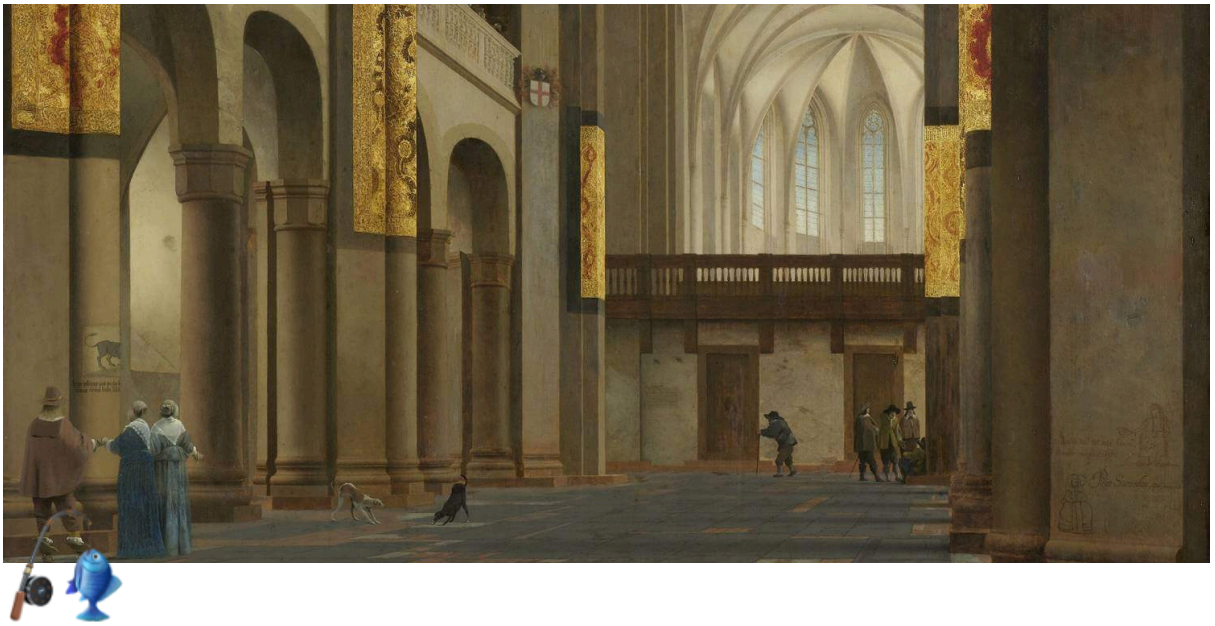


Discrete mathematics

Lecture notes for Mathematical Foundations

Mathematical and Logical Foundations of Computer Science

Semester 1 2024/25

Check Canvas for the latest version of these notes.

Manuscript version: *#1695ce* - 2024-10-30 11:31:05Z

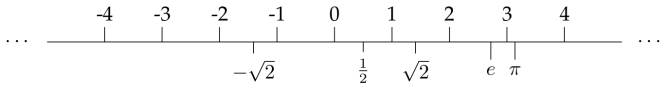
Contents

Lecture 1: Number systems

1.1 Numbers

What is a number? Rather than answering this question directly, a more modern approach is to ask 'what is a number system?'. There are many different number systems, useful for different purposes and with different structures and properties.

You are probably familiar with the idea that there is a 'number line'. We might draw it like this, with a few important numbers labelled.



Numbers that appear on the number line are called *real numbers*. The number line, or the set of all real numbers, is denoted by $\mathbb{R}$.

Definition 1.1. The *real numbers*, denoted $\mathbb{R}$, includes all the familiar numbers that appear on the number line.

What makes $\mathbb{R}$ a system of numbers? Some possible answers are as follows.

- They can be arranged on a line: if $a \neq b$ then either $a < b$ or $b < a$. We say that $\mathbb{R}$ is *totally ordered*.
- There is a natural 'centrepoint', namely 0.
- Pairs of real numbers can be added and multiplied together.
- 0 is neutral for addition, and there is an element 1 neutral for multiplication.
- Addition can be undone by subtraction, and multiplication by non-zero numbers can be undone by division.
- Addition and multiplication obey certain familiar laws of algebra — see later.

The key idea is that these are not straightforward properties of numbers as individual entities, but they are properties that make sense only in the context of a given collection of 'numbers'. In our analysis of the system of real numbers above, we have identified the following structural components.

- A *set* $\mathbb{R}$ whose elements are the numbers.

- Some *relations* between pairs of real numbers, including $<$, $\leq$, $=$. E.g., for $a, b \in \mathbb{R}$, it makes sense to ask whether $a < b$ or not.
- Some *functions* which each compute a real number from some other real numbers, including $+ : \mathbb{R}^2 \to \mathbb{R}$, $\times : \mathbb{R}^2 \to \mathbb{R}$. E.g., given $a, b \in \mathbb{R}$, we have $a + b \in \mathbb{R}$.

Remark 1.2. Perhaps the three most important concepts you will understand from this course are those of *sets*, *relations*, and *functions*. It is customary to refer to functions like $+$ and $\times$ as *operations*, and we will do so here — just remember that 'function' is the more fundamental concept and so 'operation' can always be read as 'function'.

1.2 Examples of number systems

We have discussed the real number $\mathbb{R}$ above. Several other important number systems arise as *subsets* of $\mathbb{R}$.

1.2.1 Integers

Perhaps the most obvious one is the *integers* or *whole numbers*:

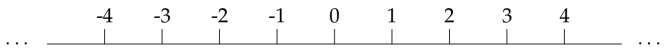
$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

(The choice of the letter Z comes from German word *Zahlen* ('numbers')). We write

$$\mathbb{Z} \subseteq \mathbb{R}$$

to indicate that the integers are a subset of the reals, i.e. that every integer is also a real number (but not conversely — there are real numbers which are not integers).

How do the integers $\mathbb{Z}$ compare to the reals $\mathbb{R}$? As they are a subset, they still arrange onto a line.



However, now there are a large 'gaps' between elements. Actually, these gaps are an artefact of our drawing the numbers on a page. A more intrinsic description of the situation is that, in the case of the reals, we can always find more real numbers between any given two numbers whereas, in the case of the integers, n and $n + 1$ have no other integers between them. So $\mathbb{Z}$ is still totally ordered, but its ordering behaves a little differently.

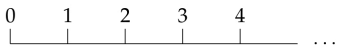
We still know that the sum and product of integers is again an integer. Moreover, integers can be subtracted from each other, but we cannot in general divide integers and get another integer. This suggests that division is less fundamental to the notion of number system than the other operations.

1.2.2 Natural numbers

Another number system is the *natural numbers*

$$\mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}.$$

These are the integers greater than or equal to 0. These are sometimes described as the 'counting numbers', since they are the numbers used for counting finite collections of discrete items. They are still totally ordered,



but now the line is infinite only in one direction. The naturals are closed under addition and multiplication, but neither division nor subtraction is generally possible. The difference $m - n$ of two naturals $m, n \in \mathbb{N}$ is a natural number iff $m \geq n$. One more interesting fact about naturals is that the ordering is easily definable in terms of the addition:

$$\forall m, n \in \mathbb{N}.\, m \leq n \iff \exists k \in \mathbb{N}.\, m + k = n.$$

1.2.3 Rational numbers

The rational numbers are a subset of the reals containing all the integers as well as all quotients of integers by integers:

$$\mathbb{Q} = \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z}, q \neq 0 \right\}.$$

These behave quite similarly to the reals in terms of admitting addition, multiplication, subtraction, and division. As you may know, and as we will see later, there are many important numbers 'missing' from the rationals, like $\sqrt{2}$ and π. On the other hand, we will see later that there is an fundamental problem with representing arbitrary real numbers on a computer, so we need an undertanding of more limited number systems too.

1.2.4 Summary of examples

So far we have seen four examples of number systems, which can be displayed as an increasing sequence as follows.

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

When you come to study linear algebra, you will also need the *complex numbers*, which are a superset of the reals:

$$\mathbb{R} \subseteq \mathbb{C}.$$

We can summarize the properties in a table.

	$\mathbb{N}$	$\mathbb{Z}$	$\mathbb{Q}$	$\mathbb{R}$	$\mathbb{C}$
admits $+$?	✓	✓	✓	✓	✓
admits $\times$?	✓	✓	✓	✓	✓
admits $-$?		✓	✓	✓	✓
admits $\div$?			✓	✓	✓
totally ordered?	✓	✓	✓	✓	

From this table, it should be clear that the things common to all number systems of interest are $+$ and $\times$, while $-$, $\div$, and $<$ are 'optional extras'.

1.3 Algebraic laws

Certain algebraic properties of addition and multiplication have names.

1. $+$ and $\times$ are *associative*:

$$(x + y) + z = x + (y + z)$$
$$(x \times y) \times z = x \times (y \times z)$$

2. $+$ and $\times$ are *commutative*:

$$x + y = y + x$$
$$x \times y = y \times x$$

3. 0 is a *neutral* element for addition, and 1 is a *neutral* element for multiplication:

$$x + 0 = x = 0 + x$$
$$x \times 1 = x = 1 \times x$$

   (Also called *identity* or *unit* elements).

4. Multiplication *distributes* over addition:

$$x \times (y + z) = x \times y + x \times z$$

When a binary operation is associative, it means we can treat it as an operation on 'lists' with no need to insert brackets. Concretely, we tend to write $x + y + z$ because it does not matter how we insert brackets. In this way we think of having k-ary operations for every $k \geq 1$. So we sometimes write a long sum as

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_{n-1} + a_n$$

where the bracketing of $a_1 + a_2 + \ldots + a_{n-1} + a_n$ does not matter. For multiplication, a long product is written

$$\prod_{i=1}^{n} a_i = a_1 a_2 \ldots a_{n-1} a_n$$

where again the bracketing does not matter. As an example, the *factorial* function '!' can be defined as

$$n! := \prod_{i=1}^{n} i.$$

When an associative operation is also commutative we can freely reorder its arguments however we like. When an associative operation has a neutral element, it makes sense to treat that neutral element as the result of the 'nullary' or 0-ary version of that operation. An empty sum is 0, but perhaps less obviously an empty product is 1. This is why we define $0! = 1$ and $n^0 = 1$.

1.4 Inverse operations

We have just reasoned by example that the operations $+$ and $\times$ are fundamental to what a number system is, unlike subtraction and division. In fact, Important for us is that subtraction and division can be seen as arising from inverses to addition and multiplication.

Definition 1.3. Let $x \in \mathbb{R}$. A *negative* (or *additive inverse*) of x is a number $x' \in \mathbb{R}$ such that

$$x + x' = 0.$$

Since this number is unique, we use the notation $-x$ for it. The operation of *subtraction* is defined by

$$x - y \coloneqq x + (-y).$$

There is a similar story for division. However, division is not really an 'operation' since it is not defined for all arguments.

Definition 1.4. Let $x \in \mathbb{R}$. A *reciprocal* (or *multiplicative inverse*) of x is a number $x' \in \mathbb{R}$ such that

$$x \times x' = 1.$$

A reciprocal exists iff $x \neq 0$, and when it does it is unique and so we use the notation x^{-1} or $\frac{1}{x}$ for it. The *quotient* of x by y is defined for $y \neq 0$ by

$$\frac{x}{y} = x \div y \coloneqq x \times y^{-1}.$$

There is a similar story for all number systems in this module.

Definition 1.5. A *field* is a number system supporting $+$ and $\times$, satisfying all the algebraic laws of the previous section, for which every number has a negative and every non-zero number has a reciprocal.

1.5 Divisibility

For natural numbers, we cannot in general perform subtraction and get a natural number. However, the situation is relatively uninteresting since the existence of a 'difference' $a - b$ reduces to whether $b \leq a$. The case for division is a bit more interesting.

Definition 1.6. For integers m, n, say that m *divides* n, written $m \mid n$, if there exists $k \in \mathbb{Z}$ such that $n = km$.

If $m \mid n$, we also say that m is a *divisor* of n. The number k in $n = km$ is the *quotient*.

- 1 divides every integer.
- Every integer divides 0.
- $0 \mid n$ iff $n = 0$.
- If $n \mid a$ and $n \mid b$, then $n \mid ax + by$ for any $x, y \in \mathbb{Z}$.

'Divisibility' is uninteresting for $\mathbb{Q}$ and $\mathbb{R}$ because we multiplicative inverses there.

Definition 1.7. A *prime number* is an integer $p \in \mathbb{Z}$ with $p > 1$ such that the only positive divisors of p are 1 and p itself.

The sequence of prime numbers begins

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, \ldots$$

Additively, 1 is the only building block we need to obtain all natural numbers. Multiplicatively, the prime numbers, in their infinite, seemingly sporadic sequence, are the building blocks. A *prime factorization* of $n \in \mathbb{N}$ means an expression

$$n = p_1 p_2 \ldots p_k$$

where $k \in \mathbb{N}$ and the p_i are all prime (not necessarily distinct).

Theorem 1.8. Every natural number $n \in \mathbb{N}$ with $n > 1$ has a prime factorization.

Proof. We prove this by giving an algorithm for finding a prime factorization.

If n is prime, then $n = n$ is itself a prime factorization. So suppose n is not prime. This means it has a divisor $a \mid n$ with $1 < a < n$. Writing $n = ab$, we must also have $1 < b < n$. Now apply the algorithm to both a and b to get prime factorizations

$$a = p_1 p_2 \ldots p_k$$
$$b = q_1 p_2 \ldots q_l.$$

Then

$$n = p_1 p_2 \ldots p_k q_1 q_2 \ldots q_l$$

is a prime factorization of n.

It remains to check that this algorithm terminates. But the recursive step of the algorithm calls itself on *smaller* arguments. Thus if it 'loops' we must have found an infinite descending chain of natural numbers, which is impossible. □

1.6 Induction

The proof above was our first example of a proof by induction. We will rewrite it shortly once we have introduced the more standard style for induction proofs.

We will introduce induction with another example.

Theorem 1.9. Let $n \in \mathbb{N}$. Then

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

Let us work through how to prove this for every $n \in \mathbb{N}$. To keep track of progress, for each particular $n \in \mathbb{N}$ write $P(n)$ for the statement that

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

When faced with a problem like this, it is not a bad idea to start calculating some basic cases by hand.

n	$1+2+\ldots+n$	$\frac{1}{2}n(n+1)$
0	0	0
1	1	$1/2 \times 1 \times 2 = 1$
2	$1+2=3$	$1/2 \times 2 \times 3 = 3$
3	$1+2+3=6$	$1/2 \times 3 \times 4 = 6$
4	$1+2+3+4=10$	$1/2 \times 4 \times 5 = 10$
5	$1+2+3+4+5=15$	$1/2 \times 5 \times 6 = 15$

Thus from our calculations we have proved $P(0)$, $P(1)$, $P(2)$, $P(3)$, $P(4)$, and $P(5)$. Note that in the middle column of our table there is quite a bit of calculation being duplicated between rows. We can reuse some of our work to make proving $P(6)$ easier:

$$\begin{aligned} 1+2+3+4+5+6 &= (1+2+3+4+5)+6 \\ &= 1/2 \times 5 \times 6 + 6 \end{aligned}$$

using $P(5)$ to make a substitution, and now it remains to rewrite this last expression $1/2 \times 6 \times 7$. We might notice that if we factorize out $1/2 \times 6$ we can do this quite easily without too much calculation.

$$\begin{aligned} 1/2 \times 5 \times 6 + 6 &= 1/2 \times 5 \times 6 + 1/2 \times 2 \times 6 \\ &= 1/2 \times 6 \times (5+2) \\ &= 1/2 \times 6 \times 7. \end{aligned}$$

This proves $P(6)$. If we try to do $P(7)$, we will notice that the calculation was almost exactly the same, this time using $P(6)$ for a substitution:

$$\begin{aligned} 1+2+3+4+5+6+7 &= (1+2+3+4+5+6)+7 \\ &= 1/2 \times 6 \times 7 + 7 \\ &= 1/2 \times 6 \times 7 + 1/2 \times 2 \times 7 \\ &= 1/2 \times 7 \times (6+2) \\ &= 1/2 \times 7 \times 8. \end{aligned}$$

This proves $P(7)$.

Now let us try to abstract the pattern. Suppose we continue this effort and have proved $P(0), P(1), \ldots P(n)$. Let us try to prove $P(n+1)$:

$$\begin{aligned} 1+2+\ldots+n+(n+1) &= (1+2+\ldots+n)+(n+1) \\ &= \tfrac{1}{2}n(n+1)+(n+1) \\ &= \tfrac{1}{2}n(n+1)+\tfrac{1}{2} \times 2(n+1) \\ &= \tfrac{1}{2}(n+1)(n+2) \end{aligned}$$

as required.

Having managed to abstract the pattern and prove that it is a valid technique for general n, we know that the tedious process of proving $P(0)$, then $P(1)$, then $P(2)$, etc.…is guaranteed to continue succeeding. This is as good as a proof of "for all $n \in \mathbb{N}$, $P(n)$".

Let us revisit our initial table for $n = 0, 1, 2, 3, 4, 5$. It was a little hard to see at first, but the abstract pattern also lets us deduce $P(1)$ from $P(0)$, and $P(2)$ from $P(1)$, etc. It does not help us with $P(0)$ (would we deduce it from '$P(-1)$'?) and this case must be done separately.

Let us write this more formally, as a model you should use for your own induction proofs.

Proof of Theorem 1.9. Let $P(n)$ stand for the proposition

$$1 + 2 + \ldots + n = \tfrac{1}{2}n(n+1).$$

We will prove $P(n)$ by induction on $n \in \mathbb{N}$.

> Base case: When $n = 0$,
>
> $$LHS = 0$$
>
> because it is an empty sum, and
>
> $$RHS = \tfrac{1}{2} \times 0 \times 1 = 0,$$
>
> so $P(0)$ holds.

> Inductive step: Supposing $P(n)$ is true, we will prove $P(n+1)$. By the induction hypothesis, we have
>
> $$1 + \ldots + n = \tfrac{1}{2}n(n+1).$$
>
> Adding $(n+1)$ to both sides, we get
>
> $$\begin{aligned} 1 + \ldots + n + (n+1) &= \tfrac{1}{2}n(n+1) + (n+1) \\ &= \tfrac{1}{2}(n+1)(n+2) \end{aligned}$$
>
> which is $P(n+1)$, as required.

Hence, by induction, $P(n)$ holds for all $n \in \mathbb{N}$. □

As a second model, let us write out another proof of Theorem 1.8. Note, this proof does not require a base case.

Proof of Theorem 1.8. Let $P(n)$ be the statement that n has a prime factorization. We will prove $P(n)$ for all $n \in \mathbb{N}$ with $n > 1$ by induction.

> Inductive step: Let $n > 1$ and suppose $P(k)$ is true for all $1 < k < n$. We will prove $P(n)$. There are two cases to consider.
>
> > n is prime: If n is prime then $n = n$ is a prime factorization, so $P(n)$ holds.
> >
> > n is composite (not prime): So $n = ab$ for some $a, b \in \mathbb{N}$ where $1 < a < n$ and $1 < b < n$. But by induction hypothesis $P(a)$ and $P(b)$ hold, so we have prime factorizations
> >
> > $$\begin{aligned} a &= p_1 p_2 \ldots p_k \\ b &= q_1 p_2 \ldots q_l. \end{aligned}$$
> >
> > Then
> >
> > $$n = p_1 p_2 \ldots p_k q_1 q_2 \ldots q_l$$
> >
> > is a prime factorization of n, so $P(n)$ holds.

Hence, by induction, $P(n)$ is true for all $n \in \mathbb{N}$, $n > 1$. □

Remark 1.10. Think carefully about why we did not require a 'base case' step in the above. Since the statement to be proved begins with $n = 2$, you might think we have to prove $P(2)$ separately. But, actually, the inductive step proves $P(n)$ from $P(2), P(3), \ldots, P(n-1)$, which in the case $n = 2$ means proving $P(2)$ from an empty list of assumptions. It would not be wrong to include a separate proof of $P(2)$ as a base case, just unnecessary.

Lecture 2: Divisors

2.1 `div` and `mod`

Given natural numbers $a, b \in \mathbb{N}$, we might not have $a \mid b$ but we can attempt division anyway and get a 'remainder'. When $a \neq 0$, there are unique $q, r \in \mathbb{N}$ with $r < a$ such that

$$b = qa + r.$$

In this expression, q is called *quotient* and r is the *residue* (or *remainder*).

