q$	p false, or p, q both true	p false and q true
biconditional	$p \Leftrightarrow q$	p, q both true or both false	p true and q false, or vice versa
xor	$p \oplus q$	p true or q true, but not both	p, q both true or both false

Summary of truth tables

Conjunction ('and')

p	q	$p \wedge q$
✓	✓	✓
✓	×	×
×	✓	×
×	×	×

Disjunction ('or')

p	q	$p \vee q$
✓	✓	✓
✓	×	✓
×	✓	✓
×	×	×

Negation ('not')

p	$\neg p$
✓	×
×	✓

Implication ('if-then')

p	q	$p \Rightarrow q$
✓	✓	✓
✓	×	×
×	✓	✓
×	×	✓

Biconditional ('if and only if')

p	q	$p \Leftrightarrow q$
✓	✓	✓
✓	×	×
×	✓	×
×	×	✓

Exclusive disjunction

p	q	$p \oplus q$
✓	✓	×
✓	×	✓
×	✓	✓
×	×	×

Algebra of logical operators

Rule name	Equivalence	Dual equivalence
Double-negation	$\neg\neg p \equiv p$	
Commutativity	$p \wedge q \equiv q \wedge p$	$p \vee q \equiv q \vee p$
Associativity	$p \wedge (q \wedge r) \equiv (p \wedge q) \wedge r$	$p \vee (q \vee r) \equiv (p \vee q) \vee r$
Distributivity	$p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$	$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$
De Morgan's laws	$\neg(p \wedge q) \equiv (\neg p) \vee (\neg q)$	$\neg(p \vee q) \equiv (\neg p) \wedge (\neg q)$
Idempotence	$p \wedge p \equiv p$	$p \vee p \equiv p$
Absorption	$p \wedge (p \vee q) \equiv p$	$p \vee (p \wedge q) \equiv p$

Summary of proof principles

Rule name	Result
Principle of excluded middle	$p \vee \neg p$ is always true.
Principle of noncontradiction	$p \wedge \neg p$ is always false.
Principle of contradiction	If q is false and $\neg p \Rightarrow q$ is true, then p is true.
Principle of contraposition	$p \Rightarrow q \equiv (\neg q) \Rightarrow (\neg p)$

Summary of quantifiers

Quantifier	Notation	Proof strategy
universal (‘for all’)	$\forall x, p(x)$	Prove $p(x)$ is true regardless of what x is ...or show no x for which $p(x)$ is false can exist
existential (‘there exists’)	$\exists x, p(x)$	Find x such that $p(x)$ is true ...or show it’s not possible that $p(x)$ is false for every x

1.2 Sets and set operations

GOAL

After working through this section you should be able to **To do: finish goal**

With a system of logical notation under our belt, we're now ready to introduce the first kind of mathematical object: a *set*. You can think of a set as just being a collection of things, though this is purely for intuitive purposes—if you take ‘a set is a collection of things’ as your definition of a set, you can't go far before you've contradicted yourself.⁵

1.2.1 Definition. A **set** is a collection of things from a specified range, called a **universe of discourse**. The collection of everything in the range is called the **universal set** (or just **universe**), denoted $\mathcal{U}$ (`LATEX` code: `\mathcal{U}`).

1.2.2 Example. If the universe of discourse is the integers, then $\mathcal{U}$ is the set of all integers.

1.2.3 Definition. An **element** of a set X is simply something which is in X . The formula $x \in X$ (`LATEX` code: `\in`) denotes the fact that x is an element of X , where the range of x is the universe of discourse. We write $x \notin X$ to mean $\neg(x \in X)$, i.e. that x is not an element of X .

1.2.4 Example. If the universe of discourse is the integers, and X is the set of all even numbers, then $6 \in X$ is a true proposition, $3 \in X$ is a false proposition, and the formula $x \in X$ is the assertion that the variable x refers an even number. The statement $\frac{1}{2} \in X$ is not a proposition because $\frac{1}{2}$ is not in the universe of discourse; the same goes for the statement ‘rabbit $\in X$ ’.

If we widened our universe of discourse to include fractions too, then $\frac{1}{2} \in X$ becomes a false proposition, and ‘rabbit $\in X$ ’ is still not a proposition.

1.2.5 Exercise. Suppose $\mathcal{U}$ is the set of states of the USA, i.e. that our universe of discourse is the states of the USA. Let X be the set of states whose name contains the letter ‘N’. Give an example of an element of X , and something which isn't an element of X . Is the statement ‘England $\in X$ ’ a proposition? If so, say whether it is true or false and why; if not, give an example of a universe of discourse for which it *is* a proposition.

Specifying a set

One way of defining a set is simply to describe it in words, like we have done up to now. There are other, more concise ways, of specifying sets, which also remove such ambiguity from the process.

⁵This is beyond the scope of this course, but if you're interested, read about *Russell's paradox*.

Lists

One way is simply to provide a **list** the elements of the set. To specify that the list denotes a set, we enclose the list with braces $\{, \}$ ([L^AT_EX code: `\{, \}`](#)). For example, if our domain of discourse is colours, we can write

$$X = \{\text{brown, red, purple, blue}\}$$

to denote the fact that X is the set whose elements are the colours brown, red, purple and blue.

Sometimes a list might be long—perhaps even infinite—or the length of the list might depend on a variable. In these cases it will be convenient to use an **implied list**, in which some elements of the list are written, and the rest are left implicit by writing an ellipsis ‘ $\dots$ ’ ([L^AT_EX code: `\dots`](#)). For example, the statement

$$X = \{1, 4, 9, \dots, n^2\}$$

means that X is the set whose elements are all the square numbers from 1 to n^2 , where n is some number.

Set-builder notation

In general, implied lists can be ambiguous, so in practice they are avoided unless the implied list is very simple, such as a set of consecutive numbers like $\{3, 4, \dots, 9\}$. In fact, many sets can’t even be listed in this way.

To get around this, we can use **set-builder notation**, which is a means of specifying a set in terms of the properties its elements satisfy. The set X of things x for which $p(x)$ is true is denoted by

$$X = \{x \mid p(x)\}$$

The bar ‘ $\mid$ ’ ([L^AT_EX code: `\mid`](#)) separates the variable name from the formula that they make true. Some authors use a colon ‘ $:$ ’ or semicolon ‘ $;$ ’ instead.

1.2.6 Example. If $\mathcal{U}$ is the set of all integers, then

$$\{n \mid n \text{ is even}\} = \{\dots, -4, -2, 0, 2, 4, 6, \dots\}$$

When using set-builder notation, keep track of the ambient universe of discourse! For example, if $\mathcal{U}$ is the set of natural numbers, then

$$\{n \mid n \text{ is even}\} = \{2, 4, 6, \dots\}$$

PROOF TIP

When a set X is expressed in set-builder notation, say $X = \{x \mid p(x)\}$, then the statement $x \in X$ is true precisely when $p(x)$ is true. In other words, to prove $x \in X$, you can prove $p(x)$. Likewise, to prove $x \notin X$, you can prove $\neg p(x)$.

1.2.7 Exercise. Express the set of all perfect squares in set-builder notation and as an implied list. You should specify your universe of discourse.

You're probably tired of worrying about ranges and universes—and so am I. We can use the language of set theory to avoid them completely by specifying the ranges of the variables we use as soon as they appear. For example, given a set X :

- The proposition $\forall x \in X, p(x)$ means that x has range X and $\forall x, p(x)$. It is equivalent to $\forall x, x \in X \Rightarrow p(x)$, so long as the range of x contains all the elements of X .
- The proposition $\exists x \in X, p(x)$ means that x has range X and $\exists x, p(x)$. It is equivalent to $\exists x, x \in X \wedge p(x)$, so long as the range of x contains all the elements of X .
- The set $\{x \in X \mid p(x)\}$ denotes the set $\{x \mid p(x)\}$, where the range of x is X .

From now on, this is the style that we will use, and the universe $\mathcal{U}$ will be assumed to include all the mathematical objects that we define or need.

We can also use set-builder notation to specify the form of the elements of a set. For example, the set

$$Z = \{3x + 2 \mid x \text{ is an integer}\}$$

denotes the set of things of the form $3x + 2$ where x is an integer. Thus

$$Z = \{\dots, -7, -4, -1, 2, 5, 8, 11, \dots\}$$

Some examples of sets

From now on our universe of discourse will, unless otherwise specified, include all mathematical objects that we define. With this in mind, there are some very important sets to be defined.

Empty set

The first set we define is very simple:

1.2.8 Definition. The **empty set**, denoted $\emptyset$ ([L^AT_EX code: \varnothing](#)), is the set with no elements. If a set contains at least one element, it is said to be **inhabited** or **nonempty**.⁶

The empty set may seem trivial—and it is—but owing to its canonicity, it arises all over the place, and will be especially important when we come to talk about functions and cardinality in Chapter 3.

1.2.9 Exercise. Let $p(x)$ be any formula. Show that the proposition $\forall x, (x \in \emptyset \Rightarrow p(x))$ is true. What does the proposition $\forall x, (x \in \emptyset \Rightarrow x \neq x)$ mean in English? Is it true?

The remaining sets are all sets of numbers; you may recognise some of the notation used if you have taken a calculus class before.

Natural numbers

1.2.10 Definition. A **natural number** is a non-negative whole number. The set of all natural numbers is denoted by $\mathbb{N}$ ([L^AT_EX code: \mathbb{N}](#)). In implied list form:

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}$$

The natural numbers are so important that a whole section is dedicated to them; see Section 1.3. A more formal definition of *natural number* is given in Definition 1.3.1.

Some authors define $\mathbb{N} = \{1, 2, 3, \dots\}$, the set of all *positive* whole numbers. The choice about whether $0 \in \mathbb{N}$ or $0 \notin \mathbb{N}$ is a matter of taste or convenience, and is merely a convention. By default we will take $0 \in \mathbb{N}$, with the philosophy that ‘the natural numbers are the sizes of finite sets’, and not just the inhabited ones.

Integers

1.2.11 Definition. An **integer** is a whole number, which may be positive or negative or zero. The set of all integers is denoted by $\mathbb{Z}$ ([L^AT_EX code: \mathbb{Z}](#)). In implied list form:

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

We will study the properties of the integers in depth when we delve into number theory in Chapter 2.

⁶I will use the word ‘inhabited’, because ‘nonempty’ means ‘doesn’t not have elements’, which is just silly.

Rational numbers

1.2.12 Definition. A **rational number** is a number which can be expressed as an integer divided by a non-zero integer. The set of all rational numbers is denoted by $\mathbb{Q}$ ([L^AT_EX code: `\mathbb{Q}`](#)). In set-builder notation:

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

The rational numbers have the structure of a *field*. We will not study fields in this course, but they are extremely important in the area of study known as *abstract algebra*. The rational numbers have some interesting properties—for example, in a sense of ‘size’ to be defined in Section 4.3, the set of rational numbers $\mathbb{Q}$ is the same size as the set of natural numbers $\mathbb{N}$.

1.2.13 Exercise. Show that every integer is a rational number. Can you give an example of a rational number that is not an integer? Think about how you would prove that this is so, or—if you can’t prove it—what additional information you need in order to give a proof.

ASIDE

The letter $\mathbb{Z}$ stands for *Zahlen*, which is the German word for *numbers*. The letter $\mathbb{Q}$ stands for *quoziente*, which is the Italian word for *quotient*. Arguably, the $\mathbb{Q}$ just stands for *quotient*, but it was first used by an Italian mathematician called Giuseppe Peano.

Real numbers

The following definition is slightly less precise than the previous three; this is because the precise definition requires more mathematical machinery than we can currently cope with.

1.2.14 Definition. A **real number** is any value that can be represented on an infinite number line. The set of all real numbers is denoted by $\mathbb{R}$ ([L^AT_EX code: `\mathbb{R}`](#)).

Intuitively, a real number is anything that could conceivably be thought of as a ‘length’, or the negative of such a thing. The real numbers are central to the area of study known as *analysis*.

FORMATTING TIP

The *blackboard bold* font ([L^AT_EX code: `\mathbb{b}`](#)) is used for the number sets to distinguish them from variables. For example, it is not uncommon to see things written like $N \in \mathbb{N}$, to denote that the letter N refers to a natural number. Without formatting, this would just

be $N \in N$, which as a statement is necessarily either ambiguous or false!

Every rational number is real, but not every real number is rational. Determining whether a given real number is rational or not is, in general, a very difficult problem.

1.2.15 Definition. An **irrational number** is a real number that is not rational.

For example, $\sqrt{2}$ and π are irrational. The proof that $\sqrt{2}$ is irrational is a classic *proof by infinite descent*, see Proposition 1.3.17. Later on, you will prove that $\sqrt{n}$ is either an integer or an irrational number. The proof that π is irrational is much harder, and requires more advanced theory than we will get to in this course.

Subsets and set equality

Much of the discussion above concerned when an element of one set is or is not an element of another. For example, in exercise 1.2.13, you proved that every integer is a rational number; that is, you proved

$$\forall n, (n \in \mathbb{Z} \Rightarrow n \in \mathbb{Q})$$

We can say this more concisely by saying that $\mathbb{Z}$ is a *subset* of $\mathbb{Q}$.

1.2.16 Definition. Let X and Y be sets. We say X is a *subset* of Y if $\forall x, (x \in X \Rightarrow x \in Y)$. We abbreviate this proposition by writing $X \subseteq Y$ ([L^AT_EX code: \subseteq](#)), and we write $X \not\subseteq Y$ ([L^AT_EX code: \not\subseteq](#)) for its negation.

PROOF TIP

A proof that X is a subset of Y typically proceeds as follows. Let $x \in X$ be arbitrary; then knowing nothing about x other than the fact that $x \in X$, prove that $x \in Y$.

1.2.17 Exercise. Let X be a set. Prove that $\emptyset \subseteq X$ and that $X \subseteq X$.

1.2.18 Example. From our previous discussions, we know that there is a chain of subsets given by:

$$\emptyset \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$$

The following proposition proves a property of subethood known as *transitivity*.

1.2.19 Proposition. Let X, Y, Z be sets. If $X \subseteq Y$ and $Y \subseteq Z$, then $X \subseteq Z$.

Strategy. The result we want to prove is an implication. Thus we *assume* $X \subseteq Y$ and $Y \subseteq Z$, and our *goal* is to derive that $X \subseteq Z$. Spelling this out slightly more, the goal is to derive $\forall x, x \in X \Rightarrow x \in Z$; so we can introduce a variable x and assume that $x \in X$. Then our goal is to use our assumptions to prove that $x \in Z$. Well, $X \subseteq Y$ means $\forall x, x \in X \Rightarrow x \in Y$. Since we're assuming $x \in X$, substituting it into this assumption yields that $x \in Y$. Likewise, the assumption that $Y \subseteq Z$ yields that $x \in Z$.

Proof. Suppose that $X \subseteq Y$ and $Y \subseteq Z$. We need to prove that every element of X is an element of Z . So let $x \in X$. Since $X \subseteq Y$, it follows that $x \in Y$; and since $Y \subseteq Z$, it follows that $x \in Z$. Hence $X \subseteq Z$. $\square$

ASIDE

Notice how in the proof of Proposition 1.2.19 we omitted many of the details of the thought process that went into coming up with the proof: decomposing the logical structure of the proposition to be proved, spelling out what our goal is at every step, and so on. We left enough of an argument to convince a mathematically literate reader that we're correct, but kept it concise enough that attention is drawn to the important steps.

Set equality

1.2.20 Exercise. Without reading further, discuss what you think ' $X = Y$ ' means, when X and Y are sets. Be as precise as you can.

There are different possible notions of 'sameness' for sets: maybe $X = Y$ when X and Y have the same elements (this is called *extensional equality*), or maybe $X = Y$ when they're described by the same criteria (this is called *intensional equality*). In mathematics, it is more useful to know when two sets have the same elements, regardless of how they are described; so we take extensional equality as our notion of sameness for sets. This doesn't mean intensional equality should be ignored—if you want to implement mathematics in a computer, the sets' *descriptions* have a much more important role to play.

1.2.21 Definition. Let X and Y be sets. We say X is *equal* to Y if $X \subseteq Y$ and $Y \subseteq X$, and we write $X = Y$. If $X \subseteq Y$ and $X \neq Y$ then we say X is a **proper subset** of Y and write $X \subsetneq Y$ ([L^AT_EX code: \subsetneqq](#)).

1.2.22 Example. Let $E = \{n \in \mathbb{Z} \mid n \text{ is even}\}$. Then:

- $E \subsetneq \mathbb{Z}$. Indeed, $E \subseteq \mathbb{Z}$ since every element of E is an element of $\mathbb{Z}$ by definition; but $E \neq \mathbb{Z}$ since, for instance, $1 \in \mathbb{Z}$ but $1 \notin E$.

- $\mathbb{N} \not\subseteq E$ since, for instance, $1 \in \mathbb{N}$ but $1 \notin E$.
- $E \not\subseteq \mathbb{N}$ since, for instance, $-2 \in E$ but $-2 \notin \mathbb{N}$.

1.2.23 Exercise. Define a set X such that:

$$\mathbb{N} \subsetneq X \quad \wedge \quad X \subsetneq \mathbb{Q} \quad \wedge \quad X \not\subseteq \mathbb{Z} \quad \wedge \quad \mathbb{Z} \not\subseteq X$$

PROOF TIP

To prove $X = Y$, you can prove that $X \subseteq Y$ and $Y \subseteq X$. This proof strategy is called *double-containment*. More specifically, such a proof is split into two parts:

- (i) Let $x \in X$; from this assumption alone, prove that $x \in Y$.
- (ii) Let $x \in Y$; from this assumption alone, prove that $x \in X$.

Set operations

In Example 1.2.22 we defined E to be the set of all even integers. What if we wanted to talk about the set of all even natural numbers instead? It would be nice if there was some expression in terms of E and $\mathbb{N}$ to denote this set. This is where *set operations* come in.

Intersection ($\cap$)

The *intersection* of two sets is the set of things which are elements of both sets.

1.2.24 Definition. Let X and Y be sets. The (**pairwise**) **intersection** of X and Y , denoted $X \cap Y$ ([L^AT_EX](#) code: `\cap`), is defined by

$$X \cap Y = \{x \mid x \in X \wedge x \in Y\}$$

1.2.25 Example. Let E be the set of all even integers. Then $E \cap \mathbb{N}$ refers to the set of things which are both even integers and natural numbers... in other words, it is the set of even natural numbers.

PROOF TIP

To prove $x \in X \cap Y$ you can give two proofs: one that $x \in X$ and one that $x \in Y$. Therefore, if $X = \{x \mid p(x)\}$ and $Y = \{x \mid q(x)\}$, then $X \cap Y = \{x \mid p(x) \wedge q(x)\}$.

1.2.26 Exercise. Let X and Y be sets. Prove that $X \subseteq Y$ if and only if $X \cap Y = X$.

On occasion, we might be working with lots of sets—maybe even infinitely many—and we want to know when something is an element of *all* of them. When we just have two sets, X and Y , to say that x is an element of both is precisely to say that $x \in X \cap Y$. This can be extended to finitely many sets; for example, $x \in (X \cap Y) \cap Z$ holds precisely when x is an element of all three sets X , Y and Z . This notation soon becomes cumbersome, and for this reason we introduce *indexed families* of sets.

1.2.27 Definition. An **indexed family** is a set F whose elements are labelled by a set I , called an **index set**, such that every element of I labels a single element of F . We typically denote an indexed family by

$$\{x_i \mid i \in I\}$$

where x_i denotes the element of the indexed family with label i .

Thus an indexed family of sets is simply an indexed family, whose elements are themselves sets!

1.2.28 Example. For each $n \in \mathbb{N}$, define $X_n = \{0, 1, \dots, n\}$; that is, X_n is the set of all natural numbers less than or equal to n . This defines an indexed family $\{X_n \mid n \in \mathbb{N}\}$; here, the index set is $\mathbb{N}$.

1.2.29 Definition. Let $F = \{X_i : i \in I\}$ be an indexed family of sets. The **(indexed) intersection** of F is the set $\bigcap_{i \in I} X_i$ (**L^AT_EX** code: `\bigcap_{i \in I} X_i`) defined by

$$x \in \bigcap_{i \in I} X_i \quad \Leftrightarrow \quad \forall i \in I, x \in X_i$$

That is, $x \in \bigcap_{i \in I} X_i$ if and only if x is an element of all the sets in the indexed family.

1.2.30 Example. In Example 1.2.28, the number 0 is an element of each of the sets X_n . Thus $0 \in \bigcap_{n \in \mathbb{N}} X_n$.

1.2.31 Example. If X_1 and X_2 are sets, then $\{X_1, X_2\}$ is a family of sets indexed by the index set $I = \{1, 2\}$. Then $x \in \bigcap_{i \in I} X_i$ if and only if $x \in X_1$ and $x \in X_2$. This proves that

$$\bigcap_{i \in \{1, 2\}} X_i = X_1 \cap X_2$$

In other words, pairwise intersection is a special case of indexed intersection.

1.2.32 Exercise. Define a family of sets indexed by $\mathbb{Z}$. Swap with a neighbour and find the intersection of their family of sets.

Union ($\cup$)

The *union* of two sets is the set of things which are elements of at least one of the sets.

1.2.33 Definition. Let X and Y be sets. The (**pairwise**) **union** of X and Y , denoted $X \cup Y$ (`\cup`), is defined by

$$X \cup Y = \{x \mid x \in X \vee x \in Y\}$$

1.2.34 Example. Let E be the set of even integers and O be the set of odd integers. Since every integer is either even or odd, $E \cup O = \mathbb{N}$. Note that $E \cap O = \emptyset$, thus $\{E, O\}$ is an example of a *partition* of $\mathbb{N}$; see **To do: insert reference**.

1.2.35 Exercise. Let X and Y be sets. Prove that $X \subseteq Y$ if and only if $X \cup Y = Y$.

1.2.36 Definition. Let $F = \{X_i : i \in I\}$ be an indexed family of sets. The (**indexed**) **union** of F is the set $\bigcup_{i \in I} X_i$ (`\bigcup_{i \in I} X_i`) defined by

$$x \in \bigcup_{i \in I} X_i \iff \exists i \in I, x \in X_i$$

That is, $x \in \bigcup_{i \in I} X_i$ if and only if x is an element of at least one of the sets in the indexed family.

1.2.37 Example. In Example 1.2.28, every natural number n is an element of the set X_n that it indexes. Moreover, every element of each of the X_n is itself a natural number, and therefore $\bigcup_{n \in \mathbb{N}} X_n = \mathbb{N}$.

1.2.38 Example. If X_1 and X_2 are sets, then $x \in \bigcup_{i \in \{1,2\}} X_i$ if and only if $x \in X_1$ or $x \in X_2$. This proves that

$$\bigcup_{i \in \{1,2\}} X_i = X_1 \cup X_2$$

In other words, pairwise union is a special case of indexed union.

1.2.39 Exercise. Define a family of sets indexed by $\mathbb{Z}$. Swap with a neighbour and find the union of their family of sets.

Relative complement ($\setminus$) and complement ($-^c$)

1.2.40 Definition. Let X and Y be sets. The **relative complement** of Y in X , denoted $X \setminus Y$ (`\setminus`), is defined by

$$X \setminus Y = \{x \in X \mid x \notin Y\}$$

If X is a set then the **complement** of X , denoted X^c (**L^AT_EX** code: $\mathbf{X^c}$), is simply the relative complement of X in the universal set: $X^c = \mathcal{U} \setminus X$.

1.2.41 Example. Let E be the set of all even integers. Then $n \in \mathbb{Z} \setminus E$ if and only if n is an integer and n is not an even integer; that is, if and only if n is odd. Thus $\mathbb{Z} \setminus E$ is the set of all odd integers.

Moreover, $n \in \mathbb{N} \setminus E$ if and only if n is a natural number and n is not an even integer. Since the even integers which are natural numbers are precisely the even natural numbers, $\mathbb{N} \setminus E$ is precisely the set of all odd natural numbers.

1.2.42 Exercise. Let X and Y be sets. Prove that $X \subseteq Y$ if and only if $Y \setminus (Y \setminus X) = X$.

Comparison with logical operators and quantifiers

The astute reader will have noticed some similarities between set operations and the logical operators and quantifiers that we saw in Section 1.1. Indeed, this can be summarised in the following table. In each row, the expressions in both columns are equivalent, where p denotes ' $x \in X$ ', q denotes ' $x \in Y$ ', and $r(i)$ denotes ' $x \in X_i$ ':

sets	logic
$x \in X \cap Y$	$p \wedge q$
$x \in X \cup Y$	$p \vee q$
$x \in \bigcap_{i \in I} X_i$	$\forall i \in I, r(i)$
$x \in \bigcup_{i \in I} X_i$	$\exists i \in I, r(i)$
$x \in X^c$	$\neg p$
$x \in X \setminus Y$	$p \wedge (\neg q)$

This translation between logic and set theory does not stop there; in fact, De Morgan's laws for logical operators and quantifiers carry straight over:

1.2.43 Theorem (De Morgan's laws for sets). Let Z be a set and let $\{X_i \mid i \in I\}$ be an indexed family of sets. Then

$$(a) \quad Z \setminus \bigcup_{i \in I} X_i = \bigcap_{i \in I} (Z \setminus X_i);$$

$$(b) \quad Z \setminus \bigcap_{i \in I} X_i = \bigcup_{i \in I} (Z \setminus X_i).$$

Proof 1 of (a). In this proof, we prove (a) directly by unpacking the definitions of relative complement, indexed union and indexed intersection.

Fix z . Note that $z \in Z \setminus \bigcup_{i \in I} X_i$ if and only if

$$z \in Z \wedge \neg \left(z \in \bigcup_{i \in I} X_i \right)$$

by definition of relative complement. This holds if and only if

$$z \in Z \wedge \neg(\exists i \in I, z \in X_i)$$

by definition of indexed union. This holds if and only if

$$z \in Z \wedge \forall i \in I, z \notin X_i$$

by De Morgan's laws for quantifiers (Theorem 1.1.43). Since the proposition $z \in Z$ doesn't depend on i , this holds if and only if

$$\forall i \in I, (z \in Z \wedge z \notin X_i)$$

which is precisely the statement that $z \in \bigcap_{i \in I} (Z \setminus X_i)$. □

Proof 2 of (a). In this proof, we prove (a) by a double-containment argument.

- $Z \setminus \bigcup_{i \in I} X_i \subseteq \bigcap_{i \in I} (Z \setminus X_i)$.

Let $z \in Z \setminus \bigcup_{i \in I} X_i$. We know that $z \in Z$ and $z \notin \bigcup_{i \in I} X_i$. We need to prove that $z \in \bigcap_{i \in I} (Z \setminus X_i)$; that is, we need to prove that, for all $i \in I$, we have $z \in Z \setminus X_i$; that is, $z \in Z$ and $z \notin X_i$. We have $z \in Z$ for free, so all we have to prove is that, for all $i \in I$, $z \notin X_i$.

So let $i \in I$. If $z \in X_i$ then $z \in \bigcup_{i \in I} X_i$, contradicting the fact that $z \notin \bigcup_{i \in I} X_i$. Therefore it must be the case that $z \notin X_i$. This finishes this half of the proof.

- $Z \setminus \bigcup_{i \in I} X_i \supseteq \bigcap_{i \in I} (Z \setminus X_i)$.

Let $z \in \bigcap_{i \in I} (Z \setminus X_i)$. We know that, for all $i \in I$, $z \in Z \setminus X_i$. Hence it's certainly true that $z \in Z$. To prove that $z \in Z \setminus \bigcup_{i \in I} X_i$, it remains to prove that $z \notin \bigcup_{i \in I} X_i$.

Suppose $z \in \bigcup_{i \in I} X_i$. Then $z \in X_i$ for some $i \in I$. Since we already know that $z \in Z$, it follows that $z \notin Z \setminus X_i$, contradicting the fact that $z \in Z \setminus X_i$ for all $i \in I$. This finishes the second half of the proof.

We have shown containment in both directions, hence equality. □

Statement (b) can be proved analogously, and it left as an exercise.

1.2.44 Corollary. Let $\{X_i : i \in I\}$ be an indexed family of sets. Then

- (a) $(\bigcup_{i \in I} X_i)^c = \bigcap_{i \in I} X_i^c$;
 (b) $(\bigcap_{i \in I} X_i)^c = \bigcup_{i \in I} X_i^c$.

Proof. Take $Z = \mathcal{U}$ in Theorem 1.2.43. □

1.2.45 Corollary. Let X, Y, Z be sets. Then

- (a) $Z \setminus (X \cup Y) = (Z \setminus X) \cap (Z \setminus Y)$;
 (b) $Z \setminus (X \cap Y) = (Z \setminus X) \cup (Z \setminus Y)$.

Proof. Take $I = \{1, 2\}$ in Theorem 1.2.43, so that the indexed families of sets consist of just two sets. □

ASIDE

Aaand breathe! All this new notation can be overwhelming at first and you'd be forgiven for wanting to give up at this point. It will get better soon and you'll reap the benefits if you keep at it. This chapter is all about teaching you a new language, because without it, any of our future pursuits will be impossible. If you're stuck, don't worry, you'll soon get the hang of it when we've started using it in context, and you can refer back to the results in this chapter for reference.

Product ($\times$, Π)

1.2.46 Definition. Let X and Y be sets. The (**pairwise**) **product** of X and Y , denoted $X \times Y$ ([L^AT_EX](#) code: `\times`), is the set of all **ordered pairs** (x, y) , where $x \in X$ and $y \in Y$. That is,

$$X \times Y = \{(x, y) \mid x \in X \wedge y \in Y\}$$

1.2.47 Example. If you have ever taken calculus, you will probably be familiar with the set $\mathbb{R} \times \mathbb{R}$.

$$\mathbb{R} \times \mathbb{R} = \{(x, y) \mid x, y \in \mathbb{R}\}$$

Formally, this is the set of ordered pairs of real numbers. Geometrically, if we interpret $\mathbb{R}$ as an infinite line, the set $\mathbb{R} \times \mathbb{R}$ is the (real) plane: an element $(x, y) \in \mathbb{R} \times \mathbb{R}$ describes the point in the plane with coordinates (x, y) .

We can investigate this further. For example, the following set:

$$\mathbb{R} \times \{0\} = \{(x, 0) \mid x \in \mathbb{R}\}$$

is precisely the x -axis. We can describe graphs as subsets of $\mathbb{R} \times \mathbb{R}$. Indeed, the graph of $y = x^2$ is given by

$$G = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = x^2\} = \{(x, x^2) \mid x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$$

1.2.48 Exercise. Let X be a set. Prove that $X \times \emptyset = \emptyset$.

There is also an indexed version of a product, which allows us to take the product of infinitely many sets:

1.2.49 Definition. Let $F = \{X_i : i \in I\}$ be an indexed family of sets. The **(indexed) product** of F is the set $\prod_{i \in I} X_i$ (**L^AT_EX** code: `\prod_{i \in I} X_i`) defined by

$$\prod_{i \in I} X_i = \{(x_i)_{i \in I} \mid \forall i \in I, x_i \in X_i\}$$

That is, the elements of $\prod_{i \in I} X_i$ are **sequences** of elements, where for each $i \in I$, the term in the sequence indexed by i is an element of the set X_i .

1.2.50 Example. The sets X_n from Example 1.2.28 were defined by $X_n = \{0, 1, \dots, n\}$ for $n \in \mathbb{N}$. The set $\prod_{n \in \mathbb{N}} X_n$ has as its elements sequences of the form

$$(x_0, x_1, x_2, \dots)$$

where $x_n \in X_n$ for each $n \in \mathbb{N}$. That is, $\prod_{n \in \mathbb{N}} X_n$ is the set of infinite sequences of natural numbers, where for each n , the number in the n^{th} position is less than or equal to n .

ASIDE

Sequences are subtly different from indexed families, in the sense that the *order* of the elements matters for sequences, but it does not for indexed families. To illustrate: let $x_1 = 7$ and $x_2 = 14$. These define both a family and a sequence indexed by the index set $\{1, 2\}$. Now

$$\{x_1, x_2\} = \{x_2, x_1\} \quad \text{but} \quad (x_1, x_2) \neq (x_2, x_1)$$

since $\{x_1, x_2\}$ is just the set whose elements are 7 and 14, but (x_1, x_2) is the sequence with 7 in its first position and 14 in its second position.

1.3 Induction and recursion

GOAL

After working through this section you should be able to prove elementary statements about natural numbers using the *weak* and *strong induction principles*.

We defined the natural numbers in Definition 1.2.10; to reiterate, they are the non-negative whole numbers

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

and we denote the set of natural numbers by $\mathbb{N}$. This was an *informal* definition: it assumed that we have an inherent notion in our minds of what a whole number is, what 0 is, and so on. And we probably *do* have such an inherent notion in our minds; it's so ingrained that you wouldn't think twice about what I mean when I write $3 + 15$ or 7×12 , even though I haven't defined what $+$ or $\times$ mean (or even what 3, 15, 7 and 12 mean).

This informal approach gets us into some trouble if we really want to be precise about what we're doing, and so Definition 1.2.10 won't suffice. The next definition characterises the natural numbers in terms of the number 0 and the operation of 'adding 1', which we call the *successor*. From these five properties, called the *Peano axioms*, everything we know about natural numbers can be derived.

1.3.1 Definition. The set of **natural numbers** is the set $\mathbb{N}$ defined by:

- (a) $\mathbb{N}$ contains a **zero element**, denoted 0;
- (b) If $n \in \mathbb{N}$ then there is an element $n^+ \in \mathbb{N}$, called the **successor** of n ;
- (c) If $n \in \mathbb{N}$ then $n^+ \neq 0$;
- (d) If $m, n \in \mathbb{N}$ and $m^+ = n^+$, then $m = n$;
- (e) If X is a set such that $0 \in X$ and, for each $n \in \mathbb{N}$, $n \in X \Rightarrow n^+ \in X$, then $\mathbb{N} \subseteq X$.

