The Language of Mathematics: A University-Level Introduction

RYAN JOO RUI AN

The Language of Mathematics:
A University-Level Introduction

Ryan Joo Rui An

Last updated: December 4, 2023

Preface

"The mathematician does not study mathematics because it is useful; he studies it because he delights in it and he delights in it because it is beautiful."

— Henri Poincaré (1854–1912)
French mathematician and theoretical physicist

About the author

At this moment of writing, I am a high school student working on my A Level studies in Singapore. I have about 11 years of participating in Mathematics competitions, including three years of experience in mental arithmetic and the rest few years in Mathematics Olympiad.

About this book

My initial purpose of writing this book was to summarise what I had learnt about university level Mathematics. The writing is intentionally kept simple, without large intimidating blocks of text, to ensure readability.

Acknowledgements

I am indebted to countless people for this work. Here is a partial (surely incomplete) list.

- Lecture notes by the University of Oxford, which can be found here.
- Lecture notes on MIT OpenCourseWare, which can be found here.
- The authors of all the books I have referred to when writing this book.

Contents

Part I

Introduction

1 Mathematical Reasoning and Logic

§1.1 Logical statements and notation

Some terminology:

- **Definition**: a precise and unambiguous description of the meaning of a mathematical term. It characterizes the meaning of a word by giving all the properties and only those properties that must be true.
- **Theorem**: a mathematical statement that is proved using rigorous mathematical reasoning. In a mathematical paper, the term theorem is often reserved for the most important results.
- **Lemma**: a minor result whose sole purpose is to help in proving a theorem. It is a stepping stone on the path to proving a theorem. Very occasionally lemmas can take on a life of their own.
- **Corollary**: a result in which the (usually short) proof relies heavily on a given theorem. We often say that "this is a corollary of Theorem A".
- **Proposition**: a proven and often interesting result, but generally less important than a theorem.
- **Conjecture**: a statement that is unproved, but is believed to be true.
- **Axiom/Postulate**: a statement that is assumed to be true without proof. These are the basic building blocks from which all theorems are proven.
- **Identity**: a mathematical expression giving the equality of two (often variable) quantities.
- **Paradox**: a statement that can be shown, using a given set of axioms and definitions, to be both true and false. Paradoxes are often used to show the inconsistencies in a flawed theory.

§1.1.1 Notation

A proposition is a sentence which has exactly one truth value, i.e. it is either true or false, but not both and not neither. A proposition is denoted by uppercase letters such as P and Q. If the proposition P depends on a variable x, it is sometimes helpful to denote it by $P(x)$.

Equivalence: $P \equiv Q$ means P and Q are logically equivalent statements.

Conjunction: $P \wedge Q$ means "P and Q".

Disjunction: $P \vee Q$ means "P or Q".

Negation: $\neg P$ means "not P".

- Double negation law:
$$P \equiv \neg(\neg P)$$
- Commutative property:
$$P \wedge Q \equiv Q \wedge P \quad P \vee Q \equiv Q \vee P$$
- Associative property for conjunction:
$$(P \wedge Q) \wedge R \equiv P \wedge (Q \wedge R)$$
- Associative property for disjunction:
$$(P \vee Q) \vee R \equiv P \vee (Q \vee R)$$
- Distributive property for conjunction across disjunction:
$$P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge Q)$$
- Distributive property for disjunction across conjunction:
$$P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R)$$
- **De Morgan's Laws**:
$$\neg(P \vee Q) \iff (\neg P \wedge \neg Q)$$
$$\neg(P \wedge Q) \iff (\neg P \vee \neg Q)$$

Example 1.1.1

Assume that x is a fixed real number. What is the negation of the statement $1 < x < 2$?

Solution. The negation of $1 < x < 2$ is "it is not the case that $1 < x < 2$". This is not useful.

Note that $1 < x < 2$ means $1 < x$ and $x < 2$. Let $P : 1 < x$ and $Q : x < 2$. Then the statement $1 < x < 2$ is $P \wedge Q$.

By De Morgan's Laws, we have $\neg(P \wedge Q) \equiv \neg P \vee \neg Q$.

Trichotomy Axiom of real numbers states that given fixed real numbers a and b, exactly one of the statements $a < b, a = b, b < a$ is true. Hence $\neg P \equiv \neg(1 < x) \equiv (x \leq 1)$ and $\neg Q \equiv \neg(x < 2) \equiv (x \geq 2)$.

Thus

$$\neg(1 < x < 2) \equiv \neg(P \wedge Q) \equiv \neg P \vee \neg Q \equiv (1 \geq x) \vee (x \geq 2).$$

Therefore the negation of $1 < x < 2$ is logically equivalent to the statement $x \leq 1$ or $x \geq 2$. $\square$

Example 1.1.2

Assume that n is a fixed positive integer. Find a useful denial of the statement

$$n = 2 \text{ or } n \text{ is odd.}$$

Solution. Using De Morgan's Laws,

$$\begin{aligned}\neg[(n = 2) \vee (n \text{ is odd})] &\equiv \neg(n = 2) \wedge \neg(n \text{ is odd})\\ &\equiv (n \neq 2) \wedge (n \text{ is even})\end{aligned}$$

where we are using the fact that every integer is either even or odd, but not both.

Thus a useful denial of the given statement is: n is an even integer other than 2. $\square$

Implication: $P \implies Q$ means "P implies Q", i.e. if P holds then Q also holds. It is equivalent to saying "If P then Q". The only case when $P \implies Q$ is false is when the hypothesis P is true and the conclusion Q is false.

- $P \implies Q \equiv (\neg P) \vee Q$.
- $\neg(P \implies Q) \equiv P \wedge (\neg Q)$.

The **converse** of $P \implies Q$ is the statement $Q \implies P$.

$$P \implies Q \not\equiv Q \implies P$$

The **contrapositive** of $P \implies Q$ is the statement $(\neg Q) \implies (\neg P)$.

$$P \implies Q \equiv (\neg Q) \implies (\neg P)$$

Bidirectional implication: $P \iff Q$ means $P \implies Q$ and $Q \implies P$. We can read this as "P if and only if Q". The letters "iff" are also commonly used to stand for 'if and only if'.

$$P \iff Q \equiv (P \implies Q) \wedge (Q \implies P)$$

- $P \iff Q$ is true exactly when P and Q have the same truth value.

Quantifiers: universal quantifier $\forall$ means "for all" or "for every", universal quantifier $\exists$ means "there exists". For example, "$\exists x \in S$ s.t. $P(x)$" can be read as "there exists x in S such that $P(x)$ holds". A common variant is $\exists!$ which means "there exists unique", implying that there is one, and only one, element with the given property.

These are versions of De Morgan's laws for quantifiers:

$$\neg \forall x\, P(x) \iff \exists x\, \neg P(x)$$
$$\neg \exists x\, P(x) \iff \forall x\, \neg P(x)$$

Example 1.1.3

Find a useful denial of the statement

$$\text{for all real numbers } x, \text{ if } x > 2, \text{ then } x^2 > 4$$

Solution. In logical notation, this statement is $(\forall x \in \mathbb{R})[x > 2 \implies x^2 > 4]$.

$$\begin{aligned} \neg\{(\forall x \in \mathbb{R})[x > 2 \implies x^2 > 4]\} &\equiv (\exists x \in \mathbb{R})\neg[x > 2 \implies x^2 > 4] \\ &\equiv (\exists x \in \mathbb{R})\neg[(x \le 2) \vee (x^2 > 4)] \\ &\equiv (\exists x \in \mathbb{R})[(x > 2) \wedge (x^2 \le 4)] \end{aligned}$$

Therefore a useful denial of the statement is:

$$\text{there exists a real number } x \text{ such that } x > 2 \text{ and } x^2 \le 4.$$

□

Remark. Regardless of how much you use these symbols in your own writing, it is important to understand and be fluent in interpreting these symbols in other people's writing.

§1.1.2 Handling logical statements

If, only if, $\implies$

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

(i) if P then Q;

(ii) P implies Q;

(iii) $P \implies Q$;

(iv) P only if Q;

(v) P is a sufficient condition for Q;

(vi) Q is a necessary condition for P;

(vii) whenever P holds, Q also holds;

(viii) if Q does not hold then P does not hold;

(ix) not Q implies not P;

(x) $\neg Q \implies \neg P$.

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

How to prove: To prove $P \implies Q$, start by assuming that P holds and try to deduce through some logical steps that Q holds too. Alternatively, start by assuming that Q does not hold and show that P does not hold (that is, we prove the contrapositive).

Remark. Note that the contrapositive is not the same as the converse. The contrapositive of $P \implies Q$ is $\neg Q \implies \neg P$; it is simply a different way of stating exactly the same thing. But the converse of $P \implies Q$ is $Q \implies P$, which means something completely different.

If and only if, iff, $\iff$

These statements are usually best thought of separately as 'if' and 'only if' statements.

How to prove: To prove $P \iff Q$, prove the statement in both directions, i.e. prove both $P \implies Q$ and $Q \implies P$. Remember to make very clear, both to yourself and in your written proof, which direction you are doing.

Quantifiers

The quantifiers $\forall$ and $\exists$ are probably the most challenging of the notation. Do practice reading statements that include these symbols and checking that you understand their meaning.

How to prove: To prove a statement of the form $\forall x \in X$ s.t. $P(x)$', start the proof with 'Let $x \in X$.' or 'Suppose $x \in X$ is given.' to address the quantifier with an arbitrary x; provided no other assumptions about x are made during the course of proving $P(x)$, this will prove the statement for all $x \in X$.

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

Remark. Read from left to right, and as new elements or statements are introduced they are allowed to depend on previously introduced elements but cannot depend on things that are yet to be mentioned.

Remark. To avoid confusion, it is a good idea to keep to the convention that the quantifiers come first, before any statement to which they relate.

Negation

For a statement P, the negated statement $\neg P$ is the statement that is false when P is true, and true when P is false. It is important to be adept at negating statements (in order to seek contradictions, for example). For a simple statement such as $x \in S$, the negation is simply $x \notin S$. For more involved statements, it can be more confusing.

For statements in the form of $P \implies Q$, the negated statement is $P \not\Longrightarrow Q$. Since $P \implies Q$ means that Q is true whenever P is true, $P \not\Longrightarrow Q$ means that (at least in some circumstance) P is true and Q is not true. Proving $P \not\Longrightarrow Q$ would typically involve demonstrating such a circumstance.

For statements involving quantifiers, the negation of $\forall x \in X, P(x)$ is $\exists x \in X, \neg P(x)$ (since, if it is not true that $P(x)$ holds for every x, then it must be the case that there is some x for which $P(x)$ does not hold). Similarly, the negation of $\exists x \in X, P(x)$ is $\forall x \in X, \neg P(x)$ (since, if it is not true that there is an x for which $P(x)$ holds, then it means that $P(x)$ does not hold for any x).[1]

[1] This is essentially an instance of De Morgan's laws.

§1.2 Proofs

§1.2.1 Direct proof

To prove $P \implies Q$ directly, we make use of P to arrive at Q through a sequence of logical reasoning. It may be that we can start from P and work directly to Q, or it may be that we make use of P along the way.

§1.2.2 Proof by contradiction

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

Example 1.2.1: Irrationality of $\sqrt{2}$

Prove that $\sqrt{2}$ is irrational.

Proof. We prove by contradiction. Suppose otherwise, that $\sqrt{2}$ is rational. Using the definition of rational numbers, we can write it as $\sqrt{2} = \frac{a}{b}$ for some $a, b \in \mathbb{Z}, b \neq 0$.

We also assume that $\frac{a}{b}$ is simplified to lowest terms, since that can obviously be done with any fraction. Notice that in order for $\frac{a}{b}$ to be in simplest terms, both a and b cannot be even; one or both must be odd, otherwise we could simplify the fraction further.

Squaring both sides gives us

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

Since RHS is even, LHS must also be even. Hence it follows that a is even. Let $a = 2k$ where $k \in \mathbb{Z}$. Substituting $a = 2k$ into the above equation and simplifying it gives us

$$b^2 = 2k^2.$$

This means that b^2 is even, from which follows again that b is even.

This is a contradiction, as we started out assuming that $\frac{a}{b}$ was simplified to lowest terms, and now it turns out that a and b both would be even. Hence proven. □

§1.2.3 Proof by induction

The following principle is sometimes quoted as a theorem, although it follows directly from our definition of the natural numbers.

Theorem 1.2.1: Principle of Induction

Let $P(n)$ be a family of statements indexed by the natural numbers. Suppose that

(i) $P(1)$ is true and

(ii) for all $m \in \mathbb{N}$, if $P(m)$ is true then $P(m+1)$ is also true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Using logic notation, this is written as

$$\{P(1) \wedge (\forall n \in \mathbb{Z}^+)[P(m) \implies P(m+1)]\} \implies (\forall n \in \mathbb{Z}^+)P(n)$$

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

Example 1.2.2

Prove that for any $n \in \mathbb{N}$,

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

Proof. Clearly $P(1)$ holds because for $n = 1$, the sum on the LHS is 1 and the expression on the RHS is also 1.

Now suppose $P(n)$ holds. Then

$$\begin{aligned}\sum_{k=1}^{n+1} k &= \sum_{k=1}^{n} k + (n+1)\\ &= \frac{n(n+1)}{2} + (n+1)\\ &= \frac{(n+1)(n+2)}{2}\end{aligned}$$

which is exactly the statement $P(n+1)$. So by induction, $P(n)$ is true for all $n \in \mathbb{N}$. □

A corollary[2] of induction is if the family of statements holds for $n \geq N$, rather than necessarily $n \geq 0$:

Corollary 1. Let N be an integer and let $P(n)$ be a family of statements indexed by integers $n \geq N$. Suppose that

(i) $P(N)$ is true and

[2] an extension of, or a consequence of, a theorem or proposition; a corollary is generally not such a major result as the theorem or proposition itself.

(ii) for any $n \geq N$, if $P(n)$ is true then $P(n+1)$ is also true.

Then $P(n)$ is true for all $n \geq N$.

Proof. This follows directly by applying the above theorem to the statement $Q(n) = P(n+N)$ for $n \in N$. □

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

Theorem 1.2.2: Strong Form of Induction

Let $P(n)$ be a family of statements indexed by the natural numbers. Suppose that

(i) $P(1)$ is true and

(ii) for all $m \in \mathbb{N}$, if for integers k with $1 \leq k \leq m$, $P(k)$ is true then $P(m+1)$ is true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Using logic notation, this is written as

$$\{P(1) \wedge (\forall m \in \mathbb{Z}^+)[P(1) \wedge P(2) \wedge \cdots \wedge P(m) \implies P(m+1)]\} \implies (\forall n \in \mathbb{Z}^+)P(n)$$

Proof. We can this it to an instance of “normal" induction by defining a related family of statements $Q(n)$.

Let $Q(n)$ be the statement “$P(k)$ holds for $k = 0, 1, \ldots, n$". Then the conditions for the strong form are equivalent to

(i) $Q(0)$ holds and

(ii) for any n, if $Q(n)$ is true then $Q(n+1)$ is also true.

It follows by induction that $Q(n)$ holds for all n, and hence $P(n)$ holds for all n. □

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

Example 1.2.3: Fundamental Theorem of Arithmetic

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

Proof. Let $P(n)$ be the statement that n may be expressed as a product of prime numbers.

Clearly $P(2)$ holds, since 2 is itself prime.

Let $n \geq 2$ be a natural number and suppose that $P(m)$ holds for all $m < n$.

- If n is prime then it is trivially the product of the single prime number n.

- If n is not prime, then there must exist some $r, s > 1$ such that $n = rs$. By the inductive hypothesis, each of r and s can be written as a product of primes, and therefore $n = rs$ is also a product of primes.

Thus, whether n is prime or not, we have have that $P(n)$ holds. By strong induction, $P(n)$ is true for all natural numbers. That is, every natural number greater than 1 may be expressed as a product of one or more primes. □

Cauchy induction

Problem 1.2.1 (FM/TJC/2023)**.** Prove by mathematical induction, for $n \geq 2$,

$$\sqrt[n]{n} < 2 - \frac{1}{n}.$$

Proof. Let $P(n)$ be the proposition that $\sqrt[n]{n} < 2 - \frac{1}{n}$ for $n \geq 2$.

When $n = 2$, $\sqrt{2} < 2 - \frac{1}{2} = 1.5$ which is true. Hence $P(2)$ is true.

Assume $P(k)$ is true for $k \geq 2, k \in \mathbb{Z}^+$, i.e.

$$\sqrt[k]{k} < 2 - \frac{1}{k} \implies k < \left(2 - \frac{1}{k}\right)^k$$

We want to prove that $P(k+1)$ is true, i.e.

$$k + 1 < \left(2 - \frac{1}{k+1}\right)^{k+1}$$

Since $k > 2$, we have

$$\begin{aligned}
\left(2 - \frac{1}{k+1}\right)^{k+1} &> \left(2 - \frac{1}{k}\right)^{k+1} \qquad \because k > 2 \\
&= \left(2 - \frac{1}{k}\right)^k \left(2 - \frac{1}{k}\right) \\
&> k\left(2 - \frac{1}{k}\right) \quad \text{by inductive hypothesis} \\
&= 2k - 1 = k + k - 1 > k - 1 \because k > 2
\end{aligned}$$

Hence $P(k+1)$ is true.

Since $P(2)$ is true and $P(k) \implies P(k+1)$, by mathematical induction $P(n)$ is true. $\square$

Problem 1.2.2. Prove that for all integers $n \geq 3$,

$$\left(1 + \frac{1}{n}\right)^n < n$$

Proof. Suppose for an integer k, we have

$$\left(1 + \frac{1}{k}\right)^k < k$$

Then

$$\left(1 + \frac{1}{k}\right)^k \left(1 + \frac{1}{k}\right) = \left(1 + \frac{1}{k}\right)^{k+1} < k\left(1 + \frac{1}{k}\right) = k + 1$$

Note

$$\left(1 + \frac{1}{k}\right)^{k+1} > \left(1 + \frac{1}{k+1}\right)^{k+1} \quad \text{since } k < k + 1 \iff \frac{1}{k} > \frac{1}{k+1}$$

The rest of the proof follows easily. □

§1.2.4 Counterexamples

Providing a counterexample is the best method for refuting, or dispoving, a conjecture.

In seeking counterexamples, it is a good idea to keep the cases you consider simple, rather than searching randomly. It is often helpful to consider "extreme" cases; for example, something is zero, a set is empty, or a function is constant.

2 Set Theory

§2.1 Basics

§2.1.1 Notation

You should, by now, be familiar with the following definitions and notation:

- A **set** S can be loosely defined as a collection of objects.
- For a set S, we write $x \in S$ to mean that x is an **element** of S, and $x \notin S$ if otherwise.
- A set can be defined in terms of some property $P(x)$ that the elements $x \in S$ satisfy, denoted by
$$\{x \in S \mid P(x)\}$$
- Some basic sets (of numbers) you should be familiar with:
  - $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ denotes the natural numbers (non-negative integers).
  - $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ denotes the integers.
  - $\mathbb{Q} = \{\frac{p}{q} \mid p, q \in \mathbb{Z}, q \neq 0\}$ denotes the rational numbers.
  - $\mathbb{R}$ denotes the real numbers, which can be expressed in terms of decimal expansion.
  - $\mathbb{C} = \{x + yi \mid x, y \in \mathbb{R}\}$ denotes the of complex numbers.
- The **empty set** is the set with no elements, denoted by $\varnothing$.
- A is a **subset** of B if every element of A is in B, denoted by $A \subseteq B$.
$$A \subseteq B \iff \forall x, x \in A \implies x \in B$$

  $\subseteq$ is transitive, i.e. if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

  Proof. Let $x \in A$. Since $A \subseteq B$ and $x \in A$, $x \in B$. Since $B \subseteq C$ and $x \in B$, $x \in C$. Hence $A \subseteq C$. □

A is a **proper subset** of B if $A \subseteq B$ and $A \neq B$, denoted by $A \subset B$.

Using this definition, we have the relationship

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

- A and B are **equal** if and only if they contain the same elements, denoted by $A = B$.

  To prove that A and B are equal, we simply need to prove that $A \subseteq B$ and $A \subseteq B$.

  Proof. We have

$$\begin{aligned} A = B &\iff (\forall x)[x \in A \iff x \in B] \\ &\iff (\forall x)[(x \in A \implies x \in B) \wedge (x \in B \implies x \in A)] \\ &\iff \{(\forall x)[x \in A \implies x \in B]\} \wedge (\forall x)[x \in B \implies x \in A)] \\ &\iff (A \subseteq B) \wedge (B \subseteq A) \end{aligned}$$

  □

- Some frequently occurring subsets of the real numbers are known as **intervals**, which can be visualised as sections of the real line:
  - Open interval
  $$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$$
  - Closed interval
  $$[a, b] = \{x \in \mathbb{R} \mid a \leq x < b\}$$
  - Half open interval
  $$(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$$

- The **power set** $\mathcal{P}(A)$ of A is the set of all subsets of A (including the set itself and the empty set).

- An **ordered pair** is denoted by (a, b), where the order of the elements matters. Two pairs (a_1, b_1) and (a_2, b_2) are equal if and only if $a_1 = a_2$ and $b_1 = b_2$.

  Similarly, we have ordered triples (a, b, c), quadruples (a, b, c, d) and so on. If there are n elements it is called an n-tuple.

- The **Cartesian product** of sets A and B, denoted by $A \times B$, is the set of all ordered pairs with the first element of the pair coming from A and the second from B:

$$A \times B = \{(a, b) \mid a \in A, b \in B\} \tag{2.1}$$

  If $A = B$, we write $A \times A$ as A^2. Note that the case where $A = B = \mathbb{R}$ is a particularly important one as $\mathbb{R}^2$ represents the two-dimensional real plane.

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

Example 2.1.1

$\mathbb{R}^2$ is the Euclidean plane, $\mathbb{R}^3$ is the Euclidean space, and $\mathbb{R}^n$ is the n-dimensional Euclidean space.

$$\begin{aligned} \mathbb{R} \times \mathbb{R} = \mathbb{R}^2 &= \{(x, y) \mid x, y \in \mathbb{R}\} \\ \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3 &= \{(x, y, z) \mid x, y, z \in \mathbb{R}\} \\ \mathbb{R}^n &= \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in \mathbb{R}\} \end{aligned}$$

§2.1.2 Algebra of Sets

Given $A \subset S$ and $B \subset S$.

- The **union** $A \cup B$ is the set consisting of elements that are in A or B (or both):

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

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

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

A and B are **disjoint** if both sets have no element in common:

$$A \cap B = \varnothing$$

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

$$\bigcup_{i \in I} A_i = \{x \mid \exists i \in I\,(x \in A_i)\}$$

and

$$\bigcap_{i \in I} A_i = \{x \mid \forall i \in I\,(x \in A_i)\}$$

- The **complement** of A, denoted by A^c or A', is the set containing elements that are not in A:

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

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

$$A \setminus B = \{x \in A \mid x \notin B\}$$

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

Proposition 2.1.1 (Double Inclusion)**.** Let $A \subset S$ and $B \subset S$. Then

$$A = B \iff A \subseteq B \text{ and } B \subseteq A \tag{2.2}$$

Proof. We prove both directions.

Forward direction:

If $A = B$, then every element in A is an element in B, so certainly $A \subseteq B$, and similarly $B \subseteq A$.

Backward direction:

Suppose $A \subseteq B$, and $B \subseteq A$. Then for every element $x \in S$, if $x \in A$ then $A \subseteq B$ implies that $x \in B$, and if $x \notin A$ then $B \subseteq A$ means $x \notin B$. So $x \in A$ if and only if $x \in B$, and therefore $A = B$. □

Proposition 2.1.2 (Distributive Laws)**.** Let $A \subset S$, $B \subset S$ and $C \subset S$. Then

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \tag{2.3}$$

$$(A \cap B) \cap C = (A \cup C) \cap (B \cup C) \tag{2.4}$$

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

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

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

The proof of the second one follows similarly and is left as an exercise. □

Proposition 2.1.3 (De Morgan's Laws)**.** Let $A \subset S$ and $B \subset S$. Then

$$(A \cup B)^c = A^c \cap B^c \tag{2.5}$$

$$(A \cap B)^c = A^c \cup B^c \tag{2.6}$$

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

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

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

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

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

Problem 2.1.1. Let A be the set of all complex polynomials in n variables. Given a subset $T \subset A$, define the *zeros* of T as the set

$$Z(T) = \{P = (a_1, \dots, a_n) \mid f(P) = 0 \text{ for all } f \in T\}$$

A subset $Y \in \mathbb{C}^n$ is called an algebraic set if there exists a subset $T \subset A$ such that $Y = Z(T)$.

Prove that the union of two algebraic sets is an algebraic set.

Proof. We would like to consider $T = \{f_1, f_2, \dots\}$ expressed as indexed sets $T = \{f_i\}$. Then $Z(T)$ can also be expressed as $\{P \mid \forall i, f_i(P) = 0\}$.

Suppose that we have two algebraic sets X and Y. Let $X = Z(S)$, $Y = Z(T)$ where S, T are subsets of A (basically, they are certain sets of polynomials). Then

$$X = \{P \mid \forall f \in S, f(P) = 0\}$$
$$Y = \{P \mid \forall g \in T, g(P) = 0\}$$

We imagine that for $P \in X \cap Y$, we have $f(P) = 0$ or $g(P) = 0$. Hence we consider the set of polynomials

$$U = \{f \cdot g \mid f \in S, g \in T\}$$

For any $P \in X \cup Y$ and for any $fg \in U$ where $f \in S$ and $f \in g$, either $f(P) = 0$ or $g(P) = 0$, hence $fg(P) = 0$ and thus $P \in Z(U)$.

On the other hand if $P \in Z(U)$, suppose otherwise that P is not in $X \cup Y$, then P is neither in X nor in Y. This means that there exists $f \in S, g \in T$ such that $f(P) \neq 0$ and $g(P) \neq 0$, hence $fg(P) \neq 0$. This is a contradiction as $P \in Z(U)$ implies $fg(P) = 0$. Hence we have $X \cup Y = Z(U)$ and thus $X \cup Y$ is an algebraic set.

Now the other direction is simpler and can actually be generalised: The intersection of arbitrarily many algebraic sets is algebraic.

The basic result is that if $X = Z(S)$ and $Y = Z(T)$ then $X \cap Y = Z(S \cup T)$.

This is the beginning of a subject called algebraic geometry, which discusses about the geometry of algebraic surfaces, namely those kinds of surfaces defined by zeroes of polynomials (conic sections for example). □

§2.1.3 Cardinality

Informally, the **cardinality** of S, denoted $|S|$, is a measure of its "size".

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

Definition 2.1.1: Finiteness and the cardinality of a finite set

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

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

Proposition 2.1.4. Let A and B be finite sets. Then $|A \cup B| = |A| + |B| - |A \cap B|$.

Proof. The proof is left as an exercise. □

Proposition 2.1.5 (Subsets of a finite set)**.** If a set A is finite with $|A| = n$, then its power set has $|\mathcal{P}(A)| = 2^n$.

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

Now suppose that $n \geq 0$ and that $|P(S)| = 2^n$ for any set S with $|S| = n$. Let A be any set with $|A| = n + 1$. By definition, this means that there is an element a and a set $A_0 = A \setminus \{a\}$ with $|A_0| = n$. Any subset of A must either contain the element a or not, so we can partition $\mathcal{P}(A) = P(A_0) \cup \{S \cup \{a\} \mid S \in P(A_0)\}$. These two sets are disjoint, and each of them has cardinality $|P(A_0)| = 2^n$ by the inductive hypothesis. Hence $|\mathcal{P}(A)| = 2^n + 2^n = 2^{n+1}$.

