Math 13 — An Introduction to Abstract Mathematics

Neil Donaldson & Alessandra Pantano

Summer Session II 2014 (Aug 22nd)

Useful Texts

- *Book of Proof*, Richard Hammack, 2nd ed 2013. Available free online! Very good on the basics: if you're having trouble with reading set notation or how to construct a proof, this book's for you! These notes are *deliberately* pitched at a high level relative to this textbook to provide contrast.
- *Mathematical Proofs: A Transition to Advanced Mathematics*, Chartrand/Polimeni/Zhang, 3rd Ed 2013, Pearson. Ostensibly the course text. Has many, many exercises; the first half is fairly straightforward while the second half is much more complex and dauntingly detailed.
- *The Elements of Advanced Mathematics*, Steven G. Krantz, 2nd ed 2002, Chapman & Hall and *Foundations of Higher Mathematics*, Peter Fletcher and C. Wayne Patty, 3th ed 2000, Brooks–Cole are old course textbooks for Math 13. Both are readable and concise with good exercises, if a little weak on the elementary material.

1 Introduction

This course has several purposes:

1. Developing the skills necessary to read and practice mathematics, especially pure mathematics.
2. Understanding the concept of proof, and of several distinct proof techniques.
3. Learning what sort of questions mathematicians ask, what excites them, and what they are looking for.
4. Introducing upper division mathematics by giving a taste of the material in several different classes.

For many students this course is a game-changer. A crucial part of the course is the acceptance that mathematics is very different to what is presented at grade-school and in the calculus sequence. What you've largely studied in the past has been the application of mathematics to real world problems (recall that the entitity of calculus is motivated by the distance-velocity problem). What distinguishes a Mathematician is a comparative lack of interest in the computation of some quantity compared to an understanding of what that quantity means, why it is important, and how it relates to other concepts. As an example, consider Calc I and what you (hopefully!) enjoyed about it. Were you were more interested in being able to compute derivatives, or by the ways in which a function can fail to be differentiable? The former is incredibly important in broad contexts, but is inherently boring to

professional mathematicians. It is the second issue that keeps us interested. If you want to succeed in and enjoy upper division mathematics, you have to appreciate the kind of questions each course is asking, and why the answers are important.

Proof

Mathematics is one of the very few disciplines in which one can categorically say that something is *true* or *false*, or that something has been *proved*. A greasy salesman in a TV advert may claim that they have *proved* that a certain cream makes you look younger; a defendant may be *proved* guilty in court; the gravitational constant *is* 9.81ms^{-2}. Ask yourself what these statements mean. The advert is just trying to sell you something, but push harder and they might come up with some justification: e.g. '100 people used the product for a month and 75 of them claim to look younger.' Is this *proof*? Is a defendant really guilty of a crime just because a court has found them so; have there never been any miscarriages of justice? Is the gravitational constant really equal to precisely 9.81 meters per second squared, or is this merely a good approximation? This kind of pedantry may seem over the top in everyday life: indeed most of us would agree that if 75% of people think a cream helps, then it probably is doing something. In mathematics and philosophy, we think very differently: the concepts of true and false and of proof are very precise.

How do mathematicians reach this blissful state where everything is either right or wrong and, once proved, is proved forever? The answer, rather disappointingly, is by cheating. *Nothing* in mathematics is true except with reference to some assumption. For example, consider the following theorem:

Theorem 1.1. *The sum of any two even integers is even.*

We all believe that this is true, but can we *prove* it? What we need is a *definition* of what an even number is.[1]

Definition 1.2. An integer is *even* if it may be written in the form $2n$ where n is an integer.

The proof of the theorem flows straight from the definition.

Proof. Let x and y be *any* two even integers. We want to show that $x + y$ is an even integer.
By definition, an integer is even if it can be written in the form $2k$ for *some* integer k. Thus there exist integers n, m such that $x = 2n$ and $y = 2m$. We compute:

$$x + y = 2n + 2m = 2(n + m).$$

Because $n + m$ is an integer, this shows that $x + y$ is an even integer. ∎

There are several important observations:

- 'Any' in the statement of the theorem means the proof must work *regardless* of what even numbers you choose. It is not good enough to simply select, for example, 4 and 16, then write $4 + 16 = 20$. This is an *example* of the theorem, not a proof.
- According to the definition, $2m$ and $2n$ together represent *all possible* even numbers.

[1] And more fundamentally what 'addition' and 'integer' mean.

- The proof makes direct reference to the definition. The vast majority of the proofs in this course are of this type. If you know the definition of everything in the theorem, you can often discover a proof simply by writing down the definition of everything in the theorem.
- It is hard to construct a proof unless you have *variables.* In this case the variables m and n come for free *once you write the definition of evenness!*

The important link between theorems and definitions is largely what learning higher-level mathematics is about. We prove theorems (and solve homework problems) because they make us use and understand the subtleties of the definitions. One does not *know* mathematics, one *does* it. Mathematics is a *practice;* an art as much as it is a science.

Conjectures

In this course, you will also discover one of the most creative and fun aspects of mathematics, that is the art of formulating, proving and disproving conjectures. To get a taste, consider the following:

Conjecture 1.3. *If n is any odd integer, then $n^2 - 1$ is a multiple of 8.*

and

Conjecture 1.4. *For every positive integer n, the integer $n^2 + n + 41$ is prime.*

Are these claims true or false? How can we decide? Well, on a first attempt, we may try to plug in some integers and test the conjecture for some small values of n.

	$n=1$	$n=3$	$n=5$	$n=7$	$n=9$
n^2-1	0	8	24	48	80

and

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
n^2+n+41	43	47	53	61	71

Because 0, 8, 24, 48 and 80 are all multiples of 8, and 43, 47, 53, 61 and 71 are all prime, both conjectures appear to be true. Would you bet $100 that this is - indeed - the case?
Is $n^2 - 1$ a multiple of 8 *for every* odd integer n? Is $n^2 + n + 41$ prime *for every* positive integer n? The only way to establish whether a conjecture is true or false is to *prove it* (by showing it must be true in all cases) or *disprove it* (by finding at least one instance in which the conjecture is false).
Let us try to work with conjecture 1.3. If n is an odd integer, then—by definition—we can write it as $n = 2k + 1$ for some integer k. Then

$$n^2 - 1 = (2k+1)^2 - 1 = (4k^2 + 1 + 4k) - 1 = 4k^2 + 4k.$$

We need to investigate whether this is *always* a multiple of 8. Because

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

is already a multiple of 4, it all comes down to deciding whether or not $k^2 + k$ contains a factor 2 (that is, $k^2 + k$ is even) for all possible choices of k. Do we believe this? Trying out some small values of k gives

	$k=-2$	$k=-1$	$k=0$	$k=1$	$k=2$
k^2+k	2	0	0	2	6

Once again, the claim seems to be true for some small value of k... but how do we know it is true for *all* k? Well, the only way is to *prove it* or *disprove it*.
How to proceed? The question here is whether or not k^2+k is *always* even. Factoring out a k, we get:

$$k^2+k = k(k+1).$$

We have expressed k^2+k as a product of two consecutive integers. This is great, because for any two consecutive integers, one is even and the other is odd, so the product is even. We have now proven that the conjecture is true. Conjecture 1.3 is—indeed—a theorem!
Let us formalize the argument above into a proof.

Theorem 1.5. *If n is any odd integer, then n^2-1 is a multiple of 8.*

Proof. Let n be any odd integer; we want to show that n^2-1 is a multiple of 8. By definition of *odd* integer, we can write $n = 2k+1$ for some integer k. Then

$$n^2-1 = (2k+1)^2-1 = (4k^2+1+4k)-1 = 4k^2+4k = 4k(k+1).$$

We distinguish two cases. If k is even, say $k = 2t$ for some integer t, then

$$n^2-1 = 4k(k+1) = 4(2t)(2t+1) = 8t(t+1).$$

Because $t(t+1)$ is an integer, n^2-1 is a multiple of 8. Similarly, if k is an odd integer, then we can write $k = 2s+1$ for some integer s and we get

$$n^2-1 = 4k(k+1) = 4(2s+1)(2t+2) = 8(2s+1)(s+1).$$

Since $(2s+1)(s+1)$ is an integer, n^2-1 is a multiple of 8. This concludes the proof. ∎

It is now time to explore conjecture 1.4. The question here is whether or not n^2+n+41 is a prime integer for *every* positive integer n. We know that when $n = 1$, 2, 3, 4 or 5 the answer is yes... but a bunch of examples do not make a proof. At this point, we do not even know whether the claim is true or false.
Let us investigate the question further. Suppose that n is any positive integer; we must ask ourselves whether it is possible to factor n^2+n+41 as a product of two integers, none of which is one. (Once again, we rely on a definition!)

When $n = 41$ such a factorization certainly exists, since we can write:

$$41^2+41+41 = 41(41+1+1) = 41 \cdot 43.$$

Our *counterexample* shows that there exists at least one value of n for which n^2+n+41 is *not* prime. Thus, we have disproved the conjecture that 'for all positive integers n, n^2+n+41 is prime'. Conjecture 1.4 was false!
The moral of the story is that to show that a conjecture is true you must prove that it holds for all the cases in consideration, but to show that it is false a single counterexample suffices!

Now we know a little of what mathematics is about, it is time to practice some of it!

2 Logic and the Language of Proofs

2.1 Propositions

Definition 2.1. A *proposition* or *statement* is a sentence that is either true or false.

Examples. 1. $17 - 24 = 7$.

2. 39^2 is an odd integer.

3. The moon is made of cheese.

4. Every cloud has a silver lining.

5. God exists.

All these propositions require to make sense is a clear *definition* of every concept they contain. There are many concepts of God in many cultures, but once it is decided *which* we are talking about, clearly they either exist or do not. This example illustrates that a question need not be indisputably answerable (by us) in order to qualify as a proposition. Indeed mostly when people argue over propositions and statements, what they are really arguing over are the defintions!

Anything that does not have a true or false answer is not a proposition. 'January 1^{st}' is not a proposition, neither is 'Green.'

Truth Tables

Often one has to deal with abstract propositions; those where you do not know the truth or falsity. In such cases it is often convenient to represent the combinations of propositions in a tabular format. For instance, if we have two propositions (P and Q), or even three (P, Q and R) then all possibile combinations of truth T and falsehood F are represented in the following tables:

P	Q
T	T
T	F
F	T
F	F

P	Q	R
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

The mathematician in you should be looking for the pattern and asking: how many rows would a truth table corresponding to n propositions have, and can I *prove* my assertion? Right now it is hard to prove that the answer is 2^n rows: induction (later in the course) makes this very easy…

Connecting Propositions: Conjunction, disjunction and negation

We now *define* how to combine propositions in certain natural ways, modeled on the words 'and' and 'or'.

Definition 2.2. Let P and Q be propositions. The *conjunction* (AND, $\wedge$), the *disjunction* (OR, $\vee$) of P and Q, and the *negation* or *denial* (NOT, $\neg$, $\sim$, $\bar{}$) of P are defined as follows:

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

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

P	$\neg P$
T	F
F	T

It is best to use 'and, or, not' when speaking about these concepts: conjunction, disjunction and negation may make you sound educated, but at the serious risk of getting them confused!

Example. Let P, Q & R be the following propositions:

P: Irvine is a city in California.

Q: Irvine is a town in Ayrshire, Scotland.

R: Irvine has seven letters.

Clearly P is true while R is false. If you happen to know someone from Scotland, you might know that Q is true…[2] We can now compute the following (increasingly grotesque) combinations…

$P \wedge Q$	$P \vee Q$	$P \wedge R$	$\neg R$	$(\neg R) \wedge P$	$\neg(R \vee P)$	$(\neg P) \vee [((\neg R) \vee P) \wedge Q]$
T	T	F	T	T	F	T

How did we establish these facts? Some are quick, and can be done in your head. Consider, for instance, the statement $(\neg R) \wedge P$. Because R is false, $\neg R$ is true. Thus $(\neg R) \wedge P$ is the disjunction of two true statements, hence it is true. Similarly, we can argue that $R \vee P$ is true (because R is false and P is true), so the negation $\neg(R \vee P)$ is false. Establishing the truth value of the proposition $(\neg P) \vee [((\neg R) \vee P) \wedge Q]$ requires more work. You may want to set up a truth table with several auxiliary columns:

P	Q	R	$\neg P$	$\neg R$	$(\neg R) \vee P$	$((\neg R) \vee P) \wedge Q$	$(\neg P) \vee [((\neg R) \vee P) \wedge Q]$
T	T	F	F	T	T	T	T

The importance of parentheses in a logical expressions cannot be stressed enough. For instance, build a truth table for the propositions $P \vee (Q \wedge R)$ and $(P \vee Q) \wedge R$. Are

Other logical connectives: Conditional and biconditional

In oder to logically set up proofs, we need to see how propositions can lead one to another.

Definition 2.3. The *conditional* ($\Rightarrow$) and *biconditional* ($\Longleftrightarrow$) *connectives* have the truth tables

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

P	Q	$P \Longleftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

[2]The first Californian antecedent of the Irvine family which gave its name to UCI was an Ulster-Scotsman named James Irvine (1827–1886).

For $P \Rightarrow Q$, P is the *hypothesis* and Q the *conclusion*.

It will take a while to get used to the first of these tables. In particular, notice that if P is false, the implication $P \Rightarrow Q$ is always true (independently of whether Q is true or false). The second table should be easy to remember: $P \iff Q$ is true precisely when P and Q are identical.

Why is $F \Rightarrow T$ considered *true?*

Here's a mathematical example to help you understand the idea that $F \Rightarrow T$ is true, written with an English translation at the side.

$$\begin{aligned} 7 = 3 &\Rightarrow 0 \cdot 7 = 0 \cdot 3 && \text{(If } 7 = 3\text{, then 0 times 7 equals 0 times 3)} \\ &\Rightarrow 0 = 0 && \text{(then 0 equals 0)} \end{aligned}$$

Thus $7 = 3 \Rightarrow 0 = 0$. Logically speaking this is a perfectly correct argument, thus the *implication* is true. $7 = 3$ is clearly false however, so the argument makes us uncomfortable.

If you want to understand connectives more deeply than this, then take a logic or philosophy course! For example, although neither statement makes the least bit of sense in English;

$$\begin{gathered} 17 \text{ is odd} \Rightarrow \text{Mexico is in China is } \textit{false,} \\ \text{whilst} \\ 17 \text{ is even} \Rightarrow \text{Mexico is in China is } \textit{true.} \end{gathered}$$

Such bizarre constructions are happily beyond the point and consideration of this course!

In terms of truth tables, connectives are not all that useful. The intuition is the following: a *theorem* is an assertion that $P \Rightarrow Q$ is true, where you assume that P is true to begin with. Looking at the truth table , the only possibility of both P and $P \Rightarrow Q$ being simultaneously true is if Q is true. This leads to one of the many synomyms for $P \Rightarrow Q$: P implies Q.

Synonyms

$\Rightarrow$ and $\iff$ can be read in English in many different sounding ways:

$P \Rightarrow Q$	$P \iff Q$
P implies Q	P if and only if Q (P iff Q)
Q if P	P and Q are equivalent
P only if Q	P is necessary and sufficient for Q
P is sufficient for Q	
Q is necessary for P	

For instance, the following propositions are all equivalent:

- If you are born in Rome, then you are Italian.
- You are Italian if you are born in Rome.
- You are born in Rome only if you are Italian.
- Being born in Rome is sufficient to be Italian.

- Being Italian is necessary for being born in Rome.

When proving a theorem it is often convenient to rewrite the statement as an implication of the form $P \Rightarrow Q$. As an example, consider the following:

Theorem 2.4. *The product of two odd integers is odd.*

We can write this theorem in terms of propositions and connectives:

- P is 'Two integers x and y are odd'. This is our *assumption*, the hypothesis.
- Q is 'The product of x and y is odd'. This is what we want to show, the conclusion.
- Showing that $P \Rightarrow Q$ is true, that (the truth of) P implies (the truth of) Q requires an argument. This is the *proof.*

Proof. Let x and y be *any* two odd integers. We want to show that product $x \cdot y$ is an odd integer. By definition, an integer is odd if it can be written in the form $2k + 1$ for *some* integer k. Thus there must be integers n, m such that $x = 2n + 1$ and $y = 2m + 1$. We compute:

$$x \cdot y = (2n + 1)(2m + 1) = 4mn + 2n + 2m + 1 = 2(2mn + n + m) + 1.$$

Because $2mn + n + m$ is an integer, this shows that $x \cdot y$ is an odd integer. ∎

The above is an example of a *direct proof*. This is the most straightforward method of proving an implication $P \Rightarrow Q$: We assume that the hypothesis (P) is true and, using definitions, we proceed to show that the conclusion (Q) is true.

The Converse and Contrapositive

The following constructions are used continually in mathematics: it is vitally important to know the difference between them!

Definition 2.5. The *converse* of an implication $P \Rightarrow Q$ is $Q \Rightarrow P$.
The *contrapositive* of $P \Rightarrow Q$ is $(\neg Q) \Rightarrow (\neg P)$.

Example. Let P be the statement 'Claudia is holding a peach' and Q be 'Claudia is holding a piece of fruit'. The implication $P \Rightarrow Q$ is true, since all peaches are fruit.
The *converse* of $P \Rightarrow Q$ is $Q \Rightarrow P$: 'If Claudia is holding a piece of fruit then she is holding a peach'. This is palpably false: Claudia could be holding an apple!
The *contrapositive* of $P \Rightarrow Q$ is: 'if Claudia is *not* holding any fruit, then she is *not* holding a peach'. This is clearly true.

What the example illustrates is that the contrapositive has *the same truth table* as the original implication. Therefore, they have the same meaning, and we say that $P \Rightarrow Q$ and $(\neg Q) \Rightarrow (\neg P)$ are *logically equivalent*.

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

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

The usefulness of this is hard to understate for it allows us to prove results in a different way. Consider another basic theorem:

Theorem 2.6. *If the sum of two integers is odd, then exactly one of them is odd.*

The statement of the theorem is an implication of the form $P \Rightarrow Q$. Here

- P: 'The sum $x + y$ of integers x and y is odd'.
- Q: 'Exactly one of x or y is odd'.

A direct proof would require that we assume P and we show Q. The problem is that it is hard to work with these propositions, especially Q. The negation of Q is, however, much easier:

- $\neg Q$: 'x and y are both even or both odd (that is, they have the same parity)'.
- $\neg P$: 'The sum $x + y$ of integers x and y is even'.

Therefore, we choose to prove the contrapositive instead: $\neg Q \Rightarrow \neg P$. (This is, after all, equivalent to proving the original implication).

Proof. There are two cases: x and y are both even, or both odd.
Case 1: Let $x = 2m$ and $y = 2n$ be even. Then $x + y = 2(m + n)$ is even.
Case 2: Let $x = 2m + 1$ and $y = 2n + 1$ be odd. Then $x + y = 2(m + n + 1)$ is even. ∎

The above is an example of a *proof by contrapositive*.

De Morgan's Laws

Two of the most famous results in logic are attributable to Augustus De Morgan, a very famous 19th century logician. They are a lot easier to understand in examples than to read in logical notation.

Theorem 2.7 (de Morgan's laws)**.** *Let P and Q be any propositions. Then:*

- $\neg(P \wedge Q) \iff \neg P \vee \neg Q$
- $\neg(P \vee Q) \iff \neg P \wedge \neg Q$

The way one reads, and proves, these statements is that the expressions on each side have the *same truth table.* Here is a proof of the first law. Try the second on your own.

Proof.

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

Simply observe that the fourth and seventh columns are identical. ∎

It is worth pausing to notice how similar the two laws are, and how concise. There is some beauty there. With an English example these laws are much easier to comprehend.

Example. (Of the first law) Suppose that of a morning you can choose (or not) to ride the subway to work, and you can choose (or not) to have a cup of coffee. The opposite (negation) of 'I rode the subway and had coffee' is 'I didn't ride the subway or I didn't have coffee'. Note that the mathematical use of 'or' includes the possibility that you neither rode the subway nor had coffee.

You will see these laws again when we think about sets.

Negating Conditionals

Often one wants to understand the negation of a statement. In particular, it is important to understand the negation of a contitional 'P implies Q' What is the opposite of this? Is it 'P doesn't imply Q'? And what could this mean? To answer the question, you can use truth tables, or just think.

Here is the truth table of $P \Rightarrow Q$ and its negation: recall that negation simply swaps T and F in the truth table.

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

The only time there is a T in the final column is when *both* P is true *and* Q is false. We have therefore proved the following:

Theorem 2.8. *$\neg(P \Rightarrow Q)$ is equivalent to $P \wedge \neg Q$ (read 'P and not Q').*

According to de Morgan's laws this says that $P \Rightarrow Q$ is equivalent to $\neg P \vee Q$.

Now think rather than calculate. What is the opposite of the implication, 'It's the morning therefore I'll have coffee'? Hopefully it is clear that the negation is,

'It's the morning *and* I *won't* have coffee.'

Warning! The negation of $P \Rightarrow Q$ is *not* the converse $Q \Rightarrow P$. Write down the truth table for the converse and compare with the above to check!

Example. What is the negation of, 'If x is even then x^2 is even'?
Written in terms of propositions, we have:
P: x is even
Q: x^2 is even
and we want the negation of $P \Rightarrow Q$. This is $P \wedge \neg Q$, or

'There is some x which is even and such that x^2 is odd.'

It should be clear both from truth tables and just by thinking that out claimed negation is the correct one. Until it is, practice with your own sentences! One way to help check is if you happen to know whether $P \Rightarrow Q$ is true or not. Here 'x even $\Rightarrow$ x^2 even' is certainly true, so it's negation must be *false.* Even reading the negation should make you a little uncomfortable.

Tautologies and Contradiction

There are two final related concepts that are helpful for understanding *why* proofs work.

Definition 2.9. A *tautology* is a logical expression that is always true, *regardless* of what the component statments might be.
A *contradiction* is a logical expression that is always false.

The easiest way to detect these is timply to construct the truth table.

Example. $P \wedge (\neg P)$ is a very simple contradiction:

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

Whatever P is, it cannot be true at the same time as its negation.

Of more importance to the methods of proof that we'll see shortly, are the following two results:

Theorem 2.10. *1.* $(P \wedge (P \Rightarrow Q)) \Rightarrow Q$ *is a tautology.*

2. $(P \wedge \neg Q) \wedge (P \Rightarrow Q)$ *is a contradiction.*

Proof. Again we compute only the first: write the truth table for the second out for yourself.

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

■

What is the meaning of this logical soup? Quite simply, if P is true and so is the implication $P \Rightarrow Q$, then Q must also be true. This is the essence of a proof: assume the hypothesis P is true, provide an argument to demonstrate the truth of the implication $P \Rightarrow Q$, and conclude that Q is true. Once you have done a couple of proofs by contradiction, you should revisit the second result.

2.2 Methods of Proof

We have already seen several proofs. A *proof* is an argument to show that $P \Rightarrow Q$ is true for some propositions P, Q.

There are four standard methods of proof—in practice, long proofs will use several of these.

Direct Assume P and logically deduce Q.

Contrapositive Assume $\neg Q$ and deduce $\neg P$. This is enough since the contrapositive $\neg Q \Rightarrow \neg P$ is logically equavalent to $P \Rightarrow Q$.

Contradiction Assume that P and $\neg Q$ are true and show *a contradiction*. Deduce that $P \wedge \neg Q$ (P and (not Q)) must be false. Because $P \wedge \neg Q$ is equivalent to $\neg(P \Rightarrow Q)$, this is enough to conclude that $P \Rightarrow Q$ must be true. (Theorem 2.8).

Induction This has a completely different flavor: we will consider it later in the course.

The direct method has the advantage of being easy to follow logically. The contrapositive has its advantage in that it may be very difficult to work directly with the propositions P, Q, especially if one or both involve the *non-existence* of something. Working with their negations might give you the existence of ingredients with which you can calculate. Proof by contradiction has a similar advantage: assuming both P and $\neg Q$ gives you two pieces of information with which you can calculate. Logically speaking there is no difference between the first three methods, beyond how one visualizes the argument.

To illustrate the difference between direct proof, proof by contrapositive, and proof by contradiction, we prove the same simple theorem in three different ways.

Theorem 2.11. *Suppose that x is an integer. If $3x + 5$ is even, then $3x$ is odd.*

Direct Proof. We show that if $3x + 5$ is even then $3x$ is odd. Assume that $3x + 5$ is even, then $3x + 5 = 2n$ for some integer n. Hence

$$3x = 2n - 5 = 2(n - 3) + 1.$$

This is clearly odd, because it is of the form 'an even number plus one'. ∎

Contrapositive Proof. We show that if $3x$ is even then $3x + 5$ is odd. Assume that $3x$ is even, and write $3x = 2n$ for some integer n. Then

$$3x + 5 = 2n + 5 = 2(n + 2) + 1 \text{ is odd, because } n + 2 \text{ is an integer.}$$

∎

Contradiction Proof. We assume that $3x + 5$ and $3x$ are both even, and we show a contradiction. Write $3x + 5 = 2n$ and $3x = 2k$ for some integers n and k. Then

$$5 = (3x + 5) - 3x = 2n - 2k = 2(n - k) \text{ is clearly even. This is a contradiction.}$$

∎

Some simple proofs

We now give several examples of simple proofs. The only notation needed to speed things along is that of some basic sets of numbers: $\mathbb{N}$ for the positive integers, $\mathbb{Z}$ for the integers, $\mathbb{R}$ for the real numbers,[3] and $\in$ for 'is a member of the set'. Thus $2 \in \mathbb{Z}$ is read as '2 is a member of the set of integers', or more concisely, '2 is an integer'.

Theorem 2.12. *Integers m and n are both odd if and only if the product mn is odd.*

There are two theorems here:

($\Rightarrow$) If integers m and n are both odd, then the product mn is odd.