Definition 2.11. We define binary operations $\texttt{div}, \texttt{mod}$ on pairs (b, a) where $a, b \in \mathbb{N}$ and $a \neq 0$, by

$$b = (b \,\texttt{div}\, a)a + (b \,\texttt{mod}\, a).$$

Proposition 2.12. For $a, b \in \mathbb{N}$ with $a \neq 0$,

$$a \mid b \iff b \,\texttt{mod}\, a = 0$$

in which case $b = (b \,\texttt{div}\, a)a$. □

2.2 Greatest common divisors

Recall that $d \in \mathbb{Z}$ is a *divisor* of $n \in \mathbb{Z}$, written $d \mid n$ iff there exists $k \in \mathbb{Z}$ such that $n = kd$.

Definition 2.13. Let $a, b \in \mathbb{Z}$. We call $d \in \mathbb{N}$ a *common divisor* of a and b if d is a divisor of a and a divisor of b, i.e. if both $d \mid a$ and $d \mid b$. It is a *greatest common divisor* if it is a common divisor and, whenever d' is also a common divisor, $d' \mid d$. When it exists, we write the greatest common divisor as $\gcd(a, b)$.

Proposition 2.14. Let $a, b \in \mathbb{Z}$. If $\gcd(a, b)$ exists, it is unique.

Proof. Let d_1, d_2 be two greatest common divisors of a and b. Then since d_1 is a greatest common divisor and d_2 is a common divisor, by definition $d_2 \mid d_1$. Similarly, since d_2 is a greatest common divisor and d_1 is a common divisor, $d_1 \mid d_2$. Therefore $d_1 = d_2$. □

In most cases a $\gcd$ will also be the 'greatest' divisor in terms of magnitude. However, can you see why $\gcd(0, 0) = 0$?

Example 2.15. For $n \in \mathbb{N}$:

- $\gcd(0, n) = n$,
- $\gcd(1, n) = 1$,
- $\gcd(n, n) = n$.

2.3 Euclid's algorithm

We want to show that $\gcd(a, b)$ exists for every pair $a, b \in \mathbb{Z}$, and that it is computable. The following lemma is a key idea.

Lemma 2.16. Suppose $\gcd(a, b)$ exists. Then, for any $k \in \mathbb{Z}$, $\gcd(a, b + ka) = \gcd(a, b)$.

Proof. For any $d \in \mathbb{N}$, d divides both a and $b + ka$ iff d divides both a and b. Thus both pairs have exactly the same set of common divisors. □

Definition 2.17 (Euclid's algorithm)**.** Given two natural numbers $a, b \in \mathbb{N}$, we produce another natural number as follows.

- Start defining a sequence of naturals (r_n) where $r_0 = \max(a, b)$ and $r_1 = \min(a, b)$.
- Given $r_0, \ldots, r_k, r_{k+1}$, if $r_{k+1} = 0$ output r_k.
- Otherwise, continue the sequence with $r_{k+2} = r_k \texttt{ mod } r_{k+1}$.

Theorem 2.18. For any $a, b \in \mathbb{N}$, Euclid's algorithm terminates and returns $\gcd(a, b)$.

Proof. We first deal with the case $a = b$. Following the algorithm, $r_0 = r_1 = a$ and $r_2 = 0$, so it does terminate and it returns $r_1 = a$. But $\gcd(a, a) = a$, so it is correct in this case.

Now we take the case $a \neq b$, and WLOG $a < b$, so that $r_0 = b$, $r_1 = a$.

► *Claim.* The sequence (r_n) is strictly decreasing for as long as it is defined.

> *Proof of Claim.* Let $k \in \mathbb{N}$. We will show that $r_{k+1} < r_k$. At the beginning of the sequence, when $k = 0$, $r_1 = a < b = r_0$ by definition, as required. Now suppose $k > 0$ and suppose $r_k \neq 0$, so that r_{k+1} is defined. Then $r_{k+1} = r_{k-1} \texttt{ mod } r_k$, which is always less than r_k by the definition of `mod`. This proves the claim. ◄

This implies that the algorithm must terminate, since there is no infinite strictly decreasing sequence of natural numbers.

► *Claim.* If the sequence continues to r_{k+1}, then $\gcd(a, b) = \gcd(r_k, r_{k+1})$ (i.e. if either side exists then so does the other and they are equal).

> *Proof of Claim.* By induction on k.
>
> Base case: When $k = 0$, $\gcd(r_0, r_1) = \gcd(b, a) = \gcd(a, b)$, so the claim is trivial.
>
> Inductive step: It suffices to show that $\gcd(r_k, r_{k+1}) = \gcd(r_{k+1}, r_{k+2})$. But $r_{k+2} = r_k - (r_k \texttt{ div } r_{k+1})r_{k+1}$, so this follows from Lemma 2.16.
>
> Hence, by induction, $\gcd(a, b) = \gcd(r_k, r_{k+1})$. ◄

In particular, if the sequence terminates after $r_N \neq 0$ with $r_{N+1} = 0$, then $\gcd(a, b) = \gcd(r_N, 0) = r_N$. But then $r_N = \gcd(a, b)$ is the output of the algorithm, as required. □

Example 2.19. Let us run Euclid's algorithm to compute $\gcd(29, 73)$.

$$\begin{aligned} r_0 &= 73 \\ r_1 &= 29 \\ r_2 &= 73 \bmod 29 \\ &= 15 \\ r_3 &= 29 \bmod 15 \\ &= 14 \\ r_4 &= 15 \bmod 14 \\ &= 1 \\ r_5 &= 14 \bmod 1 \\ &= 0. \end{aligned}$$

Since $r_5 = 0$, the output is $r_4 = 1$. So $\gcd(29, 73) = 1$. In this case there is a shortcut to that answer, since 29 and 73 are distinct prime numbers.

For another example consider $\gcd(144, 234)$.

$$\begin{aligned} r_0 &= 234 \\ r_1 &= 144 \\ r_2 &= 234 \bmod 144 \\ &= 90 \\ r_3 &= 144 \bmod 90 \\ &= 54 \\ r_4 &= 90 \bmod 54 \\ &= 36 \\ r_5 &= 54 \bmod 36 \\ &= 18 \\ r_6 &= 36 \bmod 18 \\ &= 0. \end{aligned}$$

Since r_6 is 0, the output is $r_5 = 18$. So $\gcd(144, 234) = 18$. The alternative approach is to find prime factorizations $144 = 2^4 \times 3^2$ and $234 = 2 \times 3^2 \times 13$ whence $\gcd(144, 234) = 2 \times 3^2$.

Theorem 2.20 (Bézout's Lemma)**.** Let $a, b \in \mathbb{N}$. Then there exist integers $x, y \in \mathbb{Z}$ such that

$$xa + yb = \gcd(a, b).$$

Proof. The result is trivial if $a = b$, so WLOG $a < b$. We can read them off x and y from Euclid's algorithm as follows. Consider the sequence (r_n) generated by the algorithm.

▶ *Claim.* If the sequence continues to r_k, then there exist $x, y \in \mathbb{Z}$ such that $xa + yb = r_k$.

Proof of Claim. We use induction on k.

Base cases, $k = 0$ or 1: When $k = 0$, $r_0 = b = 0 \cdot a + 1 \cdot b$. When $k = 1$, $r_1 = a = 1 \cdot a + 0 \cdot b$.

Inductive step, $k \geq 2$: Since $k \geq 2$, we have $r_{k-2} = (r_{k-2} \texttt{ div } r_{k-1})r_{k-1} + r_k$. Let us write $q_k = (r_{k-2} \texttt{ div } r_{k-2})$, so that

$$r_k = r_{k-2} - q_k r_{k-1}.$$

By induction hypothesis, there are $x_{k-2}, y_{k-2}, x_{k-1}, y_{k-1} \in \mathbb{Z}$ such that

$$r_{k-2} = x_{k-2}a + y_{k-2}b \qquad \text{and} \qquad r_{k-1} = x_{k-1}a + y_{k-1}b.$$

Hence

$$\begin{aligned} r_k &= r_{k-2} - q_k r_{k-1} \\ &= (x_{k-2}a + y_{k-2}b) - q_k(x_{k-1}a + y_{k-1}b) \\ &= (x_{k-2} - q_k x_{k-1})a + (y_{k-2} - q_k y_{k-1})b \end{aligned}$$

as required.

Hence the result follows by induction. ◀

Since $\gcd(a, b)$ is the last non-zero value of the sequence (r_n), the result follows. □

By following the argument in the lemma, we can find the the *Bézout coefficients* explicitly. Let us return to the example before.

Example 2.21. Since $\gcd(29, 73) = 1$, we can find $x, y \in \mathbb{Z}$ such that $29x + 73y = 1$.

$$\begin{aligned} 1 &= 15 - 14 \\ &= 15 - (29 - 15) \\ &= 2 \times 15 - 29 \\ &= 2 \times (73 - 2 \times 29) - 29 \\ &= 2 \times 73 - 5 \times 29 \end{aligned}$$

so we can take $x = -5$ and $y = 2$.

Let us consider the other example. Since $\gcd(144, 234) = 18$, we can find $x, y \in \mathbb{Z}$ such that $144x + 234y = 18$.

$$\begin{aligned} 18 &= 54 - 36 \\ &= 54 - (90 - 54) \\ &= 2 \times 54 - 90 \\ &= 2 \times (144 - 90) - 90 \\ &= 2 \times 144 - 3 \times 90 \\ &= 2 \times 144 - 3 \times (234 - 144) \\ &= 5 \times 144 - 3 \times 234 \end{aligned}$$

so we can take $x = 5$ and $y = -3$.

We could instead compute the x_k and y_k (and q_k) in one pass of Euclid's algorithm. We might collect the results into a table.

Example 2.22. Let us compute x, y such that $21x + 57y = \gcd(21, 57)$.

k	r_k	q_k	x_k	y_k
0	57		0	1
1	21		1	0
2	15	2	-2	1
3	6	1	3	-1
4	3	2	-8	3
5	0			

The column r_k is just the sequence from Euclid's algorithm, $q_k = r_{k-2} \texttt{ div } r_{k-1}$ is such that $r_{k-2} = q_k r_{k-1} + r_k$, and the last two columns are given by the recurrences $x_k = x_{k-2} - q_k x_{k-1}$ and $y_k = y_{k-2} - q_k y_{k-1}$. Each row maintains the invariant

$$r_k = x_k a + y_k b.$$

Indeed, $\gcd(21, 57) = 3 = (-8) \times 21 + 3 \times 57$.

This version which computes the Bézout coefficients in one pass is sometimes called 'extended Euclid's algorithm'. You are not expected to be able to reproduce this version of the algorithm from memory, as long as you know how to find the Bézout coefficients by 'working backwards' after the ordinary Euclid's algorithm.

2.4 Coprimality and prime factorization

The proofs in this section are not examinable, but you will need to be familiar with the statements of the propositions.

Definition 2.23. Let $a, b \in \mathbb{Z}$. We say that a and b are *coprime* if $\gcd(a, b) = 1$.

Obviously, if d is a common divisor of a and b, then for any $x, y \in \mathbb{Z}$,

$$d \mid ax + by.$$

Thus, Bézout's Lemma tells us that, for $a, b \in \mathbb{Z}$, it is possible to find $x, y \in \mathbb{Z}$ such that

$$ax + by = 1$$

if and only if a and b are coprime. (Bézout's Lemma as we proved it is actually only for $a, b \in \mathbb{N}$, but why does it not matter?)

The number 1 is coprime to every integer. A prime number p is a positive integer which is coprime to every positive integer except p itself.

Lemma 2.24 (Euclid's Lemma)**.** Let $a, b \in \mathbb{N}$ and let $p \in \mathbb{N}$ be a prime. If $p \mid ab$, then either $p \mid a$ or $p \mid b$.

Proof. We will suppose $p \nmid a$ and prove $p \mid b$. Since $p \nmid a$, $\gcd(p, a) = 1$. Hence, by Bézout's Lemma, we can find $x, y \in \mathbb{Z}$ such that

$$xa + yp = 1.$$

Then

$$b = xab + ypb$$

where trivially $p \mid ypb$ and $p \mid xab$ by hypothesis. Therefore $p \mid b$, their sum. □

In fact, virtually the same argument proves something slightly more general.

Lemma 2.25. Let $a, b, c \in \mathbb{N}$ with $a \mid bc$. If $\gcd(a, b) = 1$, then $a \mid c$.

Proof. By Bézout's Lemma, we can find $x, y \in \mathbb{Z}$ such that

$$xa + yb = 1.$$

Then

$$c = xac + ybc$$

where trivially $a \mid xac$ and $a \mid ybc$ by hypothesis. Therefore $a \mid c$, their sum. □

Theorem 2.26. Let $n \in \mathbb{N}$ be greater than 0. Then the factorization of n into primes is unique.

Proof. Suppose we have two prime factorizations which happen to be equal:

$$p_1 p_2 \dots p_s = q_1 q_2 \dots q_t.$$

We can cancel off all common factors, so that WLOG none of the p_i are equal to any of the q_j. If one side is not an empty product, say $s > 0$, then $p_1 \mid q_1 \dots q_t$. By repeated application of Euclid's Lemma, either $p_1 \mid q_j$ for some $1 \leq j \leq t$. But this means $p_1 = q_j$, a contradiction since we had already cancelled all common factors. □

Lecture 3: Relations, rationals, residue classes

3.1 Some set notations

3.1.1 Products of sets

If X and Y are sets, we write $X \times Y$ for the set of *ordered pairs* (x, y). That is,

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

For example,

$$\{0, 1, 2\} \times \{a, b\} = \{(0, a), (1, a), (2, a), (0, b), (1, b), (2, b)\}.$$

We also allow ourselves to talk about ordered n-tuples in this way. E.g. for $n = 3$:

$$X \times Y \times Z = \{(x, y, z) \mid x \in X, y \in Y, z \in Z\}.$$

When dealing with products of a set with itself, we use 'exponential notation'. E.g.

$$X^2 = X \times X = \{(x_1, x_2) \mid x_1, x_2 \in X\}.$$

3.1.2 Powerset

If X is a set, we write $P(X)$ for the *powerset* of X, the set of subsets of X. That is,

$$P(X) = \{S \mid S \subseteq X\}.$$

For example,

$$P(\{0, 1, 2\}) = \{\{\}, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{1, 2\}, \{2, 0\}, \{0, 1, 2\}\}.$$

Recall that the order in which we list the elements of a set does not matter.

The elements of a powerset $P(X)$ are themselves *sets*. Observe that $P(\{0, 1, 2\})$ has

- one element of size 0: $\{\}$,
- three elements of size 1: $\{0\}, \{1\}, \{2\}$,
- three elements of size 2: $\{0, 1\}, \{1, 2\}, \{2, 0\}$,

- one element of size 3: $\{0, 1, 2\}$.

In particular, $\{0, 1, 2\} \in P(\{0, 1, 2\})$. In general, $X \in P(X)$ because $X \subseteq X$.

One more thing: the empty set $\{\}$ gets a special notation, $\emptyset$ (or $\varnothing$). So, for any set X, it is true that $\emptyset \in P(X)$.

3.2 Relations

3.2.1 Binary relations on a set

Intuitively, a *binary relation* is a proposition that can be true or false about two objects. For example, if $x, y \in \mathbb{R}$, then the statement "$x < y$" can be true or false.

Definition 3.27. A *binary relation on a set* X is a subset $R \subseteq X^2$.

Intuitively, the subset $R \subseteq X^2$ is the set of pairs (x_1, x_2) for which the proposition is true. Note, the notion of relation is relative to a set being considered. So the order relations on $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ are technically different and should have different names:

- $<_{\mathbb{N}} = \{(x, y) \in \mathbb{N}^2 \mid x < y\}$
- $<_{\mathbb{Z}} = \{(x, y) \in \mathbb{Z}^2 \mid x < y\}$
- $<_{\mathbb{Q}} = \{(x, y) \in \mathbb{Q}^2 \mid x < y\}$
- $<_{\mathbb{R}} = \{(x, y) \in \mathbb{R}^2 \mid x < y\}$

Notation 3.28. If R is a binary relation on X and $x, y \in X$, then we write

$$x \; R \; y$$

to mean $(x, y) \in R$.

Since it tends not to cause confusion, we never bother giving different names to the different order relations above.

Example 3.29. How many binary relations are there on the set $\{0, 1\}$? A binary relation is just a subset of $\{0, 1\} \times \{0, 1\}$. The latter set has four elements: $(0, 0), (1, 0), (0, 1), (1, 1)$. The powerset of a four-element set has 16 elements, so there are 16 binary relations on $\{0, 1\}$.

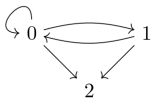
3.2.2 Graphs

A binary relation can be drawn as a graph. To do this, we think of the elements of X as nodes, and we draw an arrow $x \to y$ whenever $x \; R \; y$.

Example 3.30. Let $X = \{0, 1, 2, 3\}$ and consider the relation

$$R = \{(0, 0), (0, 1), (1, 0), (0, 2), (1, 2)\}.$$

Its graph might be drawn like this.



The position of the nodes of the graph is not significant, we can place them anywhere convenient.

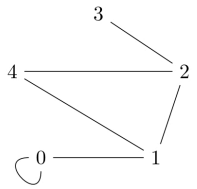
Definition 3.31. A relation R is *symmetric* if, whenever $x \ R \ y$, we also have $y \ R \ x$.

When a relation is symmetric, we can think of it as an *undirected* graph. We then draw one edge without an arrowhead between x and y whenever $x \ R \ y$ and $y \ R \ y$.

Example 3.32. Let $X = \{0, 1, 2, 3, 4\}$. Then

$$R = \{(0,0), (0,1), (1,0), (1,2), (2,1), (1,4), (4,1)(2,3), (3,2), (2,4), (4,2)\}$$

is a symmetric relation. We might draw it as an undirected graph as follows.



Remark$^+$. When thinking of graphs as symmetric relations, it is common to restrict further to *irreflexive* graphs (no self loops), but that will not concern us here.

3.2.3 Equivalence relations

Example 3.33. For any set X, *equality* is a relation on X. Let

$$R_= = \{(x, x) \mid x \in X\} \subseteq X^2.$$

Then $x \ R_= \ y$ iff $x = y$. So in fact we never use the notation $R_=$ and just use $=$ directly.

The equality relation on a set is quite special. We abstract its properties to get the notion of *equivalence relation*. This notion has three parts. One part is being symmetric. In the following, let X be a set and R a binary relation on X.

Definition 3.34. A relation R is *reflexive* if, for all $x \in X$, we have $x \ R \ x$.

Definition 3.35. A relation R is *transitive* if, for all $x, y, z \in X$:

$$x \ R \ y \text{ and } y \ R \ z \implies x \ R \ z.$$

Definition 3.36. An *equivalence relation* on X is a binary relation R on X which is reflexive, symmetric, and transitive.

Proposition 3.37. Equality is an equivalence relation.

Proof. Let X be a set and consider equality $=$ as a binary relation on X. We just go through the three axioms for an equivalence relation and check that they are true for equality. They are all fairly obvious.

- Reflexivity. Obviously, everything is equal to itself.
- Symmetry. Obviously, if $x = y$ then also $y = x$.
- Transitivity. If $x = y$ and $y = z$, then indeed $x = z$.

□

3.2.4 Partitions

An equivalence relation determines a partition of the set it is on.

Definition 3.38. Let R be an equivalence relation on a set X. For any $x \in X$, the *equivalence class* of x is

$$[x] = \{y \in X \mid x \ R \ y\}.$$

We write $[x]_R$ when we need to disambiguate between the equivalence classes for multiple different equivalence relations.

Notation 3.39. The set of R-equivalence classes is written X/R, and sometimes called the *quotient of X by R*. Note that $X/R \subseteq P(X)$.

In the following, two sets A and B are *disjoint* if they have no members in common, i.e. if $A \cap B = \emptyset$.

Proposition 3.40. Distinct equivalence classes are disjoint and the union of the equivalence classes is all of X.

Proof. ▶ *Claim.* Let $x, y \in X$. If $x \ R \ y$, then $[x] = [y]$.

> *Proof of Claim.* We first show that $[x] \subseteq [y]$. So let $z \in [x]$. By definition, this means that $x \ R \ z$. By symmetry of R, $z \ R \ x$. By transitivity (using the hypothesis that $x \ R \ y$) we have $z \ R \ y$. By symmetry, $y \ R \ z$. But now by definition of $[y]$ we have $z \in [y]$. This concludes the argument that $[x] \subseteq [y]$. The argument that $[y] \subseteq [x]$ is virtually identical. ◀

▶ *Claim.* Let $x, y \in X$. If $[x] \cap [y] \neq \emptyset$, then $x \ R \ y$ (and hence $[x] = [y]$).

