The University of Melbourne
School of Mathematics and Statistics

MAST20026
Real Analysis

Semester 2, 2018

Partial Lecture Notes

Some things to be aware of:

- These notes are not intended as a textbook. Rather they are an accompaniment to the lectures. You will need to take notes during the lectures.
- This is not an online course! You are expected to attend lectures.
- You are expected also to utilise textbooks from the library (or elsewhere).

Main Topics

Introduction

In your previous studies (e.g. MAST10007 Linear Algebra) you will have seen some simple proofs, and constructed some yourselves. For example, showing that a particular set is (or is not) a vector subspace of a bigger vector space.

This subject is partly about developing your skills in proof writing rather than about calculation! To do this we need to have a thorough understanding of how mathematical arguments work. This involves starting with logic and the way it is used in mathematics.

This subject has two main aims:

- To study how to write mathematical proofs
- To study Real Analysis (convergence, continuity, differentiation, integration)

Along the way we will be investigating questions like:

- What is a real number?
- We need a definition of real number that includes numbers like $\sqrt{2}$. Is it a rational number?
- How do we know that $x^2 - 2 = 0$ has a real solution, but $x^2 + 1 = 0$ does not?
- Is 'infinity' a real number?
- What does the following mean? Does it give a real number?

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots\cdots$$

- Does the following make sense?

$$1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \cdots\cdots$$

To answer these questions we need a *mathematical definition* of a real number.

We need a mathematical definition of what things such as

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots\cdots$$

mean.

The considerations above demonstrate the need for:

Definitions: We need *precise* definitions of the objects we are working with.

Proofs: We need to be able to *prove* statements based on the definitions as well as statements based on axioms (axioms are statements assumed to be true).

Logic and Proof

We want

1. A precise way to write mathematical definitions.
2. A precise notion of what constitutes a mathematical proof.
3. A precise notion of what constitutes a theorem.
4. Some methods of proof.

We will consider two forms of logic

1. *Propositional logic*
2. *First order logic*

Propositional Logic

Logic is about how to *infer* true things from known or assumed things.

Propositional logic is about *two valued variables* and *two valued functions* of those variables. These two values will be called “true” and “false” or sometimes 0 and 1.

Definition 2.1 (Statement)

A **statement** is a “sentence” or “expression” that is either *true* or *false*.

The *statement* takes on the role of a logical variable (the variables are **true** or **false**). We will generally use lower case letters $p, q, r, \ldots$ to represent statements.

Some books use the word **proposition** rather than *statement*.

Examples

Example 2.1 (Which of these are statements?)

- All maths lecturers are named Jesse.
- $2 + 3 = 7$
- $1 + 1 = 2$
- potato
- $x > 2$
- For all $x \in \mathbb{Z}$, $x^2 \geq 0$
- Every even number greater than 2 is the sum of two primes.

Connectives

We can combine statements to give new statements. Such statements are called *compound statements* (or logical formulae). They are constructed from:

1. primitive statements (that is, logical variables)
2. connectives
3. parentheses (to remove ambiguity)

Generally, we use upper case letters $A, B, C, \ldots$ to denote compound statements.

The connectives we will consider are:

- negation
- conjunction
- equivalence
- disjunction
- implication

Definition 2.2 (Negation)

If p is a statement, the **negation** of p is the statement "not p" and is denoted by $\sim p$. Its logical value is the opposite of p (ie., if p is true then $\sim p$ is false).
It can also be denoted by $\sim p$.

p	$\sim p$
T	F
F	T

Example 2.2

If p is the statement "the grass is green" then $\sim p$ is the statement "not [the case that] the grass is green". (We would usually say "the grass is not green".)

Example 2.3

If p is the statement "$5 > 0$" then $\sim p$ is the statement "5 is not > 0" which is the same as "$5 \leq 0$".

Example 2.4

If q is the statement "$1+1=7$" then "$\sim q$" is the statement "$1+1 \neq 7$".

In the English language, there are two different usages of the word *'or'*.

The *exclusive* use of "or": *I will ride my bike or catch the train.*

usually means that one or the other of these modes of transportation will be used, but not both. This is **not** the mathematical usage of "or".

Definition 2.3 (Disjunction)

If p and q are statements, the **disjunction** of p and q is the statement "p or q" and is denoted by $p \vee q$.

It is True if one or both of p and q are True.
It is False if both p and q are False.

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

The disjunction is an *inclusive* use of "or", that is; it allows for the possibility that both statements are true.

Example 2.5

If p is the statement "$2+5=7$" and q is the statement "$1+1=2$", then $p \vee q$ is the statement "$2+5=7$ or $1+1=2$".

Example 2.6

If p is the statement "$2+5=7$" and q is the statement "$1+1=0$", then $p \vee q$ is the statement "$2+5=7$ or $1+1=0$".

When defining the above connectives we gave their *Truth Tables*.

p	$\sim p$
T	
F	

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

We will use truth tables to define some other connectives.

Definition 2.4 (Conjunction)

If p and q are statements, the **conjunction** of p and q is the statement "p and q", denoted $p \wedge q$.

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

Conjunction accords with the natural language use of the word "and".

Example 2.7

If p is the statement "$2+5=7$" and q is the statement "$1+1=2$" then the conjunction of p and q is the statement "$2+5=7$ and $1+1=2$".

Example 2.8

If p is the statement "$2+5=7$" and q is the statement "$1+1=0$" then the conjunction of p and q is the statement "$2+5=7$ and $1+1=0$".

Example 2.9

Note that $p \wedge q$ has the same truth table as $q \wedge p$, and $p \vee q$ has the same truth table as $q \vee p$; that is, conjunction and disjunction are commutative.

Definition 2.5 (Implication)

A statement of the form "if p then q" is called an **implication** (or a **conditional** statement).
It is written as $p \Rightarrow q$.

p	q	$p \Rightarrow q$
T	T	
T	F	
F	T	
F	F	

The statements p and q have truth values that are independent of each other.

The statement p is called the **antecedent** and the statement q is called the **consequent**.

Example 2.10

$p : 1 + 1 = 2$
$q : 2^2 = 4$
$r : \pi = 3$

$p \Rightarrow q$ is True
$p \Rightarrow r$ is
$r \Rightarrow p$ is True

There are many natural language equivalent statements:

- If p then q
- p only if q
- q follows from p
- q is necessary for p
- q provided that p
- q is a necessary condition for p
- there are others.........

Example 2.11

p	q	$p \Rightarrow q$	$\sim p$	$(\sim p) \vee q$

Definition 2.6 (equivalence)

A statement of the form "p if and only if q" is called an **equivalence** (or a **biconditional** statement).
It is written as $p \Leftrightarrow q$.

p	q	$p \Leftrightarrow q$

Note that p if and only if q is the same as (ie., "logically equivalent to") $(p \Rightarrow q) \wedge (q \Rightarrow p)$.

Definition 2.7 (Compound statement)

Statements constructed from several primitives are called **compound** statements.
Usually they are written with capital letters, $P, Q, R, \ldots$.

A compound statement is also either true or false. Another way to look at it is, if you combine statements with connectives in a valid way then you have a compound statement which is either true or false, and its truth value depends on the truth values of the statements comprising it.

Example 2.12

Let p, q, r be primitive statements, then

- $p \wedge q$
- $(\sim p) \wedge (\sim q)$
- $(p \Rightarrow q) \wedge (q \Rightarrow p)$

are all *compound* statements.

Negating Compound Statements

Example 2.13

Since $p \wedge q$ is a statement, we can negate it, giving, $\sim (p \wedge q)$. The truth table is

p	q	$p \wedge q$	$\sim (p \wedge q)$

Be careful with your use of brackets. They should be used to remove any ambiguity.
If I write $\sim p \wedge q$ do I mean $(\sim p) \wedge q$ or $\sim (p \wedge q)$? They are not the same.

Example 2.14

Now consider the truth table for $(\sim p) \vee (\sim q)$

p	q	$\sim p$	$\sim q$	$(\sim p) \vee (\sim q)$

Notice that $(\sim p) \vee (\sim q)$ always gives the same truth value as $\sim (p \wedge q)$

Logical Equivalence

Definition 2.8 (Logically equivalent)

Statements which have the same truth table are said to be **logically equivalent**.
That is, p and q are logically equivalent if the statement $p \Leftrightarrow q$ is a tautology.
It is denoted $p \equiv q$.

1. $p \wedge q \equiv q \wedge p$
2. $p \vee q \equiv q \vee p$
3. $\sim (p \wedge q) \equiv (\sim p) \vee (\sim q)$
4. $\sim (p \vee q) \equiv (\sim p) \wedge (\sim q)$

Example 2.15

We show number 4 from above

p	q	$p \vee q$	$\sim (p \vee q)$	$\sim p$	$\sim q$	$(\sim p) \wedge (\sim q)$	$\sim (p \vee q) \Leftrightarrow (\sim p) \wedge (\sim q)$

First Order Logic

Mathematical theories have several ingredients that propositional logic cannot easily represent (for example, sets, functions and mathematical operations).

We use *mathematical variables* $x, y, z, \ldots$ to represent elements of the set of objects $\mathcal{U}$ (called the *universal set*).

We also have *equations* and *inequalities* containing variables:

- $x > 0$
- $x^2 - 3x + 2 = 0$.

These two formulae are not statements. They are neither true nor false, because their truth depends on the value of x.

Formulae which depend on (free) variables are called **conditions** (or **predicates**)

Conditions on x are written as $p(x), q(x), \ldots$.

Example 2.16

Some conditions on x are:

- $p(x) : x > 0$
- $q(x) : x^2 - 3x + 2 = 0$.

They are neither True nor False.

Some statements are:

- $p(2) : 2 > 0$
- $p(-5) : -5 > 0$
- $q(3) : 3^2 - 3^2 + 2 = 0$.

They can be seen as True or False.

If x is replaced by a particular value (or element of a set), then the formulae become statements.

There are (at least) two ways to convert formulae containing mathematical variables into logical statements.

1. Substitution

Pick any element of $\mathcal{U}$ and substitute into the formula:

Thus, if $\mathcal{U}$ is $\mathbb{Z}$, then

- $(x = 2)$
- $(x = -5)$
- $(x = 3)$

2. Using quantifiers

Thus, if $\mathcal{U}$ is still $\mathbb{Z}$, then

- for all $x \in \mathbb{Z}, x > 0$
- for all $x \in \mathbb{Z}, x^2 - 3x + 2 = 0$
- there exists an $x \in \mathbb{Z}$ such that $x > 0$
- there exists an $x \in \mathbb{Z}$ such that $x^2 - 3x + 2 = 0$

The phrases "for all" and "there exists" are quantifiers.

Expressions like $p(x) \wedge q(x)$ and $p(x) \Rightarrow q(x)$ are compound conditions (or predicates).

Example 2.17

- $p(x) \wedge q(x)$ is the condition "$p(x)$ and $q(x)$".
- $p(x) \vee q(x)$ is the condition "$p(x)$ or $q(x)$".
- $p(x) \Rightarrow q(x)$ is the condition "if $p(x)$ then $q(x)$".

These all become valid logic statements when a particular value for x is substituted.

Quantifiers

As we have seen, we can turn conditions on x into statements.

Consider the condition $p(x) : x^2 \geq 3$.

We can turn this into a statement in various ways:

Definition 2.9 (Universal Quantifier)

The phrase "for all" is called the **universal quantifier** and is written as $\forall$.

Now consider the condition $p(x) : x^2 - 1 = 0$.

Definition 2.10 (Existential Quantifier)

The phrase "there exists" is called the **Existential Quantifier** and is written as $\exists$.

Proving and Disproving Quantified Statements

If Q is the statement "$\forall x \in D \quad p(x)$", where D is a set and $p(x)$ is a condition on x, then Q is true only if $p(x)$ is be true for every member of D (this might be difficult to show).

However if at least one instance is false (say $p(c)$) then $p(a) \wedge p(b) \wedge p(c) \cdots$ is false.

So Q can be disproved if at least one false case can be found.

This is called a **counterexample**.

Example 2.18

$\forall x \in \mathbb{R}, \quad (x-1)^2 > 0$

Similarly, if R is the statement "$\exists x \in D \quad p(x)$", then R is true if there is one true instance of $p(x)$.

So R can be proved true if at least one primitive statement is true.

This is called an **example**.

Example 2.19

$\exists x \in \mathbb{R} \quad (x-1)^2 > 0$

In general

- Disprove a universal statement with a counterexample.
- Prove an existential statement with an example.

Writing theorems using First Order Logic

We can use First Order Logic to write theorems in a precise way.

1. Any real number has the property that its square is not negative.

2. For every even integer m, $3m^5$ will also be an even integer.

3. If x and y are real numbers, the absolute value of their sum is no greater than the sum of their absolute values.

4. $\sqrt{2}$ is irrational.

5. If n is an integer greater than or equal to 5, then 2^n is greater than n^2.

Negation of $\forall$

Since quantified conditions are statements their negation is well defined. For example:

$$\sim[\ \forall x \in D \quad p(x)\ ] \quad \text{and} \quad \sim[\ \exists x \in D \quad p(x)\ ]$$

Are these logically equivalent to anything else?

Theorem 2.1 (Negation of $\forall$)

$$\sim(\ \forall x \in D \quad p(x)) \equiv \ \exists x \in D \quad \sim p(x)$$

Negation of $\exists$

Theorem 2.2 (Negation of $\exists$)

$$\sim(\ \exists x \in D \quad p(x)\ ) \equiv \ \forall x \in D \quad \sim p(x)$$

Example 2.20

No positive real number satisfies $x^3 = -1$.

This is the same as:

Tautologies and Contradictions

Definition 2.11 (Tautology)

If the truth table of a compound statement $P(p_1, \ldots, p_n)$ is true for all values of $p_1, \ldots, p_n$ then it is called a **tautology**.

Example 2.21

p	$\sim p$	$(\sim p) \vee p$

Thus $(\sim p) \vee p$ is a tautology.

Example 2.22

$P(p, q) : p \Rightarrow p \vee q$

p	q	$p \vee q$	$p \Rightarrow (p \vee q)$

Thus P is a tautology.

Example 2.23

Show $[p \wedge (p \Rightarrow q)] \Rightarrow q$ is a tautology.

p	q	$p \Rightarrow q$	$p \wedge (p \Rightarrow q)$	$[p \wedge (p \Rightarrow q)] \Rightarrow q$

Definition 2.12 (Contradiction)

If the truth table of a compound statement $P(p_1, \ldots, p_n)$ is false for all values of $p_1, \ldots, p_n$ then it is called a **contradiction**.

Example 2.24

$P(p) : (\sim p) \wedge p$

p	$\sim p$	$(\sim p) \wedge p$

Thus P is a contradiction.

Note: There is a second type of contradiction associated with inference rules — see later.

Theorem 2.3

If p and q are statements and c is a statement that is always false then $(p \Rightarrow q) \equiv [(p \wedge \sim q) \Rightarrow c]$.

Proof.

p	q	$\sim q$	$p \wedge \sim q$	c	$(p \wedge \sim q) \Rightarrow c$

□

Converse and Contrapositive

Definition 2.13 (Converse)

The **converse** of $p \Rightarrow q$ is the statement $q \Rightarrow p$.

A conditional statement **is not** logically equivalent to its converse.

Example 2.25

p	q	$p \Rightarrow q$	$q \Rightarrow p$

Definition 2.14 (Contrapositive)

The **contrapositive** of $p \Rightarrow q$ is the statement $(\sim q) \Rightarrow (\sim p)$.

(ie., switch *and* negate p and q)

The following theorem says that the contrapositive of a conditional statement *is* logically equivalent to the original statement. This is very useful!

Theorem 2.4 (Conditional and Contrapositive)

If p and q are statements, then

$$p \Rightarrow q \equiv (\sim q) \Rightarrow (\sim p)$$

Proof.

p	q	$\sim p$	$\sim q$	$p \Rightarrow q$	$(\sim q) \Rightarrow (\sim p)$

□

This means that a conditional statement is logically equivalent to its contrapositive.

The following is a very useful form of the biconditional.

Theorem 2.5 (Biconditional and Conditional)

If p and q are statements, then

$$[p \Leftrightarrow q] \equiv [(p \Rightarrow q) \wedge (q \Rightarrow p)].$$

Proof.

p	q	$p \Leftrightarrow q$	$p \Rightarrow q$	$q \Rightarrow p$	$(p \Rightarrow q) \wedge (q \Rightarrow p)$

□

Columns 3 and 6 are the same thus the statements are logically equivalent.

For proof writing if we need to show $p \Leftrightarrow q$ is true, then we can use Theorem 2.5.

In general we show:

1. $p \Rightarrow q$ is true, and
2. $q \Rightarrow p$ is true (that is, the converse is true).

Thus $(p \Rightarrow q) \wedge (q \Rightarrow p)$ is true ($T - T$ row of truth table).

Thus $p \Leftrightarrow q$ is true.

Conversely, if we know $p \Leftrightarrow q$ is true then Theorem 2.5 tells us that $p \Rightarrow q$ is true and that $q \Rightarrow p$ is true

Example 2.26

An even number x is a number that can be written in the form $x = 2k,\ k \in \mathbb{Z}$.