($\Leftarrow$) If the product mn of integers m and n is odd, then both integers are odd.

Most of the time when there are two directions, you'll have to prove them separately.

[3] We will revisit all of these in the next section.

Proof. ($\Rightarrow$) Prove directly. Let m, n be odd. Then $m = 2k + 1$ and $n = 2l + 1$ for some $k, l \in \mathbb{Z}$. Then

$$mn = (2k + 1)(2l + 1) = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1$$

is odd, because $2kl + k + l \in \mathbb{Z}$.

($\Leftarrow$) Prove by the contrapositive. Suppose that the integers m and n are *not* both odd, that is, assume that at least one of them is even. We show that the product mn is even. Without loss of generality,[4] we can assume that n is even, so that $n = 2k$ for some integer k. Then

$$mn = m(2k) = 2(mk) \text{ is even.}$$

■

Note how we do not need to inquire whether m is even or odd: if n is even, the product mn will be even regardless.

The second part of this would have been very difficult to do directly: Assume mn is odd, then $mn = 2k + 1$, so…We are pretty much stuck!

Theorem 2.13. *If $3x + 5$ is even, then x is odd.*

We can prove this directly, by the contrapositive or by contradiction. We'll do all of them, so you can appreciate the difference.

Direct Proof. For a direct proof simply quote the two previous theorems. Because $3x + 5$ is even, $3x$ must be odd by Theorem 2.11. Now, since $3x$ is odd, both 3 and x are odd by Theorem 2.12. ■

Contrapositive Proof. For the contrapositive, suppose that x is even. Then $x = 2m$ for some integer m and we get

$$3x + 5 = 6m + 5 = 2(3m + 2) + 1.$$

Because $(3m + 2)$ is an integer, $3x + 5$ is odd. ■

Contradiction Proof. For a proof by contradiction, suppose that both $3x + 5$ and x are even. We can write $3x + 5 = 2m$ and $x = 2k$ for some integers m and k. Then

$$5 = (3x + 5) - 3x = 2m - 6k = 2(m - 3k) \text{ is even. Contradiction.}$$

■

Proving things is a matter of taste. You should be able to see the advantages and disadvantages in either approach. The direct proof is more logically straightforward, but it depends on two previous results. The contrapositive and the contradiction are quicker and more self-contained, but they require a deeper familiarity with logic.[5]

[4] See 'Potential Mistakes!' below for what this means.

[5] For even more variety, here is a direct proof of Theorem 2.13 that does not use any previous theorem. Suppose $3x + 5$ is even, so $3x + 5 = 2n$ for some integer n. Then

$$x = (3x + 5) - 2x - 5 = 2n - 2x - 5 = 2(x - n - 3) + 1 \text{ is odd.}$$

As this examples illustrates, you often have a variety of possible approaches, and that makes proving theorems even more fun.

Potential Mistakes!

There are many common mistakes that you should be careful of avoiding. Here are two incorrect 'proofs' of the $\Rightarrow$ direction of Theorem 2.12.

Fake Proof 1. $m = 3$ and $n = 5$ are both odd, so $mn = 15$ is odd. ■

This is an *example* of the theorem, not a proof. Examples are critical to helping you understand and believe what a theorem says, but they are no substitute for a proof!

Fake Proof 2. Let $m = 2k + 1$ and $n = 2k + 1$ be odd. Then, $mn = (2k + 1)(2k + 1) = 2(2k^2 + 2k) + 1$ is odd. ■

The problem with this 'proof' is that it is not sufficiently general. Indeed m and n are supposed to be *any* odd integers, but by setting both of them equal to $2k + 1$, we've chosen m and n to be the same! Notice how the correct proof uses $m = 2k + 1$ and $n = 2l + 1$ instead.

Generality and 'Without Loss of Generality'

By *generality* we mean that we must make sure to consider all possibilities suggested in the hypothesis. The phrase *Without Loss of Generality*, often shorted to WLOG, is used when a choice is made which might at first appear to restrict things but, in fact, does not.

Think back to how this was used in the the proof of Theorem 2.12. If at least one of two integers m, n is even, then we lose nothing by assuming that it is the second integer n. The labels m, n are arbitrary: if n happened not to be odd, we can simply relabel the integers; changing their order so that the second one certainly is even.

The phrase WLOG is used to pre-empt a challenge to a proof in the sense of *Fake Proof 2*, as if to say to the reader; 'I've thought about this, and see no problem.'

Here's a ludicrous 'theorem' to illustrate another potential mistake.

Theorem (Fake Theorem)**.** *The only number is zero.*

Fake Proof. Let x be any number and let $y = x$, then

$$
\begin{aligned}
x^2 &= xy && \text{(Multiply both sides of } y = x \text{ by } x\text{)} \\
\Rightarrow x^2 - y^2 &= xy - y^2 && \text{(Subtract } y^2 \text{ from both sides)} \\
\Rightarrow (x - y)(x + y) &= (x - y)y && \text{(Factorize)} \\
\Rightarrow x + y &= y && \text{(Divide both sides by } x - y\text{)} \\
\Rightarrow x &= 0
\end{aligned}
$$

■

Everything is fine up to the third equation, but then we divide by zero! Don't let yourself get so overcome by logical manipulations that you forget to check the basics.

More simple proofs

Theorem 2.14. *Suppose $x \in \mathbb{R}$. Then $x^3 + 2x^2 - 3x - 10 = 0 \implies x = 2$.*

Logically we can prove this theorem using any of the three methods. All rely on your ability to factorize the polynomial:

$$x^3 + 2x^2 - 3x - 10 = (x-2)(x^2+4x+5) = (x-2)[(x+2)^2+1],$$

and partly on your knowledge that $ab = 0 \iff a = 0$ or $b = 0$ (see the homework).

Direct Proof. If $x^3 + 2x^2 - 3x - 10 = 0$, then $(x-2)[(x+2)^2+1] = 0$. Hence $x - 2 = 0$ or $(x+2)^2 + 1 = 0$. The second case is impossible, because $(x+2)^2 \geq 0$ so $(x+2)^2 + 1 > 0$. Therefore $x = 2$ is the only solution. ■

Contrapositive Proof. Suppose $x \neq 2$. Then $x^3 + 2x^2 - 3x - 10 = (x-2)[(x+2)^2+1] \neq 0$ since none of the factors is zero. ■

Contradiction Proof. Suppose that $x^3 + 2x^2 - 3x - 10 = 0$ and $x \neq 2$. Then

$$0 = x^3 + 2x^2 - 3x - 10 = (x-2)[(x+2)^2+1].$$

Since $x + 2 \neq 0$, $(x+2)^2 + 1$ must be zero. Because $(x+2)^2 + 1 \geq 1$ for all real numbers x, this is a contradiction. ■

On balance the contrapositive proof is probably the nicest, but you may decide for yourself.

Being Excessively Logical

Let's be clear about what P and Q are in the above proof:

$$P\text{: } \text{“}x^3 + 2x^2 - 3x - 10 = 0\text{”} \qquad Q\text{: } \text{“}x = 2\text{”}$$

You can make life very hard for yourself by going over the top with logic. For instance you may wish take a third proposition R: “$x \in \mathbb{R}$,” and state the theorem as $R \Rightarrow (P \Rightarrow Q)$. This way lies the path of pain. It's easier to simply assume that you're always dealing with real numbers as a universal constraint, and ignore it logically.

Indeed, one could always append a third proposition to the front of any theorem, namely, “all math I already know.” Try to resist the temptation to be so logical that your arguments become unreadable!

Theorem 2.15. *If $n \in \mathbb{Z}$ is divisible by $p \in \mathbb{N}$, then n^2 is divisible by p^2.*

Before trying to prove this, recall what 'n is divisible by p' means: that $n = pk$ for some integer k. With the correct definition, the proof is immediate.

Proof. We prove directly. Let n be divisible by p. Then $n = pk$ for some $k \in \mathbb{Z}$. Then $n^2 = p^2k^2$, and so n^2 is divisible by p^2. ■

Remember: state the definition of everything important in the theorem and often the proof will be staring you in the face.

The next proof involves breaking things into cases.

Theorem 2.16. *If n is an integer, then n^2 has remainder 0 or 1 upon dividing by 3.*

Proof. We again prove directly. There are three cases: n has remainder 0, 1 or 2 upon dividing by 3.

(a) If n has remainder 0. Then $n = 3m$ for some $m \in \mathbb{Z}$ and so $n^2 = 9m^2$ has remainder 0.

(b) If n has remainder 1. Then $n = 3m + 1$ for some $m \in \mathbb{Z}$ and so

$$n^2 = 9m^2 + 6m + 1 = 3(3m^2 + 2m) + 1 \quad \text{has remainder 1.}$$

(c) If n has remainder 2. Then $n = 3m + 2$ for some $m \in \mathbb{Z}$ and so

$$n^2 = 9m^2 + 12k + 4 = 3(3m^2 + 4m + 1) + 1 \quad \text{has remainder 1.}$$

Thus n^2 has remainder 0 or 1 and cannot have remainder 2. ∎

Theorem 2.17. *Prove that $x^{17} + 12x^3 + 13x + 3 = 0$ has no positive solution.*

Showing that something does not exist is an excellent time for a contrapositive or contradiction proof. Here we use contradiction. We can interpret theorem as an implication: If x is a solution of the equation $x^{17} + 12x^3 + 13x + 3 = 0$ (that is, if $x^{17} + 12x^3 + 13x + 3 = 0$), then $x \geq 0$. So we take P to be '$x^{17} + 12x^3 + 13x + 3 = 0$', and Q to be 'x is not positive'. The negation of Q is simply '$x > 0$'.

Proof. If $x > 0$ satisfies $x^{17} + 12x^3 + 13x + 3 = 0$, then all powers of x are positive, and so $x^{17} + 12x^3 + 13x + 3 > 0$. A contradiction. ∎

Note how quickly the proof is written: it assumes that you, and any reader, are familiar with the underlying logic of a contradiction proof without it needing to be spelled out.

If you recall the Intermediate and Mean Value Theorems from Calculus, you should be able to prove that there is exactly one solution to the above polynomial.

Here we give several proofs of a famous result: an inequality relating the Arithmetic and Geometric Means of two or more numbers.

Theorem 2.18 (AM-GM inequality)**.** *If x, y are positive numbers, then $\frac{x+y}{2} \geq \sqrt{xy}$ with equality iff $x = y$.*

Here we can perform almost the same argument as a direct proof or as a contradiction.

Proof. **Direct** Clearly $(x - y)^2 \geq 0$ with equality $\iff x = y$. Thus

$$\begin{aligned} x^2 - 2xy + y^2 \geq 0 &\Rightarrow (x^2 + 2xy + y^2) - 4xy \geq 0 \\ &\Rightarrow x^2 + 2xy + y^2 \geq 4xy \\ &\Rightarrow (x + y)^2 \geq 4xy \\ &\Rightarrow |x + y| \geq 2\sqrt{xy} \end{aligned}$$

$$\Rightarrow x + y \geq 2\sqrt{xy}$$
$$\Rightarrow \frac{x+y}{2} \geq \sqrt{xy}.$$

Notice how the implication signs ($\Rightarrow$) are stacked to make the argument easy to read!

Contradiction Suppose that $\frac{x+y}{2} < \sqrt{xy}$. Then, since $x + y \geq 0$, we have $(x+y)^2 < 4xy$, whence multiplying out and rearranging as above we conclude that

$$(x-y)^2 < 0,$$

a contradiction. Thus $\frac{x+y}{2} \geq \sqrt{xy}$.
Clearly $x = y$ gives equality, and, if we have equality, then $(x-y)^2 = 0$, from which we recover $x = y$. Hence result.

■

The second proof uses the same argument as the first, but in a different order. This is not always possible, and you have to be careful when trying it. The 'with equality iff $x = y$' part is proved the same way in both examples, showing that a proof typically combines multiple techniques. Most readers will likely agree that the direct proof is easiest to follow.

The general AM-GM inequality

This is harder to read, and perhaps the most accessible proof is much harder than the one we've just covered. It requires a little calculus, namely the first (or second) derivative test.

Theorem 2.19. *If $x_1, \ldots, x_n > 0$ then $\frac{x_1+\cdots+x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}$, with equality iff $x_1 = x_2 = \cdots = x_n$.*

Proof. Consider the function $f(x) = e^{x-1} - x$. Its derivative is $f'(x) = e^{x-1} - 1$, which is zero iff $x = 1$. The sign of the derivative changes from negative to positive at $x = 1$, whence this is a local minimum. There are no other critical points of f, and the domain of f is the whole real line, whence $x = 1$ is the location of the *global* minimum of f.

Since $f(1) = 0$, we have $e^{x-1} \geq x$ with equality iff $x = 1$. Now let $\mu = \frac{x_1+x_2+\cdots+x_n}{n}$. Applying the previous formula to $x = \frac{x_i}{\mu}$, we have

$$\frac{x_i}{\mu} \leq \exp\left(\frac{x_i}{\mu} - 1\right)$$
$$\Rightarrow \frac{x_1 \cdots x_n}{\mu^n} = \frac{x_1}{\mu} \cdots \frac{x_n}{\mu} \leq \exp\left(\frac{x_1}{\mu} - 1 + \cdots + \frac{x_n}{\mu} - 1\right) = \exp(n-n) = 1$$

from which $\mu \leq \sqrt[n]{x_1 \cdots x_n}$ follows.

■

The style of the proof is deliberate. If you are capable of following the logic, then you likely don't need all of the algebra spelled out! It is perfectly reasonable to ask: 'How the hell would I know to start the proof that way?' The answer is you don't. When you read a proof like this, appreciate that you are reading a distillation of thousands of attempts and improvements over many years. No-one came up with this argument as a first attempt!

Combining Theorems

Sometimes it is useful to break a proof into pieces, something akin to viewing a computer program as a collection of subroutines that you put together for the finale. Mathematics does this with *Lemmas:* little theorems that are often individually unimportant but are useful for a larger purpose. Here is a famous example involving $\sqrt{2}$.

Lemma 2.20. *Suppose $n \in \mathbb{Z}$. Then n^2 is even $\iff$ n is even.*

Prove this yourself: the $\Rightarrow$ direction is easiest using the contrapositive, while the $\Leftarrow$ direction works well directly.

Theorem 2.21. *$\sqrt{2}$ is irrational.*

This is tricky for a few reasons. The theorem is seemingly not in the form $P \Rightarrow Q$. In fact it is: Q here is '$\sqrt{2}$ is irrational', while P is 'everything you already know in mathematics!' Now that we have P, it is clear that a direct proof will be hard to start (what exactly are we to use from the whole universe of mathematics?). Similarly the contrapositive might be tricky: $\neg Q$ straightforwardly states that $\sqrt{2}$ is rational, but again we are not sure what the P is that we are negating. Instead we use proof by contradiction.

Proof. Suppose that $\sqrt{2} = \frac{m}{n}$ for some $m, n \in \mathbb{N}$, where m, n have no common factors.[6]
Then $m^2 = 2n^2$ which says that m^2 is even.
By Lemma 2.20 we have that m is even.
Thus $m = 2k$ for some $k \in \mathbb{Z}$.
But now, $n^2 = 2k^2$, from which (Lemma 2.20 again) we see that n is even.
Thus m and n have a common factor of 2. This is a contradiction. ∎

Here is another famous result involving prime numbers.

Definition 2.22. A positive integer $p \geq 2$ is *prime* if its only positive divisors are itself and 1.

The first few primes are $2, 3, 5, 7, 11, 13, 17, 19, \ldots$. It follows from the definition that all positive integers ≥ 2 are either primes, or *composites* (products of primes).

Theorem 2.23. *There are infinitely many prime numbers.*

To break down the proof we first prove a lemma:

Lemma 2.24. *Suppose that $p_1, \ldots, p_n$ are integers ≥ 2. Then $\Pi := p_1 p_2 \cdots p_n + 1$ is not divisible by p_i for any i.*

Proof. Suppose that Π is divisible by p_i. Since $p_1 \cdots p_n$ is divisible by p_i, then $1 = \Pi - p_1 \cdots p_n$ is divisible by p_i. A contradiction. ∎

Proof of theorem. Again we prove by contradiction. Assume that there are exactly n prime numbers $p_1, \ldots, p_n$ and consider $\Pi := p_1 \cdots p_n + 1$. By the lemma, Π is not divisible by any of the primes $p_1, \ldots, p_n$. There are two cases:

(a) Π is prime, in which case it is a *larger* prime than any of $p_1, \ldots, p_n$.

[6] This is *no resistriction* upon m, n and not the source of the contradiction to come!

(b) Π is composite, in which case it is a divisible by a prime, which cannot be in our list $p_1, \ldots, p_n$.

In either case we've shown that there is another prime not in the list $p_1, \ldots, p_n$, and we've contradicted our assumption that we had all the primes.

■

Note how the lemma was used to make the ultimate proof of the theorem easier to read. The use of lemmas in this way is entirely a matter of personal preference.

2.3 Quantifiers

The proofs we've dealt with thusfar have been fairly straightforward. In higher mathematics, however, there are often definitions and theorems that involve many pieces, and it becomes unwieldy to write everything out in the style we've been doing. Two space-saving devices called *quantifiers* are often used to contract sentences and make the larger structure of a statement clearer.[7] Their use in formal logic is more complex, but for mathematics all you need is to be able to recognize, understand, and negate them. This last is most important for attempting contrapositive or contradiction proofs.

Definition 2.25. The *universal quantifier* $\forall$ is read 'for all'. The *existential quantifier* $\exists$ is read 'there exists'.

Many sentences in English can be restated using quantifiers:

Examples.

1. Every cloud has a silver lining: $\forall$ clouds $\exists$ a silver lining.
2. Humans have a brain: $\forall$ humans $\exists$ a brain.
3. There is an integer smaller than π: $\exists$ an integer n such that $n < \pi$.
4. (Harder) π cannot be written as a ratio of integers: $\forall$ integers m, n, we have $\frac{m}{n} \neq \pi$.

As we've observed, mathematics is something of an art form, and like with all art, different practitioners have different tastes. When writing mathematics some people write things very quickly, using only quantifiers, propositions and parentheses. Some write almost entirely in English. Most use a hybrid of quantifiers and English, aiming for a balance between brevity and clarity.

Example. The famous 'sum of four squares' theorem can be stated in three different ways

English	Every positive integer may be written as the sum of four squares
Quick	$(\forall n \in \mathbb{N})(\exists a, b, c, d \in \mathbb{Z})(n = a^2 + b^2 + c^2 + d^2)$
Hybrid	$\forall n \in \mathbb{N}$, $\exists a, b, c, d \in \mathbb{Z}$ such that $n = a^2 + b^2 + c^2 + d^2$

The insertion of a single comma and the phrase 'such that' in the Hybrid line makes a massive difference to the readability!

[7] At least that's the idea: very often they are *over-used* and achieve the opposite effect!

Negating Statements with Quantifiers

Besides the concision afforded by quantifiers, one of their benefits is a rule that allows for easy negation.

Suppose that $P(x)$ is a 'propositional functional': a proposition depending on a variable x. For example

$$P(x): \quad `x^2 > 4'$$

E.g. $P(5)$ is true, whilst $P(-1)$ is false. More generally, $P(x)$ is true for some values of x (namely $x > 2$ or $x < -2$) and false for others ($-2 \leq x \leq 2$).

Definition 2.26. A *counterexample* to $\forall x, P(x)$, is a single object t such that $P(t)$ is false.

Clearly $x = 1$ is a suitable counterexample to $\forall x, x^2 > 4$.

The following theorem tells you how to negate statements:

Theorem 2.27. *For any propositional function $P(x)$ with one variable, we have:*

- $\neg(\forall x, P(x))$ *is equivalent to* $\exists x, \neg P(x)$
- $\neg(\exists x, P(x))$ *is equivalent to* $\forall x, \neg P(x)$

Like with all theorems, to understand it you should unpack it, write it in English, and come up with an example:

> The negation of '$P(x)$ is true for all x' is 'There exists an x such that $P(x)$ is false.'
> The negation of 'There exists an x such that $P(x)$ is true' is '$P(x)$ is false for all x'.

Example. The negation of the statement 'Everyone owns a bicycle' is

'Someone does not own a bicycle.'

It certainly looks pedantic, but symbolically we might write

$$\neg\Big[(\forall \text{ people})(\text{person owns a bicycle})\Big] \quad \Longleftrightarrow \quad (\exists \text{ a person})(\text{person does not own a bicycle})$$

The best guide when negating statements is to *think* about what is meant, and not to follow the rules blindly. The rule of replacing one quantifier by the other when negating is helpful, but often dangerous: there are sometimes hidden quantifiers, or cases when things sound like quantifiers but aren't.

Example. 'All clouds have a silver lining' has negation

'There is (at least) one cloud without a silver lining'.

It is tempting to write '$\forall$ clouds $\exists$ silver lining'. This is perfectly correct if using $\exists$ as a short-hand for 'has a'. The problem is that 2.27 requires all quantifiers to be followed by a proposition, and 'silver lining' is *not* a proposition.
In this case 'has a silver lining' is really part of a propositional function

$$P(x): \quad \text{'cloud } x \text{ has a silver lining'}$$

The negation of $P(x)$ is then

$\neg P(x)$: 'cloud x does not have a silver lining'

The rogue $\exists$ should therefore not be turned into a $\forall$. Symbolically it is safer to start with

$\forall$ clouds x, x has a silver lining

which has negation

$\exists$ a cloud x without a silver lining

Notice how the introduction of a variable x makes things much easier!

Two pieces of advice when negating quantifiers

1. Think about the finished article and read it aloud: if it *sounds* like the opposite of what you started with then it probably is.
2. The symbol $\nexists$ for 'does not exist' is much abused. Very occasionally its use is appropriate, but it too often demonstrates laziness or a lack of understanding.

Multiple quantifiers

Once you're comfortable negating simple propositions and quantifiers, negating multiple quantifiers is easy. Just follow the rules, think, and take your time.

Example. Show that the following statement is false.

$$\forall x \in \mathbb{R}, \exists y \in \mathbb{R} \text{ such that } x \cdot y = 3$$

The negation of this expression follows the rules for replacing quantifiers with their opposites:

$$\exists x \in \mathbb{R}, \text{ such that } \forall y \in \mathbb{R} \text{ we have } x \cdot y \neq 3$$

Written this way it should be clear that the negation is true: if we take $x = 0$, then $x \cdot y = 0 \neq 3$, regardless of y, so $x = 0$ satisfies the negation. If the negation is true, then the original statement is false.

Continuity

For a harder example, combining multiple quantifiers and propositions, recall a part of calculus that everyone loves to hate: A function f is continuous at $x = a$ if

$$\forall \varepsilon > 0, \ \exists \delta > 0 \text{ such that } |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$$

The negation states what it means for f to be discontinuous at $x = a$:

$$\exists \varepsilon > 0 \text{ such that } \forall \delta > 0 \text{ we have } |x - a| < \delta \text{ and } |f(x) - f(a)| \geq \varepsilon.$$

The difficult part is negating the proposition $|x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$, but if you follow the above rules, and recall Theorem 2.8, you should be able to manage it.

We conclude this section with an important remark: When switching the order of two quantifiers, you should use caution. Consider for example the following two propositions:

1. For every person x, there exists a person y such that y is a friend of x.
2. There exists a person y such that, for every person x, y is a friend of x.

Assuming we are dealing with people, we can rewrite the sentences as

1. $\forall\, x$, $\exists y$ such that y is a friend of x, and
2. $\exists y$ such that, $\forall x$, y is a friend of x.

All we have done is switching the order of the two quantifiers! How does this affect the meaning? (1) translates into 'Everyone has a friend', while (2) states that 'There exists somebody who is friend with everybody'. Quite different!

Play around with the examples below. What is the meaning? Which one is true?

- $\forall$ days x, $\exists$ a person y such that y was born on day x
- $\exists$ a person y such that, $\forall$ days x, y was born on day x.
- $\forall$ circles x , $\exists$ a point y such that y is the center of x
- $\exists$ a point y such that, $\forall$ circles x, y is the center of x.
- $\forall\, x \in \mathbb{Z}$, $\exists y \in \mathbb{Z}$ such that $y < x$.
- $\exists y \in \mathbb{Z}$ such that, $\forall x \in \mathbb{N}$, $y < x$.

3 Divisibility and the Euclidean Algorithm

In this section we introduce the notion of *congruence*: a generalisation of the idea of separating all integers into 'even' and 'odd'. At its most basic it involves going back to elementary school when you first learned division and would write something similar to

$$33 \div 5 = 6\,\mathrm{r}\,3 \qquad \text{'6 remainder 3.'}$$

The study of congruence is of fundamental imporance to Number Theory, and provides some of the most straightforward examples of Groups and Rings. We will cover the basics in this section—enough to compute with—then return later for more formal observations.

3.1 Remainders and Congruence

Definition 3.1. Let m, n be integers. We say that n *divides* m and write $n \mid m$ if m is divisible by n: that is if there exists some integer k such that $m = kn$. Equivalently, we say that n is a *divisor* of m.

For example: $4 \mid 20$ and $17 \mid 51$, but $12 \nmid 8$.

When one integer does not divide another, there is a remainder left over.

Theorem 3.2 (The Division Algorithm)**.** *Let m be an integer and n a positive integer. Then there exist unique integers q (the* quotient*) and r (the* remainder*) which satisfy*

- $0 \leq r < n$
- $m = qn + r$.

For example: If $m = 23$ and $n = 7$, then $q = 3$ and $r = 2$ because '$23 \div 7 = 3$ remainder 2'. More formally, $23 = 3 \cdot 7 + 2$, with $0 \leq 2 < 7$. Similarly, if $m = -11$ and $n = 3$, then $q = -4$ and $r = 1$ because $-11 = (-4) \cdot 3 + 1$, with $0 \leq 1 < 3$. For practice, find a formula for all the integers that have remainder 4 after division by 6.