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

Another way to see this is through combinatorics: Consider the process of creating a subset. We can do this systematically by going through each of the $|A|$ elements in A and making the yes/no decision whether to put it in the subset. Since there are $|A|$ such choices, that yields $2^{|A|}$ different combinations of elements and therefore $2^{|A|}$ different subsets.

§2.2 Relations

§2.2.1 Definition

A relation is a set of ordered pairs.

Definition 2.2.1: Relation

R is a **relation** between A and B if and only if R is a subset of the Cartesian product $A \times B$, i.e. $R \subseteq A \times B$.
$a \in A$ and $b \in B$ are **related** if $(a, b) \in R$, denoted by aRb.

Visually speaking, a relation is uniquely determined by a simple bipartite graph over A and B. On the bipartite graph, this is usually represented by an edge between a and b.

Definition 2.2.2: Binary relation

A **binary relation** in A is a relation between A and itself, i.e. $R \subseteq A \times A$.

A and B are the **domain** and **range** of R respectively, denoted by $\operatorname{dom} R$ and $\operatorname{ran} R$ respectively, if and only if $A \times B$ is the smallest Cartesian product of which R is a subset.

Example 2.2.1

$R = \{(1, a), (1, b), (2, b), (3, b)\}$, then

- $\operatorname{dom} R = \{1, 2, 3\}$
- $\operatorname{ran} R = \{a, b\}$

In many cases we do not actually use R to write the relation because there is some other conventional notation:

Example 2.2.2

- The "less than or equal to" relation $\leq$ on the set of real numbers is $\{(x, y) \in \mathbb{R}^2 \mid x \leq y\}$. We write $x \leq y$ if (x, y) is in this set.
- The "divides" relation $\mid$ on $\mathbb{N}$ is $\{(m, n) \in \mathbb{N}^2 : m \text{ divides } n\}$. We write $m \mid n$ if (m, n) is in this set.
- For a set S, the "subset" relation $\subseteq$ on $\mathcal{P}(S)$ is $\{(A, B) \in \mathcal{P}(S)^2 \mid A \subseteq B\}$. We write $A \subseteq B$ if (A, B) is in this set.

§2.2.2 Properties of relations

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

- R is **reflexive** if xRx for all x in A.
- R is **symmetric** if whenever xRy then yRx.
- R is **anti-symmetric** if whenever xRy and yRx then $x = y$.
- R is **transitive** if whenever xRy and yRz then xRz.

Example 2.2.3: Less than or equal to

The relation $\leq$ on R is reflexive, anti-symmetric, and transitive, but not symmetric.

More generally, any relation on A that is reflexive, anti-symmetric, and transitive is called a **partial order**, denoted by $\leq$. It is called a **total order** if for every $x, y \in A$, either xRy or yRx (or both).

Example 2.2.4: Less than

The relation $<$ on R is not reflexive, symmetric, or anti-symmetric, but it is transitive.

Example 2.2.5: Not equal to

The relation $\neq$ on R is not reflexive, anti-symmetric or transitive, but it is symmetric.

Example 2.2.6: Congruence modulo n

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

Proof.

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

$\square$

§2.2.3 Equivalence relations, equivalence classes, and partitions

One important type of relation is an equivalence relation. An equivalence relation is a way of saying two objects are, in some particular sense, "the same".

Definition 2.2.3: Equivalence relation

A binary relation R on A is an **equivalence relation** if it is reflexive, symmetric and transitive. If R is an equivalence relation, we denote it by $\sim$.

Remark. We use the symbol $\sim$ to denote the equivalence relation R in $A \times A$, and whenever $(a, b) \in R$ we denote $a \sim b$.

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

Definition 2.2.4: Equivalence class

Given an equivalence relation $\sim$ on a set A, and given $x \in A$, the **equivalence class** of x, denoted $[x]$, is the subset

$$[x] = \{y \in A \mid y \sim x\}$$

Example 2.2.7: Congruence modulo n

For the equivalence relation of congruence modulo n, the equivalence class of 1 is the set $1 = \{\dots, -n+1, 1, n+1, 2n+1, \dots\}$; that is, all the integers that are congruent to 1 modulo n.

Properties of equivalence classes:

- Every two equivalence classes are disjoint
- The union of equivalence classes form the entire set

You can translate these properties into the point of view from the elements: Every element belongs to one and only one equivalence class.

- No element belongs to two distinct classes
- All elements belong to an equivalence class

Definition 2.2.5: Set of equivalence classes

The **set of equivalence classes** (quotient sets) are the set of all equivalence classes, denoted by $A/\sim$.

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

Definition 2.2.6: Partition

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

(i) $A_i \neq \varnothing$ for all $i \in I$ (that is, all the subsets are non-empty),

(ii) $\bigcup_{i \in I} Ai = A$ (that is, every member of A lies in one of the subsets),

(iii) $A_i \cap A_j = \varnothing$ for every $i \neq j$ (that is, the subsets are disjoint).

The subsets are called the parts of the partition.

Example 2.2.8: Odd and even natural numbers

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

Problem 2.2.1 (Modular Arithmetic)**.** Define the ring of integers modulo n:

$$\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}/\sim \text{ where } x \sim y \iff x - y \in n\mathbb{Z}.$$

The equivalence classes are called congruence classes modulo n.

(a) Define the sum of two congruence classes modulo n, $[x], [y] \in \mathbb{Z}/n\mathbb{Z}$, by

$$[x] + [y] = [x + y]$$

Show that the above definition is well-defined.

(b) Define the product of two congruence classes modulo n and show that such a definition is well-defined.

Solution.

(a) We often define such concepts by considering the **representatives** of the equivalence classes.

For example, here we define $[x] + [y] = [x + y]$ for two elements $[x]$ and $[y]$ in $\mathbb{Z}/n\mathbb{Z}$. So what we know here are the classes $[x]$ and $[y]$. But what exactly are x and y? They are just some element in the equivalence classes that was arbitrarily picked out. We then perform the sum $x+y$, and consequently, we used this to point towards the class $[x + y]$.

However, x and y are arbitrarily picked. We want to show that, regardless of which representatives are chosen from the equivalence classes $[x]$ and $[y]$, we will always obtain the same result.

In the definition itself, we have defined that, for the two representatives x and y we define $[x] + [y] = [x + y]$. So now, let's say that we take two other arbitrary representatives, $x' \in [x]$ and $y' \in [y]$. Then by definition, we have

$$[x] + [y] = [x' + y']$$

Thus, our goal is to show that $x' + y'] = [x + y]$. This expression means that the two sides of the equation are referring to the same equivalence class. Therefore, the expression above is completely equivalent to the condition.

$$x' + y' \sim x + y$$

We then check that this final expression is indeed true: Since $x' \in [x]$ and $y' \in [y]$, we have

$$\begin{aligned} & x' \sim x \text{ and } y' \sim y \\ \implies & x' - x, y' - y \in n\mathbb{Z} \\ \implies & (x' + y') - (x + y) = (x' - x) + (y' - y) \in n\mathbb{Z} \end{aligned}$$

(b) The product of two congruence classes is defined by

$$[x][y] = [xy]$$

For any other representatives x', y' we have

$$\begin{aligned}
&x'y' - xy \\
&= x'y' - xy' + xy' - xy \\
&= (x' - x)y' + x(y' - y) \in n\mathbb{Z}
\end{aligned}$$

Thus $[x'y'] = [xy]$ and the product is well-defined.

$\square$

Problem 2.2.2. Let $A = \mathbb{R}$ and for any $x, y \in A$, $x \sim y$ if and only if $x - y \in \mathbb{Z}$. For any two equivalence classes $[x], [y] \in A/\sim$, define

$$[x] + [y] = [x + y] \text{ and } -[x] = [-x]$$

(a) Show that the above definitions are well-defined.

(b) Find a one-to-one correspondence $\varphi : X \to Y$ between $X = A/\sim$ and $Y : |z| = 1$, i.e. the unit circle in $\mathbb{C}$, such that for any $[x_1], [x_2] \in X$ we have

$$\varphi([x_1])\varphi([x_2]) = \varphi([x_1 + x_2])$$

(c) Show that for any $[x] \in X$,

$$\varphi(-[x]) = \varphi([x])^{-1}$$

Solution.

(a)

$$(x' + y') - (x + y) = (x' - x) + (y' - y) \in \mathbb{Z}$$

Thus $[x' + y'] = [x + y]$

$$(-x') - (-x) = -(x' - x) \in \mathbb{Z}$$

Thus $[-x'] = [-x]$.

(b) Complex numbers in the polar form: $z = re^{i\theta}$

Then the correspondence is given by $\varphi([x]) = e^{2\pi ix}$

$$[x] = [y] \iff x - y \in \mathbb{Z} \iff e^{2\pi i(x-y)} = 1 \iff e^{2\pi ix} = e^{2\pi iy}$$

Hence this is a bijection.

Before that, we also need to show that φ is well-defined, which is almost the same as the above.

If we choose another representative x' then

$$\varphi([x]) = e^{2\pi ix'} = e^{2\pi ix} \cdot e^{2\pi i(x'-x)} = e^{2\pi ix}$$

(c) You can either refer to the specific correspondence $\varphi([x]) = e^{2\pi ix}$ or use its properties.

$$\varphi(-[x])\varphi([x]) = \varphi([-x])\varphi([x]) = \varphi([-x + x]) = \varphi([0]) = 1$$

□

Problem 2.2.3 (Set of Rational Numbers)**.** Let $\mathbb{Z}$ be the set of integers, and let $\mathbb{Z}^*$ be the set of nonzero integers. We define

$$\mathbb{Q} = \{(a, b) \mid a \in \mathbb{Z}, b \in \mathbb{Z}^*\} / \sim$$

where

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

Let $\dfrac{a}{b}$ denote the equivalence class for (a, b). Such an equivalence class is called a rational number.

(a) For any two rational numbers $\dfrac{a}{b}$ and $\dfrac{c}{d}$, their sum is determined by

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

Show that the above definition is well-defined.

(b) Define the product of two rational numbers and show that such a definition is well-defined.

(c) Prove that for every equivalence class $\dfrac{a}{b} \in \mathbb{Q}$, there exists a unique integer pair (p, q) satisfying the following properties:

$$q > 0, (p, q) = 1 \text{ and } (p, q) \in \frac{a}{b}.$$

(d) Using the partial order of $\mathbb{Z}$, define the partial order of $\mathbb{Q}$.

Solution.

(a) For this problem, we are dealing with a “hidden" equivalence class.

The expressions $\frac{a}{b}$ and $\frac{c}{d}$ themselves are derived from their representatives (a, b) and (c, d).

So suppose that we choose other representatives (a', b') and (c', d'), then the sum would be

$$\frac{a'}{b'} + \frac{c'}{d'} = \frac{a'd' + b'c'}{b'd'}$$

We now have to show that $\frac{ad+bc}{bd} = \frac{a'd'+b'c'}{b'd'}$:

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

a$\overline{\scriptstyle b=a'/b'(a,b)\sim(a',b')ab'=a'b}$

(ad+bc)b’d’ =ab’dd’+bb’cd’ =a’bdd’+bb’c’d =(a’d’+b’c’)bd

(b) The definition would be $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$.

This is actually a lot simpler to check.

$$a'c'bd = (a'b)(c'd) = (ab')(cd') = acb'd'$$

Hence $\frac{a'c'}{b'd'} = \frac{ac}{bd}$.

(c) We basically try to do this step by step as we would in simplifying fractions.

First pick b to be positive, otherwise we swap a and b with $-a$ and $-b$.

Then simplify the common factors. For this one we let $(a, b) = d$, and $a = dp, b = dq$. Then (p, q) is the pair that we need

(d) In order to define the partial order we need to account for whether the denominators are negative.

$\frac{a}{b} \leq \frac{c}{d}$, and if $b, d > 0$ then we can safely draw a connection to the expression $ad \leq bc$

In order to show that this does in fact give a partial order we check that

(a) 1: $ab \leq ab$ and hence $\frac{a}{b} \leq \frac{a}{b}$

(b) 2: If $\frac{a}{b} \leq \frac{c}{d}$ and $\frac{c}{d} \leq \frac{a}{b}$, then $ad \leq bc$ and $bc \leq ad$, hence $ad = bc$ and thus $\frac{a}{b} = \frac{c}{d}$

(c) 3: This is trickier due to complications arising from inequalities and multiplication

If $\frac{a}{b} \leq \frac{c}{d}$ and $\frac{c}{d} \leq \frac{e}{f}$, note that $b, d, f > 0$ and so $ad \leq bc$ and $cf \leq de$.

i) $e < 0$, then $c < 0$ and $a < 0$, thus $-ad \geq -bc$, $-cf \geq -de$ and we have $acdf \geq bcde$ $af \leq be(c < 0, d > 0)$ Thus $\frac{a}{b} \leq e/f$

ii) $e \geq 0$ but $a < 0$, then $af < 0 \leq be$ and thus $\frac{a}{b} < e/f$

iii) $a \geq 0$, then $c \geq 0$ and $e \geq 0$, and we have the ordinary case.

Hence proven.

□

§2.3 Functions

§2.3.1 Definition

Definition 2.3.1: Function

Given two sets X and Y, a **function** f from X to Y is a mapping of every element of X to some element of Y, denoted by $f : X \to Y$.
We call X and Y the **domain** and **codomain** of f respectively.

Remark. The definition requires that a unique element of the codomain is assigned for every element of the domain. For example, for a function $f : \mathbb{R} \to \mathbb{R}$, the assignment $f(x) = \frac{1}{x}$ is not sufficient as it fails at $x = 0$. Similarly, $f(x) = y$ where $y^2 = x$ fails because $f(x)$ is undefined for $x < 0$, and for $x > 0$ it does not return a unique value; in such cases, we say the the function is **ill-defined**. We are interested in the opposite; functions that are **well-defined**.

Definition 2.3.2: Image and pre-image

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

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

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

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

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

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

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

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

Definition 2.3.3: Restriction

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

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

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

Definition 2.3.4: Identity map

Given a set X, the **identity** $\mathrm{id}_X : X \to X$ is defined by $\mathrm{id}_X(x) = x$ for all $x \in X$.

Notation. If the domain is unambiguous, the subscript may be removed.

§2.3.2 Injectivity, Surjectivity, Bijectivity

Definition 2.3.5: Injectivity

$f : X \to Y$ is **injective** if each element of Y has at most one element of X that maps to it.

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

Proposition 2.3.1. If $f : X \to Y$ is injective and $g : Y \to Z$ is injective, then $g \circ f : X \to Z$ is injective.

Proof. Let $f : X \to Y$ and $g : Y \to Z$ be arbitrary injective functions. We want prove that the function $g \circ f : X \to Z$ is also injective.

To do so, we will prove $\forall x, x' \in X$ that

$$(g \circ f)(x) = (g \circ f)(x') \implies x = x'$$

Suppose that $(g \circ f)(x) = (g \circ f)(x')$. Expanding out the definition of $g \circ f$, this means that $g(f(x)) = g(f(x'))$.

Since g is injective and $g(f(x)) = g(f(x'))$, we know $f(x) = f(x')$.

Similarly, since f is injective and $f(x) = f(x')$, we know that $x = x'$, as required. □

Definition 2.3.6: Surjectivity

$f : X \to Y$ is **surjective** if *every* element of Y is mapped to at least one element of X.

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

Proposition 2.3.2. If $f : X \to Y$ is surjective and $g : Y \to Z$ is surjective, then $g \circ f : X \to Z$ is surjective.

Proof. Let $f : X \to Y$ and $g : Y \to Z$ be arbitrary surjective functions. We want to prove that the function $g \circ f : X \to Z$ is subjective.

To do so, we want to prove that for any $z \in Z$, there is some $x \in X$ such that $(g \circ f)(x) = z$. Equivalently, we want to prove that for any $z \in Z$, there is some $x \in X$ such that $g(f(x)) = z$.

Consider any $z \in Z$. Since $g : Y \to Z$ is surjective, there is some $y \in Y$ such that $g(y) = z$. Similarly, since $f : X \to Y$ is surjective, there is some $x \in X$ such that $f(x) = y$. This means that there is some $x \in X$ such that $g(f(x)) = g(y) = z$, as required. □

Definition 2.3.7: Bijectivity

$f : X \to Y$ is **bijective** if it is both injective and surjective: each element of Y is mapped to a unique element of X.

§2.3.3 Cardinality and countable sets

When do two sets have the same size? Cantor answered this question in the 1800s, stating that two sets have the same size when you can pair each element in one set with a unique element in the other.

Definition 2.3.8: Cardinality

X and Y have the same **cardinality** if there exists a bijection $f : X \to Y$, denoted by $|X| = |Y|$.

Definition 2.3.9: Cardinality of finite sets

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

Remark. Note that for finite sets X and Y, a function $f : X \to Y$ can only be **injective** if $|Y| \geq |X|$, since for any injective function the number of elements in the image $f(X)$, is equal to the number of elements in the domain, and $f(X) \subseteq Y$. In other words, the codomain of an injective function cannot be smaller than the domain.[1]

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

Theorem 2.3.1: Cantor–Schröder–Bernstein

If $|X| \leq |Y|$ and $|Y| \leq |X|$ then $|X| = |Y|$.

Definition 2.3.10: Countably infinite set

A set X is **countably infinite** if it has the same cardinality as the set $\mathbb{Z}^+$ or $\mathbb{N}$.

Definition 2.3.11: Countable set

A set X is **countable** if it is either finite or infinitely countable.

Example 2.3.1

$\{2n \mid n \in \mathbb{N}\}$ is countably infinite, i.e. $|\{2n \mid n \in \mathbb{N}\}| = |\mathbb{N}|$.
To prove this, define the function $f : \mathbb{N} \to \{2n \mid n \in \mathbb{N}\}$ as $f(n) = 2n$. Then, f is injective – if $f(n) = f(m)$ then $2n = 2m \implies n = m$. Furthermore, f is surjective, as if $m \in \{2n \mid n \in \mathbb{N}\}$ then $\exists n \in \mathbb{N}$ such that $m = 2n = f(n)$.

[1] This is sometimes referred to as the pigeonhole principle: if n letters are placed in m pigeonholes and $n > m$, then at least one hole must contain more than one letter; the non-injective function in that case is the assignment of pigeonholes to letters.

Example 2.3.2

$\mathbb{Z}^+$ is countable since we have a bijection $f : \mathbb{Z}^+ \to \mathbb{Z}$ given by

$$f(k) = \begin{cases} \dfrac{k}{2} & \text{if } k \text{ is even} \\ \dfrac{1-k}{2} & \text{if } k \text{ is odd} \end{cases}$$

In other words, $f(1) = 0, f(2) = 1, f(3) = -1, f(4) = 2, f(5) = -2, f(6) = -3, \ldots$ where our function f stretches in both positive and negative directions.

Theorem 2.3.2: Cantor

If A is a set, then $|A| < |\mathcal{P}(A)|$.

Proof. Define the function $f : A \to \mathcal{P}(A)$ by $f(x) = \{x\}$. Then, f is injective as $\{x\} = \{y\} \implies x = y$. Thus $|A| \le |\mathcal{P}(A)|$. To finish the proof now all we need to show is that $|A| \ne |\mathcal{P}(A)|$. We will do so through contradiction. Suppose that $|A| = |\mathcal{P}(A)|$. Then, there exists a surjection $g : A \to \mathcal{P}(A)$. We define the set B as

$$B := \{x \in A \mid x \notin g(x)\} \in \mathcal{P}(A)$$

Since g is surjective, there exists a $b \in A$ such that $g(b) = B$. There are two cases:

1. $b \in B$. Then $b \notin g(b) = B \implies b \notin B$.
2. $b \notin B$. Then $b \notin g(b) = B \implies b \in B$.

In either case we obtain a contradiction. Thus, g is not surjective so $|A| \ne |\mathcal{P}(A)|$. □

Corollary 2. For all $n \in \mathbb{N} \cup \{0\}$, $n < 2^n$.

Proof. This can be easily proven through induction. □

Lemma 2.3.1. If X is a countable set and $B \subseteq A$ then B is countable.

Lemma 2.3.2. If $\{A_1, A_2, \ldots\}$ is a collection of countably many countable sets then the set $\bigcup_{i=1}^{\infty} A_i$ is countable.

Lemma 2.3.3. If $\{A_1, A_2, \ldots, A_n\}$ is a collection of finitely many countable sets then the set $A_1 \times \cdots \times A_n$ is countable.

§2.3.4 Composition of functions and invertibility

Definition 2.3.12: Composition

Given two functions $f : X \to Y$ and $g : Y \to Z$, the composition $g \circ f : X \to Z$ is defined by

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

The composition of functions is not commutative, i.e. $f \circ g \neq g \circ f$. For example, if $f(x) = x^2$ and $g(x) = e^x$ are both maps from $\mathbb{R}$ to $\mathbb{R}$, then

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

However, composition is associative, as the following results shows:

Proposition 2.3.3 (Associativity)**.** Let $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$ be three functions. Then

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

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

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

□

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

Proposition 2.3.4. Let $f : X \to Y$ and $g : Y \to Z$ be functions.

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

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

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

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

Recalling that id_X is the identity map on X, we can define invertibility:

Definition 2.3.13: Invertibility

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

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

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

Proposition 2.3.5 (Uniqueness of inverse)**.** If $f : X \to Y$ is invertible then its inverse is unique.

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

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

□

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

Proposition 2.3.6. Let $f : X \to Y$ and $g : Y \to Z$ be functions. If f and g are invertible, then $g \circ f$ is invertible, and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

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

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

and similarly,

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

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

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

Theorem 2.3.3

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

Proof. We prove this in both directions.

Forward direction:

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

Backward direction:

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

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

Definition 2.3.14: Left and right invertibility

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

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

Part II

Linear Algebra

3 Vector Spaces

Recommended readings: “Linear Algebra Done Right" by Sheldon Axler

§3.1 Real and Complex Numbers

This text assumes that the reader should be familiar with the sets of real and complex numbers, denoted by $\mathbb{R}$ and $\mathbb{C}$ respectively.

Euclidean spaces, linear combinations and linear span, subspaces, linear independence, bases and dimension, rank of a matrix, inner products, eigenvalues and eigenvectors, diagonalisation, linear transformations between Euclidean spaces

§3.2 Definition

The motivation for the definition of a vector space comes from properties of addition and scalar multiplication in $\mathbb{F}^n$: Addition is commutative, associative, and has an identity. Every element has an additive inverse. Scalar multiplication is associative. Scalar multiplication by 1 acts as expected. Addition and scalar multiplication are connected by distributive properties.

We will define a vector space to be a set V with an addition and a scalar multiplication on V that satisfy the properties in the paragraph above.

Definition 3.2.1: Addition, scalar multiplication

n **addition** on V is a function that assigns an element $u + v \in V$ to each pair of elements $u, v \in V$.
A **scalar multiplication** on V is a function that assigns an element $\lambda v \in V$ to each $\lambda \in \mathbb{F}$ and each $v \in V$.

Now we are ready to give the formal definition of a vector space.

Definition 3.2.2: Vector space

A **vector space** is a set V along with an addition on V and a scalar multiplication on V such that the following properties **vector space axioms** hold:

V1 Commutativity: $\forall u, v \in V, u + v = v + u$

V2 Associativity: $\forall u, v, w \in V, u + (v + w) = (u + v) + w$

V3 Existence of additive identity: there exists $0 \in V$ such that $\forall v \in V, v + 0 = v = 0 + v$

V4 Existence of additive inverse: $\forall v \in V$ there exists $w \in V$ such that $v + w = 0_V = w + v$

V5 Existence of multiplicative identity: $\forall v \in V, 1v = v$

V6 Distributivity of scalar multiplication over vector addition: $\forall u, v \in V, \lambda \in \mathbb{F}, \lambda(u + v) = \lambda u + \lambda v$

V7 Distributivity of scalar multiplication over field addition: $\forall v \in V, \lambda, \mu \in \mathbb{F}, (\lambda + \mu)v = \lambda v + \mu v$

V8 Scalar multiplication interacts well with field multiplication: $\forall v \in V, \lambda, \mu \in \mathbb{F}, (\lambda\mu)v = \lambda(\mu v)$

The following geometric language sometimes aids our intuition.

Definition 3.2.3: Vector, point

Elements of a vector space are called **vectors** or **points**.

The scalar multiplication in a vector space depends on $\mathbb{F}$. Thus when we need to be precise, we will say that V is a vector space over $\mathbb{F}$ instead of saying simply that V is a vector space.

Example 3.2.1: $\mathbb{R}^n$ and $\mathbb{C}^n$

$\mathbb{R}^n$ is a vector space over $\mathbb{R}$, and $\mathbb{C}^n$ is a vector space over $\mathbb{C}$.

Definition 3.2.4: Real vector space, complex vector space

A vector space over $\mathbb{R}$ is called a **real vector space**.
A vector space over $\mathbb{C}$ is called a **complex vector space**.

Proposition 3.2.1 (Uniqueness of additive identity)**.** A vector space has a unique additive identity.

Proof. Suppose 0 and $0'$ are both additive identities for some vector space V.

Then

$$0' = 0' + 0 = 0 + 0' = 0$$

where the first equality holds because 0 is an additive identity, the second equality comes from commutativity, and the third equality holds because $0'$ is an additive identity.

Thus $0' = 0$, proving that V has only one additive identity. □

Proposition 3.2.2 (Uniqueness of additive inverse)**.** Every element in a vector space has a unique additive inverse.

Proof. Suppose V is a vector space. Let $v \in V$. Suppose w and w' are additive inverses of v. Then

$$w = w + 0 = w + (v + w') = (w + v) + w' = 0 + w' = w'$$

Thus $w = w'$, as desired. □

Because additive inverses are unique, the following notation now makes sense.

Notation. Let $v, w \in V$. Then $-v$ denotes the additive inverse of v; $w - v$ is defined to be $w + (-v)$.

Notation. For the rest of the book, V denotes a vector space over $\mathbb{F}$.

§3.3 Subspaces

Definition 3.3.1: Subspace

A subset $U \subset V$ is called a subspace of V if U is also a vector space (with the same addition and scalar multiplication as on V).

A subset U of V is a subspace of V if and only if U satisfies the following three conditions:

1. Existence of additive identity: $0 \in U$
2. Closed under addition: $u + w \in U \implies u + w \in U$
3. Closed under scalar multiplication: $a \in F$ and $u \in U$ implies $au \in U$.

Proof. If U is a subspace of V, then U satisfies the three conditions above by the definition of vector space.

Conversely, suppose U satisfies the three conditions above. The first condition above ensures that the additive identity of V is in U.

The second condition above ensures that addition makes sense on U. The third condition ensures that scalar multiplication makes sense on U. □

4 Matrices

linear transformations, kernels and images; inner products, inner product spaces, orthonormal sets, and the Gram-Schmidt process; eigenvectors and eigenvalues; matrix diagonalisation and its applications; symmetric and Hermitian matrices; quandratic forms and bilinear forms; Jordan normal form and other canonical forms.

5 Bases

§5.1 Spans and Spanning Sets

§5.2 Linear Independence

6 Dimension

7 Linear Transformations

8 Linear Maps and Matrices

9 Inner Product Spaces

Part III

Calculus

10 Single Variable Calculus

applications of calculus involving parametric, polar and vector functions polynomial approximations and convergence of series asymptotic and unbounded behavior

§10.1 Limits

§10.1.1 Informal Definition

Definition 10.1.1: Intuitive definition of limit

Suppose $f(x)$ is defined on some open interval that contains a, except possibly at a itself. Then we write

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

and say "the limit of $f(x)$, as x approaches a, equals L" if we can make the values of $f(x)$ arbitrarily close to L by restricting x to be sufficiently close to a (on either side of a) but not equal to a.

