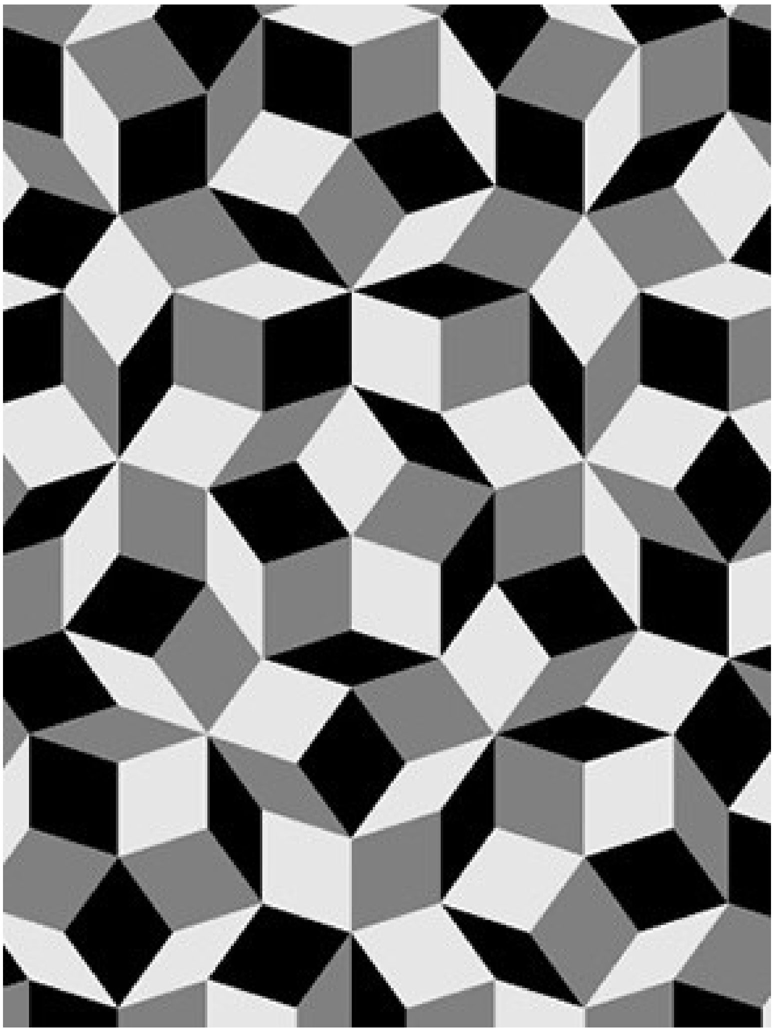
Undergraduate Mathematics



Ryan Joo Rui An

Undergraduate Mathematics

Ryan Joo Rui An

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

This is (still!) an incomplete draft. Please send corrections and comments to ryanjooruian18@gmail.com, or pull-request at https://github.com/Ryanjoo18/undergrad-math.

Typeset using LaTeX.

Last updated January 2, 2025.

Preface

Part I covers **abstract algebra**, which follows [DF04; Art11]. Chapter 3 introduces groups; Chapter 4 introduces rings.

Part II covers **linear algebra**, which follows [Axl24]. Chapter 5 introduces vector spaces, subspaces, span, linear independence, bases and dimension. Chapter 6 concerns linear maps and related concepts.

Part III covers **real analysis**, which follows [Rud76; Apo57; BS11]. [Alc14] is also a good read to get some intuition into some abstract notions. Chapter 8 introduces the real and complex number systems; Chapter 9 covers basic topology required for subsequent chapters; Chapter 10 and Chapter 14 cover numerical sequences and series, and sequences and series of functions respectively; Chapter 11 covers continuity of functions; Chapter 12 and Chapter 13 cover differentiation and Riemann–Stieljes integration respectively; Chapter 15 covers some special functions such as power series, exponential and logarithmic functions, trigonometric functions, fourier series and the gamma function.

Part IV covers **topology**, which follows [Mun18].

The reader is not assumed to have any mathematical prerequisites, although some experience with proofs may be helpful. **Preliminary topics** such as logic and methods of proofs (Chapter 1), and basic set theory (Chapter 2) are covered in the appendix.

Note on Presentation

The following are some common mathematical terms used in this book. They are neither exhaustive nor rigorous, but they should give you a good idea of what is meant when these terms are used. The following terms are **bolded** when they are used, for readability.

- **Definition**: a precise and unambiguous description of the meaning of a mathematical term. It characterises 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. It 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 (Zorn's lemma, Urysohn's lemma, Burnside's lemma, Sperner's lemma).
- **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 proved and often interesting result, but generally less important than a theorem.
- **Conjecture**: a statement that is unproved, but is believed to be true (Collatz conjecture, Goldbach conjecture, twin prime conjecture).

- **Claim**: an assertion that is then proved. It is often used like an informal lemma.
- **Axiom/Postulate**: a statement that is assumed to be true without proof. These are the basic building blocks from which all theorems are proved (Euclid's five postulates, Zermelo–Fraenkel axioms, Peano axioms).
- **Identity**: a mathematical expression giving the equality of two (often variable) quantities (trigonometric identities, Euler's identity).
- **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 (Russell's paradox). The term paradox is often used informally to describe a surprising or counterintuitive result that follows from a given set of rules (Banach–Tarski paradox, Alabama paradox, Gabriel's horn).

Important terms are **coloured** when they are first defined, and are included in the glossary at the end of the book. Less important terms are instead *italicised* when they are first defined, and are not included in the glossary.

Note on Problem Solving

Mathematics is about problem solving. In [Pól45], George Pólya outlined the following problem solving cycle.

1. **Understand the problem**

   Ask yourself the following questions:

   - Do you understand all the words used in stating the problem?
   - Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?
   - What are you asked to find or show? Can you restate the problem in your own words?
   - Draw a figure. Introduce suitable notation.
   - Is there enough information to enable you to find a solution?

2. **Devise a plan**

   A partial list of heuristics – good rules of thumb to solve problems – is included:

   - Guess and check
   - Look for a pattern
   - Make an orderly list
   - Draw a picture
   - Eliminate possibilities
   - Solve a simpler problem
   - Use symmetry
   - Use a model
   - Consider special cases
   - Work backwards
   - Use direct reasoning
   - Use a formula

- Solve an equation
- Be ingenious

3. **Execute the plan**

   This step is usually easier than devising the plan. In general, all you need is care and patience, given that you have the necessary skills. Persist with the plan that you have chosen. If it continues not to work discard it and choose another. Don't be misled, this is how mathematics is done, even by professionals.

   - Carrying out your plan of the solution, check each step. Can you see clearly that the step is correct? Can you prove that it is correct?

4. **Check and expand**

   Pólya mentions that much can be gained by taking the time to reflect and look back at what you have done, what worked, and what didn't. Doing this will enable you to predict what strategy to use to solve future problems.

   Look back reviewing and checking your results. Ask yourself the following questions:

   - Can you check the result? Can you check the argument?
   - Can you derive the solution differently? Can you see it at a glance?
   - Can you use the result, or the method, for some other problem?

Building on Pólya's problem solving strategy, Schoenfeld [Sch92] came up with the following framework for problem solving, consisting of four components:

1. **Cognitive resources**: the body of facts and procedures at one's disposal.
2. **Heuristics**: 'rules of thumb' for making progress in difficult situations.
3. **Control**: having to do with the efficiency with which individuals utilise the knowledge at their disposal. Sometimes, this is referred to as metacognition, which can be roughly translated as 'thinking about one's own thinking'.

   (a) These are questions to ask oneself to monitor one's thinking.
      - What (exactly) am I doing? [Describe it precisely.] Be clear what I am doing NOW. Why am I doing it? [Tell how it fits into the solution.]
      - Be clear what I am doing in the context of the BIG picture – the solution. Be clear what I am going to do NEXT.

   (b) Stop and reassess your options when you
      - cannot answer the questions satisfactorily [probably you are on the wrong track]; OR
      - are stuck in what you are doing [the track may not be right or it is right but it is at that moment too difficult for you].

   (c) Decide if you want to
      - carry on with the plan,

- abandon the plan, OR
- put on hold and try another plan.

4. **Belief system**: one's perspectives regarding the nature of a discipline and how one goes about working on it.

Contents

1 Mathematical Reasoning and Logic

Learning Outcomes

In this chapter, we will

- introduce basic logic;
- introduce common methods of proof.

§1.1 Zeroth-order Logic

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 $P(x)$.

We can do some algebra on propositions, which include

(i) **equivalence**, denoted by $P \iff Q$, which means P and Q are logically equivalent statements;

(ii) **conjunction**, denoted by $P \land Q$, which means "P and Q";

(iii) **disjunction**, denoted by $P \lor Q$, which means "P or Q";

(iv) **negation**, denoted by $\neg P$, which means "not P".

Here are some useful properties when handling logical statements. You can easily prove all of them using truth tables.

Proposition 1.1.

(i) Double negation law: $P \iff \neg(\neg P)$.

(ii) Commutative: $P \land Q \iff Q \land P$, $P \lor Q \iff Q \lor P$.

(iii) Conjunction is associative: $(P \land Q) \land R \iff P \land (Q \land R)$.

(iv) Disjunction is associative: $(P \lor Q) \lor R \iff P \lor (Q \lor R)$.

(v) Conjunction distributes over disjunction: $P \land (Q \lor R) \iff (P \land Q) \lor (P \land Q)$.

(vi) Disjunction distributes over conjunction: $P \lor (Q \land R) \iff (P \lor Q) \land (P \lor R)$.

Proposition 1.2 (de Morgan's laws)**.**

$$\neg(P \vee Q) \iff (\neg P \wedge \neg Q)$$

$$\neg(P \wedge Q) \iff (\neg P \vee \neg Q)$$

If, only if

Implication is denoted by $P \implies Q$, which means "P implies Q", i.e. if P holds then Q also holds. It is equivalent to saying "If P then Q". $P \implies Q$ is known as a *conditional statement*, where P is known as the *hypothesis* and Q is known as the *conclusion*. The only case when $P \implies Q$ is false is when the hypothesis P is true and the conclusion Q is false.

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 only if Q;

(iv) P is a sufficient condition for Q;

(v) Q is a necessary condition for P.

Given $P \implies Q$,

- its **converse** is $Q \implies P$; both are not logically equivalent;
- its **inverse** is $\neg P \implies \neg Q$, i.e. the hypothesis and conclusion of the statement are both negated; both are not logically equivalent;
- the **contrapositive** is $\neg Q \implies \neg P$; both are logically equivalent.

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).

If and only if, iff

Bidirectional implication is denoted by $P \iff Q$, which means both $P \implies Q$ and $Q \implies P$; $P \iff Q$ is known as a *biconditional statement*. 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".

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

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

§1.2 First-order Logic

The **universal quantifier** is denoted by $\forall$, which means "for all" or "for every". A *universal statement* takes the form $\forall x \in X, P(x)$.

The **existential quantifier** is denoted by $\exists$, which means "there exists". An *existential statement* takes the form $\exists x \in X, P(x)$, where X is known as the *domain*.

Proposition 1.3 (de Morgan's laws)**.**

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

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

Exercise

Negate the statement

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

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

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

□

Exercise

Negate surjectivity.

Solution. If $f : X \to Y$ is not surjective, then it means that there exists $y \in Y$ not in the image of X, i.e. for all x in X we have $f(x) \neq y$.

$$\begin{aligned} \neg\forall y \in Y, \exists x \in X, f(x) = y &\iff \exists y \in Y, \neg(\exists x \in X, f(x) = y) \\ &\iff \exists y \in Y, \forall x \in X, \neg(f(x) = y) \\ &\iff \exists y \in Y, \forall x \in X, f(x) \neq y \end{aligned}$$

□

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.

§1.3 Methods of Proof

A **direct proof** of $P \implies Q$ is a series of valid arguments that start with the hypothesis P and end with the conclusion Q. 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.

A **proof by contrapositive** of $P \implies Q$ is to prove instead $\neg Q \implies \neg P$.

A **disproof by counterexample** is to providing a counterexample in order to refute or disprove a conjecture. The counterexample must make the hypothesis a true statement, and the conclusion a false statement. 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.

A **proof by cases** is to first dividing the situation into cases which exhaust all the possibilities, and then show that the statement follows in all cases.

Proof by Contradiction

A **proof by contradiction** of P involves first supposing P is false, i.e. $\neg P$; to prove $P \implies Q$ by contradiction, suppose $P \wedge \neg Q$. Then show through some logical reasoning that this leads to a contradiction or inconsistency. We may arrive at something that contradicts the hypothesis 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.

Exercise (Irrationality of $\sqrt{2}$)

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

Solution. We prove by contradiction. Suppose otherwise, that $\sqrt{2}$ is rational. Then $\sqrt{2} = \frac{a}{b}$ for some $a, b \in \mathbf{Z}, b \neq 0$, a, b coprime.

Squaring both sides gives

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

Since RHS is even, LHS must also be even. Hence it follows that a is even. Let $a = 2k$ where $k \in \mathbf{Z}$. 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 contradicts the assumption that a and b coprime, so we are done. □

Exercise (Euclid)

Prove that there are infinitely many prime numbers.

Solution. Suppose otherwise, that only finitely many prime numbers exist. List them as $p_1, \ldots, p_n$. The number $N = p_1 p_2 \cdots p_n + 1$ is divisible by a prime p, yet is coprime to $p_1, \ldots, p_n$. Therefore, p does not belong to our list of all prime numbers, a contradiction. $\square$

To **prove uniqueness**, we can either assume $\exists x, y \in S$ such that $P(x) \wedge P(y)$ is true and show $x = y$, or argue by assuming that $\exists x, y \in S$ are distinct such that $P(x) \wedge P(y)$, then derive a contradiction. $\exists!$ denotes "there exists a unique". To prove uniqueness and existence, we also need to show that $\exists x \in S$ s.t. $P(x)$ is true.

Proof of Existence

To prove existential statements, we can adopt two approaches:

1. Constructive proof (direct proof)

   To prove statements of the form $\exists x \in X$ s.t. $P(x)$, find or construct **a specific example** for x. To prove statements of the form $\forall y \in Y$, $\exists x \in X$ s.t. $P(x, y)$, construct example for x in terms of y (since x is dependent on y).

   In both cases, you have to justify that your example x

   (a) belongs to the domain X, and

   (b) satisfies the condition P.

2. Non-constructive proof (indirect proof)

   Use when specific examples are not easy or not possible to find or construct. Make arguments why such objects have to exist. May need to use proof by contradiction. Use definition, axioms or results that involve existential statements.

Exercise

Prove that we can find 100 consecutive positive integers which are all composite numbers.

Proof. We can prove this existential statement via constructive proof.

Our goal is to find integers $n, n+1, n+2, \ldots, n+99$, all of which are composite.

Take $n = 101! + 2$. Then n has a factor of 2 and hence is composite. Similarly, $n + k = 101! + (k+2)$ has a factor $k + 2$ and hence is composite for $k = 1, 2, \ldots, 99$.

Hence the existential statement is proven. $\square$

Exercise

Prove that for all rational numbers p and q with $p < q$, there is a rational number x such that $p < x < q$.

Proof. We prove this by construction. Our goal is to find such a rational x in terms of p and q.

We take the average. Let $x = \frac{p+q}{2}$ which is a rational number.

Since $p < q$,

$$x = \frac{p+q}{2} < \frac{q+q}{2} = q \implies x < q$$

Similarly,

$$x = \frac{p+q}{2} > \frac{p+p}{2} = p \implies p < x$$

Hence we have shown the existence of rational number x such that $p < x < q$.

Remark. For this type of question, there are two parts to prove: firstly, x satisfies the given statement; secondly, x is within the domain (for this question we do not have to prove x is rational since $\mathbf{Q}$ is closed under addition).

□

Exercise

Prove that for all rational numbers p and q with $p < q$, there is an irrational number r such that $p < r < q$.

Proof. We prove this by construction. Similarly, our goal is to find an irrational r in terms of p and q.

Note that we cannot simply take $r = \frac{p+q}{2}$; a simple counterexample is the case $p = -1, q = 1$ where $r = 0$ is clearly not irrational.

Since p lies in between p and q, let $r = p + c$ where $0 < c < q - p$. Since $c < q - p$, we have $c = \frac{q-p}{k}$ for some $k > 1$; to make c irrational, we take k to be irrational.

Take $r = p + \frac{q-p}{\sqrt{2}}$. We need to show r is irrational and $p < r < q$.

Part 1: $p < r < q$

Since $q < p$, $r = p + (\text{positive number}) > p$. On the other hand, $\frac{q-p}{\sqrt{2}} < q - p$ so $r < p + (q - p) = q$.

Part 2: r is irrational

We prove by contradiction. Suppose r is rational. We have $\sqrt{2} = \frac{q-p}{r-p}$. Since p, q, r are all rational (and $r - p \neq 0$), RHS is rational. This implies that LHS is rational, i.e. $\sqrt{2}$ is rational, a contradiction. □

Non-constructive proof:

Exercise

Prove that every integer greater than 1 is divisible by a prime.

Proof. If n is prime, then we are done as $n \mid n$.

If n is not prime, then n is composite. So n has a divisor d_1 such that $1 < d_1 < n$. If d_1 is prime then we are done as $d_1 \mid n$. If d_1 is not prime then d_1 is composite, has divisor d_2 such that $1 < d_2 < n$.

If d_2 is prime, then we are done as $d_2 \mid d_1$ and $d_1 \mid n$ imply $d_2 \mid n$. If d_2 is not prime then d_2 is composite, has divisor d_3 such that $1 < d_3 < d_2$.

Continuing in this manner after k times, we will get

$$1 < d_k < d_{k-1} < \cdots < d_2 < d_1 < n$$

where $d_i \mid n$ for all i.

Since there can only be a finite number of d_i's between 1 and n, this process must stop after finite steps. On the other hand, the process will stop only if there is a d_i which is a prime. Hence we conclude that there must be a divisor d_i of n that is prime. □

Remark. This proof is also known as *proof by infinite descent*, a method which relies on the well-ordering principle on $\mathbf{N}$.

Exercise

Prove that the equation $x^2 + y^2 = 3z^2$ has no solutions (x, y, z) in integers where $z \neq 0$.

Proof. Suppose we have a solution (x, y, z). Without loss of generality, we may assume that $z > 0$. By the least integer principle, we may also assume that our solution has z minimal. Taking remainders modulo 3, we see that

$$x^2 + y^2 \equiv 0 \pmod 3$$

Recalling that squares may only be congruent to 0 or 1 modulo 3, we conclude that

$$x^2 \equiv y^2 \equiv 0 \implies x \equiv y \equiv 0 \pmod 3$$

Writing $x = 3a$ and $y = 3b$ we obtain

$$9a^2 + 9b^2 = 3z^2 \implies 3(a^2 + b^2) = z^2 \implies 3 \mid z^2 \implies 3 \mid z$$

Now let $z = 3c$ and cancel 3's to obtain

$$a^2 + b^2 = 3c^2.$$

We have therefore constructed another solution $(a, b, c) = \left(\frac{x}{3}, \frac{y}{3}, \frac{z}{3}\right)$ to the original equation. However $0 < c < z$ contradicts the minimality of z. □

Proof by Mathematical Induction

Induction is an extremely powerful method of proof used throughout mathematics. It deals with infinite families of statements which come in the form of lists. The idea behind induction is in showing how each statement follows from the previous one on the list – all that remains is to kick off this logical chain reaction from some starting point.

We shall assume that $\mathbf{N}$ satisfies the *well-ordering principle*: every nonempty $S \subset \mathbf{N}$ has a least element; that is, there exists $m \in S$ such that $m \leqslant k$ for all $k \in S$.

Remark. The well-ordering principle does not hold for $\mathbf{Z}$, $\mathbf{Q}$, and $\mathbf{R}$.

Lemma 1.4. Let $S \subset \mathbf{N}$. If

(i) $1 \in S$

(ii) $k \in S \implies k+1 \in S$

then $S = \mathbf{N}$.

Proof. If $S = \mathbf{N}$ then we are done. Now suppose $S \neq \mathbf{N}$. Then $\mathbf{N} \setminus S$ is not empty. By the well-ordering principle, $\mathbf{N} \setminus S$ has a least element p. Since $1 \in S$, we must have $p > 1$. By (ii), $p = (p-1)+1 \in S$. But this contradicts $p \in \mathbf{N} \setminus S$. □

Theorem 1.5 (Principle of mathematical induction)**.** Let $P(n)$ be a family of statements indexed by $\mathbf{N}$. Suppose that

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

(ii) for all $k \in \mathbf{N}$, $P(k) \implies P(k+1)$.

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

(i) is known as the **base case**, (ii) is known as the **inductive step**. Using logic notation, the above can be written as

$$\{P(1) \wedge (\forall n \in \mathbf{N})[P(k) \implies P(k+1)]\} \implies (\forall n \in \mathbf{N})P(n)$$

Proof. Apply the above lemma to the set $S = \{n \in \mathbf{N} \mid P(n) \text{ is true}\}$. □

Exercise

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

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

Proof. Let $P(n) : \displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Clearly $P(1)$ holds. Now suppose $P(k)$ holds for some $k \in \mathbf{N}$, $k \geqslant 1$; that is,

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

Adding $k+1$ to both sides,

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

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

Since $P(1)$ true and $P(k) \implies P(k+1)$ for all $k \in \mathbf{N}$, $k \geqslant 1$,
$P(n)$ is true for all $n \in \mathbf{N}$. □

Exercise (Bernoulli's inequality)

Let $x \in \mathbf{R}$, $x > -1$. Let $n \in \mathbf{Z}^+$. Then

$$(1+x)^n \geqslant 1 + nx.$$

Proof. We prove by induction on n. Fix $x > -1$. Let $P(n) : (1+x)^n \geqslant 1 + nx$.

The base case $P(1)$ is clear.

Suppose that $P(k)$ is true for some $k \in \mathbf{Z}^+$, $k \geqslant 1$. That is, $(1+x)^k \geqslant 1 + kx$. Note that $1 + x > 0$, and $kx^2 \geqslant 0$ (since $k > 0$ and $x^2 \geqslant 0$). Then

$$\begin{aligned}(1+x)^{k+1} &= (1+x)(1+x)^k \\ &\geqslant (1+x)(1+kx) \quad \text{[induction hypothesis]} \\ &= 1 + (k+1)x + kx^2 \\ &\geqslant 1 + (k+1)x \quad [\because kx^2 \geqslant 0]\end{aligned}$$

so $P(k+1)$ is true. Hence by induction, the result holds. □

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

Corollary 1.6. Let $P(n)$ be a family of statements indexed by integers $n \geqslant N$ for $N \in \mathbf{Z}$. Suppose that

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

(ii) for all $k \geqslant N$, $P(k) \implies P(k+1)$.

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

Proof. Apply Theorem 1.5 to the statement $Q(n) = P(n+N)$ for $n \in \mathbf{N}$. □

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

Theorem 1.7 (Strong induction)**.** Let $P(n)$ be a family of statements indexed by $\mathbf{N}$. Suppose that

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

(ii) for all $k \in \mathbf{N}$, $P(1) \wedge \cdots \wedge P(k) \implies P(k+1)$.

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

Proof. 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 $n \in \mathbf{N}$, $Q(n) \implies Q(n+1)$. It follows by induction that $Q(n)$ holds for all $n \in \mathbf{N}$, and hence $P(n)$ holds for all n. □

> **Exercise** (Fundamental theorem of arithmetic)
>
> Prove that every natural number greater than 1 may be expressed as a product of one or more prime numbers.

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

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

Let $n \geqslant 2$ 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.

In both cases, $P(n)$ holds. Hence by strong induction, $P(n)$ is true for all $n \in \mathbf{N}$. □

The following is also another variant on induction.

Theorem 1.8 (Cauchy induction)**.** Let $P(n)$ be a family of statements indexed by $\mathbf{N}_{\geqslant 2}$. Suppose that

(i) $P(2)$ is true;

(ii) for all $k \in \mathbf{N}$, $P(k) \implies P(2k)$ and $P(k) \implies (k-1)$.

Then $P(n)$ is true for all $n \in \mathbf{N}_{\geqslant 2}$.

Exercise (AM–GM inequality)

Given $n \in \mathbf{N}$, prove that for positive reals $a_1, a_2, dots, a_n$,

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geqslant \sqrt[n]{a_1 a_2 \cdots a_n}.$$

Proof. Let $P(n) : \dfrac{a_1 + a_2 + \cdots + a_n}{n} \geqslant \sqrt[n]{a_1 a_2 \cdots a_n}$.

Base case $P(2)$ is true because

$$\frac{a_1 + a_2}{2} \geqslant \sqrt{a_1 a_2} \iff (a_1 + a_2)^2 \geqslant 4a_1 a_2 \iff (a_1 - a_2)^2 \geqslant 0$$

Next we show that $P(n) \implies P(2n)$

$$\frac{a_1 + a_2 + \cdots + a_{2n}}{2n} = \frac{\frac{a_1 + a_2 + \cdots + a_n}{n} + \frac{a_{n+1} + a_{n+2} + \cdots + a_{2n}}{n}}{2}$$

$$\frac{\frac{a_1 + a_2 + \cdots + a_n}{n} + \frac{a_{n+1} + a_{n+2} + \cdots + a_{2n}}{n}}{2} \geqslant \frac{\sqrt[n]{a_1 a_2 \cdots a_n} + \sqrt[n]{a_{n+1} a_{n+2} \cdots a_{2n}}}{2}$$

$$\frac{\sqrt[n]{a_1 a_2 \cdots a_n} + \sqrt[n]{a_{n+1} a_{n+2} \cdots a_{2n}}}{2} \geqslant \sqrt{\sqrt[n]{a_1 a_2 \cdots a_n}\,\sqrt[n]{a_{n+1} a_{n+2} \cdots a_{2n}}}$$

$$\sqrt{\sqrt[n]{a_1 a_2 \cdots a_n}\,\sqrt[n]{a_{n+1} a_{n+2} \cdots a_{2n}}} = \sqrt[2n]{a_1 a_2 \cdots a_{2n}}$$

The first inequality follows from n-variable AM–GM, which is true by assumption, and the second inequality follows from 2-variable AM–GM, which is proven above.

Finally we show that $P(n) \implies P(n-1)$. By n-variable AM–GM, $\frac{a_1 + a_2 + \cdots + a_n}{n} \geqslant \sqrt[n]{a_1 a_2 \cdots a_n}$ Let $a_n = \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}$ Then we have

$$\frac{a_1 + a_2 + \cdots + a_{n-1} + \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}}{n} = \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}$$

So,

$$\frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1} \geqslant \sqrt[n]{a_1 a_2 \cdots a_{n-1} \cdot \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}}$$

$$\Rightarrow \left(\frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}\right)^n \geqslant a_1 a_2 \cdots a_{n-1} \cdot \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}$$

$$\Rightarrow \left(\frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1}\right)^{n-1} \geqslant a_1 a_2 \cdots a_{n-1}$$

$$\Rightarrow \frac{a_1 + a_2 + \cdots + a_{n-1}}{n-1} \geqslant \sqrt[n-1]{a_1 a_2 \cdots a_{n-1}}$$

By Cauchy induction, this proves the AM–GM inequality for n variables. $\square$

Pigeonhole Principle

Theorem 1.9 (Pigeonhole principle)**.** If $kn+1$ objects are distributed among n boxes, one of the boxes will contain at least $k+1$ objects.

Exercise (IMO 1972)

Prove that every set of 10 two-digit integer numbers has two disjoint subsets with the same sum of elements.

Solution. Let S be the set of 10 numbers. It has $2^{10}-2=1022$ subsets that differ from both S and the empty set. They are the "pigeons".

If $A \subset S$, the sum of elements of A cannot exceed $91+92+\cdots+99=855$. The numbers between 1 and 855, which are all possible sums, are the "holes".

Because the number of "pigeons" exceeds the number of "holes", there will be two "pigeons" in the same "hole". Specifically, there will be two subsets with the same sum of elements. Deleting the common elements, we obtain two disjoint sets with the same sum of elements. $\square$

Exercise (Putnam 2006)

Prove that for every set $X=\{x_1,x_2,\ldots,x_n\}$ of n real numbers, there exists a nonempty subset S of X and an integer m such that

$$\left| m+\sum_{x\in S} s \right| \leqslant \frac{1}{n+1}.$$

Solution. Recall that the fractional part of a real number x is $x-\lfloor x\rfloor$. Let us look at the fractional parts of the numbers $x_1, x_1+x_2, \ldots, x_1+x_2+\cdots+x_n$. If any of them is either in the interval $\left[0, \frac{1}{n+1}\right]$ or $\left[\frac{n}{n+1}, 1\right]$, then we are done. If not, we consider these n numbers as the "pigeons" and the $n-1$ intervals $\left[\frac{1}{n+1}, \frac{2}{n+1}\right], \left[\frac{2}{n+1}, \frac{3}{n+1}\right], \ldots, \left[\frac{n-1}{n+1}, \frac{n}{n+1}\right]$ as the "holes". By the pigeonhole principle, two of these sums, say $x_1+x_2+\cdots+x_k$ and $x_1+x_2+\cdots+x_{k+m}$, belong to the same interval. But then their difference $x_{k+1}+\cdots+x_{k+m}$ lies within a distance of $\frac{1}{n+1}$ of an integer, and we are done. $\square$

Exercises

Problem 1.1. Use the Unique Factorisation Theorem to prove that, if a positive integer n is not a perfect square, then $\sqrt{n}$ is irrational.

[The Unique Factorisation Theorem states that every integer $n > 1$ has a unique standard factored form, i.e. there is exactly one way to express $n = p_1^{k_1} p_2^{k_2} \cdots p_t^{k_t}$ where $p_1 < p_2 < \cdots < p_t$ are distinct primes and $k_1, k_2, \ldots, k_t$ are some positive integers.]

Proof. Prove by contradiction. Suppose n is not a perfect square and $\sqrt{n}$ is rational. Then $\sqrt{n} = \frac{a}{b}$ for some $a, b \in \mathbf{Z}$. Squaring both sides and clearing denominator gives

$$nb^2 = a^2. \tag{$*$}$$

Consider the standard factored forms of n, a and b:

$$n = p_1^{k_1} p_2^{k_2} \cdots p_t^{k_t}$$

$$a = q_1^{e_1} q_2^{e_2} \cdots q_u^{e_u} \implies a^2 = q_1^{2e_1} q_2^{2e_2} \cdots q_u^{2e_u}$$

$$b = r_1^{f_1} r_2^{f_2} \cdots r_v^{f_v} \implies b^2 = r_1^{2f_1} r_2^{2f_2} \cdots r_v^{2f_v}$$

i.e. the powers of primes in the standard factored form of a^2 and b^2 are all even integers.

This means the powers k_i of primes p_i in the standard factored form of n are also even by Unique Factorisation Theorem. Note that all p_i appear in the standard factored form of a^2 with even power $2c_i$, because of $(*)$. By UFT, p_i must also appear in the standard factored form of nb^2 with the same even power $2c_i$.

If $p_i \nmid b$, then $k_i = 2c_i$ which is even. If $p_i \mid b$, then p_i will appear in b^2 with even power $2d_i$. So $k_i + 2d_i = 2c_i$, and hence $k_i = 2(c_i - d_i)$, which is again even.

Hence $n = p_1^{k_1} p_2^{k_2} \cdots p_t^{k_t} = \left(p_1^{\frac{k_1}{2}} p_2^{\frac{k_2}{2}} \cdots p_t^{\frac{k_t}{2}} \right)^2$.

Since $\frac{k_i}{2}$ are all integers, $p_1^{\frac{k_1}{2}} p_2^{\frac{k_2}{2}} \cdots p_t^{\frac{k_t}{2}}$ is an integer and n is a perfect square. This contradicts the given hypothesis that n is not a perfect square. □

Problem 1.2. Prove that for every pair of irrational numbers p and q such that $p < q$, there is an irrational x such that $p < x < q$.

Proof. Consider the average of p and q: $p < \dfrac{p+q}{2} < q$.

If $\dfrac{p+q}{2}$ is irrational, take $x = \dfrac{p+q}{2}$ and we are done.

If $\dfrac{p+q}{2}$ is rational, call it r, take the average of p and r: $p < \dfrac{p+r}{2} < r < q$. Since p is irrational and r is rational, $\dfrac{p+r}{2}$ is irrational. In this case, we take $x = \dfrac{3p+q}{4}$. □

Problem 1.3. Given n real numbers $a_1, a_2, \ldots, a_n$. Show that there exists an a_i $(1 \leqslant i \leqslant n)$ such that a_i is greater than or equal to the mean (average) value of the n numbers.

Proof. Prove by contradiction.

Let $\bar{a}$ denote the mean value of the n given numbers. Suppose $a_i < \bar{a}$ for all a_i. Then

$$\bar{a} = \frac{a_1 + a_2 + \cdots + a_n}{n} < \frac{\bar{a} + \bar{a} + \cdots + \bar{a}}{n} = \frac{n\bar{a}}{n} = \bar{a}.$$

We derive $\bar{a} < \bar{a}$, which is a contradiction.

Hence there must be some a_i such that $a_i > \bar{a}$. □

Problem 1.4. Prove that the following statement is false: there is an irrational number a such that for all irrational number b, ab is rational.

Thought process: prove the negation of the statement: for every irrational number a, there is an irrational number b such that ab is irrational.

Proving technique: constructive proof (note that we can consider multiple cases and construct more than one b)

Proof. Given an irrational number a, let us consider $\dfrac{\sqrt{2}}{a}$.

Case 1: $\dfrac{\sqrt{2}}{a}$ is irrational.

Take $b = \dfrac{\sqrt{2}}{a}$. Then $ab = \sqrt{2}$ which is irrational.

Case 2: $\dfrac{\sqrt{2}}{a}$ is rational.

Then the reciprocal $\dfrac{a}{\sqrt{2}}$. Since $\sqrt{6}$ is irrational, the product $\left(\dfrac{a}{\sqrt{2}}\right)\sqrt{6} = a\sqrt{3}$ is irrational. Take $b = \sqrt{3}$, which is irrational. Then $ab = a\sqrt{3}$ which is irrational. □

Problem 1.5. Prove that there are infinitely many prime numbers that are congruent to 3 modulo 4.

Proof. Prove by contradiction.

Suppose there are only finitely many primes that are congruent to 3 modulo 4. Let $p_1, p_2, \ldots, p_m$ be the list of all the primes that are congruent to 3 modulo 4.

We construct an integer M by $M = (p_1 p_2 \cdots p_m)^2 + 2$.

We have the following observation:

(i) $M \equiv 3 \mod 4$.

(ii) Every p_i divides $M - 2$.

(iii) None of the p_i divides M. [Otherwise, together with (ii), this will imply p_i divides 2, which is impossible.]

(iv) M is not a prime number. [Otherwise, by (i), M is a prime number congruent to 3 modulo 4. But $M \neq p_i$ for all $1 \leqslant i \leqslant m$. This contradicts the assumption that $p_1, p_2, \ldots, p_m$ are all the prime numbers congruent to 3 modulo 4.]

From the above discussion, we know that M is a composite number by (iv). So it has a prime factorization $M = q_1 q_2 \cdots q_k$.

Since M is odd, all these prime factors q_j must be odd, and hence q_j must be congruent to either 1 or 3 modulo 4.

By (iii), q_j cannot be any of the p_i. So all q_j must be congruent to 1 modulo 4. Then M, which is the product of q_j, must also be congruent to 1 modulo 4.

This contradicts (i) that M is congruent to 3 modulo 4.

Hence we conclude that there must be infinitely many primes that are congruent to 3 modulo 4. □

Problem 1.6. Prove that, for any positive integer n, there is a perfect square m^2 (m is an integer) such that $n \leqslant m^2 \leqslant 2n$.

Proof. Prove by contradiction.

Suppose otherwise, that $n > m^2$ and $(m+1)^2 > 2n$ so that there is no square between n and $2n$, then

$$(m+1)^2 > 2n > 2m^2.$$

Since we are dealing with integers and the inequalities are strict, we get

$$(m+1)^2 \geqslant 2m^2 + 2$$

which simplifies to

$$0 \geqslant m^2 - 2m + 1 = (m-1)^2$$

The only value for which this is possible is $m = 1$, but you can eliminate that easily enough. □

Problem 1.7. Prove that for every positive integer $n \geqslant 4$,

$$n! > 2^n.$$

Proof. Let $P(n) : n! > 2^n$

Base case: $P(4)$

LHS: $4! = 4 \times 3 \times 2 \times 1 = 24$, RHS: $2^4 = 16 < 24$

So $P(4)$ is true.

Inductive step: $P(k) \implies P(k+1)$ for all $k \in \mathbf{N}_{\geqslant 4}$

$$\begin{aligned} k! &> 2^k \\ (k+1)k! &> 2^k(k+1) \\ &> 2^k 2 \quad \text{since from } k \geqslant 4,\ k+1 \geqslant 5 > 2 \\ &= 2^{k+1} \end{aligned}$$

hence proven $P(k) \implies P(k+1)$ for integers $k \geqslant 4$.

By PMI, we have proven $P(n)$ for all integers $n \geqslant 4$. □

Problem 1.8. Prove by mathematical induction, for $n \geqslant 2$,

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

Proof. Let $P(n) : \sqrt[n]{n} < 2 - \frac{1}{n}$ for $n \geqslant 2$.

Base case: $P(2)$

When $n = 2$, $\sqrt{2} = 1.41 \cdots < 2 - \frac{1}{2} = 1.5$ which is true. Hence $P(2)$ is true.

