DOWNING COLLEGE CAMBRIDGE
Summer School

Mathematics A

Keenan J. A. Down

July 2023

Hello and welcome to the **Downing College Super Curricular Programme**, and a special welcome to **Mathematics A**! Over the next 8 lectures we're going to explore various areas of mathematics and touch briefly on many topics you might see if you study mathematics or a mathematical subject at university, with the goal of helping you to understand what mathematics in higher education is really like.

Mathematics at university tends to be structured quite differently to mathematics at school, with a lot more of an emphasis on sharing your results, conveying a good argument, or making a point well. As part of this series, we'll try and explore many of the common themes and techniques that might be useful when transitioning to university mathematics.

At the end of the two week programme (next Friday!) there's going to be a **presentation** on one of a number of problems. I'll provide a list of problems and topics separately, and you will create a presentation to present your solution at the end of the course. The presentations are going to be in small groups of 1-3, depending on the class size! I will be judging your presentations at the end of the course.

We'll aim to structure the lectures like this:

- Sets, Functions and Relations
- Proof, Logic and Truth
- Elementary Number Theory
- Abstract Algebra
- Analysis
- Linear Algebra
- Calculus in Context
- *Mystery lecture!*

In addition to these lectures (Tues-Friday week 1, Mon-Thur week 2 at 9:00am), you'll also get some supervisions afterwards, where you should expect to have a discussion and a hands-on session exploring some mathematical questions in detail.

Let's get started on our mathematical journey!

1 Sets, Functions and Relations

1.1 Sets

Definition 1.
A **set** is a collection of objects or **elements**.

A set can consist of more or less anything, so long as it is well defined (this is harder to be sure of than you might think, but we'll come back to this later!).

Example 2.
We often express sets using curly brackets {}, possibly listing individual elements one at a time. For example, the set of numbers between 1 and 10 can be written

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}.$$

Example 3.
There are some classic examples of standard sets in mathematics, which you might have come across. For example, the **natural numbers**:

$$\mathbb{N} = \{1, 2, 3, 4, 5, \ldots\}$$

The **integers**:

$$\mathbb{Z} = \{\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$$

The **rational numbers**:

$$\mathbb{Q} = \left\{ \frac{a}{b} : a, b \in \mathbb{Z} \right\}$$

The **real numbers**: $\mathbb{R}$ (we will attempt to construct these later in the course!). The **complex numbers**:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}, i^2 = -1\}$$

And many, many more!

Sets form the basis for modern mathematics, and most mathematicians agree that all of the language of mathematics can be derived from the language of set theory! We'll talk about this a bit more in lecture 8.

Notation 4.
When we want to make it clear that something is an **element** of a set, we use the symbol $\in$. For example, 2 is an integer, so we write $2 \in \mathbb{Z}$. Similarly we can write $\sqrt{2} \in \mathbb{R}$, and since π is not rational, we could also write $\pi \notin \mathbb{Q}$.

Notation 5.
We can construct a new set from an old set using **set-builder notation**. The general format is this:

$$\{\text{elements in the new set} : \text{conditions on those elements}\}.$$

To get an idea how this works, we could write the square numbers as follows:

$$\{x^2 : x \in \mathbb{N}\} \quad = \quad \{1, 4, 9, 16, 25, 36, \ldots\}$$

This is read as *'x squared such that x is a natural number.'*

For another example, suppose we wanted to write the even integers:

$$\{2x : x \in \mathbb{Z}\} \quad = \quad \{\ldots, -6, -4, -2, 0, 2, 4, 6, \ldots\}$$

This could be read as '$2x$ *such that* x *is an integer*' or '$2x$ *such that* x *is in* $\mathbb{Z}$.'

Definition 6.
We define the **union**, $\cup$ of two sets A and B to be the set which contains elements from either A, B or both. That is:

$$A \cup B = \{\, x : x \in A \textbf{ or } x \in B\}.$$

We define the **intersection**, $\cap$ of two sets A and B to be the set which contains elements which are in both A and in B:

$$A \cap B = \{\, x : x \in A \textbf{ and } x \in B\}.$$

Example 7.
Let $A = \{1, 2, 3\}$ and $B = \{3, 4, 5\}$. Then:

$$A \cup B = \{1, 2, 3, 4, 5\}$$
$$A \cap B = \{3\}.$$

Definition 8.
We define the **set difference** $A \setminus B$ or $A - B$ to be the set of elements in A which are not in B:

$$A \setminus B = \{a \in A : a \notin B\}.$$

It is often read *'A minus B'*, *'A set minus B'* or *'A less B.'*

Example 9.
If $A = \{\text{cat, dog, horse, apple}\}$ and $B = \{\text{apple, orange, banana, pear}\}$, then

$$A \setminus B = \{\text{cat, dog horse}\}.$$

As you can probably tell, sets are incredibly useful for defining the relationships between different classes of things. They also provide a neat and concise notation for expressing a lot of complex ideas. One useful perspective on set theory is via **Euler diagrams and Venn diagrams**, which give different sets a nice geometric perspective.

For example, suppose we have two sets A and B. We can express the relationships their elements have using Venn diagrams:

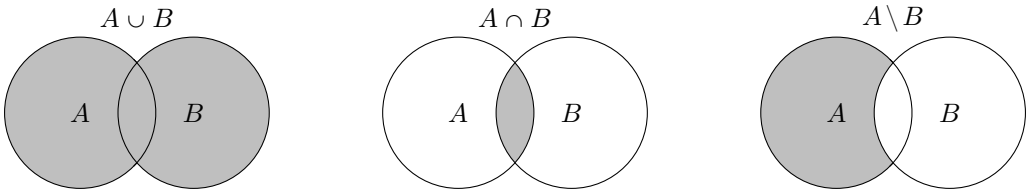

Figure 1: Venn diagrams showing set union, intersection and subtraction.

Question 1.
Including the 'outside region', how many regions can there be on a Venn diagram with n sets?

Most interesting mathematics arises out of talking about sets of interesting things, but importantly we need to know how sets can relate to one another. A set might be completely contained in another set, or even contain all of the same elements.

Notation 10.

- If two sets A and B contain all of the same elements, we write $A = B$.
- If every element in A is contained in B, we write $A \subseteq B$.

- If every element in A is contained in B, and $B \neq A$, then we write $A \subset B$.

Example 11.
Suppose $A = \{\text{John, Mary, Sue}\}$, and $B = \{\text{Sue, John, Esperanza, Anita, Mary, Mei}\}$, then $A \subset B$.

Definition 12.
The **cardinality** of a set A, written $|A|$, is the number of elements in the set A. We'll discuss the cardinality of infinite sets in the supervision!

There is one set in particular which will be very useful on many occasions, so we mention it now:

Definition 13.
The set which contains no elements is called the **empty set** and is denoted $\varnothing$.

We're going to keep using sets throughout this entire course, because they appear everywhere. We're now going to use the language of sets to discuss the notion of a *relation*.

1.2 Relations

Suppose we have a collection of similar objects, and we want to understand how they relate to each other. In order to do this, we use *relations*. You've met lots of relations already.

Example 14.
Many of the symbols you already know are relations. For example, $<, >, \leqslant, \geqslant$, and $=$ are all standard examples of relations, as they describe how objects (in this case, numbers) relate to one another. As it turns out, equality ($=$), is a special kind of a relation, called an equivalence relation.

Definition 15.
A (binary) relation $\sim$ on a set S is a property that may or may not hold between any ordered pair of objects x and y in S.
If it does hold, we write $x \sim y$.

Relations can have certain properties. These properties define part of the behaviour of the relation.

Definition 16.
Here are three standard properties a relation might have:

- A relation $\sim$ is **transitive** if for all $a, b, c \in S$ we have that $a \sim b$ and $b \sim c \implies a \sim c$.
- A relation $\sim$ is **symmetric** if for all $a, b \in S$, $a \sim b \implies b \sim a$.
- A relation $\sim$ is **reflexive** if for all $a \in S$ we have $a \sim a$.

Example 17.
The relation $\leqslant$ is transitive and reflexive, but not symmetric. For example, $1 \leqslant 2$ and $2 \leqslant 3 \implies 1 \leqslant 3$, and $1 \leqslant 1$. However, $1 \leqslant 2$ does not imply $2 \leqslant 1$.

Example 18.
The set relation $\subset$ is a relation between sets. It is transitive but not symmetric or reflexive.

Definition 19.
An **equivalence relation** is any relation $\sim$ which is transitive, symmetric and reflexive.

Example 20.
The standard example of an equivalence relation is equality between numbers. Suppose $a, b, c \in \mathbb{R}$ Then, $a = b$ and $b = c$ does imply $a = c$, so we have transitivity. We know $a = b \iff b = a$, so we have symmetry, and $a = a$ for any number $a \in \mathbb{R}$.

In general, equivalence relations can be thought in general to be like analogues of equality in different sets.

Question 2.
Which of the following relations are equivalence relations?

- Let A and B be sets. Let $A \sim B$ if A and B have identical elements.
- Let x and y be wedding guests. Let $x \sim y$ if x sat at the same table as y.
- Let α and β be different species. Let $\alpha \sim \beta$ if β is a direct descendant of α.
- Let p and q be students in Cambridge. Let $p \sim q$ if p and q are at the same college.
- Let $m, n \in \mathbb{Z}$. Let $n \sim m$ if $n - m$ is divisible by 2.

Relations can also be expressed in the language of sets. For a given relation $\sim$ on a set S, we can make a new set R which contains all of the ordered pairs for which the relation is true. E.g., let S be $\mathbb{Z}$ and consider two elements $x, y \in Z$. Let $x \sim y$ if $x - y$ is divisible by 5.

If we let

$$R = \{(x, y) \in \mathbb{Z}^2 : x - y \text{ is divisible by } 5.\}$$

Then R gives a set-theoretical way of describing the relation. The set R would contain elements like $(1, 6)$ or $(8, 3)$, for example.

1.3 Functions

We now come to one of the most useful constructions in mathematics. If we have sets, and relations to describe elements inside of a set, it would be useful to discuss the interactions between different sets. To do this, we define the notion of a function.

Definition 21.
A **function** f between two sets X and Y, written $f : X \to Y$, is a way of assigning each element of X to exactly one element of Y.

Example 22.
Here are some examples of functions.

- Let $f : \mathbb{Z} \to \mathbb{N}$ where $f(x) = x^2$. For every integer $x \in \mathbb{Z}$, this function assigns it exactly one natural number.
- Let P be the set of people at a wedding, and let T be the set of tables. Let $f : P \to T$ be the function sending every person to their table.

- Let S be the set of all possible word documents. Then let $f : S \to \mathbb{N}$ be the function sending each document to its length in characters.

Much like relations, functions can also have certain properties which describe their behaviour. We're going to describe a few properties a function might have, and then we'll provide some diagrams to understand better what they mean.

Definition 23.
Let A and B be sets. Let $a_1, a_2 \in A$. We say that a function $f : A \to B$ is **injective** if, whenever $f(a_1) = f(a_2)$, we must have $a_1 = a_2$. Injective functions are sometimes called *one-to-one*, because one element is mapped to exactly one target.

This definition looks a little bit strange to start with, but it captures the idea that different elements get sent to different places.

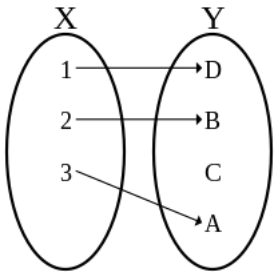

Figure 2: An injective function.

Example 24.
Suppose $f : \mathbb{Z} \to \mathbb{Z}$, where $f : x \mapsto x + 1$. (Note: the notation $\mapsto$ is read '*maps to*' and shows where individual elements are sent.) To prove that f is injective, suppose $f(a) = f(b)$. Then $a + 1 = b + 1$, so $a = b$, so the function is injective.

Example 25.
The function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ is **not** injective, as we can easily provide the counterexample $f(-1) = f(1)$.

Definition 26.
Let A and B be sets. We say that a function $f : A \to B$ is **surjective** if for all $b \in B$ there exists some $a \in A$ such that $f(a) = b$. Another name for surjective functions is *onto*, because every element in B gets mapped onto by an element in A.

Example 27.
Consider $f : \mathbb{N} \to \{0, 1\}$ where $f(x) = x \bmod 2$. I.e. $f(x) = 1$ if x is odd, and $f(x) = 0$ if x is even. Then f is a surjective function.

Definition 28.
We say that a function $f : A \to B$ is **bijective** if it is both injective and surjective.

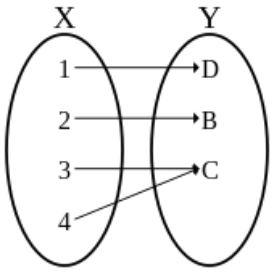

Figure 3: A surjective function.

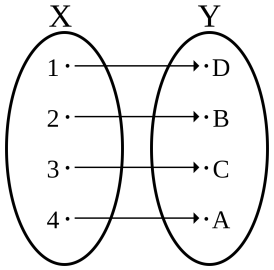

Figure 4: A bijective function.

Example 29.
Consider the function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^3$. To show f is injective, suppose $f(x) = f(y)$ for any $x, y \in \mathbb{R}$. Then $x^3 = y^3$. Since cube root is a function over $\mathbb{R}$, we must have $x = y$, so we have injectivity.

To show that f is surjective, suppose $x \in \mathbb{R}$. Then $\sqrt[3]{x} \in \mathbb{R}$ also, so there exists some $x' \in \mathbb{R}$ with $f(x') = x$.

We now come to our first proposition !

Proposition 30 (The Pigeonhole Principle)**.**
Suppose that $|A| > |B|$, and we have a function $f : A \to B$. Then f cannot be injective.

Proof. For a contradiction, suppose that f is injective. Then for every element of B there is exactly one element of A which maps to it. By pairing off elements in A and B we see we must have $|A| \leqslant |B|$. □

This result states that if you try to put m pigeons into $n < m$ boxes, then at least two pigeons must end up in the same box.

Proposition 31.
Let A and B be two sets. Then if there exists a bijective function $f : A \to B$, then $|A| = |B|$.

The rationale for this is that bijective functions necessarily need to pair elements in A up with elements in B. Moreover, f has to be defined on every element in A and has to touch every element in B (as it is surjective), so no elements in A or B go unpaired.

Figure 5: 10 pigeons in 9 holes.

Question 3.
For each of the following functions are they (a) injective, (b) surjective, (c) bijective?

- $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = e^x$.
- $f : \mathbb{R} \to \mathbb{Z}$ where $f(x) = \lfloor x \rfloor$.
- $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^3 - 2$.

2 Proof, Logic and Truth

2.1 Statements

In university mathematics, there are a few different kinds of statements.

- **Axioms** are statements we fundamentally assume to be true, without needing justification.
- **Definitions** are statements which specify what an object or property is.
- **Propositions** are general statements derived from things we already know.
- **Theorems** are big, interesting, important statements.
- **Lemmas** are smaller, intermediate statements, often used in proofs for other results.
- **Corollaries** are smaller, interesting statements that usually follow immediately from a previous theorem or result.

Statements are all pieces of information which mathematicians collect about different kinds of objects, structures, and spaces. In order to know that the statements we give are true, we use **proof** to collect logical arguments in a way that shows why a statement holds.

- **Proofs** are structured arguments which derive a new statement from something we already know or assume to be true.

Often you'll see statements paired with their proofs like this:

Proposition 32.
Let $n \in \mathbb{Z}$. Then $n^2 + n$ is even.

Proof. First we note that $n^2 + n = n(n + 1)$. Since these two factors are consecutive, either n or $n + 1$ must be even. Multiplying by an even number gives an even number, so $n(n + 1)$ must be even. □

Importantly, mathematics is all about ***conveying an argument***, and this means that, unlike the maths you might have done before, good mathematics is written *in full sentences*, supplemented with notation.

Language is the best tool for communication, so it's important to make use of it when we write mathematics. This is one of the biggest changes between school mathematics and university level mathematics. For much of the mathematics you do in school, it's perfectly fine to write a series of obtuse mathematical equations. At the university level, we use written language to explain our arguments and provide rigorous justifications.

2.2 Logic

2.2.1 Implications

Mathematics is all about how different statements relate to one another. For this purpose, we have some tools for understanding how expressions of truth can relate to one another.

Definition 33.
Let A and B be two statements. We say that A **implies** B, or A is **sufficient** for B, written $A \implies B$, if, whenever A is true, we know that B must also be true.

Example 34.

- John is a medical doctor $\implies$ John went to medical school.

- Harriet is a widow $\implies$ Harriet was married.
- Geraldine is a cat $\implies$ Geraldine is not a dog.

The reverse implication is also useful:

Definition 35.
Let A and B be two statements. We say that A **is implied by** B, or A is **necessary** for B, written $A \impliedby B$, if A is a necessary requirement for B.