Most of these properties are reasonably self-explanatory. For example, (c) says that there isn't a natural number n such that $n + 1 = 0$...if there were, then we'd have $n = -1$ but -1 isn't a natural number. And (d) says that if $m + 1 = n + 1$ then $m = n$; this makes sense because we should be able to 'subtract 1' from both sides of the equation.

The property that requires some discussion is (e). In slightly more human terms, it says: if a set X contains 0 and the successors of all its elements, then it contains $\mathbb{N}$ as a subset. Why should

this be so? Well, we know $0 \in X$. Since it's closed under successor, it contains $0 + 1$, which is 1; and it contains $1 + 1$, which is 2; and it contains $2 + 1$, which is 3; and so on.

From these five properties (a)–(e), we can recover everything we know about the natural numbers. For instance:

- **Numerals.** Define $1 = 0^+$, $2 = 1^+ (= 0^{++})$, $3 = 2^+ (= 0^{+++})$, and so on. Thus the symbols $1, 2, 3, 4, \dots$ (called *numerals*) are given meaning by saying that n is the n^{th} iterated successor of 0.

- **Addition.** We can define addition by declaring $m + 0 = m$ and $m + (n^+) = (m + n)^+$. Thus, for instance,

$$m + 1 = m + (0^+) = (m + 0)^+ = m^+$$

and, then

$$m + 2 = m + (1^+) = (m + 1)^+ = m^{++}$$

and so on.

- **Multiplication.** We can define multiplication as iterated addition. Precisely, define $m \times 0 = 0$ and $m \times (n^+) = (m \times n) + m$ (`\times`).

- **Exponentiation.** We can define exponentiation as iterated multiplication. Precisely, define $m^0 = 1$ and $m^{n^+} = (m^n) \times m$.

- **Order.** If you think about it, $m \leq n$ (`\leq`) really just means that there is some non-negative number you can add to m to obtain n . Thus we can define ' $m \leq n$ ' to mean

$$\exists k \in \mathbb{N}, m + k = n$$

and then we can define ' $m < n$ ' to mean $m \leq n \wedge m \neq n$.

The way we defined addition and multiplication is called **recursion**: we defined how they act on zero, and how they act on a successor $n + 1$ in terms of how they act on n .

1.3.2 Exercise. Using the recursive definitions of addition, multiplication and exponentiation, prove that $2^2 = 4$.

It is surprising how rich a structure emerges from something as simple as a set constructed from a single element (zero) and a single operation (successor). We will focus on the arithmetic structure—that is, addition, multiplication and exponentiation—in Chapter 2.

The remainder of this section concerns the order structure of $\mathbb{N}$, which gives rise to proof techniques known as principles of *induction*.

Weak induction

The first induction principle we encounter says that natural numbers behave like dominoes. Imagine an infinitely long line of dominoes—one for each natural number—and suppose we want to prove a statement about natural numbers, say $p(n)$. Proving $p(n)$ will correspond to the n^{th} domino falling; hence proving $p(n)$ for all $n \in \mathbb{N}$ corresponds to *all* the dominoes falling.

How do we make all the dominoes fall? Well we knock down domino 0, and from there everything is taken care of: domino 0 knocks down domino 1; then domino 1 knocks down domino 2; and so on. For $n \in \mathbb{N}$, domino n knocks down domino $n + 1$.

From a more mathematical perspective, what this means is: we prove $p(0)$; then $p(1)$ will follow from the fact that $p(0)$ is true; and $p(2)$ will follow from the fact that $p(1)$ is true; and so on. For $n \in \mathbb{N}$, $p(n + 1)$ will follow from the fact that $p(n)$ is true. In other words, provided we can prove $p(0)$ is true, and that $p(n) \Rightarrow p(n + 1)$ for each n , we've made all the dominos fall over and hence proved the proposition for all natural numbers.

Sometimes a statement might be false for a few natural numbers, but true after a certain point. For example $3n < 2^n$ is true when $n = 0$, false when $n = 1, 2, 3$, and then true for all $n \geq 4$. This isn't a problem—if all we want to do is prove that it is true for $n \geq 4$, we just knock over domino 4 first instead of domino 0!

Now let's be more precise about what we mean, and prove that we're correct.

1.3.3 Theorem (Weak induction principle). Let $p(n)$ be a logical formula concerning natural numbers, and let $b \in \mathbb{N}$. If

- (i) $p(b)$ is true; and
- (ii) $\forall n \geq b, p(n) \Rightarrow p(n + 1)$ is true;

then $\forall n \geq b, p(n)$ is true.

Proof. First suppose $b = 0$. Let $X = \{n \mid p(n)\}$ be the set of all n for which $p(n)$ is true. For a natural number n , the proposition $n \in X$ is equivalent to the proposition $p(n)$. Thus, respectively, the hypotheses of the theorem state:

- (i) $0 \in X$; and
- (ii) $\forall n \in \mathbb{N}, n \in X \Rightarrow n + 1 \in X$;

So by Definition 1.3.1(e), $\mathbb{N} \subseteq X$. Hence $\forall n \in \mathbb{N}, n \in X$, and so $\forall n \in \mathbb{N}, p(n)$.

If $b > 0$, then replace X by the set $X' = \{k \mid p(b + k)\}$; note that $X' = \mathbb{N}$ if and only if $p(n)$ is true for all $n \geq b$. □

PROOF TIP

To prove a statement $p(n)$ is true for all natural numbers $n \geq b$, you can:

- **(Base case)** Prove $p(b)$ is true;
- **(Induction step)** Fix $n \geq b$, and assume that $p(n)$ is true; from this assumption alone, derive $p(n+1)$.

The assumption $p(n)$ is called the **induction hypothesis**.

This whole process is called **proof by (weak) induction (on n)**. We won't usually use the word 'weak' unless we really need to specify it. Usually we'll also omit 'on n ' unless there is more than one variable at play, in which case we will specify.

1.3.4 Example. We will prove that $0 + 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$ for all natural numbers n , by induction.⁷ Note that since we're proving it for *all* natural numbers, our base case has $b = 0$.

Let $p(n)$ be the formula stating that $0 + 1 + \cdots + n = \frac{n(n+1)}{2}$.

- **(BC)** We prove $p(0)$ is true. Now, $p(0)$ is the expression $0 = \frac{0(0+1)}{2}$. Since the right-hand side evaluates to 0, $p(0)$ is true.
- **(IS)** Let $n \in \mathbb{N}$ and suppose $p(n)$ is true, i.e. assume

$$0 + 1 + \cdots + n = \frac{n(n+1)}{2} \quad \text{---(IH)}$$

We prove that this implies $p(n+1)$, which is the formula

$$0 + 1 + \cdots + n + (n+1) = \frac{(n+1)(n+2)}{2}$$

We proceed by calculation:

$$\begin{aligned}
0 + 1 + \cdots + n + (n+1) &= \frac{n(n+1)}{2} + (n+1) && \text{by (IH)} \\
&= (n+1) \left(\frac{n}{2} + 1 \right) && \text{by factorisation} \\
&= (n+1) \left(\frac{n}{2} + \frac{2}{2} \right) && \text{since } \frac{2}{2} = 1 \\
&= \frac{(n+1)(n+2)}{2} && \text{combining fractions}
\end{aligned}$$

⁷The L^AT_EX code for $\frac{a}{b}$ is `\frac{a}{b}`.

Hence $p(n)$ implies $p(n + 1)$. By induction, we're done. $\square$

WRITING TIP

Proofs by induction all follow the same format, so it is good to get into some good habits. These good habits make your proof more readable and better structured, and they help you to avoid silly mistakes. With reference to Example 1.3.4, here are some tips for writing proofs by induction of your own:

- **Labelling the steps.** Clearly labelling the base case and induction step helps the reader identify what part of the proof is being done. I used **BC** and **IS** to signify which is which; you are of course welcome to develop your own convention.
- **Writing down the induction hypothesis.** Writing down the induction hypothesis $p(n)$ explicitly—which I labelled by **IH**—makes it very clear what it is you are assuming. You can then refer back to it later in your proof—as I did in the first line of the calculation—to specify when you have used it.
- **Writing down the goal of the induction step.** When proving the induction step, it is common to fall down the trap of forgetting what you are actually trying to prove. Writing down $p(n + 1)$ explicitly, prefixed by something like ‘we need to prove ...’, gives you something to look back on as you complete your proof.
- **Saying when you're done.** When you have proved $p(n + 1)$ is true, it is a good idea to conclude the proof by summarising what you did. A quick statement like ‘hence $p(n)$ implies $p(n + 1)$, so by induction, we're done’ will suffice.

1.3.5 Example. We'll prove that $n^3 - n$ is divisible by 3 for all $n \in \mathbb{N}$. Thus, the statement $p(n)$ to be proved is $n^3 - n$, and the base case is when $n = 0$.

- **(BC)** We need to prove that $0^3 - 0$ is divisible by 3. Well $0^3 - 0 = 0 = 3 \times 0$, so $0^3 - 0$ is divisible by 3.
- **(IS)** Let $n \in \mathbb{N}$ and suppose that $n^3 - n$ is divisible by 3. Specifically, the induction hypothesis is:

$$n^3 - n = 3k \text{ for some } k \in \mathbb{N} \quad \text{---(IH)}$$

We need to prove that $(n + 1)^3 - (n + 1)$ is divisible by 3; in other words, we need to find some natural number ℓ such that

$$(n + 1)^3 - (n + 1) = 3\ell$$

Expanding the brackets, we obtain:

$$\begin{aligned}
 (n+1)^3 - (n+1) &= (n^3 + 3n^2 + 3n + 1) - n - 1 && \text{expand brackets} \\
 &= n^3 - n + 3n^2 + 3n + 1 - 1 && \text{rearrange terms} \\
 &= n^3 - n + 3n^2 + 3n && \text{since } 1 - 1 = 0 \\
 &= 3k + 3n^2 + 3n && \text{by (IH)} \\
 &= 3(k + n^2 + n) && \text{factorise}
 \end{aligned}$$

Thus we have expressed $(n+1)^3 - (n+1)$ in the form 3ℓ for a natural number ℓ ; specifically, $\ell = k + n^2 + n$. By induction, we're done.

The following example has a non-zero base case.

1.3.6 Example. Let $p(n)$ be the statement $3n < 2^n$. We prove $p(n)$ is true for all $n \geq 4$ by induction.

- **(BC)** $p(4)$ is the statement $3 \cdot 4 < 2^4$. This is true, since $12 < 16$.
- **(IS)** Suppose $n \geq 4$ and that $p(n)$ is true, i.e. that $3n < 2^n$ **(IH)**. We want to prove $3(n+1) < 2^{n+1}$. Well

$$\begin{aligned}
 3(n+1) &= 3n + 3 && \text{expand brackets} \\
 &< 2^n + 3 && \text{by (IH)} \\
 &< 2^n + 16 && \text{since } 3 < 16 \\
 &= 2^n + 2^4 && \text{since } 2^4 = 16 \\
 &\leq 2^n + 2^n && \text{since } n \geq 4 \\
 &= 2 \cdot 2^n && \text{since } x + x = 2x \\
 &= 2^{n+1} && \text{using laws of indices}
 \end{aligned}$$

So we have proved $3(n+1) < 2^{n+1}$, as required.

Hence $p(n)$ implies $p(n+1)$, so by induction, we're done. □

Note that the proof in Example 1.3.6 says nothing about the truth or falsity of $p(n)$ for $n = 0, 1, 2, 3$. In order to assert that these cases are false, you need to show them individually; indeed:

- $3 \times 0 = 0$ and $2^0 = 1$, hence $p(0)$ is true;

- $3 \times 1 = 3$ and $2^1 = 2$, hence $p(1)$ is false;
- $3 \times 2 = 6$ and $2^2 = 4$, hence $p(2)$ is false;
- $3 \times 3 = 9$ and $2^3 = 8$, hence $p(3)$ is false.

So we deduce that $p(n)$ is true when $n = 0$ or $n \geq 4$, and false otherwise.

Sometimes a ‘proof’ by induction might appear to be complete nonsense. The following is a classic example of a ‘fail by induction’:

1.3.7 Example. The following argument supposedly proves that every horse is the same colour.

- **(BC)** Suppose there is just one horse. This horse is the same colour as itself, so the base case is immediate.
- **(IS)** Suppose that every collection of n horses is the same colour **(IH)**. Let X be a set of $n + 1$ horses. Removing the first horse from X , we see that the last n horses are the same colour by **(IH)**. Removing the last horse from X , we see that the first n horses are the same colour. Hence all the horses in X are the same colour.

By induction, we’re done. □

1.3.8 Exercise. Write down the statement $p(n)$ that Example 1.3.7 attempted to prove for all $n \geq 1$. Convince yourself that the proof of the base case is correct, then write down—with quantifiers—exactly the proposition that the induction step is meant to prove. Explain why the argument in the induction step failed to prove this proposition.

WRITING TIP

There are several ways to avoid situations like that of Example 1.3.7 by simply putting more thought into writing the proof. Some tips are:

- **State $p(n)$ explicitly.** In the statement ‘all horses are the same colour’ it is not clear exactly what the induction variable is. However, we could have said:

Let $p(n)$ be the statement ‘every set of n horses has the same colour’.

- **Refer to the base case b in the induction step.** In Example 1.3.7, our induction hypothesis simply stated ‘assume every set of n horses has the same colour’. Had we instead said:

Let $n \geq 1$ and assume every set of n horses has the same colour.

We may have spotted the error in what was to come.

What follows are a couple more examples of proofs by weak induction.

1.3.9 Example. Given any $n \in \mathbb{N}$,

$$\sum_{k=0}^n k^3 = \left(\sum_{k=0}^n k \right)^2$$

We proved in Example 1.3.4 that $\sum_{k=0}^n k = \frac{n(n+1)}{2}$ for all $n \in \mathbb{N}$, thus it suffices to prove that

$$\sum_{k=0}^n k^3 = \frac{n^2(n+1)^2}{4}$$

for all $n \in \mathbb{N}$.

We proceed by induction.

- **(BC)** We need to prove that $0^3 = \frac{0^2(0+1)^2}{4}$. This is true since both sides of the equation are equal to 0.
- **(IS)** Fix $n \in \mathbb{N}$ and suppose that $\sum_{k=0}^n k^3 = \frac{n^2(n+1)^2}{4}$. We need to prove that $\sum_{k=0}^{n+1} k^3 = \frac{(n+1)^2(n+2)^2}{4}$. This is true since:

$$\begin{aligned} \sum_{i=0}^{n+1} k^3 &= \sum_{i=0}^n k^3 + (n+1)^3 && \text{by definition of sum} \\ &= \frac{n^2(n+1)^2}{4} + (n+1)^3 && \text{by (IH)} \\ &= \frac{n^2(n+1)^2 + 4(n+1)^3}{4} && (\text{algebra}) \\ &= \frac{(n+1)^2(n^2 + 4(n+1))}{4} && (\text{algebra}) \\ &= \frac{(n+1)^2(n+2)^2}{4} && (\text{algebra}) \end{aligned}$$

By induction, the result follows. □

1.3.10 Example. We will prove the correctness of the following formula for the sum of an *arithmetic progression*, that is a sequence of finite length such that the difference between consecutive terms is constant.

Specifically, let $a, d \in \mathbb{R}$. We will prove that

$$\sum_{k=0}^n (a + kd) = \frac{(n+1)(2a + nd)}{2}$$

for all $n \in \mathbb{N}$.

We proceed by induction.

- **(BC)** We need to prove that $\sum_{k=0}^0 (a + kd) = \frac{(0+1)(2a+0d)}{2}$. This is true, since

$$\sum_{k=0}^0 (a + kd) = a + 0d = a = \frac{2a}{2} = \frac{1 \cdot (2a)}{2} = \frac{(0+1)(2a+0d)}{2}$$

- **(IS)** Fix $n \in \mathbb{N}$ and suppose that $\sum_{k=0}^n (a + kd) = \frac{(n+1)(2a+nd)}{2}$. We need to prove:

$$\sum_{k=0}^{n+1} (a + kd) = \frac{(n+2)(2a + (n+1)d)}{2}$$

This is true, since

$$\begin{aligned} \sum_{k=0}^{n+1} (a + kd) &= \sum_{k=0}^n (a + kd) + (a + (n+1)d) && \text{by definition of sum} \\ &= \frac{(n+1)(2a + nd)}{2} + (a + (n+1)d) && \text{by (IH)} \\ &= \frac{(n+1)(2a + nd) + 2a + 2(n+1)d}{2} && (\text{algebra}) \\ &= \frac{(n+1) \cdot 2a + (n+1) \cdot nd + 2a + 2(n+1)d}{2} && (\text{algebra}) \\ &= \frac{2a(n+1+1) + (n+1)(nd + 2d)}{2} && (\text{algebra}) \\ &= \frac{2a(n+2) + (n+1)(n+2)d}{2} && (\text{algebra}) \\ &= \frac{(n+2)(2a + (n+1)d)}{2} && (\text{algebra}) \end{aligned}$$

By induction, the result follows. □

Strong induction

Sometimes it is clear that a statement can *almost* be proved by induction, but a snag appears; for example, in the following example, the truth of $p(n+1)$ seems to depend not on just $p(n)$, but also on $p(n-1)$:

1.3.11 Example. Define a sequence of numbers $(a_n)_{n \in \mathbb{N}}$ recursively by:

$$a_0 = 0, \quad a_1 = 1, \quad a_n = 3a_{n-1} - 2a_{n-2} \text{ for all } n \geq 2$$

Thus, continuing the sequence, we have

$$a_2 = 3 \cdot 1 - 2 \cdot 0 = 3, \quad a_3 = 3 \cdot 3 - 2 \cdot 1 = 7, \quad a^4 = 15, \quad \dots$$

Looking at the sequence $(0, 1, 3, 7, 15, \dots)$, you might hypothesise that $a_n = 2^n - 1$ for all $n \in \mathbb{N}$. And you would be correct! So let's try and prove that $a_n = 2^n - 1$ for all $n \in \mathbb{N}$ by induction.

The statement is demonstrably true for $n = 0, 1$, since

$$a_0 = 0 = 1 - 1 = 2^0 - 1 \quad \text{and} \quad a_1 = 1 = 2 - 1 = 2^1 - 1$$

Fix $n \geq 1$ and suppose $a_n = 2^n - 1$. If this implies that $a_{n+1} = 2^{n+1} - 1$, we'll be done by induction: indeed, induction gives that $p(n)$ is true for all $n \geq 1$, and we checked the case $n = 0$ separately.

So let's see what happens. Since $n \geq 1$, we have $n+1 \geq 2$, so we can apply the recursive formula for a_{n+1} :

$$a_{n+1} = 3a_n - 2a_{n-1}$$

Here's where we get stuck: our induction hypothesis only tells us that $a_n = 2^n - 1$, so that

$$a_{n+1} = 3(2^n - 1) - 2a_{n-1}$$

but it doesn't tell us anything at all about a_{n-1} . We need to express a_{n-1} in terms of n in order to get a reasonable formula for a_{n+1} . □

This example illustrates why weak induction is called 'weak'. But all is not lost: using the technique of weak induction, we can prove a principle of *strong induction*. The induction step in strong induction assumes not just the truth of the proposition for *one* prior step, but its truth of *all* prior steps.

Despite its name, strong induction is no stronger than weak induction; the two principles are equivalent. In fact, we'll prove the strong induction principle *by weak induction*!

1.3.12 Corollary (Strong induction principle). Let $p(x)$ be a logical formula concerning natural numbers and let $b \in \mathbb{N}$. If

(i) $p(b)$ is true; and

(ii) $\forall n \in \mathbb{N}, (\forall k \in \mathbb{N}, b \leq k \leq n \Rightarrow p(k)) \Rightarrow p(n+1)$;

then $\forall n \geq b, p(n)$ is true.

Proof. We'll prove this using weak induction. For each n , let $q(n)$ be the statement $\forall k \in \mathbb{N}, b \leq k \leq n \Rightarrow p(k)$. Notice that, for all $n \geq b$, $q(n)$ implies $p(n)$ by setting $k = n$. Thus if we can prove that $q(n)$ is true for all n , then we've proved that $p(n)$ is true for all n .

- **(BC)** $q(b)$ is equivalent to $p(b)$, since the only natural number k with $b \leq k \leq b$ is b itself; hence $q(b)$ is true by (i);
- **(IS)** Let $n \geq b$ and suppose $q(n)$ is true. By (ii), $p(n+1)$ is true. Since $q(n)$ is true, $p(k)$ is true for all $b \leq k \leq n$. Combining these facts, $p(k)$ is true for all $b \leq k \leq n+1$, which is precisely the statement that $q(n+1)$ is true.

By induction, $q(n)$ is true for all $n \geq b$. Hence $p(n)$ is true for all $n \geq b$. □

PROOF TIP

To prove a statement $p(n)$ is true for all natural numbers $n \geq b$, you can:

- **(Base case)** Prove $p(b)$ is true;
- **(Induction step)** Fix $n \geq b$, and assume that $p(k)$ is true for all $b \leq k \leq n$; from this assumption alone, derive $p(n+1)$.

The assumption that $p(k)$ is true for all $b \leq k \leq n$ is called the **induction hypothesis**.

This whole process is called **proof by (strong) induction (on n)**. We won't usually use the word 'strong' unless we really need to specify it. Usually we'll also omit 'on n ' unless there is more than one variable at play, in which case we will specify.

Strong induction is very well suited to proving formulae for sequences where subsequent terms are defined in terms of more than one previous term, as the next few examples demonstrate.

1.3.13 Example. Recall from Example 1.3.11 that we defined the sequence

$$a_0 = 0, \quad a_1 = 1, \quad a_n = 3a_{n-1} - 2a_{n-2} \text{ for all } n \geq 2$$

and we wished to prove that $a_n = 2^n - 1$ for all $n \in \mathbb{N}$. We have proved that it's true when $n = 0$, and will show that it's true for $n \geq 1$ by strong induction on n .

- **(BC)** We have already proved that $a_1 = 2^1 - 1$.
- **(IS)** Let $n \in \mathbb{N}$, and assume that $a_k = 2^k - 1$ for all $1 \leq k \leq n$. Since $a_0 = 2^0 - 1$, this in fact holds for all $k \leq n$.

We need to prove that this assumption implies that $a_{n+1} = 2^{n+1} - 1$. Well, $n \geq 1$, so $n + 1 \geq 2$ and we can apply the recursive formula to a_{n+1} . Thus

$$\begin{aligned}
 a_{n+1} &= 3a_n - 2a_{n-1} && \text{by definition of } a_{n+1} \\
 &= 3(2^n - 1) - 2(2^{n-1} - 1) && \text{since } p(k) \text{ holds for all } k \leq n \\
 &= 3 \cdot 2^n - 3 - 2 \cdot 2^{n-1} + 2 && \text{expand brackets} \\
 &= 3 \cdot 2^n - 3 - 2^n + 2 && \text{laws of indices} \\
 &= 2 \cdot 2^n - 1 && \text{simplifying} \\
 &= 2^{n+1} - 1 && \text{laws of indices}
 \end{aligned}$$

So we're done by strong induction. □

1.3.14 Example. Define a sequence recursively by $a_0 = 4$, $a_1 = 9$ and $a_n = 5a_{n-1} - 6a_{n-2}$ for all $n \geq 2$.

We will prove that $a_n = 3 \cdot 2^n + 3^n$ for all $n \in \mathbb{N}$.

We proceed by strong induction for $n \geq 1$, treating the $n = 0$ case as a second base case.

- **(BC)** The result holds when $n = 0$ and when $n = 1$, since

$$a_0 = 4 = 3 \cdot 2^0 + 3^0 \quad \text{and} \quad a_1 = 9 = 3 \cdot 2^1 + 3^1$$

- **(IS)** Fix $n \geq 1$ and suppose that $a_k = 3 \cdot 2^k + 3^k$ for all $k \leq n$. We need to prove that $a_{n+1} = 3 \cdot 2^{n+1} + 3^{n+1}$. Well,

$$\begin{aligned}
 a_{n+1} &= 5a_n - 6a_{n-1} && \text{by definition of the sequence} \\
 &= 5(3 \cdot 2^n + 3^n) - 6(3 \cdot 2^{n-1} + 3^{n-1}) && \text{by the induction hypothesis} \quad (\text{algebra}) \\
 &= (5 \cdot 3 \cdot 2 - 6 \cdot 1)2^{n-1} + (5 \cdot 3 - 6)3^{n-1} && (\text{algebra}) \\
 &= 24 \cdot 2^{n-1} + 9 \cdot 3^{n-1} && (\text{algebra}) \\
 &= 3 \cdot 2^2 \cdot 2^{n-1} + 3^2 \cdot 3^{n-1} && (\text{algebra}) \\
 &= 3 \cdot 2^{n+1} + 3^{n+1} && (\text{algebra})
 \end{aligned}$$

Hence the result we sought to prove is true.

By induction, it follows that $a_n = 3 \cdot 2^n + 3^n$ for all $n \in \mathbb{N}$. □

1.3.15 Example. Define a sequence recursively by

$$b_0 = 1 \quad \text{and} \quad b_{n+1} = 1 + \sum_{k=0}^n b_k \quad \text{for all } n \in \mathbb{N}$$

We will prove by strong induction that $b_n = 2^n$ for all $n \in \mathbb{N}$.

- **(BC)** By definition of the sequence we have $b_0 = 1 = 2^0$.
- **(IS)** Fix $n \in \mathbb{N}$, and suppose that $b_k = 2^k$ for all $k \leq n$. We need to show that $b_{n+1} = 2^{n+1}$. This is true, since

$$\begin{aligned} b_{n+1} &= 1 + \sum_{k=0}^n b_k && \text{by the recursive formula for } b_{n+1} \\ &= 1 + \sum_{k=0}^n 2^k && \text{by the induction hypothesis} \\ &= 1 + (2^{n+1} - 1) && \text{by the lemma} \\ &= 2^{n+1} \end{aligned}$$

By induction, it follows that $b_n = 2^n$ for all $n \in \mathbb{N}$. □

Well-ordering principle

In a way that we will make precise in Section 5.2, the underlying reason why we can perform induction and recursion on the natural numbers is because of the way they are ordered. Specifically, the natural numbers satisfy the *well-ordering principle*: every inhabited set of natural numbers has a least element. This sets the natural numbers apart from the other number sets; for example, $\mathbb{Z}$ has no least element, since if $a \in \mathbb{Z}$ then $a - 1 \in \mathbb{Z}$ and $a - 1 < a$.

1.3.16 Theorem (Well-ordering principle). Let X be a set of natural numbers. If X is inhabited, then X has a least element.

Strategy. Under the assumption that X is a set of natural numbers, the proposition we want to prove has the form $p \Rightarrow q$ (see [A.1.5](#)). Namely

$$X \text{ is inhabited} \quad \Rightarrow \quad X \text{ has a least element}$$

Assuming X is inhabited doesn't really give us much to work with, so let's try the contrapositive:

$$X \text{ has no least element} \Rightarrow X \text{ is empty}$$

The assumption that X has no least element *does* give us something to work with. Under this assumption we need to deduce that X is empty.

We will do this by 'forcing X up' by strong induction. Certainly $0 \notin X$, otherwise it would be the least element. If none of the numbers $0, 1, \dots, n$ are elements of X then neither can $n + 1$ be, since if it were then *it* would be the least element of X . Let's make this argument formal.

Proof. Let X be a set of natural numbers containing no least element. We prove $\forall n \in \mathbb{N}, n \notin X$ by strong induction.

- **(BC)** $0 \notin X$ since if $0 \in X$ then 0 is the least element of X .
- **(IH)** Suppose $k \notin X$ for all $0 \leq k \leq n$. If $n + 1 \in X$ then $n + 1$ is the least element of X ; indeed, if $\ell < n + 1$ then $0 \leq \ell \leq n$, so $\ell \notin X$ by the induction hypothesis. This contradicts the assumption that X has no least element, so $n + 1 \notin X$.

By strong induction, $n \notin X$ for each $n \in \mathbb{N}$. Since $X \subseteq \mathbb{N}$, it must be the case that $X = \emptyset$. $\square$

ASIDE

In Section 5.2 we will encounter more general sets with a notion of 'less than', for which any inhabited subset has a 'least' element. Any such set has an induction principle, the proof of which is more or less identical to the proof of Corollary 1.3.12. This has powerful applications in computer science, where it can be used to formally verify that a computer program containing various loops will terminate: termination of a program corresponds to a particular set having a 'least' element.

The following proof that $\sqrt{2}$ is irrational is a classic application of the well-ordering principle.

1.3.17 Proposition. The number $\sqrt{2}$ is irrational.

Strategy. Let p be the proposition. We will prove p by contradiction: assuming $\neg p$ —that is, assuming that $\sqrt{2}$ is rational—we derive something which is known to be false. In particular, we will construct an inhabited set of natural numbers with no least element, which we know cannot exist by the well-ordering principle. This proof technique is called *infinite descent*.

To prove Proposition 1.3.17 we will use the following lemma.

1.3.18 Lemma. Let $a \in \mathbb{Z}$. If a^2 is even then a is even.

Proof. We prove the contrapositive; that is, we prove that if a is odd then a^2 is odd.

Odd numbers are precisely those of the form $2k + 1$, where $k \in \mathbb{Z}$. So suppose $a = 2k + 1$ for some $k \in \mathbb{Z}$. Then

$$a^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$$

Letting $\ell = 2k^2 + 2k$ we see that $a^2 = 2\ell + 1$, and since $\ell \in \mathbb{Z}$, it follows that a^2 is odd.

By contraposition, if a^2 is even then a is even. □

Proof of Proposition 1.3.17. Suppose $\sqrt{2}$ is rational. This means that we can write

$$\sqrt{2} = \frac{a}{b}$$

where $a, b \in \mathbb{Z}$ and $b \neq 0$. Since $\sqrt{2} > 0$, we can choose a, b such that $a > 0$, so that $a \in \mathbb{N}$. (Indeed, if $a < 0$ then replace a by $-a$ and b by $-b$.)

We construct sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$, such that $\sqrt{2} = \frac{a_n}{b_n}$ and $a_0 > a_1 > a_2 > \dots$.

- Let $a_0 = a$ and $b_0 = b$.
- Suppose a_n, b_n have been defined and that $\sqrt{2} = \frac{a_n}{b_n}$ and $a_n > 0$. Then squaring both sides and rearranging gives

$$2 = \frac{a_n^2}{b_n^2} \Rightarrow 2b_n^2 = a_n^2$$

Therefore a_n^2 is even. By Lemma 1.3.18, a_n is even, so $a_n = 2k$ for some $k \in \mathbb{Z}$. Moreover, $k > 0$ since $a_n > 0$; so let $a_{n+1} = k$.

But then $2b_n^2 = a_n^2 = 4a_{n+1}^2$. Dividing by 2 gives $b_n^2 = 2a_{n+1}^2$, so that b_n is also even by Lemma 1.3.18. Thus, as before, we can define b_{n+1} by $b_n = 2b_{n+1}$.

Note also that $a_n > a_{n+1}$ and

$$\sqrt{2} = \frac{a_n}{b_n} = \frac{2a_n}{2b_n} = \frac{a_{n+1}}{b_{n+1}}$$

The set $X = \{a_n \mid n \in \mathbb{N}\}$ is nonempty and has no least element: indeed, if $x \in X$ then $x = a_n$ for some $n \in \mathbb{N}$, and then $a_{n+1} < x$ by construction of the sequence.

This contradicts the well-ordering principle. Thus, our assumption that $\sqrt{2}$ is irrational must have been false. □

WRITING TIP

In the proof of Proposition 1.3.17 we could have separately proven that a_n^2 being even implies a_n is even, and that b_n^2 being even implies b_n is even. Since this was an application of a more general result, we split it off into a separate lemma.

Proving auxiliary results (lemmas) separately and then applying them in theorems can improve the readability of the main proof, particularly when the auxiliary results are particularly technical. Doing so also helps emphasise the important steps.

1.3.19 Exercise. What goes wrong in the proof of Proposition 1.3.17 if we try instead to prove that $\sqrt{4}$ is irrational?

1.3.20 Exercise. Prove that $\sqrt{3}$ is irrational. **Hint:** Prove first that if $a \in \mathbb{Z}$ and a^2 is divisible by 3, then a is divisible by 3. You may use the fact that if $a \in \mathbb{Z}$ then a is either equal to, one more than, or one less than, a multiple of 3.

1.4 Problems

Chapter 2

Number theory

The focus of Section 1.3 was overwhelmingly on the natural numbers. Bringing negative numbers into the picture gives rise to the *integers*:

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

The integers come with a natural *arithmetic structure*, meaning that we can add, subtract and multiply them. Division of integers is not always possible—for example, the result of dividing 3 by 2 is not an integer. This means that there are interesting results to be discovered as far as division is concerned: when it is possible? Do some numbers behave more nicely under division than others? What can we say about common divisors? How much can we tell about a number knowing only its remainder upon division by some fixed integer?

There are many sets other than $\mathbb{Z}$ which are equipped with a notion of arithmetic—that is, the ability to in some sense add, subtract and multiply elements of the set together. Such ‘sets-with-arithmetic’ are called *rings*; they will be studied briefly in Section 5.1, where we will see what similarities and differences they have with the ring of integers. Although our attention in this section is focussed on $\mathbb{Z}$ alone, we will be guided by more general ring theory. In particular, this will require us to review what we really mean by the notion of a *prime number* and a *greatest common divisor*.

Since we are now dealing with both positive and negative numbers, some new notation will be useful:

2.0.1 Definition. Let $a \in \mathbb{Z}$. The **absolute value** of a , denoted $|a|$, is the natural number which can be multiplied by 1 Or -1 to give a .