Definition 10.1.2: Intuitive definition of one-sided limits

We write $\lim_{x \to a^-} f(x) = L$ and say that the **left-hand limit** of $f(x)$ as x approaches a from the left is L.
Similarly, we write $\lim_{x \to a^+} f(x) = L$ and say that the **right-hand limit** of $f(x)$ as x approaches a from the right is L.

For the limit $\lim_{x \to a} f(x)$ to exist,

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

To indicate the behavior of vertical asymptotes or infinite limits, we use the notation $\lim_{x \to a} f(x) = \infty$.

Definition 10.1.3: Intuitive definition of infinite limit

Let f be a function defined on both sides of a, except possibly at a itself. Then

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

means that $f(x)$ can be made arbitrarily large by taking x sufficiently close to, but not equal to a.

Remark. This does not mean that we are regarding ∞ as a number, nor does it mean that the limit exists; it simply expresses the particular way in which the limit does not exist: $f(x)$ can be made as large as we like by taking x close enough to 0.

A similar sort of limit, for functions that become large negative as x gets close to a, is defined below.

Definition 10.1.4

Let f be a function defined on both sides of a, except possibly at a itself. Then

$$\lim_{x\to a} f(x) = -\infty$$

means that $f(x)$ can be made arbitrarily large negative by taking x sufficiently close to, but not equal to a.

Definition 10.1.5: Vertical asymptote

The vertical line $x = a$ is a **vertical asymptote** of $y = f(x)$ if at least one of the following statements is true:

$$\lim_{x\to a} f(x) = \infty \qquad \lim_{x\to a^-} f(x) = \infty \qquad \lim_{x\to a^+} f(x) = \infty$$

$$\lim_{x\to a} f(x) = -\infty \qquad \lim_{x\to a^-} f(x) = -\infty \qquad \lim_{x\to a^+} f(x) = -\infty$$

§10.1.2 Limit Laws

Let $f(x)$ and $g(x)$ be defined for all $x \neq a$ over some open interval containing a. Assume that L and M are real numbers such that $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$, c is a constant. Then each of the following statements holds.

- **Sum law**: The limit of a sum is the sum of the limits.

$$\lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = L + M$$

- **Difference law**: The limit of a difference is the difference of the limits.

$$\lim_{x\to a}(f(x) - g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x) = L - M$$

- **Constant multiple law**: The limit of a constant times a function is the constant times the limit of the function.

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

- **Product law**: The limit of a product is the product of the limits.

$$\lim_{x \to a}(f(x) \cdot g(x)) = \lim_{x \to a} f(x) \cdot \lim_{x \to a} g(x) = L \cdot M$$

- **Quotient law**: The limit of a quotient is the quotient of the limits.

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)} = \frac{L}{M}$$

  for $M \neq 0$.

- **Power law**

$$\lim_{x \to a}(f(x))^n = (\lim_{x \to a} f(x))^n = L^n$$

  for every positive integer n.

- **Root law**

$$\lim_{x \to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x \to a} f(x)} = \sqrt[n]{L}$$

  for all L if n is odd, and for $L \geq 0$ if n is even.

§10.1.3 Evaluating Limits

Indeterminate forms of a limit include:

$$\frac{0}{0} \quad \frac{\infty}{\infty} \quad 0\times\infty \quad \infty-\infty \quad 0^0 \quad 1^\infty \quad \infty^0$$

As long as limits are in indeterminate forms, they can still be evaluated.

Methods:

- Direct substitution

  If f is a polynomial or a rational function and a is in the domain of f, then

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

- Cancel common factors
- Multiply by the conjugate of the numerator or denominator

Example 10.1.1

Evaluate the following limit:

$$\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right)$$

Solution. If we plot the graph of the function out, we see that we can try to find two functions to apply Squeeze Theorem.

Notice that

$$-1 \le \sin\left(\frac{1}{x}\right) \le 1$$

and hence

$$-x^2 \le x^2\sin\left(\frac{1}{x}\right) \le x^2$$

thus x^2 and $-x^2$ are the two functions that "sandwich" the given function.

Since $\lim_{x\to 0} x^2 = 0$ and $\lim_{x\to 0} -x^2 = 0$, applying Squeeze Theorem gives us

$$\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right) = 0$$

□

Example 10.1.2

Evaluate the following limit:

$$\lim_{x\to 3} \frac{-4x}{x-3}$$

Solution. Approaching from the left side,

$$\lim_{x \to 3^-} \frac{-4x}{x-3} = +\infty$$

Approaching from the right side,

$$\lim_{x \to 3^+} \frac{-4x}{x-3} = -\infty$$

Since $\lim_{x \to 3^-} \dfrac{-4x}{x-3} \neq \lim_{x \to 3^+} \dfrac{-4x}{x-3}$, the limit does not exist. □

Theorem 10.1.1

If $f(x) \leq g(x)$ when x is near a (except possibly at a) and the limits of f and g both exist as x approaches a, then

$$\lim_{x \to a} f(x) \leq \lim_{x \to a} g(x)$$

Theorem 10.1.2: Squeeze theorem

Suppose that $g(x) \geq f(x) \geq h(x)$ for all x in some open interval containing c except possibly at c itself. If $\lim_{x \to c} g(x) = L = \lim_{x \to c} h(x)$, then $\lim_{x \to c} f(x) = L$.

Proof. This can be proven using the epsilon-delta definition of limits.

Let $\varepsilon > 0$ be given. We are done if we find a $\delta > 0$ such that $|f(x) - L| < \varepsilon$ whenever $0 < |x - c| < \delta$.

Since $\lim_{x \to c} g(x) = L$, by definition of limits, there exists some $\delta_1 > 0$ such that for all $0 < |x - c| < \delta_1$, $|g(x) - L| < \varepsilon$. Thus,

$$-\varepsilon < g(x) - L < \varepsilon \quad \text{for all } 0 < |x - c| < \delta_1$$

so

$$L - \varepsilon < g(x) < L + \varepsilon \quad \text{for all } 0 < |x - c| < \delta_1 \tag{1}$$

Similarly, since $\lim_{x \to c} h(x) = L$, by definition of limits, there exists some $\delta_2 > 0$ such that

$$L - \varepsilon < h(x) < L + \varepsilon \quad \text{for all } 0 < |x - c| < \delta_2 \tag{2}$$

Additionally, since $g(x) \leq f(x) \leq h(x)$ for all x in some open interval containing c, there exists some $\delta_3 > 0$ such that for

$$g(x) \leq f(x) \leq h(x) \quad \text{for all } 0 < |x - c| < \delta_3 \tag{3}$$

Now, we choose $\delta = \min(\delta_1, \delta_2, \delta_3)$. Then by (1), (3), and (2), we have

$$L - \varepsilon < g(x) \leq f(x) \leq h(x) < L + \varepsilon \quad \text{for all } 0 < |x - c| < \delta.$$

Therefore, $-\varepsilon < f(x) - L < \varepsilon$ for all $0 < |x - c| < \delta$, so

$$|f(x) - L| < \varepsilon \quad \text{for all } 0 < |x - c| < \delta.$$

Hence, by definition of limits, $\lim_{x \to c} f(x) = L$. □

Theorem 10.1.3: L'Hôpital's Rule

Let $f(x)$ and $g(x)$ be differentiable on an interval I containing a, and that $g'(a) \neq 0$ on I for $x \neq a$. Suppose that

$$\lim_{x \to a} \frac{f(x)}{g(x)}$$

is in an indeterminate form.
Then as long as the limits exist, we have

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}.$$

Proof. □

§10.1.4 Precise Definition of a Limit

The intuitive definition of a limit given above is inadequate because such phrases as "x is close to 2" and "$f(x)$ gets closer and closer to L" are vague. In order to be able to prove limits conclusively, we must make the definition of a limit precise.

Definition 10.1.6: Epsilon–delta definition of limit

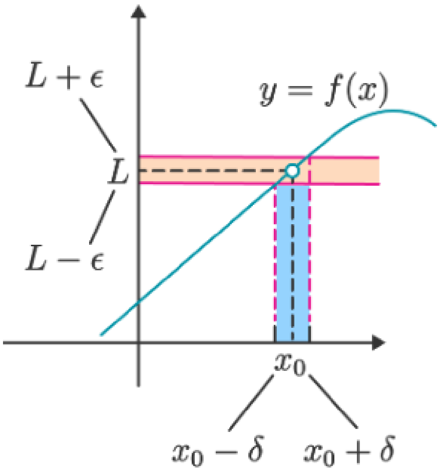
Let $f(x)$ be a function defined on an open interval around x_0. We say that the limit of $f(x)$ as x approaches x_0 is L, i.e.

$$\lim_{x \to x_0} f(x) = L$$

if $\forall \varepsilon > 0$, there exists $\delta > 0$ such that $\forall x \in \mathbb{R}$,

$$|x - x_0| < \delta \implies \left| f(x) - L \right| < \varepsilon$$

Visualising this graphically,



As ε becomes smaller and smaller, there always exists a δ that satisfies the property that for any x in the open interval $(x_0 - \delta, x_0 + \delta)$, the value of $f(x)$ lies in the interval $(L - \varepsilon, L + \varepsilon)$.

Example 10.1.3

Prove that

$$\lim_{x \to 3} 2x + 4 = 10.$$

Before the proof, we work backwards to find the value of δ in terms of ε and x_0, which we then declare in our proof.

$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall x \in \mathbb{R},$

$$|x - 3| < \delta \implies |f(x) - 10| < \varepsilon$$

Let $\varepsilon > 0$ be given.

$$|f(x) - 10| = |2x + 4 - 10| = |2x - 6| = 2|x - 3| < \varepsilon$$

Notice $|x-3| < \frac{\varepsilon}{2}$. We can thus define $\delta \coloneqq \frac{\varepsilon}{2}$. We now write our proof.

Proof. Let $\varepsilon > 0$ be given. Choose $\delta = \frac{\varepsilon}{2}$.

Then $\forall x \in \mathbb{R}$,

$$\begin{aligned} |x-3| &< \delta = \frac{\varepsilon}{2} \\ 2|x-3| &< \varepsilon \\ |2x-6| &< \varepsilon \\ |2x+4-10| &< \varepsilon \\ |f(x)-10| &< \varepsilon \end{aligned}$$

□

Example 10.1.4

Use the formal definition of the limit to verify that

$$\lim_{x\to 3} \sqrt{2x+3} = 3.$$

We must prove that $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\sqrt{2x+3} - 3 < \varepsilon$ whenever $|x-3| < \delta$.

$$\begin{aligned} \sqrt{2x+3} - 3 &= \left| \frac{(2x+3) - 3^2}{\sqrt{2x+3}+3} \right| = \left| \frac{2x-6}{\sqrt{2x+3}+3} \right| \\ &\leq \left| \frac{2(x-3)}{3} \right| \\ &= \frac{2}{3}|x-3| < \frac{2}{3}\delta \end{aligned}$$

Hence, we can define

$$\varepsilon \coloneqq \frac{2}{3}\delta$$

which we can use in our proof.

§10.1.5 Important Limits

$$\lim_{x \to 0} \frac{\sin x}{x} = 1 \tag{10.1}$$

Proof. This can be proven using the squeeze theorem, which will be discussed later. □

$$\lim_{x \to 0} \frac{1 - \cos x}{x} = 0 \tag{10.2}$$

Proof. This can be proven using the squeeze theorem, which will be discussed later. □

$$\lim_{x \to 0} \frac{\arcsin x}{x} = 1 \tag{10.3}$$

$$\lim_{x \to \pm\infty} \left(1 + \frac{1}{x}\right)^x = e \tag{10.4}$$

§10.1.6 Continuity

Definition 10.1.7: Continuity

A function $f(x)$ is continuous at $x = a$ if

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

A function is said to be continuous on the interval $[a, b]$ if it is continuous at each point in the interval.

Note that this definition is also implicitly assuming that both $f(a)$ and $\lim_{x \to a} f(x)$ exist. If either of these do not exist the function will not be continuous at $x = a$.

This definition can be turned around into the following fact.

Corollary 3. If $f(x)$ is continuous at $x = a$ then

$$\lim_{x \to a} f(x) = f(a) \qquad \lim_{x \to a^-} f(x) = f(a) \qquad \lim_{x \to a^+} f(x) = f(a)$$

A nice consequence of continuity is the following fact.

Corollary 4. If $f(x)$ is continuous at $x = b$ and $\lim_{x \to a} g(x) = b$ then

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

Example 10.1.5

Evaluate the following limit:

$$\lim_{x \to 0} e^{\sin x}$$

Solution. Since we know that exponentials are continuous everywhere we can use the fact above.

$$\lim_{x \to 0} e^{\sin x} = e^{\lim_{x \to 0} \sin x} = e^0 = \boxed{1}$$

□

Another very nice consequence of continuity is the Intermediate Value Theorem.

Theorem 10.1.4: Intermediate Value Theorem

Suppose that $f(x)$ is continuous on $[a, b]$ and let M be any number between $f(a)$ and $f(b)$. Then there exists $c \in (a, b)$ such that $f(c) = M$.

All the Intermediate Value Theorem is really saying is that a continuous function will take on all values between $f(a)$ and $f(b)$.

§10.2 Derivative

§10.2.1 Definitions

Definition 10.2.1: Derivative

The **derivative** of $f(x)$ with respect to x, denoted by $f'(x)$, is defined as

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}. \tag{10.5}$$

Definition 10.2.2: Differentiability

$f(x)$ is **differentiable** at x_0 if $f'(x_0)$ exists.
$f(x)$ is **differentiable** on an interval if the derivative exists for each and every point in the interval.

Definition 10.2.3: Continuity

$f(x)$ is **continuous** at x_0 if $f(x)$ is differentiable at $x = x_0$.

Proof.

$$\begin{aligned}\lim_{x\to x_0}(f(x) - f(x_0)) &= \lim_{x\to x_0} \frac{(f(x) - f(x_0))(x - x_0)}{x - x_0} \\ &= \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} \cdot (x - x_0) \\ &= f'(x_0) \cdot 0 = 0\end{aligned}$$

□

§10.2.2 Theorems

Theorem 10.2.1: Extreme Value Theorem

For a function f continuous on $[a, b]$, it attains its maximum and minimum values on $[a, b]$.

Proof. We prove the case that f attains its maximum value on $[a, b]$.

Since f is continuous on $[a, b]$, we know it must be bounded on $[a, b]$ by the Boundedness Theorem. Let $M = \sup f$.

If there is some $c \in [a, b]$ where $f(c) = M$ there is nothing more to show – f attains its maximum on $[a, b]$.

Suppose otherwise, that there is no such c. Then $f(x) < M$ for all $x \in [a, b]$.

We define a new function g by

$$g(x) = \frac{1}{M - f(x)}.$$

Note that $g(x) > 0$ for every $x \in [a,b]$ and g is continuous on $[a,b]$, and thus also bounded on this interval, again by the Boundedness theorem.

Given that g is bounded on $[a,b]$, there must exist some $K > 0$ such that $g(x) \leq K$ for every $x \in [a,b]$.

Consequently,

$$\frac{1}{M - f(x)} \leq K \implies f(x) \leq M - \frac{1}{K}$$

for every $x \in [a,b]$. This contradicts the assumption that M is the least upper bound.

That leaves as the only possibility that there is some c in $[a,b]$ where $f(c) = M$. That is to say, f attains its maximum on $[a,b]$.

The proof that f attains its minimum on the same interval is argued similarly and is left as an exercise for the reader. $\square$

Theorem 10.2.2: Mean Value Theorem

Let $f : [a,b] \to \mathbb{R}$ be a continuous function on the closed interval $[a,b]$ and differentiable on the open interval (a,b). Then there exists $c \in (a,b)$ such that

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

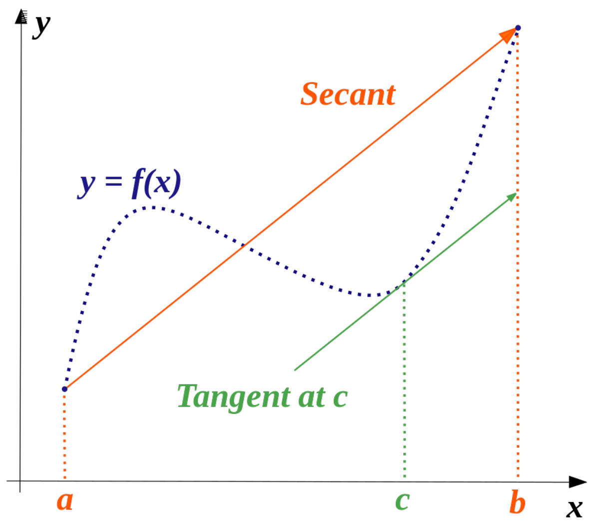


Figure 10.1: Mean value theorem

Theorem 10.2.3: Rolle's Theorem

Let $f : [a, b] \to \mathbb{R}$ be a continuous function on the closed interval $[a, b]$ and differentiable on the open interval (a, b), and $f(a) = f(b)$. Then there exists $c \in (a, b)$ such that

$$f'(c) = 0.$$

Remark. Rolle's Theorem is simply a special case of the Mean Value Theorem, where $f(a) = f(b)$.

§10.2.3 Differentiation rules

- **Scalar multiplication**
- **Addition rule**

$$(f + g)' = f' + g' \tag{10.6}$$

Proof.

$$\begin{aligned}
(f + g)'(x) &= \lim_{h\to 0} \frac{(f + g)(x + h) - (f + g)(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x + h) + g(x + h) - f(x) - g(x)}{h} \\
&= \lim_{h\to 0} \left[\frac{f(x + h) - f(x)}{h} + \frac{g(x + h) - g(x)}{h} \right] \\
&= \lim_{h\to 0} \frac{f(x + h) - f(x)}{h} + \lim_{h\to 0} \frac{g(x + h) - g(x)}{h} \\
&= f'(x) + g'(x)
\end{aligned}$$

□

- **Power rule**

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1} \tag{10.7}$$

Proof. Using implicit differentiation,

$$\begin{aligned}
y &= x^n \\
\ln y &= \ln x^n \\
\ln y &= n \ln x \\
\frac{y'}{y} &= n\frac{1}{x} \\
y' = y\frac{n}{x} &= x^n \left(\frac{n}{x}\right) = nx^{n-1}
\end{aligned}$$

□

- **Product rule**

$$(fg)' = f'g + fg' \tag{10.8}$$

Proof.

$$\begin{aligned}
(fg)'(x) &= \lim_{h\to 0} \frac{(fg)(x+h) - (fg)(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x) + f(x+h)g(x+h) - f(x+h)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x)}{h} + \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} g(x) + \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} f(x+h) \\
&= \left[\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}\right] g(x) + \left[\lim_{h\to 0} \frac{g(x+h) - g(x)}{h}\right] f(x) \\
&= f'(x)g(x) + f(x)g'(x)
\end{aligned}$$

□

- **Quotient rule**

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2} \tag{10.9}$$

Proof.

$$\begin{aligned}
\left[\frac{f(x)}{g(x)}\right]' &= \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} \\
&= \lim_{h\to 0} \frac{1}{h} \frac{f(x+h)g(x) - f(x)g(x+h)}{g(x+h)g(x)} \\
&= \lim_{h\to 0} \frac{1}{h} \frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+h)}{g(x+h)g(x)} \\
&= \lim_{h\to 0} \frac{1}{g(x+h)g(x)} \left[\frac{f(x+h)g(x) - f(x)g(x)}{h} + \frac{f(x)g(x) - f(x)g(x+h)}{h}\right] \\
&= \lim_{h\to 0} \frac{1}{g(x+h)g(x)} \left[g(x)\frac{f(x+h) - f(x)}{h} - f(x)\frac{g(x) + g(x+h)}{h}\right] \\
&= \frac{1}{g^2(x)}[g(x)f'(x) - f(x)g'(x)] \\
&= \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}
\end{aligned}$$

□

Chain rule

Theorem 10.2.4: Chain rule

If f and g are both differentiable functions and we define $F(x) = (f \circ g)(x)$, then the derivative of $F(x)$ is

$$F'(x) = f'(g(x))g'(x) \tag{10.10}$$

§10.2.4 Implicit differentiation

Implicit differentiation simply means differentiating both sides of the equation with respect to a variable.

§10.2.5 Taylor Series

A function f can be represented as a Taylor series about position a if

- it is continuous near a and
- all of its derivatives are continuous near a

Using the notation $\Delta x = x - a$:

$$f(x) = f(a) + \Delta x f'(a) + \frac{\Delta x}{2!} f''(a) + \frac{\Delta x}{3!} f^{(3)}(a) + \cdots + \frac{\Delta x}{n!} f^{(n)}(a) + \cdots$$

If infinitely many terms are used, this approximation is exact near a.

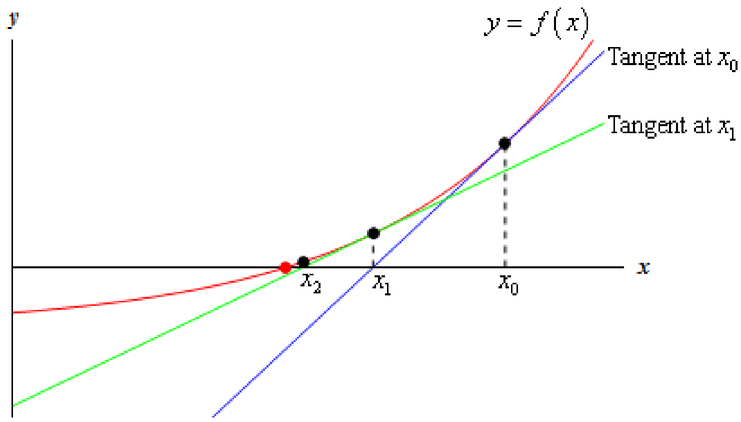
If all terms of order n and above are discarded then the error is approximately proportional to Δx^n (assuming that Δx is small). Then the approximation is said to be n-th order accurate. For example, a third order accurate approximation for $f(x)$ has error proportional to Δx^3. We say that the error is of order Δx^3 or $O(\Delta x^3)$.

$$f(x) = f(a) + \Delta x f'(a) + \frac{\Delta x}{2!} f''(a) + O(\Delta x^3)$$

Maclaurin series, determine radius and interval of convergence of a power series

§10.2.6 Newton's Method

In this section we are going to look at a method for approximating solutions to equations.



Suppose that we want to approximate the solution to $f(x) = 0$. Suppose that we have somehow found an initial rough approximation to the solution: $x = x_0$. The tangent line to $f(x)$ at $x = x_0$ is

$$y = f(x_0) + f'(x_0)(x - x_0)$$

This tangent line crosses the x-axis much closer to the actual solution to the equation than x_0 does. Let the tangent at x_0 intersect x-axis at x_1. We use this point as our new approximation to the solution. x_1 is given by:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Repeat the process; form up the tangent line at x_1 and use its root x_2 as a new approximation:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

Here is the general Newton's Method:

Theorem 10.2.5: Newton's Method

If x_n is an approximation of a solution of $f(x) = 0$ and if $f'(x_n) \neq 0$, then the next approximation is given by

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

§10.3 Integral

§10.3.1 Definition

We use the **Riemann** definition of an integral:

Definition 10.3.1: Integral

An integral is defined as an infinite sum over an interval.

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)\Delta x \tag{10.11}$$

A Riemann sum is an approximation of an integral by a finite sum.

Let f be defined on the closed interval $[a,b]$ and let Δx be a partition of $[a,b]$, with

$$a = x_1 < x_2 < \cdots < x_n < x_{n+1} = b.$$

Let Δx_i denote the length of the ith subinterval $[x_i, x_{i+1}]$ and let c_i denote any value in the ith subinterval.

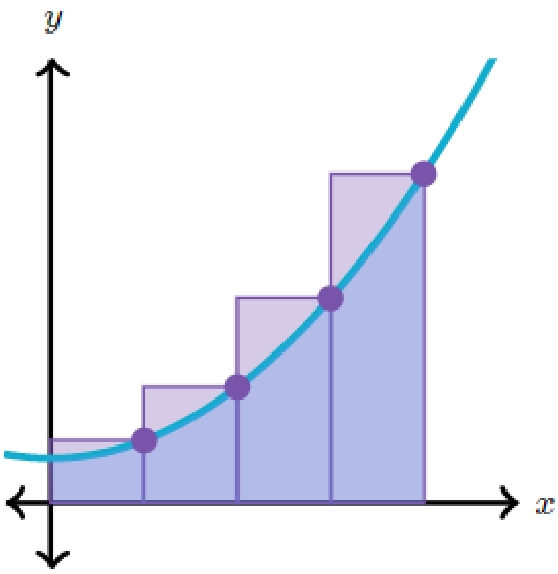
The sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i$$

is a Riemann sum of f on $[a,b]$.

As the subinterval becomes infinitesimally small,

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\Delta x\to 0}\sum_{i=1}^{n} f(x_i)\Delta x_i \tag{10.12}$$



Given a graph $y = f(x)$, and we want to find the integral on the x-interval $[0,1]$.

Split the interval $[0,1]$ into n equal subintervals

$$\left[0, \frac{1}{n}\right], \left[\frac{1}{n}, \frac{2}{n}\right], \cdots, \left[\frac{n-1}{n}, 1\right].$$

Consider the height of the rectangles. We take the right value. Hence for the kth subinterval $\left[\frac{k-1}{n}, \frac{k}{n}\right]$ where $k = 1, \dots, n$, height of rectangle is $f\left(\frac{k}{n}\right)$.

Area of kth rectangle is

$$\frac{1}{n} \cdot f\left(\frac{k}{n}\right).$$

Therefore, the integral is obtained by summing up the area of n rectangles, which gives us

$$\int_0^1 f(x)\,\mathrm{d}x = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{n} f\left(\frac{k}{n}\right) \tag{10.13}$$

where there are infinitely many rectangles, i.e. $n \to \infty$.

