Numbers and Sets

Adam Kelly (ak2316@cam.ac.uk)

June 14, 2021

'Numbers and Sets' is one of the first course in pure mathematics that you will take as an undergraduate at Cambridge. In a sense, it is the 'starting course', in that it will introduce you to the 'pure maths' way of thinking about things. This introduction will happen through the lense of thinking about objects, beginning with the natural and real numbers. You will be introduced to the 'thoughtful way' of thinking about such objects, that you can carry through to almost every other course in pure mathematics.

This article constitutes my notes for the 'Numbers and Sets' course, held in Michaelmas 2020 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas.

Contents

§1 Elementary Number Theory

Number theory is the branch of mathematics that studies the properties of *numbers*, with a particular emphasis on the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$ and occasionally the rationals $\mathbb{Q}$. In this section, we will study some of the *additive* and *multiplicative* structure of the integers, looking at divisors, primes and tools such as modular arithmetic.

One of the aims of this course (and this section in particular) is to study numbers 'from the ground up', being quite careful about what we assume. This goal immediately presents us with a question: what exactly is a 'number'?

§1.1 The Peano Axioms

What are the natural numbers? Intuitively, we might say that they are a set[1] $\mathbb{N} = \{1, 2, 3, \dots\}$, created by starting at 1 and counting forward indefinitely, each time obtaining an object distinct from all of the previous ones. This does answer our question (a natural number is any element of $\mathbb{N}$), but has created a series of other questions. For example, what does it mean to 'count forward', and how can it be done 'indefinitely'? How are we allowed to use these natural numbers, in regard to defining things like addition and multiplication?

Instead of attempting to answer these questions using our informal, intuitive definition of the natural numbers, we will instead use a definition that is more precise. Namely, we will define the natural numbers using the Peano axioms. We will state the rules or *axioms* that natural numbers satisfy, which will define the natural numbers in terms of *how they work*, rather than *what they are*. After clearly setting out this definition, we will be in a much stronger position to write concrete mathematical proofs about the natural numbers.

[1] We will look at sets later in the course, but an informal familiarity will be assumed from the beginning

Definition 1.1 (Peano Axioms)

The **natural numbers** are a set $\mathbb{N}$, along with a function $S : \mathbb{N} \to \mathbb{N}$ and an object '1' satisfying the following axioms:

1. $1 \in \mathbb{N}$.
2. If $n \in \mathbb{N}$, then $S(n) \in \mathbb{N}$.
3. $S(n) \neq 1$ for every $n \in \mathbb{N}$.
4. If $n, m \in \mathbb{N}$ and $n \neq m$, then $S(n) \neq S(m)$.
5. *Induction.* Let $P(n)$ be any property about a natural number n. Suppose that $P(0)$ is true, and suppose that whenever $P(n)$ is true, $P(S(n))$ is also true. Then $P(n)$ is true for every natural number n.

This should match with our original, informal definition of the natural numbers. We have formalized the 'counting forward' process with the *successor function* $S(n)$.

Remark. We are going to assume various things about the way we write down natural numbers using the decimal system. You can assume that when we write something like $n = 3$, we really mean $n = S(S(1))$ etc.

§1.1.1 Addition

Now we have defined natural numbers, but as of yet we can do nothing more look upon them fondly and increment them using the function $S(n)$. We will now begin to remedy that by defining addition and multiplication.

Definition 1.2 (Addition)

We define **addition** to be an operation $+$ such that for $m, n \in \mathbb{N}$, we have $m + 1 = S(m)$, and $m + (n + 1) = (m + n) + 1$.

This definition defines addition for all natural numbers by induction. We are now able to state and prove various properties of addition.

Proposition 1.3 (Properties of Addition)

For all $a, b, c \in \mathbb{N}$. Then

(i) *Addition is commutative.* $a + b = b + a$.

(ii) *Addition is associative.* $(a + b) + c = a + (b + c)$.

(iii) *Cancellation law.* If $a + b = a + c$ then $b = c$.

Proof Sketch. Use induction.[a] □

[a] The proofs for these sorts of statements tend do be slightly dull and laborious, and for this reason they have been excluded. If you wish to read them, I encourage you to consult a textbook/some other reference material.

§1.1.2 Order

We can now use addition to define an ordering on the natural numbers.

Definition 1.4 (Ordering of the Natural Numbers)

Let $n, m \in \mathbb{N}$. We say that n is **greater than or equal to** m, written $n \geq m$ if and only if $n = m$ or $n = m + a$ for some $a \in \mathbb{N}$. We say n is **strictly greater than** m if $n \geq m$ and $n \neq m$.

Proposition 1.5 (Properties of Ordering)

Let $a, b, c \in \mathbb{N}$. Then

(i) *Order is reflective.* $a \geq a$.

(ii) *Order is transitive.* If $a \geq b$ and $b \geq c$, then $a \geq c$.

(iii) *Order is anti-symmetric.* If $a \geq b$ and $b \geq a$, then $a = b$.

(iv) *Addition preserves order.* $a \geq b$ if and only if $a + c \geq b + c$.

(v) $a > b$ if and only if $a \geq b + 1$.

Proposition 1.6 (Trichotomy)

Let a and b be natural umbers. Then exactly one of the following is true: $a < b$, $a = b$ or $a > b$.

§1.1.3 Multiplication

We can also define another familiar operation, multiplication, in the same inductive/recursive fashion that we defined addition.

Definition 1.7 (Multiplication)

We define **multiplication** to be an operation $\times$ such that for $m, n \in \mathbb{N}$, $m \times 1 = m$, and $m \times (n + 1) = (m \times n) + m$.

As before, induction implies that this is defined for all natural numbers. It also guarantees that multiplying two natural numbers is a natural number.

Notation. We will use $a \times b = a \cdot b = ab$ when referring to multiplication.

Proposition 1.8 (Properties of Multiplication)

For all $a, b, c \in \mathbb{N}$. Then

(i) *Multiplication is commutative.* $a \times b = b \times a$.

(ii) *Multiplication is associative.* $(a \times b) \times c = a \times (b \times c)$.

(iii) *Distributive law.* $a \times (b + c) = a \times b + a \times c$.

(iv) *Cancellation law.* If $a \times b = a \times c$ then $b = c$.

(v) *Multiplication preserves order.* If $a < b$, then $a \times c < b \times c$.

Remark. The final two statements in the proposition above, the cancellation law and that multiplication preserves order, only hold because we are dealing with natural numbers. These properties do not hold in general if we allow a, b or c to be integers.

We could go further and define other common operations such as exponentiation, factorials and so on, all of which can be defined in the same fashion. However, in the interest of space, these definitions have been omitted.

§1.2 Strong Induction

There is a more useful form of induction that can be used, now that we have defined an ordering on the natural numbers.

Proposition 1.9 (Strong Induction)

Suppose that we have some property $P(n)$ about a natural number n. If we have $P(1)$, and for all $n \in \mathbb{N}$ we have that $P(m)$ for $m \leq n$ implies $P(n+1)$, then $P(n)$ holds for all $n \in \mathbb{N}$.

Proof. This follows from ordinary induction using the property "$P(n)$ for all $m \leq n$". □

Informally, the principle of strong induction means that whenever we are proving some property $P(n)$ holds for all $n \in \mathbb{N}$ by induction, we can feel free to assume that $P(m)$ holds for $m \in \mathbb{N}$ with $m < n$.

Remark (For Pedants)**.** Technically, we don't need to check the case $P(1)$ separately, as it is implied by the condition if suitably interpreted. Still, it's safer to just check $P(1)$.

Some other equivalent but useful[2] forms of induction are listed below. Let $P(n)$ be a property, then

- *Existence of a Minimal Counterexample*

  If $P(n)$ is false for some $n \in \mathbb{N}$, then there exists $n_0 \in \mathbb{N}$ such that $P(n_0)$ is false but $P(m)$ is true for all $m < n$.

- *The Well-Ordering Principle*

  If $P(n)$ is true for some $n \in \mathbb{N}$, then there exists a minimal n_0 such that $P(n_0)$ is true.

§1.3 The Integers and Rationals

The previous section defined the natural numbers, along with some ways that we can use them. We will now go one step further and define the *integers*, and the *rationals*.

[2] Some texts will claim that we can replace the induction axiom in Peano axioms with one of these other forms. This is incorrect, and typically one will need to add additional axioms alongside to keep the set of axioms equivalent.

Definition 1.10 (Integers)

The **integers** are a set $\mathbb{Z}$ consisting of all symbols n, $-n$ and 0, where $n \in \mathbb{N}$.

We can then define addition, multiplication and subtraction in (and subtraction) on the integers by extending our definition on the natural numbers in the obvious way. We can also check that the properties of addition we had before still hold. There are some additional properties that the integers have.

Proposition 1.11 (Algebraic Properties of the Integers)

Let $a, b \in \mathbb{Z}$. Then[a]

(i) *Identity.* $a + 0 = a$.

(ii) *Existence of an Additive Inverse.* For all $a \in \mathbb{Z}$, there exists $b \in \mathbb{Z}$ such that $a + b = 0$.

[a] These properties imply that the integers are a *group*.

Proof. $a+0 = a$ holds automatically due to our definition of addition on the integers. Then, we always have a $b \in \mathbb{Z}$ such that $a + b = 0$, as we can let $b = -a$. □

We also obtain another interesting property of multiplication.

Proposition 1.12 (Zero Product Law)

Let a and b be integers such that $a \times b = 0$. Then either $a = 0$, $b = 0$ or both.

Proof. Assume for the purposes of contradiction that both $a \neq 0$ and $b \neq 0$. Then we must have either $a > 0$ or $a < 0$ by trichotomy. If $a > 0$, then $a \times b > 0$ if $b > 0$, or $a \times b < 0$ if $b < 0$. Otherwise, if $a < 0$, then $a \times b < 0$ if $b > 0$, or $a \times b > 0$ if $b < 0$. These are all possible cases, and thus we never have that $a \times b = 0$. Thus at least one of a and b must be zero. □

Remark (Caveats)**.** We noted earlier that there was some properties of the natural numbers that don't hold over the integers. Specifically, if we have $a < b$ for $a, b \in \mathbb{Z}$, and we multiply by a negative number, then the order is no longer preserved (it is switched). Also, the cancellation law only applies when we are cancelling a non-zero integer.

We can now use the integers to define the *rationals*.

Definition 1.13 (Rationals)

The **rationals** are a set $\mathbb{Q}$ of expressions a/b for some $a, b \in \mathbb{Z}$, with $b \neq 0$. We will define equality between rationals such that $a/b = c/d \iff ad = bc$.

This definition implicitly defines $\mathbb{Q}$ using *equivalence classes*, which will be discussed later. We need to be slightly more careful in defining addition on $\mathbb{Q}$, as we will need to ensure that it respects the equality relation between rationals.

Definition 1.14 (Addition on $\mathbb{Q}$)

For a/b and $c/d \in \mathbb{Q}$, we define addition such that

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

We will now need to explicitly check that this definition is valid.

Proposition 1.15

Addition is well defined on $\mathbb{Q}$.

Proof. Let $a/b = a'/b'$ and $c/d = c'/d'$ be rationals. We show that

$$\begin{aligned} &\frac{a}{b} + \frac{c}{d} = \frac{a'}{b'} + \frac{c'}{d'} \\ \iff &\frac{ad+bc}{bd} = \frac{a'd'+b'c'}{b'd'} \\ \iff &(ad+bc)(b'd') = (a'd'+b'c')(bd) \\ \iff &(ab')dd' + (cd')bb' = (a'b)d'd + (c'd)bb', \end{aligned}$$

which follows from $ab' = a'b$ and $cd' = c'd$. □

To see why such a check was needed, consider the following example.

Example 1.16

We *cannot* define an operation on $\mathbb{Q}$ sending $a/b \to a^2/b^3$, as we would have $1/2 \to 1/8$ and $2/4 \to 4/64 = 1/16$, which is inconsistent.

We can then define multiplication in the same way as it was defined for integers, and we can check that all of the usual properties still hold. We can also define ordering, where $a/b < c/d$ if $ab < bc$. The rules from ordering $\mathbb{Z}$ still hold.

With all of these definitions in place, we can now start looking at some of the more interesting properties of numbers.

§1.4 Primes and Divisibility

We will now begin to discuss some actual number theory, beginning with the incredibly important concept of *divisibility*. You are likely to be familiar with this already, so we will jump into some definitions.

Definition 1.17 (Divisibility)

For $a, b \in \mathbb{N}$, we say that a **divides** b, written $a \mid b$ if there exists some $c \in \mathbb{N}$ such that $b = ac$.

If this is the case, we say that b is a **multiple** of a, and that a is a **divisor** of b. We can now define one of the most fundamental objects in number theory, the *primes*.

Definition 1.18 (Primes)

A natural number $n \geq 2$ is a **prime** if its only divisors are 1 and n. If a natural number is not prime, then it is **composite**.

With this definition, we can begin to state and prove some interesting results about the primes, which should hint at their importance in number theory.

Theorem 1.19

Every natural number is expressible as a product of primes.