Written as a definition:

$x \in \mathbb{Z}$ is even $\Leftrightarrow$ $\exists k \in \mathbb{Z}$ such that $x = 2k$.

Logic, Theorems and Inference

Consider the following:

Theorem 2.6

If x is an even integer, then x^2 is an even integer.

Let $p : x$ is an even integer and $q : x^2$ is an even integer.

The theorem states,

"If p then q" is true;

that is, the *conditional*

"$p \Rightarrow q$"

is true.

Using theorems: When we use the theorem we usually take p as true (ie., choose an even integer) and then use the theorem to infer q is true (ie., the square of an even integer is even).

This is justified by the truth table for the conditional:

p	q	$p \Rightarrow q$

Assuming p is true *and* having the theorem true puts us on row 1 and hence q must be true.

Thus we get

$$p \text{ true and } (p \Rightarrow q) \text{ true allows us to infer } q$$

In logic this is usually written

$$\{p, p \Rightarrow q\} \ \vDash \ q$$

where $\vDash$ is read "infers".

This is a famous "rule of inference" called **Modus Ponens**[1] or the **Law of Detachment** - q has been detached from $p \Rightarrow q$. It is valid as long as p and $(p \Rightarrow q)$ are assumed (or known) to be true.

[1] From Latin: "The method that affirms the consequent by affirming the antecedent"

Idea of Inference: By assuming some statements are true, and thus restricting rows of the truth table, rules of inference enable us conclude other statement are true.

To paraphrase:

From true statements we infer new true statements

Rules of inference are *theorems in logic* that allow us to infer that a given statement is true if a set of statements are all true.

The big advantage of inference rules is that, once they are proved, they can be used to infer new true statements *without directly using truth tables*.

Inference Contradiction

Definition 2.15 (Inference Contradiction)

If the set of statements $\{P_1, P_2, \ldots, P_n\}$ are never all true then $\{P_1, P_2, \ldots, P_n\}$ is called a **contradiction**.

That is, there is *no* row of the truth table where all the $P_1, P_2, \ldots, P_n$ are true.

Note: This is weaker than the propositional form of contradiction as it does not require a whole column to be false (but is equivalent to $P_1 \wedge P_2 \wedge \cdots \wedge P_n$ always being false).

Example: $\{p \wedge q, p \Rightarrow \sim p\}$

Primary Rules of Inference for Propositional Logic

There are many classical rules of inference. However just as there are two primary connectives in propositional logic, there are four primary rules of inference - all others can be proved (without necessarily using truth tables) from these.

Definition 2.16 (Primary Rules of Inference)

The primary Rules of Inference of propositional logic are:

1. **Contraction Rule** (infer A from $A \vee A$)
2. **Expansion Rule** (infer $B \vee A$ from A)
3. **Associative Rule** (infer $(A \vee B) \vee C$ from $A \vee (B \vee C)$)
4. **Cut Rule** (infer $B \vee C$ from $(A \vee B) \wedge (\sim A \vee C)$)

We can show that these rules are valid using truth tables.

Example 2.27

Use Primary Rules of Inference to establish Modus Ponens.

Proofs in Mathematics

Modus Ponens is a theorem in logic and the above example shows how to prove it is true.

What about mathematics theorems and proofs?

The Modus Ponens proof gives us a clue as to how to write proofs in mathematics.

The pattern of the Modus Ponens proof is:

1. Start with some statements *assumed* true
2. Infer a sequence of true statements that logically follow (ie., using rules of inference, logical equivalence, tautologies etc.) from *previous* statements
3. Arrive at true statement

Lets apply this to a simple mathematics theorem.

Example 2.28

Prove that the sum of two odd numbers is an even number.

First let's write this in a form where the logic is clear.

If (m, n are odd integers), then ($m + n$ is an even integer).

Assumed true

Inferred steps

Conclusion

Notice the big difference with the Modus Ponens proof:

The inferred steps were not *explicitly* justified using rules of inference.

Why?

Lets see what happens if you do put in *all* the inference steps.

Example 2.29

If $a, b, c \in \mathbb{N}$ then $(a + b) + c = (b + c) + a$.

Note: This level of detail of proof is *not* required for this course!

Reason: If you do insist on justifying every line with a rule of inference the proof would be dozens of slides long, take hours (days?) to write and completely obscure the ideas used in the proof.

But: Such a full inference proof does exist for example 2.28 and the important true statements of the full inference proof are precisely the statements in the short proof.

Conclusion:

In mathematics proofs are written in an abbreviated form which summarises the important statements in such a way that is clear to the reader that true statements have been inferred from previous true statements.

"Clear to the reader" means the reader accepts that a full first order logic proof *could* be written (ie., where every inferred step is justified by a rule of inference) but, in the interests of brevity and clarity, has not been so written.

This idea of theorems and proofs is due to the mathematician David Hilbert:

> *Every theorem in mathematics could, in principle, be proved starting from a set of assumed true statements (called axioms) using only first order logic and rules of inference.*

[Ref: p20 of "A Course on Mathematical Logic" by S. Srivastava]

True statements that can be used as steps in a proof *without inference* come in various flavours.

- **Premises** – statements assumed true for the purposes of a proof.
- **Axioms** – very simple statements assumed true for the purposes of building an area of mathematics.
- **Tautologies** – always true.
- **Theorems** – already proved true.
- **Definitions**
- etc...

Example 2.30

Premises:

Axioms:

Thus a **proof** is a finite sequence of statements

$$A_1, A_2, \ldots\ldots, A_n$$

such that each A_i is either

- known or assumed true or
- can be inferred from a known or assumed true statement A_j, with $j < i$ (ie., a previous statement).

The last statement, A_n, is usually called the **conclusion** and so a proof is sometimes written

$$A_1, A_2, \ldots\ldots, A_{n-1} \quad \text{proves} \quad A_n$$

There are many **methods** of proof:

- Direct proofs
- Indirect proofs
  - proof by contradiction,
  - contrapositive proof
- Proof by cases
- Others...

All these methods of proof are *shown to be valid* methods using theorems from logic.

The previous "sum of even integers is even" proof is an example of a direct proof.

Example 2.31

Prove the following: If x is an even integer then x^2 is an even integer.

Theorem 2.7 (Deduction Principle)

If $\{A_1, A_2, \ldots, A_n\}$ proves B, then $\{A_2, \ldots, A_n\}$ proves $(A_1 \Rightarrow B)$.

We use the Deduction Principle as follows: if we start by assuming A_1 is true and then show, using several steps (ie., A_2, A_3, $\ldots, A_n$), that B is true then the principle tells us the conditional $(A_1 \Rightarrow B)$ is necessarily true.

How does this help?

Answer:

The previous (incomplete) proof contained the steps showing $\{p, A_2, \ldots, A_n\}$ proves q and thus we deduce $p \Rightarrow q$ is true ie., the theorem is true.

Thus to complete the proof of example 2.31 just add the line:

The method of proof based on the Deduction Principle is called the **direct method of proof**.

The Deduction Principle is an example of a logic theorem being used to *prove* that a method of proof is valid.

Note: Many direct proofs you will read in books leave out the last Deduction Principle step!

Naive Set Theory

We are going to spend a little time thinking about the basics of set theory. Although fundamental to the way we think and write mathematics today, surprisingly, formal set theory has only been around since the late 1800's when Georg Cantor formulated the basic definitions and axioms. The most basic of these notions is that of **set**, which Cantor described simply as any collection of objects. But this quickly became problematic, giving rise for example, to Russell's Paradox.

In the early 1900's, axiomatic set theory was developed, by Von Neumann and others, in an attempt to iron out some of the crinkles in early set theory.

Recommended if you want to go beyond this course: "Naive Set Theory" by P Halmos. See Wikipedia (ZF set theory) for the logic forms of these axioms.

Russell's Paradox

In this course, we will try to get to some natural understanding of sets, with a little logic as its foundation.

The fundamental relation in set theory is that of *belonging*, written

$$x \in A$$

read "x belongs to A" or "x is contained in A" or "x is a member of A".

There are several ways to build up the mathematics of sets.

Currently most mathematicians start with the ZFC axioms: Zermelo-Fraenkel axioms + axiom of choice.[2]

This is a list of ten axioms.

To give you some idea of what is required to start building mathematics here is an informal list (for interest only):

[2] see Srivastava p12 if interested.

Axioms of Set Theory (simplified)

1. Existence. There exists a set.
2. Equality. Two set are the same (ie. equal) if they have the same elements.
3. Specification – defining subsets using conditions.
4. Pairing – For any two sets there exists a set that they both belong to.
5. Union – for any collection of sets there exists a set that contains all the elements that belong to at least one set in the collection.
6. Power set – for each set there exists a set that contains all its subsets.
7. Replacement – the range of a function defined on a set falls inside some set.
8. Inductive set – there exists a set A such that if $x \in A$ then the set $x \cup \{x\}$ is also in A. Required to prove that a set like the natural numbers exists.
9. Foundation – required to prevent Russell's paradox (there is no set of all sets).
10. Axiom of choice – from a collection of sets I can always choose one element from each to make a new set.

The primary thing we do is build new sets from given sets.

First, in more detail, our most useful axiom:

Axiom of Specification

For every non-empty set E and every condition $p(x)$ on its elements there is a set A which contains those elements of E for which $p(x)$ is true. We write

$$A = \{\ x \in E : p(x)\ \}$$

Note: This is essentially saying certain subsets of E exist: those containing the elements of E for which $p(x)$ is true.

Example 3.1

- If A is a set then the set $\{x \in A : x \neq x\}$ exists (separation axiom).

  It is called the
- $\{x \in \mathbb{R} : x^2 = 2\} =$
- $\{x \in \mathbb{Z} : x > 0\} =$
- $\{x \in \mathbb{R} : (x < 0) \vee (x = 0) \vee (x > 0)\} =$

Recall that the Axiom of Specification says that if $p(x)$ is a condition on the elements of some set E then there always exists a set A which contains those elements of E for which $p(x)$ is true. But what if p(x) is false for all elements of E?

Definition 3.1 (Empty Set)

We define the **empty set** (or *nullset*), denoted $\emptyset$, to be the set

$$\{x \in U : p(x)\}$$

where U is any set and $p(x)$ is any condition that is false for *all* elements in U.

The axiom of specification tells us such a set exists.

Relations between sets

Definition 3.2 (Equality)

Let A and B be sets. We say that $A = B$ if A and B contain the same elements.

Logic: $A = B \quad \Leftrightarrow \quad \forall x \in U \quad (x \in A \quad \Leftrightarrow \quad x \in B)$

Assuming some universal set U.

Example 3.2

$A = \{1,2\}, B = \{1,2\}, C = \{1,3\}$ and $U = \{1,2,3,4\}$.

x	$x \in A$	$x \in B$	$x \in A \Leftrightarrow x \in B$
1	T	T	
2	T	T	
3	F	F	
4	F	F	

x	$x \in A$	$x \in C$	$x \in A \Leftrightarrow x \in C$
1	T	T	
2	T	F	
3	F	T	
4	F	F	

Definition 3.3 (Subsets)

Let A and B be sets. We say that A is a **subset** of B if every element of A is also an element of B. This is written $A \subseteq B$.

Logic: $A \subseteq B \quad \Leftrightarrow \quad \forall x \in U \quad (x \in A \Rightarrow x \in B)$

If A is a subset of B and there is an element of B which is not an element of A then we say A is a **proper subset** of B.

Assuming some universal set U.

Also useful for showing A is *not* a subset of B is the negation:

$A \not\subseteq B \quad \Leftrightarrow \quad \exists x \in U$ s.t. $(x \in A) \wedge (x \notin B)$

Example 3.3

Let $A = \{1\}, B = \{1,2\}, E = \{1,2,3,4\}$

x	$x \in A$	$x \in B$	$(x \in A) \Rightarrow (x \in B)$	$(x \in B) \Rightarrow (x \in A)$
1	T	T		
2	F	T		
3	F	F		
4	F	F		

Theorem 3.1

Let A be a set. Then $\emptyset \subseteq A$.

Logic form:

"Proof".

$\square$

This is an example of an indirect proof. We will consider how to write such proofs later.

Theorem 3.2 (Equality and Subsets)

If A and B are sets then $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Logic: $A = B \;\Leftrightarrow\; (A \subseteq B) \wedge (B \subseteq A)$

How do we prove a biconditional statement?

We use

$$p \Leftrightarrow q \equiv (p \Rightarrow q) \wedge (q \Rightarrow p)$$

and write two proofs:

First the "Forward proof" showing $(p \Rightarrow q)$ is true

then the

"Converse proof" showing $(q \Rightarrow p)$ is true.

Since $(p \Rightarrow q)$ and $(q \Rightarrow p)$ are both true so is $p \Leftrightarrow q$ is true.

This is a **method of proof** - justified by the logical equivalence.

Proof

New sets from old

Given two known sets A and B we can define new sets in many ways.

Definition 3.4 (Union)

Let A and B be sets. We define the **union** of A and B, denoted $A \cup B$, to be the set that contains all elements of A and all elements of B.

$A \cup B = \{x \in U : (x \in A) \vee (x \in B)\}$

Definition 3.5 (Intersection)

Let A and B be sets. We define the **intersection** of A and B, denoted by $A \cap B$, to be the set that contains only those elements of A which are also all elements of B.

If $A \cap B = \emptyset$ we say that A and B are **disjoint**.

$A \cap B = \{x \in U : (x \in A) \wedge (x \in B)\}$

Definition 3.6 (Complement)

Let A and B be sets. We define the **complement** of B in A, denoted by $A \backslash B$, to be the set that contains only those elements of A which are not elements of B.

$A \setminus B = \{x \in U : (x \in A) \wedge (x \notin B)\}$

Theorem 3.3

Let A, B and C be subsets of some set E. Then $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$.

We usually prove set equality using Theorem 3.2

Proof.

$\square$

More notation

Definition 3.7 (Indexed Family)

Let i be an element of some nonempty set I. If for each $i \in I$ there is a corresponding set A_i, then

$$\mathcal{A} = \{A_i : i \in I\}$$

is called an **indexed family** of sets and I is called an **indexing set**.

The union of the elements in $\mathcal{A}$ is denoted by

$$\bigcup_{i \in I} A_i$$

and the intersection of the elements of I is denoted by

$$\bigcap_{i \in I} A_i.$$

Similarly, if $N = \{N_i : i \in I\}$ is set of objects N_i that can be multiplied (eg. real numbers, algebraic expressions etc.), then

$$\prod_{i \in I} N_i$$

denotes the product (multiplication) of all the elements of N.

Example 3.4

Ordered Pairs and Cartesian Products

Definition 3.8 (Ordered Pairs)

Let A and B be sets. An **ordered pair** is an object

$$(a, b)$$

where $a \in A$ and $b \in B$.

Two ordered pairs (a, b) and (c, d) are equal iff $a = c$ and $b = d$.

Definition 3.9 (Cartesian Product)

Given sets A and B, the **Cartesian product** of A and B, written $A \times B$ is the set of all ordered pairs:

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

Proposition 3.1

Let A be a set. If $(a_1, a_2) \in A \times A$ and $a_1 \neq a_2$ then $(a_1, a_2) \neq (a_2, a_1)$.

Example 3.5

Rational Numbers

Definition 3.10 (Rational Numbers)

The set of **rational numbers**, denoted $\mathbb{Q}$, is the set of ratios of integers,

$$\mathbb{Q} = \left\{ \frac{r}{s} \mid r, s \in \mathbb{Z}, s \neq 0 \right\}$$

We will investigate below the way in which the real numbers can be constructed from the rationals.

Aside:

A rational number is actually an *equivalence class* of ordered pairs of integers $(k, l), l \neq 0$. Two pairs (m, n) and (k, l) represent the same rational number iff $ml = nk$. Addition and multiplication of rational numbers is defined as

$$(k, l) + (m, n) = (kn + lm, ln)$$
$$(k, l) \cdot (m, n) = (km, ln)$$

Example 3.6

There is no rational number q satisfying $q^2 = 2$

Example 3.7

Let A and B be subsets of $\mathbb{Q}$ given by

$$A = \{q \in \mathbb{Q} : q < 0 \text{ or } q^2 < 2\} \qquad B = \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 > 2\}$$

Then A contains no largest element and B contains no smallest element.

Ordered sets and bounds

You are familiar with many basic properties properties of the real numbers.

We want to build up a list of properties that uniquely determine the real number system and see how they are related to $\mathbb{Q}$.

A familiar property of $\mathbb{Q}$ and $\mathbb{R}$ is that they are 'oredered'.

Definition 3.11 (ordered set)

Let S be a set. An **order** on S is a relation, denoted $<$, satisfying

1. If $x \in S$ and $y \in S$, then exactly one of the following holds:
   $x < y \qquad x = y \qquad y < x$
2. If $x, y, z \in S$ satisfy $x < y$ and $y < z$, then $x < z$

An **ordered set** is a set equipped with an order.

The notation $x \leq y$ is used to mean that $x < y$ or $x = y$.

We sometimes write $y > x$ in place of $x < y$.

Example 3.8

The set $\mathbb{Q}$ of rational numbers equipped with the usual order (in which $x < y$ is defined to mean that $y - x$ is positive) is an ordered set.