Theorem 10.3.1: Fundamental Theorem of Calculus

The fundamental theorem of (single variable) calculus states that if f' is continuous on $[a,b]$, then the integral of the derivative across the bounds is equal to the original function at the bounds:

$$\int_a^b f'(x)\,\mathrm{d}x = f(b) - f(a) \tag{10.14}$$

or equivalently,

$$\frac{\mathrm{d}}{\mathrm{d}x} \int_a^x f(s)\,\mathrm{d}s = f(x) \tag{10.15}$$

Using the definition of the derivative, we differentiate the following integral:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} \int_a^x f(s)\,\mathrm{d}s &= \lim_{h\to 0} \frac{\int_a^{x+h} f(s)\,\mathrm{d}s - \int_a^x f(s)\,\mathrm{d}s}{h} \\
&= \lim_{h\to 0} \frac{\int_x^{x+h} f(s)\,\mathrm{d}s}{h} \\
&= \lim_{h\to 0} \frac{hf(x)}{h} \\
&= f(x)
\end{aligned}$$

§10.3.2 Integration rules

Sum rule

$$\int f(x) + g(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x$$

Difference rule

$$\int f(x) - g(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x - \int g(x)\,\mathrm{d}x$$

Scalar multiplication

$$\int kf(x)\,\mathrm{d}x = k\int f(x)\,\mathrm{d}x$$

Power rule

$$\int x^n\,\mathrm{d}x = \frac{x^{n+1}}{n+1} + C$$

Constant rule

$$\int a\,\mathrm{d}x = ax + C$$

§10.3.3 Integration techniques

Integrals of powers and of trigonometric functions

Reciprocal rules:

$$\int \frac{1}{x}\,\mathrm{d}x = \ln|x| + C$$

$$\int \frac{1}{ax+b}\,\mathrm{d}x = \frac{1}{a}\ln(ax+b) + C$$

Exponential functions:

$$\int e^x\,\mathrm{d}x = e^x + C$$

$$\int a^x\,\mathrm{d}x = \frac{a^x}{\ln a} + C$$

Natural log rule:

$$\int \ln x\,\mathrm{d}x = x\ln x - x + C$$

Trigonometric functions:

$$\int \sin x\,\mathrm{d}x = -\cos x + C$$

$$\int \cos x\,\mathrm{d}x = \sin x + C$$

$$\int \tan x\,\mathrm{d}x = \ln|\sec x| + C$$

$$\int \operatorname{cosec} x\,\mathrm{d}x = \ln|\operatorname{cosec} x - \cot x| + C$$

$$\int \operatorname{cosec}^2 x\,\mathrm{d}x = -\cot x + C$$

$$\operatorname{cosec} x\cot x\,\mathrm{d}x = -\operatorname{cosec} x + C$$

$$\sec x\,\mathrm{d}x = \ln|\sec x + \tan x| + C$$

$$\int \sec^2 x \, \mathrm{d}x = \tan x + C$$

$$\int \sec x \tan x \, \mathrm{d}x = \sec x + C$$

$$\int \cot x \, \mathrm{d}x = \ln|\sin x| + C$$

Inverse trigonometric functions:

$$\int \frac{1}{\sqrt{1-x^2}} \, \mathrm{d}x = \sin^{-1} x + C$$

$$-\int \frac{1}{\sqrt{1-x^2}} \, \mathrm{d}x = \cos^{-1} x + C$$

$$\int \frac{x}{1+x^2} \, \mathrm{d}x = \tan^{-1} x + C$$

Substitution

$$\int f(g(x))g'(x) \, \mathrm{d}x = \int f(u) \, \mathrm{d}u \tag{10.16}$$

where $u = g(x)$.

The most common way of doing a integral by substitution, and the only way for indefinite integrals, is as follows:

1. Change variables from x to u (hence the common name “u-substitution")
2. Keep track of the relation between $\mathrm{d}x$ and $\mathrm{d}u$
3. If you chose correctly you can now do the u-integral
4. When you are done, substitute back for x

Example 10.3.1

Compute $\int \sin^n x \cos x \, \mathrm{d}x$.

Solution. Substitute $u = \sin x$ and $\mathrm{d}u = \cos x \, \mathrm{d}x$. This turns the integral into $\int u^n \, \mathrm{d}u$ which is easily valuated as $u^{n+1}/(n+1)$. Now plug back in $u = \sin x$ and you get the answer

$$\frac{\sin^{n+1} x}{n+1}.$$

□

Example 10.3.2

Compute $\int_1^2 \frac{x}{x^2+1} \, \mathrm{d}x$.

Solution. Let $u = x^2 + 1$ then $\mathrm{d}u = 2x\,\mathrm{d}x$, so the integrand becomes $(1/2)\,\mathrm{d}u\,/u$. If x goes from 1 to 2 then u goes from 2 to 5, thus the integral becomes

$$\int_2^5 \frac{1}{2}\frac{\mathrm{d}u}{u} = \frac{1}{2}(\ln 5 - \ln 2).$$

□

Example 10.3.3

Compute $\int xe^{x^2}\,\mathrm{d}x$.

Solution. To do this integral we'll use the following substitution.

$$u = x^2 \quad \mathrm{d}u = 2x \quad \mathrm{d}x \implies x\,\mathrm{d}x = \frac{1}{2}\,\mathrm{d}u$$

$$\int xe^{x^2}\,\mathrm{d}x = \frac{1}{2}\int e^u\,\mathrm{d}u = \frac{1}{2}e^u + c = \frac{1}{2}e^{x^2} + c$$

□

Integration by parts

From the product rule used for differentiation, we obtain

$$\int fg'\,\mathrm{d}x = fg - \int f'g\,\mathrm{d}x \tag{10.17}$$

Alternatively, we can rewrite this as

$$\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u \tag{10.18}$$

DI method

Partial fraction decomposition

Trigonometric substitutions

- Pythagorean identity: $\sin^2 x + \cos^2 x = 1$
- Double-angle formulae

  These can be used in the integrals of $\sin^2 x$ and $\cos^2 x$.
- Product-to-sum identities

Integrals of powers of trigonometric functions

Integrals of hyperbolic functions

Completing the square

Elimination of radicals by substitution

Weierstrass substitution

Substituting the tangent of a half-angle: $t = \tan\frac{\theta}{2}$

Through trigonometric identities and manipulation, we have

$$\sin\theta = \frac{2t}{1+t^2} \quad \cos\theta = \frac{1-t^2}{1+t^2} \quad \mathrm{d}\theta = \frac{2\,\mathrm{d}t}{1+t^2}$$

Some examples here:

Useful info here: More info to be found on Youtube.

More problems here:

Odd and even functions

An odd function $f(x)$ satisfies $f(x) = -f(-x)$ for all x. Hence for any finite a,

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = 0$$

An even function $f(x)$ satisfies $f(x) = f(-x)$ for all x. Hence for any finite a,

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = 2\int_{0}^{a} f(x)\,\mathrm{d}x$$

Reflections

This is known as **King's property**, which states that we can revere the interval of integration: to “integrate backwards".

$$\int_{a}^{b} f(x)\,\mathrm{d}x = \int_{a}^{b} f(a+b-x)\,\mathrm{d}x \tag{10.19}$$

Instead of the function being centred at 0, the function is now centred at $\frac{a+b}{2}$. Then

$$\int_{a}^{b} f(x)\,\mathrm{d}x = \frac{1}{2}\int_{a}^{b} f(x) + f(a+b-x)\,\mathrm{d}x$$

Inversions

Suppose the function f has bounded anti-derivative on $[0,\infty]$. Then via the u-substitution $x \to \frac{1}{x}$,

$$\int_{0}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{2}\int_{0}^{\infty} f(x) + \frac{f(\frac{1}{x})}{x^2}\,\mathrm{d}x$$

Inverse functions

Suppose the function f is one-to-one and increasing. Then a geometric equivalence may be established:

Feynman's integration trick

DIfferentiating under the integral sign

§10.3.4 Approximation of Integral

Trapezium rule

We can sample the integrand at regular integrals and carry out an estimate based on this. One way of doing that is to approximate the function by a sequence of straight line segments. The area between each segment and the x-axis is a *trapezium*, meaning that if the width of the interval is h, and the y-values at each end of the interval are y_i and y_{i+1}, then the area of the trapezium is

$$\frac{h}{2}(y_i + y_{i+1}).$$

The entire area between the curve and the x-axis, which is to say the integral, can be approximated by adding together several such trapezia. If there are n trapezia, and n+1 y-values (ordinates) running from y_0 to y_n, then the integral is approximately

$$T_n = \frac{h}{2}\left(y_0 + 2\,y_2 + 2\,y_2 + \cdots + 2\,y_{n-2} + 2\,y_{n-1} + y_n\right) \tag{10.20}$$

Simpson's Rule

Simpson's Rule is based on the fact that given any three points, you can find the *equation of a quadratic* through those points.

This fact inspired Simpson to approximate integrals using quadratics, as follows.

If you want to integrate $f(x)$ over the interval $[a, b]$:

1. Find $f(a)$, $f(b)$ and $f(m)$ where $m = \dfrac{a+b}{2}$.
2. Find a quadratic $P(x)$ that goes through these three points.

Since quadratics are easy to integrate, you simply need to integrate the quadratic over the interval. It turns out that the integral of the quadratic over the interval $[a, b]$ always comes out to

$$\frac{b-a}{6}[f(a) + 4f(m) + f(b)] \tag{10.21}$$

For even n subdivisions,

$$\int_a^b f(x)x' \approx \frac{\Delta x}{3}(f(x_0) + 4f(x_1) + 2f(x_2) + \cdots + 4f(x_{n-1}) + f(x_n)) \tag{10.22}$$

where $\Delta x = \dfrac{b-a}{n}$, $x_i = a + i\Delta x$.

§10.3.5 Parametric Equations and Polar Coordinates

§10.4 Ordinary Differential Equations

Definition 10.4.1: Ordinary differential equation

An ordinary differential equation (ODE) is an equation relating a variable, say x, a function, say y, of the variable x, and finitely many of the derivatives of y with respect to x.
That is, an ODE can be written in the form

$$f\left(x, y, \frac{\mathrm{d}y}{\mathrm{d}x}, \frac{\mathrm{d}^2y}{\mathrm{d}x^2}, \cdots, \frac{\mathrm{d}^ky}{\mathrm{d}x^k}\right) = 0$$

for some function f and some natural number k. Here x is the **independent variable** and the ODE governs how the **dependent variable** y varies with x.

Remark. The equation may have no, one or many functions $y(x)$ which satisfy it; the problem is usually to find the most general solution $y(x)$, a function which satisfies the differential equation.

The derivative $\frac{\mathrm{d}^ky}{\mathrm{d}x^k}$ is said to be of order k. We say that an ODE has **order** k if it involves derivatives of order k and less. Hence, a first-order differential equation involves up to the first derivative $\frac{\mathrm{d}y}{\mathrm{d}x}$, whereas a second-order differential equation involves up to the second derivative $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$.

§10.4.1 First-order differential equations

First-order differential equations take the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x, y)$$

There are several standard methods for solving first order ODEs and we look at some of these now.

Direct integration

If the ODE takes the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x)$$

in other words the derivative is a function of x only, then we can integrate directly.

Example 10.4.1

Find the general solution of

$$\frac{\mathrm{d}y}{\mathrm{d}x} = x^2 \sin x$$

Solution. By direct integration,

$$y = \int x^2 \sin x \, dx = (2 - x^2) \cos x + 2x \sin x + c$$

which is done using integration by parts. □

Separation of variables

This method is applicable when the first order ODE takes the form

$$\frac{dy}{dx} = a(x)b(y)$$

where a is a function of x and b is a function of y.

Such an equation is called **separable**. These equations can be rearranged and solved as follows. First

$$\frac{1}{b(y)} \frac{dy}{dx} = a(x)$$

and then integrating with respect to x we find

$$\int \frac{1}{b(y)} \, dy = \int a(x) \, dx$$

Here we have assumed that $b(y) \neq 0$; if $b(y) = 0$ then the solution is $y = c$ where c is a constant.

Example 10.4.2

Find the general solution to the separable differential equation

$$x(y^2 - 1) + y(x^2 - 1)\frac{dy}{dx} = 0$$

where $0 < x < 1$.

Solution. We rearrange to obtain

$$\frac{y}{y^2 - 1} \frac{dy}{dx} = -\frac{x}{x^2 - 1}$$

After integration we obtain

$$\frac{1}{2} \ln |y^2 - 1| = -\frac{1}{2} \ln |x^2 - 1| + c$$

where c is a constant. This can be arranged to give

$$(x^2 - 1)(y^2 - 1) = c.$$

Note that the constant functions $y = 1$ and $y = -1$ are also solutions of the differential equation, but are already included in the given general solution, for $c = 0$. □

Reduction to separable form by substitution

Some first order differential equations can be transformed by a suitable substitution into separable form.

Example 10.4.3

Find the general solution of

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \sin(x + y + 1)$$

Solution. Let $u(x) = x + y(x) + 1$ so that $\frac{\mathrm{d}u}{\mathrm{d}x} = 1 + \frac{\mathrm{d}y}{\mathrm{d}x}$. Then the original equation can be written as $\frac{\mathrm{d}u}{\mathrm{d}x} = 1 + \sin u$, which is separable. We have

$$\frac{1}{1 + \sin u}\frac{\mathrm{d}u}{\mathrm{d}x} = 1$$

which integrates to

$$\int \frac{1}{1 + \sin u}\,\mathrm{d}u = x + c$$

Let us evaluate the integral on the left hand side:

$$\begin{aligned}\int \frac{1}{1 + \sin u}\,\mathrm{d}u &= \int \frac{1 - \sin u}{(1 + \sin u)(1 - \sin u)}\,\mathrm{d}u \\ &= \int \frac{1 - \sin u}{1 - \sin^2 u}\,\mathrm{d}u = \int \frac{1 - \sin u}{\cos^2 u}\,\mathrm{d}u \\ &= \int \frac{1}{\cos^2 u}\,\mathrm{d}u - \int \frac{\sin u}{\cos^2 u}\,\mathrm{d}u \\ &= \tan u - \frac{1}{\cos u} + c\end{aligned}$$

Therefore

$$\tan u - \frac{1}{\cos u} = x + c$$

In terms of x and y, the solution is given by

$$\tan(x + y + 1) - \frac{1}{\cos(x + y + 1)} = x + c$$

or

$$\sin(x + y + 1) - 1 = (x + c)\cos(x + y + 1).$$

This solution, where we have not found y in terms of x, is called an **implicit solution**. □

A special group of first order differential equations is those of the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f\left(\frac{y}{x}\right)$$

These differential equations are called **homogeneous** and they can be solved with a substitution of the form

$$y(x) = xv(x)$$

to get a new equation in terms of x and the new dependent variable v. This new equation will be separable:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = v + x\frac{\mathrm{d}v}{\mathrm{d}x}$$

which becomes

$$x\frac{\mathrm{d}v}{\mathrm{d}x} = f(v) - v$$

§10.4.2 First-order differential equations

In general, a k-th order **inhomogeneous linear ODE** takes the form

$$a_k(x)\frac{\mathrm{d}^k y}{\mathrm{d}x^k} + a_{k-1}(x)\frac{\mathrm{d}^{k-1} y}{\mathrm{d}x^{k-1}} + \cdots + a_1(x)\frac{\mathrm{d}y}{\mathrm{d}x} + a_0(x)y = f(x)$$

where $a_k(x) \neq 0$. The equation is **homogeneous** if $f(x) = 0$.

First-order linear ODEs

Looking specifically at first order linear ODEs, which take the general form

$$\frac{\mathrm{d}y}{\mathrm{d}x} + P(x)y = Q(x)$$

we see that the homogeneous form, that is when $Q(x) = 0$, is separable. On the other hand, the inhomogeneous form can be solved using an **integrating factor** $I(x)$ given by

$$I(x) = e^{\int P(x)\mathrm{d}x}$$

Proof. Simply multiply the general equation for first-order linear ODEs through by the integrating factor to obtain

$$e^{\int P(x)\mathrm{d}x}\frac{\mathrm{d}y}{\mathrm{d}x} + P(x)e^{\int P(x)\mathrm{d}x}y = e^{\int P(x)\mathrm{d}x}Q(x)$$

Using the product rule for derivatives, we see that this gives

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\int P(x)\mathrm{d}x}y\right) = e^{\int P(x)\mathrm{d}x}Q(x)$$

and we can now integrate directly and rearrange, to obtain

$$y = e^{-\int P(x)\mathrm{d}x}\left[\int e^{\int P(x)dx}Q(x)\,\mathrm{d}x + c\right].$$

□

Example 10.4.4

Solve the linear differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} + 2xy = 2xe^{-x^2}.$$

Solution. We can easily see that the given differential equation is in the form of a first-order linear ODE.

First we find the integrating factor:

$$I(x) = e^{\int 2x \mathrm{d}x} = e^{x^2}$$

Multiplying the given differential equation through by the integrating factor this gives

$$e^{x^2}\frac{\mathrm{d}y}{\mathrm{d}x} + 2xe^{x^2}y = 2x$$

that is

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{x^2}y\right) = 2x$$

Integrating this gives us

$$e^{x^2}y = x^2 + c$$

so the general solution is $y = (x^2 + c)e^{-x^2}$. □

§10.4.3 Second-order differential equations

The main subject of this section is linear ODEs with constant coefficients, but before we look at these we give two theorems that are valid in the more general case.

Two theorems

Second-order homogeneous linear ODEs

§10.5 Laplace transform

Definition 10.5.1: Laplace transform

The Laplace transform of a signal (function) f is the function $F = \mathcal{L}(f)$ defined by

$$F(s) = \int_0^\infty f(t)e^{-st}\,\mathrm{d}t \tag{10.23}$$

for those $s \in \mathbb{C}$ for which the integral makes sense.

Remark. F is a complex-valued function of complex numbers. s is called the (complex) frequency variable, with units sec^{-1}; t is called the time variable (in sec); st is unitless. For now, we assume f contains no impulses at $t = 0$.

Notation. Lowercase letter denotes signal; uppercase letter denotes its Laplace transform; for example, U denotes $\mathcal{L}(u)$, V_{in} denotes $\mathcal{L}(v_{\text{in}})$.

11 Multivariable Calculus

§11.1 Introduction

§11.1.1 Vectors

From the Cartesian product in Set Theory, we know that

$$\mathbb{R}^n = (x_1, x_2, \dots, x_n)$$

which is the set of all n-tuples of real numbers x.

The elements of $\mathbb{R}^n$ are the points in n-dimensional space and are also called $n-$dimensional vectors.

Vector operations

- Scalar multiplication

  Given a vector $x = (x_1, \dots, x_n)$ in $\mathbb{R}^n$ and a scalar $\alpha \in \mathbb{R}$, the product is the vector

  $$\alpha x = (\alpha x_1, \dots, \alpha x_n).$$

- Addition and subtraction

  Another vector $y = (y_1, \dots, y_n)$ can to added to x to give a vector

  $$x + y = (x_1 + y_1, \dots, x_n + y_n).$$

  Similarly, we can also subtract vectors defining $x - y = x + (-1)y$ and then

  $$x - y = (x_1 - y1, \dots, x_n - y_n).$$

  Remark. Because elements of $\mathbb{R}^n$ can be multiplied by a scalar and added, $\mathbb{R}^n$ is a vector space.

- Magnitude

  A vector $x = (x_1, \dots, x_n)$ has a magnitude (length) of

  $$|x| = \sqrt{x_1^2 + \cdots + x_n^n}$$

Since $x - y$ goes from point y to point x, the length of this vector is the distance between the points:

$$|x - y| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

- Dot product

  The dot product of vectors x and y in $\mathbb{R}^n$ is a scalar given by

  $$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

  Then we have the following corollary:

  $$x \cdot x = |x|^2$$

§11.1.2 Functions of several variables

We are interested in functions f from $\mathbb{R}^n$ to $\mathbb{R}^m$ (or more generally from a subset $D \subset \mathbb{R}^n$ to $\mathbb{R}^m$ called the domain of the function). A function f assigns to each $x \in \mathbb{R}^n$ a point $y \in \mathbb{R}^m$ and we write

$$y = f(x)$$

The set of all such points y is the range of the function.

Each component of $y = (y_1, \dots, y_m)$ is real-valued function of $x \in \mathbb{R}^n$ written $y_i = f_i(x)$ and the function can also be written as the collection of n functions

$$y_1 = f_1(x), \dots, y_m = f_m(x)$$

If we also write out the components of $x = (x_1, \dots, x_n)$, then are function can be written as m functions of n variables each:

$$\begin{aligned} y_1 &= f_1(x_1, \dots, x_n) \\ y_2 &= f_2(x_1, \dots, x_n) \\ &\vdots \\ y_m &= f_m(x_1, \dots, x_n) \end{aligned}$$

The graph of the function is all pairs (x, y) with $y = f(x)$. It is a subset of $\mathbb{R}^{n+m}$.

Here are some special cases that are of particular interest.

1. $n = 1, m = 2$ (or $m = 3$). The function has the form

   $$y_1 = f_1(x), y_2 = f_2(x)$$

   In this case the range of the function is a curve in $\mathbb{R}^2$.

2. $n = 2, m = 1$. Then function has the form

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

   The graph of the function is a surface in $\mathbb{R}^3$.

3. $n = 2, m = 3$. The function has the form

$$\begin{aligned} y_1 &= f_1(x_1, x_2) \\ y_2 &= f_2(x_1, x_2) \\ y_3 &= f_3(x_1, x_2) \end{aligned}$$

   The range of the function is a surface in $\mathbb{R}^3$.

4. $n = 3, m = 3$. The function has the form

$$\begin{aligned} y_1 &= f_1(x_1, x_2, x_3) \\ y_2 &= f_2(x_1, x_2, x_3) \\ y_3 &= f_3(x_1, x_2, x_3) \end{aligned}$$

   The function assigns a vector to each point in space and is called a **vector field**.

§11.1.3 Limits

Consider a function $y = f(x)$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ (or possibly a subset of $\mathbb{R}^n$). Let $x_0 = (x_{01}, \ldots, x_{0n})$ be a point in $\mathbb{R}^n$ and let $y_0 = (y_{01}, \ldots, y_{0m})$ be a point in $\mathbb{R}^m$. We say that y_0 is the limit of f as x goes to x_0, written

$$\lim_{x \to x_0} f(x) = y_0 \tag{11.1}$$

if for every $\varepsilon > 0$ there exists a $\delta > 0$ so that if $|x - x_0| < \delta$ then $|f(x) - y_0| < \varepsilon$. The function is continuous at x_0 if

$$\lim_{x \to x_0} f(x) = f(x_0) \tag{11.2}$$

The function is continuous if it is continuous at every point in its domain.

§11.2 Partial Derivatives

§11.2.1 Limits

We take the limit of the function $f(x, y)$ as x approaches a and as y approaches b. This is denoted by

$$\lim_{(x,y)\to(a,b)} f(x, y)$$

Definition 11.2.1: Continuity

A function $f(x, y)$ is continuous at the point (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b)$$

§11.2.2 What it is

Definition 11.2.2: Partial derivative

Suppose f is a function from $\mathbb{R}^2$ to $\mathbb{R}$, given by $z = f(x, y)$. The **partial derivative** of f with respect to x at (x_0, y_0) is defined as

$$f_x(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} \tag{11.3}$$

if the limit exists.
Similarly, the partial derivative of f with respect to y at (x_0, y_0) is defined as

$$f_y(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}. \tag{11.4}$$

Notation. We also use the notation

$$f_x = \frac{\partial f}{\partial x} \quad \text{and} \quad f_y = \frac{\partial f}{\partial y}.$$

Notation. We can also take second partial derivatives, given by

$$f_{xx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2}$$
$$f_{yy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2}$$
$$f_{xy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x\,\partial y}$$
$$f_{yx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y\,\partial x}$$

§11.2.3 How to Do Partial Derivatives

Keep in mind that we only need to find the derivative of functions with respect to one variable by keeping the rest of the variables constant.

Example 11.2.1: First partial derivative

Find the partial derivative of $f(x,y) = 2x^2 - 4xy + y^2$ with respect to x.

Solution.

$$\begin{aligned}\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x}(2x^2 - 4xy + y^2)\\ &= 2(2x) - 4(1)y + 0\\ &= 4x - 4y\end{aligned}$$

□

Example 11.2.2: Second partial derivative

Given that $f(x,y) = 12x^2y - 3xy^2$, find f_{yx}.

Solution.

$$\begin{aligned}\frac{\partial^2 f}{\partial x\,\partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\\ &= \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}(12x^2y - 3xy^2)\right]\\ &= \frac{\partial}{\partial x}[12x^2(1) - 3x(2y)]\\ &= \frac{\partial}{\partial x}(12x^2 - 6xy)\\ &= 12(2x) - 6y(1)\\ &= 24x - 6y\end{aligned}$$

□

Theorem 11.2.1

If f_x, f_y, f_{xy}, f_{yx} exist and are continuous near (x_0, y_0), then

$$f_{xy}(x_0, y_0) = fyx(x_0, y_0)$$

§11.2.4 Directional Derivatives

To this point we've only looked at the two partial derivatives $f_x(x,y)$ and $f_y(x,y)$. Recall that these derivatives represent the rate of change of f as we vary x (holding y fixed) and as we vary y (holding x fixed) respectively.

We now discuss how to find the rate of change of f if we allow both x and y to change simultaneously. The problem here is that there are many ways to allow both x and y to change. For instance, one could be changing faster than the other and then there is also the issue of whether or not each is increasing or decreasing. So, before we get into finding the rate of change we need to get a couple of preliminary ideas taken care of first. The main idea that we need to look at is just how are we going to define the changing of x and/or y.

Let's start off by supposing that we wanted the rate of change of f at a particular point, say (x_0, y_0). Let's also suppose that both x and y are increasing and that, in this case, x is increasing twice as fast as y is increasing. So as y increases one unit of measure, x increases two units of measure.

Let's suppose that a particle is sitting at (x_0, y_0) and the particle will move in the direction given by the changing x and y. At this point, the particle can be said to be moving in the direction

$$\vec{v} = \langle 2, 1 \rangle$$

There is still a small problem with this however. There are many vectors that point in the same direction. For instance, all of the following vectors point in the same direction as $\vec{v} = \langle 2, 1 \rangle$:

$$\vec{v} = \left\langle \frac{1}{5}, \frac{1}{10} \right\rangle \quad \vec{v} = \langle 6, 3 \rangle \quad \vec{v} = \left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle$$

We need a way to consistently find the rate of change of a function in a given direction. We will do this by insisting that the vector that defines the direction of change be a unit vector. This means that for the example that we started off thinking about we would want to use

$$\vec{v} = \left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle$$

Definition 11.2.3: Directional derivative

Rate of change of $f(x, y)$ in the direction of the unit vector $\vec{u} = \langle a, b \rangle$ is called the directional derivative and is denoted by $D_{\vec{u}} f(x, y)$.
The definition of the directional derivative is

$$D_{\vec{u}} f(x, y) = \lim_{h \to 0} \frac{f(x + ah, y + bh) - f(x, y)}{h} \tag{11.5}$$

To derive an equivalent formula for taking directional derivatives, we define a new function of a single variable

$$g(z) = f(x_0 + az, y_0 + bz)$$

where x_0, y_0, a, b are some fixed numbers. Note that this really is a function of a single variable z.

Then by the definition of the derivative for functions of a single variable we have

$$g'(z) = \lim_{h \to 0} \frac{g(z + h) - g(z)}{h}$$

and the derivative at $z = 0$ is given by

$$g'(0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h}$$

If we now substitute in for $g(z)$ we get

$$g'(0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h} = \lim_{h\to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h} = D_{\vec{u}} f(x_0, y_0)$$

This gives us

$$g'(0) = D_{\vec{u}} f(x_0, y_0) \tag{1}$$

Now, let's look at this from another perspective. Let's rewrite $g(z)$ as $g(z) = f(x, y)$ where $x = x_0 + az$ and $y = y_0 + bz$. Applying chain rule,

$$g'(z) = \frac{\mathrm{d}g}{\mathrm{d}z} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}z} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}z} = f_x(x, y)a + f_y(x, y)b$$

This gives us

$$g'(z) = f_x(x, y)a + f_y(x, y)b$$

If we take $z = 0$ we get $x = x_0$ and $y = y_0$. Plugging these into the above equation gives

$$g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b \tag{2}$$

Equating (1) and (2) gives

$$D_{\vec{u}} f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b$$

Allowing x and y to be any number we get the following formula for computing directional derivatives:

$$D_{\vec{u}} f(x, y) = f_x(x, y)a + f_y(x, y)b$$

For three variables, directional derivative of $f(x, y, z)$ in the direction of the unit vector $\vec{u} = \langle a, b, c\rangle$ is given by

$$D_{\vec{u}} f(x, y, z) = f_x(x, y, z)a + f_y(x, y, z)b + f_z(x, y, z)c \tag{11.6}$$

We can write the directional derivative as a **dot product** and notice that the second vector is nothing more than the unit vector $\vec{u}$ that gives the direction of change.

$$D_{\vec{u}} f(x, y, z) = \langle f_x, f_y, f_z\rangle \cdot \langle a, b, c\rangle \tag{11.7}$$

Now let's give a name and notation to the first vector in the dot product since this vector will show up fairly regularly.