Inductive step: $P(k) \implies P(k+1)$ for all $k \in \mathbf{N}_{\geqslant 2}$

Assume $P(k)$ is true for $k \geqslant 2, k \in \mathbf{N}$, 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. □

Problem 1.9. Prove that for all integers $n \geqslant 3$,

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

Proof. **Base case:** $P(3)$

On the LHS, $\left(1+\frac{1}{3}\right)^3 = \frac{64}{27} = 2\frac{10}{27} < 3$. Hence $P(3)$ is true.

Inductive step: $P(k) \implies P(k+1)$ for all $k \in \mathbf{N}_{\geqslant 3}$

Our inductive hypothesis is

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

Multiplying both sides by $\left(1+\frac{1}{k}\right)$ (to get a $k+1$ in the power),

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

Since $k < k+1 \iff \frac{1}{k} > \frac{1}{k+1}$,

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

The rest of the proof follows easily. □

A sequence of integers F_i, where integer $1 \leqslant i \leqslant n$, is called the *Fibonacci sequence* if and only if it is defined recursively by $F_1 = 1$, $F_2 = 1$, $F_n = F_{n-1} + F_{n-2}$ for $n > 2$.

Problem 1.10. Let a_i where integer $1 \leqslant i \leqslant n$ be a sequence of integers defined recursively by initial conditions $a_1 = 1$, $a_2 = 1$, $a_3 = 3$ and the recurrence relation $a_n = a_{n-1} + a_{n-2} + a_{n-3}$ for $n > 3$.

For all $n \in \mathbf{N}$, prove that

$$a_n \leqslant 2^{n-1}.$$

Proof. Let $P(n) : a_n \leqslant 2^{n-1}$.

Given the recurrence relation, it could be possible to use $P(k), P(k+1), P(k+2)$ to prove $P(k+3)$ for all $k \in \mathbf{N}$.

Base case: $P(1), P(2), P(3)$

$P(1) : a_1 = 1 \leqslant 2^{1-1} = 1$ is true.

$P(2) : a_2 = 1 \leqslant 2^{2-1} = 2$ is true.

$P(3) : a_3 = 3 \leqslant 2^{3-1} = 4$ is true.

Inductive step: $P(k) \wedge P(k+1) \wedge P(k+2) \implies P(k+3)$ for all $k \in \mathbf{N}$

By inductive hypothesis, for $k \in \mathbf{N}$ we have $a_k \leqslant 2^k, a_{k+1} \leqslant 2^{k+1}, a_{k+2} \leqslant 2^{k+2}$.

$$\begin{aligned} a_{k+3} &= a_k + a_{k+1} + a_{k+2} \quad \text{[start from recurrence relation]} \\ &\leqslant 2^k + 2^{k+1} + 2^{k+2} \quad \text{[use inductive hypothesis]} \\ &= 2^k(1+2+2^2) \\ &< 2^k(2^3) \quad \text{[approximation, since } 1+2+2^2 < 2^3\text{]} \\ &= 2^{k+3} \end{aligned}$$

which is precisely $P(k+3) : a_{k+3} \leqslant 2^{k+3}$. □

Problem 1.11. For $m, n \in \mathbf{N}$, prove that

$$F_{n+m+1} = F_n F_m + F_{n+1} F_{m+1}.$$

Proof. For $n \in \mathbf{N}$, take $P(n) : F_{n+m+1} = F_n F_m + F_{n+1} F_{m+1}$ for all $m \in \mathbf{N}$ in the cases $k = n$ and $k = n + 1$.

So we are using induction to progress through n and dealing with m simultaneously at each stage.

To verify $P(0)$, we note that

$$F_{m+1} = F_0 F_m + F_1 F_{m+1}$$

and

$$F_{m+2} = F_1 F_m + F_2 F_{m+1}$$

for all m, as $F_0 = 0$ and $F_1 = F_2 = 1$.

For the inductive step we assume $P(n)$, i.e. that for all $m \in \mathbf{N}$,

$$\begin{aligned} F_{n+m+1} &= F_n F_m + F_{n+1} F_{m+1}, \\ F_{n+m+2} &= F_{n+1} F_m + F_{n+2} F_{m+1}. \end{aligned}$$

Then

$$\begin{aligned} F_{n+m+3} &= F_{n+m+2} + F_{n+m+1} \\ &= F_n F_m + F_{n+1} F_{m+1} + F_{n+1} F_m + F_{n+2} F_{m+1} \\ &= (F_n + F_{n+1}) Fm + (F_{n+1} + F_{n+2}) F_{m+1} \\ &= F_{n+2} F_m + F_{n+3} F_{m+1} \end{aligned}$$

thus $P(n+1)$ is true, for all $m \in \mathbf{N}$. □

2 Set Theory

Learning Outcomes

In this chapter, we will

- recap basic definitions relating to sets (excluding detailed axiomatic discussions);
- define relations and related concepts including binary relation, partial order, total order, well order, equivalence relations, equivalence relations, equivalence class, quotient set, partition of a set;
- define functions, injectivity, surjectivity, bijectivity, composition, invertibility, monotonicity;

§2.1 Basics

Definitions and Notations

A **set** S can be loosely defined as a collection of objects[1]. For a set S, we write $x \in S$ to mean that x is an **element** of S, and $x \notin S$ if otherwise.

To describe a set, one can list its elements explicitly. A set can also be defined in terms of some property $P(x)$ that the elements $x \in S$ satisfy, denoted by the following set builder notation:

$$\{x \in S \mid P(x)\}$$

Some basic sets (of numbers) you should be familiar with:

- $\mathbf{N} = \{1, 2, 3, \dots\}$ denotes the natural numbers (non-negative integers).
- $\mathbf{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ denotes the integers.
- $\mathbf{Q} = \{\frac{p}{q} \mid p, q \in \mathbf{Z}, q \neq 0\}$ denotes the rational numbers.

[1] *Russell's paradox*, after the mathematician and philosopher Bertrand Russell (1872–1970), provides a warning as to the looseness of our definition of a set. Suppose H is the collection of sets that are not elements of themselves; that is,

$$H = \{S \mid S \notin S\}.$$

The problem arises when we ask the question of whether or not H is itself in H? On one hand, if $H \notin H$ then H meets the precise criterion for being in H and so $H \in H$, a contradiction. On the other hand, if $H \in H$ then by the property required for this to be the case, $H \notin H$, another contradiction. Thus we have a paradox: H is neither in H, nor not in H.

The modern resolution of Russell's paradox is that we have taken too naive an understanding of "collection", and that Russell's "set" H is in fact not a set. It does not fit within axiomatic set theory (which relies on the so-called ZF axioms), and so the question of whether or not H is in H simply doesn't make sense.

- $\mathbf{R}$ denotes the real numbers (the construction of which using Dedekind cuts will be discussed in Chapter 8).
- $\mathbf{C} = \{x + yi \mid x, y \in \mathbf{R}\}$ denotes the complex numbers.

We have that
$$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R} \subset \mathbf{C}.$$

The **empty set** is the set with no elements, denoted by $\emptyset$.

A is a **subset** of B if every element of A is in B, denoted by $A \subset B$:
$$A \subset B \iff (\forall x)(x \in A \implies x \in B)$$

We denote $A \subsetneq B$ to explicitly mean that $A \subset B$ and $A \neq B$; we call A a **proper subset** of B.

Proposition 2.1 ($\subset$ is transitive)**.** If $A \subset B$ and $B \subset C$, then $A \subset C$.

Proof. For all $x \in A$, $x \in B$ Let $x \in A$. Since $A \subset B$ and $x \in A$, $x \in B$. Since $B \subset C$ and $x \in B$, $x \in C$. Hence $A \subset C$. □

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

Proposition 2.2 (Double inclusion)**.** Let $A \subset S$ and $B \subset S$. Then
$$A = B \iff (A \subset B) \wedge (B \subset A)$$

Proof. We have
$$\begin{aligned} A = B &\iff (\forall x)[x \in A \iff x \in B] \\ &\iff (\forall x)[(x \in A \implies x \in B) \wedge (x \in B \implies x \in A)] \\ &\iff \{(\forall x)[x \in A \implies x \in B]\} \wedge (\forall x)[x \in B \implies x \in A)] \\ &\iff (A \subset B) \wedge (B \subset A) \end{aligned}$$

□

Remark. Double inclusion is a useful tool to prove that two sets are equal.

Some frequently occurring subsets of $\mathbf{R}$ are known as **intervals**, which can be visualised as sections of the real line. We define *bounded intervals*
$$\begin{aligned} (a, b) &= \{x \in \mathbf{R} \mid a < x < b\}, \\ [a, b] &= \{x \in \mathbf{R} \mid a \leqslant x \leqslant b\}, \\ [a, b) &= \{x \in \mathbf{R} \mid a \leqslant x < b\}, \\ (a, b] &= \{x \in \mathbf{R} \mid a < x \leqslant b\}, \end{aligned}$$

and *unbounded intervals*

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

An interval of the first type (a,b) is called an *open interval*; an interval of the second type $[a,b]$ is called a *closed interval*. Note that if $a=b$, then $[a,b]=\{a\}$, while $(a,b)=[a,b)=(a,b]=\emptyset$.

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

$$\mathcal{P}(A)=\{S\mid S\subset A\}.$$

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:

$$A\times B:=\{(a,b)\mid a\in A,b\in B\}.$$

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

Example

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

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

To generalise the notion of intervals, we define a *k-cell* as

$$\{(x_1,\dots,x_n)\in\mathbf{R}^k\mid a_i\leqslant x_i\leqslant b_i(1\leqslant i\leqslant k)\}.$$

For example, a 1-cell is an interval, a 2-cell is a rectangle, and a 3-cell is a rectangular solid. In this regard, we can think of a k-cell as a higher-dimensional version of a rectangle or rectangular solid; it is the Cartesian product of k closed intervals in $\mathbf{R}$.

Algebra of Sets

We now disuss the algebra of sets. Given $A \subset S$ and $B \subset S$,

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

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

(ii) 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:

$$A \cap B = \emptyset$$

More generally, we can take unions and intersections of arbitrary numbers of sets (could be finitely or infinitely many). Given 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

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

(iii) The **complement** of A, denoted by A^c, is the set containing elements that are not in A:

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

(iv) The **set difference**, or complement of B in A, denoted by $A \setminus B$, 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 2.3 (Distributive laws)**.** Let $A, B, C \subset S$. Then

(i) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;

(ii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof.

(i) Suppose $x \in A \cup (B \cap C)$. Then

$$\begin{aligned}
x \in A \cup (B \cap C) &\iff x \in A \quad \vee \quad x \in B \cap C \\
&\iff x \in A \quad \vee \quad (x \in B) \wedge (x \in C) \\
&\iff (x \in A) \vee (x \in B) \quad \wedge \quad (x \in A) \vee (x \in C) \\
&\iff x \in A \cup B \quad \wedge \quad x \in A \cup C \\
&\iff x \in (A \cup B) \cap (A \cup C).
\end{aligned}$$

Thus $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$.

Conversely suppose that $x \in (A \cup B) \cap (A \cup C)$. Then go in the reverse direction of the above steps to show that $(A \cup B) \cap (A \cup C) \subset A \cup (B \cap C)$.

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

(ii) Similar.

□

Proposition 2.4 (de Morgan's laws)**.** Let $A, B \subset S$. Then

(i) $(A \cup B)^c = A^c \cap B^c$;

(ii) $(A \cap B)^c = A^c \cup B^c$.

Proof.

(i)

$$\begin{aligned}
x \in (A \cup B)^c &\iff x \notin A \cup B \\
&\iff x \notin A \quad \wedge \quad x \notin B \\
&\iff x \in A^c \quad \wedge \quad x \in B^c \\
&\iff x \in A^c \cap B^c
\end{aligned}$$

(ii) Similar.

□

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

$$\begin{aligned}
\left(\bigcap_{i \in I} A_i \right)^c &= \bigcup_{i \in I} {A_i}^c, \\
\left(\bigcup_{i \in I} A_i \right)^c &= \bigcap_{i \in I} {A_i}^c
\end{aligned}$$

Exercise

Prove the following:

(i) $\left(\bigcup_{i\in I} A_i\right) \cup B = \bigcup_{i\in I}(A_i \cup B)$

(ii) $\left(\bigcap_{i\in I} A_i\right) \cup B = \bigcap_{i\in I}(A_i \cup B)$

(iii) $\left(\bigcup_{i\in I} A_i\right) \cup \left(\bigcup_{j\in J} B_j\right) = \bigcup_{(i,j)\in I\times J}(A_i \cup B_j)$

(iv) $\left(\bigcap_{i\in I} A_i\right) \cup \left(\bigcap_{j\in J} B_j\right) = \bigcap_{(i,j)\in I\times J}(A_i \cup B_j)$

Exercise

Let $S \subset A \times B$. Express the set A_S of all elements of A which appear as the first entry in at least one of the elements in S.

(A_S here may be called the projection of S onto A.)

§2.2 Relations

Definition and Examples

Definition 2.5 (Relation)**.** R is a **relation** between A and B if $R \subset A \times B$; $a \in A$ and $b \in B$ are said to be *related* if $(a, b) \in R$, denoted aRb.

Remark. A relation is a set of ordered pairs.

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.

Example

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

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

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

Example

Given $R = \{(1, a), (1, b), (2, b), (3, b)\}$, then $\operatorname{dom} R = \{1, 2, 3\}$ and $\operatorname{ran} R = \{a, b\}$.

Definition 2.6 (Binary relation)**.** A **binary relation** in A is a relation between A and itself; that is, $R \subset A \times A$.

Properties of Relations

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

(i) R is **reflexive** if xRx for all $x \in A$;

(ii) R is **symmetric** if $xRy \implies yRx$;

(iii) R is **anti-symmetric** if xRy and $yRx \implies x = y$;

(iv) R is **transitive** if xRy and $yRz \implies xRz$.

> **Example** (Less than or equal to)
>
> The relation $\leqslant$ on R is reflexive, anti-symmetric, and transitive, but not symmetric.

Definition 2.7. A **partial order** on a non-empty set A is a relation on A satisfying reflexivity, anti-symmetry and transitivity.

A **total order** on A is a partial order on A such that if for every $x, y \in A$, either xRy or yRx.

A **well order** on A is a total order on A such that every non-empty subset of A has a minimal element; that is, for each non-empty $B \subset A$ there exists $s \in B$ such that $s \leqslant b$ for all $b \in B$.

> **Example**
>
> You should verify the following:
>
> - Less than: the relation $<$ on R is not reflexive, symmetric, or anti-symmetric, but it is transitive.
> - Not equal to: the relation $\neq$ on R is not reflexive, anti-symmetric or transitive, but it is symmetric.

Definition 2.8. Let the non-empty set A be partially ordered by $\leqslant$.

- A subset $B \subset A$ is called a **chain** if for all $x, y \in B$, either $x \leqslant y$ or $y \leqslant x$.
- An **upper bound** for a subset $B \subset A$ is an element $u \in A$ such that $b \leqslant u$ for all $b \in B$.
- A **maximal element** of A is an element $m \in A$ such that $m \leqslant x$ for any $x \in A$, then $m = x$.

Lemma 2.9 (Zorn's lemma)**.** If A is a non-empty partially ordered set in which every chain has an upper bound, then A has a maximal element.

It is a non-trivial result that Zorn's lemma is independent of the usual (Zermelo–Fraenkel) axioms of set theory in the sense that if the axioms of set theory are consistent, then so are these axioms together with Zorn's lemma; and if the axioms of set theory are consistent, then so are these axioms together with the negation of Zorn's lemma.

Lemma 2.10 (Axiom of choice)**.** The Cartesian product of any non-empty collection of non-empty sets is non-empty. In other words, if I is any non-empty (indexing) set and A_i is a non-empty set for all $i \in I$, then there exists a choice function from I to $\bigcup_{i \in I} A_i$.

Lemma 2.11 (Well-ordering principle)**.** Every non-empty set A has a well-ordering.

Theorem 2.12. Assuming the usual (Zermelo–Fraenkel) axioms of set theory, the following are equivalent:

(i) Zorn's lemma

(ii) Axiom of choice

(iii) Well-ordering principle

Proof. This follows from elementary set theory. We refer the reader to *Real and Abstract Analysis* by Hewitt and Stromberg, Section 3. ☐

Equivalence Relations

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 2.13 (Equivalence relation)**.** A relation R on a set A is an **equivalence relation** if it is reflexive, symmetric and transitive.

Notation. We use $\sim$ to denote the equivalence relation R in $A \times A$: we denote $a \sim b$ for $(a, b) \in R$.

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

Definition 2.14 (Equivalence class)**.** Given an equivalence relation $\sim$ on a set A, and given $x \in A$, the **equivalence class** of x is

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

Proposition 2.15. Every two equivalence classes are disjoint; thus the union of equivalence classes form the entire set.

Remark. From the point of view of the elements, every element belongs to one and only one equivalence class.

Definition 2.16 (Quotient set)**.** The **quotient set** is 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 2.17 (Partition)**.** A **partition** of a set A is a collection of subsets $\{A_i \subset A \mid i \in I\}$, where I is an indexing set, with the property that

(i) $A_i \neq \emptyset$ for all $i \in I$ (all the subsets are non-empty)

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

(iii) $A_i \cap A_j = \emptyset$ for every $i \neq j$ (the subsets are disjoint)

The subsets are called the *parts* of the partition.

Example (Modular arithmetic)

Let n be a fixed positive integer. Define a relation on $\mathbf{Z}$ by

$$a \sim b \iff n \mid (b-a).$$

Proposition 2.18. $a \sim b$ is a equivalence relation.

Proof.

(i) $a \sim a$ so $\sim$ is reflexive.

(ii) $a \sim b \implies b \sim a$ for any integers a and b, so $\sim$ is symmetric.

(iii) If $a \sim b$ and $b \sim c$ then $n \mid (a-b)$ and $n \mid (b-c)$, so $n \mid (a-b)+(b-c) = (a-c)$, so $a \sim c$ and $\sim$ is transitive.

□

Notation. We write $a \equiv b \pmod{n}$ if $a \sim b$.

Notation. For any $k \in \mathbf{Z}$ we denote the equivalence class of a by $[a]$, called the *congruence class* (or *residue class*) of a mod n, which consists of the integers which differ from a by an integral multiple of n; that is,

$$[a] = \{a + kn \mid k \in \mathbf{Z}\}.$$

There are precisely n distinct congruence classes mod n, namely

$$[0], [1], \dots, [n-1],$$

determined by the possible remainders after division by n; and these residue classes partition the integers $\mathbf{Z}$. The set of equivalence classes under this equivalence relation is denoted by $\mathbf{Z}/n\mathbf{Z}$, and called the *integers modulo n*.

Define addition and multiplication on $\mathbf{Z}/n\mathbf{Z}$ as follows: for $[a], [b] \in \mathbf{Z}/n\mathbf{Z}$,

$$[a] + [b] = [a+b]$$
$$[a][b] = [ab].$$

This means that to compute the sum / product of two elements $[a], [b] \in \mathbf{Z}/n\mathbf{Z}$, take any *representative* $a \in [a]$, $b \in [b]$, and add / multiply integers a and b as usual in $\mathbf{Z}$, then take the congruence class containing the result.

Proposition 2.19. Addition and mulltiplication on $\mathbf{Z}/n\mathbf{Z}$ are well-defined; that is, they do not depend on the choices of representatives for the classes involved. More precisely, if $a_1, a_2 \in \mathbf{Z}$

and $b_1, b_2 \in \mathbf{Z}$ with $\overline{a_1} = \overline{b_1}$ and $\overline{a_2} = \overline{b_2}$, then $\overline{a_1 + a_2} = \overline{b_1 + b_2}$ and $\overline{a_1 a_2} = \overline{b_1 b_2}$, i.e., If

$$a_1 \equiv b_1 \pmod{n}, \quad a_2 \equiv b_2 \pmod{n}$$

then

$$a_1 + a_2 \equiv b_1 + b_2 \pmod{n}, \quad a_1 a_2 \equiv b_1 b_2 \pmod{n}.$$

Proof. Suppose $a_1 \equiv b_1 \pmod{n}$, i.e., $n \mid (a_1 - b_1)$. Then $a_1 = b_1 + sn$ for some integer s. Similarly, $a_2 \equiv b_2 \pmod{n}$ means $a_2 = b_2 + tn$ for some integer t.

Then $a_1 + a_2 = (b_1 + b_2) + (s + t)n$ so that $a_1 + a_2 \equiv b_1 + b_2 \pmod{n}$, which shows that the sum of the residue classes is independent of the representatives chosen.

Similarly, $a_1 a_2 = (b_1 + sn)(b_2 + tn) = b_1 b_2 + (b_1 t + b_2 s + stn)n$ shows that $a_1 a_2 \equiv b_1 b_2 \pmod{n}$ and so the product of the residue classes is also independent of the representatives chosen. □

An important subset of $\mathbf{Z}/n\mathbf{Z}$ consists of the collection of congruence classes which have a multiplicative inverse in $\mathbf{Z}/n\mathbf{Z}$:

$$(\mathbf{Z}/n\mathbf{Z})^\times := \{[a] \in \mathbf{Z}/n\mathbf{Z} \mid \exists [c] \in \mathbf{Z}/n\mathbf{Z}, [a][c] = [1]\}.$$

Proposition 2.20. $(\mathbf{Z}/n\mathbf{Z})^\times$ is also the collection of congruence classes whose representatives are relatively prime to n:

$$(\mathbf{Z}/n\mathbf{Z})^\times = \{[a] \in \mathbf{Z}/n\mathbf{Z} \mid (a, n) = 1\}.$$

§2.3 Functions

Definitions and Examples

Definition 2.21 (Function)**.** A **function** $f : X \to Y$ is a mapping of every element of X to some element of Y; X and Y are known as 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 $f : \mathbf{R} \to \mathbf{R}$, 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 2.22. Given a function $f : X \to Y$, the **image** (or *range*) of f is

$$f(X) := \{f(x) \mid x \in X\} \subset Y.$$

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

$$f(A) := \{f(x) \mid x \in A\} \subset Y.$$

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

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

Remark. Beware the potentially confusing notation: for $x \in X$, $f(x)$ is a single element of Y, but for $A \subset X$, $f(A)$ 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).

Exercise

Prove the following statements:

(i) $f(A \cup B) = f(A) \cup f(B)$

(ii) $f(A_1 \cup \cdots \cup A_n) = f(A_1) \cup \cdots \cup f(A_n)$

(iii) $f(\bigcup_{\lambda \in A} A_\lambda) = \bigcup_{\lambda \in A} f(A_\lambda)$

(iv) $f(A \cap B) \subset f(A) \cap f(B)$

(v) $f^{-1}(f(A)) \supset A$

(vi) $f(f^{-1}(A)) \subset A$

(vii) $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$

(viii) $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$

(ix) $f^{-1}(A_1 \cup \cdots \cup A_n) = f^{-1}(A_1) \cup \cdots \cup f^{-1}(A_n)$

(x) $f^{-1}(\bigcup_{\lambda \in A} A_\lambda) = \bigcup_{\lambda \in A} f^{-1}(A_\lambda)$

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

Definition 2.23 (Restriction)**.** Given a function $f : X \to Y$ and a subset $A \subset X$, the **restriction** of f to A is the map $f|_A : A \to Y$.

Remark. 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 2.24 (Identity map)**.** Given a set X, the **identity** $\mathrm{id}_X : X \to X$ is defined by

$$\mathrm{id}_X(x) = x \quad (\forall x \in X)$$

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

Injectivity, Surjectivity, Bijectivity

Definition 2.25. Let $f : X \to Y$ be a function.

(i) f is **injective** if each element of Y has at most one element of X that maps to it:

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

(ii) f is **surjective** if every element of Y is mapped to at least one element of X:

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

(iii) f is **bijective** if it is both injective and surjective; a bijective function is termed a *bijection*.

Notation. Given two sets X and Y, write $X \sim Y$ to denote the existence of a bijection from X to Y.

Proposition 2.26 ($\sim$ is transitive)**.** If $X \sim Y$ and $Y \sim Z$, then $X \sim Z$.

Theorem 2.27 (Cantor–Schroder–Bernstein)**.** If $f : A \to B$ and $g : B \to A$ are both injections, then $A \sim B$.

Proof. □

Composition

Definition 2.28 (Composition)**.** Given $f : X \to Y$ and $g : Y \to Z$, the **composition** $g \circ f : X \to Z$ is defined by

$$(g \circ f)(x) = g(f(x)) \quad (\forall x \in X)$$

The composition of functions is not commutative. However, composition is associative, as the following results shows:

Proposition 2.29 (Associativity of composition)**.** Let $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$. Then

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

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

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

□

Proposition 2.30 (Composition preserves injectivity and surjectivity)**.**

(i) If $f : X \to Y$ is injective and $g : Y \to Z$ is injective, then $g \circ f : X \to Z$ is injective.

(ii) If $f : X \to Y$ is surjective and $g : Y \to Z$ is surjective, then $g \circ f : X \to Z$ is surjective.

Proof.

(i) Let $f : X \to Y$ and $g : Y \to Z$ be injective. To prove that $g \circ f : X \to Z$ is injective, we need to prove: for all $x, x' \in X$,

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

Suppose that $(g \circ f)(x) = (g \circ f)(x')$. Then by definition

$$g\left(f(x)\right) = g\Big(f(x')\Big).$$

Injectivity of g implies

$$f(x) = f(x'),$$

and injectivity of f implies

$$x = x'.$$

(ii) Let $f : X \to Y$ and $g : Y \to Z$ be surjective. To prove that $g \circ f : X \to Z$ is surjective, we need to prove that for any $z \in Z$, there exists $x \in X$ such that $(g \circ f)(x) = z$.

Let $z \in Z$. By surjectivity of $g : Y \to Z$, there exists $y \in Y$ such that $g(y) = z$. By surjectivity of $f : X \to Y$, there exists $x \in X$ such that $f(x) = y$. This means that there exists $x \in X$ such that $g(f(x)) = g(y) = z$, as desired.

□

Proposition 2.31. $f : X \to Y$ is injective if and only if for any set Z and any functions $g_1, g_2 : Z \to X$,

$$f \circ g_1 = f \circ g_2 \implies g_1 = g_2.$$

Proof.

$\boxed{\implies}$ Suppose f is injective, and suppose $f \circ g_1 = f \circ g_2$. Let $z \in Z$. Then we have

$$f(g_1(z)) = f(g_2(z)).$$

Injectivity of f implies

$$g_1(z) = g_2(z),$$

so $g_1 = g_2$ (since the choice of $z \in Z$ is arbitrary).

$\boxed{\impliedby}$ Pick $Z = \{1\}$, basically some random one-element set. Then for $x, y \in X$, define

$$g_1 : Z \to X, \quad g_1(1) = x,$$
$$g_2 : Z \to Y, \quad g_2(1) = y.$$

Then for $x, y \in X$,

$$f(x) = f(y) \implies f(g_1(1)) = f(g_2(1)) \implies g_1(1) = g_2(1) \implies x = y$$

which shows that f is injective. □

Proposition 2.32. $f : X \to Y$ is surjective if and only if for any set Z and any functions $g_1, g_2 : Y \to Z$,

$$g_1 \circ f = g_2 \circ f \implies g_1 = g_2.$$

Proof.

$\boxed{\implies}$ Suppose that f is surjective. Let $y \in Y$. Surjectivity of f means there exists $x \in X$ such that $f(x) = y$. Then

$$g_1 \circ f = g_2 \circ f \implies g_1(f(x)) = g_2(f(x)) \implies g_1(y) = g_2(y)$$

so $g_1 = g_2$.

$\boxed{\impliedby}$ We prove the contrapositive. Suppose f is not surjective, then there exists $y \in Y$ such that for all $x \in X$ we have $f(x) \neq y$. We then aim to construct set Z and $g_1, g_2 : Y \to Z$ such that

(i) $g_1(y) \neq g_2(y)$

(ii) $\forall y' \neq y, g_1(y') = g_2(y')$

Because if this is satisfied, then $\forall x \in X$, since $f(x) \neq y$ we have from (ii) that $g_1(f(x)) = g_2(f(x))$; thus $g_1 \circ f = g_2 \circ f$, and yet from (i) we have $g_1 \neq g_2$.

We construct $Z = Y \cup \{1, 2\}$ for some random $1, 2 \notin Y$.

Then we define

$$g_1 : Y \to Z, g_1(y) = 1, g_1(y') = y' \qquad\qquad g_2 : Y \to Z, g_2(y) = 2, g_2(y') = y'$$

Then when y is not in the image of f, these two functions will satisfy $g_1 \circ f = g_2 \circ f$ but not $g_1 = g_2$.

So conversely, if for any set Z and any functions $g_i : Y \to Z$ we have $g_1 \circ f = g_2 \circ f \implies g_1 = g_2$, such a value y that is in the codomain but not in the range of f cannot appear, and hence f must be surjective. $\square$

Invertibility

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

Definition 2.33 (Invertibility)**.** A function $f : X \to Y$ is **invertible** if there exists $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$; g is known as the *inverse* of f.

Proposition 2.34 (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

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

$\square$

Since the inverse is unique, we can give it a notation.

Notation. The inverse of f is denoted by f^{-1}

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

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

Proposition 2.35 (Inverse of composition)**.** Let $f : X \to Y$, $g : Y \to Z$. 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:

Lemma 2.36 (Invertibility criterion)**.** $f : X \to Y$ is invertible if and only if it is bijective.

Proof.

$\boxed{\Longrightarrow}$ Suppose f is invertible, so it has an inverse $f^{-1} : Y \to X$. We first show f is injective. Suppose that $f(x_1) = f(x_2)$ for some $x_1, x_2 \in X$. Then applying f^{-1} to both sides and noting that $f^{-1} \circ f = \mathrm{id}_X$ by definition,

$$x_1 = f^{-1}(f(x_1)) = f^{-1}(f(x_2)) = x_2.$$

Thus f is injective.

We now 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.

$\boxed{\Longleftarrow}$ Suppose f is bijective. We want to show that there exists $g : Y \to X$ 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 2.37. 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.

Monotonicity

Definition 2.38 (Monotonicity)**.** $f : [a, b] \to \mathbf{R}$ is called

(i) **increasing**, if any $a < x_1 \leqslant x_2 < b$, there is $f(x_1) \leqslant f(x_2)$;

(ii) **decreasing**, if any $a < x_1 \leqslant x_2 < b$, there is $f(x_1) \geqslant f(x_2)$;

f is **monotonic** if it is increasing or decreasing.

Suppose $f(x)$ is continuous in $[a, b]$. To locate the roots of $f(x) = 0$:

- If $f(a)$ and $f(b)$ have *opposite* signs, i.e. $f(a)f(b) < 0$, then there is an odd number of real roots (counting repeated) in $[a, b]$.

  Furthermore, if f is either strictly increasing or decreasing in $[a, b]$, then $f(x) = 0$ has exactly one real root in $[a, b]$.

- If $f(a)$ and $f(b)$ have *same* signs, i.e. $f(a)f(b) > 0$, then there is an even number of roots (counting repeated) in $[a, b]$.

Definition 2.39 (Convexity)**.** A function f is **convex** if for all $x_1, x_2 \in D_f$ and $0 \leqslant t \leqslant 1$, we have

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2).$$

f is *strictly convex* if the $\leqslant$ sign above is replaced with a strict inequality $<$.

Similarly, f is **concave** if for all $x_1, x_2 \in D_f$ and $0 \leqslant t \leqslant 1$, we have

$$f(tx_1 + (1-t)x_2) \geqslant tf(x_1) + (1-t)f(x_2).$$

f is *strictly concave* if the $\geqslant$ sign above is replaced with a strict inequality $>$.

§2.4 Ordered Sets and Boundedness

Let S be a set.

Definition 2.40 (Order)**.** An **order** on S is a binary relation, denoted by $<$, with the following properties:

(i) Trichotomy: $\forall x, y \in S$, one and only one of the following statements is true:

$$x < y, \quad x = y, \quad y < x.$$

(ii) Transitivity: $\forall x, y, z \in S$, if $x < y$ and $y < z$, then $x < z$.

We call $(S, <)$ an **ordered set**.

Notation. $x \leqslant y$ indicates that $x < y$ or $x = y$, without specifying which of these two is to hold. In other words, $x \leqslant y$ is the negation of $x > y$.

Definition 2.41 (Boundedness)**.** Suppose S is an ordered set, and $E \subset S$.

- If there exists $\beta \in S$ such that $x \leqslant \beta$ for all $x \in E$, we say that E is **bounded above**, and call β an **upper bound** of E.
- If there exists $\beta \in S$ such that $x \geqslant \beta$ for all $x \in E$, we say that E is **bounded below**, and call β a **lower bound** of E.

E is **bounded** in S if it is bounded above and below.

Definition 2.42 (Supremum)**.** Suppose S is an ordered set, $E \subset S$, and E is bounded above. We call $\alpha \in S$ the **supremum** of E, denoted by $\alpha = \sup E$, if it satisfies the following properties:

(i) α is an upper bound for E;

(ii) if $\beta < \alpha$ then β is not an upper bound of E, i.e. $\exists x \in S$ s.t. $x > \beta$ (least upper bound).

Definition 2.43 (Infimum)**.** We cal $\alpha \in S$ the **infimum** of E, denoted by $\alpha = \inf E$, if it satisfies the following properties:

(i) α is a lower bound for E;

(ii) if $\beta > \alpha$ then β is not a lower bound of E, i.e. $\exists x \in S$ s.t. $x < \beta$ (greatest lower bound).

Proposition 2.44 (Uniqueness of suprenum)**.** If E has a supremum, then it is unique.

Proof. Assume that M and N are suprema of E.

Since N is a supremum, it is an upper bound for E. Since M is a supremum, then it is the least upper bound and thus $M \leqslant N$.

Similarly, since M is a supremum, it is an upper bound for E; since N is a supremum, it is a least upper bound and thus $N \leqslant M$.

Since $N \leqslant M$ and $M \leqslant N$, thus $M = N$. Therefore, a supremum for a set is unique if it exists. $\square$

Definition 2.45. An ordered set S is said to have the **least-upper-bound property** (l.u.b.) if the following is true: if non-empty $E \subset S$ is bounded above, then $\sup E \in S$.

Similarly, S has the greatest-lower-bound property if the following is true: if non-empty $E \subset S$ is bounded below, then $\inf E \in S$.

We shall now show that there is a close relation between greatest lower bounds and least upper bounds, and that every ordered set with the least-upper-bound property also has the greatest-lower-bound property.

Theorem 2.46. Suppose S is an ordered set. If S has the least-upper-bound property, then S has the greatest-lower-bound property.

Proof. Suppose $B \subset S$, $B \neq \emptyset$ is bounded below. We want to show that $\inf B \in S$. To do so, let $L \subset S$ be the set of all lower bounds of B; that is, $L = \{y \in S \mid y \leqslant x \forall x \in B\}$. If we can show that $\inf B = \sup L$, then we are done.

Since B is bounded below, $L \neq \emptyset$. Since every $x \in B$ is an upper bound of L, L is bounded above. Then since S has the least-upper-bound property, we have that $\sup L \in S$.

To show that $\sup L = \inf B$, we need to show that $\sup L$ is a lower bound of B, and $\sup L$ is the greatest of the lower bounds.

Suppose $\gamma < \sup L$, then γ is not an upper bound of L. Since B is the set of upper bounds of L, $\gamma \notin B$. Considering the contrapositive, if $\gamma \in B$, then $\gamma \geqslant \sup L$. Hence $\sup L$ is a lower bound of B, and thus $\sup L \in L$.

If $\sup L < \beta$ then $\beta \notin L$, since $\sup L$ is an upper bound of L. In other words, $\sup L$ is a lower bound of B, but β is not if $\beta > \sup L$. This means that $\sup L = \inf B$. $\square$

Corollary 2.47. If S has the greatest-lower-bound property, then it has the least-upper-bound property.

Hence S has the least-upper-bound property if and only if S has the greatest-lower-bound property.

Let's explore some useful properties of sup and inf.

Proposition 2.48 (Comparison theorem)**.** Let $S, T \subset \mathbf{R}$ be non-empty sets such that $s \leqslant t$ for every $s \in S$ and $t \in T$. If T has a supremum, then so does S, and $\sup S \leqslant \sup T$.

Proof. Let $\tau = \sup T$. Since τ is a supremum for T, then $t \leqslant \tau$ for all $t \in T$. Let $s \in S$ and choose any $t \in T$. Then, since $s \leqslant t$ and $t \leqslant \tau$, then $s \leqslant t$. Thus, τ is an upper bound for S.

By the Completeness Axiom, S has a supremum, say $\sigma = \sup S$. We will show that $\sigma \leqslant \tau$. Notice that, by the above, τ is an upper bound for S. Since σ is the least upper bound for S, then $\sigma \leqslant \tau$.

Therefore,

$$\sup S \leqslant \sup T.$$

□

Let's explore some useful properties of sup and inf.

Proposition 2.49. Let S, T be non-empty subsets of $\mathbf{R}$, with $S \subset T$ and with T bounded above. Then S is bounded above, and $\sup S \leqslant \sup T$.

Proof. Since T is bounded above, it has an upper bound, say b. Then $t \leqslant b$ for all $t \in T$, so certainly $t \leqslant b$ 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 $\sup T \geqslant \sup S$ (since $\sup S$ is the least upper bound for S). □

Proposition 2.50. Let $T \subset \mathbf{R}$ 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 $t \geqslant b$ for all $t \in T$. Then $-t \leqslant -b$ for all $t \in T$, so $s \leqslant -b$ 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 $s \leqslant \sup S$ for all $s \in S$, so $t \geqslant -\sup S$ 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 $-b \geqslant \sup S$ (since $\sup S$ is the least upper bound), so $b \leqslant -\sup S$. So $-\sup S$ is the greatest lower bound.