Proof. We use induction on n. For $n = 2$ this is true, as 2 is prime. Now given $n > 2$, if n is prime, then we are done. If not, then n is composite so $n = ab$ for some $1 < a, b < n$. By our induction hypothesis, we have $a = p_1 p_2 \cdots p_k$ and $b = q_1 q_2 \cdots q_l$ for some (not necessarily distinct) primes $p_1, \cdots p_k, q_1, \cdots, q_l$, hence $ab = p_1 \cdots p_k q_1 \cdots q_l$, which is the product of primes. Thus we are done by induction. □

A nice consequence of this theorem is that the primes go on forever.

Theorem 1.20 (Euclid)

There are infinitely many primes.

Proof. Suppose there was finitely many primes, say $p_1, \cdots, p_k$. Then consider the number $N = p_1 p_2 \cdots p_k + 1$. Then N has no prime factors, contradicting the fact that n can be written as the product of primes[a]. □

[a] This theorem has an amusingly large number of proofs. A short discussion can be found in the book 'Proofs from the Book'.

Remark. There is no 'pattern' to the primes, in the sense that there is no algebraic formula for the nth prime.

So we know that a number can be written as the product of primes, but is this unique (up to some reordering)? This would seem to be true from experience, so why can't we write a number as the product of primes in two ways. For example, why can't we have $41 \times 101 = 67 \times 73$? Informally, we might think that this is impossible because we can't have 41 dividing 'a bit of' 67 and 'a bit of 73'. What we really need to show that prime factorization is unique is the following: For a prime p, if $p \mid ab$ then $p \mid a$ or $p \mid b$.

Now should this lemma be easy to prove, or hard to prove? That is, should this follow straight from definitions, or not? The answer is that it *cannot* follow straight from definitions. This is because it's about 'primes dividing things', but we define the primes with 'things diving it'. This is the wrong way round! It will take some amount of work to be able to prove this.

§1.4.1 Greatest Common Divisors

This section will begin to build some machinery that will allow us to tackle the problem of showing that prime factorizations are unique. We will begin by defining the notion of *greatest common divisor*, also known as *highest common factor*[3].

[3] This was the term used in the lectures, but I prefer greatest common divisor, so I will use that instead.

Definition 1.21 (Greatest Common Divisor)

For $a, b \in \mathbb{N}$, a natural number d is the **greatest common divisor** of a and b, written $d = \gcd(a, b)$ if

(i) $d \mid a$ and $d \mid b$;

(ii) If $c \mid a$ and $c \mid b$, then $c \mid d$.§

We want to show that the greatest common divisor always exists.

Proposition 1.22 (Divison Algorithm)

For $n, k \in \mathbb{N}$, we can write $n = qk + r$, where $q, r \in \mathbb{N}$ and $0 \leq r < k$.

Proof. We use induction on N. For $n = 1$, this is clearly true. Otherwise, given $n > 1$, we can write $n - 1 = qk + r$ by our inductive hypothesis. Then, if $r < k - 1$, we can write $n = qk + (r + 1)$, and if $r = k - 1$, we can write $n = q(k + 1)$. □

The division algorithm is related to the greatest common divisor using the following lemma.

Lemma 1.23 (Euclid)

For $a, b \in \mathbb{N}$, we can write $a = bq + r$ using the division algorithm. Then $\gcd(a, b) = \gcd(b, r)$.

Proof. If $d \mid a$ and $d \mid b$, then $d \mid a - bq = r$. Otherwise, if $d \mid b$ and $d \mid r$, then $d \mid bq + r = a$, hence the set of common divisors between a and b is the same as the set of divisors between b and r. Thus $\gcd(a, b) = \gcd(b, r)$. □

Theorem 1.24 (Euclidean Algorithm)

For $a, b \in \mathbb{N}$, if we repeatedly apply the division algorithm to get

$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1}, \end{aligned}$$

then $\gcd(a, b) = r_{n-1}$.

Proof. We repeatedly apply Lemma 1.23 until we have $\gcd(a, b) = \gcd(r_{n-1}, 0) = r_{n-1}$. □

The Euclidean Algorithm gives us both a proof that $\gcd(a, b)$ exists for any natural numbers a and b, and it also gives us an efficient way to find it.

Example 1.25

To find $\gcd(87, 52)$, we perform the following series of steps.

$$\begin{aligned} 87 &= 1 \times 52 + 35 \\ 52 &= 1 \times 35 + 17 \\ 35 &= 2 \times 17 + 1 \\ 17 &= 17 \times 1, \end{aligned}$$

thus $\gcd(87, 52) = 1$.

Notation. We will sometimes write $\gcd(87, 52) = \text{hcf}(87, 52) = (87, 52) = 1$. And when $\gcd(a, b) = 1$, we say that a and b are **coprime** or **relatively prime**.

The example above showed that $\gcd(87, 52) = 1$, so is it possible to write $1 = 87x + 52y$ for some $x, y \in \mathbb{Z}$? Looking at the computation we performed to find the greatest common divisor, we can work backwards to get

$$\begin{aligned} 1 &= 1 \times 35 - 2 \times 17 \\ &= 1 \times 35 - 2 \times (52 - 35) \\ &= -2 \times 52 + 3 \times 35 \\ &= -2 \times 52 + 3 \times (87 - 52) \\ &= 3 \times 87 - 5 \times 52, \end{aligned}$$

so it is indeed possible. In fact, there was nothing special about 87 and 52 here, we just used the sequence of steps performed in the Euclidean algorithm. We can formalize this.

Theorem 1.26

For all $a, b \in \mathbb{N}$, there exists $x, y \in \mathbb{Z}$ such that $ax + by = \gcd(a, b)$.

Proof. Run the Euclidean algorithm on a and b, say with some output r_n. Then we can write r_n as some integer linear combination of r_{n-1} and r_{n-2}. We can continue this to write any r_i as an integer linear combination of r_{i-1} and r_{i-2}. Thus eventually, we be able to write $r_n = ax + by$ for some $x, y \in \mathbb{Z}$, and as $r_n = \gcd(a, b)$, we are done. □

Alternate Proof. Let h be the least positive integer linear combination of a and b. We claim that $h = \gcd(a, b)$. If $d \mid a$ and $d \mid b$, then $d \mid ax + by$ for any $x, y \in \mathbb{Z}$, thus $d \mid h$. Now suppose that $h \nmid a$. Then $a = qh + r$ for some $q, r \in \mathbb{Z}$, with $0 < r < h$. So $r - ah = a - q(ax + by)$, which is also a linear combination of a and b. But this is impossible, as it contradicts the minimality of h. Thus $h \mid a$, and by the same argument $h \mid b$. Thus $h = \gcd(a, b)$. □

Remark. The second proof of the theorem above tells us that the greatest common divisor exists, but offers no way to find it, nor a way to find x or y.

This theorem has an interesting application: solving linear diophantine[4] equations.

[4] A diophantine equation is one where you seek only integer solutions.

Corollary 1.27 (Bezout's Lemma)

Let $a, b, c \in \mathbb{N}$. Then the equation

$$ax + by = c$$

has an integer solution if and only if $\gcd(a, b) = c$.

Proof. Let $h = \gcd(a, b)$. If we have a solution $ax + by = c$, then $h \mid a$ and $h \mid b$, thus $h \mid c$. Now, if $h \mid c$, then we have $h = ax + by$ for $a, b \in \mathbb{Z}$ by our previous theorem, and thus we can write $c = \frac{c}{h}ax + \frac{c}{h}by$. Note that $\frac{c}{h}$ is guaranteed to be an integer by our previous argument. □

§1.4.2 The Fundamental Theorem of Arithmetic

We are now in a position to prove the lemma that was discussed quite a few pages ago, which will lead us directly to the uniqueness of prime factorizations.

Theorem 1.28 (Euclid's Lemma)

Let p be a prime, and $a, b \in \mathbb{N}$. Then $p \mid ab$ implies $p \mid a$ or $p \mid b$.

Proof. Suppose that $p \nmid a$. We wish to show that $p \mid b$. Then we have $\gcd(a, p) = 1$, which implies that we can write $ax + py = 1$, for some integers x and y. Then we can multiply by b to get $abx + pby = b$. But then $p \mid ab$ by our hypothesis, and thus $p \mid abx + pby = b$, and so we are done. □

Remark. This theorem implies that if $p \mid a_1 a_2 \cdots a_k$, then p must divide at least one of a_i, for $1 \leq i \leq k$.

Theorem 1.29 (Fundamental Theorem of Arithmetic)

Every natural number $n \geq 2$ is uniquely expressible as a product of primes, up to re-ordering.

Proof. We already proved that such a product exists, so we will now prove that such a product is unique. We use induction on n. Suppose that $n = p_1 \cdots p_k = q_1 \cdots q_l$, where p_i, q_i are all primes. We wish to show that $k = l$, and after reordering, $p_i = q_i$ for all i. We have $p_1 \mid q_1 \cdots q_k$, so $p_1 \mid q_i$ for some i. We can (without loss of generality) label that prime q_1. Hence $p_1 = q_1$. So $n/p_1 = p_2 \cdots p_k = q_2 \cdots q_l$, thus by induction $k = l$ and $p_2 = q_2$, $p_3 = q_3$, etc. □

Aside: Why Unique Factorization Isn't Obvious

In the fundamental theorem of arithmetic, we took the 'things that can't be broken up' (the primes) and we broke up every number as a product of these, uniquely. The same concept makes sense in other places.

Consider instead $\mathbb{Z}[\sqrt{-3}]$, the set of complex numbers of the form $x + y\sqrt{-3}$, where

$x, y \in \mathbb{Z}$. For example, $2 + 7\sqrt{-3}$. With these objects, we can both addition and multiplication two elements in $\mathbb{Z}[\sqrt{-3}]$, and always get back an element in $\mathbb{Z}[\sqrt{-3}]$. This allows us to talk about the notions of 'divides' and 'multiple of' etc in $\mathbb{Z}[\sqrt{-3}]$.

You might think that you can take the 'things you can't break up' in $\mathbb{Z}[\sqrt{-3}]$, and take any element and break it up into those. This is obviously correct (by definition). You may then want to say that this is unique, but that would not be true. For example,

$$4 = 2 \times 2 = (1 + \sqrt{-3})(1 - \sqrt{-3})$$

and all of the terms in this product are 'things you can't break up' in $\mathbb{Z}[\sqrt{-3}]$. This allows us to conclude that unique factorization *fails* in $\mathbb{Z}[\sqrt{-3}]$.

The fundamental theorem of arithmetic has some useful applications.

- *Factors*

  Let's consider what are the factors of $2 \cdot 3^7 \cdot 11$ are. Certainly any $2^a \cdot 3^b \cdot 11^c$ where $0 \leq a \leq 3$, $0 \leq b \leq 7$ and $0 \leq c \leq 1$ is a factor. Also, there can't be any others. For example, if $7 \mid 2 \cdot 3^7 \cdot 11$, then we'd get a prime factorization involving a 7, which would contradict the uniqueness of prime factorization.

  In general, the factors of $n = p_1^{a_1} \cdots p_k^{a_k}$ are precisely numbers of the form $p_1^{b_1} \cdots p_k^{b_k}$, where $0 \leq b_i \leq a_i$, for all $1 \leq i \leq k$.

- *Greatest Common Divisors*

  If we wanted to find the common factors of $2^3 \cdot 3^7 \cdot 5 \cdot 11^3$ and $2^4 \cdot 3^2 \cdot 11 \cdot 13$, we could note that the common factors are $2^a \cdot 3^b \cdot 11^c$, where $0 \leq a \leq 3$, $0 \leq b \leq 2$ and $0 \leq c \leq 1$. So the greatest common divisor is $2^3 \cdot 3^2 \cdot 11$.

  In general the greatest common divisor of $p_1^{a_1} \cdots p_k^{a_k}$ and $p_1^{b_1} \cdots p_k^{b_k}$ (with $a_i, b_i \geq 0$) is $p_1^{\min(a_1,b_1)} \cdots p_k^{\min(a_k,b_k)}$.

- *Lowest Common Multiples*

  The common multiples of $2^3 \cdot 3^7 \cdot 5 \cdot 11^3$ and $2^4 \cdot 3^2 \cdot 11 \cdot 13$ are all of the form $2^a \cdot 3^b \cdot 5^c \cdot 11^d \cdot 13^e \cdot n$ (for a positive integer n) where $a \geq 4, b \geq 7, c \geq 1, d \geq 3$ and $e \geq 1$. We can then get the lowest common multiple as $2^4 \cdot 3^7 \cdot 5^1 \cdot 11^3 \cdot 13^1$.

  Then in general, the lowest common multiple of $p_1^{a_1} \cdots p_k^{a_k}$ and $p_1^{b_1} \cdots p_k^{b_k}$ (with $a_i, b_i \geq 0$) is $p_1^{\max(a_1,b_1)} \cdots p_k^{\max(a_k,b_k)}$.

§1.5 Modular Arithmetic

It is a common occurrence in number theory that we will consider numbers that differ by a multiple of some fixed number to be equivalent. An example of this is the value of $(-1)^n$, which depends only on whether n is odd or even – that is, the values of n that differ by a multiple of 2 will give the same result. To give another example, the last digit of two numbers will be the same when the numbers differ by some multiple of 10. Modular arithmetic (or congruence notation) is a way of expressing such equivalences, where we have two integers a and b that differ by some fixed number m.

