Introduction to University Mathematics

Preamble

The goal of this course is to introduce you to a range of mathematical ideas that are fundamental to studying degree-level mathematics. The course does not cover anything in great depth and is not foundational in the sense that we start from a list of axioms; that will happen in other courses. Rather, this course aims to provide a rapid introduction to various concepts, notation, and methods of logical reasoning, which you should find helpful as you begin studying mathematics at university.

These lecture notes draw upon material from previous notes authored by Richard Earl, Alan Lauder and Peter Neumann.

Please send corrections/queries to hewitt@maths.ox.ac.uk

Ian Hewitt, September 11, 2022

Synopsis

The natural numbers and their ordering. Induction as a method of proof, including a proof of the binomial theorem with non-negative integral coefficients.

Sets. Examples including $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, and intervals in $\mathbb{R}$. Inclusion, union, intersection, power set, ordered pairs and Cartesian product of sets. Relations. Definition of an equivalence relation. Examples.

Functions: composition, restriction; injective (one-to-one), surjective (onto) and invertible functions; images and preimages.

Writing mathematics. The language of mathematical reasoning; quantifiers: 'for all', 'there exists'. Formulation of mathematical statements with examples.

Proofs and refutations: standard techniques for constructing proofs; counter-examples. Example of proof by contradiction and more on proof by induction.

Problem-solving in mathematics: experimentation, conjecture, confirmation, followed by explaining the solution precisely.

The Greek alphabet

A, α	alpha	H, η	eta	N, ν	nu	T, τ	tau
B, β	beta	Θ, θ	theta	Ξ, ξ	xi	Y, υ	upsilon
Γ, γ	gamma	I, ι	iota	O, o	omicron	Φ, ϕ, φ	phi
Δ, δ	delta	K, κ	kappa	Π, π	pi	X, χ	chi
E, ε, ϵ	epsilon	Λ, λ	lambda	P, ρ, ϱ	rho	Ψ, ψ	psi
Z, ζ	zeta	M, μ	mu	$\Sigma, \sigma, \varsigma$	sigma	Ω, ω	omega

Contents

1 The natural numbers and induction

1.1 The natural numbers

We start by discussing the natural numbers, which we define in the following way:

Definition 1.1. A **natural number** is a non-negative integer. That is, it is a member of the sequence $0, 1, 2, 3, \ldots$, obtained by starting from 0 and adding 1 successively. We write $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ for the set of all natural numbers.

The curly bracket notation here indicates that the objects are grouped together as a *set*. A set is simply a collection of objects; we will discuss more about sets later.

There is not universal agreement on whether or not to include 0 as a natural number. When discussing foundational material it is more convenient to include 0, but you will sometimes see it not included (in which case the sequence starts at 1, as most of us more naturally start counting). This immediately highlights the importance of being clear about our definition! You will see a lot of definitions in the course of your degree - sometimes they may seem a bit pedantic, but you will quickly find that it is helpful (indeed very necessary) to have clear and precise definitions of the objects and properties with which you are working.

We are familiar with various properties of the natural numbers. For example, they can be added and multiplied. That is, if m and n are natural numbers, then we can construct $m + n$ and $m \times n$, which are also natural numbers (the multiplication symbol $\times$ is often omitted - simply writing the numbers next to each other, mn, is understood to mean multiplication. A dot may also be used, especially when handwriting; for example, you might see things like $2.3 = 6$). Addition and multiplication are examples of *binary operations*: they take a pair of elements from $\mathbb{N}$ and produce an element of $\mathbb{N}$.

Two important natural numbers are 0 and 1, which are the additive and multiplicative identities, meaning they have the properties

$$n + 0 = n \quad \text{and} \quad n \times 1 = n \quad \text{for all } n \in \mathbb{N}.$$

The symbol $\in$ is shorthand to mean 'is an element of', and we read this as 'for all n in the natural numbers', or 'for all natural numbers n'. In fact, another symbol $\forall$ is often used in place of 'for all', but for the moment I will keep writing that out in words.

Another important property of the natural numbers is that they have an ordering, so we can write things like $m \leq n$. We can carefully define this less-than-or-equal-to symbol:

Definition 1.2. Let m and n be natural numbers. We write $m \leq n$ to mean that there exists a natural number k such that $m + k = n$.

This is an example of a *relation*, which we'll discuss more later.

Notice that we have not actually *defined* addition and multiplication; I am appealing to our familiar understanding of what these operations mean. In fact, even our definition of the natural numbers is a little unsatisfactory if we want to build things up from nothing, since it implicitly relies on some notion of what the integers are, or of what it means to 'add 1'. For the purpose of this course we will rely on our basic intuition for these things, but it is possible to be more careful, building on an axiomatic description of $\mathbb{N}$ (that is, laying out some clear *axioms* - statements that we assume to be true as a starting point - and then deducing all other properties from those).

(More generally this is something to be aware of for all your courses; they will have different 'starting points'. In Analysis, you will start with the axioms for the real numbers, and will be expected to carefully deduce other statements that you might intuitively think of as 'obvious'. In more applied courses like Introductory Calculus, you may be expected to use your existing knowledge about how to differentiate functions, despite having not yet defined carefully what a derivative actually is. It can be a bit confusing to start with, working out what you are 'allowed' to assume in different contexts, but is something you should gradually become more comfortable with).

1.2 Mathematical induction

We now move on to talk about induction. The following principle is sometimes quoted as a theorem, although it follows directly from our definition of the natural numbers. In fact it can be used as an axiom when defining $\mathbb{N}$ in a more rigorous manner.

Theorem 1.3 (Principle of Induction)**.** *Let $P(n)$ be a family of statements indexed by the natural numbers. Suppose that (i) $P(0)$ is true and (ii) for any n, if $P(n)$ is true then $P(n+1)$ is also true. Then $P(n)$ is true for all natural numbers n.*

Induction is often visualised like toppling dominoes. The *inductive step* (ii) corresponds to placing each domino sufficiently close that it will be hit when the previous one falls over, and the *initial step*, (i) - often called the *base case* - corresponds to knocking over the first one.

To use induction to prove a family of statements, we simply have to demonstrate (i) and (ii). Here is a very straightforward example:

Proposition 1.4. *For any $n \in \mathbb{N}$,*

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

Proof. It is helpful to clarify what exactly the statement $P(n)$ is. In this case $P(n)$ is simply the statement that the given equality holds for that particular n. Clearly $P(0)$ holds because for $n = 0$ the sum on the LHS is 0 and the expression on the

RHS is also 0. Now suppose $P(n)$ holds. Then

$$\begin{aligned}\sum_{k=0}^{n+1} k &= \sum_{k=0}^{n} k + (n+1) \\ &= \frac{n(n+1)}{2} + (n+1) \qquad \text{[by the inductive hypothesis]} \\ &= \frac{(n+1)(n+2)}{2},\end{aligned}$$

which is exactly the statement $P(n+1)$. So by induction, $P(n)$ is true for all n. □

The small square on the right here is used to signify that we've reached the end of the proof. Historically the letters QED were used for this purpose (standing for the Latin *quad erat demonstrandum*, meaning 'what was needing to be shown'), but that has largely gone out of fashion. Other symbols, including a filled square ■, are sometimes used. It is also worth clarifying that LHS and RHS are commonly used abbreviations for 'left hand side' and 'right hand side' of an equality. When using them in your mathematical writing, make sure it is clear to which equality they relate. One other comment on the above proof is that we have noted at the side where we made use of the inductive hypothesis. Such 'commentary' on your algebraic manipulations can be extremely helpful in communicating your argument. (Some of the proofs in these notes contain a bit more commentary than is necessary - e.g. the first sentence in the proof above - because I am trying to convey some of the thought process required to develop the proof as well as the proof itself. But a well-written proof should always explain any steps that are not just routine algebra. You will probably find you need to write many more actual sentences than you have been used to writing in maths at school.)

A straightforward extension of induction is if the family of statements holds for $n \geq N$, rather than necessarily $n \geq 0$:

Corollary 1.5. *Let N be an integer and let $P(n)$ be a family of statements indexed by integers $n \geq N$. Suppose that (i) $P(N)$ is true and (ii) for any $n \geq N$, if $P(n)$ is true then $P(n+1)$ is also true. Then $P(n)$ is true for all $n \geq N$.*

Proof. This follows directly by applying Theorem 1.3 to the statements $Q(n) = P(n+N)$ for $n \in \mathbb{N}$. □

Remark 1.6. We use the word **Corollary** to mean a result that is an extension of, or a consequence of, a theorem or proposition; a corollary is generally not such a major result as the theorem or proposition itself. The words **Theorem** and **Proposition** are used somewhat interchangeably to mean a result that one has proved (unlike a *conjecture*, which is something that has not yet been proven). Theorem is typically used for more significant results, and theorems are often given a specific name. We also use the word **Lemma**, to mean a result that is going to be useful in proving a later theorem or proposition. Lemmas are typically not such exciting or major results in themselves.

Another variant on induction is when the inductive step relies on some earlier case(s) but not necessarily the immediately previous case. This is sometimes called *strong induction*:

Theorem 1.7 (Strong Form of Induction)**.** *Let $P(n)$ be a family of statements indexed by the natural numbers. Suppose that (i) $P(0)$ is true and (ii) for any n, if $P(0), P(1), \ldots, P(n)$ are true then $P(n+1)$ is also true. Then $P(n)$ is true for all natural numbers n.*

Proof. In fact this is not really anything different, and we can straightforwardly convert it to an instance of 'normal' induction by defining a related family of statements $Q(n)$. To do this, let $Q(n)$ be the statement '$P(k)$ holds for $k = 0, 1, \ldots n$'. Then the conditions for the strong form are equivalent to (i) $Q(0)$ holds and (ii) for any n, if $Q(n)$ is true then $Q(n+1)$ is also true. It follows by induction that $Q(n)$ holds for all n, and hence $P(n)$ holds for all n. □

The following example illustrates how the strong form of induction can be useful:

Proposition 1.8. *Every natural number greater than* 1 *may be expressed as a product of one or more prime numbers.*

Proof. Let $P(n)$ be the statement that n may be expressed as a product of prime numbers. Clearly $P(2)$ holds, since 2 is itself prime. Let $n \geq 2$ be a natural number and suppose that $P(m)$ holds for all $m < n$. If n is prime then it is trivially the 'product' of the single prime number n. If n is not prime, then there must exist some $r, s > 1$ such that $n = rs$. By the inductive hypothesis, each of r and s can be written as a product of primes, and therefore $n = rs$ is also a product of primes. Thus, whether n is prime or not, we have have that $P(n)$ holds. By strong induction, $P(n)$ is true for all natural numbers. That is, every natural number greater than 1 may be expressed as a product of one or more primes. □

Related to induction is the idea of *recursion* as a method of definition. For example, supposing we are happy with what it means to 'add 1', we can recursively define more general addition on the natural numbers:

Definition 1.9. Define addition on $\mathbb{N}$ by the rules that for all $m \in \mathbb{N}$, (i) $m + 0 = m$ and (ii) for any $n \in \mathbb{N}$, $m + (n+1) = (m+n) + 1$.

We can combine this with induction to prove some useful properties. For example,

Proposition 1.10 (Associativity)**.** *Addition on $\mathbb{N}$ is associative. That is, for all $x, y, z \in \mathbb{N}$,*

$$x + (y + z) = (x + y) + z.$$

Proof. We induct on z, so first suppose $z = 0$. Then, for any $x, y \in \mathbb{N}$,

$$\text{LHS} = x + (y + 0) = x + y = (x + y) + 0 = \text{RHS},$$

where we have twice used rule (i) from our definition of addition. Now for the inductive step suppose the proposition is true for $z = n$ and consider the case $z = n + 1$. Then, for any $x, y \in \mathbb{N}$,

$$\begin{aligned}
\text{LHS} &= x + (y + (n + 1)) \\
&= x + ((y + n) + 1) && \text{[rule (ii) from the definition]} \\
&= (x + (y + n)) + 1 && \text{[rule (ii) from the definition]} \\
&= ((x + y) + n) + 1 && \text{[inductive hypothesis]} \\
&= (x + y) + (n + 1) && \text{[rule (ii) from the definition]} \\
&= \text{RHS}.
\end{aligned}$$

So, by induction, the expression holds for any $z \in \mathbb{N}$. Thus addition is associative. □

We can use a similar approach to define multiplication and factorial, for example:

Definition 1.11. Define multiplication on $\mathbb{N}$ by the rules that for all $m \in \mathbb{N}$, (i) $m \times 0 = 0$ and (ii) for any $n \in \mathbb{N}$, $m \times (n + 1) = (m \times n) + m$.

Definition 1.12. Define factorial $n!$ on $\mathbb{N}$ by the rules that (i) $0! = 1$ and (ii) for any $n \in \mathbb{N}$, $(n + 1)! = n! \times (n + 1)$.