The proof of the Division Algorithm relies on the development of induction, to which we will return later. The theorem should be read as saying that n goes into m q times with r left over. The fact that the remainder is nicely defined allows us to construct an alternative form of arithmetic.

Definition 3.3. Let a, b be integers, and n a positive integer. We say that *a is congruent to b modulo n* and write $a \equiv b \pmod{n}$ if a and b have the same remainder upon dividing by n.
When the *modulus* n is clear, it tends to be dropped, and we just write $a \equiv b$.

For example: $7 \equiv 10 \pmod{3}$, since both have the same remainder (1) on dividing by 3. Can you find a formula for *all* the integers that are congruent to 10 modulo 3?

Let n be an integer. Consider the following two conjectures.

Conjecture 3.4. $n \equiv 8 \pmod{6} \Rightarrow n \equiv 2 \pmod{3}$

Conjecture 3.5. $n \equiv 2 \pmod{3} \Rightarrow n \equiv 8 \pmod{6}$

True or false?

The first conjecture is true. Indeed, if $n \equiv 8 \pmod{6}$, we can write $n = 6k + 8$ for some integer k. Then $n = 6k + 8 = 6k + 6 + 2 = 3(2k + 2) + 2$ so n has remainder 2 upon division by 3, showing that n is congruent to 2 modulo 3.
On the other hand, the second conjecture is false. Coming up with a counterexample is easy: Consider $n = 5$. Then clearly n is congruent to 2 modulo 3, but n is not congruent to 8 modulo 6 (because it has remainder 5, not 2, upon division by 6).

The following theorem is crucial, and provides an equivalent definition of congruence.

Theorem 3.6. $a \equiv b \pmod{n} \iff n \mid (a - b)$.

Proof. Let $a = q_1 n + r_1$ and $b = q_2 n + r_2$ be the expressions for a, b as given in the division algorithm. Then

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

Clearly $a - b$ is divisible by n if and only if $r_1 - r_2$ is also.
Now recall that remainders satisfy $0 \leq r < n$, whence $-n < r_1 - r_2 < n$.
Thus $r_1 - r_2$ is divisible by n if and only if it equals zero. Putting this together we have that $a - b$ is divisible by n iff $r_1 = r_2$, and a, b have the same remainder. ∎

Congruence and Divisibility
Let a be any integer and let n be a positive integer. Then

- a is divisible by n if and only if $a \equiv 0 \pmod{n}$.
- a is *not* divisible by n if and only if $a \equiv 1$ or 2 or $\ldots$ or $n-1 \pmod{n}$.

To gain familiarity with the definition of congruence, and review some proof techniques, prove the following theorem.

Theorem 3.7. *Suppose that n is an integer. Then*

$$n^2 \not\equiv n \pmod{3} \Leftrightarrow (n \not\equiv 0 \pmod{3}) \text{ and } (n \not\equiv 1 \pmod{3}).$$

That the congruence sign $\equiv$ appears similar to the equals sign $=$ is no accident. In many ways it behaves exactly the same.[8]

Here we spell out the basic rules of congruence arithmetic.[9]

Theorem 3.8. *Suppose throughout that a, b, c, d are integers. All congruences are modulo the same integer n.*

- $a \equiv b$ *and* $c \equiv d \implies ac \equiv bd$
- $a \equiv b$ *and* $c \equiv d \implies a \pm c \equiv b \pm d$

Proof. Suppose that $a \equiv b$ and $c \equiv d$. By Theorem 3.6 we have $a - b = kn$ and $c - d = ln$ for some integers k, l. Thus

$$ac = (b + kn)(d + ln) = bd + n(bl + kd + kln) \Rightarrow ac - bd = n(bl + kd + kln)$$

is divisible by n. Hence $ac \equiv bd$.
Try the second argument yourself. ∎

If we are doing nothing beyond addition, subtraction and multiplication of integers, these rules say that we can simply replace a number with its remainder for the purposes of modular arithmetic. This allows us to perform some surprising calculations. Note that integer powers are simply multiplication: you may replace the base with its remainder modulo n.

Examples. 1. Here we compute modulo $n = 6$.

$$7^9 + 14^3 \equiv 1^9 + 2^3 \equiv 1 + 8 \equiv 9 \equiv 3 \pmod{6}.$$

Otherwise said, $7^9 + 14^3 = 40356351$ has remainder 3 when divided by 6.

2. Find the remainder when $124^{12} \cdot 65^{23}$ is divided by 11. Since $124 = 11^2 + 3$ and $65 = 11 \cdot 6 - 1$, we write

$$\begin{aligned} 124^{12} \cdot 65^{23} &\equiv 3^{12} \cdot (-1)^{23} \equiv 27^4 \cdot (-1) \equiv -(5^4) && (27 \equiv 5 \pmod{11}) \\ &\equiv -(25^2) \equiv -(3^2) \equiv 2 \pmod{11} \end{aligned}$$

The remainder is therefore 2. There is no way to do this on a pocket calculator, since the original number $124^{12} \cdot 65^{23} \approx 3.7 \times 10^{88}$ is far too large to work with!

[8] It is an 'equivalence relation' (later).

[9] The usual associative, commutative and distributive laws of arithmetic

$$a + (b + c) \equiv (a + b) + c, \qquad a(bc) \equiv (ab)c, \qquad a + b \equiv b + a, \qquad ab \equiv ba, \qquad a(b + c) \equiv ab + ac$$

all follow because $x = y \Rightarrow x \equiv y \pmod{n}$, regardless of n: equal numbers have the same remainder after all!

The primary difference between the arithmetic of congruence and 'normal' arithmetic is, perhaps unsurprisingly, with regards to *division.*

Theorem 3.9. *If* $ka \equiv kb \pmod{kn}$ *then* $a \equiv b \pmod{n}$.

The modulus divides by k as well as the terms, so the meaning of $\equiv$ changes. You should be able to prove this using the definition and Theorem 3.6.

Play with the congruence notation until you are familiar with the method of calculation.

3.2 Greatest Common Divisors and The Euclidean Algorithm

Number theory (at its most basic) is about finding *integer* solutions to equations. Here is a simple-sounding pair of questions:

- Are there any *integer points*[10] on the straight line $9x - 21y = 6$?
- What about on the line $4x + 6y = 1$?

Before you do anything else, try sketching both lines (lined graph paper will help) and try to decide if there are any integer points. If there are any, how many are there? Can you find an equation for them all?

In this section we show how to answer these questions in general: when are there integer points on the line $ax + by = c$ with $a, b, c \in \mathbb{Z}$? It requires little more than the division algorithm.

Definition 3.10. Let m, n be integers. Their *greatest common divisor* $\gcd(m, n)$ is exactly what it sounds like: the largest divisor of both m and n. We say that m, n are *relatively prime* if $\gcd(m, n) = 1$.

Example. Let $m = 60$ and $n = 90$. The positive divisors of each are listed in the table:

m	1	2	3	4	5	6	10	12	15	20	<u>30</u>	60
n	1	2	3	5	6	9	10	15	18	<u>30</u>	45	90

The greatest common divisor (largest number common to both rows) is clearly $\gcd(60, 90) = 30$.

Finding the greatest common divisor by listing all the positive divisors of a number is extremely tedious. Thankfully there is a famous procedure, the Euclidean Algorithm,[11] which makes this much faster.

Algorithm. To find $\gcd(m, n)$ for two positive integers $m > n$:

1. Use the division algorithm to write $m = q_1 n + r_1$ with $0 \leq r_1 < n$.
2. If $r_1 > 0$, apply again: $n = q_2 r_1 + r_2$ where $0 \leq r_2 < r_1$.
3. Repeat process: obtain a decreasing sequence of positive integers

   $$r_1 > r_2 > r_3 > \ldots > 0$$

   Eventually the process must stop with remainder zero: $\exists r_{p+1} = 0$.

[10] Points (x, y) such that both x, y are integers.
[11] Dating at least as far back as Euclid's *Elements* c. 300 BC.

If m, n are not both positive, simply make them positive first by taking absolute values.

Theorem 3.11. *After applying the Euclidean Algorithm to $m > n$ we obtain $\gcd(m, n) = r_p$.*

For a proof see any introductory number theory book, take Math 180, or see the homework :)

Our interest in the Euclidean Algorithm comes from the fact that by turning it on its head we may find integers λ, μ such that $\gcd(m, n) = \lambda m + \mu n$. Once you've written out the algorithm, simply start at the bottom and substitute as you work your way up. This is easiest to demonstrate by example:

Example. Find $\gcd(1260, 750)$ and integers λ, μ such that $d = 1260\lambda + 750\mu$.
Apply the Euclidean algorithm:

$$\begin{aligned}
1260 &= 1 \times 750 + 510 && (m = q_1 n + r_1)\\
750 &= 1 \times 510 + 240 && (n = q_2 r_1 + r_2)\\
510 &= 2 \times 240 + 30 && (r_1 = q_3 r_2 + r_3)\\
240 &= 8 \times 30 + 0 && (r_2 = q_4 r_3\text{: algorithm terminates})
\end{aligned}$$

According to the Theorem, $\gcd(1260, 750) = 30$. Now we reverse the algorithm, starting with the third line:

$$\begin{aligned}
30 &= 510 - 2 \times 240\\
&= 510 - 2(750 - 510) = 3 \times 510 - 2 \times 750 && \text{(substitute using 2nd line)}\\
&= 3(1260 - 750) - 2 \times 750 = 3 \times 1260 - 5 \times 750 && \text{(substitute using 1st line)}
\end{aligned}$$

Thus we can choose $\lambda = 3$ and $\mu = -5$.

Indeed it is not hard to see that the following lemma follows directly from the algorithm:

Lemma 3.12. *Given any integers m, n there exist integers λ, μ such that $\gcd(m, n) = \lambda m + \mu n$.*

As an application of the above discussion, we solve our motivating problem: finding all integer points on the line $ax + by = c$ where a, b, c are integers.

Theorem 3.13. *Let a, b, c be integers and $d = \gcd(a, b)$. Then the equation $ax + by = c$ has integer solutions (x, y) iff $d \mid c$. In such a case, all integer solutions are given by*

$$x = x_0 + \frac{b}{d}t \qquad y = y_0 - \frac{a}{d}t$$

where (x_0, y_0) is any fixed integer solution, and t takes any integer value.

One uses the Euclidean Algorithm to find the initial solution (x_0, y_0) then applies the formula to obatin all of them.[12]

[12]The astute observer should recognize the similarity between this and the complementary function/particular integral method for linear differential equations: (x_0, y_0) is a 'particular solution' to the full equation $ax + by = c$, while $t(\frac{b}{d}, -\frac{a}{d})$ comprises all solutions to the 'homogeneous equation' $ax + by = 0$.

Examples. 1. Given the line $1260x + 750y = 90$, we calculated above that $\gcd(1260, 750) = 30$. Since $30 \mid 90$, we know that there are integer solutions. We also calculated that

$$d = 30 = 3 \times 1260 - 5 \times 750 \implies 90 = 9 \times 1260 - 15 \times 750,$$

whence $(x_0, y_0) = (9, -15)$ is a particular solution. The general solution is therefore

$$(x, y) = \left(9 + \frac{750}{30}t, -15 - \frac{1260}{30}t\right) = (9 + 25t, -15 - 42t), \text{ where } t \in \mathbb{Z}.$$

2. Consider the line $570x + 123y = 7$. We calculate the greatest common divisor:

$$\begin{aligned}
570 &= 4 \times 123 + 78 \\
123 &= 1 \times 78 + 45 \\
78 &= 1 \times 45 + 33 \\
45 &= 1 \times 33 + 12 \\
33 &= 2 \times 12 + 9 \\
12 &= 1 \times 9 + 3 \\
9 &= 4 \times 3 + 0
\end{aligned}$$

Thus $\gcd(570, 123) = 3$. Since $3 \nmid 7$, we conclude that the line $570x + 123y = 7$ has no integer points.

3. Repeat the above calculations for our motivating problems: what does the theorem say?

4 Sets and Functions

4.1 Set Notation and Describing a Set

We start with a very naïve definition: a set is a collection of objects.[13]

Definition 4.1. If x is an object in a set A, we write $x \in A$ and say that x is an *element* or *member* of A. Conversely, if x is a member of some other set B, but not of A, we write $x \notin A$.

Notation and Conventions

Use capital letters for sets, e.g. A, B, C, S, and lower case letters for elements, e.g. $a \in A$.
Curly brackets $\{\ ,\ \}$ are used to bookend the elements of a set: e.g. if we wrote

$$S = \{3, 5, f, \alpha, \beta\}$$

then we'd say, "S is the set whose elements are 3, 5, f, α and β."
The order of elements in a set is irrelevant: e.g. $S = \{\beta, f, 5, \alpha, 3\} = \{f, \alpha, 3, \beta, 5\}$.

[13] To get to the point where mathematicians realized that this is indeed naïve required much thinking. In particular some collections of objects cannot be considered to sets in order for the correct definition to make sense. For the present, our notion is enough.

To denote the elements of a set which satisfy some property we use the symbols "|" or ":" and read, "such that.' Which you use depends partly on taste, although the context may make one clearer to read. For example, if $S = \{3, 5, f, \alpha, \beta\}$ is the set defined above, we could write:

$$\{s \in S : s \text{ is a Greek letter}\} = \{\alpha, \beta\}$$

or

$$\{s \in S \mid s \text{ is a Greek letter}\} = \{\alpha, \beta\}.$$

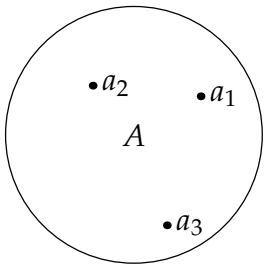
We would read: "The set of elements s in the set S such that s is a Greek letter" is $\{\alpha, \beta\}$.

Example. Let $A = \{2, 4, 6\}$ and $B = \{1, 2, 5, 6\}$. Then[14]

1. $\{a \in A : a \text{ is divisible by } 4\} = \{4\}$
2. $\{b \in B : b \text{ is odd}\} = \{1, 5\}$
3. $\{a \in A : a \in B\} = \{2, 6\}$
4. $\{a \in A : a \notin B\} = \{4\}$
5. $\{b \in B : b \text{ is odd and } b - 1 \in A\} = \{5\}$

Take your time getting on top of this notation. It is *crucial* that you can translate from set notation to English and back again, or you will be incapable of understand most higher-level mathematics.

You may find *Venn diagrams* useful when thinking abstractly about sets. A set is visualized as a region in the plane and, if necessary, members of the set can be thought of as dots in this region. This is most useful when one has to think about multiple, possibly overlapping, sets. The graphic here represents a set A with at least three elements a_1, a_2, a_3.



Sets of Numbers

Common sets of numbers are denoted using the BLACKBOARD BOLD typeface.

$$\begin{aligned}
&\mathbb{N} = \mathbb{Z}^+ = \text{natural numbers} = \{1, 2, 3, 4, \ldots\}\\
&\mathbb{N}_0 = \mathbb{W} = \mathbb{Z}_0^+ = \text{whole numbers} = \{0, 1, 2, 3, 4, \ldots\}\\
&\mathbb{Z} = \text{integers} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}\\
&\mathbb{Q} = \text{rational numbers} = \{\tfrac{m}{n} \colon m \in \mathbb{Z} \text{ and } n \in \mathbb{N}\}\\
&\mathbb{R} = \text{real numbers}\\
&\mathbb{R} \setminus \mathbb{Q} = \text{irrational numbers}\\
&\mathbb{C} = \text{complex numbers} = \{x + iy : x, y \in \mathbb{R}, \text{ where } i = \sqrt{-1}\}\\
&\mathbb{Z}_n = \text{Remainders } \pmod{n} = \{0, 1, 2, \ldots, n-1\}.
\end{aligned}$$

[14] It is conventional, though not required, to denote an abstract element of a set by the corresponding lower case letter: thus $a \in A$.

Where there are multiple choices of notation, we will tend to use the first in the list. The use of a subscript 0 to include zero and superscript $\pm$ to restrict to positive or negative numbers is standard.

For example, we could write

$$7 \in \mathbb{Z}, \qquad \pi \in \mathbb{R}, \qquad \pi \notin \mathbb{Q}, \qquad \sqrt{-5} \in \mathbb{C}.$$

There may be multiple ways to denote the same set: for example the set of even numbers (often denoted $2\mathbb{Z}$) may be written variously as

$$2\mathbb{Z} = \{2n : n \in \mathbb{Z}\} = \{n \in \mathbb{Z} : \exists k \in \mathbb{Z}, m = 2k\} = \{n \in \mathbb{Z} : n \equiv 0 \pmod 2\} = \{n \in \mathbb{Z} : 2 \mid n\}$$

where we used both congruence and divisor notation to obtain suitable descriptions.

Choice of notation

The two choices of notation for "such that" ($\mid$ vs. :) are to give you leeway in case of potential confusion. For example, the final expression (above) for the set of even numbers could have been written

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

Plainly this is foolish. Other concepts make use of colons.[15] You may use whichever notation you prefer, within the bounds of trying to be unambiguous.

Examples. 1. List the elements of the set $A = \{x \in \mathbb{R} : x^2 + 3x + 2 = 0\}$.

In English, we are looking for the set of all real number solutions to the quadratic equation $x^2 + 3x + 2 = 0$. Some simple factorizing tells us that $x^2 + 3x + 2 = (x + 1)(x + 2)$, whence $A = \{-1, -2\}$.

2. List the elements of $B = \{x \in \mathbb{N} : x \equiv 2 \text{ or } 5 \pmod{20}\}$ in a more naïve way.

We are looking for all positive integers x for which there exists some integer n such that

$$x = 2 + 20n \quad \text{or} \quad 5 + 20n.$$

In an elementary class, B might have been written

$$B = \{2, 5, 22, 25, 42, 45, 62, 65, 82, 85, \ldots\}$$

3. List the elements of the set $C = \{n \in \mathbb{Z} : n^2 - 3 \in \mathbb{Z}_{25}\}$.

Since $\mathbb{Z}_{25} = \{0, 1, 2, \ldots, 24\}$ is the set of remainders modulo 25, we see that

$$n^2 - 3 \in \mathbb{Z}_{25} \iff n^2 \in \{3, 4, 5, \ldots, 25, 26, 27\}$$

Since n must be an integer, it follows that

$$C = \{\pm 2, \pm 3, \pm 4, \pm 5\}$$

[15] E.g. functions, as we shall see shortly. For example the set of even functions with domain the real numbers might be written

$$\{f : \mathbb{R} \to \mathbb{R} : \forall x, f(x) = f(-x)\} \text{ or } \{f : \mathbb{R} \to \mathbb{R} \mid \forall x, f(x) = f(-x)\}$$

The former is certainly more confusing than the latter.

Intervals

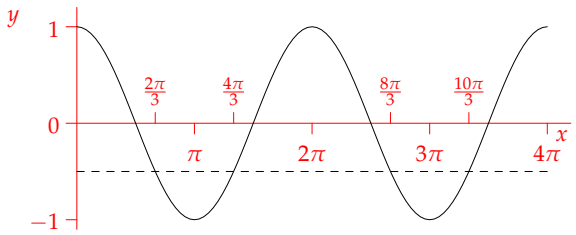
When discussing collections of *real numbers*, the interval notation is useful. For example,

$$(0,1) = \{x \in \mathbb{R} : 0 < x < 1\}$$
$$[0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$$
$$(0,1] = \{x \in \mathbb{R} : 0 < x \leq 1\}$$

When writing intervals with $\pm\infty$ use an open bracket at the infinite end(s): $[1,\infty) = \{x \in \mathbb{R} : x \geq 1\}$.

Example. Recall some basic trigonometry: the solutions of the equation $\cos x = -\frac{1}{2}$ on the interval $[0, 4\pi]$ can be written

$$\left\{x \in [0,4\pi] : \cos x = -\frac{1}{2}\right\} = \left\{\frac{2\pi}{3}, \frac{4\pi}{3}, \frac{8\pi}{3}, \frac{10\pi}{3}\right\}$$



Cardinality and the Empty Set

Definition 4.2. A set A is *finite* if it contains a finite number of elements: this number is the set's *cardinality*, written $|A|$. A is *infinite* otherwise.

Examples. 1. Let $A = \{a, b, \alpha, \gamma, \sqrt{2}\}$, then $|A| = 5$.

2. Let $B = \Big\{\{1,2\}, \{3\}\Big\}$. It is important to note that the *elements/members* of B are $\{1,2\}$ and $\{4\}$, both of which are themselves sets. Therefore $|B| = 2$. A set $\{1,2\}$ is an object in its own right, and can therefore be placed in a set with other objects.[16]

To round things off we need a symbol to denote a set that contains nothing at all!

Axiom. There exists a set $\emptyset$ with no elements (cardinality zero: $|\emptyset| = 0$). We call $\emptyset$ the *empty set*.

Axioms

An axiom is a basic assumption; something that we need to do mathematics, but cannot prove. This is the cheat by which mathematicians can be 100% sure that something is true: it is proved based on the assumption of several axioms. Particularly with reference to this axiom, it seems bizarre that we can assume the existence of some set that has nothing in it. Regardless, mathematicians have universally agreed that we need the empty set in order to do the rest of mathematics.

Note that $|A| \in \mathbb{N}$ for any finite non-empty set A. Later we shall see that there is a notion of cardinality for infinite sets, but that discussion will require a lot a preparation…

[16] The confusion here is similar to that caused by the following sentences: 'UCI *are* constructing a laboratory' and 'UCI *is* constructing a laboratory'. In the first case we are thinking of UCI as a collection of individuals, in the latter case UCI is a single object. Opinions differ in various modes of English as to which is correct…

4.2 Subsets

Now we wish to compare sets.

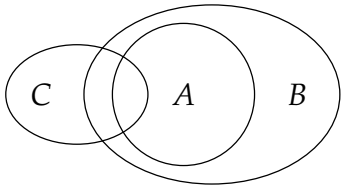
Definition 4.3. If A and B are sets such that every element of A is also an element of B, then we say that A is a *subset* of B and write $A \subseteq B$.
Sets A, B are *equal*, written $A = B$, if they have exactly the same elements. Equivalently

$$A = B \iff A \subseteq B \text{ and } B \subseteq A.$$

A is a *proper subset* of B if it is a subset which is not equal.[17]

The formulaic characterization of equality is *very* important. It is often how you *prove* that two sets are equal.

Venn diagrams are particularly useful for depicting subset relations. The graphic on the right depicts three sets A, B, C: it should be clear that the only valid subset relation between the three is $A \subseteq B$.



Restriction notation gives easy examples of subsets: our earlier examples are all subsets of the set of real numbers $\mathbb{R}$. Here are some others.

Examples. 1. $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$. This is clearly a subset of $\mathbb{Z}$.

2. $\{x \in \mathbb{R} : x^2 - 1 = 0\} \subseteq \{y \in \mathbb{R} : y^2 \in \mathbb{N}\}$
   Think for a moment what these two sets consist of: you should be tempted to write

   $$\{-1, 1\} \subseteq \{\pm\sqrt{1}, \pm\sqrt{2}, \pm\sqrt{3}, \pm\sqrt{4}, \ldots\}$$

   Ellipses ($\ldots$) are very useful for getting a feel for a set and listing elements in an intuitive way. You should, however, resist their use in formal mathematics due to their lack of clarity.

3. $\mathbb{Z}_n \subseteq \mathbb{Z}_m \iff n \leq m$.

Here we collect several results relating to subsets.

Theorem 4.4. *1. If $|A| = 0$, then $A = \emptyset$* *(Uniqueness of the empty set)*

2. For any set A, we have $\emptyset \subseteq A$ and $A \subseteq A$ *(Trivial and non-proper subsets)*

3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$ *(Transititvity of subsets)*

Proof. 1. $\emptyset$ has no members, therefore $\emptyset \subseteq A$ is trivial: there is nothing to check to see that 'all elements of $\emptyset$ are also elements of A.' $A \subseteq \emptyset$ is the identical argument.

2. $\emptyset \subseteq A$ is trivial by the argument in 1. For the second part, we must show that all elements of A are elements of A. But this is just a restatement of the same thing! If you want a more algebraic-looking argument, try

   Let $a \in A$. Then $a \in A$, so $A \subseteq A$.

[17] Some people write $A \subset B$ for proper subset. Others use $\subset$ for 'subset', whether proper or not, and write $A \subsetneq B$ to stress a proper subset. Notation is not as strict as with $\leq$ and $<$. To avoid confusion, in these notes we will always use $\subseteq$, and write 'proper subset' in English.

3. We must show that all elements of A are also elements of C. Let $a \in A$. Since $A \subseteq B$ we know that $a \in B$. Since $B \subseteq C$ we conclude that $a \in C$. Hence $A \subseteq C$.

■

As a final observation, indeed one to which we will return later, your intutition should tell you that for finite sets A and B we have

$$A \subseteq B \implies |A| \leq |B|.$$

Indeed if one simply replaces $\subseteq$ with $\leq$, $\emptyset$ with 0, and A, B, C by arbitrary non-negative integers, then the above theorem should seem completely natural. Recognizing the similarities between a new concept and a familiar one—essentially spotting patterns—is perhaps the most necessary skill in mathematics.

4.3 Unions and Intersections

For the duration of this section, suppose that $\mathcal{U}$ is some 'universal set', of which every set mentioned subsequently is a subset.[18] Given sets $A, B \subseteq \mathcal{U}$, we can form several new sets directly.