More concisely, we can define

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a \leq 0 \end{cases}$$

2.0.2 Example. Some absolute values include:

$$|3| = 3, \quad |0| = 0, \quad |-17| = 17$$

2.0.3 Exercise. Show that if $a, b \in \mathbb{Z}$ then $|a - b| = |b - a|$.

2.1 Division

This section introduces the notion of *divisibility*. As we have already mentioned, it is not always the case that one integer can divide another. As you read through this section, note that we never use fractions; everything we do is *internal* to $\mathbb{Z}$, and does not require that we ‘spill over’ to $\mathbb{Q}$ at any point. This will help us when we come to study rings in Section 5.1, and is a good practice to mimic in your own work.

The following theorem, called the division theorem, is the crux of everything that is to follow.

2.1.1 Theorem (Division theorem). Let $a, b \in \mathbb{Z}$ with $b \neq 0$. There exist unique $q, r \in \mathbb{Z}$ such that

$$a = qb + r \quad \text{and} \quad 0 \leq r < |b|$$

Strategy. Let’s look at the simple case when $a \geq 0$ and $b > 0$. We can always find q, r such that $a = qb + r$, for example $q = 0$ and $r = a$. Moreover, by increasing q we can reduce r , since

$$qb + r = (q + 1)b + (r - b)$$

We will keep doing this until the ‘remainder’ is as small as it can be without being negative. As an example, consider the case when $a = 14$ and $b = 5$. This procedure gives

$$\begin{aligned} 14 &= 0 \times 5 + 14 \\ &= 1 \times 5 + 9 \\ &= 2 \times 5 + 4 && \leftarrow \text{least non-negative remainder} \\ &= 3 \times 5 + (-1) \\ &= \dots \end{aligned}$$

This procedure shows that in this case we should take $q = 2$ and $r = 4$, since $14 = 2 \times 5 + 4$ and $0 \leq 4 < |5|$.

We can show that such a descending sequence of remainders terminates using the well-ordering principle, and then we must argue that the quotient and remainder that we obtain are unique.

Proof. We may assume that $b > 0$: if not, replace b by $-b$ and q by $-q$. We may also assume that $a \leq 0$. Otherwise, replace a by $-a$, q by $-(q + 1)$ and r by $b - r$.

Thus, what follows assumes that $a \geq 0$ and $b > 0$.

- **Existence.** We prove that such integers q, r exist by the well-ordering principle. Namely, we define a sequence $(r_n)_{n \in \mathbb{N}}$ such that $a = nb + r_n$ and $r_0 > r_1 > r_2 > \dots$, and use this sequence to find the values of q, r .

- ◇ Let $r_0 = a$. Then $a = 0b + r_0$, as required.
- ◇ Suppose r_n has been defined, and let $r_{n+1} = r_n - b$. Then

$$(n+1)b + r_{n+1} = (n+1)b + r_n - b = nb + b + r_n - b = nb + r = a$$

Since $b > 0$, we must have $r_{n+1} < r_n$ for all n .

Let $R = \mathbb{N} \cap \{r_n \mid n \in \mathbb{N}\}$. That is, R is the set of terms of the sequence which are non-negative. Since $r_0 = a \geq 0$, we have that $r_0 \in R$ and hence R is inhabited. By the well-ordering principle, R has a least element r_k for some $k \in \mathbb{N}$.

Define $q = k$ and $r = r_k$. By construction we have $a = qb + r$ and $r \geq 0$, so it remains to show that $r < b$. Well, if $r \geq b$ then $r - b \geq 0$, but $r - b = r_{k+1}$, so this would imply $r_{k+1} \in R$, contradicting minimality of r . Hence $r < b$, so q, r are as required.

- **Uniqueness.** Suppose q', r' also satisfy $a = q'b + r'$ and $0 \leq r' < b$. If we can show that $r' = r$ then this proves that $q = q'$: indeed, if $qb + r = q'b + r'$ then we can subtract r and then divide by b , since $b > 0$.

First note that $q' \geq 0$. If $q' < 0$ then $q' \leq -1$, so

$$a = q'b + r' \leq -b + r'$$

and hence $r' \geq a + b \geq b$ since $a \geq 0$. This contradicts the assumption that $r' < b$. So $q' \geq 0$.

Since $q' \geq 0$, we also know that $a = q'b + r_{q'}$, and hence $r' = r_{q'} \in R$. By minimality of r we have $r \leq r'$. It remains to show that $r = r'$. If not then $r < r'$. Thus

$$qb + r = q'b + r' > q'b + r \quad \Rightarrow \quad qb > q'b \quad \Rightarrow \quad q > q'$$

and hence $q = q' + t$ for some $t \geq 1$. But then

$$q'b + r' = a = qb + r = (q' + t)b + r = q'b + (tb + r)$$

so $r' = tb + r \geq b$, contradicting $r' < b$. So $r = r'$ as desired, and hence $q = q'$.

At long last, we are done. □

2.1.2 Definition. Let $a, b \in \mathbb{Z}$ with $b \neq 0$, and let q, r be the unique integers such that

$$a = qb + r \quad \text{and} \quad 0 \leq r < |b|$$

We say q is the **quotient** and r is the **remainder** of a divided by b .

2.1.3 Example. Some examples of division include:

$$14 = 2 \times 5 + 4, \quad -14 = -3 \times 5 + 1, \quad 15 = 3 \times 5 + 0$$

2.1.4 Definition. Let $a, b \in \mathbb{Z}$. We say b **divides** a if there exists $q \in \mathbb{Z}$ such that $a = qb$. To denote the fact that b divides a we write $b \mid a$ ([L^AT_EX code: \mid](#)). For the negation $\neg(b \mid a)$ write $b \nmid a$ ([L^AT_EX code: \nmid](#)).

Thus, when $b \neq 0$, saying $b \mid a$ is equivalent to saying that the remainder of a divided by b is 0.

2.1.5 Example. 5 divides 15 since $15 = 3 \times 5$. However, 5 does not divide 14: we know that the remainder of 14 divided by 5 is 4, not 0—and it can't be both since we proved in the division theorem that remainders are unique!

2.1.6 Exercise. Show that if $a \in \mathbb{Z}$ then $1 \mid a$, $-1 \mid a$ and $a \mid 0$. For which integers a does $a \mid 1$? For which integers a does $0 \mid a$?

We now introduce the very basic notion of a *unit*. This notion is introduced to rule out trivialities. Units become interesting when talking about general rings, but in $\mathbb{Z}$, the units are very familiar.

2.1.7 Definition. Let $u \in \mathbb{Z}$. We say u is a **unit** if $u \mid 1$; that is, u is a unit if there exists $v \in \mathbb{Z}$ such that $uv = 1$.

2.1.8 Proposition. The only units in $\mathbb{Z}$ are 1 and -1 .

Proof. First note that 1 and -1 are units, since $1 \cdot 1 = 1$ and $(-1) \cdot (-1) = 1$. Now suppose that $u \in \mathbb{Z}$ is a unit, and let $v \in \mathbb{Z}$ be such that $uv = 1$. Certainly $u \neq 0$, since $0v = 0 \neq 1$. If $u > 1$ or $u < -1$ then $v = \frac{1}{u} \notin \mathbb{Z}$. So we must have $u \in \{-1, 1\}$. $\square$

Exercise 2.1.6 shows that -1 , 0 and 1 are, from the point of view of divisibility, fairly trivial. For this reason, most of the results we discuss regarding divisibility will concern **non-zero non-units**, i.e. all integers except -1 , 0 or 1.

Greatest common divisors

2.1.9 Definition. Let $a, b \in \mathbb{Z}$. An integer d is a **greatest common divisor** of a and b if:

- (a) $d \mid a$ and $d \mid b$;
- (b) If q is another integer such that $q \mid a$ and $q \mid b$, then $q \mid d$.

2.1.10 Example. 2 is a greatest common divisor of 4 and 6; indeed:

- (a) $4 = 2 \times 2$, and $6 = 3 \times 2$, so $2 \mid 4$ and $2 \mid 6$;

- (b) Suppose $q \mid 4$ and $q \mid 6$. The divisors of 4 are $\pm 1, \pm 2, \pm 4$ and the divisors of 6 are $\pm 1, \pm 2, \pm 3, \pm 6$. Since q divides both, it must be the case that $q \in \{-2, -1, 1, 2\}$; in any case, $q \mid 2$.

Likewise, -2 is a greatest common divisor of 4 and 6.

2.1.11 Exercise. There are two greatest common divisors of 6 and 15; find both.

2.1.12 Proposition. Let $a, b \in \mathbb{Z}$. If d and d' are two greatest common divisors of a and b , then either $d = d'$ or $d = -d'$.

2.1.13 Exercise. Prove Proposition 2.1.12.

2.1.14 Definition. Let $a, b \in \mathbb{Z}$. Denote by $\gcd(a, b)$ ([L^AT_EX code: `\mathrm{gcd}\{a,b\}`](#)) the (unique!) non-negative greatest common divisor of a and b .

Our goal for the rest of this subsection is to investigate the behaviour of greatest common divisors, find out how to compute them, and look into the implications they have for solutions to certain kinds of equations.

2.1.15 Theorem. Let $a, b, q, r \in \mathbb{Z}$, and suppose that $a = qb + r$. Then

$$\gcd(a, b) = \gcd(b, r)$$

Proof. Let $d = \gcd(a, b)$. We check that d satisfies the conditions required to be a greatest common divisor of b and r .

Note that $d \mid a$ and $d \mid b$; write $a = sd$ and $b = td$.

- $d \mid b$ by definition, and $d \mid r$ since

$$r = a - qb = sd - qtd = (s - qt)d$$

- Suppose $d' \mid b$ and $d' \mid r$; say $b = ud'$ and $r = vd'$. Then $d' \mid a$, since

$$a = qb + r = qud' + vd' = (qu + v)d'$$

so $d \mid d'$ since $d = \gcd(a, b)$.

So d is a greatest common divisor of b and r . Since $d > 0$, the result is shown. $\square$

Combined with the division theorem (Theorem 2.1.1), this gives a relatively fast algorithm for computing the greatest common divisor of two integers, known as the **Euclidean algorithm**.

PROOF TIP

Euclidean algorithm. Let $a, b \in \mathbb{Z}$. To find $\gcd(a, b)$, proceed as follows.

- Set $r_0 = a$ and $r_1 = b$.
- Given r_{n-2} and r_{n-1} , define r_n to be the remainder of r_{n-2} divided by r_{n-1} .
- Stop when $r_n = 0$; then $r_{n-1} = \gcd(a, b)$.

2.1.16 Example. We will find the greatest common divisor of 148 and 28.

$$148 = 5 \times 28 + 8$$

$$28 = 3 \times 8 + 4$$

$$8 = 2 \times \boxed{4} + 0 \quad \leftarrow \text{Stop!}$$

Hence $\gcd(148, 28) = 4$. Here the sequence of remainders is given by:

$$r_0 = 148, \quad r_1 = 28, \quad r_2 = 8, \quad r_3 = 4, \quad r_4 = 0$$

2.1.17 Example. The Euclidean algorithm works surprisingly quickly, even for relatively large numbers. Consider the problem of computing $\gcd(1311, 5757)$ for example:

$$5757 = 4 \times 1311 + 513$$

$$1311 = 2 \times 513 + 285$$

$$513 = 1 \times 285 + 228$$

$$285 = 1 \times 228 + 57$$

$$228 = 4 \times \boxed{57} + 0 \quad \leftarrow \text{Stop!}$$

Hence $\gcd(1311, 5757) = 57$. Here the sequence of remainders is given by:

$$r_0 = 5757, \quad r_1 = 1311, \quad r_2 = 513, \quad r_3 = 285, \quad r_4 = 228, \quad r_5 = 57, \quad r_6 = 0$$

2.1.18 Example. Here's an example where one of the numbers is negative: we compute the value of $\gcd(-420, 76)$:

$$-420 = (-6) \times 76 + 36$$

$$76 = 2 \times 36 + 4$$

$$36 = 9 \times \boxed{4} + 0$$

Hence $\gcd(-420, 76) = 4$.

2.1.19 Exercise. Use the Euclidean algorithm to compute the greatest common divisors of the following pairs of integers

$$(12, 9), \quad (100, 35), \quad (7125, 1300), \quad (1010, 101010), \quad (-4, 14)$$

The following theorem will be useful when we study modular arithmetic in Section 2.3; it is called a ‘lemma’ for historical reasons, and is really an important result in its own right.

2.1.20 Theorem (Bézout’s lemma). Let $a, b, c \in \mathbb{Z}$, and let $d = \gcd(a, b)$. The equation $ax + by = c$ has a solution $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ if and only if $d \mid c$.

Proof. ($\Rightarrow$) Write $a = a'd$ and $b = b'd$. If there exist $x, y \in \mathbb{Z}$ such that $ax + by = z$, then

$$c = ax + by = a'dx + b'dy = (a'x + b'y)d$$

so $d \mid c$.

($\Leftarrow$) Suppose that $d \mid c$. If $d = 0$ then $a = b = c = 0$, and any pair of integers gives a solution to the equation; so suppose $d > 0$. In particular, this implies that $a \neq 0$ or $b \neq 0$. Define

$$X = \{au + bv \mid u, v \in \mathbb{Z}, au + bv > 0\}$$

That is, X is the set of positive integers of the form $au + bv$.

X is inhabited, since $a, b \in \mathbb{Z}$ and $aa + bb = a^2 + b^2 > 0$. By the well-ordering principle, X has a least element ℓ , and there exist $u, v \in \mathbb{Z}$ such that $au + bv = \ell$.

We’ll now prove that $\ell = d$. Now $d \mid \ell$ since $\ell = au + bv$ and $d \mid au + bv$; and therefore $d \leq \ell$.

It remains to prove $d \geq \ell$. By the division theorem (Theorem 2.1.1), there exist q, r such that $a = q\ell + r$, where $0 \leq r < \ell$. But then

$$r = a - q\ell = a - q(au + bv) = a(1 - qu) + b(-qv)$$

If we had $r > 0$ then we’d have $r \in X$; but $r < \ell$ so this would contradict minimality of ℓ in X . Hence $r = 0$ and $\ell \mid a$. The same argument shows that $\ell \mid b$, hence $\ell \mid d$ since $d = \gcd(a, b)$. Hence $d \geq \ell$, and so $d = \ell$, as desired.

Finally, since $d \mid c$, we can write $c = c'd$ for some $c' \in \mathbb{Z}$. Multiplying the equation $au + bv = \ell$ by c' gives

$$ac'u + bc'v + c'\ell = c'd = c$$

So $(x, y) = (c'u, c'v)$ is a solution to the equation, and we’re done. $\square$

2.1.21 Example. Here are some examples of applications of Bézout’s lemma.

- Consider the equation $1311x + 5757y = 12963$. We computed in Example 2.1.17 that $\gcd(1311, 5757) = 57$. But $57 \nmid 12963$ since $12963 = 227 \times 57 + 24$. By Bézout's lemma, the equation $1311x + 5757y = 12963$ has no solutions.
- For fixed z , the equation $4u + 6v = z$ has solutions exactly when z is even, since $\gcd(4, 6) = 2$.
- For fixed a, b , the equation $au + bv = 0$ always has solution. Indeed, setting $u = b$ and $v = -a$ gives a solution; but we knew one had to exist since by Exercise 2.1.6 we know that $d \mid 0$ for all $d \in \mathbb{Z}$.

2.1.22 Exercise. Which of the following equations have solutions?

- (a) $12u + 9v = -18$
- (b) $12u + 9v = 1$
- (c) $100u + 35v = 125$
- (d) $7125u + 1300v = 0$
- (e) $1010u + 101010v = 10101010101010$
- (f) $14u - 4v = 12$

Hint: Exercise 2.1.19 would be a good starting point.

2.1.23 Definition. Let $a, b \in \mathbb{Z}$. We say a and b are **coprime**, and write $a \perp b$ (read ‘ a is coprime to b ’) ([L^AT_EX](#) code: `\perp`), if $\gcd(a, b) = 1$.

2.1.24 Exercise. Which integers in the set $[15]$ are coprime to 15?

2.1.25 Proposition. Let $a, b \in \mathbb{Z}$. The following are equivalent:

- (1) a and b are coprime;
- (2) If $d \in \mathbb{Z}$ with $d \mid a$ and $d \mid b$, then d is a unit.

Proof. We prove that condition (1) implies condition (2), and vice versa.

- (1) $\Rightarrow$ (2). Suppose a and b are coprime, and fix $d \in \mathbb{Z}$ with $d \mid a$ and $d \mid b$. Then $d \mid \gcd(a, b) = 1$, so d is a unit.
- (2) $\Rightarrow$ (1). Suppose condition (2) above holds. We prove that 1 satisfies the conditions required to be a greatest common divisor of a and b . The fact that $1 \mid a$ and $1 \mid b$ is automatic; and the fact that if $d \mid a$ and $d \mid b$ implies $d \mid 1$ is precisely the condition (2) that we are assuming.

Hence the two conditions are equivalent. $\square$

2.1.26 Proposition. Let $a, b \in \mathbb{Z}$ and let $d = \gcd(a, b)$. The integers $\frac{a}{d}$ and $\frac{b}{d}$ are coprime.

2.1.27 Exercise. Prove Proposition 2.1.26.

The following corollary is a specialisation of Bézout's theorem to the case when a and b are coprime.

2.1.28 Corollary. Let $a, b \in \mathbb{Z}$. The equation $au + bv = 1$ has a solution if and only if a and b are coprime. Moreover, if a and b are coprime, then the equation $au + bv = z$ has a solution for all $z \in \mathbb{Z}$.

Proof. By Bézout's lemma (Theorem 2.1.20, the equation $au + bv = 1$ has a solution if and only if $\gcd(a, b) \mid 1$. But the only positive divisor of 1 is 1, so a solution exists if and only if $\gcd(a, b) = 1$, which is precisely the assertion that a and b are coprime.

If a and b are coprime, then $1 = \gcd(a, b) \mid z$ for all $z \in \mathbb{Z}$. So by Bézout's lemma again, the equation $au + bv = z$ has a solution for all $z \in \mathbb{Z}$. $\square$

A useful consequence of Bézout's lemma is the following result:

2.1.29 Proposition. Let $a, b, c \in \mathbb{Z}$. If a and b are coprime and $a \mid bc$, then $a \mid c$.

Proof. By Bézout's lemma (Theorem 2.1.20) there exist integers u and v such that $au + bv = 1$. Multiplying by c gives $acu + bcv = c$. Since $a \mid bc$, we can write $bc = ka$ for some $k \in \mathbb{Z}$, and so $acu + kav = c$. But then

$$(cu + kv)a = c$$

which proves that $a \mid c$. $\square$

2.2 Prime numbers

Thinking of divisibility as a way of *breaking down* an integer, for example $12 = 2 \times 2 \times 3$, our goal now is to show that:

- There are numbers which are *atomic*: they cannot be broken down. . .
- . . . and every non-zero non-unit can be written as a product of these atomic numbers;
- . . . and this product is *essentially unique*.

There are a couple of fairly vague terms used here: ‘atomic’ and ‘essentially unique’. Respectively, we will soon make these precise as ‘prime’ and ‘unique up to reordering and multiplication by units’.

Primes and irreducibles

2.2.1 Definition. Let $p \in \mathbb{Z}$. We say p is **prime** if p is a non-zero non-unit and, given $a, b \in \mathbb{Z}$

$$p \mid ab \quad \Rightarrow \quad p \mid a \text{ or } p \mid b$$

2.2.2 Example. Here are some examples of prime and non-prime numbers:

- 2 is prime. Suppose not; then there exist $a, b \in \mathbb{Z}$ such that $2 \mid ab$ but 2 divides neither a nor b . Thus a and b are both odd, meaning that ab is odd. . . but this contradicts the assumption that $2 \mid ab$.
- 6 is not prime. Indeed, $6 \mid 2 \times 3$ but 6 divides neither 2 nor 3.

2.2.3 Exercise. Using Definition 2.2.1, prove that 3 and 5 are prime and that 4 is not prime.

2.2.4 Example. Let $k \in \mathbb{Z}$ with $0 < k < 5$. We’ll show that $5 \mid \binom{5}{k}$.

Well, by Theorem 4.2.28 we know that

$$5! = \binom{5}{k} k! (5 - k)!$$

By Proposition 4.2.26, we have

$$\underbrace{5 \times 4!}_{=5!} = \binom{5}{k} \times \underbrace{1 \times \cdots \times k}_{=k!} \times \underbrace{1 \times \cdots \times (5 - k)}_{=(5-k)!}$$

Since 5 is prime, it must divide one of the factors on the right-hand side of this equation. Thus, either 5 divides $\binom{5}{k}$, or it divides ℓ for some $1 \leq \ell \leq k$ or $1 \leq \ell \leq 5 - k$. But $k < 5$ and $5 - k < 5$, so it cannot divide any of these values of ℓ —if it did, it would imply $5 \leq \ell \leq k$ or $5 \leq \ell \leq 5 - k$, which is nonsense. Hence 5 must divide $\binom{5}{k}$.

2.2.5 Exercise. Let $p \in \mathbb{Z}$ be a positive prime and let $0 < k < p$. Show that $p \mid \binom{p}{k}$.

ASIDE

Most people are introduced to primes with a definition along the lines of ‘ p is prime if p has exactly two positive divisors’. We have avoided this to elucidate the fact that the integers together with their arithmetic structure are the canonical example of a mathematical object called a *ring*. The notion of a *prime element* can be defined in any ring as in Definition 2.2.1. Secondly, these two definitions are equivalent in $\mathbb{Z}$, but not in all rings.

2.2.6 Definition. Let $a \in \mathbb{Z}$. We say a is **reducible** if $a = mn$ for some non-zero non-units m, n ; otherwise it is **irreducible**.

2.2.7 Proposition. A non-zero non-unit p is irreducible if and only if the only divisors of p are $p, -p, 1$ and -1 .

Proof. Suppose p is irreducible and that $a \mid p$. Then $p = ab$ for some $b \in \mathbb{Z}$. Since p is irreducible, either a or b is a unit. If a is a unit then $b = \pm p$, and if b is a unit then $a = \pm p$. So the only divisors of p are ± 1 and $\pm p$.

Conversely, suppose that the only divisors of p are ± 1 and $\pm p$, and that $p = ab$. If $a = \pm 1$ then $b = \pm p$, in which case $p \mid b$; and if $a = \pm p$ then $b = \pm 1$, in which case $p \mid a$. Hence p is irreducible. $\square$

2.2.8 Example. A couple of examples of reducible and irreducible numbers are:

- 2 is irreducible: if $2 = mn$ then either m or n is even, otherwise we’d be expressing an even number as the product of two odd numbers. We may assume m is even, say $m = 2k$; then $2 = 2kn$, so $kn = 1$ and hence n is a unit.
- 6 is reducible since $6 = 2 \times 3$ and both 2 and 3 are non-zero non-units.

2.2.9 Exercise. Prove that if $p \in \mathbb{Z}$ is prime then p is irreducible.

2.2.10 Lemma. Let $a \in \mathbb{Z}$ be a non-zero non-unit. Then there are irreducibles $p_1, \dots, p_n$ such that $a = p_1 \times \dots \times p_n$.

Proof. We may assume $a > 0$, since if $a < 0$ we can just multiply by -1 .

We proceed by strong induction on $a \geq 2$. The base case has $a = 2$ since we consider only non-units.

- **(BC)** We have shown that 2 is irreducible, so setting $p_1 = 2$ yields a product of primes.
- **(IS)** Let $a \geq 2$ and suppose that each integer k with $2 \leq k \leq a$ has an expression as a product of irreducibles. If $a + 1$ is irreducible then we're done; otherwise we can write $a + 1 = st$, where $s, t \in \mathbb{Z}$ are non-zero non-units. We may assume further that s and t are positive. Moreover, $s < a + 1$ and $t < a + 1$ since $s, t \geq 2$.

By the induction hypothesis, s and t have expressions as products of irreducibles. Write

$$s = p_1 \times \cdots \times p_m, \quad t = q_1 \times \cdots \times q_n$$

This gives rise to an expression of a as a product of irreducibles:

$$a = st = \underbrace{p_1 \times \cdots \times p_m}_{=s} \times \underbrace{q_1 \times \cdots \times q_n}_{=t}$$

By induction, we're done. □

2.2.11 Proposition. Let $p \in \mathbb{Z}$. Then p is prime if and only if p is irreducible.

Proof. We prove the two directions separately.

- **Prime $\Rightarrow$ irreducible.** This was Exercise 2.2.9.
- **Irreducible $\Rightarrow$ prime.** Suppose p is irreducible. Let $a, b \in \mathbb{Z}$ and suppose $p \mid ab$. We need to show that $p \mid a$ or $p \mid b$. It suffices to show that if $p \nmid a$ then $p \mid b$.

So suppose $p \nmid a$. Let $d = \gcd(p, a)$. Since $d \mid p$ and p is irreducible, we must have $d = 1$ or $d = p$ by Proposition 2.2.7. Since $p \nmid a$ and $d \mid a$, we must therefore have $d = 1$.

By Bézout's lemma (Theorem 2.1.20), there exist $u, v \in \mathbb{Z}$ such that $au + pv = 1$. Multiplying by b gives $abu + pbv = b$. Since $p \mid ab$, there exists $k \in \mathbb{Z}$ such that $pk = ab$. Then

$$b = abu + pbv = pku + pbv = p(ku + bv)$$

so $p \mid b$, as required.

So we're done. □

Since primes and irreducibles are the same thing in $\mathbb{Z}$, we will refer to them as 'primes', unless we need to emphasise a particular aspect of them.

Prime factorisation

Having described prime numbers in two ways, each of which emphasises a their nature of being ‘unbreakable’ by multiplication, we will extend Lemma 2.2.10 to prove that every integer can be expressed as a product of primes in an essentially unique way.

2.2.12 Theorem (Fundamental theorem of arithmetic). Let $a \in \mathbb{Z}$ be a non-zero non-unit. There exist primes $p_1, \dots, p_k \in \mathbb{Z}$ such that

$$a = p_1 \times \cdots \times p_k$$

Moreover, this expression is essentially unique: if $a = q_1 \times \cdots \times q_\ell$ is another expression of a as a product of primes, then $k = \ell$ and, re-ordering the q_i if necessary, for each i there is a unit u_i such that $q_i = u_i p_i$.

Proof. We showed that such a factorisation exists in Lemma 2.2.10, with the word ‘prime’ replaced by the word ‘irreducible’. It remains to prove (essential) uniqueness.

Let k be least such that there is an expression of a as a product of k primes, namely $a = p_1 \times \cdots \times p_k$. Let $a = q_1 \times \cdots \times q_\ell$ be any other such expression. We prove by induction on k that $\ell = k$ and, after re-ordering if necessary, for each i there is a unit u_i such that $q_i = u_i p_i$.

- **(BC)** If $k = 1$ then $a = p_1$ is itself prime. Then we have $p_1 = q_1 \times \cdots \times q_\ell$. Since p_1 is prime, $p_1 \mid q_j$ for some j ; by swapping q_1 and q_j we may take $j = 1$, so that $p_1 \mid q_1$. By irreducibility of q_1 we have $q_1 = u_1 p_1$ for some unit u_1 .
- **(IS)** Let $k \geq 1$ and suppose that any integer which can be expressed as a product of k primes is (essentially) uniquely expressible in such a way. Suppose a has an expression as a product of $k + 1$ primes, and that $k + 1$ is the least such number. Suppose also that

$$a = p_1 \times \cdots \times p_k \times p_{k+1} = q_1 \times \cdots \times q_\ell$$

Note that $\ell \geq k + 1$. Since p_{k+1} is prime we must have $p_{k+1} \mid q_j$ for some j ; by swapping q_j and q_ℓ if necessary, we may take $j = \ell$, so that $p_{k+1} \mid q_\ell$. As before, $q_\ell = u_{k+1} p_{k+1}$ for some unit u_{k+1} . Dividing through by p_{k+1} gives

$$p_1 \times \cdots \times p_k = q_1 \times \cdots \times q_{\ell-1} \times u_{k+1}$$

Replacing $q_{\ell-1}$ by $q_{\ell-1} u_{k+1}$, which is still prime, we can apply the induction hypothesis to obtain $k = \ell - 1$, so $k + 1 = \ell$, and, after reordering if necessary $q_i = u_i p_i$ for all $i \leq k$. Since this also holds for $i = k + 1$, we’re done.

By induction, we’re done. □

2.2.13 Example. Here are some examples of numbers written as products of primes:

- $12 = 2 \times 2 \times 3$. We could also write this as $2 \times 3 \times 2$ or $(-2) \times (-3) \times 2$, and so on.
- $53 = 53$ is an expression of 53 as a product of primes.
- $-1000 = 2 \times 5 \times (-2) \times 5 \times 2 \times 5$.

2.2.14 Exercise. Express the following numbers as products of primes:

$$16 \quad -240 \quad 5050 \quad 111111 \quad -123456789$$

To make things slightly more concise, we introduce a standard way of expressing a number as a product of primes:

2.2.15 Definition. The **canonical prime factorisation** of a non-zero non-unit $a \in \mathbb{Z}$ is the expression in the form

$$a = up_1^{j_1} \cdots p_r^{j_r}$$

where:

- $u = 1$ if $a > 0$, and $u = -1$ if $a < 0$;
- The numbers p_i are all positive primes;
- $p_1 < p_2 < \cdots < p_r$;
- $j_i \geq 1$ for all i .

We call j_i the **multiplicity** of p_i in the factorisation of a , and we call u the **sign** of a .

Typically we omit u if $u = 1$, and just write a minus sign $(-)$ if $u = -1$.

2.2.16 Example. The canonical prime factorisations of the integers given in Example 2.2.13 are:

- $12 = 2^2 \cdot 3$.
- $53 = 53$.
- $-1000 = -2^3 \cdot 5^3$.

2.2.17 Exercise. Write out the canonical prime factorisations of the numbers from Exercise 2.2.14, which were:

$$16 \quad -240 \quad 5050 \quad 111111 \quad -123456789$$

Distribution of primes

So far we have seen several examples of prime numbers; to name a few, we've seen 2, 3, 5 and 53. It might seem like the prime numbers go on forever, but proving this is less than obvious.

2.2.18 Exercise. Let X be an inhabited finite set of prime numbers and let m be the product of all the elements of X . That is, for some $n \geq 1$ we can write

$$X = \{p_1, \dots, p_n\}, \quad m = p_1 \times \dots \times p_n$$

Using the fundamental theorem of arithmetic, show that $m + 1$ has a prime divisor which is not an element of X .

2.2.19 Theorem. There are infinitely many primes.

Proof. If X is any finite set of primes, then by Exercise 2.2.18 we can find a prime number not contained in X . Hence the set of all primes cannot be finite. $\square$

This is one proof of many, which is due to Euclid around 2300 years ago. We might hope that a proof of the infinitude of primes gives some insight into how the primes are *distributed*. That is, we might ask questions like: how frequently do primes occur? How fast does the sequence of primes grow? How likely is there to be a prime number in a given set of integers?

As a starting point, Euclid's proof gives an algorithm for writing an infinite list of primes:

- Let $p_1 = 2$; we know that 2 is prime;
- Given $p_1, \dots, p_n$, let p_{n+1} be the smallest positive prime factor of $p_1 \times \dots \times p_n + 1$.

The first few terms produced would be:

- $p_1 = 2$ by definition;
- $2 + 1 = 3$, which is prime, so $p_2 = 3$;
- $2 \times 3 + 1 = 7$, which is prime, so $p_3 = 7$;
- $2 \times 3 \times 7 + 1 = 43$, which is prime, so $p_4 = 43$;
- $2 \times 3 \times 7 \times 43 + 1 = 1807 = 13 \times 139$, so $p_5 = 13$;
- $2 \times 3 \times 7 \times 43 \times 13 + 1 = 23479 = 53 \times 443$, so $p_6 = 53$;
- ... and so on.

The sequence obtained, called the *Euclid–Mullin sequence*, is a bit bizarre:

2, 3, 7, 43, 13, 53, 5, 6221671, 38709183810571, 139, 2801, 11, 17, 5471, . . .

Big primes like 38709183810571 often appear before small primes like 11. It remains unknown whether or not every positive prime number appears in this list!

The chaotic nature of this sequence makes it difficult to extract information about how the primes are distributed: the numbers $p_1 \times \cdots \times p_n + 1$ grow very quickly—indeed, it must be the case that $p_1 \times \cdots \times p_n + 1 > 2^n$ for all n —so the upper bounds for the sequence grow at least exponentially.

An investigation into other proofs of the infinitude of primes and their implications for the distribution of primes would make a good final project, see Project ?? on page ??.

2.3 Modular arithmetic

It turns out that much arithmetic can be done by considering only the *remainders* of integers when divided by a fixed integer. Here is a simple example:

2.3.1 Example. Suppose a_1 has remainder r_1 and a_2 has remainder r_2 when divided by 7. That is, there exist $q_1, q_2 \in \mathbb{Z}$ such that

$$a_1 = 7q_1 + r_1 \quad \text{and} \quad a_2 = 7q_2 + r_2$$

Then $a_1 + a_2$ has the same remainder as $r_1 + r_2$ when divided by 7. Indeed, suppose $a_1 + a_2 = 7q + r$, where $0 \leq r < 7$. Then

$$\begin{aligned} r_1 + r_2 &= (a_1 - 7q_1) + (a_2 - 7q_2) \\ &= (a_1 + a_2) - 7(q_1 + q_2) \\ &= (7q + r) - 7(q_1 + q_2) \\ &= 7(q - q_1 - q_2) + r \end{aligned}$$

An example of this in action: $41 = 5 \times 7 + 6$ and $240 = 34 \times 7 + 2$, so the remainders of 41 and 240 when divided by 7 are 6 and 2, respectively. Now

$$41 + 240 = 281 = 40 \times 7 + 1 \quad \text{and} \quad 6 + 2 = 8 = 1 \times 7 + 1$$

which demonstrates that $41 + 240$ and $6 + 2$ have the same remainder when divided by 7.