Here is another important property of the natural numbers that we can prove using induction:

Theorem 1.13 (Well-ordering property of the natural numbers)**.** *Every non-empty subset of* $\mathbb{N}$ *has a least element.*

[We have not yet defined a subset, but maybe you can guess what it means. S is a subset of $\mathbb{N}$ if every element of S is also an element of $\mathbb{N}$. Non-empty means it contains one or more elements.]

Proof. We prove this by contradiction. Suppose, for a contradiction, that there is a non-empty subset S that does *not* have a least element (a least element is sometimes called a minimal element, or a least member - various phrasings mean the same thing). We define S^* to be the set of natural numbers that are not in S (the *complement* of S), and aim to show (by induction) that in fact $S^* = \mathbb{N}$, since that means S is empty, which provides the contradiction.

Let $P(n)$ be the statement that S^* contains n. For the initial step, note that 0 is not in S (we can write $0 \notin S$), since if it were, then S would have a least element (namely 0). So $0 \in S^*$ and therefore $P(0)$ holds. Now suppose $P(0), \ldots, P(n)$ hold. Then $n + 1$ cannot be in S, because if it were then it would be the least element of S (since by the inductive hypothesis all the smaller elements of $\mathbb{N}$ are not in S). Hence $n + 1 \in S^*$, and therefore $P(n + 1)$ holds. By strong induction, $n \in S^*$ for all $n \in \mathbb{N}$, and therefore S is empty. This contradicts our initial assumption and therefore proves the result. □

Here we have laid out the proof by carefully defining the statements $P(n)$ involved in the inductive argument. We can often get away without doing this quite so explicitly, and that is acceptable so long as the logic is clear. Indeed, it is not a good idea to make a proof unnecessarily long-winded. But when developing a feel for how such proofs work, and especially if you start to find yourself confused about your argument, you may find it helpful to be explicit like this.

The well-ordering property of the natural numbers is one that you may well think is 'obvious' (though note that the same property is not true of the real numbers, for example, so it is not an entirely trivial property). Here, we have used the principle of induction to prove it. In fact, the well-ordering property is essentially equivalent to the principle of induction; it is also possible to work the other way and use it to prove the principle of induction. Here is a proof of Theorem 1.3 based only on the well-ordering property:

Proof of Theorem 1.3. Let $S = \{n \in \mathbb{N} : P(n) \text{ is false}\}$. We aim to show that S is empty.

Suppose, for a contradiction, that S were not empty. Then the well-ordering property means that S has a least element. That least element cannot be 0 since $P(0)$ is true by the initial step (i). Therefore we can write the least element as $n+1$ for some $n \in \mathbb{N}$. Since $n+1$ is the least element in S it must be the case that $P(n)$ holds. But then the inductive step (ii) implies that $P(n+1)$ also holds, which contradicts $n+1$ being in S.

Thus S must be empty, and therefore $P(n)$ holds for all $n \in \mathbb{N}$. □

1.3 The binomial theorem

We next aim to prove the **binomial theorem**, which provides the rule for how to expand a product of the form $(x+y)^n$. First, we define some notation:

Definition 1.14. For natural numbers n and k, we define the **binomial coefficient**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \qquad \text{for } 0 \leqslant k \leqslant n.$$

This is read as 'n choose k' and is also sometimes denoted by ${}^nC_k$. By convention $\binom{n}{k} = 0$ if $k > n$.

The binomial coefficients appear in many areas of mathematics. They represent the number of ways of choosing k elements from a set of size n. They can famously be laid out as an array called Pascal's triangle, in which the nth row contains each of the non-zero $\binom{n}{k}$:

$$\begin{array}{ll}
n=0 & \qquad\qquad\qquad\qquad 1 \\
n=1 & \qquad\qquad\qquad 1 \qquad 1 \\
n=2 & \qquad\qquad 1 \qquad 2 \qquad 1 \\
n=3 & \qquad\quad 1 \qquad 3 \qquad 3 \qquad 1 \\
n=4 & \qquad 1 \qquad 4 \qquad 6 \qquad 4 \qquad 1 \\
n=5 & \quad 1 \qquad 5 \qquad 10 \qquad 10 \qquad 5 \qquad 1
\end{array}$$

The following result is an algebraic expression of the defining feature of Pascal's triangle, that each entry is the sum of the two entries most immediately above it.

Lemma 1.15 (Pascal's Triangle)**.** *Let n and k be natural numbers with $1 \leq k \leq n$. Then*

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

Proof. This is simply a question of putting in the definitions and playing with the algebra. Putting the left hand side over a common denominator we obtain

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

which is equal to the right hand side. □

This lemma can be used to show (by induction) that the binomial coefficients are integers rather than just rational numbers (a fact that is perhaps not immediately obvious from the definition).

We are now in a position to prove the binomial theorem (for non-negative integer exponents):

Theorem 1.16 (Binomial Theorem)**.** *Let x and y be real (or complex) numbers, and n be any natural number. Then*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof. We use induction on n. First we check that the expression holds for $n = 0$. This is true, since the left hand side is 1 in that case, and the right hand side is also 1 (because $\binom{0}{0} = 1$ and any number raised to the power 0 is 1). Now assume the expression holds for n and consider the case for $n+1$,

$$(x+y)^{n+1} = (x+y)(x+y)^n = (x+y)\left(\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\right),$$

where we have made use of the inductive hypothesis in this last step. Continuing to expand the brackets gives

$$\sum_{k=0}^{n} \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n+1-k} = x^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=1}^{n} \binom{n}{k} x^k y^{n+1-k} + y^{n+1},$$

where we have taken out the last term from the first sum and the first term from the second sum. In the first sum we now make a change of indexing variable; we set

$k = l - 1$, noting that as k ranges over $0, 1, ..., n-1$ then l ranges over $1, 2, ..., n$. So the above equals

$$\begin{aligned}
&x^{n+1} + \sum_{l=1}^{n} \binom{n}{l-1} x^l y^{n+1-l} + \sum_{k=1}^{n} \binom{n}{k} x^k y^{n+1-k} + y^{n+1} \\
&= x^{n+1} + \sum_{k=1}^{n} \left\{ \binom{n}{k-1} + \binom{n}{k} \right\} x^k y^{n+1-k} + y^{n+1} \qquad \text{[relabelling } l \text{ as } k\text{]} \\
&= x^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} x^k y^{n+1-k} + y^{n+1} \qquad \text{[using Lemma 1.15]} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k},
\end{aligned}$$

which shows that the expression holds for $n+1$. Thus, by induction, the expression holds for all n. □

2 Sets

2.1 Definitions, notation, and examples

Sets are amongst the most primitive objects in mathematics, so primitive that it is not straightforward to give a precise definition. We will work with the naive definition that a set is a collection of objects. The objects are called the **elements**, or members, of the set.

We might ask what exactly is meant by a 'collection', or by 'objects', and there are problems with taking too broad an interpretation of these terms (see the example of Russell's paradox below). But hopefully we have some intuitive understanding of what is meant, and we will see examples that make it clearer. The looseness of our definition is something to be aware of, but not to get too hung up about at this stage. We *can* make things more precise, and you could take the Part B Set Theory course if you want to explore this more deeply.

For a set S, we write $x \in S$ to mean that x is an element of S, and we write $x \notin S$ to mean that x is not an element of S. Two sets S and T are equal if and only if they contain the same elements, and in that case we can write $S = T$.

Definition 2.1. The **empty set** is the set with no elements, and is denoted by $\varnothing$. (Note that this is not quite the same symbol as the Greek letter ϕ).

Definition 2.2. A set A is said to be a **subset** of a set S if every element of A is also an element of S. We write $A \subseteq S$. If $A \subseteq S$ and $A \neq S$, we call A a **proper subset** of S.

The symbol $\subseteq$ can be read as 'is a subset of' or 'is contained in'. The symbol $\subset$ is commonly used to mean the same thing, and does not necessarily imply that the subset is proper (as you might otherwise imagine by analogy with $\leq$ and $<$ symbols).

As we have already seen, curly brackets are used to denote sets. The set with elements $a_1, a_2, \ldots a_n$ is written $\{a_1, a_2, \ldots a_n\}$. (If nothing is written at the end of a string of dots, such as $\{a_1, a_2, \ldots\}$, it typically indicates an infinite list of elements, although this is not a hard and fast rule. It is helpful practice to write the final element, as done here, to indicate if such a list is finite). The order of the elements inside the curly brackets is not important, and repeating the elements does not mean anything extra, so $\{0, 1\}$ and $\{1, 0\}$, and $\{0, 1, 1\}$ are all the same set (some additional distinguishing notation needs to be introduced if you want to include multiple 'copies' of the same element).

It is common to want to define a set in terms of some property $P(x)$ that the elements x satisfy. Typically the set is a subset of some larger set S, say, and we write $\{x \in S : P(x)\}$, or $\{x \in S \mid P(x)\}$, to mean the set of elements $x \in S$ that have property $P(x)$. This is read as 'the set of x in S such that P holds'. For example, the set of even natural numbers is $\{n \in \mathbb{N} : n \text{ is divisible by } 2\}$. The set $\{n \in \mathbb{N} : n^2 < 0\}$ is equal to the empty set $\varnothing$, since no elements of $\mathbb{N}$ satisfy the given property.

Remark 2.3. Don't confuse a with $\{a\}$, since they are quite different objects. One is the element a and the other is the set containing the single element a. For example, if $a = \varnothing$, the empty set, then a is a set with no elements, but $\{a\}$ is a set with one element (namely $\varnothing$).

We have already seen the natural numbers, $\mathbb{N} = \{0, 1, 2, \ldots\}$, as an example of a set. Some other important examples are:

Definition 2.4. The set of **integers** (or whole numbers), $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. The notation stems from the German word 'Zahlen', for 'number'.

Definition 2.5. The set of **rational numbers** (or simply 'rationals'), $\mathbb{Q}$, is the set comprising all fractions where the numerator and denominator are both integers. That is,

$$\mathbb{Q} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z}, n > 0 \right\}.$$

Definition 2.6. The set of **real numbers**, $\mathbb{R}$, is the set containing numbers with a decimal expansion. These will be more formally introduced in Analysis I.

Definition 2.7. The set of **complex numbers**, $\mathbb{C} = \{a + b\mathrm{i} : a, b \in \mathbb{R}\}$, where $\mathrm{i} = \sqrt{-1}$.

Note that 'infinity' (∞) is not included in any of these sets. They are sometimes 'extended' to include a notion of infinity, and you will likely come across such extensions later in your degree, but that requires care about what exactly is meant by infinity.

The symbols for each of these sets are written in 'blackboard bold' font. When writing them by hand there is no need to make them look quite so fancy - simply adding an extra line somewhere in the capital letter is sufficient.

Some frequently occurring subsets of the real numbers are **intervals**, which can be visualised as sections of the real line:

Definition 2.8. Given real numbers a, b with $a \leq b$ we define **bounded intervals**

$$\begin{aligned}(a,b) &= \{x \in \mathbb{R} : a < x < b\}, \\ [a,b] &= \{x \in \mathbb{R} : a \leqslant x \leqslant b\}, \\ [a,b) &= \{x \in \mathbb{R} : a \leqslant x < b\}, \\ (a,b] &= \{x \in \mathbb{R} : a < x \leqslant b\},\end{aligned}$$

and **unbounded intervals**

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

An interval of the first type (a, b) is called an **open interval** and an interval of the second type $[a, b]$ is called a **closed interval**. Note that if $a = b$, then $[a, b] = \{a\}$, while $(a, b) = [a, b) = (a, b] = \varnothing$.

All of the examples above are sets of numbers (scalars). But there is no reason why the objects could not be other types of things, like words, shapes, or people. In Linear Algebra, you will work with sets of matrices, such as:

Definition 2.9. The set $M_{mn}(\mathbb{R})$ is the set of m by n matrices with real coefficients, that is

$$M_{mn}(\mathbb{R}) = \left\{ \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} : a_{ij} \in \mathbb{R} \right\}.$$

In some cases, the objects themselves are sets. In particular:

Definition 2.10. The **power set** of a set A, denoted $\mathcal{P}(A)$, is the set of all subsets of A.

Example 2.11. For $A = \{0, 1\}$, we have $\mathcal{P}(A) = \{\varnothing, \{0\}, \{1\}, \{0, 1\}\}$.

If we have two objects a and b we can combine them together to make a set $\{a, b\}$, but we could also combine them to make an **ordered pair** (a, b). The distinction is that in an ordered pair the order matters. Thus two pairs (a_1, b_1) and (a_2, b_2) are equal if and only if $a_1 = a_2$ and $b_1 = b_d$. We can similarly create ordered triples (a, b, c), quadruples (a, b, c, d) and so on. If there are n elements it is called an n-tuple.