These two implications only differ by their direction.

Definition 36.
Let A and B be two statements. If A is **necessary and sufficient** for B, written $A \iff B$ (normally read A if and only if B), then A and B are said to be **equivalent statements**.

When two statements are equivalent, one is true if and only if the other is true. This means that if $A \iff B$, then evaluating the truth of A will give you the truth of B. In order to show that two statements A and B are equivalent, it is enough to show that $A \implies B$ and that $B \implies A$.

Question 4.
Recall the definition of equivalence relation from the first lecture. In the set of all mathematical statements, is $\iff$ an equivalence relation?

Note that $A \implies B$ says absolutely nothing about B if A is not true. The statement

$$p \text{ is a prime number larger than } 2 \implies p \text{ is odd.}$$

Gives you absolutely no information about whether or not p is odd if p is not prime, for example.

Example 37.

- A matrix X is singular $\iff \det(X) = 0$.
- Let n be an integer. Then n is even $\iff n$ is not odd.
- $x = 2x \iff x = 0$.
- Zorn's lemma $\iff$ The axiom of choice.

Example 38.
Let $n \in \mathbb{N}$. Then $\sqrt{n} \in \mathbb{Q} \iff n$ is a square number.

Proof. We first prove the forward implication ($\implies$). Suppose that $\sqrt{n} \in \mathbb{Q}$. Then by definition of $\mathbb{Q}$ we must have that $\sqrt{n} = a/b$ for $a, b \in \mathbb{N}$, $b \neq 0$, where a and b can be taken to share no factors.

Hence we have $n = a^2/b^2 \implies nb^2 = a^2$. Since n is not a square number, there must be some prime factor of n which appears an odd number of times. Let this prime be p. Then $n = p^{2m+1} \cdot c$ for some $m \in \mathbb{N}$ and $c \in \mathbb{N}$, where c contains no factors of p.

Let $x \in \mathbb{N}$ be the power of p appearing in a, and let $y \in \mathbb{N}$ be the power of p appearing in b. Then, counting the number of times p appears on both sides, we have that

$$(2m + 1) + 2x = 2y$$

Which is impossible, since the left hand side is odd and the right hand side is even.

Now we prove the reverse implication ($\impliedby$), although relatively trivial. By definition of a square number, there exists some integer $m \in \mathbb{Z}$ such that $m^2 = n$. Since $m \in \mathbb{Z}$, m is also rational, so $\sqrt{n} \in \mathbb{Q}$. □

We'll now talk a little bit about another useful tool in mathematical logic - the notion of **quantifier**.

2.2.2 Quantifiers

If we want to say that something is true, we have to say what kind of things it is true for. For example, if a property is true *for all* elements in a set, or if it is true for *at least one*.

Notation 39.
We use the symbol $\forall$ as shorthand for *'for all'*. This is called the **universal quantifier**.

Hence the proposition at the beginning of the lecture can also be written:

$$\forall n \in \mathbb{Z}, \quad n^2 + n \text{ is even.}$$

Here are a couple of uninteresting examples:

Example 40.

- $\forall n \in \mathbb{Z}\,[n^2 \in \mathbb{N}]$.
  'For all n in $\mathbb{Z}$, we have n^2 is in $\mathbb{N}$'.
- $\forall n \in \mathbb{N}\,[n^2 + n \bmod 2 = 0]$.
  'For all natural numbers n, $n^2 + n$ is even.
- $\forall x, y \in \mathbb{R}\,[x + y \in \mathbb{R}]$.
  'For all x and y in $\mathbb{R}$, $x + y$ is also in $\mathbb{R}$.

To make our language more powerful, we will introduce another quantifier.

Notation 41.
We use the symbol $\exists$ as shorthand for *'there exists (at least one)'*. This is called the **existential quantifier**. Note that after the exists statement, we also add an implicit *such that* before the statement which follows.

Note: sometimes this implicit *such that* is written with a colon :, but it can also just be read at the start of the statement, written in brackets []. Be careful though, because sometimes colons will still represent functions and mappings.

Here are some slightly more interesting examples.

Example 42.

- $\forall x \in \mathbb{R}, \exists c \in \mathbb{C} : c^2 = x$.
  'For every real number x, there exists a complex number c such that $c^2 = x$.'
- $\forall x \in \mathbb{R}, \exists f : \mathbb{R} \to \mathbb{R} \, [f(x) = x]$.
  'For every x in $\mathbb{R}$, there exists a function mapping $\mathbb{R}$ to $\mathbb{R}$ such that $f(x) = x$.'
- **(Bézout's Identity)**: $\forall m, n \in \mathbb{Z} \; \exists \, a, b \in \mathbb{Z} \, [am + bn = \gcd(m, n)]$.
  'For all integers m and n, there exist two integers a and b such that $am + bn$ is equal to the greatest common divisor of m and n.

We've deliberately used slight variations on these statements, because many authors make slight changes like omitting commas, using colons or using round brackets.

These kinds of statements are really useful when you start learning university mathematics, because they give you an idea about how you should think to structure your argument. Consider the following definition for continuity, written in pure logic.

Definition 43 (Continuous Function)**.**
Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that f is **continuous** at a point $x \in \mathbb{R}$ if:

$$\forall y \in \mathbb{R}, \forall \varepsilon > 0, \; \exists \, \delta > 0 \, [\, |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon \,] \, .$$

At first glance this definition looks very complicated, but if you break it down, the idea that it captures is actually quite intuitive. The idea that the definition captures is that a function is continuous, if whenever you take two inputs and put them arbitrarily close together, then the output of the function can also be arbitrarily close together.

Don't worry too much if this definition is a bit intimidating. It's one of the first logical statements you might meet in **analysis** at university, which is the rigorous study of calculus.

Example 44.
The function $f(z) = 2z$ is continuous at all points $x \in \mathbb{R}$.

Proof. Let $\varepsilon > 0$, and let $y \in \mathbb{R}$. Then $|f(x) - f(y)| = |2x - 2y| \leqslant 2|x - y|$. Then take $\delta = \frac{\varepsilon}{2}$. Then:

$$|x - y| < \delta \implies 2|x - y| < 2 \cdot \frac{\varepsilon}{2} \implies |f(x) - f(y)| < \varepsilon.$$

□

In this case, we allowed ε to be any number greater than 0, and we found an appropriate δ. That is, we showed that **for all** $\varepsilon > 0$ we could have chosen that **there exists** a $\delta > 0$ which we calculated such that the statement is true. Hence we proved the statement.

Question 5.
Prove that $f(x) = 3x + 2$ is continuous for all $x \in \mathbb{R}$.

2.2.3 And and Or

Given two or more logical statements, you can connect them to make a more complex logical statement.

Definition 45.
Let A and B be statements. Then $A \wedge B$ (A *and* B) is the statement that both A and B are true. Similarly, $A \vee B$ (A *or* B) is the statement that at least one of A or B is true.

Example 46.
Consider these statements and how they give their implications:

$$(\text{John is a mathematician}) \wedge (\text{All mathematicians are logical}) \implies (\text{John is logical}).$$

$$(\text{Janet works in medicine}) \vee (\text{Janet has a PhD}) \implies (\text{Janet is a doctor}).$$

Here's a more mathematical example:

$$\forall\, x \in \mathbb{R}\,[(x > 0) \vee (x \leqslant 0)].$$

2.2.4 Negations

When you have a logical statement, you can also negate that statement. For example,

'Harry is a bachelor' becomes *'Harry is not a bachelor.'*

We can do the same thing with mathematical statements. When we have a statement S that we want to negate, we can add the symbol $\neg$ in front of it, giving $\neg S$ (read *'not S'*).

Example 47.
Suppose we have some set M and a function $f : M \to \mathbb{R}$. Here is one possible statement and its negation:

$$\begin{aligned} S: &\quad \forall\, x \in M[f(x) > 0]. \\ \neg S: &\quad \exists\, x \in M[f(x) \leqslant 0]. \end{aligned}$$

Note that the *for all* symbol, $\forall$, flipped and became a *there exists* symbol, $\exists$. This is because in order to show that something is *not true for all* elements of a set, it suffices to show that *there exists* some thing for which it is not true.

Example 48.
Here is a more complex example, using the definition for continuity from before. Suppose again we have a function $f : \mathbb{R} \to \mathbb{R}$ and a point $x \in \mathbb{R}$.

$$\begin{aligned} S: &\quad \forall\, y \in \mathbb{R},\, \forall\, \varepsilon > 0,\, \exists\, \delta > 0\,[\,|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon\,]. \\ \neg S: &\quad \exists\, y \in \mathbb{R},\, \exists\, \varepsilon > 0,\, \forall\, \delta > 0\,[\,|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon\,]. \end{aligned}$$

Notice again that all of the $\forall$ symbols flipped to $\exists$. Note also that in order to show that a given variable with the desired property does not exist, it suffices to show that for all variables that could exist, they do not have the desired property. E.g. to show that there is no man who is a bachelor, it is enough to show that every man is not a bachelor.

Hence we have seen we can negate statements by flipping the quantifiers and then negating the statement!

Question 6.
Can you negate the following statement?

$$\forall\, x \in \mathbb{R}, \exists\, y \in \mathbb{R},\ \text{s.t. } \forall\, z \in \mathbb{R}\, [\sqrt{x} + z = y + z]$$

Question 7.
What happens to the symbols $\wedge$ and $\vee$ when we negate statements?

2.3 Proof

Proof is the structured argument that is required in mathematics to justify your claims. Mathematics is one of the only subjects where we are able to demonstrate that our hypotheses and ideas are definitely true, so mathematicians hold proof as an important standard.

Almost all work in mathematics will be done with justifications in the form of proof. Mathematicians are very good at spotting logical gaps in arguments!

2.3.1 Strategies of proof

Direct proof is the most simple form of proof. It is usually done via a series of steps leading *directly* to the finished result.

Here is an example:

Example 49.
Let $n \in \mathbb{N}^+$ (sometimes authors write $\mathbb{N}^+$ to signal that 0 is not included!). Then $m = n(n+1)(n+2)(n+3)$ is divisible by 24.

Proof. Since there are 4 consecutive numbers being multiplied, exactly two of them contain a factor of 2, one of them contains an additional factor of 2 (as it is a multiple of 4), and at least one of them contains a factor of 3. Hence m must be divisible by $2 \cdot 2 \cdot 2 \cdot 3 = 24$. □

Proof by contradiction (reductio ad absurdum) is the method of proving a statement S to be true, by demonstrating that the negation $\neg S$ is false. This is a very common method of proof, and can make certain problems a lot easier to frame. The most classic example of this method of proof is the following:

Example 50.
The square root of 2 is irrational.

Proof. Suppose for a contradiction that $\sqrt{2}$ is rational. Then $\sqrt{2} = \frac{a}{b}$ for some $a, b \in \mathbb{N}$ both coprime (not sharing factors), with $b \neq 0$. Then we see:

$$\sqrt{2} = \frac{a}{b} \implies 2b^2 = a^2.$$

Counting the prime factors on both sides of the equation shows us that there is an odd number on the left and an even number on the right, which is impossible. Hence our assumption that $\sqrt{2}$ is rational is false, so it must be irrational. □

2.3.2 Proof by induction

Proof by induction is a powerful technique of proof which can demonstrate the truth of a countable number of statements. Sometimes we find that we need to prove a statement for an infinite list of statements. In these cases, we obviously cannot prove every statement, so we use induction to complete the proof for us.

Suppose these statements can be listed $P(1), P(2), \ldots, P(n), \ldots$, so that we have infinitely many cases to prove. Induction relies on showing two things:

- **The base case**: We start by proving that the first statement $P(1)$ is true.
- **The inductive step**: We demonstrate that $P(n) \implies P(n+1)$.

Note that in the inductive step we are allowed to assume that $P(n)$ is true, and use this to prove that $P(n+1)$ must also be true.

If we have $P(1)$, then by the inductive step we must also have $P(2)$. Similarly $P(2)$ gives us $P(3)$ and so on:

$$P(1) \implies P(2) \implies P(3) \implies P(4) \implies \cdots$$

So we get to prove an infinite list of statements by just proving one! Here is an example:

Proposition 51.
Let $n \in \mathbb{N}$. Then the sum of all natural numbers up to n can be expressed as follows:

$$\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n - 1 + n = \frac{n(n+1)}{2}.$$

Proof. **Base case**: Let $n = 1$. Then the left hand side is $\sum_{i=1}^{1} i = 1$, and the right hand side is $\frac{1(1+1)}{2} = 1$, so the statement is true for $n = 1$.

Inductive step: Assume that the statement is true for $n = 1$. Then

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

Then we have that

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

Which is the statement for $n + 1$. Hence we have shown that $P(n) \implies P(n+1)$, so we're done! □

2.3.3 Other methods

Disproof by counterexample is the quickest way to show that a statement is not true. All you need to do is find an example which immediately disproves the statement.

Example 52.
'All natural numbers are prime' **cannot** be true as 4 is a positive natural number which is not prime.

Some massive open conjectures in mathematics (such as the **Riemann Hypothesis**) could also be easily disproved if there exists a counterexample. The problem is finding one!

Proof by contrapositive relies on an intriguing tautological fact (i.e. something that is always true) that $[A \implies B] \iff [\neg B \implies \neg A]$. That is to say, if you need to prove a forward implication, you can negate the statements and prove the reverse implication!

Example 53.
Consider the statement *'All prime numbers greater than 2 are odd.'*. Let p be prime. Then this statement is equivalent to the statement ($p > 2 \implies p$ is odd).

This entire statement is equivalent to saying (p is even $\implies p \leqslant 2$).

Sometimes using the contrapositive can make proving a statement much easier!

2.3.4 Truth tables

One great tool for illustrating logical identities is **truth tables**. In a truth table, all of the possible truth values of different variables are written out in full. For example, here is a truth table for the *and* ($\wedge$) operation we saw before:

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

Figure 6: Truth table for **and**.

Similarly, here is one for the *or* ($\vee$) operation:

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

Figure 7: Truth table for **or**.

Here is a truth table for **implication** ($\implies$):

A	B	$A \implies B$
T	T	T
T	F	F
F	T	T
F	F	T

Figure 8: Truth table for **implication**.

Note that the third and fourth rows seem initially quite counterintuitive. To see why the implication is true, recall that the implication makes no statement about the truth status of B when A is false. It only states that *if* A is true, then B must also be true.

Here is a much more complex example, showing why the **contrapositive** of an implies statement can be called a tautology.

Definition 54.
A **tautology** is a mathematical statement which is *always true*.

A	B	$\neg A$	$\neg B$	$A \implies B$	$\neg B \implies \neg A$	$[A \implies B] \iff [\neg B \implies \neg A]$
T	T	F	F	T	T	T
T	F	F	T	F	F	T
F	T	T	F	T	T	T
F	F	T	T	T	T	T

Figure 9: Truth table for the **contrapositive**.

You can see from the table that regardless of the truth value of A or B, the contrapositive identity is always true! Hence it is an example of a tautology.

Someone once described mathematics as the 'study of interesting tautologies.' Now you can go out and find your own interesting tautologies!

Question 8.
Write out the truth table for $\neg[A \wedge B] \iff [\neg A \vee \neg B]$.

3 Elementary Number Theory

In this lecture we're going to have a look at some different concepts from elementary number theory. By the end we will cover some of the most fundamental concepts in number theory:

- The Fundamental Theorem of Arithmetic;
- The Division Algorithm;
- Modular Arithmetic.

3.1 The Division Algorithm

We're going to start by looking at a useful algorithm for quickly finding the highest common factor of two integer numbers m and n. We're going to use the algorithm (also called Euclid's algorithm) to see some interesting identities and then later demonstrate a key theorem in number theory called the **Fundamental Theorem of Arithmetic**, which guarantees that any natural number can be uniquely factorised into a product of prime numbers.

3.1.1 LCM and HCF

Notation 55.
We use the symbol | to mean '*divides*'. That is to say, if $x|y$, then $y = ax$ for some $a \in \mathbb{Z}$. Equivalently, x is a factor of y.

Definition 56.
Let $m, n \in \mathbb{N}$. The **lowest common multiple** (lcm) of m and n, $\text{lcm}(m, n)$, is the smallest number $k \in \mathbb{N}$ such that both $m|k$ and $n|k$.

Definition 57.
Let $m, n \in \mathbb{N}$. The **highest common factor** (hcf) of m and n, $\text{hcf}(m, n)$, is the largest number $l \in \mathbb{N}$ such that both $l|m$ and $l|n$.

The highest common factor is also sometimes called the **greatest common divisor** (gcd).

Note that these two definitions are very similar. Here's an important result you might want to try proving:

Question 9.
Let $m, n \in \mathbb{N}$. Prove that

$$\text{lcm}(m, n)\,\text{hcf}(m, n) = mn.$$

Because of this result, we can calculate the lowest common multiple from the highest common factor and vice versa.