In this section we will study the extent to which we can do arithmetic with integers knowing only their remainders upon division by a given integer.

2.3.2 Definition. Fix $n \in \mathbb{Z}$. Given integers $a, b \in \mathbb{Z}$, we say a is **congruent** to b **modulo** n , and write

$$a \equiv b \pmod{n} \quad (\text{L\AA T\_{E}X code: } \texttt{a \equiv b \pmod{n}})$$

if a and b have the same remainder when divided by n . If a is not congruent to b modulo n , write

$$a \not\equiv b \pmod{n} \quad (\text{L\AA T\_{E}X code: } \texttt{\not\equiv \pmod{n}})$$

That is, $a \equiv b \pmod{n}$ if and only if there exist $q_1, q_2 \in \mathbb{Z}$ and $r \in \mathbb{Z}$ with $0 \leq r < |n|$ such that

$$a = q_1 n + r \quad \text{and} \quad b = q_2 n + r$$

2.3.3 Example. Some examples of congruence modulo n are as follows:

- $16 \equiv 30 \pmod{2}$ since 16 and 30 are both even, so they both leave the same remainder—namely 0—when divided by 2.

- $44 \equiv 20 \pmod{6}$ since 44 and 20 both have remainder 2 when divided by 6.
- We saw in Example 2.3.1 that if $a_1 \equiv r_1 \pmod{7}$ and $a_2 \equiv r_2 \pmod{7}$ then $a_1 + a_2 \equiv r_1 + r_2 \pmod{7}$.

2.3.4 Exercise. Show that if $a, b \in \mathbb{Z}$ with $a, b \geq 0$ then $a \equiv b \pmod{10}$ if and only if the decimal expressions of a and b end in the same digit.

There is a slightly simpler characterisation of congruence modulo n :

2.3.5 Proposition. Fix $n \in \mathbb{Z}$. Given $a, b \in \mathbb{Z}$, $a \equiv b \pmod{n}$ if and only if $n \mid a - b$.

Strategy. When we've introduced the variables n, a, b , the statement we have to prove is of the form $p \Leftrightarrow q$, where p is the statement ' $a \equiv b \pmod{n}$ ' and q is the statement ' $n \mid a - b$ '. We will prove $p \Rightarrow q$ directly, and $q \Rightarrow p$ by contraposition.

Proof. We prove each direction separately.

- ($\Rightarrow$) Suppose $a \equiv b \pmod{n}$, and let $q_1, q_2, r \in \mathbb{Z}$ be such that

$$a = q_1n + r, \quad b = q_2n + r \quad \text{and} \quad 0 \leq r < |n|$$

Then $a - b = (q_1 - q_2)n$, which proves that $n \mid a - b$.

- ($\Leftarrow$) Suppose that $a \not\equiv b \pmod{n}$. Then there exist $q_1, q_2, r_1, r_2 \in \mathbb{Z}$, with $r_1 \neq r_2$, such that

$$a = q_1n + r_1, \quad b = q_2n + r_2, \quad 0 \leq r_1 < |n| \quad \text{and} \quad 0 \leq r_2 < |n|$$

We may assume $r_1 > r_2$, otherwise swap the roles of a and b . Then we have

$$a - b = (q_1 - q_2)n + r_1 - r_2$$

Then $r_1 - r_2 > 0$ since $r_1 > r_2$, and $r_1 - r_2 \leq r_1 < |n|$, so the remainder of $a - b$ when divided by $|n|$ is nonzero. Hence $|n| \nmid a - b$.

So we're done. □

For example, we saw that $44 \equiv 20 \pmod{6}$, and indeed $44 - 20 = 24 = 6 \times 4$, so $6 \mid 44 - 20$.

The following theorem tells us that, in a very limited sense, the $\equiv$ symbol can be treated as a $=$ symbol for the purposes of doing addition, subtraction and multiplication.

2.3.6 Theorem (Modular arithmetic). Fix $n \in \mathbb{Z}$, and let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ be such that

$$a_1 \equiv b_1 \pmod{n} \quad \text{and} \quad a_2 \equiv b_2 \pmod{n}$$

Then the following congruences hold:

- (a) $a_1 + a_2 \equiv b_1 + b_2 \pmod{n}$;
- (b) $a_1 a_2 \equiv b_1 b_2 \pmod{n}$.

Proof. We know from Proposition 2.3.5 that $n \mid a_1 - b_1$ and $n \mid a_2 - b_2$, so there exist $q_1, q_2 \in \mathbb{Z}$ such that

$$a_1 - b_1 = q_1 n \quad \text{and} \quad a_2 - b_2 = q_2 n$$

This implies that

$$(a_1 + a_2) - (b_1 + b_2) = (a_1 - b_1) + (a_2 - b_2) = q_1 n + q_2 n = (q_1 + q_2)n$$

so $n \mid (a_1 + a_2) - (b_1 + b_2)$. This proves (a).

The algebra for (b) is slightly more involved:

$$\begin{aligned} a_1 a_2 - b_1 b_2 &= (q_1 n + b_1)(q_2 n + b_2) - b_1 b_2 \\ &= q_1 q_2 n^2 + b_1 q_2 n + b_2 q_1 n + b_1 b_2 - b_1 b_2 \\ &= q_1 q_2 n^2 + b_1 q_2 n + b_2 q_1 n \\ &= (q_1 q_2 n + b_1 q_2 + b_2 q_1)n \end{aligned}$$

This shows that $n \mid a_1 a_2 - b_1 b_2$, thus proving (b). □

2.3.7 Corollary. Fix $n \in \mathbb{Z}$. If $a, b \in \mathbb{Z}$ with $a \equiv b \pmod{n}$ then for all $c \in \mathbb{Z}$ we have:

- (a) $ca \equiv cb \pmod{n}$;
- (b) $c - a \equiv c - b \pmod{n}$.

Proof. (a) is immediate from Theorem 2.3.6(b), since $c \equiv c \pmod{n}$ and $a \equiv b \pmod{n}$.

It then follows that $-a \equiv -b \pmod{n}$ by setting $c = -1$ in (a). Then $c - a \equiv c - b \pmod{n}$ follows from Theorem 2.3.6(a), since $c \equiv c \pmod{n}$ and $-a \equiv -b \pmod{n}$. □

This corollary shows that Theorem 2.3.6 allows us to perform algebraic manipulations with congruences as if they were equations, provided all we're doing is adding, multiplying and subtracting.

2.3.8 Example. We will solve the congruence $3x - 5 \equiv 2x + 3 \pmod{7}$ for x :

$$\begin{aligned} 3x - 5 &\equiv 2x + 3 \pmod{7} \\ \Leftrightarrow x - 5 &\equiv 3 \pmod{7} && \text{subtract } 2x \\ \Leftrightarrow x &\equiv 8 \pmod{7} && \text{add } 5 \\ \Leftrightarrow x &\equiv 1 \pmod{7} && \text{since } 8 \equiv 1 \pmod{7} \end{aligned}$$

So the integers x for which $3x - 5$ and $2x + 3$ leave the same remainder when divided by 7, are precisely the integers x which leave a remainder of 1 when divided by 7:

$$3x - 5 \equiv 2x + 3 \pmod{7} \quad \Leftrightarrow \quad x = 7q + 1 \text{ for some } q \in \mathbb{Z}$$

2.3.9 Exercise. For which integers x does the congruence $5x + 1 \equiv x + 8 \pmod{3}$ hold? Characterise such integers x in terms of their remainder when divided by 3.

So far this all feels like we haven't done very much: we've just introduced a new symbol $\equiv$ which behaves just like equality... but does it really? The following exercises should expose some more ways in which congruence *does* behave like equality, and some in which it *doesn't*.

2.3.10 Exercise. Fix $n \in \mathbb{Z}$. Is it true that

$$a \equiv b \pmod{n} \quad \Rightarrow \quad a^k \equiv b^k \pmod{n}$$

for all $a, b \in \mathbb{Z}$ and $k \in \mathbb{N}$? If so, prove it; if not, provide a counterexample.

2.3.11 Exercise. Fix $n \in \mathbb{Z}$. Is it true that

$$k \equiv \ell \pmod{n} \quad \Rightarrow \quad a^k \equiv a^\ell \pmod{n}$$

for all $k, \ell \in \mathbb{N}$ and $a \in \mathbb{Z}$? If so, prove it; if not, provide a counterexample.

2.3.12 Exercise. Fix $n \in \mathbb{Z}$. Is it true that

$$qa \equiv qb \pmod{n} \quad \Rightarrow \quad a \equiv b \pmod{n}$$

for all $a, b, q \in \mathbb{Z}$ with $q \not\equiv 0 \pmod{n}$? If so, prove it; if not, provide a counterexample.

COMMON ERROR

The false sense of security that Theorem 2.3.6 induces often leads students new to all this to the belief that $\equiv$ and $=$ are interchangeable concepts. This is emphatically *not* the case. In particular:

- Fractions don't make sense in modular arithmetic; for instance, it is invalid to say $2x \equiv 1 \pmod{5}$ implies $x \equiv \frac{1}{2} \pmod{5}$.
- Square roots don't make sense in modular arithmetic; for instance, it is invalid to say

$$x^2 \equiv 3 \pmod{4} \text{ implies } x \equiv \pm\sqrt{3} \pmod{4}.$$

- Numbers in exponents cannot be replaced by congruent numbers; for instance, it is invalid to say $x^3 \equiv 2^3 \pmod{4}$ implies $x \equiv 2 \pmod{4}$.

Multiplicative inverses

I made a big deal about the fact that fractions don't make sense in modular arithmetic. That is, it is invalid to say

$$2x \equiv 1 \pmod{5} \quad \Rightarrow \quad x \equiv \frac{1}{2} \pmod{5}$$

Despite this, we can still make sense of 'division', provided we change what we mean when we say 'division'. Indeed, the congruence $2x \equiv 1 \pmod{5}$ has a solution:

$$\begin{aligned} 2x &\equiv 1 \pmod{5} \\ \Leftrightarrow 6x &\equiv 3 \pmod{5} && \text{multiply by 3} \\ \Leftrightarrow x &\equiv 3 \pmod{5} && \text{since } 6 \equiv 1 \pmod{5} \end{aligned}$$

Here we didn't divide by 2, but we still managed to cancel the 2 by instead multiplying through by 3. For the purposes of solving the equation this had the same effect as division by 2 would have had if we were allowed to divide. The key here was that $2 \times 3 \equiv 1 \pmod{5}$.

2.3.13 Definition. Fix $n \in \mathbb{Z}$. Given $a \in \mathbb{Z}$, a **multiplicative inverse** for a modulo n is an integer u such that $au \equiv 1 \pmod{n}$.

2.3.14 Example. Some examples of multiplicative inverses are as follows:

- 2 is a multiplicative inverse of itself modulo 3, since $2 \times 2 \equiv 4 \equiv 1 \pmod{3}$.
- 2 is a multiplicative inverse of 3 modulo 5, since $2 \times 3 \equiv 6 \equiv 1 \pmod{5}$.
- 7 is also a multiplicative inverse of 3 modulo 5, since $3 \times 7 \equiv 21 \equiv 1 \pmod{5}$.
- 3 has no multiplicative inverse modulo 6. Indeed, suppose $u \in \mathbb{Z}$ with $3u \equiv 1 \pmod{6}$. Then $6 \mid 3u - 1$, so $3u - 1 = 6q$ for some $q \in \mathbb{Z}$. But then

$$1 = 3u - 6q = 3(u - 2q)$$

which implies that $3 \mid 1$, which is nonsense.

Knowing when multiplicative inverses exist is very important for solving congruences: if u is a multiplicative inverse for a modulo n , then we can solve equations of the form $ax \equiv b \pmod{n}$ extremely easily:

$$ax \equiv b \pmod{n} \quad \Rightarrow \quad x \equiv ub \pmod{n}$$

2.3.15 Exercise. For $n = 7, 8, 9, 10, 11, 12$, either find a multiplicative inverse for 6 modulo n , or show that no multiplicative inverse exists. Can you spot a pattern?

2.3.16 Proposition. Let $a, n \in \mathbb{Z}$. Then a has a multiplicative inverse modulo n if and only if $a \perp n$.

Proof. Note that a has a multiplicative inverse u modulo n if and only if there is a solution (u, v) to the equation $au + nv = 1$. Indeed, $au \equiv 1 \pmod{n}$ if and only if $n \mid au - 1$, which occurs if and only if there is some $q \in \mathbb{Z}$ such that $au - 1 = nq$. Setting $q = -v$ and rearranging yields the desired equivalence.

By Bézout's lemma (Theorem 2.1.20), such a solution (u, v) exists if and only if $\gcd(a, n) \mid 1$. This occurs if and only if $\gcd(a, n) = 1$, i.e. if and only if $a \perp n$. $\square$

PROOF TIP

To solve a congruence of the form $ax \equiv b \pmod{n}$ when $a \perp n$, first find a multiplicative inverse u for a modulo n , and then simply multiply through by u to obtain $x \equiv ub \pmod{n}$.

2.3.17 Corollary. Let $a, p \in \mathbb{Z}$, where p is prime. If $p \nmid a$ then a has a multiplicative inverse modulo p .

Proof. Suppose $p \nmid a$, and let $d = \gcd(a, p)$. Since $d \mid p$ and p is prime we have $d = 1$ or $d = p$. Since $d \mid a$ and $p \nmid a$ we can't have $d = p$; therefore $d = 1$. By Proposition 2.3.16, a has a multiplicative inverse modulo p . $\square$

2.3.18 Example. 11 is prime, so each of the integers a with $1 \leq a \leq 10$ should have a multiplicative inverse modulo 11. And indeed, the following are all congruent to 1 modulo 11:

$$\begin{array}{cccccc} 1 \times 1 = 1 & 2 \times 6 = 12 & 3 \times 4 = 12 & 4 \times 3 = 12 & 5 \times 9 = 45 \\ 6 \times 2 = 12 & 7 \times 8 = 56 & 8 \times 7 = 56 & 9 \times 5 = 45 & 10 \times 10 = 100 \end{array}$$

2.3.19 Exercise. Find all integers x such that $25x - 4 \equiv 4x + 3 \pmod{13}$.

Orders and totients

For any $n \in \mathbb{Z}$ there are only finitely possible remainders modulo n . A nice consequence of this finiteness is that, when $a \perp n$, we can choose some power of a to be its multiplicative inverse, as proved in the following exercise.

2.3.20 Exercise. Let $a, n \in \mathbb{Z}$ with $a \perp n$. Prove that there exists $k \geq 1$ such that $a^k \equiv 1 \pmod{n}$.

2.3.21 Definition. Let $a, n \in \mathbb{Z}$ with $a \perp n$. The **order** of a modulo n is the least $k \geq 1$ such that $a^k \equiv 1 \pmod{n}$.

Note that this definition makes sense by Exercise 2.3.20 and the well-ordering principle.

2.3.22 Example. The powers of 7 modulo 100 are:

- $7^1 = 7$, so $7^1 \equiv 7 \pmod{100}$;
- $7^2 = 49$, so $7^2 \equiv 49 \pmod{100}$;
- $7^3 = 343$, so $7^3 \equiv 43 \pmod{100}$;
- $7^4 = 2401$, so $7^4 \equiv 1 \pmod{100}$.

Hence the order of 7 modulo 100 is 4, and 7^3 and 43 are multiplicative inverses of 7 modulo 100.

Our focus turns to computing specific values of k such that $a^k \equiv 1 \pmod{n}$, whenever $a, n \in \mathbb{Z}$ and $a \perp n$. We first focus on the case when n is prime; then we develop the machinery of *totients* to study the case when n is not prime.

2.3.23 Lemma. Let $a, b \in \mathbb{Z}$ and let $p \in \mathbb{Z}$ be a positive prime. Then $(a + b)^p \equiv a^p + b^p \pmod{p}$.

Proof. By the binomial theorem

$$(a + b)^p = \sum_{k=0}^p \binom{p}{k} a^k b^{p-k}$$

By Exercise 2.2.5, $p \mid \binom{p}{k}$ for all $0 < k < p$, and hence $\binom{p}{k} a^k b^{p-k} \equiv 0 \pmod{p}$ for all $0 < k < p$. Thus

$$(a + b)^p \equiv \binom{p}{0} a^0 b^{p-0} + \binom{p}{p} a^p b^{p-p} \equiv a^p + b^p \pmod{p}$$

as desired. □

2.3.24 Theorem (Fermat's little theorem). Let $a, p \in \mathbb{Z}$ with p prime. Then $a^{|p|} \equiv a \pmod{p}$.

Proof. We may assume $p > 0$; otherwise, replace p by $-p$ in the proof. We may also assume that $a \geq 0$, otherwise replace a by its remainder modulo p .

We will prove that $a^{|p|} \equiv a \pmod{p}$ by induction on a .

- **(BC)** Since $p \neq 0$ we have $0^p = 0$, hence $0^p \equiv 0 \pmod{p}$.
- **(IS)** Fix $a \geq 0$ and suppose $a^p \equiv a \pmod{p}$. Then $(a+1)^p \equiv a^p + 1^p \pmod{p}$ by Lemma 2.3.23. Now $a^p \equiv a \pmod{p}$ by the induction hypothesis, and $1^p = 1$, so we have $(a+1)^p \equiv a+1 \pmod{p}$.

By induction, we're done. □

2.3.25 Corollary. Let $a, p \in \mathbb{Z}$ with p prime and $p \nmid a$. then $a^{|p|-1} \equiv 1 \pmod{p}$.

Proof. Since $p \nmid a$, it follows that $a \perp p$. Fermat's little theorem (Theorem 2.3.24) tells us that $a^{|p|} \equiv a \pmod{p}$. By Proposition 2.3.16, a has a multiplicative inverse b modulo p . Hence

$$a^{|p|}b \equiv ab \pmod{p}$$

But $a^{|p|}b \equiv a^{|p|-1}ab \pmod{p}$, and $ab \equiv 1 \pmod{p}$, so we get

$$a^{|p|-1} \equiv 1 \pmod{p}$$

as required. □

This can be useful for computing remainders of humongous numbers when divided by smaller primes.

2.3.26 Example. We compute the remainder of 2^{1000} when divided by 7. By Fermat's little theorem (Theorem 2.3.24), we know that $2^7 \equiv 2 \pmod{7}$. Since $7 \nmid 2$ it has a multiplicative inverse, so we can cancel it from both sides to obtain $2^6 \equiv 1 \pmod{7}$. Now $1000 = 166 \times 6 + 4$, so

$$2^{1000} \equiv 2^{166 \times 6 + 4} \equiv (2^6)^{166} \cdot 2^4 \equiv 2^4 \equiv 16 \equiv 2 \pmod{7}$$

so the remainder of 2^{1000} when divided by 7 is 2.

2.3.27 Exercise. Find the remainder of 3^{244886} when divided by 13.

Unfortunately, the hypothesis that p is prime in Fermat's little theorem is necessary. For example, 6 is not prime, and $5^{6-1} = 5^5 = 3125 = 520 \times 6 + 5$, so $5^5 \equiv 5 \pmod{6}$.

2.3.28 Definition. Let $n \in \mathbb{Z}$. The **totient** of n is the natural number $\varphi(n)$ ([L^AT_EX](#) code: `\varphi(n)`) defined by

$$\varphi(n) = |\{k \in [|n|] \mid k \perp n\}|$$

That is, $\varphi(n)$ is the number of integers from 1 to $|n|$ which are coprime to n .

2.3.29 Example. Here are some examples of totients:

- The elements of $[6]$ which are coprime to 6 are 1 and 5, so $\varphi(6) = 2$.
- If p is a positive prime, then every element of $[p]$ is coprime to p except for p itself. Hence if p is a positive prime then $\varphi(p) = p - 1$. More generally, if p is prime then $\varphi(p) = |p| - 1$.

2.3.30 Exercise. Prove that if p is a positive prime and $k \geq 1$ then

$$\varphi(p^k) = p^k - p^{k-1}$$

2.3.31 Theorem (Euler's theorem). Let $a, n \in \mathbb{Z}$ with $a \perp n$. Then

$$a^{\varphi(n)} \equiv 1 \pmod{n}$$

Proof. By definition of totient, the set $X = \{k \in [n] \mid k \perp n\}$ has $\varphi(n)$ elements. List the elements as

$$X = \{x_1, x_2, \dots, x_{\varphi(n)}\}$$

Note that $ax_i \perp n$ for all i , so let y_i be the (unique) element of X such that $ax_i \equiv y_i \pmod{n}$.

Note that if $i \neq j$ then $y_i \neq y_j$. Indeed, since $a \perp n$, by Proposition 2.3.16, a has a multiplicative inverse, say b . Then

$$y_i = y_j \pmod{n} \Rightarrow ax_i \equiv ax_j \pmod{n} \Rightarrow bax_i \equiv bax_j \pmod{n} \Rightarrow x_i \equiv x_j \pmod{n}$$

and $x_i \equiv x_j \pmod{n}$ if and only if $x_i = x_j$. Thus

$$X = \{x_1, x_2, \dots, x_{\varphi(n)}\} = \{y_1, y_2, \dots, y_{\varphi(n)}\}$$

Pure more simply, $(y_1, \dots, y_{\varphi(n)})$ is a permutation of X : it's all the same numbers, just perhaps in a different order.

Hence

$$x_1 \times \dots \times x_{\varphi(n)} \equiv y_1 \times \dots \times y_{\varphi(n)} = a^{\varphi(n)} \times x_1 \times \dots \times x_n$$

and hence

$$x_1 \times \dots \times x_{\varphi(n)} \equiv a^{\varphi(n)} \times x_1 \times \dots \times x_n \pmod{n}$$

Since each x_i is coprime to n , we can cancel the x_i terms to obtain

$$a^{\varphi(n)} \equiv 1 \pmod{n}$$

as required. □

2.3.32 Example. Some examples of Euler's theorem in action are as follows:

- We have seen that $\varphi(6) = 2$, and we know that $5 \perp 6$. And, indeed,

$$5^{\varphi(6)} = 5^2 = 25 = 4 \times 6 + 1$$

so $5^{\varphi(6)} \equiv 1 \pmod{6}$.

- By Exercise 2.3.30, we have

$$\varphi(121) = \varphi(11^2) = 11^2 - 11^1 = 121 - 11 = 110$$

Moreover, given $a \in \mathbb{Z}$, $a \perp 121$ if and only if $11 \nmid a$. Hence $a^{110} \equiv 1 \pmod{121}$ whenever $11 \nmid a$.

2.3.33 Exercise. Use the inclusion–exclusion principle to show that $\varphi(100) = 40$. Use this fact to prove that the last two digits of 3^{79} are ‘67’.

2.4 Problems

Chapter 3

Relations and functions

So far, our focus has been on studying mathematical objects in isolation: what properties are true of natural numbers, integers, and so on. However, most mathematically interesting results we've seen so far rely not on objects in isolation, but their interactions with other objects. For example, induction principles rely on the order of the natural numbers with respect to one another, and all results regarding divisibility, primality and greatest common divisors concern relationships between integers with respect to divisibility. This chapter focuses the study of mathematical objects with respect to one another—not just to themselves.

First we will study **relations**, which are ways that mathematical objects can interact with one another. We will classify certain kinds of relations in terms of their behaviour. For example, congruence modulo n is a relation which behaves like equality in a way that we will make precise as an *equivalence relation*. Furthermore, the relations of 'less than' and 'divisibility' have very similar behaviours, in a way that we will make precise as *partial order* relations.

We will then move on to **functions**, which provide a way of transforming one kind of mathematical object into another in a determined fashion. The value of functions is impossible to understate; functions lie at the core of many, if not most, areas of modern mathematics. Our first application of functions will be in Chapter 4, when we use them to compare the sizes of (potentially infinite) sets. We will then study the interaction between functions and structure in Chapter 5.

3.1 Relations

3.1.1 Definition. Let X and Y be sets. A **(binary) relation** from X to Y is a logical formula $R(x, y)$ with two free variables x, y , where x has range X and y has range Y . We call X the **domain** of R and Y the **codomain** of R .

Given $x \in X$ and $y \in Y$, if $R(x, y)$ is true then we say ‘ x is **related** to y by R ’, and write $x R y$ ([L^AT_EX](#) code: `x\; R\; y`).

A relation from a set X to itself is called a relation **on** X .

In more human terms, a relation from X to Y is a statement about a general element $x \in X$ and a general element $y \in Y$, which can be true or false depending on the values of x and y .

3.1.2 Example. We have seen many examples of relations so far. For example:

- Given a set X , equality between elements of X ($x = y$) is a relation on X .
- Divisibility ($x \mid y$) is a relation on $\mathbb{Z}$.
- Fix $n \in \mathbb{Z}$. Congruence modulo n ($x \equiv y \pmod{n}$) is a relation on $\mathbb{Z}$.
- Order ($x \leq y$) is a relation on $\mathbb{N}$, or on $\mathbb{Z}$, or on $\mathbb{Q}$, or on $\mathbb{R}$.
- Given sets X and Y , there is an **empty relation** $\emptyset_{X,Y}$ from X to Y , which is false for all $x \in X$ and $y \in Y$.

3.1.3 Exercise. Define a relation R on $\mathbb{Z}$ which is not on the list given in Example 3.1.2.

It is possible, and extremely useful, to represent relations as sets. We do this by defining the *graph* of a relation, which is the set of all pairs of elements which are related by the relation.

3.1.4 Definition. Let X and Y be sets, and let R be a relation from X to Y . The **graph** of R is the set Γ_R ([L^AT_EX](#) code: `\Gamma_{\mathbf{R}}`) of pairs $(x, y) \in X \times Y$ for which $x R y$. That is

$$\Gamma_R = \{(x, y) \in X \times Y \mid x R y\} \subseteq X \times Y$$

3.1.5 Example. Consider the relation of divisibility on $\mathbb{Z}$, that is $R(x, y)$ is the statement $x \mid y$. The graph Γ_R of R is the set whose elements are all pairs (m, n) where $m, n \in \mathbb{Z}$ and $m \mid n$. For example, $(2, 6) \in \Gamma_R$ since $2 \mid 6$, but $(2, 7) \notin \Gamma_R$ since $2 \nmid 7$.

Since $m \mid n$ if and only if $n = qm$ for some $q \in \mathbb{Z}$, we thus have

$$\Gamma_R = \{(m, qm) \mid m, q \in \mathbb{Z}\} \subseteq \mathbb{Z} \times \mathbb{Z}$$

3.1.6 Exercise. Let X and Y be sets. What is the graph of the empty relation from X to Y ?

The graph of a relation R from a set X to a set Y can often be represented graphically: draw a pair of axes, with the horizontal axis representing the elements of X and the vertical axis representing the elements of Y , and plot the point (x, y) if and only if $R(x, y)$ is true.

3.1.7 Example. Consider the relation S from $\mathbb{R}$ to $\mathbb{R}$ defined by $S(x, y) \Leftrightarrow x^2 + y^2 = 1$. Then

$$\Gamma_S = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = 1\}$$

Plotting Γ_S on a standard pair of axes yields a circle with radius 1 centred at the point $(0, 0)$.

3.1.8 Example. The graph of the equality relation on a set X is very simple:

$$\Gamma_{=} = \{(x, y) \in X \times X \mid x = y\} = \{(x, x) \mid x \in X\} \subseteq X \times X$$

This set is often denoted Δ_X ([L^AT_EX code: `\Delta\_{\mathbf{X}}`](#)), and called the *diagonal subset* of $X \times X$. The reason for the word ‘diagonal’ is because—provided the horizontal and vertical axes have the same ordering of the elements of X —the points plotted are precisely those on the diagonal line.

Since we defined relations as particular logical formulae, and we have not defined a notion of equality between logical formulae, if we want to say that two relations are equal then first we need to define what we mean by *equal*. As with sets, this raises some subtleties: should two relations be equal when they’re described by the same formula? Or should two relations be equal when they relate the same elements, even if their underlying descriptions are somewhat different? As with equality between sets (Definition 1.2.21), our notion of equality between relations will be *extensional*: for the purposes of deciding whether two relations are equal, we forget their descriptions and look only at whether or not they relate the same pairs elements.

3.1.9 Definition. Let X and Y be sets, and let R and S be relations from X to Y . We say R and S are **equal**, and write $R = S$, if

$$\forall x \in X, \forall y \in Y, (x R y \Leftrightarrow x S y)$$

That is, $R = S$ if they relate the same pairs of elements.

Note that two relations R and S are equal as relations if and only if their graphs Γ_R and Γ_S are equal as sets. This fact, together with the correspondence between relations from a set X to a set Y and subsets of $X \times Y$ given by Theorem 3.1.10 below, is incredibly convenient, because it makes the notion of a relation more concrete.

3.1.10 Theorem. Let X and Y be sets. Any subset $\Gamma \subseteq X \times Y$ is the graph of exactly one relation R from X to Y .

Proof. Fix $\Gamma \subseteq X \times Y$. Define a relation R by

$$\forall x \in X, \forall y \in Y, x R y \Leftrightarrow (x, y) \in \Gamma$$

Then certainly $\Gamma = \Gamma_R$.

Moreover, if S is a relation from X to Y such that $\Gamma = \Gamma_S$, then, for all $x \in X$ and $y \in Y$

$$x S y \Leftrightarrow (x, y) \in \Gamma_S \Leftrightarrow (x, y) \in \Gamma_R \Leftrightarrow x R y$$

so $S = R$. Hence there is exactly one relation from X to Y whose graph is Γ . $\square$

3.1.11 Exercise. Let X and Y be finite sets with $|X| = m$ and $|Y| = n$. Prove that there are 2^{mn} relations from X to Y .

Relations on a set

For the rest of this section, we will focus on relations on a single set X , rather than relations between two different sets. We have seen many such relations so far, such as: equality, congruence modulo n , divisibility, inclusion of subsets ($\subseteq$), and comparison of size ($\leq$). Remarkably, each of these can be characterised in one of two ways: either as an *equivalence relation* or as a *partial order*.

Equivalence relations are those that behave in some sense like equality, and partial orders are those that behave in some way like $\leq$.

• **Equality.** If X is any set, then equality on X satisfies:

- ◊ Given $x \in X$, we have $x = x$;
- ◊ Given $x, y \in X$, if $x = y$, then $y = x$;
- ◊ Given $x, y, z \in X$, if $x = y$ and $y = z$, then $x = z$.

Note that these are all true if we replace X by $\mathbb{N}$ and $\cdot = \cdot$ by $\cdot \equiv \cdot \pmod n$ for some fixed $n > 0$.

• **Order.** If $X = \mathbb{N}$ (or $\mathbb{Z}$ or $\mathbb{Q}$ or $\mathbb{R}$), then the order relation $\leq$ on X satisfies:

- ◊ Given $x \in X$, we have $x \leq x$;
- ◊ Given $x, y \in X$, if $x \leq y$ and $y \leq x$, then $x = y$;
- ◊ Given $x, y, z \in X$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

Note that these are all true if we replace $(X, \leq)$ by $(\mathcal{P}(X), \subseteq)$ or $(\mathbb{N}, |)$.

For both equality and order, the first condition states that every element is related to itself, and the third condition states that in some sense we can cut out intermediate steps. These conditions are known as *reflexivity* and *transitivity*. The second condition for equality states

that the direction of the relation doesn't matter; this condition is called *symmetry*. The second condition for the order relation states that the only way two objects can be related to each other in both directions is if they are equal; this condition is called *antisymmetry*.

We will classify relations in terms of reflexivity, symmetry, antisymmetry and transitivity, and then focus on *equivalence relations*, which are 'equality-like', and *partial orders*, which are 'order-like'.

Reflexive relations are those that relate everything to itself.

3.1.12 Definition. Let X be a set. A relation R on X is **reflexive** if $x R x$ for all $x \in X$.

3.1.13 Example. Given a set X , the equality relation on X is reflexive since $x = x$ for all $x \in X$.

3.1.14 Example. The divisibility relation on $\mathbb{N}$, or on $\mathbb{Z}$, is reflexive. Given $n \in \mathbb{Z}$ we have $n = 1 \times n$, and so $n \mid n$.

The following exercise demonstrates the importance of specifying the (co)domain of a relation: it shows that a logical formula may define a reflexive relation on one set, but not on another.

3.1.15 Exercise. Prove that coprimality (' $x \perp y$ ') is not a reflexive relation on $\mathbb{Z}$, but that it is a reflexive relation on the set $\{-1, 1\}$.

As such, it doesn't make sense to say 'coprimality is a reflexive relation' or 'coprimality is not a reflexive relation': we must specify on which set we are considering the divisibility relation.

The result of the next exercise characterises reflexive relations in terms of their graph.

3.1.16 Exercise. Let X be a set and let R be a relation on X . Prove that R is reflexive if and only if $\Delta_X \subseteq \Gamma_R$, where Δ_X is the diagonal subset of $X \times X$ (see Example 3.1.8). Deduce that if X is finite and $|X| = n$, then there are $2^{n(n-1)}$ reflexive relations on X .

Symmetric relations are those for which the *direction* of the relation doesn't matter.

3.1.17 Definition. Let X be a set. A relation R on X is **symmetric** if, for all $x, y \in X$, $x R y$ implies $y R x$.

3.1.18 Example. Some examples of symmetric relations include:

- Equality is a symmetric relation on any set X . Indeed, if $x, y \in X$ and $x = y$, then $y = x$.
- Coprimality is a symmetric relation on $\mathbb{Z}$, since if $a, b \in \mathbb{Z}$ then $a \perp b$ if and only if $b \perp a$.
- Divisibility is not a symmetric relation on $\mathbb{Z}$, since for instance $1 \mid 2$ but $2 \nmid 1$. However, divisibility *is* a symmetric relation on $\{-1, 1\}$, since $1 \mid -1$ and $-1 \mid 1$.

3.1.19 Exercise. Let X be a finite set with $|X| = n$. Prove that there are $2^{\binom{n}{2}} \cdot 2^n$ symmetric relations on X .