You are probably familiar with such objects in the context of **vectors** or coordinates. If the elements are real numbers, for example, then (a, b) is a two-dimensional vector, and can be thought of as representing the x-y coordinates of a point on a plane (in that context it is clear that $(1, 2)$ is quite different from $(2, 1)$). Although we will often use ordered pairs in this way, the definition is much more general, since there is no reason why the elements a and b need come from the same sets or even be similar 'types' of object.

Combining elements as ordered pairs provides a way of combining sets:

Definition 2.12. Given sets A and B, the **Cartesian product**, denoted $A \times B$, is the set of all ordered pairs with the first element of the pair coming from A and the second from B. That is,

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

If $A = B$, we can also write $A \times A$ as A^2.

More generally, we define $A_1 \times A_2 \times \ldots \times A_n$ to be the set of all ordered n-tuples $(a_1, a_2, \ldots, a_n)$, where $a_i \in A_i$ for $1 \leq i \leq n$. If all the A_i are the same, we write the product as A^n.

The case when $A = B = \mathbb{R}$ is a particularly important one that you will see a lot of in Geometry. In that case the Cartesian product is $\mathbb{R}^2$ (usually read as 'r two' rather than 'r squared'), and represents the two-dimensional plane.

Remark 2.13. The following example is known as Russell's paradox (after the mathematician and philosopher Bertrand Russell, 1872-1970). It provides a warning as to the looseness of our definition of a set. Suppose

$$H = \{\text{sets } S : S \notin S\}.$$

That is, H is the collection of sets S that are not elements of themselves. All the sets we have come across seem to be in H (for example, $\mathbb{N}$ is in H since the elements of $\mathbb{N}$ are individual numbers and clearly none of them is the set $\mathbb{N}$ itself). The problem arises when we ask the question of whether or not H is itself in H?

On the one hand, if $H \notin H$ then H meets the precise criterion for being in H and so $H \in H$, a contradiction. On the other hand, if $H \in H$ then by the property required for this to be the case, $H \notin H$, another contradiction. Thus we have a paradox: H seems to be neither in H nor not in H.

The modern resolution of Russell's Paradox is that we have taken too naive an understanding of 'collection', and that Russell's 'set' H is in fact not a set. It does not fit within axiomatic set theory (which relies on the so-called ZF axioms), and so the question of whether or not H is in H simply doesn't make sense.

2.2 Algebra of sets

Definition 2.14. Given subsets A and B of a set S, the **union** $A \cup B$ is the set consisting of those elements that are in A or B (or both), that is:

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

The **intersection** $A \cap B$ is the set consisting of those elements that are in both A and B, that is:

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

The **complement** of A, written A^c or sometimes A', is the subset consisting of those elements that are not in A, that is:

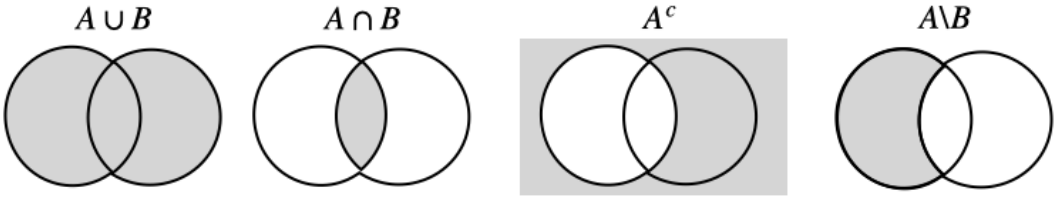
$$A^c = \{x \in S : x \notin A\}.$$

The **set difference**, or complement of B in A, written $A\backslash B$, is the subset consisting of those elements that are in A and not in B, that is:

$$A\backslash B = \{x \in A : x \notin B\}.$$

Note that $A\backslash B = A \cap B^c$.

A useful way of visualising these is using a Venn diagram. The following is an example, where A and B are the regions inside the two circles, respectively:



Definition 2.15. Two sets A and B are said to be **disjoint** if $A \cap B = \varnothing$, that is the two subsets have no element in common.

More generally, we can take unions and intersections of arbitrary numbers of sets, even infinitely many. If we have a family of subsets $\{A_i : i \in I\}$, where I is called an **indexing set**, we write

$$\bigcap_{i\in I} A_i = \{x \in S : x \in A_i \text{ for all } i \in I\},$$

and

$$\bigcup_{i\in I} A_i = \{x \in S : x \in A_i \text{ for at least one } i \in I\}.$$

Often I might a subset of $\mathbb{N}$, in which case we write things like

$$\bigcap_{i=1}^{n} A_i, \qquad \bigcup_{i=1}^{\infty} A_i,$$

but it could be an even 'larger' set such as $\mathbb{R}$ (see the discussion of cardinality below).

Example 2.16. Let S be the set of all students at Oxford, $A \subseteq S$ be the of set students studying mathematics, and $B \subseteq S$ be the set of students at your college. Then $A \cap B$ is the set of students studying mathematics at your college, $A \cup B$ is the set of students either at your college or studying mathematics (this is probably the set where many of your friends will come from), B^c is the the set of all students at other colleges, $A\backslash B$ is the set of students studying mathematics at other colleges.

The following result may well seem obvious, but it provides quite an important recipe for how to show that two sets are the same, as we will see below.

Proposition 2.17 (Double Inclusion)**.** *Let A and B be two subsets of a set S. Then $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.*

Proof. If $A = B$, then every element in A is an element in B, so certainly $A \subseteq B$, and similarly $B \subseteq A$. Conversely, suppose $A \subseteq B$, and $B \subseteq A$. Then for every element $x \in S$, if $x \in A$ then $A \subseteq B$ implies that $x \in B$, and if $x \notin A$ then $B \subseteq A$ means $x \notin B$. So $x \in A$ if and only if $x \in B$, and therefore $A = B$. □

Here is an example of how we can make use of double inclusion to show that two sets are equal:

Proposition 2.18 (Distributive Laws)**.** *Let A, B, C be subsets of a set S. Then*

(i) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(ii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Proof. We first prove (i).

Suppose x is in the LHS of (i), that is $x \in A \cup (B \cap C)$. This means that $x \in A$ or $x \in B \cap C$ (or both). Thus either $x \in A$ or x is in both B and C (or x is in all three sets). If $x \in A$ then $x \in A \cup B$ and $x \in A \cup C$, and therefore x is in the RHS. If x is in both B and C then similarly x is in both $A \cup B$ and $A \cup C$. Thus every element of the LHS is in the RHS, which means we have shown $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$.

Conversely suppose that $x \in (A \cup B) \cap (A \cup C)$. Then x is in both $A \cup B$ and $A \cup C$. Thus either $x \in A$ or, if $x \notin A$, then $x \in B$ and $x \in C$. Thus $x \in A \cup (B \cap C)$. Hence $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$.

By double inclusion, $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$.

The proof of (ii) follows similarly and is left as an exercise. □

Remark 2.19. In the proof above there were two separate things to show (that is, LHS $\subseteq$ RHS, and RHS $\subseteq$ LHS, which combine to give the required result). When laying out a proof like this it is helpful to separate the two things out clearly, both to aid your own understanding and that of the reader. Here we have done that by starting a new paragraph, while the smaller individual steps of the argument were written as sentences. When hand-writing a proof, some people tend to put each step of the logic on a new line, in which case leaving a larger gap or using different indentation may help to distinguish the separate sections.

Here is another important example, about how set complements work:

Proposition 2.20 (De Morgan's Laws)**.** *Let A and B be subsets of a set S. Then*

$$(A \cup B)^c = A^c \cap B^c \qquad \text{and} \qquad (A \cap B)^c = A^c \cup B^c.$$

Proof. For the first one, suppose $x \in (A \cup B)^c$. Then x is not in either A or B. Thus $x \in A^c$ and $x \in B^c$, and therefore $x \in A^c \cap B^c$.

Conversely, suppose $x \in A^c \cap B^c$. Then $x \notin A$ and $x \notin B$, so x is in neither A nor B, and therefore $x \in (A \cup B)^c$.

By double inclusion, the first result holds. The second result follows similarly and is again left as an exercise. □

De Morgan's laws extend naturally to any number of sets, so if $\{A_i : i \in I\}$ is a family of subsets of S, then

$$\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c \qquad \text{and} \qquad \left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c.$$

2.3 Truth tables

Another way of proving set-theoretical identities is via **truth tables**. These provide a systematic way of cataloguing all the different cases for whether or not a given element is in each set. Here is an example:

A	B	$A \cap B$	$A \cup B$	$A \backslash B$
F	F	F	F	F
F	T	F	T	F
T	F	F	T	T
T	T	T	T	F

The different cases are listed in the rows, one row for each, and various sets are placed along the columns. We put a T or an F in each column to indicate whether it is true or false that the given element is in that set for each case. There will be different numbers of cases to consider depending on the number of sets involved; in this example with two sets, there are four (the first two columns effectively define these four different cases, and the entries in the other columns then follow from those).

Truth tables like this are more often used to catalogue the cases of 'true' or 'false' for a series of logical statements. These truth tables for sets are a particular instance, where the statements are of the form '$x \in A$', '$x \in B$', '$x \in A \cap B$', and so on.

Here is an alternative proof of De Morgan's laws using a truth table:

Proof of Proposition 2.20. We list the four combinations of cases for whether or not $x \in A$ and $x \in B$:

A	B	$A \cap B$	$(A \cap B)^c$	$A \cup B$	$(A \cup B)^c$	A^c	B^c	$A^c \cup B^c$	$A^c \cap B^c$
F	F	F	T	F	T	T	T	T	T
F	T	F	T	T	F	T	F	T	F
T	F	F	T	T	F	F	T	T	F
T	T	T	F	T	F	F	F	F	F

Comparing columns, the fact that $(A \cap B)^c$ and $A^c \cup B^c$ are the same in every case shows that those two sets are the same. Similarly, the fact that $(A \cup B)^c$ and $A^c \cap B^c$ are the same in every case shows that those two sets are the same. □

2.4 Cardinality

Informally, the cardinality of a set S, denoted $|S|$, is a measure of its 'size'. For finite sets, there is little ambiguity about this - it is simply the number of distinct elements in the set (we give a formal definition below, which will also clarify what it

means to say that a set is finite). But for infinite sets, things are more interesting. For example, one might be tempted to think that the set of even natural numbers is in some sense 'smaller' than the set of natural numbers - there might reasonably seem to be 'fewer' of them, since we have left the odd numbers out. But if we simply divide every element in that set by 2, we don't change the number of elements and we are left with the set of natural numbers itself, so by this logic it seems that these two sets ought to be the same size. Indeed, these two sets do have the same cardinality, $\aleph_0$ (pronounced 'aleph-null'). This is the smallest infinite cardinal (the concept of cardinals was invented and investigated widely by Georg Cantor, 1845-1918). Perhaps surprisingly, the rational numbers $\mathbb{Q}$ also have the same cardinality $\aleph_0$. But the cardinality of the real numbers $\mathbb{R}$ is larger, as will be discussed more in the Analysis I course.

We will be able to give a nicer definition of cardinality later, once we have discussed bijections, but the following provides a recursive definition of the cardinality for a finite set:

Definition 2.21 (Finiteness and the cardinality of a finite set)**.** The empty set $\varnothing$ is finite with $|\varnothing| = 0$. A set S is finite with $|S| = n + 1$, if there exists $s \in S$ such that $|S \backslash \{s\}| = n$ for some $n \in \mathbb{N}$. We call $|S|$ the **cardinality** of S. Any set that is not finite is said to be infinite.

It is not hard to see that this means that if $S = \{x_1, x_2, \ldots x_n\}$, and $x_i \neq x_j$ whenever $i \neq j$, then $|S| = n$. Conversely, if $|S| = n$ then S is a set with n elements.

Proposition 2.22. *Let A and B be finite sets. Then $|A \cup B| = |A| + |B| - |A \cap B|$.*

The proof is left as an exercise (see problem sheet).

Proposition 2.23 (Subsets of a finite set)**.** *If a set A is finite with $|A| = n$, then its power set has $|\mathcal{P}(A)| = 2^n$.*

Proof. We use induction. For the initial step, note that if $|A| = 0$ then $A = \varnothing$ has no elements, so there is a single subset, $\varnothing$, and therefore $|\mathcal{P}(A)| = 1 = 2^0$.

Now suppose that $n \geq 0$ and that $|\mathcal{P}(S)| = 2^n$ for any set S with $|S| = n$. Let A be any set with $|A| = n + 1$. By definition, this means that there is an element a and a set $A' = A \backslash \{a\}$ with $|A'| = n$. Any subset of A must either contain the element a or not, so we can partition $\mathcal{P}(A) = \mathcal{P}(A') \cup \{S \cup \{a\} : S \in \mathcal{P}(A')\}$. These two sets are disjoint, and each of them has cardinality $|\mathcal{P}(A')| = 2^n$ by the inductive hypothesis. Hence $|\mathcal{P}(A)| = 2^n + 2^n = 2^{n+1}$.