As it turns out, there's a relatively quick way of computing the highest common factors of any two numbers $m, n \in \mathbb{N}$. This algorithm is called the **Division Algorithm** or **Euclid's Algorithm**. The algorithm depends on an important mathematical fact, called **Euclid's Division Lemma**:

Lemma 58 (Euclid's Division Lemma)**.**
Let $m, n \in \mathbb{N}$. Then there exist unique $q, r \in \mathbb{N}$ such that $0 \leqslant r < n$ and

$$m = qn + r$$

Hard Question 10.
The **well-ordering principle** (WOP) states that every non-empty subset of the positive natural

numbers has a smallest element. Using the well-ordering principle, prove existence and uniqueness for Euclid's Division Lemma.

Lemma 59.
Let $m, n \in \mathbb{N}$. Let $q, r \in \mathbb{N}$ be the divisor and remainder of m and n as per Euclid's division lemma. Then

$$\text{hcf}(m, n) = \text{hcf}(n, r).$$

Proof. Suppose that d is the highest common factor of m and n. Then in particular

$$d \,|\, m \implies d \,|\, (qn + r)$$

Since $d \,|\, n$, we must also have $d \,|\, r$. As a result we must have $\text{hcf}(n, r) \geqslant \text{hcf}(m, n)$.

Now suppose $d' = \text{hcf}(n, r)$. Then $d' \,|\, n$ and $d' \,|\, r$, so $d' \,|\, m$. Hence $\text{hcf}(m, n) \geqslant \text{hcf}(n, r)$.

Combining these facts we must have $\text{hcf}(m, n) = \text{hcf}(n, r)$. □

The strategy we used to show that the two hcfs were the same is very common in mathematics. Often the easiest way to show two numbers f and g are the same is to show in two steps that $f \leqslant g$ and $g \leqslant f$. This strategy also works for things like sub-objects and subsets (where you can use $\subseteq$ and $\supseteq$).

Algorithm 60 (Euclid's Division Algorithm)**.**
Select two numbers $m, n \in \mathbb{N}$ for which you want to calculate $\text{hcf}(m, n)$. We'll start with $m_1 = m$ and $n_1 = n$.

1. Compute q and r for the pair (m_i, n_i) as per Euclid's Division Lemma.
2. If $r = 0$ then terminate the algorithm with $\text{hcf}(m, n) = n_i$.
3. Go back to step 1 using $m_{i+1} = n_i$ and $n_{i+1} = r$.

Proof. The algorithm will terminate as at every step r must be strictly less than n. Hence r will decrease with every step until it is eventually zero.

Moreover, due to lemma 59, we know that the highest common factor $\text{hcf}(m_i, n_i) = \text{hcf}(m_{i+1}, n_{i+1})$ is preserved at each step. In the last step, where $r = 0$, we can stop as $\text{hcf}(k, 0) = k$ for any integer k. □

Hard Question 11.
In practice, the algorithm is actually must faster than linear. Explain why the value of r decreases at least exponentially on average.

Euclid's algorithm is incredibly useful for quickly calculating the highest common factor of even very large numbers. Because the magnitude of the numbers decreases so quickly, it is actually a surprisingly efficient algorithm. We're now going to use Euclid's algorithm to look at a special result.

3.1.2 Bézout's Identity

We can use the division algorithm to calculate the highest common factor of two numbers $m, n \in \mathbb{N}$, but we can also use the division algorithm to find a special expression relating them to their highest common factor. The result is called **Bézout's Identity**.

Theorem 61.
Let $m, n \in \mathbb{Z}$. Then there exist some $a, b \in \mathbb{Z}$ such that

$$am + bn = \text{hcf}(m, n).$$

Proof. Every step of Euclid's algorithm gives a new expression of the (m_i) in terms of (m_{i+1}, m_{i+2}), where m_{i+2} has coefficient 1.

Hence we can substitute m_j with an expression in terms of m_{j-2}, m_{j-1}. This means we can take the final equation where $m_k = \text{hcf}(m, n)$, and substitute backwards into $m_1 = m$ and $n_1 = n$ to get the required expression. □

Example 62.
Let $m = 29$ and $n = 13$. Then at each step we obtain:

$$\begin{aligned} 29 &= 2 \cdot 13 + 3 \\ 13 &= 4 \cdot 3 + 1 \\ 3 &= 3 \cdot 1 + 0. \end{aligned}$$

Moving each expression so that only the remainder is on the right hand side gives:

$$\begin{aligned} 29 - 2 \cdot 13 &= 3 \\ 13 - 4 \cdot 3 &= 1 \end{aligned}$$

Then we substitute the top equation into the bottom to give

$$13 - 4 \cdot (29 - 2 \cdot 13) = 1$$

That is we have

$$-4 \cdot 29 + 9 \cdot 13 = 1.$$

Hard Question 12.
Note that this means that any two numbers $m, n \in \mathbb{N}$ which are coprime have some expression

$$am + bn = 1.$$

Is the converse also true? If such an expression exists, are m and n necessarily prime?

3.2 The Fundamental Theorem of Arithmetic

Definition 63.
Let $p \in \mathbb{N}^+$, with $p \geqslant 2$. We say that p is **prime** if it is only evenly divisible by 1 and p. That is, p has no factors besides 1 and itself.

Prime numbers have fascinated mathematicians for millenia and continue to be an important topic of research today. For example, an important encryption algorithm, **RSA** is built using two large prime numbers.

Besides encryption, the distribution of prime numbers captures the complexity of the interplay between addition and multiplication. While all natural numbers are generated additively using only 1, all of the prime numbers are required to generate $\mathbb{N}$ multiplicatively.

One large unsolved problem in mathematics, the **Riemann Hypothesis**, is predominantly concerned with studying the **Riemann Zeta Function**, whose zeros encode the locations of prime numbers. Proving the Riemann Hypothesis has been one of the most evasive challenges in modern mathematics - so much so that in the year 2000, the Clay Mathematics Institute named it as one of seven of their *Millenium Prize Problems*. Each of these problems are so important that the institute announced a total prize of 1 million dollars for the first correct proof!

For now, we're going to state one of the oldest and most useful theorems in number theory. So useful, in fact, that it has the grandiose name **The Fundamental Theorem of Arithmetic**, or FTA. To prove it, we're going to first need a lemma.

Lemma 64 (Euclid's Lemma)**.**
Let $p \in \mathbb{N}$ be a prime number and let $a, b \in \mathbb{N}$, and suppose that $p|ab$. Then $p|a$ or $p|b$.

Proof. It suffices by symmetry to show that if $p|ab$ but $p \nmid a$, then $p|b$, so assume $p \nmid a$. Hence p and a must be coprime (their highest common factor is 1).

Using Bézout's identity from the previous subsection, we hence know that there exists some $m, n \in \mathbb{Z}$ such that

$$np + ma = 1.$$

Multiplying by b gives

$$nbp + mab = b.$$

Since we know $p|ab$, this means that both terms on the left are divisible by p. Hence The left hand side is divisible by p, and hence $p|b$. □

Theorem 65 (The Fundamental Theorem of Arithmetic)**.**
Every $n \in \mathbb{N}$ greater than one can be expressed uniquely as the product of prime factors, up to the order of the factors.

This theorem has two parts - *existence* and *uniqueness*. This kind of structure comes up repeatedly in mathematics, so normally we prove existence first and then we show uniqueness. To show uniqueness, the general strategy is to assume for a contradiction that two distinct objects with the given property exist, and then show that they are, in fact, the same. We'll use this strategy in this proof.

Proof. **Existence**.
Let $n \in \mathbb{N}$. If n is prime then it is already a product of prime factors and we are done, so assume that n is composite.

We'll continue by (strong) induction. Assume (for the inductive step) that all numbers $1 < k < n$ have the desired factorisation into primes. Then n is composite, so n can be written as the product of two numbers $n = ab$ with $1 < a, b < n$.

By our assumption both a and b have a factorisation into primes, so putting these together will give the factorisation of n into primes. Hence n can be factorised into prime numbers.

Uniqueness.
Suppose that n has two prime factorisations $n = p_1p_2p_3 \dots p_k = q_1q_2q_3 \dots q_l$, where p_i and q_j are all prime numbers, possibly repeated.

We see that $p_1|q_1q_2 \dots q_k$, as $p_1|n$. Using Euclid's Lemma (lemma 64), we know that $p_1|q_j$ for some prime q_j. But this can only happen if $p_1 = q_j$. Hence we have

$$\frac{n}{p_1} = p_2 \dots p_k = q_1 \dots q_{j-1}q_{j+1} \dots q_l$$

Doing the same trick with the remaining p_i we see that for every p_i we have $p_i = q_{\sigma(i)}$ for some permutation σ (a bijective function from $\{1, \dots, j\}$ to itself). Hence the prime factors are unique up to reordering. □

It's hard to understate the importance of the fundamental theorem of arithmetic. It provides so much useful structure and interesting mathematics that the idea behind it never stops appearing in number theory. In general, when a theorem forms the basis for an entire field of study, it is often labelled as a fundamental theorem. E.g. the fundamental theorem of calculus, the fundamental theorem of Galois theory etc.

3.3 Modular Arithmetic

Most of the time when we count we think of the number line as extending infinitely far in both directions, but there are other systems of arithmetic which are finite which have interesting and important mathematics. The most standard example in this domain is the way humans tell time:

we do not count how many seconds have elapsed since the beginning of time, rather we break time up into 24 hour cycles and do most of our calculations with this.

When we use a clock, time loops around endlessly. If I ask you what time it is 4 hours after 22:00, then the clock will loop back around on itself and go back down to 02:00. This notion of *looping around* is the essence of modular arithmetic. Instead of counting on an infinite number line, we work with a (usually) finite cycle of numbers. In the same way that we add 4 to 22 to get 26, and then drop the superfluous 24 hours, we can pick any number, called the **modulus**, which we can also drop.

Notation 66.
When we use modular arithmetic, we select a number called the **modulus**. This is usually written at the end of an equation as 'mod n' or '(mod n)'.

Example 67.
Calculating in mod 7 we have:

$$4 + 6 = 10 \equiv 3 \bmod 7.$$

Some authors use $\equiv$ rather than $=$ once we've performed the reduction, to signify that any representative (e.g. any $3 + 7n$) could in theory could be used as notation. You might also see $\overline{3}$, which captures the same idea.

Definition 68.
Let $a \in \mathbb{Z}$ and $n \in \mathbb{N}$. Consider the set $\overline{a} \bmod n = \{a + nx : x \in \mathbb{Z}\}$. Every element in this set is equivalent in modular arithmetic, so we can choose any of these elements to represent the set. The element we choose (usually the smallest) is called a **representative** of the set.

You might see that computing the value of $m \bmod n$ is equivalent to computing the remainder when dividing m by n.

Modular arithmetic works because of the following two facts:

Proposition 69.
Let $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}^+$, then the following two statements both hold:

- $(a \bmod n) + (b \bmod n) = (a + b) \bmod n$;
- $(a \bmod n) \cdot (b \bmod n) = (ab) \bmod n$.

Proof. We'll show both properties individually.

- Consider two representatives $a + xn$ and $b + yn$ for $x, y \in \mathbb{Z}$. Then $(a + xn) + (b + yn) = a + b + (x + y)n \equiv a + b \bmod n$.
- Taking the same representatives $a + xn$ and $b + yn$, we note that $(a + xn) \cdot (b + xn) = ab + (ax + bx + x^2 n)n \equiv ab \bmod n$.

□

One way of phrasing this property (which you might see later into a mathematics degree) is that addition and multiplication *commute* with reduction $\bmod n$. This means that you can do either operation first and the result will not change.

3.3.1 Modular Arithmetic with Exponents

Because reduction $\bmod n$ commutes with multiplication, we are able to use this to our advantage to compute the remainder when we divide very large numbers. What is the final digit of 2023^{2023}?

At first this seems like a really difficult question, but using modular arithmetic this becomes a very accessible problem. We'll give some examples to demonstrate how this can be done.

Example 70.
We're going to compute $93^{28} \bmod 5$. To do this, note that:

$$93^{28} \equiv 3^{28} \bmod 5.$$

Immediately we can discard the extra 90 which we don't need. Now we notice that

$$3^{28} = 3^{2\cdot 2\cdot 7} = 9^{2\cdot 7} = 4^{2\cdot 7} \bmod 5,$$

where in the last step we've again discarded an unnecessary 5. Continuing in this fashion we see

$$4^{2\cdot 7} = 16^7 = 1^7 \equiv 1 \bmod 5.$$

Hence we now that 93^{28} is, in fact, one more than a multiple of 5!

Question 13.
What *is* the final digit of 2023^{2023}?

Modular arithmetic is an incredibly useful tool, and the concept it represents (discarding some uninteresting data and keeping some useful detail) is given by the slightly intimidating concept of **quotients** in abstract algebra. We'll get a chance to talk about quotients in the next lecture, where the concept is generalised to all sorts of unfamiliar objects. Taking quotients is one of the most powerful tools in mathematics and will reappear so often that the basic concepts you've just learned in modular arithmetic might just reappear one hundred times in a hundred different forms. After the next lecture you'll have a new insight on why this seemingly simple game of *reading clocks* can actually be incredibly powerful, and has been the gateway to some of the richest mathematics of modern times.

4 Abstract Algebra

One deep insight that resulted from 19th and 20th century mathematics is that many different kinds of objects behave in ways which are suspiciously similar. Consider for example the set $\{1, i, -1, -i\}$. Multiplying any of these elements together will result in another element in the set, and drawing these numbers on an Argand diagram looks *suspiciously similar* to something like algebra $\bmod 4$. How can we discuss the connection between these two systems? Why should there be a connection at all?

As we've seen, even certain kinds of proofs appear to have some structural similarity, and this desire to see a deeper connection between complex objects which, despite perhaps looking quite different, behave in very similar ways, is a central and deep concept in modern algebra.

To explore these kinds of connections, we're going to take a look at a kind of mathematical object called a **group**, which can be thought of as representing various different kinds of symmetry. To start with, we'll need some definitions.

4.1 Groups

Definition 71.
A **binary operation** $*$ is any kind of function from pairs in a set S back to S. I.e. $* : S \times S \to S$.

Example 72.
You are already familiar with many different binary operations. For example:

- Addition $(+)$ is a binary operation.
- Multiplication $(\cdot)$ is a binary operation.
- Subtraction $(-)$ is a binary operation in $\mathbb{Z}, \mathbb{Q}$ or $\mathbb{R}$, but not in $\mathbb{N}$.
- Division $(\div)$ is a binary operation in $\mathbb{Q}^*$ and $\mathbb{R}^*$.

Note that the asterisk in $\mathbb{Q}^*$ and $\mathbb{R}^*$ signals that we exclude zero, as we cannot divide by zero in general.

Question 14.
Why isn't $\div$ a binary operation on $\mathbb{Z}^*$?

Binary operations can have many properties, but one of the most useful for calculations is called *associativity*.

Definition 73.
A binary operation $*$ on S is **associative** if for all $x, y, z \in S$ we have

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

That is, it does not matter in which order we do the operations.

Now that we've given these two definitions, we're ready to talk about one of the most fundamental objects in mathematics. The **group**. Group theory, the study of groups, is an enormous field of study with many far reaching consequences.

Definition 74 (Group)**.**
Let S be a set and let $*$ be a binary operation on S. Then $(S, *)$ is a **group** if it has the following properties:

- **Closure**: For all $x, y \in S$, $x * y \in S$ (note this is implicit from the definition of binary operation, but often stated).
- **Associativity**: For all $x, y, z \in S$, we have $(x * y) * z = x * (y * z)$.
- **Identity**: There exists some element $e \in S$ such that for all $x \in S$, $e * x = x * e = x$.
- **Inverses**: For every element $x \in S$ there exists some x^{-1} such that $x * x^{-1} = x^{-1} * x = e$.

Note in particular that we do not require in general that $a * b = b * a$. This property is called **commutativity**. If a group has a commutative operation, it is known as an **abelian group**.

Notation 75.
You will find that often that, where the group operation is known explicitly, then it will often be simply excluded. This notation, called **multiplicative notation**, simply places symbols next to each other. For example,

$$x * y \quad \text{might be written as} \quad xy.$$

In groups like this we often just denote the identity by 1 or $\mathbb{1}$.

Moreover, when the group operation is commutative, we might use **additive notation**, whereby the group operation is represented with a $+$. In this notation the identity is usually written as $\mathbb{0}$.