Definition 1.30 (Integers Modulo n)

Let $n \in \mathbb{N}$. The **integers modulo** n, written $\mathbb{Z}/n\mathbb{Z}$ consist of the integers where we regard two to be the same if they differ by a multiple of n.

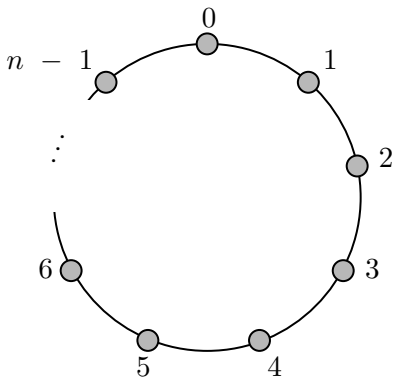
Definition 1.31 (Congruence Notation)

If x and y are integers that are the same modulo n, then we write

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

This notation implies that $x \equiv y \pmod{n}$ if and only if $n \mid x-y$, that is, if $x = y+kn$ for some $k \in \mathbb{Z}$. Note that no two of $0, 1, \dots, n-1$ are congruent $\pmod{n}$, and every $x \in \mathbb{Z}$ is congruent to exactly one of these modulo n (this follows from the division algorithm).

We can view $\mathbb{Z}/n\mathbb{Z}$ as the following picture (indeed this is the 'correct picture' of $\mathbb{Z}/n\mathbb{Z}$).



With this mental picture, we can begin to build up some properties of modular arithmetic.

Proposition 1.32 (Arithmetic Modulo n)

If $a \equiv a' \pmod{n}$ and $b \equiv b' \pmod{n}$, then $a+b \equiv a'+b' \pmod{n}$ and $ab \equiv a'b' \pmod{n}$.

Proof. We have $a' = a + kn$ and $b' = b + jn$, for some $k, j \in \mathbb{Z}$. So $a' + b' = a+b+(k+j)n \equiv a+b \pmod{n}$, and $a'b' = (a+kn)(b+jn) = ab+(bk+aj+kjn)n \equiv ab \pmod{n}$. □

Many of the other rules of arithmetic are inherited from $\mathbb{Z}$, for example we have $a+b \equiv b+a \pmod{n}$ as $a+b = b+a$ in $\mathbb{Z}$. Also some of the number-theoretic facts we have previously established can be expressed in modular arithmetic.

Example 1.33 (Euclid's Lemma in Modular Arithmetic)

The statement that for a prime p, '$p \mid ab \implies p \mid a$ or $p \mid b$' is equivalent saying that $ab \equiv 0 \pmod{p}$ implies that $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{0}$.

§1.5.1 Modular Inverses

We will now begin to look at the multiplicative structure of $\mathbb{Z}/n\mathbb{Z}$.

Definition 1.34 (Modular Inverses)

For $a, b \in \mathbb{Z}$, we say that b is an **inverse** of a if $ab \equiv 1 \pmod{n}$.[a]

[a] This should be reminiscent of the notion of inverses from group theory.

Example 1.35

In $\mathbb{Z}/10\mathbb{Z}$, the inverse of 3 is 7, as $3 \times 7 = 21 \equiv 1 \pmod{10}$. The inverse of 4 does not exist, as for all $x \in \mathbb{Z}$, $4x \not\equiv 1 \pmod{10}$, as $4x$ is even.

This example shows that inverses do not always exist for an arbitrary modulus n.

Notation. We write a^{-1} to mean the inverse of a (modulo some n).

Proposition 1.36 (Properties of Modular Inverses)

In $\mathbb{Z}/n\mathbb{Z}$,

(i) If a modular inverse exists, then it is unique modulo n.

(ii) If a has an inverse, and $ab \equiv ac \pmod{n}$, then $b \equiv c \pmod{n}$. That is, we can cancel invertible elements.[a]

[a] We *cannot* cancel elements in general.

Proof. To show (i), suppose there exists $b, c \in \mathbb{Z}$ such that for some a, $ab \equiv ac \equiv n \pmod{n}$. Then $b(ac) \equiv bab \implies c \equiv b \pmod{n}$. Then to show (ii), we can just multiply both sides by a^{-1}. □

When we are working modulo a prime, things tend to be 'nicer'.

Proposition 1.37

Let p be a prime, then every $a \neq 0 \pmod{p}$ has an inverse modulo p.

Proof. We have $\gcd(a, p) = 1$, so we can write $ax + py = 1$ for some $x, y \in \mathbb{Z}$. But then $ax = 1 - py$, and thus $ax \equiv 1 \pmod{p}$. □

Alternate Proof. In $\mathbb{Z}/n\mathbb{Z}$, consider the multiplies of a: $0 \times a$, $1 \times a$, $2 \times a$, …, $(p-1) \times a$. We wish to show that one of them is 1. We have p items written down, and I claim these are all *distinct* in $\mathbb{Z}/p\mathbb{Z}$. If $ia = ja \implies (i-j)a \equiv 0$, which implies $a \equiv 0$ or $i - j \equiv 0 \implies i \equiv j \pmod{p}$. Hence, there must be $0, 1, \ldots, p-1$ in this list, in some order. Thus $xa \equiv 1 \pmod{p}$, for some $x \in \mathbb{Z}/n\mathbb{Z}$. □

We can generalize this proposition by considering which properties of the primes we used in the previous proof.

Proposition 1.38

Let $n \in \mathbb{N}$. Then a has an inverse modulo n if and only if $\gcd(a, n) = 1$.

Proof. We have $\gcd(a, n) = 1 \iff ax + ny = 1$, for some $x, y \in \mathbb{Z}$. Then $ax = 1 - ny \equiv 1 \pmod{n}$. □

There is a commonly used function to count how many elements have an inverse modulo some n.

Definition 1.39 (Euler's Totient Function)

Euler's totient function $\phi(n)$ is the number of numbers $1, 2, \ldots, n$ that are relatively prime to n. This function counts the number of invertibles or **units** in $\mathbb{Z}/n\mathbb{Z}$.

Example 1.40

If p is a prime, then $\phi(p) = p - 1$, $\phi(p^2) = p^2 - p$ (we cannot have $p, 2p, \ldots, pp$). If p and q are distinct primes, then $\phi(pq) = pq - p - q + 1$.

§1.5.2 Exponentiation

Consider the powers of 2 in $\mathbb{Z}/7\mathbb{Z}$. We have the sequence

$$\begin{aligned} 2^1 &\equiv 2 \pmod{7} \\ 2^2 &\equiv 4 \pmod{7} \\ 2^3 &\equiv 1 \pmod{7} \\ 2^4 &\equiv 2 \pmod{7} \\ &\vdots \end{aligned}$$

This sequence will repeat periodically, with a period of 3. It turns out that in general, this will always occur. This can be seen in the following theorem, which turns out to be an incredibly important result which is used all the time. This theorem has many proofs, but we present one particularity nice one, which emphasizes the theory that we have developed so far.

Theorem 1.41 (Fermat's Little Theorem)

Let p be a prime, then for $a \not\equiv 0 \pmod{p}$, we have

$$a^{p-1} \equiv 1 \pmod{p}.$$

Proof. Consider the numbers $1 \cdot a$, $2 \cdot a$, $\ldots$, $(p-1) \cdot a$ modulo p. They are distinct, as if $ai \equiv aj$ then $i \equiv j \pmod{p}$, because a is invertible. They are also non-zero. So we must have the numbers $1, 2, \ldots, p-1 \pmod{p}$ in some order. Then, if we multiply together we have

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod{p},$$

so we can cancel $(p-1)!$ as it's the product of invertibles, to obtain $a^{p-1} \equiv 1 \pmod{p}$. □

As before, we can consider what happens in the case of a non-prime modulus.

Theorem 1.42 (Fermat-Euler Theorem)

Let $n \in \mathbb{N}$. Then for $a \not\equiv 0 \pmod{n}$, we have

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

Proof. We copy the proof of Fermat's little theorem. Let the units in $\mathbb{Z}/n\mathbb{Z}$ be $x_1, \dots, x_{\phi(n)}$. Consider $a \cdot x_1$, $a \cdot x_2$, $\dots$, $a \cdot x_{\phi(n)}$. They are distinct and invertible as before, so they must be $x_1, \dots, x_{\phi(n)}$ in some order. Multiplying them all together,

$$a^{\phi(n)} x_1 x_2 \cdots x_{\phi(n)} \equiv x_1 x_2 \cdots x_{\phi(n)} \pmod{n},$$

and then we can cancel each x_i to obtain $a^{\phi(n)} \equiv 1 \pmod{n}$. □

We can pause for a moment and notice that in the proof of Fermat's little theorem, we could cancel $(p-1)! \pmod{p}$ because it was non-zero. So what exactly was it? We can try an example of $p = 5$, then we have $4! = 1 \equiv 24 \equiv -1 \pmod{5}$. We can also try $p = 7$ to get $6! = 720 \equiv -1 \pmod{7}$. It can be proved that this is always the case.

First, we prove a small lemma.

Lemma 1.43

Let p be prime, then $x^2 \equiv 1 \pmod{p}$ implies $x \equiv 1$ or $x \equiv -1 \pmod{p}$.[a]

[a] Note that this is only true because p is prime. For example, $1^2 \equiv 3^2 \equiv 5^2 \equiv 7^2 \equiv 1 \pmod{8}$.

Proof. $x^2 \equiv 1 \implies x^2 - 1 \equiv 0 \pmod{p}$, which we can factor as $(x+1)(x-1) \equiv 0 \pmod{p}$ and thus $x \equiv 1$ or $x \equiv -1 \pmod{p}$. □

Remark. It turns out that for a prime p, a non-zero polynomial in $\mathbb{Z}/p\mathbb{Z}$ of degree k always has at most k roots.

We can now prove our result about the value of $(p-1)! \pmod{p}$.

Theorem 1.44 (Wilson's Theorem)

Let p be a prime, then

$$(p-1)! \equiv -1 \pmod{p}.$$

Proof. Note that this is true for $p = 2$, so we may assume that $p \geq 2$. Consider the numbers $1, 2, \dots, p-1$ modulo p. We can pair up each a with its inverse a^{-1} for $a \neq a^{-1}$. But $a = a^{-1} \iff a^2 = 1$, so the only elements that are their own inverse are 1 and -1. Thus $1, 2, \dots, p-1$ consists of some pairs a, a^{-1} and 1 and -1. Thus when we multiply, $(p-1)! \equiv 1^{\frac{p-3}{2}} \cdot 1 \cdot -1 \equiv -1 \pmod{p}$. □

To see how we can use both Wilson's theorem and Fermat's little theorem together, we

can consider another question that follows naturally from Lemma 1.43.

Lemma 1.45 ($x^2 + 1$ Lemma)

Let p be an odd prime. Then $x^2 \equiv -1 \pmod{-1}$ has a solution if and only if $p \equiv 1 \pmod 4$.

Proof. If $p = 4k + 3$, suppose that $x^2 \equiv -1 \pmod p$. Then we have from Fermat's little theorem that $x^{4k+2} \equiv 1 \pmod p$, but $x^{4k+2} \equiv (x^2)^{2k+1} \equiv (-1)^{2k+1} \equiv -1$, which is a contradiction.

We now construct an example to show that $p \equiv 1 \pmod 4$ works. By Wilson's theorem, $(4k!) \equiv -1 \pmod p$. Then, noting that $4k - j \equiv -j - 1 \pmod p$, $(2k)!^2 \equiv (2k!)^2 \cdot (-1)^{2k} \equiv (4k)! \equiv -1 \pmod p$, so a solution exists. □

§1.5.3 Congruence Equations and the Chinese Remainder Theorem

We now return back to the topic of linear diophantine equations, this time looking specifically at linear congruence equations. Let's look at an example.

Example 1.46 (Solving a Linear Congruence Equation)

Solve $7x \equiv 4 \pmod{30}$.

Solution. First we *find a solution.*

We have $\gcd(7, 30) = 1$, so we can run the Euclidean algorithm on 7 and 30 to obtain $7 \times 13 - 30 \times 3 = 1$. So $7 \times 13 \equiv 1 \pmod{30}$, hence $7 \times 52 \equiv 4 \pmod{30}$. So $x \equiv 52 \pmod{30}$ is a solution.

Now we show there is *no more solutions.*

If x' is a solution, then we have $7x \equiv 4 \pmod{30}$ and $7x' \equiv 4 \pmod{30}$. Thus $7x \equiv 7x' \pmod{30}$, which implies that $x \equiv x' \pmod{30}$ as 7 is invertible. So $x \equiv 52 \pmod{30}$ is the only solution. □

There is a shorter method that can be used too, if we can quickly spot the inverse of the coefficient of our variable.

Example 1.47

Solve $3x \equiv 9 \pmod{28}$.

Solution. $3x \equiv 9 \pmod{28} \iff 19 \cdot 3x \equiv 19 \cdot 9 \pmod{28}$, as 19 is invertible. Thus $x \equiv 171 \equiv 3 \pmod 8$. □

If we don't have that the coefficient and the modulus is coprime, it's typically easier to write the equation as a standard linear diophantine equation.