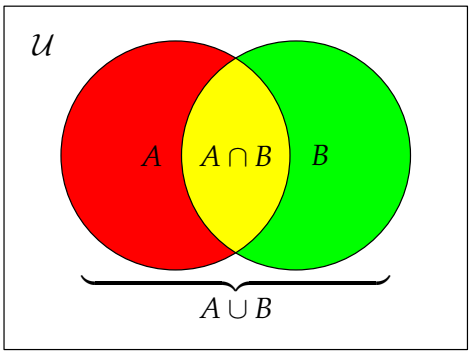
Definition 4.5. The *union* of two sets A, B is defined by

$$A \cup B = \{x \in \mathcal{U} : x \in A \text{ or } x \in B\}.$$

The *intersection* of A, B is defined by

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

The Venn diagram of both is shown where A, B are depicted as overlapping circles.



We say that A, B are *disjoint* if $A \cap B = \emptyset$.

'Or' is used in the logical sense: $A \cup B$ is the collection of all elements in A, in B, or in both. Now observe the pattern with the notation: $\cup$ and $\cap$ look very similar to $\vee$ and $\wedge$.

Examples. 1. Let $\mathcal{U} = \{\text{fish, dog, cat, hamster}\}$, $A = \{\text{fish, cat}\}$, and $B = \{\text{dog, cat}\}$. Then,

$$A \cup B = \{\text{fish, dog, cat}\}, \qquad A \cap B = \{\text{cat}\}.$$

2. Let $A = (-\infty, 3)$ and $B = [-2, \infty)$ in interval notation. Then $A \cup B = \mathbb{R}$ and $A \cap B = [-2, 3)$. In this example, it seems reasonable to assume that $\mathcal{U} = \mathbb{R}$: the universal set is rarely explicit.

Theorem 4.6. *Let A, B, C be sets. Then:*

1. $\emptyset \cup A = A$ *and* $\emptyset \cap A = \emptyset$
2. $A \cap B \subseteq A \subseteq A \cup B$
3. $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$

[18] Partly this makes formulæ make sense, partly it is necessary in axiomatic set theory.

4. $A \cup (B \cup C) = (A \cup B) \cup C$ *and* $A \cap (B \cap C) = (A \cap B) \cap C$

5. $A \cup A = A \cap A = A$

6. $A \subseteq B \Rightarrow A \cup C \subseteq B \cup C$ *and* $A \cap C \subseteq B \cap C$

We will only prove one of these: some others may appear in the homework and you should try them yourself. Observe how we use the definitions of union, intersection and subset.

Proof of 2. There are two results here. We show each separately, along with some thinking.
Suppose that $x \in A \cap B$. (Must show $x \in A \cap B \Rightarrow x \in A$)
Then $x \in A$ and $x \in B$. (Definition of intersection)
But then $x \in A$, whence $A \cap B \subseteq A$ (Definition of subset)
Now let $y \in A$. (Must show $y \in A \Rightarrow y \in A \cup B$)
Then '$y \in A$ or $y \in B$' is true, from which we conclude that $y \in A \cup B$.
Thus $A \subseteq A \cup B$.
■

Once you get comfortable, you can strip away all the comments and write the proof more quickly.

Complements

If union and intersection are similar to 'or/and' in logic, then the analogue of negation is the following.

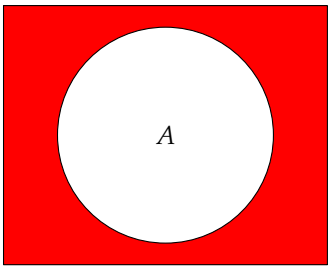
Definition 4.7. Let $A \subseteq \mathcal{U}$ be a set. The *complement* of A is the set

$$A^{\mathsf{C}} = \{x \in \mathcal{U} : x \notin A\}$$

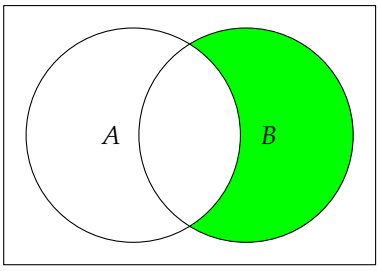
This can also be written $U \setminus A$, $U - A$, A'.
If $B \subseteq \mathcal{U}$ is another set, then the *complement of A relative B* is

$$B \setminus A = \{x \in B : x \notin A\}$$

The Venn diagrams for the two complements of A are colored below.



A^{C}: everything not in A



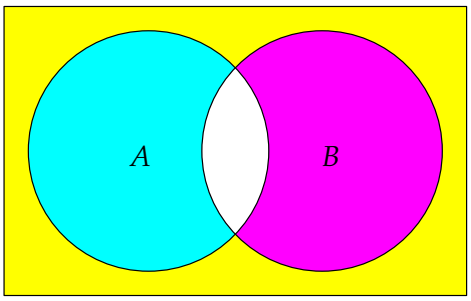
$B \setminus A$: everything not in B but not in A

In many cases the two notations mean the same thing: if A is a subset of B, then one might as well assume that $B = \mathcal{U}$ and the pictures are identical. The difference between the above definitions is that for $B \setminus A$, A is not required to be a subset of B. For example, if $U = \{1,2,3,4,5\}$, $A = \{1,2,3\}$, $B = \{2,3,4\}$, then

$$A^{\mathsf{C}} = U \setminus A = \{4,5\}, \quad B^{\mathsf{C}} = U \setminus B = \{1,5\}, \quad B \setminus A = \{4\}, \quad A \setminus B = \{1\}.$$

Theorem 4.8. *Let A, B be sets. Then:*

1. $(A \cup B)^{\mathsf{C}} = A^{\mathsf{C}} \cap B^{\mathsf{C}}$
2. $(A \cap B)^{\mathsf{C}} = A^{\mathsf{C}} \cup B^{\mathsf{C}}$
3. $(A^{\mathsf{C}})^{\mathsf{C}} = A$
4. $A \setminus B = A \cap B^{\mathsf{C}}$
5. $A \subseteq B \iff B^{\mathsf{C}} \subseteq A^{\mathsf{C}}$



Result 2: Add the colored regions

It is not important to memorize these laws. You should be able to *visualize* them by drawing Venn diagrams. A Venn diagram does not constitute a formal proof, though it is extremely helpful for clarification. For example:

Proof of 1. Let $x \in (A \cup B)^{\mathsf{C}}$. Then $x \notin A \cup B$: otherwise said, x is not in 'A or B'. Then $x \notin A$ *and* $x \notin B$, whence $x \in A^{\mathsf{C}} \cap B^{\mathsf{C}}$. ∎

Parts (1) and (2) of the theorem are known as *de Morgan's laws*, just as the equivalent statements in logic: Theorem 2.7. Indeed, we could rephrase our proof in that language.

Alternative Proof of 1. Let P, Q be the statements

$$P: \text{“}x \in A\text{”} \qquad Q: \text{“}x \in B\text{”}$$

Then $P \vee Q$ is "$x \in A$ or $x \in B$," equivalently "$x \in A \cup B$." By de Morgan's laws, its negation is

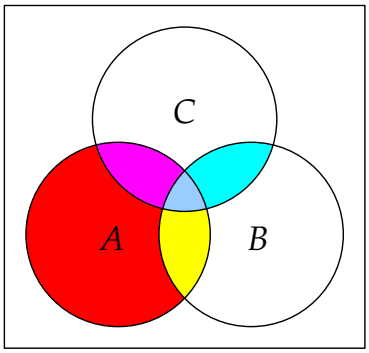
$$\neg P \wedge \neg Q: \text{“}x \notin A \text{ and } x \notin B\text{”}$$

Otherwise said: $x \in A^{\mathsf{C}} \cap B^{\mathsf{C}}$. ∎

Theorem 4.9 (Distributive laws)**.** *For any sets A, B, C:*

1. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

We will prove only the second. Try the first yourself. The method is the standard approach in such cases: show that each side is a subset of the other.



2: Add the colored regions

Proof. ($\subseteq$) Let $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in B \cap C$. There are two cases:

- If $x \in A$, then $x \in A \cup B$ and $x \in A \cup C$ by Theorem 4.6, part 2.
- If $x \in B \cap C$, then $x \in B$ and $x \in C$. It follows that $x \in A \cup B$ and $x \in A \cup C$, again by Theorem 4.6.

In both cases $x \in (A \cup B) \cap (A \cup C)$.

($\supseteq$) Let $y \in (A \cup B) \cap (A \cup C)$. Then $y \in A \cup B$ and $y \in A \cup C$. There are again two cases:

- If $y \in A$, then we are done, for $y \in A \cup (B \cap C)$ by Theorem 4.6.
- If $y \notin A$, then $y \in B$ and $y \in C$. Hence $y \in B \cap C$. In particular $y \in A \cup (B \cap C)$.

In both cases $y \in A \cup (B \cap C)$.

■

4.4 Introduction to Functions

You will have been used to using functions for a long time. A formal definition in terms of relations will be given later. We will just use the following:

Definition 4.10. Let A, B be sets. A *function from A to B* is a rule f that assigns an element of B to each element of A.
The *domain* of f, written $\mathrm{dom}(f)$, is the set A.
The *range* of f is the subset of B consisting of all the elements assigned by f.

Notation

If f is a function from A to B we write $f : A \to B$.
If $a \in A$, we write $b = f(a)$ for the the element of B assigned to a by f. We can also write $f : a \mapsto b$, read "f maps a to b."
If $U \subseteq A$ is a subset then the *image* of U is the following subset of B

$$f(U) = \{f(u) \in B : u \in U\}$$

The image of A is precisely the range of f,

$$\mathrm{range}(f) = \{f(a) \in B : a \in A\}$$

Examples. 1. Let $f : [-3, 2) \to \mathbb{R}$ be defined by $f : x \mapsto x^2$. Then $\mathrm{dom}(f) = [-3, 2)$, and $\mathrm{range}(f) = [0, 9]$. Indeed we could also calculate

$$f((-1, 2]) = [0, 4]$$

2. Define $f : \mathbb{Z} \to \mathbb{Z}_3$ by $f : n \mapsto n^2 \pmod 3$, where we take the remainder of n^2 modulo 3. Clearly $\mathrm{dom}(f) = \mathbb{Z}$, but what is the range? Trying a few examples, we see the following:

n	0	1	2	3	4	5	6	7	8	9	10
$f(n)$	0	1	1	0	1	1	0	1	1	0	1

It looks like the range is simply $\{0, 1\}$. To prove it, recall modular arithmetic as in Section 3.1. Since we are squaring n, which only involves multiplication,[19] we need only consider three cases:

If $n \equiv 0$, then $n^2 \equiv 0 \pmod 3$

[19] An operation which behaves normally in modular arithmetic.

If $n \equiv 1$, then $n^2 \equiv 1 \pmod 3$

If $n \equiv 2$, then $n^2 \equiv 4 \equiv 1 \pmod 3$

Thus $n^2 \equiv 0, 1 \pmod 3$, and $\text{range}(f) = \{0, 1\}$. Compare this to the proof of Theorem 2.16 to observe the benefit of modular calculations.

3. Let $f : \mathbb{Z}_{10} \to \mathbb{Z}_{10}$ be defined by $f : n \mapsto 3n \pmod{10}$. To help understand this function, list the elements: the domain only has 10 elements after all.

n	0	1	2	3	4	5	6	7	8	9
$f(n)$	0	3	6	9	2	5	8	1	4	7

It should be obvious that $\text{range}(f) = \mathbb{Z}_{10}$.

4. Let $f : \mathbb{Z}_{10} \to \mathbb{Z}_{10}$ be defined by $f : n \mapsto 4n \pmod{10}$. This time the function is

n	0	1	2	3	4	5	6	7	8	9
$f(n)$	0	4	8	2	6	0	4	8	2	6

with $\text{range}(f) = \{0, 2, 4, 6, 8\}$.

Injections, surjections and bijections

Definition 4.11. A function $f : A \to B$ is *1–1* (one-to-one), *injective*, or an *injection*, if f never takes the same value twice. Equivalently[20]

$$\forall a_1, a_2 \in A,\ f(a_1) = f(a_2) \Rightarrow a_1 = a_2$$

$f : A \to B$ is *onto*, *surjective*, or a *surjection*, if $B = \text{range}(f)$. Equivalently

$$\forall b \in B, \exists a \in A \text{ such that } f(a) = b$$

$f : A \to B$ is *invertible*, *bijective*, or a *bijection*, if it is both 1–1 and onto.

Examples. First we consider our examples above: for examples 2, 3 and 4, make sure you understand why the answer is correct, and, if necessary, construct an argument similar to example 1 to explain it.

1. Neither 1–1 nor onto. Indeed:

(1–1) $f(-1) = f(1)$ shows that f is not 1–1.

(Onto) $81 \in \mathbb{R}$, yet there is no $x \in [-3, 2)$ such that $f(x) = 81$. Thus f is not onto.

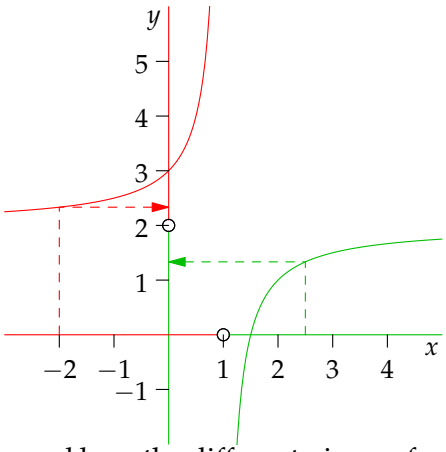
2. Neither 1–1 nor onto.

3. A bijection: this function is an example of a *permutation*, a bijection from a set onto itself.

4. Neither 1–1, nor onto.

[20] This is the contrapositive of the English definition.

5. Prove that $f : \mathbb{R} \setminus \{1\} \to \mathbb{R} \setminus \{2\}$ defined by $f(x) = 2 + \frac{1}{1-x}$ is bijective.



(1–1) Suppose that $f(x_1) = f(x_2)$. Then $2 + \frac{1}{1-x_1} = 2 + \frac{1}{1-x_2}$. A little elementary algebra shows that $x_1 = x_2$, whence f is 1–1.

(Onto) Let $y \in \mathbb{R} \setminus \{2\}$ and define $x = 1 - \frac{1}{y-2}$. This makes sense since $y \neq 2$. But then

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

whence f is onto.

The graphic is colored so that you can see the range and domain, and how the different pieces of range and domain correspond bijectively. The argument for onto is sneaky: how did we know to choose $x = 1 - \frac{1}{y-2}$? The answer is scratch work: just solve $y = 2 + \frac{1}{1-x}$ for x. Essentially we've shown that f has inverse function $f^{-1}(x) = 1 - \frac{1}{x-2}$.

Inverse Functions

You should recall finding the inverse of a 1–1 function from calculus. Indeed if f is invertible, then there is an *inverse function* $f^{-1} : B \to A$.

In calculus we saw that any 1–1 function has an inverse. To square this with our definition, consider $f : [0,3] \to \mathbb{R} : x \mapsto x^2$. This is 1–1 but not onto. If we restrict $B = \mathbb{R}$ to the range of the function, then we get a bijective function: indeed

$$g : [0,3] \to [0,9] : x \mapsto x^2 \text{ has inverse } g^{-1} : [0,9] \to [0,3] : x \mapsto \sqrt{x}$$

In calculus we didn't nitpick like this; we'd simply go straight to $f^{-1}(x) = \sqrt{x}$.
In general, if $f : A \to B$ is any 1–1 function, then $g : A \to f(A) : x \mapsto f(x)$ is automatically bijective. More conceptually, given $b \in B$, onto and 1–1 (in that order) merely encapsulate the existence and uniqueness of an element a satisfying $f(a) = b$.

Injections/Surjections and Cardinality

An important observation for finite sets is the following:

Theorem 4.12. *Let A, B be finite sets. Then*

$$\begin{aligned} |A| \leq |B| &\iff \exists f : A \to B \text{ injective} \\ &\iff \exists g : B \to A \text{ surjective} \end{aligned}$$

The proof illustrates a commonly used technique: when showing that multiple statements are equivalent, it is enough to prove in a circle: in our case with three statements, the proof requires three arguments,

$$① \implies ② \implies ③ \implies ①$$

Proof. The proof relies on the fact that A, B are finite. Suppose that $|A| = m$ and $|B| = n$ throughout and list the elements as

$$A = \{a_1, a_2, \ldots, a_m\} \qquad B = \{b_1, b_2, \ldots, b_n\}$$

- Assume that $m \leq n$. Define $f : A \to B$ by $f(a_k) = b_k$. This is injective.
- Suppose that $f : A \to B$ is injective. Without loss of generality (why?) we may assume that elements of A, B are labeled such that $f(a_k) = b_k$. Now define $g : B \to A$ by

$$g(b_k) = \begin{cases} a_k & \text{if } k \leq m \\ a_1 & \text{if } k > m \end{cases}$$

  Then g is surjective.
- Finally suppose that $g : B \to A$ is surjective. WLOG we may assume that, with the elements of A, B labeled as above, we have $a_k = g(b_k)$ for $1 \leq k \leq m$. Thus $n \geq m$.

■

Corollary 4.13. *If A, B are finite sets, then $|A| = |B| \iff \exists f : A \to B$ bijective.*

Using injective and surjective functions to compare cardinalities becomes a *definition* when one works with infinite sets. Such pleasures will have to wait…

Composition of functions

Definition 4.14. Suppose that $f : A \to B$ and $g : B \to C$ are functions. The *composition* $g \circ f : A \to C$ is the function defined by $(g \circ f)(a) = g(f(a))$.

Implied Domains

You should be careful to make sure that $\text{range}(f) \subseteq \text{dom}(g)$ before composing functions. In calculation-based classes, the domain and range are not always explicitly mentioned, and at times some restriction of the domain is implied. For example, in calculus you might happily have written

$$f(x) = x^2, \quad g(x) = \frac{1}{x-1} \implies (g \circ f)(x) = \frac{1}{x^2 - 1}$$

without mentioning domains. According to our definition, the domains of f and $g \circ f$ should be identical. The implied domain of $g \circ f$ is $\mathbb{R} \setminus \{\pm 1\}$. Because we are composing with g, we should really view f as a function $f : \mathbb{R} \setminus \{\pm 1\} \to \mathbb{R}$.

Theorem 4.15. *Let $f : A \to B$ and $g : B \to C$ be functions. Then:*

1. *If f and g are injective, then $g \circ f$ is injective.*
2. *If f and g are surjective, then $g \circ f$ is surjective.*

Proof. 1. Suppose that f and g are injective and let $a_1, a_2 \in A$ satisfy $(g \circ f)(a_1) = (g \circ f)(a_2)$. We are required to show that $a_1 = a_2$. However,

$$\begin{aligned}(g \circ f)(a_1) = (g \circ f)(a_2) &\implies g\big(f(a_1)\big) = g\big(f(a_2)\big) \\ &\implies f(a_1) = f(a_2) && \text{(since } g \text{ is injective)} \\ &\implies a_1 = a_2 && \text{(since } f \text{ is injective)}\end{aligned}$$

■

Part (2) is left to the homework.

It follows that the composition of bijective functions is also bijective.

It is an interesting exercise to note that the converse of this theorem is *false.* Assuming that a composition is injective or surjective only requires one of the component functions to be so.

Theorem 4.16. *Suppose that $f : A \to B$ and $g : B \to C$ are functions.*

1. *If $g \circ f$ is injective, then f is injective.*
2. *If $g \circ f$ is surjective, then g is surjective.*

Here we give a sketch proof of the first result. Try to prove the second yourself.

Proof of 1. Suppose that $f(a_1) = f(a_2)$. Then $(g \circ f)(a_1) = (g \circ f)(a_2)$. But if $g \circ f$ is injective, we conclude that $a_1 = a_2$, whence f is injective.

■

The interesting part is producing examples. For instance, take

$$\begin{aligned}&f : [0,2] \to [-4,4] : x \mapsto x^2 && \text{(injective only)} \\ &g : [-4,4] \to [0,16] : x \mapsto x^2 && \text{(surjective only)} \\ &g \circ f : [0,2] \to [0,16] : x \mapsto x^4 && \text{(bijective!)}\end{aligned}$$

5 Mathematical Induction and Well-ordering

5.1 Proof by Induction

Induction is the mathematical equivalent of a domino rally; toppling the nth domino causes the $(n+1)$th domino to fall, to knock all the dominos over it is enough merely to topple the first.

In mathematics, we think of each domino as being a *proposition.* The nth domino is a proposition $P(n)$ (depending on n). Induction demonstrates the truth of *every* $P(n)$ by doing two things:

1. Proving that $P(1)$ is true (knock over the first domino)
2. Proving that $\forall n \in \mathbb{N},\ P(n) \implies P(n+1)$ (the nth domino knocks over the $(n+1)$th)

Putting these together we have: $P(1)$ is true, therefore $P(2)$ is true, therefore $P(3)$ is true, etc. More logically: $P(1) \implies P(2) \implies P(3) \implies P(4) \implies P(5) \implies \cdots$.

As a motivating example, suppose you are asked to prove the following theorem.

Theorem 5.1. *The sum of the first n positive integers is given by the formula*

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

We could attempt a direct proof, à la Gauss, by rearranging the order of the terms to make pairs summing to $n+1$:

$$1+2+\cdots+(n-1)+n = [1+n]+[2+(n-1)]+[3+(n-2)]+\cdots$$

Deciding on the final term of this sequence involves tediously dealing with the separate cases of n being even or odd. Instead we prove by Induction.

Proof. (Define $P(n)$) Let $P(n)$ be the proposition: $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$.

(Prove $P(1)$) Clearly $\sum_{i=1}^{1} i = 1 = \frac{1}{2}1(1+1)$, so $P(1)$ is true.

(Prove $P(n) \implies P(n+1)$) Assume that $P(n)$ is true for some fixed n. The sum of the first $n+1$ positive integers is

$$\begin{aligned}\sum_{i=1}^{n+1} i &= (n+1) + \sum_{i=1}^{n} i \\ &= (n+1) + \frac{1}{2}n(n+1) \qquad \text{(by assumption of } P(n)\text{)} \\ &= \left(1+\frac{1}{2}n\right)(n+1) = \frac{1}{2}(n+2)(n+1) \\ &= \frac{1}{2}(n+1)\Big([n+1]+1\Big).\end{aligned}$$

This last says that $P(n+1)$ is true.

(Conclusion) By mathematical induction, $P(n)$ is true for all $n \in \mathbb{N}$. ∎

Note how we grouped $\frac{1}{2}(n+1)\Big([n+1]+1\Big)$ so that it is obviously the right hand side of $P(n+1)$.

With a little practice, you'll find there is no need to give all the logical details. One can get away without explicitly mentioning $P(n)$. While learning induction, however, it is a good idea to write proofs in the style of the above model.

Here is an example in the same vein, but done a little faster. If the logic seems unclear, try rewriting exactly as above.

Theorem 5.2. *Prove that $n(n+1)(2n+1)$ is divisible by 6 for any natural number n.*

Proof. Let $P(n)$ be the proposition: $n(n+1)(2n+1)$ is divisible by 6.
Clearly $1\cdot(1+1)\cdot(2\cdot 1+1) = 6$ is divisible by 6, hence $P(1)$ is true.
Now assume that $P(n)$ is true for some fixed $n \in \mathbb{N}$. Then

$$\begin{aligned}(n+1)(n+2)(2(n+1)+1) - n(n+1)(2n+1) &= (n+1)((n+2)(2n+3) - n(2n+1)) \\ &= (n+1)(2n^2+7n+6-2n^2-n) \\ &= 6(n+1)^2.\end{aligned}$$

This is divisible by 6, and hence, by the induction hypothesis,[21] so is $(n+1)(n+2)(2(n+1)+1)$. Thus $P(n+1)$ is true.
By induction $P(n)$ is true for all n.

■

Here is another simple example: this time we don't explicitly name $P(n)$.

Theorem 5.3. $2+5+8+\cdots+(3n-1)=\frac{1}{2}n(3n+1)$.

Proof. For $n=1$ we have $2=2$, hence the proposition holds. Now suppose the proposition holds for some $n\in\mathbb{N}$. Then

$$\begin{aligned}2+5+\cdots+(3(n+1)-1) &= (2+5+\cdots+(3n-1))+3n+2\\ &= \frac{1}{2}n(3n+1)+3n+2=\frac{1}{2}(3n^2+7n+4)\\ &= \frac{1}{2}(n+1)(3n+4)=\frac{1}{2}(n+1)(3(n+1)+1).\end{aligned}$$

But this says that the proposition holds for $n+1$. By mathematical induction the proposition holds for all n.

■

5.2 Well-ordering and the Principle of Mathematical Induction

Simple induction proofs always follow exactly the same pattern as the above examples. For harder proofs it is typically the *induction step* (the argument for $P(n)\Rightarrow P(n+1)$) that gets longer. The logical structure remains the same. Before seeing more examples, it is worth understanding *why* induction works. It in fact depends on a fundamental property of the natural numbers.

Definition 5.4. A set of real numbers A is *well-ordered* if every non-empty subset of A has a least element.

Examples.
1. $A=\{4,-7,\pi,19,\ln 2\}$ is a well-ordered set. There are 31 non-empty subsets[22] of A, each of which has a minimum element.
2. The interval $[3,10)$ is not well-ordered. Indeed $(3,4)$ is a non-empty subset which has no minimum element.
3. The integers $\mathbb{Z}$ are not well-ordered, since there is no minimum integer.

In general, every finite set of numbers is well-ordered, and intervals are not. Are there *infinite* well-ordered sets? The answer is yes. Indeed it is part of the standard definition[23] of the natural numbers that $\mathbb{N}$ is such a set.

Axiom. $\mathbb{N}$ is well-ordered.

Armed with this axiom, we can justify Induction.

Theorem 5.5 (Principle of Mathematical Induction)**.** *Let $P(n)$ be a proposition for each $n\in\mathbb{N}$. Suppose:*

[21] The *induction hypothesis* is the assumption in the middle of an induction argument that $P(n)$ is true.
[22] Can you justify this *without* listing them?
[23] Peano's Axioms.