Example 76.
You already know lots of examples of groups! Here are some of the most common:

- $(\mathbb{Z}, +)$. In this group the identity is 0, and the inverse of $x \in \mathbb{Z}$ is $-x$. Note that $x + (-x) = 0 = (-x) + x$.
- $(\mathbb{R}^*, \cdot)$. This group is multiplicative, so the identity is 1, and the inverse of x is x^{-1}. Hence $x \cdot x^{-1} = 1$. Note that we exclude 0 because it does not have a multiplicative inverse.
- $(\mathbb{Z}_n, +_n)$. This group represents the integers $\bmod n$ with addition $\bmod n$. For example with $n = 7$, here $5 + 4 = 2 \bmod 7$. The identity is 0, and the additive inverse of x is $7 - x$. E.g. $5 + 2 = 7 \equiv 0 \bmod 7$.

Note that we don't even have to use numbers to construct groups. It is perfectly valid to construct a group by describing how each symbolic element interacts with every other element. We'll give one important classical example in the form of a **dihedral group**.

Example 77 (Dihedral Group, D_8)**.**
Consider the square:

Here you can see that there are 8 geometric symmetries of the square. We have the four rotations $\rho_0, \ldots, \rho_3$ and the four reflections $\sigma_1, \ldots, \sigma_4$. Note that the numbering is slightly different between these two types of symmetry. The reason for this is that there is a *zero rotation*, ρ_0, which leaves the square completely unchanged. As this element leaves the vertices unmoved, it is the identity element in this group.

We call this the **dihedral group of order 8**, as it has 8 elements:

$$\{\rho_0, \rho_1, \rho_2, \rho_3, \sigma_1, \sigma_2, \sigma_3, \sigma_4\}.$$

If we keep track of the numbers on the vertices, then we see that

$$(1, 2, 3, 4) \xrightarrow{\rho_1} (4, 1, 2, 3) \xrightarrow{\sigma_2} (4, 3, 2, 1).$$

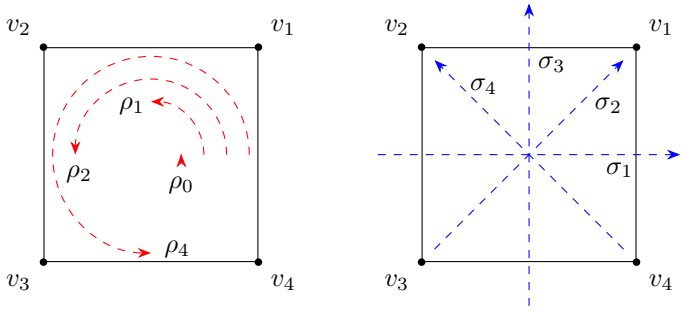

Figure 10: The rotational and reflectional symmetries of the square.

Since doing both of these steps is equivalent to performing the single reflection σ_4:

$$(1,2,3,4) \xrightarrow{\sigma_1} (4,3,2,1),$$

we can allow ourselves to define the multiplication such that $\sigma_2\rho_1 = \sigma_1$ [Note: it is commonplace to place the first action on the right and then subsequent actions on the left - this is because when we compose functions $f \circ g$, this can also be written $f(g(x))$ and this would mean that g is actually performed first].

In fact, for every combination of steps we could apply to the vertices of the square, there is always one action in the group which does the equivalent permutation of the vertices. Hence we can specify the group operation in D_8 to be given by this reduction of steps.

Here is a table to illustrate the multiplication:

$\cdot$	ρ_0	ρ_1	ρ_2	ρ_3	σ_1	σ_2	σ_3	σ_4
ρ_0	ρ_0	ρ_1	ρ_2	ρ_3	σ_1	σ_2	σ_3	σ_4
ρ_1	ρ_1	ρ_2	ρ_3	ρ_0	σ_4	σ_1	σ_2	σ_3
ρ_2	ρ_2	ρ_3	ρ_0	ρ_1	σ_3	σ_4	σ_1	σ_2
ρ_3	ρ_3	ρ_0	ρ_1	ρ_2	σ_2	σ_3	σ_4	σ_1
σ_1	σ_1	σ_2	σ_3	σ_4	ρ_0	ρ_1	ρ_2	ρ_3
σ_2	σ_2	σ_3	σ_4	σ_1	ρ_3	ρ_0	ρ_1	ρ_2
σ_3	σ_3	σ_4	σ_1	σ_2	ρ_2	ρ_3	ρ_0	ρ_1
σ_4	σ_4	σ_1	σ_2	σ_3	ρ_1	ρ_2	ρ_3	ρ_0

Note that the action on the left is performed first and the action on top is performed second. This is called a **Cayley table**, and it represents all of the multiplication combinations that are possible in the group.

Note in general that $ab \neq ba$ in D_8, so the group is not abelian.

Definition 78.
The **order** of a group G, written $|G|$ or $\text{ord}(g)$, is the cardinality (size) of its underlying set.

Example 79.

- The dihedral group D_8, has $|D_8| = 8$.
- The integers $\bmod n$ has $|\mathbb{Z}_n| = n$.
- The group of rotational and reflective symmetries of an m-gon has $2m$ elements.

4.2 Subgroups

4.2.1 Definition and Examples

You might have noticed that in the case of D_8, we could have removed all of the reflective symmetries σ_i and still been left with a group consisting of only rotations ρ_j. This object, a smaller group contained in a larger group, is known as a **subgroup**.

Definition 80.
Let G be a group. H is a **subgroup** of G, written $H \leqslant G$, if we have that $H \subseteq G$ and H is a group with respect to the same group operation.

Definition 81.
Let G be a group, and let $x \in G$ be an element. The set $\langle x \rangle = \{\ldots, x^{-2}, x^{-1}, 1, x, x^2, \ldots\}$ is called the **subgroup generated by** x.

Note that this definition can also be expressed in additive notation, in which case we have

$$\langle x \rangle = \{\ldots, -2x, -x, 0, x, 2x, \ldots\}.$$

Example 82.

- In the dihedral group of order 8, D_8, we can see that there exists a subgroup consisting only of rotations ρ_i. This subgroup is equivalent to $\langle \rho_1 \rangle$.
- In the additive group $(\mathbb{Z}, +)$, all of the subgroups are of the form $n\mathbb{Z} = \{nz : z \in \mathbb{Z}\}$. For example, the even numbers $2\mathbb{Z}$ are an additive group under addition.
- Consider the group $S = \{c \in \mathbb{C} : |c| = 1\}$. This is a group under multiplication. Inside of this group, the set of n-th roots of unity forms a subgroup for all $n \in \mathbb{N}$.

Subgroups are incredibly important to the study of groups, and, in general, whenever you want to study a new kind of objects in mathematics, it's always wise to look at a few things:

- **Sub-objects**: how is the object structured in terms of its parts.
- **Maps between objects**: what are the sensible ways to relate objects to each other.

We might come back to the maps (called *homomorphisms*) in supervisions, but we're going to discuss first some interesting results that can be obtained by thinking about subgroups.

4.2.2 Cosets

In the previous subsection we saw that the even numbers are a subgroup in the integers $\mathbb{Z}$. This is quite interesting, but leaves one question hanging in the air: *what about the odd numbers?* Indeed, most of the time we're quite comfortable thinking about odd and even numbers as being equally important, so it might seem strange at first to realise that the odd numbers *do not* form an additive subgroup of $\mathbb{Z}$ (for instance, because they do not contain 0).

In fact, the correct language for discussing the odd numbers is via **cosets**. Cosets will reappear many times in university level algebra, and are fundamental to one idea we introduced earlier: *taking quotients*.

Definition 83.
Let $H \leqslant G$, and let $g \in G$. Then the set:

$$gH = \{gh : h \in H\}$$

is called a **left coset**. Similarly, any set of the form

$$Hg = \{hg : h \in H\}$$

is called a **right coset**. If the group is abelian (and the group operation is commutative), then the left and right cosets are equal for each $g \in G$.

Notation 84.
Note that we've written the definition for cosets using multiplicative notation. Recall that, in general, we only use additive notation when the group is abelian. This means that given any $g \in G$ and $h \in H$, that $g + h = h + g$. As such, there is no notion of left or right cosets, but just *coset*. Given an element $g \in G$, the additive version looks like this:

$$g + H = \{g + h : h \in H\}.$$

Example 85.
Here are some examples of cosets:

- With $G = \mathbb{Z}$ and $H = 2\mathbb{Z} \leqslant G$, we have the coset $1 + 2\mathbb{Z} = \{1 + 2x : x \in \mathbb{Z}$ (the odd numbers).
- Recall that $\langle \rho_1 \rangle \leqslant D_8$. The coset $\sigma_1 \langle \rho_1 \rangle = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$.

There are two important results which make cosets so powerful, and lead to the central theorem of this section. We'll state them now.

Lemma 86.
Let $H \leqslant G$ and let $g \in G$. Then $|gH| = |H|$. That is to say, every coset of H in G has the same cardinality as H.

Proof. Consider the function $f(h) : h \mapsto gh$, that is, f multiplies an element $h \in H$ by g on the left. We know by the definition of function that the output $f(H) = \{f(h) : h \in H\}$ (also known as the **image** of f) must have $|f(H)| \leqslant |H|$. We will show that these are, in fact, equal.

Suppose for a contradiction that $|f(H)| < |H|$. Then this means that f was not injective, so two distinct elements h_1 and h_2 were sent to the same place by f. Explicitly this means that $gh_1 = gh_2$. Multiplying on the left by g^{-1} (which exists by the definition of group), we see $h_1 = h_2$. Hence f was in fact injective, and $|f(H)| = |H|$ as needed. □

The above result is really quite important, because it tells you that the size of every coset is always the same inside of a group. This is not necessarily something that is immediately obvious when you see the definition, but it ends up being incredibly useful for proving some nice results.

Question 15.
How many cosets does the subgroup $n\mathbb{Z}$ have inside of the integers $(\mathbb{Z}, +)$?

Lemma 87.
Let $g_1, g_2 \in G$, and let $H \leqslant G$. Let g_1H and g_2H be two cosets of H in G. Then either $g_1H = g_2H$ or $g_1H \cap g_2H = \varnothing$.

Proof. We will demonstrate that if $g_1H \cap g_2H \neq \varnothing$, then $g_1H = g_2H$. Firstly consider the coset hH for some $h \in H$. Since hH is a coset, we have $|hH| = |H|$. Moreover, since H is a subgroup, $hh_i \in H$ for all $h_i \in H$ by the closure property of groups, so we have $hH = H$ as a set.

Suppose then that $g_1H \cap g_2H \neq \varnothing$. Then there exists some $h_1, h_2 \in H$ such that $g_1h_1 = g_2h_2$. As a result we must have that

$$g_1h_1H = g_2h_2H$$

However, we saw that applying h_i to every element in H leaves H unchanged, so we must have that

$$g_1 H = g_2 H.$$

□

This lemma again provides more of an intriguing picture of how cosets fit together to build a group. In fact, they partition up a group into equal-sized cosets. Because of this, it seems reasonable to have the following defintion:

Definition 88.
Let $H \leqslant G$. Then the **index** of H in G, denoted $[G : H]$ is the number of cosets of H in G.

We've now done enough work to prove our main result.

Theorem 89 (Lagrange's Theorem)**.**
Let $H \leqslant G$. Then

$$|G| = [G : H]|H|.$$

In particular, the order of H divides the order of G.

Proof. We note that due to the previous two lemmas, every element $g \in G$ belongs to a disjoint coset of size $|H|$. This means that if there are $[G : H]$ cosets all of size $|H|$, then $|G|$ must equal $[G : H]|H|$. □

Example 90.
Recall that in D_8 we saw we had a subgroup $H = \langle \rho_1 \rangle$. There were hence two cosets of H in D_8, and $|H| = 4$, which divides 8.

Example 91.
Consider the integers modulo 13, $(\mathbb{Z}_{13}, +_{13})$. In this group we have $|\mathbb{Z}_{13}| = 13$, which is a prime number. Since the only divisors of 13 are 1 and 13, the only subgroups of $\mathbb{Z}_{13}$ are $\mathbb{Z}_{13}$ and $\{0\}$.

Definition 92.
The (sub)group consisting of only the identity $\{e\}$ is called the **trivial** (sub)group.

Corollary 93.
Let G be a group with $|G| = p$ where p is prime. Then G has no proper nontrivial subgroups.

Proof. The only divisors of p are 1 and p. If $|H| = |G|$ then $H = G$, and if $|H| = 1$, then H must be the trivial subgroup as every group must contain an identity. □

It's hard to overstate the usefulness of Lagrange's theorem. It's one of the first theorems that gives a real hint at how groups as objects tend to behave and how they relate to the subgroups they consist of. Group theory as a field goes much much further in exploring this relationship, producing many useful results of a similar calibre.

In fact, one result, the **Classification of finite simple groups**, classifies and describes every single type of finite group which can exist using a handful of categories. Proving this result took about 50 years of work from hundreds of authors, and the result has been described as one of the 'crowning achievements of modern mathematics' (Hiroshi Ogori).

Many other types of mathematical object exist, and groups are just the beginning. I hope you've enjoyed this brief introduction to university algebra!

5 Analysis

At university, **analysis** is the rigorous study of concepts relating to calculus. This means that it deals with concepts like *limits*, *continuity*, *differentiation* and *integration*, among many others. If you take a mathematics-heavy subject such as physics or computer science, you'll find that analysis might be one of the first things you study.

By the end of this lecture, we'll have discussed:

- **Sequences**. We'll explore sequences of numbers and what it means for a sequence to converge.
- **Limits and continuity**. From this, we'll explore the notion of a continuous function a little bit deeper.
- **Derivatives**. Lastly, we'll examine the formal definition of a derivative.

As students move towards university mathematics, analysis is often used as a good starting point for learning to construct strong and coherent arguments. The reason for this is that many of the concepts we study in analysis are quite intuitive, so it gives us a good place to begin exploring how to relate an intuitive idea to a rigorous mathematical formalism.

Let's get started!

5.1 Sequences

Sequences of numbers appear all over mathematics, and studying their behaviour is a great way to get a grasp on some high-level concepts in analysis. For the rest of this lecture we will take $0 \notin \mathbb{N}$.

Definition 94.
A **sequence** $(a_n)_{n\in\mathbb{N}}$ is an ordered list of numbers.

Definition 95.
A **series** is a sequence $(a_n)_{n\in\mathbb{N}}$ such that each element in the sequence can be expressed as a partial sum:

$$a_n = \sum_{k=1}^{n} f(k)$$

for some $f(k)$.

Example 96.
Here are some examples of sequences and series for a different a_n:

- $a_n = n$

$$1, 2, 3, 4, \ldots$$

- $a_n = \frac{1}{n}$

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$$

- **The Fibonacci Sequence**: $a_1, a_2 = 1$, $a_n = a_{n-1} + a_{n-2}$.

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

- **The Harmonic Series**: $a_n = \sum_{k=1}^{n} \frac{1}{k}$

$$1, \quad 1 + \frac{1}{2}, \quad 1 + \frac{1}{2} + \frac{1}{3}, \quad 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}, \quad \ldots$$

$$1, \quad 1.5, \quad 1.8\dot{3}, \quad 2.08\dot{3}, \quad \ldots$$

You'll notice that the second sequence, $a_n = 1/n$, will slowly get closer and closer to 0 as we go further and further in the sequence. Is the same true of the Harmonic Series? Does this sequence of numbers converge eventually or does it get infinitely large, like the other two sequences?

In order to explore this idea, we'll need to give a definition which captures this behaviour of 'tending to a number.'

Definition 97.
Let $(a_n)_{n\in\mathbb{N}}$ be a sequence with $a_n \in \mathbb{R}$. We say that a_n converges to $a \in \mathbb{R}$ if

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}, N > 0 \, [\, \forall n \geqslant N \, [\, |a_n - a| < \varepsilon \,] \,].$$

We might also write $a_n \to 0$. If a sequence does not converge to a finite number, then the sequence is said to be **divergent**.

The idea behind this definition is this: pick any distance $\varepsilon > 0$ which can be any arbitrary positive number. If I can find a point in the sequence where I'm never further than ε away from the target a for $n \geqslant N$, then the sequence converges.

In particular, note that if we make the distance ε smaller and smaller, then N might have to get larger and larger so that the sequence can still get close enough to a.

Example 98.
The sequence $a_n = \frac{1}{n}$ converges to 0.

Proof. Let $\varepsilon > 0$. Then pick some natural number $N > \frac{1}{\varepsilon}$. Then for all $n > N$ we have

$$|a_n - 0| = \left|\frac{1}{n}\right| = \frac{|1|}{|n|} < \frac{|1|}{|1/\varepsilon|} = \varepsilon.$$

□

Example 99.
Let $a_n = n$. We'll show that a_n does not converge to any number. To prove this, we'll demonstrate that, for all values $a \in \mathbb{R}$, the negation is true:

$$\exists \varepsilon > 0, \forall N \in \mathbb{N}, N > 0, [\, \exists \, n \geqslant N \, [\, |a_n - a| \geqslant \varepsilon \,] \,].$$

Take $\varepsilon = 1$. Then suppose $N \in \mathbb{N}$ with $N \neq 0$. Setting $n = \max(a + 1, N + 1)$ we have $n \geqslant N$ and $|a_n - a| \geqslant |a + 1 - a| = 1 \geqslant \varepsilon$.