> *Proof of Claim.* Since the intersection is non-empty, we can pick some element $z \in [x] \cap [y]$. So $x \ R \ z$ and $y \ R \ z$. But now by the equivalence relation axioms, $x \ R \ y$. ◀

This suffices for the first part of the proposition (we have shown the contrapositive, that non-disjoint equivalence classes are equal). To say that the union of the equivalence classes is all of X is just to say that every $x \in X$ is in some equivalence class. But of course this is the case, since by reflexivity $x \in [x]$. □

Definition 3.41. Let X be a set. A *partition* of X is a subset $P \subseteq P(X)$ of the powerset of X such that

- each $S \in P$ is non-empty,
- if $S, T \in P$ and $S \cap T \neq \emptyset$, then $S = T$, and
- for every $x \in X$, there is some $S \in P$ with $x \in S$.

The proposition above shows that X/R is a partition of X. Given a partition P, the elements of P are called *partition classes*. When passing from an equivalence relation to a partition, the equivalence classes become the partition classes. We can also pass backwards: every partition determines an equivalence relation whose equivalence classes are the partition classes.

Proposition 3.42. Let X be a set and let P be a partition on X. Define a relation R on X by $x \ R \ y$ iff x and y are in the same P-partition class. Then R is an equivalence relation, and P is the set of equivalence classes of R.

Proof. Exercise. □

3.3 Congruence modulo m

Let $m \in \mathbb{N}$ with $m > 0$. We extend $n \,\texttt{div}\, m$ and $n \,\texttt{mod}\, m$ to allow $n \in \mathbb{Z}$ by insisting that $n \,\texttt{mod}\, m$ is still a natural number less than m, but now $n \,\texttt{div}\, m$ can be negative, so that the equation

$$n = (n \,\texttt{div}\, m)m + (n \,\texttt{mod}\, m)$$

still holds

We define a binary relation R_m on $\mathbb{Z}$ by

$$R_m = \{(x, y) \in \mathbb{Z}^2 \mid x \,\texttt{mod}\, m = y \,\texttt{mod}\, m\}.$$

Equivalently, $x \; R_m \; y$ iff $m \mid (x - y)$. This relation is pronounced "x and y are congruent modulo m". The equivalence classes are sometimes called "residue classes modulo m".

Proposition 3.43. R_m is an equivalence relation.

Proof. We check the three axioms of an equivalence relation in turn.

- Reflexivity. Let $x \in \mathbb{Z}$. Then every number is equal to itself, so in particular

  $$x \,\texttt{mod}\, m = x \,\texttt{mod}\, m$$

  so we have $x \; R_m \; x$ as required.

- Symmetry. Suppose $x \; R_m \; y$. By definition, this means that $x \,\texttt{mod}\, m = y \,\texttt{mod}\, m$. Then since equality is symmetric, we also have $y \,\texttt{mod}\, m = y \,\texttt{mod}\, m$, as required.

- Transitivity. Suppose $x \; R_m \; y$ and $y \; R_m \; z$. Then

  $$x \,\texttt{mod}\, m = y \,\texttt{mod}\, m = z \,\texttt{mod}\, m$$

  so $x \,\texttt{mod}\, m = z \,\texttt{mod}\, m$, so $x \; R_m \; z$.

□

Notation 3.44. The standard notation for congruence modulo m is

$$x \equiv y \pmod{m}.$$

Unfortunately this does not quite fit our general pattern of notation, since it is hard to say what the name of the subset of $\mathbb{Z}^2$ is, but it is convenient and we will use it. When we need a name for it as a subset of $\mathbb{Z}^2$, we can write $(\mathrm{mod}\ m) \subseteq \mathbb{Z}^2$.

There is some scope for confusing $(\mathrm{mod}\ m)$ with the operation $\,\texttt{mod}\, m$. In these notes we will use typewriter font for the operation defined earlier, and serif font for the relation defined just now. It should not cause too many problems.

3.3.1 The residue classes

For any $n \in \mathbb{Z}$, there are m possibilities for the value of $n \,\texttt{mod}\, m$, namely

$$0, 1, 2, \ldots, m - 1.$$

If $n \in \mathbb{Z}$ Thus there are m equivalence classes for congruence modulo m.

Definition 3.45. The term *residue class* $[(\mathrm{mod}\ m)]$ is just a special name for the equivalence classes of $(\mathrm{mod}\ m)$.

Notation 3.46. $\mathbb{Z}_m$ is defined to be $\mathbb{Z}/(\mathrm{mod}\ m)$, the set of residue classes $(\mathrm{mod}\ m)$.

When we need to be clear about the modulus, we write the residue classes as $[i]_m$ instead of $[i]_{(\mathrm{mod}\ m)}$. Usually we can get away with writing $[i]$, i.e.

$$\mathbb{Z}_m = \{[0], [1], \ldots, [m-1]\}.$$

Another notation sometimes used for the equivalence classes of $(\mathrm{mod}\ m)$ is

$$m\mathbb{Z}, 1 + m\mathbb{Z}, 2 + m\mathbb{Z}, \ldots, (m-1) + m\mathbb{Z}.$$

The idea is that

$$m\mathbb{Z} = \{mn \mid n \in \mathbb{Z}\} = \{\ldots, -2m, -m, 0, m, 2m, \ldots\}$$

and

$$a + m\mathbb{Z} = \{a + mn \mid n \in \mathbb{Z}\} = \{\ldots, a-2m, a-m, a, a+m, a+2m, \ldots\}.$$

Remark$^+$. A more standard notation for the set of equivalence classes of $(\mathrm{mod}\ m)$ would be $\mathbb{Z}/m\mathbb{Z}$. To fully explain the framework in which this makes sense would take us far further into algebra than the scope for this course. Our notation is shorter and is also used by some authors. Unfortunately, there is another family of number systems, called the p-adic numbers, which have the standard notation $\mathbb{Z}_p$ (only for p prime). We mention this just in case you see this in books.

3.3.2 Modular arithmetic

Proposition 3.47. Let $a, b, c, d \in \mathbb{Z}$. Suppose

$$a \equiv c \pmod{m} \qquad\qquad b \equiv d \pmod{m}.$$

Then

- $a + b \equiv c + d \mod m$, and
- $ab \equiv cd \mod m$.

Proof. Let's use the alternative characterization of $x \equiv y \pmod{m}$, i.e. $m \mid x - y$. We are given

$$m \mid a - c \qquad\qquad m \mid b - d.$$

Hence we can write

$$a - c = mh \qquad\qquad b - d = mk$$

for some $h, k \in \mathbb{Z}$. But now

$$\begin{aligned}(a+b) - (c+d) &= (a-c) + (b-d)\\ &= m(h-k)\end{aligned}$$

so $m \mid (a+b) - (c+d)$, i.e. $a + b \equiv c + d \pmod{m}$. Also

$$\begin{aligned}ab - cd &= (c+mh)(d+mk) - cd\\ &= m(hd + ck + mhk)\end{aligned}$$

which is a mutliple of m, so $ab \equiv cd \pmod{m}$. □

This is what makes $\mathbb{Z}_m$ a number system: the fact that addition and multiplication make sense on residue classes. Explicitly, it means that we can define addition and multiplication of residue classes in a natural way:

$$[x] + [y] = [x + y] \qquad\qquad [x][y] = [xy].$$

The neutral elements for addition and multiplication are $[0]$ and $[1]$ respectively.

Proposition 3.48. $\mathbb{Z}_m$ admits negatives.

Proof. Let $x \in \mathbb{Z}_m$. Then $x = [n]$ for some $n \in \mathbb{Z}$. Now the negative of x is $[-n]$, because

$$x + [-n] = [n] + [-n] = [n + (-n)] = [0].$$

□

Actually, it is typically much easier to phrase modular arithmetic in terms of congruence between integers, rather than equality between residue classes, and we qualify terms such as 'negative' or 'reciprocal' with $(\mathrm{mod}\ m)$. So the above proposition could be written as follows.

Proposition 3.49. Let $n \in \mathbb{Z}$. Then n admits a negative $(\mathrm{mod}\ m)$.

Proof. Then negative $(\mathrm{mod}\ m)$ of n is just $-n$, since

$$n + (-n) \equiv 0 \pmod{m}.$$

□

3.4 Rational numbers

Recall that the rational numbers are

$$\mathbb{Q} = \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z}, q \neq 0 \right\}.$$

This means that every rational number is represented by a pair $(p, q) \in \mathbb{Z}^2$ where $q \neq 0$. However, rational numbers do not precisely correspond to such pairs, since each rational number has many possible representations. Consider how $\frac{1}{2} = \frac{2}{4} = \frac{3}{6} = \ldots$.

Let us write $\mathbb{Z}_{\neq 0}$ for $\{n \in \mathbb{Z} \mid n \neq 0\}$. Let us define a relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}_{\neq 0}$ by

$$(a, b) \sim (c, d) \iff ad = bc.$$

We can make a new definition of rational numbers.

Definition 3.50. The rational numbers $\mathbb{Q}$ are the equivalence classes $(\mathbb{Z} \times \mathbb{Z}_{\neq 0})/\sim$.

3.4.1 Operations on fractions

The rules for adding and multiplying fractions are

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \qquad\qquad \frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}.$$

We can view these as operations on $\mathbb{Z} \times \mathbb{Z}_{\neq 0}$:

$$(a, b) + (c, d) = (ad + bc, bd) \qquad\qquad (a, b) \times (c, d) = (ac, bd).$$

Proposition 3.51. These $+$ and $\times$ operations respect the equivalence relation $\sim$.

Proof. Suppose $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$. This just means that $ab' = a'b$ and $cd' = c'd$.

▶ *Claim.* $(ad + bc, bd) \sim (a'd' + b'c', b'd')$

Proof of Claim. We just calculate:

$$\begin{aligned}(ad + bc)b'd' &= adb'd' + bcb'd' \\ &= a'bdd' + bb'c'd \\ &= (a'd' + b'c')bd\end{aligned}$$

as required. ◀

▶ *Claim.* $(ac, bd) \sim (a'c', b'd')$.

Proof of Claim. Just calculate:

$$\begin{aligned}(ac)(b'd') &= (ad')(b'c) \\ &= (a'd)(bc') \\ &= (a'c')(bd)\end{aligned}$$

as required. ◀

□

3.5 Normal forms

Let R be an equivalence relation on a set X. A *choice of representatives* is a choice of an element $x_C \in C$ for each equivalence class $C \in X/R$. Sometimes there is a very natural choice of representatives. A choice of representatives that seems particularly canonical might be called a choice of *normal forms*.

3.5.1 Residues

For arithmetic modulo m, it is natural to choose $0, 1, 2, \ldots, m - 1$ as normal forms. These are the possible values of $n \bmod m$, so we call them the *residues* $(\mathrm{mod}\ m)$. These are representatives because no two of them are congruent modulo m, and every integer is congruent to one of them. However, arithmetic operations are not very convenient on residues. For example, if we are working $(\mathrm{mod}\ m)$ then the integer sum or product of two residues is not necessarily another residue, We can still compute the sum and product in terms of residues, but with an extra step of mapping an integer to its residue $(\mathrm{mod}\ m)$ They would be:

- $a +' b = (a + b) \bmod m$, and
- $a \times' b = (ab) \bmod m$.

Now it is a bit less clear that $+'$ is associative, which means that

$$((a + b) \bmod m + c) \bmod m = (a + ((b + c) \bmod m)) \bmod m$$

3.5.2 Reduced fractions

We will consider more carefully some normal forms for representing rational numbers.

Definition 3.52. Let $a, b \in \mathbb{Z}$ with $b \neq 0$. The pair (a, b) is a *reduced fraction* if

- $b > 0$, and
- $\gcd(a, b) = 1$.

We more often say that $\frac{a}{b}$, rather than (a, b), is the reduced fraction. But this is a slight abuse of notation, since so far we have used $\frac{a}{b}$ to denote an underlying rational number whereas it is the *presentation* of the rational number that is or is not reduced.

Remark 3.53. If $\frac{0}{b}$ is a reduced fraction, then $b > 0$ and $1 = \gcd(0, b) = b$. So $\frac{0}{1}$ is the only reduced fraction corresponding to the rational number 0.

Proposition 3.54. Every rational number can be represented by a reduced fraction.

Proof. Let $\frac{p}{q}$ be a rational number, so $p, q \in \mathbb{Z}$, $q \neq 0$. WLOG $q > 0$, since otherwise we could replace (p, q) with $(-p, -q)$, because $\frac{p}{q} = \frac{-p}{-q}$. Let $d = \gcd(p, q)$. Then

$$p = p'd \qquad\qquad q = q'd$$

for some $p', q' \in \mathbb{Z}$ with $q' > 0$. But now

$$\frac{p}{q} = \frac{p'd}{q'd} = \frac{p'}{q'}.$$

This last fraction is in reduced form, since $\gcd(p', q') = 1$ (which is because $(\gcd(p', q')d) \mid d$). □

To prove that the reduced fractions are a sensible choice of normal forms for rational numbers, we need to check that whenever two reduced fractions present the same rational number, then they are actually equal as pairs of integers. This might seem obvious but it is actually a subtle argument using the strengthened form of Euclid's Lemma (or uniqueness of prime factorization, which is roughly the same thing). We include it just for interest and completeness.

Proposition 3.55. Let $(a, b), (c, d) \in \mathbb{Z} \times \mathbb{Z}_{\neq 0}$. Suppose that (a, b) and (c, d) are reduced fractions and that $(a, b) \sim (c, d)$. Then $(a, b) = (c, d)$.

Proof. We are given that:

- $b > 0$, $d > 0$,
- $\gcd(a, b) = 1$, $\gcd(c, d) = 1$, and
- $ad = bc$.

Firstly, a and c are either both positive, both negative, or both 0, since $b, d > 0$. If they are 0, then by the remark above $b = d = 1$ and so both pairs are equal to $(0, 1)$.

WLOG a and c are both positive. Since $ad = bc$, $a \mid bc$. Since $\gcd(a, b) = 1$, by Lemma 2.25, we see that $a \mid c$. By symmetry (i.e. repeating the argument but swapping the roles of (a, b) and (c, d)), we get that $c \mid a$. Since they are both positive, $a = c$. By cancellation, $b = d$, as required. □

3.5.3 Normal forms vs equivalence relations

We now have two ways to think about representing $\mathbb{Q}$ (on a computer, say). We can either think of rational numbers as being represented by:

- pairs $(a, b) \in \mathbb{Z} \times \mathbb{Z}_{\neq 0}$ — but each rational is represented by many different pairs, or
- *reduced* pairs (a, b), $(b > 0, \gcd(a, b) = 1)$ — precisely one representation per rational.

Similarly with the residue classes modulo m. An element of $\mathbb{Z}_m$ is either represented by:

- an integer n — but each residue class is represented by many different integers, or
- an integer n with $0 \leq n < m$ — precisely one representation per residue class.

There is a trade-off between the two choices. Usually, the more redundant representation has simpler formulas for the basic operations. This comes at the cost of occasionally needing to do some work to test for equality. On the other hand, using normal forms trivializes the equality test, but means that basic operations have an additional step to get a normal form.

For example, if we write $+'$ for the addition operation which returns normal forms, we say before that congruence we would have

$$x +' y = (x + y) \bmod m$$

and for rationals we would have

$$(a, b) +' (c, d) = ((ad + bc)/\gcd(ad + bc, bd), bd/\gcd(ad + bc, bd)).$$

(Could you show directly that $+'$ is associative?)

Lecture 4: Modular arithmetic, orders

4.1 Modular arithmetic

Modular arithmetic means treating the sets $\mathbb{Z}_m$ as 'number systems'. Throughout we write m to mean a natural number > 0.

4.1.1 Inverses

We have already seen that $\mathbb{Z}_m$ admits negatives for all of its members, because

$$n + (-n) \equiv 0 \pmod{m}.$$

In terms of normal forms/residues, the negative of i is $m - i$ (unless $i = 0$, in which case the negative is 0 since m is not a residue).

What about reciprocals? A residue class $[n]$ admits a reciprocal in $\mathbb{Z}_m$ (or we say n admits a reciprocal $\pmod m$) iff there exists $k \in \mathbb{Z}$ with

$$kn \equiv 1 \pmod{m}.$$

What we are calling 'reciprocals' are commonly called 'inverses', which is short for 'multiplicative inverse' (not to be confused with 'additive inverses', which we are calling 'negatives').

Proposition 4.56. n admits a reciprocal $\pmod m$ iff $\gcd(n, m) = 1$ (n and m are coprime).

Proof. Since this is an 'if and only if' proposition, we break this proof into two halves.

$\underline{\Longrightarrow}$ Suppose n admits a reciprocal $\pmod m$. So there is $k \in \mathbb{Z}$ such that $nk \equiv 1 \pmod m$. But this means that $m \mid (nk - 1)$, so there is $l \in \mathbb{Z}$ such that

$$nk - 1 = ml.$$

▶ *Claim.* $\gcd(n, m) = 1$.

Proof of Claim. Let d be a common divisor of m and n. Then $d \mid nk$ and $d \mid ml$, so $d \mid (nk - ml) = 1$. ◀

This completes the forward direction.

$\Longleftarrow$ Suppose n and m are coprime. Then by Bézout's Lemma there are $k, l \in \mathbb{Z}$ such that

$$kn + lm = 1.$$

But now

$$\begin{aligned} kn &= 1 - lm \\ &\equiv 1 \pmod{m} \end{aligned}$$

which says that k is a reciprocal of $n \pmod{m}$.

□

Example 4.57. Let us compute $7^{-1} \in \mathbb{Z}_{100}$. 7 and 10 are coprime, so this is possible. First, we find Bézout coefficients.

$$\begin{aligned} 100 &= 14 \times 7 + 2 \\ 7 &= 3 \times 2 + 1 \\ 2 &= 2 \times 1 + 0 \end{aligned}$$

So

$$\begin{aligned} 1 &= 7 - 3 \times 2 \\ &= 7 - 3 \times (100 - 14 \times 7) \\ &= 43 \times 7 - 3 \times 100. \end{aligned}$$

Now we see that

$$\begin{aligned} 43 \times 7 &= 301 \\ &\equiv 1 \pmod{100}. \end{aligned}$$

So $7^{-1} = 43$ in $\mathbb{Z}_{100}$.

Example 4.58. Let us try to compute $64^{-1} \in \mathbb{Z}_{100}$. If we had a reciprocal k, then we would have $100 \mid 64k - 1$. This obviously does not work, since 100 is even but $64k - 1$ is odd. In more general terms, 2 is a common divisor of 100 and 64, so we would deduce that $2 \mid 1$, a contradiction.

We might try to find Bézout coefficients.

$$\begin{aligned} 100 &= 64 + 36 \\ 64 &= 36 + 28 \\ 36 &= 28 + 8 \\ 28 &= 3 \times 8 + 4 \\ 8 &= 2 \times 4 + 0 \end{aligned}$$

So

$$\begin{aligned} 4 &= 28 - 3 \times 8 \\ &= 28 - 3 \times (36 - 28) \\ &= 4 \times 28 - 3 \times 36 \\ &= 4 \times (64 - 36) - 3 \times 36 \\ &= 4 \times 64 - 7 \times 36 \\ &= 4 \times 64 - 7 \times (100 - 64) \\ &= 11 \times 64 - 7 \times 100. \end{aligned}$$

We have found that $11 \times 64 \equiv 4 \pmod{100}$, and this is 'as close' as we can get to finding a reciprocal, since $\gcd(64, 100) = 4$.

4.1.2 $\mathbb{Z}_p$ as a field

An interesting special case is when the modulus is a prime number.

Proposition 4.59. Let p be a prime number. Then $\mathbb{Z}_p$ is a field.

Proof. The only thing to check (other than $0 \neq 1$) is that every non-zero element has a reciprocal. Let n represent a non-zero residue class in $\mathbb{Z}_p$. Then $n \not\equiv 0 \pmod{p}$ means that $p \not| n$, and p is prime so $\gcd(n, p) = 1$. Thus by the argument above, n has an reciprocal $\pmod{p}$. □

As a consequence of this, solving linear equations in $\mathbb{Z}_p$ is fairly simple.

Example 4.60. Let us find all $x \in \mathbb{Z}$ such that $6x \equiv 5 \pmod{11}$. We start by finding a reciprocal $6^{-1} \in \mathbb{Z}_{11}$. Either spot it by guessing or use Euclid's algorithm:

$$\begin{aligned} 11 &= 6 + 5 \\ 6 &= 5 + 1 \\ 5 &= 5 \times 1 + 0. \end{aligned}$$

so

$$\begin{aligned} 1 &= 6 - 5 \\ &= 6 - (11 - 6) \\ &= 2 \times 6 - 11 \end{aligned}$$

so $2 \times 6 = 12 \equiv 1 \pmod{11}$.