Thus, by induction, the result holds for all n. □

An alternative, perhaps easier, way to see why the size of the power set should be $2^{|A|}$ for a finite set A, is to consider the process of creating a subset. We can do this systematically by going through each of the $|A|$ elements in A and making the yes/no decision whether to put it in the subset. Since there are $|A|$ such choices, that yields $2^{|A|}$ different combinations of elements and therefore $2^{|A|}$ different subsets.

3 Relations

3.1 Definition and examples

In mathematics a **relation** (sometimes *binary relation*) is something like $\leq$ or $\subseteq$ that tells us how two objects compare to each other. $=$ is also an example of a relation. Formally, we define it as the set of ordered pairs (a, b) for which the relation holds:

Definition 3.1. A **relation** R on a set S is a subset of $S \times S$. If $(a, b) \in R$, we write aRb.

It may seem odd to think of $\leq$ as a subset like this, but this definition turns out to be a convenient way to describe such a relation as a mathematical object. We could alternatively think of a relation as a binary operation that takes two input elements and returns a 'True' or 'False', depending on whether the two elements 'relate' according to that relation. But this is effectively the same thing as our definition above, since the subset R is simply those elements of $S \times S$ that return 'True' under that operation.

Example 3.2. In many cases we don't actually use R to write the relation because there is some other conventional notation:

(i) The 'less than or equal to' relation $\leq$ on the set of real numbers is $\{(x, y) \in \mathbb{R}^2 : x \leq y\}$. We write $x \leq y$ if (x, y) is in this set.

(ii) The 'divides' relation $|$ on $\mathbb{N}$ is $\{(m, n) \in \mathbb{N}^2 : \ m \text{ divides } n\}$. We write $m|n$ if (m, n) is in this set.

(iii) For a set S, the 'subset' relation $\subseteq$ on $\mathcal{P}(S)$ is $\{(A, B) \in \mathcal{P}(S)^2 : A \subseteq B\}$. We write $A \subseteq B$ if (A, B) is in this set.

We have defined a relation between elements that are in the same set S, and these are the most common. But it is also possible to have a relation between elements that are in *different* sets. In that case R is simply a subset of $A \times B$, where A and B are the two sets. Indeed there are very many different kinds of relations that we come across in mathematical and everyday life:

Example 3.3. If S is the set of all students at Oxford, and C is the set of all Oxford colleges, we could define $R = \{(s, c) \in S \times C : \text{student } s \text{ is a member of college } c\}$. Then saying sRc simply means s is studying at c (I have not strictly followed the definition here; I have converted 'is a member of' to 'is studying at', but in theory at least, these ought to be the same!)

3.2 Reflexivity, symmetry, anti-symmetry, and transitivity

Definition 3.4. Let S be a set, R a relation on S and $x, y, z \in S$. We say that

(i) R is **reflexive** if xRx for all x in S,

(ii) R is **symmetric** if whenever xRy then yRx,

(iii) R is **anti-symmetric** if whenever xRy and yRx then $x = y$,

(iv) R is **transitive** if whenever xRy and yRz then xRz.

For example, the relation $\leq$ on $\mathbb{R}$ is reflexive, anti-symmetric, and transitive, but not symmetric. More generally, any relation on a set S that is reflexive, anti-symmetric, and transitive is called a **partial order**. It is called a **total order** if for every $x, y \in S$, either xRy or yRx (or both).

As further examples, the relation $<$ on $\mathbb{R}$ is not reflexive, symmetric, or antisymmetric, but it is transitive. The relation $\neq$ on $\mathbb{R}$ is not reflexive, antisymmetric or transitive, but it is symmetric.

Here is another quite important example:

Example 3.5. Let $n \geq 2$ be an integer, and define R on $\mathbb{Z}$ by saying aRb if and only if $a - b$ is a multiple of n. Then R is reflexive, symmetric and transitive.

Proof. Reflexivity: For any $a \in \mathbb{Z}$ we have aRa as 0 is a multiple of n.
Symmetry: If aRb then $a - b = kn$ for some integer k. So $b - a = -kn$, and hence bRa.
Transitivity: If aRb and bRc then $a - b = kn$ and $b - c = ln$ for integers k, l. So then $a - c = (a - b) + (b - c) = (k + l)n$, and hence aRc. □

3.3 Equivalence relations, equivalence classes, and partitions

Example 3.5 provides an example of a particularly important type of relation, an **equivalence relation**. An equivalence relation provides a way of saying two objects are, in some particular sense, 'the same':

Definition 3.6. A relation R on a set S is an **equivalence relation** if it is reflexive, symmetric and transitive. If R is an equivalence relation, we denote it by $\sim$ (various other symbols, including $\equiv$, are sometimes used).

Example 3.7. The following are all examples of equivalence relations:

(i) $S = \mathbb{C}$, with $z \sim w \Leftrightarrow |z| = |w|$;
(ii) S is the set of polygons in $\mathbb{R}^2$, and $\sim$ is congruence;
(iii) S is the set of differentiable functions on $\mathbb{R}$, and $f \sim g \Leftrightarrow f'(x) = g'(x)$;
(iv) The relation on $\mathbb{Z}$ defined in Example 3.5; in this case $\sim$ represents **congruence** modulo n. It is the basis for modular arithmetic, and you may often see $a \sim b$ expressed as '$a = b \bmod n$' in this case.

The following example is quite an important equivalence relation that you will see in Linear Algebra (ignore this example if you do not yet know about matrices):

Example 3.8. Let $S = M_n(\mathbb{R})$, the set of $n \times n$ matrices with real coefficients, and define $\sim$ on S by saying $A \sim B$ if and only if there exists an invertible matrix P such that $A = P^{-1}BP$. Then $\sim$ is an equivalence relation. This is called *similarity* of matrices; if $A \sim B$ then A and B are said to be *similar*.

Proof. Reflexivity: we see that $A \sim A$ for all A as $A = I^{-1}AI$, where I denotes the identity matrix.
Symmetry: if $A \sim B$ then $A = P^{-1}BP$ for some P and so $B = PAP^{-1} = (P^{-1})^{-1}AP^{-1}$ showing $B \sim A$.

Transitivity: if $A \sim B$ and $B \sim C$ then $A = P^{-1}BP$ and $B = Q^{-1}CQ$ for invertible P, Q. Then $A = P^{-1}Q^{-1}CQP = (QP)^{-1}C(QP)$, showing that $A \sim C$, since QP is invertible.

□

An equivalence relation provides a way of grouping together elements that can be viewed as being the same:

Definition 3.9. Given an equivalence relation $\sim$ on a set S, and given $x \in S$, the **equivalence class** of x, denoted $\overline{x}$ (or sometimes $[x]$), is the subset

$$\overline{x} = \{y \in S : y \sim x\}.$$

For example, with the equivalence relation defined in Example 3.5 (congruence modulo n) the equivalence class of 1 is the set $\overline{1} = \{\ldots, -n+1, 1, n+1, 2n+1, \ldots\}$; that is, all the integers that are congruent to 1 modulo n.

Grouping the elements of a set into equivalence classes provides a **partition** of the set, which we define as follows:

Definition 3.10. A **partition** of a set S is a collection of subsets $\{A_i \subseteq S : i \in I\}$, where I is an indexing set, with the property that

(i) $A_i \neq \varnothing$ for all $i \in I$ (that is, all the subsets are non-empty),
(ii) $\bigcup_{i \in I} A_i = S$ (that is, every member of S lies in one of the subsets),
(iii) $A_i \cap A_j = \varnothing$ for every $i \neq j$ (that is, the subsets are disjoint).

The subsets are called the parts of the partition.

For example, $\{\{n \in \mathbb{N} : n \text{ is divisible by } 2\}, \{n \in \mathbb{N} : n+1 \text{ is divisible by } 2\}\}$ forms a partition of the natural numbers, into evens and odds.

The fact that the equivalence classes for any equivalence relation form a partition is something that will be proved in the Groups and Group Action course later in the year (though it is not particularly difficult and you may like to have a go at doing so). Conversely, we can use any given partition to define an equivalence relation, by saying that $x \sim y$ if and only if x and y are elements of the same part of the partition (you may like to check that indeed this definition satisfies the conditions to be an equivalence relation: reflexivity, symmetry and transitivity). Thus, there is a natural correspondence between equivalence relations and partitions of a set. For example:

Example 3.11. Suppose S is the set of students at Oxford. This set can be partitioned according to colleges; that is, the partitioning subsets are the sets of students at each college, which form a partition since (i) there are no colleges with no students, (ii) every student is a member of a college, and (iii) you can't be at more than one college (there may be some strange exceptions, but we'll ignore those). Then the equivalence relation $\sim$ induced by this partition says that $x \sim y$ if and only if x and y are at the same college.

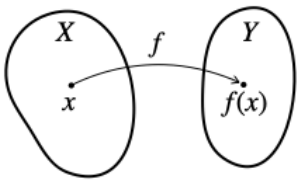
This is a good example of the fact that being 'equivalent' does not mean the elements are actually the same! Every one of you is wonderfully unique... but there might be some purpose for which it is convenient to view all the students in a college as essentially the same, and the equivalence class provides a way of representing that mathematically.

4 Functions

4.1 Definitions and examples

Definition 4.1. Let X and Y be sets. A **function** $f\colon X \to Y$ is an assignment of an element $f(x) \in Y$ for each $x \in X$. Functions are also referred to as **maps** or **mappings**. The set X is called the **domain** of f and the set Y is called the **codomain** of f.

So a function takes an input element from the set X, and maps it to an output in the set Y. The input is often referred to as the *argument* of the function (the word 'argument' has many different meanings!).



Sometimes the rule for how to convert input to output may involve some straightforward algebraic manipulation, such as the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. But a function need not be expressible in such convenient terms; the definition just requires that an output value is assigned for each possible input value (you may have come across the term 'black box', which would describe a function where the details of how input is converted to output are something of a mystery).

Example 4.2. We can define a function $f\colon \mathbb{R} \to \mathbb{Z}$ that takes any real number x and returns the least integer that is greater than or equal to x. (This is called the **ceiling** function, and is denoted by $\lceil x \rceil$ - you may be familiar with it as 'rounding up'; there is a related **floor** function, $\lfloor x \rfloor$).

Remark 4.3. The definition requires that a unique element of the codomain is assigned for every element of the domain, so our 'recipe' for defining a function needs to take account of this. For example if we want to define a function $f\colon \mathbb{R} \to \mathbb{R}$, the assignment $f(x) = 1/x$ is not sufficient, since it fails at $x = 0$. Similarly, the recipe '$f(x) = y$ where $y^2 = x$' fails for two reasons: one is that $f(x)$ is undefined for $x < 0$, and the other is that for $x > 0$ it does not return a unique value - does $f(4)$ equal 2 or -2? In such cases, we say the the function is **ill-defined**. We are interested in the opposite; functions that are **well-defined**. When we start to define more complicated functions, it may not be so immediately obvious whether this is the case, so some effort is often required to demonstrate that a given recipe produces a well-defined function.

Another example of what might seem to be a 'function', but which is not well-defined (and therefore not actually a function), is if we try to let $f\colon \mathbb{Q} \to \mathbb{Z}$ be given by $f(m/n) = n$. The problem here is that there is not a unique way of expressing an element of $\mathbb{Q}$ as m/n, so this assignment gives multiple different values for the same argument (for example $f(2/3) = 3$ and $f(4/6) = 6$, but to be well-defined we need these to be the same).

Remark 4.4. You may come across the slightly confusing term 'multi-valued function' or 'multifunction' to describe something that is similar to a function, but which assigns multiple values of $y \in Y$ as the 'output' (for example, $x^{1/2}$ or $\tan^{-1} x$ might be described in this way). These are not functions. But you could think of them as a family of functions, and you can produce a well-defined function by specifying which of the multiple outputs to take (for example, by restricting the codomain).

Example 4.5. As the remarks above highlight, the definition of a function needs to make clear the domain and codomain, not just the 'formula'. The following are all different functions:

$$\begin{aligned} f_1\colon \mathbb{R} \to \mathbb{R} \qquad &\text{given by} \qquad f_1(x) = x^2. \\ f_2\colon \mathbb{R} \to [0,\infty) \qquad &\text{given by} \qquad f_2(x) = x^2. \\ f_3\colon [0,\infty) \to \mathbb{R} \qquad &\text{given by} \qquad f_3(x) = x^2. \\ f_4\colon [0,\infty) \to [0,\infty) \qquad &\text{given by} \qquad f_4(x) = x^2. \end{aligned}$$

Definition 4.6. Given a function $f\colon X \to Y$, the **image** or **range** of f is

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

More generally, given $A \subseteq X$, the **image** of A (under f) is

$$f(A) = \{f(x) : x \in A\} \subseteq Y.$$

Given $B \subseteq Y$, the **pre-image** of B (under f) is

$$f^{-1}(B) = \{x : f(x) \in B\} \subseteq X.$$

For example, for the function f_1 defined above in Example 4.5, the image is $[0,\infty)$, and $f_1([0,1]) = [0,1]$, $f_1^{-1}([0,1]) = [-1,1]$, and $f_1^{-1}((-\infty,0]) = \{0\}$. For f_4, the image is $[0,\infty)$, and $f_4([0,1]) = [0,1]$, $f_4^{-1}([0,1]) = [0,1]$, and $f_4^{-1}((-\infty,0])$ is not defined, since $(-\infty,0]$ is not a subset of the codomain in this case.