Definition 100.
We say that a sequence $(a_n)_{n\in\mathbb{N}}$ **tends to infinity** if

$$\forall a \in \mathbb{R}, \, \exists N \in \mathbb{N}, \, N > 0 \, [\, \forall n \geqslant N, [\, a_n > a \,] \,].$$

This definition captures the idea that a sequence grows without bound. More specifically, for every upper bound $a \in \mathbb{R}$ you could think of, eventually the sequence grows larger than a and stays larger than a.

One useful tool for proving that a sequence is convergent or divergent is a **comparison test**. In this test, we can use sequences and series that we already know are convergent or divergent and compare individual terms.

Proposition 101.
Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be real-valued sequences. If a_n tends to infinity, and for all $N > 0$ there exists some $n > 0$ with $b_n \geqslant a_n$, then b_n is divergent.

Hard Question 16.
Using the definitions for tending to infinity and convergence, can you prove the proposition?

Theorem 102.
The Harmonic Series is divergent.

Proof. Let $(a_n)_{n\in\mathbb{N}}$ be the Harmonic series. Grouping the series together we see that we can compare each of the terms in the sequence with another sequence (b_n), which we construct by swapping out elements at every $n = 2^k$ for $k \in \mathbb{N}^+$ in this fashion:

$$1,\quad 1+\frac{1}{2},\qquad \ldots,\quad 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4},\qquad \ldots,\quad 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8},\qquad \ldots$$
$$1,\quad 1+\frac{1}{2},\qquad \ldots,\quad 1+\frac{1}{2}+\left(\frac{1}{4}+\frac{1}{4}\right),\qquad \ldots,\quad 1+\frac{1}{2}+\left(\frac{1}{4}+\frac{1}{4}\right)+\left(\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}\right),\qquad \ldots$$

We see that for $n = 2^k$ for $k \geqslant 1$ we have $b_n = 1 + \frac{k}{2}$. Hence (b_n) must be a divergent sequence tending to infinity. Hence, as $a_n \geqslant b_n$ for all $n \in \mathbb{N}$, we must also have that (a_n) is divergent. □

Question 17.
Consider each of the following sequences. Are they convergent or divergent?

1. $a_n = \sin(n)$.
2. $a_n = \cos(1/n)$.
3. $a_n = \frac{n^2+3}{n^3}$.
4. $a_n = \frac{n^2}{n^2+1} + \sin(n) + \cos(n + \pi/2)$.

If they converge, what are they converging to?

The theory of sequences and series actually goes a lot deeper than we've had space to explore here, but they provide a useful starting point for exploring more complex topics such as continuity and limits.

5.2 Limits and Continuity

Earlier in the course we briefly met the definition of continuity while learning about quantifiers. To explore this idea a little bit more, we're going to give a new definition, the notion of a **limit**, and use it to rephrase the definition of continuity into something a little bit more intuitive.

In the previous subsection we captured the notion of a *sequence* tending towards a specific number. This is interesting, but sequences are very discrete. What if we have a function $f : \mathbb{R} \to \mathbb{R}$, rather than a sequence $a : \mathbb{N} \to \mathbb{R}$? How can we adapt our definition to capture the idea that f gets closer and closer to a fixed number? The answer, of course, is limits.

Limits capture the idea that as the input to a function f gets very close to a certain point p, then the output of that function gets very close to a certain point q. Sometimes we also use limits to explain the behaviour of a function as its input tends to infinity.

Definition 103.
Let $f : \mathbb{R} \to \mathbb{R}$. We say that the **limit** of $f(x)$ as x tends to c is a if

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

If such an $a \in \mathbb{R}$ exists, we say that the limit exists and write

$$\lim_{x\to c} f(x) = a.$$

If no such a exists, then we say that the limit **does not exist**.

This definition captures the idea that for any distance ε you could name between $f(x)$ and a, there is some small distance δ such that when x is within δ of c, then $f(x)$ is within ε of a.

That is to say, as x gets close to c, $f(x)$ gets close to a.

Example 104.

- The following limit exists:
$$\lim_{x\to 10} x^2 = 100.$$
- This limit does not exist:
$$\lim_{x\to 5} \frac{1}{|x-5|}.$$
- Consider the function $f : \mathbb{R} \to \mathbb{R}$ such that
$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geqslant 0 \end{cases}.$$
Then the limit
$$\lim_{x\to 0} f(x)$$
does not exist.

Limits are lots of fun and give us a way to notate a relatively complex idea. As a shorthand it's incredibly useful as it allows us to avoid using epsilon and delta every time we want to get the same idea across. Moreover, once you're familiar with them, there are several useful limit laws which make computing with limits even easier.

Proposition 105.
Let $f, g : \mathbb{R} \to \mathbb{R}$. We have all of the following limit laws wherever the limits exist:

- $\lim_{x\to c} f(x) + g(x) = \lim_{x\to c} f(x) + \lim_{x\to c} g(x)$.
- $\lim_{x\to c} f(x)g(x) = [\lim_{x\to c} f(x)][\lim_{x\to c} g(x)]$.
- $\lim_{x\to c} f(x)/g(x) = [\lim_{x\to c} f(x)]/[\lim_{x\to c} g(x)]$, provided $\lim_{x\to c} g(x) \neq 0$.

We omit the proofs as they simply more variations on the same style of ε-δ proofs we've seen.

Hard Question 18.
Can you prove all of the identities above?

Question 19.
Evaluate the following limits, or state if they do not exist:

1. $$\lim_{x\to \pi} \sin(x).$$
2. $$\lim_{x\to \pi} \frac{\sin(x)}{x}.$$
3. $$\lim_{x\to e} \frac{1}{\ln(x) - 1}.$$

4.

$$\lim_{x\to 3} x(x^2 + \log_3(x)).$$

The limit in 2. is called an **indeterminate form**, as both the top and the bottom tend to 0. If you want a much easier way to evaluate it, you can read about L'Hôpital's rule online.

We're going to briefly restate the definition of continuity we gave earlier, so that we can restate it in terms of limits.

Definition 106 (Continuous Function ($\varepsilon - \delta$ version))**.**
Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that f is **continuous** at a point $x \in \mathbb{R}$ if:

$$\forall y \in \mathbb{R}, \forall \varepsilon > 0, \exists\, \delta > 0 \left[\, |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon \,\right].$$

Here is the same definition, written in limits:

Definition 107 (Continuous Function (limit version))**.**
Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that f is **continuous** at a point $c \in \mathbb{R}$ if

$$\lim_{x\to c} f(x) = f(c).$$

and the limit exists.

You might agree that the limit version of the definition is a lot easier to digest. These two definitions are equivalent, and you won't be surprised that the limit version is a lot more intuitive and easier to use for more complex proofs.

Interestingly, this isn't the most powerful version of this definition that exists. The most powerful version is defined in terms of special sets called open sets, which are used in the definition to capture the idea that two points are close together. The reason this version is so strong is that these open sets can be defined in many different ways, called **topologies**. In fact, sometimes certain questions are easier defined and answered in more unusual topological spaces, and choosing an appropriate topology can make the notion of 'continuity' very useful in many abstract ways.

Question 20.
Using the limit version of continuity, prove that:

- $f(x) = x^3 + 2x - 3$ is continuous.
- The function

$$f(x) = \begin{cases} x^2 & x < 0 \\ 1 - x^2 & x \geqslant 0 \end{cases}$$

  is not continuous.

Hard Question 21.
Let $f : \mathbb{R} \to \mathbb{Q}$ be **Thomae's function**, where

$$f(x) = \begin{cases} \left(\frac{a}{b}\right) = \frac{1}{b} & a \text{ and } b \text{ are coprime integers with } b > 0 \\ 0 & \text{otherwise} \end{cases}.$$

For which values of $\mathbb{R}$ is f continuous?

5.3 Derivatives

Now that we've built the tools to understand limits, we now have the language to explore the formal definition of the derivative. As you might have seen, the derivative of a continuous function $f : \mathbb{R} \to \mathbb{R}$ is given at a point x by drawing a tangent at x and computing the gradient of the tangent line.

In practice, drawing a tangent line can be achieved through a limiting process whereby we draw an arc connecting two points separated by a distance δ in the x-axis, and then gradually letting δ get smaller and smaller.

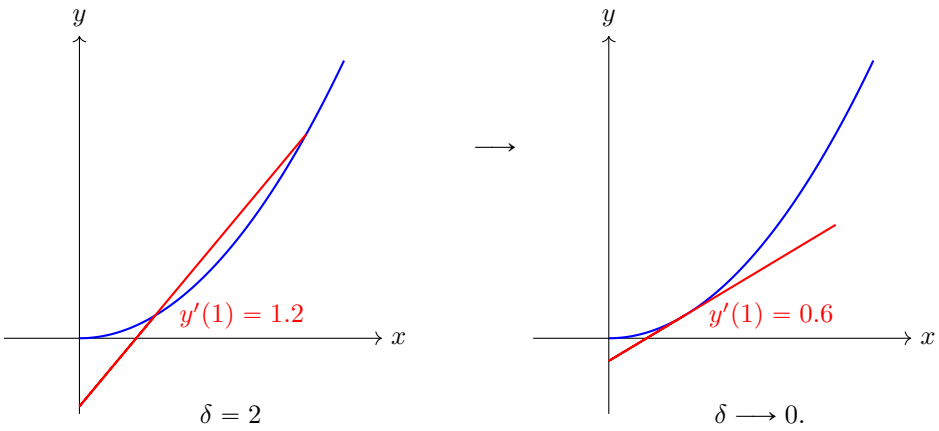

Figure 11: Drawing arcs across a decreasing distance $\delta \to 0$ will give a line with gradient equal to the tangent in the limit.

This limiting process is what we're going to use to construct the derivative. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Note that the gradient of the arc between x and $x + \delta$ can be expressed as

$$\text{gradient} = \frac{f(x+\delta) - f(x)}{(x+\delta) - x} = \frac{f(x+\delta) - f(x)}{\delta}.$$

Now we have the language to use limits, we can use one to construct the derivative of a function.

Definition 108.
Let $f : \mathbb{R} \to \mathbb{R}$ be a single-valued continuous function. The **derivative** of f with respect to x is given by

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\delta \to 0} \frac{f(x+\delta) - f(x)}{\delta},$$

provided this limit exists. If the limit exists, we say that f is **differentiable** at x.

Example 109.
The function $f(x) = x^2$ is differentiable at all $x \in \mathbb{R}$.

Proof. Firstly we have the expression

$$\frac{f(x+\delta) - f(x)}{\delta} = \frac{(x+\delta)^2 - x^2}{\delta} = \frac{2x\delta - \delta^2}{\delta}$$

Cancelling the factor and taking the limit gives

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\delta \to 0}(2x + \delta) = \lim_{\delta \to 0} 2x + \lim_{\delta \to 0} \delta = 2x.$$

□

Question 22.
Using the definition and limits, compute the derivative of $f(x) = x^n$.

Hard Question 23.
Using the definition and limits, show that

- $(f + g)' = f' + g'$ (the addition rule).
- $(fg)' = f'g + fg'$ (the product rule).

You might be able to think of some standard examples of functions which are not differentiable, such as $|x|$ or $\sin(1/x)$ as $x \to 0$. These cases are both interesting, but the vast majority of the points are still differentiable. Much harder to build is a function which is *nowhere differentiable* but still continuous - where regardless of where you look, there is never a derivative. One classic example of this is the Weierstrass function, which has exactly this property!

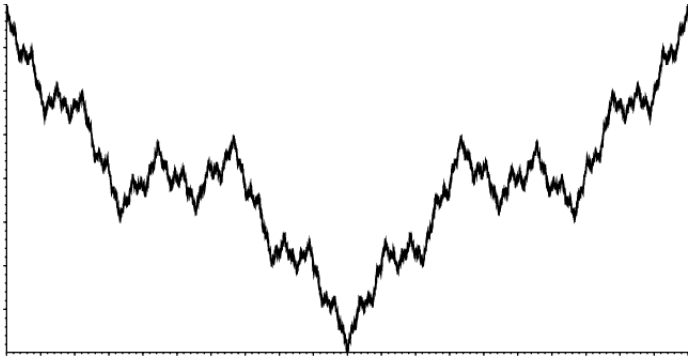

Figure 12: The Weierstrass function for $a = 2, b = 3$: a continuous, nowhere differentiable function. Credit: Siess Vincent.

6 Linear Algebra

Linear algebra is the study of linear systems and their behaviour. Often in mathematics many questions can be reduced to nice, linear, behaviour. When we say that a function is **linear**, we usually mean that the behaviour tends to follow straight lines or scales its inputs in some way. For example, the function

$$f : \mathbb{R} \to \mathbb{R} \quad f(x) = 2x$$

is a **linear** function, as it simply scales its input. However, the function

$$f : \mathbb{R} \to \mathbb{R} \quad f(x) = 6 + x^2$$

is **not linear**, as the transformation warps and bends the input.

Linear maps are, in fact, so well behaved and easy to work with that inside of algebra we have an entire body of knowledge called **linear algebra**, which concerns itself with studying these kinds of linear maps and their properties.

Since linear behaviour is so common, many seemingly complex problems can often be reduced to problems in linear algebra, where we can then draw on the large body of knowledge it provides.

6.1 Vector Spaces

We'd like to understand in what kinds of contexts we might be able to apply linear algebra and matrix algebra to solve a problem, so we're going to take some time now to explore the kinds of spaces in which we can apply the tools of linear algebra. These spaces are called **vector spaces** or **linear spaces**, and they appear in many unexpected places in mathematics.

Definition 110.
A **vector space** (or linear space) over $\mathbb{R}$ or $\mathbb{C}$ is a set V, whose elements are called **vectors** with two binary operations, called **vector addition** and **scalar multiplication**.

- **Vector addition** takes two vectors $\boldsymbol{v}, \boldsymbol{w} \in V$ and returns a new vector $\boldsymbol{v} + \boldsymbol{w} \in V$. This operation must have the following properties:
  - **Associativity**: $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$.
  - **Commutativity**: $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
  - **Identity**: There exists some $\mathbf{0}_V$ with $\mathbf{0}_V + \boldsymbol{v} = \boldsymbol{v} + \mathbf{0}_V = \boldsymbol{v}$.
  - **Inverses**: For each $\boldsymbol{v} \in V$, there exists some $-\boldsymbol{v} \in V$.
- **Scalar multiplication** takes one scalar $c \in \mathbb{R}$ (or $\mathbb{C}$) and one vector $\boldsymbol{v} \in V$, and returns a vector $c\boldsymbol{v} \in V$. This operation must have the following properties:
  - **Associativity**: $(ab)\boldsymbol{v} = a(b\boldsymbol{v})$.
  - **Identity**: There is an identity $1 \in \mathbb{R}, \mathbb{C}$ such that $1\boldsymbol{v} = \boldsymbol{v}$.
  - **Distributivity over scalars**: $(a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$.
  - **Distributivity over vectors**: $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$.

This definition seems exceedingly complicated, but the idea that it captures is quite intuitive (I promise!). In essence, we have an additive group of vectors with the ability to scale them up and down, and these two things play well with each other in a way you would expect.

Example 111.
Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors in $\mathbb{R}^2$, with

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}.$$

Then we have

$$\boldsymbol{a}+\boldsymbol{b}=\begin{bmatrix}4\\7\end{bmatrix},\quad \text{and}\quad 7\boldsymbol{a}=\begin{bmatrix}7\\14\end{bmatrix}.$$

Moreover, we can find the additive inverses of both $\boldsymbol{a}$ and $\boldsymbol{b}$ by just making all of the entries negative:

$$-\boldsymbol{a}=\begin{bmatrix}-1\\-2\end{bmatrix},\quad -\boldsymbol{b}=\begin{bmatrix}-3\\-5\end{bmatrix}.$$

Moreover we know that adding $\boldsymbol{a}$ to $-\boldsymbol{a}$ gives us

$$\boldsymbol{a}-\boldsymbol{a}=\begin{bmatrix}0\\0\end{bmatrix}=\boldsymbol{0_V},$$

the additive identity.

Note that while we've taken these vector spaces to have scalars in $\mathbb{R}$ or $\mathbb{C}$, in practice, many of the results we use here can be applied to more general sets of scalars. In fact, given a set $\mathbb{F}$ where we can add, subtract, multiply and divide in a sensible way, we can use $\mathbb{F}$ instead. Sets with this property are called **fields**, and they are a special class of **rings**, the next algebraic object you're likely to encounter at university after groups. We won't discuss rings or fields in this course, but they are some of the most fascinating objects that exist in algebra, and they form the basis of many powerful and interesting results.