Definition 3.12 (bounded above)

Let S be an ordered set and $A \subseteq S$ a subset. If there exists $\beta \in S$ such that

$$\forall x \in A, \ x \leq \beta$$

then we say that A is **bounded above** and call β an **upper bound** for A.

Example 3.9

The set A from Example 3.7 is bounded above and any element of B is an upper bound for A.

Definition 3.13 (least upper bound)

Let S be an ordered set and $A \subseteq S$ a subset that is bounded above. An element $\alpha \in S$ is called a **least upper bound** (or **supremum**) of A if it has the following properties:

1. α is an upper bound for A
2. If γ is an upper bound for A, then $\alpha \leq \gamma$

It is clear from the definition that there can be at most one least upper bound for a given set A. We denote it by $\sup A$.

Example 3.10

1. $\{q \in \mathbb{Q} : q \leq \frac{1}{2}\} \subseteq \mathbb{Q}$ has least upper bound $\frac{1}{2}$
2. $\{q \in \mathbb{Q} : q < \frac{1}{2}\} \subseteq \mathbb{Q}$ has least upper bound $\frac{1}{2}$
3. The set $A \subseteq \mathbb{Q}$ as in Example 3.7 is bounded above but has *no least upper bound*

The defintions of **lower bound** and **greatest lower bound** (or **infimum**) are defined in exactly the same manner, but with $\leq$ replaced by $\geq$.

Definition 3.14

Let S be an ordered set and $A \subseteq S$ a subset. If there exists $\beta \in S$ such that

$$\forall x \in A, \ x \geq \beta$$

then we say that A is **bounded below** and call β a **lower bound** for A.

An element $\alpha \in S$ is called a **greatest lower bound** (or **infimum**) of A if it has the following properties:

1. α is a lower bound for A
2. If γ is a lower bound for A, then $\alpha \geq \gamma$

It is clear from the definition that there can be at most one greatest lower bound for a given set A. We denote it by $\inf A$.

Example 3.11

1. $E = \{\frac{1}{n} : n = 1, 2, 3, \dots\} \subset \mathbb{Q}$ is bounded above and bounded below.
   We have $\sup E = 1$ and $\inf E = 0$.
   Notice that, in this example, the infimum is not itself an element of E.

2. The set $B \subseteq \mathbb{Q}$ of Example 3.7 is bounded below but has *no greatest lower bound*

Definition 3.15

An ordered set S is said to have the **least-upper-bound property** if the following holds: If $A \subseteq S$ is not empty and is bounded above, then A has a least upper bound in S.

We have seen in the examples above that $\mathbb{Q}$ does *not* have the least-upper-bound property.

Fields

Definition 3.16 (Field)

A **field** is a set F equipped with two binary operations called addition (denoted $+$) and multiplication (denoted using $\times$ or simply by concatenation) that satisfy the following axioms (A1– A5, M1–M5, D):

Field axioms for addition

(A1) $\forall x, y \in F, \quad x + y \in F$ (closure)

(A2) $\forall x, y \in F, \quad x + y = y + x$ (commutivity)

(A3) $\forall x, y, z \in F, \quad (x + y) + z = x + (y + z)$ (associativity)

(A4) $\exists 0 \in F, \forall x \in F, \quad x + 0 = x$ (additive identity)

(A5) $\forall x \in F, \exists\, y \in F, \quad x + y = 0$ (additive inverse)

Definition (Field, continued)

Field axioms for multiplication

(M1) $\forall x, y \in F, \quad xy \in F$ (closure)

(M2) $\forall x, y \in F, \quad xy = yx$ (commutivity)

(M3) $\forall x, y, z \in F, \quad (xy)z = x(yz)$ (associativity)

(M4) $\exists 1 \in F \setminus \{0\}, (\forall x \in F, \quad 1x = x)$ (multiplicative identity)

(M5) $\forall x \in F \setminus \{0\}, \quad (\exists y \in F, \text{ s.t. } xy = 1)$ (multiplicative inverse)

Distributive law

(D) $\forall x, y, z \in F, \quad x(y + z) = xy + xz$

Example 3.12

1. $\mathbb{Q}$ equipped with the usual operations (defined above) forms a field.
2. $\mathbb{Z}$ (with the usual operations) is not a field.

Definition 3.17 (Ordered Field)

An **ordered field** is a field F that is also an ordered set and satisfies:

1. $\forall x, y, z \in F, \ (y < z \Rightarrow x + y < x + z)$
2. $\forall x, y \in F, \ ((x > 0) \wedge (y > 0) \Rightarrow xy > 0)$

Example 3.13

$\mathbb{Q}$ is an ordered field

The real field

We have seen that $\mathbb{Q}$ is an ordered field but does not have the least-upper-bound property.

Theorem 3.4

There exists and ordered field $\mathbb{R}$ that has the least-upper-bound property. Moreover, $\mathbb{R}$ contains $\mathbb{Q}$ as a subfield.

The proof of this theorem constructs $\mathbb{R}$ from $\mathbb{Q}$ using 'Dedekind cuts'

Positive and negative numbers

Notation:

- The set $\mathbb{R}^+ = \{x \in \mathbb{R} : 0 < x\}$ is called the set of **positive** real numbers.
- The set $\mathbb{R}^- = \{x \in \mathbb{R} : x < 0\}$ is called the set of **negative** real numbers.
- Note, zero is in neither of these sets.
- Less common is the notation: $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$ and $\mathbb{R}_0^- = \mathbb{R}^- \cup \{0\}$.

Addition and Cancellation

Theorem 3.5 (Addition and Cancellation)

$$\forall a, b, c \in \mathbb{R}, \quad (a = b) \quad \Leftrightarrow \quad (a + c = b + c).$$

Inequalities

Theorem 3.6

$$\forall x \in \mathbb{R}, \quad (0 < x) \Rightarrow (-x < 0).$$

Indirect Proofs

Up until now we have used "direct proofs":

- Start with true premises and infer a true conclusion using valid arguments.

Indirect proofs are different. We will consider two types:

- Proof using the contrapositive.
- Proof by contradiction.

Contrapositive Proofs

Instead of proving

$$P \Rightarrow Q$$

we use the logical equivalence (Theorem 2.4)

$$(P \Rightarrow Q) \equiv ((\sim Q) \Rightarrow (\sim P))$$

and prove

$$(\sim Q) \Rightarrow (\sim P)$$

using any method of proof.

Theorem 3.7

$(\forall x \in \mathbb{Z})$, *$x^2$ is even iff x is even.*

Proof by Contradiction

Recall the proof that $\emptyset \subseteq A$ for any set A:

A_1:		$\emptyset \nsubseteq A$	(contradiction) premise
A_2:	$\Rightarrow$	$\exists x \in \emptyset$ s.t. $x \notin A$	neg. of subset def. (def 3.3)
A_3:	$\Rightarrow$	$\emptyset$ is not empty	
A_4:		$\emptyset$ is empty	def of $\emptyset$ (def 3.1)
	$\Rightarrow$	$A_3 \wedge A_4$ is a contradiction	

Thus, by the method of "proof by contradiction", $\emptyset \nsubseteq A$ is not true, that is $\emptyset \subseteq A$ is true. $\square$

Previously we said in a proof true assumptions always lead to other true statements. But in the above "proof" we have a manifestly false statement $A_3 \wedge A_4$.

What is going on?

We can repeat $\emptyset \subseteq A$ as two proofs. First put theorem in negated form:

$\sim(P \Rightarrow Q) \equiv P \wedge \sim Q$:

Now the two proofs:

Thus,

From now on we combine the two proofs into a single contradiction "proof"

Thus, if a theorem is in the form of a conditional:

Theorem: If P then Q

that is, we need to prove $P \Rightarrow Q$ is true, then a proof by contradiction generally takes the form:

Proof (by contradiction):

P	Premise (ie. assume P true)
$\sim Q$	Contradiction premise (ie. assume $\sim Q$ true)
Main steps of proof ...	
Negation of some previous statement	(ie. a contradiction)
Thus Q is true	Contradiction Inference
Thus $P \Rightarrow Q$	Deduction Inference. QED.

Example 3.14

Prove by contradiction that,

Theorem: If x is a rational number and y is an irrational number, then $x + y$ is irrational.

First place theorem into logic form:

Proof (by contradiction):

Example 3.15

If A, B and C are non-empty sets such that $A \subseteq B$ and $B \cap C = \emptyset$ then $A \cap C = \emptyset$.

First place theorem into logic form:

Proof (by contradiction):

Absolute value and inequalities

The absolute value will be used extensively in the course so we need to know how to use it in proofs. It is an example of a situation where it may be easier to prove a "$p \Rightarrow q$" theorem by special cases rather than all cases at once.

Definition 3.18

The **modulus** or **absolute value** of $x \in \mathbb{R}$, denoted $|x|$, is defined as

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

Geometrically, $|x|$ is the distance from x to the origin on the real line.

The following theorem gives us a way of avoiding using $|x|$ in some cases (or vice versa).

Theorem 3.8 (Absolute-Interval)

Let $x \in \mathbb{R}$ and $a \in \mathbb{R}^+$. Then $|x| \leq a$ if and only if $-a \leq x \leq a$.

The following of results are useful when using absolute value in proofs.

Theorem 3.9

1. $\forall x \in \mathbb{R} \quad 0 \leq |x|$
2. $\forall x \in \mathbb{R} \quad x \leq |x|$
3. $\forall x \in \mathbb{R} \quad -x \leq |x|$

The triangle inequality

Another useful result about inequalities (which you should remember from Linear Algebra) is:

Theorem 3.10 (Triangle Inequality)

$\forall x, y \in \mathbb{R} \quad |x + y| \leq |x| + |y|$

Corollary 3.1 (reverse triangle inequality)

$\forall x, y \in \mathbb{R} \quad \big||x| - |y|\big| \leq |x - y|$

Bounded sets in $\mathbb{R}$

Geometrically, we would expect that any number bigger than an upper bound, is also an upper bound of the set E.

Theorem 3.11

Let $s, t \in \mathbb{R}$ and $E \subseteq \mathbb{R}$. If s is an upper bound of E and $s < t$ then t is an upper bound of E.

Proof:

How do we show w is **not** an upper bound of $E \subseteq \mathbb{R}$?

Proposition 3.2 (Not an Upper Bound)

Let $w \in \mathbb{R}$ and $E \subseteq \mathbb{R}$. Then w is not an upper bound of E if and only if there exists $x \in E$ such that $x > w$.

Thus to prove a number, w, is *not* an upper bound of a subset, we need to find at least one number *in the subset* that is greater than w.

Definition 3.19

(i) If $a < b$ then the subset $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ is called a **closed interval**.

(ii) If $a < b$ then the subset $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ is called an **open interval**.

Note that: the intervals $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ and $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$ are *neither* open *nor* closed.

Both of the sets $[a, b]$ and (a, b) are bounded above, however there is a *fundamental* difference between them.

Question Does there exist $s \in (a, b)$ such that s is an upper bound of (a, b)? More generally, do all subsets E which are bounded above contain an upper bound of themselves?

To answer this we will need the following theorem:

Theorem 3.12

$$\forall x, y \in \mathbb{R} \quad x < y \Rightarrow x < \frac{x+y}{2} < y.$$

Geometrically:

This theorem states that between **any** two real numbers, there exists at least one other real number.

Note: This statement is also true in $\mathbb{Q}$ but not in $\mathbb{Z}$!

Proof:

Proposition 3.3

Let $w \in \mathbb{R}$. Then w is an upper bound of (a, b) if and only if $w \geq b$.

Thus the subset (a, b) does not contain an upper bound of itself.

Proof:

Proposition 3.4

Let $s, t \in \mathbb{R}$. If s is a lower bound of E and $t < s$ then t is a lower bound of E.

Proposition 3.5

(i) *If $a, b \in \mathbb{R}$ with $a < b$, then a is a lower bound of (a, b).*
(ii) *If $a, b \in \mathbb{R}$ with $a < b$, then a is a lower bound of $[a, b]$.*

Note: $a \notin (a, b)$ and $a \in [a, b]$.

Proof: Similar to upper bound theorem.

We also have the result

$$\forall s \in \mathbb{R} \quad (s \text{ is not a lower bound of } E) \quad \Leftrightarrow \quad \exists w \in E \quad \text{s.t.} \quad w < s.$$

We now combine the idea of upper and lower bounds:

Definition 3.20 (Bounded)

If $E \subseteq \mathbb{R}$ and E is bounded above *and* below, we say that E is a **bounded** set. If E is not bounded it is called an **unbounded** set.

Proposition 3.6

1. $\mathbb{R} = \mathbb{R}^- \cup \{0\} \cup \mathbb{R}^+$.
2. *If $x \in \mathbb{R}^-$ and $y \in \mathbb{R}^+$ then $x < y$.*

Proof: - exercise -

The subsets $\mathbb{R}^+$ and $\mathbb{R}^-$ are important examples of unbounded subsets.

Proposition 3.7

(i) *$\mathbb{R}^+$ is unbounded.*
(ii) *$\mathbb{R}^-$ is unbounded.*

Proof:

Recall that b is an upper bound of (a, b) and $[a, b]$ but $b \notin (a, b)$ and $b \in [a, b]$. The situation in which a set contains its upper bound is special.

Definition 3.21 (Greatest element)

Let $E \subseteq \mathbb{R}$. If $\gamma \in E$ and γ is an upper bound of E, then γ is called the **greatest element** of E.

Definition 3.22 (Least element)

Let $E \subseteq \mathbb{R}$. If $\lambda \in E$ and λ is a lower bound of E, then λ is called the **least element** of E.

Example 3.16

$E_1 = \{1, 2, 3, 4\}$.

Note: The greatest element and least element, if they exist, are unique. That is, a set can have at most one greatest element and can have at most one least element.

Example 3.17

$E_2 = (a, b)$.

Example 3.18

$E_3 = [a, b]$.

Example 3.19

$E_4 = \left\{ -\frac{1}{n} : n \in \mathbb{N} \right\}$.

Summary:

Subset	Least Element	Greatest Element

So, some subsets have greatest elements or least elements and some don't. However, consider the following two subsets:

Definition 3.23 (Set of Upper/Lower bounds)

Let $E \subseteq \mathbb{R}$.

$$U_E = \{x \in \mathbb{R} : x \text{ is an upper bound of } E\}$$

$$L_E = \{x \in \mathbb{R} : x \text{ is a lower bound of } E\}$$

Example 3.20

That every nonempty subset of $\mathbb{R}$ that is bounded above (below) has a supremum (infimum) is ensured by the least-upper-bound property. This is sometimes referred to as the completeness axiom and stated in the following form:

Axiom (Completeness). Every non-empty set of upper bounds of $E \subseteq \mathbb{R}$ has a least element.

The following result is sometimes useful when establishing suprema.

Proposition 3.8

γ is the supremum of the set E if and only if for any $\epsilon > 0$ no element of E is greater than $\gamma + \epsilon$ and there is an element of E greater than $\gamma - \epsilon$.

Logic:

$$\gamma = \sup E \quad \Leftrightarrow \quad \forall \epsilon > 0 \quad ((\forall x \in E \quad x \leq \gamma + \epsilon) \wedge (\exists x \in E \quad \text{s.t.} \quad x > \gamma - \epsilon))$$

Example: If $A = \{-\frac{1}{n} \text{ s.t. } n \in \mathbb{N}\}$ then $\sup A = 0$.

Since we expect to add two real numbers to get another real, or multiply two reals to get a real, we should be able to add and multiply sup's to get new sup's (since all reals are sup's of some set).

In fact sup's should satisfy all the axioms of reals.

How adding sup's works is seen from the following theorem:

Theorem 3.13 (Sup Addition)

Let A and B be nonempty subsets of $\mathbb{R}$ and

$$C = \{x + y \in \mathbb{R} : x \in A \text{ and } y \in B\}.$$

If A and B have suprema, then C has a supremum and

$$\sup C = \sup A + \sup B.$$

Proof:

Mathematical Induction

We now consider our last method of proof - Induction.

Denote the set of positive integers by $\mathbb{N} = \{1, 2, 3 \ldots\}$.

Suppose $p(n)$ is a condition on $n \in \mathbb{N}$ and that we wish to prove theorems such as

$$\forall n \in \mathbb{N}, \quad p(n).$$

That is, we want to prove that $p(n)$ is true for all $n \in \mathbb{N}$. For example,

1. $p(n) : 1 + 2 + 3 + \ldots + n = \frac{1}{2}n(n+1)$
2. $p(n) : 2^n > n$
3. $p(n) : (1 + x)^n \geq 1 + nx, \quad x \in \mathbb{R}, x > -1$

To do this we use the *Principle of Mathematical Induction*.

Mathematical Induction

The following inference rule allows us to replace showing $p(n)$ is true (for all n) with showing a *conditional*, $p(n) \Rightarrow p(n+1)$ is true (for all n) - generally much easier.

Theorem 3.14 (Principle of Mathematical Induction)

If $p(n), \quad n \in \mathbb{N}$ are statements such that:

- *$p(1)$ is true, and*
- *$\forall k \in \mathbb{N} \quad \big(p(k) \Rightarrow p(k+1)\big)$ is true, then*
- *$\forall n \in \mathbb{N} \quad p(n)$ is true.*

This means that all the statements $p(1), p(2), \ldots$ are true.