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

(b) $\forall n \in \mathbb{N},\ P(n) \implies P(n+1)$.

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

Proof. Argue by contradiction. Let $S = \{m \in \mathbb{N} : P(m) \text{ is false}\}$ and suppose that S is non-empty. Since S is a non-empty subset of the well-ordered set $\mathbb{N}$, it follows that S has a least element. Call this N.
By (a) we have $P(1)$ true. Clearly $N \neq 1$, therefore $N \geq 2$ and so $N - 1 \in \mathbb{N}$.
Since $N = \min S$ it follows that $P(N-1)$ is true.
But (b) then forces $P(N)$ to be true. A contradiction.
We conclude that $S = \emptyset$ and $P(n)$ is true for all $n \in \mathbb{N}$. ∎

An induction argument need not begin with $n = 1$. By proving Theorem 5.5 it should be clear where we used the well-ordering of $\mathbb{N}$. Indeed induction works to prove any well-ordered set of propositions. Since, for any fixed $m \in \mathbb{Z}$, the set

$$\{n \in \mathbb{Z} : n \geq m\} = \{n \in \mathbb{Z} : n - m \in \mathbb{N}_0\}$$

is well-ordered, the following modification of the induction principle is immediate:

Corollary 5.6 (Principle of Mathematical Induction)**.** *Fix* $m \in \mathbb{Z}$. *Let* $P(n)$ *be a proposition for each integer* $n \geq m$. *Suppose:*

(a) $P(m)$ *is true.*

(b) $\forall n \geq m,\ P(n) \implies P(n+1)$.

Then $P(n)$ *is true for all* $n \geq m$.

Here is an example where we begin with $n = 0$.

Theorem 5.7. $3 \mid (2^n + 2^{n+1})$ *for all non-negative integers* $n \in \mathbb{N}_0$.

Proof. If $n = 0$ we have $2^n + 2^{n+1} = 3$, so the proposition is true.
Now suppose that $3 \mid (2^n + 2^{n+1})$ for some $n \in \mathbb{N}_0$. Then

$$\exists k \in \mathbb{Z} \text{ such that } 2^n + 2^{n+1} = 3k.$$

But then

$$2^{n+1} + 2^{(n+1)+1} = 2(2^n + 2^{n+1}) = 6k$$

is divisible by 3, so the proposition is true for $n + 1$.
By induction $3 \mid (2^n + 2^{n+1})$ for all $n \in \mathbb{N}_0$. ∎

You should try to prove the above directly; it's much easier than induction!

The next example is reminiscent of sequences and series from elementary calculus. Compare the direct proof of such a formula given in a calculus text with the following.

Theorem 5.8. *Prove that* $\forall n \in \mathbb{N}$ *such that* $n \geq 3$ *we have* $\displaystyle\sum_{i=3}^{n} \frac{1}{i(i-2)} = \frac{3}{4} - \frac{2n-1}{2n(n-1)}$.

Proof. The initial case is $n = 3$. But this simply reads $\displaystyle\sum_{i=3}^{3} \frac{1}{i(i-2)} = \frac{1}{3} = \frac{3}{4} - \frac{5}{12}$, which is true.
Now assume that the formula is true for some n. Then

$$\begin{aligned}\sum_{i=3}^{n+1} \frac{1}{i(i-2)} &= \sum_{i=3}^{n} \frac{1}{i(i-2)} + \frac{1}{(n+1)(n-1)} = \frac{3}{4} - \frac{2n-1}{2n(n-1)} + \frac{1}{(n+1)(n-1)} \\ &= \frac{3}{4} + \frac{1}{2(n+1)n(n-1)}\left[-(2n-1)(n+1) + 2n\right] \\ &= \frac{3}{4} + \frac{1+n-2n^2}{2(n+1)n(n-1)} = \frac{3}{4} + \frac{(2n+1)(1-n)}{2(n+1)n(n-1)} \\ &= \frac{3}{4} - \frac{2n+1}{2(n+1)n}\end{aligned}$$

which is exactly the formula with n replaced with $n+1$. By induction the expression is true for all $n \geq 3$.

■

Induction can also be used to prove results for all integers, sometimes by doing induction in two directions, other times by using induction as part of a larger proof. In this example[24] we begin with the case $n = 0$, use induction to prove for $n \geq 0$, and use properties of matrices to prove for $n < 0$.

Theorem 5.9. *If* $A = \begin{pmatrix} 7 & 12 \\ -2 & -3 \end{pmatrix}$, *prove that*

$$A^n = \begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^n \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}, \ \forall n \in \mathbb{Z}.$$

Here $A^{-n} = (A^n)^{-1}$.

Proof. For $n = 0$, the formula is trivially true. Now suppose it is true for some n. Then

$$\begin{aligned}A^{n+1} = A^n A &= \left(\begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^n \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}\right)\begin{pmatrix} 7 & 12 \\ -2 & -3 \end{pmatrix} \\ &= \begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^n \begin{pmatrix} 9 & 18 \\ -3 & -6 \end{pmatrix} \\ &= \begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^{n+1} \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}.\end{aligned}$$

By induction we have established the formula for all $n \geq 0$.
Now calculate

$$\begin{aligned}&\left(\begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^{-n} \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}\right)\left(\begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + 3^n \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}\right) \\ &\qquad = \begin{pmatrix} -2 & -6 \\ 1 & 3 \end{pmatrix} + \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix} + 3^n \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + 3^{-n} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\end{aligned}$$

whence the formula also holds for $n < 0$.

■

[24] If you've not done Math 3A, don't worry about this problem!

5.3 Recurrence Relations and Sequences

Induction is intimately related to the way we think about sequences. Often one has a simple recurrence relation but not general formula. I.e. our sequence is defined by a rule $x_{n+1} = f(x_n)$ where x_1 is given. Induction can allow us to prove general formulas for such sequences, and prove general results, such as convergence, without finding a general formula. Here is a simple example.

Example. Consider the operation whereby you take a stack of paper, cut all sheets in half, then stack both halves together. If a sheet of paper has thickness 0.1 mm, how many times would you have to repeat the operation until the stack of paper reached to the sun? (Approx 150 million kilometers).

If h_n is the height of the stack after n operations, then it is clear that we have a sequence $(h_n)_{n=0}^{\infty}$ satisfying the recurrence relation

$$\begin{cases} h_{n+1} = 2h_n, \\ h_0 = 0.1\,\text{mm}. \end{cases}$$

It is not hard to hypothesize that, after n such operations, the stack of paper will have height

$$h_n = 2^n \times 0.1\,\text{mm} = 2^n \times 10^{-4}\,\text{m}.$$

How do we *know* our guess is correct? By a very simple induction argument:

If $n = 0$, then $h_0 = 2^0 \times 0.1 = 0.1\,\text{mm}$.
If $h_n = 2^n \times 0.1$ for some n, then $h_{n+1} = 2h_n = 2^{n+1} \times 0.1\,\text{mm}$.

Finishing the problem is then easy: We need to find $n \in \mathbb{N}$ such that

$$h_n = 2^n \times 10^{-4} \geq 150 \times 10^9 \text{m}.$$

Indeed $n \geq \frac{14+\log_{10} 15}{\log_{10} 2} \approx 50.4$, so 51 iterations are sufficient.

The Tower of Hanoi

The famous *Tower of Hanoi* problem involves three pegs on which are stacked circular disks of decreasing radii. A disk may only be placed on top of a larger disk, or on an empty peg. The idea is to move all the disks from one peg to another. How many moves does this take? This can be solved fairly easily by constructing a recurrence relation, and proved to be correct by induction.

Suppose that the minimum number of moves required to move a stack of height n to another peg is r_n. Now suppose that we have $n+1$ disks on the first peg. In order to move the bottom disk we need to have moved all the n disks above it onto another peg. This takes r_n moves. Then we move the bottom disk; 1 move. Then we move the top n disks onto the bottom disk; r_n further moves. Thus $r_{n+1} = 2r_n + 1$. By inspecting this formula and thinking about small values of n ($n = 1$ clearly requires only 1 move, while $n = 2$ requires 3, etc...) we hypothesize the following:

Theorem 5.10. *The Tower of Hanoi with n disks can be completed in a minimum of $r_n = 2^n - 1$ moves.*

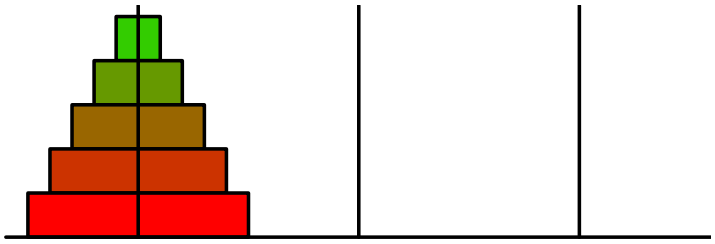
Proof. The formula is clearly true for $n = 1$. Now suppose that it is true for some n. Then

$$r_{n+1} = 2r_n + 1 = 2(2^n - 1) + 1 = 2^{n+1} - 2 + 1 = 2^{n+1} - 1.$$

Thus the formula holds for $n+1$. By induction the formula holds for all $n \in \mathbb{N}$. ∎

As an illustration of how ridiculously time-consuming the Tower becomes, if you were able to move one disk per second, Towers consisting of the following numbers of disks would take you approximately the following times to move:

Disks	Time
5	31sec
10	17min 3sec
15	9hr 6min 7sec
20	12days 3hrs 16min 15sec
25	$\sim$ 1yr 23days
30	$\sim$ 34yrs 9days



Animation of five disks: click…

The induction arguments in the above examples are so simple that they hardly seem worth mentioning. In other situations things can be much harder.
Bob

Example. Recall the monotone convergence theorem from sequences. If (x_n) is an increasing (decreasing) sequence bounded above (below), then it is convergent. Here we use this theorem to prove that the following sequence converges to $\frac{1}{2}$:

$$\begin{cases} x_{n+1} = \frac{1}{3}(x_n + 1) + (x_n - \frac{1}{2})^2, \\ x_1 = 1. \end{cases}$$

You can try hunting for a general formula for x_n (if you find one, let us know…). Instead we observe the first few terms of the sequence: $(x_n) = (1, \frac{11}{12}, \frac{13}{16}, \frac{539}{768}, \ldots)$ and hypothesize:

Conjecture: (x_n) is a decreasing sequence and $x_n > \frac{1}{2}$ for all $n \in \mathbb{N}$.
We prove by induction.
Certainly $x_1 = 1 > \frac{1}{2}$. Now if $x_n > \frac{1}{2}$, we have $x_n - \frac{1}{2} \neq 0$, whence

$$x_{n+1} > \frac{1}{3}(x_n + 1) > \frac{1}{3}\left(\frac{1}{2} + 1\right) = \frac{1}{3} \cdot \frac{3}{2} = \frac{1}{2}.$$

Thus all $x_n > \frac{1}{2}$ by induction.

Given this, we can now see that

$$x_{n+1} - x_n = \frac{1}{3}(1 - 2x_n) - \left(x_n - \frac{1}{2}\right)^2 < 0,$$

thus (x_n) is decreasing. Since (x_n) is also bounded below (by $\frac{1}{2}$), the monotone convergence theorem says that the sequence converges.
Call the limit x. Clearly $x \geq 1$. But then

$$x = \frac{1}{3}(x + 1) + \left(x - \frac{1}{2}\right)^2 \iff x = \frac{1}{2} \text{ or } \frac{7}{6}.$$

Since (x_n) is decreasing from 1, it is clear that $\lim_{n\to\infty} x_n = \frac{1}{2}$.
Note that it is essential that we establish the existence of the limit before calculating it: the same sequence but with initial value $x_1 = 2$ is *divergent* to ∞.

5.4 Strong induction

The principle of mathematical induction is sometimes known as *weak* induction. In weak induction, the induction step requires only that one proposition $P(n)$ is true to demonstrate the truth of $P(n+1)$. In strong induction we assume that some, or even *all*, of the earlier propositions are true.

Theorem 5.11 (Principle of Strong induction)**.** *Let $P(n)$ be a proposition for each $n \in \mathbb{N}$ and fix $m \in \mathbb{N}$. Suppose:*

(a) $P(1), P(2), \ldots, P(m)$ *are true.*

(b) $\forall n \geq m,\ (P(1) \wedge P(2) \wedge \cdots \wedge P(n)) \implies P(n+1)$.

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

In particular strong and weak induction are logically equivalent.
Here are two examples: in both cases it is enough to take $m = 2$ in the above Theorem.

Definition 5.12. The Fibonacci numbers are defined by the recurrence relation

$$f_{n+1} = f_n + f_{n-1} \text{ for } n \geq 2, \text{ where } f_1 = f_2 = 1.$$

Theorem 5.13. *Prove that $f_n < 2^n$ for all n.*

Proof. Let $P(n)$ be the proposition $f_n < 2^n$. Clearly $P(n)$ is true for $n = 1, 2$.
Suppose now that $P(1), \ldots, P(n)$ are true for some $n \geq 2$. Then

$$f_{n+1} = f_n + f_{n-1} < 2^n + 2^{n-1} < 2^n + 2^n = 2^{n+1}.$$

But this is $P(n+1)$. By strong induction $P(n)$ is true for all n. ∎

In the proof we had to assume that the relation $f_n < 2^n$ was true for two values of n.

Theorem 5.14. *Suppose that a sequence of integers $(a_n)_{n=0}^{\infty}$ is defined by*

$$\begin{cases} a_n = 5a_{n-1} - 6a_{n-2}, & n \geq 2 \\ a_0 = 0,\ a_1 = 1 \end{cases}$$

Then $a_n = 3^n - 2^n$ for all $n \in \mathbb{N}_0$.

Proof. The formula is clearly true for $a_0 = 0 = 3^0 - 2^0$ and $a_1 = 1 = 3^1 - 2^1$.
Suppose it is true for all integers $\leq n$. Then

$$\begin{aligned} a_{n+1} &= 5a_n - 6a_{n-1} = 5(3^n - 2^n) - 6(3^{n-1} - 2^{n-1}) \\ &= (15-6)3^{n-1} + (10-6)2^{n-1} = 3^{n+1} - 2^{n+1}. \end{aligned}$$

By induction the formula is true for all $n \in \mathbb{N}_0$. ∎

6 Set Theory, Part II

Now we return to sets, where we consider several more-advanced constructions.

6.1 Cartesian Products

You've been working with Cartesian products for years. Referring to a point in the plane by its *Cartesian co-ordinate* (x, y) is the classic example. The basic idea is that each of the co-ordinates x and y is a member of the set $\mathbb{R}$.

Definition 6.1. Let A, B be sets. Their *Cartesian product* is the set[25]

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

The Cartesian product is exactly the set of ordered pairs.

Examples. 1. The Cartesian product of $\mathbb{R}$ with itself is what we call the xy-plane: rather than writing $\mathbb{R} \times \mathbb{R}$ which is unwieldy, we write $\mathbb{R}^2$.

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x, y \in \mathbb{R}\}.$$

More generally, $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \mathbb{R}}_{n \text{ times}}$ is the set of n-tuples of real numbers:

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

2. Suppose you go to a restaurant where you have a choice of Main Courses and Sides. Mathematically we might say:

$$\text{Mains} = \{\text{fish, steak, eggplant, pasta}\}$$
$$\text{Sides} = \{\text{asparagus, salad, potatoes}\}$$

The Cartesian product Mains×Sides is the set of all possible meals made up of one main and one side. It should be obvious that there are 4×3 possible meal choices.

This last example illustrates the following theorem. Indeed it partly explains the use of the word 'product' in the definition.

Theorem 6.2. *If A and B are finite sets, then $|A \times B| = |A| \cdot |B|$.*

Proof. Simply label the elements of each set and list the elements of $A \times B$ lexicographically. If $|A| = m$ and $|B| = n$, then we have:

$$\begin{array}{cccccc} (a_1, b_1) & (a_1, b_2) & (a_1, b_3) & \cdots & (a_1, b_n) \\ (a_2, b_1) & (a_2, b_2) & (a_2, b_3) & \cdots & (a_2, b_n) \\ \vdots & \vdots & \vdots & & \vdots \\ (a_m, b_1) & (a_m, b_2) & (a_m, b_3) & \cdots & (a_m, b_n) \end{array}$$

It should be clear that every element of $A \times B$ is listed exactly once. There are m rows and n columns, thus $|A \times B| = mn$. ∎

As an example of a basic set relationship involving Cartesian products, we prove a theorem:

[25] A strict set-theoretic definition requires you to build the ordered pair (a, b) as a set: typically $\{a, \{a, b\}\}$. One then proves that $(a, b) = (c, d) \iff a = b$ and $c = d$. Such details are beyond the scope of this class.

Theorem 6.3. *Let A, B, C, D be sets. Then $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.*

Proof. Since we are dealing with a Cartesian product of two sets, it is easier to write the general element in the form (x, y).
Let $(x, y) \in (A \times B) \cup (C \times D)$. Then

$$(x, y) \in A \times B \qquad \text{or} \qquad (x, y) \in C \times D.$$

But then

$$(x \in A \text{ and } y \in B) \qquad \text{or} \qquad (x \in C \text{ and } y \in D).$$

Clearly $x \in A$ or $x \in C$, so $x \in A \cup C$.
Similarly $y \in B$ or $y \in D$, so $y \in B \cup D$.
Therefore $(x, y) \in (A \cup C) \times (B \cup D)$, as required. ∎

A careful reading of the proof should convince you that the two sets in the theorem are not equal: if $x \in A$ and $y \in D$, then (x, y) is an element of the right hand side, but not the left.

6.2 Power Sets

Given a set A, we often want to consider all possible subsets of A. If we collect these subsets together we want to call the collection a set.

Definition 6.4. The *power set* of A is the set $\mathcal{P}(A)$ of all subsets of A:

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

Otherwise said: $B \in \mathcal{P}(A) \iff B \subseteq A$.

Example. Let $A = \{1, 3, 7\}$. Then A has the following subsets:

0-element :	$\emptyset$
1-element:	$\{1\}, \{3\}, \{7\}$
2-element:	$\{1,3\}, \{1,7\}, \{3,7\}$
3-element:	$\{1,3,7\}$

Gathering these together, we have

$$\mathcal{P}(A) = \Big\{\emptyset, \{1\}, \{3\}, \{7\}, \{1,3\}, \{1,7\}, \{3,7\}, \{1,3,7\}\Big\}.$$

Be careful with notation here: in the above example, we can write both $1 \in A$ and $\{1\} \in \mathcal{P}(A)$.
The following theorem is mostly an exercise in following notation: read it carefully.

Theorem 6.5. *If $A \subseteq B$, then $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.*

Proof. Suppose that $A \subseteq B$ and $C \in \mathcal{P}(A)$. We must show that $C \in \mathcal{P}(B)$. Now $C \subseteq A$. But subsets are transitive (Theorem 4.4), whence

$$C \subseteq A \subseteq B \implies C \subseteq B,$$

which is exactly $C \in \mathcal{P}(A)$. ∎

If you know a little about combinations, it should be clear that a set A with n elements has precisely ${}^nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$ distinct subsets with r elements.

Theorem 6.6. *Suppose that A is a finite set. Then $|\mathcal{P}(A)| = 2^{|A|}$.*

The basic idea of the proof is that every set with $n+1$ elements is the disjoint union of a set with n elements and asingle-element set. We will prove by induction: supposing that $|A| = n$, we must show that $|\mathcal{P}(A)| = 2^n$. The trick here is to do induction on the cardinalty of A: thus *all* sets A with the same cardinality are dealt with at once.

Proof. Let $S(n)$ be the proposition: "If $|A| = n$ then $|\mathcal{P}(A)| = 2^n$."
Initial case: If $n = 0$, then $A = \emptyset$. But then $\mathcal{P}(A) = \{\emptyset\}$, whence $|\mathcal{P}(A)| = 1 = 2^0$. Therefore $S(0)$ is true.
Now assume that $S(n)$ is true for some fixed n, and let B be any set containing $n+1$ elements. Choose one of the elements $b \in B$ and define $A = B \setminus \{b\}$. Since $b \notin A$, we have that $B = A \cup \{b\}$ is a disjoint union.
To any subset $X \subseteq A$ there correspond exactly two subsets of B, namely X and $X \cup \{b\}$. Therefore

$$|\mathcal{P}(B)| = 2\,|\mathcal{P}(A)| = 2 \cdot 2^n = 2^{n+1},$$

by $S(n)$. By induction, $S(n)$ is true for all n. ∎

An alternative proof[26] also uses induction to prove that $\sum_{r=0}^{n} \binom{n}{r} = 2^n$.

Power sets and Cartesian products can be the source of great confusion, so take your time with them! As an example, consider the power set of a Cartesian product. Suppose that $A = \{x\}$ and $B = \{y, z\}$. Then

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

The power set $\mathcal{P}(A \times B)$ therefore contains $2^2 = 4$ elements:

$$\mathcal{P}(A \times B) = \Big\{\emptyset,\ \{(x,y)\},\ \{(x,z)\},\ \{(x,y),(x,z)\}\Big\}.$$

$\mathcal{P}(A) = \{\emptyset, \{x\}\}$, and $\mathcal{P}(B) = \{\emptyset, \{y\}, \{z\}, \{y,z\}\}$ have 2 and 4 elements respectively. The Cartesian product of the power sets therefore has $2 \times 4 = 8$ elements:

$$\mathcal{P}(A) \times \mathcal{P}(B) = \Big\{(\emptyset,\emptyset),\ (\emptyset,\{y\}),\ (\emptyset,\{z\}),\ (\emptyset,\{y,z\}),\ (\{x\},\emptyset),\ (\{x\},\{y\}),\ (\{x\},\{z\}),\ (\{x\},\{y,z\})\Big\}.$$

It should be clear from this example not only that $\mathcal{P}(A \times B) \neq \mathcal{P}(A) \times \mathcal{P}(B)$, but that the elements of the two sets look completely different. In fact there is no general relationship between the sets!

[26] The induction step requires you to use the identity $\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}$.

6.3 Indexed Collections of Sets

An indexed family of sets is a collection of sets A_i, one for each i in some indexing set I. It is often the case that $I = \mathbb{N}, \mathbb{Z}$ or $\mathbb{R}$, and the label for the index is often chosen accordingly: e.g. $n \in \mathbb{N}$ or $x \in \mathbb{R}$, etc.

Examples. 1. Let $A_n = [-n, n] \subseteq \mathbb{R}$, for each $n \in \mathbb{N}$. For example $A_1 = [-1, 1]$, $A_2 = [-2, 2]$, etc.

2. Let $A_n = (n, n+1] \subseteq \mathbb{R}$, for each $n \in \mathbb{Z}$. E.g. $A_{-17} = (-17, -16]$.

3. Let $A_n = [0, \frac{1}{n}) \subseteq \mathbb{R}$, for each $n \in \mathbb{N}$. E.g. $A_{1000} = [0, \frac{1}{1000})$.

4. Let $A_n = (0, \frac{1}{n}) \subseteq \mathbb{R}$, for each $n \in \mathbb{N}$.

5. Let $A_n = \{x \in \mathbb{R} : |x^2 - 1| < \frac{1}{n}\}$, for each $n \in \mathbb{N}$. Here $A_3 = \left(-\sqrt{\frac{4}{3}}, -\sqrt{\frac{2}{3}}\right) \cup \left(\sqrt{\frac{2}{3}}, \sqrt{\frac{4}{3}}\right)$.

Definition 6.7. Given a family of indexed sets $\mathcal{A} = \{A_i : i \in I\}$, we may form the *union* and *intersection* of the A_i:

$$\cup\mathcal{A} = \bigcup_{i \in I} A_i = \{x : x \in A_i \text{ for some } i \in I\}$$
$$\cap\mathcal{A} = \bigcap_{i \in I} A_i = \{x : x \in A_i \text{ for all } i \in I\}.$$

A collection $\mathcal{A} = \{A_i : i \in I\}$ is *pairwise disjoint* if $A_i \cap A_j = \emptyset$ whenever $i \neq j$.

When the indexing set is $\mathbb{N}$ or $\mathbb{Z}$, it is also common to write as in summation ($\sum$) notation: for instance $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n=1}^{\infty} A_n$.

Let us return to our examples and calculate some unions and intersections. When no working is given, see if you can provide an argument yourself.

Examples. 1. If $A_n = [-n, n]$, then it should be clear that $n \leq m \implies A_n \subseteq A_m$. It follows that

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots, \text{whence } \bigcap_{n \in \mathbb{N}} A_n = A_1 = [-1, 1].$$

The union is a little harder. If $x \geq 0$, then $x \leq \lceil x \rceil$, whence $x \in A_{\lceil x \rceil}$. Similarly, if $x < 0$, then $x \in A_{-\lfloor x \rfloor}$. For example,

$$-3.124 \in A_{-\lfloor -3.124 \rfloor} = A_{-(-4)} = A_4.$$

It follows that all real numbers x are in at least one of these sets A_n, and so $\bigcup_{n \in \mathbb{N}} A_n = \mathbb{R}$.

2. $\bigcup_{n \in \mathbb{Z}} A_n = \mathbb{R}$, and $\bigcap_{n \in \mathbb{Z}} A_n = \emptyset$. These sets are pairwise disjoint.