So $\inf T$ exists and $\inf T = -\sup S$. □

Proposition 2.51 (Approximation property)**.** Let $S \subset \mathbf{R}$ be non-empty and bounded above. For any $\varepsilon > 0$, there is $s_\varepsilon \in S$ such that $\sup S - \varepsilon < s_\varepsilon \leqslant \sup S$.

Proof. Take $\varepsilon > 0$.

Note that by definition of the supremum we have $s \leqslant \sup S$ for all $s \in S$. Suppose, for a contradiction, that $\sup S - \varepsilon \geqslant s$ 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 $\sup S - \varepsilon < s_\varepsilon$. □

If we are dealing with rational numbers, the sup/inf of a set may not exist. For example, a set of numbers in $\mathbf{Q}$, 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 $\mathbf{Q}$.

By ZFC, we have the Completeness Axiom, which states that any non-empty set $A \subset \mathbf{R}$ 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 $M \in \mathbf{R}$ such that $M = \sup A$.

Exercise

Consider the set

$$\left\{\frac{1}{n} \mid n \in \mathbf{Z}^{+}\right\}.$$

(a) Show that $\max S = 1$.

(b) Show that if d is a lower bound for S, then $d \leqslant 0$.

(c) Use (b) to show that $0 = \inf S$.

Exercise

Find, with proof, the supremum and/or infimum of $\{\frac{1}{n}\}$.

Solution. For the suprenum,

$$\sup\left\{\frac{1}{n}\right\} = \max\left\{\frac{1}{n}\right\} = 1.$$

For the infinum, for all positive a we can pick $n = [\frac{1}{a}] + 1$, then $a > \frac{1}{n}$. Hence

$$\inf\left\{\frac{1}{n}\right\} = 0.$$

□

Exercise

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

$$0 \leqslant \{(2\pi - 6) \cdot m(b-a)\} - \left(2 - \frac{\pi}{2}\right) < \delta$$

Picking $k = m(b-a)$ thus gives

$$\begin{aligned} 2k\pi + \frac{\pi}{2} &= 6k + (2\pi - 6)k + \frac{\pi}{2} \\ &= 6k + [(2\pi - 6)k] + 2 + (2\pi - 6)k - \left(2 - \frac{\pi}{2}\right) \end{aligned}$$

Thus $n = 6k + [(2\pi - 6)k] + 2$ satisfies $\left|2k\pi + \frac{\pi}{2} - n\right| < \delta$

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.)

$$\begin{aligned} \sin n &= \sin\left(n - \left(2k\pi + \frac{\pi}{2}\right) + \left(2k\pi + \frac{\pi}{2}\right)\right) \\ &= \cos\left(n - \left(2k\pi + \frac{\pi}{2}\right)\right) \\ &= \cos\theta \end{aligned}$$

$$1 - \sin n = 2\sin^2\frac{\theta}{2} = 2\sin^2\left|\frac{\theta}{2}\right| \leqslant \frac{\theta^2}{2} < \delta$$

Hence we simply pick $\delta = \varepsilon$ to ensure that $1 - \sin n < \varepsilon$, and we're done. $\square$

§2.5 Cardinality

This section is about formalising the notion of the "size" of a set.

Definition 2.52. Two sets A and B said to be **equivalent** (or have the same *cardinal number*), denoted by $A \sim B$, if there exists a bijection $f : A \to B$.

Notation. For $n \in \mathbf{Z}^+$, let

$$\begin{aligned} \mathbf{Z}^+ &= \{i \in \mathbf{Z} \mid i \geqslant 1\}, \\ \mathbf{Z}_n^+ &= \{i \in \mathbf{Z}^+ \mid 1 \leqslant i \leqslant n\}, \\ n\mathbf{Z}^+ &= \{ni \mid i \in \mathbf{Z}^+\}. \end{aligned}$$

Definition 2.53. For any set A, we say

- A is **finite** if $A \sim \mathbf{Z}_n^+$ for some integer n, the **cardinality** of A is $|A| = n$; A is **infinite** if A is not finite;
- A is **countable** if $A \sim \mathbf{Z}^+$; A is **uncountable** if A is neither finite nor countable; A is **at most countable** if A is finite or countable.

Finite Sets

For finite sets, we can do some arithmetic with their cardinalities.

Proposition 2.54 (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 $A = \emptyset$ has no elements, so there is a single subset $\emptyset$, and therefore $|\mathcal{P}(A)| = 1 = 2^0$.

Now suppose that $n \geqslant 0$ 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. $\square$

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.

Theorem 2.55 (Cantor's Theorem)**.** For a set A, finite or infinite,

$$|A| < |\mathcal{P}(A)|.$$

Proof. Suppose, for a contradiction, that $|A| = |\mathcal{P}(A)|$. Then there exists a bijection $f : A \to \mathcal{P}(A)$. Put

$$B = \{x \in A \mid x \notin f(A)\}.$$

Now consider any $x \in A$. In the first case, $x \in f(A)$, then

$$x \in f(A) \iff x \notin B,$$

thus $f(A) \neq B$. In the second case, $x \notin f(A)$, then

$$x \notin f(A) \iff x \in B,$$

thus $f(x) \neq B$. Hence f is not surjective, which is a contradiction. □

Corollary 2.56. For all $n \in \mathbf{Z}_0^+$,

$$n < 2^n.$$

Proof. This can be easily proven through induction. □

Proposition 2.57. Let A and B be finite sets. Then $|A \cup B| = |A| + |B| - |A \cap B|$.

Proof. The proof is left as an exercise. □

Theorem 2.58 (Principle of Inclusion and Exclusion)**.** Let S_i be finite sets. Then

$$\left|\bigcup_{i=1}^n S_i\right| = \sum_{i=1} |S_i| - \sum_{1\leqslant i<j\leqslant n} |S_i \cap S_j| + \sum_{1\leqslant i<j<k\leqslant n} |S_i \cap S_j \cap S_k| + \cdots + (-1)^{n+1} \left|\bigcap_{i=1}^n S_i\right|. \tag{2.1}$$

Proof. By induction. □

Alternative proof. Let U be a finite set (interpreted as the universal set), and $S \subset U$. Define the characteristic/indicator function of S by

$$\chi_S(x) = \begin{cases} 1 & (x \in S) \\ 0 & (x \notin S) \end{cases}$$

In other words,

$$x \in S \iff \chi_S(x) = 1$$

and equivalently,

$$x \notin S \iff \chi_S(x) = 0.$$

Let $S_1, S_2 \subset U$ be given. Then for any $x \in U$ it holds that

$$\chi_{S_1 \cap S_2}(x) = \chi_{S_1}(x) \cdot \chi_{S_2}(x)$$

where $\cdot$ denotes ordinary multiplication.

Similarly,

$$\chi_{S_1\cup S_2}(x) = 1 - (1-\chi_{S_1}(x))\cdot(1-\chi_{S_2}(x)).$$

In general, for any $x \in U$ it holds that

$$\chi_{S_1\cup\cdots\cup S_n}(x) = 1 - (1-\chi_{S_1}(x))\cdots(1-\chi_{S_n}(x))$$

for any $S_1,\ldots,S_n \subset U$.

Since $x \in S$ if and only if $\chi_S(x) = 1$, it follows that

$$|S| = \sum_{x\in U}\chi_S(x).$$

To prove the PIE, we calculate

$$\begin{aligned}
&|S_1\cup\cdots\cup S_n| \\
&= \sum_{x\in U}\chi_{S_1\cup\cdots\cup S_n}(x) \\
&= \sum_{x\in U} 1-(1-\chi_{S_1}(x))\cdots(1-\chi_{S_n}(x)) \\
&= (\chi_{S_1}(x)+\cdots+\chi_{S_n}(x)) - \Big(\chi_{S_1}(x)\chi_{S_2}(x)+\cdots+\chi_{S_{n-1}}(x)\chi_{S_n}(x)\Big)+\cdots+(-1)^{n+1}\chi_{S_1}(x)\cdots\chi_{S_n}(x) \\
&= (\chi_{S_1}(x)+\cdots+\chi_{S_n}(x)) - \Big(\chi_{S_1\cap S_2}(x)+\cdots+\chi_{S_{n-1}\cap S_n}(x)\Big)+\cdots+(-1)^{n+1}\chi_{S_1\cap\cdots\cap S_n}(x) \\
&= \sum_{i=1}^{n}|S_i| - \sum_{J\subset\{1,\ldots,n\},|J|=2}\left|\bigcap_{j\in J}S_j\right| + \cdots + (-1)^{k+1}\sum_{J\subset\{1,\ldots,n\},|J|=k}\left|\bigcap_{j\in J}S_j\right| + \cdots + (-1)^{n+1}\left|\bigcap_{i=1}^{n}S_i\right|.
\end{aligned}$$

□

Countability

For two finite sets A and B, we evidently have $A \sim B$ if and only if A and B contain the same number of elements. For infinite sets, however, the idea of "having the same number of elements" becomes quite vague, whereas the notion of bijectivity retains its clarity.

Proposition 2.59. $n\mathbf{Z}^+$ is countable.

Proof. Let $f : \mathbf{Z}^+ \to n\mathbf{Z}^+$ be given by

$$f(k) = nk.$$

For any $k_1, k_2 \in \mathbf{Z}^+$, $nk_1 = nk_2$ implies $k_1 = k_2$ so f is injective. For any $x \in n\mathbf{Z}^+$, $x = ni$ for some $i \in \mathbf{Z}^+$, thus $\frac{x}{n} = i \in \mathbf{Z}^+$ so f is surjective. Hence f is bijective, $n\mathbf{Z}^+ \sim \mathbf{Z}^+$ and we are done. □

Proposition 2.60. $\mathbf{Z}$ is countable.

Proof. Consider the following arrangement of the elements of $\mathbf{Z}$ and $\mathbf{Z}^+$:

$$\begin{aligned} \mathbf{Z}: &\quad 0, 1, -1, 2, -2, 3, -3, \dots \\ \mathbf{Z}^+: &\quad 1, 2, 3, 4, 5, 6, 7, \dots \end{aligned}$$

The function $f : \mathbf{Z}^+ \to \mathbf{Z}$ defined as

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

is bijective. $\square$

Proposition 2.61. Every infinite subset of a countable set A is countable.

Proof. Suppose $E \subset A$, and E is infinite. Arrange the elements $x \in A$ in a sequence $\{x_n\}$ of distinct elements. Construct a sequence $\{n_k\}$ as follows: Let n_1 be the smallest positive integer such that $x_{n_1} \in E$. Having chosen $n_1, \dots, n_{k-1}$ ($k = 2, 3, 4, \dots$), let n_k be the smallest integer greater than n_{k-1} such that $x_{n_k} \in E$.

Let $f : \mathbf{Z}^+ \to E$ be defined as

$$f(k) = x_{n_k},$$

which is bijective. Hence $E \sim \mathbf{Z}^+$, E is countable. $\square$

This shows that countable sets represent the "smallest" infinity: No uncountable set can be a subset of a countable set.

Proposition 2.62. Let (E_n) be a sequence of countable sets, put

$$S = \bigcup_{n=1}^{\infty} E_n.$$

Then S is countable.

Proof. Let every set E_n be arranged in a sequence $\{x_{n_k}\}$ ($k = 1, 2, 3, \dots$), and consider the infinite array

$$\begin{array}{ccccc} x_{11} & x_{12} & x_{13} & x_{14} & \cdots \\ x_{21} & x_{22} & x_{23} & x_{24} & \cdots \\ x_{31} & x_{32} & x_{33} & x_{34} & \cdots \\ x_{41} & x_{42} & x_{43} & x_{44} & \cdots \\ \vdots & & & & \end{array}$$

in which the elements of E_n form the n-th row. The array contains all elements of S. These elements can be arranged in a sequence

$$x_{11}, x_{21}, x_{12}, x_{31}, x_{22}, x_{13}, x_{41}, x_{32}, x_{23}, x_{14}, \ldots$$

If any two of the sets En have elements in common, these will appear more than once in (17). Hence there is a subset T of the set of all positive integers such that S T, which shows that S is at most countable (Theorem 2.8). Since $E_1 \subset S$, and E_1 is infinite, S is infinite, and thus countable. □

Corollary 2.63. Suppose A is at most countable, and, for every $\alpha \in A$, B_α is at most countable. Put

$$T = \bigcup_{\alpha \in A} B_\alpha.$$

Then T is at most coutable.

Proposition 2.64. Let A be a countable set, and let B_n be the set of all n-tuples $(a_1, \ldots, a_n)$, where $a_i \in A$. Then B_n is countable.

Corollary 2.65. $\mathbf{Q}$ is countable.

Proposition 2.66. The set of all algebraic numbers is countable. (Exercise 2)

Proposition 2.67. Let A be the set of all sequences whose elements are the digits 0 and 1. This set A is uncountable.

The idea of the above proof was first used by Cantor, and is called Cantor's diagonal process; for, if the sequences $s_1, s_2, s_3, \ldots$ are placed in an array like (16), it is the elements on the diagonal which are involved in the construction of the new sequence.

Corollary 2.68. $\mathbf{R}$ is uncountable.

Proof. This follows from the binary representation of the real numbers. □

Infinite Sets

A consequence of Cantor's Theorem (Theorem 2.55) is that there is no largest infinity. Since there are an infinite number of different sizes of infinity, it makes sense for us to order them from smallest onwards.

The **aleph numbers** are a sequence of numbers used to represent the cardinality of infinite sets, given by

$$\aleph_0, \aleph_1, \aleph_2, \aleph_3, \ldots,$$

where $\aleph_0 = |\mathbf{N}|$.

Another set of infinite cardinals is the set of **beth numbers**,

$$\beth_0, \beth_1, \beth_2, \beth_3, \ldots,$$

where

$$\begin{aligned}\beth_0 &= \aleph_0,\\ \beth_1 &= 2^{\aleph_0} = |\mathcal{P}(\mathbf{N})| = |\mathbf{R}|,\end{aligned}$$

and in general, for all n, we can recursively define

$$\beth_{n+1} = 2^{\beth_n}.$$

A natural question to ask is if the aleph numbers and beth numbers line up. We have defined $\beth_0 = \aleph_0$, but is $\beth_1 = \aleph_1$? Another way to ask this question is whether

$$|\mathcal{P}(\mathbf{N})| = |\mathbf{R}|.$$

This is called the **continuum hypothesis**. In fact it has been show that the continuum hypothesis can neither be proved nor disproved using the standard ZFC set theory axioms. The generalised continuum hypothesis is as follows:

$$\begin{aligned}2^{\aleph_n} &= \aleph_{n+1} \quad (\forall n)\\ \aleph_n &= \beth_n \quad (\forall n)\end{aligned}$$

A restatement of the above is that there is no set S such that

$$\aleph_n < |S| < 2^{\aleph_n}.$$

Exercises

Problem 2.1. 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 $Y \in \mathbf{C}^n$ 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 $X \cap Y = Z(S \cup T)$. □

Problem 2.2. Let $A = \mathbf{R}$ and for any $x, y \in A$, $x \sim y$ if and only if $x - y \in \mathbf{Z}$. 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 $\phi : X \to Y$ between $X = A/\sim$ and $Y : |z| = 1$, i.e. the unit circle in $\mathbf{C}$, such that for any $[x_1], [x_2] \in X$ we have

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

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

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

Solution.

(a)

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

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

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

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

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

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

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

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

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

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

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

□

Problem 2.3 (Complex Numbers)**.** Let $\mathbf{R}[x]$ denote the set of real polynomials. Define

$$\mathbf{C} = \mathbf{R}[x]/(x^2+1)\mathbf{R}[x]$$

where

$$f(x) \sim g(x) \iff x^2+1 \text{ divides } f(x)-g(x).$$

The complex number $a+bi$ is defined to be the equivalence class of $a+bx$.

(a) Define the sum and product of two complex numbers and show that such definitions are well-defined.

(b) Define the reciprocal of a complex number.

I

Abstract Algebra

The following is an excerpt from [Pin10]:

Thus, we are led to the modern notion of algebraic structure. An *algebraic structure* is understood to be an arbitrary set, with one or more operations defined on it. And algebra, then, is defined to be *the study of algebraic structures.*

It is important that we be awakened to the full generality of the notion of algebraic structure. We must make an effort to discard all our preconceived notions of what an algebra is, and look at this new notion of algebraic structure in its naked simplicity. *Any* set, with a rule (or rules) for combining its elements, is already an algebraic structure. There does not need to be any connection with known mathematics. For example, consider the set of all colors (pure colors as well as color combinations), and the operation of mixing any two colors to produce a new color. This may be conceived as an algebraic structure. It obeys certain rules, such as the commutative law (mixing red and blue is the same as mixing blue and red).

3 Groups

§3.1 Introduction to Groups

Definitions and Examples

Definition 3.1 (Binary operation)**.** A **binary operation** $*$ on a set G is a function $*: G \times G \to G$. For any $a, b \in G$, we write $a * b$ for the image of (a, b) under $*$.

$*$ is **associative** on G if $(a * b) * c = a * (b * c)$ for all $a, b, c \in G$.

$*$ is **commutative** on G if $a * b = b * a$ for all $a, b \in G$.

Definition 3.2 (Group)**.** A **group** $(G, *)$ consists of a set G and a binary operation $*$ on G satisfying the following group axioms:

(i) Associativity: $a * (b * c) = (a * b) * c$ for all $a, b, c \in G$.

(ii) Identity: there exists identity element $e \in G$ such that $a * e = e * a = a$ for all $a \in G$.

(iii) Invertibility: for all $a \in G$, there exists inverse $c \in G$ such that $a * c = c * a = e$.

G is **abelian**[1] if the operation is commutative; it is **non-abelian** if otherwise.

Remark. When verifying that $(G, *)$ is a group we have to check (i), (ii), (iii) above and also that $*$ is a binary operation – that is, $a * b \in G$ for all $a, b \in G$; this is sometimes referred to as closure.

Notation. We simply denote a group $(G, *)$ by G if the operation is clear.

Notation. We abbreviate $a * b$ to just ab if the operation is clear.

Notation. Since the operation $*$ is associative, we can omit unnecessary parentheses and write $(ab)c = a(bc) = abc$.

Notation. For any $a \in G$, $n \in \mathbf{Z}^+$ we abbreviate $a^n = \underbrace{a \cdots a}_{n \text{ times}}$.

Notation. We write $(\mathbf{Z}, +)$, $(\mathbf{Q}, +)$, $(\mathbf{R}, +)$, $(\mathbf{C}, +)$ as simply $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$.

> **Example**
>
> The following are some examples of groups.
>
> - $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$ are groups, with identity 0 and (additive) inverse $-a$ for all a.
> - $\mathbf{Q} \setminus \{0\}$, $\mathbf{R} \setminus \{0\}$, $\mathbf{C} \setminus \{0\}$, $\mathbf{Q}^+$, $\mathbf{R}^+$ are groups under $\times$, with identity 1 and (multiplicative)

[1] after the Norwegian mathematician Niels Abel (1802-1829)

inverse $\frac{1}{a}$ for all a; $\mathbf{Z} \setminus \{0\}$ is not a group under $\times$, because all elements except for ± 1 do not have an inverse in $\mathbf{Z} \setminus \{0\}$.

- For $n \in \mathbf{Z}^+$, $\mathbf{Z}/n\mathbf{Z}$ is an abelian group under $+$.
- For $n \in \mathbf{Z}^+$, $(\mathbf{Z}/n\mathbf{Z})^\times$ is an abelian group under multiplication.

Definition 3.3 (Product group)**.** Let $(G, *_G)$ and $(H, *_H)$ be groups. Then the operation $*$ is defined on $G \times H$ by

$$(g_1, h_1) * (g_2, h_2) = (g_1 *_G g_2, h_1 *_H h_2)$$

for all $g_1, g_2 \in G$, $h_1, h_2 \in H$. $(G \times H, *)$ is called the **product group** of G and H.

Proposition 3.4. The product group is a group.

Proof.

(i) Since $*_G$ and $*_H$ are both associative binary operations, it follows that $*$ is also an associative binary operation on $G \times H$.

(ii) We also note

$$e_{G\times H} = (e_G, e_H), \quad (g, h)^{-1} = (g^{-1}, h^{-1})$$

as for any $g \in G$, $h \in H$,

$$(e_G, e_H) * (g, h) = (g, h) = (g, h) * (e_G, e_H).$$

(iii) As for identity,

$$(g^{-1}, h^{-1}) * (g, h) = (e_G, e_H) = (g, h) * (g^{-1}, h^{-1}).$$

$\square$

Proposition 3.5. Let G be a group. Then

(i) the identity of G is unique,

(ii) for each $a \in G$, a^{-1} is unique,

(iii) $(a^{-1})^{-1} = a$ for all $a \in G$,

(iv) $(ab)^{-1} = b^{-1}a^{-1}$,

(v) for any $a_1, \dots, a_n \in G$, $a_1 \cdots a_n$ is independent of how we arrange the parantheses (generalised associative law).

Proof.

(i) Suppose otherwise, then e and e' are identites of G. We have

$$e = ee' = e'$$

where the first equality holds as e' is an identity, and the second equality holds as e is an identity. Since $e = e'$, the identity is unique.

(ii) Suppose otherwise, then b and c are both inverses of a. Let e be the identity of G. Then $ab = e$, $ca = e$. Thus
$$c = ce = c(ab) = (ca)b = eb = b.$$
Hence the inverse is unique.

(iii) To show $(a^{-1})^{-1} = a$ is exactly the problem of showing that a is the inverse of a^{-1}, which is by definition of the inverse (with the roles of a and a^{-1} interchanged).

(iv) Let $c = (ab)^{-1}$. Then $(ab)c = e$, or $a(bc) = e$ by associativity, which gives $bc = a^{-1}$ and thus $c = b^{-1}a^{-1}$ by multiplying b^{-1} on both sides.

(v) The result is trivial for $n = 1, 2, 3$. For all $k < n$ assume that any $a_1 \cdots a_k$ is independent of parantheses. Then
$$(a_1 \cdots a_n) = (a_1 \cdots a_k)(a_{k+1} \cdots a_n).$$
Then by assumption both are independent of parentheses since $k, n - k < n$ so by induction we are done.

□

Notation. Since the inverse is unique, we denote the inverse of $a \in G$ by a^{-1}.

Proposition 3.6 (Cancellation law)**.** Let $a, b \in G$. Then the equations $ax = b$ and $ya = b$ have unique solutions for $x, y \in G$. In particular, we can cancel on the left and right.

Proof. That $x = a^{-1}b$ is unique follows from the uniqueness of a^{-1} and the same for $y = ba^{-1}$. □

Definition 3.7 (Order of a group)**.** The cardinality $|G|$ of a group G is called the **order** of G. We say that a group G is finite if $|G|$ is finite.

One way to represent a finite group is by means of the group table or Cayley table[2]. Let $G = \{e, g_2, g_3, \ldots, g_n\}$ be a finite group. The Cayley table (or group table) of G is a square grid which contains all the possible products of two elements from G. The product $g_i g_j$ appears in the i-th row and j-th column of the Cayley table.

Remark. Note that a group is abelian if and only if its Cayley table is symmetric about the main (top-left to bottom-right) diagonal.

Example (Dihedral groups)

An important family of groups is the **dihedral groups**. For $n \in \mathbf{Z}^+$, $n \geqslant 3$, let D_{2n} be the set of symmetries[a] of a regular n-gon.

Remark. Here "D" stands for "dihedral", meaning two-sided.

To visualise this, we first choose a labelling of the n vertices. Then each symmetry S can be

[2] after the English mathematician Arthur Cayley (1821 – 1895)

described uniquely by the corresponding permutation σ of $\{1, 2, \ldots, n\}$ where if the symmetry s puts vertex i in the place where vertex j was originally, then σ is the permutation sending i to j.

We now make D_{2n} into a group. For $S, T \in D_{2n}$, define the binary operation ST to be the symmetry obtained by first applying T then S to the n-gon (this is analagous to function composition). If S and T effect the permutations σ and τ respectively on the vertices, then ST effects $\sigma \circ \tau$.

(i) The binary operation on D_{2n} is associative since the composition of functions is associative.

(ii) The identity of D_{2n} is the identity symmetry, which leaves all vertices fixed, denoted by 1.

(iii) The inverse of $S \in D_{2n}$ is the symmetry which reverses all rigid motions of S (so if S effects permutation σ on the vertices, S^{-1} effects σ^{-1}).

Let r be the rotation clockwise about the origin by $\frac{2\pi}{n}$ radians, let s be the reflection about the line of symmetry through the first labelled vertex and the origin.

Proposition 3.8.

(i) $|r| = n$

(ii) $|s| = 2$

(iii) $s \neq r^i$ for all i

(iv) $sr^i \neq sr^j$ for all $i \neq j$ $(0 \leqslant i, j \leqslant n-1)$, so

$$D_{2n} = \{1, r, \ldots, r^{n-1}, s, sr, \ldots, sr^{n-1}\}$$

and thus $|D_{2n}| = 2n$.

(v) $rs = sr^{-1}$

(vi) $r^i s = sr^{-i}$

Proof.

(i) It is obvious that $1, r, r^2, \ldots, r^{n-1}$ are all distinct and $r^n = 1$, so $|r| = n$.

(ii) This is fairly obvious: either reflect or do not reflect.

(iii) This is also obvious: the effect of any reflection cannot be obtained from any form of rotation.

(iv) Just cancel on the left and use the fact that $|r| = n$. We assume that $i \not\equiv j \pmod{n}$.

(v) Omitted.

(vi) By (5), this is true for $i = 1$. Assume it holds for $k < n$. Then $r^{k+1}s = r(r^k s) = rsr^{-k}$. Then $rs = sr^{-1}$ so $rsr^{-k} = sr^{-1}r^{-k} = sr^{-k-1}$ so we are done.

□

A presentation for the dihedral group D_{2n} using generators and relations is

$$D_{2n} = \langle r, s \mid r^n = s^2 = 1, rs = sr^{-1} \rangle.$$

[a] a symmetry is any rigid motion of the n-gon which can be effected by taking a copy of the n-gon, moving this copy in any fashion in 3-space and then placing the copy back on the original n-gon so it exactly covers it. A symmetry can be a reflection or a rotation.

Example (Permutation groups)

Let S be a non-empty set. A bijection $S \to S$ is called a **permutation** of S; the set of permutations of S is denoted by $\mathrm{Sym}(S)$.

$\mathrm{Sym}(S)$ is a group under function composition $\circ$. We show that the group axioms hold for $(\mathrm{Sym}(S), \circ)$:

(i) $\circ$ is a binary operation on $\mathrm{Sym}(S)$ since if $\sigma : S \to S$ and $\tau : S \to S$ are both bijections, then $\sigma \circ \tau$ is also a bijection from S to S.

(ii) Since function composition is associative in general, $\circ$ is associative.

(iii) The identity of $\mathrm{Sym}(S)$ is 1, defined by $1(a) = a$ for all $a \in S$.

(iv) For every permutation σ, there is a (2-sided) inverse function $\sigma^{-1} : S \to S$ satisfying $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = 1$.

$(\mathrm{Sym}(S), \circ)$ is called the **symmetric group** on S. In the special case where $S = \{1, 2, \ldots, n\}$, the symmetric group on S is denoted S_n, the symmetric group of degree n.

Proposition 3.9. If $|S| \geqslant 3$ then $\mathrm{Sym}(S)$ is non-abelian.

Proof. Let $S = \{x_1, x_2, x_3\}$ where three elements are distinct. □

Proposition 3.10. The order of S_n is $n!$.

Proof. Obvious, since there are $n!$ permutations of $\{1, 2, \ldots, n\}$. □

Example (Matrix groups)

A field is denoted by $\mathbf{F}$; $\mathbf{F}^\times = \mathbf{F} \setminus \{0\}$.

For $n \in \mathbf{Z}^+$, let $GL_n(\mathbf{F})$ be the set of all $n \times n$ invertible matrices whose entries are in $\mathbf{F}$:

$$GL_n(\mathbf{F}) = \{A \mid A \in M_{n\times n}(\mathbf{F}), \det(A) \neq 0\}.$$

We show that $GL_n(\mathbf{F})$ is a group under matrix multiplication:

(i) Since $\det(AB) = \det(A) \cdot \det(B)$, it follows that if $\det(A) \neq 0$ and $\det(B) \neq 0$, then $\det(AB) \neq 0$, so $GL_n(\mathbf{F})$ is closed under matrix multiplication.

(ii) Matrix multiplication is associative.

(iii) $\det(A) \neq 0$ if and only if A has an inverse matrix, so each $A \in GL_n(\mathbf{F})$ has an inverse $A^{-1} \in GL_n(\mathbf{F})$ such that
$$AA^{-1} = A^{-1}A = I$$
where I is the $n \times n$ identity matrix.

We call $GL_n(\mathbf{F})$ the **general linear group** of degree n.

Example (Quaternion group)

The **Quaternion group** Q_8 is defined by
$$Q_8 = \{1, -1, i, -i, j, -j, k, -k\}$$
with product $\cdot$ computed as follows:

- $1 \cdot a = a \cdot 1 = a$ for all $a \in Q_8$
- $(-1) \cdot (-1) = 1$
- $(-1) \cdot a = a \cdot (-1) = -a$ for all $a \in Q_8$
- $i \cdot i = j \cdot j = k \cdot k = -1$
- $i \cdot j = k$, $j \cdot i = -k$, $j \cdot k = i$, $k \cdot j = -i$, $k \cdot i = j$, $i \cdot k = -j$

Note that Q_8 is a non-abelian group of order 8.

An important (if rather elementary) family of groups is the *cyclic groups.*

Definition 3.11 (Cyclic group)**.** A group G is called **cyclic** if there exists $g \in G$ such that
$$G = \{g^k \mid k \in \mathbf{Z}\}.$$
Then g is called a **generator** of G.

Notation. If G is generated by x, we write $G = \langle x \rangle$.

Remark. A cyclic group may have more than one generator. For example, if $G = \langle x \rangle$, then also $G = \langle x^{-1} \rangle$ because $(x^{-1})^n = x^{-n} \in G$ for $n \in \mathbf{Z}$ so does $-n$, so that
$$\{x^n \mid n \in \mathbf{Z}\} = \{(x^{-1})^n \mid n \in \mathbf{Z}\}.$$

> **Example**
>
> $\mathbf{Z}$ is a cyclic group with generators 1 and -1.

Proposition 3.12. Cyclic groups are abelian.

Proof. Let G be a cyclic group. For $g^i, g^j \in G$, by the laws of exponents,

$$g^i g^j = g^{i+j} = g^j g^i.$$

□

Proposition 3.13. If $G = \langle x \rangle$, then $|G| = |x|$ (where if one side of this equality is infinite, so is the other):

(i) if $|G| = n < \infty$, then $x^n = 1$ and $1, x, x^2, \ldots, x^{n-1}$ are all the distinct elements of G;

(ii) if $|G| = \infty$, then $x^n \neq 1$ for all $n \neq 0$, and $x^a \neq x^b$ for all $a, b \in \mathbf{Z}$, $a \neq b$.

Proposition 3.14. Let G be an arbitrary group, $x \in G$ and let $m, n \in \mathbf{Z}$. If $x^n = 1$ and $x^m = 1$, then $x^d = 1$ where $d = \gcd(m, n)$. In particular, if $x^m = 1$ for some $m \in \mathbf{Z}$, then $|x|$ divides m.

Theorem 3.15. Any two cyclic groups of the same order are isomorphic:

(i) if $n \in \mathbf{Z}^+$ and $\langle x \rangle$ and $\langle y \rangle$ are both cyclic groups of order n, then the map $\phi : \langle x \rangle \to \langle y \rangle$ which maps $x^k \mapsto y^k$ is well-defined and is an isomorphism.

(ii) if $\langle x \rangle$ is an infinite cyclic group, the map $\phi : \mathbf{Z} \to \langle x \rangle$ which maps $k \mapsto x^k$ is well-defined and is an isomorphism.

Notation. For each $n \in \mathbf{Z}^+$, C_n denotes the cyclic group of order n:

$$C_n = \{e, g, g^2, \ldots, g^{n-1}\}$$

which satisfy $g^n = e$. Thus given two elements in C_n, we define

$$g^i * g^j = \begin{cases} g^{i+j} & (0 \leqslant i + j < n) \\ g^{i+j-n} & (n \leqslant i + j \leqslant 2n - 2) \end{cases}$$

Subgroups

Definition 3.16 (Subgroup)**.** Let G be a group. $H \subset G$, $H \neq \emptyset$ is a **subgroup** of G, denoted $H \leqslant G$, if the group operation $*$ restricts to make a group of H; that is,

(i) $e \in H$;

(ii) $xy \in H$ for all $x, y \in H$;

(iii) $x^{-1} \in H$ for all $x \in H$.

Remark. Observe that if $*$ is an associative (respectively, commutative) binary operation on G and $*$ is restricted to some $H \subset G$ is a binary operation on H, then $*$ is automatically associative (respectively, commutative) on H as well.

Lemma 3.17 (Subgroup criterion)**.** Let G be a group. $H \subset G$, $H \neq \emptyset$ is a subgroup of G if and only if $xy^{-1} \in H$ for all $x, y \in H$.

Furthermore, if H is finite, then it suffices to check that H is non-empty and closed under multiplication.

Proof. If H is a subgroup of G, then we are done, by definition of subgroup.

Conversely, we want to prove that for $H \neq \emptyset$, if $xy^{-1} \in H$ for all $x, y \in H$, then $H \leqslant G$:

(i) Since $H \neq \emptyset$, take $x \in H$, let $y = x$, then $1 = xx^{-1} \in H$, so H contains the identity of G.

(ii) Since $1 \in H$, $x \in H$, then $x^{-1} \in H$ so H is closed under taking inverses.

(iii) For any $x, y \in H$, $x, y^{-1} \in H$, so by (ii), $x(y^{-1})^{-1} = xy \in H$, so H is closed under multiplication.

Hence H is a subgroup of G.

For the last part, suppose that H is finite and closed under multiplication. Take $x \in H$. Then there are only finitely many distinct elements among $x, x^2, x^3, \ldots$ and so $x^a = x^b$ for $a, b \in \mathbf{Z}$ with $a < b$. If $n = b - a$, then $x^n = 1$ so in particular every element $x \in H$ is of finite order. Then $x^{n-1} \in x^{-1} \in H$, so H is closed under inverses. $\square$

We now introduce some important families of subgroups of an arbitrary group G. Let $A \subset G$, $A \neq \emptyset$.

Example (Centraliser)

The **centraliser** of A in G is defined by

$$C_G(A) := \{g \in G \mid \forall a \in A, gag^{-1} = a\}.$$

Since $gag^{-1} = a$ if and only if $ga = ag$, $C_G(A)$ is the set of elements of G which commute with every element of A.

Proposition 3.18. $C_G(A)$ is a subgroup of G.

Notation. In the special case when $A = \{a\}$ we simply write $C_G(a)$ instead of $C_G(\{a\})$. In this case $a^n \in C_G(a)$ for all $n \in \mathbf{Z}$.

Example (Center)

The **center** of G is the set of elements commuting with all the elements of G:

$$Z(G) := \{g \in G \mid \forall x \in G, gx = xg\}.$$

Proposition 3.19. $Z(G)$ is a subgroup of G.

Proof. Note that $Z(G) = C_G(G)$, so the argument above proves $Z(G) \leqslant G$ as a special case. $\square$

Example (Normaliser)

Define $gAg^{-1} = \{gag^{-1} \mid a \in A\}$. The **normaliser** of A in G is

$$N_G(A) := \{g \in G \mid gAg^{-1} = A\}.$$

Proposition 3.20. $N_G(A)$ is a subgroup of G.

Proof. Notice that if $g \in C_G(A)$, then $gag^{-1} = a \in A$ for all $a \in A$ so $C_G(A) \leqslant N_G(A)$. $\square$

The fact that the normaliser of A in G, the centraliser of A in G, and the center of G are all subgroups are special cases of results on group actions.

Example (Stabiliser)

If G is a group acting on a set S, $s \in S$, then the **stabiliser** of s in G is

$$G_s := \{g \in G \mid g \cdot s = s\}.$$

Proposition 3.21. G_s is a subgroup of G.

Cosets

Definition 3.22 (Order)**.** Let G be a group, $g \in G$. If there is a positive integer k such that $g^k = e$, then the **order** of g is defined as

$$o(g) := \min\{m > 0 \mid g^m = e\}.$$

Otherwise we say that the order of g is infinite.

Example

Some examples to illustrate the above concept.

- An element of a group has order 1 if and only if it is the identity.
- In the additive groups $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$, every non-zero (i.e. non-identity) element has infinite order.
- In the multiplicative groups $\mathbf{R} \setminus \{0\}$ or $\mathbf{Q} \setminus \{0\}$, the element -1 has order 2 and all other non-identity elements have infinite order.
- In $\mathbf{Z}/9\mathbf{Z}$, the element $\overline{6}$ has order 3. (Recall that in an additive group, the powers of an

element are integer multiples of the element.)

- In $(\mathbf{Z}/7\mathbf{Z})^\times$, the powers of the element $\overline{2}$ are $\overline{2}, \overline{4}, \overline{8} = \overline{1}$, the identity in this group, so 2 has order 3. Similarly, the element $\overline{3}$ has order 6, since 3^6 is the smallest positive power of 3 that is congruent to 1 mod 7.

Proposition 3.23. If G is finite, then $o(g)$ is finite for each $g \in G$.

Proof. Consider the list

$$g, g^2, g^3, g^4, \dots \in G.$$

As G is finite, then this list must have repeats. Hence there are integers $i > j$ such that $g^i = g^j$. So $g^{i-j} = e$ showing that $\{m > 0 \mid g^m = e\}$ is non-empty and so has a minimal element. □

Proposition 3.24. If $g \in G$ and $o(g)$ is finite, then $g^n = e$ if and only if $o(g) \mid n$.