The proof relies on the *well ordering* property of $\mathbb{N}$ which is not in the course.

To use the theorem, we need to

- Show $p(1)$ is true.
- Show that, if we assume $p(k)$ true then we can deduce $p(k+1)$ is true.
- Then, the deduction principle tells us that $p(k) \Rightarrow p(k+1)$ is true.
- Then the Induction principle tells us that $p(1), p(2), \ldots$ are all true

This gives the proof template:

Example 3.21

$p(n) : 1 + 2 + 3 + \ldots + n = \frac{1}{2}n(n+1)$

Example 3.22

$p(n) : 2^n > n$

Example 3.23

$p(n) : (1+x)^n \geq 1 + nx$ and $x > -1$ with $x \in \mathbb{R}$

Generalisations

Some more general forms of induction exist.

- Shift from $n \geq 1$ to $n \geq -m \;\; (m > 0)$;

  that is, $p(-4),\; p(-3),\; p(-2), \ldots$.

  Show $p(-4)$ is true then $p(k) \Rightarrow p(k+1)$

- Another form of induction (Strong Mathematical Induction) allows you to assume *all* of $p(k), p(k+1), \ldots p(n)$ are true in order to prove $p(n+1)$
  - Show $p(k)$ is true for $k \leq n \in \mathbb{N}$
  - Show $p(k), p(k+1), \ldots p(n) \Rightarrow p(n+1)$
  - Conclude $p(n)$ is true for all integers $n \geq k$.

The Archimedean Property

An important consequence of the Completeness Axiom is called the *Archimedean Property*.

Theorem 3.15 (Archimedean Property - Form I)

The set $\mathbb{N}$ of natural numbers is unbounded above in $\mathbb{R}$

Proof.

□

The two other theorems equivalent to the Archimedean Property are:

Theorem 3.16 (Archimedean Property - Form II)

If $x \in \mathbb{R}^+$ then $\exists n \in \mathbb{N}$ s.t. $n - 1 < x \leq n$

This gives us the picture that their is always an integer to the left and to the right of any real number.

Theorem 3.17 (Archimedean Property - Form III)

If $x, y \in \mathbb{R}^+$ then $\exists n \in \mathbb{N}$ s.t. $y < nx$

This is usually paraphrased as: "No matter how small a measuring stick it is always possible to place it end-to-end to extend beyond any distance".

Density of the Rationals

Theorem 3.18

If x and y are real numbers with $x < y$, then there exists a rational number r such that $x < r < y$.

Proof.

□

An easy consequence of Theorem 3.18 tells us that we can also find an irrational number between any two reals. This means that wherever you look along the real line - every interval, no matter how small, contains a rational number and an irrational number.

Functions

Recall the *informal* definition of a function:

A function (or map) $f : A \to B$ from the set A, called the **domain**, to the set B called the **codomain**, is a "rule" that assigns to every $x \in A$, a unique element $y \in B$, called the **image** of x. The subset of all images is called the **range** of f.

Example 4.1

This informal definition is not quite precise enough for our purposes, since the question of when something is, or isn't, a *rule* is not easy to answer. To help us get a handle on the precise idea of a function, we use the concept of a *relation*.

Definition 4.1 (Relation)

A **relation** R between two sets A and B, is a subset of the Cartesian product $A \times B$.

Thus the "rule" is defined by the pairs.

Example 4.2

Definition 4.2 (Function)

A **function** $f : A \to B$ is a triple (A, B, f), where $f \subseteq A \times B$ is a relation, such that

1. if $(x, y_1) \in f$ and $(x, y_2) \in f$ then $y_1 = y_2$,
2. if $x \in A$ then $\exists z \in B$ such that $(x, z) \in f$.

Example 4.3

If $A = \{-1, 0, 1\}, \; B = \{T, F\}$ then

Special types of functions

Injective:

Surjective:

Bijective:

We will discuss functions again later in the course. For now we are interested in functions with domain $A = \mathbb{N}$.

Sequences

Consider a function whose domain is $\mathbb{N}$ (or sometimes a subset of $\mathbb{N}$).

We can represent sequences with diagrams in two ways:

1. Cartesian Plot and 2. Points on real line.

Definition 4.3 (Sequence)

A **sequence** is a function $f : \mathbb{N} \to B$ (usually $B = \mathbb{R}$) and is written

$$f(1), f(2), f(3), \ldots$$

or

$$f_1, f_2, f_3, \ldots$$

or

$$\{f_n\}$$

or

$$\{f_n\}_{n\geq 1},$$

or

$$(f_n)$$

or

$$(f_n)_{n\geq 1},$$

where the image of n, $f(n)$, is called the n^{th} **term of f**.

Convergence of Sequences

In real analysis we are mostly interested in how the terms of a sequence approximate some real number, L.

Question: As n gets larger are the terms f_n a better approximation to L than previous terms?

Think of decimal approximations to π. Terminating the expansion gives a rational number:

$$f_1 = 3.1, \quad f_2 = 3.14, \quad f_3 = 3.142, \quad f_4 = 3.1416, \quad f_5 = 3.14159, \ldots$$

or

$$f_1 = \frac{31}{10}, \quad f_2 = \frac{314}{100}, \quad f_3 = \frac{3142}{1000}, \quad f_4 = \frac{31416}{10000}, \ldots$$

Is each successive rational a better approximation to π?

Some sequences do provide ever better approximations to some real L, others do not.

Those that do we will end up calling: **convergent**.

Possible long-term behaviours of a sequence

Question: How can we consider f_1, f_2, f_3, f_4,.. as successively better approximations to some real number L?

Some sequences may "jump around" but still give increasingly better approximations

Some sequences do not approximate any $L \in \mathbb{R}$.

Use distance to measure the accuracy: $\epsilon \in \mathbb{R}^+$ is a way to "measure" the accuracy of an approximation.

Two requirements for a sequence to approximate some L:

– Must be "arbitrarily accurate":

– Must be "increasingly accurate":

In summary: a sequence $\{f_n\}$ approximates some number $L \in \mathbb{R}$ if:

Definition of Convergence of a Sequence

Definition 4.4 (Sequence Limit)

Let $L \in \mathbb{R}$, $\epsilon \in \mathbb{R}^+$, $n \in \mathbb{N}$, and $f : \mathbb{N} \to \mathbb{R}, n \mapsto f_n$ be a sequence.

We say that L is the **limit** of the sequence f_n, if

for all $\epsilon > 0$ there exists an $M \in \mathbb{N}$ such that

for all $n > M$, we have $|f_n - L| < \epsilon$.

This is written

$$\lim_{n \to \infty} f_n = L.$$

If so, we say that the sequence f_n **converges** to L.

If we write this using logic symbols, we have

$$\forall \epsilon > 0 \quad \exists M \in \mathbb{N} \text{ s.t. } \quad \forall n > M, \ \ |f_n - L| < \epsilon$$

Thinking geometrically, this says:

f_n converges to L if:

1. for any distance $\epsilon > 0$ we choose,
2. we can always find an $M \in \mathbb{N}$,
3. such that for all $n > M$,
4. the distance of every f_n to L is smaller than ϵ.

Definition 4.5 (Sequence Convergence)

If there is some L such that $\lim\limits_{n\to\infty} = L$ (i.e., if such an L exists) then we say that the sequence f_n **converges**.

If there is no such L then we say that the sequence f_n **diverges**.

For the previous example, $f_n = 1 - 1/n$:

Question: Given $\epsilon > 0$ what value of M do we need?

Sometimes we can answer this geometrically.

The previous calculation gives us a candidate for M (which depends on ϵ). We now need to prove that it "does the job", that is, it will enable us to prove that

$$\forall \epsilon > 0 \quad \big(\exists M \in \mathbb{N} \text{ s.t. } (\forall n > M, \ \ |f_n - L| < \epsilon)\big)$$

is true?

Proof

Note the following notation:

- The phrase "For all $\epsilon > 0 \quad p(\epsilon)$" will be written:

$$\forall \epsilon > 0 \quad p(\epsilon)$$

  where $p(\epsilon)$ is some condition on n.

  This is an *abbreviation* for the logic statement:

$$\forall \epsilon \in \mathbb{R}^+ \quad p(\epsilon).$$

- Similarly, the phrase "For all $n > M \quad p(n)$" will be written:

$$\forall n > M \quad p(n)$$

  where $p(n)$ is some condition on n.

  This is an *abbreviation* for the logic statement:

$$\forall n \in \{M+1, M+2, \dots\} \quad p(n)$$

We will need the following function and theorem.

Definition 4.6 (Max)

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

$$\max(a,b) = \begin{cases} a, & \text{if } a \geq b \\ b, & \text{if } a < b. \end{cases}$$

Theorem 4.1

Let $a, b, y \in \mathbb{R}$. If $\max(a,b) < y$ then $a < y$ and $b < y$.

Example 4.4

$f : \mathbb{N} \to \mathbb{R}$, $f_n = \dfrac{2^n - 1}{2^n}$

Example 4.5

$g : \mathbb{N} \to \mathbb{R}$, $g_n = 1 + \left(\frac{-1}{2}\right)^n$

Note: Not all sequences behave like f_n and g_n ie. "get closer to some number".

Example 4.6

$a_n = (-1)^n$

Luckily the proof that a sequence converges to L is almost always of the same form. We have a simple template:

The integer M **does not** have to be the smallest integer that works.

The difficult part is finding M (which depends on ϵ) and then filling in the middle.

Usually this is done before writing the proof. In the proof you have to work out how to get from your already computed M to $|f_n - L| < \epsilon$.

Sometimes the previous calculation can be "reversed". Other times a different route is required.

For proofs about convergence of sequences, we often need to use the ceiling function.

Definition 4.7 (Ceiling)

Let $x \in \mathbb{R}$. We define the **ceiling** of x, $\lceil x \rceil$, to be the smallest integer greater than or equal to x. Equivalently,

$$\lceil x \rceil = \inf\{k \in \mathbb{Z} : x \leq k\}.$$

Geometrically $\lceil x \rceil$ it is the closest integer to the right (or equal to) of x.

Note, inf is only defined for non-empty sets. The fact that $\{k \in \mathbb{Z} : x \leq k\}$ is non-empty is a consequence of the Archimedean Property (Form II).

Example 4.7

We will frequently rely on the following simple theorem:

Theorem 4.2

Let $x \in \mathbb{R}$, $n, M \in \mathbb{Z}$. If $M = \lceil x \rceil$ and $M < n$ then $x < n$.

Geometrically the theorem is clear, however a picture is *not* a proof so it is a good exercise to try prove it.

Example 4.8

Show that the sequence $\left(\frac{n}{2n+1}\right)_{n\in\mathbb{N}}$ converges to $\frac{1}{2}$.

Example 4.9

For the sequence $f_n = \dfrac{1}{n^2 + 6n + 4}$ which converges to $L \in \mathbb{R}$,

find an $M \in \mathbb{N}$ such that, if $\epsilon > 0$, and $n \in \mathbb{N}$ with $n > M$, then $|f_n - L| < \epsilon$.

Chain of inequalities

Chain of inequalities to simplify expressions.

Notes:

- Most books use "N" in place of "M", in which case the proof of the existence of a limit is called an "$\epsilon - N$ proof".
- Many books use a "challenge-response" explanation of the idea convergence.

Consider a two player game:

Player A chooses any $\epsilon > 0$ and then challenges player B to find a natural number M such that $(\forall n > M), \ |f_n - L| < \epsilon$.

If player B can do this for any $\epsilon > 0$ then B wins and L is the limit of the sequence.

Written as a proof, player B needs $M(\epsilon)$ - hence producing an M for any $\epsilon > 0$ - and also needs the calculation starting from $M(\epsilon)$ and showing that $|f_n - L|$ is indeed less than ϵ for all $n > M$. The proof shows that no matter what $\epsilon > 0$ is, for the subset

$$M_\epsilon = \{n \in \mathbb{N} : n > M\}$$

it is true that there exists an $L \in \mathbb{R}$ such that $|f_n - L| < \epsilon$.

Note: This approach appears to put time into the proof ie. if the the challenge-response goes on "forever" then the theorem is proved - this is **not** the case. The logic statement only requires showing existence of integer (M) for every positive real number (ϵ) such that a certain condition is satisfied ($|f_n - L| < \epsilon$) - this has nothing to do with some process going on forever.

Divergent Sequences

To show that (f_n) **does not** converge to L, we need to show that

$$\sim \left[\lim_{n\to\infty} f_n = L\right]$$

is true. That is,

$$\sim [\forall \epsilon > 0 \quad (\exists M \in \mathbb{N} \text{ s.t. } (\forall n > M, \ |f_n - L| < \epsilon))] \tag{1}$$

$$\equiv \quad \exists \epsilon > 0 \text{ s.t. } \left(\forall M \in \mathbb{N} \quad (\exists n > M, \text{ s.t. } |f_n - L| \geq \epsilon)\right) \tag{2}$$

This is equivalent to saying that:

we can find an example of $\epsilon > 0$ so that,

for any $M \in \mathbb{N}$,

we can always find some $n > M$ such that $|f_n - L| \geq \epsilon$.

Note:

- We only need to find *one* such $\epsilon > 0$, since the statement is now a "there exists" statement.
- We only need to find *one* term (ie. there exists n) such that $|f_n - L| \geq \epsilon$, for some $n > M$.

Example 4.10

Show that $(-1)^n$ does not converge to $L = 1$.

Proof:

Example 4.11

Show that $\frac{1}{n}$ does not converge to $L = 1$, but it does converge to $L = 0$.

Proof:

Not convergent

Beware the distinction between "not convergent to L" and "not convergent".

To prove (f_n) is not convergent, we need to prove that, for any $L \in \mathbb{R}$,

$$\lim_{n\to\infty} f_n \neq L.$$

This can usually be done by cases on L: choose ranges of L where $\lim_{n\to\infty} f_n \neq L$ is easier to prove.

There is a theorem for proving divergence that generally gives a simpler proofs. To use the theorem we need the idea of "subsequences". A **subsequence** of the sequence $\{f_n\}_{n\geq 1}$ is a sequence $f_{n_1}, f_{n_2}, f_{n_3}, f_{n_4}, \ldots$ where $n_1 < n_2 < n_3 < \ldots$ ie. pick a subset (without changing the order) of the terms of $\{f_n\}_{n\geq 1}$. The subsequence is usually written in the form $\{f_{n_k}\}_{k\geq 1}$.

Example 4.12

The following are subsequences of $\{1/n\}$:

Theorem 4.3 (Divergence Criteria for Sequences)

A sequence $\{f_n\}_{n\geq 1}$ is divergent if any of the following hold:

1. *$\{f_n\}_{n\geq 1}$ has two subsequences $\{f_{n_k}\}_{k\geq 1}$ and $\{f_{n_p}\}_{p\geq 1}$ that converge to two different limits.*
2. *$\{f_n\}_{n\geq 1}$ has a subsequence that is divergent.*
3. *$\{f_n\}_{n\geq 1}$ is unbounded.*

Example 4.13

The sequence $f_n = (-1)^n$ is divergent.

Proof:

Special Types of Sequences

For certain types of sequences we do not need full $\epsilon - M$ proofs to show convergence.

Definition 4.8

The following sequences are called **monotonic** sequences.

A sequence (f_n) is:

- **Increasing** if: $\forall n \in \mathbb{N} \;\; f_{n+1} > f_n$
- **Non-Decreasing** if: $\forall n \in \mathbb{N} \;\; f_{n+1} \geq f_n$
- **Decreasing** if: $\forall n \in \mathbb{N} \;\; f_{n+1} < f_n$
- **Non-Increasing** if: $\forall n \in \mathbb{N} \;\; f_{n+1} \leq f_n$

Example 4.14

Bounded Sequences

Definition 4.9 ((Un)Bounded Above)

- A sequence (f_n) is **bounded above** if there exists an $s \in \mathbb{R}$ such that for all $n \in \mathbb{N}$, $f_n \leq s$.
- A sequence (f_n) is **unbounded above** if no such s exists.

Similarly,

Definition 4.10 ((Un)Bounded Below)

- A sequence (f_n) is **bounded below** if there exists an $s \in \mathbb{R}$ such that for all $n \in \mathbb{N}$, $f_n \geq s$.
- A sequence (f_n) is **unbounded below** if no such s exists.

Note it can be proved (do it as an exercise) that unbounded sequences do not converge and so sometimes also called divergent.

Definition 4.11 (Bounded)

A sequence is called **bounded** if it is both bounded above and bounded below.

Example 4.15

Prove that the sequence (2^n) is not bounded.

Proof:

Combining the ideas of bounded and monotonic gives a very useful theorem:

Theorem 4.4 (Non-decreasing and convergence)

Let (f_n) be a non-decreasing sequence. If (f_n) is bounded above, then it is convergent.

Note: This theorem tells us that (f_n) has a limit, but not what that limit is.

(The limit is actually $\sup A$, where A is a set containing all the terms of (f_n)).

There is a similar theorem for non-increasing, bounded below sequences. When we combine the two, we get:

Theorem 4.5 (Bounded monotonic sequence)

Every bounded, monotonic sequence is convergent.

To use this theorem we need to show:

- The sequence is bounded above.
- The sequence is bounded below.
- The sequence is monotonic.