3. Let $A_n = [0, \frac{1}{n})$ for $n \in \mathbb{N}$. Then $m \leq n \implies \frac{1}{n} \leq \frac{1}{m} \implies A_n \subseteq A_m$. Since the sets A_n are nested, the union is simply the largest such:

$$\bigcup_{n=1}^{\infty} A_n = A_1 = [0, 1).$$

Now for the intersection. First consider a finite intersection. Fix $m \in \mathbb{N}$: since the sets are nested, we have

$$\bigcap_{n=1}^{m} A_n = A_m = \left[0, \frac{1}{m}\right),$$

which is non-empty *for every* m. What about the infinite intersection? You might be tempted to take a limit and make an argument such as

$$\bigcap_{n=1}^{\infty} A_n = \lim_{m\to\infty} \bigcap_{n=1}^{m} A_n = \lim_{m\to\infty} \left[0, \frac{1}{m}\right) = [0,0).$$

Quite apart from the issue that $[0,0)$ is ugly and could only mean the empty set,[27] we should worry about whether this is a legitimate use of limits. It isn't!

We claim: $\bigcap_{n=1}^{\infty} A_n = \{0\}$. Here is a proof:

If $x \in A_n$ for any n, then clearly $x \geq 0$. Certainly $0 \in A_n$ for all $n \in \mathbb{N}$, hence $0 \in \bigcap_{n=1}^{\infty} A_n$.

Now suppose, for contradiction, that $\exists x > 0$ such that $x \in \bigcap_{n=1}^{\infty} A_n$. Then $\forall n \in \mathbb{N}$,

$$x \in A_n \iff x < \frac{1}{n}.$$

Since $\lim_{n\to\infty} \frac{1}{n} = 0$, we can certainly choose[28] N large enough so that $\frac{1}{N} \leq x$. Hence $x \notin A_N$. A contradiction. Thus the intersection contains no positive elements, and we conclude that

$$\bigcap_{n=1}^{\infty} A_n = \{0\}.$$

4. There is no difference between this and example 3, *except* that $0 \notin A_n$. Therefore

$$\bigcup_{n=1}^{\infty} A_n = A_1 = (0,1), \quad \text{and} \quad \bigcap_{n=1}^{\infty} A_n = \emptyset.$$

The moral of these two examples is that you can't naïvely apply limits to sets, and that you have to take your time and be very careful with infinite unions and intersections. You intuition can easily lead you astray: stick rigidly to the definitions!

5. $\bigcup_{n\in\mathbb{N}} A_n = (-\sqrt{2}, 0) \cup (0, \sqrt{2})$, and $\bigcap_{n\in\mathbb{N}} A_n = \{1, -1\}$.

Observe that $\bigcup_{i\in I} A_i = \emptyset \iff \forall i \in I,\ A_i = \emptyset$, while $\bigcap_{i\in I} A_i$ may easily be empty, even if all A_i are non-empty (e.g. example 4). This links in with the following proposition:

[27] $x \in [0,0) \iff x \geq 0$ and $x < 0 \ldots$

[28] Explicitly, you may choose choose $N = \lceil \frac{1}{x} \rceil$, or anything larger…

Theorem 6.8. *Let $\mathcal{A} = \{A_i : i \in I\}$ and let $j \in I$. Then*

$$A_j \subseteq \bigcup_{i\in I} A_i \quad \textit{and} \quad \bigcap_{i\in I} A_i \subseteq A_j.$$

Proof. 1. Let $x \in A_j$. Then $\exists i \in I$ such that $x \in A_i$, and hence $x \in \bigcup_{i\in I} A_i$.

2. Let $x \in \bigcap_{i\in I} A_i$. Then $\forall i \in I$ we have $x \in A_i$. In particular, $x \in A_j$. ∎

A more complex example: finite decimals

Here is another example where 'taking the limit' is not the right thing to do.

For each $n \in \mathbb{N}$, let $A_n = \{\text{length } n \text{ decimals } 0.a_1a_2\ldots a_n\}$, where each $a_i \in \{0, 1, 2, \ldots, 9\}$. For example $0.134 \in A_3$, but $0.134 \notin A_2$. Since $0.134 = 0.1340$, we also have $0.134 \in A_4$. Generalizing, it should be clear that

$$m \leq n \implies A_m \subseteq A_n.$$

It should be clear from this that the infinite intersection is simply

$$\bigcap_{n\in\mathbb{N}} A_n = A_1 = \{0, 0.1, \ldots, 0.9\}.$$

Consider first a finite union: if $m \in \mathbb{N}$, then[29]

$$\bigcup_{n=1}^{m} A_n = \{x \in [0, 1) : x \text{ has a decimal representation of length} \leq m\} = A_m.$$

If one were to take the limit as $m \to \infty$ of the *property* 'length m decimal,' it seems like the infinite union should be the whole[30] interval $[0, 1]$. This is incorrect: one cannot take the limit of a property! Instead use the definiton:

$$\begin{aligned} x \in \bigcup_{n\in\mathbb{N}} A_n &\iff \exists n \in \mathbb{N} \text{ such that } x \in A_n \\ &\iff \exists n \in \mathbb{N} \text{ such that } x \text{ has a length } n \text{ decimal representation.} \end{aligned}$$

It follows that

$$\bigcup_{n\in\mathbb{N}} A_n = \{x \in [0, 1) : x \text{ has a } \textit{finite} \text{ decimal representation}\}$$

Not only does this mean that there are no irrational numbers in $\bigcup_{n\in\mathbb{N}} A_n$, but many rational numbers are also ruled out. For example $\frac{1}{3} = 0.3333\ldots$ is not in any set A_n and so is therefore not in the union.

[29] For a little proof practice, prove both directly and by induction that the cardinality of A_m is 10^m.

[30] $1 = 0.9999\ldots$ after all.

Indexed Unions: Confusing Sets and Elements

It is easy to confuse, but important to distinguish between the sets

$$\mathcal{A} = \{A_i : i \in I\} \qquad \text{and} \qquad \cup\mathcal{A} = \bigcup_{i\in I} A_i.$$

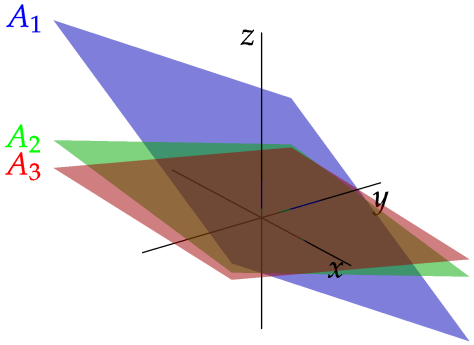
$\mathcal{A}$ is a set whose *elements* are themselves sets. The second is the collection of all elements in *any* set A_i. Consider the following examples.

Examples. 1. For each $n \in \{1,2,3\}$, let A_n be the plane $\{(x,y,z) : x + ny + n^2z = 1\} \subseteq \mathbb{R}^3$.
$\mathcal{A} = \{A_1, A_2, A_3\}$ has *three* elements: each of the planes A_1, A_2, A_3 is an object in its own right.
The union $\cup\mathcal{A} = A_1 \cup A_2 \cup A_3$ is an *infinite* set consisting of all the *points* on the three planes.
For the intersection, a little work with simultaneous equations should convince you that

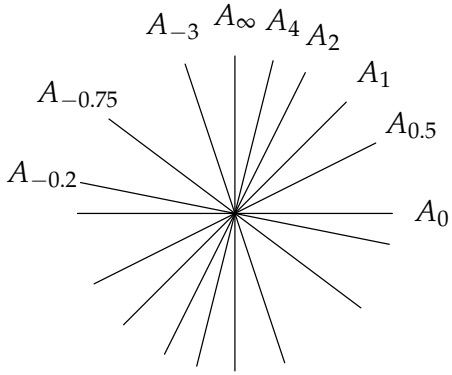
$$(x,y,z) \in \bigcap_{n\in\{1,2,3\}} A_n \iff \begin{cases} x+y+z=1 \\ x+2y+4z=1 \\ x+3y+9z=1 \end{cases} \iff (x,y,z) = (1,0,0).$$

The planes are drawn below.

2. For each $m \in \mathbb{R} \cup \{\infty\}$, let A_m be the line[31] through the origin in $\mathbb{R}^2$ with gradient m.
Each element of $\mathcal{A}$ is a *line*: there is one for each possible direction through the origin.
$\cup\mathcal{A}$ is all of the *points* that lie on *any* line through the origin. Since every point can be joined to the origin with a straight line, we have $\cup\mathcal{A} = \mathbb{R}^2$ being all the points in the plane.
It should be clear that all the lines intersect at the origin, and so $\cap\mathcal{A} = \{(0,0)\}$.
The collection of lines $\mathcal{A}$ is the famous *projective space* $\mathbb{P}(\mathbb{R}^2) = \{A_m : m \in \mathbb{R} \cup \{\infty\}\}$, certainly a different set from $\mathbb{R}^2$! This example also shows that indexing sets don't have to be simple sets of numbers.



Ex 1: Three elements, or an infinite number?



Ex 2: Elements in $\mathbb{P}(\mathbb{R}^2)$

The Cantor Set

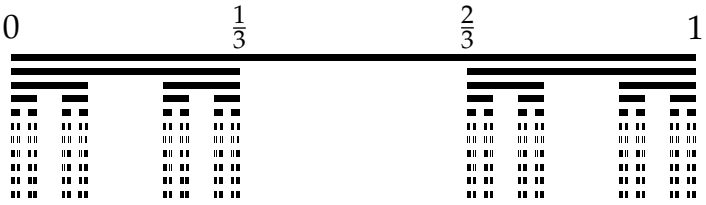
For a bit of fun, we can use infinite intersections to create self-similar sets, or *fractals*. Here is a famous example: the *Cantor middle-third set*.

[31] We include the vertical line A_∞.

Construct a sequence of sets C_n for $n \in \mathbb{N}_0$ by repeatedly removing the middle third of each of the intervals at each step, starting with $[0, 1]$.

$C_0 = [0, 1]$,
$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$,
$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, etc.

The sequence is drawn up to C_{10}.



Define the *Cantor set* $\mathcal{C}$ to be the infinte intersection $\mathcal{C} = \bigcap_{n \in \mathbb{N}_0} C_n$. This set has several interesting properties.

Zero Measure (length): Intuitively, the *length* of a set of real numbers is the sum of the lengths of all the intervals contained in the set. Since we remove $\frac{1}{3}$ of the length of C_n to obtain C_{n+1}, it should be clear that

$$\text{length}(C_n) = \left(\frac{2}{3}\right)^n.$$

As $n \to \infty$ this goes to zero, so the Cantor set contains no intervals: it is purely made up of individual points. This at least seems reasonable from the picture.

Non-emptiness: The Cantor set contains the endpoints of every interval removed at any stage of its construction. In particular, $\frac{1}{3^n} \in \mathcal{C}$ for all $n \in \mathbb{N}_0$. $\mathcal{C}$ is therefore an infinite set. Indeed it is more than just infinite, it is *uncountably* so, as we shall see later.

Self-similarity: If $\frac{\mathcal{C}}{3}$ means 'take all the numbers in $\mathcal{C}$ and divide them by three', then

$$\mathcal{C} = \frac{\mathcal{C}}{3} \cup \left(\frac{2}{3} + \frac{\mathcal{C}}{3}\right). \qquad (*)$$

Otherwise said, $\mathcal{C}$ is made up of two shrunken copies of itself, a classic property of fractals.

To get further with the Cantor set, it is necessary to understand exactly what the elements of the set are. This can be accomplished using the *ternary representation.* It can be shown that every number $x \in [0, 1]$ may be written in the form[32]

$$x = [0.a_1a_2a_3\cdots]_3 = \sum_{n=1}^{\infty} a_n \cdot 3^{-n} = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \cdots$$

where each $a_n \in \{0, 1, 2\}$. For example:

$$[0.12]_3 = \frac{1}{3} + \frac{2}{3^2} = \frac{5}{9}, \qquad \frac{64}{243} = \frac{2}{3^2} + \frac{1}{3^3} + \frac{1}{3^5} = [0.02101]_3, \qquad 1 = [0.22222\cdots]_3.$$

For this last, use the formula for the sum of a geometric series to calculate $\sum_{n=1}^{\infty} 2\left(\frac{1}{3}\right)^n = 2 \cdot \frac{1/3}{1-1/3} = 1$
The only possibility whereby x can have two ternary expansions is if one of them terminates. The

[32] Analogous to a decimal representation $x = \sum_{n=1}^{\infty} a_n \cdot 10^{-n} = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots$ for some $a_n \in \{0, 1, 2, \ldots, 9\}$.

other becomes a sequence of repeating 2's. For example:[33]

$$[0.0222222\cdots]_3 = [0.1]_3 = \frac{1}{3} \quad \text{and} \quad [0.10122222\cdots]_3 = [0.102]_3 = \frac{1}{3} + \frac{2}{27} = \frac{11}{27}.$$

Theorem 6.9. *C_n is the set of all numbers $x \in [0,1]$ with a ternary expansion whose first n digits are only 0 or 2. It follows that $\mathcal{C}$ is exactly the set of $x \in [0,1]$ with a ternary expansion containing only 0 and 2.*

Proof. We prove by induction.
Initial case: The proposition is clearly true for $C_0 = [0,1]$, as there is nothing to check.
Induction step: Assume that the proposition is true for some fixed n. Analogously to $(*)$ above, observe that

$$C_{n+1} = \left\{\frac{1}{3}x : x \in C_n\right\} \cup \left\{\frac{1}{3}x + \frac{2}{3} : x \in C_n\right\} = \frac{1}{3}C_n \cup \left(\frac{1}{3}C_n + \frac{2}{3}\right).$$

Multiplication by $\frac{1}{3}$ shifts a ternary representation one position to the right.[34]
Addition of $\frac{2}{3}$ adds $[0.2]_3$ to the representation: inserts 2 in the (now empty) first ternary place.
Thus if C_n contains only 0's and 2's in its first n entries, C_{n+1} contains only 0's and 2's in its first $n+1$ entries.
By induction the proposition is true for all $n \in \mathbb{N}$. ∎

As the Theorem shows, the Cantor set contains a lot of elements. For example:

$$[0.020202020\cdots]_3 = 2\sum_{n=1}^{\infty} 3^{-2n} = \frac{2/9}{1-\frac{1}{9}} = \frac{1}{4} \in \mathcal{C}.$$

What is strange is that $\frac{1}{4}$ is not the endpoint of any one of the open intervals deleted during the construction of $\mathcal{C}$.

Generalizations and related concepts include Cantor dust $\mathcal{C} \times \mathcal{C}$, the Sierpiński carpet and gasket, and the von Koch snowflake.

7 Relations and Partitions

The mathematics of sets is rather basic, at least until one has a notion of how to relate elements of sets with each other. We are already familiar with examples of this:

1. The usual *order* of the natural numbers (e.g. $3 < 7$) is a way of relating/comparing two elements of $\mathbb{N}$. Recall that, as sets, order doesn't matter: $\{3,7\} = \{7,3\}$. As *ordered pairs* however, $(3,7) \neq (7,3)$.

2. A *function* $f : A \to B$ relates elements in the set A with those in B.

It turns out that the concept of an ordered pair is essential to relating elements.

[33] This is ticklish to prove, as is the corresponding result for decimals: consider $1 = 0.99999999\cdots$.
[34] Compare multiplication by $\frac{1}{10}$ and a decimal…

7.1 Relations

Definition 7.1. Let A and B be sets. A *(binary) relation* R from A to B is a set of ordered pairs

$$R \subseteq A \times B.$$

A relation on A is a relation from A to itself.
If $(x, y) \in R$ we can also write $x \, R \, y$. Similarly $x \not R y$ means $(x, y) \notin R$.

Examples. 1. $R = \{(1,3), (2,2), (2,3), (3,2), (4,1), (5,2)\}$ is a relation from $\mathbb{N}$ to $\mathbb{N}$.

2. $R = [1,3) \times (3,4] \cup \{(2t+1, t^2) : t \in [0.5, 2]\}$ is a relation from $\mathbb{R}$ to $\mathbb{R}$. Be careful: it is easy to confuse interval notation with the notation for ordered pair!

3. Equality is a relation on any set A: i.e. $(x, y) \in R \iff x = y$ defines a relation.

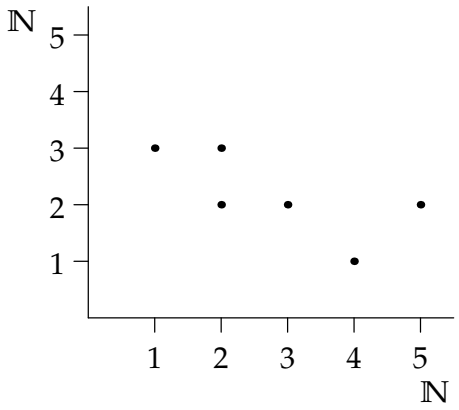
4. If $A = \{\text{all humans}\}$, we may define $R \subseteq A \times A$ by

   $$(a_1, a_2) \in R \iff a_1, a_2 \text{ have a parent-child, or sibling relationship.}$$

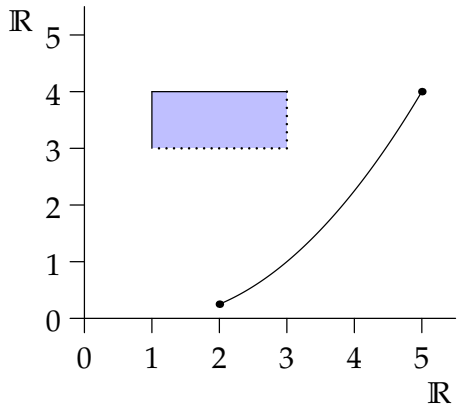
   In this sense, the mathematical use of the word relation is identical to that in English. I am related (both ways) to my sister.

5. If A is a set, then $\subseteq$ is a relation on the power set $\mathcal{P}(A)$.

When R is a relation between sets of numbers, we can often graph the relation. Examples 1 and 2 above would be graphed as follows:



Example 1.



Example 2.

Not all relations between sets of numbers can be graphed: for instance, how would you represent $\mathbb{Q} \times \mathbb{Q}$?

Definition 7.2. If $R \subseteq A \times B$ is a relation, then its *inverse* $R^{-1} \subseteq B \times A$ is the set

$$R^{-1} = \{(y, x) \in B \times A : (x, y) \in R\}.$$

Suppose $A = B$. We say that R is *symmetric* if $R = R^{-1}$.

The following results should seem natural, even if some of the proofs may not be obvious.

Theorem 7.3. *Given any relations $R, S \subseteq A \times A$:*

1. $(R^{-1})^{-1} = R$
2. $R \subseteq S \iff R^{-1} \subseteq S^{-1}$
3. $(R \cup S)^{-1} = R^{-1} \cup S^{-1}$
4. $(R \cap S)^{-1} = R^{-1} \cap S^{-1}$
5. $R \cup R^{-1}$ *is symmetric*
6. $R \cap R^{-1}$ *is symmetric*

Proof. Here are two of the arguments. Try the others yourself.

2. Assume that $R \subseteq S$, and suppose that $(x, y) \in R^{-1}$. We must prove that $(x, y) \in S^{-1}$. However,

$$\begin{aligned}(x, y) \in R^{-1} &\implies (y, x) \in R \implies (y, x) \in S \\ &\implies (x, y) \in S^{-1}.\end{aligned}$$

   Thus $R^{-1} \subseteq S^{-1}$. For the converse, use 1. to see that

$$R^{-1} \subseteq S^{-1} \implies (R^{-1})^{-1} \subseteq (S^{-1})^{-1} \implies R \subseteq S.$$

   $R \subseteq S \iff (x, y) \in R \implies (x, y) \in R^{-1} \subseteq S^{-1}$

5. By 3, $(R \cup R^{-1})^{-1} = R^{-1} \cup (R^{-1})^{-1} = R^{-1} \cup R = R \cup R^{-1}$ is symmetric.

■

Be careful! Several of the above results look suspiciously similar to earlier results (e.g. 3 and 4 look almost like de Morgan's laws, except that $\cup$ and $\cap$ do not switch over). This is why it is important to be able to prove and come up with examples of such statements. For example, you might expect that

$$(R \cup S)^{-1} = \begin{cases} R^{-1} \cup S^{-1} & \text{or,} \\ R^{-1} \cap S^{-1}. \end{cases}$$

Now that you have two sensible guesses, you should be able to decide the correct one by thinking about examples and, if necessary, prove it!

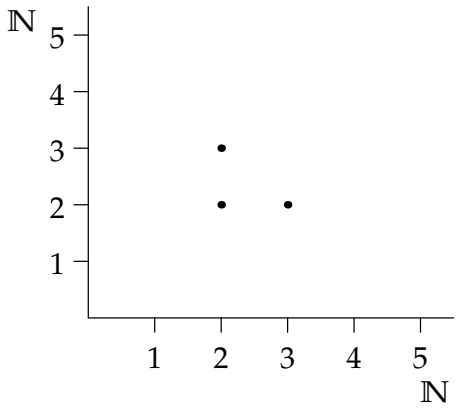
Example. Consider our example $R = \{(1,3), (2,2), (2,3), (3,2), (4,1), (5,2)\}$ from earlier. This is clearly not symmetric since $(1,3) \in R$ but $(3,1) \notin R$. However

$$R^{-1} = \{(3,1), (2,2), (3,2), (2,3), (1,4), (2,5)\},$$

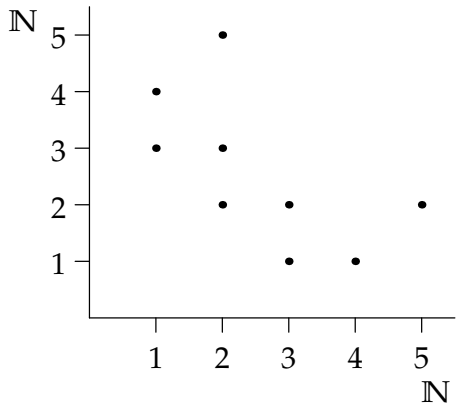
so that

$$\begin{aligned}R \cap R^{-1} &= \{(2,2), (2,3), (3,2)\} \quad \text{and} \\ R \cup R^{-1} &= \{(1,3), (3,1), (2,2), (2,3), (3,2), (4,1), (1,4), (5,2), (2,5)\}\end{aligned}$$

are both symmetric.



The relation $R \cap R^{-1}$



The relation $R \cup R^{-1}$

The above pictures should confirm something: if you can graph a symmetric relation, then the graph has symmetry about the line $y = x$.

7.2 Functions revisited

Now that we have the language of relations, we can properly define functions. If a function $f : A \to B$ is a rule that assigns an element of B to each element of A, then we may view the function as a collection of ordered pairs $\{(a, f(a)) \in A \times B\}$: this set is nothing more than the *graph* of the function, and it is a relation.

Definition 7.4. Let $f \subseteq A \times B$ be a relation from A to B. The *domain* and *range* of f are the sets

$$\text{dom}(f) = \{a \in A : (a, b) \in f \text{ for some } b \in B\},$$
$$\text{range}(f) = \{b \in B : (a, b) \in f \text{ for some } a \in A\}.$$

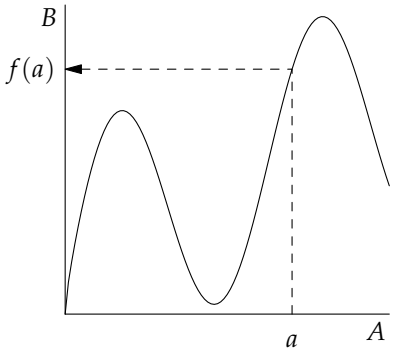
A *function* from A to B is a relation $f \subseteq A \times B$ satisfying the following conditions:

1. $\text{dom}(f) = A$,
2. $(a, b_1), (a, b_2) \in f \implies b_1 = b_2$.

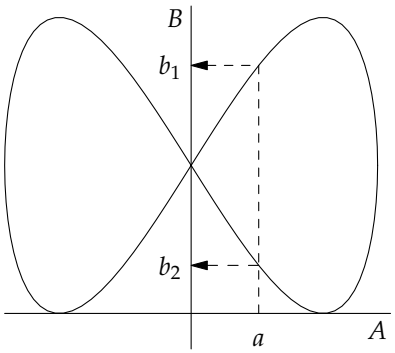
The two conditions can be thought of as saying:

1. Everything in A gets mapped to something in B: i.e. $\forall a \in A,\ f(a) \in B$,
2. Each element of A is mapped to a single element of B.

This second is the vertical line test, familiar from calculus.



$b_1 = b_2 = f(a)$: a function



$b_1 \neq b_2$: not a function

Definition 7.5. The *inverse* of a function $f \subseteq A \times B$ is the inverse relation $f^{-1} \subseteq B \times A$.

In general, the inverse of a function is just a relation and not a function in its own right. Theorem 7.6 will discuss when the inverse is a function.

The inverse of a function is usually written in set notation, thus if $b \in B$ we have the *inverse image of b*

$$f^{-1}(b) = \{a \in A : f(a) = b\},$$

and, more generally, if $V \subseteq B$, the *inverse image of V*

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

Both are *subsets* of A. Only when f^{-1} is a function are we entitled to write $f^{-1}(b) = a$.

Examples. 1. Let $A = B = \{1,2,3\}$ and $f = \{(1,3),(2,1),(3,3)\}$.
Note that $\mathrm{dom}(f) = \{1,2,3\} = A$, and that each element of A appears exactly once as the first element in a pair $(a,b) \in f$. This relation therefore satisfies both conditions necessary to be a function.
We find it easier to write $f(1) = 3$, $f(2) = 1$, and $f(3) = 3$.
The inverse relation $f^{-1} = \{(3,1),(1,2),(3,3)\}$ is not a function, since $(3,1),(3,3) \in f^{-1}$ but $1 \neq 3$.

2. Let $A = B = \mathbb{R}$ and $f = \{(x,x^2) : x \in \mathbb{R}\}$. This is just the function $f(x) = x^2$.
The inverse in this case is not a function:

$$f^{-1} = \{(x^2,x) : x \in \mathbb{R}\} = \{(y,\pm\sqrt{y}) : y \geq 0\},$$

so that, e.g., $f^{-1}(4) = \{-2,2\}$.