Example 1.48

Solve $10x \equiv 12 \pmod{34}$.

Solution. $10x \equiv 12 \pmod{34} \iff 10x = 12 + 34y$ for some $y \in \mathbb{Z}$. Then $10x = 12 + 34y \iff 5x = 6 + 17y$, which we can write as $5x \equiv 6 \pmod{17}$. We can then solve as before. □

Now consider a simultaneous congruence equation. Specifically, consider the equations

$$\begin{aligned} x &\equiv 6 \pmod{17} \\ x &\equiv 2 \pmod{19}. \end{aligned}$$

Do we expect a solution to this? We might guess yes, as 17 and 19 are coprime, so intuitively 'modulo 17 and modulo 19 should be independent of each other'. What about these equations:

$$\begin{aligned} x &\equiv 6 \pmod{34} \\ x &\equiv 11 \pmod{36}. \end{aligned}$$

We shouldn't expect solutions to these equations. Even if you try and figure out if x is even or odd, you'll run into issues. This intuition can be formalized into a theorem.

Theorem 1.49 (Chinese Remainder Theorem)

Let m and n be relatively prime positive integers. Then for any integers a and b, there is solution to

$$\begin{aligned} x &\equiv a \mod m \\ x &\equiv b \mod n, \end{aligned}$$

and this solution is unique modulo mn.

Proof. First we prove existence. We have $ms + nt = 1$ for some $s, t \in \mathbb{Z}$ by Bezout's lemma. Then $ms \equiv 0 \mod m$ and $1 \pmod{n}$ and $nt \equiv 0 \pmod{n}$ and $1 \pmod{m}$. Hence $x = a(nt) + b(ms)$ has $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$.

To show uniqueness, certainly any $x' \equiv x \pmod{mn}$ is also a solution. Conversely, suppose x' has $x' \equiv a \pmod{m}$ and $b \pmod{n}$. Then $x' \equiv x \pmod{m}$ and $x' \equiv x \pmod{n}$, that is, $m \mid x' - x$ and $n \mid x' - x$. Thus $mn \mid x' - x$, as $\gcd(m, n) = 1$, and thus $x' \equiv x \pmod{mn}$. □

This theorem generalizes directly.

Theorem 1.50 (General Chinese Remainder Theorem)

Let $m_1, \ldots, m_k$ be pairwise relatively prime positive integers, and let

$$M = m_1 m_2 \cdots m_k.$$

Then for all integers $x_1, \ldots, x_k$, the equations

$$\begin{aligned} x &\equiv x_1 \pmod{m_1} \\ x &\equiv x_2 \pmod{m_2} \\ &\vdots \\ x &\equiv x_k \pmod{m_k} \end{aligned}$$

have a solution x that is unique modulo M.

Proof Sketch. Reduce the number of equations by applying the two variable Chinese remainder theorem to the last two equations. □

§1.5.4 An Application of the Fermat-Euler Theorem: RSA Encryption

RSA is an algorithm that underpins a huge amount of modern cryptography and security.

Normally, to send a coded message to someone, you would encode the message, which you could then send to the intended receiver who would decode it. This decoding would be performing the encoding operation in reverse – they are inverse operations. Because of this, it would seem obvious in this way that knowing how to encode is the same as knowing how to decode. For example, if your encoding is 'add 1 to each letter', a *Caesar cipher*, then you immediately know how to decode: 'take 1 from each letter'. For another example, if your encoding is 'add 1 to each letter then reverse each word', then to decode you 'reverse each word and subtract 1 to each letter'. This belief, that knowing how to encrypt implies knowing how to decrypt, was held for centuries, but it isn't quite correct. We will see this in the RSA encryption system.

Begin with two large distinct primes p and q, and let $n = pq$. Let's say we want to encrypt some element $x \in \mathbb{Z}/n\mathbb{Z}$ – this may involve taking the message we want to send and splitting it up in some way. We will do that in the following way.

Fix a 'coding exponent' e. To encode a message $x \in \mathbb{Z}/n\mathbb{Z}$, raise it to the power e.

$$x \longmapsto x^e.$$

Now let's figure out how to decode this message. We want some $d \in \mathbb{Z}/n\mathbb{Z}$ so that $(x^e)^d \equiv x \pmod{n}$. First, assume that x and n are coprime (ensure that the message x is constructed to have this condition). We know from the Fermat-Euler theorem that $x^{\phi(n)} \equiv 1 \pmod{n}$, and also that $x^{k\phi(n)} \equiv 1 \pmod{n}$, for all $k \in \mathbb{Z}$. So we have

$$x^{k\phi(n)+1} \equiv x \pmod{n}.$$

So it suffices to find a d such that $de = k\phi(n) + 1$, for some $k \in \mathbb{Z}$. That is, we want an inverse of e modulo $\phi(n)$. We can do this by running the Euclidean algorithm on the numbers e and $\phi(n)$. This implies we needed to have picked e coprime with $\phi(n)$.

With the encryption and decryption methods defined, we can now consider what information is needed to perform each step.

To encode, we take $x \longmapsto x^e \pmod{n}$[5]. This means we need to know that will need to know n (so you know what modulus to use) and also the power e.

To decode, we take $y \longmapsto y^d \pmod{n}$, so we need to know n and d. But we worked out d as the inverse of e modulo $\phi(n)$, so we really need n, e and $\phi(n)$. Recall that

$$\phi(n) = n - p - q + 1.$$

So to compute $\phi(n)$, we need to know the factors of n. However, the problem of factoring n is thought to be incredibly hard, and the best known methods of today would take

[5]This process can be done quite quickly by squaring the number whenever possible.

longer than the age of the universe to factor n into two 100 digit long primes. This means that if e and n were published, then anyone could encode a message, but only you (or someone who knows the factors of n) would be able to decode. So, with RSA encryption, anyone can encrypt a message and send it to you, but only you can decrypt it.

§2 The Reals

Moving on from number theory, this chapter will center on the questions of *what is a real number* and *what can we assume like them.* Unlike when we defined the natural numbers, in this chapter we will not just define the reals formally and then proceed as normal; instead, we will try and understand some of the more subtle aspects of the real numbers, that will eventually be carried through in 'Analysis'.

§2.1 Why We Need Real Numbers

Recall that in section 1, we began by defining the natural numbers $\mathbb{N}$, extended them to the integers $\mathbb{Z}$, and extended them to the rationals $\mathbb{Q}$. So why would we not stop there? In the following example, we will show that for some (many) purposes, the rationals are not adequate.

Proposition 2.1

There is no $x \in \mathbb{Q}$ such that $x^2 = 2$.

Proof. Assume that $x \geq 0$. If there was such a rational, we could write

$$x = \frac{a}{b},$$

where $a, b \in \mathbb{Z}$. We can square this to get

$$x^2 = \frac{a^2}{b^2} = 2 \implies a^2 = 2b^2,$$

but this is a contradiction of the fundamental theorem of arithmetic, as the power of 2 in a^2 would have to be odd[a]. □

Alternate Proof. Suppose that for some $x = a/b \in \mathbb{Q}$ where $a, b \in \mathbb{N}$, we have $x^2 = 2$. Note that for any integers c and d, the number $cx + d$ is of the form e/b, for some $e \in \mathbb{Z}$. So, if $cx + d > 0$, then $cd + d > \frac{1}{b}$. But $x^2 = 2 \implies 0 < x - 1 < 1$, we can say $0 < (x-1)^k < 1/b$ for large enough k. But this is a contradiction, because $(x-1)^k$ is of the form $cx + d$. □

[a] The same proof shows that if there exists $x \in \mathbb{Q}$ so that $x^2 = n$, for $n \in \mathbb{N}$, then n must be a square number. This is because each exponent in the prime factorisation of n must be even.

Informally, what this result shows is that $\mathbb{Q}$ has a 'gap' at $\sqrt{2}$, and in constructing the reals, we will try formally define a number system that has no such 'gaps'. This immediately raises the question of how we can define such a system, formalizing our rather vague notion of a 'gap'. This must be done while only making statements about the rationals.

Consider again our example that there is no rational x such that $x^2 = 2$. With this in mind, define the set S as follows:

$$S = \{x \in \mathbb{Q} \ : \ x^2 < 2\}.$$

We will now think about bounds on the elements in this set.

Definition 2.2

A set S is **bounded above** if there exists $x \in \mathbb{R}$ with $x \geq y$ for all $y \in S$. Such an x is a **least upper bound** of S if x is an upper bound for S, and every upper bound x' of S satisfies $x < x'$.

It should be clear that 1.5 is an upper bound on the elements in this set. 1.42 is also an upper bound. By setting such upper bounds, we can get arbitrarily close to a number x such that $x^2 = 1$, that is there cannot be a *least* upper bound. If there was, it would have to be a number so that $x^2 = 2$, but we know there isn't such a rational. It's using this idea of 'least upper bounds' that we can define the reals.

Definition 2.3 (The Reals)

The **reals** are a set $\mathbb{R}$ with elements 0 and 1 (where $0 \neq 1$), along with operations $+$ and $\times$ and an ordering $<$ such that:

1. $+$ is commutative, associative, has an identity 0, and every $x \in \mathbb{R}$ has an inverse.
2. $\times$ is commutative, associative, has an identity 1, and every $x \in \mathbb{R}$ where $x \neq 0$ has an inverse.
3. $\times$ is distributive over $+$.
4. $\forall a, b$, exactly one of $a < b$, $a = b$ and $b < a$ holds. Also if $a < b$ and $b < c \implies a < c$.
5. $\forall a, b, c \in \mathbb{R}$, $a < b \implies a + c < b + c$ and $a < b \implies ac < bc$ if $c > 0$.
6. *Least Upper Bound.* Every non-empty set of reals that is bounded above has a least upper bound.

It's worth noting that the conditions in axiom 6 of the set of reals being 'non-empty' and 'bounded above' is needed because if there was no upper bound, then there is certainty no *least* one. Also, if the set was empty then every $x \in \mathbb{R}$ would be an upper bound, and there would be no least one.

From these axioms, it's straightforward to check some very basic facts, for example that $0 < 1$.

Example 2.4

We have $0 < 1$.

Proof. If that were not the case, as we have $0 \neq 1$ as an axiom, we would need $1 < 0$. This would imply $0 < -1$ and thus it is 'positive', so we can multiply by the 'positive' -1 to get $0 < 1$, which contradicts our assumption. □

We can also (as you may expect) view $\mathbb{Q}$ as contained in $\mathbb{R}$, by identifying $a/b \in \mathbb{Q}$ with $a, b^{-1} \in \mathbb{R}$. In this system, axioms 1 to 5 hold, but axiom 6 does not hold. We have defined the reals here by specifying the axioms that they satisfy, but that is not the only way. It is possible to construct $\mathbb{R}$ from $\mathbb{Q}$, in a way that is (vaguely) similar to the way we constructed $\mathbb{Q}$ from $\mathbb{Z}$. This is, surprisingly, not as helpful as one might imagine, so we will skip over this topic for now.

§2.1.1 Least Upper Bounds

The notion of least upper bounds is so central to the real numbers that we will consider some examples of different sets and their least upper bounds, in order for you to get see some of the properties. We will use the following notation for the least upper bound.

Definition 2.5 (Supremum)

The least upper bound of a set S is also known as the **supremum** of S, written $\sup S$.

Example 2.6 (Sets and Least Upper Bounds)

1. The set $S = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$ has a least upper bound $x = 1$. This is because 1 is an upper bound, and also every upper bound y has $y \geq 1$, as $1 \in S$.
2. Consider the set $S = \{x \in \mathbb{R} : 0 < x < 1\}$. We have 2 as an upper bound, as $x \in S$ has $x < 1 < 2$. Also $3/4$ is not an upper bound, as $7/8 \in S$, and $3/4 < 7/8$. Now the least upper bound of S is 1, because 1 is an upper bound, and no upper bound c has $c < 1$, as otherwise, $c \in S$, and we can take $c + (1 - c)/2 \in S$.
3. $S = \left\{1 - \frac{1}{n} : n \in \mathbb{N}\right\} = \left\{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\}$. Clearly 1 is an upper bound, and we will return to this example to show that it is the least upper bound.

Theorem 2.7 (Axiom of Archimedes)

$\mathbb{N}$ is not bounded above in $\mathbb{R}$.

Proof. Let's suppose that it was bounded above, in which case we would have $c = \sup \mathbb{N}$, for some $c \in \mathbb{R}$. So $c - 1$ is *not* an upper bound for $\mathbb{N}$. That implies that there is a natural number n with $n > c - 1$. But then $n + 1 \in \mathbb{N}$, and $n + 1 > c$, which is a contradiction. □

Corollary 2.8

For each $t > 0$, there exists $n \in \mathbb{N}$ with $\frac{1}{n} < t$.

Proof. We have some $n \in \mathbb{N}$ with $n > \frac{1}{t}$ by the previous result, so $\frac{1}{n} < t$. □

Remark. Informally, these two results are telling us that $\mathbb{R}$ does not contain any 'infinitely big' or 'infinitesimally small' elements.