Remark 4.7. Beware the potentially confusing notation: for $x \in X$, $f(x)$ is a single *element* of Y, but for $A \subseteq X$, $f(A)$ is a *set* (a subset of Y). Note also that $f^{-1}(B)$ should be read as 'the pre-image of B' and not as 'f-inverse of B'; the pre-image is defined even if no inverse function exists (in which case f^{-1} on its own has no meaning; we discuss invertibility of a function below).

If a function is defined on some larger domain than we care about, it may be helpful to restrict the domain:

Definition 4.8. Given a function $f\colon X \to Y$ and a subset $A \subseteq X$, the **restriction** of f to A is the map $f|_A\colon A \to Y$ defined by $f|_A(x) = f(x)$ for all $x \in A$.

The restriction is almost the same function as the original f - just the domain has changed.

Another rather trivial but nevertheless important function is the identity map:

Definition 4.9. Given a set X, the **identity** $\mathrm{id}_X\colon X \to X$ is defined by $\mathrm{id}_X(x) = x$ for all $x \in X$. Other notation is sometimes used, such as 1_X, and if the domain is unambiguous, the subscript may be removed.

4.2 Injectivity and surjectivity

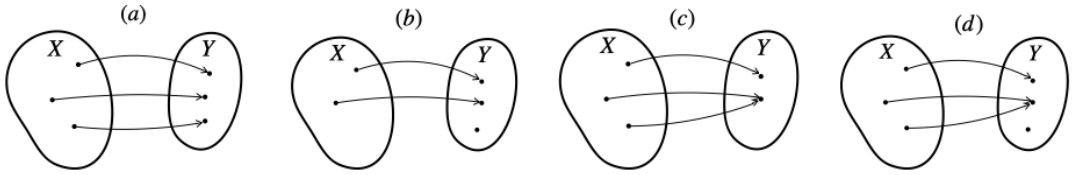
Definition 4.10. Let $f\colon X \to Y$ be a function.

(i) We say that f is **injective**, or **one-to-one** (1-1), if whenever $f(x_1) = f(x_2)$ then $x_1 = x_2$.

(ii) We say that f is **surjective**, or **onto**, if for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$.

(iii) We say that f is **bijective** if it is both injective and surjective. A bijective function is termed a **bijection**.

The figure below shows schematic representations of functions $f\colon X \to Y$ that are (a) both injective and surjective, (b) injective but not surjective, (c) surjective but not injective, (d) neither injective nor surjective:



Example 4.11. The ceiling function $\lceil x \rceil$ defined in Example 4.2 is surjective, because for any integer $n \in \mathbb{Z}$, we have $\lceil n \rceil = n$. But it is not injective, because there are other elements in $\mathbb{R}$ (infinitely many of them in fact) that will return the same output; for example $\lceil n - \frac{1}{2} \rceil = n$.

Example 4.12. Returning to the functions defined in Example 4.5, we see that the function f_1 is not injective (because, for example, $f(-1) = f(1) = 1$), and it is not surjective (because there are no $x \in \mathbb{R}$ that give $x^2 = y$ for $y < 0$). The function f_2 is also not injective but *is* surjective (because $y < 0$ is not in the codomain this time, and for every $y \in [0, \infty)$ there *is* some $x \in \mathbb{R}$ with $x^2 = y$). The function f_3 is again not surjective, but this time it is injective (because negative values are now excluded from the domain). The function f_4 is both injective and surjective (and is therefore a bijection).

Remark 4.13. For real-valued functions $f\colon X \to Y$, where $X, Y \subseteq \mathbb{R}$, thinking about the graph of the function can be helpful. An injective function will have a graph that is **monotonic** (either never decreasing or never increasing).

Having introduced injective and surjective functions, we can give an alternative and more intuitive definition of the cardinality of finite sets:

Definition 4.14. The empty set $\varnothing$ is finite and has cardinality $|\varnothing| = 0$. A non-empty set S is said to finite and have cardinality $|S| = n \in \mathbb{N}$ if and only if there exists a bijection from S to the set $\{1, 2, \dots n\}$.

The bijection provides a way of counting the elements of S. You might like to convince yourself that this definition is equivalent to the inductive definition given earlier.

Note that for finite sets X and Y, a function $f\colon X \to Y$ can only be injective if $|Y| \geq |X|$, since for any injective function the number of elements in the image $f(X)$, is equal to the number of elements in the domain, and $f(X) \subseteq Y$. In other words, the codomain of an injective function cannot be smaller than the domain. This is sometimes referred to as the **pigeonhole principle** (so called from the observation that if n letters are placed in m pigeonholes and $n > m$, then at least one hole must contain more than one letter; the non-injective function in that case is the assignment of pigeonholes to letters[1]).

Similarly, a function $f\colon X \to Y$ can only be surjective if $|Y| \leq |X|$. Hence if f is bijective, then $|X| = |Y|$; that is, the domain and codomain of a bijection have equal cardinality. (These results hold true for infinite sets too, though less obviously).

4.3 Composition of functions and invertibility

Definition 4.15. Given two functions $f\colon X \to Y$ and $g\colon Y \to Z$, the **composition** $g \circ f\colon X \to Z$ is defined by

$$(g \circ f)(x) = g(f(x)) \qquad \text{for all } x \in X.$$

If $Z = X$ then we can similarly define $f \circ g\colon Y \to Y$, but in general $f \circ g \neq g \circ f$ (indeed, if $X \neq Y$, these functions have different domain and codomain, but even in the case $X = Y$ the functions will generally not be the same). For example, if $f(x) = x^2$ and $g(x) = e^x$ are both maps from $\mathbb{R}$ to $\mathbb{R}$, then

$$(f \circ g)(x) = e^{2x} \neq e^{x^2} = (g \circ f)(x).$$

This shows that composition of functions is not *commutative*. However, composition is *associative*, as the following results shows:

Proposition 4.16. *Let $f\colon X \to Y$, $g\colon Y \to Z$, $h\colon Z \to W$ be three functions. Then*

$$f \circ (g \circ h) = (f \circ g) \circ h.$$

[1] College porters are no longer afforded this mathematical insight, since so few letters are sent. The same problem could occur with rainforest-sourced parcels, of which there are sometimes so many that they may well exceed the number of pigeonholes; but since they usually don't fit in the pigeonholes, the analogy rather breaks down... we might say that the the mapping from parcels to pigeonholes is ill-defined in that case, since it fails to properly assign a pigeonhole to each parcel.

Proof. Let $x \in X$. Then, by the definition of composition, we have

$$(f \circ (g \circ h))(x) = f((g \circ h)(x)) = f(g(h(x))) = (f \circ g)(h(x)) = ((f \circ g) \circ h)(x).$$

□

The following proposition addresses the extent to which composition of functions preserves injectivity and surjectivity:

Proposition 4.17. *Let $f\colon X \to Y$ and $g\colon Y \to Z$ be functions.*

(i) If f and g are injective then so is $g \circ f$. Conversely, if $g \circ f$ is injective, then f is injective, but g need not be.

(ii) If f and g are surjective then so is $g \circ f$. Conversely, if $g \circ f$ is surjective, then g is surjective, but f need not be.

Proof (+commentary). We prove (i), and leave the proof of (ii) as an exercise.

It is helpful to clarify for each part of the proposition what exactly we are told (the hypotheses), and what exactly we need to show. For the first part of (i), we can take it that f and g are injective, and need to show that $g \circ f$ is injective. From the definition of injectivity, that means we need to show that for any $x_1, x_2 \in X$, if $(g \circ f)(x_1) = (g \circ f)(x_2)$ then $x_1 = x_2$. So let's suppose $x_1, x_2 \in X$ and $(g \circ f)(x_1) = (g \circ f)(x_2)$, and aim to show $x_1 = x_2$, making use of what we know. From the injectivity of g we know that if $g(f(x_1)) = g(f(x_2))$ then $f(x_1) = f(x_2)$, so this must be the case here. Then from the injectivity of f we know that this means $x_1 = x_2$. So we have indeed shown what is needed for $g \circ f$ to be injective.

For the second part of (i), we are told that $g \circ f$ is injective, and we need to show that f is injective; that is, we need to show that if $f(x_1) = f(x_2)$ then $x_1 = x_2$. So let's suppose that $f(x_1) = f(x_2)$ and aim to show this, making use of what we know about $g \circ f$ this time. Applying g to both sides gives $g(f(x_1)) = g(f(x_2))$, and then we see that the injectivity of $g \circ f$ immediately tells us that $x_1 = x_2$. So we have shown that f is injective.

An alternative approach here could have been to use contradiction. If we start with the supposition that f is *not* injective, then it means there exist some $x_1, x_2 \in X$ for which $x_1 \neq x_2$ but $f(x_1) = f(x_2)$. Then $g(f(x_1)) = g(f(x_2))$, so in fact this means there exist some $x_1, x_2 \in X$ for which $x_1 \neq x_2$ but $(g \circ f)(x_1) = (g \circ f)(x_2)$. But that would contradict the definition of $g \circ f$ being injective, so our supposition that f was not injective was incorrect, and we have therefore shown that f *is* injective.

To show that g need not be injective, we should give a counterexample. A bit of thought may lead to the observation that g could have a larger domain than the image of f. An extreme example is to take $X = Z = \{0\}$ and $Y = \mathbb{R}$, and have f and g defined by $f(0) = 0$ and $g(y) = 0$ for all $y \in \mathbb{R}$. Then $(g \circ f)\colon X \to X$ is injective (it simply maps 0 to 0). But clearly g is not injective. □

I have written much more than is needed in the proof above because I am spelling out the thought process as well as the logic. Having worked out what to do, we could streamline it:

Proof. For the first part of (i), suppose $(g \circ f)(x_1) = (g \circ f)(x_2)$ for some $x_1, x_2 \in X$. From the injectivity of g we know that $g(f(x_1)) = g(f(x_2))$ implies $f(x_1) = f(x_2)$, and then from the injectivity of f we know that this implies $x_1 = x_2$. So $g \circ f$ is injective.

For the second part of (i), suppose $f(x_1) = f(x_2)$ for some $x_1, x_2 \in X$. Then applying g gives $g(f(x_1)) = g(f(x_2))$, and by the injectivity of $g \circ f$ this means $x_1 = x_2$. So f is injective. To see that g need not be injective, a counterexample is $X = Z = \{0\}$, $Y = \mathbb{R}$, with $f(0) = 0$ and $g(y) = 0$ for all $y \in \mathbb{R}$. □

Recalling that id_X is the identity map on a set X, we are now in a position to define invertibility:

Definition 4.18. A function $f\colon X \to Y$ is **invertible** if there exists a function $g\colon Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. The function g is the **inverse** of f, and we write $g = f^{-1}$.

Note that directly from the definition, if f is invertible then f^{-1} is also invertible, and $(f^{-1})^{-1} = f$.

An immediate concern we might have is whether there could be multiple such functions g, in which case the inverse f^{-1} would not be well-defined. This is resolved by the following result:

Proposition 4.19. *If $f\colon X \to Y$ is invertible then its inverse is unique.*

Proof. Let g_1 and g_2 be two functions for which $g_i \circ f = \mathrm{id}_X$ and $f \circ g_i = \mathrm{id}_Y$. Using the fact that composition is associative, and the definition of the identity maps, we can write

$$g_1 = g_1 \circ \mathrm{id}_Y = g_1 \circ (f \circ g_2) = (g_1 \circ f) \circ g_2 = \mathrm{id}_X \circ g_2 = g_2.$$

□

The following result shows how to invert the composition of invertible functions:

Proposition 4.20. *Let $f\colon X \to Y$ and $g\colon Y \to Z$ be functions between sets X, Y, Z. If f and g are invertible, then $g \circ f$ is invertible, and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.*

Proof. Making repeated use of the fact that function composition is associative, and the definition of the inverses f^{-1} and g^{-1}, we note that

$$\begin{aligned}(f^{-1} \circ g^{-1}) \circ (g \circ f) &= \big((f^{-1} \circ g^{-1}) \circ g\big) \circ f = \big(f^{-1} \circ (g^{-1} \circ g)\big) \circ f \\ &= \big(f^{-1} \circ \mathrm{id}_Y\big) \circ f = f^{-1} \circ f = \mathrm{id}_X,\end{aligned}$$

and similarly,

$$\begin{aligned}(g \circ f) \circ (f^{-1} \circ g^{-1}) &= g \circ \big(f \circ (f^{-1} \circ g^{-1})\big) = g \circ \big((f \circ f^{-1}) \circ g^{-1}\big) \\ &= g \circ \big(\mathrm{id}_Y \circ g^{-1}\big) = g \circ g^{-1} = \mathrm{id}_Z,\end{aligned}$$

which shows that $f^{-1} \circ g^{-1}$ satisfies the properties required to be the inverse of $g \circ f$. □

The following result provides an important and useful criterion for invertibility:

Theorem 4.21. *A function $f\colon X \to Y$ is invertible if and only if it is bijective.*

Proof. First suppose f is invertible, so there it has an inverse $f^{-1}\colon Y \to X$. To show f is injective, suppose that for some $x_1, x_2 \in X$ we have $f(x_1) = f(x_2)$. Then applying f^{-1} to both sides and noting that by definition $f^{-1} \circ f = \mathrm{id}_X$, we see that $x_1 = f^{-1}(f(x_1)) = f^{-1}(f(x_2)) = x_2$. So f is injective. To show that f is surjective, let $y \in Y$, and note that $f^{-1}(y) \in X$ has the property that $f(f^{-1}(y)) = y$. So f is surjective. Therefore f is bijective.