If L_u and L_l are the upper and lower bounds, then

$$L_l \leq \lim f_n \leq L_u$$

that is, the bounds give an estimate of the limit. The better the bounds, the better the estimate.

Algebra of Limits

Now that we have precise definitions of convergence, we can prove some limit laws.

Theorem 4.6

If $f_n \to \alpha$ and $g_n \to \beta$ then:

1. $(f_n + g_n) \to \alpha + \beta$.
2. $(f_n \cdot g_n) \to \alpha \cdot \beta$.
3. $\dfrac{f_n}{g_n} \to \dfrac{\alpha}{\beta}$, *for* $\beta \neq 0$.
4. $k \cdot f_n \to k \cdot \alpha$, *for any* $k \in \mathbb{R}$.

Many of these (and other) algebra theorems use the following:

Recall "$\forall \epsilon > 0$" is an abbreviation for "$\forall \epsilon \in \mathbb{R}^+$". It does not matter how the elements of $\mathbb{R}^+$ are "represented" so long as they give all of $\mathbb{R}^+$. This allows us to change the ϵ to some other function, $g(\epsilon)$, so long as the following condition holds:

$$\{g(\epsilon) : \epsilon > 0\} = \mathbb{R}^+$$

Examples of valid functions:

1. $g_1(\epsilon) = \epsilon/2$
2. $g_2(\epsilon) = \epsilon^2$
3. $g_3(\epsilon) = \log(\epsilon + 1)$

Examples of invalid functions:

1. $g_4(\epsilon) = 1 + \epsilon$
2. $g_5(\epsilon) = \epsilon^2 - 1$
3. $g_6(\epsilon) = \log(\epsilon)$

If $g(\epsilon)$ is valid then we get logical equivalence. For example, all the statements below are logically equivalent.

1. $\forall \epsilon > 0 \quad \exists M \in \mathbb{N}$ s.t. $\forall n > M \quad |f_n - L| < \epsilon$
2. $\forall \epsilon > 0 \quad \exists M \in \mathbb{N}$ s.t. $\forall n > M \quad |f_n - L| < \epsilon/2$
3. $\forall \epsilon > 0 \quad \exists M \in \mathbb{N}$ s.t. $\forall n > M \quad |f_n - L| < \epsilon^2$

In the above three examples $|f_n - L|$ still has to be less than some positive real number. What will in general change is $M(\epsilon)$.

Proof of $(f_n + g_n) \to \alpha + \beta$

The downside of the convergence definition is that you *need to know the limit*. Can we get around this for an arbitrary (convergent) sequence?

Geometrically, we have

This suggests that if (f_n) converges, then $|f_n - f_m| \to 0$. What we would like is the converse: "if $|f_n - f_m| \to 0$ then (f_n) converges".

Cauchy Sequences

Definition 4.12 (Cauchy Sequence)

A sequence f_n of real numbers is called a **Cauchy Sequence** if for each $\varepsilon > 0$ there exists an $M \in \mathbb{N}$ such that whenever $n, m > M$ we have $|f_n - f_m| < \varepsilon$.

Example 4.16

Lemma 4.1 (Convergent-Cauchy)

Every convergent sequence is a Cauchy sequence.

Proof.

□

Theorem 4.7 (Cauchy-Bounded)

Every Cauchy sequence is bounded.

Proof: Omitted - it's rather long – see just about any real analysis text book.

Cauchy's Convergence Criterion

Theorem 4.8 (Cauchy's Convergence Criterion)

A sequence (f_n) converges if and only if for every $\epsilon > 0$ there exists $M \in \mathbb{N}$ such that for all $n, m > M$,

$$|f_n - f_m| < \epsilon.$$

This is known as **Cauchy's Convergence Criterion**.

Proof:

Example 4.16 continued:

Limit Points

We now generalise the idea of a point γ being $\sup E$.

Definition 5.1 (Limit Points)

Let $E \subseteq \mathbb{R}$ and $\ell \in \mathbb{R}$. Then ℓ is called a **limit point** of E if, given any real number $\delta > 0$ there is a point $x \in E$ such that

$$0 < |x - \ell| < \delta.$$

Logic: (ℓ a limit point of E) $\Leftrightarrow$ $\forall \delta > 0 \quad \exists x \in E \quad \text{s.t.} \quad 0 < |x - \ell| < \delta$

Note that:

1. The condition $0 < |x - \ell|$ excludes $x = \ell$.
2. E *may* or *may not* contain the limit point.
3. Limit points are also called **accumulation points** or **cluster points** in some books.

Example 5.1

If $E = \left\{x \in \mathbb{R} : \sin\left(\frac{1}{x}\right) = \pm 1\right\}$ then $\ell = 0$ is a limit point of E (and $0 \notin E$).

Definition 5.2 (Deleted neighbourhood)

Let $\ell \in \mathbb{R}$ and $\delta > 0$. The set defined as

$$N_0(\ell, \delta) = \{x \in \mathbb{R} : 0 < |x - l| < \delta\}$$

is called a **deleted neighbourhood** of ℓ.

We can extend Definition 5.1 to allow $\pm\infty$ to be limit points.

Definition 5.3 (Limit point at Infinity)

Let $E \subseteq \mathbb{R}$. Then if, given any $\alpha > 0$, there is an element $x \in E$ such that $x > \alpha$, we say that E has a **limit point at ∞** (or if $x < \alpha$, E has a limit point at $-\infty$).

This is equivalent to saying that E is unbounded above (respectively unbounded below).

Note that;

1. "$\pm\infty$" are **not** elements of $\mathbb{R}$, but **are** limit points of $\mathbb{R}$.
2. The only limit point of $\mathbb{N}$ is ∞.

Continuous Functions

We would like to generalise the idea of convergence of sequences to convergence of functions to some limit L. That is, we will think about functions with domain $E \subseteq \mathbb{R}$ where E is not $\mathbb{N}$.

This means we want to know what happens to $f(x)$ as x gets close to some point ℓ.

1. $n \to \infty$ becomes $x \to \ell \in \mathbb{R}$.
2. $f_n \to L$ becomes $f(x) \to L$.

Similarly for functions

1. ℓ is not generally in E ,
2. f may not be defined at $x = \ell$.

So we would like to define $f \to L$ without assuming that ℓ is in the domain of f. This requires us to think about what "gets close to" means.

- $\ell \notin E$ requires us to define the *deleted neighbourhood*.
- f not defined at $x = \ell$ will be avoided by using $|f(x) - L| < \varepsilon$ with $x \neq \ell$.

Example 5.2

We can now generalise the idea of sequence limits to function limits:

Definition 5.4 (Limits of Functions)

Let f be the function $f : E \to \mathbb{R}$. Let ℓ be a limit point of E and f be defined on a deleted neighbourhood of ℓ (that is, $N_0(\ell, \delta) \subseteq E$). We say that **the limit of f as x tends to ℓ is $L \in \mathbb{R}$**, if: for any $\epsilon > 0$ there is a $\delta > 0$ such that for all x satisfying

$$0 < |x - l| < \delta$$

we have that

$$|f(x) - L| < \epsilon.$$

We write this as

$$\lim_{x \to \ell} f = L$$

or $f \to L$ as $x \to \ell$. We also say that f **converges** to L as x tends to ℓ.

Logic Form:

$$\forall \epsilon > 0 \ \left(\exists \delta > 0 \ \text{ s.t. } \ \left(\forall x \in E \ (0 < |x - \ell| < \delta \Rightarrow |f(x) - L| < \epsilon)\right)\right).$$

To motivate the conditional

$$(0 < |x-\ell| < \delta) \Rightarrow (|f(x) - L| < \epsilon)$$

consider the following example:

$$f : \mathbb{R} \setminus \{2\} \to \mathbb{R} \quad ; \quad f(x) = \frac{x^2-4}{x-2}.$$

How does the value of $f(x)$ change as the value of x gets closer to $x = 2$?

Logic Form:

$$\forall \epsilon > 0 \; \left(\exists \delta > 0 \text{ s.t. } \left(\forall x \in E \; (0 < |x-\ell| < \delta \Rightarrow |f(x) - L| < \epsilon)\right)\right).$$

It might be easier to read when written in the form:

$$\forall \epsilon > 0 \quad \exists \delta > 0 \quad \text{s.t.} \quad \forall x \in E \quad p(x, \delta, \epsilon)$$

where

$$p(x, \delta, \epsilon) : \quad (0 < |x-\ell| < \delta) \Rightarrow (|f(x) - L| < \epsilon).$$

Thus: for every $\epsilon > 0$ we need to show there exists a $\delta > 0$ such that $p(x, \delta, \epsilon)$ is true.

$p(x, \delta, \epsilon)$ is true if: whenever x is in the deleted neighbourhood, (ie. $0 < |x-\ell| < \delta$) it is the case that $f(x)$ is within the ϵ strip (ie. $|f(x) - L| < \epsilon$).

Recall the logical equivalence: If $A \subset B$ then,

$$\forall x \in B \quad (x \in A \Rightarrow p(x)) \quad \equiv \quad \forall x \in A \quad p(x).$$

This enables us to rewrite the limit definition as

$$\forall \epsilon > 0 \quad \Big(\exists \delta > 0 \quad \text{s.t.} \quad \big(\forall x \in N_0(\ell, \delta) \quad |f(x) - L| < \epsilon\big)\Big)$$

where $N_0(\ell, \delta)$ is the subset $0 < |x-\ell| < \delta$.

This form of the logic then reads as:

For every $\epsilon > 0$ we need to show there exists a $\delta > 0$ such that for every x in the deleted neighbourhood it is the case that $|f(x) - L| < \epsilon$.

Notes:

- f is never evaluated at $x = \ell$ so it does not matter if $\ell \notin E$.
- $N_0(\ell, \delta) \subseteq E$ is a strong restriction on the domain of E.
- The definition requires $\ell \in \mathbb{R}$ but can be extended to allow for limit points at $\pm\infty$.
- We specify ϵ first and then find the deleted neighbourhood of ℓ such that $f(x)$ is always within ϵ of L.

Note the similarity with the $\epsilon - M$ convergence definition:

- Thus a proof that $f \to L$ as $x \to \ell$ has a very similar template to the one we used for sequences.
- *Before the proof* (in the strategy section) we use $|f(x) - L| < \epsilon$ to find a suitable $\delta(\epsilon)$. That is, we start with $|f(x) - L| < \epsilon$ and try to get to $0 < |x - l| < \delta(\epsilon)$.
- *In the proof* we start with $0 < |x - l| < \delta(\epsilon)$ and try to get to $|f(x) - L| < \epsilon$.

Proof Template:

Geometrically we have the picture:

Example 5.3

If $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$, prove that $\lim_{x \to 2} f(x) = 4$

We will need:

Lemma 5.1

Let $a, x, y \in \mathbb{R}$. If $a < \min(x, y)$ then $a < x$ and $a < y$.

Proof of previous example:

Beware bad points

Functions which are not defined near limit points or are unbounded near limit points need special care. Choosing δ small enough can help to avoid these points.

Example 5.4

Prove that

$$\lim_{x \to 1} \frac{x^2 - 1}{2x - 1} = 0$$

Proof:

One-sided limits

The definition of one-sided limits is the same as limits, *except* the deleted neighbourhood $N_0(\ell, \delta)$ is replaced by left or right deleted neighbourhoods.

Left and Right Limits

The **left limit** is written:

$$\lim_{x \to \ell^-} f(x) = L$$

"x approaches ℓ from below (or from the left)".

The **right limit** is written:

$$\lim_{x \to \ell^+} f(x) = L$$

"x approaches ℓ from above (or from the right)".

The following theorem is useful for showing a limit does not exist.

Theorem 5.1 (Left-Right Limits)

Let $f : E \to \mathbb{R}$ be a function. Then

$$\lim_{x \to \ell} f(x) = L$$

if and only if

$$\lim_{x \to \ell^-} f(x) = L \textit{ and } \lim_{x \to \ell^+} f(x) = L.$$

Example 5.5

Prove that $f(x) = \begin{cases} x - 1 & x \leq 2 \\ x + 1 & x > 2 \end{cases}$

does not have a limit as $x \to 2$.

The limit points $\pm\infty$

We can now think about what happens to a function as:

1. "$x \to \pm\infty$"
2. $x \to \ell$ but "$f(x) \to \pm\infty$".

f bounded.

In the first case, as with sequences, we are really asking, what happens to f as x gets arbitrarily large (positive or negative)? That is, what values can f take in the neighbourhood

$$N_{\infty}(\delta) = \{x \in \mathbb{R} : x > \delta\}$$

or similarly in the neighbourhood

$$N_{-\infty}(\delta) = \{x \in \mathbb{R} : x < \delta\}?$$

So we are looking at what happens on an unbounded subset of $\mathbb{R}$.

Thus the logic form of the definition of $\lim\limits_{x\to\infty} f(x) = L$ is:

$$\forall \epsilon > 0 \quad \big(\exists \delta > 0 \quad \text{s.t.} \quad (\forall x \in E \quad (x > \delta \Rightarrow |f(x) - L| < \epsilon))\big).$$

where E is the domain of f. Note, E must be unbounded above for $\lim\limits_{x\to\infty} f(x)$ to be defined. This is logically equivalent to:

$$\forall \epsilon > 0 \quad \big(\exists \delta > 0 \quad \text{s.t.} \quad (\forall x \in N_{\infty}(\delta) \quad |f(x) - L| < \epsilon\big).$$

Geometrically this gives:

f unbounded.

In the second case, if f is unbounded as $x \to \ell \in \mathbb{R}$ we say, $\lim\limits_{x\to\ell} f(x) = +\infty$ if

$$\forall \alpha > 0 \quad \big(\exists \delta > 0 \quad \text{s.t.} \quad (\forall x \in E \quad (0 < |x - l| < \delta \Rightarrow f(x) > \alpha))\big).$$

where E is the domain of f. We also say that f is *divergent* to ∞.

Note that there are analogous results for $-\infty$.

Remember that: ∞ is *not* a real number, but a convenient notation for unbounded subsets of $\mathbb{R}$ or as limit points.

Example 5.6

Limit Laws

Now that we have a precise definition of a limit, we can prove the standard limit laws.

Theorem 5.2

If $f \to \alpha$ and $g \to \beta$ as $x \to \ell$ where $\alpha, \beta \in \mathbb{R}$, then

1. $f + g \to \alpha + \beta$
2. $f \cdot g \to \alpha \cdot \beta$
3. $\dfrac{f}{g} \to \dfrac{\alpha}{\beta}$, $\beta \neq 0$

These are proved in a very similar way to the sequence limit laws and are left as an exercise.

When a limit does not exist at $x = \ell$.

There are generally two ways to show that a function has no limit as $x \to \ell$.

1. Prove the negation of the limit definition is true.
2. Use some theorems, for example left and right limits must both exist and be equal.

Example 5.7

Continuity

Question: When is $\lim_{x \to \ell} f(x) = f(\ell)$? That is, when can we just substitute to get the value of a limit?

Consider the following functions at $x = 0$:

Definition 5.5 (Continuous)

Let f be a real valued function with domain $E \subseteq \mathbb{R}$. Suppose $\ell \in E$ and E contains a neighbourhood of ℓ. Then f is said to be **continuous at $x = \ell$**, if

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

If $D \subseteq E$ and f is continuous at all $\ell \in D$, then f is said to be **continuous on D**.

A very useful consequence of this is that, if we know a function is continuous at $x = \ell$ then we immediately know the value of the limit as $x \to \ell$. We get the value of the limit by substituting $x = \ell$, that is $f(\ell)$.

So continuity at $x = \ell$ requires:

1. $\lim_{x \to \ell} f(x) \in \mathbb{R}$
2. $f(\ell)$ is defined (that is, $\ell \in E$)
3. the limit equals the value of f at ℓ.

The definition requires $N(\ell, \delta) \subseteq E$, so that $x = \ell$ is in the domain of f (so $f(\ell)$ must be defined). Thus if f is not defined at $x = \ell$ it *cannot* be continuous at $x = \ell$.

Definition 5.6

A **neighbourhood** of $x = \ell$ is the open interval

$$N(\ell, \delta) = \{x \in \mathbb{R} : |x - l| < \delta\}$$

for some $\delta > 0$.

Note the difference between a *neighbourhood* and a *deleted neighbourhood*. Here, $x = \ell$ is included.

Thus the answer to the original question is:

If f is continuous at $x = \ell$ then its limit as $x \to \ell$ equals the value of the function.

This is only really useful if we know a function is continuous at $x = \ell$ *without* proving the limit exists (via $\epsilon - \delta$) - which is where the continuity laws come in.

Theorem 5.3

Let f and g be continuous at $x = \ell$. Then the following functions are continuous at $x = \ell$.

1. $f + g$
2. fg
3. $\dfrac{f}{g}$, *as long as* $g(\ell) \neq 0$
4. $f \circ g$, *as long as f is continuous at* $x = g(\ell)$.

These can all be proved using the modified $\epsilon - \delta$ Definition 5.7, or using the limit laws.