6.2 Linear maps and matrices

We've now developed an idea behind the notion of a vector space, so now we might want to consider again those *linear* maps and functions between vector spaces.

More often than not, we'll want to use linear algebra to discuss mappings which might have many inputs and many outputs. For example,

$$f:\mathbb{R}^3\to\mathbb{R}^3\quad f(x,y,z)=(6x+2y+3z,x+y+z,7z).$$

This function takes three inputs and gives three outputs, where each of the arguments is a linear function of the input. You might notice though that writing out mappings in this way can get muddled and complicated very quickly, so we often use **matrices** instead.

Notation 112.
A **matrix** is a rectangular array of numbers. Often we denote the set of $m\times n$ matrices by $M_{m\times n}(\mathbb{R})$, where the $\mathbb{R}$ specifies that the entries are taken from $\mathbb{R}$. Here is one such matrix:

$$\begin{bmatrix}2 & 3 & 1\\ \pi & -4 & 1.2\\ \sqrt{3} & 1 & -42\end{bmatrix}$$

Note that in general, the first number m specifies the number of **rows** of the matrix, and the second number n specifies the number of **columns** of the matrix. This might be different to what you expect.

Notation 113.
We can use matrices to store information, but generally we would like that matrices can **interact** with each other. For that reason, we define a way to multiply two matrices (**matrix multiplication**).

Let $\boldsymbol{A}=(a_{ij})$ be an $l\times m$ matrix where $a_{ij}\in\mathbb{R}$. Similarly let $\boldsymbol{B}=(b_{ij})$ be an $m\times n$ matrix with $b_{ij}\in\mathbb{R}$. Then the $l\times n$ matrix $\boldsymbol{C}=(c_{ij})=\boldsymbol{AB}$ is given by

$$c_{ij}=\sum_{k=1}^{m}a_{ik}b_{kj}.$$

In practise, this can be seen to work as follows:

$$\begin{bmatrix} 1 & 2 & 3 \\ \dots & \dots & \dots \end{bmatrix} \cdot \begin{bmatrix} x & \dots \\ y & \dots \\ z & \dots \end{bmatrix} = \begin{bmatrix} 1x + 2y + 3z & \dots \\ \dots & \dots \end{bmatrix}.$$

I.e. we take the rows of the first matrix, take the dot product with the columns of the second matrix, and place the result in the relevant location in the output matrix. Note in particular that in this case $\boldsymbol{BA}$ is not even defined, so we we know that in general that matrix multiplication **is not commutative**.

Example 114.
Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be as above. I.e. $f(x, y, z) = (6x + 2y + 3z, x + y + z, 7z)$. Let

$$\boldsymbol{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Then we can express the behaviour of f using a matrix

$$f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 6 & 2 & 3 \\ 1 & 1 & 1 \\ 0 & 0 & 7 \end{bmatrix}.$$

Question 24.
Let $\boldsymbol{A_1}$ be an $i \times j$ matrix, and let $\boldsymbol{A_{10}}$ be a $k \times l$ matrix. If I have a string of matrices $\boldsymbol{A_1}, \dots, \boldsymbol{A_{10}}$, which I can multiply together, what will the dimensions of the output matrix be?

What is the shape of $\boldsymbol{A_1 A_2 A_3 A_4 A_5 A_6 A_7 A_8 A_9 A_{10}}$?

Definition 115.
The **identity matrix** $\boldsymbol{I_n}$ is the square matrix with 1 on the diagonal and 0 elsewhere. That is to say, $a_{ij} = 1$ if $i = j$ and $a_{ij} = 0$ if $i \neq j$.

For example, here is $\boldsymbol{I_4}$.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The identity matrix leaves elements on the left and right completely unchanged, provided they have matching dimensions. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}.$$

This behaviour sounds familiar, and you'd be correct in realising that the identity matrix $\boldsymbol{I_n}$ is in fact the identity element inside of given group (recall the lecture on abstract algebra!). We haven't defined this group yet, but we'll mention it after we've explored the **determinant** a little bit more in the next section.

Example 116.
Consider the system of simultaneous equations

$$10 = 2x + y$$
$$17 = 3x + 2y.$$

This can be re-expressed using matrices:

$$\begin{bmatrix} 10 \\ 17 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

What would be super useful would be if we could find a new matrix which could solve this system for us.

6.3 Determinants

If we have a linear map such as the one we introduced at the beginning of the lecture, you might have noticed that it is going to *stretch* the input to the output:

$$f(x) = 2x$$

This effect can also be seen for matrices which describe linear maps with multiple inputs and multiple outputs. Consider the following matrix:

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

If you consider what happens to the two points $(1, 0)$ and $(0, 1)$, you'll also notice that the **area** of the region that they trap is stretched by a factor of 4 when we multiply by the matrix.

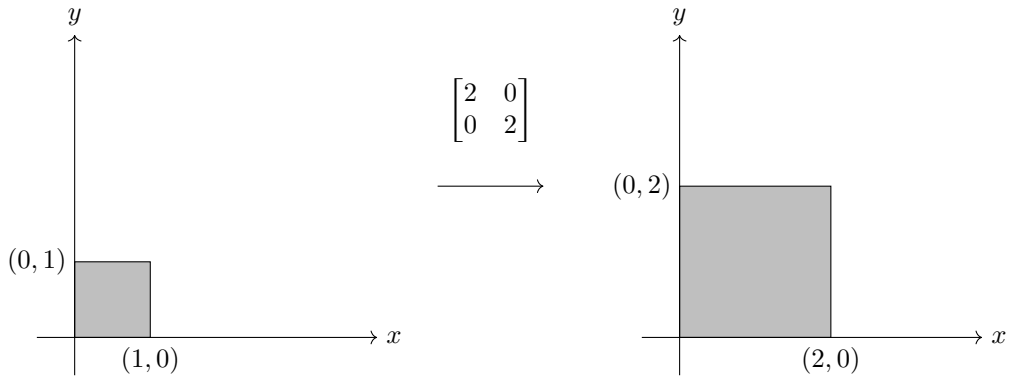

Figure 13: Applying the linear map scales the area here by a factor of 4.

We have a special name for the number which represents how much the area, volume, or space is scaled under the action of a linear map.

Definition 117.
Let M be a matrix representing a linear map $\mathbb{R}^n \to \mathbb{R}^m$ (note that M has shape $m \times n$. Then the **determinant**, $\det M$ or $|M|$, is the area of a unit volume (or area, etc.) under the action of the matrix.

Looking at figure 13, you can see that the total area of the unit square between $(0, 0), (0, 1), (1, 0)$ and $(1, 1)$ has been scaled up by a factor of 4 under the action of $\boldsymbol{M} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. This means that the determinant of the matrix $\boldsymbol{M}$ is 4.

We're only going to give the method of calculating the determinant for 2×2 matrices, as the method for computing the determinant for larger matrices gets a lot more involved, even though the interpretation above remains the same. We restate the formula without proof here.

Proposition 118.
Let $\boldsymbol{M}$ be a 2×2 matrix where

$$\boldsymbol{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then

$$\det(\boldsymbol{M}) = ad - bc.$$

Definition 119.
Let $\boldsymbol{M}$ be a matrix. Then $\boldsymbol{M}$ is **singular** if $\det(\boldsymbol{M}) = 0$.

Proposition 120.
Let $\boldsymbol{M}$ be a matrix with entries in $\mathbb{R}$ or $\mathbb{C}$. Then

$$\boldsymbol{M} \text{ is invertible } \iff \det(\boldsymbol{M}) \neq 0.$$

That is to say, there is some element $\boldsymbol{M}^{-1}$ such that

$$\boldsymbol{M}\boldsymbol{M}^{-1} = \boldsymbol{M}^{-1}\boldsymbol{M} = \boldsymbol{I},$$

the identity matrix, whenever $\boldsymbol{M}$ is non-singular.

This is an intriguing fact, but doesn't alone help us know how to calculate the inverse. For that purpose, we'll give one more result without proof which can help with calculations on 2×2 matrices.

Proposition 121.
Let $\boldsymbol{M}$ be a 2×2 matrix over $\mathbb{R}$ or $\mathbb{C}$ (or any field), with

$$\boldsymbol{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then

$$\boldsymbol{M}^{-1} = \frac{1}{\det \boldsymbol{M}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

6.4 The General Linear Group and Special Linear Group

6.4.1 The General Linear Group

If you recall our lecture on groups, you might be trying to work out what exactly the connection is to these non-singular matrices. We've just gone to the effort of explaining when exactly a matrix is invertible, so perhaps we should use these non-singular matrices to try and construct a group? We give a definition.

Definition 122.
The **General Linear Group** of $n \times n$ matrices over $\mathbb{R}$, $\mathrm{GL}_n(\mathbb{R})$ or $\mathrm{GL}(n)$, is the group consisting of all real non-singular square matrices of size $n \times n$, with the group operation given by matrix multiplication.

Question 25.
Using the definition of group from lecture 4, can you prove that GL_n is in fact a group?

As it happens, the general linear group is a good example of a non-abelian group. Recall that an abelian group is a group with a commutative group operation. In this instance, it is quite easy to see that matrix multiplication is not commutative. For example:

$$\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} = \begin{bmatrix}4 & 6\\3 & 4\end{bmatrix} \neq \begin{bmatrix}1 & 3\\3 & 7\end{bmatrix} = \begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}.$$

Because matrix multiplication is not commutative, we have to always be careful to make sure that we multiply on the right or on the left. Many of the most interesting groups are non-abelian groups, and, as it turns out, all finite groups can actually be represented as a group of matrices. This is one of the key results in **linear representation theory**, which studies mappings from group elements into vector spaces.

6.4.2 The Special Linear Group

As it turns out, given two linear maps $f : \mathbb{R}^x \to \mathbb{R}^y$ and $g : \mathbb{R}^y \to \mathbb{R}^z$ with corresponding matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ respectively, we can construct the linear map $g \circ f : \mathbb{R}^x \to \mathbb{R}^z$ by simply multiplying the two matrices together to get $\boldsymbol{GF}$.

What happens to the determinant of the corresponding map? As you might expect, if you stretch space with a scale factor of $\det(\boldsymbol{F})$ and then stretch space again with a scale factor of $\det(\boldsymbol{G})$, you'll have stretched the entire space by a factor of $\det(\boldsymbol{G})\det(\boldsymbol{F})$. That is to say, we have the following useful fact about determinants:

Proposition 123 (The determinant is multiplicative)**.**
Let $\boldsymbol{F}$ be an $k \times l$ matrix and let $\boldsymbol{G}$ be a $j \times k$ matrix. Then we have

$$\det(\boldsymbol{GF}) = \det(\boldsymbol{G}) \cdot \det(\boldsymbol{F}).$$

This property turns out to be immensely useful, and we're going to use it to define another group of interest.

Definition 124.
The **Special Linear Group** on $n \times n$ matrices over $\mathbb{R}$, $\mathrm{SL}_n(\mathbb{R})$ or $\mathrm{SL}(n)$, is the group given by the set of matrices

$$\{\boldsymbol{M} \in \mathrm{GL}_n(\mathbb{R}) : \det(\boldsymbol{M}) = 1\}$$

under matrix multiplication.

Question 26.
Can you prove that $\mathrm{SL}_n(\mathbb{R})$ is a group? [Hint: use the multiplicative property of det to demonstrate closure.]

Hard Question 27.
Using Bézouts identity, can you construct an element in $\mathrm{SL}_2(\mathbb{Z})$? This group and its generalisations are of great interest in number theory!

7 Calculus in Context

For this lecture, we're going to explore various topics in analysis and calculus. Some of the results we'll see are some of the most important and well known results in mathematics, and we'll finish the lecture by talking about the **Fundamental Theorem of Calculus**, which describes the relationship between differentiation and integration.

For lack of time we'll briefly illustrate how some of these results are proven, but we won't go into enormous detail.

7.1 Iterative Methods for Solving Equations

You might find often in mathematics that finding an explicit solution to an equation is often very difficult or impossible. One example, which we'll talking about again soon, is that any degree 5 polynomial doesn't have an explicit formula for its roots. That is to say, there is no explicit formula for solving the equation

$$a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

in general for $a_i \in \mathbb{R}$.

We're going to look at two ways of computing roots (zeros) to equations iteratively.

7.1.1 Creating Iterative Formulae

In some cases it is possible to take an equation of the form

$$f(x) = 0$$

and change it into an iterative expression in the form

$$g(x_n) = x_{n+1}.$$

Provided that you're fortunate enough to have a solution a such that $g(a) = a$ and $f(a) = 0$, then you may find that this expression will converge iteratively to a solution.

Example 125.
Consider the equation $x^2 - 2x - 1 = 0$. We can rewrite this expression as

$$x = \sqrt{1 + 2x} = g(x).$$

Starting with $x_1 = 1$, and rounding each time to 3 decimal places, we obtain $x_2 = 1.732$, $x_3 = 2.112$, $x_4 = 2.285$, $x_5 = 2.360$, $x_6 = 2.391$, $x_7 = 2.404$, $x_8 = 2.410$, $x_9 = 2.412$, and so on. As a sequence, $(x_n)_{n\in\mathbb{N}}$ converges to the real solution $1 + \sqrt{2}$.

Question 28.
Suppose that we have $\left|\frac{\mathrm{d}g}{\mathrm{d}x}\right| < 1$ for all $x \in \mathbb{R}$. Can you explain why this means that this method will converge to a solution?

Example 126.
For another example, consider the equation

$$x^5 + 4x^2 - 2x - 1 = 0$$

We can rewrite this expression to make a new iterative formula

$$g(x_n) = \sqrt[5]{1 + 2x_n - 4x_n^2} = x_{n+1}.$$

One tool you can use to explore these kinds of iterative formulae is **spider diagrams**. In this kind of a diagram, we plot $y = g(x)$ and the line $y = x$ and then draw a web by moving vertically from $y = x$ to $y = g(x)$ and then across again to $y = x$.

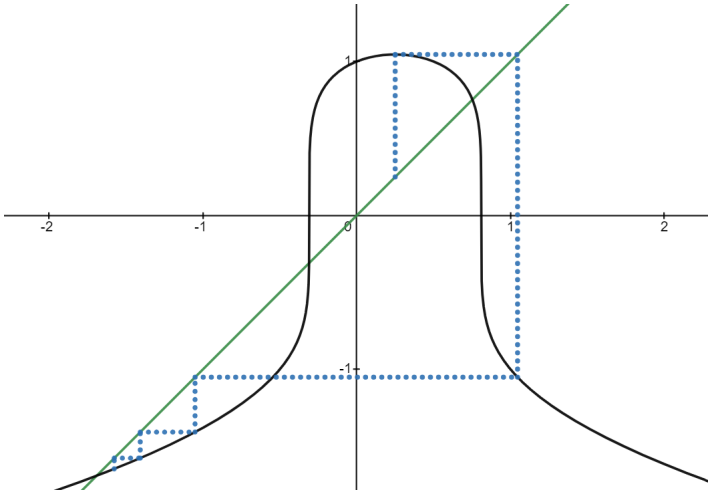

Figure 14: A spider diagram for $g(x)$ above with the starting point $x_1 = 0.25$.

7.1.2 The Newton-Raphson Method

You might be wondering if there is a guaranteed method for creating these kinds of iterative functions. As with all things in life that seem too good to be true - it is. There is, however, one very powerful method that you might have met before, called **Newton's method**, or the **Newton-Raphson method**, which does give us a more structured way of trying to produce these kinds of iterative formulae.

The Newton-Raphson method can only work on a differentiable function f provided that the starting point x_1 is chosen sufficiently carefully. To find the next value for x, we use the formula

$$x_{n+1} = x_n - \frac{f(x)}{f'(x)}.$$

In fact, if it will converge, the Newton-Raphson method will actually converge exceedingly quickly - certainly much quicker than some of the iterative formulae we saw in the first subsection.

The Newton-Raphson method can fail in many different ways, for example, if at any point $f'(x_n) = 0$, then the next value x_{n+1} will not even be defined. Worse still, it is quite possible that the values x_n enter an endless loop, never converging, or even blow up to be infinitely large.

Is it possible to find some conditions under which the Newton-Raphson method is definitely able to work? We state one possible approach to constraining the function and the starting value.

Proposition 127.
Given any differentiable function $f : \mathbb{R} \to \mathbb{R}$ with a continuous derivative and some root $x \in \mathbb{R}$ of f (i.e. $f(x) = 0$) where $f'(x) \neq 0$, there exists some $\varepsilon > 0$ such that for every starting value x_1 in the neighbourhood $\{a \in \mathbb{R} : |a - x| < \varepsilon\}$, the Newton-Raphson method will eventually converge on x.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Proving this proposition is considerably beyond the scope of what we'll cover in this course, but you might be able to see for yourself the intuition behind why it might be true. To see it, let's go back and look at the geometric behaviour of Newton's method.