Proof. If $n = ko(g)$ then

$$g^n = \left(g^{o(g)}\right)^k = e^k = e.$$

Conversely, if $g^n = e$ then, by the division algorithm, there are integers q, r such that $n = qo(g) + r$ where $0 \leqslant r < o(g)$. Then

$$g^r = g^{n-qo(g)} = g^n \left(g^{o(g)}\right)^{-q} = e.$$

By the minimality of $o(g)$ then $r = 0$ and so $n = qo(g)$. □

Proposition 3.25. If $\phi : G \to H$ is an isomorphism and $g \in G$ then $o(\phi(g)) = o(g)$.

Proof. We have

$$(\phi(g))^k = e_H \iff \phi(g^k) = e_H \iff g^k = e_G$$

as ϕ is injective. □

We introduce left cosets and right cosets of a subgroup.

Definition 3.26 (Coset)**.** Let $H \leqslant G$. For $g \in G$, a **left coset** of H in G is

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

Similarly, for $g \in G$, a **right coset** of H in G is

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

Any element of a coset is called a **representative** for the coset.

The set of left cosets is given by

$$(G/H)_l := \{gH \mid g \in G\}.$$

Similarly, the set of right cosets is given by

$$(G/H)_r := \{Hg \mid g \in G\}.$$

Proposition 3.27. Let $H \leqslant G$. Given $g, g' \in G$, two (left) cosets gH and $g'H$ are either disjoint or equal; that is, $(G/H)_l$ form a partition of G.

Proof. We want to prove: if the cosets gH and $g'H$ have an element in common, then they are equal.

Suppose $gh = g'h'$ for some $h, h' \in H$. Then $g = g'h'h^{-1}$. But $h'h^{-1} \in H$, so $gH = g'(h'h^{-1})H = g'H$, since $h'h^{-1}H = H$. $\square$

The following result shows that H partitions G into equal-sized parts.

Lemma 3.28. The cosets of H in G are the same size as H; that is, for all $a \in G$, $|aH| = |H|$.

Proof. Let $f : H \to aH$ which sends $h \mapsto ah$. For $h_1, h_2 \in H$,

$$\begin{aligned} f(h_1) = f(h_2) &\implies ah_1 = ah_2 \\ &\implies a^{-1}ah_1 = a^{-1}ah_2 \\ &\implies h_1 = h_2 \end{aligned}$$

thus f is an injective mapping. Note that f is surjective by the definition of aH. Since f is bijective, $|H| = |aH|$. $\square$

An important result relating the order of a group with the orders of its subgroups is Lagrange's theorem.

Theorem 3.29 (Lagrange's theorem)**.** If G is a finite group, $H \leqslant G$, then $|H|$ divides $|G|$, and the number of left cosets of H in G equals $\frac{|G|}{|H|}$.

Proof. Since $|G| < \infty$, let

$$(G/H)_l = \{a_1H, a_2H, \ldots, a_nH\}.$$

Since G is the disjoint union of $a_1H, \ldots, a_nH$, we have that

$$\begin{aligned} |G| &= \sum_{i=1}^{n} |a_iH| \\ &= \sum_{i=1}^{n} |H| \\ &= nH. \end{aligned}$$

Thus

$$n = \frac{|G|}{|H|} \in \mathbf{N}$$

as desired. $\square$

We call $|G : H| := \frac{|G|}{|H|}$ the **index** of H in G.

Theorem 3.30 (Fermat's little theorem)**.** For every finite group G, for all $a \in G$, $a^{|G|} = e$.

Proof. Consider the subgroup H generated by a; that is,

$$H = \{a^i \mid i \in \mathbf{Z}\}.$$

Since G is finite and $|H| < |G|$, H must be finite, so the infinite sequence $a^0 = e, a^1, a^2, a^3, \ldots$ must repeat, say $a^i = a^j$ $(i < j)$. Let $k = j - i$. Multiplying both sides by $a^{-i} = \left(a^{-1}\right)^i$, we get $a^{j-i} = a^k = e$. 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

$$a^{|G|} = (a^k)^{\frac{|G|}{k}} = e.$$

□

Theorem 3.31 (Fermat–Euler Theorem (or Euler's totient theorem))**.** If a and N are coprime, then $a^{\phi(N)} \equiv 1 \pmod{N}$, where ϕ is Euler's totient function.

§3.2 Homomorphisms and Isomorphisms

In this section, we make precise the notion of when two groups "look the same"; that is, they have the same group-theoretic structure. This is the notion of an *isomorpism* between two groups.

Definitions

Definition 3.32 (Homomorphism)**.** Let $(G, *)$ and $(H, \diamond)$ be groups. A map $\phi : G \to H$ is called a **homomorphism** if, for all $x, y \in G$,

$$\phi(x * y) = \phi(x) \diamond \phi(y).$$

When the group operations for G and H are not explicitly written, the homomorphism condition becomes simply

$$\phi(xy) = \phi(x)\phi(y)$$

but it is important to keep in mind that the product on the LHS is computed in G, and the product on the RHS is computed in H.

Definition 3.33 (Isomorphism)**.** $\phi : G \to H$ is called an **isomorphism** if

(i) ϕ is a homomorphism;

(ii) ϕ is a bijection.

Then G and H are said to be **isomorphic**, denoted by $G \cong H$.

In other words, the groups G and H are isomorphic if there is a bijection between them which preserves the group operations. Intuitively, G and H are the same group except that the elements and the operations may be written differently in G and H.

We also have the following terminology: An **automorphism** of a group G is an isomorphism from G to G. The automorphisms of G form a group $\mathrm{Aut}(G)$ under composition. An endomorphism of G is a homomorphism from G to G. (Rarely used) A **monomorphism** is an injective homomorphism and an **epimorphism** is a surjective homomorphism.

Example

For any group G, $G \cong G$ as the identity map provides an isomorphism from G to itself. (Exercise: prove that the identity map is the *only* isomorphism from G to itself.)

$\mathbf{Z} \cong 10\mathbf{Z}$ as the map $\phi : \mathbf{Z} \to 10\mathbf{Z}$ by $x \mapsto 10x$ is a homomorphism and a bijection.

Exercise

Prove that $(\mathbf{R}, +) \cong (\mathbf{R}^+, \times)$.

Proof. The exponential map $\exp : \mathbf{R} \to \mathbf{R}^+$ defined by $\exp(x) = e^x$, where e is the base of the natural logarithm, is an isomorphism from $(\mathbf{R}, +)$ to $(\mathbf{R}^+, \times)$.

(i) exp is a bijection since it has an inverse function (namely ln).

(ii) exp preserves the group operations since $e^{x+y} = e^x e^y$.

We see that both the elements and the operations are different yet the two groups are isomorphic, that is, as groups they have identical structures. □

Proposition 3.34. Let $\phi : G \to H$ be a homomorphism between groups and let $g \in G$, $n \in \mathbf{Z}$. Then

(i) $\phi(e_G) = e_H$;

(ii) $\phi(g^{-1}) = (\phi(g))^{-1}$;

(iii) $\phi(g^n) = (\phi(g))^n$.

Proof.

(i) We have
$$\phi(e_G) = \phi(e_G e_G) = \phi(e_G)\phi(e_G).$$
Now apply $\phi(e_G)^{-1}$ to both sides. Since $\phi(e_G)\phi(e_G)^{-1} = e_H$, we have
$$e_H = \phi(e_G)e_H,$$
so $\phi(e_G) = e_H$.

(ii)
$$\phi(g)\phi(g^{-1}) = \phi(gg^{-1}) = \phi(e_G) = e_H.$$

(iii) Note more generally that we can show $\phi(g^n) = (\phi(g))^n$ for $n > 0$ by induction and then for $n = -k < 0$ we have
$$\phi(g^n) = \phi((g^{-1})^k) = (\phi(g^{-1}))^k = (\phi(g)^{-1})^k = \phi(g)^n.$$

□

Proposition 3.35. If $\phi : G \to H$ is an isomorphism, then

(i) $|G| = |H|$;

(ii) G is abelian if and only if H is abelian;

(iii) $|x| = |\phi(x)|$ for all $x \in G$.

Kernel and Image

Definition 3.36 (Kernel, image)**.** If ϕ is a homomorphism $\phi : G \to H$, the **kernel** of ϕ is
$$\ker \phi := \{g \in G \mid \phi(g) = e_H\} \subset G.$$

The **image** of G under ϕ is
$$\operatorname{im} \phi := \phi(G) = \{\phi(g) \mid g \in G\} \subset H.$$

Remark. $\operatorname{im} \phi$ is the usual set theoretic image of ϕ.

Definition 3.37 (Normal subgroup)**.** Let G be a group, $H \leqslant G$. H is said to be a **normal subgroup** of G, denoted by $H \triangleleft G$, if
$$gH = Hg \quad (\forall g \in G)$$
or equivalently if
$$g^{-1}hg \in H \quad (\forall g \in G, h \in H)$$

Remark. This does *not* mean that $gh = hg$ for all $g \in G$, $h \in H$ or that G is abelian. Although we can easily see that all subgroups of abelian groups are normal.

Proposition 3.38. Let $\phi : G \to H$ be a homomorphism between groups. Then $\ker \phi \leqslant G$. In fact, $\ker \phi \triangleleft G$.

Proposition 3.39. Let $\phi : G \to H$ be a homomorphism between groups. Then $\operatorname{im} \phi \leqslant H$.

Quotient Groups

Definition 3.40 (Quotient group)**.**

Isomorphism Theorems

Theorem 3.41 (First isomorphism theorem)**.** Let $\phi : G \to H$ be a homomorphism of groups. Then $G/\ker\phi \cong \operatorname{im}\phi(G)$.

Corollary 3.42. Let $\phi : G \to H$ be a homomorphism of groups.

(i) ϕ is injective if and only if $\ker\phi = 1$.

(ii) $|G : \ker\phi| = |\phi(G)|$.

Theorem 3.43 (Second isomorphism theorem)**.**

Theorem 3.44 (Third isomorphism theorem)**.**

Theorem 3.45 (Fourth isomorphism theorem)**.**

§3.3 Group Actions

We move now, from thinking of groups in their own right, to thinking of how groups can move sets around – for example, how S_n permutes $\{1, 2, \ldots, n\}$ and matrix groups move vectors.

Definition 3.46 (Left action)**.** A **left action** of a group G on a set S is a map $\rho : G \times S \to S$ such that

(i) $\rho(e, s) = s$ for all $s \in S$;

(ii) $\rho(g, \rho(h, s)) = \rho(gh, s)$ for all $s \in S$, $g, h \in G$.

Notation. We will normally write $g \cdot s$ for $\rho(g, s)$ and so (i) and (ii) above would now read as:

(i) $e \cdot s = s$ for all $s \in S$;

(ii) $g \cdot (h \cdot s) = (gh) \cdot s$ for all $s \in S$, $g, h \in G$.

Remark. We will think of $g \cdot s \in S$ as the point that s is moved to by g.

Definition 3.47 (Right action)**.** A **right action** of a group G on a set S is a map $\rho : S \times G \to S$ such that

(i) $\rho(s, e) = s$ for all $s \in S$;

(ii) $\rho(\rho(s, h), g) = \rho(s, hg)$ for all $s \in S$, $g, h \in G$.

4 Rings

§4.1 Introduction to Rings

Definitions and Examples

Definition 4.1 (Ring)**.** A **ring** $(R, +, \times, 0, 1)$ consists of a set R, $0, 1 \in R$, together with two binary operations addition and multiplication, denoted $+$ and $\times$, satisfying the following axioms:

(i) $(R, +)$ is an abelian group with additive identity 0.

(ii) $\times$ is associative with multiplicative identity 1.

(iii) $\times$ distributes over $+$: for all $a, b, c \in R$,

$$a \times (b + c) = (a \times b) + (a \times c),$$
$$(a + b) \times c = (a \times c) + (b \times c).$$

Notation. We simply write ab rather than $a \times b$ for $a, b \in R$.

A ring is said to be a **commutative ring** if $\times$ is commutative.

Remark. It is also worth noting that some texts require an additional axiom asserting that $1 \neq 0$. In fact it is easy to see from the other axioms that if $1 = 0$ then the ring has only one element. We will refer to this ring as the “zero ring”. While it is a somewhat degenerate object, it seems unnecessary to me to exclude it.

Definition 4.2. A ring R with identity 1, where $1 \neq 0$, is called a **division ring** if every $a \in R$, $a \neq 0$ has a multiplicative inverse, i.e. there exists $b \in R$ such that $ab = ba = 1$.

A commutative division ring is called a **field**.

$$\begin{tikzcd}
 & \text{commutative ring} \arrow[dr, "\exists a^{-1}"] & \\
\text{ring} \arrow[ur, "ab=ba"] \arrow[dr, "\exists a^{-1}"'] & & \text{field} \\
 & \text{division ring} \arrow[ur, "ab=ba"'] &
\end{tikzcd}$$

Example

$\mathbf{Z}$ under usual addition and multiplication is a commutative ring with identity 1.

$\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$ are field.

$\mathbf{Z}/n\mathbf{Z}$ is a commutative ring with identity $\bar{1}$ under addition and multiplication of residue classes.

Proposition 4.3. Let R be a ring. Then

(i) $0a = a0 = 0$ for all $a \in R$.

(ii) $(-a)b = a(-b) = -(ab)$ for all $a, b \in R$.

(iii) $(-a)(-b) = ab$ for all $a, b \in R$.

(iv) if R has identity 1, then the identity is unique and $-a = (-1)a$.

Proof. These all follow from the distributive laws and cancellation in the additive group $(R, +)$.

(i) $0a = (0 + 0)a = 0a + 0a$ then add the additive identity of $0a$ to both sides to get $0a = 0$. Similarly, $a0 = a(0 + 0) = a0 + a0$ then add the additive identity of $a0$ to both sides to get $a0 = 0$.

(ii)

(iii)

(iv)

□

Definition 4.4 (Subring)**.** $S \subset R$ is a **subring** of ring R if S is a subgroup of R that is closed under multiplication.

Lemma 4.5 (Subring criterion)**.** Let R be a ring, $S \subset R$. Then S is a subring of R if and only if

(i) $1 \in S$;

(ii) $s_1 s_2 \in S$ and $s_1 - s2 \in S$ for all $s_1, s_2 \in S$.

Proof.

$\boxed{\Longrightarrow}$

$\boxed{\Longleftarrow}$ The condition that $s_1 - s_2 \in S$ for all $s_1, s_2 \in S$ implies that S is an additive subgroup by the subgroup test (note that as $1 \in S$ we know that S is nonempty). The other conditions for a subring hold directly. □

When studying any kind of algebraic object, it is natural to consider maps between those kind of objects which respect their structure. For example, for vector spaces the natural class of maps are linear maps, and for groups the natural class are the group homomorphisms. The natural class of maps to consider for rings are defined similarly:

Definition 4.6. Let R and S be rings. $\phi : R \to S$ is a **homomorphism** if it satisfies

(i) $\phi(1_R) = 1_S$;

(ii) $\phi(r_1 + r_2) = \phi(r_1) + \phi(r_2)$ for all $r_1, r_2 \in R$;

(iii) $\phi(r_1 r_2) = \phi(r_1)\phi(r_2)$ for all $r_1, r_2 \in R$.

A bijective ring homomorphism is called an **isomorphism**, denoted by $R \cong S$.

Recall that in a ring we do not require that nonzero elements have a multiplicative inverse. Nevertheless, because the multiplication operation is associative and there is a multiplicative identity, the elements which happen to have multiplicative inverses form a group:

Definition 4.7 (Unit)**.** Let R be a ring. $a \in R$ is called a **unit** in R if there exists $b \in R$ such that $ab = ba = 1$.

Proposition 4.8. The units in a ring R form a group under multiplication.

Definition 4.9 (Group of units)**.** Let R be a ring. The subset

$$R^{\times} = \{r \in R \mid \exists s \in R, rs = 1\}$$

is called the **group of units** in R; it is a group under multiplication $\times$ with identity element 1.

Polynomial Rings

§4.2 Basic Properties

Integral Domains

Definition 4.10. Let R be a ring. $a \in R \setminus \{0\}$ is called a **zero divisor** if there exists $b \in R \setminus \{0\}$ such that $ab = 0$.

A ring which is not the zero ring and has no zero divisors is called an **integral domain**. Thus if a ring is an integral domain and a.b = 0 then one of a or b is equal to zero.

The absence of zero divisors in integral domains give these rings a cancellation property:

Proposition 4.11. Let R be a ring. $a, b, c \in R$, a is not a zero divisor. If $ab = ac$, then either $a = 0$ or $b = c$. In particular, for any a, b, c in an integral domain and $ab = ac$, then either $a = 0$ or $b = c$.

Corollary 4.12. Any finite integral domain is a field.

The Field of Fractions

§4.3 Ideals and Quotients

From now on we will assume all our rings are commutative. In this section we study the basic properties of ring homomorphisms, and establish an analogue of the "first isomorphism theorem" which you have seen already for groups. Just as for homomorphisms of groups, homomorphisms of rings have kernels and images.

Definition 4.13. Let $\phi : R \to S$ be a ring homomorphism. The **kernel** of ϕ is

$$\ker \phi := \{r \in R \mid \phi(r) = 0\},$$

and the **image** of ϕ is

$$\operatorname{im} \phi := \{s \in S \mid \exists r \in R, \phi(r) = s\}.$$

If $\operatorname{im} \phi = S$, we say that ϕ is surjective.

Definition 4.14. Let R be a ring. A subset $I \subset R$ is called an **ideal** in R, denoted by $I \triangleleft R$, if

(i) I is a subgroup of $(R, +)$;

(ii) $ar \in I$ for all $a \in I$, $r \in R$.

Lemma 4.15. If $\phi : R \to S$ is a ring homomorphism, then $\ker \phi$ is an ideal. Moreover $I \subset R$ is an ideal if and only if it is nonempty, closed under addition, and closed under multiplication by arbitrary elements of R.

Proof. This is immediate from the definitions. For the moreover part, we just need to check that I is closed under taking additive inverses. But this follows from the fact that it is closed under multiplication by any element of R since $-x = (-1)x$ for any $x \in R$. □

Note that if I is an ideal of R which contains 1, then $I = R$. We will shortly see that in fact any ideal is the kernel of a homomorphism. First let us note a few basic properties of ideals:

§4.4 Domains

Euclidean Domains

Principal Ideal Domains

Unique Factorisation Domains

§4.5 Polynomial Rings

II

Linear Algebra

5 Vector Spaces

§5.1 Definition of Vector Space

Notation. A field is denoted by $\mathbf{F}$, which can mean either $\mathbf{R}$ or $\mathbf{C}$. $\mathbf{F}^n$ is the set of n-tuples whose elements belong to $\mathbf{F}$:

$$\mathbf{F}^n := \{(x_1, \dots, x_n) \mid x_i \in \mathbf{F}\}$$

For $(x_1, \dots, x_n) \in \mathbf{F}^n$ and $i = 1, \dots, n$, we say that x_i is the i-th coordinate of $(x_1, \dots, x_n)$.

Definition 5.1 (Vector space)**.** V is a **vector space** over $\mathbf{F}$ if the following properties hold:

(i) Addition is commutative: $u + v = v + u$ for all $u, v \in V$

(ii) Addition is associative: $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$

Multiplication is associative: $(ab)v = a(bv)$ for all $v \in V$, $a, b \in \mathbf{F}$

(iii) Additive identity: there exists $\mathbf{0} \in V$ such that $v + \mathbf{0} = v$ for all $v \in V$

(iv) Additive inverse: for every $v \in V$, there exists $w \in V$ such that $v + w = \mathbf{0}$

(v) Multiplicative identity: $1v = v$ for all $v \in V$

(vi) Distributive properties: $a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b, \in \mathbf{F}$ and $u, v \in V$

Notation. For the rest of this text, V denotes a vector space over $\mathbf{F}$.

> **Example**
>
> $\mathbf{R}^n$ is a vector space over $\mathbf{R}$, $\mathbf{C}^n$ is a vector space over $\mathbf{C}$.

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

The scalar multiplication in a vector space depends on $\mathbf{F}$. Thus when we need to be precise, we will say that V is a vector space over $\mathbf{F}$ instead of saying simply that V is a vector space. For example, $\mathbf{R}^n$ is a vector space over $\mathbf{R}$, and $\mathbf{C}^n$ is a vector space over $\mathbf{C}$. A vector space over $\mathbf{R}$ is called a *real vector space*; a vector space over $\mathbf{C}$ is called a *complex vector space*.

Proposition 5.2 (Uniqueness of additive identity)**.** A vector space has a unique additive identity.

Proof. Suppose otherwise, then $\mathbf{0}$ and $\mathbf{0}'$ are additive identities of V. Then

$$\mathbf{0}' = \mathbf{0}' + \mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}$$

where the first equality holds because $\mathbf{0}$ is an additive identity, the second equality comes from commutativity, and the third equality holds because $\mathbf{0}'$ is an additive identity. Thus $\mathbf{0}' = \mathbf{0}$. □

Proposition 5.3 (Uniqueness of additive inverse)**.** Every element in a vector space has a unique additive inverse.

Proof. Suppose otherwise, then for $v \in V$, w and w' are additive inverses of v. Then

$$w = w + \mathbf{0} = w + (v + w') = (w + v) + w' = \mathbf{0} + w' = w'.$$

Thus $w = w'$. □

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 $w + (-v)$.

We now prove some seemingly trivial facts.

Proposition 5.4.

(i) For every $v \in V$, $0v = \mathbf{0}$.

(ii) For every $a \in \mathbf{F}$, $a\mathbf{0} = \mathbf{0}$.

(iii) For every $v \in V$, $(-1)v = -v$.

Proof.

(i) For $v \in V$, we have
$$0v = (0 + 0)v = 0v + 0v.$$
Adding the additive inverse of $0v$ to both sides of the equation gives $\mathbf{0} = 0v$.

(ii) For $a \in \mathbf{F}$, we have
$$a\mathbf{0} = a(\mathbf{0} + \mathbf{0}) = a\mathbf{0} + a\mathbf{0}.$$
Adding the additive inverse of $a\mathbf{0}$ to both sides of the equation gives $\mathbf{0} = a\mathbf{0}$.

(iii) For $v \in V$, we have
$$v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = \mathbf{0}.$$
Since $v + (-1)v = \mathbf{0}$, $(-1)v$ is the additive inverse of v.

□

Example

$\mathbf{F}^\infty$ is defined to be the set of all sequences of elements of $\mathbf{F}$:

$$\mathbf{F}^\infty := \{(x_1, x_2, \dots) \mid x_i \in \mathbf{F}\}$$

- Addition on $\mathbf{F}^\infty$ is defined by

$$(x_1, x_2, \dots) + (y_1, y_2, \dots) = (x_1 + y_1, x_2 + y_2, \dots)$$

- Scalar multiplication on $\mathbf{F}^\infty$ is defined by

$$\lambda(x_1, x_2, \dots) = (\lambda x_1, \lambda x_2, \dots)$$

Verify that $\mathbf{F}^\infty$ becomes a vector space over $\mathbf{F}$. Also verify that the additive identity in $\mathbf{F}^\infty$ is $\mathbf{0} = (0, 0, \dots)$.

Our next example of a vector space involves a set of functions.

Example

If S is a set, $\mathbf{F}^S := \{f \mid f : S \to \mathbf{F}\}$.

- Addition on $\mathbf{F}^S$ is defined by

$$(f + g)(x) = f(x) + g(x) \quad (\forall x \in S)$$

  for all $f, g \in \mathbf{F}^S$.

- Multiplication on $\mathbf{F}^S$ is defined by

$$(\lambda f)(x) = \lambda f(x) \quad (\forall x \in S)$$

  for all $\lambda \in \mathbf{F}$, $f \in \mathbf{F}^S$.

Verify that if S is a non-empty set, then $\mathbf{F}^S$ is a vector space over $\mathbf{F}$.

Also verify that the additive identity of $\mathbf{F}^S$ is the function $0 : S \to \mathbf{F}$ defined by

$$0(x) = 0 \quad (\forall x \in S)$$

and for $f \in \mathbf{F}^S$, additive inverse of f is the function $-f : S \to \mathbf{F}$ defined by

$$(-f)(x) = -f(x) \quad (\forall x \in S)$$

Remark. $\mathbf{F}^n$ and $\mathbf{F}^\infty$ are special cases of the vector space $\mathbf{F}^S$; think of $\mathbf{F}^n$ as $\mathbf{F}^{\{1,2,\dots,n\}}$, and $\mathbf{F}^\infty$ as $\mathbf{F}^{\{1,2,\dots\}}$.

Example (Complexification)

Suppose V is a real vector space. The *complexifcation* of V, denoted by $V_{\mathbf{C}}$, equals $V \times V$. An element of $V_{\mathbf{C}}$ is an ordered pair (u, v), where $u, v \in V$, which we write as $u + iv$.

- Addition on $V_{\mathbf{C}}$ is defined by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$

  for all $u_1, v_2, u_2, v_2 \in V$.

- Complex scalar multiplication on $V_{\mathbf{C}}$ is defined by

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

  for all $a, b \in \mathbf{R}$ and all $u, v \in V$.

You should verify that with the defnitions of addition and scalar multiplication as above, $V_{\mathbf{C}}$ is a (complex) vector space.

§5.2 Subspaces

Whenever we have a mathematical object with some structure, we want to consider subsets that also have the same structure.

Definition 5.5 (Subspace)**.** $U \subset V$ is a **subspace** of V if U is also a vector space (with the same addition and scalar multiplication as on V). We denote this as $U \leqslant V$.

The sets $\{\mathbf{0}\}$ and V are always subspaces of V. The subspace $\{\mathbf{0}\}$ is called the *zero subspace* or *trivial subspace*. Subspaces other than V are called *proper subspaces*.

The following result is useful in determining whether a given subset of V is a subspace of V.

Lemma 5.6 (Subspace test)**.** Suppose $U \subset V$. Then $U \leqslant V$ if and only if U satisfies the following conditions:

(i) Additive identity: $\mathbf{0} \in U$

(ii) Closed under addition: $u + w \in U$ for all $u, w \in U$

(iii) Closed under scalar multiplication: $\lambda u \in U$ for all $\lambda \in \mathbf{F}$, $u \in U$

Proof.

$\boxed{\implies}$ If $U \leqslant V$, then U satisfies the three conditions above by the definition of vector space.

$\boxed{\impliedby}$ Suppose U satisfies the three conditions above. (i) ensures that the additive identity of V is in U. (ii) ensures that addition makes sense on U. (iii) ensures that scalar multiplication makes sense on U.

If $u \in U$, then $-u = (-1)u \in U$ by (iii). Hence every element of U has an additive inverse in U.

The other parts of the definition of a vector space, such as associativity and commutativity, are automatically satisfied for U because they hold on the larger space V. Thus U is a vector space and hence is a subspace of V. $\square$

Proposition 5.7. Suppose $U \leqslant V$. Then

(i) U is a vector space over $\mathbf{F}$. In fact, the only subsets of V that are vector spaces over $\mathbf{F}$ are the subspaces of V;

(ii) if $W \leqslant U$, then $W \leqslant V$ ("a subspace of a subspace is a subspace").

Proof.

(i) We first check that we have legitimate operations. Since U is closed under addition, the operation $+$ restricted to U gives a map $U \times U \to U$. Likewise since U is closed under scalar multiplication, that operation restricted to U gives a map $\mathbf{F} \times U \to U$.

We now check that U satisfies the vector space axioms.

  (i) Commutativity and associativity of addition are inherited from V.
  (ii) There is an additive identity (by the subspace test).
  (iii) There are additive inverses: if $u \in U$ then multiplying by $-1 \in \mathbf{F}$ and shows that $-u = (-1)u \in U$.
  (iv) The remaining four properties are all inherited from V. That is, they apply to general vectors of V and vectors in U are vectors in V.

(ii) This is immediate from the definition of a subspace.

□

Definition 5.8 (Sum of subsets)**.** Suppose $U_1, \dots, U_n \subset V$. The sum of $U_1, \dots, U_n$ is the set of all possible sums of elements of $U_1, \dots, U_n$:

$$U_1 + \cdots + U_n := \{u_1 + \cdots + u_n \mid u_i \in U_i\}.$$

> **Example**
> Suppose that $U = \{(x, 0, 0) \in \mathbf{F}^3 \mid x \in F\}$ and $W = \{(0, y, 0) \in \mathbf{F}^3 \mid y \in \mathbf{F}\}$. Then
>
> $$U + W = \{(x, y, 0) \mid x, y \in \mathbf{F}\}.$$
>
> Suppose that $U = \{(x, x, y, y) \in \mathbf{F}^4 \mid x, y \in \mathbf{F}\}$ and $W = \{(x, x, x, y) \in \mathbf{F}^4 \mid x, y \in \mathbf{F}\}$. Then
>
> $$U + W = \{(x, x, y, z) \in \mathbf{F}^4 \mid x, y, z \in \mathbf{F}\}.$$

The next result states that the sum of subspaces is a subspace, and is in fact the smallest subspace containing all the summands.

Proposition 5.9. Suppose $U_1, \dots, U_n \leqslant V$. Then $U_1 + \cdots + U_n$ is the smallest subspace of V containing $U_1, \dots, U_n$.

Proof. It is easy to see that $\mathbf{0} \in U_1 + \cdots + U_n$ and that $U_1 + \cdots + U_n$ is closed under addition and scalar multiplication. Hence by the subspace test, $U_1 + \cdots + U_n \leqslant V$.

Let M be the smallest subspace of V containing $U_1, \dots, U_n$. We want to show that $U_1 + \cdots + U_n = M$. To do so, we show double inclusion: $U_1 + \cdots + U_n \subset M$ and $M \subset U_1 + \cdots + U_n$.

(i) For all $u_i \in U_i$ $(1 \leqslant i \leqslant n)$,

$$u_i = \mathbf{0} + \cdots + \mathbf{0} + u_i + \mathbf{0} + \cdots + \mathbf{0} \in U_1 + \cdots + U_n,$$

where all except one of the u's are $\mathbf{0}$. Thus $U_i \subset U_1 + \cdots + U_n$ for $1 \leqslant i \leqslant n$. Hence $M \subset U_1 + \cdots + U_n$.

(ii) Conversely, every subspace of V containing $U_1, \dots, U_n$ contains $U_1 + \cdots + U_n$ (because subspaces must contain all finite sums of their elements). Hence $U_1 + \cdots + U_n \subset M$.

□

Definition 5.10 (Direct sum)**.** Suppose $U_1, \dots, U_n \leqslant V$. If each element of $U_1 + \cdots + U_n$ can be written in only one way as a sum $u_1 + \cdots + u_n$, $u_i \in U_i$, then $U_1 + \cdots + U_n$ is called a **direct sum**. In this case, we denote the sum as

$$U_1 \oplus \cdots \oplus U_n.$$

Example

Suppose that $U = \{(x, y, 0) \in \mathbf{F}^3 \mid x, y \in \mathbf{F}\}$ and $W = \{(0, 0, z) \in \mathbf{F}^3 \mid z \in \mathbf{F}\}$. Then $\mathbf{F}^3 = U \oplus W$.

Suppose U_i is the subspace of $\mathbf{F}^n$ of those vectors whose coordinates are all 0 except for the i-th coordinate; that is, $U_i = \{(0, \dots, 0, x, 0, \dots, 0) \in \mathbf{F}^n \mid x \in \mathbf{F}\}$. Then $\mathbf{F}^n = U_1 \oplus \cdots \oplus U_n$.

Lemma 5.11 (Condition for direct sum)**.** Suppose $V_1, \dots, V_n \leqslant V$, let $W = V_1 + \cdots + V_n$. Then the following are equivalent:

(i) Any element in W can be uniquely expressed as the sum of vectors in $V_1, \dots, V_n$.

(ii) If $v_i \in V_i$ satisfies $v_1 + \cdots + v_n = \mathbf{0}$, then $v_1 = \cdots = v_n = \mathbf{0}$.

(iii) For $k = 2, \dots, n$, $(V_1 + \cdots + V_{k-1}) \cap V_k = \{\mathbf{0}\}$.

Proof.

(i) $\iff$ (ii) First suppose W is a direct sum. Then by the definition of direct sum, the only way to write $\mathbf{0}$ as a sum $u_1 + \cdots + u_n$ is by taking $u_i = \mathbf{0}$.

Now suppose that the only way to write $\mathbf{0}$ as a sum $v_1 + \cdots + v_n$ by taking $v_1 = \cdots = v_n = \mathbf{0}$. For $v \in V_1 + \cdots + V_n$, suppose that there is more than one way to represent v:

$$\begin{aligned} v &= v_1 + \cdots + v_n \\ v &= v_1' + \cdots + v_n' \end{aligned}$$

for some $v_i, v_i' \in V_i$. Substracting the above two equations gives

$$\mathbf{0} = (v_1 - v_1') + \cdots + (v_n - v_n').$$

Since $v_i - v_i' \in V_i$, we have $v_i - v_i' = \mathbf{0}$ so $v_i = v_i'$. Hence there is only one unique way to represent $v_1 + \cdots + v_n$, thus W is a direct sum.

(ii) $\iff$ (iii) First suppose if $v_i \in V_i$ satisfies $v_1 + \cdots + v_n = \mathbf{0}$, then $v_1 = \cdots = v_n = \mathbf{0}$. Let $v_k \in (V_1 + \cdots + V_{k-1}) \cap V_k$. Then $v_k = v_1 + \cdots + v_{k-1}$ where $v_i \in V_i$ $(1 \leqslant i \leqslant k-1)$. Thus

$$\begin{aligned} v_1 + \cdots + v_{k-1} - v_k &= \mathbf{0} \\ v_1 + \cdots + v_{k-1} + (-v_k) + \mathbf{0} + \cdots + \mathbf{0} &= \mathbf{0} \end{aligned}$$

by taking $v_{k+1} = \cdots = v_n = \mathbf{0}$. Then $v_1 = \cdots = v_k = \mathbf{0}$.

Now suppose that for $k = 2, \dots, n$, $(V_1 + \cdots + V_{k-1}) \cap V_k = \{\mathbf{0}\}$.

$$v_1 + \cdots + v_n = \mathbf{0}$$
$$v_1 + \cdots + v_{n-1} = -v_n$$

where $v_1 + \cdots + v_{n-1} \in V_1 + \cdots + V_{n-1}$, $-v_n \in V_n$. Thus

$$v_1 + \cdots + v_{n-1} = -v_n \in (V_1 + \cdots + V_{n-1}) \cap V_n = \{\mathbf{0}\}$$

so $v_1 + \cdots + v_{n-1} = \mathbf{0}$, $v_n = \mathbf{0}$. Induction on n gives $v_1 = \cdots = v_{n-1} = v_n = \mathbf{0}$. □

Proposition 5.12. Suppose $U, W \leqslant V$. Then $U + W$ is a direct sum if and only if $U \cap W = \{\mathbf{0}\}$.

Proof.

$\boxed{\implies}$ Suppose that $U + W$ is a direct sum. If $v \in U \cap W$, then $\mathbf{0} = v + (-v)$, where $v \in U$, $-v \in W$. By the unique representation of $\mathbf{0}$ as the sum of a vector in U and a vector in W, we have $v = \mathbf{0}$. Thus $U \cap W = \{\mathbf{0}\}$.

$\boxed{\impliedby}$ Suppose $U \cap W = \{\mathbf{0}\}$. Suppose $u \in U$, $w \in W$, and $0 = u + w$. $u = -w \in W$, thus $u \in U \cap W$, so $u = w = \mathbf{0}$. By Lemma 5.11, $U + W$ is a direct sum. □

§5.3 Span and Linear Independence

Definition 5.13 (Linear combination)**.** v is a **linear combination** of vectors $v_1, \dots, v_n \in V$ if there exists $a_1, \dots, a_n \in \mathbf{F}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Definition 5.14 (Span)**.** The **span** of $\{v_1, \dots, v_n\}$ is the set of all linear combinations of $v_1, \dots, v_n$:

$$\operatorname{span}(v_1, \dots, v_n) := \{a_1 v_1 + \cdots + a_n v_n \mid a_i \in \mathbf{F}\}.$$

The span of the empty list ( ) is defined to be $\{\mathbf{0}\}$.

We say that $v_1, \dots, v_n$ *spans* V if $\operatorname{span}(v_1, \dots, v_n) = V$.

If $S \subset V$ is such that $\operatorname{span}(S) = V$, then we say that S *spans* V, and that S is a *spanning set* for V:

$$\operatorname{span}(S) := \{a_1 v_1 + \cdots + a_n v_n \mid v_i \in S, a_i \in \mathbf{F}\}.$$

Proposition 5.15. $\operatorname{span}(v_1, \dots, v_n)$ in V is the smallest subspace of V containing $v_1, \dots, v_n$.

Proof. First we show that $\operatorname{span}(v_1, \dots, v_n) \leqslant V$, using the subspace test.

(i) $\mathbf{0} = 0v_1 + \cdots + 0v_n \in \operatorname{span}(v_1, \dots, v_n)$

(ii) $(a_1v_1 + \cdots + a_nv_n) + (c_1v_1 + \cdots + c_nv_n) = (a_1 + c_1)v_1 + \cdots + (a_n + c_n)v_n \in \operatorname{span}(v_1, \dots, v_n)$, so $\operatorname{span}(v_1, \dots, v_n)$ is closed under addition.

(iii) $\lambda(a_1v_1 + a_nv_n) = (\lambda a_1)v_1 + \cdots + (\lambda a_n)v_n \in \operatorname{span}(v_1, \dots, v_n)$, so $\operatorname{span}(v_1, \dots, v_n)$ is closed under scalar multiplication.

Let M be the smallest vector subspace of V containing $v_1, \dots, v_n$. We claim that $M = \operatorname{span}(v_1, \dots, v_n)$. To show this, we show that (i) $M \subset \operatorname{span}(v_1, \dots, v_n)$ and (ii) $M \supset \operatorname{span}(v_1, \dots, v_n)$.