Now we 'divide' both sides of equation by 6.

$$\begin{aligned} & 6x \equiv 5 \pmod{11} \\ \iff & 2 \times 6x \equiv 2 \times 5 \pmod{11} \\ \iff & x \equiv 10 \pmod{11}. \end{aligned}$$

So x is a solution to the equation iff $x \equiv 10 \pmod{11}$. In other words, the solutions are all $x \in \mathbb{Z}$ of the form

$$x = 10 + 11k$$

for some $k \in \mathbb{Z}$.

Remark 4.61. We mentioned before that reciprocals are unique when they exist. Let us write out a proof of this for arithmetic modulo m (though the same argument works for more general number systems). Let $a \in \mathbb{Z}$ and suppose that $x, y \in \mathbb{Z}$ are both reciprocals of a modulo m. By definition of reciprocal, this means we have both

$$xa \equiv 1 \pmod{m}, \qquad ya \equiv 1 \pmod{m}.$$

But now we can calculate as follows.

$$\begin{aligned} x &\equiv x \times 1 \pmod{m} \\ &\equiv x \times (y \times a) \pmod{m} \\ &\equiv y \times (x \times a) \pmod{m} \\ &\equiv y \pmod{m}. \end{aligned}$$

So $x \equiv y \pmod{m}$. Thus any two reciprocals of a modulo m belong to the same residue class modulo m.

4.1.3 Simultaneous congruences

Here is a technique for solving problems of the following particular form. We are given integers $a, b \in \mathbb{Z}$, non-zero naturals $m, n \in \mathbb{N}$ with $\gcd(m, n) = 1$. The task is to find all integers $x \in \mathbb{Z}$ with

$$x \equiv a \pmod{m} \qquad \text{and} \qquad x \equiv b \pmod{n}.$$

To do this, find Bézout coefficients $h, k \in \mathbb{Z}$:

$$hm + kn = 1.$$

Now consider

$$x_0 = bhm + akn.$$

► *Claim.* x_0 is one solution to the problem.

Proof of Claim.

$$\begin{aligned} x_0 &= bhm + akn \\ &\equiv akn \pmod{m} \\ &\equiv a(1 - hm) \pmod{m} \\ &\equiv a \pmod{m}. \end{aligned}$$

Also

$$\begin{aligned} x_0 &= bhm + akn \\ &\equiv bhm \pmod{n} \\ &\equiv b(1 - kn) \pmod{n} \\ &\equiv b \pmod{n}. \end{aligned}$$

◄

Now suppose x is any (other) solution. Then

$$x - x_0 \equiv 0 \pmod{m} \qquad x - x_0 \equiv 0 \pmod{n}$$

i.e. $m \mid (x - x_0)$ and $n \mid (x - x_0)$. But m and n are coprime, so $mn \mid (x - x_0)$. So if x is a solution, then $x \equiv x_0 \pmod{mn}$. It is easy to check that the converse is true, that all numbers of the form $x_0 + cmn$ for $c \in \mathbb{Z}$ are solutions. This fully characterizes the solutions.

Example 4.62. Find all $x \in \mathbb{Z}$ such that

$$x \equiv 3 \pmod{10} \qquad x \equiv 1 \pmod{3}.$$

We notice easily that

$$1 = 10 - 3 \times 3.$$

Now consider

$$10 - 3 \times 3 \times 3 = -17.$$

Then -17 is a solution, and for any solution x,

$$x + 17 \equiv 3 + 7 \equiv 0 \pmod{10}$$

and

$$x + 17 \equiv 1 + 2 \equiv 0 \pmod{3}.$$

Since 3 and 10 are coprime, we have $30 \mid x + 17$. Hence the solutions are among the integers x of the form

$$x = -17 + 30k$$

for $k \in \mathbb{Z}$. But is easy to check that all of these are solutions. Equivalently, the solutions are all integers of the form

$$x = 13 + 30k$$

for $k \in \mathbb{Z}$.

Remark 4.63. When solving equations, we normally assume we have a solution x and then try to narrow down what x might be. This means proving a statement like "if x is a solution, then $x = n$", which does not logically imply that "if $x = n$, then x is a solution". Thus it is necessary to check the second statement separately. Alternatively, one can take care to argue directly a statement like "x is a solution iff $x = n$" by using a chain of equivalences. For very simple problems this is almost automatic. For solving simultaneous congruences as above, it can be done but it is slightly easier just to check that the potential solutions are actual solutions afterwards.

Remark 4.64. One can also consider simultaneous congruences where the moduli are not coprime using a refinement of this technique.

4.1.4 Polynomial equations

When working over $\mathbb{R}$, a quadratic polynomial has at most two roots ('zeroes'). The same is true working over $\mathbb{Z}_p$ for p prime. (For general $\mathbb{Z}_m$, the situation is more complicated.)

Example 4.65. Let us solve the congruence $x^2 \equiv 1 \pmod{5}$.

Approach 1: If we let $x' = x + 5$, then

$$(x')^2 = (x+5)^2 = x^2 + 10x + 25 \equiv x^2 \pmod{5}.$$

Thus, whether $x \in \mathbb{Z}$ is a solution depends solely on the residue $x \bmod 5 \in \{0, 1, 2, 3, 4\}$. Since 5 is small, it is not so time consuming to check all of these:

$$\begin{aligned}
0^2 &= 0 \\
&\not\equiv 1 \pmod{5} \\
1^2 &= 1 \\
&\equiv 1 \pmod{5} \\
2^2 &= 4 \\
&\not\equiv 1 \pmod{5} \\
3^2 &= 9 \\
&\equiv 4 \pmod{5} \\
&\not\equiv 1 \pmod{5} \\
4^2 &= 16 \\
&\equiv 1 \pmod{5}
\end{aligned}$$

So x is a solution iff $x \equiv 1 \pmod 5$ or $x \equiv 4 \pmod 5$. So the solutions are all $x \in \mathbb{Z}$ of the form

$$x = 1 + 5k \qquad \text{or} \qquad x = 4 + 5k$$

for some $k \in \mathbb{Z}$.

Approach 2: We rearrange and factorize:

$$\begin{aligned} & x^2 \equiv 1 \pmod 5 \\ \iff & x^2 - 1 \equiv 0 \pmod 5 \\ \iff & (x-1)(x+1) \equiv 0 \pmod 5 \\ \iff & 5 \mid (x-1)(x+1). \end{aligned}$$

But 5 is a prime number, so this is iff $5 \mid x-1$ or $5 \mid x+1$. So the solutions are all integers x of the form

$$5k \pm 1$$

for some $k \in \mathbb{Z}$. (This is just a different way of writing the same set of solutions we arrived at before).

4.2 Ordering

Last time we looked at relations and in particular the notion of *equivalence relation*. Now let us look at another kind of relation.

In the following, let X be a set and R a binary relation on X.

Definition 4.66. A relation R is *antisymmetric* if, whenever both $x \ R \ y$ and $y \ R \ y$, we have $x = y$.

Definition 4.67. A *partial order* on X is a binary relation R on X which is reflexive, antisymmetric, and transitive.

Remark 4.68. The formal 'difference' between a partial order and an equivalence relation is that, where an equivalence relation is required to be symmetric, a partial order is required to be antisymmetric instead.

Definition 4.69. A *total order* is a partial order R for which, whenever $x, y \in X$, at least one of $x \ R \ y$ or $y \ R \ x$ holds.

Example 4.70. If X is any set, then the equality relation on X is a partial order. (This is called a 'discrete' partial order). It is not a total order unless X is empty or has precisely 1 element.

Example 4.71. Let X be any of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$. Then the usual meaning of $\leq$ is a partial order on X. Writing out the axioms, we just have to convince ourselves that:

- we always have $x \leq x$ (the relation is called 'less than or equal to'),
- whenever $x \leq y$ and $y \leq x$ we have $x = y$, and
- if $x \leq y$ and $y \leq z$ then also $x \leq z$.

In addition, $\leq$ is a total order.

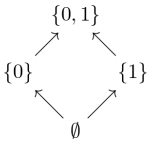
Example 4.72. Let X be any set. Then the subset relation $\subseteq$ is a partial order on the powerset $P(X)$.

- Every set is a subset of itself, $S \subseteq S$.
- If $S \subseteq T$ and $T \subseteq S$, then S and T have exactly the same members, so $S = T$.
- If $S \subseteq T$ and $T \subseteq U$, then obviously $S \subseteq U$.

It is not a total order if X has at least two elements.

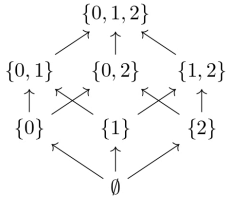
4.2.1 Diagrams of partial orders

Just like we drew numbers on a number line, we can draw a picture of a partial order. It will not be a line unless it is a total order. This is a little like how we depicted a general relation as a graph but with some differences. Firstly, we tend to arrange the elements of the set so that all the arrows roughly point upwards (or maybe left to right). For the subset relation $\subseteq$ on $P(\{0,1\})$, the picture looks like this.



When $x \ R \ y$, we draw an arrow from x to y. The second difference is that we do not draw every arrow: we omit any arrow which is implicitly there by transitivity or reflexivity. For example, above we did not draw an arrow $\emptyset \to \{0,1\}$ despite $\emptyset \subseteq \{0,1\}$, because it would clutter the diagram. Since we already have $\emptyset \to \{0\}$ and $\{0\} \to \{0,1\}$, there is no need for it.

We can also draw $\subseteq$ on $P(\{0,1,2\})$.



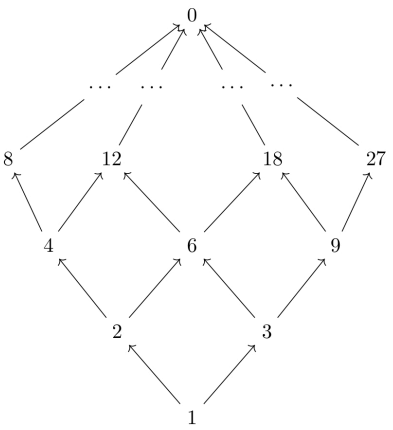
If we did not omit the arrows that were consequences of transitivity, this would be even messier.

Example 4.73. The divisibility relation on $\mathbb{N}$ is a partial order.

- For any $n \in \mathbb{N}$, $n \mid n$.
- If $m, n \in \mathbb{N}$ and both $m \mid n$ and $n \mid m$, then $m = n$,
- If $a \mid b$ and $b \mid c$, then we can write $b = ah$ and $c = bk$, so $c = ahk$, i.e. $a \mid c$.

It is not a total order since, for example, $2 \not| 3$ and $3 \not| 2$.

This is an infinite partial order, but we can draw part of it.



If we continued the diagram, we would have all the prime numbers forming a row just above 1. For each prime number p, there is chain of prime powers straight up and capped by 0: $1 \to p \to p^2 \to p^3 \to \ldots 0$. Note that 1 is the 'least' element in this order, and 0 is the 'greatest'. (Does this explain the 'greatest' in 'greatest common divisor' given that $\gcd(0,0) = 0$ despite every natural number dividing 0?)

Lecture 5: Functions

5.1 Sets

To help us with our study of relations and functions, let us set a few more set notations. Fix a set X. Given subsets $A, B \subseteq X$, we define the *intersection* of A and B to be

$$A \cap B = \{x \in X \mid x \in A \wedge x \in B\},$$

where, recall, the symbol $\wedge$ is 'logical and'. So $A \cap B$ is the set of all elements of X that are in *both* of A and B. We define the *union* of A and B to be

$$A \cup B = \{x \in X \mid x \in A \vee x \in B\},$$

where, recall, the symbol $\vee$ is 'logical or'. So $A \cup B$ is the set of all elements of X that are in *either* A or B (or both). We define the *difference* to be

$$A \backslash B = \{x \in X \mid x \in A \wedge x \notin B\}.$$

So $A \backslash B$ is the set of all elements of A which are not in B. The *symmetric difference* is

$$A \triangle B = \{x \in X \mid (x \in A \wedge x \notin B) \vee (x \notin A \wedge x \in B)\}.$$

(The condition in the definition of $A \triangle B$ could be written as $x \in A \,\mathrm{XOR}\, x \in B$ if we use the boolean 'exclusive or'). So $A \triangle B$ is the set of elements of X which are in *precisely one* of A or B.

When X is a finite set, we write $|X|$ for the number of elements in X.

5.2 Functions

5.2.1 General relations

So far we have only considered *binary* relations that are on a *single* set. We can generalize relations to any number of possibly different sets. We will not need the fully general notion, we will only need to consider binary relations from one set to another.

Definition 5.74. Let X and Y be sets. A (*binary*) *relation from X to Y* is a subset $R \subseteq X \times Y$.

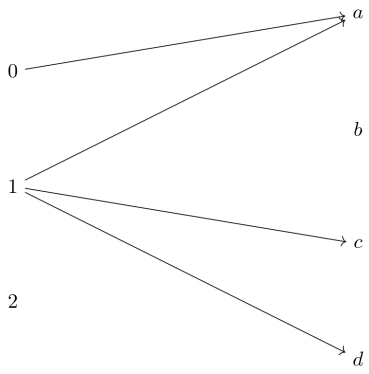
Thus, what we have been calling a 'binary relation on X' is the same as a 'binary relation from X to X'.

When we draw a relation from X to Y, we have to draw the elements of both sets. Then all the arrows we draw will point from the X-nodes to the Y-nodes.

Example 5.75. Let $X = \{0, 1, 2\}$ and $Y = \{a, b, c, d\}$ be sets. Consider the relation R from X to Y given by

$$R = \{(0, a), (1, a), (1, c), (1, d)\}.$$

We can depict this as directed, (bipartite) graph as follows.



5.2.2 Functions

Let X and Y be sets.

Definition 5.76. A *function f from X to Y*, written $f : X \to Y$, is a relation f from X to Y which is:

- *total*: for all $x \in X$, there is a $y \in Y$ with $x \; f \; y$; and
- *single-valued*: if $x \; f \; y_1$ and $x \; f \; y_2$ then $y_1 = y_2$.

More concisely, a relation f from X to Y is a function if for every $x \in X$ *there is a unique* $y \in Y$ with $x \; f \; y$.

Notation 5.77. If $f : X \to Y$ is a function, we write $f(x) = y$ instead of $(x, y) \in f$ or $x \; f \; y$.

When talking about a function $f : X \to Y$, the set X is called the *domain* of f and Y is called the *codomain* of f.

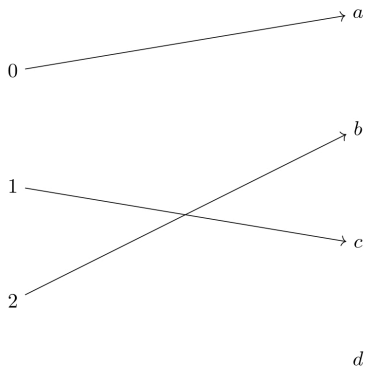
Remark$^+$. As we have introduced them, functions (and also relations) are always defined with respect to a particular domain and codomain, and we can only say something is a function with reference to a choice of domain and codomain. Not all authors follow this convention. Also, *single-valued* is often used to mean the conjunction of what we call 'total' and 'single-valued'. But (a) this is also called 'functional' and is just what it means to be a function, (b) that usage is mainly useful when needing to contrast with 'multivalued functions' (which we will not consider here), and (c) our usage of 'total' is completely standard.

Example 5.78. The relation from Example 5.75 is not a function $\{0, 1, 2\} \to \{a, b, c, d\}$. It is not total because there is no arrow pointing out of 2. It is also not single-valued because there are multiple arrows pointing out of 1.

Example 5.79. With $X = \{0, 1, 2\}$ and $Y = \{a, b, c, d\}$ as before, consider the relation f from X to Y given by

$$f = \{(0, a), (1, c), (2, b)\}.$$

This is indeed a function $X \to Y$. We can draw it as a graph like this:



Definition 5.80. Given a function $f : X \to Y$, the *image* (or *range*) of f is the set

$$\begin{aligned} f[X] &= \{f(x) \mid x \in X\} \\ &= \{y \in Y \mid \exists x \in X.(x, y) \in f\}. \end{aligned}$$

The image of f is always a subset of the codomain $f[X] \subseteq Y$. When $x \in X$, we also say that $f(x)$ is the 'image' of the element x.

Example 5.81. Continuing Example 5.78, the image of f is the set

$$\{a, b, c\}.$$

Definition 5.82. Given a function $f : X \to Y$ and a subset $B \subseteq Y$ of the codomain, the *preimage* of B is the subset $f^{-1}(B)$ of the domain X given by

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

When B is a singleton set $\{y\}$, we sometimes call $f^{-1}(\{y\})$ the 'preimage of y' rather than 'preimage of $\{y\}$'. (In some settings, this is also called the *fibre* of f over y).

Definition 5.83. For any set X, the *identity function on* X is the function $id_X : X \to X$ where

$$\mathsf{id}_X = \{(x, x) \mid x \in X\}.$$

In other words, id_X is given by $\mathsf{id}_X(x) = x$.

Definition 5.84. Let X be a set and let $A \subseteq X$ be any subset. The *characteristic function of* A is the function $\chi_A : X \to \{0, 1\}$ given by

$$\chi_A = \{(a, 1) \mid a \in A\} \cup \{(x, 1) \mid x \in X \backslash A\}.$$

In other words, $\chi_A : X \to \{0, 1\}$ is given by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases}$$

The symbol χ is the Greek letter 'chi'. A suggested pronunciation is /kʌɪ/ ("kigh").

5.2.3 Defining particular functions

A function is by definition a certain kind of binary relation, so a function $X \to Y$ can be specified by explicitly writing down its elements as a subset of $X \times Y$. For functions whose domain is infinite, we have to make use of abbreviations.

Example 5.85. There is a function $\texttt{double} : \mathbb{N} \to \mathbb{N}$ given by

$$\texttt{double} = \{(n, 2n) \mid n \in \mathbb{N}\}.$$

A simpler way to write functions is probably already familiar to you.

Example 5.86. The function `double` can equivalently be defined as the unique function $\texttt{double} : \mathbb{N} \to \mathbb{N}$ satisfying

$$\texttt{double}(n) = 2n$$

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

Another notation for describing functions is the 'maps to' arrow: $\mapsto$. Compare this with the ordinary function arrow. We write

$$f : X \to Y,$$

without the small vertical bar on the tail of the arrow, to mean that f is a function with domain X and codomain Y. When we have a particular function in mind, we write

$$x \mapsto y$$

to mean that the element x is sent to the element y by the function, i.e. $f(x) = y$ or $(x, y) \in f$. We use this notation both for particular instances of f and for giving the general pattern.

Example 5.87. The function f from Example 5.78 can be defined as the function $f : \{0, 1, 2\} \to \{a, b, c, d\}$ given by

$$0 \mapsto a \qquad 1 \mapsto c \qquad 2 \mapsto b.$$

In this case, because $0, 1, 2$ are 'constants' rather than 'variables', i.e. particular elements of $X = \{0, 1, 2\}$, so the expression $0 \mapsto a$ tells us that $(0, a) \in f$, etc., so that

$$f = \{(0, a), (1, c), (2, b)\}.$$

Example 5.88. We can set out the definition of the function $\texttt{double} : \mathbb{N} \to \mathbb{N}$ from Example 5.86 as like this:

$$\begin{aligned} \texttt{double} : \mathbb{N} &\to \mathbb{N} \\ n &\mapsto 2n. \end{aligned}$$

In this case '$n \mapsto 2n$' presents the general pattern of the function `double`, i.e. that

$$\texttt{double} = \{(n, 2n) \mid n \in \mathbb{N}\}.$$

The 'maps to' arrow is common in mathematical texts. In Computer Science contexts, it also common to use 'lambda' notation for functions.

Example 5.89. The function $\texttt{double} : \mathbb{N} \to \mathbb{N}$ from Example 5.86 can be defined as

$$\texttt{double} = \lambda n.2n$$

An expression "$\lambda x.M$" means "the function which, given an argument x, returns the value of the expression M". Here M is usually an expression involving the variable x.

The λ-expression, and the expression with $\mapsto$ giving the general case, can be used to describe a function without picking a name for the function. Hence λ-expressions are also called 'anonymous functions'.

5.2.4 Sets of functions

Let X and Y be sets.

Definition 5.90. The set Y^X is defined to be

$$Y^X = \{f \subseteq X \times Y \mid f \text{ is a function } X \to Y\}.$$

Remark 5.91. Other notations for the set of function $X \to Y$ include $X \to Y$, $X \Rightarrow Y$, $[X, Y]$, $\hom(X, Y)$, $\mathsf{Set}(X, Y)$, ...

Proposition 5.92. Let X and Y be finite sets. Suppose X has m elements and Y has n elements. Then Y^X has n^m elements.