A related condition a relation may possess is *antisymmetry*. A word of warning here is that ‘antisymmetric’ does not mean the same thing as ‘not symmetric’—indeed, we will see, equality is both symmetric and antisymmetric, and many relations are neither symmetric nor antisymmetric.

3.1.20 Definition. Let X be a set. A relation R on X is **antisymmetric** if, for all $x, y \in X$, if $x R y$ and $y R x$, then $x = y$.

3.1.21 Example. Some examples of antisymmetric relations include:

- Let X be a set. The equality relation on X is antisymmetric: it is immediate that if $x, y \in X$ and $x = y$ and $y = x$, then $x = y$.
- The relation $\leq$ on the set $\mathbb{N}$ (or $\mathbb{Z}$ or $\mathbb{Q}$ or $\mathbb{R}$) is antisymmetric: if $m, n \in \mathbb{N}$ and $m \leq n$ and $n \leq m$, then $m = n$.
- The divisibility relation on $\mathbb{N}$ is antisymmetric. Indeed, let $m, n \in \mathbb{N}$ and suppose $m \mid n$ and $n \mid m$. Then $n = km$ for some $k \in \mathbb{Z}$ and $m = \ell n$ for some $\ell \in \mathbb{Z}$. It follows that $n = k\ell n$. If $n = 0$ then $m = n$ trivially; otherwise, we have $k\ell = 1$. Hence k is a unit; moreover, since $m, n \geq 0$ and $n = km$, we must have $k = 1$. Hence $m = n$.

3.1.22 Exercise. Show that the divisibility relation on $\mathbb{Z}$ is not antisymmetric.

3.1.23 Exercise. Let X be a set and let R be a relation on X . Prove that R is both symmetric and antisymmetric if and only if $\Gamma_R \subseteq \Delta_X$, where Δ_X is the diagonal subset of $X \times X$ (see Exercise 3.1.8). Deduce that the only reflexive, symmetric and antisymmetric relation on a set X is the equality relation on X .

3.1.24 Exercise. Let X be a finite set with $|X| = n$. Prove that there are $3^{\binom{n}{2}} \cdot 2^n$ antisymmetric relations on X .

Transitivity is the property of $\leq$ that allows us to deduce, for example, that $0 \leq 4$, from the information that $0 \leq 1 \leq 2 \leq 3 \leq 4$.

3.1.25 Definition. Let X be a set. A relation R on X is **transitive** if, for all $x, y, z \in X$, if $x R y$ and $y R z$, then $x R z$.

3.1.26 Example. Some examples of transitive relations include:

- Equality is a transitive relation on any set X , since it is immediate that if $x, y, z \in X$ with $x = y$ and $y = z$, then $x = z$.

- Divisibility is a transitive relation on $\mathbb{N}$, or on $\mathbb{Z}$. Indeed, if $a, b, c \in \mathbb{N}$ with $a \mid b$ and $b \mid c$, then there exist $k, \ell \in \mathbb{Z}$ such that $b = ka$ and $c = \ell b$. Then $c = (k\ell)a$, so $a \mid c$.
- Inclusion is a transitive relation on $\mathcal{P}(X)$, for any set X . Indeed, Proposition 1.2.19 implies that if $U, V, W \subseteq X$ with $U \subseteq V$ and $V \subseteq W$, then $U \subseteq W$.

A fundamental property of transitive relations is that we can prove two elements a and b are related by finding a chain of related elements starting at a and finishing at b . This is the content of the following proposition.

3.1.27 Proposition. Let R be a relation on a set X . Then R is transitive if and only if, for any finite sequence $x_0, x_1, \dots, x_n$ of elements of X such that $x_{i-1} R x_i$ for all $i \in [n]$, we have $x_0 R x_n$.

Proof. For the sake of abbreviation, let $p(n)$ be the assertion that, for any $n \geq 1$ and any sequence $x_0, x_1, \dots, x_n$ of elements of X such that $x_{i-1} R x_i$ for all $i \in [n]$, we have $x_0 R x_n$.

We prove the two directions of the proposition separately.

- $(\Rightarrow)$ Suppose R is transitive. For $n \geq 1$. We prove $p(n)$ is true for all $n \geq 1$ by induction.
 - ◊ (BC) When $n = 1$ this is immediate, since we assume that $x_0 R x_1$.
 - ◊ (IS) Fix $n \geq 1$ and suppose $p(n)$ is true. Let $x_0, \dots, x_n, x_{n+1}$ is a sequence such that $x_{i-1} R x_i$ for all $i \in [n+1]$. We need to prove that $x_0 R x_n$.
By the induction hypothesis we know that $x_0 R x_n$. By definition of the sequence we have $x_n R x_{n+1}$. By transitivity, we have $x_0 R x_{n+1}$.

So by induction, we have proved the $\Rightarrow$ direction.

- $(\Leftarrow)$ Suppose $p(n)$ is true for all $n \geq 1$. Then in particular $p(2)$ is true, which is precisely the assertion that R is transitive.

So we're done. □

That is, Proposition 3.1.27 states that for a transitive relation R on a set X , if $x_0, x_1, \dots, x_n \in X$ and

$$x_0 R x_1 R \cdots R x_n$$

then $x_0 R x_n$.

Equivalence relations

We will now study what it is for a relation to be *equality-like*.

3.1.28 Definition. A relation R on a set X is an **equivalence relation** if R is reflexive, symmetric and transitive.

When we talk about arbitrary equivalence relations, we usually use a symbol like ‘ $\sim$ ’ ([L^AT_EX code: `\sim`](#)) or ‘ $\equiv$ ’ ([L^AT_EX code: `\equiv`](#)) or ‘ $\approx$ ’ ([L^AT_EX code: `\approx`](#)) instead of ‘ R ’.

3.1.29 Example. We have seen two main examples of equivalence relations so far.

- Equality is an equivalence relation. The proof is immediate.
- Fix $n > 0$. Congruence modulo n is an equivalence relation on $\mathbb{Z}$; spelling out the details is almost a triviality:
 - ◊ **Reflexivity.** If $a \in \mathbb{Z}$ then $a \equiv a \pmod{n}$ since a leaves the same remainder as itself when divided by n .
 - ◊ **Symmetry.** If $a, b \in \mathbb{Z}$ and $a \equiv b \pmod{n}$, then a and b leave the same remainder when divided by n . This means that b and a leave the same remainder when divided by n , so $b \equiv a \pmod{n}$.
 - ◊ **Transitivity.** If $a, b, c \in \mathbb{Z}$, and $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then a and b , and b and c , leave the same remainder when divided by n ; hence a and c leave the same remainder when divided by n , so $a \equiv c \pmod{n}$.

3.1.30 Exercise. Define a relation $\sim$ on $\mathbb{Z}$ by declaring, for $m, n \in \mathbb{Z}$,

$$m \sim n \iff \varphi(m) = \varphi(n)$$

Prove that $\sim$ is an equivalence relation.

Equivalence relations are useful because they allow us to ignore irrelevant information about elements of a set. As an example, to prove that, for $a \in \mathbb{Z}$, if $3 \nmid a$ then $a^2 \equiv 1 \pmod{3}$, we could write $a = 3k \pm 1$ for some $k \in \mathbb{Z}$ and deduce that $a^2 = 3(3k^2 \pm 2k) + 1$. This requires us to use more information than we need: the value of k doesn’t make any difference to the truth of the result, and the expression $3(3k^2 \pm 2k) + 1$ is frankly fairly ugly.

3.1.31 Definition. Let X be a set and let $\sim$ be an equivalence relation on X . The **$\sim$ -equivalence class** of $x \in X$ is the set

$$[x]_{\sim} = \{y \in X \mid x \sim y\}$$

The **quotient** of X by $\sim$, denoted $X/\sim$, is the set of all $\sim$ -equivalence classes of elements of X ; that is

$$X/\sim = \{[x]_{\sim} \mid x \in X\}$$

FORMATTING TIP

Putting braces (`{` and `}`) around a symbol like $\sim$ tells $\text{\LaTeX}$ to consider the symbol on its own, rather than in the context of its surrounding variables. Typing `X/\sim = Y` confuses $\text{\LaTeX}$ slightly, since it thinks you're saying ' $X/$ is related to is equal to Y ', which clearly makes no sense; whereas typing `X/{\sim} = Y` signifies to $\text{\LaTeX}$ that the $\sim$ symbol isn't relating anything. Compare:

$$X/\sim = Y \quad \text{vs.} \quad X/{\sim} = Y$$

3.1.32 Example. Let $\sim$ be the relation of congruence modulo 5 on the set of integers. Then

$$[0]_{\sim} = \{a \in \mathbb{Z} \mid a \sim 0\}$$

Now, $a \sim 0$ if and only if $5 \mid a$, so we can also write

$$[0]_{\sim} = \{\dots, -10, -5, 0, 5, 10, \dots\} = \{5k \mid k \in \mathbb{Z}\}$$

So in fact $[0]_{\sim} = [5k]_{\sim}$ for any $k \in \mathbb{Z}$. And likewise

$$[r]_{\sim} = [r + 5k]_{\sim}$$

for all $r, k \in \mathbb{Z}$

When the equivalence relation in question is congruence modulo n on the set $\mathbb{Z}$ of integers, we call the equivalence class of $a \in \mathbb{Z}$ the **congruence class** of a modulo n , denoted $[a]_n$, and we write $\mathbb{Z}/n\mathbb{Z}$ to denote the quotient of $\mathbb{Z}$ by congruence modulo n .

3.1.33 Exercise. Prove that if $n > 0$ then $\mathbb{Z}/n\mathbb{Z}$ is finite and $|\mathbb{Z}/n\mathbb{Z}| = n$.

Exercise 3.1.33 doesn't tell us much more than we already know: namely, that there are only finitely many possible remainders modulo n . But it makes our lives significantly easier for doing modular arithmetic, because now there are only finitely many objects to work with. We will come back to this when discussing functions in Section 3.2.

One last word on equivalence relations is that they are essentially the same thing as partitions (see Definition 4.1.12).

3.1.34 Exercise. If $\sim$ be an equivalence relation on X , then $X/\sim$ partitions X .

3.1.35 Proposition. Let X be a set and let $\mathcal{U}$ be a partition of X . Then $\mathcal{U} = X/\sim$ for exactly one equivalence relation $\sim$ on X .

Proof. Define a relation $\sim$ by

$$x \sim y \iff \exists U \in \mathcal{U}, x \in U \text{ and } y \in U$$

for all $x, y \in X$. That is, $x \sim y$ if and only if x and y are elements of the same set of the partition. We check that $\sim$ is an equivalence relation.

- **Reflexivity.** Let $x \in X$. Then $x \in U$ for some $U \in \mathcal{U}$ since $\bigcup_{U \in \mathcal{U}} U = X$. Hence $x \sim x$.
- **Symmetry.** Let $x, y \in X$ and suppose $x \sim y$. Then there is some $U \in \mathcal{U}$ with $x \in U$ and $y \in U$. But then it is immediate that $y \sim x$.
- **Transitivity.** Let $x, y, z \in X$ and suppose that $x \sim y$ and $y \sim z$. Then there exist $U, V \in \mathcal{U}$ with $x, y \in U$ and $y, z \in V$. Thus $y \in U \cap V$. Since $\mathcal{U}$ is a partition of X , its elements are pairwise disjoint; thus if $U \neq V$ then $U \cap V = \emptyset$. Hence $U = V$. Thus $x \in U$ and $z \in U$, so $x \sim z$.

The definition of $\sim$ makes it immediate that $X/\sim = \mathcal{U}$.

To prove that $\sim$ is the only such relation, suppose $\approx$ is another equivalence relation on X for which $X/\approx = \mathcal{U}$. Then, given $x, y \in X$, we have:

$$\begin{aligned} x \sim y &\iff [x]_{\sim} = [y]_{\sim} \\ &\iff \exists U \in \mathcal{U}, x \in U \wedge y \in U \\ &\iff [x]_{\approx} = [y]_{\approx} \\ &\iff x \approx y \end{aligned}$$

So $\sim = \approx$. □

Partial orders

We will now study what it is for a relation to be *order-like*.

3.1.36 Definition. A relation R on a set X is a **partial order** if R is reflexive, antisymmetric and transitive. A set X together with a partial order on X is called a **partially ordered set**, or **poset** for short, and is denoted (X, R) . If (X, R) is a poset, we say X is **partially ordered** by R .

When we talk about partial orders, we usually use a symbol like ' $\preceq$ ' (`\preceq`) or ' $\sqsubseteq$ ' (`\sqsubseteq`). When we do so, we will use $\prec$ (`\prec`) to denote the relation

$$x \prec y \iff x \preceq y \text{ and } x \neq y$$

and likewise for $\sqsubset$ (`\sqsubset`).

3.1.37 Example. We have seen many examples of posets so far:

- Any of the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$, with the usual order relation $\leq$.
- Given a set X , its power set $\mathcal{P}(X)$ is partially ordered by $\subseteq$. Indeed:
 - ◊ **Reflexivity.** If $U \in \mathcal{P}(X)$ then $U \subseteq U$.
 - ◊ **Antisymmetry.** If $U, V \in \mathcal{P}(X)$ with $U \subseteq V$ and $V \subseteq U$, then $U = V$ by definition of set equality.
 - ◊ **Transitivity.** If $U, V, W \in \mathcal{P}(X)$ with $U \subseteq V$ and $V \subseteq W$, then $U \subseteq W$ by 1.2.19.
- The set $\mathbb{N}$ of natural numbers is partially ordered by divisibility (see Examples 3.1.14, 3.1.21 and 3.1.26). However, by Exercise 3.1.22, the set $\mathbb{Z}$ of integers is not partially ordered by divisibility, since divisibility is not antisymmetric on $\mathbb{Z}$.
- Any set X is partially ordered by its equality relation. This is called the **discrete order** on X .

We will study posets in greater depth in Section 5.2. However, there are some basic properties of posets we have already encountered that we are prepared to study already.

3.1.38 Definition. Let $(X, \preceq)$ be a poset. A **$\preceq$ -least element** of X is an element $\perp \in X$ (`\bot`) such that $\perp \preceq x$ for all $x \in X$. A **$\preceq$ -greatest element** of X is an element $\top \in X$ (`\top`) such that $x \preceq \top$ for all $x \in X$.

3.1.39 Example. Some examples of least and greatest elements that we have already seen are:

- In $(\mathbb{N}, \leq)$ the least element is 0; there is no greatest element.
- Let $n \in \mathbb{N}$ with $n > 0$. The least element of $([n], \leq)$ is 1 and the greatest element is n .
- $(\mathbb{Z}, \leq)$ has no greatest or least elements.

3.1.40 Exercise. Let X be a set. The poset $(\mathcal{P}(X), \subseteq)$ has a least element and a greatest element; find both.

3.1.41 Definition. Let $(X, \preceq)$ be a poset and let $A \subseteq X$. A **$\preceq$ -supremum** of A is an element $s \in X$ such that

- $a \preceq s$ for each $a \in A$; and
- If $s' \in X$ with $a \preceq s'$ for all $a \in A$, then $s \preceq s'$.

A **$\preceq$ -infimum** of A is an element $i \in X$ such that

- $i \preccurlyeq a$ for each $a \in A$; and
- If $i' \in X$ with $i' \preccurlyeq a$ for all $a \in A$, then $i' \preccurlyeq i$.

3.1.42 Example. The well-ordering principle states that if $U \subseteq \mathbb{N}$ is inhabited then U has a $\leq$ -infimum, and moreover the infimum of U is an element of U .

3.1.43 Exercise. Let X be a set, and let $U, V \in \mathcal{P}(X)$. Prove that the $\subseteq$ -supremum of $\{U, V\}$ is $U \cup V$, and the $\subseteq$ -infimum of $\{U, V\}$ is $U \cap V$.

3.1.44 Exercise. Let $a, b \in \mathbb{N}$. Find a supremum and an infimum of $\{a, b\}$ with respect to divisibility.

The following proposition proves that suprema and infima are unique, provided they exist.

3.1.45 Proposition. Let $(X, \preccurlyeq)$ is a poset, and let $A \subseteq X$.

- (i) If $s, s' \in X$ are suprema of A , then $s = s'$;
- (ii) If $i, i' \in X$ are infima of A , then $i = i'$.

Proof. Suppose s, s' are suprema of A . Then:

- $a \preccurlyeq s'$ for all $a \in A$, so $s' \preccurlyeq s$ since s is a supremum of A ;
- $a \preccurlyeq s$ for all $a \in A$, so $s \preccurlyeq s'$ since s' is a supremum of A .

Since $\preccurlyeq$ is antisymmetric, it follows that $s = s'$. This proves (i).

The proof of (ii) is almost identical and is left as an exercise to the reader. □

We will study posets in far more detail in Section 5.2; their true nature comes to light when we study how posets interact with each other.

3.2 Functions

A function from a set X to a set Y is a relation that behaves like a *transformation*: in some sense, it turns elements of X into elements of Y in a deterministic way.

3.2.1 Definition. Let X and Y be sets. A **function** f from X to Y is an assignment to each element $x \in X$ a unique element of Y , called the **value** of f at x and denoted $f(x)$. To denote the fact that f is a function from X to Y we write $f : X \rightarrow Y$ (**L^AT_EX** code: `f : X \to Y`). We call X the **domain** of f and Y the **codomain** of f . The set of all functions $X \rightarrow Y$ is denoted by Y^X .

3.2.2 Example. There are many examples of functions already at our fingertips. For instance:

- Let X be any set. There is a function $\text{id}_X : X \rightarrow X$ called the **identity function** on X , defined by $\text{id}_X(x) = x$ for all $x \in X$.
- Let $\mathcal{F}$ be the set of all finite sets in some fixed universe $\mathcal{U}$. There is a function $f : \mathcal{F} \rightarrow \mathbb{N}$ defined by $f(X) = |X|$.
- Fix $n \in \mathbb{Z}$ with $n > 0$. By the division theorem, there is a function $r_n : \mathbb{Z} \rightarrow \{0, 1, \dots, n-1\}$, where for all $a \in \mathbb{Z}$, $r_n(a)$ is the remainder of a upon division by n .
- In Definition 2.3.28, we defined the *totient* $\varphi(n)$ of an arbitrary integer n . This defines a function $\varphi : \mathbb{Z} \rightarrow \mathbb{N}$, called **Euler's totient function**.
- Given any set X , there is an **empty function** $\emptyset_X : \emptyset \rightarrow X$, which assigns... nothing, since $\emptyset$ has no elements.
- Addition defines a function **plus** : $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$, defined for $(x, y) \in \mathbb{N} \times \mathbb{N}$ by

$$\text{plus}(x, y) = x + y$$

Likewise for multiplication and exponentiation. If we replace $\mathbb{N}$ by $\mathbb{Z}$, we can also include subtraction.

- By the well-ordering principle, there is a function

$$\min : \mathcal{P}(\mathbb{N}) \setminus \{\emptyset\} \rightarrow \mathbb{N}$$

defined for $X \subseteq \mathbb{N}$ with $X \neq \emptyset$ by declaring $\min(X)$ to be the least element of X .

COMMON ERROR

Exposure to calculus at a young age often instills bad habits as far as notation and terminology are concerned. To clarify some points, for a function $f : X \rightarrow Y$:

- ‘ f ’ denotes the function itself;
- ‘ $f(x)$ ’ denotes the value of the function f at an element $x \in X$ (which should typically be introduced), and is an element of the codomain Y .

Thus, for example, it is incorrect to refer to $f(x)$ as a function—unless the codomain of the function is itself a set of functions!

Functions are often described as ‘machines’, which take as inputs elements of their domain and give as outputs elements of their codomain. This intuition leaves us thinking somehow as a function as being an *algorithm*, a set of rules for turning one thing into another. This is a useful intuition, but it leaves to question what it means for two functions to be the *same*... and as with sets (Definition 1.2.21) and relations (Definition 3.1.9), we return to the question: should functions be equal if they have the same definition (*intensional equality*), or if they produce the same ‘outputs’ on the same ‘inputs’ (*extensional equality*)? As with sets and relations, we take extensional equality as our notion of sameness.

3.2.3 Definition. Let $f : X \rightarrow Y$ and $g : U \rightarrow V$ be functions. We say f is **equal** to g , and write $f = g$, if $X = U$, $Y = V$ and $f(x) = g(x)$ for all $x \in X$.

3.2.4 Definition. Let X and Y be sets and let $f : X \rightarrow Y$. The **graph** of f is the subset Γ_f of $X \times Y$ defined by

$$\Gamma_f = \{(x, f(x)) \mid x \in X\}$$

3.2.5 Example. Graphs of functions with domain and codomain $\mathbb{R}$ are a central object of study in early calculus courses. For example, if $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined by $f(x) = x^2$ for all $x \in \mathbb{R}$, then

$$\Gamma_f = \{(x, x^2) \mid x \in \mathbb{R}\}$$

When plotted on a pair of axes, Γ_f is the shape of a parabola.

3.2.6 Theorem. Let X and Y be sets, and suppose $\Gamma \subseteq X \times Y$ is a set such that, for all $x \in X$, there is exactly one $y \in Y$ such that $(x, y) \in \Gamma$. Then Γ is the graph of exactly one function $f : X \rightarrow Y$.

3.2.7 Exercise. Prove Theorem 3.2.6.

For example, the set

$$\Gamma_1 = \{(1, 1), (2, 3), (3, 1)\}$$

is the graph of the function $f : [3] \rightarrow [3]$, defined by $f(1) = 1$, $f(2) = 3$ and $f(3) = 1$. However, the set

$$\Gamma_2 = \{(1, 1), (2, 3), (2, 1), (3, 1)\}$$

is not the graph of a function, since $(2, 3) \in \Gamma_2$ and $(2, 1) \in \Gamma_2$ but $3 \neq 1$.

3.2.8 Exercise. Let X and Y be finite sets, with $|X| = m$ and $|Y| = n$. Prove that $|Y^X| = n^m$.

Images and preimages

3.2.9 Definition. Let $f : X \rightarrow Y$ be a function, and let $U \subseteq X$. The **image** of U under f is the set $f_*(U) \subseteq Y$, defined by

$$f_*(U) = \{y \in Y \mid \exists x \in U, f(x) = y\} = \{f(x) \mid x \in U\}$$

That is, the image of U under f is the set of elements of Y which are of the form $f(x)$ for some $x \in U$.

3.2.10 Example. Let $f : \mathbb{Z} \rightarrow \mathbb{Z}$ be the function defined by $f(x) = x^2$ for all $x \in X$. Then

- $f_*(\{1, 2, 3, 4\}) = \{1, 4, 9, 16\}$;
- $f_*(\mathbb{Z}) = f_*(\mathbb{N})$. The $\supseteq$ direction is immediate; for the $\subseteq$ direction, note that if $a \in \mathbb{Z} \setminus \mathbb{N}$ then $-a \in \mathbb{N}$, and $f(-a) = (-a)^2 = a^2 = f(a)$.

3.2.11 Example. Let $f : X \rightarrow Y$ be a function and let $U, V \subseteq X$. Then

- $f_*(U \cup V) \subseteq f_*(U) \cup f_*(V)$. To see this, let $y \in f_*(U \cup V)$. Then $y = f(x)$ for some $x \in U \cup V$ by definition of image. By definition of union, $x \in U$ or $x \in V$.
 - ◊ If $x \in U$ then $y \in f_*(U)$, and hence $y \in f_*(U) \cup f_*(V)$.
 - ◊ If $x \in V$ then $y \in f_*(V)$, and hence $y \in f_*(U) \cup f_*(V)$.

In any case, $y \in f_*(U) \cup f_*(V)$.

- $f_*(U) \cup f_*(V) \subseteq f_*(U \cup V)$. To see this, let $y \in f_*(U) \cup f_*(V)$. By definition of union, $y \in f_*(U)$ or $y \in f_*(V)$.
 - ◊ If $y \in f_*(U)$ then $y = f(x)$ for some $x \in U$; but then $x \in U \cup V$, so $y \in f_*(U \cup V)$.
 - ◊ If $y \in f_*(V)$ then $y = f(x)$ for some $x \in V$; but then $x \in U \cup V$, so $y \in f_*(U \cup V)$.

In any case, $y \in f_*(U \cup V)$.

It follows that $f_*(U \cup V) = f_*(U) \cup f_*(V)$.

3.2.12 Exercise. Let $f : X \rightarrow Y$ be a function. Show that $f_*(\emptyset) = \emptyset$, and give an example to demonstrate that it is not necessarily true that $f_*(X) = Y$.

3.2.13 Definition. Let $f : X \rightarrow Y$ be a function, and let $V \subseteq Y$. The **preimage** of V under f is the set $f^*(V) \subseteq X$, defined by

$$f^*(V) = \{x \in X \mid f(x) \in V\}$$

3.2.14 Example. Let $f : \mathbb{Z} \rightarrow \mathbb{Z}$ be the function defined by $f(x) = x^2$ for all $x \in \mathbb{Z}$. Then

- $f^*(\{1, 4, 9, 16\}) = \{-4, -3, -2, -1, 1, 2, 3, 4\}$;
- $f^*([16]) = \{-4, -3, -2, -1, 1, 2, 3, 4\}$ too, since the other elements of $[16]$ are not perfect squares, and hence not of the form $f(x)$ for $x \in \mathbb{Z}$;
- $f^*(\mathbb{N}) = \mathbb{Z}$, since for any $x \in \mathbb{Z}$ we have $f(x) \in \mathbb{N}$.

3.2.15 Example. Let $f : X \rightarrow Y$ be a function and let $U, V \subseteq Y$. Then

- $f^*(U \cup V) \subseteq f^*(U) \cup f^*(V)$. To see this, let $x \in f^*(U \cup V)$. Then $f(x) \in U \cup V$. By definition of union, $f(x) \in U$ or $f(x) \in V$.

◊ If $f(x) \in U$ then $x \in f^*(U)$, so $x \in f^*(U) \cup f^*(V)$.

◊ If $f(x) \in V$ then $x \in f^*(V)$, so $x \in f^*(U) \cup f^*(V)$.

In any case, $x \in f^*(U) \cup f^*(V)$.

- $f^*(U) \cup f^*(V) \subseteq f^*(U \cup V)$. To see this, let $x \in f^*(U) \cup f^*(V)$. By definition of union, $x \in f^*(U)$ or $x \in f^*(V)$.

◊ If $x \in f^*(U)$ then $f(x) \in U$, so $f(x) \in U \cup V$, and hence $x \in f^*(U \cup V)$.

◊ If $x \in f^*(V)$ then $f(x) \in V$, so $f(x) \in U \cup V$, and hence $x \in f^*(U \cup V)$.

In any case, $x \in f^*(U \cup V)$.

It follows that $f^*(U \cup V) = f^*(U) \cup f^*(V)$.

3.2.16 Exercise. Let $f : X \rightarrow Y$ be a function. Show that $f^*(\emptyset) = \emptyset$ and that $f^*(Y) = X$.

Composition and restriction

When the codomain of a function f and the domain of another function g are equal, we can apply g to the values of f . The effect of doing this defines a new function, whose domain is the same as that of f and whose codomain is the same as that of g .

3.2.17 Definition. Let X, Y, Z be sets, and let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be functions. The **composite** of f and g is the function $g \circ f : X \rightarrow Z$ (**L^AT_EX** code: `g \circ f`) defined by

$$(g \circ f)(x) = g(f(x))$$

for all $x \in X$.

Note that in order for the composite $g \circ f$ to be defined, the codomain of f must be equal to the domain of g .

ASIDE

Given sets X, Y, Z , the composition operation defines a function

$$\text{comp} : Y^X \times Z^Y \rightarrow Z^X$$

defined by $\text{comp}(f, g) = g \circ f$ for all $f \in Y^X$ and $g \in Z^Y$.

The following example demonstrates that composition depends on the order that the functions are composed, even when their domains and codomains all coincide:

3.2.18 Example. Let $f : \mathbb{Z} \rightarrow \mathbb{Z}$ and $g : \mathbb{Z} \rightarrow \mathbb{Z}$ be defined by

$$f(x) = x + 2 \quad \text{and} \quad g(x) = x^2$$

for all $x \in \mathbb{Z}$. Then for all $x \in \mathbb{Z}$, we have

$$(g \circ f)(x) = g(f(x)) = [f(x)]^2 = (x + 2)^2 = x^2 + 4x + 4$$

and

$$(f \circ g)(x) = f(g(x)) = g(x) + 2 = x^2 + 2$$

3.2.19 Exercise. Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be functions.

- Prove that $(g \circ f)_*(U) = g_*(f_*(U))$ for all $U \subseteq X$.
- Prove that $(g \circ f)^*(V) = f^*(g^*(V))$ for all $V \subseteq Z$.

Sometimes we may be presented with a function on a larger set than we'd like to think about. For example, we might want to consider the behaviour of the function $f : \mathbb{Z} \rightarrow \mathbb{N}$ defined by $f(n) = n^2$ only when $n \geq 0$. If we make the domain smaller, we're defining a new function.

3.2.20 Definition. Let $f : X \rightarrow Y$ be a function, and let $U \subseteq X$. The **restriction** of f to U is the function $f \upharpoonright U : U \rightarrow Y$ (read ' f restricted to U ') ([L^AT_EX code: `\upharpoonright`](#)), defined by $(f \upharpoonright U)(u) = f(u)$ for all $u \in U$.

Equivalently, $f \upharpoonright U = f \circ i$, where $i : U \rightarrow X$ is defined by $i(u) = u$ for all $u \in U$.

3.2.21 Example. Let $f : \mathbb{Z} \rightarrow \mathbb{N}$ be defined by $f(n) = n^2$. Then $f \upharpoonright \mathbb{N}$ is a function $\mathbb{N} \rightarrow \mathbb{N}$ with the same equational definition as f .

3.2.22 Exercise. Let $f : X \rightarrow Y$ be a function, let $U \subseteq X$ and let $V \subseteq Y$. Prove that

$$(f \upharpoonright U)^*(V) = f^*(V) \cap U$$

Inverse functions

Thinking of functions as types of transformation, it is often useful to know when they can be, in some sense, done in reverse.

One way we might wish to reverse a function is to be able to recover an element of the domain from the knowledge of its value in the codomain. That is, we might want to be able to *undo* f .

3.2.23 Example. Suppose we want to encode English words as natural numbers. Formally, this can be done by defining a function $f : X \rightarrow \mathbb{N}$, where X is a set of English words. For this to be of much use, we'd like to be able to recover a word $x \in X$ from its encoding $f(x) \in \mathbb{N}$.

We could do this as follows: define $g : \mathbb{N} \rightarrow X$ by declaring $g(n) = x$ if the natural number $n \in \mathbb{N}$ encodes a word x , and $g(n) = \text{'donkey'}$ otherwise. By definition of g we have $g(f(x)) = x$ for all $x \in X$ —the values of g at natural numbers which don't encode English words is irrelevant.

3.2.24 Definition. Let $f : X \rightarrow Y$ be a function. A **left inverse** for f is a function $g : Y \rightarrow X$ such that $g \circ f = \text{id}_X$.

3.2.25 Example. A function may or may not have a left inverse. For example:

- The function $f : \mathbb{Z} \rightarrow \mathbb{Z}$ defined by $f(x) = 3x + 1$ has a left inverse. To see this, define $g : \mathbb{Z} \rightarrow \mathbb{Z}$ as follows:
 - ◊ If $a \in \mathbb{Z}$ and $a \equiv 1 \pmod{3}$, then $a = 3k + 1$ for a exactly one $k \in \mathbb{Z}$. Define $g(a) = k$ for this value of k .
 - ◊ If $a \in \mathbb{Z}$ and $a \not\equiv 1 \pmod{3}$, then define $g(a) = 0$.

Then g is a left inverse for f . Indeed, if $x \in \mathbb{Z}$ then $f(x) = 3x + 1 \equiv 1 \pmod{3}$. Thus $g(f(x))$ is the integer k such that $f(x) = 3k + 1$... but this means that $k = x$, so $g(f(x)) = x$.

- The function $f : \mathbb{Z} \rightarrow \mathbb{Z}$ defined by $f(x) = x^2$ has no left inverse. To see this, let $g : \mathbb{Z} \rightarrow \mathbb{Z}$ be a function. If g is a left inverse for f then since $f(-1) = (-1)^2 = 1 = 1^2 = f(1)$, we have

$$-1 = g(f(-1)) = g(1) = g(f(1)) = 1$$

and so $-1 = 1$. This is nonsense, so f must have no left inverse.

3.2.26 Exercise. Define $f : \mathbb{Z} \rightarrow \mathbb{N}$ be defined by

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

Find a left inverse for f .

3.2.27 Definition. A function $f : X \rightarrow Y$ is **injective** if, for all $x, x' \in X$,

$$f(x) = f(x') \quad \Rightarrow \quad x = x'$$

An injective function is also called an **injection**.

By contraposition, $f : X \rightarrow Y$ being injective is equivalent to saying that if $x, x' \in X$ and $x \neq x'$, then $f(x) \neq f(x')$.

3.2.28 Proposition. Let $f : X \rightarrow Y$ be a function and suppose X is inhabited. Then f has a left inverse if and only if f is injective.

Proof. We prove each direction separately.

- ($\Rightarrow$) Suppose f has a left inverse $g : Y \rightarrow X$. Suppose $x, x' \in X$ with $f(x) = f(x')$. Then

$$\begin{aligned} x &= g(f(x)) && \text{since } g \text{ is a left inverse for } f \\ &= g(f(x')) && \text{since } f(x) = f(x') \\ &= x' && \text{since } g \text{ is a left inverse for } f \end{aligned}$$

Hence $f(x) = f(x')$ implies $x = x'$, so that f is injective.

- ($\Leftarrow$) Suppose f is injective. Define $g : Y \rightarrow X$ as follows. Since X is inhabited, it has at least one element—let $x_0 \in X$ be such an element. Now fix $y \in Y$; we must define the value of $g(y)$:

- ◊ If $y \in f_*(X)$, let $x \in X$ be such that $y = f(x)$. There is exactly one such value, since if $x' \in X$ and $y = f(x')$, then $f(x) = f(x')$, so $x = x'$. Thus we can define $g(y) = x$ for this single value of x .
- ◊ If $y \notin f_*(X)$, then define $g(y) = x_0$.