Conversely, suppose that f is bijective. We aim to show that there is a well-defined $g\colon Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. Since f is surjective, we know that for any $y \in Y$, there is an $x \in X$ such that $f(x) = y$. Furthermore, since f is injective, we know that this x is unique. So for each $y \in Y$ there is a unique $x \in X$ such that $f(x) = y$. This recipe provides a well-defined function $g(y) = x$, for which we have $g(f(x)) = x$ for any $x \in X$ and $f(g(y)) = y$ for any $y \in Y$. So g satisfies the property required to be an inverse of f and therefore f is invertible. □

It is also possible to define left-inverse and right-inverse functions as functions that partially satisfy the definition of the inverse:

Definition 4.22. A function $f\colon X \to Y$ is **left invertible** if there exists a function $g\colon Y \to X$ such that $g \circ f = \mathrm{id}_X$, and is **right invertible** if there exists a function $h\colon Y \to X$ such that $f \circ h = \mathrm{id}_Y$.

As may be somewhat apparent from the previous proof, being left- and right-invertible is equivalent to being injective and surjective, respectively. We leave this as an exercise to show.

5 Mathematical reasoning and logic

5.1 Logical statements and notation

We have already seen that we must deal with a lot of precise statements and assertions, ranging from very simple ones like '$n = 2$' to more involved ones like 'for all $x \in \mathbb{R}, x^2 \geq 0$', or 'there exist $x, y, z \in \mathbb{N}\backslash\{0\}$ such that $x^{2022} + y^{2022} = z^{2022}$'. (You will see much more complicated examples too!) A lot of mathematical reasoning comes from careful manipulation of such statements, perhaps re-expressing them in a different way, or observing that they imply some other statement. The statements can just as well be true or false (perhaps you recognise the final example just given, which is famously false, by Fermat's last Theorem). Often during a mathematical argument we might not (yet) know if a given statement is true or false. Nevertheless we can work with it; following through the logical consequences of a statement may eventually lead us to the conclusion that it is false, for example.

To discuss such logic it is helpful to denote the statements by symbols; we have already seen examples of this. If P is the statement '$x \geq 2$' and Q is '$x^2 \geq 4$' (in

the context of $\mathbb{R}$), we can then say things like 'P implies Q' (which is, itself, another logical statement - one that we know to be true in this case). If the statement depends on a variable, it is sometimes helpful (depending on the context) to write things like $P(x)$ to clarify this; the example just given is really $P(x)$, and is true or false depending on the value of x (nevertheless, 'P implies Q' is a statement that does not depend on x in this example, since it is true regardless of the value of x).

We can combine logical statements using connecting words like 'and', and 'or', and we can negate a statement P by writing 'not P'. Symbols are sometimes used for these: $P \wedge Q$ means 'P **and** Q', while $P \vee Q$ means 'P **or** Q', and $\neg P$ is the same as '**not** P'.

For example, if P is '$n = 2$' and Q is 'n is even' (in the context of $\mathbb{N}$), then '$P \vee Q$' is equivalent to 'n is even', '$P \wedge Q$' means '$n = 2$', and $\neg P$ is the same as '$n \neq 2$'.

Remark 5.1. In regular English, the word 'or' is often interpreted as an *exclusive* or; that is, it may carry an implicit meaning of 'one but not the other' (as in 'you can have a piece of cake or an ice cream'). This is not the case in mathematical usage, where 'P or Q' should be interpreted to mean that P holds or Q holds or both do. If we mean to use an 'exclusive or', we should say something extra to indicate that (like '... but not both'). There are a number of less standard symbols that are sometimes used for an exclusive or, but we could also do it by writing $(P \wedge \neg Q) \vee (\neg P \wedge Q)$ (convince yourself this says the right thing).

There is a direct analogy between the symbols $\vee$, $\wedge$ and $\neg$, and the symbols $\cup$, $\cap$, and ^{c} for set union, intersection and complement. To make this analogy clear, suppose $A, B \subseteq S$ are sets and let P and Q be the statements '$x \in A$' and '$x \in B$'. Then $x \in A \cup B$ is clearly equivalent to $P \vee Q$, $x \in A \cap B$ is equivalent to $P \wedge Q$, and $x \in A^c$ is equivalent to $\neg P$.

These logical symbols therefore obey the same distributive laws as for sets (Proposition 2.18) and also De Morgan's laws (Proposition 2.20), which in this context are

$$\neg(P \vee Q) \;\Leftrightarrow\; (\neg P) \wedge (\neg Q) \qquad \text{and} \qquad \neg(P \wedge Q) \;\Leftrightarrow\; (\neg P) \vee (\neg Q).$$

These rules for how to negate 'P or Q' or 'P and Q' are hopefully quite intuitive (if we don't have P *or* Q holding then that means we don't have P holding *and* we don't have Q holding). But when it comes to more complicated statements it is easy to confuse ourselves, so having these clear rules to fall back on may be useful.

Remark 5.2. In regular English, we use the phrasing 'neither... nor...' to negate 'either... or...'. So the opposite of 'either P or Q' is 'neither P nor Q'. The word 'nor' may be a bit confusing because this really means 'not P *and* not Q'.

We often use the shorthand $P \Rightarrow Q$ to mean 'P **implies** Q'. This means that if P holds then Q also holds. It is equivalent to saying '**If** P **then** Q', or to saying that 'P is sufficient for Q'. An important thing to note is that it does not mean it is *necessary* for P to hold in order for Q to hold. If P is '$n = 2$' and Q is 'n is even', then $P \Rightarrow Q$, but $Q \not\Rightarrow P$.

We can write $P \Leftrightarrow Q$ to mean $P \Rightarrow Q$ *and* $Q \Rightarrow P$. We can read this as 'P **if and only if** Q', or 'P is necessary and sufficient for Q', or 'P is equivalent to Q'. Some people don't like using this notation, and its use in a proof needs some caution, as we'll discuss later (it has a logical implication both forwards and backwards, whereas our thought process tends to work in one direction most of the time!). The letters 'iff' are also commonly used to stand for 'if and only if'.

Remark 5.3. In regular English usage, 'implies' or 'if ... then...' tends to be understood to indicate a degree of *causation*. So if I say X implies Y I would usually mean that X had something to do with Y ('if you had not fallen asleep then you would have got more out of this lecture'). In mathematical usage this does not need to be the case (although it usually *is* the case for most useful statements). So if we say '$P \Rightarrow Q$' or 'If P then Q', we simply mean that whenever P is true Q is also true. So 'If Paris is the capital of France then the Thames flows through London' is a true statement, despite the fact that there is obviously no connection between these two facts. Similarly, 'If Oxford is on Mars then Cambridge is on Venus' is also a true statement (because, since the 'P' in this case is *never* true, it actually does not matter what we say afterwards - the statement '$P \Rightarrow Q$' will still be true. Such a statement is, however, completely useless). A statement like this, where the 'P' is never true, is said to be *vacuously true*. A similarly useless statement might begin 'for all $x \in \varnothing$, ...'.

The symbol $\forall$ denotes '**for all**' or 'for every', and can simply be used as shorthand for those words. The symbol $\exists$ denotes '**there exists**', and can similarly replace those words. Typically a phrase like 'there exists $x \in S$' is followed by a statement $P(x)$ about what specific property x has (otherwise it doesn't tell us anything other than that S is non-empty), and it is quite common for $\exists$ to stand for the following 'such that' as well as the 'there exists'. For example, '$\exists x \in S \quad P(x)$' can be read as 'there exists x in S such that $P(x)$ holds'. Personally I prefer to include the letters 's.t.' (or a colon ':') to stand in place of the 'such that', so I write '$\exists x \in S$ s.t. $P(x)$'. The symbols $\forall$ and $\exists$ are known as **quantifiers** (*universal* and *existential* quantifiers, respectively). A variant that is also common is $\exists!$ which means 'there exists unique', implying that there is one, and only one, element with the given property.

Using these symbols we can write things like

$$\forall x \in \mathbb{R} \quad x^2 \geq 0 \quad \text{but} \quad \exists x \in \mathbb{C} \text{ s.t. } x^2 < 0,$$

which you should read as 'for all x in the real numbers, x^2 is greater than or equal to zero, but there exists x in the complex numbers such that x^2 is less than zero'. Similarly, the definitions of what it means for $f\colon X \to Y$ to be injective and surjective can be written as

$$\forall x_1, x_2 \in X,\ f(x_1) = f(x_2) \Rightarrow x_1 = x_2,$$

and

$$\forall y \in Y,\ \exists x \in X \text{ s.t. } f(x) = y.$$

It is helpful to practice 'translating' such statements into English sentences, and vice versa.

Using the logical notation introduced in this section can often help make a proof more concise. Here is an alternative proof of the double inclusion principle:

Proof of Proposition 2.17. We argue, via a sequence of equivalent statements, that $A = B$ is the same as ($A \subseteq B$ and $B \subseteq A$):

$$\begin{aligned} A = B \quad &\Leftrightarrow \quad \forall x \in S \quad (x \in A \Leftrightarrow x \in B) \\ &\Leftrightarrow \quad \forall x \in S \quad (x \in A \Rightarrow x \in B \quad \text{and} \quad x \in B \Rightarrow x \in A) \\ &\Leftrightarrow \quad \forall x \in S \quad (x \in A \Rightarrow x \in B) \quad \text{and} \quad \forall x \in S \quad (x \in B \Rightarrow x \in A) \\ &\Leftrightarrow \quad A \subseteq B \quad \text{and} \quad B \subseteq A. \end{aligned}$$

□

Notice that bracketing is sometimes necessary in order to make clear what the statements are to which the quantifiers refer. Intelligent use of spacing on the page can also be helpful, especially when writing by hand.

To use $\Leftrightarrow$ like this we need to make sure that the logic works both ways, both forwards and backwards. When constructing such a proof, it can be helpful to first work forwards, with each statement implying the next one, and then separately check whether the argument works backwards (i.e. first write it with the arrows as $\Rightarrow$, before checking if each one can be converted to $\Leftrightarrow$).

Remark 5.4. It is to some extent a matter of personal taste how much to use symbolic notation rather than writing things out in words. Personally I never use the symbols $\vee$, $\wedge$, and $\neg$ but rather write out 'and', 'or', and 'not'. I often use $\forall$ and $\exists$ but sometimes write these out in words too; it depends on the circumstances. Regardless of how much you use them in your own writing, it is important to understand and be fluent in interpreting these symbols in other people's writing.

5.2 Handling logical statements

Below we make some comments about how to handle different types of statements. You may find it helpful to re-read this section later in the year, once you have more experience of seeing and using them yourself.

If, only if, $\Rightarrow$

Statements of this form are probably the most common, although they may sometimes appear quite differently. The following all mean the same thing:

(i) if P then Q;
(ii) P implies Q;
(iii) $P \Rightarrow Q$;
(iv) P only if Q;

(v) P is a sufficient condition for Q;
(vi) Q is a necessary condition for P;
(vii) whenever P holds, Q also holds;
(viii) if Q does not hold then P does not hold;
(ix) not Q implies not P;
(x) $\neg Q \Rightarrow \neg P$.

The last three of these are known as the **contrapositive**.

In order to prove a statement of this form, we typically start by assuming that P holds and try to deduce through some logical steps that Q holds too. Alternatively, we can start by assuming that Q does not hold and show that P does not hold (that is, we prove the contrapositive).

Remark 5.5. Note that the contrapositive is not the same as the **converse**. The contrapositive of 'if P then Q' is 'if not Q then not P', and it is simply a different way of stating exactly the same thing. But the *converse* of 'if P then Q' is 'if Q then P', which means something completely different. (The *negation* of 'if P then Q' is something different again; that simply means that 'if P then Q' is not true.)