3. Let $A = B = \mathbb{R}$ and $f = \{(x,x^3) : x \in \mathbb{R}\}$. This is the function $f(x) = x^3$.
The inverse this time is also a function, $f^{-1}(y) = \sqrt[3]{y}$:

$$f^{-1} = \{(x^3,x) : x \in \mathbb{R}\} = \{(y,\sqrt[3]{y}) : y \in \mathbb{R}\}.$$

4. Let $A = \mathbb{R}$, $B = \mathbb{Q}$ and $f = \{(x,x) : x \in \mathbb{Q}\} \cup \{(x,0) : x \notin \mathbb{Q}\}$. Otherwise said,

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

The inverse f^{-1} is certainly not a function, for example $f^{-1}(0) = \{0\} \cup (\mathbb{R} \setminus \mathbb{Q})$.

The following theorem is very important but it's proof is quick. You should recall the meanings of injective and surjective from section 4.4.

Theorem 7.6. *f^{-1} is a function $\iff$ f is both injective and surjective.*

Proof. f^{-1} is a function $\iff \begin{cases} \mathrm{dom}(f^{-1}) = B, \\ (b,a_1),(b,a_2) \in f^{-1} \implies a_1 = a_2. \end{cases}$
The first of these is equivalent to $\mathrm{range}(f) = B$, i.e. that f is surjective. The second is equivalent to $(a_1,b),(a_2,b) \in f \implies a_1 = a_2$, which is injectivity. ∎

Equality of functions

There are two notions of equality of functions, dependent on what definition you take as fundamental.

1. $f = g$ means that f and g are the same subset of the *same* $A \times B$. This notion is preferred by set theorist because it sticks rigidly to the idea that a function is a *relation*, which requires both A and B to be explicit.
2. $f = g$ means that $f \subseteq A \times B$, $g \subseteq A \times C$, and $(a, b) \in f \iff (a, b) \in g$. This notion considers fundamental the notion of what a function *does*, rather than its strict status as a relation; if two functions do the same thing to the same domain then they are the same. As such, this notion is used more often.

Unfortunately the second notion, while intuitive, has a problem. For example, let

$$f : \mathbb{R} \to \mathbb{R}, \quad \text{and} \quad g : \mathbb{R} \to [-1, 1] \quad \text{satisfy} \quad f(x) = g(x) = \sin x.$$

Applying the first notion, these are *different functions.* Under the second notion, they are the *same.* However, g is surjective while f is not, so don't we want f and g to be different?!

The same problem does not arise when considering domains. For example, if

$$f : \mathbb{R} \to \mathbb{R}, \quad \text{and} \quad g : \mathbb{R} \setminus \{1\} \to \mathbb{R} \quad \text{satisfy} \quad f(x) = g(x) = x^2 + 2,$$

then $f \neq g$, since they have *different domains.*

7.3 Equivalence Relations

In mathematics, the notion of *equality* is not as simple as one might think. Two numbers being equal is straightforward, but suppose we want to consider two paths between given points as equal iff they have the same length? Since two 'equal' paths might look very different, is this a good notion of equality? It is very common for mathematicians to want to group together collections of objects that have some common property and then treat them as if they were a single object. This is done using equivalence relations and equivalence classes.

First recall the alternative notation for relation: if $R \subseteq A \times B$ is a relation, then $x\, R\, y$ has the same meaning as $(x, y) \in R$. We might read $x\, R\, y$ as "x is R-related to y." This notation is more appropriate for equivalence relations, although there is nothing stopping you using the former.

Definition 7.7. A relation R on a set A is:

Reflexive if: $\forall x \in A,\ x\, R\, x$ (every element of A is related to itself)

Symmetric if: $\forall x, y \in A,\ x\, R\, y \implies y\, R\, x$ (if x is related to y then y is related to x)

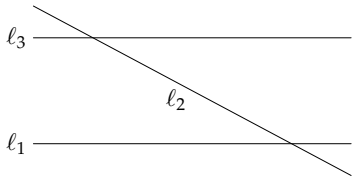
Transitive if: $x\, R\, y$ and $y\, R\, z \implies x\, R\, z$ (if x is related to y, and y to z, then x is related to z)

Symmetry is exactly the same notion at in Definition 7.2.

Examples. 1. Let $A = \mathbb{R}$ and let R be $\leq$. Thus $2 \leq 3$, but $7 \not\leq 4$. We check each of the above notions.

Reflexivity: **True.** $\forall x \in \mathbb{R},\ x \leq x$.
Symmetry: **False.** For example, $2 \leq 3$ but $3 \not\leq 2$.
Transitivity: **True.** If $x \leq y$ and $y \leq z$, then $x \leq z$.

2. Let A be the set of lines in the plane and define $\ell_1 \, R \, \ell_2 \iff \ell_1$ and ℓ_2 intersect.



Reflexivity: **True.** Every line intersects itself, so $\ell \, R \, \ell$ for all $\ell \in A$.
Symmetry: **True.** If ℓ_1 intersects ℓ_2, then ℓ_2 intersects ℓ_1.
Transitivity: **False.** Let ℓ_1 and ℓ_3 be parallel lines, and ℓ_2 cross both of these. Then $\ell_1 \, R \, \ell_2$ and $\ell_2 \, R \, \ell_3$, but $\ell_1 \not R \, \ell_3$.

You should get used to checking each of the above properties whenever you see a relatio, precisely because of the following definition.

Definition 7.8. An *equivalence relation* is a relation $\sim$ which is reflexive, symmetric and transitive.

The symbol $\sim$ is almost universally used for a general equivalence relation. It can be read as 'related to,' 'tilde,' or 'twiddles.' The two examples above are *not* equivalence relations because they fail one of the three conditions. Here is the easiest equivalence relation:

Example. Equals '=' is an equivalence relation on any set, hence the name!

Read the definitions of reflexive, symmetric and transitive until you are certain of this fact. There are countless other equivalence relations: here are a few.

Examples. 1. For all $x, y \in \mathbb{Z}$, let $x \sim y \iff x - y$ is even. We claim that $\sim$ is an equivalence relation on $\mathbb{Z}$.

Reflexivity: $x - x = 0$ is even, hence $x \sim x$.
Symmetry: $x \sim y \implies x - y$ is even $\implies y - x$ is even $\implies y \sim x$.
Transitivity: If $x \sim y$ and $y \sim z$, then $x - y$ and $y - z$ are even. But the sum of two even numbers is even, hence $x - z = (x - y) + (y - z)$ is even, and so $x \sim z$.

2. Let $A = \{\text{all math 13 students}\}$. For all $x, y \in A$, let $x \sim y \iff x$ achieves the same letter-grade as y. Then $\sim$ is an equivalence relation on A.

Reflexivity: $x \sim x$ since everyone scores the same as themself!
Symmetry: $x \sim y \implies x$ achieves the same letter-grade as y
$\implies y$ achieves the same letter-grade as x
$\implies y \sim x$
Transitivity: If $x \sim y$ and $y \sim z$, then x achieves the same as y who achieves the same as z, whence x achieves the same as z. Thus $x \sim z$.

3. For all $x, y \in \mathbb{Z}$, let $x \sim y \iff x^2 \equiv y^2 \pmod 5$. Then $\sim$ is an equivalence relation on $\mathbb{Z}$.

Reflexivity: $x \sim x$ since x^2 is always congruent to itself!
Symmetry: $x \sim y \implies x^2 \equiv y^2 \pmod 5$
$\implies y^2 \equiv x^2 \pmod 5$
$\implies y \sim x$

Transitivity: If $x \sim y$ and $y \sim z$, then $x^2 \equiv y^2$ and $y^2 \equiv z^2 \pmod 5$. But then $x^2 \equiv z^2 \pmod 5$ and so $x \sim z$.

The proofs of the three properties are often very simple.

The important observation with these examples is that an equivalence relation separates elements of a set into subsets where elements share a common property (e.g., even/oddness, or letter-grade). The next definition formalizes this idea.

Definition 7.9. Let $\sim$ be an equivalence relation on X. The *equivalence class* of x is the set

$$[x] = \{y \in X : x \sim y\}.$$

$X/_\sim$ is the set of all equivalence classes: the *quotient* of X by $\sim$, or 'X mod $\sim$.'

Returning to our examples:

Examples.

1. $[0] = \{y \in \mathbb{Z} : 0 \sim y\} = \{y \in \mathbb{Z} : y \text{ is even}\}$ is the set of even numbers. This is also equal to $[2] = [4] = [6] = \cdots$.
   The other equivalence class is $[1] = \{y \in \mathbb{Z} : 1 - y \text{ is even}\}$, which is the set of odd numbers.
   The quotient set is $\mathbb{Z}/_\sim = \big\{[0], [1]\big\} = \big\{\{\text{even numbers}\}, \{\text{odd numbers}\}\big\}$.

2. There is one equivalence class for each letter grade awarded. Each class contains all the students awarded the same letter-grade. If we call the equivalence classes $\mathrm{A}^+, \mathrm{A}, \mathrm{A}^-, \mathrm{B}^+, \ldots, \mathrm{F}$, where, say, $\mathrm{B} = \{\text{students obtaining a B-grade}\}$, then

   $$\{\text{Students}\}/_\sim = \{\mathrm{A}^+, \mathrm{A}, \mathrm{A}^-, \mathrm{B}^+, \ldots, \mathrm{F}\}.$$

3. The equivalence classes for this example are a little tricky. First observe that $x \equiv y \implies x^2 \equiv y^2 \pmod 5$, so we only need to consider the equivalence classes of $0, 1, 2, 3$ and 4. If we square each of these, modulo 5, we obtain

x	0	1	2	3	4
$x^2 \pmod 5$	0	1	4	4	1

   Notice that $1 \sim 4$, so they share an equivalence class. It follows that the distinct equivalence classes are

   $$[0] = \{x \in \mathbb{Z} : x \equiv 0 \pmod 5\}$$
   $$[1] = \{x \in \mathbb{Z} : x \equiv 1, 4 \pmod 5\}$$
   $$[2] = \{x \in \mathbb{Z} : x \equiv 2, 3 \pmod 5\}$$

7.4 Partitions

Consider the above examples. Notice that every element of the original set X ends up in exactly one equivalence class: for instance, every integer is either even or odd but not both. The equivalence classes *partition* X in the same way that cutting a cake partitions the crumbs: each crumb ends up in one slice.

Here is another example: Partition the set $X = \{1,2,3,4,5\}$ into subsets $A_1 = \{1,3\}$, $A_2 = \{2,4\}$ and $A_3 = \{5\}$ and consider the relation

$$R = \{(1,1),(1,3),(3,1),(3,3),(2,2),(2,4),(4,2),(4,4),(5,5)\}.$$

What does R have to do with the partition? R was constructed by insisting that

$$x\,R\,y \iff x, y \text{ are in the } \textit{same} \text{ subset } A_i.$$

With a little checking, you should be able to confirm that R is an equivalence relation. Moreover, the equivalence classes of R are exactly the sets A_1, A_2, A_3.

This is an example of a general result that we shall state shortly: *partitions and equivalence relations are essentially the same thing*. To get further, we need to properly define partitions.

Definition 7.10. Let X be a set and $\mathcal{A} = \{A_i : i \in I\}$ a collection of subsets $A_i \subseteq X$. Then X is *partitioned by* $\mathcal{A}$ if

1. $A_i \cap A_j = \emptyset$, $\forall i \neq j$. (the A_i are pairwise disjoint[35])
2. $X = \bigcup_{i \in I} A_i$. (the A_i together make up X)

Examples. 1. The integers can be partitioned dependent on their remainder modulo 3:

$$\mathbb{Z} = A_0 \cup A_1 \cup A_2 \quad \text{where} \quad A_i = \{n \in Z : n \equiv i \pmod 3\}.$$

2. More generally, if $n \in \mathbb{N}$, then the set of integers $\mathbb{Z}$ is partitioned into n sets $A_0, \ldots, A_{n-1}$ where $A_i = \{x \in \mathbb{Z} : x \equiv i \pmod n\}$ is the set of integers with remainder i upon dividing by n.
3. The real numbers $\mathbb{R}$ are partitioned by the sets of rational and irrational numbers.

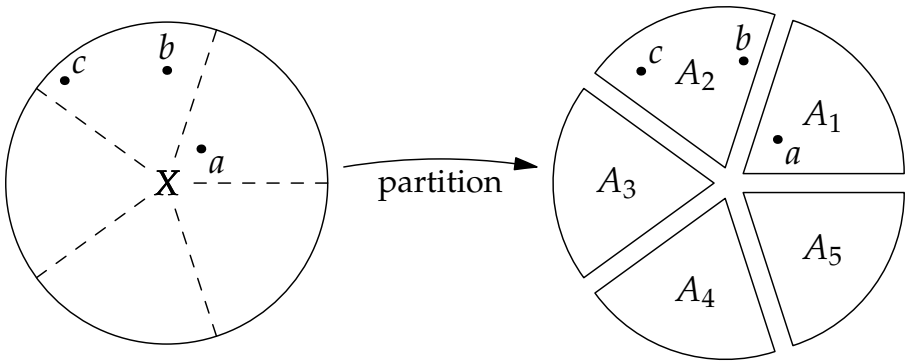
The fact that partitions and equivalence relations are intimately related is made clear by the following important result.

Theorem 7.11. *Let X be a set.*

1. *If $\sim$ is an equivalence relation on X, then X is partitioned by the equivalence classes of $\sim$.*
2. *If $\{A_i\}_{i \in I}$ is a partition of X, then the relation $\sim$ on X defined by*

   $$x \sim y \iff \exists A_i \text{ such that } x, y \in A_i$$

   is an equivalence relation.



[35] Recall that two sets A, B are *disjoint* if $A \cap B = \emptyset$: see Definition 4.5.

Each element of X ends up in exactly one subset. In the language of the Theorem, we have

$$A_1 = [a], \quad A_2 = [b] = [c], \quad b \sim c, \quad a \nsim b, \quad a \nsim c.$$

Two things are important when reading this proof:

- Keep your eyes on the picture: it's where your intuition comes from, and it's how you should remember the result. The algebra merely confirms that the picture is telling a legitimate story.
- Look for where the assumptions about $\sim$ are used, and think why the theorem fails if $\sim$ is not reflexive, symmetric or transitive.

Proof. 1. Assume that $\sim$ is an equivalence relation on X, and recall that $x \sim y \iff y \in [x]$. We must show that every element of X is in some equivalence class and that the equivalence classes of $\sim$ are pairwise disjoint.
For the first, observe that reflexivity gives us $x \sim x \implies x \in [x]$, so that every element of X is in the equivalence class defined by itself.
Now suppose that $[x] \cap [y] \neq \emptyset$. Then there is some element $z \in [x] \cap [y]$. But then,

$$\begin{aligned} x \sim z \text{ and } y \sim z &\implies x \sim z \text{ and } z \sim y && \text{(Symmetry)} \\ &\implies x \sim y && \text{(Transitivity)} \end{aligned}$$

Now let $\hat{y} \in [y]$. Then

$$x \sim y \text{ and } y \sim \hat{y} \implies x \sim \hat{y} \implies \hat{y} \in [x]. \qquad \text{(Transitivity)}$$

Therefore $[y] \subseteq [x]$. The argument is entirely symmetric, so we also have $[x] \subseteq [y]$. Hence $[x] \cap [y] \neq \emptyset \iff [x] = [y]$, the equivalence classes are pairwise disjoint, and we've proved 1.

2. Now suppose that $\{A_i\}_{i \in I}$ is a partition and that $x \sim y \iff \exists A_i$ such that $x, y \in A_i$. We must prove the reflexivity, symmetry and transitivity of $\sim$.

   Reflexivity: Every $x \in X$ is in some A_i. Thus $x \sim x$ for all $x \in X$.
   Symmetry: If $x \sim y$, then $\exists A_i$ such that $x, y \in A_i$. But then $y, x \in A_i$ and so $y \sim x$.
   Transitivity: Let $x \sim y$ and $y \sim x$. Then $\exists A_j, A_k$ such that $x, y \in A_j$ and $y, z \in A_k$. Since $\{A_i\}_{i \in I}$ is a partition and $y \in A_j \cap A_k$, we have $j = k$. Thus $x, z \in A_j$ and so $x \sim z$.

   Thus $\sim$ is an equivalence relation. ∎

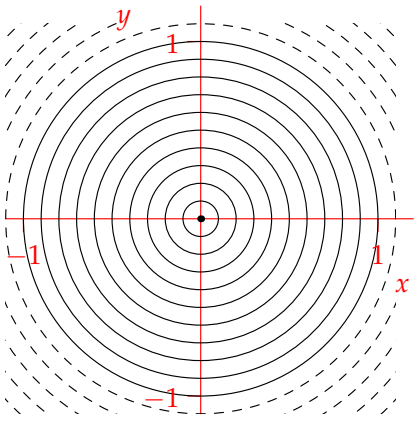
Examples of partitions are especially easy to see with curves in the plane.

Example. For each real number $r \geq 0$, define the set

$$A_r = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = r^2\}.$$

This is simply the circle of radius r centered at the origin. We check that $\{A_r\}_{r \in \mathbb{R}_0^+}$ is a partition of $\mathbb{R}^2$.

- $(x, y) \in A_{\sqrt{x^2+y^2}}$ since $\sqrt{x^2 + y^2}$ is the distance of (x, y) from the origin. Thus $\mathbb{R}^2 = \bigcup_{r \in \mathbb{R}_0^+} A_r$.
- If $r_1 \neq r_2$, then the circles A_{r_1} and A_{r_2}, being concentric, do not intersect.



Now define an equivalence relation $\sim$ on $\mathbb{R}^2$ via $(x,y) \sim (v,w) \iff x^2+y^2 = v^2+w^2$. The equivalence classes of this relation are exactly the sets A_r. Indeed

$$[(v,w)] = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 = v^2+w^2\} = A_{\sqrt{v^2+w^2}}$$

is just the circle of radius $\sqrt{v^2+w^2}$.

Geometric Examples

The language of equivalence relations and partitions is used heavily in geometry and topology to describe complex shapes. Here are a few examples.

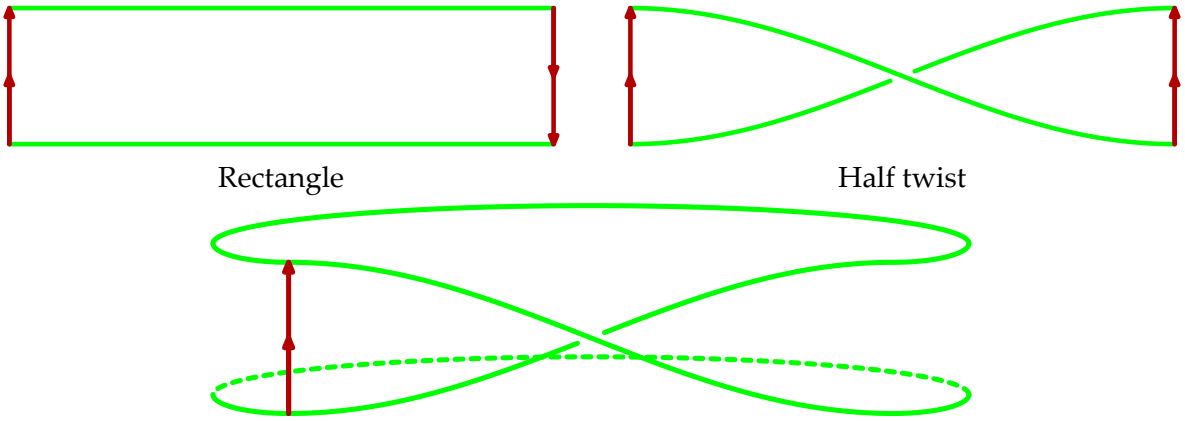
The Möbius Strip Take a rectangle $X = [0,4] \times [0,1]$ and partition into the following subsets.

- If a point does not lie on the left or right edge of the rectangle, place it in a subset by itself: $\{(x,y)\}$ for $x \neq 0,4$,
- If a point does lie on the left or right edge of the rectangle, place it in a subset with one point from the other edge: $\{(0,y),(4,1-y)\}$ for any y.

The rectangle is drawn below, where the points on the left and right edges are colored red. The arrows indicate how the edges are paired up. For example $(0,0.8)$ (high on the left) is paired with $(4,0.2)$ (low on the right).

These subsets clearly partition the rectangle X. The partitions define an equivalence relation $\sim$ on X in accordance with Theorem 7.11. How can we interpret the set $X/\sim$ of euivalence classes?

This is easier to visualize than to state explicitly: the left and right edges are identified, but in the opposite directions, so we imagine holding X like a strip of paper, giving one side a twist, then gluing the two ends together. This is the classic construction of a Möbius strip.



Rectangle Half twist

Glue arrows to get Möbius strip

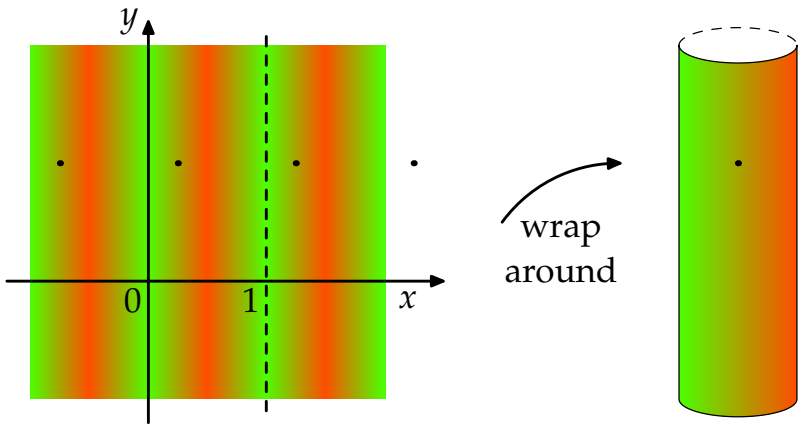
The Cylinder One could construct a cylinder similarly to the Möbius strip, simply by identifying edges of the rectangle *without* applying the half-twist. Instead we do something a little different.

Let $X = \mathbb{R}^2$ with equivalence relation $\sim$ defined by

$$(a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \textbf{ and } b = d.$$

The equivalence classes are horizontal strings of points with the same y co-ordinate. If we imagine wrapping $\mathbb{R}^2$ repeatedly around a cylinder of circumference 1, all of the points in a given equivalence class will now line up. The set of equivalence classes $\mathbb{R}^2/_\sim$ can therefore be vizualized as the cylinder.

Alternatively, you may imagine piercing a roll of toilet paper and unrolling it. The single pucture now becomes a row of holes (almost[36]) equally spaced.



More complex shapes can be created by other partitions/realtions. If you want a challenge in visualization, consider why the equivalence relation

$$(a,b) \sim (c,d) \iff a-c \in \mathbb{Z} \textbf{ and } b-d \in \mathbb{Z}$$

on $\mathbb{R}^2$ defines a torus (doughnut with a hole).

7.5 Well-definition, Rings and Congruence

Now we return to our discussion of congruence (recall Section 3.1). The important observation (Example 2, page 63 in conjunction with Theorem 7.11) is that congruence modulo n is an equivalence relation on $\mathbb{Z}$, each equivalence class being the set of all integers sharing a remainder modulo n. Explicitly,

$$\begin{aligned} a \equiv b \pmod{n} &\iff a-b \text{ is divisible by } n \\ &\iff [a]=[b]=\{x \in \mathbb{Z} : x-a \text{ is divisible by } n\} \end{aligned}$$

If the meaning of the above notation is not clear, go back and read the previous two sections—they are critically important!

In Section 3.1 we gave a naïve definition of the ring $\mathbb{Z}_n$. Here we give the correct one. The set of equivalence classes of $\mathbb{Z}$ modulo n is the quotient[37]

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

This set is what will become $\mathbb{Z}_n$. To do so we need to define addition and multiplication of the objects in $\mathbb{Z}/_\sim$. That is, we want to understand how to add and multiply *equivalence classes.*

[36] Unfortunately for the analogy, toilet paper has purposeful thickness!

[37] The notation here is complpicated. If you prefer, write $a \sim b$ for $a \equiv b \pmod{n}$, then the quotient is $\mathbb{Z}/_\sim$.

Define operations $+_n, \cdot_n$ on $\mathbb{Z}/_{\sim}$ by

$$[x] +_n [y] := [x + y], \qquad [x] \cdot_n [y] := [x \cdot y].$$

Note that $+_n$ is not the same operation as $+$: we are *defining* $+_n$ using $+$. The former is adding equivalence classes, while the latter sums integers.

There is something to prove before we can start using these operations. Consider the process for computing $[x] +_n [y]$:

1. *Choose* $x \in [x]$ and $y \in [y]$.
2. Add $x + y$.
3. Take the equivalence class of $[x + y]$.

There are many (indeed an infinity of) elements in each equivalence class $[x]$. If our definition is to make sense, the process 1,2,3 must result in the *same* equivalence class $[x + y]$ *regardless* of the choice of representative elements we made in step 1. Given that we have already defined the operations $+_n, \cdot_n$, the process of checking that the definition is independent of choice is known as checking *well-definition*.

Theorem 7.12. *The operations $+_n, \cdot_n$ are well-defined.*

Proof. All elements of $[x]$ and $[y]$ have the form $x + kn$ and $y + ln$ for some $k, l \in \mathbb{Z}$. It therefore suffices to check that

$$\forall k, l \in \mathbb{Z}, \quad (x + kn) + (y + ln) \in [x + y] \quad \text{and} \quad (x + kn)(y + ln) \in [xy].$$

Indeed $(x + kn) + (y + ln) = x + y + (k + l)n \equiv x + y \pmod{n}$ gives the first result, and the second is similar. You should re-read Theorem 3.8 until you are comfortable that this is merely a corollary of that result. ∎

Definition 7.13. The *ring* $\mathbb{Z}_n$ is the quotient set $\mathbb{Z}/_{\equiv} = \{[0], [1], \ldots, [n-1]\}$ together with the operations $+_n, \cdot_n$ defined above.