Proof (sketch). Let us write $X = \{x_1, x_2, \ldots, x_m\}$ where the x_i are distinct. Then a function $f : X \to Y$ is determined by $f(x_1), f(x_2), \ldots, f(x_m)$. There are n choices for each of those values. We can make the choices independently of each other, so the number of functions is

$$\underbrace{n \times \ldots \times n}_{m} = n^m.$$

□

Example 5.93. The notation X^2, for the set of ordered pairs, can be seen as a special case of the function set notation. If we imagine 2 as standing for a set with two elements, say $\{0, 1\}$, then the set of function $X^{\{0,1\}}$ is 'essentially the same' as the set of ordered pairs, because a function $f : \{0, 1\} \to X$ is determined by the elements $f(0), f(1) \in X$ in a fixed order.

5.3 Operations on functions

5.3.1 Restriction

Let $f : X \to Y$ be a function. Suppose $A \subseteq X$ is a subset of the domain.

Definition 5.94. The *restriction* of f to A is the function $f \restriction_A : A \to Y$ given by

$$f \restriction_A = f \cap A \times Y.$$

(Another common notation for the restriction is $f|_A$).

To spell this out some more, the restriction $f \restriction_A$ is a function $A \to Y$ which satisfies $f \restriction_A (x) = f(x)$ for all $x \in A$.

5.3.2 Composition

Let $f : X \to Y$ and $g : Y \to Z$ be two functions, where the *codomain* of f is equal to the *domain* of g. We sometimes call f and g a *composable* pair.

Definition 5.95. The *composite* of f and g is the function $g \circ f : X \to Z$ given by

$$g \circ f = \{(x, z) \in X \times Z \mid \exists y \in Y. (x, y) \in f \wedge (y, z) \in g\}.$$

In more simple terms, the composition $g \circ f$ is the function $X \to Z$ satisfying

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

5.4 Types of function

Let $f : X \to Y$ be a function.

Definition 5.96. f is an *injection* (or f is *injective*) if

$$f(x) = f(y) \implies x = y.$$

Definition 5.97. f is a *surjection* (or f is *surjective*) if

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

Equivalently, the image is all of Y: $f[X] = Y$.

Having an injection or a surjection immediately tells us something about the relative sizes of the two sets.

- If f is an injection, then $|X| \leq |Y|$.
- If f is a surjection, then $|Y| \leq |X|$.

Definition 5.98. f is a *bijection* (or f is *bijective*) if it is both an injection and a surjection.

If there is a bijection $f : X \to Y$ then this means that X and Y have exactly the same number of elements.

Definition 5.99. An *inverse* for f is a function $g : Y \to X$ (N.B. the domain and codomain are swapped relative to X) such that $f \circ g = \mathsf{id}_Y$ and $g \circ f = \mathsf{id}_X$. In other words, g satisfies $f(g(y)) = y$ and $g(f(x)) = x$.

Remark 5.100. When an inverse to f exists, often we denote it f^{-1}. Do not confuse this with the notation $f^{-1}(B)$ for the preimage of $B \subseteq Y$, which does not require f to have an inverse $f^{-1} : Y \to X$.

Proposition 5.101. Let $f : X \to Y$ be a function. Then f has an inverse iff f is a bijection.

Proof. $\underline{\implies}$ Let $g : Y \to X$ be an inverse for f.

Injectivity Suppose $f(x_1) = f(x_2)$. Then $g(f(x_1)) = g(f(x_2))$. But g is an inverse to f, so

$$x_1 = g(f(x_1)) = g(f(x_2)) = x_2,$$

as required.

Surjectivity Let $y \in Y$. We need to find $x \in X$ such that $f(x) = y$. So let $x = g(y)$. Then, since g is an inverse,

$$f(x) = f(g(y)) = y,$$

as required.

Thus, since f is both injective and surjective, it is a bijection.

$\underline{\impliedby}$ Suppose f is a bijection. We need to exhibit an inverse to f. Either of the following approaches is fine.

Approach 1 We define a function $g : Y \to X$ as follows. For $y \in Y$, since f is surjective there is at least one $x \in X$ with $f(x) = y$. Let $g(y)$ be an arbitrarily chosen such x. By making such a choice for each $y \in Y$, we define a function g.
Now we show that g is an inverse. By construction, $f(g(y)) = y$. We need to show that $g(f(x)) = x$. Observe that

$$f(g(f(x))) = f(x)$$

by definition of $g(f(x))$. But f is injective, so we deduce

$$g(f(x)) = x$$

as required.

Approach 2 We define a relation g from Y to X by

$$g = \{(y, x) \in Y \times X \mid (x, y) \in f\}.$$

► *Claim.* g is a function.

Proof of Claim. g is total because if $y \in Y$, then by surjectivity of f there is an $x \in X$ with $(x, y) \in f$, hence $(y, x) \in g$. g is single-valued since if $(y, x_1), (y, x_2) \in g$, then $(x_1, y), (x_2, y) \in f$, and then since f is injective we have $x_1 = x_2$. ◄

Now we need to check that $g \circ f = \mathsf{id}_X$ and $f \circ g = \mathsf{id}_Y$. For the former, we have $(x, f(x)) \in f$ and $(f(x), x) \in g$, so $(x, x) \in g \circ f$. For the latter, to see that $(y, y) \in f \circ g$ we need there to be an $x \in X$ such that $(y, x) \in g$ and $(x, y) \in f$, but f is surjective (or g is total) so this exists. □

Remark 5.102. In the proof above, you will probably have found Approach 1 much easier to follow. However, Approach 2 is included for interest since it demonstrates how the totality axiom for functions is 'dual' to surjectivity and the single-valuedness axiom is 'dual' to injectivity.

5.4.1 Examples of bijections

Characteristic functions

The key to saying that two kinds of things are 'essentially the same' is to construct a bijection between two sets. In Definition 5.84, we showed how a subset $A \subseteq X$ gives rise to a function $\chi_A : X \to \{0, 1\}$. We will show that this makes "subsets of X" and "functions $X \to \{0, 1\}$" essentially the same thing.

Proposition 5.103. $A \mapsto \chi_A$ defines a bijection $P(X) \to \{0, 1\}^X$.

Proof. It is obvious that $A \mapsto \chi_A$ defines a function $P(X) \to \{0, 1\}^X$.

► *Claim.* $\chi_{(-)}$ is injective.

Proof of Claim. Suppose $\chi_A = \chi_B$, where $A, B \subseteq X$. Then, for any $x \in X$, we have the following chain of equivalences:

$$x \in A \iff \chi_A(x) = 1 \iff \chi_B(x) = 1 \iff x \in B.$$

Hence A and B have precisely the same elements, so $A = B$. ◄

► *Claim.* $\chi_{(-)}$ is surjective.

Proof of Claim. Let $f : X \to \{0,1\}$ be a function. Then we need to show that $f = \chi_A$ for some $A \subseteq X$. To see this, let $A = f^{-1}(\{1\})$ be the preimage of 1. Then

$$\chi_A(x) = 1 \iff x \in \{x \in X \mid f(x) = 1\} \iff f(x) = 1$$

and

$$\chi_A(x) = 0 \iff x \notin \{x \in X \mid f(x) = 1\} \iff f(x) = 0.$$

Thus $\chi_A(x) = f(x)$ for all $x \in X$, so $f = \chi_A$. ◀

Since $\chi_{(-)}$ is both injective and surjective, it is bijective. □

Residues and residue classes

Proposition 5.104. Let $m \in \mathbb{N}$ with $m > 0$. Then

$$\begin{aligned} \{0, 1, \ldots, m-1\} &\to \mathbb{Z}_m \\ i &\mapsto [i]_m \end{aligned}$$

is a bijection.

Proof. It is evidently a function.

Injectivity Let $i, j \in \{0, 1, \ldots, m-1\}$. To have $[i]_m = [j]_m$ is to have $i \equiv j \pmod m$. This means that $m \mid i - j$. But $-m < i - j < m$ by the assumption that $i, j \in \{0, 1, \ldots, m-1\}$, so we must have $i - j = 0$, i.e. $i = j$.

Surjectivity Let $[n]_m \in \mathbb{Z}_m$ be an arbitrary residue class. Then $n \equiv n \bmod m \pmod m$, i.e. $[n]_m = [n \bmod m]_m$, where $n \bmod m \in \{0, 1, \ldots, m-1\}$, as required.

□

5.5 Counting

Functions can be used to break down counting problems. The formal rule that can help us count the elements of a set is:

$$|X| = \sum_{y \in Y} |f^{-1}(\{y\})|$$

for any function $f : X \to Y$. A special case of this is when f is a bijection, for then each $|f^{-1}(\{y\})| = 1$ and we get

$$|X| = |Y|.$$

Example 5.105. Count the number of equivalence relations on $X = \{0, 1, 2\}$. Write E for the set of equivalence relations on that set. Let $f : E \to \mathbb{N}$ map an equivalence relation R to $|X/R|$, the number of R-equivalence classes. Then $f^{-1}(\{n\}) = \emptyset$ for $n = 0$ or $n \geq 4$. Now

$$|f^{-1}(\{1\})| = 1$$

because it contains only the maximal relation, and

$$|f^{-1}(\{3\})| = 1$$

because it contains only the equality relation. It remains to compute $|(f^{-1}(\{2\})|$. If there are two equivalence classes, then one must be a singleton and the other must have two elements. Knowing the singleton fixes the whole thing, and there are the 3 choices, so this preimage has size 3. Thus the final answer is $1 + 1 + 3 = 5$.

Example 5.106. An alternative proof of Proposition 5.92: for finite sets X, Y, show that $|Y^X| = |Y|^{|X|}$. We use induction on $|X|$. The base case is easy: $|Y|^0 = 1$ for any Y, and there is a unique function $\emptyset \to Y$ for any Y. For the inductive step, suppose $|X| > 0$, let $x_0 \in X$ be any element, and let

$$\begin{aligned} f : Y^X &\to Y \times Y^{X \setminus \{x_0\}} \\ \phi &\mapsto (f(x_0), f \restriction_{X \setminus \{x_0\}}) \end{aligned}$$

It is easy to see that f is a bijection. Thus

$$|Y^X| = |Y| \times |Y^{X \setminus \{x_0\}}|.$$

But, by induction hypothesis,

$$|Y^{X \setminus \{x_0\}}| = |Y|^{|X \setminus \{x_0\}|} = |Y|^{|X|-1}$$

whence the result follows.


BREAK

Lecture 6: Reals

6.1 Discussion

Recall that we write $\mathbb{R}$ for the set of real numbers and that this a superset of the naturals, integers, and rationals.

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

Furthermore, $\mathbb{R} \subseteq \mathbb{C}$, the complex numbers. We will not use the complex numbers in this part of the module, but you will need to use them in Linear Algebra.

The term 'real' number arose from the contrast with general complex numbers. The latter are 'complex' because they consist of two parts: one 'real' and one 'imaginary'. The idea is that the real numbers have an interpretation as physical magnitudes such as distances and displacements whereas properly complex numbers (those with non-zero imaginary part) do not. The terminology is a bit problematic. Firstly, it obscures the fact that complex numbers are extremely useful in engineering and natural sciences. Secondly, the real numbers are actually a highly idealized abstraction with which we only interact indirectly through rational approximations.

Before trying to answer the question of what the real numbers are, it is worth considering what sort of answer we have come up with for the other number systems that we have considered up to now: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{Z}_m$. For $\mathbb{N}$ and $\mathbb{Z}$, we did not really explain at all what they are. Instead, I trusted that you would have some intuitions about what I meant when I wrote

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

We did link $\mathbb{N}$ and $\mathbb{Z}$ together with $\mathbb{N}$ definable as a subset of $\mathbb{Z}$. For $\mathbb{Q}$, at first we relied on prior intuitions and took it to be a definite subset of $\mathbb{R}$, but knowing that we would be struggling to define $\mathbb{R}$ later, we gave an alternative definition as a *concrete construction* based on $\mathbb{Z}$ (equivalence classes of pairs of integers). For $\mathbb{Z}_m$, we only defined it by a concrete construction.

It is typical in formal sciences that we rely on intuition until such time as it causes us to run into confusion or paradoxes. Arguably, the natural numbers are the safest number system to rely on our intuition for. We do interact with them relatively directly in physical space: the natural numbers are the numbers we use to describe the sizes of finite sets. We might be less sure of the integers. The following is a typical construction that might be found in a standard course on the foundations of mathematics or set theory. Consider the relation $\sim$ on $\mathbb{N}^2$ given by $(a, b) \sim (c, d)$ iff $a + d = b + c$. As an exercise, show this is an equivalence relation. If you have sufficient trust in your intuition for $\mathbb{Z}$, show that $[(a, b)]_\sim \mapsto a - b$ gives a well-defined bijection $\mathbb{N}^2/\sim \to \mathbb{Z}$. If you are unsure what $\mathbb{Z}$ is, you can take $\mathbb{N}^2/\sim$ as a definition. So we can reduce $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{Z}_m$ to concrete constructions relative to $\mathbb{N}$, only relying on prior intuitions for $\mathbb{N}$.

Aside from relying on intuition or giving a concrete construction, a third kind of answer we might have given is to use the axiomatic method. This means specifying some propositions we expect our object to satisfy (axioms), which should be enough to determine our desired object uniquely, and then deriving further propositions from the axioms. For example, we could have defined $\mathbb{Q}$ to be as 'the smallest field containing $\mathbb{Z}$', and we could have defined $\mathbb{Z}$ as 'the smallest ring containing $\mathbb{N}$'. We can even define $\mathbb{N}$ this way as a model of a certain logical theory. Axiomatic definitions are nice because they give you elegant properties of the object that you can work with immediately. However, it is not obvious that axiomatic definitions refer to mathematical objects that necessarily exist: it is necessary to prove separately that a 'smallest field generated by $\mathbb{Z}$' is something that exists, and this is done using a concrete construction.

Both axiomatic and concrete definitions of the real numbers exist. The approach we take here will be far from comprehensive. Roughly speaking, we will take an informal axiomatic approach. For us, the real numbers will be an ordered field containing $\mathbb{Q}$ and satisfying an extra condition — the 'Shrinking Interval Property' (SIP). You will not be examined on knowledge of the SIP, but we will discuss it in order to motivate the use of interval bisection algorithms for calculating specific real numbers which you should be able to perform. The exercise sheets will contain fragments of concrete constructions that can be used. We will not at all cover the proofs that any axiomatic definitions single out an essentially unique concrete model.

6.2 The real numbers as an ordered field

The real numbers are a field. Recall that this means they have operations $+$ and $\times$, which are associative and commutative, which admit neutral elements 0 and 1 respectively, and for which $\times$ distributes over $+$, and moreover all real numbers have a negative and all non-zero real numbers have a reciprocal. In other words, they support all the usual laws of algebra that you would expect.

Like its subsets $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$, the reals are *ordered*. The usual 'less than or equal to' relation $\leq$ is a partial order on $\mathbb{R}$. This partial order is also a total order, recall that this means that for all $a, b \in \mathbb{R}$ we have either $a \leq b$ or $b \leq a$ (this corresponds to the fact that we can draw the real numbers on a line). More than just a partial order, $\leq$ is compatible with the algebraic operations of $\mathbb{R}$ in ways you are probably familiar with.

- If $a \leq a'$ and $b \leq b'$, then $a + b \leq a' + b'$.
- If $0 \leq a \leq a'$ and $0 \leq b \leq b'$, then $0 \leq ab \leq a'b'$.

When multiplying numbers that are possibly negative we have to be careful. For example, $-2 \leq 1$ and $-3 \leq 4$, but

$$(-2) \times (-3) = 6 > 4 = 1 \times 4$$

so here the order is reversed.

A field which is totally-ordered in a way compatible with the algebraic operations, like $\mathbb{R}$, is sometimes called an *ordered field*.

6.2.1 The absolute value and distances

Definition 6.107. Let $x \in \mathbb{R}$ (or any ordered field). We write $|x|$ for the *absolute value* or *magnitude* of x, defined by

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

The absolute value of x is defined so that $|x| \geq 0$. Moreover, $x = |x|$ iff $x \geq 0$, and $x = -|x|$ iff $x \leq 0$.

Lemma 6.108. For $x, y \in \mathbb{R}$, $|xy| = |x||y|$.

Proof. Exercise (check the different cases). □

Lemma 6.109. For $x, y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$, with equality iff x and y have the same sign (both ≥ 0 or both ≤ 0).

Proof. Exercise. □

Thinking of real numbers as points on the number line, it is useful to think about the *distance* between two real numbers.

Given that $a \leq b$, the distance between a and b is $b - a$. When we have two real numbers a and b, we very commonly want to consider the distance between them without regard to which is the greater. So you will often see an expression such as

$$|a - b|$$

since this represents the distance between a and b regardless of the ordering of a and b.

6.2.2 Intervals

Here are some useful notations for certain subsets of the real line.

Definition 6.110. For $a, b \in \mathbb{R}$ with $a \leq b$, the *closed interval* $[a, b]$ is the subset of $\mathbb{R}$ defined by

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}.$$

Definition 6.111. For $a, b \in \mathbb{R}$ with $a < b$, the *open interval* (a, b) is the subset of $\mathbb{R}$ defined by

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}.$$

Remark 6.112. The notation for open intervals is the same as that for ordered pairs. It almost always clear from context which is meant.

Remark 6.113. The difference between open and closed intervals is that closed intervals contain their endpoints whereas open intervals omit their endpoints.

Definition 6.114. For $a, b \in \mathbb{R}$ with $a < b$, the *half-open intervals* $[a, b)$ and $(a, b]$ are the two subsets of $\mathbb{R}$ defined by

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

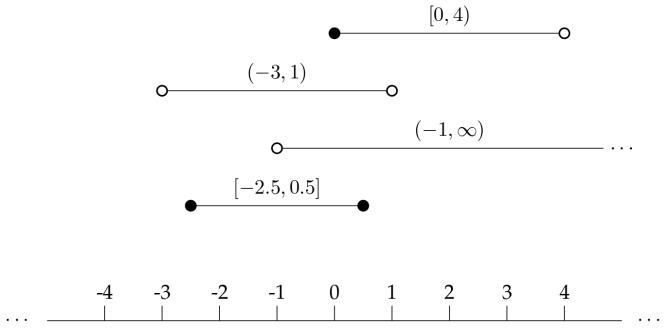
For an interval $[a, b]$, the *width* or *length* of the interval is $|b - a| = b - a$. We define the width similarly for open and half-open intervals (the absence of endpoints is not considered to affect the width).

Definition 6.115. Furthermore, we include the symbol ∞ in the open and half-open intervals to mean the following sets.

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

We have no need for notations such as '$[0, \infty]$' which would suggest they contained an element '∞'.

We sometimes depict intervals using a filled circle to indicate that the endpoint is contained in the set and a hollow circle for when the endpoint is not contained in the set.



Example 6.116. Intervals are useful for expressing a *range* of values when we make an imprecise measurement. For example, suppose we time a run with a stopwatch, recording a time of 55.78 seconds. Since our reaction time affects the time recorded, maybe by up to half a second, all we are really fairly sure about is that the true time is in the interval $[55.28, 56.28]$.

Example 6.117. A day is a half-open interval of time. By convention, this interval is closed on the left and open on the right: the exact moment of midnight is 00:00, and is the first instant of the new day. There is no last instant of a day: if the clock shows 23:59 at the current instant, there is another instant in the (near) future at which it will still show 23:59 and not yet be showing 00:00. We could not have a convention where days were always open intervals or always closed intervals without leaving some bits of time not contained within a day or having days overlapping each other.

Remark 6.118. Our notation for intervals will always refer to subsets of $\mathbb{R}$. However, we can talk about 'intervals' in any partially ordered set. The existing notation lets us talk about intervals in $\mathbb{Q}$ just by taking intersections with the intervals in $\mathbb{R}$, e.g.:

$$[a, b] \cap \mathbb{Q} = \{x \in \mathbb{Q} \mid a \leq x \leq b\}.$$

6.2.3 Injectivity of strictly increasing functions

Let $f : E \subseteq \mathbb{R}$ be a function on some interval $E \subseteq \mathbb{R}$. (It does not matter if E is closed, open, half-open, etc.).

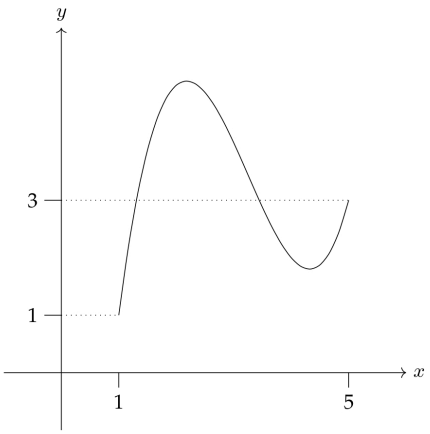
Definition 6.119. f is *increasing* if whenever $x, y \in E$ and $x \leq y$, then also $f(x) \leq f(y)$. It is *strictly increasing* if moreover $x < y$ implies $f(x) < f(y)$.