Given some differentiable function $f : \mathbb{R} \to \mathbb{R}$, the equation of the tangent at a point $(a, f(a))$ is given by

$$y - f(a) = f'(a)(x - a)$$

Since we're searching for a root, it makes sense to jump to the place where this tangent meets the x-axis (where $y = 0$). Substituting this into the expression we obtain

$$-f(a) = f'(a)(x - a) \implies x = a - \frac{f(a)}{f'(a)}.$$

Since we expect the tangent line to get closer to the root, we can set $a = x_n$ and use the output $x_{n+1} = x$ to create an iterative formula which stands a good chance at getting nearer to the zero of f. Have a look at the diagram in figure 15 to see how this works.

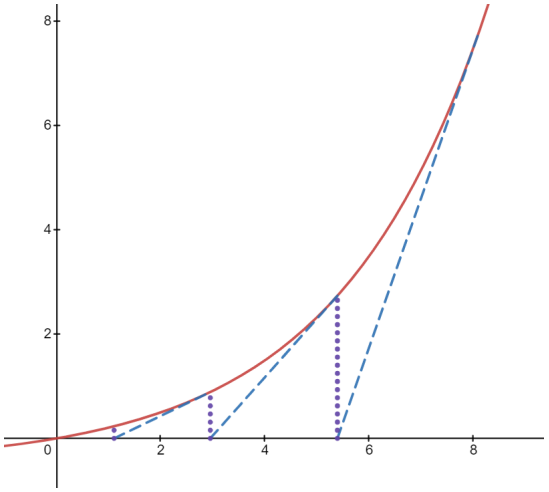

Figure 15: The Newton-Raphson method on $f(x) = \sqrt{2}^{(x-2)} - 1/2$, using the starting value $x_1 = 8$.

You might be able to see that if the function is sufficiently well behaved around the root point, then the derivative $f'(x)$ is going to be very similar to the derivative at neighbouring points $f'(x + \delta x)$. This means that in a small neighbourhood of the root the next point will get even closer to the root, and we can repeat the argument. If we were going to prove proposition 127, this would be one strategy for doing so. For now, we're going to talk a little bit more about some other important ideas in analysis and calculus!

7.2 The Completeness of $\mathbb{R}$

In the first lecture we talked about the *size* of different sets, and we saw later on that all of $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are all **countably infinite** sets. That is to say, all of these sets have elements which can be written in an infinite list.

This property is not true of the real numbers $\mathbb{R}$. In particular, you might have seen a version of **Cantor's diagonal argument** in one of the supervisions. This argument is one of the most straightforward ways of showing that the real numbers *cannot* be countable, and if you haven't seen it, it is definitely worth seeing.

The argument goes like this. Suppose we have a list of every real number $r_n \in \mathbb{R}$ where $n \in \mathbb{N}$. Then we can write these into a table. This table can be thought of as a function $f : \mathbb{N} \to \mathbb{R}$. The goal is to demonstrate that f cannot be surjective, i.e. there is at least one number which cannot be in the list.

For each element in the list, we select a digit in a unique position, and construct the new number in such a way that this digit must be different. For example, in the table above we have that the 1st digit of the 1st entry is 1. If we construct our new number to have a digit 2 at this location, then we know that the new number cannot be equal to the first entry.

We continue in this way. For example, in the 3rd row above, the third element in the list has digit 0 in the 3rd position. We can increase this by 1 and write it down to ensure that the new number we construct is distinct from the 3rd item. Continuing in this fashion for all digits and elements in the list, we create a new number which cannot already be included in the list.

	2	5.	1	3	9	3	4	4	$\cdots$
1	$\boxed{1}$	2.	8	6	4	2	3	1	$\cdots$
2	0	$\boxed{4.}$	0	0	1	0	0	1	$\cdots$
3	0	0.	$\boxed{0}$	7	2	2	3	3	$\cdots$
4	0	5.	1	$\boxed{2}$	3	4	9	8	$\cdots$
5	3	4.	8	8	$\boxed{8}$	8	8	8	$\cdots$
6	2	2.	1	1	1	$\boxed{2}$	2	7	$\cdots$
7	0	1.	0	9	2	8	$\boxed{3}$	7	$\cdots$
8	3	3.	3	3	3	3	3	$\boxed{3}$	$\cdots$
$\vdots$									

Figure 16: Cantor's argument on an infinite list.

Because of this, f cannot be surjective. Hence we must have that $|\mathbb{R}| > |\mathbb{N}|$ in some way. While this sounds incredibly counterintuitive, as they are both infinite sets, it demonstrates one peculiar mathematical fact - that infinite sets *can* have different sizes (the term **cardinality** is often preferred). So what's missing from $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$ to make them somehow 'much smaller' than $\mathbb{R}$?

One way of thinking about countably infinite sets is that they don't occupy space in the same way as an uncountable set. It's quite clear that $\mathbb{N}$ and $\mathbb{Z}$ leave large *gaps* to be filled, but this isn't immediately clear from looking at $\mathbb{Q}$. As it turns out though, $\mathbb{Q}$ does leave out many numbers (in fact, almost every number!). So how can we add these numbers back in?

We move from $\mathbb{Q}$ to $\mathbb{R}$ via a process called **completion**. To get a handle on the idea, consider the following sequence of numbers:

$$(a_n)_{n\in\mathbb{N}} \text{ where } a_n = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots + (-1)^n \frac{1}{2n+1}.$$

This is called the Madhava-Leibniz series. If we consider a similar sequence $(b_m)_{m\in\mathbb{N}}$ such that $b_m = a_{2m}$, then this gives us now a strictly increasing sequence of numbers:

$$b_m = \left(1 - \frac{1}{3}\right) + \left(\frac{1}{5} - \frac{1}{7}\right) + \left(\frac{1}{9} - \frac{1}{11}\right) + \cdots + \left(\frac{1}{2m+1} - \frac{1}{2m+3}\right).$$

If we let m increase indefinitely, what happens to b_m? Does it tend to a limit b? As it happens, it does, and, somewhat remarkably, we have that

$$\lim_{m\to\infty} b_m = \frac{\pi}{4}.$$

The constant π seems to appear from nowhere in this context, but something else rather odd has happened as well. At every step of the sequence b_m, we know that, adding all of the fractions together, b_m is still a rational number ($b_m \in \mathbb{Q}$). However, somehow in the limit, we appear to have reached a quantity which is irrational - as π itself is irrational.

This means that we had an infinite sequence $b_m \in \mathbb{Q}$ such that $b \notin \mathbb{Q}$. Here we will give one version of the notion of **completeness** which shall make the reason for this peculiarity clear, although several equivalent definitions exist.

Definition 128.
A totally ordered set S under $\leqslant$ is **complete** if every bounded subset $T \subseteq S$ has a least upper bound (also known as a **supremum**) in S. I.e. there exists some $a \in S$ such that $t \leqslant a$ for all $t \in T$.

Notation 129.
To denote the least upper bound (**supremum**) of a set S or sequence a_n, we can write

$$\sup_{n\in\mathbb{N}}(a_n) \quad \text{or} \quad \sup(S).$$

Similarly, to denote the greatest lower bound (**infimum**) of a set S or a sequence a_n, we can write

$$\inf_{n\in\mathbb{N}}(a_n) \quad \text{or} \quad \inf(S).$$

Since b_m is strictly increasing and tends towards $\frac{\pi}{4}$ in the limit, this means that $\frac{\pi}{4}$ is an upper bound for b_m. Hence the rational numbers $\mathbb{Q}$ cannot be complete, as this would imply that there is a smallest rational number larger than $\frac{\pi}{4}$.

The way we construct $\mathbb{R}$ is by *completing* $\mathbb{Q}$. We simply 'add back in' all of those missing numbers and allow them to be part of the set. This means that every real number $x \in \mathbb{R}$ can be thought of as being equivalent to some sequence of rational numbers $a_n \in \mathbb{Q}$, where x is a least upper bound of a_n.

Proposition 130.
Let S be the set of all countably infinite sequences in $\mathbb{Q}$. Let $\boldsymbol{x}, \boldsymbol{y} \in S$ be countably infinite sequences. Then the relation

$$x \sim y \iff \sup(\boldsymbol{x}) = \sup(\boldsymbol{y})$$

is an equivalence relation.

This notion actually demonstrates one way you can think about constructing the real numbers. In fact, we can be so explicit and say

$$\mathbb{R} = \{\boldsymbol{x} : \boldsymbol{x} \text{ is an equivalence class in } S \text{ under } \sim\}.$$

Completeness as a property gives us a lot of power for proving certain kinds of results that we might not have been able to prove before. We'll see some of them in the next subsection.

Question 29.
Prove the proposition: show that $\sim$ is an equivalence relation.

7.3 The Intermediate Value Theorem

7.3.1 Intermediate Values on $\mathbb{R}$

One of the most important theorems in mathematics is the intermediate value theorem. This theorem is so important that it is often simply known as the **IVT**. What the theorem says seems like a reasonably obvious assumption, but proving the statement actually requires the completeness of $\mathbb{R}$.

Theorem 131 (The Intermediate Value Theorem)**.**
Let f be a continuous real-valued function which is defined on some interval $[a, b] \subset \mathbb{R}$ (i.e. $\{x \in \mathbb{R} : a \leqslant x \leqslant b\}$). If $f(a) \leqslant f(b)$ then let $I = [f(a), f(b)]$. If $f(b) < f(a)$ then let $I = [f(b), f(a)]$.

Then for every $q \in I$ there exists some $p \in [a, b]$ such that $f(p) = q$.

We will not give the complete proof, as the proof is lengthy and quite hard to digest. Importantly, the proof requires the completeness of the real numbers. To see how it makes use of this property, we note that the set

$$S = \{x \in [a, b] : f(x) < q\}.$$

is bounded above by b, so by completeness it must have a least upper bound $p = \sup(S)$. The remainder of the proof is simply demonstrating that in fact $f(p) = q$.

Example 132.
Consider the function $f(x) = x^2 - 2x$. This is a continuous function, and it is well defined on all

of $\mathbb{R}$. In particular, the function is defined on the set $[-1, 1]$ and its output has the range $[-1, 3]$. Because of the intermediate value theorem, we know that there must exist some $x \in [-1, 1]$ such that $f(x) = 0$. That is, the function must have a root in this domain.

Question 30.
Consider the function $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2 - 2$. Compute $f(0)$ and $f(2)$ and hence deduce that there is a root (zero) in the range $[0, 2]$.

Consider the function $f : \mathbb{Q} \to \mathbb{Q}$ with $f(x) = x^2 - 2$ (i.e. the same function over the rational numbers). Explain why there is no root (zero) in the domain between $[0, 2]$.

7.4 Riemann Integration

In an earlier lecture on analysis we saw the formal definition of the derivative. Generally, the definition of the derivative is much more straightforward than the definition of the integral, and several different versions of integrability exist. The first you might meet is called **Riemann integration**. We won't have the time or space to fully develop this concept, but we'll give a brief overview about how this definition is constructed.

If you've met integrals before, you will probably know that they can be used to compute the area underneath a curve. If we were to attempt to do this without integration, it would be quite hard to do if the curve's function is not particularly simple. Riemann constructed a method using sums we now called **Riemann sums**, which successively estimate the area underneath a curve. In the limit, as we increase the number of *strips* in the approximation, the area underneath the curve should become exact.

Here is a parabola with some approximating *strips* as per Riemann's method. Here is a given curve $f(x)$ with 4 strips to approximate the area between 0 and 8:

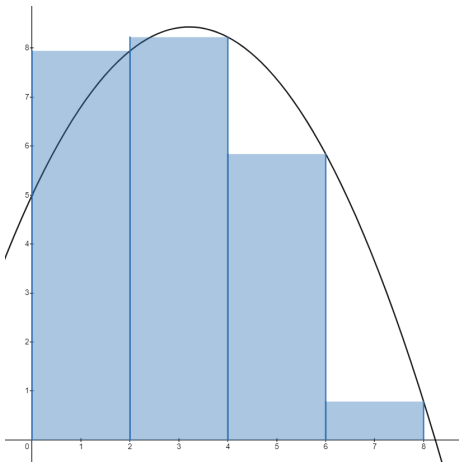

Figure 17: $f(x) = 5 + 2.14x - x^2/2$ with 4 strips.

Calculating the area in the figure corresponds to the sum:

$$S_4 = 15.89\dot{3} + 16.45\dot{3} + 11.680 + 1.57\dot{3} = 45.6$$

We can make the sum even more accurate by drawing 8 strips,
This corresponds to the sum

$$S_8 = 6.80\dot{6} + 7.94\dot{6} + 8.420 + 8.22\dot{6} + 7.36\dot{6} + 5.840 + 3.64\dot{6} + 0.78\dot{6} = 49.04$$

For n strips on a general function $f(x)$ between a and b, one possible Riemann sum is given by

$$S_n = \frac{b-a}{n} \sum_{k=1}^{n} f\left(a + \frac{b-a}{n} \cdot k\right)$$

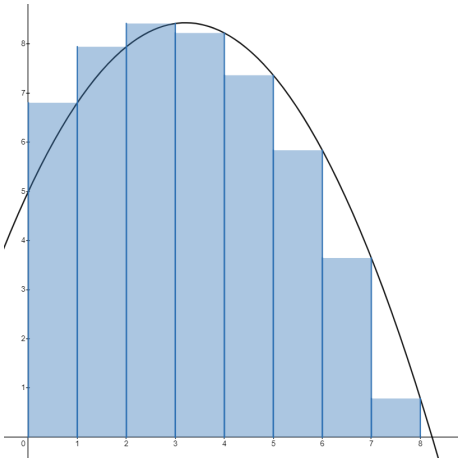

Figure 18: $f(x) = 5 + 2.14x - x^2/3$ with 8 strips.

where $(b-a)/n$ is the width of the strip, and we add up the function at n locations between a and b.

Definition 133.
Suppose that the sum S_n exists for all n. We say that $f(x)$ is **Riemann integrable** if the corresponding sum S_n converges as n tends to infinity, and write

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{n\to\infty} S_n.$$

Example 134.
For a simple example, consider

$$\begin{aligned}\int_0^{10} x\,\mathrm{d}x &= \lim_{n\to\infty} \frac{10}{n}\left(\sum_{k=1}^{n} \frac{10}{n}\cdot k\right)\\ &= \lim_{n\to\infty} \frac{100}{n^2}\left(\sum_{k=1}^{n} k\right)\\ &= \lim_{n\to\infty} \frac{100}{n^2}\cdot\frac{n(n+1)}{2}\\ &= \lim_{n\to\infty} 50\cdot\frac{n^2+n}{n^2} = 50.\end{aligned}$$

This definition doesn't match the one you might see at university exactly, but it gives a good idea of how we can begin define the integral. With the integral and the derivative in hand, we now have an understanding of the two fundamental operations of calculus. You will probably have learnt that the integral and the derivative are *opposites* of each other. The reason this idea is justified is made precise by the **fundamental theorem of calculus**, which we will talk about in the next section.

7.5 The Fundamental Theorem of Calculus

The fundamental theorem of calculus (or FTC) is the reason that all of calculus works so neatly together. It gives the basis on which we can rigorously think of derivatives and integrals (or antiderivatives, to give them a better name) as opposites. The FTC can be thought of as consisting of two parts, which we will state but not prove. The two parts we give here are taken from Apostol (1967).

Theorem 135 (First Part)**.**
Let $f : \mathbb{R} \to \mathbb{R}$ be a Riemann integrable function on $[a, x]$ for each x in $[a, b]$. Let c be such that $a \leqslant c \leqslant b$ and define a new function F:

$$F(x) = \int_c^x f(t)\,\mathrm{d}t \quad \text{for} \quad a \leqslant x \leqslant b.$$

Then the derivative $F'(x)$ exists at each point x in the open interval (a, b) where f is continuous, and for such x we have

$$F'(x) = f(x).$$

This part of the theorem essentially states that we can undo integration using differentiation. We now give the other notion.

Theorem 136 (Second Part)**.**
Let $f : \mathbb{R} \to \mathbb{R}$ be continuous on the closed interval $[a, b]$ and let F be a continuous function on $[a, b]$ which is an antiderivative of f in (a, b):

$$F'(x) = f(x).$$

Then if f is Riemann integrable on $[a, b]$ we have

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

These two theorems relate the behaviour of integration and differentiation to each other, and they form the basis for the rigorous study of calculus.

Thank you for joining us on this journey into analysis!

8 It Gets Weirder

Welcome to the last lecture of the Downing College supercurricular programme! Throughout this course we've done a lot of really hard mathematics and seen some pieces from all sorts of different areas. Some you'll have liked more than others. Some you'll find harder than others. Mathematics is hard, and that's the deal. In exchange for your hard work and your dedication, mathematics can reward you with beautiful patterns and intricate, understated elegance. But **you will get stuck** - that's part of being a mathematician.