Given the usefulness of $\mathbb{Z}_n$, and the cumbersome nature of the above notation, it is customary to drop the square brackets and subscripts and simply write

$$\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}, \qquad x + y := x + y \mod n, \qquad x \cdot y := xy \mod n.$$

When using this description of $\mathbb{Z}_n$, you should realize that we are working with equivalence classes, not numbers. It might appear that $-3 \in \mathbb{Z}_8$ makes no sense, but we really mean $[-3] \in \mathbb{Z}_8$. This is percectly fine, since $[-3] = [5]$ in $\mathbb{Z}_8$.

7.6 Functions and Partitions

To complete our discussion of partitions and equivalence relations, we consider how to define functions on equivalence classes. Take congruence as our motivating example.

Suppose we want to define a function $f : \mathbb{Z}_4 \to \mathbb{Z}_6$. Say $f(x) = 3x \pmod 6$. This certainly looks like a function, but is it? Remember that 'x' and '$3x$' here are really equivalence classes, so we should say

$$f([x]) = [3x].$$

Is *this* a function? To make sure, we need to check that *any* representative $a \in [x]$ gives the same result. I.e. we need to prove that

$$a \equiv b \pmod 4 \implies 3a \equiv 3b \pmod 6.$$

This is not so hard:

$$\begin{aligned} a \equiv b \pmod 4 &\implies \exists n \in \mathbb{Z} \text{ such that } a = b + 4n \\ &\implies 3a = 3b + 12n \implies 3a \equiv 3b \pmod 6. \end{aligned}$$

It might look like a small difference, but attempting to define $g : \mathbb{Z}_4 \to \mathbb{Z}_6$ by $g(x) = 2x \pmod 6$ is *not* a function. If it were, then we must have

$$a \equiv b \pmod 4 \implies 2a \equiv 2b \pmod 6,$$

which is simply not true: for example $4 \equiv 0 \pmod 4$, but $8 \not\equiv 0 \pmod 6$.

Just as in the previous section, the process of verifying that a rule really is a function is called checking *well-definition.* In general, the problem is this: if we are defining a function whose domain is a quotient, i.e.

$$f : {}^{X}\!/_{\sim} \to A,$$

it is often necessary to construct f by saying what happens to a representative of an equivalence class:

$$f([x]) = \text{ 'do something to } x\text{'}.$$

We need to make sure that the 'something' is *independent of the choice of element* x.

Definition 7.14. Suppose that $f : {}^{X}\!/_{\sim} \to A$ is a rule. We say that f is *well-defined* if

$$[x] = [y] \implies f([x]) = f[(y)].$$

Examples. 1. Show that $f : \mathbb{Z}_n \to \mathbb{Z}_n$ defined by $f(x) = x^2 + 4 \pmod n$ is well-defined.
We must check that $x \equiv y \pmod n \implies x^2 + 4 \equiv y^2 \equiv 4 \pmod n$. But this is trivial!

2. For what integers k is the rule $f : \mathbb{Z}_4 \to \mathbb{Z}_6$ defined by $f(x) = kx \pmod 6$ a well-defined function?
We require $x \equiv y \pmod 4 \implies kx \equiv ky \pmod 6$. Now

$$x \equiv y \pmod 4 \implies \exists n \in \mathbb{Z} \text{ such that } x - y = 4n \implies kx - ky = 4kn.$$

Since things must be as independent as possible of the choices of x, y, we conclude that f is well-defined iff $6 \mid 4kn$ *for all* $n \in \mathbb{Z}$. Otherwise said,

$$f \text{ is well-defined} \iff 6 \mid 4k \iff 3 \mid 2k \iff 3 \mid k.$$

Since equivalent k modulo 6 won't change f, it follows that there are only *two* such functions $f : \mathbb{Z}_4 \to \mathbb{Z}_6$, namely $f_0(x) = 0$ and $f_3(x) = 3x$. In tabluar form, here they are:

x	0	1	2	3
$f_0(x)$	0	0	0	0

x	0	1	2	3
$f_3(x)$	0	3	0	3

It is instructive to play with another value of k, say $k = 5$, and attempt to construct a table:

x	0	1	2	3	4	5	$\cdots$
$f_5(x)$	0	5	4	3	2	1	$\cdots$

The problem is that $4 \equiv 0 \pmod 4$, yet $f_5(4) \not\equiv f_5(0) \pmod 6$. In order to be a function, the second row must repeat with period four.

Functions on the Cylinder and Torus

Recall our construction of the cylinder (page 65), where we viewed the cylinder as the set $\mathbb{R}^2/_\sim$ where $(a,b) \sim (c,d) \iff a - c \in \mathbb{Z}$ and $b = d$. In order to define a function f whose domain is the cylinder, *well-definition* will require us to check that our formula satisfies

$$(a,b) \sim (c,d) \implies f\big([(a,b)]\big) = f\big([(c,d)]\big).$$

The condition $a - c \in \mathbb{Z}$ forces f to be periodic in x with period 1: thus

$$f\big([(x,y)]\big) = y^2 \sin(2\pi x)$$

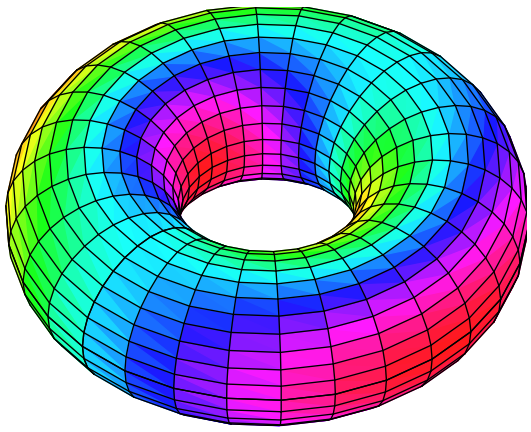
would be a suitable choice of function $f : \mathbb{R}^2/_\sim \to \mathbb{R}$.
More generally, to define a function whose domain is the torus

$$T^2 = \mathbb{R}^2/_\sim \quad \text{where} \quad (a,b) \sim (c,d) \iff a - c \in \mathbb{Z} \text{ and } b - d \in \mathbb{Z},$$

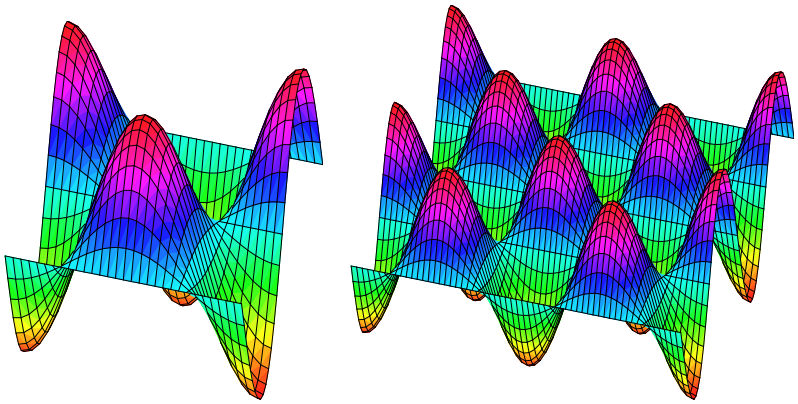
requires a function which has periodic 1 in *both* x and y. The function $f\big([(x,y)]\big) = \sin(2\pi x)\cos(2\pi y)$ is plotted below, with the color on the torus indicating the value of f. It is easier for us to simply consider the function

$$F : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto \sin(2\pi x)\cos(2\pi y).$$

This is also plotted, where the z-axis corresponds (with color) to the value. To see the periodicity, F is plotted on the fundamental domain $[0,1] \times [0,1]$ and on a larger set.



The function f: domain T^2



The function F: domains $0 \leq x, y \leq 1$ and $0 \leq x, y \leq 2$ respectively

The Canonical Map

To do this justice, and to give you a taste for the details which are necessary in pure mathematics, here is the important definition.

Definition 7.15. Suppose that $\sim$ is an equivalence relation on a set X. The function $\gamma : X \to {}^{X}/_{\sim}$ defined by $\gamma(x) = [x]$ is the *canonical map*.[38]

The canonical map has only one purpose; to allow us to construct functions $f : {}^{X}/_{\sim} \to A$.

Theorem 7.16. *Suppose that $\sim$ is an equivalence relation on X.*

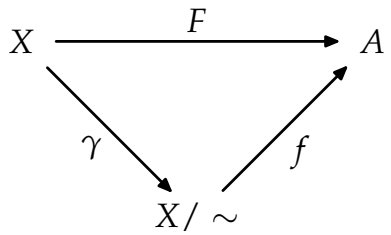
1. *If $f : {}^{X}/_{\sim} \to A$ is a function, then $F : X \to A$ defined by $F = f \circ \gamma$ satisfies $x \sim y \implies F(x) = F(y)$.*
2. *If $F : X \to A$ satisfies $x \sim y \implies F(x) = F(y)$, then there is a unique function $f : {}^{X}/_{\sim} \to A$ satisfiying $F = f \circ \gamma$.*

Proof. 1. This is trivial: $x \sim y \implies [x] = [y] \implies \gamma(x) = \gamma(y)$
$\implies f(\gamma(x)) = f(\gamma(y)) \implies F(x) = F(y)$.

2. *Define* $f : {}^{X}/_{\sim} \to A$ by $f([x]) = F(x)$. We must show that this is well-defined:
$[x] = [y] \implies x \sim y \implies F(x) = F(y) \implies f([x]) = f([y])$.

■

The proof, like much of mathematics, is a masterpiece in concision that seems to be doing nothing at all. The following picture should clear things up.
Defining $f : {}^{X}/_{\sim} \to A$ is equivalent to defining $F : X \to A$ such that $F = f \circ \gamma$: that is $x \sim y \implies F(x) = F(y)$.



This result will be resurrected when you study Groups Rings & Fields as part of the famous *First Isomorphism Theorem.*

[38] Canonical, in mathematics, just means natural or obvious.

8 Cardinalities of Infinite Sets

8.1 Cantor's notion of Cardinality

We have already naïvely dealt with cardinalities of finite sets: recall Theorem 4.12. This result essentially forms Cantor's *definition* of cardinality for any set.

Definition 8.1. The *cardinalities* of two sets A, B are denoted $|A|$ and $|B|$. We compare cardinals as follows:

- $|A| = |B| \iff \exists f : A \to B$ bijective.
- $|A| \leq |B| \iff \exists f : A \to B$ injective.

We write $|A| < |B| \iff A \leq B$ and $A \neq B$. That is $\exists f : A \to B$ injective but no bijective $g : A \to B$.

What is important to realize about this definition is that cardinality is merely an abstract *property.* The cardinality of a finite set can be viewed as the number of elements in that set, but for infinite sets cardinality is simply a method whereby two sets can be *compared.* Strictly speaking we have:

Theorem 8.2. *On any collection of sets, the relation $A \sim B \iff |A| = |B|$ is an equivalence relation.*

Definition 8.3. The cardinality of the set of natural numbers $\mathbb{N}$ is denoted $\aleph_0$, read *aleph-nought* or *aleph-null.*[39]

Examples.
1. If the set of positive even numbers is written $2\mathbb{N} = \{2, 4, 6, 8, 10, \ldots\}$, then the function $f : \mathbb{N} \to 2\mathbb{N} : n \mapsto 2n$ is a bijection. It follows that $|2\mathbb{N}| = |\mathbb{N}| = \aleph_0$.
2. Similarly, the function $f(n) = n + 1$ is a bijection $f : \mathbb{N} \to \{n \in \mathbb{N} : n \geq 2\} = \{2, 3, 4, 5, 6, \ldots\}$, whence the latter set also has cardinality $\aleph_0$.

These examples shows one of the strange properties of infinite sets: $2\mathbb{N}$ is a *proper subset* of $\mathbb{N}$, and yet the two sets are in bijective correspondence with one another!

$\aleph_0$ versus ∞

It is tempting to think of $\aleph_0$ and ∞ as being the same thing, but this isn't the case. It might seem so because people often naïvely talk about an 'infinite number' of objects: if there aren't an infinite number of natural numbers, how many are there? The challenge is to appreciate that this question is meaningless! We certainly have the following:

Theorem 8.4. *If A is finite, then $|A| < \aleph_0$.*

Proof. Suppose that $|A| = n$, so that we may write $A = \{a_1, a_2, \ldots, a_n\}$. But then $f : A \to \mathbb{N}$ defined by $f(a_n) = n$ is injective, and so $|A| \leq \aleph_0$.
Now suppose that $|A| = \aleph_0$. Then $\exists g : \mathbb{N} \to A$ bijective. By Dirichlet's pigeonhole principle, two of the values $g(1), \ldots, g(n+1)$ must be equal and so g is not injective. A contradiction. ∎

The Theorem says that $\aleph_0$ is 'larger than all integers.' Should this not be infinity? The problem is twofold:

[39] $\aleph$ is the first letter of the Hebrew alphabet.

1. As we shall see shortly, there are infinite sets with cardinality different to $\aleph_0$: in the naïve sense, there are multiple infinities, so we will need more symbols…

2. The symbol ∞ is mostly used in a limiting sense: for example $\lim_{x\to 1^-} \frac{1}{1-x} = \infty$ is simply a shorthand for the notion that the function $f(x) = \frac{1}{1-x}$ gets unboundedly larger as x approaches 1 from below. We are not saying that anything is 'equal to infinity;' a calculus student will likely not understand when you correct them for writing $f(1) = \infty$.

8.2 Countably Infinite Sets

Definition 8.5. A set A is *countably infinite*[40] or *denumerable* if $|A| = \aleph_0$.

We have already seen that the set of even numbers $2\mathbb{N}$ is denumerable. Here is another.

Theorem 8.6. *The integers $\mathbb{Z}$ are denumerable.*

Proof. We must construct a bijective function $f : \mathbb{N} \to \mathbb{Z}$. By experimenting, you may feel it is enough simply to write down the first few terms of a suitable function:

n	1	2	3	4	5	6	7	8	$\cdots$
$f(n)$	0	1	-1	2	-2	3	-3	4	$\cdots$

With a bit of thinking, it should be obvious what the function is doing, and that it is bijective. For a bit more formality, we can write

$$f(n) = \begin{cases} \frac{1}{2}n & \text{if } n \text{ even,} \\ -\frac{1}{2}(n-1) & \text{if } n \text{ odd.} \end{cases}$$

Now we check that this is bijective:

Injectivity: Suppose $f(m) = f(n)$. Without loss of generality, there are three cases.

m, n **even** Then $\frac{m}{2} = \frac{n}{2} \implies m = n$.

m, n **odd** Then $-\frac{1}{2}(m-1) = -\frac{1}{2}(n-1) \implies m = n$.

m **even,** n **odd** Then $\frac{m}{2} = -\frac{1}{2}(n-1) \implies m + n = 1$. But $m, n \in \mathbb{N}$, so $m + n \geq 2$. A contradiction.

Thus f is injective.

Surjectivity: With a little calculation, you should be able to see that, for any $M \in \mathbb{Z}$,

$$M = \begin{cases} f(2M) & \text{if } M > 0, \\ f(1-2M) & \text{if } M \leq 0. \end{cases}$$

Clearly f is surjective. This is in fact the construction of the inverse function f^{-1}. ∎

[40] Sometimes this is shortened to *countable,* although some authors use countable to mean 'finite or denumerable,' i.e. any A for which $|A| \leq \aleph_0$. Use countably infinite or denumerable to avoid confusion.

As you build up examples, you no longer have to compare denumerable sets directly with $\mathbb{N}$. A set A is denumerable iff $\exists f : A \to B$ bijective where B is *any other* denumerable set. This holds because *bijectivity is transitive* (Theorem 4.15).

Theorem 8.7. *The rational numbers* $\mathbb{Q}$ *are denumerable.*

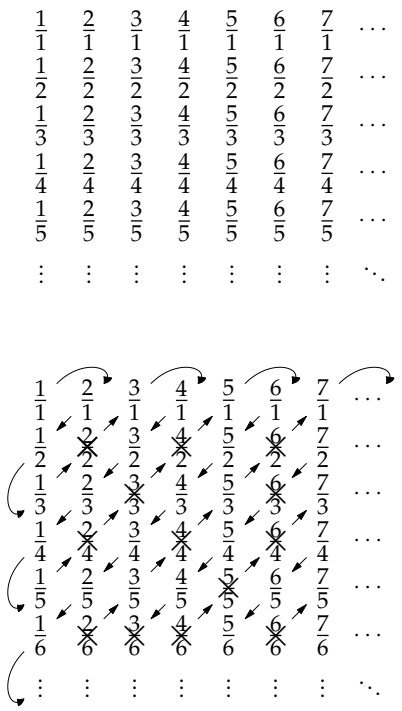
Proof. We do this in stages. First we construct a bijection between the positive rationals $\mathbb{Q}^+$ and the natural numbers $\mathbb{N}$.
List all fractions $\frac{a}{b}$ where $a, b \in \mathbb{N}$ in an infinite square as shown. Clearly every positive rational number appears somewhere in this list (indeed many times). Now define the *ordered set*

$$A = \left\{\tfrac{1}{1}, \tfrac{2}{1}, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{3}{1}, \tfrac{4}{1}, \tfrac{3}{2}, \tfrac{2}{3}, \tfrac{1}{4}, \tfrac{1}{5}, \ldots\right\} = \{a_1, a_2, \ldots\},$$

by tracing diagonals and deleting any number that has already appeared in the list ($\frac{2}{2} = \frac{1}{1}$, etc.). It is clear that A is simply a reordering of the positive rational numbers $\mathbb{Q}^+$. Hence $f : \mathbb{N} \to \mathbb{Q}^+$ defined by $f(n) = a_n$ is a bijection.
To finish things off, consider extending the function to all rational numbers by

$$g : \mathbb{Z} \to \mathbb{Q} : n \mapsto \begin{cases} f(n) & \text{if } n > 0, \\ 0 & \text{if } n = 0, (n) \\ -f(-n) & \text{if } n < 0, \end{cases}$$

Now $g : \mathbb{Z} \to \mathbb{Q}$ is a bijection, from which $|\mathbb{Q}| = |\mathbb{Z}| = \aleph_0$.



■

If you don't think that this result is strange then you're not thinking hard enough! At the very least it should teach you why the concept of cardinality is necessary and that the phrase 'an infinite number' should never be used! Any sensible person should feel that there are far, far more rational numbers than integers and yet the two sets have the same cardinality. Bizarre.

There are many more countable sets that appear to be even larger. For example, $\mathbb{N} \times \mathbb{N}$ is countable (almost the same proof except that there are no repeats to delete). For a much larger-seeming countable set, consider the *algebraic numbers:* the set

$$\{x \in \mathbb{R} : \exists \text{ polynomial } p \text{ with integer coefficients and for which } p(x) = 0\}.$$

Clearly every rational number $\frac{a}{b}$ is algebraic, since $x = \frac{a}{b}$ satisfies $bx - a = 0$. There are many more algebraic numbers: e.g. $\sqrt[5]{2} - 3$ is algebraic since it is a root of the polynomial $(x+3)^5 - 2 = 0$. Numbers like π and e are not algebraic: these are termed *transcendental.*

8.3 Uncountability

You might think, since $\mathbb{Q}$ seems so large, that there couldn't be any sets with strictly larger cardinality. But we haven't thought about the set of real numbers yet…

Definition 8.8. A set A is *uncountably infinite* if $|A| > \aleph_0$, i.e. if there exists an injection $f : \mathbb{N} \to A$ but no bijection $g : \mathbb{N} \to A$.

Theorem 8.9. *The interval* $[0,1]$ *of real numbers is uncountable.*

Proof. First note that $f : \mathbb{N} \to [0,1] : n \mapsto \frac{1}{n}$ is an injective function. Thus $\aleph_0 \leq |[0,1]|$.
Now we argue by contradiction. Suppose that $g : \mathbb{N} \to [0,1]$ is a bijection. Since g is surjective, it follows that all of the numbers in $[0,1]$ are in the following list of decimals:

$$\begin{aligned} g(1) &= 0.b_{11}b_{12}b_{13}b_{14}b_{15}b_{16}\cdots \\ g(2) &= 0.b_{21}b_{22}b_{23}b_{24}b_{25}b_{26}\cdots \\ g(3) &= 0.b_{31}b_{32}b_{33}b_{34}b_{35}b_{36}\cdots \\ g(4) &= 0.b_{41}b_{42}b_{43}b_{44}b_{45}b_{46}\cdots \\ g(5) &= 0.b_{51}b_{52}b_{53}b_{54}b_{55}b_{56}\cdots \\ &\vdots \end{aligned} \qquad \text{where each } b_{ij} \in \{0,\ldots,9\}.$$

Here we choose the terminating decimal expansion if a number has one.[41] It follows that the decimal representation of each $x \in [0,1]$ is unique, and so there are no repeats in the list. Now consider the decimal

$$c = 0.c_1c_2c_3c_4c_5\cdots \qquad \text{where } c_i = \begin{cases} 0 & \text{if } b_{nn} \neq 0, \\ 1 & \text{if } b_{nn} = 0. \end{cases}$$

Since c disagrees with $g(n)$ at the nth decimal place, we have $c \neq g(n)$, $\forall n$. Hence c is not in the above list. But $c \in [0,1]$ and g is surjective, so we have a contradiction. ∎

The interval $[0,1]$ thus has a strictly larger cardinality than the set of integers; we write $\mathfrak{c}$, or $2^{\aleph_0}$ for its cardinality.[42]. It can be shown that both the real numbers and irrational numbers are also uncountable (see the homework). More amazingly, the Cantor 1/3-set (page 53) also has cardinality $\mathfrak{c}$, despite seeming vashishingly small.

8.4 More advanced ideas

Our countable and uncountable examples are merely scratching the foothills of a truly weird subject. Here are a couple more ideas.

The following theorem is very useful for being able to compare cardinals. It allows us to compute the cardinality of a set without constructing bijective functions. Injective functions are usually much easier to find. The theorem seems like it should be obvious, but the proof is far from it.

Theorem 8.10 (Cantor–Schröder–Bernstein)**.** $|A| \leq |B|$ *and* $|B| \leq |A| \implies |A| = |B|$.

Otherwise said: given injective functions $f : A \to B$ and $g : B \to A$, there exists a bijective $h : A \to B$. The proof is far too hard for this course. It can be found in any book on set theory, or even on the web. Here is an example of its use:

Theorem 8.11. *The interval* $(0,1)$ *has cardinality* $\mathfrak{c} = |[0,1]|$.

[41] Similar to our discussion of ternary representaions and the Cantor set (page 54), a decimal representation of a number x is unique unless x has a termination representation. In such a case we can subtract 1 from the final digit, and insert an infinite sequence of 9's. For example, $0.317 = 0.3169999\cdots$.

[42] $\mathfrak{c}$ for 'continuum.'

It would be extremely messy to attempt to define a bijection $h : (0,1) \to [0,1]$. Instead we construct two injections.

Proof. $f : (0,1) \to [0,1] : x \mapsto x$ is clearly an injection. Now define

$$g : [0,1] \to (0,1) : x \mapsto \frac{1}{2}x + \frac{1}{4}.$$

g is certainly injective (in fact it is bijctive onto $[\frac{1}{4}, \frac{3}{4}] \subseteq (0,1)$). By Cantor–Schröder–Bernstein, the sets $(0,1)$ and $[0,1]$ have the same cardinality. ∎

For a final idea, we show that we can always build sets with larger cardinality.

Theorem 8.12. $|A| \lneq |\mathcal{P}(A)|$.

Proof. We must show that there is an injective function $f : A \to \mathcal{P}(A)$, but no bijective function $g : A \to \mathcal{P}(A)$.
First note that $f : a \mapsto \{a\}$ is certainly injective.
Now suppose for a contradiction that $\exists g : A \to \mathcal{P}(A)$ bijective. Consider the set

$$X = \{a \in A : a \notin g(a)\}.$$

This is a difficult set to think about. Before proceeding, let us consider an example. Suppose that $g : \{1,2\} \to \mathcal{P}(\{1,2\})$ is defined by

$$g(1) = \{1,2\}, \qquad g(2) = \{1\}.$$

Then $1 \in g(1)$ and $2 \notin g(2)$, whence the above set is exactly $X = \{2\}$. Since we are trying to show that the g in the proof does not exist, it is important to note that our example g here is *not bijective!*

Since g is bijective, it is certainly surjective, and so X *is in the image of* g. That is, $\exists \hat{a} \in A$ such that $g(\hat{a}) = X$. We ask whether $\hat{a}$ is an element of X. If it is, then $\hat{a} \notin g(\hat{a})$, which is X. We thus conclude:

$$\hat{a} \in X \iff \hat{a} \notin X.$$

This is clearly a contradiction! ∎

What this result means is that there is no 'largest set.' Given any set, its power set is strictly larger. For example, the set $\mathcal{P}(\mathbb{R})$ of subsets of $\mathbb{R}$ has a larger cardinality even than $\mathbb{R}$!

This theorem in part motivated the push to axiomatizing set theory. Here is a conundrum: If a 'set' is just a collection of objects, then there should be nothing stopping us considering the 'set of all sets.' Call this X. Now consider the power set of X. Since this is a set of sets, it must be a subset of X, whence $|\mathcal{P}(X)| \leq |X|$. However, by the theorem, we have

$$|X| \lneq |\mathcal{P}(X))| \implies |\mathcal{P}(X)| \lneq |\mathcal{P}(X)|.$$

This is a palpable contadiction. The remedy is a thorough definition of 'set' which prevents us from considering X to be a set: this is *axiomatic set theory.*