Proposition 6.120. Let $f : E \to \mathbb{R}$ be a strictly increasing function. Then f is injective.

Proof. Let $x, y \in E$ and suppose $x \neq y$. We need to show that $f(x) \neq f(y)$. But since $x, y \in \mathbb{R}$, either $x < y$ or $y < x$. Hence either $f(x) < f(y)$ or $f(y) < f(x)$ because f is strictly increasing. Either way, $f(x) \neq f(x)$. □

6.2.4 Intermediate value property of continuous functions

We do not have time to build a proper theory of continous functions, so we will use the notion intuitively. Suppose we are given a 'continuous' function $f : [a, b] \to \mathbb{R}$ on a closed interval. For example, here is an example of a continuous function $f : [1, 5] \to \mathbb{R}$ with $f(1) = 1$ and $f(5) = 3$.



From a visual inspection, the curve passes through every y-value between 1 and 3. More formally,

$$\forall y \in [1, 3]. \exists x \in [1, 5]. f(x) = y.$$

There is no guarantee of uniquesness for the x's, since this f is not strictly increasing. Moreover, there are y's outside of $[1, 3]$ which are in the image of f. But the important thing is that there is no clever way to draw a continuous curve connecting the points $(1, 1)$ and $(5, 3)$ in the plane which avoids passing through every y-value between 1 and 3. In a full course on real analysis, this property would be a theorem about continuous functions, called *the intermediate value theorem*.

We can use this observation to reason, for example, about the surjectivity of certain functions $\mathbb{R} \to \mathbb{R}$. For example, consider $x \mapsto x^5 - x^2$. When $|x|$ is very large, the x^5 term

dominates, and so $x^5 - x^2$ 'goes to infinity' as we make x larger. When we send x further towards $-\infty$, $x^5 - x^2$ also 'goes to $-\infty$'.

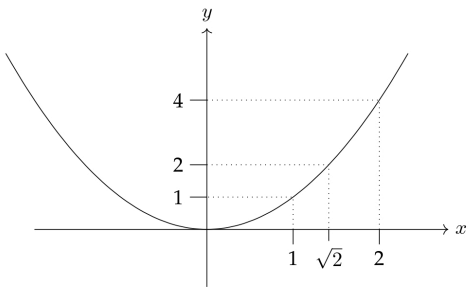
It follows that $f(x) = x^5 - x^2$ is a surjective function $\mathbb{R} \to \mathbb{R}$. If we want to show that some value y_0 is in the image, we just need to find a positive real c such that $f(-c) < -y_0$ and $y_0 < f(c)$. Then, by the discussion above, since f is continuous there is $x_0 \in [-c, c]$ such that $f(x_0) = y_0$.

6.3 Square roots

If $a \in \mathbb{R}$, a *square root* of a is a number $x \in \mathbb{R}$ such that $x^2 = a$. We can also ask for square roots in any number system (which has a multiplication). Real numbers have the property that, for any $x \in \mathbb{R}$, $x^2 \geq 0$. Hence negative real numbers cannot have a real square root. However, all positive numbers will have two square roots, since if $x^2 = a$ then also $(-x)^2 = a$.

6.3.1 Existence of square roots

Let us sketch the curve $y = x^2$. (We will draw the axes on different scales, because the curve grows quite fast).



From a visual inspection of the curve, and using the intermediate value property of the continuous function $x \mapsto x^2$, we see that finding a square root of a real number a can be done when the horizontal line crossing the y-axis at a also intersects the parabola. When a is positive, it will intersect the parabola in two places, when $a = 0$ in one place, and if $a < 0$ not at all. The uniqueness of the positive square root is due to the fact that $x \mapsto x^2$ is strictly increasing on $[0, \infty)$. The uniqueness of the negative square root can be seen from the fact that the function is strictly *decreasing* on $(-\infty, 0]$.

Definition 6.121. For $x \in \mathbb{R}$ with $x \geq 0$, the *principal square root* of x is the unique number $\sqrt{x} \geq 0$ with

$$(\sqrt{x})^2 = x.$$

Thus $x \mapsto \sqrt{x}$ is a function $[0, \infty) \to \mathbb{R}$ (or, optionally, a function $[0, \infty) \to [0, \infty)$).

Taking $\sqrt{2}$ as an example, because $1^2 = 1 < 2 < 4 = 2^2$, it follows that, whatever $\sqrt{2}$ is, it must lie between 1 and 2. This is a consequence of $x \mapsto x^2$ being strictly increasing for $x \geq 0$.

6.3.2 Irrationality of $\sqrt{2}$

Before explaining how to construct and specify real numbers, let us first show how the rational numbers by themselves are inadequate. If we tried to consider the rational numbers as constituting the number line, we would find that there were many 'holes'. Here is a famous example.

Proposition 6.122. There is no rational number x with $x^2 = 2$.

Proof. Suppose for contradiction that there is. Then we can write $x = \frac{a}{b}$ as a reduced fraction, so $b \in \mathbb{N}$, $b > 0$ and $\gcd(a, b) = 1$. By hypothesis

$$\left(\frac{a}{b}\right)^2 = 2.$$

Rearranging, we get

$$a^2 = 2b^2.$$

▶ *Claim.* $2 \mid a$.

> *Proof of Claim.* We have $2 \mid 2b^2 = a^2$. By Euclid's Lemma, when 2 divides a product it divides one of the factors, hence $2 \mid a \cdot a$ implies $2 \mid a$. ◀

▶ *Claim.* $2 \mid b$.

> *Proof of Claim.* We have just seen that $2 \mid a$. Therefore $a = 2a'$ for some $a' \in \mathbb{Z}$. Then
>
> $$4(a')^2 = 2b^2$$
>
> whence
>
> $$2(a')^2 = b^2.$$
>
> Now we apply the same reasoning as before. We have $2 \mid b^2$, so by Euclid's Lemma $2 \mid b$. ◀

We have just shown that 2 is a common divisor of a and b, but this contradicts the coprimality of a and b. □

This proposition is a 'negative' one: "show that $\sqrt{2}$ is irrational", i.e. "there does not exist a rational number x with $x^2 = 2$". It is sometimes said that you cannot 'prove a negative'. In mathematics, we often can. As you will learn in Logic, to prove $\neg P$, it suffices to show that supposing P leads to a contradiction/absurdity. In the proof above, supposing that there were a rational number whose square is 2 led to the absurd statement that there are coprime integers a, b with a common divisor of 2.

6.3.3 Rational approximations to $\sqrt{2}$

For now we will accept the argument above that all non-negative reals have a non-negative square root. We just saw that the square root of 2 is not a rational number. However, we can still find rational numbers $x \in \mathbb{Q}$ with with x^2 as close to 2 as we like.

When an interval (usually closed) $[a, b]$ has both of its endpoints rational $a, b \in \mathbb{Q}$, we sometimes call $[a, b]$ a *rational interval*. We will define a sequence of rational intervals $([a_n, b_n])_n$, each of which contains $\sqrt{2}$. Moreover, this sequence will be nested:

$$[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$$

and the widths of the intervals will tend to 0.

We define the intervals $([a_n, b_n])_n$ with a *bisection* technique. We begin with $a_0 = 1$, $b_0 = 2$. Note that

$$a_0^2 = 1 \leq 2 \leq 4 = b_0^2.$$

We choose subsequent intervals so as to maintain the property

$$a_n^2 \leq 2 \leq b_n^2$$

while also shrinking the distance between a_n and b_n.

Suppose the rational interval $[a_k, b_k]$ has been given, and $a_k^2 \leq 2 \leq b_k^2$. Then let $c_k = \frac{1}{2}(a_k + b_k)$ be the midpoint of the interval, another rational number. Then we case-split:

- Case 1: $c_k^2 < 2$. Define $a_{k+1} \coloneqq c_k$, $b_{k+1} \coloneqq b_k$.
- Case 2: $2 < c_k^2$. Define $a_{k+1} \coloneqq a_k$, $b_{k+1} \coloneqq c_k$.

Example 6.123. Let us calculate the first few intervals. By construction, the first interval is $[a_0, b_0] = [1, 2]$. The midpoint c_0 is $\frac{3}{2}$. We calculate

$$\left(\frac{3}{2}\right)^2 = \frac{9}{4} > 2$$

so the next interval is $[a_1, b_1] = [1, \frac{3}{2}]$. Now, $c_1 = \frac{1}{2}(\frac{3}{2} + 1) = \frac{1}{2}(\frac{5}{2}) = \frac{5}{4}$, and

$$c_1^2 = \left(\frac{5}{4}\right)^2 = \frac{25}{16} < 2,$$

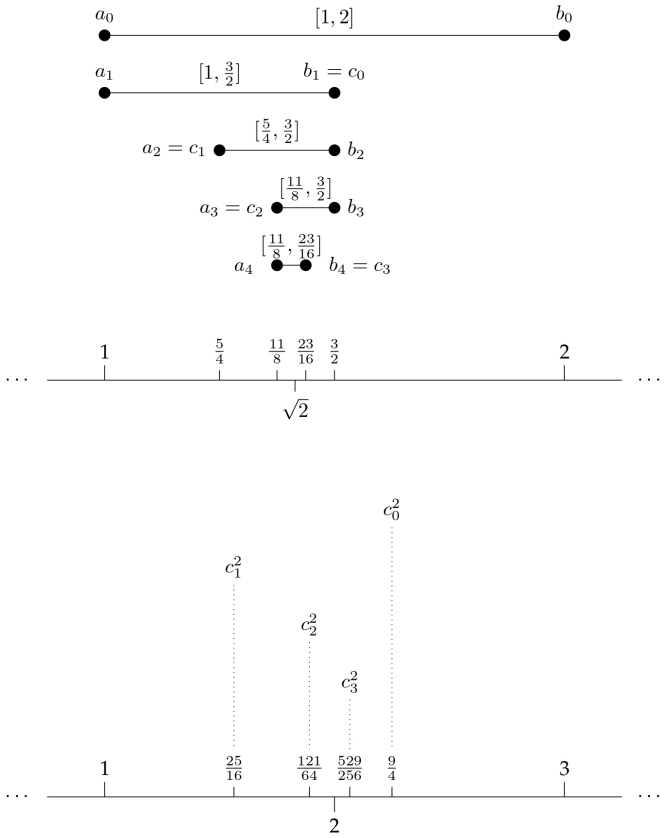
so $[a_2, b_2] = [\frac{5}{4}, \frac{3}{2}]$. Now, $c_2 = \frac{1}{2}(\frac{5}{4} + \frac{3}{2}) = \frac{1}{2}(\frac{11}{4}) = \frac{11}{8}$, and

$$c_2^2 = \left(\frac{11}{8}\right)^2 = \frac{121}{64} < 2$$

so $[a_3, b_3] = [\frac{11}{8}, \frac{3}{2}]$. Now, $c_3 = \frac{1}{2}(\frac{11}{8} + \frac{3}{2}) = \frac{1}{2}(\frac{23}{8}) = \frac{23}{16}$, and

$$c_3^2 = \left(\frac{23}{16}\right)^2 = \frac{529}{256} > 2$$

so $[a_4, b_4] = [\frac{11}{8}, \frac{23}{16}]$.



From the diagrams above we can draw some qualitative inferences about the process. As an observation about the definition of the process: if c_k^2 is to the *left* of 2, we take $[a_{k+1}, b_{k+1}]$ to be the *right* half of $[a_k, b_k]$ and, conversely, if c_k^2 is to the *right* of 2, we take $[a_{k+1}, b_{k+1}]$ to be the *left* half of $[a_k, b_k]$.

The intervals $[a_n, b_n]$ are getting smaller, the width halving each time. Note that the intervals are getting smaller and smaller, narrowing down on the point $\sqrt{2}$. The values of c_k^2 are closing in on 2, in the sense that each one is within the interval spanned by the previous two closest values: e.g. $c_2^2 \in [c_1^2, c_0^2]$ and $c_3^2 \in [c_2^2, c_0^2]$. However, looking closely, it is apparent that the sequence of values c_k^2 does not uniformly get closer to 2: c_1^2 is actually further from 2 than c_0^2 was. However, we are guaranteed that eventually the values of c_k^2 will all be closer to 2 than c_0^2 was

Proposition 6.124. For all n, $1 \leq a_n < b_n \leq 2$, and $a_n^2 < 2 < b_n^2$, and $|b_n - a_n| = 2^{-n}$.

Proof. We use induction on n.

$n = 0$: When $n = 0$, the $1 = a_0 < b_0 = 2$ and $a_0^2 = 1 < 2 < 4 = 2^2 = b_0^2$. For the third claim:

$$\begin{aligned} |b_0 - a_0| &= |2 - 1| \\ &= 1 \\ &= 2^{-0} \end{aligned}$$

as required.

Inductive step: Suppose the proposition is true when $n = k$, so that $1 \leq a_k < b_k \leq 2$, and $a_k^2 < 2 < b_k^2$, and $|b_k - a_k| = 2^{-k}$. Then by definition $c_k = \frac{1}{2}(a_k + b_k)$, and this satisfies

$$1 \leq a_k < c_k < b_k \leq 2.$$

To continue, we have to case split on whether c_k^2 is greater or less than 2.

1) $c_k^2 < 2$: In this case $a_{k+1} = c_k$ and $b_{k+1} = b_k$, so the first required inequality follows from $1 \leq c_k < b_k \leq 2$ above. For the second, $a_{k+1}^2 = c_k^2 < 2$ by assumption and $b_{k+1}^2 = b_k^2 > 2$ by induction hypothesis. Finally,

$$\begin{aligned} |b_{k+1} - a_{k+1}| &= |b_k - \tfrac{1}{2}(a_k + b_k)| \\ &= |\tfrac{1}{2}(b_k - a_k)| \\ &= \tfrac{1}{2}|b_k - a_k| \\ &= 2^{-(k+1)} \end{aligned}$$

as required.

2) $c_k^2 > 2$: Exercise (very similar to the previous case).

□

From the above we see that for all $n \in \mathbb{N}$,

$$a_n < \sqrt{2} < b_n,$$

while also $|b_n - a_n|$ can be made as small as we like. We can treat this as a bound on the error of a_n and b_n as an approximation to $\sqrt{2}$:

$$\begin{aligned} |a_n - \sqrt{2}| &< |b_n - a_n| \\ &= 2^{-n} \end{aligned}$$

and

$$\begin{aligned} |b_n - \sqrt{2}| &< |b_n - a_n| \\ &= 2^{-n}. \end{aligned}$$

Thus the difference between a_n and $\sqrt{2}$ can be made as small as we like, and similarly for b_n and $\sqrt{2}$. We say that the sequence of rationals

$$a_0, a_1, a_2, \ldots$$

tends to $\sqrt{2}$ from below, and the sequence

$$b_0, b_1, b_2, \ldots$$

tends to $\sqrt{2}$ from above.

Error in the square

There is a slight difference between (a) finding a positive rational $x \in \mathbb{Q}$ which is close to $\sqrt{2}$ and (b) finding a positive rational $x \in \mathbb{Q}$ such that x^2 is close to 2. There is something of a correlation however. The following shows that our two sequences of approximations $a_0, a_1, a_2, \ldots$ and $b_0, b_1, b_2, \ldots$ have sequences of squares that tend to 2, albeit with less precision than the way the unsquared seuences tend to $\sqrt{2}$.

Proposition 6.125. For all n, $|c_n^2 - 2| \leq 2^{1-n}$.

Proof. This is not by induction, but rather we deduce it from the previous lemma. The single inequality is equivalent to having both

$$c_n^2 - 2 \leq 2^{1-n} \qquad\qquad 2 - c_n^2 \leq 2^{1-n}.$$

since $|x| \leq y$ iff both $x \leq y$ and $-x \leq y$. We calculate

$$\begin{aligned} c_n^2 - 2 &= \left(\frac{a_n + b_n}{2}\right)^2 - 2 \\ &\leq \left(\frac{a_n + b_n}{2}\right)^2 - a_n^2 \\ &= \frac{1}{4}(b_n - a_n)(b_n + 3a_n) \\ &\leq \frac{1}{4} \cdot 2^{-n} \cdot (2 + 3 \times 2) \\ &= 2^{1-n} \end{aligned}$$

using the upper bound $a_n, b_n \leq 2$. But also

$$\begin{aligned} 2 - c_n^2 &= 2 - \left(\frac{a_n + b_n}{2}\right)^2 \\ &\leq b_n^2 - \left(\frac{a_n + b_n}{2}\right)^2 \\ &= \frac{1}{4}(b_n - a_n)(3b_n + a_n) \\ &\leq 2^{1-n}. \end{aligned}$$

□

In applications, you should check whether you are trying to solve problem (a) or problem (b) above. In these notes, we will mainly be interested only in problem (a).

6.3.4 The Shrinking Interval Property

This subsection and anything relying on the Shrinking Interval Property (SIP) is optional and non-examinable.

We have so far justified the existence of square roots through the intermediate value property of continuous functions. However, we have time neither to prove this property nor to define continuous functions. We will briefly look at a more basic property of real numbers, which can be used to actually characterize $\mathbb{R}$.

We will call the following principle *the Shrinking Interval Property* (*SIP*), though perhaps 'the Shrinking *Rational* Interval Property' would be more appropriate. The terminology is not standardized, and you do not have to remember it anyway.

Principle 6.126 (SIP)**.** Given any sequence of rational intervals $([a_n, b_n])_n$ which is

- *nested*: $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$ (equivalently, $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$), and
- *shrinking*: $|b_n - a_n|$ tends to 0,

There is a unique real number $\alpha \in \mathbb{R}$ satisfying

$$a_n \leq \alpha \leq b_n$$

for all $n \in \mathbb{N}$. Moreover, all real numbers arise in this way.

You are not expected to learn or to be able to use this principle (we have not even explained what 'tending to 0' means), but we will give an example of how this implies the existence of $\sqrt{2}$ independently of appeals to continuity.

Proposition 6.127. There is a unique positive real number α such that $\alpha^2 = 2$.

Proof. We first prove existence. The sequence of (positive) rational intervals $([a_n, b_n])_n$ above is nested and shrinking, and hence defines a (positive) real number α by the Shrinking Interval Property. Since multiplication of positive numbers is order-preserving,

$$a_n^2 \leq \alpha^2 \leq b_n^2.$$

But $([a_n^2, b_n^2])_n$ is also a nested, shrinking sequence of rational intervals, and, by construction,

$$a_n^2 \leq 2 \leq b_n^2.$$

Hence $\alpha^2 = 2$, by the uniqueness part of the Shrinking Interval Property.

For uniqueness, suppose β is also a positive real with $\beta^2 = 2$. Then

$$\begin{aligned} 0 &= 2 - 2 \\ &= \alpha^2 - \beta^2 \\ &= (\alpha - \beta)(\alpha + \beta) \end{aligned}$$

where $\alpha + \beta > 0$, hence we have $\alpha - \beta = 0$, i.e. $\alpha = \beta$.

For a different proof of uniqueness, suppose α, β are two positive reals with the property. If they are not equal, then WLOG $\alpha < \beta$. Then

$$2 = \alpha^2 < \beta^2 = 2$$

a contradiction $\lightning$. □

Remark$^+$**.** We could give a complete axiomatic definition of the real numbers as follows: $\mathbb{R}$ is a totally ordered set containing $\mathbb{Q}$ and satisfying the SIP. From this it is possible to define $+$ and $\times$ of real numbers. For example, if $\alpha, \beta \in \mathbb{R}$, take shrinking sequences of rational intervals with $\alpha \in [a_n, b_n]$, $\beta \in [c_n, d_n]$. Then $([a_n + c_n, b_n + d_n])_n$ is a shrinking rational interval and hence determines a unique element of $\mathbb{R}$ by the SIP. We define $\alpha + \beta$ to be this real number. We can furthermore prove that this $\mathbb{R}$ is an ordered field, which admits all square roots of non-negative numbers etc.

6.4 Representations of real numbers

6.4.1 Decimals

A decimal representation in the form

$$N.a_1a_2a_3\ldots$$

where $N \in \mathbb{N}$, $a_i \in \{0, 1, \ldots, 9\}$. can be thought of as an infinite sum

$$N + 10^{-1}a_1 + 10^{-2}a_2 + 10^{-3}a_3 + \ldots.$$

Thus truncating the sum at any point gives a rational number $N.a_1 \ldots a_k$ which underapproximates the real number.