Definition 11.2.4: Gradient vector

he gradient vector of f is defined to be

$$\nabla f = \langle f_x, f_y, f_z\rangle \tag{11.8}$$

With the definition of the gradient we can now say that the directional derivative is given by

$$D_{\vec{u}} f = \nabla f \cdot \vec{u}$$

Theorem 11.2.2

Maximum value of $D_{\vec{u}} f(\vec{x})$ (and hence then the maximum rate of change of the function $f(\vec{x})$) is given by $\left\|\nabla f(\vec{x})\right\|$ and will occur in the direction given by $\nabla f(\vec{x})$.

Proof. □

§11.3 Partial differential equations

§11.3.1 Definitions and Terminology

Definition 11.3.1: Partial differential equation

An equation involving a function and/or its partial derivatives.

For example,

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2}$$

where $f(x,t)$ is a function of multiple variables.

We can classify PDEs based on:

- **Order.**

  The order is the number corresponding to the order of the highest partial derivative in the equation.

  For instance, the order of the following PDE is 2.

  $$\frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t}$$

  This also applies to mixed partial derivatives. For instance, the order of the following PDE is 3.

  $$\frac{\partial^3 f}{\partial x^2\,\partial y} = \frac{\partial f}{\partial t}$$

- **Number of independent variables.**

  An independent variable is what we differentiate with respect to.

- **Linearity.**

  A linear PDE is one in which the *dependent* variable (the one being differentiated) appears only in a linear fashion.

  For instance, the two PDEs above are linear as the partial derivatives are not being raised to a power or multiplied with each other.

  The following PDE is non-linear.

  $$f\frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t}$$

- **Homogeneity.**

  A homogenous PDE is one in which every term only involves the dependent variable and/or its derivatives.

  The first two PDEs above are homogenous as every term contains f or its derivatives.

The following PDE is non-homogenous as there are two terms that do not contain f.

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t} + x^2 + \tan t$$

- **Coefficient type.**

  The coefficient here refers to the coefficient of the term involving the dependent variable and its derivatives. It can be either constant or variable.

  For instance, the coefficients of the terms in the first two examples are 1. We say that these two PDEs have constant coefficients.

  The following PDE has variable coefficients.

$$\tan x \frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t}$$

- **Parabolic, Hyperbolic, or Elliptic.**

  We can do this classification for linear 2nd order PDEs which take the form of

$$A\frac{\partial^2 f}{\partial x^2} + B\frac{\partial^2 f}{\partial x\, \partial y} + C\frac{\partial^2 f}{\partial y^2} + D\frac{\partial f}{\partial x} + E\frac{\partial f}{\partial y} + Ff = G$$

  where the coefficients are generally functions of x or y.

  For a **hyperbolic** PDE, $B^2 - 4AC > 0$. Using variable substitutions to change x and y to η and ε respectively, we can reduce the PDE to

$$\frac{\partial^2 f}{\partial \eta^2} - \frac{\partial^2 f}{\partial \varepsilon^2} + g = 0$$

  where g denotes the first and lower order terms. This is similar to the equation of a hyperbola: $x^2 - y^2 = 1$.

  For a **parabolic** PDE, $B^2 - 4AC = 0$. Using variable substitutions, we can reduce the PDE to

$$\frac{\partial^2 f}{\partial \eta^2} + g = 0.$$

  This is similar to the equation of a parabola: $x^2 + y = 0$.

  For an **elliptic** PDE, $B^2 - 4AC < 0$. Using variable substitutions, we can reduce the PDE to

$$\frac{\partial^2 f}{\partial \eta^2} + \frac{\partial^2 f}{\partial \varepsilon^2} + g = 0.$$

  This is similar to the equation of an ellipse: $x^2 + y^2 = 1$.

  Note that if the coefficients are constants, the PDE can be hyperbolic, parabolic or elliptic. However, if the coefficients are variables, then it is possible for the PDE to be hyperbolic in some regions, and elliptic or parabolic in some regions.

§11.3.2 Solutions and Auxiliary Conditions

There are a lot of solutions to a given PDE, hence it is important for us to know the auxiliary conditions, i.e. boundary and initial conditions, which dictate which technique we use to solve the PDE.

- A boundary condition expresses the behavior of a function on the boundary (border) of its area of definition. An initial condition is like a boundary condition, but then for the time-direction.

§11.4 Double integrals

We want to integrate a function of two variables, $f(x,y)$. With functions of one variable we integrated over an *interval* (i.e. a one-dimensional space) and so it makes some sense then that when integrating a function of two variables we will integrate over a *region* of $\mathbb{R}^2$ (two-dimensional space).

We will start out by assuming that the region in $\mathbb{R}^2$ is a rectangle which we will denote as follows,

$$R = [a,b] \times [c,d]$$

This means that the ranges for x and y are $a \leq x \leq b$ and $c \leq y \leq d$.

§11.5 Line integrals

§11.5.1 Vector fields

A vector field is basically what you get when associating each point in space with a vector.

Definition 11.5.1: Vector field

A **vector field** in $\mathbb{R}^n$ is a function $F : \mathbb{R}^n \to \mathbb{R}^n$ that assigns to each $x \in \mathbb{R}^n$ a vector $F(x)$. A vector field in $\mathbb{R}^n$ with domain $U \subset \mathbb{R}^n$ is called a vector field on U.

The standard notation for the function $\vec{F}$ is:

$$\begin{aligned}\vec{F}(x,y) &= P(x,y)\hat{\imath} + Q(x,y)\hat{\jmath}\\ \vec{F}(x,y,z) &= P(x,y,z)\hat{\imath} + Q(x,y,z)\hat{\jmath} + R(x,y,z)\hat{k}\end{aligned}$$

depending on whether or not we're in two or three dimensions. The functions P, Q, R are called **scalar functions**.

Example 11.5.1

Sketch the following vector field:

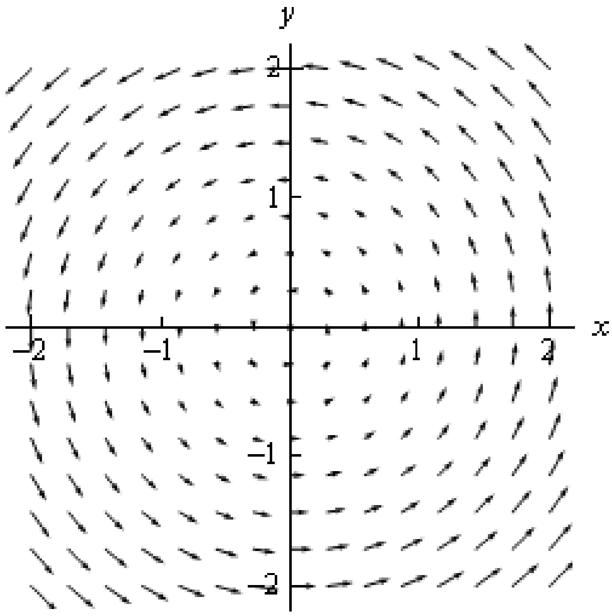
$$\vec{F}(x,y) = -y\hat{\imath} + x\hat{\jmath}$$

Solution. To graph the vector field we need to get some "values" of the function. This means plugging in some points into the function. Here are a couple of evaluations:

$$\begin{aligned}\vec{F}\left(\frac{1}{2},\frac{1}{2}\right) &= -\frac{1}{2}\hat{\imath} + \frac{1}{2}\hat{\jmath}\\ \vec{F}\left(\frac{1}{2},-\frac{1}{2}\right) &= -\left(-\frac{1}{2}\right)\hat{\imath} + \frac{1}{2}\hat{\jmath} = \frac{1}{2}\hat{\imath} + \frac{1}{2}\hat{\jmath}\\ \vec{F}\left(\frac{3}{2},\frac{1}{4}\right) &= -\frac{1}{4}\hat{\imath} + \frac{3}{2}\hat{\jmath}\end{aligned}$$

So what do these evaluations tell us? The first one tells us that at the point $\left(\frac{1}{2},\frac{1}{2}\right)$ we plot the vector $-\frac{1}{2}\hat{\imath} + \frac{1}{2}\hat{\jmath}$.

Plotting points gives us the following sketch of the vector field:



□

Definition 11.5.2: Gradient vector

Given a function $f(x, y, z)$, the gradient vector is defined by

$$\nabla f = \langle f_x, f_y, f_z \rangle \tag{11.9}$$

This is a vector field and is often called a gradient vector field.

§11.5.2 Types of line integrals

§11.5.3 Fundamental Theorem for Line Integrals

Theorem 11.5.1: Fundamental Theorem of Line Integrals

Suppose that C is a smooth curve from points A to B parameterised by $\mathbf{r}(t)$ for $t \in [a, b]$. Let f be a differentiable function whose domain includes C and whose gradient vector ∇f is continuous on C. Then

$$\int_C \nabla f \, \mathrm{d}\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a)) = f(B) - f(A) \tag{11.10}$$

Remark. Similar to the fundamental theorem of calculus, the primary change is that gradient ∇f takes the place of the derivative f'.

§11.5.4 Conservative Vector Fields

§11.5.5 Green's Theorem

to compute arc lengths, areas of curves

applications of integrals to find area and volume

12 Fourier Analysis

What is fourier analysis?

Fourier analysis is the study of how general functions can be decomposed into trigonometric or exponential functions with definite frequencies. There are two types of Fourier expansions:

- Fourier series: If a (reasonably well-behaved) function is periodic, then it can be written as a discrete sum of trigonometric or exponential functions with specific frequencies.
- Fourier transform: A general function that is not necessarily periodic (but that is still reasonably well-behaved) can be written as a continuous integral of trigonometric or exponential functions with a continuum of possible frequencies.

§12.1 Fourier Trigonometric Series

Fourier's theorem states that any (reasonably well-behaved) function can be written in terms of trigonometric or exponential functions, which we will eventually prove this theorem later. What we will do is derive what the coefficients of the sinusoidal functions must be, under the assumption that any function can in fact be written in terms of them.

Consider a function $f(x)$ that is periodic on the interval $0 \le x \le L$. Fourier's theorem works even if $f(x)$ is not continuous, although an interesting thing happens at the discontinuities, which we will talk about later. Other conventions for the interval are $-L \le x \le L$, or $0 \le x \le 1$, or $-\pi \le x \le \pi$, etc. There are many different conventions, but they all lead to the same general result in the end. If we assume $0 \le x \le L$ periodicity, then Fourier's theorem states that $f(x)$ can be written as

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi n x}{L} + b_n \sin\left(\frac{2\pi n x}{L}\right)\right)\right] \tag{12.1}$$

where coefficients a_i and b_i take on certain values that we will calculate below. This expression is the **Fourier trigonometric series** for the function $f(x)$. We could alternatively not separate out the a_0 term, and instead let the sum run from $n = 0$ to ∞, because $cos(0) = 1$ and $sin(0) = 0$. But the normal convention is to isolate the a_0 term.

With the 2π included in the arguments of the trig functions, the $n = 1$ terms have period L, the $n = 2$ terms have period $\frac{L}{2}$, and so on. So for any integer n, an integral number of oscillations fit into the period L. The expression in eq. (12.1) therefore has a period of (at most) L, which is a necessary requirement, of course, for it to equal the original periodic function $f(x)$. The period can be shorter than L if, say, only the even n's have non-zero coefficients (in which case the period is L/2). But it can't be longer than L; the function repeats at least as often as with period L.

We're actually making two statements in eq. (12.1). The first statement is that any periodic function can be written this way. This is by no means obvious, and it is the part of the theorem that we're accepting here. The second statement is that coefficients a_i and b_i take on particular values, assuming that the function f(x) can be written this way. It's reasonably straightforward to determine what these values are, in terms of f(x), and we'll do this below. But we'll first need to discuss the concept of orthogonal functions.

§12.2 Fourier Exponential Series

§12.3 Fourier Transform

§12.4 Special functions

§12.4.1 Gaussian

§12.4.2 Exponential, Lorentzian

§12.4.3 Square wave, sinc

§12.5 The delta function

§12.6 Gibbs phenomenon

§12.7 Convergence

§12.8 Relation between transforms and series

Part IV

Abstract Algebra

13 Group Theory

Readings:

- Group Theory by J.S. Milne
- Introduction to Groups, Rings and Fields by Oxford
- Math 33300: Group Theory
- Math 179: Graph Theory
- Groups, Fields and Polynomials

§13.1 Group Axioms

A **group** is an algebraic structure that captures the idea of symmetry without an object.

Definition 13.1.1: Group

A **group** is a pair $(G, *)$, where G is a set and $*$ is a binary operation on G satisfying the following **group axioms** for all $a, b, c \in G$:

G1 Associativity: $a * (b * c) = (a * b) * c$

G2 Identity: There exists an identity element $1_G \in G$ such that for all $a \in G$, $a * 1_G = 1_G * a = a$

G3 Invertibility: For all $a \in G$, there exists a unique inverse $a^{-1} \in G$ such that $a * a^{-1} = a^{-1} * a = 1_G$

G4 Closure: $a * b \in G$

G is **abelian**[1] if the operation is commutative; it is **non-abelian** if otherwise.

Notation. A group $(G, *)$ is usually simply denoted by G.

Example 13.1.1: Additive integers

The pair $(\mathbb{Z}, +)$ is a group. Note that

- The element $0 \in \mathbb{Z}$ is an identity: $a + 0 = 0 + a = a$ for any a.
- Every element $a \in \mathbb{Z}$ has an additive inverse: $a + (-a) = (-a) + a = 0$.

We call this group $\mathbb{Z}$.

Example 13.1.2: Non-zero rationals

Let $\mathbb{Q}^\times$ be the set of nonzero rational numbers. The pair $(\mathbb{Q}^\times, \cdot)$ is a group: the set is Q× and the associative operation is multiplication.

- The element $1 \in \mathbb{Q}^\times$ is an identity: for any rational number, $a \cdot 1 = 1 \cdot a = a$.
- For any rational number $x \in \mathbb{Q}^\times$, we have an inverse x^{-1}, such that $x \cdot x^{-1} = x^{-1} \cdot x = 1$.

Example 13.1.3: Addition mod n

Let $n > 1$ be an integer, and consider the residues (remainders) modulo n. These form a group under addition. We call this the cyclic group of order n, and denote it as $\mathbb{Z}/n\mathbb{Z}$, with elements $0, 1, \ldots, n-1$. The identity is 0.

[1] After the Norwegian mathematician Niels Abel (1802–1829)

Definition 13.1.2: Order

The **order** of a finite group G is the number of elements in G, denoted by $|G|$. The order of $a \in G$ is the least k such that $a^k = 1_G$. This is consistent with the definition of order of a group, as the order of a is the order of the subgroup generated by a.

Definition 13.1.3: Subgroup

A **subgroup** H of a group G is a *subset* of G which is a group under the operation of G restricted to H. We write $H \leq G$. In particular, a subset $H \subseteq G$ is a subgroup if it is closed under the operation of G.

Definition 13.1.4: Coset

A (left) **coset** of a subgroup $H \leq G$ is a set $aH = \{ah \mid h \in H\}$.

Two (left) cosets aH and bH are either disjoint or equal.

Since multiplication is injective, the cosets of H are the same size as H, and thus H partitions G into equal-sized parts.

§13.1.1 Properties of groups

We abbreviate $a * b$ to just ab. Also, since the operation $*$ is associative, we omit unnecessary parentheses: $(ab)c = a(bc) = abc$.

For any $g \in G$ and $n \in \mathbb{N}$ we abbreviate $g^n = \underbrace{g * \cdots * g}_{n \text{ times}}$.

Proposition 13.1.1 (Uniqueness of identity element)**.** The identity of a group is unique.

Proof. Suppose otherwise, that ${1_G}^{-1}$ is a second identity element, then ${1_G}^{-1} = 1_G {1_G}^{-1} = 1_G$, thus proving the uniqueness of the identity element 1_G. □

Proposition 13.1.2 (Uniqueness of inverse)**.** The inverse of a is unique.

Proof. Suppose otherwise, that b and c are the inverses of a. Then $ba = ac = 1_G$.

$$b = b1_G = b(ac) = (ba)c = 1_G c = c$$

Hence the element a^{-1} in item **G3** is uniquely determined by a. □

Proposition 13.1.3. Cancellation laws hold in groups.

Proof. By item **G3**,

$$ab = ac \implies b = c, \quad ba = ca \implies b = c$$

by multiplying a^{-1} on LHS or RHS. □

Proposition 13.1.4 (Inverse of products)**.** For $a, b \in G$, $(ab)^{-1} = b^{-1}a^{-1}$.

Proof. Direct computation. We have

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aa^{-1} = 1_G.$$

Similarly,

$$(b^{-1}a^{-1})(ab) = 1_G.$$

Hence equating both gives us $(ab)^{-1} = b^{-1}a^{-1}$. □

Proposition 13.1.5 (Left multiplication is a bijection)**.** For a group G, pick a $g \in G$. Then the map $G \to G$ given by $x \mapsto gx$ is a bijection.

Proof. Check this by showing injectivity and surjectivity directly. □

§13.2 Isomorphism

Definition 13.2.1: Isomorphism

Let $G = (G, *)$ and $H = (H, \star)$ be. A bijection $\varphi : G \to H$ is called an **isomorphism** if, for all $g_1, g_2 \in G$,

$$\varphi(g1 * g2) = \varphi(g1) \star \varphi(g2).$$

If there exists an isomorphism from G to H, G and H are **isomorphic**, denoted by $G \cong H$.

Remark. Note that in this definition, the left-hand side $\varphi(g1 * g2)$ uses the operation of G while the right-hand side $\varphi(g1) \star \varphi(g2)$ uses the operation of H.

Example 13.2.1: $\mathbb{Z} \cong 10\mathbb{Z}$

Consider the two groups

$$\mathbb{Z} = (\{\dots, -2, -1, 0, 1, 2, \dots\}, +)$$

and

$$10\mathbb{Z} = (\{\dots, -20, -10, 0, 10, 20, \dots\}, +).$$

These groups are “different", but only superficially so — you might even say they only differ in the names of the elements.
Formally, the map

$$\varphi : \mathbb{Z} \to 10\mathbb{Z} \text{ by } x \mapsto 10x$$

is a bijection of the underlying sets which respects the group operation. In symbols,

$$\varphi(x + y) = \varphi(x) + \varphi(y).$$

In other words, φ is a way of re-assigning names of the elements without changing the structure of the group.

§13.3 Lagrange's theorem

An important result relating the order of a group with the orders of its subgroups is Lagrange's theorem.

Theorem 13.3.1: Lagrange's theorem

If G is a finite group and H is a subgroup of G, then $|H|$ divides $|G|$.

Groups of small order (up to order 8). Quaternions. Fermat-Euler theorem from the group-theoretic point of view.

Theorem 13.3.2: Fermat's Little Theorem

For every finite group G, for all $a \in G$, $a^{|G|} = 1_G$.

Proof. Consider the subgroup H generated by a: $H = \{a^i \mid i \in \mathbb{Z}\}$. Since G is finite, the infinite sequence $a^0 = 1_G, a^1, a^2, a^3, \ldots$ must repeat, say $a^i = a^j, i < j$. Let $k = j - i$. Multiplying both sides by $a^{-i} = (a^{-1})^i$, we get $a^{j-i} = a^k = 1_G$. Suppose k is the least positive integer for which this holds. Then $H = \{a_0, a_1, a_2, \ldots, a^{k-1}\}$, and thus $|H| = k$. By Lagrange's Theorem, $k divides |G|$, so $a^{|G|} = (a^k)^{\frac{|G|}{k}} = 1_G$. □

§13.4 Group actions

Group actions; orbits and stabilizers. Orbit-stabilizer theorem. Cayley's theorem (every group is isomorphic to a subgroup of a permutation group). Conjugacy classes. Cauchy's theorem.

§13.5 Quotient groups

Normal subgroups, quotient groups and the isomorphism theorem

§13.6 Matrix groups

The general and special linear groups; relation with the M¨obius group. The orthogonal and special orthogonal groups. Proof (in R3) that every element of the orthogonal group is the product of reflections and every rotation in R3 has an axis. Basis change as an example of conjugation.

§13.7 Permutations

Permutations, cycles and transpositions. The sign of a permutation. Conjugacy in Sn and in An. Simple groups; simplicity of A5.

14 Ring Theory

Readings:

- Ring Theory by Brilliant
- Ring Theory (Math 113) by UC Berkeley

§14.1 Definition

A ring is just a set where you can add, subtract, and multiply. In some rings you can divide, and in others you can't. There are many familiar examples of rings, the main ones falling into two camps: "number systems" and "functions".

Definition 14.1.1: Ring

A ring is a set R endowed with two binary operations, addition and multiplication, denoted $+$ and $\cdot$, such that

R1 R is an **abelian** group with respect to $+$

R2 **Associativity of multiplication**: $\forall a, b, c \in R$,

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c$$

R3 Left **distributivity**: $\forall a, b, c \in R$,

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

Right distributivity: $\forall a, b, c \in R$,

$$(a + b) \cdot c = a \cdot c + b \cdot c$$

R4 **Multiplicative identity**: there exists an element, denoted 1, which has the property that $\forall a \in R$,

$$a \cdot 1 = 1 \cdot a = a$$

R5 **Commutativity of multiplication**: $\forall a, b \in R$,

$$a \cdot b = b \cdot a$$

Examples of rings:

- $\mathbb{Z}$: the integers $\ldots, -2, -1, 0, 1, 2, \ldots$ with usual addition and multiplication, form a ring. Note that we cannot always divide, since $1/2$ is no longer an integer.
- $2\mathbb{Z}$: the even integers $\ldots, -4, -2, 0, 2, 4, \ldots$
- $\mathbb{Z}[x]$: this is the set of polynomials whose coefficients are integers.

  It is an extension of $\mathbb{Z}$, in the sense that we allow all the integers, plus an "extra symbol" x, which we are allowed to multiply and add, giving rise to x^2, x^3, etc., as well as $2x$, $3x$, etc. Adding up various combinations of these gives all the possible integer polynomials.
- $\mathbb{Z}[x, y, z]$: polynomials in three variables with integer coefficients.

  This is an extension of the previous ring. In fact you can continue adding variables to get larger and larger rings.

- $\mathbb{Z}/n\mathbb{Z}$: integers mod n.

  These are equivalence classes of the integers under the equivalence relation "congruence mod n". If we just think about addition (and subtraction), this is exactly the cyclic group of order n. However, when we call it a ring, it means we are also using the operation of multiplication.

Ideals, homomorphisms, quotient rings, isomorphism theorems. Prime and maximal ideals. Fields. The characteristic of a field. Field of fractions of an integral domain. Factorization in rings; units, primes and irreducibles. Unique factorization in principal ideal domains, and in polynomial rings. Gauss' Lemma and Eisenstein's irreducibility criterion. Rings $\mathbb{Z}[\alpha]$ of algebraic integers as subsets of $\mathbb{C}$ and quotients of $\mathbb{Z}[x]$. Examples of Euclidean domains and uniqueness and non-uniqueness of factorization. Factorization in the ring of Gaussian integers; representation of integers as sums of two squares. Ideals in polynomial rings. Hilbert basis theorem

15 Field Theory

§15.1 Field Axioms

The **field axioms** are as follows:

Definition 15.1.1: Field

There exists two operations — addition and multiplication — satisfying the following **field axioms** for all $x, y, z \in \mathbb{F}$:

F1 Commutative laws: $x + y = y + x$ and $xy = yx$

F2 Associative laws: $x + (y + z) = (x + y) + z$ and $x(yz) = (xy)z$

F3 Distributive laws: $x(y + z) = xy + xz$

F4 Existence of identity elements: There exists two distinct elements, denoted by 0 and 1, such that for all x, $x + 0 = x$ and $x1 = x$

F5 Existence of negatives: For all x there exists a unique element $-x$ such that $x + (-x) = 0$

F6 Existence of reciprocals: For all $x \neq 0$ there exists a unique element x^{-1} such that $xx^{-1} = 1$

Example 15.1.1: $\mathbb{Z}^+$

The set of **positive integers** $\mathbb{Z}^+$ is not a field because, for example, 0 is not a positive integer, for no positive integer n is $-n$ a positive integer, for no positive integer n except 1 is n^{-1} a positive integer.

Example 15.1.2: $\mathbb{Z}$

The set of **integers** $\mathbb{Z}$ is not a field because for an integer n, n^{-1} is not an integer unless $n = 1$ or $n = -1$.

Example 15.1.3: $\mathbb{Q}$

The set of **rational numbers** $\mathbb{Q}$ is a field.

From the field axioms we can deduce all the usual laws of elementary algebra.

16 Galois Theory

Readings:

- Notes by Tom Leinster

17 Category Theory

Readings:

- Basic Category Theory, by Tom Leinster

Part V

Real Analysis

18 Properties of the real numbers

§18.1 Construction of the real numbers

This book assumes familiarity with the rational numbers $\mathbb{Q}$, i.e. numbers of the form $\dfrac{m}{n}$, where m, n are integers and $n \neq 0$).

$\mathbb{Q}$ contains *gaps* at irrational numbers such as $\sqrt{2}$ and π. In this section, we aim to construct $\mathbb{R}$ from $\mathbb{Q}$.

In 1872, German mathematician Richard Dedekind pointed out that a real number x can be determined by its lower set A and upper set B:

$$A \coloneqq \{a : \mathbb{Q} \mid a < x\}$$

$$B \coloneqq \{b : \mathbb{Q} \mid x < b\}$$

He defined a "real number" as a pair of sets of rational numbers, the lower and upper sets shown above. Such a pair of sets of rational numbers are known as a **Dedekind cut**.

- A is a **lower set**: $\forall a, b \in \mathbb{R}$, if $a < b$ where $b \in A$, then $a \in A$.
- B is an **upper set**: $\forall a, b \in \mathbb{R}$, if $a < b$ where $a \in B$, then $b \in B$.

Definition 18.1.1: Dedekind cut

Given that B is the complement of A in the reals, a non-empty subset $(A, B) \subset \mathbb{Q}$ is a Dedekind cut if:

D1 A is non-empty

$$A \neq \varnothing$$

D2 A and B are disjoint

$$A \cup B = \mathbb{Q}$$

D3 A is closed downwards

$$\forall x, y \in \mathbb{Q} \text{ with } x < y, y \in A \implies x \in A$$

D4 A does not contain a greatest element

$$\forall x \in A, \exists y \in A \text{ such that } x < y$$

Perhaps a not-so-intuitive fact here is that there are two possible things happening to B:

1. B contains a least element
2. B does not contain a least element

Case 1 and 2 are known as rational and irrational Dedekind cuts respectively.

Definition 18.1.2: Real numbers

The set of real numbers $\mathbb{R}$ is defined to be the set of all Dedekind cuts.

Remark. The way we think about this is that Dedekind cuts are real numbers, and real numbers are Dedekind cuts.

§18.1.1 Order relations

Given real numbers α and β, let $\alpha = (A, B)$ and $\beta = (C, D)$. Then

$$\alpha < \beta \iff A \subset C$$

Remark. Since B is the complement of A, α is completely determined by A itself.