If using $\Rightarrow$, note that the symbol stands for both the 'if' and the 'then'. We shouldn't use it to stand for just the 'then' as, for example, in 'if $x = -1 \Rightarrow x^2 = 1$'. This would mean 'if $x = -1$ implies that $x^2 = 1$' and would need to be followed by 'then ...' (to match the 'if'). (Since indeed $x = -1$ does imply $x^2 = 1$, whatever follows would always be true and the whole phrase would serve no purpose in this case).

If and only if, iff, $\Leftrightarrow$

These statements are usually best thought of separately as 'if' and 'only if' statements. So to prove 'P if and only if Q', we should first prove 'if P then Q', and then separately prove 'if Q then P' (or vice versa). Sometimes we may find that essentially the same argument used for the first direction also works in reverse, but sometimes quite a different method of argument may be required. One thing to be wary of is that having *assumed* P to deduce Q, and then having changed to assuming Q with a view to *deducing* P, it is all to easy to keep making use of P (or parts of P), forgetting that that is no longer assumed. It is a good idea to make very clear, both to yourself and in your written proof, which direction you are doing.

Because the logical flow of an argument is usually followed in one direction, the $\Leftrightarrow$ symbol is best used with some caution, making sure that both directions really do work. In some instances however, particularly when P and Q are more obviously related, it may be easier to think of 'if and only if' as 'is equivalent to', rather than splitting into the two directions. In such situations, using $\Leftrightarrow$ may be an efficient way of presenting the argument. The proof of Proposition 2.17 in the previous section is an example of this.

Quantifiers

The quantifiers $\forall$ and $\exists$ are probably the most challenging of the notation introduced in this section. It is a good idea to practice reading statements that include these symbols and checking that you understand their meaning. For example, if $\mathbb{P}$ is the set of prime numbers, then

$$\forall p \in \mathbb{P}, \text{ if } p > 2 \text{ then } p \text{ is odd},$$

is a way of stating 'every prime number greater than 2 is odd', and

$$\forall x \in \mathbb{R} \quad \big(\, x < 0 \quad \text{or} \quad \exists y \in \mathbb{R} \text{ s.t. } y^2 = x \, \big),$$

is a way of stating 'every non-negative real number has a real square root'.

The quantifiers should include a specification of the set over which they range ($\mathbb{P}$ or $\mathbb{R}$ in these examples). However, there are situations when this is so obvious from the context that it becomes cumbersome to keep writing this. In particular, you'll often see '$\forall \varepsilon > 0$', in which it is understood that ε is a real number.

To prove a statement of the form '$\forall x \in X \ P(x)$', it is always a good idea to start the proof with 'Let $x \in X$.' or 'Suppose $x \in X$ is given.' This 'addresses' the quantifier with an arbitrary x, which should then be treated as fixed for the rest of the proof. Provided no other assumptions about x are made during the course of proving $P(x)$, this will prove the statement for all $x \in X$.

To prove a statement of the form '$\exists x \in X$ s.t. $P(x)$', there is not such a clear steer about how to start. Somehow you need to show the existence of an x with the right properties. It might be that you can simply spot one. It could be that you're able to demonstrate logically that such an x must exist because of some earlier assumption, or it may be that you can show 'constructively' how to find one. Or you may be able to try a proof by contradiction, supposing that there is no such x and consequently arriving at some inconsistency.

It is important to note that the order of quantifiers matters. Returning to an earlier example, if S is the set of students at Oxford, and C is the set of colleges, we can say

$$\forall s \in S, \ \exists c \in C \text{ s.t. } s \in c,$$

which says that for every student there is a college of which they are a member; this is true. If we changed the order of the quantifiers and wrote

$$\exists c \in C \text{ s.t. } \forall s \in S, \ s \in c,$$

this says that there is one particular college of which every student at Oxford is a member; that is a completely different statement, and it is not true.

Importantly, we must read from left to right, and as new elements or statements are introduced they are allowed to depend on previously introduced elements but can't depend on things that are yet to be mentioned. So in the first of the statements above, the $c \in C$ that exists according to the second quantifier can (and does) depend on the specific $s \in S$ from the first quantifier. In the second statement, the specific c identified by the first quantifier needs to work for all s in the second one.

One reason why this kind of logical notation is so helpful is that English itself is sometimes ambiguous; we rely on context and common sense in order to parse certain phrases. For example, the statement

'For all natural numbers x, $x < y$ for some natural number y'

could mean 'for all $x \in \mathbb{N}$, there exists $y \in \mathbb{N}$ such that $x < y$', or 'there exists $y \in \mathbb{N}$ such that for all $x \in \mathbb{N}$, $x < y$'. The English wording could justifiably be interpreted either way, but clearly only the first of these is true. In symbolic notation, we should write

$$\forall x \in \mathbb{N},\ \exists y \in \mathbb{N} \text{ s.t. } x < y,$$

and it is unambiguous that y is allowed to depend on x.

Remark 5.6. To avoid confusion, it is a good idea to keep to the convention that the quantifiers come first, before any statement to which they relate. However, many authors (including this one!) don't stick rigidly to this if there's a last 'for all'. For example, if $f : \mathbb{R} \to \mathbb{R}$ is a bounded function, you may see something like

$$\exists M \in \mathbb{R} \text{ s.t. } |f(x)| < M\ \forall x \in \mathbb{R}.$$

Negation

For any statement P, the negated statement 'not P' is the statement that is false when P is true, and true when P is false. It is important to be adept at negating statements (in order to seek contradictions, for example). For a simple statement like '$x \in S$', the negation is straightforward: '$x \notin S$'. For more involved statements, it can be more confusing.

Firstly, if the statement is of the form '$P \Rightarrow Q$' then the negated statement is '$P \not\Rightarrow Q$'. Since $P \Rightarrow Q$ means that Q is true whenever P is true, $P \not\Rightarrow Q$ means that (at least in some circumstance) P is true and Q is not true. Proving $P \not\Rightarrow Q$ would typically involve demonstrating such a circumstance.

Secondly, if the statement involves quantifiers, we should note that the negation of '$\forall x \in X,\ P(x)$' is '$\exists x \in X$ s.t. not $P(x)$' (since, if it is not true that $P(x)$ holds for every x, then it must be the case that there is some x for which $P(x)$ does not hold). Similarly, the negation of '$\exists x \in X$ s.t. $P(x)$' is '$\forall x \in X,$ not $P(x)$' (since, if it is not true that there is an x for which $P(x)$ holds, then it means that $P(x)$ does not hold for any x). So the rule for negating statements that involve quantifiers is that we can move the negation 'through' the quantifiers, provided that we change $\forall$ to $\exists$ and $\exists$ to $\forall$ (this is essentially an instance of De Morgan's laws that we discussed earlier).

Recalling the previous example of students and colleges, the negation of

$$\forall s \in S,\ \exists c \in C \text{ s.t. } s \in c,$$

is

$$\exists s \in S \text{ s.t. } \forall c \in C,\ s \notin c.$$

So if we wanted to disprove the original statement, this is what we would need to show; that is, we would need to find a student who is not a member of any college.

5.3 Formulation of mathematical statements

In order to prove or use a theorem it is important to correctly understand its logical form. In fact this is the case with most mathematical problems we may want to solve.

Most theorems are ultimately of the form 'if P then Q', although the P and the Q may themselves be quite complicated statements that are combinations of other statements. In this context, the 'P' is the **hypothesis** and the 'Q' is the **conclusion**. (The word 'hypothesis' has different uses; its meaning here differs from its regular English usage as a conjecture that warrants further investigation. The latter is confusingly also used in mathematics, as in the 'Riemann hypothesis', for example.) As an example, consider:

> If n is a non-zero natural number, then n has a unique prime factorisation.

Here, the hypothesis is 'n is a non-zero natural number', and the conclusion is 'n has a unique prime factorisation'. In this example, the theorem is explicitly stated in the form 'if P then Q', so understanding the hypothesis and conclusions is very easy. Sometimes a theorem may be stated in a way that makes this less obvious. For example:

> Every prime number greater than 2 is odd.

Faced with such a statement, it may be helpful to think through carefully what the hypothesis and conclusion are, and to re-state it in a way that makes this more transparent. Thus, another way of saying the same thing is:

> Let p be a prime number greater than 2. Then p is odd.

The first sentence is the hypothesis, and the second is the conclusion.

As a more involved example, here is a rather poor statement of the intermediate value theorem (IVT), which you will come across in Analysis:

> Whenever f is a continuous function on $\mathbb{R}$, a, b are real numbers such that $a < b$, $f(a) < 0$ and $f(b) > 0$, $f(c) = 0$ for some $c \in (a, b)$.

Although all the correct ingredients of the theorem are here, it is not at all clear how to split the statement into hypothesis and conclusion. A better version is:

> Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous function, and suppose $a, b \in \mathbb{R}$ are such that $a < b$, $f(a) < 0$, and $f(b) > 0$. Then there exists a real number $c \in (a, b)$ such that $f(c) = 0$.

Splitting the theorem into shorter sentences has helped to clearly separate the hypothesis (which is a combination of various sub-statements in this case) from the conclusion. As a general rule, using words like 'Let' and 'Suppose' is a good way of 'setting up' the hypotheses, and 'Then' is a good way of signalling that what follows is the conclusion.

(The intermediate value theorem is stating the intuitively obvious result that if the graph of a continuous function is below the axis somewhere and above the axis somewhere else, then it must *cross* the axis somewhere in between).

6 Constructing proofs and problem solving

6.1 Methods of proof

We have already seen various methods for proving and refuting mathematical statements. We now make some comments on the general classes of methods. To illustrate the discussion, we'll consider the following simple result about the arithmetic and geometric means of non-negative real numbers:

Theorem 6.1 (AM-GM Inequality)**.** *Let x, y be non-negative real numbers. Then*

$$\sqrt{xy} \leq \tfrac{1}{2}(x+y),$$

with equality if and only if $x = y$. (The left hand side of this inequality is the geometric mean and the right hand side is the arithmetic mean.)

Direct proof

To prove a statement of the form 'if P then Q' directly, we make use of P to arrive at Q through a sequence of logical reasoning. It may be that we can start from P and work directly to Q, or it may be that we make use of P along the way (as in the example below). Sometimes some creativity and imagination may be required.

A direct proof of Theorem 6.1 could look like:

Proof of Theorem 6.1. Since $x, y \in \mathbb{R}$, we know that $(x-y)^2 \geq 0$, with equality if and only if $x = y$. Expanding the brackets, and then adding $4xy$ yields

$$\begin{aligned} x^2 - 2xy + y^2 &\geq 0, \\ x^2 + 2xy + y^2 &\geq 4xy, \\ \frac{1}{4}(x+y)^2 &\geq xy. \end{aligned}$$

Taking the square root (noting that $x, y \geq 0$) gives the required result. □

In this case, the algebraic steps are straightforward, but the starting point and the 'adding $4xy$' are perhaps not entirely obvious (having said that, this type of argument - starting with the fact that the square of a real number is non-negative - is surprisingly useful in many situations).

Proof by contradiction

To prove a statement 'if P then Q' by contradiction, we suppose that Q is not true and show through some logical reasoning (making use of the hypotheses P) that this leads to a contradiction or inconsistency. We may arrive at something that contradicts the hypotheses P, or something that contradicts the initial supposition that Q is not true, or we may arrive at something that we know to be universally false.

A proof by contradiction of Theorem 6.1 could look like:

Proof of Theorem 6.1. Suppose the opposite, that is $\sqrt{xy} > \frac{1}{2}(x+y)$. Then squaring both sides (noting that $x, y \geq 0$), and rearranging, yields

$$\begin{aligned} 4xy &> 4x^2 + 2xy + y^2, \\ 0 &> x^2 - 2xy + y^2, \\ 0 &> (x-y)^2, \end{aligned}$$

which is a contradiction, since the square of a real number is non-negative. Hence $\sqrt{xy} \leq \frac{1}{2}(x+y)$. The same steps with $>$ replaced by $=$ show that equality holds if and only if $x = y$. □

One of the useful aspects of proving things by contradiction is that by negating the statement Q, we immediately give ourselves something extra to work with. So if we cannot see a way to make any progress by starting directly from P, then supposing a contradiction may be a good thing to try instead. In the proof above, for example, no imagination was required; we simply started with the negation of the result we were aiming for and manipulated the statement to find the contradiction.

Proof by contradiction is sometimes referred to by the Latin phrase *reductio ad absurdum* ('reduction to an absurdity').