(i) Each v_i is a linear combination of $v_1, \dots, v_n$, as

$$v_i = 0 \cdot v_1 + \cdots + 0 \cdot v_{i-1} + 1 \cdot v_i + 0 \cdot v_{i+1} + \cdots + 0 \cdot v_n,$$

so by the definition of the span as the collection of all linear combinations of $v_1, \dots, v_n$, we have that $v_i \in \operatorname{span}(v_1, \dots, v_n)$. But M is the smallest vector subspace containing $v_1, \dots, v_n$, so

$$M \subset \operatorname{span}(v_1, \dots, v_n).$$

(ii) Since $v_i \in M$ $(1 \leqslant i \leqslant n)$ and M is a vector subspace (closed under addition and scalar multiplication), it follows that

$$a_1v_1 + \cdots + a_nv_n \in M$$

for all $a_i \in \mathbf{F}$ (i.e. M contains all linear combinations of $v_1, \dots, v_n$). So

$$\operatorname{span}(v_1, \dots, v_n) \subset M.$$

□

Definition 5.16 (Finite-dimensional vector space)**.** V is **finite-dimensional** if there exists some list of vector $\{v_1, \dots, v_n\}$ that spans V; otherwise, it is *infinite-dimensional.*

Remark. Recall that by defnition every list of vectors has finite length.

Remark. From this definition, infinite-dimensionality is the negation of finite-dimensionality (i.e. *not* finite-dimensional). Hence to prove that a vector space is infinite-dimensional, we prove by contradiction; that is, first assume that the vector space is finite-dimensional, then try to come to a contradiction.

Exercise

For positive integer n, $\mathbf{F}^n$ is finite-dimensional.

Proof. Suppose $(x_1, x_2, \dots, x_n) \in \mathbf{F}^n$, then

$$(x_1, x_2, \dots, x_n) = x_1(1, 0, \dots, 0) + x_2(0, 1, \dots, 0) + \cdots + x_n(0, 0, \dots, 1)$$

so

$$(x_1, \dots, x_n) \in \operatorname{span}\left((1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, \dots, 0, 1)\right).$$

The vectors $(1,0,\dots,0),(0,1,\dots,0),\dots,(0,\dots,0,1)$ spans $\mathbf{F}^n$, so $\mathbf{F}^n$ is finite-dimensional. □

Definition 5.17 (Linear independence)**.** A list of vectors $v_1,\dots,v_n$ is **linearly independent** in V if the only choice of $a_1,\dots,a_n \in \mathbf{F}$ that makes

$$a_1v_1+\cdots+a_nv_n=\mathbf{0}$$

is $a_1=\cdots=a_n=0$; otherwise, it is *linearly dependent.*

We say that $S \subset V$ is linearly independent if every finite subset of S is linearly independent.

Proposition 5.18 (Compare coefficients)**.** Let $v_1,\dots,v_n$ be linearly independent in V. Then

$$a_1v_1+\cdots+a_nv_n=b_1v_1+\cdots+b_nv_n$$

if and only if $a_i=b_i$ $(1 \leqslant i \leqslant n)$.

Proof. Exercise. □

The following result will often be useful; it states that given a linearly dependent set of vectors, one of the vectors is in the span of the previous ones; furthermore we can throw out that vector without changing the span of the original set.

Lemma 5.19 (Linear dependence lemma)**.** Suppose $v_1,\dots,v_n$ are linearly dependent in V. Then there exists v_k such that the following hold:

(i) $v_k \in \operatorname{span}(v_1,\dots,v_{k-1})$

(ii) $\operatorname{span}(v_1,\dots,v_{k-1},v_{k+1},\dots,v_n)=\operatorname{span}(v_1,\dots,v_n)$

Proof.

(i) Since $v_1,\dots,v_n$ are linearly dependent, there exists $a_1,\dots,a_n \in \mathbf{F}$, not all 0, such that

$$a_1v_1+\cdots+a_nv_n=0.$$

Take $k=\max\{1,\dots,n\}$ such that $a_k \neq 0$. Then

$$v_k=-\frac{a_1}{a_k}v_1-\cdots-\frac{a_{k-1}}{a_k}v_{k-1},$$

which means that v_k can be written as a linear combination of $v_1,\dots,v_{k-1}$, so $v_k \in \operatorname{span}(v_1,\dots,v_{k-1})$ by definition of span.

(ii) Now suppose k is such that $v_k \in \operatorname{span}(v_1,\dots,v_{k-1})$. Then there exists $b_1,\dots,b_{k-1} \in \mathbf{F}$ be such that

$$v_k=b_1v_1+\cdots+b_{k-1}v_{k-1}. \tag{1}$$

Suppose $u \in \operatorname{span}(v_1, \ldots, v_n)$. Then there exists $c_1, \ldots, c_n \in \mathbf{F}$ such that

$$u = c_1 v_1 + \cdots + c_n v_n. \tag{2}$$

In (2), we can replace v_k with the RHS of (1), which gives

$$\begin{aligned}
u &= c_1 v_1 + \cdots + c_{k-1} v_{k-1} + c_k v_k + c_{k+1} v_{k+1} + \cdots + c_n v_n \\
&= c_1 v_1 + \cdots + c_{k-1} v_{k-1} + c_k (b_1 v_1 + \cdots + b_{k-1} v_{k-1}) + c_{k+1} v_{k+1} + \cdots + c_n v_n \\
&= c_1 v_1 + \cdots + c_{k-1} v_{k-1} + c_k b_1 v_1 + \cdots + c_k b_{k-1} v_{k-1} + c_{k+1} v_{k+1} + \cdots + c_n v_n \\
&= (c_1 + b c_k) v_1 + \cdots + (c_{k-1} + b_{k-1} c_k) v_{k-1} + c_{k+1} v_{k+1} + \cdots + c_n v_n.
\end{aligned}$$

Thus $u \in \operatorname{span}(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n)$. This shows that removing v_k from $v_1, \ldots, v_n$ does not change the span of the list.

□

The following result says that no linearly independent set in V is longer than a spanning set in V.

Proposition 5.20. In a finite-dimensional vector space, the length of every linearly independent set of vectors is less than or equal to the length of every spanning set of vectors.

Proof. Suppose $A = \{u_1, \ldots, u_m\}$ is linearly independent in V, $B = \{w_1, \ldots, w_n\}$ spans V. We want to prove that $m \leqslant n$.

Since B spans V, if we add any other vector from V to the list B, we will get a linearly dependent list, since this newly added vector can, by the definition of a span, be expressed as a linear combination of the vectors in B. In particular, if we add $u_1 \in A$ to B, then the new list

$$\{u_1, w_1, \ldots, w_n\}$$

is linearly dependent. By the linear independence lemma, we can remove one of the w_i's from B, so that the remaining list of n vectors still spans V. For the sake of argument, let's say we remove w_n (we can always order the w_i's in the list so that the element we remove is at the end). Then we are left with the revised list

$$B_1 = \{u_1, w_1, \ldots, w_{n-1}\}.$$

We can repeat this process m times, each time adding the next element u_i from list A and removing the last w_i. Because of the linear dependence lemma, we know that there must always be a w_i that can be removed each time we add a u_i, so there must be at least as many w_i's as u_i's. In other words, $m \leqslant n$ which is what we wanted to prove. □

Remark. We can use this result to show, without any computations, that certain lists are not linearly independent and that certain lists do not span a given vector space.

Our intuition suggests that every subspace of a finite-dimensional vector space should also be finite-dimensional. We now prove that this intuition is correct.

Proposition 5.21. Every subspace of a finite-dimensional vector space is finite-dimensional.

Proof. Suppose V is finite-dimensional, $U \leqslant V$. To show that U is finite-dimensional, we need to find a spanning set of vectors in U. We prove by construction of this spanning set.

Step 1 If $U = \{\mathbf{0}\}$, then U is finite-dimensional and we are done. Otherwise, choose $v_1 \in U$, $v_1 \neq \mathbf{0}$ and add it to our list of vectors.

Step k Our list so far is $\{v_1, \dots, v_{k-1}\}$. If $U = \operatorname{span}(v_1, \dots, v_{k-1})$, then U is finite-dimensional and we are done. Otherwise, choose $v_k \in U$ such that $v_k \notin \operatorname{span}(v_1, \dots, v_{k-1})$ and add it to our list.

After each step, we have constructed a list of vectors such that no vector in this list is in the span of the previous vectors; by the linear dependence lemma, our constructed list is a linearly independent set.

By Proposition 5.20, this linearly independent set cannot be longer than any spanning set of V. Thus the process must terminate after a finite number of steps, and we have constructed a spanning set of U. Hence U is finite-dimensional. $\square$

§5.4 Bases

Definition 5.22 (Basis)**.** $B = \{v_1 \dots, v_n\}$ is a **basis** of V if

(i) B is linearly independent in V;

(ii) B is a spanning set of V.

> **Example** (Standard basis)
>
> Let $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ where the i-th coordinate is 1. $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis of $\mathbf{F}^n$, known as the *standard basis* of $\mathbf{F}^n$.

Lemma 5.23 (Criterion for basis)**.** Let $B = \{v_1, \dots, v_n\}$ be a list of vectors in V. Then B is a basis of V if and only if every $v \in V$ can uniquely expressed as a linear combination of $v_1, \dots, v_n$.

Proof.

$\boxed{\Longrightarrow}$ Let $v \in V$. Since B is a basis of V, there exists $a_1, \dots, a_n \in \mathbf{F}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n. \tag{1}$$

To show that the representation is unique, suppose that $c_1, \dots, c_n \in \mathbf{F}$ also satisfy

$$v = c_1 v_1 + \cdots + c_n v_n. \tag{2}$$

Subtracting (2) from (1) gives

$$\mathbf{0} = (a_1 - c_1)v_1 + \cdots + (a_n - c_n)v_n.$$

Since $v_1, \dots, v_n$ are linearly independent, we have $a_i - c_i = 0$, or $a_i = c_i$ for all i $(1 \leqslant i \leqslant n)$. Thus the representation of v as a linear combination of $v_1, \dots, v_n$ is unique.

$\boxed{\Longleftarrow}$ Suppose that every $v \in V$ can be uniquely expressed as a linear combination of $v_1, \dots, v_n$. This implies that B spans V. To show that B is linearly independent, suppose that $a_1, \dots, a_n \in \mathbf{F}$ satisfy

$$a_1 v_1 + \cdots + a_n v_n = \mathbf{0}.$$

Since $\mathbf{0}$ can be uniquely expressed as a linear combination of $v_1, \dots, v_n$, we have $a_1 = \cdots = a_n = 0$, thus B is linearly independent. Since B is linearly independent and spans V, B is a basis of V. □

A spanning set in a vector space may not be a basis because it is not linearly independent. Our next result says that given any spanning set, some (possibly none) of the vectors in it can be discarded so that the remaining list is linearly independent and still spans the vector space.

Lemma 5.24. Every spanning set in a vector space can be reduced to a basis of the vector space.

Proof. Suppose $B = \{v_1, \dots, v_n\}$ spans V. We want to remove some vectors from B so that the remaining vectors form a basis of V. We do this through the multistep process described below.

Step 1 If $v_1 = \mathbf{0}$, delete v_1 from B. If $v_1 \neq \mathbf{0}$, leave B unchanged.

Step k If $v_k \in \operatorname{span}(v_1, \dots, v_{k-1})$, delete v_k from B. If $v_k \notin \operatorname{span}(v_1, \dots, v_{k-1})$, leave B unchanged.

Stop the process after step n, getting a list B. Since we only delete vectors from B that are in the span of the previous vectors, by the linear dependence lemma, the list B still spans V.

The process ensures that no vector in B is in the span of the previous ones. By the linear dependence lemma, B is linearly independent.

Since B is linearly independent and spans V, B is a basis of V. □

Corollary 5.25. Every finite-dimensional vector space has a basis.

Proof. We prove by construction. Suppose V is finite-dimensional. By definition, there exists a spanning set of vectors in V. By Lemma 5.24, the spanning set can be reduced to a basis. □

Now we show that given any linearly independent set, we can adjoin some additional vectors so that the extended list is still linearly independent but also spans the space.

Lemma 5.26. Every linearly independent set of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.

Proof. Suppose $u_1, \dots, u_m$ are linearly independent in V, $w_1, \dots, w_n$ span V. Then the list

$$\{u_1, \dots, u_m, w_1, \dots, w_n\}$$

spans V. By Lemma 5.24, we can reduce this list to a basis of V consisting $u_1, \dots, u_m$ (since $u_1, \dots, u_m$ are linearly independent, $u_i \notin \operatorname{span}(u_1, \dots, u_{i-1})$ for all i, so none of the u_i's are deleted in the process), and some of the w_i's. □

We now show that every subspace of a finite-dimensional vector space can be paired with another subspace to form a direct sum of the whole space.

Corollary 5.27. Suppose V is finite-dimensional, $U \leqslant V$. Then there exists $W \leqslant V$ such that $V = U \oplus W$.

Proof. Since V is finite-dimensional and $U \leqslant V$, by Proposition 5.21, U is finite-dimensional, so U has a basis B, by Corollary 5.25; let $B = \{u_1, \dots, u_n\}$. Since B is linearly independent, by Lemma 5.26, B can be extended to a basis of V, say

$$\{u_1, \dots, u_n, w_1, \dots, w_n\}.$$

Take $W = \operatorname{span}(w_1, \dots, w_n)$. We claim that $V = U \oplus W$. To show this, by Lemma 5.11, we need to show that (i) $V = U + W$, and (ii) $U \cap W = \{\mathbf{0}\}$.

(i) Suppose $v \in V$. Since $\{u_1, \dots, u_n, w_1, \dots, w_n\}$ spans V, there exists $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbf{F}$ such that

$$v = a_1 u_1 + \cdots + a_n u_n + b_1 w_1 + \cdots + b_n w_n.$$

Take $u = a_1 u_1 + \cdots + a_n u_n \in U$, $w = b_1 w_1 + \cdots + b_n w_n \in W$. Then $v = u + w \in U + W$, so $V = U + W$.

(ii) Suppose $v \in U \cap W$. Since $v \in U$, v can be written as a linear combination of $u_1, \dots, u_n$:

$$v = a_1 u_1 + \cdots + a_n u_n. \tag{1}$$

Since $v \in W$, v can be written as a linear combination of $w_1, \dots, w_n$:

$$v = b_1 w_1 + \cdots + b_n w_n. \tag{2}$$

Subtracting (2) from (1) gives

$$\mathbf{0} = a_1 u_1 + \cdots + a_n u_n - b_1 w_1 - \cdots - b_n w_n.$$

Since $u_1, \dots, u_n, w_1, \dots, w_n$ are linearly independent, we have $a_i = b_i = 0$ for all i ($1 \leqslant i \leqslant n$). Thus $v = \mathbf{0}$, so $U \cap W = \{\mathbf{0}\}$.

$\square$

§5.5 Dimension

Lemma 5.28. Any two bases of a finite-dimensional vector space have the same length.

Proof. Suppose V is finite-dimensional, let B_1 and B_2 be two bases of V. By definition, B_1 is linearly independent in V, and B_2 spans V, so by Proposition 5.20, $|B_1| \leqslant |B_2|$.

Similarly, by definition, B_2 is linearly independent in V and B_1 spans V, so $|B_2| \leqslant |B_1|$.

Since $|B_1| \leqslant |B_2|$ and $|B_2| \leqslant |B_1|$, we have $|B_1| = |B_2|$, as desired. □

Since any two bases of a finite-dimensional vector space have the same length, we can formally define the dimension of such spaces.

Definition 5.29 (Dimension)**.** The **dimension** of V is the length of any basis of V, denoted by $\dim V$.

Proposition 5.30. Suppose V is finite-dimensional, $U \leqslant V$. Then $\dim U \leqslant \dim V$.

Proof. Since V is finite-dimensional and $U \leqslant V$, U is finite-dimensional. Let B_U be a basis of U, and B_V be a basis of V.

By definition, B_U is linearly independent in V, and B_V spans V. By Proposition 5.20, $|B_U| \leqslant |B_V|$, so

$$\dim U = |B_U| \leqslant |B_V| = \dim V,$$

since $|B_U| = \dim U$ and $|B_V| = \dim V$ by definition. □

To check that a list of vectors is a basis, we must show that it is linearly independent and that it spans the vector space. The next result shows that if the list in question has the right length, then we only need to check that it satisfies one of the two required properties.

Proposition 5.31. Suppose V is finite-dimensional. Then

(i) every linearly independent set of vectors in V with length $\dim V$ is a basis of V;

(ii) every spanning set of vectors in V with length $\dim V$ is a basis of V.

Proof.

(i) Suppose $\dim V = n$, $\{v_1, \dots, v_n\}$ is linearly independent in V. By Lemma 5.26, $\{v_1, \dots, v_n\}$ can be extended to a basis of V. However, every basis of V has length n (by definition of dimension), which means that no elements are adjoined to $\{v_1, \dots, v_n\}$. Hence $\{v_1, \dots, v_n\}$ is a basis of V, as desired.

(ii) Suppose $\dim V = n$, $\{v_1, \dots, v_n\}$ spans V. By Lemma 5.24, $\{v_1, \dots, v_n\}$ can be reduced to a basis of V. However, every basis of V has length n, which means that no elements are deleted from $\{v_1, \dots, v_n\}$. Hence $\{v_1, \dots, v_n\}$ is a basis of V, as desired.

□

Corollary 5.32. Suppose V is finite-dimensional, $U \leqslant V$. If $\dim U = \dim V$, then $U = V$.

Proof. Let $\dim U = \dim V = n$, let $\{u_1, \ldots, u_n\}$ be a basis of U. Then $\{u_1, \ldots, u_n\}$ is linearly independent in V (because it is a basis of U) of length $\dim V$. From Proposition 5.31, $\{u_1, \ldots, u_n\}$ is a basis of V. In particular every vector in V is a linear combination of $u_1, \ldots, u_n$. Thus $U = V$. □

Lemma 5.33 (Dimension of sum)**.** Suppose V is finite-dimensional, $U_1, U_2 \leqslant V$. Then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Proof. Let $\{u_1, \ldots, u_m\}$ be a basis of $U_1 \cap U_2$; thus $\dim(U_1 \cap U_2) = m$. Since $\{u_1, \ldots, u_m\}$ is a basis of $U_1 \cap U_2$, it is linearly independent in U_1. By Lemma 5.26, $\{u_1, \ldots, u_m\}$ can be extended to a basis $\{u_1, \ldots, u_m, v_1, \ldots, v_j\}$ of U_1; thus $\dim U_1 = m + j$. Similarly, extend $\{u_1, \ldots, u_m\}$ to a basis $\{u_1, \ldots, u_m, v_1, \ldots, v_k\}$ of U_2; thus $\dim U_2 = m + k$.

We will show that

$$\{u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k\}$$

is a basis of $U_1 + U_2$. This will complete the proof because then we will have

$$\begin{aligned}\dim(U_1 + U_2) &= m + j + k \\ &= (m + j) + (m + k) - m \\ &= \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).\end{aligned}$$

We just need to show that $\{u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k\}$ is linearly independent. To prove this, suppose

$$a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j + c_1w_1 + \cdots + c_kw_k = \mathbf{0}, \tag{1}$$

where $a_i, b_i, c_i \in \mathbf{F}$. We need to show that $a_i = b_i = c_i = 0$ for all i. (1) can be rewritten as

$$c_1w_1 + \cdots + c_kw_k = -a_1u_1 - \cdots - a_mu_m - b_1v_1 - \cdots - b_jv_j,$$

which shows that $c_1w_1 + \cdots + c_kw_k \in U_1$. But actually all the w_i's are in U_2, so $c_1w_1 + \cdots + c_kw_k \in U_2$, thus $c_1w_1 + \cdots + c_kw_k \in U_1 \cap U_2$. Then we can write

$$c_1w_1 + \cdots + c_kw_k = d_1u_1 + \cdots + d_mu_m$$

for some $d_i \in \mathbf{F}$. But $u_1, \ldots, u_m, w_1, \ldots, w_k$ are linearly independent, so $c_i = d_i = 0$ for all i. Thus our original equation (1) becomes

$$a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j = \mathbf{0}.$$

Since $u_1, \ldots, u_m, v_1, \ldots, v_j$ are linearly independent, we have $a_i = b_i = 0$ for all i, as desired. □

Exercises

Problem 5.1 ([Axl24] 1C Q12)**.** Suppose W is a vector space over $\mathbf{F}$, V_1 and V_2 are subspaces of W. Show that $V_1 \cup V_2$ is a vector space over $\mathbf{F}$ if and only if $V_1 \subset V_2$ or $V_2 \subset V_1$.

Solution. The backward direction is trivial. We focus on proving the forward direction.

Supppse otherwise, then $V_1 \setminus V_2 \neq \emptyset$ and $V_2 \setminus V_1 \neq \emptyset$. Pick $v_1 \in V_1 \setminus V_2$ and $v_2 \in V_2 \setminus V_1$. Then

$$\begin{aligned} v_1, v_2 \in V_1 \cup V_2 &\implies v_1 + v_2 \in V_1 \cup V_2 \\ &\implies v_2, v_1 + v_2 \in V_2 \\ &\implies v_1 = (v_1 + v_2) - v_2 \in V_2 \end{aligned}$$

which is a contradiction. □

Problem 5.2 ([Axl24] 1C Q13)**.** Suppose W is a vector space over $\mathbf{F}$, V_1, V_2, V_3 are subspaces of W. Then $V_1 \cup V_2 \cup V_3$ is a vector space over $\mathbf{F}$ if and only if one of the V_i contains the other two.

Solution. We prove the forward direction. Suppose otherwise, then $v_1 \in V_1 \setminus (V_2 + V_3)$, $v_2 \in V_2 \setminus (V_1 + V_3)$, $v_3 \in V_3 \setminus (V_1 + V_2)$. Consider

$$\{v_1 + v_2 + v_3, v_1 + v_2 + 2v_3, v_1 + 2v_2 + v_3, v_1 + 2v_2 + 2v_3\} \subset V_1 \cup V_2 \cup V_3$$

Then

$$\begin{aligned} &(v_1 + v_2 + 2v_3) - (v_1 + v_2 + v_3) = v_3 \notin V_1 + V_2 \\ &\implies v_1 + v_2 + v_3 \notin V_1 + V_2 \quad \text{or} \quad v_1 + v_2 + 2v_3 \notin V_1 + V_2 \\ &\implies v_1 + v_2 + v_3 \in V_3 \quad \text{or} \quad v_1 + v_2 + 2v_3 \in V_3 \\ &\implies v_1 + v_2 \in V_3 \end{aligned}$$

Similarly,

$$\begin{aligned} &(v_1 + 2v_2 + 2v_3) - (v_1 + 2v_2 + v_3) = v_3 \notin V_1 + V_2 \\ &\implies v_1 + 2v_2 + v_3 \notin V_1 + V_2 \quad \text{or} \quad v_1 + 2v_2 + 2v_3 \notin V_1 + V_2 \\ &\implies v_1 + 2v_2 + v_3 \in V_3 \quad \text{or} \quad v_1 + 2v_2 + 2v_3 \in V_3 \\ &\implies v_1 + 2v_2 \in V_3 \end{aligned}$$

This implies $(v_1 + 2v_2) - (v_1 + v_2) = v_2 \in V_3$, a contradiction. □

Problem 5.3 ([Axl24] 2A Q12)**.** Suppose $\{v_1, \ldots, v_n\}$ is linearly independent in V, $w \in V$. Prove that if $\{v_1 + w, \ldots, v_n + w\}$ is linearly dependent, then $w \in \operatorname{span}(v_1, \ldots, v_n)$.

Solution. If $\{v_1 + w, \ldots, v_n + w\}$ is linearly dependent, then there exists $a_1, \ldots, a_n \in \mathbf{F}$, not all zero, such that

$$a_1(v_1 + w) + \cdots + a_n(v_n + w) = 0,$$

or

$$a_1v_1 + \cdots + a_nv_n = -(a_1 + \cdots + a_n)w.$$

Suppose otherwise, that $a_1 + \cdots + a_n = 0$. Then

$$a_1v_1 + \cdots + a_nv_n = \mathbf{0},$$

but the linear independence of $\{v_1, \dots, v_n\}$ implies that $a_1 = \cdots = a_n = 0$, which is a contradiction. Hence we must have $a_1 + \cdots + a_n \neq 0$, so we can write

$$w = -\frac{a_1}{a_1 + \cdots + a_n}v_1 - \cdots - \frac{a_n}{a_1 + \cdots + a_n}v_n,$$

which is a linear combination of $v_1, \dots, v_n$. Thus by definition of span, $w \in \operatorname{span}(v_1, \dots, v_n)$. □

Problem 5.4 ([Axl24] 2A Q14)**.** Suppose $\{v_1, \dots, v_n\} \subset V$. Let

$$w_i = v_1 + \cdots + v_i \quad (i = 1, \dots, n)$$

Show that $\{v_1, \dots, v_n\}$ is linearly independent if and only if $\{w_1, \dots, w_n\}$ is linearly independent.

Solution. Write

$$\begin{aligned} v_1 &= w_1 \\ v_2 &= w_2 - w_1 \\ v_3 &= w_3 - w_2 \\ &\vdots \\ v_n &= w_n - w_{n-1}. \end{aligned}$$

$\boxed{\implies}$

$$a_1w_1 + \cdots + a_nw_n = \mathbf{0}$$

for some $a_i \in \mathbf{F}$. Expressing w_i's as v_i's,

$$a_1v_1 + a_2(v_1 + v_2) + \cdots + a_n(v_1 + \cdots + v_n) = 0,$$

or

$$(a_1 + \cdots + a_n)v_1 + (a_2 + \cdots + a_n)v_2 + \cdots + a_nv_n = \mathbf{0}.$$

Since $v_1, \dots, v_n$ are linearly independent,

$$\begin{aligned} a_1 + a_2 + \cdots + a_n &= 0 \\ a_2 + \cdots + a_n &= 0 \\ &\vdots \\ a_n &= 0 \end{aligned}$$

on solving simultaneously gives $a_1 = \cdots = a_n = 0$.

$\boxed{\Longleftarrow}$ Similar to the above. $\square$

Problem 5.5 ([Axl24] 2A Q18)**.** Prove that $\mathbf{F}^\infty$ is infinite-dimensional.

Solution. To prove that $\mathbf{F}^\infty$ has no finite spanning sets, we prove by contradiction. Suppose otherwise, that there exists a finite spanning set of $\mathbf{F}^\infty$, say $\{v_1, \dots, v_n\}$.

Let

$$\begin{aligned} e_1 &= (1, 0, \dots) \\ e_2 &= (0, 1, 0, \dots) \\ e_3 &= (0, 0, 1, 0, \dots) \\ &\vdots \\ e_{n+1} &= (0, \dots, 0, 1, 0, \dots) \end{aligned}$$

where e_i has a 1 at the i-th coordinate, and 0's for the remaining coordinates. Let

$$a_1 e_1 + \cdots + a_{n+1} e_{n+1} = \mathbf{0}$$

for some $a_i \in \mathbf{F}$. Then

$$(a_1, a_2, \dots, a_{n+1}, 0, 0, \dots) = \mathbf{0}$$

so $a_1 = a_2 = \cdots = a_{n+1} = 0$. Thus $\{e_1, \dots, e_{n+1}\}$ is a linearly independent set, of length $n+1$. However, $\{v_1, \dots, v_n\}$ is a spanning set of length n. By Proposition 5.20, we have reached a contradiction. $\square$

Problem 5.6 ([Axl24] 2B Q5)**.** Suppose V is finite-dimensional, $U, W \leqslant V$ such that $V = U + W$. Prove that V has a basis in $U \cup W$.

Solution. Let $\{v_i\}_{i=1}^n$ denote the basis for V. By definition we have $v_i = u_i + w_i$ for some $u_i \in U$, $w_i \in W$. Then we have the spanning set of the vector space V $\sum_{i=1}^n a_i(u_i + w_i)$, which can be reduced to a basis by the lemma. $\square$

Problem 5.7 ([Axl24] 2B Q7)**.** Suppose $\{v_1, v_2, v_3, v_4\}$ is a basis of V. Prove that

$$\{v_1 + v_2, v_2 + v_3, v_3 + v_4, v_4\}$$

is also a basis of V.

Solution. We know that $\{v_1, v_2, v_3, v_4\}$ is linearly independent and spans V. Then there exist $a_i \in \mathbf{F}$ such that

$$a_1(v_1 + v_2) + a_2(v_2 + v_3) + a_3(v_3 + v_4) + a_4 v_4 = 0 \implies a_1 = a_2 = a_3 = a_4 = 0.$$

Write

$$\begin{aligned} &a_1(v_1 + v_2) + a_2(v_2 + v_3) + a_3(v_3 + v_4) + a_4 v_4 \\ &= a_1 v_1 + (a_1 + a_2) v_2 + (a_2 + a_3) v_3 + (a_3 + a_4) v_4, \end{aligned}$$

this shows the linear independence. To prove spanning, let $v \in V$, then

$$\begin{aligned} v &= a_1 v_1 + a_2 v_2 + a_3 v_3 + a_4 v_4 \\ &= a_1(v_1 + v_2) + (a_2 - a_1)(v_2 + v_3) + (a_3 - a_2)(v_3 + v_4) + (a_4 - a_3) v_4, \end{aligned}$$

which is a linear combination of $v_1 + v_2, v_2 + v_3, v_3 + v_4, v_4$. □

Problem 5.8 ([Axl24] 2B Q10)**.** Suppose $U, W \leqslant V$ such that $V = U \oplus W$. Suppose also that $\{u_1, \dots, u_m\}$ is a basis of U, $\{w_1, \dots, w_n\}$ is a basis of W. Prove that

$$\{u_1, \dots, u_m, w_1, \dots, w_n\}$$

is a basis of V.

Solution. We know that this set is linearly independent (otherwise violating the direct sum assumption) so it suffices to prove the spanning. Let $v \in V$, then

$$v = u + w = \sum_{i=1}^{m} a_i u_i + \sum_{i=1}^{n} b_j w_j.$$

□

Problem 5.9 ([Axl24] 2C Q8)**.**

Problem 5.10 ([Axl24] 2C Q16)**.**

Problem 5.11 ([Axl24] 2C Q17)**.** Suppose that $V_1, \dots, V_n \leqslant V$ are finite-dimensional. Prove that $V_1 + \cdots + V_n$ is finite-dimensional, and

$$\dim(V_1 + \cdots + V_n) \leqslant \dim V_1 + \cdots + \dim V_n.$$

Solution. We prove by induction on n. The base case is trivial. Assume the statement holds for k. Then for $k+1$, denoting $V_1 + \cdots + V_k = M_k$, we have that

$$\dim(M_k + V_{k+1}) \leqslant \dim V_1 + \cdots + \dim V_{k+1},$$

which is finite. $\square$

6 Linear Maps

§6.1 Vector Space of Linear Maps

Definition 6.1 (Linear map)**.** A **linear map** from V to W is a function $T : V \to W$ satisfying the following properties:

(i) Additivity: $T(v + w) = Tv + Tw$ for all $v, w \in V$

(ii) Homogeneity: $T(\lambda v) = \lambda T(v)$ for all $\lambda \in \mathbf{F}$, $v \in V$

Notation. The set of linear maps from V to W is denoted by $\mathcal{L}(V, W)$; the set of linear maps on V (from V to V) is denoted by $\mathcal{L}(V)$.

The existence part of the next result means that we can find a linear map that takes on whatever values we wish on the vectors in a basis. The uniqueness part of the next result means that a linear map is completely determined by its values on a basis.

Lemma 6.2 (Linear map lemma)**.** Suppose $\{v_1, \dots, v_n\}$ is a basis of V, and $w_1, \dots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that

$$Tv_i = w_i \quad (i = 1, \dots, n)$$

Proof. First we show the existence of a linear map T with the desired property. Define $T : V \to W$ by

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n,$$

for some $c_i \in \mathbf{F}$. Since $\{v_1, \dots, v_n\}$ is a basis of V, by Lemma 5.23, each $v \in V$ can be uniquely expressed as a linear combination of $v_1, \dots, v_n$, thus the equation above does indeed define a function $T : V \to W$. For i ($1 \leqslant i \leqslant n$), take $c_i = 1$ and the other c's equal to 0, then

$$T\left(0v_1 + \cdots + 1v_i + \cdots + 0v_n\right) = 0w_1 + \cdots + 1w_i + \cdots + 0w_n$$

which shows that $Tv_i = w_i$.

We now show that $T : V \to W$ is a linear map:

(i) For $u, v \in V$ with $u = a_1v_1 + \cdots + a_nv_n$ and $c_1v_1 + \cdots + c_nv_n$,

$$\begin{aligned} T(u+v) &= T\left((a_1+c_1)v_1 + \cdots + (a_n+c_n)v_n\right) \\ &= (a_1+c_1)w_1 + \cdots + (a_n+c_n)w_n \\ &= (a_1w_1 + \cdots + a_nw_n) + (c_1w_1 + \cdots + c_nw_n) \\ &= Tu + Tv. \end{aligned}$$

(ii) For $\lambda \in \mathbf{F}$ and $v = c_1v_1 + \cdots + c_nv_n$,

$$\begin{aligned} T(\lambda v) &= T(\lambda c_1v_1 + \cdots + \lambda c_nv_n) \\ &= \lambda c_1w_1 + \cdots + \lambda c_nw_n \\ &= \lambda(c_1w_1 + \cdots + c_nw_n) \\ &= \lambda Tv. \end{aligned}$$

To prove uniqueness, now suppose that $T \in \mathcal{L}(V, W)$ and $Tv_i = w_i$ for $i = 1, \ldots, n$. Let $c_i \in \mathbf{F}$. The homogeneity of T implies that $T(c_iv_i) = c_iw_i$. The additivity of T now implies that

$$T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n.$$

Thus T is uniquely determined on $\text{span}\{v_1, \ldots, v_n\}$. Since $\{v_1, \ldots, v_n\}$ is a basis of V, this implies that T is uniquely determined on V. □

Proposition 6.3. $\mathcal{L}(V, W)$ is a vector space, with the operations addition and scalar multiplication defined as follows: suppose $S, T \in \mathcal{L}(V, W)$, $\lambda \in \mathbf{F}$,

(i) $(S + T)(v) = Sv + Tv$

(ii) $(\lambda T)(v) = \lambda(Tv)$

for all $v \in V$.

Proof. Exercise. □

Definition 6.4 (Product of linear maps)**.** $T \in \mathcal{L}(U, V)$, $S \in \mathcal{L}(V, W)$, then the **product** $ST \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu) \quad (\forall u \in U)$$

Remark. In other words, ST is just the usual composition $S \circ T$ of two functions.

Remark. ST is defined only when T maps into the domain of S.

Proposition 6.5 (Algebraic properties of products of linear maps)**.**

(i) Associativity: $(T_1T_2)T_3 = T_1(T_2T_3)$ for all linear maps T_1, T_2, T_3 such that the products make sense (meaning that T_3 maps into the domain of T_2, T_2 maps into the domain of T_1)

(ii) Iidentity: $TI = IT = T$ for all $T \in \mathcal{L}(V, W)$ (the first I is the identity map on V, and the second I is the identity map on W)

(iii) Distributive: $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$ for all $T, T_1, T_2 \in \mathcal{L}(U, V)$ and $S, S_1, S_2 \in \mathcal{L}(V, W)$

Proof. Exercise. $\square$

Proposition 6.6. Suppose $T \in \mathcal{L}(V, W)$. Then $T(\mathbf{0}) = \mathbf{0}$.

Proof. By additivity, we have

$$T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0}).$$

Add the additive inverse of $T(\mathbf{0})$ to each side of the equation to conclude that $T(\mathbf{0}) = \mathbf{0}$. $\square$

§6.2 Kernel and Image

Definition 6.7 (Kernel)**.** Suppose $T \in \mathcal{L}(V, W)$. The **kernel** of T is the subset of V consisting of those vectors that T maps to $\mathbf{0}$:

$$\ker T := \{v \in V \mid Tv = \mathbf{0}\} \subset V.$$

Proposition 6.8. Suppose $T \in \mathcal{L}(V, W)$. Then $\ker T \leqslant V$.

Proof. By Lemma 5.6, we check the conditions of a subspace:

(i) By Proposition 6.6, $T(\mathbf{0}) = \mathbf{0}$, so $\mathbf{0} \in \ker T$.

(ii) For all $v, w \in \ker T$,

$$T(v + w) = Tv + Tw = \mathbf{0} \implies v + w \in \ker T$$

so $\ker T$ is closed under addition.

(iii) For all $v \in \ker T$, $\lambda \in \mathbf{F}$,

$$T(\lambda v) = \lambda Tv = \mathbf{0} \implies \lambda v \in \ker T$$

so $\ker T$ is closed under scalar multiplication.

□

Definition 6.9 (Injectivity)**.** Suppose $T \in \mathcal{L}(V, W)$. T is **injective** if

$$Tu = Tv \implies u = v.$$

Proposition 6.10. Suppose $T \in \mathcal{L}(V, W)$. Then T is injective if and only if $\ker T = \{\mathbf{0}\}$.

Proof.

$\boxed{\implies}$ Suppose T is injective. Let $v \in \ker T$, then

$$Tv = \mathbf{0} = T(\mathbf{0}) \implies v = \mathbf{0}$$

by the injectivity of T. Hence $\ker T = \{\mathbf{0}\}$ as desired.