We also have the following theorem:

Theorem 5.4

The following functions are all continuous on the stated domains:

- *Polynomials:* $\sum_{n=0}^{m} a_n x^n$; $D = \mathbb{R}$, $a_n \in \mathbb{R}$
- e^x, $\cos(x)$, $\sin(x)$; $D = \mathbb{R}$
- $\log(x)$; $D = \mathbb{R}^+$
- $x^{1/p}$; $D = \mathbb{R}^+, p \in \mathbb{N}$ *(that is, pth root).*

Since the limit point ℓ is required to be in the domain of f, a deleted neighbourhood is not required in the $\epsilon - \delta$ definition, which gives the following equivalent definition:

Definition 5.7 (Continuity at a point)

Let f be a real valued function with domain E and $\ell \in E$. Then f is **continuous** at $x = \ell$ if: for any $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that for all x satisfying

$$|x - l| < \delta(\epsilon)$$

we have

$$|f(x) - f(\ell)| < \epsilon.$$

Note: The original $\epsilon - \delta$ definition of a limit requires a deleted neighbourhood, but Definition 5.7 does not.

Example 5.8

Show that $f : \mathbb{R} \to \mathbb{R}$ is continuous at $x = 0$, where

$$f(x) = \begin{cases} x \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0. \end{cases}$$

Bounded and Monotonic Functions

As with sequences, (where $E = \mathbb{N}$) we have similar bounded and monotonic properties for functions.

Assume that $E, D \subseteq \mathbb{R}$.

Definition 5.8 (Bounded Functions)

Let $f : E \to D$.

1. f is **bounded above** if $\exists \alpha \in \mathbb{R}$ such that $(\forall x \in E \quad f(x) \leq \alpha)$.
2. f is **bounded below** if $\exists \alpha \in \mathbb{R}$ such that $(\forall x \in E \quad f(x) \geq \alpha)$.
3. f is bounded above and bounded below (on E) if and only if f is **bounded** on E.

Definition 5.9 (Monotonic Functions)

Let $f : E \to D$. Then f is **monotonic on E** if and only if it satisfies one of the following conditions:

(i) f is **increasing**; that is, $\big(\forall x, y \in E \quad x < y \Rightarrow f(x) < f(y)\big)$.

(ii) f is **non-decreasing**; that is, $\big(\forall x, y \in E \quad x < y \Rightarrow f(x) \leq f(y)\big)$.

(iii) f is **decreasing**; that is, $\big(\forall x, y \in E \quad x < y \Rightarrow f(x) > f(y)\big)$.

(iv) f is **non-increasing**; that is, $\big(\forall x, y \in E \quad x < y \Rightarrow f(x) \geq f(y)\big)$.

If either (i) or (iii) apply, then we say f is **strictly monotonic**.

Monotonicity is *very* useful for the existence of inverse functions.

Question: If $f : E \to D$ when can we define $g : D \to E$ such that $f(g(x)) = x$?

Theorem 5.5 (Injectivity, Continuity & Monotinicity)

Let f be a continuous, real-valued, injective function on an interval I. Then f is either increasing or decreasing on I.

The theorem leads to the following useful relationship between continuity and inverse functions.

Theorem 5.6 (Continuity, Injectivity & Inverse,)

Let f be a continuous, real-valued, injective function on an interval I. Then an inverse function f^{-1} exists on I and is a continuous function.

Note that I can be $[a, b], [a, b), (a, b]$ or (a, b).

Example 5.9

The sine function. The existence of an inverse depends on the domain.

Bounded and Continuous Functions

Boundedness and continuity are related as shown in the following theorem.

Theorem 5.7 (Bounded and Continuous)

Let $f : E \to \mathbb{R}$, $E \subseteq \mathbb{R}$, be a continuous function on $[a, b]$.
Then f is bounded on $[a, b]$.

The following example shows why the interval $[a, b]$ of the Bounded and Continuous theorem must be *closed*.

To account for continuity at points were the domain of the function only contains half of any deleted neighbourhood, the definition of continuity can be weakened to left and right continuity.

Definition 5.10 (One sided Continuity)

A function $f : A \to B$ is **left continuous** at $x = \ell$ if

$$\lim_{x \to \ell^-} f(x) = f(\ell)$$

and is **right continuous** at $x = \ell$ if

$$\lim_{x \to \ell^+} f(x) = f(\ell)$$

Example 5.10

Continuity and Solving Polynomials

Question: How do I know that

$$5x^6 - 3x^4 + 2x^3 - 2 = 0$$

has at least one real solution?

Example 5.11

Intermediate Value Theorem

We now come to a classic theorem of real analysis.

Theorem 5.8

Let g be a continuous real valued function on the closed interval $I = [a, b]$.

Then g takes on every value in the interval

$$J = [\inf_I g, \sup_I g].$$

Conventionally, this is stated in a weaker form.

Theorem 5.9 (Intermediate Value Theorem)

Let $f : [a, b] \to \mathbb{R}$ be a continuous function and let $y \in \mathbb{R}$ such that $f(a) < y < f(b)$ or $f(a) > y > f(b)$.

Then there exists $c \in (a, b)$ such that $f(c) = y$.

Open and closed subsets of $\mathbb{R}$

Recall that $N(x, \epsilon) = \{y \mid d(x, y) < \epsilon\}$.

Definition 5.11

A subset $E \subset \mathbb{R}$ is called **open** if for all $x \in E$ there exists an $\epsilon > 0$ such that $N(x, \epsilon) \subseteq E$. A subset if called **closed** is its complement is open.

The union of any collection of open sets is open.

Proposition 5.1

Let $\mathcal{A} = \{A_i \mid i \in I\}$ be a collection of sets. If all the A_i are open, then $\cup_{i \in I} A_i$ is open.

Example 5.12

The intersection of a collection of open sets need not be open.

For $i \in \mathbb{N}$ define $A_i = (-1/i, 1/i) \subset \mathbb{R}$. Then $\cap_{i \in \mathbb{N}} A_i = \{0\}$ and is not open.

However, the intersection of a finite number of open sets is open.

Proposition 5.2

Let $A_1, A_2, \ldots A_n \subset \mathbb{R}$ be a finite collection of open sets. Then $A_1 \cap A_2 \cap \cdots \cap A_n$ is open.

Corollary 5.1

Let $\mathcal{C} = \{C_i \mid i \in I\}$ be a collection of closed sets. Then $\cap_{i \in I} C_i$ is closed. Let $\mathcal{C}' = \{C_1, \ldots, C_N\}$ be a finite collection of closed sets. Then $\cup_{i=1}^{N} C_i$ is closed.

Compact subsets

Definition 5.12

A subset $K \subset \mathbb{R}$ is **compact** if every infinite sequence $(x_i)_{i \in \mathbb{N}}$ with $x_i \in K$ has a convergent subsequence which converges to a point in K.

(This is in fact what is called 'sequential compactness'.)

Example 5.13

Let $a \in \mathbb{R}$ be positive.

- $[0, a)$ is not compact. (We'll show this now.)
- $[0, a]$ is compact. (This is the Heine-Borel theorem below.)

Example 5.14

$[0, \infty)$ is closed but not compact.

Theorem 5.10 (Heine-Borel)

Let E be a subset of $\mathbb{R}$. Then E is compact iff E is both closed and bounded.

Proof: Suppose first that $E \subset \mathbb{R}$ is compact.

To see that E is bounded

Now to show that E is closed.

Suppose that it were not. Then $\mathbb{R} \setminus E$ is not open, therefore there exists $x \in \mathbb{R} \setminus E$ such that for all $\epsilon > 0$, $(x - \epsilon, x + \epsilon) \cap E \neq \emptyset$.

For the converse, suppose now that $E \subset \mathbb{R}$ is closed and bounded. We must show that E is compact.

$\square$

One of the central features of compact sets is that a continuous function on a compact set has a maximum value.

Lemma 5.2

Let $K \subset \mathbb{R}$ be compact. Then $\sup K \in K$.

Proof:

Lemma 5.3

Let $f : \to \mathbb{R}$ be a contiuous function, and let $O \subset \mathbb{R}$ be open. Then the set $f^{-1}(O) = \{a \in \mathbb{R} : f(a) \in O\}$ is open.

Proposition 5.3

Let f be a continuous function and $K \subset \mathbb{R}$ a compact set. Then there is an $a \in K$ such that $f(a) \geq f(b)$ for all $b \in K$.

Proof: By the previous lemma it suffices to show that $f(K)$ is compact.

Differentiability

In this topic, I represents any of the intervals $[a, b], (a, b), [a, b), (-\infty, a), (-\infty, a]$, etc, where $a < b$.

Refresher: Let $f : I \to \mathbb{R}$. The *chord function*

Since we now have an $\epsilon - \delta$ definition of a limit, we can prove that $f'(x_0)$ exists (as a real number).

Definition 6.1 (Differentiability)

Let I be an interval of $\mathbb{R}$ and $f : I \to \mathbb{R}$. The **derivative of f at x_0** denoted $f'(x_0)$, is defined as

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

If $f'(x_0) \in \mathbb{R}$ (that is, exists) we say that f is **differentiable at x_0**. If f is differentiable for all $x \in I$ we say that f is **differentiable**. If f is differentiable we can define the function

$$f' : I \to \mathbb{R}, x \mapsto f'(x).$$

Note:

- $f'(x_0)$ is the limiting value of the chord function, $\mathcal{C}_{x_0}(h)$ as $h \to 0$.

  Thus if $\mathcal{C}_{x_0}(h)$ is not defined in a deleted neighbourhood $N_0(x_0, \delta)$, then the limit defining $f'(x_0)$ cannot exist.

  In particular $f(x_0)$ must be defined.

What is the relationship between continuity and differentiability?

Theorem 6.1 (Differentiability and Continuity)

If $f : I \to \mathbb{R}$ is differentiable at $x_0 \in I$ then f is continuous at x_0.

Proof:

Uses the trick

$$f(x_0 + h) - f(x_0) = h \cdot \frac{f(x_0 + h) - f(x_0)}{h}, h \neq 0.$$

Also, if $f'(x_0)$ exists then $\mathcal{C}_{x_0}(h)$ exists in some deleted neighbourhood $(N_0(0, \delta))$ so $f(x_0)$ is defined.

Note: Since $\lim f = \lim g \not\Rightarrow f = g$, the converse of the theorem is not true. That is, continuity does not imply differentiability.

Example 6.1

The function $f(x) = |x|$ is continuous for all $x \in \mathbb{R}$ but not differentiable at $x = 0$.

Relationship between differentiability and monotonicity

If f is differentiable at x_0 we can get some information about the "local" monotonicity properties of f. That is, if we are close enough to x_0 then we have the following result:

Theorem 6.2 (Differentiability and Monotonicity)

Let f be differentiable at x_0. Then if $f'(x_0) > 0$ we have:

- $\exists \epsilon > 0$ *such that*
$$\forall h, \; (0 < h < \epsilon) \Rightarrow f(x_0 + h) > f(x_0).$$
- $\exists \epsilon > 0$ *such that*
$$\forall h, \; (-\epsilon < h < 0) \Rightarrow f(x_0) > f(x_0 + h).$$

with similar statements for $f'(x_0) < 0$.

Rolle's Theorem

We now come to another classic theorem of real analysis.

Theorem 6.3 (Rolle's Theorem)

Let f be continuous on $[a, b]$, differentiable on (a, b) and suppose $f(a) = f(b) = 0$. Then there exists a point $c \in (a, b)$ such that $f'(c) = 0$.

That is, there must exist at least one point of (a, b) where the tangent is horizontal. If f goes upwards somewhere after $x = a$ it must "turn around" to get back down to $x = b$.

Note carefully the conditions on the intervals.

It must be continuous on the closed interval, differentiable on the open interval and c must be in the open interval.

Proof:

Mean Value Theorem

The primary application is to prove another classic theorem, the Mean Value theorem, which in turn is used to prove results for Taylor polynomials - see later.

Rolle's theorem can be generalised to remove the condition $f(a) = f(b) = 0$, in which case the slope becomes non-zero, as given by:

Theorem 6.4 (Mean Value Theorem)

Let f be continuous on $[a, b]$, differentiable on (a, b). Then there exists $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Note that, $f'(c)$ is the slope of the chord through $f(a)$, $f(b)$.

Thus the Mean Value theorem says

The proof uses a trick to define a function $g(x)$ that can be used with Rolle's theorem:

$$g(x) = f(x) - f(a) - \left(\frac{f(b) - f(a)}{b - a}\right) \cdot (x - a).$$

Note that $g(a) = g(b) = 0$.

Proof:

Notes:

- You may use all the usual rules for differentiating functions:
  - The sum rule, difference rule, scalar multiple rule.
  - The product rule, quotient rule, chain rule.
  - Standard derivatives for polynomials, sin, cos, log, etc.

  These are all straightforward to prove, and are left as an exercise.
- The main use of the MVT is for proving things about Taylor polynomials.

Riemann Integration

(There is a more general form of integration, called Lebesgue Integration, but this requires Measure Theory.)

Recall: Integration has two ideas

- The idea of an antiderivative, the opposite operation to differentiation.
- The area under a curve.

The two ideas are linked via the Fundamental Theorem of Calculus. We want a precise definition of:

$$\int_a^b f(x)\,dx$$

Generalising the first idea is too restrictive, some functions are not differentiable, but can be integrated. So we will try to make the second idea precise.

To do this we will show that for some functions, approximating the area under the curve by rectangles can be a very good approximation to the actual area.

There are two considerations:

- How do we divide up $[a, b]$?
- What height do we use for each rectangle?

In the first instance, we allow arbitrary positions:

Definition 7.1 (Partition)

A **partition** P of $[a, b]$ is a finite ordered set, written

$$P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}.$$

Thus, the partition fixes the widths of the rectangles, but what height should we use?

There are many possibilities

But there is a *natural choice*:

Definition 7.2 (Riemann Sums)

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. For each subinterval $[x_{k-1}, x_k]$ of the partition $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$, define

$$m_k = \inf\{f(x) : x \in [x_{k-1}, x_k]\}$$

$$M_k = \sup\{f(x) : x \in [x_{k-1}, x_k]\}$$

and then define *two* approximations to the area: the **Lower Riemann Sum**

$$L(f, P) = \sum_{k=1}^{n} m_k(x_k - x_{k-1})$$

and the **Upper Riemann Sum**:

$$U(f, P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1}).$$

Informally,

- $U(f, P)$ uses the maximum value of f on each subinterval.
- $L(f, P)$ uses the minimum value of f on each subinterval.

Note:

1. We only assume f is *bounded* (so we can ensure that sup and inf exist), not continuous.
2. Riemann sums depend on P. Changing the number of points, or the position of points will change the sum.

The useful property of sup and inf is that $m_k \leq M_k$ for *all* subintervals, thus we have:

Theorem 7.1 (Sum ordering: same partition)

$$L(f, P) \leq U(f, P)$$

for all partitions of $[a, b]$.

Thus, $L(f, P)$ is *always* a lower bound and $U(f, P)$ is *always* an upper bound on the "area".

Question

How do L and U change as we add new points to the partition P? That is, what if we keep the current n points fixed, but add new points between the existing ones to give a new partition $\hat{P}$ so that

$$P \subseteq \hat{P}?$$

For example,

$$P = \left\{1, \tfrac{3}{2}, 2\right\} \text{ and } \hat{P} = \left\{1, \tfrac{5}{4}, \tfrac{3}{2}, 2\right\}.$$

Definition 7.3 (Refinement)

If $P \subseteq \hat{P}$ then $\hat{P}$ is called a **refinement** of P (that is, $\hat{P}$ has the same points and *more*).

How L and U change under refinement is given by the following theorem:

Theorem 7.2 (Sum ordering: refinement)

If $\hat{P}$ is a refinement of P then

$$L(f, P) \leq L(f, \hat{P})$$

and

$$U(f, \hat{P}) \leq U(f, P).$$

So increasing the number of subintervals increases L and decreases U. Hence, it improves the estimates.

Intuitively we can see this from the geometry:

The proof relies on the following theorem:

Theorem 7.3 (Subintervals and Sup's/Inf's)

Let $f : [a, b] \to \mathbb{R}$ be bounded on $[a, b]$, with $c \in (a, b)$. Then

1. $\sup_{[a,c]} f \leq \sup_{[a,b]} f$
2. $\sup_{[c,b]} f \leq \sup_{[a,b]} f$

with similar (reversed) inequalities for infimum.

The proof is left as an easy exercise.

We now prove Theorem 7.2.

What happens as we change the position of the points in P? If we compare U and L (rather than U and U) then we have:

Theorem 7.4 (Sum Ordering: different partitions)

If P and Q are any partitions of $[a, b]$, then

$$L(f, P) \leq U(f, Q).$$

Note this is a generalisation of Theorem 7.1.

Proof:

Thus as we refine the partition (that is, increase points), L is non-decreasing and U is non-increasing.

Question: Do they meet in the middle somewhere?

Answer: Sometimes... it depends on f.

Riemann Integrable

In the *special* case that L and U are equal in the limit, we define

Definition 7.4 (Riemann Integrable)

Let $\tilde{P}$ be the set of *all* possible partitions of $[a, b]$.
Define the **Lower Riemann Integral** of f on $[a, b]$ as:

$$L(f) = \sup\{L(f, P) : P \in \tilde{P}\}$$

and the **Upper Riemann Integral** of f on $[a, b]$ as:

$$U(f) = \inf\{U(f, P) : P \in \tilde{P}\}.$$

We say that f is **Riemann integrable** on $[a, b]$ if $U(f) = L(f)$

In this case we define the integral of f by

$$\int_a^b f = U(f) = L(f).$$

Example 7.1

Here is a bounded function that is *not* integrable.