We can return to the third example above, and we can now show that $\sup S = 1$, as if there was some $c < 1$ that was an upper bound, then $1 - \frac{1}{n} < c$ for every $n \in \mathbb{N}$, so $1 - c < \frac{1}{n}$, which is impossible (by the corollary we just proved).

Remark (Warning)**.** If a set S has a greatest element (like $[0, 1]$), then that greatest element is $\sup S$, so $\sup S \in \S$. But if S has no greatest element (like $(0, 1)$), then $\sup S \notin S$.

We have been discussing upper bounds of sets, but the least upper bound axiom also implies some facts about lower bounds.

Definition 2.9

A set S is **bounded below** if there exists $x \in \mathbb{R}$ with $x \leq y$ for all $y \in S$. Such an x is a **greatest lower bound** of S if x is a lower bound for S, and every lower bound $x' \neq x$ of S satisfies $x < x'$. We write $x = \inf S$.

Proposition 2.10

If a set S is non-empty and has a lower bound, then it has a greatest lower bound.

Proof. Let the set $-S = \{-y : y \in S\}$. It is non-empty and bounded above, so it has a least upper bound c. So $-c$ is the greatest lower bound of S, $\inf S$. □

§2.1.2 Rationals and Irrationals

Let's return to our original question which prompted our discussion of the real numbers: that there is no $x \in \mathbb{Q}$ with $x^2 = 2$. If we have in any way succeeded, we should be able to do somewhat better than that in the reals.

Proposition 2.11

There is some $x \in \mathbb{R}$ such that $x^2 = 2$.

Proof. Let $S = \{x \in \mathbb{R}, x^2 < 2\}$. We have S is nonempty, and it is bounded above, so $c = \sup S$ for some $c \in \mathbb{R}$. We claim that $c^2 = 2$. If this wasn't the case, then either $c^2 < 2$ or $c^2 > 2$. If $c^2 < 2$, then for $t < \frac{2-c^2}{4}$ we have $(c + t)^2 < 2$, which contradicts c being an upper bound for S. If $c^2 > 2$, then we can let $t < \frac{c^2-2}{4}$ and then $(c - t)^2 > 2$, which contradicts that c is the *least* upper bound for S. So we must have $c^2 = 2$. □

This same proof also shows that $\sqrt[n]{x}$ exists for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$. We regularly distinguish between a real number that is also a rational number, and a real number that is not also a rational.

Definition 2.12

A real that is not rational is called an **irrational**.

Example 2.13

$\sqrt{2}$, $\sqrt{3}$, $\sqrt{6}$ and $\sqrt{15}$ are all irrational, as is $2+3\sqrt{5}$.

It's worth noting that the rationals are **dense** in the reals.

Theorem 2.14 ($\mathbb{Q}$ is dense in $\mathbb{R}$)

The rationals are **dense** in $\mathbb{R}$, that is, for $a, b \in \mathbb{R}$, there exists $q \in \mathbb{Q}$ such that $a \leq q \leq b$.

Proof. Without loss of generality, assume that $a, b \geq 0$. Then choose some $n \in \mathbb{N}$ such that $\frac{1}{n} < b - a$. Then among $\frac{0}{n}, \frac{1}{n}, \ldots$, there is a *final* one that is less than or equal to a, say $\frac{k}{n}$. So $a < \frac{q+1}{n} < b$. □

Slightly more interesting is the following result.

Theorem 2.15

The irrationals are dense, that is, for $a, b \in \mathbb{R}$ with $a < b$, there exists an irrational c such that $a < c < b$.

Proof. This follows from the rational case, as we can find a rational c between $\sqrt{2}a$ and $\sqrt{2}b$, then $c/\sqrt{2}$ is irrational, and lies between a and b. □

§2.2 Sequences

We are now going to discuss one of the major uses of the least upper bound statement – sequences. We are going to consider some questions like

$$\text{What should } 1 + \frac{1}{2} + \frac{1}{4} + \cdots = 2 \text{ mean?}$$

and

$$\text{What should } 0.3333\cdots = \frac{1}{3} \text{ mean?}$$

For the first question, presumably the first sequence should 'tend to' 2, and the second should 'tend to' 1/3. So what does it mean for a sequence $x_1, x_2, \ldots$ to 'tend to' some c? It should not mean that x_n is getting closer to c, for example, we would not want $1/2, 2/3, 3/4, \ldots$ to tend to 17. We also don't mean that x_n gets arbitrarily close to c, as for example, the sequence $1/2, 10, 2/3, 10, \ldots$ should not tend to 1. The correct way to think of it is that we will get arbitrarily close *and stay* within some value of c.

Definition 2.16

We say that a sequence of reals $x_1, x_2, \ldots$ **tends to** $c \in \mathbb{R}$ if $\forall \varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$,

$$|x_n - c| < \epsilon.$$

We will also say that c is the **limit** of $x_1, x_2, \ldots$.

Remark (Informal)**.** We can think of this say saying '$\forall \varepsilon > 0$, x_n eventually is within ε of c'.

There is quite a lot of subtlety to this definition, so we will spend some time considering it's consequences.

Notation. If x_n tends to c, we will write $x_n \to c$, or $x_n \to c$ as $n \to \infty$, or $\lim_{n\to\infty} x_n = c$.

Example 2.17

Consider the sequence

$$\frac{1}{2},\ \frac{1}{2}+\frac{1}{4},\ \frac{1}{2}+\frac{1}{4}+\frac{1}{8},\dots,$$

this is the sequence $x_1, x_2, \dots$ where $x_n = 1 - \frac{1}{2^n}$ (inductively). We claim $x_n \to 1$.

Proof. Given $\epsilon > 0$, choose $N \in \mathbb{N}$ with $N > \frac{1}{\epsilon}$. Then $\forall n \geq N$, $|x_n - 1| = \frac{1}{2^n} \leq \frac{1}{n} \leq \frac{1}{N} < \epsilon$. □

Example 2.18

The constant sequence $c, c, c, \dots$ (where $x_n = c$ for all n), then $x_n \to c$.

Proof. Given $\epsilon > 0$, we have $|x_n - c| < \epsilon$ for all n. □

Example 2.19

The sequence $-1, 1, -1, 1, \dots$ where $x_n = (-1)^n$ has no $c \in \mathbb{R}$ with $x_n \to c$.

Proof. Suppose that $x_n \to c$, then choose $\epsilon = 1$. So there exists $N \in \mathbb{N}$ such that for $n \geq N$, we have $|x_n - c| < 1$, in particular, $|1 - c|$ and $|(-1) - c| < 1$, so $|1 - (-1)| < 2$ by the triangle inequality, which is a contradiction. □

Example 2.20

The sequence $1, 0, 1/3, 0, 1/5, \dots$ with x_n given by

$$x_n = \begin{cases} 1 & \text{if } n \equiv 0 \pmod 5 \\ 5 & \text{if } n \equiv 1 \pmod 5 \\ \frac{1}{n} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases},$$

then $x_n \to 0$.

Proof. Given $\epsilon > 0$, choose $N \in \mathbb{N}$ with $\frac{1}{N} < \epsilon$. Then $\forall n \geq N$, $x_n = \frac{1}{n}$ or 0, so $|x_n - 0| \leq \frac{1}{n} \leq \frac{1}{N} < \epsilon$. □

Definition 2.21 (Convergent and Divergent Sequences)

If $x_c \to c$ for some c, we say that the sequence $x_1, x_2, x_3 \dots$ is **convergent**. If a sequence is not convergent, we say it is **divergent**.

We now show that if a sequence converges, then it converges to a unique value.

Proposition 2.22 (Limits Are Unique)
If a sequence $(x_n) \to c$, and $(x_n) \to d$, then $c = d$.

Proof. Suppose $c \neq d$, and choose $\epsilon = \frac{1}{2}|c-d|$. then $\exists N \in N$ with $|x_n - c| < \epsilon$ for all $n \geq N$, and some $M \in \mathbb{N}$ with $|x_n - d| < \epsilon$ for all $n \geq M$. But now for any $n \geq \max(M, N)$, we have $|x_n - c|$, $|x_n - d| < \epsilon$, so $|c - d| < 2\epsilon$, which is a contradiction. □

Some special notation can be used if a sequence is written in a particular form.

Definition 2.23 (Series)
A sequence given in the form x_1, $x_1 + x_2$, $x_1 + x_2 + x_3 + \dots$ is called a **series**, and can be written as

$$\sum_{n=1}^{\infty} x_n.$$

The kth term of this sequence is $\sum_{n=1}^{k} x_n$, the kth **partial sum** of this series.

If the series x_1, $x_1 + x_2$, $x_1 + x_2 + x_3$, $\dots$ is convergent say to c, then we will write

$$\sum_{n=1}^{\infty} x_n = c.$$

Remark (Warning)**.** We *cannot* write $\sum_{n=1}^{\infty} x_n$ unless we know that the limit of the partial sums exist.

Limits do generally behave as we would expect.

Proposition 2.24
If $x_n \to c$ and $y_n \to d$, then $x_n + y_n \to c + d$.

Proof. Given $\epsilon > 0$, then as $x_n \to c$, so there exists $N \in \mathbb{N}$ such that $|x_n - c| < \epsilon/2 \forall n \geq N$. Also, as $y_n \to d$, so there exists $M \in \mathbb{N}$ such that $|x_m - d| < \epsilon/2$ for $m \geq N$. Then we can take $m \geq \max(M, N)$, and $|(x_n + y_n) - (c + d)| \leq |x_n - c| + |y_n - d| < \epsilon/2 + \epsilon/2 = \epsilon$. □

In the proofs above, we have only been claiming what the limit is, and then proving that it converges to that. This isn't always necessary, and sometimes you an prove that a sequence converges without knowing the limit that it converges to.

Definition 2.25 (Increasing Sequence)
A sequence x_1, x_2, $\dots$ is **increasing** if $x_{n+1} \geq x_n$ for all n.

Theorem 2.26 (Monotone Convergence Theorem)

If an increasing sequence is bounded above, then it converges.[a]

[a] This also holds for a decreasing sequence, bounded below.

Proof. Let the sequence be x_1, x_2, ..., and let $c = \sup\{x_1, x_2, \dots\}$. I claim that this sequence converges to c. Then there exists N such that $x_N > c - \epsilon$ (as otherwise $c - \epsilon$ would be an upper bound, which is a contradiction). So for all $n \geq N$, we have $c - \epsilon < x_n \leq x_n < c$, hence $|x_n - c| < \epsilon$. □

To see how this can be used to prove that a series converges (or diverges) without finding the limit, consider the following two examples.

Proposition 2.27

(i) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

(ii) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.

Proof.

(i) We have $1/3 + 1/4 \geq 1/2$, and $1/5 + 1/6 + 1/7 + 1/8 \geq 1/2$. In general,

$$\frac{1}{2^n+1} + \frac{1}{2^n+2} + \cdots + \frac{1}{2^{n-1}} \geq \frac{2^n}{2^{n+1}} = \frac{1}{2}, \ \forall n.$$

Hence the partial sum of $\sum_{n=1}^{\infty} \frac{1}{n}$ is unbounded, and thus the series is not convergent.

(ii) We have $1/2^2 + 1/3^2 \leq 2/2^2 = 1/2$, and $1/4^2 + 1/5^2 + 1/6^2 + 1/7^2 \leq 4/4^2 = 1/4$, and in general,

$$\frac{1}{(2^n)^2} + \frac{1}{(2^n+1)^2} + \cdots + \frac{1}{(2^{n+1}-1)^2} \leq \frac{2^n}{(2^n)^2} = \frac{1}{2^n}, \ \forall n.$$

Hence the partial sums of $\sum_{n=1}^{\infty} \frac{1}{n^2}$ are bounded by $1 + 1/2 + 1/4 + \cdots = 2$, and thus this series converges by the monotone convergence theorem.

□

The value of the second series (the value that it converges to) is $\pi^2/6$. Also, the first series is quite important, and it has it's own name.

Definition 2.28 (Harmonic Series)

The series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

is the **harmonic series**.

The monotone convergence theorem also tells us something about decimal expansions.

Let the decimal expansion

$$0.a_1a_2a_3\dots$$

with $0 \leq a_n < 9$ for each n to be

$$\sum_{n=1}^{\infty} \frac{a_n}{10^n}$$

This must converge by the monotone convergence theorem, as all of the terms are at least zero and the partial sums are bounded by (for example 1). The converse of this is true, that given some $x \in \mathbb{R}$ with $0 < x < 1$, we can write it as $0.a_1a_2\dots$ for $a_i \in \{0,\dots,0\}$.

Begin by choosing the greatest $a_1 \in \{0,1,\dots,9\}$ such that $\frac{a_1}{10} \leq x$ thus $0 \leq x - \frac{a_1}{10} < \frac{1}{10}$. Then we can take the greatest $a_2 \in \{0,1,\dots,9\}$ such that $\frac{a_1}{10} + \frac{a_2}{100} \leq x$, so $0 \leq x - \frac{a_1}{10} - \frac{a_2}{100} < \frac{1}{100}$. We can continue this inductively to obtain $a_1, a_2, \dots \in \{0,1,\dots,0\}$ such that

$$0 \leq x - \sum_{n=1}^{k} \frac{a_n}{10^n} < \frac{1}{10^n},$$

for all k. Thus

$$\sum_{n=1}^{\infty} \frac{a_n}{10^n} = x.$$

Remark. We call a decimal expansion $0.a_1a_2\dots$ **recurrent** if there exists some $a_{n+k} = a_n$ for all $n \geq N$, for some N and k. For example, $0.3178426426426426\dots$ is recurrent. It can be checked that a number has a recurrent decimal expansion if and only if that number is rational.