$\boxed{\impliedby}$ Suppose $\ker T = \{\mathbf{0}\}$. Let $u, v \in V$ such that $Tu = Tv$. Then

$$T(u - v) = Tu - Tv = \mathbf{0}.$$

By definition of kernel, $u - v \in \ker T = \{\mathbf{0}\}$, so $u - v = \mathbf{0}$, which implies that $u = v$. Hence T is injective, as desired. □

Definition 6.11 (Image)**.** Suppose $T \in \mathcal{L}(V, W)$. The **image** of T is the subset of W consisting of those vectors that are of the form Tv for some $v \in V$:

$$\operatorname{im} T := \{Tv \mid v \in V\} \subset W.$$

Proposition 6.12. Suppose $T \in \mathcal{L}(V, W)$. Then $\operatorname{im} T \leqslant W$.

Proof.

(i) $T(\mathbf{0}) = \mathbf{0}$ implies that $\mathbf{0} \in \operatorname{im} T$.

(ii) For $w_1, w_2 \in \operatorname{im} T$, there exist $v_1, v_2 \in V$ such that $Tv_1 = w_1$ and $Tv_2 = w_2$. Then

$$w_1 + w_2 = Tv_1 + Tv_2 = T(v_1 + v_2) \in \operatorname{im} T \implies w_1 + w_2 \in \operatorname{im} T.$$

(iii) For $w \in \operatorname{im} T$ and $\lambda \in \mathbf{F}$, there exists $v \in V$ such that $Tv = w$. Then

$$\lambda w = \lambda T v = T(\lambda v) \in \operatorname{im} T \implies \lambda w \in \operatorname{im} T.$$

□

Definition 6.13 (Surjectivity)**.** Suppose $T \in \mathcal{L}(V, W)$. T is **surjective** if $\operatorname{im} T = W$.

Fundamental Theorem of Linear Maps

Theorem 6.14 (Fundamental theorem of linear maps)**.** Suppose V is finite-dimensional, $T \in \mathcal{L}(V, W)$. Then $\operatorname{im} T$ is finite-dimensional, and

$$\dim V = \dim \ker T + \dim \operatorname{im} T. \tag{6.1}$$

Proof. Let $\{u_1, \dots, u_m\}$ be basis of $\ker T$, then $\dim \ker T = m$. The linearly independent list $u_1, \dots, u_m$ can be extended to a basis

$$\{u_1, \dots, u_m, v_1, \dots, v_n\}$$

of V, thus $\dim V = m + n$. To simultaneously show that $\operatorname{im} T$ is finite-dimensional and $\dim \operatorname{im} T = n$, we prove that $\{Tv_1, \dots, Tv_n\}$ is a basis of $\operatorname{im} T$. Thus we need to show that the set (i) spans $\operatorname{im} T$, and (ii) is linearly independent.

(i) Let $v \in V$. Since $\{u_1, \dots, u_m, v_1, \dots, v_n\}$ spans V, we can write

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_n v_n,$$

for some $a_i, b_i \in \mathbf{F}$. Applying T to both sides of the equation, and noting that $Tu_i = \mathbf{0}$ since $u_i \in \ker T$,

$$\begin{aligned} Tv &= T\left(a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_n v_n\right) \\ &= a_1 \underbrace{Tu_1}_{\mathbf{0}} + \cdots + a_m \underbrace{Tu_m}_{\mathbf{0}} + b_1 Tv_1 + \cdots + b_n v_n \\ &= b_1 Tv_1 + \cdots + b_n Tv_n \in \operatorname{im} T. \end{aligned}$$

Since every element of $\operatorname{im} T$ can be expressed as a linear combination of $Tv_1, \dots, Tv_n$, we have that $\{Tv_1, \dots, Tv_n\}$ spans $\operatorname{im} T$.

Moreover, since there exists a set of vectors that spans $\operatorname{im} T$, $\operatorname{im} T$ is finite-dimensional.

(ii) Suppose there exist $c_1, \dots, c_n \in \mathbf{F}$ such that

$$c_1 Tv_1 + \cdots + c_n Tv_n = \mathbf{0}.$$

Then

$$T(c_1 v_1 + \cdots + c_n v_n) = T(\mathbf{0}) = \mathbf{0},$$

which implies $c_1 v_1 + \cdots + c_n v_n \in \ker T$. Since $\{u_1, \dots, u_m\}$ is a spanning set of $\ker T$, we can write

$$c_1 v_1 + \cdots + c_n v_n = d_1 u_1 + \cdots + d_m u_m$$

for some $d_i \in \mathbf{F}$, or

$$c_1 v_1 + \cdots + c_n v_n - d_1 u_1 - \cdots - d_m u_m = \mathbf{0}.$$

Since $u_1, \dots, u_m, v_1, \dots, v_n$ are linearly independent, $c_i = d_i = 0$. Since $c_i = 0$, $\{Tv_1, \dots, Tv_n\}$ is linearly independent.

□

We now show that no linear map from a finite-dimensional vector space to a “smaller” vector space can be injective, where “smaller” is measured by dimension.

Proposition 6.15. Suppose V and W are finite-dimensional vector spaces, $\dim V > \dim W$. Then there does not exist $T \in \mathcal{L}(V, W)$ such that T is injective.

Proof. We first prove the following claim.

Claim. $\dim \operatorname{im} T \leqslant \dim W$.

Proof. This follows directly from Proposition 5.30, since W is finite-dimensional and $\operatorname{im} T \leqslant W$. □

Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \ker T &= \dim V - \dim \operatorname{im} T && (1) \\ &\geqslant \dim V - \dim W && (2) \\ &> 0 \end{aligned}$$

where (1) follows from the fundamental theorem of linear maps, (2) follows from the above claim.

Since $\dim \ker T > 0$. This means that $\ker T$ contains some $v \in V \setminus \{\mathbf{0}\}$. Since $\ker T \neq \{\mathbf{0}\}$, T is not injective. □

The next result shows that no linear map from a finite-dimensional vector space to a "bigger" vector space can be surjective, where "bigger" is also measured by dimension.

Proposition 6.16. Suppose V and W are finite-dimensional vector spaces, $\dim V < \dim W$. Then there does not exist $T \in \mathcal{L}(V, W)$ such that T is surjective.

Proof. Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \operatorname{im} T &= \dim V - \dim \ker T && (1) \\ &\leqslant \dim V && (2) \\ &< \dim W, \end{aligned}$$

where (1) follows from the fundamental theorem of linear maps, (2) follows since the dimension of a vector space is non-negative so $\dim \ker T \geqslant 0$.

Since $\dim \operatorname{im} T < \dim W$, $\operatorname{im} T \neq W$ so T is not surjective. $\square$

§6.3 Matrices

Representing a Linear Map by a Matrix

Definition 6.17 (Matrix)**.** Suppose $m, n \in \mathbf{N}$. An $m \times n$ **matrix** A is a rectangular array with m rows and n columns:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

where $a_{ij} \in \mathbf{F}$ denotes the entry in row i, column j. We also denote $A = (a_{ij})_{m\times n}$, and drop the subscript if there is no ambiguity.

Notation. i is used for indexing across the m rows, j is used for indexing across the n columns.

Notation. $\mathcal{M}_{m\times n}(\mathbf{F})$ denotes the set of $m \times n$ matrices with entries in $\mathbf{F}$.

As we will soon see, matrices provide an efficient method of recording the values of Tv_j's in terms of a basis of W.

Definition 6.18 (Matrix of a linear map)**.** Suppose $T \in \mathcal{L}(V, W)$, $\mathcal{V} = \{v_1, \dots, v_n\}$ is a basis of V, $\mathcal{W} = \{w_1, \dots, w_m\}$ is a basis of W. The matrix of T with respect to these bases is the $m \times n$ matrix $\mathcal{M}(T)$ whose entries a_{ij} are defined by

$$Tv_j = \sum_{i=1}^{m} a_{ij} w_i.$$

That is, the j-th column of $\mathcal{M}(T)$ consists of the scalars $a_{1j}, \dots, a_{mj}$ needed to write Tv_j as a linear combination of the bases of W.

Notation. If the bases of V and W are not clear from the context, we adopt the notation $\mathcal{M}_{\mathcal{V},\mathcal{W}}(T)$.

Matrix Operations

Definition 6.19 (Matrix operations)**.**

(i) Addition: the sum of two matrices of the same size is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} = \begin{pmatrix} a_{11}+c_{11} & \cdots & a_{1n}+c_{1n} \\ \vdots & & \vdots \\ a_{m1}+c_{m1} & \cdots & a_{mn}+c_{mn} \end{pmatrix}.$$

(ii) Scalar multiplication: the product of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

Proposition 6.20. Suppose $S, T \in \mathcal{L}(V, W)$. Then

(i) $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$;

(ii) $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$ for $\lambda \in \mathbf{F}$.

Proof.

(i)

(ii)

□

Proposition 6.21. With addition and scalar multiplication defined as above, $\mathcal{M}_{m\times n}(\mathbf{F})$ is a vector space of dimension mn.

Definition 6.22 (Matrix multiplication)**.** Given $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{n\times p}$. Then

$$AB = \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right)_{m\times p}$$

Notation. $A_{i,\cdot}$ denotes the row vector corresponding to the i-th row of A; $A_{\cdot,j}$ denotes the column vector corresponding to the j-th column of A.

Proposition 6.23. Suppose A is $m \times n$ matrix, B is $n \times p$ matrix. Then

$$(AB)_{j,k} = A_{j,\cdot} B_{\cdot,k}$$

for $i \leqslant j \leqslant m$, $1 \leqslant k \leqslant p$. In other words, the entry in row j, column k of AB equals (row j of A) times (column k of B).

Proposition 6.24. Suppose A is $m \times n$ matrix, B is $n \times p$ matrix. Then

$$(AB)_{\cdot,k} = AB_{\cdot,k}$$

for $1 \leqslant k \leqslant p$. In other words, column k of AB equals A times column k of B.

Proposition 6.25 (Linear combination of columns)**.** Suppose A is an $m \times n$ matrix, $b = \begin{pmatrix} b_1 \\ \dots \\ b_n \end{pmatrix}$. Then

$$Ab = b_1 A_{\cdot,1} + \cdots + b_n A_{\cdot,n}.$$

In other words, Ab is a linear combination of the columns of A, with the scalars that multiply the columns coming from b.

Proposition 6.26. If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.

Definition 6.27 (Column rank, row rank)**.** Given $A \in \mathcal{M}_{m\times n}(\mathbf{F})$, column vectors are

$$\{A(\cdot, 1), \dots, A(\cdot, n)\}$$

and row vectors are

$$\{A(1, \cdot), \dots, A(m, \cdot)\}.$$

Then we define the **column rank** c by

$$c = \dim \operatorname{span}\{A(\cdot, k) \mid 1 \leqslant k \leqslant n\}$$

and **row rank** r by

$$r = \dim \operatorname{span}\{A(k, \cdot) \mid 1 \leqslant k \leqslant m\}.$$

Definition 6.28 (Transpose)**.** Suppose $A \in \mathcal{M}_{m\times n}(\mathbf{F})$, then the **transpose** $A^T \in M_{n\times m}(\mathbf{F})$ is given by

$$A^T(i, j) = A(j, i).$$

Proposition 6.29 (Column-rank factorisation)**.** Suppose $A \in \mathcal{M}_{m\times n}(\mathbf{F})$, column rank $c \geqslant 1$. Then $A = CR$ where $C \in M_{m\times c}(\mathbf{F})$, $R \in M_{c\times n}(\mathbf{F})$.

Proposition 6.30 (Column rank equals row rank)**.** Suppose $A \in \mathcal{M}_{m\times n}(\mathbf{F})$. Then the column rank of A equals to row rank of A.

Since column rank equals row rank, we can dispense with the terms "column rank" and "row rank", and just use the simpler term "rank".

Definition 6.31 (Rank)**.** The **rank** of a matrix $A \in \mathcal{M}_{m\times n}(\mathbf{F})$ is the column rank of A.

§6.4 Invertibility and Isomorphism

Notation. $I_V \in \mathcal{L}(V)$ denotes the identity map on V; the subscript is omitted if there is no ambiguity. For all $v \in V$, $Iv = v$.

Definition 6.32 (Invertibility)**.** $T \in \mathcal{L}(V,W)$ is **invertible** if there exists $S \in \mathcal{L}(W,V)$ such that $ST = I_V$, $TS = I_W$; S is known as the *inverse* of T.

Proposition 6.33 (Uniqueness of inverse)**.** The inverse of an invertible linear map is unique.

Proof. Suppose $T \in \mathcal{L}(V,W)$ is invertible, $S_1, S_2 \in \mathcal{L}(W,V)$ are inverses of T. Then

$$\begin{aligned} S_1 &= S_1 I_W \\ &= S_1(TS_2) \\ &= (S_1T)S_2 \\ &= I_V S_2 \\ &= S_2. \end{aligned}$$

Thus $S_1 = S_2$. □

Now that we know that the inverse is unique, we can give it a notation.

Notation. If T is invertible, then its inverse is denoted by T^{-1}.

Proposition 6.34. Suppose $T \in \mathcal{L}(V,W)$. Then T is invertible if and only if it is injective and surjective.

Proof.

$\boxed{\Longrightarrow}$ Suppose $T \in \mathcal{L}(V,W)$, where $Tu = Tv$, is invertible. Then there exists an inverse T^{-1} such that

$$u = T^{-1}Tu = T^{-1}Tv = v.$$

To show T is surjective, we have that for any $w \in W$, $w = T(T^{-1}w)$.

$\boxed{\Longleftarrow}$ Define $S \in \mathcal{L}(W,V)$ such that for each $w \in W$, $S(W)$ is the unique element such that $T(S(w))w$ (we can do this due to injectivity and surjectivity). Then we have that $T(ST)v = (TS)Tv = Tv$ and thus $STv = v$ so $ST = I$. It is easy to show that S is a linear map. □

Proposition 6.35. Suppose that V and W are finite-dimensional, $\dim V = \dim W$, and $T \in \mathcal{L}(V,W)$. Then

$$T \text{ is invertible} \iff T \text{ is injective} \iff T \text{ is surjective}.$$

Corollary 6.36. Suppose V and W are finite-dimensional, $\dim V = \dim W$, $S \in \mathcal{L}(W,V)$, $T = \mathcal{L}(V,W)$. Then $ST = I$ if and only if $TS = I$.

Definition 6.37 (Isomorphism)**.** An **isomorphism** is an invertible linear map. V and W are **isomorphic**, denoted by $V \cong W$, if there exists an isomorphism $T \in \mathcal{L}(V,W)$.

Remark. Think of an isomorphism $T : V \to W$ as relabeling $v \in V$ as $Tv \in W$. This viewpoint explains why two isomorphic vector spaces have the same vector space properties. Isomorphism essentially means that two vector spaces are essentially the same.

The following result shows that we need to look at only at the dimension to determine whether two vector spaces are isomorphic.

Lemma 6.38. Two finite-dimensional vector spaces V and W are isomorphic if and only if they have the same dimension:

$$V \cong W \iff \dim V = \dim W.$$

Proposition 6.39. Suppose $\{v_1, \ldots, v_n\}$ is a basis of V, $\{w_1, \ldots, w_m\}$ is a basis of W. Then M is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$.

Corollary 6.40. Suppose V and W are finite-dimensional. Then $\mathcal{L}(V, W)$ is finite-dimensional and

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Previously we defned the matrix of a linear map. Now we defne the matrix of a vector.

Definition 6.41 (Matrix of a vector)**.** Suppose $v \in V$, $\{v_1, \ldots, v_n\}$ is a basis of V. The matrix of v with respect to this basis is the $n \times 1$ matrix

$$\mathcal{M}(v) = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

where $b_1, \ldots, b_n$ are scalars such that

$$v = b_1 v_1 + \cdots + b_n v_n.$$

Remark. The matrix $\mathcal{M}(v)$ of a vector $v \in V$ depends on the basis $\{v_1, \ldots, v_n\}$ and v. We can think of elements of V as relabelled to $n \times 1$ matrices, i.e., $V \to \mathbf{F}^{n,1}$.

Lemma 6.42. $T \in \mathcal{L}(V, W)$, $\{v_1, \ldots, v_n\}$ is basis of V, $\{w_1, \ldots, w_m\}$ is basis of W. Then $\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(Tv_k)$ for $k = 1, \ldots, n$.

The following result shows that linear maps act like matrix multiplication.

Proposition 6.43. $T \in \mathcal{L}(V, W)$, $v \in V$. $\{v_1, \ldots, v_n\}$ is basis of V, $\{w_1, \ldots, w_m\}$ is basis of W. Then

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v).$$

Proof. Applying the previous result,

$$\begin{aligned}\mathcal{M}(Tv) &= b_1\mathcal{M}(Tv_1) + \cdots + b_n\mathcal{M}(Tv_n)\\ &= b_1\mathcal{M}(T)_{\cdot,1} + \cdots + b_n\mathcal{M}(T)_{\cdot,n}\\ &= \mathcal{M}(T)\mathcal{M}(v).\end{aligned}$$

□

Proposition 6.44. Suppose V and W are finite-dimensional, $T \in \mathcal{L}(V, W)$. Then

$$\dim\ker T = \operatorname{rank}\mathcal{M}(T).$$

Definition 6.45 (Identity matrix)**.** For positive integer n, the $n \times n$ matrix

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

is called the **identity matrix**, denoted by I_n.

Definition 6.46 (Invertibility)**.** A square matrix A is called invertible if there is a square matrix B of the same size such that $AB = BA = I$; we call B the inverse of A and denote it by A^{-1}.

§6.5 Products and Quotients of Vector Spaces

Definition 6.47 (Product)**.** Suppose $V_1, \dots, V_n$ are vector spaces over $\mathbf{F}$. The **product** $V_1 \times \cdots \times V_n$ is defined by

$$V_1 \times \cdots \times V_n = \{(v_1, \dots, v_n) \mid v_i \in V_i\}.$$

Proposition 6.48. $V_1 \times \cdots \times V_n$ is a vector space over $\mathbf{F}$, with addition and scalar multiplication defined by

$$\begin{aligned}(u_1, \dots, u_n) + (v_1, \dots, v_n) &= (u_1 + v_1, \dots, u_n + v_n)\\ \lambda(v_1, \dots, v_n) &= (\lambda v_1, \dots, \lambda v_n)\end{aligned}$$

The following result shows that the dimension of a product is the sum of dimensions.

Proposition 6.49. Suppose $V_1, \dots, V_n$ are finite-dimensional vector spaces. Then $V_1 \times \cdots \times V_n$ is finite-dimensional and

$$\dim(V_1 \times \cdots \times V_n) = \dim V_1 + \cdots + \dim V_n.$$

Products are also related to direct sums, by the following result.

Lemma 6.50. Suppose that $V_1, \dots, V_n$ are subspaces of V. Define a linear map $\Gamma : V_1 \times \cdots \times V_n \to V_1 + \cdots + V_n$ by

$$\Gamma(v_1, \dots, v_n) = v_1 + \cdots + v_n.$$

Then $V_1 + \cdots + V_n$ is a direct sum if and only if Γ is injective.

The next result says that a sum is a direct sum if and only if dimensions add up.

Proposition 6.51. Suppose V is finite-dimensional and $V_1, \dots, V_n$ are subspaces of V. Then $V_1 + \cdots + V_n$ is a direct sum if and only if

$$\dim(V_1 + \cdots + V_n) = \dim V_1 + \cdots + \dim V_n.$$

Definition 6.52 (Coset)**.** Suppose $v \in V$, $U \subset V$. Then $v + U$ is called a **coset** of U, defined by

$$v + U := \{v + u \mid u \in U\}.$$

Definition 6.53 (Quotient space)**.** Suppose $U \leqslant V$. Then the **quotient space** V/U is the set of all cosets of U:

$$V/U := \{v + U \mid v \in V\}.$$

> **Example**
> If $U = \{(x, 2x) \in \mathbf{R}^2 \mid x \in \mathbf{R}\}$, then $\mathbf{R}^2/U$ is the set of all lines in $\mathbf{R}^2$ that have gradient of 2.

It is obvious that two translates of a subspace are equal or disjoint. We shall now prove this.

Proposition 6.54. Suppose $U \leqslant V$, and $v, w \in V$. Then

Proposition 6.55. Suppose $U \leqslant V$. Then V/U is a vector space, with addition and scalar multiplication defined by

$$(v + U) + (w + U) = (v + w) + U$$
$$\lambda(v + U) = (\lambda v) + U$$

for all $v, w \in V$, $\lambda \in \mathbf{F}$.

Definition 6.56 (Quotient map)**.** Suppose $U \leqslant V$. The **quotient map** $\pi : V \to V/U$ is the linear map defined by

$$\pi(v) = v + U$$

for all $v \in V$.

Proposition 6.57 (Dimension of quotient space)**.** Suppose V is finite-dimensional, $U \leqslant V$. Then

$$\dim V/U = \dim V - \dim U.$$

Definition 6.58. Suppose $T \in \mathcal{L}(V, W)$. Define $\tilde{T} : V/\ker T \to W$ by

$$\tilde{T}(v + \ker T) = Tv.$$

§6.6 Duality

Dual Space and Dual Map

Linear maps into the scalar field $\mathbf{F}$ play a special role in linear algebra, and thus they get a special name.

Definition 6.59 (Linear funtional)**.** A **linear functional** on V is a linear map from V to $\mathbf{F}$; that is, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

The vector space $\mathcal{L}(V, \mathbf{F})$ also gets a special name and special notation.

Definition 6.60 (Dual space)**.** The **dual space** of V, denoted by V^*, is the vector space of all linear functionals on V; that is, $V^* = \mathcal{L}(V, \mathbf{F})$.

Lemma 6.61. Suppose V is finite-dimensional. Then V^* is also finite-dimensional, and

$$\dim V^* = \dim V.$$

Proof. By , we have

$$\dim V^* = \dim \mathcal{L}(V, \mathbf{F}) = (\dim V)(\dim \mathbf{F}) = \dim V$$

as desired. □

Definition 6.62 (Dual basis)**.** If $\{v_1, \dots, v_n\}$ is a basis of V, then the **dual basis** of $\{v_1, \dots, v_n\}$ is the list $\{\phi_1, \dots, \phi_n\}$ of elements of V^*, where each ϕ_i is the linear functional on V such that

$$\phi_i(v_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The following result states that dual basis gives coefficients for linear combination.

Proposition 6.63. Suppose $\{v_1, \dots, v_n\}$ is a basis of V, and $\{\phi_1, \dots, \phi_n\}$ is the dual basis. Then for each $v \in V$,

$$v = \phi_1(v)v_1 + \cdots + \phi_n(v)v_n.$$

The following result states that the dual basis is a basis of the dual space.

Proposition 6.64. Suppose V is finite-dimensional. Then the dual basis of a basis of V is a basis of V^*.

Definition 6.65 (Dual map)**.** Suppose $T \in \mathcal{L}(V, W)$. The **dual map** of T is the linear map $T^* \in \mathcal{L}(V, W)$ defined for each $\phi \in W^*$ by

$$T^*(\phi) = \phi \circ T.$$

Proposition 6.66 (Algebraic properties of dual map)**.** Suppose $T \in \mathcal{L}(V, W)$. Then

(1) $(S+T)^* = S^* + T^*$ for all $S \in \mathcal{L}(V,W)$

(2) $(\lambda T)^* = \lambda T^*$ for all $\lambda \in \mathbf{F}$

(3) $(ST)^* = T^* S^*$ for all $S \in \mathcal{L}(V,W)$

Kernel and Image of Dual of Linear Map

Our goal in this subsection is to describe $\ker T^*$ and $\operatorname{im} T^*$ in terms of $\operatorname{im} T$ and $\ker T$. To do this, we will need the next definition.

Definition 6.67 (Annihilator)**.** For $U \subset V$, the **annihilator** of U is defined by

$$U^\circ \coloneqq \{\phi \in V^* \mid \phi(u) = 0, \forall u \in U\}.$$

From the definition, you can probably see that the name is rather fitting.

Proposition 6.68. $U^\circ \leqslant V$.

Proposition 6.69 (Dimension of annihilator)**.** Suppose V is finite-dimensional, $U \leqslant V$. Then

$$\dim U^\circ = \dim V - \dim U.$$

The following are conditions for the annihilator to equal $\{\mathbf{0}\}$ or the whole space.

Proposition 6.70. Suppose V is finite-dimensional, $U \leqslant V$. Then

(i) $U^\circ = \{\mathbf{0}\} \iff U = V$;

(ii) $U^\circ = V^* \iff U = \{\mathbf{0}\}$.

The following result concerns $\ker T^*$. Note that the proof of (1) does not use the hypothesis that V and W are finite-dimensional.

Proposition 6.71. Suppose V and W are finite-dimensional, $T \in \mathcal{L}(V,W)$. Then

(i) $\ker T^* = \operatorname{Ann}(\operatorname{im} T)$;

(ii) $\dim \ker T^* = \dim \ker T + \dim W - \dim V$.

The next result can be useful because sometimes it is easier to verify that T^* is injective than to show directly that T is surjective.

Proposition 6.72. Suppose V and W are finite-dimensional, $T \in \mathcal{L}(V,W)$. Then

$$T \text{ is surjective} \iff T^* \text{ is injective.}$$

The following result concerns $\operatorname{im} T^*$.

Proposition 6.73. Suppose V and W finite-dimensional, $T \in \mathcal{L}(V, W)$. Then

(i) $\dim \operatorname{im} T^* = \dim \operatorname{im} T$;

(ii) $\dim T^* = \operatorname{Ann}(\ker T)$.

Proposition 6.74. Suppose V and W are finite-dimensional, $T \in \mathcal{L}(V, W)$. Then

$$T \text{ is injective } \iff T^* \text{ is surjective.}$$

Matrix of Dual of Linear Map

Proposition 6.75. Suppose V and W are finite-dimensional, $T \in \mathcal{L}(V, W)$. Then

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^t.$$

Problems

Problem 6.1 ([Axl24] 3A)**.** Suppose $b, c \in \mathbf{R}$. Define $T : \mathbf{R}^3 \to \mathbf{R}^2$ by

$$T(x, y, z) = (2x - 4y + 3z + b, 6x + cxyz).$$

Show that T is linear if and only if $b = c = 0$.

Problem 6.2 ([Axl24] 3A Q11)**.** Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$. Prove that T is a scalar multiple of the identity if and only if $ST = TS$ for all $S \in \mathcal{L}(V)$.

Problem 6.3 ([Axl24] 3D)**.** Suppose $T \in \mathcal{L}(V, W)$ is invertible. Show that T^{-1} is invertible and

$$\left(T^{-1}\right)^{-1} = T.$$

Solution. T^{-1} is invertible because there exists T such that $TT^{-1} = T^{-1}T = I$. So

$$T^{-1}T = TT^{-1} = I$$

thus $\left(T^{-1}\right)^{-1} = T$. □

Problem 6.4 ([Axl24] 3D)**.** Suppose $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$ are both invertible linear maps. Prove that $ST \in \mathcal{L}(U, W)$ is invertible and that $(ST)^{-1} = T^{-1}S^{-1}$.

Solution.

$$(ST)(T^{-1}S^{-1}) = S(TT^{-1})S^{-1} = I = T^{-1}S^{-1}ST.$$

□

Problem 6.5 ([Axl24] 3D)**.** Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that the following are equivalent:

(i) T is invertible;

(ii) $\{Tv_1, \dots, Tv_n\}$ is a basis of V for every basis $\{v_1, \dots, v_n\}$ of V;

(iii) $\{Tv_1, \dots, Tv_n\}$ is a basis of V for some basis $\{v_1, \dots, v_n\}$ of V.

Solution.

(i) $\Longrightarrow$ (ii) It only suffices to prove linear independence. We can show this

$$a_1Tv_1 + \cdots + a_nTv_n = 0 \iff a_1v_1 + \cdots + a_nv_n = 0$$

since T is injective and thus the only solution is all a_i are identically zero.

(ii) $\Longrightarrow$ (iii) Trivial.

(iii) $\Longrightarrow$ (i) By the linear map lemma, there exists $S \in \mathcal{L}(V)$ such that $S(Tv_i) = v_i$ for all i. Such S is the inverse of T (one can verify) and thus T is invertible. □

Problem 6.6 ([Axl24] 3E)**.** Suppose $U \leqslant V$, V/U is finite-dimensional. Prove that $V \cong U \times (V/U)$.

Solution.

$$\dim V = \dim U + \big(\dim V - \dim U\big) = \dim U + \dim(V/U).$$

$\square$

7 Eigenvalues and Eigenvectors

§7.1 Invariant Subspaces

Eigenvalues

Definition 7.1 (Operator)**.** A linear map from a vector space to itself is called an **operator**.

Definition 7.2 (Invariant subspace)**.** Suppose $T \in \mathcal{L}(V)$. $U \leqslant V$ is called **invariant** under T if $Tu \in U$ for all $u \in U$.

Definition 7.3 (Eigenvalue and eigenvector)**.** Suppose $T \in \mathcal{L}(V)$. $\lambda \in \mathbf{F}$ is called an **eigenvalue** of T if there exists $v \in V \setminus \{\mathbf{0}\}$ such that $Tv = \lambda v$; v is called an **eigenvector** of T corresponding to λ.

Lemma 7.4 (Equivalent conditions to be an eigenvalue)**.** Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{F}$. Then the following are equivalent:

(i) λ is an eigenvalue of T.

(ii) $T - \lambda I$ is not injective.

(iii) $T - \lambda I$ is not surjective.

(iv) $T - \lambda I$ is not invertible.

Proof.

$\boxed{\text{(i)} \iff \text{(ii)}}$ $Tv = \lambda v$ is equivalent to the equation $(T - \lambda I)v = \mathbf{0}$, so $T - \lambda I$ is not injective.

$\boxed{\text{(ii)} \iff \text{(iii)} \iff \text{(iv)}}$ This directly follows from □

Proposition 7.5 (Linearly independent eigenvectors)**.** Suppose $T \in \mathcal{L}(V)$. Then every list of eigenvectors of T corresponding to distinct eigenvalues of T is linearly independent.

Proposition 7.6. Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Polynomials Applied to Operators

Notation. Suppose $T \in \mathcal{L}(V)$, $n \in \mathbf{Z}^+$. $T^n \in \mathcal{L}(V)$ is defined by $T^n = \underbrace{T \cdots T}_{m \text{ times}}$. T^0 is defined to be the identity operator I on V. If T is invertible with inverse T^{-1}, then $T^{-n} \in \mathcal{L}(V)$ is defined by $T^{-n} = \left(T^{-1}\right)^n$.

§7.2 The Minimal Polynomial

§7.3 Upper-Triangular Matrices

§7.4 Diagonalisable Operators

§7.5 Commuting Operators

III

Real Analysis

8 Real and Complex Number Systems

Learning Outcomes

In this chapter, we will

- Discuss the construction and properties of the real field $\mathbf{R}$.
- Discuss the construction and properties of the complex field $\mathbf{C}$.
- Discuss the construction and properties of the Euclidean space $\mathbf{R}^n$.

§8.1 Real Numbers

$\mathbf{Q}$ has some problems, the first of which being *algebraic incompleteness*: there exists equations with coefficients in $\mathbf{Q}$ but do not have solutions in $\mathbf{Q}$ (in fact $\mathbf{R}$ has this problem too, but $\mathbf{C}$ is algebraically complete, by the Fundamental Theorem of Algebra).

Lemma 8.1. $x^2 - 2 = 0$ has no solution in $\mathbf{Q}$.

Proof. Suppose, for a contradiction, that $x^2 - 2 = 0$ has a solution $x = \frac{p}{q}$, $q \neq 0$. We also assume $\frac{p}{q}$ is in lowest terms; that is, p, q are coprime. Squaring both sides gives $\frac{p^2}{q^2} = 2$, or $p^2 = 2q^2$. Observe that p^2 is even, so p is even; let $p = 2m$ for some integer m. Then this implies $4m^2 = 2q^2$, or $2m^2 = q^2$. Similarly, q^2 is even so q is even.

Since p and q share a common factor of 2, we have reached a contradiction. $\square$

The second problem is *analytic incompleteness*: there exists a sequence of rational numbers that approach a point that is not in $\mathbf{Q}$; for example, the sequence

$$1, 1.4, 1.41, 1.414, 1.4142, \dots$$

tends to the the irrational number $\sqrt{2}$.

Continuing from the above lemma,

Lemma 8.2. Let

$$\begin{aligned} A &= \{p \in \mathbf{Q} \mid p > 0, p^2 < 2\}, \\ B &= \{p \in \mathbf{Q} \mid p > 0, p^2 > 2\}. \end{aligned}$$

Then A contains no largest number, and B contains no smallest number.

Proof. Prove by construction. We associate with each rational $p > 0$ the number

$$q = p - \frac{p^2 - 2}{p + 2} = \frac{2p + 2}{p + 2}$$

and so

$$q^2 - 2 = \frac{2(p^2 - 2)}{(p + 2)^2}.$$

For any $p \in A$, $q > p$ and $q \in A$ since $q^2 < 2$, so A has no largest number.

For any $p \in B$, $q < p$ and $q \in B$ since $q^2 > 2$, so B has no smallest number. □

A direct consequence of this is that $\mathbf{Q}$ does not have the least-upper-bound property, for $A \subset \mathbf{Q}$ is bounded above but A has no least upper bound in $\mathbf{Q}$ [B is the set of all upper bounds of A, and B does not have a smallest element].

Real Field

Definition 8.3 (Ordered field)**.** A field F is an **ordered field** if there eists an order $<$ on F such that for all $x, y, z \in F$,

(i) if $y < z$ then $x + y < x + z$;

(ii) if $x > 0$ and $y > 0$ then $xy > 0$.

Proposition 8.4 (Basic properties)**.** The following statements are true in every ordered field.

(i) If $x > 0$ then $-x < 0$, and vice versa.

(ii) If $x > 0$ and $y < z$ then $xy < xz$.

(iii) If $x < 0$ and $y < z$ then $xy > xz$.

(iv) If $x \neq 0$ then $x^2 > 0$. In particular, $1 > 0$.

(v) If $0 < x < y$ then $0 < \frac{1}{y} < \frac{1}{x}$.

Theorem 8.5 (Existence of real field)**.** There exists an ordered field $\mathbf{R}$ that

(i) contains $\mathbf{Q}$ as a subfield, and

(ii) has the least-upper-bound property (also known as the completeness axiom).

Proof. We prove by contruction, as follows. □

One method to construct $\mathbf{R}$ from $\mathbf{Q}$ is Dedekind cuts.

Definition (Dedekind cut)**.** A **Dedekind cut** $\alpha \subset \mathbf{Q}$ satisfies the following properties:

(i) $\alpha \neq \emptyset$, $\alpha \neq \mathbf{Q}$;

(ii) if $p \in \alpha$, $q \in \mathbf{Q}$ and $q < p$, then $q \in \alpha$;

(iii) if $p \in \alpha$, then $p < r$ for some $r \in \alpha$.

Note that (iii) simply says that α has no largest member; (ii) implies two facts which will be used freely:

- If $p \in \alpha$ and $q \notin \alpha$ then $p < q$.
- If $r \notin \alpha$ and $r < s$ then $s \notin \alpha$.

Example

Let $r \in \mathbf{Q}$ and define

$$\alpha_r := \{p \in \mathbf{Q} \mid p < r\}.$$

We now check that this is indeed a Dedekind cut.

(i) $p = 1 + r \notin \alpha_r$ thus $\alpha_r \neq \mathbf{Q}$. $p = r - 1 \in \alpha_r$ thus $\alpha_r \neq \emptyset$.

(ii) Suppose that $q \in \alpha_r$ and $q' < q$. Then $q' < q < r$ which implies that $q' < r$ thus $q' \in \alpha_r$.

(iii) Suppose that $q \in \alpha_r$. Consider $\dfrac{q+r}{2} \in \mathbf{Q}$ and $q < \dfrac{q+r}{2} < r$. Thus $\dfrac{q+r}{2} \in \alpha_r$.

This example shows that every rational r corresponds to a Dedekind cut α_r.

Example

$\sqrt[3]{2}$ is not rational, but it is real. $\sqrt[3]{2}$ corresponds to the cut

$$\alpha = \{p \in \mathbf{Q} \mid p^3 < 2\}.$$

(i) Trivial.

(ii) If $q < p$, by the monotonicity of the cubic function, this implies that $q^3 < p^3 < 2$ thus $q \in \alpha$.

(iii) If $p \in \alpha$, consider $\left(p + \frac{1}{n}\right)^3 < 2$.

Definition. The set of real numbers, denoted by $\mathbf{R}$, is the set of all Dedekind cuts:

$$\mathbf{R} := \{\alpha \mid \alpha \text{ is a Dedekind cut}\}.$$

Proposition. $\mathbf{R}$ has an order, where $\alpha < \beta$ is defined to mean that $\alpha \subset \beta$.

Proof. Let us check if this is a valid order (check for transitivity and trichotomy).

(i) For $\alpha, \beta, \gamma \in \mathbf{R}$, if $\alpha < \beta$ and $\beta < \gamma$ it is clear that $\alpha < \gamma$. (A proper subset of a proper subset is a proper subset.)

(ii) It is clear that at most one of the three relations

$$\alpha < \beta, \quad \alpha = \beta, \quad \beta < \alpha$$

can hold for any pair α, β.

To show that at least one holds, assume that the first two fail. Then α is not a subset of β. Hence there exists some $p \in \alpha$ with $p \in \beta$.

If $q \in \beta$, it follows that $q < p$ (since $p \notin \beta$), hence $q \in \alpha$, by (ii). Thus $\beta \subset \alpha$. Since $\beta \neq \alpha$, we conclude that $\beta < \alpha$.

Thus $\mathbf{R}$ is an ordered set. □

Proposition. The ordered set $\mathbf{R}$ has the least-upper-bound property.

Proof. Let $A \neq \emptyset$, $A \subset \mathbf{R}$. Assume that $\beta \in \mathbf{R}$ is an upper bound of A.