The greatest mathematicians are the greatest challengers; they will get stuck, over and over, until the problems they face fall apart in their hands, and the golden kernel of truth that remains can be kept forever. It's a bit dramatic, yes, but the principle is true. Half of mathematics is the belief that you **can** find the solution; the other half is just thinking and time.

Today I want to show you all sorts of reasons why, regardless of what you study at university, mathematics should be a source of joy and intrigue. Applied carefully, weird and peculiar mathematics can pop up all over the place - from protein folding in biology, to colouring in maps, to tornadoes and the weather.

We're going to look at some examples of really unexpected and powerful ideas that maths is able to capture. I hope you enjoy this mathematical adventure!

8.1 Cartographic Colouring Book

Have you ever tried to colour in a map using different colours, so that touching countries always have a different colour? You might have drawn something like this:

Figure 19: A colouring of the world map.

But have you ever asked yourself how many colours you actually need to do this? Clearly 1 isn't enough, as then every country would be touching a like-coloured country. We can think of a pretty quick counterexample to 2 and 3 to show that this also isn't enough colours:

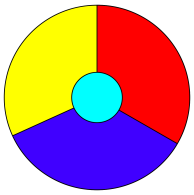

Figure 20: A map requiring at least 4 colours.

So we know that we definitely need at least 4 colours. As it turns out, mathematicians have been thinking about this question for a really, really long time, and it wasn't until recently that we managed to actually prove the following theorem:

Theorem 137 (Four Colour Theorem)**.**

Any map in a plane can be coloured using only 4 colours.

This theorem was finally proven in 1976 using computers by Appel and Haken. The proof reduced the problem to initially 1,936 different cases, which were then checked individually. The large number of cases made the 'proof' very long, and it took one single computer more than 1000 hours to check the cases at the time.

Needless to say, mathematicians were quite skeptical of using a computer to prove a theorem. The search for a human-readable version of the proof has been ongoing.

For higher surfaces like tori, we know some information about a lower bound for the number of colours necessary depending on a number called the **genus**.

Theorem 138 (Ringel-Youngs Theorem)**.**
For any surface of a given **genus** $g > 0$, the minimum number of colours necessary to colour a map is bounded below by

$$\gamma(g) = \left\lfloor \frac{7 + \sqrt{1 + 48g}}{2} \right\rfloor.$$

The only exception is the Klein bottle.

Here is a colouring of the world map using only 4 colours.

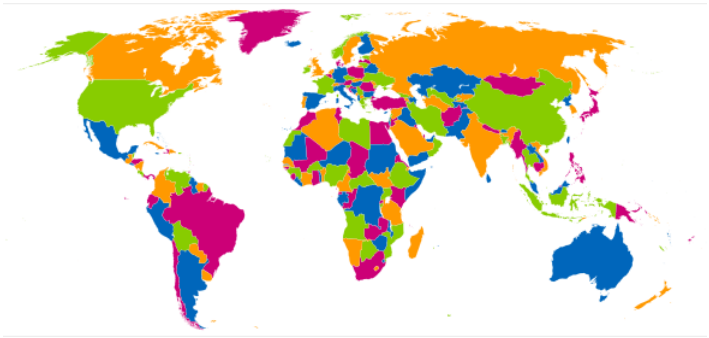

Figure 21: A colouring of the world map using only 4 colours.

8.2 Topological Meteorology

Topology is generally thought of as the study of shapes and spaces. In topology, the way we relate different objects is by continuous, smooth deformations, without pinching or breaking.

One famous topology joke is that coffee mugs and doughnuts are the same thing. This is because one can be continuously deformed into the other. Topologically they are both genus 1 surfaces (and hence need at least 7 colours for maps!).

Figure 22: How to turn a coffee mug into a doughtnut.

Because many different objects can be continuously deformed into a sphere, the sphere and its higher dimensional analogues are really important to the study of topology. In particular there are two theorems that are renowned for their remarkable consequences.

Theorem 139 (Hairy Ball Theorem)**.**
On the surface of a 2-dimensional sphere, there is no non-vanishing continuous tangent vector field. [Translation to English: if you try to comb a hairy ball, there will always be a tuft or a cowlick somewhere.]

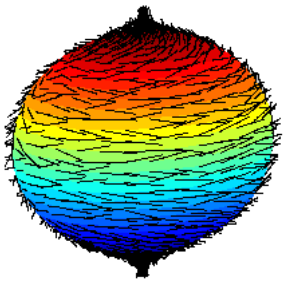

Figure 23: A hairy ball which could not be combed.

One interesting implication of this is meteorological. Looking at the wind on the surface of the earth, the direction of the wind can be thought of as being represented by the direction of a 'hair' at that point. The theorem then states there must always be a location which has zero wind, which can almost always be found at the *center of a cyclone* (a 'tuft' of wind).

Here's another really intriguing theorem:

Theorem 140 (Borsuk-Ulam Theorem)**.**
Every continuous function from the sphere to real space must have two antipodal points being sent to the same place. I.e. if $f(x) : S^n \to \mathbb{R}^n$ then there exists a point p and its antipode $\bar{p}$ such that $f(p) = f(\bar{p})$.

Normally we think of the 1-sphere S^1 as a circle, and S^2 as a sphere, but we can also generalise to higher dimensions. In the case of $n = 1$, the theorem states that given any loop around the earth (say, the equator), any continuous function must be the same at two antipodal points on this circle. For example, there are two points on the equator on opposite sides of the planet with the same **wind speed**, **temperature**, or **humidity**.

This is already a really cool result, but we can go further and apply Borsuk-Ulam to the whole of a sphere S^2. In this case we can view the function

$$f : S^2 \to \mathbb{R}^2 \quad f(p) = (\text{temperature at } p, \text{humidity at } p).$$

Then the theorem still guarantees there are two antipodal points on the planet where the temperature and humidity must be the same!

8.3 The Choice to... Mess Everything Up?

In the lecture on elementary number theory we saw Euclid's division lemma.

Lemma 141 (Euclid's Division Lemma)**.**
Let $m, n \in \mathbb{N}$. Then there exist unique $q, r \in \mathbb{N}$ such that $0 \leqslant r < n$ and

$$m = qn + r$$

In the proof, we had to make use of something called the **Well-Ordering Principle** which states that:

Every non-empty subset of the natural numbers has a least element.

Example 142.
This seems like a really obvious statement. For example, consider the set of increasing numbers

$$\{4, 7, 8, 9, 10, 11, \ldots\}.$$

The well-ordering principle simply states that 4 is the smallest thing in this set.

As it turns out, the well-ordering principle has some baggage attached. It's an incredibly useful assumption to make, but making this assumption leads you to some really inconvenient and difficult truths. So why is this statement so problematic?

If we go all the way back to the beginning of mathematics, most mathematicians make 9 fundamental assumptions. These are called the **Zermelo-Fraenkel** axioms, and they form a *foundation* of mathematics from which everything else, in theory, can be derived. Here are the axioms:

Axioms 143 (Zermelo-Fraenkel)**.**
The Zermelo-Fraenkel axioms are:

1. **The axiom schema of specification**
2. **The axiom schema of replacement**
3. **The axiom of extensionality**
4. **The axiom of regularity**
5. **The axiom of infinity**
6. **The axiom of pairing**
7. **The axiom of union**
8. **The axiom of power set**
9. **The well-ordering axiom**

They sound fancy, but most of them actually say very standard, comprehensible things in terms of sets and **quantifiers**. For example, the axioms of specification say that we can filter sets and select elements from them to make a new set.

The last of these axioms, the well-ordering axiom, is logically equivalent to another fundamental assumption called **the axiom of choice**. However this axiom is most peculiar, as using the axiom of choice gives us a way to prove some really unusual and unexpected results. The most famous of which is the Banach-Tarski paradox.

Theorem 144 (Banach-Tarski Paradox)**.**
Given a solid 3-dimensional ball, the points of the ball can be split up into a finite number of pieces (it can be done with 5) and re-arranged to produce two spheres of equal size to the first.

As it turns out, the theorem can also be used to demonstrate that a small object can be broken into finitely many pieces and put back together to make a larger object. This is sometimes presented as 'You can cut up a pea and put it back together to make the sun.'

Figure 24: A ball being broken up and pieced back together to construct two balls identical to the first.

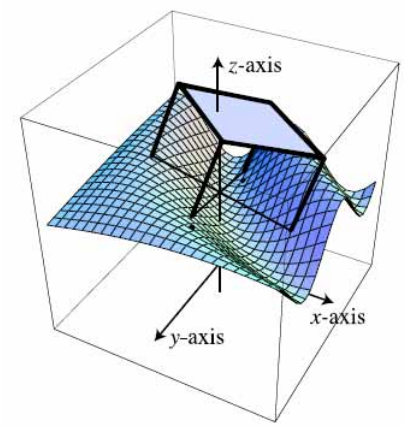

Figure 25: A wobbly table. Credit: Martin Gardner.

This means that our seemingly intuitive assumption that we can pick a least element of subsets of $\mathbb{N}$ has led us to a completely unintuitive result, all because it is logically equivalent to the axiom of choice.

$$\text{The Well-Ordering Principle} \iff \text{The Axiom of Choice}$$

8.4 Turning Tilty Tables

As a fun aside, we're going to take a quick look now at a fun result which is a little bit closer to your day-to-day life.

Theorem 145 (The Wobbly-Table Theorem)**.**
Suppose we have a wobbly, rectangular four-legged table, where the length of the legs is equal. Then we can stabilise the table by rotating it by less than $90°$.

Question 31.
See if you can explain why you can stabilise a table by rotating it by less than $90°$.

The theorem requires that the centre of the table stays roughly in the same location, but in reality you don't need to be too careful. This helpful tip will stop you from having to awkwardly find a piece of paper to stuff under one of the legs for the rest of your life. You're welcome!

8.5 Polynomial Problems

Earlier in the course we mentioned in passing that there is no equation to compute the general solutions of the fifth degree polynomial

$$a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

with each $a_i \in \mathbb{C}$. This is probably quite surprising, as every student leaves school knowing the equation for solving

$$ax^2 + bx + c = 0$$

with $a, b, c \in \mathbb{C}$. That is,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

As it turns out, there is also a formula for solving cubic equations and another one for quartic (degree 4) equations. Unfortunately, for some bizarre reason, the is no formula for computing the roots of the general 5th degree polynomial. It's not that we haven't tried or just haven't found it yet. In fact, it's **impossible** to have a formula.

One way to demonstrate this fact is using **Galois theory**, which uses the language of algebra, most notably groups, to explore questions about objects called **field extensions**. As it turns out, to every polynomial $p(x)$ we can assign a field extension, and then we can study the field extension to draw conclusions about the polynomial.

Importantly, for every field extension we can associate to it a special kind of **group** (like the ones we've seen). These groups are usually finite and are called **automorphism groups** or **Galois groups** in a certain context.

Galois theory is able to tell us that when this group has a special property, called **solubility**, then the corresponding polynomial must also be soluble in radicals. I.e. there is an algorithm to describe the roots in terms of addition, multiplication, subtraction, division, n-th roots and n-th powers.

One special class of groups are called the **symmetric groups**, and they mark all of the permutations (ways of swapping) of a given number of objects n. These permutation groups are incredibly important, as the Galois groups can all be thought of as containing permutations of the roots of the polynomial.

Question 32.
Let S_n be the group of all permutations of n objects (i.e. ways of mixing up n things). Then S_n is a group under composition of permutations.

What is $|S_n|$? How many permutations are there of n things?

Theorem 146 (Galois)**.**
S_5 is not a soluble group, and hence there is no general formula for solving a degree 5 equation in radicals.

Figure 26: Evariste Galois, aged 15.

The most remarkable thing about Galois is that he proved this remarkable result (that polynomials are soluble by radicals if and only if their automorphism groups are soluble) while he was still a teenager. The work that he did forms the basis of Galois theory and much of modern group theory.

Unfortunately, Galois died unforgivably young from injuries sustained in a duel. The night before, quite aware of the possibility of his death, he wrote most of his results down in a letter to his friend Chevalier. He died aged 20, having completely revolutionised modern algebra.

8.6 Confoundingly Complex: The Collatz Conjecture

The last topic we're going to briefly look at is called the **Collatz conjecture**. It is an enormous problem in mathematics, mostly because it is ridiculously simple to state, but nobody has managed to prove it.

Consider the function

$$f(x) = \begin{cases} \frac{x}{2} & x \text{ is even} \\ 3x+1 & x \text{ is odd} \end{cases}$$

Conjecture 147 (The Collatz Conjecture)**.**
Let $n \in \mathbb{N}$. Then there exists some $m \in \mathbb{N}$ such that

$$f^m(n) = 1.$$

That is to say, for every starting input n, iterating by applying f over and over again will eventually lead back to 1.

Question 33.
Iterate with the Collatz function starting with $n = 7$. How many steps does it take to get back to 1?

An alternative way of thinking about the same problem is to start at 1 and then work backwards. At each step you can ask 'which numbers could lead me here' and then go back in this fashion. If you do this, you can actually draw some really pretty 'Collatz Trees.'

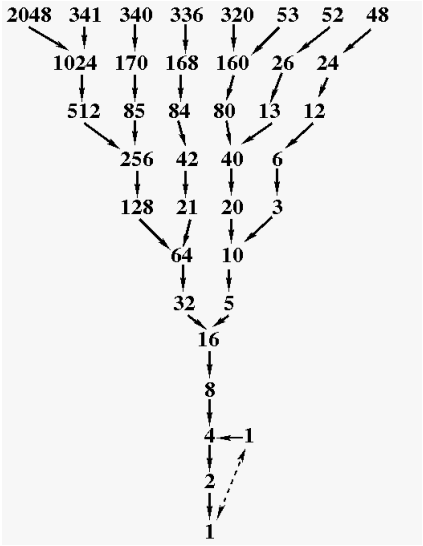

Figure 27: A Collatz Tree for a few different starting numbers.

The Collatz conjecture comes with a warning to new mathematicians. Many people try and prove the Collatz conjecture and, so far, all have failed. The internet is, annoyingly, full of strange

'proofs' which have the most deranged arguments. It's quite entertaining to look at these attempts for an hour or so, but after that it just gets quite depressing.

The closest we now have to a solution to this problem is the following, due to Terence Tao (2019).

Theorem 148 (Tao, 2019)**.**
Let $\mathrm{Col}_{\min}(n)$ be the smallest value obtained from a number n. In other words, let $\mathrm{Col}_{\min}(n) = \min\{f^m(n) : m \in \mathbb{N}\}$.

Suppose further that $g : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{n\to\infty} g(n) = +\infty$. Then

$$\mathrm{Col}_{\min}(n) < g(n)$$

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

As interesting as the Collatz conjecture is, it's probably best to avoid it in serious mathematical work. If we could prove the Collatz conjecture, however, the mathematics we would have to develop along the way would likely be far more interesting than the result itself. Often times in mathematics, the methods and techniques used to solve a problem are more valuable than its direct consequences.

Conclusion

I'm a little sorry to say we've reached the end of our mathematical journey together. After these two weeks, I really hope that you saw some good reasons that mathematics shouldn't just be studied - but cherished, enjoyed, and explored. Mathematics **is hard**, but, with enough patience and determination, you can train yourself to see all of the beautiful, elegant ideas it contains.

The two weeks we've spent together have been a wild mathematical romp through concepts in a first-year mathematics course. The ideas and skills you've picked up will help you in making the transition to studying this challenging and rewarding subject - be it for computer science, psychology, engineering, or just for its own sake. Along the way we saw everything from sets to proof, numbers to groups, sequences to matrices and back again. I hope you saw unexpected connections between these different concepts. Mathematics is full of secret tunnels.

Mathematics at university looks very different to mathematics in school up to this point. The pace of this course has been fast, and we've covered a lot of new material in a short time. You should expect that mathematics and similar subjects will be fast-paced and intense in a similar way at the university level. You will see a lot of really interesting and remarkable ideas, but you'll often have to put the effort in to see their beauty. Often at a university level you won't be given many examples, so you might have to search for your own examples. As my old lecturer used to say to me: '**mathematics is not a spectator sport**.' Do lots of practice, and always search for the **why** behind a definition or a result. Asking this question will always lead you to fresh green pastures of mathematical knowledge!

I hope you've enjoyed this two week course into mathematics and that it gave you a new perspective on a very old subject. While our party has disbanded for this leg of the journey, I leave it to you to continue to lead your own mathematical quest.

If you have any comments, questions or suggestions on this course please let me know at my email:

kjad2@cam.ac.uk

I've had loads of fun teaching you all, and good luck on your mathematical adventures!
Keenan