Remark (Uniqueness of Expansions)**.** Decimal expansions need not be unique. For example, $0.370000\dots = 0.36999\dots$. However, this is the only case of non-uniqueness (where we have a 'terminating' decimal or a string of 9s).

Theorem 2.29 (e Exists)

We define

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \dots = \sum_{n=0}^{\infty} \frac{1}{n!},$$

and we claim that it exists.

Proof. The sequence of partial sums converge as all terms are positive, and the partial sums are bounded by $1 + 1 + \frac{1}{2} + \frac{1}{4} + \dots = 3$, since $1/n! \leq \frac{1}{2^{n-1}}$ (which is true by induction). □

We can also show that this number e is irrational.

Theorem 2.30 (Irrationality of e)

e is not rational.

Proof. Suppose that e was rational, then write $e = p/q$ where $p, q \in \mathbb{N}$, and also

suppose that $q > 1$. So, multiplying by $q!$ we have

$$\sum_{n=0}^{\infty} \frac{q!}{n!} = p(q-1)!,$$

which is an integer. This works, as if $x_n \to c$, then $kx_n \to kc$ for $k \in \mathbb{R}$. So

$$p(q-1)! = \underbrace{q! + \frac{q!}{1!} + \frac{q!}{2!} + \cdots + \frac{q!}{q!}}_{\in\mathbb{Z}} + \frac{q!}{(q+1)!} + \cdots$$

But in general,

$$\frac{q!}{(q+n)!} \leq \frac{1}{(q+1)^n}.$$

so we can bound part of the series with

$$\sum_{n=q+1}^{\infty} \frac{q!}{(q+n)!} \leq \frac{1}{q+1} + \frac{1}{(q+1)^2} + \cdots = \frac{1}{q} < 1,$$

and thus we have a contradiction, as it implies $p(q-1)!$ is not an integer. □

§2.3 Algebraic Numbers

Now we know that numbers can be rational or irrational, but we can also divide up numbers based on if they are the root of some integer polynomial.

Definition 2.31 (Algebraic)

We say that a real number x is **algebraic** if it is a root of a (non-zero) polynomial with integer coefficients. A number that is not algebraic is **trancendental**.

When we talk about algebraic numbers, we typically care about *irrational* algebraic numbers, because of the following propoisition.

Proposition 2.32

Every rational is algebraic.

Proof. The number p/q for $p, q \in \mathbb{Z}$ with $q \neq 0$ is a root of $qx - p = 0$. □

Let's have a look at some irrational algebraic numbers.

Example 2.33

$\sqrt{2}$ is algebraic as it satisfies $x^2 - 2 = 0$.

A natural question to ask is whether there are any non-algebraic, that is, transcendental numbers. We are going to construct one.

Theorem 2.34 (Example of a Trancendental Number)

The following number c is transcendental[a].

$$c = \sum_{n=1}^{\infty} \frac{1}{10^{n!}} = 0.11000100000000000000000010000\ldots$$

[a] This is going to be the hardest proof on the course.

To prove that this number is indeed trancendental, we are going to need recall some facts about polynomials.

Proposition 2.35 (Properties of Polynomials)

(i) For any polynomial p, there exists a constant k such that $|p(x)-p(y)| \leq k|x-y|$ for all $0 \leq x, y \leq 1$.

(ii) A non-zero polynomial of degree d has at most d roots.

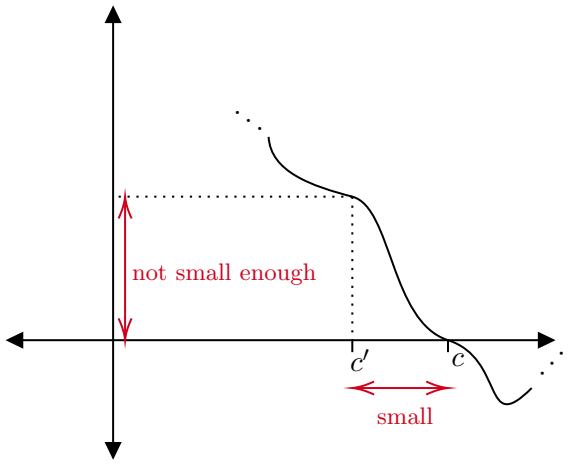
Proof.

(i) Say $p(x) = a_d x^d + \cdots + a_0$. Then $p(x) - p(y) = a_d(x^d - y^d) + \cdots + a_1(x_y) = (x-y)\left[a_d(x^{d-1} + x^{d-2}y + \cdots + y^{d-1}) + \cdots + a_1\right]$. so $|p(x) - p(y)| \leq |x - y|\left(|a_d|d + |a_{d-1}|d + \cdots + |a_1|d\right)$, because $0 \leq x, y \leq 1$.

(ii) Given a polynomial p of degree d, if p has no roots, then we are done. Otherwise, it has a root, say a, then $p(x) = (x-a)q(x)$ for some polynomial q of degree $d-1$. So each root of p is either a or a root of q, but q has at most $d-1$ roots by induction.

□

We are now ready to prove that our number c is transcendental. The idea of the proof is summarized in the image below



Proof of Theorem 2.34. Write $c_n = \sum_{k=0}^{n} \frac{1}{10^{k!}}$, so $c_n \to c$. Suppose the polynomial $p = a_d x^d + \cdots + a_0$ of degree d has c as a root. Then there exists k such that

$|p(x) - p(y)| \leq k|x - y|$, for all $0 \leq, x, y, \leq 1$. Now

$$|c - c_n| = \sum_{k=1}^{\infty} \frac{1}{10^{k!}} \leq \frac{2}{10^{(n+1)!}}.$$

Also, $c_n = \frac{1}{10^{n!}}$ for some $a \in \mathbb{Z}$, so $p(c_n) = \frac{b}{10^{dn!}}$, for some $b \in \mathbb{Z}$ (since $p(\frac{s}{t}) = \frac{q}{t^d}$ for some integer q whenever $s, t \in \mathbb{Z}$). But, for n large enough, c_n is *not* a root of p (as it has finitely many roots). So $|p(c_n)| \geq \frac{1}{10^{dn!}}$.

This implies that $|p(c_n) - p(c)| \geq \frac{1}{10^{dn!}}$. Thus $\frac{1}{10^{dn!}} \leq k\frac{2}{10^{(n+1)!}}$, which is a contradiction for n sufficiently large. □

Remark. The same proof shows that any real x such that $\forall n$, there exists $\frac{p}{q} \in \mathbb{Q}$ with $0 < |x - \frac{p}{q}| < \frac{1}{q^n}$ is transcendental. Informally, 'x has very good rational approximations'. Such x are often called **Liouville numbers**, and this says that 'every Liouville number' is transcendental.

§3 Sets and Functions

In a sense, this is a chapter of notation. We really just get some terminology, notation and straightforward concepts and make sure we agree with them. It's got some theorems, but there isn't any amazing theorems or proofs that are really hard, and the theorems are going to be rather unsurprising.

However, the concepts will be important, as is the way of thinking that we will develop.

§3.1 Sets

Let's begin with a definition, that will have a few caveats.

Definition 3.1 (Set)

A **set** is any[a] collection of (mathematical) objects.

[a] we are going to change this later

Example 3.2

$\mathbb{R}$, $\mathbb{N}$, $\mathbb{Q}$ are all sets.

We say that two sets with the same members are the same (that is, a set is determined by it's members). Symbolically, if A and B are sets, then if $\forall x$, $x \in A \iff x \in B$, then $A = B$. For example, the sets $\{1, 3, 7\}$ and $\{1, 7, 3\}$ equal, as are $\{3, 4, 4, 6\}$ and $\{3, 4, 6\}$.

So that's what a set is, but what's more important is how we can build sets. We are going to enter a series of sections which informally can be grouped as 'creating new sets from old sets'.

§3.1.1 Subsets

When we have a set, we are able to choose certain elements of that set with some condition, to create another set.

Axiom 3.3 (Subset Selection)
Given a set A and a property $p(x)$, we can form $\{x \in A : p(x)\}$, the set of all members of A with property p.

Definition 3.4 (Subsets)
We say that B is a **subset** of A if $\forall x$ where $x \in B$ implies $x \in A$. This is written $B \subseteq A$.

Note that this implies that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

§3.1.2 Unions and Intersections

We have two main ways of joining two sets.

Definition 3.5 (Union and Intersection)
Given sets A and B, we can form their **union** $A \cup B = \{x : x \in A \text{ or } x \in B\}$, and their **intersection** $A \cap B = \{x : x \in A \text{ and } x \in B\}$.

Definition 3.6 (Disjoint)
If the intersection of two sets A and B the empty set, $A \cap B = \emptyset$, then we say that A and B are **disjoint**.

Note that union and intersection are commutative and associative. Also, union is distributive over intersection, that is, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. This can be proven by checking definitions.

We can also take a bigger union.

Example 3.7 (Infinite Union)
If $A_n = \{n^2, n^3\}$ for each $n \in \mathbb{N}$, then $A_1 \cup A_2 \cup \cdots$ is going to be the set of all squares or cubes. We can write this as

$$A_1 \cup A_2 \cup \cdots = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n \in \mathbb{Z}} A_n.$$

Of course, unlike our previous use of similar notation, this is not a limiting process but an *indexing process*.

In general, given a set I and sets A_i where $i \in I$, then we can form

$$\bigcup_{i \in I} A_i = \{x : x \in A_i \text{ for some } i \in I\},$$

and similarity,

$$\bigcap_{i\in I} A_i = \{x : x \in A_i \ \forall i \in I\}.$$

However, this section set can only be defined for a *non-empty index set*, for reasons that we will see later.

§3.1.3 Ordered Pairs

Definition 3.8 (Ordered Pair)

For any a, b, we can form the **ordered pair** (a, b), such that $(a, b) = (c, d) \iff a = c$ and $b = d$.

Using the idea of an ordered pair, we can form the *product* of sets

Definition 3.9 (Cartesian Product)

For sets A and B, their **cartesian product** $A \times B = \{(a, b) : a \in A, b \in B\}$.

Remark. If we wished, we could define ordered pairs using sets, $(a, b) = \{\{a\}, \{a, b\}\}$, and then we can check that the equality case is the same.

§3.1.4 Power Set

Definition 3.10

For any set X, we can form the **power set** $\mathcal{P}(X)$ consisting of all subsets of X.

$$\mathcal{P}(X) = \{Y : Y \subseteq X\}.$$

Example 3.11

If $X = \{1, 2\}$, then $\mathcal{P}(X) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

§3.1.5 Defining Sets

For a set A, we can form $\{x \in A : p(x)\}$. However, we *cannot* form $\{x, p(x)\}$. Indeed, if we could, then consider the set

$$X = \{x : x \notin X\}.$$

So is $X \in X$? If yes, then $X \notin X$, but this is a contradiction. But if no, then $x \notin X$, but that would imply that it is in X! So this doesn't make any sense. This is Russell's paradox.

Similarity, there is no 'universal' set Y, such that $\forall x$, $x \in Y$. This is because by subset selection, we could create the group $X = \{x : x \notin X\}$, which would be a contradiction.

Thus to guarantee that a given set exists, we must obtain it in some way (for example with the rules given above) from known sets.

§3.2 Finite Sets

Now let's have a look at the sizes of sets. We are going to write $\mathbb{N}_0 = \mathbb{N} \cup \{0\} = \{0, 1, \dots, \}$.

Definition 3.12 (Size)

For $n \in \mathbb{N}_0$, a set A has **size** n if it can be written $A = \{a_1, a_2, \dots, a_n\}$ with the a_i all distinct.

Definition 3.13 (Finite)

We will say that a set A is **finite** if there exists some n such that the size of A is n. Otherwise, we say that the size of the set is **infinite**.

Proposition 3.14 (Counting Subsets)

A set of size n has exactly 2^n subsets.

Proof. Assume that our set is $\{1, \dots, n\}$. To pecify a subset S, we must say if $1 \in S$, or $1 \notin S$, then if $2 \in S$ or if $2 \notin S$, and so on. So the number of independent choices needed to form S is $\underbrace{2 \times 2 \times \cdots \times 2}_{n \text{ times}} = 2^n$.

Alternate Proof. We do induction on n. This is true for $n = 0$, and given some $n > 0$, for $T \subseteq \{1, 2, \dots, n-1\}$, then we can count how many $S \subseteq \{1, 2, \dots, n\}$ have $S \cap \{1, \dots, n-1\} = T$. There is exactly 2, namely T and $T \cup \{n\}$. Hence the number of subsets of $\{1, \dots, n\}$ is twice the number of subsets of $\{1, \dots, n-1\}$. Then we are done by induction. □

Remark. We could view the second proof as a 'more formal version' of the first proof.