This ordering on the real numbers satisfies the following properties:

- $x < y$ and $y < z \implies x < z$
- Exactly one of $x < y$, $x = y$ or $x > y$ holds
- $x < y \implies x + z < y + z$

Property 18.1.1 (Ordering)**.** For any two real numbers α and β, one of the following must hold:

$$\alpha < \beta \quad \alpha = \beta \quad \alpha > \beta$$

Proof. We prove by contradiction.

Note that $\alpha \leq \beta \iff A \subseteq C$ ($A = C$ is possible).

Suppose otherwise, that all three of the above are false, then neither of the sets A and C can be a subset of the other.

We pick two rational numbers from each set: Pick p where $p \in A$, $p \notin C$, pick q where $q \in C$, $q \notin A$

- Obviously we cannot have $p = q$.
- If $p < q$, then since $q \in C$, according to property 3, we have $p \in C$, a contradiction.
- Similarly for $p > q$, we would find that $q \in A$, a contradiction.

Hence our assumption is false.

$\therefore$ One of the three cases $\alpha < \beta$, $\alpha = \beta$, $\alpha > \beta$ must hold. $\square$

§18.1.2 Addition

Property 18.1.2 (Addition)**.** Let $\alpha = (A, B)$, $\beta = (C, D)$, then $\alpha + \beta = (X, Y)$ where

$$X = \{a + c \mid a \in A, c \in C\}$$

Proof. To show that (X, Y) is a Dedekind cut, we simply need to check the conditions for Dedekind cuts.

- Property 1 is trivial.
- Property 2 is by definition.
- Property 3:

  Let $x, y \in X$ satisfy $x < y$, $y \in X$.

  Let $y = a + c$, $a \in A$, $c \in C$.

  Let $\varepsilon = y - x$.

  Let $a' = a - \dfrac{\varepsilon}{2}$, $c' = c - \dfrac{\varepsilon}{2}$.

  Then
  $$a' + c' = a + c - \varepsilon = x$$
  $a' < a, a \in A \implies a' \in A$. Similarly, $c' \in C$.
  $\therefore x = a' + c' \in X$.

- Property 4:

  $\forall a + c \in X, a \in A, c \in C$, $\exists a' \in A, c' \in C$ such that $a < a', c < c'$.

  $\therefore a' + c' \in X$ satisfies $a + c < a' + c'$.

□

Property 18.1.3 (Commutativity)**.** Addition is **commutative**:

$$\alpha + \beta = \beta + \alpha$$

Proof. The proof is trivial. □

Property 18.1.4 (Associativity)**.** Addition is **associative**:

$$\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$$

Proof. Let $\alpha = (A, A')$, $\beta = (B, B')$, $\gamma = (C, C')$

$$\beta + \gamma = (B + C, (B + C)')$$

In this notation we only need to show that $A + (B + C) = (A + B) + C$.

$$\begin{aligned}
&x \in A + (B + C) \\
&\iff \exists a \in A, p \in B + C \text{ such that } x = a + p \\
&\iff \exists a \in A, b \in B, c \in C, \text{ such that } x = a + b + c \\
&\iff x \in (A + B) + C
\end{aligned}$$

Hence proven. □

Example 18.1.1

Prove that
$$\alpha + 0 = \alpha = 0 + \alpha$$

Proof. Let $0 = (O, O')$ where $O = \{x \mid x < 0\}, O' = \{x \mid x \geq 0\}$.

Let $\alpha = (A, B)$, then $\alpha + 0 = (C, D)$ where

$$\begin{aligned} C &= \{a + \varepsilon \mid a \in A, \varepsilon < 0\} \\ &= \{a - \varepsilon \mid a \in A, \varepsilon > 0\} \end{aligned}$$

$$a - \varepsilon < a, a \in A \implies a - \varepsilon \in A \implies C \subseteq A$$

According to Property 4, $\forall a \in A, \exists a' \in A$ such that $a < a'$.

Let $\varepsilon = a' - a > 0$, then
$$a = a' - \varepsilon, a' \in A, \varepsilon > 0 \implies a \in C$$

So $A = C$.

$\therefore \alpha + 0 = \alpha$ $\square$

Example 18.1.2

Express $-\alpha$ in terms of α; show

$$\alpha + (-\alpha) = 0 = (-\alpha) + \alpha$$

Proof. We split this into two cases.

Case 1: α is a rational number, then $\alpha = (A, B)$ where $A = \{x \mid x < \alpha\}$, $B = \{x \mid x \geq \alpha\}$.

Let $-\alpha = (A', B')$, where $A' = \{x \mid x < -\alpha\}$, $B' = \{x \mid x \geq -\alpha\}$. We see that $\alpha + (-\alpha) \leq 0$ is obvious.

On the other hand, since $0 = (O, O')$, for any $\varepsilon < 0$ we have

$$\varepsilon = \left(\alpha + \frac{\varepsilon}{2}\right) + \left(-\alpha + \frac{\varepsilon}{2}\right) \in A + A'$$

Hence $\alpha + (-\alpha) = 0$.

Case 2: α is irrational, let $\alpha = (A, B)$ where B does not have a lowest value. Then $-B = \{-x \mid x \in B\}$ does not have a highest value.

We wish to define $-\alpha = (-B, -A)$, but first we need to show that this is well-defined by checking through all the conditions.

- Property 1: This is trivial.

- Property 2: Prove that $-A$ and B are disjoint.

  Note that $\forall x \in \mathbb{R}$, if $x = -y$, then exactly one out of $y \in A$ and $y \in B$ is true $\implies$ exactly one out of $x \in -B$ and $x \in -A$ is true.

- Property 3: Prove $-B$ is closed downwards.

  Suppose otherwise, that $x < y, y \in -B$ but $x \notin -B$. Then $-y \in B$, $-x \notin B$. Since A is the complement of B, $-y \notin A$, $-x \in A$. But $-y < -x$, which is a contradiction.

- Property 4 is already guaranteed by the irrationality of α.

All of these properties imply that the real numbers form a commutative group by addition.

□

§18.1.3 Negation

Given any set $X \subset \mathbb{R}$, let $-X$ denote the set of the negatives of those rational numbers. That is $x \in X$ if and only if $-x \in -X$.

If (A, B) is a Dedekind cut, then $-(A, B)$ is defined to be $(-B, -A)$.

This is pretty clearly a Dedekind cut. - proof

§18.1.4 Signs

A Dedekind cut (A, B) is **positive** if $0 \in A$ and **negative** if $0 \in B$. If (A, B) is neither positive nor negative, then (A, B) is the cut representing 0.

If (A, B) is positive, then $-(A, B)$ is negative. Likewise, if (A, B) is negative, then $-(A, B)$ is positive. The cut (A, B) is non-negative if it is either positive or 0.

§18.1.5 Multiplication

Positive multiplication

Let $\alpha = (A, B)$ and $\beta = (C, D)$ where α, β are both non-negative.

We define $\alpha \times \beta$ to be the pair (X, Y) where

X is the set of all products ac where $a \in A, c \in C$ and at least one of the two numbers is non-negative. Y is the set of all products bd where $b \in B, d \in D$.

General Multiplication

Intermediate Value Theorem

Bolzano-Weiersstrass Theorem

Connectedness of $\mathbb{R}$

§18.2 Supremum and Infimum

§18.2.1 Ordered sets

Let A be a set.

Definition 18.2.1: Order

An **order** on A is a relation, denoted by $<$, with the following two properties:

O1 $\forall x, y \in A$, one and only one of the following statements is true:

$$x < y, \quad x = y, \quad y < x$$

O2 $\forall x, y, z \in A$, if $x < y$ and $y < z$, then $x < z$.

Notation. The notation $x \leq y$ indicates that $x < y$ or $x = y$, without specifying which of these two is to hold. In other words, $x \leq y$ is the negation of $x > y$.

Definition 18.2.2: Ordered set

An **ordered set** is a set S in which an order is defined.

For example, $\mathbb{Q}$ is an ordered set if $r < s$ is defined to mean that $s - r$ is a positive rational number.

§18.2.2 Boundedness

Let $A \subset \mathbb{R}$.

Definition 18.2.3: Bounded

A is **bounded from above** if there exists an **upper bound** $M \in \mathbb{R}$ such that $x \leq M$ for all $x \in A$.
A is **bounded from below** if there exists a **lower bound** $m \in \mathbb{R}$ such that $x \geq m$ for all $x \in A$.
A is **bounded** in the real numbers if it is bounded above and below.

Definition 18.2.4: Supremum

The **supremum** of A, denoted by $\sup A$, is defined as the smallest real number M such that $x \leq M$ for all $x \in A$.

(i) M is an upper bound for A.

(ii) If N is an upper bound for A, then $M \leq N$.

The suprenum is also known as the *least upper bound*.

Remark. If $M \in A$, then M is the **maximum value** of A.

The following proposition is convenient in working with suprema.

Proposition 18.2.1. Let A be a nonempty subset of $\mathbb{R}$ that is bounded above. Then $M = \sup A$ if an only if

(i) $x \leq M$ for all $x \in A$

(ii) For any $\varepsilon > 0$, there exists $a \in A$ such that $M - \varepsilon < a$.

Proof. Suppose first that $M = \sup A$. Then clearly (i) holds (since this is identical to condition (1) in the definition of supremum). Now let $\varepsilon > 0$. Since $M - \varepsilon < a$, condition (ii) in the definition of supremum implies that $M - \varepsilon$ is not an upper bound of A. Therefore, there must exist an element a in A such that $M - \varepsilon < a$, as desired. □

Definition 18.2.5: Infimum

The **infimum** of A, denoted by $\inf A$, is defined as the largest real number m such that $x \geq m$ for all $x \in A$.

(i) m is a lower bound for A.

(ii) If n is a lower bound for A, then $m \geq n$.

The infimum is also known as the *greatest lower bound.*

Remark. If $m \in A$, then m is the **minimum value** of A.

Proposition 18.2.2 (Uniqueness of suprenum)**.** If a set $A \subset \mathbb{R}$ has a supremum, then it is unique.

Proof. Assume that M and N are suprema of a set A.

Since N is a supremum, it is an upper bound for A. Since M is a supremum, then it is the least upper bound and thus $M \leq N$.

Similarly, since M is a supremum, it is an upper bound for A; since N is a supremum, it is a least upper bound and thus $N \leq M$.

Since $N \leq M$ and $M \leq N$, thus $M = N$. Therefore, a supremum for a set is unique if it exists. □

Theorem 18.2.1: Comparison Theorem

Let $S, T \subset \mathbb{R}$ be non-empty sets such that $s \leq t$ for every $s \in S$ and $t \in T$. If T has a supremum, then so does S, and $\sup S \leq \sup T$.

Proof. Let $\tau = \sup T$. Since τ is a supremum for T, then $t \leq \tau$ for all $t \in T$. Let $s \in S$ and choose any $t \in T$. Then, since $s \leq t$ and $t \leq \tau$, then $s \leq t$. Thus, τ is an upper bound for S.

By the Completeness Axiom, S has a supremum, say $\sigma = \sup S$. We will show that $\sigma \le \tau$. Notice that, by the above, τ is an upper bound for S. Since σ is the least upper bound for S, then $\sigma \le \tau$. Therefore,

$$\sup S \le \sup T.$$

□

Let's explore some useful properties of sup and inf.

Proposition 18.2.3. Let S, T be non-empty subsets of $\mathbb{R}$, with $S \subseteq T$ and with T bounded above. Then S is bounded above, and $\sup S \le \sup T$.

Proof. Since T is bounded above, it has an upper bound, say b. Then $t \le b$ for all $t \in T$, so certainly $t \le b$ for all $t \in S$, so b is an upper bound for S.

Now S, T are non-empty and bounded above, so by completeness each has a supremum. Note that $\sup T$ is an upper bound for T and hence also for S, so $\sup T \ge \sup S$ (since $\sup S$ is the least upper bound for S). □

Proposition 18.2.4. Let $T \subseteq \mathbb{R}$ be non-empty and bounded below. Let $S = \{-t \mid t \in T\}$. Then S is non-empty and bounded above. Furthermore, $\inf T$ exists, and $\inf T = -\sup S$.

Proof. Since T is non-empty, so is S. Let b be a lower bound for T, so $t \ge b$ for all $t \in T$. Then $-t \le -b$ for all $t \in T$, so $s \le -b$ for all $s \in S$, so $-b$ is an upper bound for S.

Now S is non-empty and bounded above, so by completeness it has a supremum. Then $s \le \sup S$ for all $s \in S$, so $t \ge -\sup S$ for all $t \in T$, so $-\sup S$ is a lower bound for T.

Also, we saw before that if b is a lower bound for T then $-b$ is an upper bound for S. Then $-b \ge \sup S$ (since $\sup S$ is the least upper bound), so $b \le -\sup S$. So $-\sup S$ is the greatest lower bound.

So $\inf T$ exists and $\inf T = -\sup S$. □

Proposition 18.2.5 (Approximation Property)**.** Let $S \subseteq \mathbb{R}$ be non-empty and bounded above. For any $\varepsilon > 0$, there is $s_\varepsilon \in S$ such that $\sup S - \varepsilon < s_\varepsilon \le \sup S$.

Proof. Take $\varepsilon > 0$.

Note that by definition of the supremum we have $s \le \sup S$ for all $s \in S$. Suppose, for a contradiction, that $\sup S - \varepsilon \ge s$ for all $s \in S$.

Then $\sup S - \varepsilon$ is an upper bound for S, but $\sup S - \varepsilon < \sup S$, which is a contradiction.

Hence there is $s_\varepsilon \in S$ with $\sup S - \varepsilon < s_\varepsilon$. □

Problem 18.2.1. Consider the set $\{\frac{1}{n} \mid n \in \mathbb{Z}^+\}$.

(a) Show that $\max S = 1$.

(b) Show that if d is a lower bound for S, then $d \le 0$.

(c) Use (b) to show that $0 = \inf S$.

Proof. □

If we are dealing with rational numbers, the sup/inf of a set may not exist. For example, a set of numbers in $\mathbb{Q}$, defined by $\{[\pi \cdot 10^n]/10^n\}$. 3,3.1,3.14,3.141,3.1415,3.14159,... But this set does not have an infimum in $\mathbb{Q}$.

By ZFC, we have the Completeness Axiom, which states that any non-empty set $A \subset \mathbb{R}$ that is bounded above has a supremum; in other words, if A is a non-empty set of real numbers that is bounded above, there exists a $M \in \mathbb{R}$ such that $M = \sup A$.

Problem 18.2.2. Find, with proof, the supremum and/or infimum of $\{\frac{1}{n}\}$.

Proof. sup1/n = max1/n = 1 inf1/n=0 as for all positive a, we can pick n=[1/a]+1, then a>1/n □

Problem 18.2.3. Find, with proof, the supremum and/or infimum of $\{\sin n\}$.

Proof. The answer is easy to guess: ±1

For the supremum, we need to show that 1 is the smallest we can pick, so for any $a = 1 - \varepsilon < 1$ we want to find an integer n close enough to $2k\pi + \dfrac{\pi}{2}$ so that $\sin n > a$.

Whenever we want to show the approximations between rational and irrational numbers we should think of the **pigeonhole principle**.

$$2k\pi + \frac{\pi}{2} = 6k + (2\pi - 6)k + \frac{\pi}{2}$$

Consider the set of fractional parts $\{(2\pi - 6)k\}$. Since this an infinite set, for any small number δ there is always two elements $\{(2\pi - 6)a\} < \{(2\pi - 6)b\}$ such that

$$|\{(2\pi - 6)b\} - \{(2\pi - 6)a\}| < \varepsilon$$

Then $\{(2\pi - 6)(b - a)\} < \delta$

We then multiply by some number m (basically adding one by one) so that

$$0 \le \{(2\pi - 6) \cdot m(b - a)\} - \left(2 - \frac{\pi}{2}\right) < \delta$$

Picking $k = m(b - a)$ thus gives

$$\begin{aligned} 2k\pi + \frac{\pi}{2} &= 6k + (2\pi - 6)k + \frac{\pi}{2} \\ &= 6k + [(2\pi - 6)k] + 2 + (2\pi - 6)k - (2 - \frac{\pi}{2}) \end{aligned}$$

Thus $n = 6k + [(2\pi - 6)k] + 2$ satisfies $\left|2k\pi + \dfrac{\pi}{2} - n\right| < \delta$

Now we're not exactly done here because we still need to talk about how well $\sin n$ approximates to 1.

We need one trigonometric fact: $\sin x < x$ for $x > 0$. (This simply states that the area of a sector in the unit circle is larger than the triangle determined by its endpoints.)

$$
\begin{aligned}
\sin n &= sin\left(n - \left(2k\pi + \frac{\pi}{2}\right) + \left(2k\pi + \frac{\pi}{2}\right)\right) \\
&= \cos\left(n - \left(2k\pi + \frac{\pi}{2}\right)\right) \\
&= \cos\theta
\end{aligned}
$$

$$
1 - \sin n = 2\sin^2\frac{\theta}{2} = 2\sin^2\left|\frac{\theta}{2}\right| \le \frac{\theta^2}{2} < \delta
$$

Hence we simply pick $\delta = \varepsilon$ to ensure that $1 - \sin n < \varepsilon$, and we're done. □

Theorem 18.2.2: Archimedean Principle

If $a, b \in \mathbb{R}$ with $a > 0$, then there exists $n \in \mathbb{N}$ such that $na > b$.

Proof. Suppose that the Archimedean Property is false. Then there exists $a, b \in \mathbb{R}, a > 0$ such that $na \le b$ for all $n \in \mathbb{N}$.

For these particular a and b, we can say that b is an upper bound of $S \coloneqq \{na \mid n \in \mathbb{N}\}$. From the completeness axiom, $s_0 \coloneqq \sup S$ exists. Let $n \in \mathbb{N}$, we have $n + 1 \in \mathbb{N}$. So $s_0 \ge (n+1)a = na + a$.

Then we have $s_0 - a \ge na$. This is true for all $n \in \mathbb{N}$. So $s_0 - a$ is an upper bound of S. However, $s_0 - a < s_0$, which contradicts that s_0 is the least upper bound of S. This contradiction shows that the Archimedean Property is true. □

§18.3 Completeness

§18.3.1 Completeness axiom

Theorem 18.3.1: Completeness axiom for the real numbers

Let A be a non-empty subset of $\mathbb{R}$ that is bounded above. Then A has a supremum.

Any set in the reals bounded from above/below must have a supremum/infimum.

Proof. We prove this using Dedekind cuts.

Let S be a real number set. We consider the rational number set $A = \{x \in \mathbb{Q} \mid \exists y \in S\}$. Set B is defined to be the complement of A in $\mathbb{Q}$.

We go through the definitions to check that $(A|B)$ is a Dedekind cut.

1. Since $S \neq \varnothing$, pick $y \in S$, then $[y] - 1$ is a real number smaller than some element in S, hence $[y] - 1 \in A$ and thus $A \neq \varnothing$.

   Since we're given that S is bounded, $\exists M > 0$ as the upper bound for S, thus $B \neq \varnothing$.

   (Note that an upper bound is simply a number that is bigger than anything from the set, and is not the supremum

2. We defined B to be the complement of A in $\mathbb{Q}$, so this condition is trivial.

3. For any $x, y \in A$, if $x < y$ and $y \in A$, then $\exists z \in S$ such that $y < z \implies x < z \implies x \in A$.

4. Suppose otherwise that $x \in A$ is the largest element in A, then $\exists y \in S$ such that $x < y$ We then pick a rational number z between x and y. Since we still have $z < y$, we have $z \in A$ but $z > x$, contradictory to z being the largest.

   Now there's actually an issue with the proof for property 4 here How exactly are we finding z?

   First $x \in \mathbb{Q}$. Then $y \in \mathbb{R}$ so we rewrite it as $y = (C|D)$ via definition.

   $x < y$ translates to the fact that $x \in C$.

   Since y is real, by definition we know that C must not have a largest element.

   In particular, x is not largest and we can pick $z \in C$ such that $z > x$. This is in fact the z that we need

Now that all the properties of a real number are validated, we may finally conclude that $\alpha = (A|B)$ is indeed a real number.

Now we need to show that $\alpha = \sup S$.

Let $x \in S$. If x is not the maximum value of S, i.e. $\exists y \in S, x < y$, then $x \in A$ and thus $x < \alpha$.

If x is the maximum value of S, then for any rational number $y < x$ we have $y \in A$, and for any rational number $y \geq x$ we have $y \in B$. Thus $x = (A|B) = \alpha$.

In conclusion, $x \leq \alpha$ for all $x \in S$.

For any upper bound x of S, since $\forall y \in S, x \geq y$ we have $x \in B$ and thus $x \geq \alpha$.

$\therefore$ α is the smallest upper bound of S and thus $\sup S = \alpha$ exists. $\square$

Theorem 18.3.2: Archimedean property of $\mathbb{N}$

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

Proof. Suppose, for a contradiction, that $\mathbb{N}$ is bounded above. Then $\mathbb{N}$ is non-empty and bounded above, so by completeness (of $\mathbb{R}$) $\mathbb{N}$ has a supremum.

By the Approximation property with $\varepsilon = \frac{1}{2}$, there is a natural number $n \in \mathbb{N}$ such that $\sup \mathbb{N} - \frac{1}{2} < n \leq \sup \mathbb{N}$.

Now $n + 1 \in \mathbb{N}$ and $n + 1 > \sup \mathbb{N}$. This is a contradiction. $\square$

§18.4 Order properties of the real numbers

§18.5 Topological properties of the real numbers

19 Numerical Sequences and Series

§19.1 Limit of a sequence

Definition 19.1.1: Convergence of sequence

Let $\{a_n\}$ be a real sequence, let $L \in \mathbb{R}$. We say that $\{a_n\}$ **converges** to L as $n \to \infty$ if

$$\forall \varepsilon 0 \exists N \in \mathbb{N} \text{ such that } \forall n \geq N, |a_n - L| < \varepsilon.$$

In this case we write $a_n \to L$ as $n \to \infty$, and we say that L is the **limit** of $\{a_n\}$. If $\{a_n\}$ does not converge, then we say that it **diverges**.

Remark. Take note of the use of logical statements

- ε is independent, so it is literally for all $\varepsilon > 0$.
- N is dependent on ε; if ε is very small we would expect the sequence x_n to get close enough to x further down the line.
- The order of the quantifiers matters.

Example 19.1.1

What do we really mean by saying that $\frac{1}{n} \to 0$ as $n \to \infty$?

We mean that the sequence of numbers $\frac{1}{n}$ converges to 0, proven as follows:

Proof. $\forall \varepsilon > 0$, pick $N = \frac{1}{\varepsilon} + 1$. Then $\forall n > N$,

$$\frac{1}{n} < \frac{1}{N} < \frac{1}{\frac{1}{\varepsilon}} = \varepsilon$$

□

Intuitively, the first thousand or million terms of a sequence shouldn't affect whether it converges. We'll prove a result that makes this precise, but first we need a quick definition.

Definition 19.1.2: Tail

Let $\{a_n\}$ be a sequence. A **tail** of $\{a_n\}$ is a sequence $\{b_n\}$, where for some natural number k we have $b_n = a_n + k$ for $n \geq 1$. That is, $\{b_n\}$ is the sequence obtained by deleting the first k terms of $\{a_n\}$.

Lemma 19.1.1 (Tails Lemma)**.** Let $\{a_n\}$ be a sequence.

Lemma 19.1.2. If a sequence in a metric space converges to a limit, this limit is unique.

This is false in a general topological space. We will discuss the properties of a topological space that will guarantee a sequence has a unique limit.

We can formulate the notion of the convergence of a sequence without mentioning the limit point. In this case, we want that the points of a sequence become arbitrarily close to each other (whereas above, we demanded that the points become arbitrarily close to a given point p)

Characteristics of limits:

1. Given a sequence of points $\{x_k\}$ and a point $x \in \mathbb{R}^n$, x_k converges to x if and only if all neighbourhoods of x "eventually" contain all x_k.

   By eventually we mean something similar to the definition above: there exists some large N such that the property is satisfied for all $n > N$.

   Proof. **Forward direction:**

   If $\{x_k\}$ converges to x, we wish to prove: given any neighbourhood U of x, U eventually contains all x_k.

   Since U is a neighbourhood of x, we pick a ball of radius ε centered at x, $B(x, \varepsilon)$, so that $B(x, \varepsilon)$ is contained in U.

   Then since $B(x, \varepsilon)$ is precisely the set of points whose distance to x is no larger than ε, we then apply the fact that $\{x_k\}$ converges to x.

   So for this particular ε, we take a natural number N so that $|x_k - x| < \varepsilon$, or $x_k \in B(x, \varepsilon)$, for all $k > N$.

   Then simultaneously x_k are in U since $B(x, \varepsilon)$ is a subset of U, thus we've shown that U will contain all x_k after a certain point N.

   Backward direction:

   Suppose that all neighbourhoods of x will eventually contain all x_k, then in particular for any $\varepsilon > 0$, since $B(x, \varepsilon)$ is a neighbourhood of x, it will also eventually contain all x_k.

   This then easily translates to the fact that $\{x_k\}$ converges to x. □

2. **Uniqueness of the limit**

   Suppose that $\{x_k\}$ converges to both x and x', then $x = x'$.

Proof. $\forall \varepsilon > 0$, we know that the terms in $\{x_k\}$ must be less than ε away from its limit after a certain point.

However, this certain point may not be the same for both limits; for the two limits x and x', we must first assume two separate numbers N and N' so that $|x_k - x| < \varepsilon$ when $k > N$, and $|x_k - x'| < \varepsilon$ when $k > N'$.

Now if you look at the book here, it says that we have a stronger requirement: $|x_k - x| < \varepsilon/2$ when $k > N$, $|x_k - x'| < \varepsilon/2$ when $k > N'$. This is simply because we want to prove certain statements strictly by definition

There is an important detail to take note, regarding $\max\{N, N'\}$.

We're taking the larger one of these, so it means that, after this certain point, we in fact have $|x_k - x| < \frac{\varepsilon}{2}$ and $|x_k - x'| < \frac{\varepsilon}{2}$ at the same time.

Therefore by triangle inequality,

$$|x - x'| \leq |x_k - x| + |x_k - x| < \varepsilon$$

The choice of k actually vanished in the final statement; you can think of this as if picking this particular choice of k helps us to establish some kind of property for the original objects

Finally, since we've in fact proven that $|x - x'| < \varepsilon$ holds for any given positive $\varepsilon > 0$, we must have $|x - x'| = 0$ and therefore $x = x'$.

Strictly speaking, for the first part we need to explain why $a < \varepsilon$ for any positive ε implies that $a \leq 0$. This is very easy to prove (by contradiction) so let's not be too redundant The second part simply relies on the fact that |x-y| is the Euclidean metric and so by positive definiteness |x-y|=0 if and only if x=y. $\square$

3. **Boundedness of converging sequences**

   If $\{x_k\}$ converges, then $\{x_k\}$ is bounded.

   Obviously this doesn't work the other way around

   We simply take the limit x and note that the sequence is eventually contained in some ball centered at x, say $B(x, 1)$.

   There are several outlying points prior to this, but since there are only a finite number of these, it doesn't change the fact that the sequence (viewed as a set) is bounded nevertheless.

   This argument is precisely expressed by the construction of r given in the book: let $|x_k - x| < 1$ whenever $k > N$, then $\{x_k\}$ is in $B(x, r)$ where $r = \max\{1, |x_1 - x|, \ldots, |x_N - x|\}$

4. We talk about the relationship between the limit of a sequence and the limit points of a set.

   Generally, limit points are a weaker construction.