Question: Is it possible to determine if f is integrable *without* testing all partitions?

Answered by Riemann:

Theorem 7.5 (Integrability and Boundedness)

Let $f : [a, b] \to \mathbb{R}$ be bounded. Then f is integrable if and only if, for all $\epsilon > 0$ there exists a partition P of $[a, b]$ such that,

$$U(f, P) - L(f, P) < \epsilon.$$

This is similar to an $\epsilon - \delta$ definition except finding $\delta(\epsilon)$ is replaced by finding $P(\epsilon)$.

In the special case when f is continuous, we can say more (without considering any partitions at all).

Theorem 7.6 (Integrability and Continuity)

If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, then f is integrable on $[a, b]$.

Notes

1. The converse is NOT true: Integrable $\not\Rightarrow$ Continuous.
2. Functions which are not continuous on $[a, b]$ may still be integrable, for example, $f : \mathbb{R} \to [-1, 1]$
$$f(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0. \end{cases}$$
3. The proof is not difficult but requires "uniform continuity" which is not in this course.

Theorem 7.7 (Fundamental Theorem of Calculus)

- *If $f : [a, b] \to \mathbb{R}$ is integrable and $F : [a, b] \to \mathbb{R}$ satisfies $F'(x) = f(x)$ for all $x \in [a, b]$ then*
$$\int_a^b f(x)\, dx = F(b) - F(a)$$
- *Let $g : [a, b] \to \mathbb{R}$ be integrable and define $G : [a, b] \to \mathbb{R}$ by*
$$G(x) = \int_a^x g(u)\, du, \qquad G(a) = 0.$$
Then G is continuous on $[a, b]$. If g is continuous at $c \in [a, b]$ then G is differentiable at c and $G'(c) = g(c)$.

The theorem says:

1. If f has an antiderivative, then we can evaluate $\int_a^b f(x)\, dx$.
2. If we can evaluate $\int_a^x f(u)\, du$, then we get the antiderivative (which is also continuous).

Algebraic Properties of the Integral

All these usual properties can be proved from the definition:

Theorem 7.8 (Algebra of Integrals)

If f and g are integrable on $[a, b]$ and $c \in [a, b]$ then:

1. $\int_a^c f$ *and* $\int_c^b f$ *exist*
2. $\int_a^b f = \int_a^c f + \int_c^b f$
3. $\int_a^b (f + g) = \int_a^b f + \int_a^b g$
4. $\int_a^b \lambda f = \lambda \int_a^b f, \qquad \lambda \in \mathbb{R}$

Improper Integrals

Question: What if f is unbounded at some point $c \in [a, b]$? Without loss of generality, we can take $c = a$ or b since we can use Theorem 7.8 part 3.

As with all things that might be unbounded, we consider the limit of finite things!

Definition 7.5 (Improper Integral: unbounded integrand)

Let $f : (a, b] \to \mathbb{R}$ be such that f is integrable on $[c, b]$ for all $c \in (a, b)$. If the limit
$$\lim_{c \to a^+} \int_c^b f$$
exists we call it the **improper integral** of f on $(a, b]$.

Beware: the same notation

$$\int_a^b f$$

is generally used . You need to check the function f carefully to see if

$$\lim_{c \to a^+} \int_c^b f$$

is meant. That is, check to see if a is actually in the domain of f.

We can make a similar definition for *unbounded* intervals:

Definition 7.6 (Improper integral: unbounded interval)

Let $f : [a, \infty) \to \mathbb{R}$ be integrable on $[a, c]$ for every $c > a$. If the limit

$$\lim_{c \to \infty} \int_a^c f$$

exists we call it an **improper integral**, written

$$\int_a^\infty f.$$

Note: there is a similar definition for $\displaystyle\int_{-\infty}^a f = \lim_{c \to -\infty} \int_c^a f$.

See practice class sheets for examples.

Series

For this topic the focus is on applying theorems rather than proving them.

Consider the following method[1] of approximating the area under the parabola $y = 1 - x^2$

[1] Due to Archimedes over 2000 years ago.

The (finite) *partial sums*

$$A_k = 1 + \frac{1}{4} + \cdots + \frac{1}{4^k}$$

provide an increasingly better approximation to the area.

This example shows very clearly that:

- LHS is an increasingly better approximation to the area
- The approximation is always less than $\frac{4}{3}$

$$A_k < \frac{4}{3}, \qquad \forall k \in \mathbb{N}$$

  and so is bounded above
- Less obviously, if I choose any number less than $\frac{4}{3}$, I can always find a $k \in \mathbb{N}$ such that A_k is larger than it, that is,

$$\frac{4}{3} = \sup\{A_k : k \in \mathbb{N}\}.$$

Sound familiar?

$$\forall \epsilon > 0 \quad \exists k_0 \in \mathbb{N} \quad \text{s.t.} \quad \forall k > k_0 \quad \left| A_k - \frac{4}{3} \right| < \epsilon.$$

Recall the definition of the convergence of a sequence f_n:

$$\forall \epsilon > 0 \quad \exists M \in \mathbb{N} \quad \text{s.t.} \quad \forall n > M, \ |f_n - L| < \epsilon.$$

Thus the **partial sums** behave like the terms of a sequence

$$A_1, A_2, A_3, \ldots.$$

So we say that, A_k converges to $\frac{4}{3}$ and write it as

$$1 + \frac{1}{4} + \frac{1}{16} + \ldots = \frac{4}{3}.$$

This is NOT the addition of an "infinite" number of terms: it is a convenient notation for saying the partial sums converge to $\frac{4}{3}$. That is, the approximations have a finite limit.

Question: Why can't we think of it as the sum of an infinite number of terms?

Clearly A_k **is** a sum of terms!

If it were a summation, it would need to have *all* the properties of the binary operation "+", that is:

- Associative $\quad (1+2)+(3+4) = 1+(2+3)+4$
- Commutative $\quad 1+2+3 = 3+2+1$
- etc.

Example 8.1

Consider $1 - 1 + 1 - 1 + 1 - 1 + \ldots = \quad ?$

The essential idea is to consider the partial sums, A_k, as defining a sequence (A_k) and then consider the *convergence of the sequence* (A_k).

Note: Historically things happened the other way around: Problems with "Infinite sums" (particularly those from Fourier series) were finally resolved once convergence was understood.

Beware: "$a_1 + a_2 + a_3 + \ldots$" is **not** the same mathematical object as "$a_1 + a_2 + \ldots + a_k$" (ie. a finite sum).

The notation is, however, convenient as **sometimes** $a_1 + a_2 + a_3 + \ldots$ has some of the properties of a finite sum.

Definition 8.1 (Series)

The expression "$a_1 + a_2 + a_3 + \dots$" is called **a series** and is usually written as

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

or

$$\sum_{n>0} a_n.$$

The number a_n is called the **n-th term** of the series.

Furthermore,

$$A_k = a_1 + a_2 + \dots + a_k$$

is called the k-th **partial sum** of the series and (A_k) the **sequence of partial sums**.

Thus we have

$$a_1 + a_2 + a_3 \dots$$

Since the partial sums give a sequence of approximations to the series, in the sense of Archimedes, we are motivated to define:

Definition 8.2 (Convergent series)

The series

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

is said to be **convergent** if the sequence (A_k)

$$A_k = \sum_{n=1}^{k} a_n$$

is convergent. If $\lim_{k\to\infty} A_k = L$, we call L the **sum** (or value) of the series and write:

$$a_1 + a_2 + \dots = L, \quad \text{or} \quad \sum_{n>0} a_n = L.$$

The series is **divergent** if (A_k) is divergent.

Again we have the standard algebra:

Theorem 8.1 (Algebra of Series)

If $\sum_{n>0} a_n = \alpha$ *and* $\sum_{n>0} b_n = \beta$ *then*

$$\sum_{n=1}^{\infty}(a_n + b_n) = \alpha + \beta$$

and

$$\sum_{n=1}^{\infty}(\lambda a_n) = \lambda\alpha, \quad \lambda \in \mathbb{R}.$$

Proof: Follows from the algebra of sequences.

Warning: Do not confuse the sequence of terms (a_n) with the sequence of partial sums (A_k).

Given **any** sequence (a_n) we can define a series via its partial sums:

The convergence of (A_k) (and hence the value of the series "$a_1 + a_2 + \ldots$") is related to the sequence of terms in an important way:

Theorem 8.2 (Series and Term convergence)

If $\sum_{n=1}^{\infty} a_n$ converges then $\lim_{n\to\infty} a_n = 0$.

The **converse is false**: $\lim_{n\to\infty} a_n = 0$ does not imply that $\sum_{n=1}^{\infty} a_n$ converges.

The classic counterexample is the **Harmonic Series**:

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

which diverges, even though

$$\frac{1}{n} \to 0.$$

The Harmonic series is the $s = 1$ case of the famous Riemann Zeta Function:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

which is the function of the famous Riemann Hypothesis.

Proof of Theorem 8.2:

Even though the converse is false, the **contrapositive** (which is necessarily true) is very useful:

$$\sim (\lim_{n\to\infty} a_n = 0) \Rightarrow \sim \left(\sum_{n>0} a_n \text{ converges} \right)$$

which gives the equivalent theorem:

Theorem 8.3 (Terms & Series Divergence)

If $\lim_{n\to\infty} a_n \neq 0$ then $\sum_{n>0} a_n$ diverges.

This is sometimes called the *divergence test*.

Example 8.2

1. $\sum_{n>0} 2^n$ diverges.
2. $\sum_{n>0} \frac{n^3}{n^2+4}$ diverges.
3. $\sum_{n>0} (-1)^{n+1}$ diverges.

Back to the original problem:

Question: How do we know when the series $a_1 + a_2 + a_3 + \ldots$ behaves like a finite sum, that is, we can commute or associate the terms without changing its value?

Theorem 8.4 (Associativity)

If $\sum_{n>0} a_n$ converges then any series obtained by grouping the terms without changing their order converges to the same value.

That is; convergence $\Rightarrow$ associative (regrouping) of terms is allowed.

Again the converse is false. If you split the terms of $a_1 + a_2 + \ldots$ into parts, then the series may or may not converge.

But the contrapositive of Theorem 8.4 is still useful. Informally:

Theorem 8.5 (Associativity and Divergence)

Regrouped $\sum_{n>0} a_n$ diverges $\Rightarrow \sum_{n>0} a_n$ diverges.

We can use this to show that the harmonic series diverges.

What about the convergence of $\sum_{n=1}^{\infty} \frac{1}{n^2}$ or $\sum_{n=1}^{\infty} \frac{1}{n^p}$ etc?

Previous contrapositive theorems relate one *divergent series* to another *divergent series*.

For convergence, do we always have to use the definition? That is, show the sequence of partial sums converges via $\epsilon - M$?

The answer depends on the type of series (that is, the properties of its terms).

Don't get lost (summary of some of the tests)

Positive Series

Series with all positive terms are generally easier to test for convergence.

Definition 8.3 (Positive series)

A series $\sum_{n>0} a_n$ is **positive** if $a_n \geq 0$ for all $n \in \mathbb{N}$.

Theorem 8.6 (Positive Bounded Series)

A positive series $\sum_{n>0} a_n$ converges if and only if there exists $K > 0$ such that

$$\forall n \in \mathbb{N} \quad A_n \leq K,$$

where $A_n = \sum_{m=1}^{n} a_m$

Note:

1. if *all* $a_n \leq 0$ then you can factor out -1 to get a positive series
2. if some of the terms equal zero then 're-indexing' changes it to the case $a_n > 0$ for all n.

Example 8.3

Show that $\sum_{n>0} \frac{1}{n^2}$ converges.

Comparison Test

Theorem 8.6 still requires knowledge of the partial sums $\{A_k\}$.

Question: Can we determine the convergence of a series using only the terms a_n?

Answer: If we can compare the series with some other series known to converge, then, yes:

Theorem 8.7 (Comparison Test)

Let $\sum_{n>0} a_n$ be a series and $\sum_{n>0} b_n$ a series convergent to L.

If there exists $K > 0$ such that a_n and b_n satisfy

$$0 \leq a_n \leq K\, b_n,$$

for all $n \in \mathbb{N}$, then $\sum_{n>0} a_n$ converges and $\sum_{n>0} a_n \leq KL$.

Conditions:

The theorem is similar to the sandwich theorem for limits. If the terms of a positive series $\sum_{n=1}^{\infty} a_n$ are always bounded above by the terms of a convergent positive series (and bounded below by zero) then we get the convergence of $\sum_{n=1}^{\infty} a_n$.

We also have the contrapositive of the theorem

Theorem 8.8 (Comparison test - divergence)

If there exists $K > 0$ such that for all $n \in \mathbb{N}$, $\quad 0 \leq b_n \leq Ka_n$ and $\sum_{n>0} b_n$ diverges, then $\sum_{n>0} a_n$ diverges.

Example 8.4

Does $\displaystyle\sum_{n>0} \frac{2}{n^2+1}$ converge?

Ratio Test

A test that does not rely on the convergence of another series is:

Theorem 8.9 (Ratio Test - Non-limit form)

Let $\sum_{n>0} a_n$ be a positive series. Let r and R be real numbers with $0 < r < 1$ and $R > 1$. Then

1. *If there exists $M_1 \in \mathbb{N}$ such that for all $n > M_1$,*
$$\frac{a_{n+1}}{a_n} < r, \qquad \text{then } \sum_{n>0} a_n \text{ converges.}$$
2. *If there exists $M_2 \in \mathbb{N}$ such that for all $n > M_2$,*
$$R < \frac{a_{n+1}}{a_n}, \qquad \text{then } \sum_{n>0} a_n \text{ diverges.}$$

Conditions:

Thus, if the ratio of consecutive terms is bounded by a positive constant, we can determine if $\sum_{n>0} b_n$ converges.

On occasion it is easier to use the limit form of the Ratio Test.

Theorem 8.10 (Ratio Test - Limit Form)

Let $\sum_{n>0} a_n$ be a series such that the sequence $\left|\frac{a_{n+1}}{a_n}\right|$ converges to a limit r.

1. *If $r < 1$ then $\sum_{n>0} |a_n|$ converges (as does $\sum_{n>0} a_n$).*
2. *If $r > 1$ then $\sum_{n>0} a_n$ diverges.*

Conditions:

Example 8.5

Does $\sum_{n>0} \frac{1}{5^n}$ converge?

Example 8.6

More generally, when does $\sum_{n\geq 0} r^n$ converge?

Example 8.7

Show that $\sum_{n>0} \frac{n^2}{2^n}$ converges.

What happens when $\left|\frac{a_{n+1}}{a_n}\right| \longrightarrow 1$ as $n \to \infty$?

There are many convergence tests. A few more are given below (without proof). You many use any of them to prove convergence.

Theorem 8.11 (The Integral Test)

Let f be a continuous function defined on $[N, \infty)$, $N \in \mathbb{N}$, which is positive and decreasing. Then the series

$$\sum_{n=N}^{\infty} f(n)$$

converges if and only if

$$\lim_{n\to\infty}\left(\int_N^n f(x)dx\right)$$

exists.

Conditions:

Theorem 8.12 (p-series)

The p-series is the series $\sum_{n>0} \frac{1}{n^p}$ where $p > 0$.

1. *If $p > 1$, then the series converges.*
2. *If $0 < p \leq 1$ then the series diverges.*

p-series can be used to test convergence by using it as part of other tests eg. comparison test.

Theorem 8.13 (The Limit Comparison Test)

Let $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ be series with positive terms.

Consider

$$\lim_{k\to\infty} \frac{a_k}{b_k}.$$

1. *If the limit is a positive number, then $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges.*
2. *If the limit is 0 and $\sum_{n=1}^{\infty} b_n$ converges then $\sum_{n=1}^{\infty} a_n$ converges.*
3. *If the limit is ∞ (ie. unbounded) and $\sum_{n=1}^{\infty} b_n$ diverges then $\sum_{n=1}^{\infty} a_n$ diverges.*

Conditions:

Alternating Series

What about series that have some negative terms? The simplest type in this class are those with terms whose signs alternate.

Definition 8.4

Alternating series are those of the form

$$\sum_{n=1}^{\infty} a_n \qquad \text{where } a_n = (-1)^{n+1} b_n \text{ and } b_n > 0.$$

We use $(-1)^{n+1}$ rather than $(-1)^n$ so that the first term is positive.

For example,

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}.$$

Alternating series have a *very special* relationship between (A_k) and (a_n):

Theorem 8.14 (Alternating Series Test)

Let (b_n) be a non-increasing sequence with $b_n > 0$ and $b_n \to 0$. Then the alternating series

$$\sum_{n=1}^{\infty} (-1)^{n+1} b_n$$

converges.

Conditions:

Thus for alternating series it is necessary *and* sufficient for convergence, that $b_n \to 0$ (see Theorem 6.3).

Example 8.8

The series

$$\sum_{n>0} \frac{(-1)^{n+1}}{n}$$

converges.