As a bit of notation, if A has size n, then we write $|A| = n$. Using this notation, our counting subsets proposition says that

$$|A| = n \implies |\mathcal{P}(A)| = 2^n.$$

§3.3 Finite Sets and their Sizes

Given a set A, we are now going to think of how many subsets of A there is of size k.

Definition 3.15 (Binomial Coefficients)

For $n \in \mathbb{N}_0$ and $0 \leq k \leq n$, we write $\binom{n}{k}$ for the number of subsets of an n-element set that are of size k:

$$\binom{n}{k} = |\{S \subset \{1, 2, \dots, n\} : |S| = k\}|.$$

Example 3.16

The 2-element sets in a 4-element set, say $\{1, 2, 3, 4\}$ are $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$,

$\{2,4\}$, $\{3,4\}$. So $\binom{4}{2} = 6$.

We will also write $\binom{n}{0} = 1$, $\binom{n}{n} = 1$, and note that $\binom{n}{1} = n$.

With this definition of binomial coefficients, we can deduce some nice properties.

Proposition 3.17 (Properties of Binomial Coefficients)

(i) $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$.

(ii) $\binom{n}{k} = \binom{n}{n-k}$, for all $n \in \mathbb{N}_0$ with $0 \leq k \leq n$.

(iii) $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$, for all $n \geq 1$, and $0 < k < n$.[a]

[a] This property can be used to compute small binomial coeficients with 'Pascal's triangle'.

Proof. (i) Each side counts the number of subsets of $\{1, \ldots, n\}$.

(ii) Picking a k element subset from an n element set, is the same as choosing $n-k$ elements to not pick.

(iii) The number of subsets of $\{1, 2, \ldots, n\}$ with k elements is made up of the k-element subsets of $\{1, 2, \ldots, n-1\}$ and also the $k-1$-element subsets of $\{1, 2, \ldots n-1\}$, along with n.

□

Sometimes we will have to compute binomial coefficients, and indeed there is a formula for binomial coefficients. However, what's more important is where the formula comes from.

Proposition 3.18 (Binomial Coefficient Formula)

We have for integers n and k,

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}.$$

Proof. The number of ways to have a k-set is $n(n-1)\cdots(n-k+1)$, as there is n ways to name one element, $n-1$ ways to name a different element, and so on. This does grossly overcount though. The number of times a given k-set was named is $k(k-1)\cdots 1$. Hence the number of k-sets is

$$\frac{n(n-1)\cdots(n-k+1)}{k!}.$$

□

Example 3.19

We have

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

Remark. You could, if you wanted, write

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

but this is rather odd, as we have just multiplied my loads of terms that will cancel.

The formula we derived is also good for approximating the size of binomial coefficients. For example,

$$\binom{n}{3} = \frac{n(n-1)(n-2)}{6} \approx \frac{n^3}{6}$$

for large n.

An important application of binomial coefficients (and why we call them binomial coefficients) is the binomial theorem.

Theorem 3.20 (Binomial Theorem)

For a positive integer n,

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{0}b^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}.$$

Proof. When we expand $(a+b)^n = (a+b)(a+b)\cdots(a+b)$, we obtain terms of the form $a^k b^{n-k}$, where $0 \leq k \leq n$. The number of such terms $a^k b^{n-k}$ is the number of ways of choosing k brackets with an a, out of n. That is, the number of times the term $a^k b^{n-k}$ occurs is $\binom{n}{k}$. So

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

□

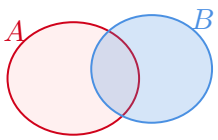
Example 3.21

We can expand $(1+x)^n$ using the binomial theorem as

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2}x^2 + \cdots + nx^{n-1} + x^n,$$

so for small x, a good approximation for $(1+x)^n$ is $1+nx$. For example, $1.00001^8 \approx 1.0008$. A better approximation could be obtained by taking an extra term.

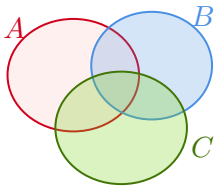
So this describes sizes of a single finite set, but what about the sizes of unions and intersections of finite sets? For example, how big is $|A \cup B|$?



We could try and write down $|A \cup B| = |A| + |B|$, but then we would count every element that's in both sets twice. That is, we have double counted $A \cap B$. So we really have

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

For three sets, we could try and write down $|A \cup B \cup C| = |A| + |B| + |C|$, but then we have over counted many elements.



Specifically, we would have overcounted every element that's in A and B, every element in A and C, and every element in B and C. So we need to subtract and get $|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C|$. But we have then undercounted elements in all 3 sets! We really have

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

In general, we have the following theorem.

Theorem 3.22 (Inclusion-Exclusion Theorem)

Let $S_1, \dots, S_n$ be finite sets. Then

$$|S_1 \cup \cdots \cup S_n| = \sum_{|A|=1} |S_A| - \sum_{|A|=2} |S_A| + \sum_{|A|=3} |S_A| - \cdots + (-1)^{n+1} \sum_{|A|=n} |S_A|,$$

where S_A is $\bigcap_{i \in A} S_i$, and $\sum_{|A|=r}$ is taken over all $A \cap \{1, \dots, n\}$ of size k.

Proof. Let x be some element in one of the sets. We wish to show that it has been counted exactly once. Suppose that contained in exactly k of the sets. Then the number of A such that $|A| = r$ and $x \in S_A$ is $\binom{k}{r}$, as we must choose r of the sets of which x is an element. So the number of times x is counted on the right hand side is

$$k - \binom{k}{2} + \binom{k}{3} - \cdots + (-1)^{k+1} \binom{n}{k},$$

for $r \leq k$. But $(1 + (-1))^k$ is exactly this by the binomial theorem, so the number of times x is counted is exactly $1 - (1 + (-1))^k = 1$. □

§3.4 Functions

For some sets A and B, we would intuitively think of a function as some rule that each $a \in A$ is assigned to a point $f(a) \in B$. We can define this notion more precisely.

Definition 3.23 (Function)

A **function** from A to B is a set $f \subset A \times B$ such that for each $x \in A$ there exists a unique $y \in B$ with $(x, y) \in f$.

If $(x, y) \in f$, we write $f(x) = y$.

Example 3.24 (Examples of Functions)

$f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ is a function.

$f : \mathbb{R} \to \mathbb{R}$ where $f(x) = 1$ if x is rational, and $f(x) = 0$ otherwise is a function.

The second example given is slightly subtle. Using the intuitive notion of a function, one might wonder whether the 'rule' of $f(x) = 1$ if x is rational and $f(x) = 0$ otherwise is valid. For example, we know that $f(2) = 1$ and $f(e) = 0$, but what about $f(e + \pi)$? Well the definition of a function tells us that we only need to make sure that this is a 'well defined' thing, and indeed as every real number is either rational or irrational, this is a valid function. Let's state some not-so valid functions.

Example 3.25 (Non-Examples of Functions)

$f : \mathbb{R} \to \mathbb{R}$ with $f(x) = \frac{1}{x}$ is *not* a function because it doesn't specify the value of $f(0)$.

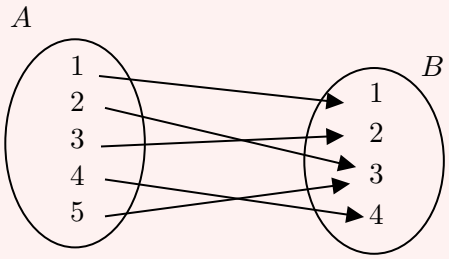
$f : \mathbb{R} \to \mathbb{R}$ with $f(x) = \pm\sqrt{|x|}$ is *not* a function because it assigns multiple values to every value of $|x|$.

Let's continue looking at examples of valid functions.

Example 3.26

The following are all functions.

(i) Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 2, 4, 5\}$. Then $f : A \to B$ given by:



is a function.

(ii) f from $\{1, 2, 3\}$ to $\{1, 2, 3\}$ with $1 \mapsto 1$, $2 \mapsto 3$, and $3 \mapsto 2$ is a function.

We frequently deal with functions that have certain interesting properties, which we will describe in the following definitions.

Definition 3.27 (Injectivity)

A function $f : A \to B$ is **injective** if for any $a, a' \in A$, then $a \neq a' \implies f(a) \neq f(a')$.

Definition 3.28 (Surjectivity)

A function $f : A \to B$ is **surjective** if for all $b \in B$, there exists $a \in A$ such that $f(a) = b$.

Definition 3.29 (Bijectivity)

A function $f : A \to B$ is **bijective** if it is injective and surjective.

We also name certain parts of a function.

Definition 3.30

For $f : A \to B$, we say A is the **domain**, and B is the **range**. We also say the **image** of f is the set $\{f(a) \mid a \in A\}$.

Also, when specifying a function, you must say what the domain and range are. For example, is the function $f(x) = x^2$ injective? The answer is that this is meaningless, as we haven't defined what the domain or range are. If $f : \mathbb{N} \to \mathbb{N}$, then it is injective, but it is not injective if $f : \mathbb{R} \to \mathbb{R}$.

When we are dealing with functions on finite sets, some obvious facts are true.

Proposition 3.31

For A and B finite, the following are true:

(i) There is no surjection from $A \to B$ if $|B| > |A|$.

(ii) There is no injection from $A \to B$ if $|A| > |B|$.

(iii) For $f : A \to A$, if f is injective then f is also surjective (and vice versa).

(iv) There is no bijection from A to any proper subset of A.

Proof Sketch. Check definitions. □

However, this *does not* hold when A and/or B is not finite. For example, consider the function $f_0 : \mathbb{N} \to \mathbb{N}$ with $f(x) = x+1$. Then f_0 is injective but not surjective. Similarity, define $f_0 : \mathbb{N} \to \mathbb{N}$, so that $f(1) = 1$ and $f(x) = x - 1$ if $x > 1$. Then this function is surjective but not injective. Lastly, define $g : \mathbb{N} \to \mathbb{N}\backslash\{0\}$ with $x \mapsto x + 1$. Then this is a bijection between a proper subset of $\mathbb{N}$ to $\mathbb{N}$.

Let's look at more functions!

Example 3.32

The following are functions.

(i) For any set X, the **identity function** $1_X : X \to X$ such that $1_X(x) = x$ for all $x \in X$.

(ii) For any set X and $A \subset X$ the **indicator function** or **characteristic function** $\chi_A : X \to \{0, 1\}$ with $\chi_A(x) = 1$ if $x \in A$ and 0 otherwise.

This function is important because it carries all of the information about what is in the set A.

(iii) A sequence $x_1, x_2, \ldots$ of reals is a function from $\mathbb{N} \to \mathbb{R}$ with $n \to x_n$.

(iv) The operation $+$ on $\mathbb{N}$ is a function from $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$.

(v) A set X has size $|n|$ if there is a bijective function from $\{1, 2, \dots, n\} \to X$.

One thing that we regularly do with functions is *compose* them, that is, apply one after another.

Definition 3.33

Given $f : A \to B$ and $g : B \to C$, the **composition** $g \circ f : A \to C$ is defined by

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

Example 3.34

If $f : \mathbb{R} \to \mathbb{R}$ is defined as $f(x) = 2x$, and $g : \mathbb{R} \to R$ is defined as $g(x) = x + 1$, then

$$(g \circ f)(x) = 2x + 1, \quad \text{and} \quad (f \circ g)(x) = 2x + 2.$$

In general (and in the example above), composition isn't commutative. However, composition is associative.

Composition is useful in talking about the idea of inverting functions.

Definition 3.35

We say a function $f : A \to B$ is **invertible** if there exists a function $g : B \to A$ such that $g \circ f = 1_A$ and $f \circ g = 1_B$.

Remark. It is not sufficient to show just that $g \circ f = 1_A$ or $f \circ g = 1_B$. Both must be checked.

Proposition 3.36

A function $f : A \to B$ is invertible if and only if it is bijective.

Proof. Let's say that g is inverse to f. Then for any $b \in B$, we have $f(g(b)) = b$, so f is surjective. Also, $f(a) = f(a')$ implies $g(f(a)) = g(f(a'))$, and $a = a'$, so f is injective and thus bijective.

Now let's suppose that f is bijective, and let $g(b)$ be the unique $a \in A$ with $f(a) = b$. Then g is the inverse of f. □

§3.5 Equivalence Relations

The last item we will discuss in this chapter is equivalence relations. These appear throughout mathematics, and the induce phenomena that quite commonly appear when we define something.

First, we must define what a relation is.

Definition 3.37 (Relation)

A **relation** on a set X is a subset R of $X \times X$.

We normally write aRb for $(a, b) \in R$.

Example 3.38

The following are all relations.

(i) On $\mathbb{N}$, aRb if $a \cong b \pmod 5$. For example, $2R12$ but not $2R11$.

(ii) On $\mathbb{N}$, aRb if $a \mid b$.

(iii) On $\mathbb{N}$, aRb if $a = b \pm 1$.

Some properties a relation might satsify are the following.

Definition 3.39 (Reflexive)

A relation R is **reflexive** if for all $x \in X$, xRx.

Definition 3.40 (Symmetric)

A relation R is **symmetric** if for all $x, y \in X$, then $xRy \implies yRx$.