   Suppose that $\{x_k\}$ converges to x If we view $\{x_k\}$ as a set, then x will be a limit point of this set

   The converse, however, is not true

Exercise 1: Construct a sequence in R that is bounded and contains a single limit point but is divergent (not convergent)

The thing about convergence of a series is that, unlike for limit points where we only require that there are other points that get arbitrarily close, but moreover we have to ensure that this pattern ensues for each and every term in the sequence

Me:Suppose that $\{x_k\}$ converges to x If we view $\{x_k\}$ as a set, then x will be a limit point of this set" - - - - - - - - - - - - - - - - Sorry I forgot something crucial about this : There is the strange possibility that the sequence $\{x_k\}$ is constant : (or at least eventually constant) : Then in fact x by definition is not a limit point of x_k because you can find a ball around x that only contains the element x itself, since that point is merely what the entire sequence $\{x_k\}$ amounts to : Anyways, we simply can't say that a sequence $\{x_k\}$ converges to x if we're only provided with the fact that x is a limit point of $\{x_k\}$

However, we can say the following: (d) If x is a limit point of E, then there exists a sequence $\{x_n\}$ in $E \setminus x$ such that $\{x_n\}$ converges to x

In fact this is correct in both ways so let's rewrite this as follows: (d) x is a limit point of E, if and only if there exists a sequence $\{x_n\}$ in $E \setminus x$ such that $\{x_n\}$ converges to x

($E \setminus x$ is important here, otherwise we simply pick the constant sequence $x_k = x$)

$\rightarrow$: If x is a limit point, then for all $\varepsilon > 0$, $B_0(x, \varepsilon)$ contains points in E We then construct such a sequence $\{x_k\}$ in $E \setminus x$: pick any $x_k \in E$ so that x_k is contained in $B_0(x, 1/k)$

Then it is easy to show that $\{x_k\}$ is a sequence in $E \setminus x$ which converges to x.

$\leftarrow$: Suppose that there exists a sequence $\{x_n\}$ in $E \setminus x$ such that $\{x_n\}$ converges to x We wish to show that $B_0(x, \varepsilon)$ contains points in E for all $\varepsilon > 0$

Since $\{x_n\}$ converges to x, for all $\varepsilon > 0$ the sequence is eventually contained in $B(x, \varepsilon)$ However because we have the precondition that $\{x_n\}$ has to be in $E \setminus x$, the sequence is in fact eventually contained in $B_0(x, \varepsilon)$.

topological aspects of limits.

§19.2 Subsequences

Properties:

1. $\{x_k\}$ converges to x if and only if every subsequence of $\{x_k\}$ converges to x.

   We only need to prove this in the forwards direction Every subsequence of $\{x_k\}$ can be written in the form $\{x_{k_i}\}$ where $k_1 < k_2 < \ldots$ is a strictly increasing sequence of natural numbers

   Intuitively, if every neighbourhood of x eventually contains all x_k, then since $\{x_{k_i}\}$ is just a subset of $\{x_k\}$ they should all be contained in the neighbourhood eventually as well. For every $\varepsilon > 0$, pick N such that for $k > N$, $|x_k - x| < \varepsilon$. Pick M such that $k_M > N$, then for all $i > M$ we have $|x_(k_i) - x| < \varepsilon$.

2. Subsequential limits of a sequence are precisely the limit points of the sequence (viewed as a set)

   This is just part (d) of the previous section.

   Again, to make this work, we need to assume that nothing funny is going on at subsequential limits If the limits appear due to eventually constant subsequences, then they need not be limit points of the original sequence when viewed as a set

   3.6, 3.7 are precisely the statements we've prepared for last week

3. If $\{x_n\}$ is a sequence in a compact set (bounded closed set), then there exists a convergent subsequence of $\{x_n\}$ This is Weierstrass-Bolzano together with part (b)

   Ah yes, regarding compact sets I need to emphasize this again, but the definition that we are currently using for compact sets is not the actual definition

   I've sent a video before the lesson which talks about the real definition for compact sets Essentially, compact sets satisfies the property akin to the statement in Heine-Borel: Given a topological space (X, τ), a compact set K in X is a set satisfying that, given any open covering $\{U_i\}$ of X, there exists a finite open cover $\{U_1, \ldots, U_n\}$ of X

   This is difficult to process at this stage Since we're currently only working with Euclidean spaces it would be more beneficial if you consider the Heine-Borel Theorem as a property first It would be a lot easier to accept the definition after you're more accustomed to applying the theorem

4. (Rudin 3.7) Subsequential limits form a closed subset

   Actually we've done this two weeks before, it is simply saying that A" is a subset of A'.

   (A" is not always A'; consider the set in R² given by (1/n,1/m)|n,m in N Then (1,0),(0,1) are in A' but not in A"

§19.3 Cauchy Sequences

Definition 19.3.1: Cauchy sequence

A sequence $\{x_k\}$ in $\mathbb{R}^n$ is a **Cauchy sequence**, if the distances between any two terms is sufficiently small after a certain point.
Formally, this is given by: $\forall \varepsilon > 0$, there exists integer N such that

$$\forall k, l > N, |x_k - x_l| < \varepsilon.$$

It is easy to prove that a converging sequence is Cauchy using the triangle inequality. The idea is that, if all the points are becoming arbitrarily close to a given point p, then they are also becoming close to each other. The converse is not always true, however.

Lemma 19.3.1. A sequence $\{x_k\}$ in $\mathbb{R}^n$ is convergent if and only if it is Cauchy.

Proof. **Forward direction:**

Suppose that $\{x_k\}$ converges to x, then there exists N such that for $k > N$, $|x_k - x| < \dfrac{\varepsilon}{2}$
Then for $k, l > N$,

$$|x - k - x_l| \leq |x_k - x| + |x_l - x| < \varepsilon$$

Backward direction:

First, we show that $\{x_k\}$ must be bounded. Pick N such that for all $k, l > N$ we have $|x_k - x_l| < 1$. Centered at x_k, we show that $\{x_k\}$ is bounded; to do this we pick

$$r = \max\{1, |x_k - x_1|, \ldots, |x_k - x_N|\}$$

Then the sequence x_k is in $B(x_k, r)$ and thus is bounded.

Since $\{x_k\}$ is bounded, by the collolary of Bolzano-Weierstrass we know that $\{x_k\}$ contains a subsequence $\{x_{k_i}\}$ that converges to a limit x.

Then for all $\varepsilon > 0$, pick N_1 such that for all $k, l > N$, $|x_k - x_l| < \dfrac{\varepsilon}{2}$. Simultaneously, since $\{x_{k_i}\}$ converges to x, pick M such that for $i > M$, $|x_{k_i} - x| < \dfrac{\varepsilon}{2}$.

Now, since $k_1 < k_2 < \ldots$ is a sequence of strictly increasing natural numbers, we can pick $i > M$ such that $k_i > N$. Then for all $k > N$, by setting $l = k_i$ we obtain

$$|x_k - x_{k_i}| < \frac{\varepsilon}{2}, \quad |x_{k_i} - x| < \frac{\varepsilon}{2}$$

and hence

$$|x_k - x| \leq |x_k - x_{k_i}| + |x_{k_i} - x| < \varepsilon$$

□

§19.4 Upper and Lower Limits

§19.5 Limits of multiple sequences

20 Continuity

§20.1 Limit of Functions

Definition 20.1.1: Limit

Let X and Y be metric spaces; suppose $E \subset X$, $f : E \to Y$ and p is a limit point of E. We write $f(x) \to q$ as $x \to p$, or

$$\lim_{x \to p} f(x) = q$$

if there is a point $q \in Y$ with the following property: $\forall \varepsilon > 0 \exists \delta > 0$ such that

$$d_Y(f(x), q) < \varepsilon$$

for all points $x \in E$ for which

$$0 < d_X(x, p) < \delta.$$

Remark. The symbols d_X and d_Y refer to the distances in X and Y respectively. If X and/or Y are replaced by the real line, the complex plane, or some euclidean space $\mathbb{R}^k$, the distances d_X and d_Y are replaced by absolute values, or by norms of differences.

We can recast this definition in terms of limits of sequences:

$$\lim_{n \to \infty} f(p_n) = q$$

for every sequence $(p_n) \in E$ so that $p_n \neq p$ and $\lim_{n \to \infty} p_n = p$.

By the same proofs as for sequences, limits are unique, and in R they add/multiply/divide as expected.

Definition 20.1.2: Continuity

f is continuous at p if

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

In the case where p is not a limit point of the domain E, we say f is continuous at p. If f is continuous at all points of E, then we say f is continuous on E.

§20.2 Continuous Functions

§20.3 Continuity and Compactness

§20.4 Continuity and Connectedness

§20.5 Discontinuities

§20.6 Monotonic Functions

§20.7 Infinite Limits and Limits at Infinity

21 Differentiation

Definition 21.0.1: Derivative

Let f be defined on $[a, b]$. Then for $x \in [a, b]$, the **derivative** is defined as

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \tag{21.1}$$

If f' is defined at a point/set, we say f is **differentiable** at that point/set.

Series - Taylor series, Fourier series Integration - Riemann integration, Lebesgue integration and measure

22 Riemann-Stieltjes Integral

23 Sequences and Series of Functions

24 Metric Spaces

§24.1 Structures on Euclidean Space

§24.1.1 Vector and Metric Spaces

Definition 24.1.1: Bounded set

E is bounded if there exists M and q so that $\forall p \in E, d(p,q) < M$.

Definition 24.1.2: Limit and isolated point

A point p is a limit point of E if every neighborhood of p contains a point $q \neq pinE$.
If p is not a limit point but is in E, then p is an isolated point.

Definition 24.1.3: Closed set

E is closed if every limit point of E is in E. Intuitively, this means E "contains all its edges".
The closure $\bar{E}$ of E is the union of E and the set of its limit points.

Definition 24.1.4: Interior point

A point p is an interior point of E if there is a neighborhood N of p such that $N \subset E$. Note that interior points must be in E itself, while limit points need not be.

Definition 24.1.5: Open set

E is open if every point of E is an interior point of E. Intuitively, E "doesn't have edges".

Definition 24.1.6: Dense set

E is dense in X if every point of X is a limit point of E or a point of E, or both.

Definition 24.1.7: Interior

The interior E^0 of E is the set of all interior points of E, or equivalently the union of all open sets contained in E.

§24.1.2 Norms and Scalar Product

From the example above, observe that $d_2(x, y)$ only depends on the $x - y$. In particular, by defining $|x|_2 \coloneqq \sqrt{\sum_{i=1}^n x_i{}^2}$, we recover $d(x, y) = |x - y|_2$.

Norms are required to satisfy the following fundamental properties:

1. Positive Definiteness: For any vector x, $\|x\| \geq 0$, and $\|x\| = 0 \iff x = 0$.
2. Absolute Homogeneity: For any vector x and scalar a, $\|ax\| = |a| \cdot \|x\|$
3. Subadditivity (Triangle Inequality): For any two vectors x and y, $\|x + y\| \leq \|x\| + \|y\|$

Definition 24.1.8: Normed vector space

For $x \in V$, a pair $(V, \|x\|)$ is a **normed vector space** if

(i) V is a $\mathbb{K}$-vector space.

(ii) $\|x\| : V \to [0, \infty)$ is a norm on V, which satisfies the properties of a norm above.

The norm of the Euclidean space $\|\cdot\|$ is a real-valued function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$. Given a vector $x = (x_1, \ldots, x_n)^T$ in $\mathbb{R}^n$, the **norm** of x is given by

$$\|x\| = \sqrt{\sum_{i=1}^n x_i{}^2} = \sqrt{x_1{}^2 + \cdots + x_n{}^2}.$$

Definition 24.1.9: Inner product

Let V be a vector space. Then

$$\langle \cdot, \cdot \rangle = V \times V \to \mathbb{K}, (x, y) \mapsto \langle x, y \rangle$$

is called an **inner product** if

(i) $\langle x, y \rangle = \langle y, x \rangle \forall x, y \in V$.

(ii) $\langle x + \alpha y, z \rangle = \langle x, z \rangle + \alpha \langle y, z \rangle$ and $\langle x, y + \alpha z \rangle = \langle x, y \rangle + \bar{\alpha} \langle x, z \rangle \forall x, y, x \in V, \forall \alpha \in \mathbb{K}$.

(iii) $\langle x, x \rangle \geq 0 \forall x \in V$ and $\langle x, x \rangle = 0 \iff x = 0$. The pair $(V, \langle \cdot, \cdot \rangle)$ is called an **inner product space**.

Theorem 24.1.1: Cauchy-Schwarz

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Then $\forall x, y \in V$,

$$|\langle x,y\rangle| \le \sqrt{\langle x,x\rangle}\sqrt{\langle y,y\rangle} := \|x\|\|y\|$$

§24.1.3 Some Concepts in Euclidean Space

Definition 24.1.10: Bounded set

A set E in $\mathbb{R}^n$ is a **bounded set** if there exists $M > 0$ such that $\forall x \in E$, $\|x\| \le M$.

Problem 24.1.1. E, F in $\mathbb{R}^n$ and real k, define

$$kE = \{kx \mid x \in E\}$$

$$E + F = \{x + y \mid x \in E, y \in F\}$$

(a) Show that if E is bounded, then kE is bounded;

(b) Show that if E and F are bounded, then $E + F$ is bounded

Definition 24.1.11: Diameter of set

iven a set $E \subset \mathbb{R}^n$, the **diameter** of E is defined as

$$\operatorname{diam} E = \sup_{x,y\in E} d(x,y).$$

Problem 24.1.2. Find the diameter of the open unit ball in $\mathbb{R}^n$ given by

$$B = \{x \in \mathbb{R}^n \mid \|x\| < 1\}$$

Solution. First note that

$$d(x,y) = \|x - y\| \le \|x\| + \|-y\| = \|x\| + \|y\| < 1 + 1 = 2$$

On the other hand, for any $\varepsilon > 0$, we pick

$$x = (1 - \frac{\varepsilon}{4}, 0, \ldots, 0), y = (-\left(1 - \frac{\varepsilon}{4}\right), 0, \ldots, 0)$$

Then

$$d(x,y) = 2 - \frac{\varepsilon}{2} > 2 - \varepsilon$$

Therefore $\operatorname{diam} B = 2$. □

Problem 24.1.3. Given a set E in $\mathbb{R}^n$, show that E is bounded if and only if $\operatorname{diam} E < +\infty$.

Solution.
Forward direction:

If E is bounded, then there exists $M > 0$ such that $\forall x \in E, \|x\| \leq M$.

Thus $\forall x, y \in E$,

$$d(x,y) = \|x - y\| \leq \|x\| + \|y\| \leq 2M.$$

Thus $\operatorname{diam} E = \sup d(x,y) \leq 2M < +\infty$

Backward direction:

Suppose that $\operatorname{diam} E = r$.

Pick a random point $x \in E$, suppose that $\|x\| = R$

Then for any other $y \in E$,

$$\|y\| = \left\|x + (y - x)\right\| \leq \|x\| + \|y - x\| \leq R + r$$

Thus, by picking $M = R + r$, we obtain $\|y\| \leq M \forall y \in E$, and we're done.

Basically you use x to confine E within a ball, which is then confined within an even bigger ball centered at the origin. ☐

Definition 24.1.12: Distance between sets

Given two sets $E, F \subset \mathbb{R}^n$, the **distance between sets** E and F is defined as

$$d(E,F) = \inf_{x \in E, y \in F} \|x - y\|.$$

Obviously $d(E,F) > 0$ implies that E and F are disjoint, but E and F may still be disjoint even if $d(E,F) = 0$, e.g. the closed intervals $E = (-1,0), F = (0,1)$.

Problem 24.1.4. Suppose that E and F are sets in $\mathbb{R}^n$ where F is finite, then E and F are disjoint if and only if $d(E,F) > 0$.

Topology in Euclidean Space

Before we move on, we need to talk about how we think about topology. The concept first begins with an attempt to say that two points are close to one another.

Of course, we did define the metric earlier But as it turns out, this particular notion can be made extremely abstract

Specifically speaking, we could theoretically define closeness simply with set theory

Imagine that in some random set X, there is a predetermined family of subsets scrA in P(X) (scr: script; cursive)

Now for some element x in X, suppose that we can pick a set in $\mathscr{A}$ containing x.

We may denote this set as U(x) Then, from the perspective of U(x), a point y in X would seem to be close to x if y also lies in U(x)

Ah actually the family of subsets is usually denoted as $\mathscr{N}$

The family $\mathscr{N}$ is called the neighbourhood system

There is also the notion of the neighbourhood system of a particular point,

$$\mathscr{N}(x) = \{U \in \mathscr{N} \mid x \in U\}$$

Now, here's the easiest part to confuse : The word 'system' in the above terminology is actually quite crucial It is not named 'neighbourhood set' or 'neighbourhood family' for a reason : That's because the terminology of 'neighbourhood' is used as follows: We say that a subset of X, let's say N, is a neighbourhood of x, if there is some neighbourhood $U \in \mathscr{N}$ such that $x \in U$ and $U \subset N$. : Ah I'm very sorry but I messed up the terminology

According to wikipedia, the neighbourhood system actually refers to all neighbourhoods What I was talking about earlier should've been called a neighbourhood basis

: The neighbourhood basis is denoted by scrB

Okay let's redo the entire thing

1. Neighbourhood Basis Given a set X, we define a family of subsets in X, denoted by $scrB$, to describe points close to each other; points that belong to the same set U in $scrB$ are considered to be close to each other with respect to U.

2. Neighbourhood Given a point x in X, we use the term **neighbourhood** to describe a particular construction for x; N is said to be a neighbourhood of x, if there exists U in $\mathscr{B}$ containing x such that $U \subset N$.

2'. Neighbourhood System Given a point x in X, the **neighbourhood system** of x, denoted $\mathscr{N}(x)$, is the set of all neighbourhoods of x.

These are the axioms for the neighbourhood systems

1. $\mathscr{N}(x)$ is nonempty, and $\forall U \in \mathscr{N}(x), x \in U$
2. If $U, V \in \mathscr{N}(x)$, then $\exists W \in \mathscr{N}(x)$ s.t. $W \subset U \cap V$
3. If $U \in \mathscr{N}(x)$ and $y \in U$, then $\exists V \in \mathscr{N}(y)$ s.t. $V \subset U$

As for the Euclidean plane, we have a natural way of defining the neighbourhood systems First we pick the neighbourhood basis to be

$$\mathscr{B} = \{B(x, \varepsilon) \mid x \in \mathbb{R}^n, \varepsilon > 0\}$$

Then we say that N is a neighbourhood of x if there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subset N$. $B(x, \varepsilon)$ represents the points close to x, whereas a neighbourhood N of x should contain all the points close to x, at least from the perspective of $B(x, \varepsilon)$

Once we have neighbourhood systems, we can then define the two most important kinds of sets in topology, open and closed sets.

Part VI

Topology

25 Metric Spaces

§25.1 Definition

Definition 25.1.1: Metric space

A **metric space** is a pair (X, d) consisting of a set of **points** X and a **metric** $d : X \times X \to \mathbb{R}_{\geq 0}$. For all $x, y, z \in X$, the distance function d satisfies the following conditions:

M1 **Positive definitive**: $d(x, y) \geq 0$ with equality if and only if $x = y$.

M2 **Symmetric**: $d(x, y) = d(y, x)$

M3 **Triangle inequality**: $d(x, z) \leq d(x, y) + d(y, z)$

Notation. We usually abbreviate (X, d) as just X.

Example 25.1.1: Metric spaces of $\mathbb{R}$

(a) The real line $\mathbb{R}$ is a metric space under the metric $d(x, y) = |x - y|$.

(b) The interval $[0, 1]$ is also a metric space with the same distance function.

(c) In fact, any subset $S \subseteq \mathbb{R}$ can be made into a metric space in this way.

Example 25.1.2: Metric spaces of $\mathbb{R}^2$

(a) We can make $\mathbb{R}^2$ into a metric space by imposing the Euclidean distance function

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

(b) Just like with the first example, any subset $S \subseteq \mathbb{R}^2$ can be made into a metric space, such as the unit disk, unit circle, and the unit square $[0, 1]^2$.

Example 25.1.3: Metric spaces of $\mathbb{R}^n$

To generalise the above examples, for positive integer n,

(a) Let $\mathbb{R}^n$ be the metric space whose points are points in n-dimensional Euclidean space, and whose metric is the Euclidean metric

$$d((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = \sqrt{(a_1 - b_1)^2 + \cdots + (a_n - b_n)^2}$$

This is the n-dimensional **Euclidean space**.

(b) The open unit ball B^n is the subset of $\mathbb{R}^n$ consisting of the points $(x_1, \ldots, x_n)$ such that $x_1^2 + \cdots + x_n^2 < 1$.

Notation. We will refer to $\mathbb{R}^n$ with the Euclidean metric by just $\mathbb{R}^n$; if we wish to take the metric space for a subset $S \subseteq \mathbb{R}^n$ with the inherited metric, we will just write S.

§25.2 Convergence

Since we can talk about the distance between two points, we can talk about what it means for a sequence of points to converge. This is the same as the typical epsilon–delta definition, with absolute values replaced by the distance function.

Definition 25.2.1: Convergence

Let $(x_n)_{n\geq 1}$ be a sequence of points in a metric space X. We say that x_n **converges** to x if the following condition holds: for all $\varepsilon > 0$, there exists an integer N (depending on ε) such that $d(x_n, x) < \varepsilon$ for each $n \geq N$. This is written as

$$\lim_{n\to\infty} x_n = x.$$

We say that a sequence converges in X if it converges to a point in X.

§25.3 Continuity

From calculus, the ε–δ definition of a continuous function is

> A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point $p \in \mathbb{R}$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|x - p| < \delta \implies |f(x) - f(p)| < \varepsilon$.

For the definition in metric space, all we have do is replace the absolute values with the more general distance functions: this gives us a definition of continuity for any function $M \to N$.

Definition 25.3.1: Continuity

For metric spaces $X = (X, d_X)$ and $Y = (Y, d_Y)$, a function $f : X \to Y$ is **continuous** at a point $p \in X$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$d_X(x, p) < \delta \implies d_Y(f(x), f(p)) < \varepsilon.$$

Moreover, the function f is continuous if it is continuous at every point $p \in X$.

Here is an equivalent condition for sequences.

Theorem 25.3.1: Sequential continuity

A function $f : X \to Y$ of metric spaces is **continuous** at a point $p \in X$ if and only if the following property holds: if $x_1, x_2, \ldots$ is a sequence in X converging to p, then the sequence $f(x_1), f(x_2), \ldots$ in Y converges to $f(p)$.

Proposition 25.3.1 (Composition of continuous functions is continuous)**.** Let $f : X \to Y$ and $g : Y \to Z$ be continuous maps of metric spaces. Then their composition $g \circ f$ is continuous.

§25.4 Homeomorphisms

§25.5 Extended example/definition: product metric

§25.6 Open sets

Continuity is really about what happens “locally": how a function behaves “close to a certain point p". One way to capture this notion of “closeness" is to use metrics as we have done above. In this way we can define an r-neighborhood of a point.

Definition 25.6.1: Neighbourhood

For metric space X and point $p \in X$, an **r-neighborhood** of p, denoted by $N_r(p)$, is the set of all q with $d(p,q) < r$ for some radius $r > 0$.

$$N_r(p) = \{q \in X \mid d(p,q) < r\}$$

Remark. Others define a neighborhood as any set that contains one of these neighborhoods, which are instead called "the open ball of radius r about p".

Such an open ball is sometimes referred to as the open neighborhood of p of radius r.

Open balls are instances of open sets.

Definition 25.6.2: Open set

A subset $U \subset X$ is open if, for every point $x \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$.

The idea is that, in a open set, there exists a "safety margin" around every point. Given a point p, one can *move around in the set a certain distance and remain* in the sense.

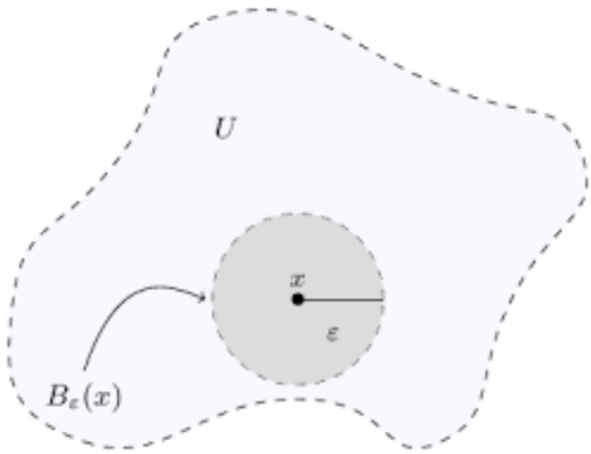


Figure 25.1: Open set

Some basic properties of open sets are

1. Open balls are open. This is a basic consequence of the triangle inequality.
2. $\varnothing$ is open (vacuously).
3. X is open (as all open balls are contained in X).
4. The arbitrary union of open sets is open (even infinitely many). This follows easily from the definition.
5. The intersection of finitely many open sets is open.

§25.7 Closed sets

§25.8 Terminology

Let X be a metric space. All points and sets mentioned below are understood to be elements and subsets of X respectively.

- A **ball** in $\mathbb{R}^n$ is determined by its center $x \in \mathbb{R}^n$ and its radius $r > 0$, and is denoted by $B(x, r)$.
$$B(x,r) = \{p \in X \mid d(x,p) < r\}$$
A **punctured ball** in $\mathbb{R}^n$ is a ball excluding its center, and is denoted by $B_0(x, r)$.
- A containing x is a **neighborhood** of x if $B(x, \varepsilon) \subset A$ for some $\varepsilon > 0$.
- The **complement** of A, denoted by A^c, is the set of all points $x \in X$ such that $x \notin A$.
- A point $x \in A$ is an **interior point** of A if A is a neighbourhood of x.
The **interior** of A, denoted by A°, is the set of all interior points in A.
A is **open** if every point of A is an interior point of A, i.e. $A^\circ = A$.
- A point $x \in A$ is a **limit point** of A if every neighborhood of x contains a point $y \neq x$ such that $y \in A$.
This means $B_0(x, \varepsilon) \cap A \neq \varnothing$ for all $\varepsilon > 0$.
The **induced set** of A, denoted by A', is the set of all limit points of A.
The **closure** of A, denoted by $\bar{A}$, is the union set $A \cup A'$. A is **closed** if all limit points of A are contained in A, i.e. $\bar{A} = A$.
- A point $x \in A$ is an **isolated point** of A if x is not a limit point of A.
- The **boundary** of a set A, denoted by ∂A, is the set difference $\bar{A} \setminus A^\circ$.
A point x is a **boundary point** of A if $x \in \partial A$.
- A point x is an **exterior point** of A if it is an interior point of A^c.
- A is **perfect** if A is closed and if every point of A is a limit point of A.
- A is **bounded** if there is a real number M and a point $p \in X$ such that $d(x, p) < M$ for all $x \in A$.
- A is **compact** if it is a bounded closed set.
- A is **dense** in X if every point of X is a limit point of A, or a point of A (or both).
- A subset $B \subset A$ is a **dense subset** of A if $\bar{B} = A$.
- A is **nowhere dense** if its closure has no interior, i.e. $(\bar{A})^\circ = \varnothing$.

Remark. It is important to take note that the terminology of neighbourhood in Rudin is actually just a ball here.

This is actually standard in calculus, but I am using the terminology as you would see them in general point-set topology.