Thus we have defined a function $g : Y \rightarrow X$ in such a way that $g(f(x)) = x$ for all $x \in X$. So f has a left inverse.

□

Another way we might wish to reverse a function is to be able to find, for a given element of its codomain, an element of the domain mapping to that element.

3.2.29 Example. Let $\mathbb{S}^1 = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = 1\}$. Geometrically, $\mathbb{S}^1$ is the circle in the (real) plane with radius 1, centred at the origin. There is a function $p : \mathbb{R} \rightarrow \mathbb{S}^1$ given by

$$p(t) = (\cos 2\pi t, \sin 2\pi t) \text{ for all } t \in \mathbb{R}$$

For example, $p(0) = (1, 0)$ and $p(\frac{1}{2}) = (-1, 0)$. Intuitively, for $n \in \mathbb{Z}$, as t traces the real numbers between n and $n + 1$, $g(t)$ traces the circle anticlockwise exactly once.

Given a point (x, y) on the circle, we seek a value of t for which $p(t) = (x, y)$. So define $f : \mathbb{S}^1 \rightarrow \mathbb{R}$ by

$$f(x, y) = \begin{cases} \frac{1}{2\pi} \tan^{-1} \frac{y}{x} & \text{if } x > 0 \\ \frac{1}{2\pi} (\tan^{-1} \frac{y}{x} + \pi) & \text{if } x < 0 \\ \frac{1}{4} & \text{if } (x, y) = (0, 1) \\ \frac{3}{4} & \text{if } (x, y) = (0, -1) \end{cases}$$

It is a straightforward exercise in trigonometry to check that $p(f(x, y)) = (x, y)$ in each case. That is, $f(x, y)$ is a real number mapped by p to (x, y) .

3.2.30 Definition. Let $f : X \rightarrow Y$ be a function. A **right inverse** for f is a function $g : Y \rightarrow X$ such that $f \circ g = \text{id}_Y$.

3.2.31 Example. A function may or may not have a right inverse. For example:

- The function $f : \mathbb{Z} \rightarrow \mathbb{Z}$ defined by $f(x) = 3x + 1$ does not have a left inverse. Indeed, let $g : \mathbb{Z} \rightarrow \mathbb{Z}$ be a function. If g is a right inverse for f , then we should have $f(g(0)) = 0$. But $f(g(0)) = 3g(0) + 1$, so $f(g(0)) \equiv 1 \pmod{3}$. Since $0 \not\equiv 1 \pmod{3}$, we must have $f(g(0)) \neq 0$. So g is not a right inverse for f .
- The absolute value function $v : \mathbb{Z} \rightarrow \mathbb{N}$ defined by $v(n) = |n|$ for all $n \in \mathbb{Z}$ has a right-inverse. Indeed, define $i : \mathbb{N} \rightarrow \mathbb{Z}$ by $i(n) = n$ for all $n \in \mathbb{N}$. Then

$$v(i(n)) = |i(n)| = |n| = n$$

for all $n \in \mathbb{N}$, so that $v \circ i = \text{id}_{\mathbb{N}}$.

3.2.32 Definition. A function $f : X \rightarrow Y$ is **surjective** if, for all $y \in Y$, there exists $x \in X$ such that $f(x) = y$. A surjective function is also called a **surjection**.

3.2.33 Proposition. Let $f : X \rightarrow Y$ be a function. Then f has a right inverse if and only if f is surjective.

Proof. We prove each direction separately.

- ($\Rightarrow$) Suppose f has a right inverse $g : Y \rightarrow X$. Fix $y \in Y$, and let $x = g(y)$. Then $f(x) = f(g(y)) = y$ since g is a right inverse for f . Hence f is surjective.
- ($\Leftarrow$) Suppose f is surjective. Define $g : Y \rightarrow X$ as follows. For each $y \in Y$ there exists $x \in X$ such that $f(x) = y$; fix some $x \in X$ such that $f(x) = y$, and define $g(y) = x$. This defines a function $g : Y \rightarrow X$, and by construction it follows that if $y \in Y$ then $f(g(y)) = y$, so that g is a right inverse for f .

So we're done. □

ASIDE

Although we didn't mention it, the $\Leftarrow$ direction of Proposition 3.2.33 made use of the *axiom of choice*, and is in fact equivalent to it. In the case when Y is an infinite set, we had to make infinitely many choices in the step where we chose, for each $y \in Y$, an (arbitrary) element $x \in X$ such that $f(x) = y$ —the axiom of choice is what enables us to do make infinitely many choices simultaneously in this way.

Although most mathematicians accept the axiom of choice, and in many cases even *rely* on it, some mathematicians consider it to be one step too far removed from reality and dislike its non-constructive character. This isn't all bad, though: there is a wealth of fascinating research by mathematicians working in foundational areas like set theory concerning the consequences of rejecting the axiom of choice.

3.2.34 Definition. Let $f : X \rightarrow Y$ be a function. An **inverse**, or **two-sided inverse**, for f is a function $g : Y \rightarrow X$ which is both a left inverse and a right inverse for f .

3.2.35 Theorem. Let $f : X \rightarrow Y$ be a function. If f has a left inverse and a right inverse, then they are equal.

Proof. Let $g : Y \rightarrow X$ be a left inverse for f , and let $h : Y \rightarrow X$ be a right inverse for f . We will prove that $g = h$, so that g is an inverse for f . So fix $y \in Y$. Then

$$\begin{aligned} g(y) &= g(f(h(y))) && \text{since } y = f(h(y)) \\ &= h(y) && \text{since } g(f(z)) = z \text{ for all } z \in X \\ &&& \dots \text{in particular, when } z = h(y) \end{aligned}$$

So by definition of function equality, we have $g = h$. □

3.2.36 Definition. A function $f : X \rightarrow Y$ is **bijective** if it is injective and surjective. A bijective function is also called a **bijection**.

3.2.37 Example. • There are two bijections $f_1, f_2 : \{1, 2\} \rightarrow \{A, B\}$, defined by

$$f_1(1) = A, f_1(2) = B; \quad f_2(1) = B, f_2(2) = A$$

- Given any finite set X , a permutation of X gives rise to a bijection $f : X \rightarrow X$. Indeed, writing $X = \{x_1, x_2, \dots, x_n\}$, where $n = |X|$, we can define $f(x_i)$ to be the i^{th} component in the list. More generally, we can (re)define a **permutation** of a general set X to simply be a bijection $X \rightarrow X$.

3.2.38 Exercise. Define $f : \mathbb{N} \rightarrow \mathbb{N}$ and $g : \mathbb{Z} \rightarrow \mathbb{Z}$ by $f(x) = x + 1$ for all $x \in \mathbb{N}$ and $g(x) = x + 1$ for all $x \in \mathbb{Z}$. For each, determine if it is bijective or not bijective.

A consequence of Theorem 3.2.35 is the following result:

3.2.39 Corollary. Let $f : X \rightarrow Y$ be a function. Then f has a two-sided inverse if and only if f is bijective.

Proof. We prove each direction separately. First consider the case when X is inhabited.

($\Rightarrow$) Suppose f has an inverse $g : Y \rightarrow X$. Then g is a left inverse for f , so f is injective by Proposition 3.2.28; and g is a right inverse for f , so f is surjective by Proposition 3.2.33. Hence f is bijective.

($\Leftarrow$) Suppose f is bijective. Then it is injective, so has a left inverse by 3.2.28, and surjective, so has a right inverse by 3.2.33. By Theorem 3.2.35, f has an inverse.

If $X = \emptyset$ then both conditions imply that $Y = \emptyset$. Indeed, if f has an inverse $g : Y \rightarrow X$ and there exists $y \in Y$, then $g(y) \in \emptyset$; and if f is bijective and there exists $y \in Y$, then there exists $x \in \emptyset$ such that $f(x) = y$ by surjectivity of f . But there is only one function $\emptyset \rightarrow \emptyset$, namely the identity function $\text{id}_{\emptyset} : \emptyset \rightarrow \emptyset$. This is vacuously injective and surjective, and is its own inverse. $\square$

Some basic results about bijections that will prove useful later are contained in the following three exercises.

3.2.40 Exercise. Let X be a set. The identity function $\text{id}_X : X \rightarrow X$ is a bijection.

3.2.41 Exercise. Let $f : X \rightarrow Y$ be a bijection. Then $f^{-1} : Y \rightarrow X$ is a bijection.

3.2.42 Exercise. Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be bijections. Then $g \circ f : X \rightarrow Z$ is a bijection and find its inverse.

3.3 Problems

Chapter 4

Finite and infinite sets

This chapter focuses on the set of **natural numbers**, which is remarkable for a number of reasons. First we will use its properties to allow definitions by **recursion** and to exploit a new proof technique called **induction**. This will be used widely in this course and in pretty much every mathematics course you take from now on.

We then focus our attention to **combinatorics**, which is the study of discrete structures. It is a huge area of mathematics with a lot of active research, and the methods used have applications in statistics, optimisation, management science, algorithm design, data mining... the list goes on. Our entry-level focus will be on the combinatorics of *finite sets*; specifically, we will apply what we learnt in Chapter 1 to find and compare sizes of finite sets with different specifications.

The first combinatorics topic we will discuss is **counting**. Counting principles are tools for finding the number of elements in a finite set which is built out of other finite sets in one way or another. This is usually done through the lens of counting the number of ways of doing something, or the number of things of a certain kind.

We will then introduce **binomial coefficients** and **factorials**. These arise in combinatorics, number theory and algebra, and are useful for expressing large numbers in a meaningful way.

4.1 Counting principles

Counting principles provide ways of measuring the number of elements of a finite set. This leads to useful formulae for finding the size of large finite sets built up from smaller ones, say, as the product or union of the smaller sets. Before we can begin discussing the sizes of finite sets, we need to make precise what we mean by ‘finite’ and ‘size’. First we’ll introduce some useful notation:

4.1.1 Definition. For $n \in \mathbb{N}$, define

$$[n] = \{k \in \mathbb{N} \mid 1 \leq k \leq n\}$$

That is, $[0] = \emptyset$ and, if $n \geq 1$, then $[n] = \{1, 2, \dots, n\}$.

4.1.2 Definition. A set X is **finite** if there is a bijection $[n] \rightarrow X$ for some $n \in \mathbb{N}$, called the **size** of X . Write $|X|$ for the size of X . If X is not finite we say it is **infinite**.

In more intuitive terms: a set X is finite if the number of elements of X is a natural number; the size X is simply the number of elements of X .

Finding a bijection $f : [n] \rightarrow X$ is equivalent to expressing X as an indexed family with index set $[n]$: writing

$$X = \{x_1, x_2, \dots, x_n\}$$

is equivalent to defining a function $f : [n] \rightarrow X$ by $f(i) = x_i$ for all $i \in [n]$; then f is a bijection if and only if $|X| = n$. Since they are equivalent, we will use bijections $[n] \rightarrow X$, and indexings of X by $[n]$, interchangeably.

4.1.3 Example. For each $n \in \mathbb{N}$ the set $[n]$ is finite, and $|[n]| = n$. This is because the identity function $\text{id}_{[n]} : [n] \rightarrow [n]$ is a bijection.

4.1.4 Lemma. Every inhabited finite set X of natural numbers has a greatest element.

Proof. We prove this lemma by induction on the size of X . That is, by induction on $n \geq 1$, we show that if $|X| = n$ then X has a greatest element.

- **(BC)** If $|X| = 1$ then $X = \{x\}$ for some $x \in \mathbb{N}$. Since x is the only element of X , it is certainly the greatest element!
- **(IS)** Let $n \in \mathbb{N}$ and suppose that every set of natural numbers of size n has a greatest element **(IH)**.

Let $X = \{x_1, x_2, \dots, x_n, x_{n+1}\}$. We wish to show that X has a greatest element.

To do this, let $Y = X \setminus \{x_{n+1}\}$. Since $|Y| = n$, by **(IH)** it has a greatest element, say x_i . If $x_i > x_{n+1}$ then x_i is the greatest element of X ; otherwise, x_{n+1} is the greatest element of X . In either case, X has a greatest element.

By induction, we're done. □

4.1.5 Theorem. The set $\mathbb{N}$ is infinite.

Proof. We proceed by contradiction. Suppose $\mathbb{N}$ is finite. Then $|\mathbb{N}| = n$ for some $n \in \mathbb{N}$, and hence $\mathbb{N}$ is either empty (nonsense) or, by Lemma 4.1.4, it has a greatest element g . But $g + 1 \in \mathbb{N}$ since every natural number has a successor, contradicting maximality of g . Hence $\mathbb{N}$ is infinite. □

For the remainder of this section we will only concern ourselves with finite sets; we will begin to worry about infinite sets and their sizes in Section 4.3.

4.1.6 Theorem. Let X and Y be finite sets. Then the following sets are finite:

- (a) Any subset $Z \subseteq X$, and $|Z| \leq |X|$;
- (b) $X \cap Y$;
- (c) $X \cup Y$, and $|X \cup Y| = |X| + |Y| - |X \cap Y|$;
- (d) $X \times Y$, and $|X \times Y| = |X| \cdot |Y|$.

Proof. These are all immediate if $X = \emptyset$ or $Y = \emptyset$, so assume that X and Y are inhabited.

- (a) Let $X = \{x_1, \dots, x_n\}$ and suppose $Z \subseteq X$. Then $Z = \{x_{i_1}, \dots, x_{i_k}\}$ for some subset $\{i_1, \dots, i_k\} \subseteq [n]$. But this is precisely an indexing of Z by $[k]$ for some $k \leq n$, demonstrating that Z is finite.
- (b) $X \cap Y \subseteq X$, so this is immediate from (a).
- (c) Let $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$. By (b), the set $X \cap Y$ is finite, so we can write $X \cap Y = \{z_1, \dots, z_k\}$. By rearranging elements, we may suppose that $z_i = x_i = y_i$ for all $1 \leq i \leq k$. But this implies that

$$X \cup Y = \underbrace{\{z_1, \dots, z_k\}}_{X \cap Y} \underbrace{\{x_{k+1}, \dots, x_n\}}_{X \setminus (X \cap Y)} \underbrace{\{y_{k+1}, \dots, y_m\}}_{Y \setminus (X \cap Y)}$$

Write $z_i = x_i$ for $k+1 \leq i \leq n$ and $z_i = y_{i-n+k}$ for $n+1 \leq i \leq m+n-k$. Then the foregoing expression becomes

$$X \cup Y = \underbrace{\{z_1, \dots, z_k\}}_{X \cap Y} \underbrace{\{z_{k+1}, \dots, z_n\}}_{X \setminus (X \cap Y)} \underbrace{\{z_{n+1}, \dots, z_{n+m-k}\}}_{Y \setminus (X \cap Y)}$$

so $X \cup Y$ is finite and $|X \cup Y| = n + m - k = |X| + |Y| - |X \cap Y|$.

(d) Let $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$. Then

$$X \times Y = \left\{ \begin{array}{cccc} (x_1, y_1), & (x_1, y_2), & \dots, & (x_1, y_m), \\ (x_2, y_1), & (x_2, y_2), & \dots, & (x_2, y_m), \\ \vdots & \vdots & \ddots & \vdots \\ (x_n, y_1), & (x_n, y_2), & \dots, & (x_n, y_m) \end{array} \right\}$$

This expresses $X \times Y$ as an $n \times m$ grid, which has nm components. Labelling these from top-left to bottom-right yields an expression of $X \times Y$ as a set indexed by $[nm]$. Hence $X \times Y$ is finite and $|X \times Y| = nm = |X| \cdot |Y|$.

□

With this theorem under our belt, we can now explore its consequences, known as **counting principles**.

Multiplication principle

Intuitively, the multiplication principle says that if you want to do two things and there are a ways of doing one and b ways of doing the other, then there are $a \times b$ ways of doing both. More precisely, it says that the size of a product of sets is the product of the sizes of the individual sets.

4.1.7 Corollary (Multiplication principle). Let $\{X_1, \dots, X_n\}$ be a family of finite sets, with $n \geq 1$. Then $\prod_{i=1}^n X_i$ is finite, and

$$\left| \prod_{i=1}^n X_i \right| = |X_1| \cdot |X_2| \cdot \dots \cdot |X_n|$$

Proof. We proceed by induction on n .

- (BC) When $n = 1$, $\prod_{i=1}^1 X_i = X_1$, so the result is immediate.
- (IS) Let $\{X_1, \dots, X_n, X_{n+1}\}$ be a family of sets, and suppose

$$\left| \prod_{i=1}^n X_i \right| = |X_1| \cdot |X_2| \cdot \dots \cdot |X_n| \quad \text{---(IH)}$$

Now

$$\prod_{i=1}^{n+1} X_i = \left(\prod_{i=1}^n X_i \right) \times X_{n+1}$$

so by Theorem 4.1.6(d), it follows that

$$\left| \prod_{i=1}^{n+1} X_i \right| = \left| \prod_{i=1}^n X_i \right| \cdot |X_{n+1}|$$

and by (IH) it follows that

$$\left| \prod_{i=1}^{n+1} X_i \right| = |X_1| \cdot |X_2| \cdot \cdots \cdot |X_n| \cdot |X_{n+1}|$$

By induction, we're done. □

4.1.8 Example. You go to an ice cream stand selling the following flavours:

vanilla, strawberry, chocolate, rum and raisin, mint choc chip, toffee crunch

You can have your ice cream in a tub, a regular cone or a chocolate-covered cone. You can only afford a single scoop. How many options do you have?

Let F be the set of flavours and let C be the set of cone options. By simply counting the options presented above, we see that $|F| = 6$ and $|C| = 3$. Selecting an ice cream flavour and cone option is the same as selecting an element of $F \times C$, for example (chocolate, regular cone) is one such element. Thus the number of options you have is

$$|F \times C| = |F| \cdot |C| = 6 \cdot 3 = 18$$

4.1.9 Example. We will count the number of subsets of $[7]$.

Specifying a subset $S \subseteq [7]$ is the same as choosing, for each element $n \in [7]$, whether $n \in S$ or $n \notin S$ ('in or out'). For example, the subset $\{1, 2, 6\} \subseteq [7]$ corresponds to the choices

1 in, 2 in, 3 out, 4 out, 5 out, 6 in, 7 out

Thus specifying a subset of $[7]$ is the same as specifying an element of the set $\prod_{i=1}^7 \{\text{in, out}\}$. Hence the number of subsets is

$$\left| \prod_{i=1}^7 \{\text{in, out}\} \right| = \underbrace{|\{\text{in, out}\}| \cdot |\{\text{in, out}\}| \cdot \cdots \cdot |\{\text{in, out}\}|}_{7 \text{ times}} = 2^7 = 128$$

4.1.10 Exercise. Prove that the number of subsets of a finite set X is $2^{|X|}$.

4.1.11 Exercise. Since September 2001, number plates on the island of Great Britain have taken the form **XX NN XXX**, where each **X** can be any letter of the alphabet except for 'I', 'Q' or 'Z', and **NN** is the last two digits of the year.¹ How many possible number plates are there? Number plates of vehicles registered in the region of Yorkshire begin with the letter 'Y'. How many Yorkshire number plates can be issued in a given year?

¹This is a slight simplification of what is really the case, but let's not concern ourselves with *too* many details!

Addition principle

Intuitively, the addition principle says that if you want to do one of two things (but not both) and there are a ways of doing one and b ways of doing the other, then there are $a + b$ ways of doing both. More precisely, it says that the size of a union of sets with no elements in common is the sum of the sizes of the individual sets.

4.1.12 Definition. Sets X and Y are **disjoint** if $X \cap Y = \emptyset$. More generally, a family of sets $\{X_i \mid i \in I\}$ is **pairwise disjoint** if $X_i \cap X_j = \emptyset$ for all $i, j \in I$ with $i \neq j$. A **partition** of a set X is a pairwise disjoint family of sets $\{X_i \mid i \in I\}$ such that $\bigcup_{i \in I} X_i = X$.

4.1.13 Corollary. If X and Y are disjoint then $|X \cup Y| = |X| + |Y|$.

Proof. This is immediate from Theorem 4.1.6(c), since $|X \cap Y| = |\emptyset| = 0$. $\square$

4.1.14 Corollary (Addition principle). Let $\{X_1, \dots, X_n\}$ be a pairwise disjoint family of finite sets, with $n \geq 1$. Then $\bigcup_{i=1}^n X_i$ is finite, and

$$\left| \bigcup_{i=1}^n X_i \right| = |X_1| + |X_2| + \dots + |X_n|$$

4.1.15 Exercise. Prove Corollary 4.1.14. The proof follows the same pattern as that of the multiplication principle (Corollary 4.1.7).

4.1.16 Example. We will count the number of inhabited subsets of $[7]$ which either contain only even numbers, or contain only odd numbers.

Let O denote the set of inhabited subsets of $[7]$ containing only odd numbers, and let E denote the set of inhabited subsets of $[7]$ containing only even numbers. We need to compute $|O \cup E|$.

- A subset containing only odd numbers must be a subset of $\{1, 3, 5, 7\}$. By Exercise 4.1.10 there are $2^4 = 16$ such subsets. Thus the number of *inhabited* subsets of $[7]$ containing only odd numbers is 15, since we must exclude the empty set. That is, $|O| = 15$.
- A subset containing only even numbers must be a subset of $\{2, 4, 6\}$. Again by Exercise 4.1.10 there are $2^3 = 8$ such subsets. Hence there are 7 inhabited subsets of $[7]$ containing only even numbers. That is, $|E| = 7$.

Clearly O and E are disjoint: if a set contains an even number then it cannot consist entirely of odd numbers, and vice versa. Hence

$$|O \cup E| = |O| + |E| = 15 + 7 = 22$$

And here they are:

$$\begin{array}{cccccc}
\{1\} & \{3\} & \{5\} & \{7\} & \{1, 3\} & \{1, 5\} \\
\{1, 7\} & \{3, 5\} & \{3, 7\} & \{5, 7\} & \{1, 3, 5\} & \{1, 3, 7\} \\
\{1, 5, 7\} & \{3, 5, 7\} & \{1, 3, 5, 7\} & \{2\} & \{4\} & \{6\} \\
\{2, 4\} & \{2, 6\} & \{4, 6\} & \{2, 4, 6\} & &
\end{array}$$

4.1.17 Exercise. Pick your favourite integer $n > 1000$. For this value of n , how many inhabited subsets of $[n]$ contain either only even or only odd numbers? (You need not evaluate exponents.)

Inclusion–exclusion principle

The addition principle is useful only for counting *disjoint* sets, i.e. ones with no overlaps. We saw in Theorem 4.1.6(c) how to compute the size of a union of two sets which do overlap:

$$|X \cup Y| = |X| + |Y| - |X \cap Y|$$

So far so good. But what if we have three or four sets instead of just two?

4.1.18 Exercise. Let X, Y, Z be sets. Show that

$$|X \cup Y \cup Z| = |X| + |Y| + |Z| - |X \cap Y| - |X \cap Z| - |Y \cap Z| + |X \cap Y \cap Z|$$

Let W be another set. Derive a similar formula for $|W \cup X \cup Y \cup Z|$.

4.1.19 Theorem (Inclusion–exclusion principle). Let $n \geq 2$ and let $X_1, X_2, \dots, X_n$ be sets. Then

$$\left| \bigcup_{i=1}^n X_i \right| = \sum_{J \subseteq [n]} (-1)^{|J|+1} \left| \bigcap_{j \in J} X_j \right|$$

where for the purposes of the formula we take $\bigcap_{j \in \emptyset} X_j = \emptyset$.

Proof. We proceed by induction.

- **(BC)** The proof for the case $n = 2$ was Theorem 4.1.6(c).
- **(IS)** Fix $n \geq 2$ and suppose, for any sets $X_1, X_2, \dots, X_n$, that

$$\left| \bigcup_{i=1}^n X_i \right| = \sum_{J \subseteq [n]} (-1)^{|J|+1} \left| \bigcap_{j \in J} X_j \right| \quad \text{---(IH)}$$

We need to prove that, for any sets $X_1, X_2, \dots, X_n, X_{n+1}$, that

$$\left| \bigcup_{i=1}^{n+1} X_i \right| = \sum_{J \subseteq [n+1]} (-1)^{|J|+1} \left| \bigcap_{j \in J} X_j \right|$$

Write $U = \bigcup_{i=1}^n X_i$. We know that

$$|U \cup X_{n+1}| = |U| + |X_{n+1}| - |U \cap X_{n+1}| \quad \text{---}(\star)$$

Now

◇ By **(IH)** we know $|U|$ straight away:

$$|U| = \sum_{J \subseteq [n]} (-1)^{|J|+1} \left| \bigcap_{j \in J} X_j \right|$$

This covers the sizes of all the $J \subseteq [n+1]$ for which $n+1 \notin J$.

◇ Note that $U \cap X_{n+1} = \bigcap_{i=1}^n X_i \cap X_{n+1}$. Applying **(IH)** again we get

$$\begin{aligned} -|U \cap X_{n+1}| &= - \sum_{J \subseteq [n]} (-1)^{|J|+1} \left| \left(\bigcap_{j \in J} X_j \right) \cap X_{n+1} \right| && \text{by (IH)} \\ &= - \sum_{J \subseteq [n]} (-1)^{|J \cup \{n+1\}|} \left| \bigcap_{j \in J \cup \{n+1\}} X_j \right| && \text{re-indexing the sum} \\ &= \sum_{J \subseteq [n]} (-1)^{|J \cup \{n+1\}|+1} \left| \bigcap_{j \in J \cup \{n+1\}} X_j \right| && \text{distributing the } - \text{ sign} \end{aligned}$$

This covers the sizes of all the $J \subseteq [n+1]$ for which $n+1 \in J$ and which contain some element of $[n]$.

◇ The only subset of $[n+1]$ not covered by the above two sums is $\{n+1\}$, and $(-1)^{|\{n+1\}|+1} = (-1)^2 = 1$, so that

$$(-1)^{|\{n+1\}|+1} |X_{n+1}| = |X_{n+1}|$$

Together with $(\star)$, these three facts yield the equation we wanted to prove was true.

By induction, we're done. □

PROOF TIP

To find the size of a union of $\bigcup_{i=1}^n X_i$:

- Add the sizes of the individual sets X_i ;
- Subtract the sizes of the double-intersections $X_i \cap X_j$;
- Add the sizes of the triple-intersections $X_i \cap X_j \cap X_k$;
- Subtract the sizes of the quadruple-intersections $X_i \cap X_j \cap X_k \cap X_\ell$;
- ... and so on ...

Keep alternating until the intersection of all the sets is covered.

4.1.20 Example. We count how many subsets of $[12]$ contain a multiple of 3. Precisely, we count the number of elements of the set

$$X_3 \cup X_6 \cup X_9 \cup X_{12}$$

where $X_k = \{S \subseteq [12] \mid k \in S\}$. We will apply the inclusion–exclusion principle:

- (i) An element $S \in X_3$ is precisely a set of the form $\{3\} \cup S'$, where $S' \subseteq [12] \setminus \{3\}$. Since $[12] \setminus \{3\}$ has 11 elements, there are 2^{11} such subsets. So $|X_3| = 2^{11}$, and likewise $|X_6| = |X_9| = |X_{12}| = 2^{11}$.
- (ii) An element $S \in X_3 \cap X_6$ is a set of the form $\{3, 6\} \cup S'$, where $S' \subseteq [12] \setminus \{3, 6\}$. Thus there are 2^{10} such subsets, so $|X_3 \cap X_6| = 2^{10}$. And likewise

$$|X_3 \cap X_9| = |X_3 \cap X_{12}| = |X_6 \cap X_9| = |X_6 \cap X_{12}| = |X_9 \cap X_{12}| = 2^{10}$$

- (iii) Reasoning as in the last two cases, we see that

$$|X_3 \cap X_6 \cap X_9| = |X_3 \cap X_6 \cap X_{12}| = |X_3 \cap X_9 \cap X_{12}| = |X_6 \cap X_9 \cap X_{12}| = 2^9$$

- (iv) ... and $|X_3 \cap X_6 \cap X_9 \cap X_{12}| = 2^8$.

Thus the number of subsets of $[12]$ which contain a multiple of 3 is

$$\underbrace{4 \times 2^{11}}_{\text{by (i)}} - \underbrace{6 \times 2^{10}}_{\text{by (ii)}} + \underbrace{4 \times 2^9}_{\text{by (iii)}} - \underbrace{2^8}_{\text{by (iv)}}$$

which is equal to 3840.

4.1.21 Exercise. How many natural numbers less than 1000 are multiples of 2, 3, 5 or 7?

4.2 Binomials and factorials

We now introduce two kinds of number which are extremely useful for counting. The first of these is *binomial coefficients* $\binom{n}{k}$, which count the number of ways of choosing k elements from a set of size n —for example, the number of ways of forming soccer team from a population of 4000 students is $\binom{4000}{11}$. The second is *factorials* $n!$, which count the number of ways a set of n things can be rearranged—for example, the number of ways shuffling a standard deck of cards is $52!$. (The exclamation point is part of the notation.)

The value behind this notation is that it allows us to express huge numbers in a concise and meaningful way. For example,

$$\binom{4000}{11} = 103\,640\,000\,280\,154\,258\,645\,590\,429\,564\,000$$

Although these two numbers are equal, their *expressions* are very different; the expression on the left is meaningful, but the expression on the right is completely meaningless out of context.

Subsets of finite sets: binomial coefficients

4.2.1 Definition. Let X be a set. The set of all subsets of X is called the **power set** of X , denoted $\mathcal{P}(X)$ (`LATEX` code: `\mathcal{P}\{X\}`).

Note that ‘ $U \in \mathcal{P}(X)$ ’ means precisely the same thing as ‘ $U \subseteq X$ ’.

4.2.2 Proposition. If X is finite then $\mathcal{P}(X)$ is finite and $|\mathcal{P}(X)| = 2^{|X|}$.

Proof. This was Exercise 4.1.10. □

4.2.3 Example. We know that $\mathcal{P}([3])$ has $2^3 = 8$ elements. They are:

$$\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$$

4.2.4 Exercise. Write out $\mathcal{P}(\emptyset)$ and $\mathcal{P}([4])$.

4.2.5 Exercise. Let X be a set. Show that $\emptyset \in \mathcal{P}(X)$ and $X \in \mathcal{P}(X)$.

4.2.6 Definition. Let X be a set and let $k \in \mathbb{N}$. A **k -element subset** of X is a subset $U \subseteq X$ such that $|U| = k$. The set of all k -element subsets of X is denoted $\binom{X}{k}$ (read: ‘ X choose k ’) (`LATEX` code: `\binom{X}{k}`).

Intuitively, $\binom{X}{k}$ is the set of ways of picking k elements from X , without repetitions, in such a way that order doesn’t matter. (If order mattered, the elements would be *sequences* instead of *subsets*.)

4.2.7 Example. We find $\binom{[4]}{k}$ for all $k \in \mathbb{N}$.

- $\binom{[4]}{0} = \{\emptyset\}$ since the only set with 0 elements is $\emptyset$;
- $\binom{[4]}{1} = \{\{1\}, \{2\}, \{3\}, \{4\}\}$;
- $\binom{[4]}{2} = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$;
- $\binom{[4]}{3} = \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}\}$;
- $\binom{[4]}{4} = \{\{1, 2, 3, 4\}\}$;
- If $k \geq 5$ then $\binom{[4]}{k} = \emptyset$, since by Theorem 4.1.6(a), no subset of $[4]$ can have more than 4 elements.

4.2.8 Proposition. If X is a finite set, then $\mathcal{P}(X) = \bigcup_{k \leq |X|} \binom{X}{k}$.

Proof. Let $U \subseteq X$. By Theorem 4.1.6(a), U is finite and $|U| \leq |X|$. Thus $U \in \binom{X}{|U|}$, and hence $U \in \bigcup_{k \leq |X|} \binom{X}{k}$. This proves that $\mathcal{P}(X) \subseteq \bigcup_{k \leq |X|} \binom{X}{k}$.

The fact that $\bigcup_{k \leq |X|} \binom{X}{k} \subseteq \mathcal{P}(X)$ is immediate, since elements of $\binom{X}{k}$ are defined to be subsets of X , and hence elements of $\mathcal{P}(X)$. $\square$

4.2.9 Exercise. Show that Proposition 4.2.8 fails to be true if we remove the hypothesis that X is finite.

4.2.10 Definition. Let $n, k \in \mathbb{N}$. Denote by $\binom{n}{k}$ (read: ‘ n choose k ’) ([L^AT_EX](#) code: `\binom{n}{k}`) the number of k -element subsets of a set of size n . That is, $\binom{n}{k} = \left| \binom{[n]}{k} \right|$. The numbers $\binom{n}{k}$ are called **binomial coefficients**.

Intuitively, $\binom{n}{k}$ is the number of ways of selecting k things from n , without repetitions, in such a way that order doesn’t matter.

4.2.11 Example. In Example 4.2.7 we proved that:

$$\binom{4}{0} = 1, \binom{4}{1} = 4, \binom{4}{2} = 6, \binom{4}{3} = 4, \binom{4}{4} = 1$$

and that $\binom{4}{k} = 0$ for all $k \geq 5$.

4.2.12 Exercise. Fix $n \in \mathbb{N}$. Prove that $\binom{n}{0} = 1$, $\binom{n}{1} = n$ and $\binom{n}{n} = 1$.

We can use the counting principles from Section 4.1 to prove identities about the binomial coefficients. This is an important task: many theorems in algebra, probability theory and statistics rely on our ability to work with binomial coefficients.