Definition 3.41 (Transitive)

A relation R is **transitive** if for all $x, y, z \in X$, then xRy and yRz implies xRz.

When all of these properties are satisfied, we have an equivalence relation.

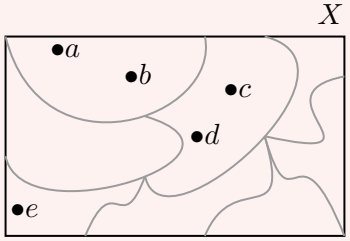
Definition 3.42 (Equivalence Relation)

An equivalence relation is a relation that is reflexive, symmetric and transitive.

To get a feel for what equivalence relations actually are, consider the following slightly natural yet slightly contrived example.

Example 3.43

Imagine we had a map of the world X, divided up into countries.



There are people in the world, and let's suppose each person is in just one country and there are no border disputes.

The relation 'is in the same country as' is an equivalence relation.

This relation is reflexive (you're in the same country as yourself), symmetric (if a is in the same country as b, then surely b is in the same country as a), and it's also

transitive (if x and y are in the same country, and y and z are in the same country, then x and y are in the same country).

We can phrase this example in a more sensible way.

Example 3.44

Let X be a set, and let $\{C_i \mid i \in I\}$ be a **partition** of X, such that each C_i is non-empty, they are disjoint, and $\bigcup_{i\in I} C_i = X$.

Then aRb if there exists i such that $a, b \in C_i$ is an equivalence relation on X.

It actually turns out that all examples look like this! We will need a definition first.

Definition 3.45 (Equivalence Class)

For an equivalence relation R on a set X, and $x \in X$, the **equivalence class**, $[x]$, is the set $\{y \in X \mid yRx\}$. That is, the set of everything in X that is related to x.

Proposition 3.46 (Equivalence Classes Partition a Set)

Let R be an equivalence relation on a set X. Then the equivalence classes of R partition X.

Proof. Each equivalence class $[x]$ is non-empty (as $x \in [x]$), and $\cup_{x\in X}[x] = X$, as $x \in [x]$ for all $x \in X$. Thus it suffices to show that all equivalence classes are either disjoint or equal.

Given x, y with $[x] \cap [y] \neq \emptyset$. We wish to show $[x] = [y]$. Take $z \in [x] \cap [y]$. Then zRx and zRy so xRy. Thus $[x] = [y]$. □

Definition 3.47 (Quotient and Quotient Map)

Given an equivalence relation R on a set X, the **quotient** of x by R is $X/R = \{[x] \mid x \in X\}$.

The map $q : X \to X/R$ is the **quotient map**, $x \mapsto [x]$. It is also known as the **projection map**.

§4 Countability

In this chapter we will discuss something we completely ignored in the previous discussion of sets – the sizes of infinite sets. Of course, we are going to have to introduce some ways of talking about the sizes of infinite sets...

Definition 4.1 (Countability)

We say that a set X is **countable** if X is finite or if there is a bijection with $\mathbb{N}$. Equivalently, X is countable if and only if we can list X as $a_1, a_2, \ldots$ which might terminate.

§4.1 Countable Sets

Let's use this definition to look at some countable sets.

Example 4.2 (Trivially Countable Sets)

Any finite set is countable, and $\mathbb{N}$ is countable, by definition.

Example 4.3 ($\mathbb{Z}$ is Countable)

$\mathbb{Z}$ is countable, as we can list it as $0, 1, -1, 2, -2, \dots$.

It's sometimes easier to work with the following characterization of countability.

Proposition 4.4

A set X is countable if and only if there exists an injection $f : X \to \mathbb{N}$.

Proof. If X is finite or bijects with $\mathbb{N}$, then of course X injects in $\mathbb{N}$.

Now assume that X is not finite (if it was finite, then it's already countable), and has an injection $f : X \to \mathbb{N}$. Then we have a bijection from X with its image $f(X) = \{f(x) \mid x \in X\}$.

Set a_1 to be $\min f(X)$, a_2 to be $\min f(X)\backslash\{a_1\}$, and so on. Then $f(X) = \{a_1, a_2, \dots\}$ – each $a \in X$ is a_n for some n, because $a = a_n$ for some $n \leq a$. So $f(x)$ is countable. □

Remark (A Warning)**.** In $\mathbb{R}$, let $X = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\} \cup \{1\}$. Then X is countable, as we can list it as $1, 1/2, 2/3, \dots$. But if we counted by 'least element', 1 would not be on the list. So the counting method used in the proof above does not always work.

Let's look at some more countable sets.

Theorem 4.5

$\mathbb{N} \times \mathbb{N}$ is countable.

Proof. Define $a_1 = (1, 1)$, and inductively letting $a_{n-1} = (p, q)$,

$$a_n = \begin{cases} (p-1, q+1) & \text{if } p \neq 1, \\ (1, p+1) & \text{if } p = 1. \end{cases}$$

Then a_n hits every point $(x, y) \in \mathbb{N} \times \mathbb{N}$, by induction on $x + y$, so we have listed $\mathbb{N} \times \mathbb{N}$, and thus it is countable. □

Alternate Proof. We will look for an injection $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Indeed, $f(x, y) = 2^x 3^y$ works, and thus $\mathbb{N} \times \mathbb{N}$ is countable. □

The same proof shows the following theorem.

Theorem 4.6

Let $A_1, A_2, \ldots$ be countable sets. Then $A_1 \cup A_2 \cup \ldots$ is countable.

Proof. For each i, we have A_i countable, so we can list A_i as $a_{i,1}, a_{i,2}, \ldots$, which might terminate. Then define $f : \bigcup_{n\in\mathbb{N}} A_n \to \mathbb{N}$, so that $x \to 2^i 3^j$, where $x = a_{ij}$ for the least such i. This is an injection. □

This is a quite powerful way of showing that sets are countable, as we shall see.

Example 4.7

$\mathbb{Q}$ is countable, as $\mathbb{Q} = \mathbb{Z} \cup \frac{1}{2}\mathbb{Z} \cup \frac{1}{3}\mathbb{Z} \cup \ldots$, where $\frac{1}{k}\mathbb{Z} = \{\frac{x}{k} \mid x \in \mathbb{Z}\}$. So Q is countable.

Example 4.8

The set $\mathbb{A}$ of all algebraic numbers is countable.

Indeed, it's enough to show that the set of all integer polynomials is countable, as each has finitely many roots, so $\mathbb{A}$ would be a countable union of finite sets.

So, to show that the collection of integer polynomials is countable, it's enough to show that for each d, the set of all integer polynomials of degree d is countable. But this set injects into $\mathbb{Z}^{d+1}$ with it's coefficients, and $\mathbb{Z}^n$ is countable for any n, by our previous theorem.

§4.2 Uncountable Sets

All of the examples we have given so far have been countable sets, but don't start thinking that all sets are countable!

Definition 4.9 (Uncountability)

A set X is **uncountable** if it is not countable.

Theorem 4.10

$\mathbb{R}$ is uncountable.

Proof. It suffices to show that $(0, 1)$ is uncountable. Given a sequence $r_1, r_2, \ldots$ in $(0, 1)$, we wish to find some $s \in (0, 1)$ that's not in this list.

For each r_n, we have a decimal expansion $r_n = 0.r_{n,1}r_{n,2}\ldots$. Define $s = 0.s_1s_2\ldots$ by

$$s_i = \begin{cases} 5 & \text{if } r_{i,i} \neq 5, \\ 6 & \text{if } r_{i,i} = 5. \end{cases}.$$

Then s cannot be on our list (as $s \neq r_n$ as they differ at decimal digit n). □

The proof is known as a **diagonal argument**, or **cantor's diagonal argument**.

Interestingly, now that we know $\mathbb{R}$ is uncountable, but $\mathbb{A}$ is countable, there must be a transcendental number! Indeed, 'most' numbers are transcendental, in that $\mathbb{R}\backslash\mathbb{A}$ is uncountable.

Let's look at another uncountable set, which we will show is uncountable using a diagonal argument.

Theorem 4.11

The power set of $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$ is uncountable.

Proof. Suppose $\mathcal{P}(\mathbb{N})$ is listed as $S_1, S_2, \ldots$. We wish to find some $S \subset \mathbb{N}$ that's not on this list. For every $i \in \mathbb{N}$, we take $i \in S$ iff $i \notin S_i$. Then S cannot be on this list, since $S \neq S_n$ for all n, as they differ at element n. Hence $\mathcal{P}(\mathbb{N})$ is uncountable. □

In fact, our result about the power set of $\mathbb{N}$ shows the following theorem.

Theorem 4.12

For any set X, there is no bijection from X to $\mathcal{P}(X)$.

Proof. Given any function $f : X \to \mathcal{P}(X)$, we will show f is not surjective.

Let $S = \{x \in X \mid x \notin f(x)\}$. Then S does not belong to the image of f, since for every x, $S \neq f(x)$, as S and $f(x)$ differ at element x. □

Remark. This proof is vaguely reminiscent of Russell's paradox. In fact, this theorem proves that there is no universal set, just as Russell's paradox did. This is because otherwise, we would have $\mathcal{P}(V) \subset V$, where certain that V would surject to $\mathcal{P}(V)$.

Now we're going to look at an example of how we can think about countability.

Example 4.13

Let A_i, $i \in I$ be a family of open intervals that are pairwise disjoint. Must the family be countable?

The answer is yes. Each A_i contains a rational, and $\mathbb{Q}$ is countable. So this family is countable.

Another proof is that we can note the set $\{i \in I \mid A_i \text{has length} \geq \frac{1}{k}\}$ is countable (it injects into $\mathbb{Z}$). So the family be countable.

In general, to show a set X is uncountable, we had two approaches.

1. Run the diagonal argument on X.
2. Inject your favorite uncountable set into X.

To show a set X is countable, we could do something like

1. List it (which can be ugly)
2. Inject it into $\mathbb{N}$
3. Use 'countable union of countable sets is countable'. This is usually the best one.
4. If in or near $\mathbb{R}$, sometimes using the countability of $\mathbb{Q}$ helps.

§4.3 Bijections

Intuitively, we can think of 'A bijects with B' as saying that A and B are the same size. Similarly, we can think of 'A injects into B' as sort of saying that A is at most as large as B. Similarly we can think of 'A surjects to B' as saying A is at least as large as B.

For these to make sense, we want for A and B nonempty that there exists an injection $f : A \to B$ if and only if there exists a surjection $g : B \to A$.

This is easy to show. Fix some $a_0 \in A$, and define $g : B \to A$ so that b maps to the unique $a \in A$ with $f(a) = b$, and a_0 if that does not exist. Also, if g is surjective, then for each $a \in A$ we have $a' \in B$ with $g(a') = a$. Then let $f(a) = a'$. Then f is injective.

Theorem 4.14 (Schröder-Bernstein Theorem)

If $f : A \to B$ and $g : B \to A$ are injections, then there exists a bijection $h : A \to B$.

Proof. For $a \in A$, we write $g^{-1}(a)$ for the $b \in B$ such that $g(b) = a$, if it exists. We write $f^{-1}(b)$ for $b \in B$ similarly.

The *ancestor sequence* of $a \in A$ is $g^{-1}(a)$, $f^{-1}(g^{-1}(a)), \dots$, which may terminate.

Let A_0 be the set of $a \in A$ such that the ancestor sequence stops in A, and A_1 the set of $a \in A$ such that the sequence stops in B. Lastly define A_∞ for the $a \in A$ whose ancestor sequence does not stop. Define B_0, B_1 and B_∞ similarity.

Then f bijects A_0 with B_1 (noting that every $b \in B_1$ is $f(a)$ for some $a \in A_0$). Similarity g bijects A_1 with B_0. Finally, f or g bijects A_∞ with B_∞.

Thus the function $h : A \to B$ with

$$h(a) = \begin{cases} f(a) & \text{if } a \in A_0, \\ g^{-1}(a) & \text{if } a \in A_1, \\ f(a) & \text{if } a \in A_\infty \end{cases}$$

is a bijection. □

Example 4.15

Do $[0, 1]$ and $[0, 1] \cup [2, 3]$ biject?

By Schröder-Bernstein this is trivial. Consider the identity function $f : [0, 1] \to [0, 1] \cup [2, 3]$, and the function $g : [0, 1] \cup [2, 3] \to [0, 1]$ with $g(x) = x/3$ are both injections, so there is a bijection between these sets.

In the same vein, a nice result to have would be 'for any sets A and B, either A injects into B or B injects into A'. This is true, but is not proved in this course.

§4.4 Sizes of Sets

Going back to the sizes of sets, we have seen that we can have finite sets, countable sets and uncountable sets. We also know that in the realm of uncountable sets, taking power sets makes you bigger. So the sequence $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$, $\mathcal{P}(\mathcal{P}(\mathbb{N}))$, ... may make us wonder does every set X inject into one of these?

This is not true. For example, take the set $X = \mathbb{N} \cup \mathcal{P}(\mathbb{N}) \cup \mathcal{P}(\mathcal{P}(\mathbb{N})) \cup \dots$. This isn't the biggest set either, as we can take $X' = X \cup \mathcal{P}(X) \cup \dots$. And this isn't biggest either, as we could just do the same thing all over again, and again, and again.