Proof by induction

We saw numerous examples of induction earlier on; it is useful for proving results that can be indexed by the natural numbers. So it would not be useful as a method to prove Theorem 6.1 itself, but it could be used to prove the generalisation:

Theorem 6.2. *Let $n \geq 2$. If $x_1, x_2, \ldots, x_n$ are non-negative real numbers, then*

$$(x_1 x_2 \ldots x_n)^{1/n} \leq \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Proof. We argue by induction, noting that the case $n = 2$ was proven (twice!) above. Suppose that the result holds for n numbers and consider the case for $n+1$ numbers, $x_1, x_2, \ldots, x_{n+1}$. To simplify the algebra, let $\overline{x}$ denote their arithmetic mean (so the inequality we are looking to show can be written as $x_1 x_2 \ldots x_{n+1} \leq \overline{x}^{n+1}$). By ordering the x_i appropriately we can assume, without loss of generality, that $x_{n+1} \geq \overline{x}$ and $x_n \leq \overline{x}$. Thus $(x_{n+1} - \overline{x})(\overline{x} - x_n) \geq 0$, which rearranges to give $x_n + x_{n+1} - \overline{x} \geq x_{n+1}x_n/\overline{x}$ (note that $\overline{x}$ is zero only in the case when all x_i are zero, in which case the result holds trivially). By the definition of $\overline{x}$, we have $(n+1)\overline{x} = x_1 + \ldots + x_n + x_{n+1}$, so rearranging and making use of the inequality just noted allows us to write

$$n\overline{x} \geq x_1 + \ldots + x_{n-1} + x_n x_{n+1}/\overline{x}.$$

The right hand side here is a sum of n numbers, on which we can apply the inductive hypothesis, to give

$$n\overline{x} \geq n(x_1 x_2 \ldots x_n x_{n+1}/\overline{x})^{1/n}.$$

Rearranging then gives the required inequality $x_1 x_2 \ldots x_n x_{n+1} \leq \overline{x}^{n+1}$.

Hence, by induction, the results holds for all $n \geq 2$. □

The phrase 'without loss of generality', which appears in this proof, is quite common; it is sometimes contracted to 'w.l.o.g.'.

Counterexamples

Providing a counterexample is the best method for *refuting*, or *dispoving*, a conjecture. In seeking counterexamples, it is a good idea to keep the cases you consider simple, rather than searching randomly. It is often helpful to consider 'extreme' cases, in which something is zero, a set is empty, or a function is constant, for example. If you are relaxing one of the hypotheses of a theorem and contemplating whether the conclusion still holds, make sure to consider cases that contravene the relaxed hypothesis (else you already know that they won't provide the desired counterexample!)

For example, suppose it were claimed that the requirement for x and y to be non-negative could be removed from Theorem 6.1 if we simply put a modulus sign inside the square root:

Claim. *Let x, y be real numbers. Then $\sqrt{|xy|} \leq \frac{1}{2}(x+y)$, with equality if and only if $x = y$.*

Refutation. This claim is not true. A counterexample is $x = 1$, $y = -1$.

There is no need to expand with additional arguments about why the counterexample exists - providing a single counterexample is sufficient to disprove the claim.

6.2 General advice

When seeking to prove or disprove a result, the following suggestions may be helpful:

- Make sure you are clear about the hypotheses and conclusions.
- 'Unpack' any definitions and re-state what exactly it is you know and what it is that you need to show (either in your head or, if it's helpful, write it down).
- If you need to show something 'for all $\varepsilon > 0$', start with 'Let $\varepsilon > 0$ be given.'
- If you can't see a way to start, consider 'seeking a contradiction' and suppose the result is not true to give yourself more to work with.
- If you need to show uniqueness, suppose there are two of whatever it is, and try to show that they are equal.
- If you haven't used all the hypotheses in your proof, you've probably missed something. (This is not *necessarily* true; it's possible more hypotheses were given than were needed, but more often than not they will all be needed).
- Look for extreme/simple cases as counterexamples.
- Don't be afraid to experiment, but have in mind what you're aiming for. If you're not making progress, try a different approach.
- Use sketches and diagrams to help gain intuition.
- If you get stuck, take a break. Look at it again with fresh eyes.

- Re-read your final proof. Be critical, and check that you are convinced by what you've written. (This is probably the most important of all these suggestions!)

6.3 Examples

In this section we discuss some example problems to illustrate aspects of proof and problem-solving.

The following example explores how the images and preimages of set intersections behave. It is presented as a simple true/false question, to which it should be understood that we need to provide reasoning for our answer. So we first need to decide whether we think they are true or false (some experimenting and thought may be required); then if true, we should prove it, and if false, we should provide a counterexample. A counterexample may well have been found anyway as part of our initial investigation to decide whether it is true or false, or conversely we may have gained insight into how a proof could work.

Example. (Images and pre-images). Let $f\colon X \to Y$ be a mapping and let $A, B \subseteq X$ and $C, D \subseteq Y$. Are the following statements true or false?

(i) $f(A \cap B) = f(A) \cap f(B)$,

(ii) $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.

Solution (+commentary). For both of these it might be reasonable to consider some extreme cases in case we stumble across an immediate counterexample. In this situation, such an extreme case might be if A and B, or C and D, are disjoint. In that case the sets on the left hand sides are both the empty set (the image and the pre-image of the empty set are the empty set). In the first case this immediately suggests the possibility of a counterexample, since $f(A)$ and $f(B)$ could easily intersect. If the function were constant, for example (that is, if the function assigns the same output to every input), then $f(A) = f(B)$, so provided A and B are not themselves empty their intersection will not be the empty set.

So we claim (i) is false. A counterexample would be $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 0$ for all $x \in \mathbb{R}$. Then if $A = \{0\}$, $B = \{1\}$, then $f(A \cap B) = \varnothing$, but $f(A) \cap f(B) = \{0\}$.

For (ii), the same thinking does not yield a counterexample, since if C and D are disjoint then their pre-images are also disjoint. It also seems unlikely that the question setter has given us two false statements (this is of course a hopeless *mathematical* argument, but a reasonable psychological one). So we claim that (ii) is true and aim to prove it:

Proof. If $x \in$ LHS, then $f(x) \in C \cap D$, so $f(x) \in C$ and $f(x) \in D$. Thus $x \in f^{-1}(C)$ and $x \in f^{-1}(D)$, and therefore $x \in$ RHS. Conversely, suppose $x \in$ RHS. Then $x \in f^{-1}(C)$ and $x \in f^{-1}(D)$, so $f(x) \in C$ and $f(x) \in D$, and therefore $x \in$ LHS. So each side is a subset of the other, and the sets are therefore equal. □

The next example relates to ideas that are covered in the Groups and Group Action course. This example is phrased in a way that is similar to a problem-sheet or exam-style question, with related sub-parts. The way the two questions in (ii) are

phrased strongly suggests that there must be a difference between the case when n is prime and when n is arbitrary. A 'hint' is given at the end, and it would be foolish not to make use of this, so we should think about how we could relate that to the question that is asked.

Example. (Modular arithmetic). Let $n \geq 2$ be an integer and let $\mathbb{Z}_n$ be the set of equivalence classes $\{\overline{0}, \overline{1}, , \ldots, \overline{n-1}\}$ defined by congruence modulo n on $\mathbb{Z}$.

(i) Show that the operation $\otimes$ on $\mathbb{Z}_n$ defined by

$$\overline{x} \otimes \overline{y} = \overline{x \times y},$$

is well-defined, where $x \times y$ denotes standard muliplication on $\mathbb{Z}$.

(ii) If $\overline{x} \neq \overline{0}$, a multiplicative inverse $\overline{y}$ has the property that $\overline{x} \otimes \overline{y} = \overline{1}$. Is there a multiplicative inverse for every $\overline{x} \neq \overline{0}$? What if n is prime?

[*You may assume Bezout's lemma, which says that if integers a and b are coprime, there exist integers k and l such that $a \times k + b \times l = 1$.*]

Solution (+commentary). Firstly, it is helpful to recall the precise definition of the equivalence relation 'congruence modulo n'; that is $x \sim y \Leftrightarrow y - x$ is a multiple of n.

For (i), we should consider why the given definition might *not* be well-defined. The concern must be that the same equivalence class can be represented in terms of different x and y (e.g. $\overline{0} = \overline{n}$, etc.). Since the given definition depends on x and y themselves, it might give a different answer if we represent the elements $\overline{x}$ and $\overline{y}$ using different values of x and y. So, supposing $\overline{x_1} = \overline{x_2}$ and $\overline{y_1} = \overline{y_2}$, we need to show that $\overline{x_1} \otimes \overline{y_1} = \overline{x_2} \otimes \overline{y_2}$. By definition of the equivalence classes $x_2 - x_1 = kn$ and $y_2 - y_1 = ln$ for some integers k and l. So

$$x_2 \times y_2 = (x_1 + kn) \times (y_1 + ln) = x_1 \times y_1 + (x_1 l + y_1 k + kln)n,$$

(we have omitted some of the $\times$ symbols to save space). Since the final term is a multiple of n this means that $\overline{x_2 \times y_2} = \overline{x_1 \times y_1}$, so indeed we have $\overline{x_1} \otimes \overline{y_1} = \overline{x_2} \otimes \overline{y_2}$. Hence, $\otimes$ is well-defined.

For (ii), the question seems to be suggesting things may be different when n is prime. So consider first a case where n is not prime and experiment a little to see if there are any simple counterexamples. If $n = 4$, there are only four equivalence classes, $\overline{0}$, $\overline{1}$, $\overline{2}$, and $\overline{3}$. We observe that $\overline{1}$ and $\overline{3}$ are their own inverse, but $\overline{2}$ does not have one. So we have found a counterexample: $n = 2$ and $\overline{x} = \overline{2}$. So the answer is No, there is not necessarily a multiplicative inverse for every $\overline{x} \neq \overline{0}$.

If n is prime, we must be supposed to use the hint. If n is prime then for any $0 < x < n$, x and n will be coprime, so the lemma implies that there are integers k and l such that $x \times k + n \times l = 1$. But this means that $\overline{x \times k} = \overline{1}$ so, following the definition, $\overline{x} \otimes \overline{k} = \overline{1}$. So this $\overline{k}$, which we know exists according to the lemma, is the multiplicative inverse of $\overline{x}$. So if n is prime, the answer becomes Yes, every $\overline{x} \neq \overline{0}$ does have a multiplicative inverse.

This example has shown that if p is prime, then every non-zero element of $\mathbb{Z}_p$ has a multiplicative inverse. This goes part way to showing that $\mathbb{Z}_p$ is a finite *field*

(meaning that it behaves in many respects similar to $\mathbb{R}$).

The final example below relates to things you will see in the Analysis II course. We are given a definition that involves a complicated-looking statement involving quantifiers, which is describing rigorously what it means for a function to decay to zero at infinity. We are being asked to apply this definition to two particular functions. Hopefully we have some immediate intuition (from the shape of their graphs, for example) that the first function does decay and the second one doesn't.

Example. (Limits). A continuous function $f\colon \mathbb{R} \to \mathbb{R}$ tends to zero as $x \to \infty$ if

$$\forall \varepsilon > 0,\ \exists X \in \mathbb{R} \text{ s.t. } \forall x \in \mathbb{R},\ \text{if } x > X \text{ then } |f(x)| < \varepsilon.$$

Prove or disprove whether the following functions tend to zero as $x \to \infty$:
(i) $f(x) = e^{-x}$;
(ii) $f(x) = \cos x$.

Solution (+commentary). For (i), we aim to show that the definition holds. Since the statement starts with $\forall \varepsilon$, we should start our proof by letting an arbitrary $\varepsilon > 0$ be given. Then we need to show that there exists an X such that for all $x > X$, $|f(x)| < \varepsilon$. Since the function $f(x) = e^{-x}$ is decreasing, and is always positive, this can achieved by taking $X = -\ln \varepsilon$. Then for $x > X$, $|f(x)| = |e^{-x}| < |e^{-X}| = \varepsilon$. So we have shown that the definition holds, and $f(x) = e^{-x}$ does tend to zero as $x \to \infty$.

For (ii), we aim to show that the definition does not hold, so we need to prove its negation. Following the rules for how to negate quantifiers, that is,

$$\exists \varepsilon > 0 \text{ s.t. } \forall X \in \mathbb{R},\ \exists x \in \mathbb{R} \text{ s.t. } x > X \text{ and } |f(x)| \geq \varepsilon.$$

(The original statement here is of the form '$\forall\ \exists\ \forall\ P(x)$', where P is itself of the form $Q \Rightarrow R$, in which Q is '$x > X$' and R is '$|f(x)| < \varepsilon$'. So the negated statement is of the form '$\exists\ \forall\ \exists$ not $P(x)$', and not $P(x)$ has been expressed as 'Q and not R'.) To see that this negated statement is true, we can observe that if $\varepsilon = \frac{1}{2}$ then for any $X \in \mathbb{R}$ there is a multiple of 2π, say $2\pi n$, that is larger than X, and for which $\cos 2\pi n = 1 \geq \varepsilon$. Hence $f(x) = \cos x$ does not tend to zero as $x \to \infty$.