Here is one such identity:

4.2.13 Proposition. Let $n, k \in \mathbb{N}$ with $n \geq k$. Then

$$\binom{n}{k} = \binom{n}{n-k}$$

Proof. Specifying a subset $U \subseteq [n]$ of size k can be done in two ways: choose k elements from n to *put into* U , or choose $n - k$ elements from n to *leave out of* U . There are $\binom{n}{k}$ ways of doing the former, and $\binom{n}{n-k}$ ways of doing the latter, hence these two quantities are equal. $\square$

This proof may seem informal—and it is—but it can be made formal if necessary. A more formal version would show that $\binom{[n]}{k}$ and $\binom{[n]}{n-k}$ have the same number of elements, since there is a correspondence matching $U \in \binom{[n]}{k}$ with $[n] \setminus U \in \binom{[n]}{n-k}$. At this elementary level, we can afford to cut back on the technicalities in favour of making the proofs understandable.

4.2.14 Proposition. Let $n \in \mathbb{N}$. Then $2^n = \sum_{k=0}^n \binom{n}{k}$.

Proof. We know that $|\mathcal{P}([n])| = 2^n$ by Proposition 4.2.2 and that $\mathcal{P}([n]) = \bigcup_{k=0}^n \binom{[n]}{k}$ by Proposition 4.2.8. Moreover, the sets $\binom{[n]}{k}$ are pairwise disjoint, so by the addition principle it follows that

$$2^n = |\mathcal{P}([n])| = \left| \bigcup_{k=0}^n \binom{[n]}{k} \right| = \sum_{k=0}^n \left| \binom{[n]}{k} \right| = \sum_{k=0}^n \binom{n}{k}$$

$\square$

PROOF TIP

In the proof of Proposition 4.2.14 we proved that 2^n and $\sum_{k=0}^n \binom{n}{k}$ were equal by showing that they each represent the size of the same set.

This proof technique is called **double-counting**, and is very useful for proving identities regarding numbers that have a combinatorial interpretation (especially binomial coefficients and factorials, which will be introduced later).

4.2.15 Proposition. Let $n, k, \ell \in \mathbb{N}$ with $n \geq k \geq \ell$. Then

$$\binom{n}{k} \binom{k}{\ell} = \binom{n}{\ell} \binom{n-\ell}{k-\ell}$$

Proof. Let's home in on the left-hand side of the equation. By the multiplication principle, $\binom{n}{k} \binom{k}{\ell}$ is the number of ways of selecting a k -element subset of $[n]$ and an ℓ -element subset of $[k]$. Equivalently, it's the number of ways of selecting a k -element subset of $[n]$ and then an ℓ -element subset of the k -element subset that we just selected. To make this slightly more concrete, let's put it this way:

$\binom{n}{k} \binom{k}{\ell}$ is the number of ways of painting k balls red from a bag of n balls, and painting ℓ of the red balls blue. This leaves us with ℓ blue balls and $k - \ell$ red balls.

Now we need to find an equivalent interpretation of $\binom{n}{\ell} \binom{n-\ell}{k-\ell}$. Well, suppose we pick the ℓ elements to be coloured red first. To make up the rest of the k -element subset, we now have to select $k - \ell$ elements, and there are now $n - \ell$ to choose from. Thus

$\binom{n}{\ell} \binom{n-\ell}{k-\ell}$ is the number of ways of painting ℓ balls from a bag of n balls blue, and painting $k - \ell$ of the remaining balls red.

Thus, both numbers represent the number of ways of painting ℓ balls blue and $k - \ell$ balls red from a bag of n balls. Hence they are equal. $\square$

4.2.16 Exercise. Show that if $n, k \in \mathbb{N}$ with $n \geq k$, then

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

Hint: How many ways can you select $k + 1$ animals from a set containing n cats and a dog?

The identity in Exercise 4.2.16 gives rise to an algorithm for computing binomial coefficients: they fit into a diagram known as **Pascal's triangle**, with each binomial coefficient computed as the sum of the two lying above it (with zeroes omitted):

$$\begin{array}{ccccccc} & & & \binom{0}{0} & & & \\ & & \binom{1}{0} & \binom{1}{1} & & & \\ & \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & & & \\ \binom{3}{0} & \binom{3}{1} & \binom{3}{2} & \binom{3}{3} & & & \\ \binom{4}{0} & \binom{4}{1} & \binom{4}{2} & \binom{4}{3} & \binom{4}{4} & & \\ \binom{5}{0} & \binom{5}{1} & \binom{5}{2} & \binom{5}{3} & \binom{5}{4} & \binom{5}{5} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \end{array} = \begin{array}{ccccccc} & & & 1 & & & \\ & & 1 & 1 & & & \\ & 1 & 2 & 1 & & & \\ & 1 & 3 & 3 & 1 & & \\ & 1 & 4 & 6 & 4 & 1 & \\ 1 & 5 & 10 & 10 & 5 & 1 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \end{array}$$

4.2.17 Exercise. Write down the next two rows of Pascal's triangle.

4.2.18 Theorem (Binomial theorem). Let $n \in \mathbb{N}$ and $x, y \in \mathbb{R}$. Then

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

Proof. We proceed by induction on n .

- **(BC)** $(x + y)^0 = 1$ and $\binom{0}{0} x^0 y^{0-0} = 1 \cdot 1 \cdot 1 = 1$, so the statement is true when $n = 0$.
- **(IS)** Fix $n \in \mathbb{N}$ and suppose that

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \quad \text{---(IH)}$$

We need to show that $(x + y)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k}$. Well,

$$\begin{aligned} (x + y)^{n+1} &= (x + y)(x + y)^n && \text{by laws of indices} \\ &= (x + y) \cdot \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} && \text{by (IH)} \\ &= x \cdot \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} + y \cdot \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} && \text{by expanding } (x + y) \\ &= \sum_{k=0}^n \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^n \binom{n}{k} x^k y^{n-k+1} && \text{distributing } x \text{ and } y \\ &= \sum_{k=1}^{n+1} \binom{n}{k-1} x^k y^{n-k+1} + \sum_{k=0}^n \binom{n}{k} x^k y^{n-k+1} && \text{renaming } k \text{ to } k-1 \text{ in first sum} \end{aligned}$$

If $1 \leq k \leq n$ then the coefficient of $x^k y^{n-k+1}$ is $\binom{n}{k} + \binom{n}{k-1}$, which by Exercise 4.2.16 is equal to $\binom{n+1}{k}$. The coefficient of $x^0 y^{n+1}$ is $\binom{n}{0} = 1 = \binom{n+1}{0}$ from the first sum; and the coefficient of $x^{n+1} y^0$ is $\binom{n}{n} = 1 = \binom{n+1}{n+1}$ from the second sum. Hence for each $0 \leq k \leq n+1$ the coefficient of $x^k y^{n-k+1}$ is $\binom{n+1}{k}$, we thus have

$$(x + y)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n-k+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k}$$

By induction, we're done. □

4.2.19 Example. We will use the binomial theorem to show that the alternating sum of binomial coefficients is zero. That is, for $n \in \mathbb{N}$, we have

$$\sum_{k=0}^n (-1)^k \binom{n}{k} = 0$$

This is immediate by substituting $x = -1$ and $y = 1$ into the statement of the theorem; indeed

$$0 = (-1 + 1)^n = \sum_{k=0}^n \binom{n}{k} (-1)^k 1^{n-k} = \sum_{k=0}^n (-1)^k \binom{n}{k}$$

Verifying this in the case $n = 4$ gives

$$\binom{4}{0} - \binom{4}{1} + \binom{4}{2} - \binom{4}{3} + \binom{4}{4} = 1 - 4 + 6 - 4 + 1 = 0$$

4.2.20 Exercise. Use the binomial theorem to prove the identity

$$\sum_{k=0}^n \binom{n}{k} 2^k = 3^n$$

Prove this identity again using a double-counting argument. Verify that the identity holds when $n = 4$.

Permutations of finite sets: factorials

4.2.21 Definition. Let X be a finite set. A **permutation** of X is an ordered list of all the elements of X , with no repeats. Denote the set of permutations of X by S_X ([L^AT_EX](#) code: `S_X`),² and write $S_{[n]} = S_n$ for $n \in \mathbb{N}$.

4.2.22 Example. The permutations of the set $\{X, Y, Z\}$ are

$$(X, Y, Z) \quad (X, Z, Y) \quad (Y, X, Z) \quad (Y, Z, X) \quad (Z, X, Y) \quad (Z, Y, X)$$

4.2.23 Exercise. List all the permutations of the set $[4]$.

4.2.24 Definition. Let $n \in \mathbb{N}$. Denote by $n!$ (read: ‘ n factorial’) the number of permutations of a set of size n . That is, $n! = |S_n|$. The numbers $n!$ are called **factorials**.

4.2.25 Example. Example 4.2.22 shows that $3! = 6$.

There is a very useful recursive formula for factorials:

²The ‘ S ’ comes from ‘symmetry’. The set S_X comes with the natural structure of a *group*; see Section 5.3.

4.2.26 Proposition. $0! = 1$ and if $n \in \mathbb{N}$ then $(n + 1)! = (n + 1) \cdot n!$.

Proof. The only permutation of $\emptyset$ is $()$, the empty list. Hence $S_0 = \{()\}$ and $0! = |S_0| = 1$.

Let $n \in \mathbb{N}$. A permutation of $[n + 1]$ is a list of length $n + 1$. Specifying such a list is equivalent to carrying out the following procedure:

- Choose the position in the list for the number $n + 1$; there are $n + 1$ such choices.
- Choose the positions for the elements of $[n]$; since the position of $n + 1$ is fixed, there are n positions available, meaning that this is equivalent to choosing a permutation of $[n]$. Hence there are $n!$ such choices.

By the multiplication principle, $(n + 1)! = |S_{n+1}| = (n + 1) \cdot n!$. □

By induction, this yields the familiar formula for factorials when $n \geq 1$:

$$n! = 1 \times 2 \times \cdots \times n$$

ASIDE

Many authors define $n!$ using the formula $n! = 1 \times 2 \times \cdots \times n$. I avoid this for two reasons: first, it clouds the combinatorial nature of the factorial operator; and second, it makes the reason why $0!$ should be equal to 1 (and not 0) much hazier.

4.2.27 Example. We count the number of ways we can shuffle a standard deck of cards in such a way that the colour of the cards alternate between red and black.

A procedure for choosing the order of the cards is as follows:

- (i) Choose the colour of the first card. There are 2 such choices. This then determines the colours of the remaining cards, since they have to alternate.
- (ii) Choose the order of the red cards. There are $26!$ such choices.
- (iii) Choose the order of the black cards. There are $26!$ such choices.

By the multiplication principle, there are $2 \cdot (26!)^2$ such rearrangements. This number is huge, and in general there is no reason to write it out. Just for fun, though:

$$2 \cdot (26!)^2 = 325\,288\,005\,235\,264\,929\,014\,077\,766\,819\,257\,214\,042\,112\,000\,000\,000\,000$$

WRITING TIP

When expressing the sizes of finite sets described combinatorially, it is best to *avoid* evaluating the expression; that is, leave in the powers, products, sums, binomial coefficients and factorials! The reason for this is that performing the calculations takes the meaning away from the expressions.

4.2.28 Theorem. Let $n, k \in \mathbb{N}$ with $n \geq k$. Then

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Proof. We will actually prove that $n! = \binom{n}{k} \cdot k! \cdot (n-k)!$; the result then follows by dividing through by $k!(n-k)!$. We prove this equation by counting the number of elements of S_n .

A procedure for defining an element of S_n is as follows:

- (i) Choose which elements will appear in the first k positions of the list. There are $\binom{n}{k}$ such choices.
- (ii) Choose the order of these k elements. There are $k!$ such choices.
- (iii) Choose the order of the remaining $n-k$ elements. There are $(n-k)!$ such choices.

By the multiplication principle, $n! = \binom{n}{k} \cdot k! \cdot (n-k)!$. □

The proof of Theorem 4.2.28 relied only on the combinatorial definitions of binomial coefficients and factorials; we didn't need to know how to compute them at all!

4.2.29 Example. In Example 4.2.11 we showed that $\binom{4}{2} = 6$. Let's verify this:

$$\binom{4}{2} = \frac{4!}{2! \cdot 2!} = \frac{1 \times 2 \times 3 \times 4}{(1 \times 2) \times (1 \times 2)} = \frac{24}{4} = 6$$

ASIDE

You may have seen the formula in Theorem 4.2.28 given as the *definition* of binomial coefficients, rather than an identity proved about them. This is a bad idea for combinatorial purposes, since $\binom{n}{k}$ has meaning even when $n < k$, even though the formula doesn't hold.

We can use Theorem 4.2.28 to give algebraic proofs of many familiar combinatorial identities.

4.2.30 Exercise. Give an algebraic proof that $\binom{n}{k} = \binom{n}{n-k}$ for all $n, k \in \mathbb{N}$ with $n \geq k$.

4.2.31 Example. What follows is an algebraic proof of Proposition 4.2.15, which said that if $n, k, \ell \in \mathbb{N}$ with $n \geq k \geq \ell$ then $\binom{n}{k} \binom{k}{\ell} = \binom{n}{\ell} \binom{n-\ell}{k-\ell}$.

$$\begin{aligned}
 \binom{n}{k} \binom{k}{\ell} &= \frac{n!}{k!(n-k)!} \cdot \frac{k!}{\ell!(k-\ell)!} && \text{by Theorem 4.2.28} \\
 &= \frac{n!k!}{k!\ell!(n-k)!(k-\ell)!} && \text{combine fractions} \\
 &= \frac{n!}{\ell!(n-k)!(k-\ell)!} && \text{cancel } k! \\
 &= \frac{n!(n-\ell)!}{\ell!(n-k)!(k-\ell)!(n-\ell)!} && \text{multiply by } \frac{(n-\ell)!}{(n-\ell)!} \\
 &= \frac{n!}{\ell!(n-\ell)!} \cdot \frac{(n-\ell)!}{(k-\ell)!(n-k)!} && \text{separate fractions} \\
 &= \frac{n!}{\ell!(n-\ell)!} \cdot \frac{(n-\ell)!}{(k-\ell)!((n-\ell)-(k-\ell))!} && \text{since } n-k = (n-\ell) - (k-\ell) \\
 &= \binom{n}{\ell} \binom{n-\ell}{k-\ell} && \text{by Theorem 4.2.28}
 \end{aligned}$$

□

In general, we will avoid such algebraic proofs whenever possible. They can be useful for proving by brute force that particular results are *true*, but they don't give any combinatorial meaning to the identities being proved... which, after all, is the whole point!

4.3 Cardinality

Sizes of finite sets revisited

To do: Rewrite We can use the three kinds of functions we’ve seen so far—injections, surjections and bijections—to study the relative *size* of sets. This makes intuitive sense:

- If $f : X \rightarrow Y$ is injective, then Y must have enough elements that distinct elements of X can be sent to distinct elements of Y . In this sense, the ‘size’ of X must be smaller or equal to that of Y .
- If $f : X \rightarrow Y$ is surjective, then X must have enough elements that every element of Y is the value of some element of X . In this sense, the ‘size’ of X must be greater or equal to that of Y .
- If $f : X \rightarrow Y$ is bijective, then f puts the elements of X and Y into correspondence, so the ‘size’ of X must be equal to that of Y .

4.3.1 Example. There is an injection $f : [3] \rightarrow [5]$ defined by $f(1) = 1$, $f(2) = 2$ and $f(3) = 3$. However, there is no surjection $[3] \rightarrow [5]$, since if $g : [3] \rightarrow [5]$ is a function, then $|g_*([3])| \leq 3$. This means that $g_*([3]) \neq X$ by Theorem 4.1.6(a), so there exists $x \in X$ with $x \notin g_*([3])$. That is, there is no $k \in [3]$ such that $g(k) = x$, so that g is not surjective.

This intuition works in the finite case.

4.3.2 Theorem. Let X and Y be finite sets.

- (a) There exists an injection $f : X \rightarrow Y$ if and only if $|X| \leq |Y|$;
- (b) There exists a surjection $f : X \rightarrow Y$ if and only if $|X| \geq |Y|$;
- (c) There exists a bijection $f : X \rightarrow Y$ if and only if $|X| = |Y|$.

Proof. Let $|X| = m$ and $|Y| = n$, and write

$$X = \{x_1, x_2, \dots, x_m\} \quad \text{and} \quad Y = \{y_1, y_2, \dots, y_n\}$$

(a) We prove each direction separately.

- ($\Rightarrow$) Suppose there exists an injection $f : X \rightarrow Y$. Then $|f_*(X)| = m$. Indeed,

$$f_*(X) = \{z_1, z_2, \dots, z_m\}$$

where $z_i = f(x_i)$ for each $i \in [m]$, and by injectivity of f we have $z_i \neq z_j$ whenever $i \neq j$. Since $f_*(X) \subseteq Y$, we have $m \leq n$ by Theorem 4.1.6(a).

- ($\Leftarrow$) Suppose $m \leq n$. Define $f : X \rightarrow Y$ by

$$f(x_i) = y_i \quad \text{for all } i \in [n]$$

Then f is injective since if $i \neq j$ then $f(i) = y_i \neq y_j = f(j)$.

(b) We prove each direction separately.

- ($\Rightarrow$) Suppose there exists a surjection $f : X \rightarrow Y$. Then $f_*(X) = Y$, so that $|f_*(X)| = n$. But $|f_*(X)| \leq m$, since we can write

$$f_*(X) = \{z_1, z_2, \dots, z_m\}$$

where $z_i = f(x_i)$ for all $i \in [m]$. Hence $n \leq m$.

- ($\Leftarrow$) Suppose $m \geq n$. Define $f : X \rightarrow Y$ by

$$f(x_i) = \begin{cases} y_i & \text{if } 1 \leq i \leq n \\ y_m & \text{if } i > n \end{cases}$$

Then f is surjective: if $y \in Y$ then $y = y_i$ for some $i \in [m]$, so $y = f(x_i)$.

(c) This is immediate from (a) and (b).

This completes the proof. □

Recall Definition 4.1.2, where we said a set X is *finite* if its elements can be labelled by the elements of $[n]$ for some $n \in \mathbb{N}$. With our new language of functions, we could rephrase this as saying that X is finite if there is a bijection³ $f : [n] \rightarrow X$ for some $n \in \mathbb{N}$. Then if $x \in X$, there is some $k \in [n]$ such that $f(k) = x$ by surjectivity of f , so that x is labelled by k .

The next definition takes this one step further.

4.3.3 Definition. Let X be a set. We say X is **countably infinite** if there exists a bijection $\mathbb{N} \rightarrow X$, and X is **countable** if it is countably infinite or finite.

That is, a set X is countably infinite if its elements can be *listed*, with one entry in the list for each natural number.

4.3.4 Example. We have already seen many examples of countably infinite sets.

- The set $\mathbb{N}$ is countably infinite, since by Exercise 3.2.40, $\text{id}_{\mathbb{N}} : \mathbb{N} \rightarrow \mathbb{N}$ is a bijection.

³In fact, a *surjection* would do just fine here.

- The function $f : \mathbb{Z} \rightarrow \mathbb{N}$ defined for $x \in \mathbb{Z}$ by

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

is a bijection. Indeed, it has an inverse is given by

$$f^{-1}(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ -\frac{x+1}{2} & \text{if } x \text{ is odd} \end{cases}$$

Hence the set of integers $\mathbb{Z}$ is countably infinite. The corresponding list of integers is given by

$$0, -1, 1, -2, 2, -3, 3, -4, 4, \dots$$

The fact that f^{-1} is a bijection means that each integer appears on this list exactly once.

4.3.5 Exercise. Prove that the function $p : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ defined by $p(x, y) = 2^x(2y + 1) - 1$ is a bijection. Deduce that if X and Y are countably infinite sets, then $X \times Y$ is countably infinite.

Exercise 4.3.5 allows us to prove that the product of finitely many countably infinite sets are countably infinite.

4.3.6 Proposition. Let $n \geq 1$ and let $X_1, \dots, X_n$ be countably infinite sets. Then the product $\prod_{i=1}^n X_i$ is countably infinite.

Proof. We proceed by induction on n .

- **(BC)** When $n = 1$ the assertion is trivial: if X_1 is countably infinite then X_1 is countably infinite.
- **(IS)** Fix $n \geq 1$ and suppose that for any sets $X_1, \dots, X_n$, the product $\prod_{i=1}^n X_i$ is countably infinite. Fix sets $X_1, \dots, X_{n+1}$. Then $\prod_{i=1}^n X_i$ is countably infinite by the induction hypothesis, and X_{n+1} is countably infinite by assumption. Hence by Exercise 4.3.5, the set

$$\prod_{i=1}^{n+1} X_i = \left(\prod_{i=1}^n X_i \right) \times X_{n+1}$$

is countably infinite, as required.

By induction, we're done. □

Finding a *bijection* $\mathbb{N} \rightarrow X$, or equivalently $X \rightarrow \mathbb{N}$, can be a bit of a hassle. However, in order to prove that a set X is countable, it suffices to find either a surjection $\mathbb{N} \rightarrow X$ or an injection $X \rightarrow \mathbb{N}$.

4.3.7 Theorem. Let X be an inhabited set. The following are equivalent:

- (i) X is countable;
- (ii) There exists a surjection $f : \mathbb{N} \rightarrow X$;
- (iii) There exists an injection $f : X \rightarrow \mathbb{N}$.

Proof. We'll prove (i) $\Leftrightarrow$ (ii) and (i) $\Leftrightarrow$ (iii).

- (i) $\Rightarrow$ (ii). Suppose X is countable. If X is countably infinite, then there exists a bijection $f : \mathbb{N} \rightarrow X$, which is a surjection. If X is finite then there exists a bijection $g : [m] \rightarrow X$, where $m = |X| \geq 1$. Define $f : \mathbb{N} \rightarrow X$ by

$$f(n) = \begin{cases} g(n) & \text{if } n \leq m \\ g(1) & \text{if } n = 0 \text{ or } n > m \end{cases}$$

Then f is surjective: if $x \in X$ then there exists $n \in [m]$ such that $g(n) = x$, and then $f(n) = g(n) = x$.

- (ii) $\Rightarrow$ (i). Suppose there exists a surjection $f : \mathbb{N} \rightarrow X$. To prove that X is countable, it suffices to prove that if X is infinite then it is countably infinite. So suppose X is infinite, and define a sequence recursively by

- ◊ $a_0 = 0$;
- ◊ Fix $n \in \mathbb{N}$ and suppose $a_0, \dots, a_n$ have been defined. Define a_{n+1} to be the least natural number for which $f(a_{n+1}) \notin \{f(a_0), f(a_1), \dots, f(a_n)\}$.

Define $g : \mathbb{N} \rightarrow X$ by $g(n) = f(a_n)$ for all $n \in \mathbb{N}$. Then

- ◊ g is injective, since if $m \leq n$ then $f(a_m) \neq f(a_n)$ by construction of the sequence $(a_n)_{n \in \mathbb{N}}$.
- ◊ g is surjective. Indeed, given $x \in X$, by surjectivity there exists $m \in \mathbb{N}$ which is least such that $f(m) = x$, and we must have $a_n = m$ for some $n \leq m$ by construction of the sequence $(a_n)_{n \in \mathbb{N}}$. So $x = g(a_n)$, and hence g is surjective.

So g is a bijection, and X is countable.

- (i) $\Rightarrow$ (iii). Suppose X is countable. If X is countably infinite, then there exists a bijection $f : \mathbb{N} \rightarrow X$, so $f^{-1} : X \rightarrow \mathbb{N}$ is bijective and hence injective. If X is finite then there exists a bijection $g : [m] \rightarrow X$, where $m = |X| \geq 1$. Then $g^{-1} : X \rightarrow [m]$ is injective. Let $i : [m] \rightarrow \mathbb{N}$ be defined by $i(k) = k$ for all $k \in [m]$. Then $i \circ g^{-1}$ is injective; indeed, for $x, x' \in X$ we have

$$i(g^{-1}(x)) = i(g^{-1}(x')) \Rightarrow g^{-1}(x) = g^{-1}(x') \Rightarrow x = x'$$

The first implication is by definition of i , and the second is by injectivity of g^{-1} . So there exists an injection $X \rightarrow \mathbb{N}$.

- (iii) $\Rightarrow$ (i). Suppose there exists an injection $f : X \rightarrow \mathbb{N}$. To prove that X is countable, it suffices to prove that if X is infinite then it is countably infinite. Define a sequence $(a_n)_{n \in \mathbb{N}}$ recursively as follows:

- ◊ Let a_0 be the least element of $f_*(X)$;
- ◊ Fix $n \in \mathbb{N}$ and suppose $a_0, \dots, a_n$ have been defined. Let a_{n+1} be the least element of $f_*(X) \setminus \{a_0, \dots, a_n\}$. This exists since f is injective, so $f_*(X)$ is infinite.

Define $g : \mathbb{N} \rightarrow X$ by, for each $n \in \mathbb{N}$, letting $g(n)$ be the unique value of x for which $f(x) = a_n$. Then

- ◊ g is injective. By construction $a_m \neq a_n$ whenever $m \neq n$. Let $x, y \in X$ be such that $f(x) = a_m$ and $f(y) = a_n$. Since f is injective, we must have $x \neq y$, and so $g(m) = x \neq y = g(n)$.
- ◊ g is surjective. Fix $x \in X$. Then $f(x) \in f_*(X)$, so there exists $m \in \mathbb{N}$ such that $f(x) = a_m$. Hence $g(m) = x$.

So g is a bijection, and X is countably infinite.

Hence the equivalences have been proved. □

In fact, we needn't even use $\mathbb{N}$ as the domain of the surjection or the codomain of the injection; we can in fact use any countable set C .

4.3.8 Corollary. Let X be an inhabited set. The following are equivalent:

- (a) X is countable;
- (b) There exists a surjection $f : C \rightarrow X$ for some countable set C ;
- (c) There exists an injection $f : X \rightarrow C$ for some countable set C .

4.3.9 Exercise. Prove Corollary 4.3.8.

This is useful for proving the countability of many other sets. For example, we can now prove that the set of all rational numbers is countable.

4.3.10 Example. $\mathbb{Q}$ is countable. Indeed, by Exercises 4.3.4 and 4.3.5, the set $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is countable. Moreover, there exists a surjection $q : \mathbb{Z} \times (\mathbb{Z} \setminus \{0\}) \rightarrow \mathbb{Q}$ defined by

$$q(a, b) = \frac{a}{b}$$

By Corollary 4.3.8, $\mathbb{Q}$ is countable.

4.3.11 Exercise. Let X be a countable set. Prove that $\binom{X}{k}$ is countable for each $k \in \mathbb{N}$.

4.3.12 Theorem. A countable union of countable sets is countable. More precisely, let $\{X_n \mid n \in \mathbb{N}\}$ be a family of countable sets. Then the set X defined by

$$X = \bigcup_{n \in \mathbb{N}} X_n$$

is countable.

Proof. We may assume that the sets X_n are all inhabited, since the empty set does not contribute to the union.

For each $n \in \mathbb{N}$ there is a surjection $f_n : \mathbb{N} \rightarrow X_n$. Define $f : \mathbb{N} \times \mathbb{N} \rightarrow X$ by $f(m, n) = f_m(n)$ for all $m, n \in \mathbb{N}$. Then f is surjective: if $x \in X$ then $x \in X_m$ for some $m \in \mathbb{N}$. Since f_m is surjective, it follows that $x = f_m(n)$ for some $n \in \mathbb{N}$. But then $x = f(m, n)$. Since $\mathbb{N} \times \mathbb{N}$ is countable, it follows from Corollary 4.3.8 that X is countable. $\square$

4.3.13 Example. Let X be a countable set. The set of all finite subsets of X is countable. Indeed, the set of all finite subsets of X is equal to $\bigcup_{k \in \mathbb{N}} \binom{X}{k}$, which is a union of countably many countable sets by Exercise 4.3.11, so is countable by Theorem 4.3.12.

We can also use some clever trickery to prove that certain sets are *uncountable*. The proof of the following theorem is known as **Cantor's diagonal argument**.

4.3.14 Theorem. The set $\{0, 1\}^{\mathbb{N}}$ is uncountable.

Proof. Let $f : \mathbb{N} \rightarrow \{0, 1\}^{\mathbb{N}}$ be a function. We will prove that f is not surjective by constructing a sequence which is not contained in the image of $\mathbb{N}$ under f .

Define an element $b \in \{0, 1\}^{\mathbb{N}}$, i.e. a function $b : \mathbb{N} \rightarrow \{0, 1\}$, by

$$b(n) = 1 - f(n)(n)$$

Then $b(n) \neq f(n)(n)$ for all $n \in \mathbb{N}$. If $b = f(m)$ for some m , then by definition of function equality we must have $b(m) = f(m)(m)$; but we just saw that this is necessarily false. Hence $b \notin f_*(\mathbb{N})$, so f is not surjective.

Hence there does not exist a surjective function $\mathbb{N} \rightarrow \{0, 1\}^{\mathbb{N}}$. By Theorem 4.3.7, the set $\{0, 1\}^{\mathbb{N}}$ is uncountable. $\square$

This result can be used to show that the set $\mathbb{R}$ of all real numbers is uncountable, though this relies on features of the real numbers that we have not developed.

Cardinality

4.3.15 Definition. Let X and Y be sets. Write $X \leq Y$ if there exists an injection $X \rightarrow Y$, and write $X \cong Y$ ([L^AT_EX](#) code: `\cong`) if there exists a bijection $X \rightarrow Y$.

We will use these definitions to generalise the notion of ‘size’ to sets that are not necessarily finite. In particular, for a fixed universe $\mathcal{U}$ of sets:

- $\cong$ is an equivalence relation on $\mathcal{U}$ (see Theorem 4.3.16), and $\cong$ -equivalence classes represent the *size* (called *cardinality*) of the sets they contain (see Definition 4.3.17).
- $\leq$ induces to a partial order on $\mathcal{U}/\cong$ (see Theorem 4.3.19).

Exercises 3.2.40, 3.2.41 and 3.2.42 did most of the legwork for defining cardinality.

4.3.16 Theorem. Let $\mathcal{U}$ be a given universe of sets. Then $\cong$ is an equivalence relation on $\mathcal{U}$.

Proof. We need to prove that $\cong$ is reflexive, symmetric and transitive.

- **Reflexivity.** Fix $X \in \mathcal{U}$. Then $X \cong X$ since, by Exercise 3.2.40, the identity function $\text{id}_X : X \rightarrow X$ is a bijection.
- **Symmetry.** Fix $X, Y \in \mathcal{U}$ and suppose $X \cong Y$. Then there is a bijection $f : X \rightarrow Y$. By Exercise 3.2.41, the inverse function $f^{-1} : Y \rightarrow X$ is a bijection, so $Y \cong X$.
- **Transitivity.** Fix $X, Y, Z \in \mathcal{U}$ and suppose $X \cong Y$ and $Y \cong Z$. Then there are bijections $f : X \rightarrow Y$ and $g : Y \rightarrow Z$. By Exercise 3.2.42, the composite $g \circ f : X \rightarrow Z$ is a bijection. Hence $X \cong Z$.

So $\cong$ is an equivalence relation on $\mathcal{U}$. $\square$

4.3.17 Definition. Fix a universe of sets $\mathcal{U}$. Given $X \in \mathcal{U}$, the **cardinality** of X is the equivalence class $[X]_{\cong}$. We denote the cardinality of X by $|X|$. When X is finite, we identify its cardinality with its size (as in Definition 4.1.2).

4.3.18 Example. We can phrase much of what we've mentioned so far in terms of cardinality:

- A set X is finite if $|X| = n$ for some $n \in \mathbb{N}$.
- A set X is countably infinite if $|X| = |\mathbb{N}|$. Thus for example

$$|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$$

4.3.19 Theorem (Cantor–Schröder–Bernstein theorem). Let X and Y be sets. If $X \leq Y$ and $Y \leq X$ then $X \cong Y$.

Proof. Let X and Y be sets. We have already seen that this is true when X and Y are finite, so suppose X and Y are infinite. Let $f : X \rightarrow Y$ and $g : Y \rightarrow X$ be injections, and suppose further that X and Y are disjoint.

We'll construct a bijection $h : X \rightarrow Y$ first by defining an equivalence relation on $X \cup Y$.

Given $z, z' \in X \cup Y$, say that z' is **reachable** from z if z' can be obtained from z by applying f and g a finite number of times. Define a relation $\sim$ on $X \cup Y$ by declaring $z \sim z'$ if z is reachable from z' , or z' is reachable from z . Then $\sim$ is an equivalence relation:

To do: Details; they're intuitive but horrific!

Hence the set $(X \cup Y)/\sim$ is a partition of $X \cup Y$. Given $z \in X \cup Y$, there are three possibilities:

- (a) There exists $x \in X$ such that each $w \in [z]_{\sim}$ is reachable from x ; in this case there is a bijection

$$h_w : [z]_{\sim} \cap X \rightarrow [z]_{\sim} \cap Y$$

defined by $h_w(u) = f(u)$.

- (b) There exists $y \in Y$ such that each $w \in [z]_{\sim}$ is reachable from y ; in this case there is a bijection

$$h_w : [z]_{\sim} \cap X \rightarrow [z]_{\sim} \cap Y$$

defined by $h_w^{-1}(u) = g(u)$.

- (c) There does not exist an element of $[z]_{\sim}$ from which every other element of $[z]_{\sim}$ is reachable. In this case, pick a fixed $w \in [z]_{\sim} \cap X$; then there is a bijection

$$h_w : [z]_{\sim} \cap X \rightarrow [z]_{\sim} \cap Y$$

defined by $h_w(u) = f(u)$.

To do: Details that these are all bijections.

Finally, define $h : X \rightarrow Y$ by letting $h(x) = h_w(x)$, where $w \in [x]_{\sim}$ is specified as above.