Define β to be the union of all $\alpha \in A$; in other words, $p \in \gamma$ if and only if $p \in \alpha$ for some $\alpha \in A$. We shall prove that $\gamma \in \mathbf{R}$ by checking the definition of Dedekind cuts:

(i) Since A is not empty, there exists an $\alpha_0 \in A$. This α_0 is not empty. Since $\alpha_0 \subset \gamma$, γ is not empty.

Next, $\gamma \subset \beta$ (since $\alpha \subset \beta$ for every $\alpha \in A$), and therefore $\gamma \neq \mathbf{Q}$.

(ii) Pick $p \in \gamma$. Then $p \in \alpha_1$ for some $\alpha_1 \in A$. If $q < p$, then $q \in \alpha_1$, hence $q \in \gamma$.

(iii) If $r \in \alpha_1$ is so chosen that $r > p$, we see that $r \in \gamma$ (since $\alpha_1 \subset \gamma$).

Next we prove that $\gamma = \sup A$.

(i) It is clear that $\alpha \leqslant \gamma$ for every $\alpha \in A$.

(ii) Suppose $\delta < \gamma$. Then there is an $s \in \gamma$ and that $s \notin \delta$. Since $s \in \gamma$, $s \in \alpha$ for some $\alpha \in A$. Hence $\delta < \alpha$, and δ is not an upper bound of A.

□

Definition (Addition)**.** Given $\alpha, \beta \in \mathbf{R}$, addition is defined as

$$\alpha + \beta := \{r \in \mathbf{Q} \mid r = a + b, a \in \alpha, b \in \beta\}.$$

Proposition (Addition on $\mathbf{R}$ is closed)**.** For all $\alpha, \beta \in \mathbf{R}$, $\alpha + \beta \in \mathbf{R}$.

Proof. We check that $\alpha + \beta$ is a Dedekind cut:

(i) $\alpha \neq \emptyset$ and $\beta \neq \emptyset$ implies there exists $a \in \alpha$ and $b \in \beta$. Hence $r = a + b \in \alpha + \beta$ so $\alpha + \beta \neq \emptyset$.

Since $\alpha \neq \mathbf{Q}$ and $\beta \neq \mathbf{Q}$, there is $c \neq \alpha$ and $d \neq \beta$. $r' = c + d > a + b$ for any $a \in \alpha, b \in \beta$, so $r' \notin \alpha + \beta$. Hence $\alpha + \beta \neq \mathbf{Q}$.

(ii) Suppose that $r \in \alpha + \beta$ and $r' < r$. We want to show that $r' \in \alpha + \beta$.

$r = a + b$ for some $a \in \alpha, b \in \beta$. $r' - a < b$. Since $\beta \in \mathbf{R}$, $r' - a \in \beta$ so $r' - a = b_1$ for some $b_1 \in \beta$. Hence $r' = a + b_1 \in \alpha + \beta$.

(iii) Suppose $r \in \alpha + \beta$, so $r = a + b$ for some $a \in \alpha, b \in \beta$. There exists $a' \in \alpha, b' \in \beta$ with $a < a'$ and $b < b'$. Then $r = a + b < a' + b' \in \alpha + \beta$. We define $r' = a' + b' \in \alpha + \beta$ with $r < r'$.

□

Proposition.

(i) Addition on $\mathbf{R}$ is commutative: $\forall \alpha, \beta \in \mathbf{R}$, $\alpha + \beta = \beta + \alpha$.

(ii) Addition on $\mathbf{R}$ is associative: $\forall \alpha, \beta, \gamma \in \mathbf{R}$, $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$.

(iii) Define $0^* := \{p \in \mathbf{Q} \mid p < 0\}$. Then $\alpha + 0^* = \alpha$.

(iv) Fix $\alpha \in \mathbf{R}$, define $\beta = \{p \in \mathbf{Q} \mid \exists r > 0 \text{ s.t. } -p - r \notin \alpha\}$. Then $\alpha + \beta = 0^*$

Proof.

(i) We need to show that $\alpha + \beta \subset \beta + \alpha$ and $\beta + \alpha \subset \alpha + \beta$.

Let $r \in \alpha + \beta$. Then $r = a + b$ for $a \in \alpha$ and $b \in \beta$. Thus $r = b + a$ since $+$ is commutative on $\mathbf{Q}$. Hence $r \in \beta + \alpha$. Therefore $\alpha + \beta \subset \beta + \alpha$.

Similarly, $\beta + \alpha \subset \alpha + \beta$.

Therefore $\alpha + \beta = \beta + \alpha$.

(ii) Let $r \in \alpha + (\beta + \gamma)$. Then $r = a + (b + c)$ where $a \in \alpha, b \in \beta, c \in \gamma$. Thus $r = (a + b) + c$ by associativity of $+$ on $\mathbf{Q}$. Therefore $r \in (\alpha + \beta) + \gamma$, hence $\alpha + (\beta + \gamma) \subset (\alpha + \beta) + \gamma$.

Similarly, $(\alpha + \beta) + \gamma \subset \alpha + (\beta + \gamma)$.

(iii) Let $r \in \alpha + 0^*$. Then $r = a + p$ for some $a \in \alpha, p \in 0^*$. Thus $r = a + p < a + 0 = a$ by ordering on $\mathbf{Q}$ and identity on $\mathbf{Q}$. Hence $\alpha + 0^* \subset \alpha$.

Let $r \in \alpha$. Then there exists $r' > p$ where $r' \in \alpha$. Thus $r - r' < 0$, so $r - r' \in 0^*$. We see that

$$r = \underbrace{r'}_{\in \alpha} + \underbrace{(r - r')}_{\in 0^*}.$$

Hence $\alpha \subset \alpha + 0^*$.

(iv) We first need to show that β is a Dedekind cut.

(i) If $s \notin \alpha$ and $p = -s - 1$, then $-p - 1 \notin \alpha$, hence $p \in B$, so $\beta \neq \emptyset$. If $q \in \alpha$, then $-q \notin \beta$ so $\beta \neq \mathbf{Q}$.

(ii) Pick $p \in \beta$ and pick $r > 0$ such that $-p - r \notin \alpha$. If $q < p$, then $-q - r > -p - r$, hence $-q - r \notin \alpha$. Thus $q \in \beta$.

(iii) Put $t = p + \frac{r}{2}$. Then $t > p$, and $-t - \frac{r}{2} = -p - r \notin \alpha$, so $t \in \beta$.

Hence $\beta \in \mathbf{R}$.

If $r \in \alpha$ and $s \in \beta$, then $-s \notin \alpha$, hence $r < -s$ so $r + s < 0$. Thus $\alpha + \beta \subset 0^*$.

To prove the opposite inclusion, pick $v \in 0^*$, put $w = -\frac{v}{2}$. Then $w > 0$, and there exists $n \in \mathbf{N}$ such that $nw \in \alpha$ but $(n+1)w \notin \alpha$, by the Archimedean property on $\mathbf{Q}$. Put $p = -(n+2)w$. Then $p \in \beta$, since $-p - w \notin \alpha$, and

$$v = nw + p \in \alpha + \beta.$$

Thus $0* \subset \alpha + \beta$. We conclude that $\alpha + \beta = 0^*$.

This β will of course be denoted by $-\alpha$.

□

We say that a Dedekind cut α is *positive* if $0 \in \alpha$ and negative if $0 \notin \alpha$. If α is neither positive nor negative, then $\alpha = 0^*$.

Multiplication is a little more bothersome than addition in the present context, since products of negative rationals are positive. For this reason we confine ourselves first to $\mathbf{R}^+$, the set of all $\alpha \in \mathbf{R}$ with $\alpha > 0*$.

For all $\alpha, \beta \in \mathbf{R}^+$, we define multiplication as

$$\alpha \cdot \beta := \{p \in \mathbf{Q} \mid p \leqslant ab, a \in \alpha, b \in \beta, a, b > 0\}.$$

We define $1^* := \{q \in \mathbf{Q} \mid q < 1\}$.

Proposition (Multiplication on $\mathbf{R}$ is closed)**.** For all $\alpha, \beta \in \mathbf{R}$, $\alpha \cdot \beta \in \mathbf{R}$.

Proof.

(i) $\alpha \neq \emptyset$ means there exists $a \in \alpha, a > 0$. Similarly, $\beta \neq \emptyset$ means there exists $b \in \beta, b > 0$. Then $a \cdot b \in \mathbf{Q}$ and $ab \leqslant ab$, so $ab \in \alpha \cdot \beta \neq \emptyset$.

$\alpha \neq \mathbf{Q}$ means there exists $a' \notin \alpha, a' > a$ for all $a \in \alpha$. $\beta \neq \mathbf{Q}$ means there exists $b' \in \beta, b' > b$ for all $b \in \beta$. Then $a'b' > ab$ for all $a \in \alpha, b \in \beta$, so $a'b' \notin \alpha \cdot \beta$, thus $\alpha \cdot \beta \neq \mathbf{Q}$.

(ii) $p < \alpha \cdot \beta$ means $p \leqslant a \cdot b$ for some $a \in \alpha, b \in \beta, a, b > 0$.

For $q < p$, $q < p \leqslant a \cdot b$ so $q \in \alpha \cdot \beta$.

(iii) $p \in \alpha \cdot \beta$ means $p \leqslant a \cdot b$ for some $a \in \alpha, b \in \beta, a, b > 0$. Pick $a' \in \alpha$ and $b' \in \beta$ with $a' > a$ and $b' > b$. Form $a'b' > ab \geqslant p$, $a'b' \leqslant a'b'$ means $a'b' \in \alpha \cdot \beta$.

Hence $\alpha \cdot \beta$ is a Dedekind cut. □

We complete the definition of multiplication by setting $\alpha 0^* = 0^* = 0^* \alpha$, and by setting

$$\alpha \cdot \beta = \begin{cases} (-\alpha)(-\beta) & a < 0^*, \beta < 0^*, \\ -[(-\alpha)\beta] & a < 0^*, \beta > 0^*, \\ -[\alpha \cdot (-\beta)] & \alpha > 0^*, \beta < 0^*. \end{cases}$$

Properties of R

We now discuss properties of $\mathbf{R}$.

Theorem 8.6 ($\mathbf{R}$ is archimedian)**.** For any $x \in \mathbf{R}^+$ and $y \in \mathbf{R}$, there exists $n \in \mathbf{N}$ such that $nx > y$.

Proof. Suppose, for a contradiction, that $nx \leqslant y$ for all $n \in \mathbf{N}$. Then y is an upper bound of $A = \{nx \mid n \in \mathbf{N}\}$. Since $\mathbf{R}$ has the least-upper-bound property and $A \subset R$ is bounded above, $M = \sup A \in \mathbf{R}$.

Consider $M - x$. Since $M - x < M = \sup A$, $M - x$ is not an upper bound of A. Then there exists $n_0 \in \mathbf{N}$ such that $M - x \leqslant n_0 x$, or $M \leqslant (n_0 + 1)x$, which is a contradiction. □

Corollary 8.7. Let $\varepsilon > 0$. Then there exists $n \in \mathbf{N}$ such that $0 < \frac{1}{n} < \varepsilon$.

Proof. Take $x = \varepsilon$ and $y = 1$. □

Theorem 8.8 ($\mathbf{Q}$ is dense in $\mathbf{R}$)**.** For any $x, y \in \mathbf{R}$ with $x < y$, there exists $p \in \mathbf{Q}$ such that $x < p < y$.

Proof. Prove by construction.

Since $x < y$, we have $y - x > 0$. By the archimedian property, there exists $n \in \mathbf{N}$ such that

$$n(y - x) > 1.$$

Apply the archimedian property again to obtain $m_1, m_2 \in \mathbf{N}$ such that $m_1 > nx$ and $m_2 > -nx$. Then

$$-m_2 < nx < m_1.$$

Hence there exists $m \in \mathbf{N}$ (with $-m_2 \leqslant m \leqslant m_1$) such that

$$m - 1 \leqslant nx < m.$$

If we combine there inequalities, we obtain

$$nx < m \leqslant 1 + nx < ny.$$

Since $n > 0$, it follows that

$$x < \frac{m}{n} < y.$$

Take $p = \frac{m}{n}$, and we are done. $\square$

Theorem 8.9 (**R** is closed under taking roots)**.** For every $x \in \mathbf{R}^+$ and every $n \in \mathbf{N}$, there exists a unique $x \in \mathbf{R}^+$ so that $y^n = x$.

Proof. That there is at most one such y is clear, since $0 < y_1 < y_2$ implies $y_1^n < y_2^n$. Let

$$E = \{t \in \mathbf{R}^+ \mid t^n < x\}.$$

We first show that E has a supremum:

(i) If $t = \frac{x}{1-x}$ then $0 \leqslant t < 1$. Hence $t^n \leqslant t < x$. Thus $t \in E$, and $E \neq \emptyset$.

(ii) If $t > 1 + x$ then $t^n \geqslant t > x$, so that $t \notin E$. Thus $1 + x$ is an upper bound of E.

Hence E has a supremum; let $y = \sup E$.

To prove that $y^n = x$ we will show that each of the inequalities $y^n < x$ and $y^n > x$ leads to a contradiction. The identity $b^n - a^n = (b-a)\left(n^{n-1} + b^{n-2}a + \cdots + a^{n-1}\right)$ yields the inequality

$$b^n - a^n < (b-a)nb^{n-1}$$

when $0 < a < b$.

Assume $y^n < x$. Choose h so that $0 < h < 1$ and

$$h < \frac{x - y^n}{n(y+1)^{n-1}}.$$

Put $a = y$, $b = y + h$. Then

$$(y+h)^n - y^n < hn(y+h)^{n-1} < hn(y+1)^{n-1} < x - y^n.$$

Thus $(y+h)^n < x$, and $y + h \in E$. Since $y + h > y$, this contradicts the fact that y is an upper bound of E.

Now assume $y^n > x$. Put

$$k = \frac{y^n - x}{ny^{n-1}}.$$

Then $0 < k < y$. If $t \geqslant y - k$, we conclude that

$$y^n - t^n \leqslant y^n - (y-k)^n < kny^{n-1} = y^n - x.$$

Thus $t^n > x$, and $t \notin E$. It follows that $y - k$ is an upper bound of E. But $y - k < y$, which contradicts the fact that y is the *least* upper bound of E.

Hence $y^n = x$, and the proof is complete. $\square$

Notation. This number y is written $\sqrt[n]{x}$ or $x^{\frac{1}{n}}$.

Corollary 8.10. If $a, b \in \mathbf{R}^+$ and $n \in \mathbf{N}$, then

$$(ab)^{\frac{1}{n}} = a^{\frac{1}{n}} b^{\frac{1}{n}}.$$

Proof. Put $\alpha = a^{\frac{1}{n}}$, $\beta = b^{\frac{1}{n}}$. Then

$$ab = \alpha^n \beta^n = (\alpha\beta)^n$$

since multiplication is commutative. The uniqueness assertion of the above result shows that

$$(ab)^{\frac{1}{n}} = \alpha\beta = a^{\frac{1}{n}} b^{\frac{1}{n}}.$$

□

Proposition 8.11. Real numbers can be represented by decimal expansions.

Proof. Let $x \in \mathbf{R}^+$. Let n_0 be the largest integer such that $n_0 \leqslant x$. (Note that the existence of n_0 depends on the archimedian property of $\mathbf{R}$.) Having chosen $n_0, n_1, \dots, n_{k-1}$, let n_k be the largest integer such that

$$n_0 + \frac{n_1}{10} + \dots + \frac{n_k}{10^k} \leqslant x.$$

Let

$$E = \left\{ n_0 + \frac{n_1}{10} + \dots + \frac{n_k}{10^k} \;\middle|\; k = 0, 1, 2, \dots \right\}.$$

Then $x = \sup E$. The decimal expansion of x is

$$n_0.n_1n_2n_3 \cdots .$$

Conversely, for any infinite decimal, E is bounded above, and $n_0.n_1n_2n_3 \cdots$ is the decimal expansion of $\sup E$. □

Extended Real Number System

Definition 8.12 (Extended real number system)**.** We add two symbols $+\infty$ and $-\infty$ to $\mathbf{R}$, and denote the union

$$\overline{\mathbf{R}} = \mathbf{R} \cup \{\pm\infty\},$$

known as the **extended real number system**. We preserve the original order in $\mathbf{R}$, and define

$$-\infty < x < +\infty$$

for every $x \in \mathbf{R}$.

Proposition 8.13. Any non-empty $E \subset \overline{\mathbf{R}}$ has a supremum and infimum.

Proof. If E is bounded above in $\mathbf{R}$, then we are done. If E is not bounded above in $\mathbf{R}$, then $\sup E = +\infty$ in the extended real number system.

Exactly the same remarks apply to lower bounds. □

The extended real number system does not form a field, but it is customary to make the following conventions:

(i) If x is real then
$$x + \infty = +\infty, \quad x - \infty = -\infty, \quad \frac{x}{+\infty} = \frac{x}{-\infty} = 0.$$

(ii) If $x > 0$ then $x \cdot (+\infty) = +\infty$, $x \cdot (-\infty) = -\infty$.

(iii) If $x < 0$ then $x \cdot (+\infty) = -\infty$, $x \cdot (-\infty) = +\infty$.

When it is desired to make the distinction between real numbers on the one hand and the symbols $+\infty$ and $-\infty$ on the other quite explicit, the former are called *finite*.

§8.2 Complex Field

Consider the Cartesian product

$$\mathbf{R}^2 := \mathbf{R} \times \mathbf{R} = \{(a,b) \mid a,b \in \mathbf{R}\}.$$

A *complex number* is an ordered pair $(a,b) \in \mathbf{R}^2$. Let $x = (a,b)$, $y = (c,d)$ be two complex numbers. We write $x = y$ if and only if $a = c$ and $b = d$. Define addition and multiplication on $\mathbf{R}^2$ as

$$\begin{aligned} x + y &= (a+c, b+d), \\ xy &= (ac - bd, ad + bc). \end{aligned}$$

Proposition 8.14. $\mathbf{R}^2$, with addition and multiplication defined as above, is a field, with additive identity $(0,0)$ and multiplicative identity $(1,0)$. We call this structure $\mathbf{C}$, the **complex field**.

Proof. Check the field axioms. □

Proposition 8.15. For any $a,b \in \mathbf{R}$, we have

$$\begin{aligned} (a,0) + (b,0) &= (a+b, 0), \\ (a,0)(b,0) &= (ab, 0). \end{aligned}$$

Proof. Exercise. □

The above result shows that the complex numbers of the form $(a,0)$ have the same arithmetic properties as the corresponding real numbers a. We can therefore identify $(a,0) \in \mathbf{C}$ with $a \in \mathbf{R}$. This identification gives us $\mathbf{R}$ as a subfield of $\mathbf{C}$.

The reader may have noticed that we have defined the complex numbers without any reference to the mysterious square root of -1. We now show that the notation (a, b) is equivalent to the more customary $a + bi$.

Definition 8.16 (Imaginary number)**.** $i = (0, 1)$.

Proposition 8.17. $i^2 = -1$.

Proof.

$$i^2 = (0, 1)(0, 1) = (-1, 0) = -1.$$

□

Proposition 8.18. For $a, b \in \mathbf{R}$, $(a, b) = a + bi$.

Proof.

$$\begin{aligned} a + bi &= (a, 0) + (b, 0)(0, 1) \\ &= (a, 0) + (0, b) \\ &= (a, b). \end{aligned}$$

□

Definition 8.19. For $a, b \in \mathbf{R}$, $z = a + bi$, we call a and b the *real part* and *imaginary part* of z respectively, denoted by $a = \operatorname{Re}(z)$, $b = \operatorname{Im}(z)$; $\overline{z} = a - bi$ is called the *conjugate* of z.

Proposition 8.20. For $z, w \in \mathbf{C}$,

(i) $\overline{z + w} = \overline{z} + \overline{w}$

(ii) $\overline{zw} = \overline{zw}$

(iii) $z + \overline{z} = 2\operatorname{Re}(z)$, $z - \overline{z} = 2i\operatorname{Im}(z)$

(iv) $z\overline{z} \in \mathbf{R}$ and $z\overline{z} > 0$ (except when $z = 0$)

Definition 8.21. For $z \in \mathbf{C}$, its *absolute value* is

$$|z| \coloneqq (z\overline{z})^{\frac{1}{2}}.$$

Proposition 8.22. For $z, w \in \mathbf{C}$,

(i) $|z| > 0$ unless $z = 0$, $|0| = 0$

(ii) $|\overline{z}| = |z|$

(iii) $|zw| = |z||w|$

(iv) $|\operatorname{Re}(z)| \leqslant |z|$

(v) $|z+w| \leqslant |z|+|w|$

Proof.

(i) The square root is non-negative, by definition.

(ii) The conjugate of $\overline{z}$ is z, and the rest follows by the definition of absolute value.

(iii) Let $z = a+bi$, $w = c+di$ with $a,b,c,d \in \mathbf{R}$. Then

$$|zw|^2 = (ac-bd)^2 + (ad-bc)^2 = (a^2+b^2)(c^2+d^2) = |z|^2|w|^2,$$

or $|zw|^2 = (|z||w|)^2$ so the desired result follows.

(iv) Let $z = a+bi$. Note that $a^2 \leqslant a^2+b^2$, hence

$$|\operatorname{Re}(z)| = |a| = \sqrt{a^2} \leqslant \sqrt{a^2+b^2} = |z|.$$

(v) Let $z, w \in \mathbf{C}$. Note that the conjugate of $z\overline{w}$ is $\overline{z}w$, so $z\overline{w}+\overline{z}w = 2\operatorname{Re}(z\overline{w})$. Hence

$$\begin{aligned}
|z+w|^2 &= (z+w)(\overline{z+w}) \\
&= (z+w)(\overline{z}+\overline{w}) \\
&= z\overline{z} + z\overline{w} + \overline{z}w + w\overline{w} \\
&= |z|^2 + 2\operatorname{Re}(z\overline{w}) + |w|^2 \\
&\leqslant |z|^2 + 2|z\overline{w}| + |w|^2 \\
&= |z|^2 + 2|z||w| + |w|^2 \\
&= (|z|+|w|)^2
\end{aligned}$$

and taking square roots yields the desired result.

□

Theorem 8.23 (Schwarz inequality)**.** If $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbf{C}$, then

$$\left|\sum_{i=1}^n a_i b_i\right|^2 \leqslant \sum_{i=1}^n |a_i|^2 \sum_{i=1}^n |b_i|^2. \tag{8.1}$$

Proof. Let $A = \sum |a_i|^2$, $B = \sum |b_i|^2$, $C = \sum a_i \overline{b_i}$. If $B = 0$, then $b_1 = \cdots = b_n = 0$, and the conclusion is trivial. Assume therefore that $B > 0$. Then we have

$$\begin{aligned}
\sum |Ba_i - Cb_i|^2 &= \sum (Ba_i - Cb_i)(\overline{Ba_i - Cb_i}) \\
&= \sum (Ba_i - Cb_i)(B\overline{a_i} - \overline{Cb_i}) \\
&= B^2 \sum |a_i|^2 - B\overline{C} \sum a_i \overline{b_j} - BC \sum \overline{a_i} b_i + |C|^2 \sum |b_i|^2 \\
&= B^2 A - B|C|^2 \\
&= B(AB - |C|^2).
\end{aligned}$$

Since each term in the first sum is non-negative, we see that

$$B(AB - |C|^2) \geqslant 0.$$

Since $B > 0$, it follows that $AB - |C|^2 \geqslant 0$. This is the desired inequality.

(when does equality hold?) $\square$

§8.3 Euclidean Spaces

For $n \in \mathbf{Z}^+$,

$$\mathbf{R}^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbf{R}\}$$

where $\mathbf{x} = (x_1, \dots, x_n)$, x_i's are called the coordinates of $\mathbf{x}$. The elements of $\mathbf{R}^n$ are called *points*, or *vectors*.

Define addition and scalar multiplication on $\mathbf{R}^n$ as follows: for $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, $\alpha \in \mathbf{R}$,

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1 + y_1, \dots, x_n + y_n), \\ \alpha \mathbf{x} &= (\alpha x_1, \dots, \alpha x_n). \end{aligned}$$

Proposition 8.24. $\mathbf{R}^n$, with addition and scalar multiplication defined above, is a vector space over $\mathbf{R}$, where the zero element of $\mathbf{R}^n$ (sometimes called the origin or the null vector) is the point $\mathbf{0}$, all of whose coordinates are 0.

Proof. These two operations satisfy the commutative, associatives, and distributive laws (the proof is trivial, in view of the analagous laws for the real numbers). □

We define the *inner product* of $\mathbf{x}$ and $\mathbf{y}$ by

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^{n} x_i y_i,$$

and the *norm* of $\mathbf{x}$ by

$$\|\mathbf{x}\| := (\mathbf{x} \cdot \mathbf{x})^{\frac{1}{2}} = \left(\sum_{i=1}^{n} {x_i}^2 \right)^{\frac{1}{2}}.$$

The structure now defined (the vector space $\mathbf{R}^n$ with the above inner product and norm) is called the **Euclidean n-space**.

Proposition 8.25. Suppose $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{R}^n$, $\alpha \in \mathbf{R}$. Then

(i) $\|\mathbf{x}\| \geqslant 0$

(ii) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$

(iii) $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$

(iv) $\|\mathbf{x} \cdot \mathbf{y}\| \leqslant \|\mathbf{x}\| \|\mathbf{y}\|$

(v) $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$

(vi) $\|\mathbf{x} - \mathbf{z}\| \leqslant \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathbf{z}\|$

Proof.

(i) Obvious from definition.

(ii)

$$\begin{aligned}\|\mathbf{x}\| = 0 &\iff \left(\sum_{i=1}^{n} {x_i}^2\right)^{\frac{1}{2}} = 0 \\ &\iff \sum_{i=1}^{n} {x_i}^2 = 0 \\ &\iff x_1 = \cdots = x_n = 0 \\ &\iff \mathbf{x} = (0, \dots, 0) = \mathbf{0}\end{aligned}$$

since ${x_i}^2 \geqslant 0$.

(iii) Obvious from definition.

(iv) This is an immediate consequence of the Cauchy–Schwarz inequality.

(v) By (iv) we have

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\| &= (\mathbf{x}+\mathbf{y}) \cdot (\mathbf{x}+\mathbf{y}) \\ &= \mathbf{x}\cdot\mathbf{x} + 2\mathbf{x}\cdot\mathbf{y} + \mathbf{y}\cdot\mathbf{y} \\ &\leqslant \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 .\end{aligned}$$

(vi) This follows directly from (v) by replacing $\mathbf{x}$ by $\mathbf{x}-\mathbf{y}$ and $\mathbf{y}$ by $\mathbf{y}-\mathbf{z}$.

□

Exercises

Problem 8.1 ([Rud76] Ch.1 Q1)**.** If $r \in \mathbf{Q} \setminus \{0\}$ and $x \in \mathbf{R} \setminus \mathbf{Q}$, prove that $r + x \in \mathbf{R} \setminus \mathbf{Q}$ and $rx \in \mathbf{R} \setminus \mathbf{Q}$.

Solution. We prove by contradiction. Suppose $r + x$ is rational, then $r + x = \dfrac{m}{n}, m, n \in \mathbf{Z}$, and m, n have no common factors. Then $m = n(r + x)$. Let $r = \frac{p}{q}, p, q \in \mathbf{Z}$, the former equation implies that $m = n\left(\frac{p}{q} + x\right)$, i.e., $qm = n(p + qx)$, giving

$$x = \frac{mq - np}{nq},$$

which says that x can be written as the quotient of two integers, so x is rational, a contradiction.

The proof for the case rx is similar. □

Problem 8.2 ([Rud76] Ch.1 Q4)**.** Let E be a nonempty subset of an ordered set; suppose α is a lower bound of E and β is an upper bound of E. Prove that $\alpha \leqslant \beta$.

Solution. Let $x \in E$. By definition of lower and upper bounds, $\alpha \leqslant x \leqslant \beta$. □

Problem 8.3 ([Rud76] Ch.1 Q5)**.** Let $A \subset \mathbf{R}$, $A \neq \emptyset$ be bounded below. Let $-A$ be the set of all numbers $-x$, where $x \in A$. Prove that

$$\inf A = -\sup(-A).$$

Solution. Let $\alpha = \inf A$. If $x \in (-A)$ then $-x \in A$, so $\alpha \leqslant -x$, and so $-\alpha \leqslant x$. This implies that $-\alpha$ is an upper bound for $-A$.

If $\beta < -\alpha$ then $-\beta > \alpha$, and there exists $x \in A$ such that $x < -\beta$. Then $-x \in (-A)$, and $-x > \beta$. This shows that $-\alpha = \sup(-A)$, and we are done. □

Problem 8.4 ([Rud76] Ch.1 Q8)**.** Proe that no order can be defined in $\mathbf{C}$ that turns it into an ordered field.

Solution. By Proposition 1.18d, an ordering $<$ that makes $\mathbf{C}$ an ordered field would have to satisfy $-1 = i^2 > 0$, contradicting $1 > 0$. □

Problem 8.5 ([Rud76] Ch.1 Q9, lexicographic order)**.** Suppose $z = a + bi$, $w = c + di$. Define an order on $\mathbf{C}$ as follows:

$$z < w \iff \begin{cases} a < c, \text{ or} \\ a = c, b < d. \end{cases}$$

Prove that this turns $\mathbf{C}$ into an ordered set. Does this ordered set have the least upper bound property?

Problem 8.6 ([Rud76] Ch.1 Q10)**.** Suppose $z = a + bi$, $w = u + iv$, and

$$a = \left(\frac{|w| + u}{2}\right)^{\frac{1}{2}}, \quad b = \left(\frac{|w| - u}{2}\right)^{\frac{1}{2}}.$$

Prove that $z^2 = w$ if $v \geqslant 0$ and that $\overline{z}^2 = w$ if $v \leqslant 0$. Conclude that every complex number (with one exception!) has two complex square roots.

Solution. We have

$$a^2 - b^2 = \frac{|w| + u}{2} - \frac{|w| - u}{2} = u,$$

and

$$2ab = (|w| + u)^{\frac{1}{2}} (|w| - u)^{\frac{1}{2}} = \left(|w|^2 - u^2\right)^{\frac{1}{2}} = \left(v^2\right)^{\frac{1}{2}} = |v|.$$

Hence

$$z^2 = \left(a^2 - b^2\right) + 2abi = u + |v|i = w$$

if $v \geqslant 0$, and

$$\overline{z}^2 = \left(a^2 - b^2\right) - 2abi = u - |v|i = w$$

if $v \leqslant 0$. Hence every non-zero w has two square roots $\pm z$ or $\pm\overline{z}$. Of course, 0 has only one square root, itself. □

Problem 8.7 ([Rud76] Ch.1 Q11)**.** If $z \in \mathbf{C}$, prove that there exists $r \geqslant 0$ and $w \in \mathbf{C}$ with $|w| = 1$ such that $z = rw$. Are w and r always uniquely determined by z?

Problem 8.8 ([Rud76] Ch.1 Q12, triangle inequality)**.** If $z_1, \ldots, z_n \in \mathbf{C}$, prove that

$$|z_1 + \cdots + z_n| \leqslant |z_1| + \cdots + |z_n|.$$

Solution. By the triangle inequality, $|z_1 + z_2| \leqslant |z_1| + |z_2|$. Assume the statement holds for $n - 1$. Then

$$|z_1 + \cdots + z_{n-1} + z_n| \leqslant |z_1 + \cdots + z_{n-1}| + |z_n| \leqslant |z_1| + \cdots + |z_n|,$$

which establishes the claim by induction. □

Problem 8.9 ([Rud76] Ch.1 Q13)**.** If $x, y \in \mathbf{C}$, prove that

$$||x| - |y|| \leqslant |x - y|.$$

Solution. By the triangle inequality,

$$|x| = |(x - y) + y| \leqslant |x - y| + |y|$$

so that

$$|x| - |y| \leqslant |x - y|.$$

Interchanging the roles of x and y in the above, we also have

$$|y| - |x| \leqslant |x - y|$$

so that

$$\big||x| - |y|\big| \leqslant |x - y|.$$

$\square$

9 Basic Topology

This chapter discusses basic notions of point set topology, which focuses on the metric space and its related structures. Then we introduce compactness and prove three major results (Theorem 9.34, Theorem 9.36, Theorem 9.37). We also briefly talk about perfect sets, and connectedness of sets.

§9.1 Metric Space

Definitions and Examples

Definition 9.1 (Metric space)**.** A **metric space** is a set X with an associated *metric* $d : X \times X \to \mathbf{R}$, which satisfies the following properties for all $p, q \in X$:

(i) Positive definitiveness: $d(p,q) \geqslant 0$, where equality holds if and only if $p = q$;

(ii) Symmetry: $d(p,q) = d(q,p)$;

(iii) Triangle inequality: $d(p,q) \leqslant d(p,r) + d(r,q)$ for any $r \in X$.

For the rest of the chapter, X is taken to be a metric space, unless specified otherwise.

Example (Metrics on $\mathbf{R}^n$)

Each of the following functions define metrics on $\mathbf{R}^n$.

$$
\begin{aligned}
d_1(x,y) &= \sum_{i=1}^{n} |x_i - y_i|; \\
d_2(x,y) &= \sqrt{\sum_{i=1}^{n} (x_i - y_i)} \\
d_\infty(x,y) &= \max_{i \in \{1,2,\dots,n\}} |x_i - y_i|.
\end{aligned}
$$

These are called the ℓ^1-, ℓ^2- (or Euclidean) and ℓ^∞-distances respectively.

The proof that each of d_1, d_2, d_∞ is a metric is mostly very routine, with the exception of proving that d_2, the Euclidean distance, satisfies the triangle inequality. To establish this, recall that the Euclidean norm $\|x\|_2$ of a vector $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ is

$$
\|x\|_2 := \left(\sum_{i=1}^{n} {x_i}^2 \right)^{\frac{1}{2}} = \langle x, x \rangle^{\frac{1}{2}},
$$

where the inner product is given by

$$\langle x, y\rangle := \sum_{i=1}^{n} x_i y_i.$$

Then $d_2(x,y) = \|x-y\|_2$, and so the triangle inequality is the statement that

$$\|w-y\|_2 \leqslant \|w-x\|_2 + \|x-y\|_2 \,.$$

This follows immediately by taking $u = w - x$ and $v = x - y$ in the following lemma.

Lemma 9.2. If $u, v \in \mathbf{R}^n$ then $\|u+v\|_2 \leqslant \|u\|_2 + \|v\|_2$.

Proof. Since $\|u\|_2 \geqslant 0$ for all $u \in \mathbf{R}^n$, squaring both sides of the desired inequality gives

$$\|u+v\|_2{}^2 \leqslant \|u\|_2{}^2 + 2\|u\|_2\|v\|_2 + \|v\|_2{}^2.$$

But since

$$\|u+v\|_2{}^2 = \langle u+v, u+v\rangle = \|u\|_2{}^2 + 2\langle u, v\rangle + \|v\|_2^2 \,,$$

this inequality is immediate from the Cauchy–Schwarz inequality, that is to say the inequality

$$|\langle u, v\rangle| \leqslant \|u\|_2\|v\|_2 \,.$$

□

The following are a few interesting examples of metrics.

Example (Discrete metric)

The **discrete metric** on an arbitrary set X is defined as follows:

$$d(x,y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

Example (2-adic metric)

On $\mathbf{Z}$, define $d(x,y)$ to be 2^{-m}, where 2^m is the largest power of two dividing $x - y$. The triangle inequality holds in the following stronger form, known as the ultrametric property:

$$d(x,z) \leqslant \max\{d(x,y), d(y,z)\}.$$

Indeed, this is just a rephrasing of the statement that if 2^m divides both $x - y$ and $y - z$, then 2^m divides $x - z$.

This metric is very unlike the usual distance. For example, $d(999, 1000) = 1$, whilst $d(0, 1000) = \frac{1}{8}$.

The role of 2 can be replaced by any other prime p, and the metric may also be extended in a natural way to the rationals $\mathbf{Q}$.

Example (Path metric)

Let G be a graph, that is to say a finite set of vertices V joined by edges. Suppose that G is connected, that is to say that there is a path joining any pair of distinct vertices. Define a distance d as follows: $d(v, v) = 0$, and $d(v, w)$ is the length of the shortest path from v to w. Then d is a metric on V, as can be easily checked.

Example (Word metric)

Let G be a group, and suppose that it is generated by elements a, b and their inverses. Define a distance on G as follows: $d(v, w)$ is the minimal k such that $v = wg_1 \cdots g_k$, where $g_i \in \{a, b, a^{-1}, b^{-1}\}$ for all i.

Example (Hamming distance)

Let $X = \{0, 1\}^n$ (the boolean cube), the set of all strings of n zeroes and ones. Define $d(x, y)$ to be the number of coordinates in which x and y differ.

Example (Projective space)

Consider the set $P(\mathbf{R}^n)$ of one-dimensional subspaces of $\mathbf{R}^n$, that is to say lines through the origin. One way to define a distance on this set is to take, for lines L_1, L_2, the distance between L_1 and L_2 to be

$$d(L_1, L_2) = \sqrt{1 - \frac{|\langle v, w\rangle|^2}{\|v\|^2 \|w\|^2}},$$

where v and w are any non-zero vectors in L_1 and L_2 respectively.

When $n = 2$, the distance between two lines is $\sin \theta$ where θ is the angle between those lines.

A metric space (X, d) naturally induces a metric on any of its subsets.

Definition 9.3 (Subspace)**.** Suppose (X, d) is a metric space, $Y \subset X$. Then the restriction of d to $Y \times Y$ gives Y a metric so that $(Y, d_{Y\times Y})$ is a metric space. We call Y equipped with this metric a **subspace**.

Example

Subspaces of $\mathbf{R}$ include $[0, 1]$, $\mathbf{Q}$, $\mathbf{Z}$.