Theorem 25.8.1

Every neighborhood is an open set.

Proof. Consider a neighborhood $E = N_r(p)$, and let q be any point of E. Then there is a positive real number h such that

$$d(p,q) = r - h.$$

For all points s such that $d(q,s) < h$, we have then

$$d(p,s) \leq d(p,q) + d(q,s) < r - h + h = r,$$

so $s \in E$. Hence q is an interior point of E. □

These are certain properties regarding open and closed sets in $\mathbb{R}^n$:

P1 A is open if and only if A^c is closed

Proof. **Forward direction**: Let A be open, we consider the punctured balls of $x \notin A$ (if $x \notin A$, we consider the punctured balls centered at x).

Our goal is to show that $B_0(x,r)$ always intersects with A^c

So suppose otherwise that $B_0(x,\varepsilon)$ is a subset of A for some $\varepsilon > 0$

Ah no sorry, we consider x not in A^c

The thing is we want to show that A^c is closed, i.e. all limit points of A^c are in A^c

So suppose otherwise that x is a limit point of A^c that is not in A^c

x is a limit point of A^c, hence for all $\varepsilon > 0$, $B_0(x,\varepsilon)$ always intersects with A^c

This is equivalent to saying that $B_0(x,\varepsilon)$ is never a subset of $(A^c)^c = A$

However, x is not in $x \notin A^c$, so $x \in A$.

But if A is open, then there exists $\varepsilon > 0$ such that $B(x,\varepsilon)$ is a subset of A, a contradiction

Backward direction: Let A^c be closed. Suppose otherwise that A is not open, i.e. there is a point $x \in A$ such that $B(x,\varepsilon)$ is never a subset of A; that is to say, $B(x,\varepsilon)$ always intersects with A^c

Since $x \in A$, then $B(x,\varepsilon) \cap A^c = B_0(x,\varepsilon) \cap A^c$

But this means that $B_0(x,\varepsilon) \cap A^c$ is never empty, hence x is a limit point of A^c.

However, $x \in A$, contradictory to A^c being closed and thus should contain all of its limit points □

P2 An arbitrary union of open sets is open; a finite intersection of open sets is open.

Proof. Let A be an arbitrary union of open sets $\{U_i\}_{i \in I}$.

Then for any $x \in A$, suppose that $x \in U_i$, then since U_i is open we can pick $B(x, \varepsilon)$ subset of U_i subset of A

On the other hand, let U and V be open sets and let $x \in U \cap V$. Since U and V are open, we can pick ε_1 and ε_2 such that $B(x, \varepsilon_1)$ is in U whereas $B(x, \varepsilon_2)$ is in V. Then we simply pick $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ so that $B(x, \varepsilon)$ is in $U \cap V$. □

P3 An arbitrary intersection of closed sets is closed; a finite union of closed sets is closed.

Proof. This follows from de Morgan's Law on P1 and P2. □

Problem 25.8.1. Compare the following sets:

1. $(A \cup B)^\circ$, $A^\circ \cup B^\circ$
2. $(A \cap B)^\circ$, $A^\circ \cap B^\circ$
3. $\overline{A \cup B}$, $\bar{A} \cup \bar{B}$
4. $\overline{A \cap B}$, $\bar{A} \cap \bar{B}$

Compare their sizes, i.e. determine if they are equal, or if one set may be a subset of the other.

Proof.

1. $(A \cup B)^\circ$ may be bigger

   In $\mathbb{R}$ we consider the intervals $A = (-1, 0]$ and $B = [0, 1)$, then

   $$A^\circ \cup B^\circ = (-1, 0) \cup (0, 1), \quad (A \cup B)^\circ = (-1, 1)$$

   For $x \in A^\circ \cup B^\circ$, we have either $x \in A^\circ$ or $x \in B^\circ$, so there is some ball centered at x that is contained in either A or B and thus must be contained in $A \cup B$ as well.

2. Equal

   If $x \in (A \cap B)^\circ$, then there exists a ball U centered at x such that U is in both A and B, so x is in both A° and B°.

   On the other hand, $A^\circ \cap B^\circ$ is a subset of $A \cap B$; taking the interior of both sides, then since the intersection between two open sets is open we find that $A^\circ \cap B^\circ$ is a subset of $(A \cap B)^\circ$.

3. Equal

4. $\bar{A} \cap \bar{B}$ may be bigger

□

Problem 25.8.2. Prove that the set of exterior points, $(A^c)^\circ$ is the same as $(\bar{A})^c$.

Proof.

$$\begin{aligned} & x \in (A^c)^\circ \\ & \iff \exists \varepsilon > 0 \text{ such that } B(x,\varepsilon) \subset A^c \\ & \iff B(x,\varepsilon) \cap A = \varnothing \\ & \iff x \notin A \text{ and } B_0(x,\varepsilon) \cap A = \varnothing \\ & \iff x \notin A \cup A' = \bar{A} \\ & \iff x \in (\bar{A}^c) \end{aligned}$$

□

Problem 25.8.3. Regarding alternative descriptions:

1. A is a neighbourhood of x if and only if there exists an open set U such that x is in U, U is subset of A (trivial except you'll actually need to prove that balls are open sets).
2. If x is a limit point of A, then in fact for any $\varepsilon > 0$, $B(x,\varepsilon)$ contains infinitely many elements of A (you don't need to mention the punctured ball here because of obvious reasons; converse is trivial but a good and intuitive description).
3. x is a boundary point of A if and only if for all $\varepsilon > 0$, $B(x,\varepsilon)$ intersects with both A and A^c.

Proof.

1. We show that $B(x,\varepsilon)$ is open:
   $\forall y \in B(x,\varepsilon)$,
   $$|y - x| < \varepsilon$$
   $\forall z \in B(y, \varepsilon - |y - x|)$,
   $$|z - x| \le |z - y| + |y - x| < \varepsilon - |y - x| + |y - x| = \varepsilon$$
   $\therefore B(y, \varepsilon - |y - x|) \subset B(x,\varepsilon)$
2. We construct a sequence $\{x_n\}$ recursively as follows:
   - Pick $x_1 \in B_0(x,\varepsilon) \cap A$
   - Pick $x_{n+1} \in B_0(x, |x_n - x|) \cap A$

It is easy to see that the balls above are getting smaller so all x_n are both mutually distinct and all contained in $B(x, \varepsilon)$.

3. x is a boundary point if and only if $x \in \bar{A} \setminus A^\circ$

   Forward direction:

   We consider two cases

   - $x \in A$, then all $B(x, \varepsilon)$ intersects with A at x, but since x is not in A° they must always intersect with A^c as well.
   - $x \notin A$, then all $B(x, \varepsilon)$ intersect with A^c at x, but since $x \in \bar{A}$, x is a limit point of A and thus $B(x, \varepsilon)$ always intersects with A.

   Backward direction:

   We consider two cases

   - $x \in A$, then since $B(x, \varepsilon)$ always intersects with A^c, x cannot be in A°.
   - $x \notin A$, then since $B(x, \varepsilon)$ always intersects with A, x must be in $\bar{A}$.

   In fact we can describe the closure without referring to punctured balls and induced sets: $x \in \bar{A}$ if and only if $B(x, \varepsilon)$ always intersects with A

   Also as a side note, $A \circ \cup dA \cup (A^c) \circ = \mathbb{R}^n$

□

Problem 25.8.4. Regarding closures (The following properties are relatively nontrivial compared to its 'open-set' counterparts):

(a) A' is closed.

(b) $\bar{A}$ is closed, i.e. bar(barA)=barA

Proof.

(a) In order to show that A' is closed, we need to show that if x is a limit point of A', then $x \in A'$, i.e. x is a limit point of A.

   So we need to show that limit points of A' are always limit points of A: Let x be a limit point of A', then for all $\varepsilon > 0$, $B_0(x, \varepsilon/2)$ intersects with A' and we may pick $y \in B_0(x, \varepsilon/2) \cap A'$

   Now here's the tricky part Since $y \in A'$, y is a limit point of A, hence $B_0(y, |y - x|)$ intersects with A and thus we may pick $z \in B_0(y, |y - x|) \cap A$.

   We show that $z \in B_0(x, \varepsilon)$:

   $$|z - x| \leq |z - y| + |y - x| < 2|y - x| < \varepsilon,$$

   hence $z \in B(x, \varepsilon)$.

   $$|z - y| < |x - y|,$$

   hence $z \neq x$

   $\therefore z \in B_0(x, \varepsilon)$

(b) As for 5-2, it is just 5-1 and 2-3

□

For homework, you'll work out some properties regarding dense sets

1. A is a dense set in X if and only if A intersects with all open sets in X 2. If A is dense in X and B is dense in A, then B is dense in X 3. If A and B are dense in X where A is open, then A∩$BisdenseinX$

§25.9 Compactness

Definition 25.9.1: Open cover

By an **open cover** of a set A in a metric space X we mean a collection $\{G_\alpha\}$ of open subsets of X such that $A \subset \bigcup_\alpha G_\alpha$.

Definition 25.9.2: Compact set

A subset K of a topological (or metric) space is compact if every open cover of K has a *finite* subcover.

An open cover of A is a collection of open sets that collectively cover A.

A subcover is a subcollection of these open sets that still collectively cover A.

This means that any infinite collection of open sets that together cover a compact set always "overcovers" it.

The simplest kind of compact set is just a finite set: a collection of finitely many points.

§25.10 Some theorems

Theorem 25.10.1: Cantor's Intersection Theorem

Given a decreasing sequence of compact sets $A_1 \supset A_2 \supset \cdots$, there exists a point $x \in \mathbb{R}^n$ such that x belongs to all A_i. In other words, $\bigcap_{i=1}^{\infty} A_i \neq \varnothing$. Moreover, if for all $i \in \mathbb{N}$ we have $\operatorname{diam} A_{i+1} \leq c \cdot \operatorname{diam} A_k$ for some constant $c < 1$, then such a point must be unique, i.e. $\bigcap_{i=1}^{\infty} A_k = \{x\}$ for some $x \in \mathbb{R}^n$.

Theorem 25.10.2: Heine-Borel Theorem

A set $A \subset \mathbb{R}^n$ is compact if and only if every open covering has a finite subcover, i.e. for any family of open sets $\mathscr{U} = \{U_i\}_{i \in I}$ satisfying $A \subset \bigcup_{i \in I} U_i$, there exists $\{U_1, \ldots, U_n\} \subset \mathscr{U}$ such that $A \subset \bigcup_{i=1}^{n} U_i$.

Theorem 25.10.3: Bolzano-Weierstrass Theorem

Infinite bounded sets in $\mathbb{R}^n$ must contain limit points.

Proof. We will follow a very specific sequence of steps to prove them:

1. Cantor Intersection for $n = 1$
2. Bolzano-Weierstrass for $n = 1$
3. Bolzano-Weierstrass for general n
4. Cantor Intersection for general n
5. Heine-Borel for general n

Proof:

1. Suppose that there is a decreasing sequence of compact sets $A_1, A_2, \ldots$ in the real numbers

   Since A_k are bounded, we may let $a_k = \inf A_k$ Also since A_k are closed, $a_k \in A_k$

   Note that since A_k is a decreasing sequence of sets we have $a_1 \leq a_2 \leq \ldots$

   Also, whenever we have $n > k$, we have $a_n \in A_n$, but $A_n \subset A_k$ and thus $a_n \in A_k$.

   Let $b_1 = \sup A_1$, then $a_k \in A_1$ and thus $a_k \leq b_1$ for all k.

   This tells us that the sequence $\{a_k\}$ is bounded above, and thus we may let $a = \sup a_k$.

   Our goal is to show that the number a appears in all A_k, thus showing that the entire intersection $\bigcap A_k$ contains a and thus must be non-empty.

   Now we split this in two cases, which asks whether a is simply made from isolated points, or if it is actually some nontrivial point obtained from the boundaries of A_k

Case 1: $a_k = a$ for some k In this case we see that $a_k \le a_n \le a$ for all $n > k$ and thus $a_n = a$ in this case, therefore a is an element in A_n for all n

In this case you can imagine that there is a possibility where a is an isolated minimum point of A_n which stays there forever in the decreasing sequence of sets

Case 2: $a_k < a$ for all k; in this case we see that a is the limit point of the increasing sequence $\{a_k\}$

Exercise 1: Show that a is a limit point of each A_k

Note that a_n is in A_k for each $n > k$, and since $a = \sup\{a_k\}$ where a_k is increasing, we can actually show that a is a limit point of $\{a_n \mid n \le k\}$: For every $\varepsilon > 0$, we pick n_0 such that $0 < a - a_{n_0} < \varepsilon$ Pick $n\prime > \max\{k, n_0\}$, then $\mathrm{a}'_n a_{n_0} andso 0 < a - a_n\prime \le a_{n_0} < \varepsilon$ This shows that there exists a'_n in $B_0(a,\varepsilon) \cap \{a_n \mid n > k\}$ for all ε, and so a is a limit point of $\mathrm{a}_n | n > k : Nowsince \{\mathrm{a}_n | n \ge k\}$ is a subset of A_k we also see that a is a limit point of A_k Finally, since A_k is closed, we conclude that a is in A_k for all k, and we are done

: Wait hold on, I forgot about the second part : Now we consider a decreasing sequence of compact sets $\mathrm{A}_1, A_2, ... suchthatdiam A_{k+1} \le cdiam \mathrm{A}_k$ for c<1 : Suppose otherwise that there exists x, y in $\mathrm{A}_k : Y oucanimaginethatthiswillformafixeddistancebetweentwopoints, andthusthe$ $diam A_k |x - y| > 0 forallk : Butthiscannotbetruebecause \mathrm{diam}\, A_{k+1} \le c\, \mathrm{diam}\, A_k$ and so the diameter is controlled by a decreasing geometric sequence: diam $\mathrm{A}_{k+1} \le c^k diam A_1$

So we can simply pick a natural number k such that k>$\log_c(|x - y|/diam A_1)$

We consider an infinite bounded set A in the real numbers. Since A is bounded, we can pick a closed interval $[a_1, b_1]$ containing A.

We then perform a series of binary cuts: Consider the two halves of $[a_1, b_1]$. We know that at least one of these two must contain infinitely many elements in A, otherwise A cannot be infinite. We pick this half of the interval and denote it by $[a_2, b_2]$. We continue this to pick a decreasing sequence of closed intervals $[a_n, b_n]$.

Now $\mathrm{diam}[a_{n+1}, b_{n+1}] = \frac{1}{2}\,\mathrm{diam}[a_n, b_n]$, so by the Cantor Intersection Theorem, there exists a unique real number c in the intersection $\bigcap[a_n, b_n]$.

We show that this c is in fact a limit point of A.

For any $\varepsilon > 0$, we need to show that $B_0(c,\varepsilon) \cap A \ne \varnothing$, i.e. we need to find an element $x \ne c$ in A that is less than ε apart from c.

We then realize that we can simply exploit the decreasing sequence $[a_n, b_n]$ Since $\mathrm{diam}[a_n, b_n]$ is controlled by a decreasing sequence:

$$\mathrm{diam}[a_{n+1}, b_{n+1}] \le 1/2^n\, \mathrm{diam}[a_1, b_1]$$

We take a sufficiently large n so that $b_n - a_n < \varepsilon$ Since c is in $[a_n, b_n]$, for all x in $[a_n, b_n]$ we have $|x - c| \le b_n - a_n < \varepsilon$ and therefore $[a_n, b_n]$ is within $B_0(c,\varepsilon)$.

Here's the funny part: $[a_n, b_n]$ contains infinitely many elements of A, so it must contain at least one element in A that is not c : Therefore this element xc is in B(c,$\varepsilon) Imadeatypo, [a_n, b_n] issupposed$

Now we have an infinte bounded set A in $\mathbb{R}^n$

The idea here is to consecutively come up with better and better sequences of points in A. We denote x_i to be the i-th coordinate in $\mathbb{R}^n$.

Our first wish is to pick some elements in A so that they sort of converge at x_1.

Because such considerations of 'restricting to a single coordinate' is important here, we define the projection map to the i-th coordinate by

$$f_i(x_1, \dots, x_n) = x_i$$

So, we look at $f_i(A)$ and try to apply BW for the case where $n = 1$.

However, the problem is that $f_i(A)$ need not be infinite. For example, the set $\{(0,0),(0,1),(0,2),\dots\}$ projected onto the first coordinate is simply $\{0\}$.

This forces us to consider two cases : Exercise 2: Show that f$_i(A) isbounded This issimple:$
$1. f_1(A) isinfinite, then we can apply BW (n = 1) to find a real number c_1 which is a limit point in f_1(A)$

: Here we can construct a sequence of points x$^{(1),1}, x^{(1),2}, ... so that their first coordinates satisfy |x_1^{(1),n} -$
$c_1| < 1/n for all natural number n (I know this notation is cumbersome but the problem is that we need multiple$
$2. f_1(A) is finite, then by the Pigeonhole Principle there exists a real number c_1 such that its preimage f_1^{-1}(c_1)$

In this case we can randomly pick a sequence x$^{(1),1}, x^{(1),2}, ... in A so that their first coordinate is equal to c_1$

I forgot to mention something that is implied, but we actually do have the need to emphasize that the sequence x$^{(1),1}, x^{(1),2}, ... can be chosen to contain mutually distinct entries$

Now that we have a sequence that behaves nice on the first coordinate, we may then move on to the second coordinate

Let A$_1 = x^{(1),1}, x^{(1),2}, ... We again consider f_2(A_1) in two cases, infinite or finite$

In any case, we are able to find a subsequence x$^{(2),1}, x^{(2),2}, ..., where x^{(2),k} = x^{(1),n_k} for some strictly increas$

So that, for the limit point/point with infinite preimage c_2, this sequence satisfies

$$|f_2(x^{(2),n}) - c_2| < \frac{1}{n}$$

Note that the property we have for the second case (we in fact have $f_2(x^{(2),n}) = c_2$) is just a better version of this.

Now, take note that picking this subsequence does no harm whatsoever towards the first coordinate (if anything it would turn out to be better) since

$$|f_1(x^{(2),k}) - c_1| = |f_1(x^{(1),n_k} - c_1| < \frac{1}{n_k} \le \frac{1}{k}$$

($n_1 < \dots < n_k$ is a strictly increasing sequence of natural numbers so $n_k \ge k$)

This continues on until we obtain a sequence of points $\{x^{(n),1}, x^{(n),2}, \dots\}$ in A so that

$$|f_i(x^{(n),k} - c_i| < \frac{1}{k} \quad \forall i, k$$

As we can see, the point $c = (c_1, \dots, c_n)$ is in fact a limit point of A as we can always choose a big enough k so that $x^{(n),k}$ is in $B(c, \varepsilon) \cap A$.

Since $\{x^{(n),k}\}$ was always chosen to be a sequence of distinct entries, there is no danger for this sequence to always be c, and so c must be a limit point of A.

We may now return to the general case of Cantor.

Suppose that there is a sequence of decreasing compact sets $A_1, A_2, \ldots$ in $\mathbb{R}^n$. Note that every point is contained in A_1, so boundedness will never be an issue here.

Since A_k are all nonempty, we can simply pick any element a_k from A_k.

For the uncannily specific case that there are only finitely many $\{a_k\}$ chosen, we simply note that, again by Pigeonhole Principle, one of the a_k appears infinitely often; thus for each A_n we simply pick $n_k > n$ so that A_{n_k} contains a_k, then a_k is in A_{n_k} which is a subset of A_n.

Otherwise, we can then note that $\{a_k\}$ is an infinite bounded set of points, so there must exist a limit point a of $\{a_k\}$.

We can now see that a is always an element of A_k: Using the same technique as Exercise 1, we see that a is a limit point of $\{a_n \mid n > k\}$ and so is a limit point of A_k, therefore a is in A_k as A_k is closed.

: This proves the first part of the statement The second part is completely identical to the second part of the n=1 case so we don't need to waste our time there either :

We now consider a compact set A with some open covering $\mathscr{U}$.

This theorem is proved by contradiction: Suppose otherwise that set A cannot be covered by any finite collection of open sets in $\mathscr{U}$

Since A is compact, we may enclose it in a closed cube Q_1 (whose edges are parallel to the axes)

Now, for each step, we partition Q into 2^n cubes by cutting it in half from each direction.

Then, starting from Q_1, there must exist one of these smaller cubes, denoted by Q_2, such that $A \cap Q_2$ cannot be covered by a finite collection of open sets in $\mathscr{U}$. Otherwise, if each $A \cap Q$ has a finite cover, then we simply collect all of these open sets together to form a finite cover of A, which violates our assumption.

We continue on to partition Q_n and pick Q_{n+1} so that A_{n+1} has no finite cover (denote $A_n = A \cap Q_n$).

Note that A and Q_n are both compact, so A_n is compact Also we see that there is a decreasing sequence $A_1, A_2, \ldots$ (we can't exactly obtain a relation between $\operatorname{diam} A_n$ and $\operatorname{diam} A_{n+1}$ here)

By Cantor Intersection Theorem we can always find a point x in A located in the intersection $\bigcap A_k$.

Now, since $\mathscr{U}$ is an open covering of A, there exists an open set U in $\mathscr{U}$ such that $x \in U$.

The final key step is to exploit the sequence of decreasing cubes Q_n. So even though there isn't a clear cut way to control the sizes of $\operatorname{diam} A_n$, we do in fact have the property that $\operatorname{diam} Q_{n+1} = \frac{1}{2^n} \operatorname{diam} Q_1$.

Therefore, by picking a sufficiently large n, we can obtain Q_n that is contained in U.

But this is a contradiction. This is because we've specifically chosen the sequence A_n to be sets that do not possess any finite cover $\{U_1, ..., U_n\}$ in $\mathscr{U}$. But here A_n simply would have a one-element cover $\{U\}$.

This completes our proof.

□

26 Euclidean n-space Topology

§26.1 n-dimensional Euclidean space

$\mathbb{R}^n$, as a set, is defined as the set of vertical vectors with n coordinates in the real numbers.

Algebraically, $\mathbb{R}^n$ is an n-dimensional vector space over $\mathbb{R}$. Vectors in $\mathbb{R}^n$ are expressed as vertical vectors

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

We usually express the above vector compactly as follows:

$$x = (x_1, \ldots, x_n)^T$$

§26.1.1 Properties

Algebraically, $\mathbb{R}^n$ is a vector space over $\mathbb{R}$. This means that $\mathbb{R}^n$ has the following extra properties.

These properties make up the algebraic structure of $\mathbb{R}^n$, which may then be further expanded in linear algebra However, for now we will continue on with the analytical/-topological aspects of Euclidean space.

Addition

For any two vectors x, y, we may perform addition

$$x + y = (x_1 + y_1, \ldots, x_n + y_n)^T$$

Properties of addition:

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

3. The zero vector $0 = (0, \ldots, 0)$ satisfies $x + 0 = 0 + x = x$
4. For any vector x, its negative $-x$ satisfies $x + (-x) = (-x) + x = 0$

Scalar multiplication

For any vector x and real number (scalar) k, we may perform scalar multiplication

$$kx = (kx_1, \ldots, kx_n)^T$$

Properties of scalar multiplication:

1. $0 \cdot x = 0$, $1 \cdot x = x$
2. $(kl)x = k(lx) = l(kx)$
3. $k(x + y) = kx + ky$
4. $(k + l)x = kx + lx$

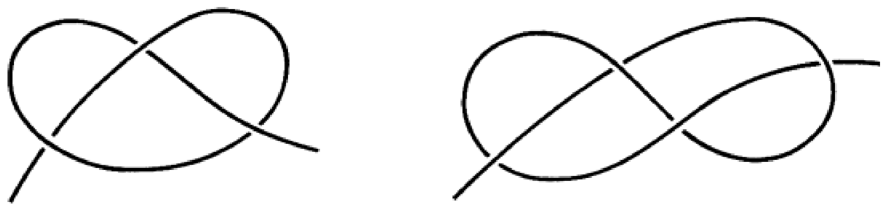
27 Knot Theory

Readings:

- Knot Theory by Stanford University
- The Knot Book by Colin C. Adams

§27.1 Knot and Knot Types

Almost everyone is familiar with at least the simplest of the common knots: the overhand knot and the figure-eight knot.



A little experimenting with a piece of rope will convince anyone that these two knots are different: one cannot be transformed into the other without passing a loop over one of the ends, i.e., without "tying" or "untying". Nevertheless, the failure to change the figure-eight into the overhand by hours of patient twisting is no proof that it can't be done. The problem that we shall consider is the problem of showing mathematically that these knots (and many others) are distinct from one another.

Before this, we must have a mathematical definition of what a knot is, and another mathematical definition of when two knots are to be considered the same.

Definition 27.1.1: Knot

K is a knot if there exists a homeomorphism of the unit circle C into 3-dimensional space $\mathbb{R}^3$ whose image is K.
The circle C is the set of points (x, y) in the plane $\mathbb{R}^2$ which satisfy the equation $x^2 + y^2 = 1$.

Part VII

Complex Analysis

28 Complex Numbers

Readings:

- Complex Analysis by Serge Lang

Part VIII

Discrete Mathematics

29 Graph Theory

30 Game Theory

Recommended readings: “An Introduction to Game Theory" by Osborne

Game Theory is the study of strategically interdependent behaviour.

§30.1 Strict Dominance

§30.1.1 Prisoner’s Dilemma

To start off, we will take a look at the **Prisoner’s Dilemma**, which goes as follows:

Two thieves plan to rob a store, but the police arrest them for trespassing. The police suspect that they planned to break in but lack the evidence to support such an accusation. They require a confession to charge the suspects. The police offer them the following deal:

- If no one confesses, both are charged a *one month* jail sentence each for trespassing.
- If a rat confesses and the other does not, the rat is not charged but the other is charged a *twelve month* jail sentence for robbery.
- If both confess, both are charged an *eight month* jail sentence each.

If both criminals are self-interested and only care about minimising their jail time, should they take the interrogator’s deal?

We condense the above information into a **payoff matrix** as shown below, where we have two players, A and B. The horizontal rows represent A’s choices, while the vertical columns represent B’s choices, and each cell contains a combination of their payoffs.

	quiet	confess
quiet	−1, −1	−12, 0
confess	0, −12	−8, −8

§30.1.2 Split or Steal

The game goes as follows:

> Each of two players, Sarah and Steve, has to pick one of two balls: inside one ball appears the word '**split**' and inside the other the word '**steal**' (each player is first asked to secretly check which of the two balls in front of him/her is the split ball and which is the steal ball). They make their decisions simultaneously.

The possible outcomes are shown in the figure below, where each row is labelled with a possible choice for Sarah and each column with a possible choice for Steven. Each cell in the table thus corresponds to a possible pair of choices and the resulting outcome is written inside the cell.

		Steven	
		Split	Steal
Sarah	Split	Sarah gets \$50,000 Steven gets \$50,000	Sarah gets nothing Steven gets \$100,000
	Steal	Sarah gets \$100,000 Steven gets nothing	Sarah gets nothing Steven gets nothing

§30.2 Nash Equilibrium

Nash Equilibrium is a set of optimal strategies that work against *all* counter-steategies. This means that if any given player were told the strategies of all their opponents, they still would choose to retain their original strategy.

§30.2.1 Matrix games

§30.3 Fair Division

§30.3.1 Rental harmony problem

Sperner's lemma

https://www.cs.cmu.edu/ arielpro/15896/docs/paper19b.pdf

Part IX

Differential Geometry

Readings:

- Introduction to Differentiable Manifolds and Riemannian Geometry
- Differential Geometry of Curves and Surfaces