To do: Details that h is injective. □

4.3.20 Exercise. Fix a universe of sets $\mathcal{U}$, and define

$$\mathbf{Card} = \{|X| \mid X \in \mathcal{U}\}$$

That is, $\mathbf{Card}$ is the set of cardinalities of all the sets in $\mathcal{U}$. Define a relation $\leq$ on $\mathbf{Card}$ by

$$|X| \leq |Y| \quad \Leftrightarrow \quad X \leq Y$$

Prove that $\leq$ is a partial order on $\mathbf{Card}$.

To do: Waffle about how this yields a new notion of ‘infinite number’.

4.3.21 Exercise. Does there exist a set X such that $|\mathbb{N}| < |X| < |\mathcal{P}(\mathbb{N})|$?

To do: Continuum hypothesis.

To do: Cardinal arithmetic. Contrast to finitary arithmetic.

4.4 Problems

4.4.1 Problem. Let X be a set. Show that $\mathcal{P}(X)$ is closed under unions, intersections and relative complements in X . That is, show that if $\{U_i \mid i \in I\}$ is a family of elements of $\mathcal{P}(X)$ then $\bigcup_{i \in I} U_i \in \mathcal{P}(X)$ and $\bigcap_{i \in I} U_i \in \mathcal{P}(X)$; and if $U \in \mathcal{P}(X)$ then $X \setminus U \in \mathcal{P}(X)$.

Chapter 5

Sets with structure

5.1 Rings: sets with arithmetic

In Chapter 2 we examined the integers by pushing their *arithmetic structure* to their limits. Everything we did—including the definition and study of divisibility—was done in terms of addition, subtraction and multiplication.

Given a set R , provided we can make sense of addition, subtraction and multiplication, we can make all the same basic definitions—like divisibility, greatest common divisors, and so on—and ask ourselves to what extent the results that held true of integers hold for R .

This motivates the definition of a *ring*.

5.1.1 Definition. A **(commutative) ring (with unity)** is a set R equipped with:

- An **addition** function $a : R \times R \rightarrow R$; we write $a(r, s) = r + s$ for $r, s \in R$;
- A **negation** function $n : R \rightarrow R$; we write $n(r) = -r$ for $r \in R$;
- A **multiplication** function $m : R \times R \rightarrow R$; we write $m(r, s) = r \cdot s$ for $r, s \in R$;
- An **additive identity** element $0_R \in R$;
- A **multiplicative identity** element $1_R \in R$

such that the following conditions hold:

- Properties of addition
 - ◊ (Associativity) If $r, s, t \in R$ then $(r + s) + t = r + (s + t)$;
 - ◊ (Identity) If $r \in R$ then $r + 0_R = 0_R + r = r$;
 - ◊ (Inverse) If $r \in R$ then $r + (-r) = (-r) + r = 0_R$;
 - ◊ (Commutativity) If $r, s \in R$ then $r + s = s + r$.
- Properties of multiplication
 - ◊ (Associativity) If $r, s, t \in R$ then $(r \cdot s) \cdot t = r \cdot (s \cdot t)$;
 - ◊ (Identity) If $r \in R$ then $r \cdot 1_R = 1_R \cdot r = r$;
 - ◊ (Commutativity) If $r, s \in R$ then $r \cdot s = s \cdot r$.
- Relationship between addition and multiplication
 - ◊ (Distributivity) If $r, s, t \in R$ then $r \cdot (s + t) = (r \cdot s) + r \cdot t$.

5.1.2 Example. The number sets $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are all rings with their usual notions of addition and multiplication and so on. The ring $\mathbb{Z}$ of integers was our focus in Chapter 2.

5.1.3 Example. Let X be a set. Then the power set $\mathcal{P}(X)$ is a ring, with:

- $U + V = U \Delta V$; that is, addition is given by the symmetric difference.
- $-U = U$;
- $U \cdot V = U \cap V$; that is, multiplication is given by intersection;
- $0_{\mathcal{P}(X)} = \emptyset$;
- $1_{\mathcal{P}(X)} = X$.

5.1.4 Exercise. Verify that $\mathcal{P}(X)$, as described in Example 5.1.3, is a ring.

5.1.5 Example. There are many other rings. Some commonly occurring ones are:

- Given a ring R , there is a ring $R[x]$ of **polynomials over R** . That is, expressions of the form

$$r_0 + r_1x + \cdots + r_nx^n$$

where $n \in \mathbb{N}$ and $r_0, \dots, r_n \in R$.

- If $n \in \mathbb{Z}$ is not a perfect square, we can define

$$\mathbb{Z}[\sqrt{n}] = \{a + b\sqrt{n} \mid a, b \in \mathbb{Z}\}$$

with addition and multiplication defined as you'd expect; in particular

$$(a + b\sqrt{n})(c + d\sqrt{n}) = (ac + bdn) + (ad + bc)\sqrt{n}$$

This is defined even when $n < 0$. For example in $\mathbb{Z}[\sqrt{-5}]$ we have

$$(1 - \sqrt{-5})(1 + \sqrt{-5}) = (1 + 5) + (1 - 1)\sqrt{-5} = 6$$

- Given $n > 0$, the set $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$ can be given the structure of a ring by declaring $a + b$ to be the remainder of $a + b$ when divided by n , and likewise for $a \cdot b$.

Many of the definitions that we provided for integers can then be carried over to arbitrary rings; we'll spend most of the rest of this section comparing them.

Units and zero divisors

5.1.6 Definition. Let R be a ring.

- Let $r, s \in R$. We say r **divides** s , and write $r \mid s$ if there exists $q \in R$ such that $s = qr$.
- $u \in R$ is a **unit** if $u \mid 1_R$.

5.1.7 Example. Some examples of divisors and units in rings are as follows:

- We showed above that $1 + \sqrt{5} \mid 6$ in $\mathbb{Z}[\sqrt{-5}]$.
- The number $1 + \sqrt{2}$ is a unit in $\mathbb{Z}[\sqrt{2}]$: indeed, $1 + \sqrt{2} \mid 1$ since
$$(1 + \sqrt{2})(-1 + \sqrt{2}) = (-1 + 2) + (-1 + 1)\sqrt{2} = 1$$
- In $\mathbb{Q}$ and $\mathbb{R}$, every non-zero element is a unit. Indeed, if $x \in \mathbb{Q}$ and $x \neq 0$ then $\frac{1}{x} \in \mathbb{Q}$, so $x \cdot \frac{1}{x} = 1$, so $x \mid 1$. Likewise with $\mathbb{R}$.
- We know that if $a \perp n$ then a has a multiplicative inverse modulo n ; this is precisely the assertion that $a \perp n$ if and only if a is a unit in $\mathbb{Z}_n$. (Contrast this to the situation of $\mathbb{Z}$, where the only units are -1 and 1 .)

One definition that we *didn't* give for the integers, because it would have been silly, is that of a *zero divisor*.

5.1.8 Definition. Let R be a ring. A **zero divisor** in R is an element $r \in R$ such that $r \mid 0_R$. We say R is an **integral domain** if the only zero divisor in R is 0_R itself.

So for example $\mathbb{Z}$ is an integral domain, since if $r \in \mathbb{Z}$ and $r \mid 0$, then $r = 0$.

5.1.9 Proposition. Let $n > 1$. Then $\mathbb{Z}_n$ is an integral domain if and only if n is prime.

Proof. Suppose n is composite, say $n = ab$ where $0 < a < n$ and $0 < b < n$. Then $ab = 0$ in $\mathbb{Z}_n$, so $a \mid 0$, even though $a \neq 0$. Hence $\mathbb{Z}_n$ is not an integral domain.

Conversely, suppose n is prime. Then $a \perp n$ for all $a \in \mathbb{Z}_n$ with $a \neq 0$, so a has a multiplicative inverse $b \in \mathbb{Z}_n$ modulo n . That is, $ab = 1$ in $\mathbb{Z}_n$. Suppose, furthermore, that a is a zero divisor. Then $ac = 0$ for some $c \in \mathbb{Z}_n$ with $c \neq 0$. Hence

$$c = 1 \cdot c = (ab)c = (ac)b = 0b = 0$$

contradicting the fact that $c \neq 0$. So a must not be a zero divisor. So $\mathbb{Z}_n$ is an integral domain. $\square$

Primes and irreducibles

The definitions of primes and irreducibles in $\mathbb{Z}$ carry over to arbitrary integral domains.

5.1.10 Definition. Let R be an integral domain and let $p \in R$. Then

- p is **prime** if p is a non-zero non-unit and, for all $r, s \in R$, if $p \mid rs$ then $p \mid r$ or $p \mid s$;
- p is **irreducible** if p is a non-zero non-unit and, for all $r, s \in R$, if $p = rs$ then either r is a unit or s is a unit.

Primes and irreducibles coincide in the ring $\mathbb{Z}$, as we proved in Proposition 2.2.11, but this is not necessarily the case in an arbitrary ring.

5.1.11 Example. In $\mathbb{Z}[\sqrt{-5}]$, the element 2 is irreducible. Moreover, since $(1 + \sqrt{-5})(1 - \sqrt{-5}) = 6$, we have

$$2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

However, $2 \nmid 1 + \sqrt{-5}$ and $2 \nmid 1 - \sqrt{-5}$. Indeed, if $2 \mid 1 + \sqrt{-5}$, then there will exist $a, b \in \mathbb{Z}$ such that

$$1 + \sqrt{-5} = 2(a + b\sqrt{-5}) = 2a + 2b\sqrt{-5}$$

But this implies that $1 = 2a$, and hence $2 \mid 1$ in $\mathbb{Z}$, which is nonsense. Likewise for $1 - \sqrt{-5}$.

Hence 2 is not prime in $\mathbb{Z}[\sqrt{-5}]$. However 2 is irreducible in $\mathbb{Z}[\sqrt{-5}]$, though the proof of this is omitted here because it's a little involved.

We can classify some rings in which primes and irreducibles *do* coincide, however.

Principal ideal domains

5.1.12 Definition. Let R be a ring. An **ideal** in R is a subset $I \subseteq R$ such that

- If $x, y \in I$ then $x - y \in I$; and
- If $x \in I$ and $r \in R$ then $rx \in I$.

5.1.13 Example. We have seen **To do: reference; PS4 Q3** that if $I \subseteq \mathbb{Z}$ is an ideal then

$$I = d\mathbb{Z} = \{dn \mid n \in \mathbb{Z}\}$$

for some $d \in \mathbb{Z}$.

The proof of Exercise 5.1.13 relies fundamentally on the division theorem. However, we cannot state the division theorem for an arbitrary ring because it used the order relation $\leq$ on $\mathbb{Z}$, which might not exist in an arbitrary ring.

5.1.14 Example. There is an ideal in $\mathbb{Z}[x]$ defined by

$$(2, x) = \{2p(x) + xq(x) \mid p(x), q(x) \in \mathbb{Z}[x]\} = \{2a + xq(x) \mid a \in \mathbb{Z}, q(x) \in \mathbb{Z}[x]\}$$

Ideals in a ring R give rise to a nice kind of an equivalence relation, called a **congruence relation**. Define $\equiv_I$ by $r \equiv_I s$ if and only if $r - s \in I$. The quotient $R/\equiv_I$ is denoted by R/I , and the equivalence class $[r]_{\equiv_I}$ is denoted by $r + I$.

5.1.15 Example. In $\mathbb{Z}$, if $I = n\mathbb{Z}$ then, given $a, b \in \mathbb{Z}$, we have

$$a \equiv_{n\mathbb{Z}} b \iff a - b \in n\mathbb{Z} \iff n \mid a - b \iff a \equiv b \pmod{n}$$

Hence $a + n\mathbb{Z} = [a]_n$ and $\mathbb{Z}/n\mathbb{Z}$ is what we previously called $\mathbb{Z}/n\mathbb{Z}$.

Moreover, in any ring R with ideal I , the relation $\equiv_I$ is compatible with the operations of R ; in other words, if $r + s = r' + s'$ in R then

$$(r + s) + I = (r' + s') + I$$

and likewise for multiplication. Hence we have a form of modular arithmetic that we can perform in any ring.

5.1.16 Example. Consider the ring $\mathbb{R}[x]$ of polynomials with real coefficients. Define

$$I = \langle x^2 + 1 \rangle = \{(x^2 + 1)f(x) \mid f(x) \in \mathbb{R}[x]\}$$

Then I is an ideal in $\mathbb{R}[x]$, and the quotient ring $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ is ‘the same’ as the set of complex numbers; that is, the real numbers extended by allowing $\sqrt{-1}$ to exist.

Principal ideal domains

5.1.17 Definition. An ideal I is **principal** if there exists $d \in R$ such that

$$I = \{rd \mid r \in R\}$$

In this case we write $I = (d)$. A ring R is called a **principal ideal domain** if every ideal in R is principal.

A consequence of 5.1.13 is:

5.1.18 Proposition. $\mathbb{Z}$ is a principal ideal domain.

Nonetheless, there are other examples of principal ideal domains, such as $\mathbb{Q}[x]$ and $\mathbb{Z}[\sqrt{-1}]$.

5.1.19 Theorem. Let R be a ring. If R is a principal ideal domain, then every irreducible element $p \in R$ is prime.

Proof. Let $p \in R$ and suppose p is irreducible.

First note that if $I \subseteq R$ is an ideal and $p \in I$, then either $I = (p)$ or $I = R$. Certainly $(p) \subseteq I$ since p , and hence all multiples of p are elements of I . Since R is a principal ideal domain, $I = (x)$ for some $x \in R$. Since $p \in I$ we must have $p = rx$ for some $r \in R$. Since p is irreducible, either x is a unit or r is a unit. If x is a unit then $I = R$. If r is a unit then $qr = 1$ for some $q \in R$, so $x = qrx = qp \in (p)$, so $I \subseteq (p)$, and hence $I = (p)$. So either $I = (p)$ or $I = R$.

Now fix $r, s \in R$ and suppose that $p \mid rs$. We need to prove that $p \mid r$ or $p \mid s$. N

□

5.2 Posets: sets with order

5.3 Groups: sets with symmetry

5.4 Metric spaces: sets with distance

Appendix A

Proof techniques

A.1 Structuring a proof

The following subsections give ways of proving propositions in terms of their component parts. They all have a *direct proof*, which usually has a constructive character.

These tips can be combined. For example, the statement

Every even natural number greater than two can be written as the sum of two primes.

has the form

$$\forall n \geq 2, \underbrace{\underbrace{(n \text{ even})}_{q(n)} \Rightarrow \underbrace{\exists a, b \in \mathbb{N}, \overbrace{a, b \text{ prime}}^{s(n,a,b)} \wedge \overbrace{a + b = n}^{t(n,a,b)}}_{r(n)}}_{p(n)}$$

One way we could structure a proof of this statement is as follows:

- Fix $n \geq 2$. (A.1.7 on $\forall n \geq 2, p(n)$)
 - ◊ We now need prove $p(n)$, using the variable n freely, assuming that $n \geq 2$.
- Now $p(n)$ has the form $q(n) \Rightarrow r(n)$
- Suppose n is even. (A.1.5 on $q(n) \Rightarrow r(n)$)
 - ◊ We now need to prove $r(n)$, using the assumptions that $n \geq 2$ and that n is even.

Now $r(n)$ has the form $\exists a, b \in \mathbb{N}, s(n, a, b) \wedge t(n, a, b)$

- Let $a = [\text{define } a \text{ here}]$ and $b = [\text{define } b \text{ here}]$. (A.1.8 on $\exists a, b \in \mathbb{N}, r(n, a, b)$)

◊ Now we need to prove $r(n, a, b)$ is true, where a and b are now defined by whatever values you gave them.

Now $r(n, a, b)$ has the form $s(n, a, b) \wedge t(n, a, b)$.

- First we prove that a and b are prime. *[Insert proof that a and b are prime here.]* Now we prove that $a + b = n$. *[Insert proof that $a + b = n$ here.]* (A.1.2 on $s(n, a, b) \wedge t(n, a, b)$) $\square$

A.1.1 Simple propositions: p

Double-negation

- Prove $\neg(\neg p)$; see A.1.4.

Possible proof format. Suppose $\neg p$. *[Insert proof of a statement q which is known or assumed to be false.]* This contradicts the [fact/assumption] that q is false. Hence p is true. $\square$

A.1.2 Conjunctions: $p \wedge q$

Direct proof

- With no change to assumptions, prove p .
- With no change to assumptions, prove q .

Possible proof format. First we prove p . *[Insert proof of p here.]*

Now we prove q . *[Insert proof of q here.]* $\square$

A.1.3 Disjunctions: $p \vee q$

Direct proof

- With no change to assumptions, either prove p , or prove q .

Proof as an implication

This uses the fact that $p \vee q$ is equivalent to $(\neg p) \Rightarrow q$.

- Add $\neg p$ to the list of assumptions, and prove q .

A.1.4 Negations: $\neg p$

Direct proof (proof by contradiction)

- Add p to list of assumptions, and prove something which is known to be false.

A.1.5 Implications: $p \Rightarrow q$

Direct proof

- Add p to list of assumptions, and prove q .

Proof as a disjunction

- With no change to assumptions, prove $(\neg p) \vee q$; see A.1.3 and A.1.4.

A.1.6 Biconditionals: $p \Leftrightarrow q$

Direct proof

- With no change to assumptions, prove $p \Rightarrow q$.
- With no change to assumptions, prove $q \Rightarrow p$.

A.1.7 Universal quantifiers: $\forall x \in X, p(x)$

Direct proof

- Introduce the variable x , add $x \in X$ to list of assumptions, and prove $p(x)$.

Proof using De Morgan's laws

- Prove $\neg \exists x \in X, \neg p(x)$; see [A.1.8](#)

Proof by weak induction

For this proof technique, we assume $X = \{n \in \mathbb{N} \mid n \geq b\}$ for some $b \in \mathbb{N}$. That is, the statement we are proving is of the form $\forall n \geq b, p(n)$.

- With no change to assumptions, prove $p(b)$.
- Introduce the variable n , add $n \geq b$ and $p(n)$ to the list of assumptions, and prove $p(n+1)$.

Proof by strong induction

For this proof technique, we assume $X = \{n \in \mathbb{N} \mid n \geq b\}$ for some $b \in \mathbb{N}$. That is, the statement we are proving is of the form $\forall n \geq b, p(n)$.

- With no change to assumptions, prove $p(b)$.
 - Introduce the variable n , add the following to the list of assumptions:
 - ◊ $n \geq b$;
 - ◊ For all k with $b \leq k \leq n$, $p(k)$ is true.
- ... and then prove $p(n+1)$.

Proof by infinite descent

For this proof technique, we assume $X = \mathbb{N}$. That is, the statement we are proving is of the form $\forall n \in \mathbb{N}, p(n)$.

- Introduce the variable n . Add $n \in \mathbb{N}$ and $\neg p(n)$ to the list of assumptions. Prove:
 - ◊ $n > 0$;
 - ◊ $\exists m < n, p(m)$; see [A.1.8](#).

A.1.8 Existential quantifiers: $\exists x \in X, p(x)$

Direct proof

- With no change to assumptions, find some element $x \in X$ which makes the formula $p(x)$ true, and prove that this is so.

Appendix B

Typesetting mathematics in L^AT_EX

Being able to type up your mathematical writing is a beneficial skill to have; unfortunately, most Office-style WYSIWYG (‘what you see is what you get’) text editors are not designed for this task—they just can’t cope with all the notation.

L^AT_EX¹ is a markup language that allows you to input both text and mathematical notation, the latter in the form of code, which is then beautifully typeset. What follows is a brisk intro to L^AT_EX, that should suffice for the purposes of this course.

B.1 Finding the software

There are several good L^AT_EX editors that you can install on your computer—I recommend Texmaker (<http://www.xmlmath.net/texmaker/>) if you’re new to it, because it’s cross-platform and fairly intuitive.

There are also online editors that you can use if you want to avoid installing new software; I highly recommend ShareLaTeX (<http://www.sharelatex.com>), which is free to use and stores your `.tex` files on the cloud.

There is some faffing around to be done with document headers and so on. So that you can avoid this, I’ve uploaded template `.tex` files to Blackboard for you to use in your homework and final project write-ups.

¹The word L^AT_EX is pronounced like ‘lay-tek’ or ‘lah-tek’; some people pronounce the ‘k’ like the German ‘ch’ sound, meant to resemble the Greek letter chi (χ). It doesn’t really matter, but if you pronounce it like ‘lay-teks’ then people will think you’re talking about something somewhat different.

B.2 Text mode and math mode

Before we get into the nitty-gritty, I should mention the difference between ‘text mode’ and ‘math mode’.

- **Text mode** is the default mode: the stuff you type will appear as text, and this is the mode you should use when writing anything that isn’t mathematical notation.
- You should use **math mode** when you’re typing anything which is mathematical notation, including variables, numbers, fractions, square roots, powers, sums, products, binomial coefficients, and so on.

To enter math mode, enclose whatever mathematical notation you are writing with dollar signs (\$). For example, if I type `$E=mc^2$` then L^AT_EX shows $E = mc^2$. Sometimes it is convenient to put longer expressions on their own line, in which case you can enclose it with double-dollar signs (\$\$); for example, if I type `$$a^2+b^2+c^2=ab+bc+ca$$` then L^AT_EX displays

$$a^2 + b^2 + c^2 = ab + bc + ca$$

on a line all of its own.

If you need to type text inside math mode (enclosed by \$ signs), you can do that using `\text{...}`, for example the code

```
$$\sum_{i=1}^n i = \frac{n(n+1)}{2} \text{ for all } n \in \mathbb{N}$$
```

gives

$$\sum_{i=1}^n i = \frac{n(n+1)}{2} \text{ for all } n \in \mathbb{N}$$

Note the spaces before and after ‘for all’; had I left those out of the code, they would not appear because L^AT_EX ignores spacing in math mode. You can force a space by putting a backslash before a space, for example `$a b$` gives ab but `$a\ b$` gives $a b$.

All mathematical notation should be in math mode, including single variables. Notice the difference between the following two lines:

If a and b are both even then so is a+b.

If a and b are both even then so is $a + b$.

While the first is written entirely in text mode, the second is written using math mode for the variables and + sign.

B.3 Mathematical symbols

The following table lists—I hope—all of the mathematical symbols you will need in this course. If you come across other symbols, please let me know.

Logic		
conjunction, disjunction	$\wedge, \vee$	<code>\wedge, \vee</code>
negation	$\neg$	<code>\neg</code>
implication, biconditional	$\Rightarrow, \Leftrightarrow$	<code>\Rightarrow, \Leftrightarrow</code>
exclusive disjunction	$\oplus$	<code>\oplus</code>
true, false (in truth table)	$\checkmark, \times$	<code>\checkmark, \times</code>
quantifiers (universal, existential)	$\forall, \exists$	<code>\forall, \exists</code>
Set theory		
element, subset	$\in, \subseteq$	<code>\in, \subseteq</code>
not equal, proper subset	$\neq, \subsetneq$	<code>\neq, \subsetneq</code>
intersection, (indexed)	$\cap, \bigcap_{i=1}^n$	<code>\cap, \bigcap_{i=1}^n</code>
union, (indexed)	$\cup, \bigcup_{i=1}^n$	<code>\cup, \bigcup_{i=1}^n</code>
relative complement, complement	$X \setminus Y, X^c$	<code>\setminus, X^c</code>
product, (indexed)	$\times, \prod_{i=1}^n$	<code>\times, \prod_{i=1}^n</code>
implied lists	$\{1, \dots, n\}$	<code>\{ 1, \dots, n \}</code>
indexed sets	$\{x_i \mid i \in I\}$	<code>\{ x_i \mid i \in I \}</code>
set-builder notation	$\{x \mid p(x)\}$	<code>\{ x \mid p(x) \}</code>
empty, universal set	$\emptyset, \mathcal{U}$	<code>\varnothing, \mathcal{U}</code>
number sets	$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$	<code>\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}</code>
Numbers and combinatorics		
multiplication	$m \times n, m \cdot n$	<code>\times, \cdot</code>
fractions, exponents	$\frac{m}{n}, m^n$	<code>\frac{m}{n}, m^n</code>
order relations	$\leq, \geq$	<code>\leq, \geq</code>
divisibility, (non-)	$m \mid n, m \nmid n$	<code>\mid, \nmid</code>
binomial coefficient	$\binom{n}{k}$	<code>\binom{n}{k}</code>
indexed sum, product	$\sum_{i=1}^n a_i, \prod_{i=1}^n a_i$	<code>\sum_{i=1}^n a_i, \prod_{i=1}^n a_i</code>
modular arithmetic	$a \equiv b \pmod{n}, \not\equiv$	<code>a \equiv b \pmod{n}, \not\equiv</code>
Functions and relations		
functions	$f : X \rightarrow Y, X \xrightarrow{f} Y$	<code>\to, X \xrightarrow{f} Y</code>
composition	$g \circ f$	<code>\circ</code>
isomorphism	$\cong$	<code>\cong</code>
equivalence relations	$\sim, \approx$	<code>\sim, \approx</code>
Structured sets		
order relation	$\preceq, \prec$	<code>\preceq, \prec</code>
group operations	$\cdot, \star, \circ$	<code>\cdot, \star, \circ</code>

B.4 Organisation and formatting

When typing up solutions to problems, and particularly when typing up your final project, organisation can be the difference between a masterpiece and an unreadable heap of notation. Here are some tips to help you organise your work:

Sections and paragraphs

You can split your work up into sections, subsections, subsubsections, and even subsubsubsections. To do this, use `\section{Section title}` or `\section*{Section title}`; the former includes a section number, and the latter omits it. To start a new paragraph, simply make two new lines in the code.

Bulleted and enumerated lists

Sometimes it is useful to use bullet points or give an enumerated list. For example, in these notes, I separate the base case from the induction step in proofs by induction by using bullet points.

For a bulleted list you can use the `itemize` environment:

<pre>\begin{itemize} \item Something here\dots \item You can do lists within lists: \begin{itemize} \item Like this. \item Isn't it crazy! \end{itemize} \item Well, not that crazy. \end{itemize}</pre>	• Something here... • You can do lists within lists: ◊ Like this. ◊ Isn't it crazy! • Well, not that crazy.
--	--

For an enumerated list, you can use the `enumerate` environment. You can play around with different methods of enumeration, which you specify in square brackets [...]; personally I like (1), (i) and (a) the best:

<pre>\begin{enumerate}[(a)] \item Here's the first thing; \item Here's the second thing; \item And here's the third thing. \end{enumerate}</pre>	(a) Here's the first thing; (b) Here's the second thing; (c) And here's the third thing.
--	--

Definitions and results

If you use the provided templates, you can make definitions, and state and prove theorems, using the following environments:

`definition`, `example`, `proposition`, `theorem`, `lemma`, `corollary`, `proof`

They are given a number **m.n**, where m is the section number and n is the position within the section.

Here's an example of a theorem appearing in the third section of a document, in which five definitions, results or examples come before it:

```
\begin{theorem}
If  $n \in \mathbb{N}$  then  $n \geq 0$ .
\end{theorem}

\begin{proof}
This is really obvious.
\end{proof}
```

Theorem 3.6. If $n \in \mathbb{N}$ then $n \geq 0$.
Proof. This is really obvious. $\square$

Note that the box ($\square$) designating the end of the proof is inserted automatically when you close the `proof` environment.

Labels

As you change the contents of a document, the numbering of the definitions, examples and results might change. To refer to a specific result, instead of typing the number and having to change it each time the number changes, you can use the `\label` and `\ref` commands.

An example of this in action is as follows:

```
\begin{definition}
\label{defDivides}
Say  $a \textbf{ divides } b$  if there
exists  $k \in \mathbb{Z}$  such that
 $ka=b$ .
\end{definition}

We will use Definition \ref{defDivides}
for absolutely nothing.
```

Definition 2.11. Say a **divides** b if there exists $k \in \mathbb{Z}$ such that $ka = b$.

We will use Definition 2.11 for absolutely nothing.

Formatting

To put the icing on the cake, you might want to make some words **bold** or *italicised*. This is simple: for bold text type `\textbf{text here}` and for italic text type `\textit{text here}`. In Texmaker and ShareLaTeX you can press **Ctrl+B** and **Ctrl+I** to avoid having to type all this out.

Aligned equations

Occasionally a proof may require you to demonstrate that two terms are equal by proving a sequence of intermediate equations. This can be done using the `align*` environment. New lines are introduced by inserting a double-backslash (`\\`), and the two columns are separated by an ampersand (`&`). The left column is aligned right, and the right column is aligned left. Here's an example:

<pre>\begin{align*}</pre>	
<pre>(n+1)! - n! &= (n+1)n! - n! \\</pre>	$(n+1)! - n! = (n+1)n! - n!$
<pre>&= n \cdot n! + n! - n! \\</pre>	$= n \cdot n! + n! - n!$
<pre>&= n \cdot n!</pre>	$= n \cdot n!$
<pre>\end{align*}</pre>	

Note that the `align*` environment automatically enters into math mode, so to enter text you should use the `\text` command.

More advanced formatting

I should take a moment to emphasise that what matters in this course is your ability to communicate mathematical arguments clearly and correctly. The $\text{\LaTeX}$ tools discussed so far in this section are more than sufficient for the purposes of this course.

However, if you are interested in pushing your $\text{\LaTeX}$ skills further or there is a formatting feature you're unsure about how to implement, then I recommend browsing or searching one of the following websites:

- <http://tex.stackexchange.com>
- <https://en.wikibooks.org/wiki/LaTeX>

B.5 Practice page

Use the provided L^AT_EX template `template.tex` to re-create the following page:

Squarefree integers

Carl Friedrich Gauss, Wednesday 14th September 1831

Introduction

When you've written this page, you will be unstoppable, at least as far as typesetting mathematics is concerned. You will need to implement:

- Text mode stuff: sections, paragraphs, text formatting, labels and references, lists;
- Math mode stuff: definitions and results, aligned equations, etc.

So let's get on with it!

1 Squarefree integers

1.1 Definition and an elementary result

Definition 1.1. An integer a is **squarefree** if it is divisible by no perfect square other than 1. That is, if n^2 divides a then $n^2 = 1$.

Proposition 1.2. A non-zero non-unit a is squarefree if and only if

$$a = p_1 \times p_2 \times \cdots \times p_n$$

for distinct primes $p_1, p_2, \dots, p_n$.

Proof. We leave the proof as an exercise to the reader. □

1.2 Some examples

Example 1.3. Some concrete examples include:

(i) 5610 is squarefree by Proposition 1.2, since

$$\begin{aligned} 5610 &= 10 \times 561 \\ &= (2 \times 5) \times (11 \times 17) \end{aligned}$$

(ii) 12 is not squarefree since $4 \mid 12$ and $4 = 2^2$.

1

Index

- absolute value, 53
- addition principle, 109
- antisymmetric relation, 86

- biconditional, 12, *142*
- bijective function, 101
- binomial coefficient, 114
- bound variable, 19

- canonical prime factorisation, 67
- Cantor's diagonal argument, 127
- cardinality, 129
- codomain
 - of a function, 93
 - of a relation, 82
- complement, 31
- composite of functions, 96
- congruence, 70
- congruence class, 89
- conjunction, 4, *141*
- contradiction, 13
- contraposition, 14
- coprime, 61
- corollary, 9
- countable set, 123
- counting principle, 107

- De Morgan's laws
 - for logical operators, 9
 - for quantifiers, 18
 - for sets, 32
- diagonal subset, 83
- disjoint, 109
- disjunction, 5, *141*
 - exclusive, 12

- division, **57**
- division theorem, 55
- domain
 - of a function, 93
 - of a relation, 82
- double-counting, 115

- element, 22
- empty function, 93
- empty relation, 82
- empty set, 25
- equivalence class, 88
- equivalence relation, 88
- Euclidean algorithm, 58
- Euler's theorem, 78
- excluded middle, 12

- factorial, 118
- Fermat's little theorem, 76
- finite set, 105
- free variable, 15
- function, 93
 - identity function, 93
 - injective, 99
 - surjective, 100

- graph
 - of a function, 94
 - of a relation, 82
- greatest common divisor, 57
- greatest element of a poset, 91

- image, 95
- implication, 10, *142*
- inclusion–exclusion principle, 110

- indexed family, 30
- induction principle
 - strong, 45
 - weak, 38
- infimum, 91
- infinite set, 105
- inhabited set, 25
- injective function, 99
- integer, 25
- intersection
 - indexed, 30
 - pairwise, 29
- inverse, 101
 - left inverse, 98
 - right inverse, 100
- irrational number, 27
- irreducible number, 64

- least element of a poset, 91
- lemma, 9
- logical equivalence, 7
- logical formula, 16
- logical operator, 3

- modular arithmetic, 72
- modulo, 70
- multiplication principle, 107
- multiplicity
 - of a prime, 67

- natural number, 25, **36**
- negation, 7, *142*
- non-zero non-unit, 57

- ordered pair, 34

- parameter, *see* free variable
- partial order, 90
- partition, 109
- Pascal's triangle, 116
- permutation, 101, 118
- poset, 90
- power set, 113

- preimage, 95
- prime
 - canonical prime factorisation, 67
- prime number, 63
- product of sets
 - indexed, 35
 - pairwise, 34
- proof, 2
- proposition, 2

- quantifier
 - existential, 18, *144*
 - universal, 16, *142*
- quotient
 - of a set by an equivalence relation, 88
 - of numbers, 56

- range of a variable, 16
- rational number, 26
- real number, 26
- recursion, 37
- reducible number, 64
- reflexive relation, 85
- relation, 82
 - antisymmetric, 86
 - reflexive, 85
 - symmetric, 85
 - transitive, 86
- remainder, 56
- restriction, 97
- ring, 133

- sequence, 35
- set, 22
 - universal, 22
- set equality, 28
- set-builder notation, 23
- sign, 67
- size
 - of a finite set, 105
- subset, 27
 - k -element subset, 113
- substitution, 15

- supremum, 91
- surjective function, 100
- symmetric relation, 85

- theorem, 9
- totient, 77
- transitive relation, 86
- truth table, 5

- union
 - indexed, 31
 - pairwise, 31
- unit, 57
- universe of discourse, 22

- well-ordering principle, 48

- xor, 12