Another way to look at decimals is in terms of intervals. The decimal presentation above means the unique real number common to all rational intervals in the sequence $([c_n, d_n])_n$ where

$$c_n = N.a_1 \ldots a_n \qquad\qquad d_n = N.a_1 \ldots a_n + 10^{-n}$$

(which exists and is unique by the SIP). So, if $\alpha \in \mathbb{R}$ is represented as a decimal by '$N.a_1a_2a_3\ldots$', then

$$\begin{aligned} |\alpha - N| &\leq 1 \\ |\alpha - (N + 10^{-1}a_1)| &\leq 10^{-1} \\ |\alpha - (N + 10^{-1}a_1 + 10^{-2}a_2)| &\leq 10^{-2} \end{aligned}$$

and so on.

For convenience let us suppose $N = 0$ and that the sequence (a_n) is not eventually all 0's. Then we can write this as

$$\begin{aligned} |\alpha| &< 1 \\ \left|\alpha - \frac{b_1}{10}\right| &< 10^{-1} \\ \left|\alpha - \frac{b_2}{10^2}\right| &< 10^{-2} \end{aligned}$$

etc, where

$$b_k = \sum_{i=1}^{k} 10^{k-i}a_i \in \mathbb{N}.$$

In this case, the b_i is the *unique* integer with

$$0 < \alpha - \frac{b_i}{10^i} < 10^{-i}.$$

From this point of view, finding the decimal digits of a real number means finding the optimal rational approximations from below with fixed denominators of the form 10^i.

6.4.2 Non-uniqueness of decimals

One problem with decimal presentations is that they are not unique for certain rational numbers. You should be aware of this as a fact, and for interest we will show how it follows from the SIP.

Proposition 6.128. $0.99999\ldots = 1$.

Proof. The expression on the left means the unique real number common to all the intervals $([a_n, b_n])_n$ where

$$a_n = 0.\underbrace{9\ldots9}_{n} = 1 - 10^{-n} \qquad\qquad b_n = a_n + 10^{-n} = 1.$$

But $a_n \leq 1 \leq b_n$, so this real number is just the rational number 1. □

Remark 6.129. The real numbers that admit two different decimal representations are precisely those non-zero rationals in whose reduced representation $\frac{p}{q}$ the denominator q has the form $2^m \times 5^n$. I.e., those rational numbers which can be written as $\frac{m}{10^n}$ for some $n \in \mathbb{N}$, $m \in \mathbb{Z}$, $m \neq 0$. Together with 0, these are the only real numbers which have 'terminating' decimal representation, i.e. one which ends in an infinite sequence of 0's.

Remark 6.130. A real number is rational iff its decimal representation '$N.a_1a_2a_3\ldots$' is eventually periodic, i.e. eventually it follows a pattern where some fixed block of digits repeats over and again.

Remark 6.131. For other number bases we have an analogous story.

6.4.3 Binary digits of $\sqrt{2}$

We have used the Shrinking Interval Property to show that $\sqrt{2}$ exists and along the way found that $\sqrt{2} \in [a_n, b_n]$ for every $n \in \mathbb{N}$, i.e. that $\sqrt{2}$ is contained in every interval we constructed by the interval bisection process. As we noted before, this means that the distance of $\sqrt{2}$ from a_n or b_n is bounded by the length of the interval $[a_n, b_n]$. So

$$|a_n - \sqrt{2}| = \sqrt{2} - a_n < b_n - a_n = 2^{-n}$$

and similarly $|b_n - \sqrt{2}| < 2^{-n}$. Looking at the particular values of a_n we constructed, we can now write down some of the binary representation of $\sqrt{2}$. Recall the first few values.

n	0	1	2	3	4
a_n	1	1	$5/4$	$11/8$	$11/8$
b_n	2	$3/2$	$3/2$	$3/2$	$23/16$

From this we deduce the following.

- The smallest integer n_0 with $\sqrt{2} - n_0 < 1$ is $n_0 = 1$.
- The smallest integer n_1 with $\sqrt{2} - 2^{-1}n_1 < 2^{-1}$ is $n_1 = 2$.
- The smallest integer n_2 with $\sqrt{2} - 2^{-2}n_2 < 2^{-2}$ is $n_2 = 5$.
- The smallest integer n_3 with $\sqrt{2} - 2^{-3}n_3 < 2^{-3}$ is $n_3 = 11$.
- The smallest integer n_4 with $\sqrt{2} - 2^{-4}n_4 < 2^{-4}$ is $n_4 = 22$.

To get the first few binary digits of $\sqrt{2}$, we just write any of the integers $1, 2, 5, 11, 22$ in base 2.

$$1 = 1_2$$
$$2 = 10_2$$
$$5 = 101_2$$
$$11 = 1011_2$$
$$22 = 10110_2$$

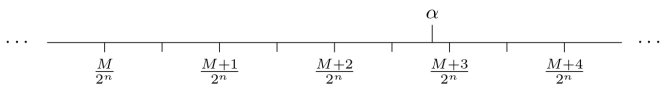
So, for example, because $5 = 101_2$, we know the binary expansion of $\sqrt{2}$ begins '$1.01\ldots$'.

Remark 6.132. As an aside, this is not the same as 'rounding to 2 binary places' (e.g. by analogy with 'rounding to 2 decimal places'). To round to 2 binary places, we need to inspect the next digit to see if we should round up or down. Since the next digit is a 1, the correct way to write $\sqrt{2}$ to 2 binary places is as 1.10_2. In this module we will not phrase questions in these terms.

6.4.4 Optimal dyadic approximations to $\sqrt{2}$

We mention one more interpretation of the numbers a_n, b_n above. A *dyadic rational* means a number of the form $\frac{m}{2^n}$ where $m \in \mathbb{Z}$, $n \in \mathbb{N}$.

For a given irrational number $\alpha \in \mathbb{R}$, we can ask for the closest possible dyadic rational with a given denominator. More precisely, for $n \in \mathbb{N}$, there is an integer $m \in \mathbb{Z}$ which minimizes $|\alpha - \frac{m}{2^n}|$. Consider the following picture.



Since the intervals $[\frac{M+i}{2^n}, \frac{M+i+1}{2^n}]$ cover the entire real line, α must lie in one of them. Moreover, since α is irrational, it cannot equal any of the endpoints, so it lies in a unique one of these intervals, and further still it cannot lie precisely at the midpoint of the interval it is in. Therefore it is within a distance of *half the width* of these intervals of an endpoint of one of these intervals. From the picture, α is within $2^{-(n+1)}$ of $\frac{M+3}{2^n}$. Thus the error of the approximation $\frac{M+3}{2^n}$ is one order of magnitude less than that of the denominator of the fraction.

Let us work this out concretely for $\sqrt{2}$.

- Both $a_0 = 1$ and $b_0 = 2$ are within a distance 1 of $\sqrt{2}$. But the midpoint $c_0 = \frac{3}{2}$ is greater than $\sqrt{2}$, so $a_0 = 1$ is the unique integer with $|\sqrt{2} - 1| < 2^{-1}$.
- Both $a_1 = 1$ and $b_1 = \frac{3}{2}$ are within a distance 2^{-1} of $\sqrt{2}$. But the midpoint $c_1 = \frac{5}{4}$ is less than $\sqrt{2}$, so $b_1 = \frac{3}{2}$ is the unique rational with denominator 1 or 2 satisfying $|\sqrt{2} - \frac{3}{2}| < 2^{-2}$.
- We have $\sqrt{2} \in [\frac{3}{2}, \frac{5}{4}]$, and the midpoint $c_2 = \frac{11}{8}$ is less than $\sqrt{2}$, so $\frac{3}{2}$ is the unique rational with denominator a divisor of 2^2 with $|\sqrt{2} - \frac{3}{2}| < 2^{-3}$.
- The midpoint of $[\frac{11}{8}, \frac{3}{2}]$ is $c_3 = \frac{23}{16}$ which is greater than $\sqrt{2}$, so $\frac{11}{8}$ is the unique rational with denominator a divisor of 2^3 with $|\sqrt{2} - \frac{11}{8}| < 2^{-4}$.
- We have $\sqrt{2} \in [\frac{11}{8}, \frac{23}{16}]$, but to work out which endpoint is the optimal dyadic approximation for precision 2^{-5} we have to go one step further and test the next midpoint. So
$$c_4 = \frac{1}{2}\left(\frac{11}{8} + \frac{23}{16}\right) = \frac{45}{32}$$
and
$$c_4^2 = \left(\frac{45}{32}\right)^2 = \frac{2025}{1024} < 2.$$
So the closer endpoint is $\frac{23}{16}$. That is, $\frac{23}{16}$ is the unique rational number with denominator a divisor of 2^4 such that $|\sqrt{2} - \frac{23}{16}| < 2^{-5}$.

6.5 Extra example: $\sqrt{3}$

There is no new content in this section, we will just go through some of the same working for $\sqrt{3}$ to show how you might set out your answers to problems.

Example 6.133. Find the first three digits of $\sqrt{3}$ in binary. Since $1^2 < 3 < 2^2$, we set $a_0 = 1$ and $b_0 = 2$ as before. This time we will set out our working in a table, where at each stage $c_n = \frac{1}{2}(a_n + b_n)$ is the midpoint of the current interval.

n	a_n	b_n	$\lvert b_n - a_n \rvert$	c_n	c_n^2	c_n^2 vs 3
0	1	2	1	$\frac{3}{2}$	$\frac{9}{4}$	$<$
1	$\frac{3}{2}$	2	2^{-1}	$\frac{7}{4}$	$\frac{49}{16}$	$>$
2	$\frac{3}{2}$	$\frac{7}{4}$	2^{-2}			

We can stop here because we have $\sqrt{3} \in [\frac{3}{2}, \frac{7}{4}]$, an interval of width 2^{-2}. Thus, $\sqrt{3}$ truncated to the three binary digits is just $\frac{3}{2}$. We have $3 = 11_2$ so the first three digits of $\sqrt{3}$ are 1.10_2.

Example 6.134. Find an integer m such that $|\sqrt{3} - 2^{-5}m| < 2^{-6}$. We continue the table from above.

n	a_n	b_n	$\lvert b_n - a_n \rvert$	c_n	c_n^2	c_n^2 vs 3
2	$\frac{3}{2}$	$\frac{7}{4}$	2^{-2}	$\frac{13}{8}$	$\frac{169}{64}$	$<$
3	$\frac{13}{8}$	$\frac{7}{4}$	2^{-3}	$\frac{27}{16}$	$\frac{729}{256}$	$<$
4	$\frac{27}{16}$	$\frac{7}{4}$	2^{-4}	$\frac{55}{32}$	$\frac{3025}{1024}$	$<$
5	$\frac{55}{32}$	$\frac{7}{4}$	2^{-5}	$\frac{111}{64}$	$\frac{12321}{4096}$	$>$

We can stop here because $\sqrt{3} \in [\frac{55}{32}, \frac{7}{4}]$, an interval of width 2^{-5}, and we know that it is in the left half not the right. Thus, $\sqrt{3}$ is within 2^{-6} of $\frac{55}{32}$. The final answer is $\underline{55}$.

Remark 6.135. Because we started with $[a_0, b_0] = [1, 2]$ where 1 and 2 are the closest integers to $\sqrt{3}$, the final column, headed '$c_n^2 - 3$', allows us to read off the binary digits of $\sqrt{3}$ following the 'binary point'. We read a 1 where there is a $<$ and 0 where there is $>$. So $\sqrt{3} = 1.101110\ldots_2$.

6.6 Definitions of $\mathbb{R}$

This section consists of an extension discussion purely for interest and not examinable.

We have mentioned already that it is possible to give *axiomatic* definition of the reals number in terms of the Shrinking Interval Property, which we mentioned should really be the 'shrinking *rational* intervals property'.

Remark$^+$. What I have called the SIP is not quite the same as 'nested intervals theorem' described on Wikipedia[1], at the time of writing, as weaker than necessary and requiring supplementation by the 'Archimedean property'. For us, the SIP requires that the intervals $[a_n, b_n]$ have rational endpoints and that $b_n - a_n \to 0$ as a variable in the ordered field $\mathbb{Q}$.

[1] https://en.wikipedia.org/wiki/Completeness_of_the_real_numbers

The Wikipedia notion is talking about $b_n - a_n \to 0$ in the ordered field $\mathbb{R}$. The assumption of the Archimedean property allows for $b_n - a_n \to 0$ in $\mathbb{Q}$ to imply the same in $\mathbb{R}$.

A more common axiomatic definition is in terms of 'least upper bounds' or 'suprema'. Least upper bounds can be defined in any poset — though they may not exist. One way to define the reals is as a (totally) ordered field in which every non-empty set S which has an upper bound has a *least* upper bound. To define $\sqrt{2}$ in this system, we would let

$$S = \{x \in \mathbb{R} \mid x^2 \leq 2\}.$$

Then S is certainly non-empty, since $0^2 = 0 \leq 2$. It has an upper bound, e.g. 2, because if $x > 2$ then $x^2 > 4 > 2$. So there it has a *least* upper bound α. By definition, this means that

$$\forall y \in \mathbb{R}. (\forall x \in S. x \leq y) \implies \alpha \leq y.$$

It will turn out that this α satisfies $\alpha^2 = 2$ (but the argument involves showing that poor approximations to $\sqrt{2}$ can be improved slightly, just like we did with the shrinking interval method).

Axiomatic definitions are ideally accompanied by constructions. The most popular concrete construction of $\mathbb{R}$ is probably the method of 'Dedekind cuts'. This is based on the idea that "a real number α divides $\mathbb{Q}$ into the set of rationals x with $x < \alpha$ and the set of rationals y with $\alpha \leq y$". In this construction, real numbers are identified with certain pairs (A, B) where $A \cup B = \mathbb{Q}$, $A \cap B = \emptyset$, and some other conditions hold, such as

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

Another concrete construction, the method of 'Cauchy sequences', is based on the idea that "a real number is something that can be approximated by a convergent sequence of rational numbers".

Lecture 7: Countability

7.1 Sets

We have so far used the term 'set' informally. As a reminder, a *set* is a collection X for which it makes sense to ask, for any x, whether $x \in X$ or $x \notin X$. There is no distinction between the members in terms of 'order' or being present 'multiple times'.

Sets are assumed to satisfy the principle of *extensionality*. which says that two sets are equal iff they have the same members.

Principle 7.136 (Set extensionality)**.** Let X, Y be sets. Then $X = Y$ iff

$$\forall a.\, a \in X \iff a \in Y.$$

It should be clear from the set extensionality principle that the following expressions all denote equal sets,

$$\{1, 1, 2\} = \{1, 2\} = \{2, 1\} = \{2, 2, 1\}$$

where we interpret the curly bracket expression as denoting a set X for which $a \in X$ holds for prcisely those a listed inside the brackets.

As a generalization of this notation, we write expressions such as

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

or alternatively

$$X = \{x : P(x)\}$$

to mean that X is a set where $x \in X$ iff $P(x)$ holds. This is called a *set comprehension*. For reasons discussed at the end of this chapter, we cannot make the existence of set comprehensions a principle.

A *subset* of a set X is a set A whose members are also members of X.

$$A \subseteq X \iff \forall a.\, a \in A \implies a \in X.$$

The converse relation, of being a *superset*, is written with a flipped symbol:

$$Y \supseteq X \iff \forall a.\, a \in X \implies a \in Y.$$

7.1.1 Unions

We have previously discussed the union $A \cup B$ where A and B are both subsets of some other set X. We can also talk about the union of two sets given abstractly, not necessarily as subsets of some other set. For example,

$$\{0, \pi, \emptyset, \{\sqrt{2}\}, \text{'a'}\} \cup \{0, 1, 2, \{\sqrt{2}, \sqrt{3}\}\} = \{0, 1, 2, \pi, \emptyset, \{\sqrt{2}\}, \{\sqrt{2}, \sqrt{3}\}, \text{'a'}\}.$$

So the union is just the set of things that are in either or both of the two sets.

We will also need to consider unions of larger families of sets, not necessarily subsets of a common superset. The notation for this is

$$\bigcup_{i \in I} A_i = \{a \mid \exists i \in I.\, a \in A_i\}$$

where $(A_i \mid i \in I)$ is a family of sets A_i indexed by a set I. Typically, $I = \mathbb{N}$, which we can write informally as

$$\bigcup_{i \in \mathbb{N}} A_i = A_0 \cup A_1 \cup A_2 \cup \dots.$$

Definition 7.137. The set of (*finite*) *binary strings*

$$\{0,1\}^* = \{v \mid \exists N \in \mathbb{N}.\, v \in \{0,1\}^N\}$$

7.2 Finite and infinite sets

For each $n \in \mathbb{N}$, the set

$$\{x \in \mathbb{N} \mid x < n\}$$

of natural numbers less than n has precisely n elements. We often write the set as

$$\{0, 1, 2, \dots, n-1\}$$

since this sets the general pattern, if the list of elements is understood to be empty when $n = 0$ and other appropriate readings for $n = 1$ and $n = 2$.

Definition 7.138. A set X is *finite* if there exists $n \in \mathbb{N}$ and a bijection $X \to \{x \in \mathbb{N} \mid x < n\}$. A set is *infinite* if it is not finite.

Notation 7.139. We sometimes write

$$X \cong Y$$

to mean there is a bijection between sets X and Y. But this is notation is quite context-dependent, so we will avoid it.

The following proposition is hopefully intuitive. You can just assume it, but we might also try to justify it by induction.

Proposition 7.140. For any $n \in \mathbb{N}$, every subset of $\{x \in \mathbb{N} \mid x < n\}$ is finite.

Proof. By induction on n.

Base case: When $n = 0$, $\{x \in \mathbb{N} \mid x < 0\} = \emptyset$, so the only subset is the empty set, which is finite.

Inductive step: Suppose the proposition is true for $n = k$. Then any subset S of $\{x \in \mathbb{N} \mid x < k+1\}$ can be separated into two parts:

$$S_1 = S \cap \{x \in \mathbb{N} \mid x < k\} \qquad S_2 = S \cap \{k+1\}.$$

where $S_1 \subseteq \{x \in \mathbb{N} \mid x < k\}$ is finite by induction hypothesis, and $S_2 \subseteq \{k+1\}$ must either be empty or the singleton $\{k+1\}$. In the former case, $S = S_1$ so S is finite, In the latter case, given a bijection $f : S_1 \to \{x \in \mathbb{N} \mid x < m\}$ for some $m \in \mathbb{N}$, we extend it to a bijection $g : S \to \{x \in \mathbb{N} \mid x < m+1\}$ by $g(s) = f(s)$ for $s < k$ and $g(k) = m$.

□

Proposition 7.141. A set X is finite iff there exists $n \in \mathbb{N}$ and an injection

$$X \to \{x \in \mathbb{N} \mid x < n\}.$$

Proof. $\underline{\Longrightarrow}$ If X is finite, then by definition there exists $n \in \mathbb{N}$ and a bijection $f : X \to \{x \in \mathbb{N} \mid x < n\}$. But a bijection is in particular an injection.

$\underline{\Longleftarrow}$ Suppose we are given an injection

$$f : X \to \{x \in \mathbb{N} \mid x < n\}.$$

An injection gives a bijection from X to a subset of the codomain, but by the previous proposition this is finite.

□

7.2.1 The infinitude of primes

To show that a set is infinite set, we have to show that there is no bijection with a finite set.

Example 7.142. Let $P = \{p \in \mathbb{N} \mid p \text{ is prime}\}$. Then P is an infinite set.

Proof. Suppose for contradiction that P were finite. Then $P = \{p_0, p_1, \ldots, p_{n-1}\}$ for some n. Now let

$$N = p_0 p_1 \ldots p_{n-1} + 1,$$

the product of all the primes plus 1. Clearly $N \in \mathbb{N}$ and $N > 1$, so N has a prime divisor $p \mid N$. But p cannot be any of the p_i, since if $p_i \mid N$ then

$$p_i \mid N - p_0 p_1 \ldots p_{n-1} = 1,$$

a contradiction. □

We can apply the same kind of idea to other sets to show their infinitude, though it is much more obvious in those cases. For example, $\mathbb{N}$ is infinite because if it consisted of only $x_0, \ldots, x_{n-1}$, we could write

$$N = x_0 + x_1 + \ldots + x_{n-1} + 1$$

which would be another natural number larger than all of them.

7.3 Cardinalities

Recall that if X and Y are finite sets, can deduce relationships between their cardinalities from the existence of certain functions as follows.

- If there exists an injection $f : X \to Y$, then $|X| \leq |Y|$.
- If there exists a surjection $f : X \to Y$, then $|X| \geq |Y|$.
- If there exists a bijection $f : X \to Y$, then $|X| = |Y|$.