Absolute Convergence

We have seen that "convergence implies regrouping is ok" but what about commutativity - rearranging the order of terms? This depends on the following property:

Definition 8.5 (Absolute Convergence)

The series $\sum_{n>0} a_n$ is called **absolutely convergent** if $\sum_{n>0} |a_n|$ converges.

Absolute Convergence allows us the arbitrarily change the signs of the terms:

Theorem 8.15 (The Absolute Convergence Test)

If $\sum_{n>0} a_n$ is absolutely convergent, then $\sum_{n>0} a_n$ converges.

Conditions:

The converse is false.

Note: Series for which $\sum_{n>0} a_n$ converges, but $\sum_{n>0} |a_n|$ does not, are called *conditionally convergent*.

For example,

$\sum_{n>0} \frac{(-1)^{n+1}}{n}$ converges, but $\sum_{n>0} \frac{1}{n}$ does not.

Example 8.9

Does the series $\sum_{n=1}^{\infty} \frac{\sin n}{n^2}$ converge?

We can now finally state the *rearrangement* condition:

Theorem 8.16 (Rearrangement)

Let $\sum_{n>0} a_n$ be an absolutely convergent series. Then every rearrangement of $\sum_{n>0} a_n$ converges absolutely to the same value.

This says that, absolute convergence $\Rightarrow$ the terms commute.

So we have

"Absolute convergence $\Rightarrow$ rearranging is allowed"
and
"Convergence $\Rightarrow$ regrouping is allowed"

Since

"Absolute convergence $\Rightarrow$ convergence"
then
Absolutely convergent series can be both rearranged and regrouped!

Aside

We need to define "rearrangement".

Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a bijection. If $\sum_{n>0} a_n$ and $\sum_{n>0} b_n$ are two series such that

$$\forall n \in \mathbb{N} \quad a_n = b_{\sigma(n)},$$

then $\sum_{n>0} a_n$ is a rearrangement (or permutation) of $\sum_{n>0} b_n$.

Power Series

A power series is an infinite polynomial. These series can be added, subtracted, multiplied, differentiated and integrated to give new power series.

We will study some of these series with a particular interest on if and when they converge.

Consider the *geometric series*

$$1 + r + r^2 + r^3 + \dots, \qquad r \in \mathbb{R}.$$

Example 8.10

What does it converge to?

Theorem 8.17 (Geometric Series)

The geometric series $\sum_{n\geq 0} r^n$ *converges if* $|r| < 1$ *and*

$$1 + r + r^2 + r^3 + \dots = \frac{1}{1-r}.$$

If $|r| \geq 1$ *then the geometric series diverges.*

We can generalise the geometric series by changing the coefficients of each term:

Definition 8.6 (Power Series)

A (real) **power series**, (about $r = 0$), is a series of the form

$$a_0 + a_1 r + a_2 r^2 + a_3 r^3 + \dots, \qquad a_n, r \in \mathbb{R}.$$

The sequence $(a_n)_{n\geq 0}$ is the **sequence of coefficients**.

Note, a power series conventionally starts with a_0 as then the coefficient index is the same as the exponent of r:

$$\sum_{n=0}^{\infty} a_n r^n.$$

Question: The geometric series ($a_n = 1, \ \forall n \in \mathbb{N}$) converges for $|r| < 1$. How do the coefficients change this?

Obviously convergence depends on (a_n). Thus we refine the question: Given the coeffcients (a_n) how does the convergence depend on r?

Notice that the partial sums are polynomials

$$A_k = a_0 + a_1 r + a_2 r^2 + \dots + a_k r^k,$$

thus we can think of the partial sums as *polynomial functions of* r:

$$P_k : \mathbb{R} \to \mathbb{R}, r \mapsto a_0 + a_1 r + \dots + a_k r^k.$$

So we can think of the sequence of partial sums $(A_k) = (P_k(r))$ as a sequence of functions and hence the power series

$$f(r) = \lim_{k\to\infty} P_k(r)$$

as a function obtained as a "limit of polynomial functions".

Question: How does the convergence of $(P_k(r))$ depend on r?

Question: Polynomials are continuous and differentiable for all $r \in \mathbb{R}$. What about $f(r)$?

Question: Can known functions (for example, sin, log) be expressed as power series?

First some examples.

$$f_1(r) = r - \frac{r^2}{2} + \frac{r^3}{3} - \frac{r^4}{4} + \dots \qquad a_n = \frac{(-1)^{n+1}}{n}$$

$$f_3(r) = 1 + r + \frac{1}{2!}r^2 + \frac{1}{3!}r^3 + \dots \qquad a_n = \frac{1}{n!}$$

$$f_2(r) = r - \frac{1}{3!}r^3 + \frac{1}{5!}r^5 + \dots \qquad a_n = \begin{cases} \frac{(-1)^k}{n!} & n = 2k+1 \text{ (ie. odd)} \\ 0 & n \text{ even} \end{cases}$$

Given the coefficients $(a_n)_{n\geq 0}$, how does the convergence of the series change with r?

Consider the set of all values of r for which $f(r) = \sum_{n=0}^{\infty} a_n r^n$ converges:

Let

$$S(f) = \{r \in \mathbb{R} : f(r) \text{ converges } \}.$$

- Since $\sum_{n=0}^{\infty} a_n r^n$ converges if $r = 0$ the set S is non-empty.
- If $\sum_{n=0}^{\infty} a_n r^n$ converges absolutely for some $r = s$ (ie. $\sum_{n=0}^{\infty} |a_n s^n|$ converges) then for all r satisfying $0 \leq |r| \leq |s|$ it must also converge.

Why?

Thus we can consider the "largest possible $|s|$":

$$\rho(f) = \sup\{\, |s| \, : s \in S(f)\}.$$

• If $\sum_{n=0}^{\infty} a_n r^n$ converges for all $|r|$ then of course ρ does not exist, but if not, then ρ certainly exists as S contains $s = 0$. Thus we define:

Definition 8.7 (Radius of convergence)

Let $S(f)$ be the set of all $r \in \mathbb{R}$ for which $f(r) = \sum_{n=0}^{\infty} a_n r^n$ converges.

- If S is unbounded then the series is said to have **infinite radius of convergence**.
- If S is bounded, then let

$$\rho(f) = \sup\{\, |s| \, : s \in S(f)\}.$$

and $\rho(f)$ is called **the radius of convergence** of the power series $f(r)$.

Note:

1. ρ is the sup of the *absolute values* $|s|$ so therefore $\rho \geq 0$.
2. It is called the *radius* as r can be generalised to a complex number and then ρ is the radius of a circle in the complex plane (and $\sum_{n=0}^{\infty} a_n r^n$ converges for all values inside the circle).
3. If $|r| > \rho$, then by definition of ρ, the power series diverges.
4. The set $S(f)$ may or may not contain $\pm\rho$. Thus is is always necessary to check if $f(r)$ converges when $r = \pm\rho$. For example the power series $g(r) = \sum_{n\geq 0} r^n$ has $\rho = 1$ but $g(1)$ does not converge.
5. $S(f)$ unbounded means the series converges for all $r \in \mathbb{R}$.

The radius of convergence marks the divide between values of r for which $\sum_{n=0}^{\infty} a_n r^n$ certainly converges and does not converge:

Theorem 8.18 (Power series convergence)

Let ρ be the radius of convergence of the power series $\sum_{n=0}^{\infty} a_n r^n$. Then if $|r| < \rho$ the series is absolutely convergent; if $|r| > \rho$ it is divergent.

Proof:

Example 8.11

For what values of x does

$$\sum_{n\geq 0} \frac{x^n}{n!}$$

converge?

Example 8.12

For what values of x does

$$\sum_{n\geq 0} \frac{(-1)^{n+1}}{n} x^n$$

converge?

Example 8.13

For what values of x does

$$\sum_{n\geq 0} \frac{x^n}{5^n}$$

converge?

Note that Power Series are easily generalised to the form:

$$\sum_{n=0}^{\infty} a_n(x-b)^n, \quad b \in \mathbb{R},$$

in which case the convergence condition for $|r|$ is replaced by one for $|x-b|$.

Example 8.14

$$\sum_{n\geq 0} \frac{(x-2)^n}{5^n}$$

We first define a series related to $f(x)$ and then consider if the series converges to f.

The idea is to match the derivatives of f to those of the partial sums.

Assume f is infinitely differentiable (that is, all derivatives exist). Then, if

$$A_k = \sum_{n=0}^{k} a_k x^k$$

we require

$$\left.\frac{d^n A_k}{dx^n}\right|_{x=0} = \left.\frac{d^n f}{dx^n}\right|_{x=0} \text{ for } 0 < n \leq k.$$

This will give unique values to the coefficients of $\sum_{n\geq 0} a_k x^n$.

Example 8.15

Let $f(x) = e^x$, then

Thus we get the series

$$\sum_{n\geq 0} a_n x^n = 1 + x + \tfrac{1}{2!}x^2 + \tfrac{1}{3!}x^3 + \tfrac{1}{4!}x^4 + \dots$$

Question: Does this series converge to e^x, and if so, what is the radius of convergence?

We generalise this to an arbitrary point $x = b$ rather than $x = 0$.

Definition 8.8 (Taylor Series)

Let $b \in \mathbb{R}$ and J be an open interval centred on $x = b$. Let $f : J \to \mathbb{R}$ be infinitely differentiable at $x = b$.

The **Taylor Series of f about $x = b$** is the series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(b)}{n!}(x-b)^n.$$

The partial sums

$$A_k(x) = \sum_{n=0}^{k} \frac{f^{(n)}(b)}{n!}(x-b)^n$$

are called the **Taylor Polynomials of f about $x = b$.**

Example 8.16

The Taylor polynomials for e^x are

But, does $\sum_{n\geq 0} \frac{1}{n!}x^n$ converge to e^x for $x \in (-\rho, \rho)$?

In other words, does $\lim_{k\to\infty} A_k(x) = e^x$?

To prove convergence (which has to be done every time we change f) we need to look at the error in the approximation

$$|f(x) - A_k(x)|.$$

That is, $\forall \epsilon > 0, \exists M$ such that $\forall n > M \quad |f(x) - A_n| < \epsilon$.

To help with this it would be useful if A_k, or better, $f(x) - A_k$ could be bounded:

Theorem 8.19 (Taylor's Theorem)

Let $f : (-\rho + b, b + \rho) \to \mathbb{R}$ be a $k+1$ times differentiable function on $(-\rho + b, b + \rho)$ and let

$$A_k(x) = \sum_{n=0}^{k} \frac{f^{(n)}(b)}{n!}(x - b)^n$$

be the kth partial sum of the Taylor Series of f. Then for any $x \in (-\rho + b, b + \rho)$ there exists a real number c between b and x such that

$$f(x) - A_k(x) = \frac{f^{(k+1)}(c)}{(k+1)!}(x - b)^{k+1}.$$

Notes:

1. Instead of having to work with

$$f(x) - \sum_{n=0}^{k} \frac{f^{(n)}(b)}{n!}(x - b)^n$$

   we only need to know the $(n+1)$th derivative of f.
2. $f(x) - A_k(x)$ is a measure of how well $A_k(x)$ approximates $f(x)$ - compare with the Archimedes parabola area.
3. For fixed k the approximation gets worse as x moves away from b.
4. The only requirement is that f is $k+1$ times differentiable in an open neighbourhood of $x = b$.

It is sometimes possible to estimate the remainder $R_n(x)$. This is a very useful result.

Theorem 8.20 (The Remainder Estimation Theorem)

If there is a positive constant M such that $|f^{(k+1)}(t)| \leq M$ for all t between x and b, then the remainder term $R_k(x)$ in Taylor's Theorem satisfies

$$|R_k(x)| \leq M\frac{|x - b|^{(k+1)}}{(k+1)!}.$$

If this condition holds for every $k \in \mathbb{N}$ and all conditions of Taylor's Theorem are satisfied by f, then the series converges to $f(x)$.

We have seen how Taylor series can be used to approximate functions, locally, by polynomials. We now turn our attention to Fourier series, which are useful for approximating functions on a wider interval, and can sometimes be used for discontinuous functions.

Partial Sums

In this case, our partial sums look like

$$F_k(x) = \sum_{n=0}^{k} (a_n \cos(nx) + b_n \sin(nx)).$$

Thus if

$$\lim_{k \to \infty} F_k(x)$$

exists (for $x \in J$) then so does the series

$$\sum_{n \geq 0} (a_n \cos(nx) + b_n \sin(nx)).$$

Note:

The function $\sin(x)$ is an odd function:

$$\sin(-x) = -\sin(x)$$

and $\cos(x)$ is an even function:

$$\cos(-x) = \cos(x).$$

Thus $\sum_{n=0}^{k} a_n \cos(nx)$ is even and $\sum_{n=0}^{k} b_n \sin(nx)$ is odd.

Historically:

Early 1800's, Fourier proposed a solution to heat flow in a thin infinite slab:

as a solution to the heat equation - a partial differential equation for the temperature in the $x - y$ plane: $T(t; x, y)$.

The boundary conditions are:

$T(x, 0) = f(x)$, where f is an even function.

$T(-1, y) = T(1, y) = 0$; that is, constant temperature on left and right sides.

The simplest case is $f(x) = 1$.

Fourier proposed the time independent solution

$$T(x, y) = \sum_{n \geq 0} a_n e^{-(2n-1)\frac{\pi y}{2}} \cos\left(\frac{(2n-1)\pi}{2} x\right)$$

where the a_n's depend on $f(x)$.

Thus, along the x-axis, he proposed a solution of the form:

$$f(x) = \sum_{n\geq 0} a_n \cos\left(\frac{n\pi}{2}x\right)$$

and for $f(x) = 1$, his method gave

$$1 = \frac{4}{\pi}\sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos\left(\frac{(2n-1)\pi}{2}x\right), \qquad -1 \leq x \leq 1$$

or

$$\frac{\pi}{4} = \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos\left(\frac{(2n-1)\pi}{2}x\right).$$

Let's check the partial sums:

$$F_k(x) = \sum_{n=0}^{k} \frac{(-1)^{n-1}}{2n-1} \cos\left(\frac{(2n-1)\pi}{2}x\right).$$

Note: $F_k(\pm 1)$ does fit with the temperature of the plate on the two sides, however, mathematically we are trying to find an infinite series that converges to 1 for $x \in [-1, 1]$, perhaps this is not possible (with a Fourier series) and it only converges to 1 for $x \in (-1, 1)$?

Let's use the computer to plot the partial sums; that is, the approximations to $f(x) = 1$.

$$F_k(x) = \frac{4}{\pi}\Big[\cos\left(\frac{\pi x}{2}\right) - \frac{1}{3}\cos\left(\frac{3\pi x}{2}\right) + \dots \\ \dots + \frac{(-1)^{n-1}}{2n-1}\cos\left(\frac{(2n-1)\pi}{2}x\right)\Big]$$

Computing the coefficients was given by Fourier:

Definition 8.9

Let $f : [-\pi, \pi] \to \mathbb{R}$ be a function. Then the **Fourier Coefficients** are defined as:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\ dx$$
$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\ dx$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\ dx$$

and the **Fourier Series** as

$$a_0 + \sum_{n=1}^{\infty}(a_n\cos(nx) + b_n\sin(nx)).$$

Note: The assumed property of f is that all the integrals

$$\int_{-\pi}^{\pi} f(x)\sin(nx)\ dx, \qquad \int_{-\pi}^{\pi} f(x)\cos(nx)\ dx$$

exist; that is, $f(x)\sin(nx)$ and $f(x)\cos(nx)$ are integrable.

Some books replace a_0 with $\frac{a_0}{2}$ and then *all* the integrals have a $\frac{1}{\pi}$ factor.

The big question is what does the series converge to?

The answer was given by Dirichlet:

Theorem 8.21

Let $f : [-\pi,\pi] \to \mathbb{R}$ be a bounded, piecewise continuous and piecewise monotonic function. Furthermore, assume that f is periodic with period 2π and that

$$f(x) = \frac{1}{2}\left(\lim_{h\to 0^+} f(x+h) + \lim_{h\to 0^-} f(x+h)\right)$$

for every value of x. Then

$$f(x) = a_0 + \sum_{k=1}^{\infty}(a_n\cos(nx) + b_n\sin(nx))$$

where (a_n) and (b_n) are given by Definition 8.9.

There are several assumptions for f:

1. f is integrable on $[-\pi,\pi]$
2. f is periodic with period 2π
3. If f has a discontinuity its value must be the average of the left and right limits.
4. Piecewise continuous (can be discontinuous at isolated points)
5. Piecewise monotonic

The rather strange condition

$$f(x) = \frac{1}{2}\left(\lim_{h\to 0^+} f(x+h) + \lim_{h\to 0^-} f(x+h)\right)$$

is only important where the function is *discontinuous*. If the function is continuous at x, then the left and right limits are equal:

$$\lim_{h\to 0^+} f(x+h) = \lim_{h\to 0^-} f(x+h) = f(x).$$

and the equation is redundant.

Note:

There is a more general version (where item 5 is relaxed and only requires that f is a bounded variation; that is, there exists increasing functions, g and h such that $f(x) = g(x) - h(x)$ for all $x \in [-\pi,\pi]$).

Thus Dirichlet's Theorem explains Fourier's solution with $f(x) = 1$.

Example 8.17

Example 8.18