Remark 7.143. The converse of the second point is not true. If Y is empty, then $|\{0\}| \geq |\emptyset|$ but there is no function $\{0\} \to \emptyset$, never mind surjection.

For arbitrary sets, we essentially take this as a definition of the 'cardinality' relation.

Definition 7.144. For sets X and Y, we say

- $|X| \leq |Y|$ if there exists an injection $X \to Y$, and
- $|X| = |Y|$ if there exists a bijection $X \to Y$.

These formulas can be read as "X has cardinality less than or equal to the cardinality of Y" and "X and Y have the same cardinality". But note that for us '$|X|$' is just a suggestive notation, and not actually defined as an entity.

Proposition 7.145. "Having the same cardinality" is an 'equivalence relation' between sets.

Proof. We check the three axioms.

- Reflexivity. For any set X, the identity $id_x : X \to X$ is a bijection, so $|X| = |X|$.
- Symmetry. Suppose $|X| = |Y|$. Then there exists a bijection $f : X \to Y$. But a function is a bijection iff it has an inverse, so $f^{-1} : Y \to X$ gives us $|Y| = |X|$.
- Transitivity. Suppose $|X| = |Y|$ and $|Y| = |Z|$. Then we have bijections $f : X \to Y$ and $g : Y \to Z$. But then $g \circ f$ is a bijection $X \to Z$, so $|X| = |Z|$. □

Remark 7.146. The reason for the scare quotes around "equivalence relation" is because, as we will discuss at the end of this chapter, there is no 'set of all sets', so we cannot treat $|X| = |Y|$ as a relation on a set as defined in this module.

Definition 7.147. For a set X, the *cardinality* of X means the 'equivalence class' of X in the 'there exists a bijection between' relation.

For finite sets, we continue to treat the cardinality as an actual natural number n to mean the equivalence of the set $\{x \in \mathbb{N} \mid x < n\}$.

7.3.1 The Cantor-Schröder-Bernstein Theorem

The following theorem tells us that "$|X| \leq |Y|$" defines a 'partial order' on cardinalities.

Theorem 7.148 (The Cantor-Schröder-Bernstein Theorem)**.** Let X and Y be two sets. Then $|X| \leq |Y|$ and $|Y| \leq |X|$ implies $|X| = |Y|$.

The proof is not so difficult. Given injections $X \to Y$ and $Y \to X$ there is a recipe by which you can construct a bijection between X and Y. It is 'constructive' in a certain sense, but not quite enough to be useful for us to go through here.

Remark$^+$. This CSB theorem tells us that cardinalities are partially ordered. In fact, in suitable foundations, cardinalities are totally ordered. This means that for any two sets X and Y there exists at least one injection either $X \to Y$ or $Y \to X$. The proof of this fact is much more difficult, and is even less constructive.

7.4 Countable and uncountable sets

The dichotomy between finite and infinite sets is in many ways mirrored by a dichotomy between countable and uncountable sets.

Definition 7.149. A set X is *countable* in case either

- X is finite, or
- there exists a bijection $X \to \mathbb{N}$.

In the second case, we say that X is *countably infinite*.

Definition 7.150. A set is *uncountable* if it is not countable.

Remark 7.151. A set X is finite if it makes sense to write it as

$$X = \{x_0, x_1, \ldots, x_{n-1}\}$$

for some $n \in \mathbb{N}$. It is countably infinite if it makes sense to write it as

$$X = \{x_0, x_1, x_2, \ldots\}.$$

An uncountable set is a set which is so large that we cannot list its elements as a sequence like this. We will see some uncountable sets shortly.

Proposition 7.152. A set X is countable iff there exists an injection $X \to \mathbb{N}$.

Proof sketch. The $\implies$ direction is trivial. For the $\impliedby$ direction, given an injection $f : X \to \mathbb{N}$, either the image $f[X] \subseteq \mathbb{N}$ is bounded by some natural number, in which case X is finite, or it is unbounded in $\mathbb{N}$. But an infinite subset of $\mathbb{N}$ can be written as an increasing sequence $a_0 < a_1 < a_2 < \ldots$, which tells us how to count it.

Alternatively, given an injection $X \to \mathbb{N}$, if X is infinite then there is an injection $\mathbb{N} \to X$, so by the CSB Theorem there is also a bijection between X and $\mathbb{N}$. □

Example 7.153. The set $\mathbb{N}$ is countable, since $\mathrm{id}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ is a bijection.

Example 7.154. The set $[0]_2$ of even numbers is countable. Obviously, it injects into $\mathbb{N}$. We can also give a bijection

$$\begin{aligned} \mathbb{N} &\to [0]_1 \\ n &\mapsto 2n. \end{aligned}$$

Example 7.155. The set $[1]_2$ of odd numbers is countable. Obviously, it injects into $\mathbb{N}$. We can also give a bijection

$$\begin{aligned} \mathbb{N} &\to [1]_1 \\ n &\mapsto 2n + 1. \end{aligned}$$

Example 7.156. The set $\mathbb{Z}$ of integers is countable. For example,

$$\begin{aligned} \mathbb{N} &\to \mathbb{Z} \\ n &\mapsto \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ -\frac{n+1}{2} & \text{if } n \text{ is odd} \end{cases} \end{aligned}$$

is a bijection (exercise). Alternatively, we can inject $\mathbb{Z}$ into $\mathbb{N}$ with

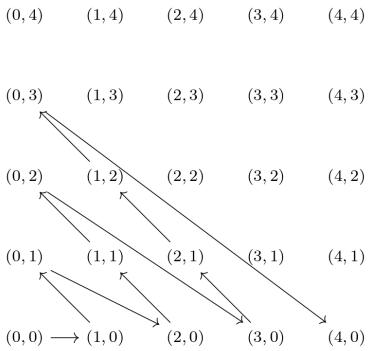
$$\begin{aligned} \mathbb{Z} &\to \mathbb{N} \\ n &\mapsto \begin{cases} 2^n & \text{if } n \geq 0 \\ 3^{-n} & \text{if } n < 0. \end{cases} \end{aligned}$$

7.4.1 Rational numbers

Perhaps surprisingly, there are precisely the same number of rational numbers as integers. That is to say, $\mathbb{Q}$ is countable.

Proposition 7.157. $\mathbb{N} \times \mathbb{N}$ is countable.

Proof sketch. If we draw a diagram and start enumerating the points of the plane that have natural number coordinates,



it is fairly clear that we can continue this process to define a bijection $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$. □

Remark 7.158. This bijection can be given by a formula, $(m, n) \mapsto \frac{1}{2}(m+n)(m+n+1)+n$ (there is no need to learn this).

Corollary 7.159. $\mathbb{Q}$ is countable.

Proof. Since $\mathbb{Z}$ is countable, $\mathbb{Z}^2$ is countable by the previous proposition. We can identify $\mathbb{Q}$ with a subset of $\mathbb{Z}^2$ using reduced fractions, so $\mathbb{Q}$ is a subset of a countable set, and hence countable. □

7.4.2 Countable unions of countable sets

Our discussion of the rational numbers can be generalized a little. For a start, we have the following.

Corollary 7.160. If X and Y are countable, then $X \times Y$ is countable. □

Now consider the following fact about finite sets.

Proposition 7.161. Let $n \in \mathbb{N}$ and let $X_0, \ldots, X_{n-1}$ be finite sets. Then

$$\bigcup_{i<n} X_i = X_0 \cup \ldots \cup X_{n-1}$$

is a finite set. □

We have something analogous for countable sets.

Proposition 7.162. Let $X_0, X_1, \ldots$ be a sequence of countable sets. Then

$$\bigcup_{i \in \mathbb{N}} X_i = X_0 \cup X_1 \cup \ldots$$

is a countable set. □

Proof. We may assume that the sets X_i are pairwise disjoint. For each $i \in \mathbb{N}$, take an injection $f_i : X_i \to \mathbb{N}$. Now we construct an injection $\bigcup_{i \in \mathbb{N}} X_i \to \mathbb{N}^2$ by

$$x \mapsto (i, f_i(x)) \text{ if } x \in X_i.$$

Thus $\bigcup_{i \in \mathbb{N}} X_i$ injects into a countable set, so it is itself countable. □

We summarize this situation by the slogan "a countable union of countable sets is countable".

Example 7.163. The set $\{0,1\}^*$ of finite binary strings is countable. To see this, observe that each set $\{0,1\}^n$ is finite and therefore countable, and so

$$\{0,1\}^* = \bigcup_{N \in \mathbb{N}} \{0,1\}^N$$

is a countable union of countable sets, which is countable.

Example 7.164. Similarly, $\mathbb{N}^*$ is countable, where

$$\mathbb{N}^* = \bigcup_{i \in \mathbb{N}} \mathbb{N}^i$$

and this time each $\mathbb{N}^i$ is countably infinite.

7.4.3 Uncountable sets

We will now see our first example of an uncountable set.

Theorem 7.165 (Cantor's diagonal argument)**.** The powerset of the natural numbers, $P(\mathbb{N})$, is uncountable.

Proof. We suppose for contradiction that it is countable. It is clearly infinite, so we can take a bijection $f : \mathbb{N} \to P(\mathbb{N})$. We will now construct a 'paradoxical' subset of $\mathbb{N}$. Let $C \subseteq \mathbb{N}$ be given by

$$C = \{n \in \mathbb{N} \mid n \notin f(n)\}.$$

This C is carefully constructed to obtain the following.

▶ *Claim.* C is not in the image of f.

| *Proof of Claim.* Suppose for contradiction that $C = f(i)$. Then let us consider whether

$i \in C$.

$$\begin{aligned} & i \in C \\ \iff & \quad [\text{since } C = f(i)] \\ & i \in f(i) \\ \iff & \quad [\text{by definition of } C] \\ & i \notin C \end{aligned}$$

So $i \in C$ iff $i \notin C$, which is absurd. ◀

So if $C \notin f[\mathbb{N}]$, this contradicts the assumption that f was surjective. □

Example 7.166. The set $\{0,1\}^{\mathbb{N}}$ of infinite binary sequences is uncountable. To see this, we can either repeat a version of the argument above, or use the fact that there is abijection $P(\mathbb{N} \to \{0,1\}^{\mathbb{N}}$ given by taking the characteristic function:

$$A \mapsto \chi_A.$$

Remark 7.167. The reason for calling the above 'the diagonal argument' becomes clear if we consider the version of it for $\{0,1\}^{\mathbb{N}}$. We first suppose we have a bijection $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$. We can draw this as an infinite table whose rows are the sequences $f(0), f(1), \ldots$:

n	$f(n)_0$	$f(n)_1$	$f(n)_2$	$f(n)_3$	$f(n)_4$	$\ldots$
0	$\underline{1}$	0	0	1	0	$\ldots$
1	1	$\underline{0}$	1	1	1	$\ldots$
2	0	1	$\underline{1}$	1	0	$\ldots$
3	0	0	0	$\underline{1}$	0	$\ldots$
4	0	1	0	0	$\underline{0}$	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

where here we have invented some illustrative possible values for f. The diagonal of this table is another sequence beginning $10110\ldots$. The idea now is to define a sequence $c \in \{0,1\}^{\mathbb{N}}$ by flipping the bit in every diagonal entry, so c will begin $c = 01001\ldots$. But now c cannot appear as a row in the table, since by construction it differs from every row in at least one place. Specifically, $c_i \neq f(i)_i$ for every $i \in \mathbb{N}$.

Remark 7.168. The diagonal argument shares some structural similarity with the proof of the infinitude of primes. We first supposed that we could enumerate the set of primes/subsets as a finite/countable sequence. We then used this sequence to construct a new prime/sequence that could not have been in the original sequence.

7.5 The cardinality of $\mathbb{R}$

7.5.1 Uncountability of $\mathbb{R}$

Theorem 7.169. The set $\mathbb{R}$ of real numbers is uncountable.

Proof sketch. It is sufficient to give an injection from an uncountable set into the reals. We define $\{0,1\}^{\mathbb{N}} \to \mathbb{R}$ by

$$(a_n)_n \mapsto \sum_{i \in \mathbb{N}} \left(\frac{2}{3}\right)^i a_i.$$

The infinite sum can be made precise (e.g. using the Shrinking Intervals Property), and it can be shown that this is an injection. □

Remark 7.170. One could consider defining the injection $\{0,1\}^{\mathbb{N}} \to \mathbb{R}$ by

$$(a_n)_n \mapsto \sum_{i \in \mathbb{N}} \left(\frac{1}{2}\right)^i a_i$$

which sends a sequence $(a_n)_n$ to the real number with binary expansion $a_0.a_1a_2a_3\ldots$. However, this is not an injection, since some real numbers have two binary expansions. The argument can still be made to work this way.

Remark 7.171. Sometimes the uncountability of the reals is proved using a version of the diagonal argument which considers either the binary or decimal expansion of each real number and constructs a new real number whose digits differ from every sequence in the table. This argument can be made to work, but it has a similar problem to the variant above, due to non-uniqueness of binary and decimal expansions.

Remark$^+$. Cantor's original proof of the uncountability of the real numbers is an elegant application of the Shrinking Intervals Property, and can still be seen as a diagonal argument. Here is a version of that argument. Suppose $\alpha_0, \alpha_1, \alpha_2, \ldots$ is an enumeration of the reals in the interval $[0,1]$. Then at least one of the intervals $[0, \frac{1}{3}], [\frac{1}{3}, \frac{2}{3}], [\frac{2}{3}, 1]$ does not contain α_0. Let $[a_0, b_0]$ be such an interval. Again divide this new interval into thirds, and let $[a_1, b_1]$ be one of those subintervals that does not contain α_1. Continue this process to get a nested shrinking sequence $[a_0, b_0] \supseteq [a_1, b_1] \supseteq \ldots$ of intervals, which by the SIP determines a real number β with $\beta \in [a_n, b_n]$ for all $n \in \mathbb{N}$. But, by construction, $\alpha_n \notin [a_n, b_n]$, so $\beta \neq \alpha_n$ for any n. This contradicts the assumption that $\alpha_0, \alpha_1, \alpha_2, \ldots$ enumerates all of $[0,1]$.

7.5.2 Comparison with binary sequences

So far we know that both $\{0,1\}^{\mathbb{N}}$ and $\mathbb{R}$ are uncountable. In fact, we used an injection $\{0,1\}^{\mathbb{N}} \to \mathbb{R}$ to see that $\mathbb{R}$ is uncountable, so actually we know that

$$|\{0,1\}^{\mathbb{N}}| \leq |\mathbb{R}|.$$

In fact these two sets have the same cardinality.

Lemma 7.172. $|\mathbb{R}| \leq |\{0,1\}^{\mathbb{N}}|$.

Proof sketch. We can inject $\mathbb{R}$ into the interval $[0,1]$, e.g. by $x \mapsto \frac{1}{\pi}\arctan(x) + \frac{1}{2}$, or by $x \mapsto \frac{1}{e^x+1}$, etc. So it suffice to inject $[0,1]$ into $\{0,1\}^{\mathbb{N}}$. We can do this by mapping x to a sequence $(a_n)_n$ where $0.a_0a_1a_2\ldots$ is a binary expansion of x, and where there are two choices of expansion we choose one of them, say the one that is eventually all 1's. □

Corollary 7.173. $|\mathbb{R}| = |\{0,1\}^{\mathbb{N}}|$.

Proof. This now follows from the Cantor-Schröder-Bernstein Theorem. □

7.6 Optional further discussion of sets

This section contains optional further discussion on some points mentioned earlier and the outlook for a finer classification of infinite sets.

7.6.1 Russell's paradox

Here is a problem that arises if we take set comprehension as a principle.

Define R to be the set

$$R = \{X \mid X \text{ is a set, and } X \notin X\},$$

that is, R is the set of all sets that are not members of themselves. Now let us ask whether $R \in R$.

- If $R \in R$, then by definition R is a set (we knew that already) and $R \notin R$. But that contradicts $R \in R$.
- If $R \notin R$, then by definition, since R is certainly a set, $\neg(R \notin R)$, i.e. $R \in R$, but that is also a contradiction.

So we cannot have either $R \in R$ or $R \notin R$, since both lead to a contradiction!

7.6.2 Set-building principles

Formal set theories have been developed to try avoid paradoxes such as Russell's. The most popular formal set theory is ZFC, which stands for Zermelo-Fraenkel set theory extended with the Axiom of Choice. We have not covered any formal set theories in this module but instead we used some of the same ideas informally.

The key feature of formal set theories is they provide a precise list of principles (or 'axioms') for which sets exist. These principles were refined over time, adding extra power to ensure we have all the sets we need, and removing power when paradoxes are discovered.

We can summarize the basic principles we have used so far.

- Certain basic sets are already understood and are assumed to exist, e.g. $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$, and the empty set $\emptyset$.
- If we can define a (possibly infinite) sequence of objects $a_0, a_1, a_2, \ldots$, then we can collect them into a set $\{a_0, a_1, a_2, \ldots\}$.
- **Pairs**: if X and Y are sets, already known to exist, then there exists a set $X \times Y$ whose members are all ordered pairs (x, y) with $x \in X$, $y \in Y$.
- **Subsets**: if X is a set, we can define new sets as subsets of X (all those elements which satisfy some condition).
- **Powerset**: if X is a given set, then there exists a set $P(X)$ whose members are precisely the subsets of X.
- **Unions**: if $(X_i : i \in I)$ is family of sets indexed by a set I, then there exists a set $\bigcup_{i \in I} X_i$ whose members are those a for which $\exists i \in I.\, a \in X_i$.

The existence of sets of functions Y^X follows from these principles already, since the set of functions $X \to Y$ is a subset of $P(X \times Y)$.

Proposition 7.174. There is no set of all of sets, i.e., there is no set $\mathcal{U}$ satisfying

$$\forall x.\, x \in \mathcal{U} \iff x \text{ is a set.}$$

Proof. Suppose there exists such a set $\mathcal{U}$. Then by the subset principle, we can define a set R where

$$R = \{x \in \mathcal{U} \mid x \notin x\} \subseteq \mathcal{U}.$$

But as $\mathcal{U}$ contains all sets, this is precisely the set R from Russell's paradox. □

Corollary 7.175. We cannot in general form unrestricted set comprehensions such as

$$X = \{x \mid p(x) \text{ is true}\}.$$

Proof. If we could, we could define $\mathcal{U} = \{x \mid x \text{ is a set}\}$, or even define the Russell set directly. □

One way to think about this is that the 'set of all sets' is too large to make sense. To get hold of bigger sets, we have to construct them iteratively using the principles that give us bigger sets, mainly the powerset principle but also the union principle. The fact that we have to iterate prevents us making the jump to a set that would be too large.

Remark$^+$. Referring to our list of principles above, the second principle is a special case of ZFC's 'replacement axiom'. The technical name for subset principle is the 'separation axiom'. Set theories such as ZFC have a reductivist ontology which does not include ordered pairs. Instead, ordered pairs are encoded as sets, and our version of the pairs principle is a consequence of other axioms. The subset principle allows us to take unions of arbitrary families of sets which happen to be subsets of some given set. For us the union principle is only important for how we define sets like $\{0,1\}^*$. We could do without it at the cost of having to choose an encoding of $\{0,1\}^*$ as a subset of $\{0,1,2\}^{\mathbb{N}}$, say.

7.6.3 Different uncountable infinities

We have given a trichotomy between finite, countably infinite, and uncountable sets. This is far from the end of the matter.

Cantor's diagonal argument tells that, for any set X, $P(X)$ has strictly greater cardinality. Thus

$$|\mathbb{N}| < |P(\mathbb{N})| < |P(P(\mathbb{N}))| < \dots$$

is an infinite, strictly increasing chain of 'infinities', where the first is countable and the rest are increasing levels of uncountable.

It is still only a countably infinite set of infinities. By the union principle,

$$P^\omega(\mathbb{N}) = \bigcup_{i \in \mathbb{N}} P^i(\mathbb{N})$$

is a set, and clearly it does not inject into any of the $P^i(\mathbb{N})$, so it is another strictly bigger set.

In fact, we can use a version of the diagonal argument to see that there must be an uncountable number of uncountable infinities. Suppose there were only countably many cardinalities, represented by sets $X_0, X_1, X_2, \dots$, say. In particular, for every $i \in \mathbb{N}$, there is a $j \in \mathbb{N}$ with $|P(X_i)| = |X_j|$. Then let

$$Y = \bigcup_{i \in \mathbb{N}} X_i.$$

But now Y does not inject into any of the X_i, since by the observation above, every $P(X_i)$ injects into Y.

In fact, if we work in ZFC foundations, then there is a canonical choice of representative for each cardinality, and the collection of all these representatives is, like the non-existent set $\mathcal{U}$ of all sets, too large a collection to be contained in any set. In short, there are more different infinities than any given infinity.