Proposition 9.4 (Product space)**.** If (X, d_X) and (Y, d_Y) are metric spaces, set

$$d_{X\times Y}\left((x_1, y_1), (x_2, y_2)\right) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}.$$

for $x_1, x_2 \in X$, $y_1, y_2 \in Y$. Then $d_{X\times Y}$ gives a metric on $X \times Y$; we call $X \times Y$ the *product space.*

Proof. Reflexivity and symmetry are obvious. Less clear is the triangle inequality. We need to prove that

$$\begin{aligned}&\sqrt{d_X(x_1,x_3)^2+d_Y(y_1,y_3)^2}+\sqrt{d_X(x_3,x_2)^2+d_Y(y_3,y_2)^2}\\&\geqslant\sqrt{d_X(x_1,x_2)^2+d_Y(y_1,y_2)^2}\end{aligned} \tag{1}$$

Write $a_1 = d_X(x_2,x_3)$, $a_2 = d_X(x_1,x_3)$, $a_3 = d_X(x_1,x_2)$ and similarly $b_1 = d_Y(y_2,y_3)$, $b_2 = d_Y(y_1,y_3)$ and $b_3 = d_Y(y_1,y_2)$. Thus we want to show

$$\sqrt{{a_2}^2+{b_2}^2}+\sqrt{{a_1}^2+{b_1}^2}\geqslant\sqrt{{a_3}^2+{b_3}^2}. \tag{2}$$

To prove this, note that from the triangle inequality we have $a_1+a_2\geqslant a_3$, $b_1+b_2\geqslant b_3$. Squaring and adding gives

$${a_1}^2+{b_1}^2+{a_2}^2+{b_2}^2+2(a_1a_2+b_1b_2)\geqslant {a_3}^2+{b_3}^2.$$

By Cauchy–Schwarz,

$$a_1a_2+b_1b_2\leqslant\sqrt{{a_1}^2+{b_1}^2}\sqrt{{a_2}^2+{b_2}^2}.$$

Substituting this into the previous line gives precisely the square of (2), and (1) follows. □

Balls and Boundedness

Definition 9.5 (Balls)**.** The **open ball** centred at $x\in X$ with radius $r>0$ is defined to be the set

$$B_r(x):=\{y\in X\mid d(x,y)<r\}.$$

Similarly the **closed ball** centred at x with radius r is

$$\overline{B}_r(x):=\{y\in X\mid d(x,y)\leqslant r\}.$$

The **punctured ball** is the open ball excluding its centre:

$$B_r(x)\setminus\{x\}=\{y\in X\mid 0<d(x,y)<r\}.$$

> **Example**
>
> Considering $\mathbf{R}^3$ with the Euclidean metric, $B_1(0)$ really is what we understand geometrically as a ball (minus its boundary, the unit sphere), whilst $\overline{B}_1(0)$ contains the unit sphere and everything inside it.

Remark. We caution that this intuitive picture of the closed ball being the open ball "together with its boundary" is totally misleading in general. For instance, in the discrete metric on a set X, the open ball $B_1(a)$ contains only the point a, whereas the closed ball $\overline{B}_1(a)$ is the whole of X.

Definition 9.6 (Bounded)**.** $E\subset X$ is said to be **bounded** if E is contained in some open ball; that is, there exists $M\in\mathbf{R}$ and $q\in X$ such that $d(p,q)<M$ for all $p\in E$.

Proposition 9.7. Let $E\subset X$. Then the following are equivalent:

(i) E is bounded;

(ii) E is contained in some closed ball;

(iii) The set $\{d(x,y) \mid x,y \in E\}$ is a bounded subset of $\mathbf{R}$.

Proof.

(1) $\implies$ (2) This is obvious.

(2) $\implies$ (3) This follows immediately from the triangle inequality.

(3) $\implies$ (1) Suppose E satisfies (iii), then there exists $r \in \mathbf{R}$ such that $d(x,y) \leqslant r$ for all $x,y \in E$. If $E = \emptyset$, then E is certainly bounded. Otherwise, let $p \in E$ be an arbitrary point. Then $E \subset B_{r+1}(p)$. □

Open and Closed Sets

Definition 9.8 (Neighbourhood)**.** $N \subset X$ is called a **neighbourhood** of $p \in X$ if $B_\delta(p) \subset N$ for some $\delta > 0$.

Definition 9.9 (Open set)**.** $E \subset X$ is **open** (in X) if it is a neighbourhood of each of its elements; that is, for all $x \in E$, $B_\delta(x) \subset E$ for some $\delta > 0$.

Proposition 9.10. Any open ball is open.

Proof. Let $B_r(x)$ be an open ball. Then for any point $y \in B_r(x)$, there is $d(y,x) < r$. Take $\delta = r - d(y,x)$, which is positive.

Consider the ball $B_\delta(y)$. We shall show it lives in $B_r(x)$. For this, take any point $z \in B_\delta(y)$. By the triangle inequality, we have

$$\begin{aligned} d(z,x) &\leqslant d(z,y) + d(y,x) \\ &< \delta + d(y,x) \\ &= r. \end{aligned}$$

and so $z \in B_r(x)$. Since for all $y \in B_r(x)$ there exists $\delta > 0$ such that $B_\delta(y) \subset B_r(x)$, we have that $B_r(x)$ is open. □

Proposition 9.11. (i) Both $\emptyset$ and X are open.

(ii) For any indexing set I and collection of open sets $\{E_i \mid i \in I\}$, $\bigcup_{i \in I} E_i$ is open.

(iii) For any *finite* indexing set I and collection of open sets $\{E_i \mid i \in I\}$, $\bigcap_{i \in I} E_i$ is open.

Proof.

(i) Obvious by definition.

(ii) If $x \in \bigcup_{i \in I} E_i$ then there is some $i \in I$ with $x \in E_i$. Since E_i is open, there exists $\delta > 0$ such that $B_\delta(x) \subset E_i$ and hence $B_\delta(x) \in \bigcup_{i \in I} E_i$.

(iii) Suppose that I is finite and that $x \in \bigcap_{i \in I} E_i$. For each $i \in I$, we have $x \in E_i$ and so there exists δ_i such that $B_{\delta_i}(x) \subset E_i$. Set $\delta = \min_{i \in I} \delta_i$, then $\delta > 0$ (here it is, of course, crucial that I be finite), and $B_\delta(x) \subset B_{\delta_i}(x) \subset E_i$ for all i. Therefore $B_\delta(x) \subset \bigcap_{i \in I} E_i$.

□

Remark. (1) is in fact a special case of (2) and (3), taking I to be the empty set.

Remark. It is extremely important to note that, whilst the indexing set I in (2) can be arbitrary, the indexing set in (3) must be finite. In general, an arbitrary intersection of open sets is not open; for instance, the intervals $E_i = \left(-\frac{1}{i}, \frac{1}{i}\right)$ are all open in $\mathbf{R}$, but their intersection $\bigcap_{i=1}^{\infty} E_i = \{0\}$, which is not an open set.

Proposition 9.12. Suppose Y is a subspace of X. $E \subset Y$ is open relative to Y if and only if $E = Y \cap G$ for some open subset G of X.

Proof.

$\boxed{\Longrightarrow}$ Suppose E is open relative to Y. Then for each $p \in E$ there exists $r_p > 0$ such that the conditions $d(p,q) < r_p$, $q \in Y$ imply $q \in E$.

For each $p \in E$, let the open ball

$$V_p = \{q \in X \mid d(p,q) < r_p\},$$

and define

$$G = \bigcup_{p \in E} V_p.$$

Since G is an intersection of open balls and open balls are open sets, by Proposition 9.11, G is an open subset of X. Since $p \in V_p$ for all $p \in E$, it is clear that $E \subset G \cap Y$.

To show the opposite containment, by our choice of V_p, we have $V_p \cap Y \subset E$ for every $p \in E$, so that $G \cap Y \subset E$. Hence $E = G \cap Y$.

$\boxed{\Longleftarrow}$ Conversely, if G is open in X and $E = G \cap Y$, every $p \in E$ has a neighbourhood $V_p \cap Y \subset E$. Hence E is open relative to Y. □

The complement of an open set is a closed set.

Definition 9.13 (Closed set)**.** $E \subset X$ is **closed** if its complement $E^c = X \setminus E$ is open.

Proposition 9.14. Any closed ball is closed.

Proof. To prove that $\overline{B}_r(x) = \{y \in X \mid d(x,y) \leqslant r\}$ is closed, we need to show that its complement $\overline{B}_r(x)^c = \{y \in X \mid d(x,y) > r\}$ is open. To do so, we need to show that for all $z \in \overline{B}_r(x)^c$, there exists $\delta > 0$ such that $B_\delta(z) \subset \overline{B}_r(x)^c$.

Take $\delta > 0$ such that $r + \delta < d(x, z)$; that is, $\delta < d(x, z) - r$.

Pick $y \in B_\delta(z)$. Then $d(y, z) < \delta$. But $r + d(y, z) < d(x, z)$ so $r < d(x, z) - d(y, z) \leqslant d(x, y)$ by triangle inequality. Hence we have $r < d(x, y)$, thus $y \in \overline{B}_r(x)^c$ and so $B_\delta(z) \subset \overline{B}_r(x)^c$. Therefore $\overline{B}_r(x)^c$ is open, so $\overline{B}_r(x)$ is closed. □

Proposition 9.15. (i) Both $\emptyset$ and X are closed.

(ii) For any indexing set I and collection of closed sets $\{F_i \mid i \in I\}$, $\bigcap_{i \in I} F_i$ is closed.

(iii) For any *finite* indexing set I and collection of closed sets $\{F_i \mid i \in I\}$, $\bigcup_{i \in I} F_i$ is closed.

Proof. From Proposition 9.11, simply take complements and apply de Morgan's laws. □

Remark. As above, the indexing set in (3) must be finite; for instane, the closed intervals $F_i = \left[-1 + \frac{1}{n}, 1 - \frac{1}{n}\right]$ are all closed in $\mathbf{R}$, but their union $\bigcup_{i=1}^{\infty} F_i = (-1, 1)$ is open, not closed.

Interiors, Closures, Limit Points

Definition 9.16. The **interior** of $E \subset X$, denoted by E°, is defined to be the union of all open subsets of X contained in E.

The **closure** of E, denoted by $\overline{E}$, is defined to be the intersection of all closed subsets of X containing E.

The set $\overline{E} \setminus E^\circ$ is known as the **boundary** of E, denoted by ∂E. p is a **boundary point** of E if $p \in \partial E$.

A set $E \subset X$ is said to be **dense** if $\overline{E} = X$.

Since an arbitrary union of open sets is open, E° is itself an open set, and it is clearly the unique largest open subset of X contained in E. If E is itself open then evidently $E = E^\circ$.

Since an arbitrary intersection of closed sets is closed, $\overline{E}$ is the unique smallest closed subset of X containing E. If E is itself closed then evidently $E = \overline{E}$.

If $x \in E^\circ$ we say that x is an **interior point** of E. One can also phrase this in terms of neighbourhoods: the interior of E is the set of all points in E for which E is a neighbourhood.

> **Example**
>
> Consider the closed interval $[a, b]$ in $\mathbf{R}$; its interior is just the open interval (a, b).
>
> The rationals $\mathbf{Q}$ are a dense subset of $\mathbf{R}$.

Let us give a couple of simple characterisations of the closure of a set.

Proposition 9.17. Suppose $E \subset X$. $p \in \overline{E}$ if and only if every open ball $B_\delta(p)$ contains a point of E.

Proof.

$\boxed{\Longrightarrow}$ Suppose that $p \in \overline{E}$. Suppose, for a contradiction, that there exists some open ball $B_\delta(p)$ that does not meet E, then $B_\delta(p)^c$ is a closed set containing E. Therefore $B_\delta(p)^c$ contains $\overline{E}$, and hence it contains p, which is obviously nonsense.

$\boxed{\Longleftarrow}$ Suppose that every ball $B_\delta(p)$ meets E. Suppose, for a contradiction, that $p \notin \overline{E}$. Then since $\overline{E}^c$ is open, there is a ball $B_\delta(p)$ contained in $\overline{E}^c$, and hence in E^c, contrary to assumption. □

Remark. A particular consequence of this is that $E \subset X$ is dense if and only if it meets every open set in X.

We now introduce the notion of limit points.

Definition 9.18 (Limit point)**.** $p \in E$ is a **limit point** (or *accumulation point*) of E if every neighbourhood of p contains some $q \neq p$ such that $q \in E$.

The **induced set** of E, denoted by E', is the set of all limit points of E in X.

> **Example**
>
> Consider the metric space $\mathbf{R}$, a and b are limit points $(a, b]$. The limit point set of $(a, b]$ is $[a, b]$, which is also the closure $(a, b]$.
>
> Consider the metric space $\mathbf{R}^2$. The limit point set of any open ball $B_r(x)$ is the closed ball $\bar{B}_r(x)$, which is also the closure of $B_r(x)$.
>
> Consider $\mathbf{Q} \subset \mathbf{R}$. $\mathbf{Q}' = \bar{\mathbf{Q}} = \mathbf{R}$.

Note that we do not necessarily have $E \subset E'$, that is to say it is quite possible for a point $p \in E$ not to be a limit point of E. This occurs if there exists some ball $B_\delta(p)$ such that $B_\delta(p) \cap E = \{p\}$; in this case we say that p is an **isolated point** of E.

Proposition 9.19. If p is a limit point of E, then every ball of p contains infinitely many points of E.

Proof. Prove by contradiction. Suppose there exists $B_r(p)$ which contains only a finite number of points of E: $q_1, \dots, q_n$, where $q_i \neq p$ for $i = 1, \dots, n$. Define

$$r = \min\{d(p, q_1), \dots, d(p, q_n)\}.$$

The minimum of a finite set of positive numbers is clearly positive, so that $r > 0$.

$B_r(p)$ contains no point $q \in E, q \neq p$ so that p is not a limit point of E, a contradiction. □

Corollary 9.20. A finite point set has no limit points.

Proposition 9.21. Suppose $E \subset X$. E' is a closed subset of X.

Proof. To prove that E' is closed, we need to show that the complement $(E')^c$ is open.

Suppose $p \in (E')^c$. Then exists a ball $B_\varepsilon(p)$ whose intersection with E is either empty or $\{p\}$. We claim that $B_{\frac{\varepsilon}{2}}(p) \subset (E')^c$. Let $q \in B_{\frac{\varepsilon}{2}}(p)$. If $q = p$, then clearly $q \in (E')^c$. If $q \neq p$, there is some ball about q which is contained in $B_\varepsilon(p)$, but does not contain p: the ball $B_\delta(q)$ where $\delta = \min\left(\frac{\varepsilon}{2}, d(p,q)\right)$ has this property. This ball meets E in the empty set, and so $q \in (E')^c$ in this case too. $\square$

Proposition 9.22. Suppose $E \subset X$. Then $\overline{E} = E \cup E'$.

Proof. We first show the containment $E \cup E' \subset \overline{E}$. Obviously $E \subset \overline{E}$, so we need only show that $E' \subset \overline{E}$. Suppose $p \in \overline{E}^c$. Since $\overline{E}^c$ is open, there is some ball $B_\varepsilon(p)$ which lies in $\overline{E}^c$, and hence also in E^c, and therefore a cannot be a limit point of E. This concludes the proof of this direction.

Now we look at the opposite containment $\overline{E} \subset E \cup E'$. If $p \in \overline{E}$, we saw in Lemma 5.1.5 that there is a sequence (x_n) of elements of E with $x_n \to p$. If $x_n = p$ for some n then we are done, since this implies that $p \in E$. Suppose, then, that $x_n \neq p$ for all n. Let $\varepsilon > 0$ be given, for sufficiently large n, all the x_n are elements of $B_\varepsilon(p) \setminus \{p\}$, and they all lie in E. It follows that p is a limit point of E, and so we are done in this case also. $\square$

Proposition 9.23. Suppose $E \subsetneq \mathbf{R}$, $E \neq \emptyset$ be bounded above. Let $y = \sup E$. Then $y \in \overline{E}$. Hence $y \in E$ if E is closed.

Proof. If $y \in E$ then $y \in \overline{E}$. Assume $y \notin E$. For every $h > 0$ there exists then a point $x \in E$ such that $y - h < x < y$, for otherwise $y - h$ would be an upper bound of E. Thus y is a limit point of E. Hence $y \in \overline{E}$. $\square$

§9.2 Compactness

The following is a useful analogy to visualise the concept of compactness:

> Compactness is like a well-contained space where nothing "escapes" or goes off to infinity.
>
> An open cover is a collection of open sets that completely cover the compact set (think of a bunch of overlapping circles covering a shape).
>
> The key feature of compact sets is that from any open cover, you can always select a finite number of sets from the cover that still manage to cover the entire space.

Definition 9.24. Let $\mathcal{U} = \{U_i \mid i \in I\}$ be a collection of open subsets of X. We say that $\mathcal{U}$ is an **open cover** of a set K if

$$K \subset \bigcup_{i \in I} U_i.$$

If $I' \subset I$ and $K \subset \bigcup_{i \in I'} U_i$, we say that $\{U_i \mid i \in I'\}$ is a **subcover** of $\mathcal{U}$. If moreover, I' is finite, then it is called a **finite subcover**.

Definition 9.25 (Compactness)**.** $K \subset X$ is said to be **compact** if every open cover of K contains a finite subcover.

That is, if $\mathcal{U} = \{U_i \mid i \in I\}$ is an open cover of K, then there are finitely many indices $i_1, \dots, i_n$ such that

$$K \subset \bigcup_{k=1}^{n} U_{i_k}.$$

Exercise

Every finite set is compact.

Solution. Let E be finite. Let $\mathcal{U} = \{U_i \mid i \in I\}$ be an open cover of E, then we have that $E \subset \bigcup_{i \in I}$.

For each point $x \in E$, take i_x such that $x \in U_{i_x}$. Let $\mathcal{V} = \{U_{i_x} \mid x \in E\}$. By construction, since $x \in \mathcal{V}$ for all $x \in E$, $E \subset \mathcal{V}$ so $\mathcal{V}$ is an open cover of E. Since there are finitely many x, $\mathcal{V}$ is thus a finite subcover of E, and hence E is compact. □

Proposition 9.26. Suppose Y is a subspace of X, and $K \subset Y$. Then K is compact relative to X if and only if K is compact relative to Y.

Proof.

$\boxed{\Longrightarrow}$ Suppose K is compact relative to X. Let $\mathcal{U}$ be an open cover of K; that is, $\mathcal{U} = \{U_i \mid i \in I\}$ is a collection of sets open relative to Y, such that $K \subset \bigcup_{i \in I} U_i$. By Proposition 9.12, for all $i \in I$, there exist V_i open relative to X such that $U_i = Y \cap V_i$. Since K is compact relative to X, we have

$$K \subset \bigcup_{k=1}^{n} V_{i_k} \tag{1}$$

for some choice of finitely many indices $i_1, \dots, i_n$. Since $K \subset Y$, (1) implies that

$$K \subset \bigcup_{k=1}^{n} U_{i_k}. \tag{2}$$

This proves that K is compact relative to Y.

$\boxed{\Longleftarrow}$ Suppose K is compact relative to Y, let $\mathcal{V} = \{V_i \mid i \in I\}$ be a collection of open subsets of X which covers K, and put $U_i = Y \cap V_i$. Then (2) will hold for some choice of $i_1, \dots, i_n$; and since $U_i \subset V_i$, (2) implies (1). ☐

Proposition 9.27. Compact subsets of metric spaces are closed.

Proof. Let $K \subset X$ be compact. To prove that K is closed, we need to show that K^c is open.

Suppose $p \in X$, $p \neq K$. If $q \in K$, let V_q and W_q be neighbourhoods of p and q respectively, of radius less than $\frac{1}{2}d(p,q)$. Since K is compact, there exists finite many points $q_1, \dots, q_n \in K$ such that

$$K \subset \bigcup_{k=1}^{n} W_{q_k} = W.$$

If $V = \bigcap_{k=1}^{n} V_{q_k}$, then V is a neighbourhood of p which does not intersect W. Hence $V \subset K^c$, so p is an interior point of K^c. The theorem follows. ☐

Proposition 9.28. Closed subsets of compact sets are compact.

Proof. Suppose $F \subset K \subset X$, F is closed (relative to X), and K is compact.

Let $\mathcal{V} = \{V_i \mid i \in I\}$ be an open cover of F. If F^c is adjoined to $\mathcal{V}$, we obtain an open cover Ω of K. Since K is compact, there is a finite subcollection Φ of Ω which covers K, and hence F. If F^c is a member of Φ, we may remove it from Φ and still retain an open cover of F. We have thus shown that a finite subcollection of $\mathcal{V}$ covers F. ☐

Corollary 9.29. If F is closed and K is compact, then $F \cap K$ is compact.

Proposition 9.30. If E is an infinite subset of a compact set K, then E has a limit point in K.

Proposition 9.31. If (I_n) is a sequence of intervals in $\mathbf{R}$ such that $I_i \supset I_{i+1}$, then $\bigcap_{i=1}^{\infty} I_n \neq \emptyset$.

Proposition 9.32. If (I_n) is a sequence of k-cells such that $I_n \supset I_{n+1}$, then $\bigcap_{n=1}^{\infty} \neq \emptyset$.

Proof. Let I_n consist of all points $\mathbf{x} = (x_1, \dots, x_k)$ such that ☐

Proposition 9.33. Every k-cell is compact.

Theorem 9.34 (Cantor's Intersection Theorem)**.** Given a decreasing sequence of compact sets $A_1 \supset A_2 \supset \cdots$, there exists a point $x \in \mathbf{R}^n$ such that x belongs to all A_i. In other words, $\bigcap_{i=1}^{\infty} A_i \neq \emptyset$. Moreover, if for all $i \in \mathbf{N}$ we have $\operatorname{diam} A_{i+1} \leqslant c \cdot \operatorname{diam} A_k$ for some constant $c < 1$, then such a point must be unique, i.e. $\bigcap_{i=1}^{\infty} A_k = \{x\}$ for some $x \in \mathbf{R}^n$.

Proposition 9.35. If $E \subset \mathbf{R}^n$ has one of the following three properties, then it has the other two:

(i) E is closed and bounded.

(ii) E is compact.

(iii) Every infinite subset of E has a limit point in E.

Theorem 9.36 (Heine–Borel Theorem)**.** $E \subset \mathbf{R}^n$ is compact if and only if E is closed and bounded.

Proof. □

Theorem 9.37 (Bolzano–Weierstrass Theorem)**.** Every bounded infinite subset of $\mathbf{R}^n$ has a limit point in $\mathbf{R}^n$.

Proof. □

§9.3 Perfect Sets

Definition 9.38 (Perfect set)**.** E is **perfect** if E is closed and if every point of E is a limit point of E.

Proposition 9.39. Let P be a non-empty perfect set in $\mathbf{R}^n$. Then P is uncountable.

Corollary 9.40. Every interval $[a, b]$ is uncountable. In particular, $\mathbf{R}$ is uncountable.

The set which we are now going to construct shows that there exist perfect sets in $\mathbf{R}$ which contain no segment.

Let

$$E_0 = [0, 1].$$

Remove the segment $\left(\frac{1}{3}, \frac{2}{3}\right)$ to give

$$E_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

Remove the middle thirds of these intervals to give

$$E_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{3}{9}\right] \cup \left[\frac{6}{9}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

Repeating this process, we obtain a monotonically decreasing sequence of compact sets (E_n), where $E_1 \supset E_2 \supset \cdots$ and E_n is the union of 2^n intervals, each of length 3^{-n}.

The **Cantor set** is defined as

$$P \coloneqq \bigcap_{n=1}^{\infty} E_n.$$

Proposition 9.41. P is compact.

Proposition 9.42. P is not empty.

Proof. This follows from Theorem 2.36. □

Proposition 9.43. P contains no segment.

Proof. No segment of the form

$$\left(\frac{3k+1}{3^m}, \frac{3k+2}{3^m}\right),$$

where $k, m \in \mathbf{Z}^+$, has a point in common with P. Since every segment (α, β) contains a segment of the above form, if

$$3^{-m} < \frac{\beta - \alpha}{6},$$

P contains no segment. □

Proposition 9.44. P is perfect.

Proof. To show that P is perfect, it is enough to show that P contains no isolated point. Let $x \in P$, and let S be any segment containing x. Let I_n be that interval of E_n which contains x. Choose n large enough, so that $I_n \subset S$. Let x_n be an endpoint of I_n, such that $x_n \neq x$.

It follows from the construction of P that $x_n \in P$. Hence x is a limit point of P, and P is perfect. $\square$

§9.4 Connectedness

Definition 9.45 (Connectedness)**.** A and B are said to be **separated** if $A \cap \overline{B} = \emptyset$ and $\bar{A} \cap B = \emptyset$; that is, no point of A lies in the closure of B and no point of B lies in the closure of A.

$E \subset X$ is said to be **connected** if E is not a union of two non-empty separated sets.

Remark. Separated sets are of course disjoint, but disjoint sets need not be separated. For example, the interval $[0,1]$ and the segment $(1,2)$ are not separated, since 1 is a limit point of $(1,2)$. However, the segments $(0,1)$ and $(1,2)$ are separated.

The connected subsets of the line have a particularly simple structure:

Proposition 9.46. *$E \subset \mathbf{R}^1$ is connected if and only if it has the following property: if $x, y \in E$ and $x < z < y$, then $z \in E$.*

Proof.

$\boxed{\Longleftarrow}$ If there exists $x, y \in E$ and some $z \in (x, y)$ such that $z \notin E$, then $E = A_z \cup B_z$ where

$$A_z = E \cap (-\infty, z), \quad B_z = E \cap (z, \infty).$$

Since $x \in A_z$ and $y \in B_z$, A and B are non-empty. Since $A_z \subset (-\infty, z)$ and $B_z \subset (z, \infty)$, they are separated. Hence E is not connected.

$\boxed{\Longrightarrow}$ Suppose E is not connecetd. Then there are non-empty separated sets A and B such that $A \cup B = E$. Pick $x \in A$, $y \in B$, and WLOG assume that $x < y$. Define

$$z := \sup(A \cap [x, y].)$$

By □

Definition 9.47. We say that a metric space is **disconnected** if we can write it as the disjoint union of two nonempty open sets. We say that a space is **connected** if it is not disconnected.

If X is written as a disjoint union of two nonempty open sets U and V then we say that these sets **disconnect** X.

> **Example**
>
> If $X = [0,1] \cup [2,3] \subset \mathbf{R}$ then we have seen that both $[0,1]$ and $[2,3]$ are open in X. Since X is their disjoint union, X is disconnected.

The following lemma gives some equivalent ways to formulate the concept of connected space.

Lemma 9.48. The following are equivalent:

(i) X is connected.

(ii) If $f : X \to \{0, 1\}$ is a continuous function then f is constant.

(iii) The only subsets of X which are both open and closed are X and $\emptyset$.

(Here the set $\{0, 1\}$ is viewed as a metric space via its embedding in $\mathbf{R}$, or equivalently with the discrete metric.)

Proof. □

Frequently one has a metric space X and a subset E of it whose connectedness or otherwise one wishes to ascertain. To this end, it is useful to record the following lemma.

Lemma 9.49. Let $E \subset X$, considered as a metric space with the metric induced from X. Then E is connected if and only if the following is true: if U, V are open subsets of X, and $U \cap V \cap E = \emptyset$, then $E \subset U \cup V$ implies either $E \subset U$ or $E \subset V$.

Proof. □

We now turn to some basic properties of the notion of connectedness. These broadly conform with one's intuition about how connected sets should behave.

Lemma 9.50 (Sunflower lemma)**.** Let $\{E_i \mid i \in I\}$ be a collection of connected subsets of X such that $\bigcap_{i \in I} E_i \neq \emptyset$. Then $\bigcup_{i \in I} E_i$ is connected.

Proof. □

10 Numerical Sequences and Series

Tthis chapter will deal primarily with sequences and series in $\mathbf{R}$ (and also $\mathbf{C}$). The basic facts about convergence, however, are just as easily explained in a more general setting. The first three sections will therefore be concerned with sequences in euclidean spaces, or even in metric spaces.

As usual, let (X,d) be a metric space.

§10.1 Sequences

Convergent Sequences

Definition 10.1 (Convergence)**.** A sequence (x_n) **converges** to $x \in X$, denoted by $x_n \to x$ or $\lim_{n\to\infty} x_n = x$, if

$$\forall \varepsilon > 0, \quad \exists N \in \mathbf{N}, \quad \forall n \geqslant N, \quad d(x_n, x) < \varepsilon.$$

We call x the *limit* of (x_n).

If (x_n) does not converge, it is said to *diverge*.

Exercise

Define what it means for $x_n \not\to x$.

Solution. Basically negate the definition for convergence:

$$\exists \varepsilon > 0, \quad \forall N \in \mathbf{N}, \quad \exists n \geqslant N, \quad d(x_n, x) \geqslant \varepsilon.$$

□

Exercise

Show that $\dfrac{1}{n} \to 0$ as $n \to \infty$.

Solution. Fix $\varepsilon > 0$. Then by the Archimedian property, there exists $N \in \mathbf{N}$ such that $\frac{1}{N} < \varepsilon$. Then for all $n \geqslant N$,

$$0 < \frac{1}{n} \leqslant \frac{1}{N} < \varepsilon$$

which implies $\left|\frac{1}{n} - 0\right| < \varepsilon$. Hence $\frac{1}{n} \to 0$ as $n \to \infty$. □

Exercise

Let $a_n = 1 + (-1)^n \frac{1}{\sqrt{n}}$. Show that $a_n \to 1$ as $n \to \infty$.

We want to find N such that if $n \geqslant N$ then

$$\begin{aligned}
|a_n - 1| &< \varepsilon \\
\left| \left(1 + (-1)^n \frac{1}{\sqrt{n}}\right) - 1 \right| &< \varepsilon \\
\frac{1}{\sqrt{n}} &< \varepsilon \\
\frac{1}{\varepsilon} &< \sqrt{n}
\end{aligned}$$

so we take $N = \left\lceil \frac{1}{\varepsilon^2} \right\rceil + 1$.

Solution. Let $\varepsilon > 0$ be given. Take $N = \left\lceil \frac{1}{\varepsilon} \right\rceil + 1$. If $n \geqslant N$, then

$$\begin{aligned}
n &> \frac{1}{\varepsilon^2} \\
\sqrt{n} &> \frac{1}{\varepsilon} \\
\frac{1}{\sqrt{n}} &< \varepsilon \\
|a_n - 1| &< \varepsilon.
\end{aligned}$$

Hence $a_n \to 1$ as $n \to \infty$. □

Exercise

Let $a_n = \dfrac{n\cos\left(n^3+1\right)}{5n^2+1}$ for $n \geqslant 1$. Then $a_n \to 0$ as $n \to \infty$.

Definition 10.2 (Bounded sequence)**.** The set of all points x_n is the **range** of (x_n):

$$\{x_n \mid n \in \mathbf{N}\};$$

the range of a sequence may be a finite set or it may be infinite.

(x_n) is said to be **bounded** if its range is bounded.

We now outline some important properties of convergent sequences in metric spaces.

Proposition 10.3. Let (x_n) be a sequence in metric space X.

(i) $x_n \to x$ if and only every neighbourhood of x contains x_n for all but finitely many n.

(ii) Uniqueness of limit: if $x_n \to x$ and $x_n \to x'$ for $x, x' \in X$, then $x' = x$.

(iii) Boundedness of convergent sequences: if (x_n) converges, then (x_n) is bounded.

(iv) Suppose $E \subset X$. Then x is a limit point of E if and only if there exists a sequence (x_n) in $E \setminus \{x\}$ such that $x_n \to x$.

Proof.

(i) $\boxed{\Longrightarrow}$ Suppose $x_n \to x$. We want to prove that any neighbourhood U of x eventually contains all x_n.

Since U is a neighbourhood of x, pick a ball $B_\varepsilon(x) \subset U$. Corresponding to this ε, there exists $N \in \mathbf{N}$ such that $n \geqslant N$ implies $d(x_n, x) < \varepsilon$. Thus $n \geqslant N$ implies $x_n \in U$.

$\boxed{\Longleftarrow}$ Suppose every neighbourhood of x contains all but finitely many of the x_n. Fix $\varepsilon > 0$, pick a ball $B_\varepsilon(x)$. Since $B_\varepsilon(x)$ is a neighbourhood of x, it will also eventually contain all x_n. By assumption, there eists $N \in \mathbf{N}$ such that $x_n \in B_\varepsilon(x)$ if $n \geqslant N$. Thus $d(x_n, x) < \varepsilon$ if $n \geqslant N$, hence $x_n \to x$.

(ii) Let $\varepsilon > 0$ be given. There exists $N, N' \in \mathbf{N}$ such that

$$n \geqslant N \implies d(x_n, x) < \frac{\varepsilon}{2}$$

and

$$n \geqslant N' \implies d(x_n, x') < \frac{\varepsilon}{2}.$$

Take $N_1 := \max\{N, N'\}$. Hence if $n \geqslant N_1$ we have $d(x_n, x) < \frac{\varepsilon}{2}$ and $d(x_n, x') < \frac{\varepsilon}{2}$ at the same time. By triangle inequality,

$$d(x, x') \leqslant d(x, x_n) + d(x_n, x') < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since ε was arbitrary (i.e. holds for all $\varepsilon > 0$), we must have $d(x, x') = 0$ and thus $x = x'$.

(iii) Suppose $x_n \to x$. Then there exists $N \in \mathbf{N}$ such that $n \geqslant N$ implies $d(x_n, x) < 1$. Take

$$r := \max\{1, d(x_1, x), \dots, d(x_N, x)\}.$$

Then $d(x_n, x) \leqslant r$ for $n = 1, 2, \dots, N$, so (x_n) is in $B_r(x)$.

(iv) $\boxed{\Longrightarrow}$ If x is a limit point, then for all $\varepsilon > 0$, $B_\varepsilon(x) \setminus \{x\}$ contains points in E. We then construct such a sequence (x_n) in $E \setminus \{x\}$: pick any $x_n \in E$ so that x_n is contained in $B_{\frac{1}{n}}(x) \setminus \{x\}$. Then it is easy to show that (x_n) is a sequence in $E \setminus \{x\}$ which converges to x.

$\boxed{\Longleftarrow}$ Suppose that there exists a sequence (x_n) in $E \setminus \{x\}$ such that $x_n \to x$. We wish to show that $B_\varepsilon(x) \setminus \{x\}$ contains points in E for all $\varepsilon > 0$.

Since (x_n) converges to x, for all $\varepsilon > 0$ the sequence is eventually contained in $B_\varepsilon(x)$. However because we have the precondition that (x_n) has to be in $E \setminus \{x\}$, the sequence is in fact eventually contained in $B_\varepsilon(x) \setminus \{x\}$.

□

Lemma 10.4. If (a_n) and (b_n) are two convergent sequences, and $a_n \leqslant b_n$, then $\lim_{n\to\infty} a_n \leqslant \lim_{n\to\infty} b_n$.

Remark. Even if you have $a_n < b_n$, you cannot say that $\lim_{n\to\infty} a_n < \lim_{n\to\infty} b_n$. For example, $-\frac{1}{n} < \frac{1}{n}$ but their limits are both 0.

Proof. Let $A = \lim_{n\to\infty} a_n$, $B = \lim_{n\to\infty} b_n$. Suppose otherwise that $A > B$, take $\varepsilon = A - B > 0$.

Since $\frac{\varepsilon}{2} > 0$, then there exists N_1 such that for $n \geqslant N_1$ we have $|a_n - A| < \frac{\varepsilon}{2}$; and there exists N_2 such that for $n \geqslant N_2$ we have $|b_n - B| < \frac{\varepsilon}{2}$.

Let $N = \max\{N_1, N_2\}$, then for any $n \geqslant N$, the two inequalities above will hold simultaneously. But then we would have

$$a_n > A - \frac{\varepsilon}{2}, \quad b_n < B + \frac{\varepsilon}{2}$$

and thus

$$a_n - b_n > A - B - \varepsilon = 0$$

so $a_n > b_n$, a contradiction. □

Proposition 10.5 (Arithmetic properties)**.** Suppose (a_n) and (b_n) are convergent seqeunces of real numbers, $k \in \mathbf{R}$. Then

(i) Scalar multiplication: $\lim_{n\to\infty} ka_n = k \lim_{n\to\infty} a_n$

(ii) Addition: $\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$

(iii) Multiplication: $\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$

(iv) Division: $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$ $(b_n \neq 0, \lim_{n\to\infty} b_n \neq 0)$

Proof. Let $a = \lim_{n\to\infty} a_n$, $b = \lim_{n\to\infty} b_n$.

(i) The proof is left as an exercise. You will need to consider three cases, when k is positive, negative or 0.

(ii) Let $\varepsilon > 0$ be given. Since $a = \lim_{n\to\infty} a_n$, there exists $N_1 \in \mathbf{N}$ such that for all $n \geqslant N_1$,

$$|a_n - a| < \frac{\varepsilon}{2}. \tag{1}$$

Similarly, since $b = \lim_{n\to\infty} b_n$, there exists $N_2 \in \mathbf{N}$ such that for all $n \geqslant N_2$,

$$|b_n - b| < \frac{\varepsilon}{2}. \tag{2}$$

Let $N = \max\{N_1, N_2\}$, then for all $n \geqslant N$, (1) and (2) hold simultaneously. by the triangle inequality, we have

$$\begin{aligned}\left|(a_n+b_n)-(a+b)\right| &\leqslant |a_n-a|+|b_n-b| \\ &< \frac{\varepsilon}{2}+\frac{\varepsilon}{2} \\ &= \varepsilon.\end{aligned}$$

This means that $\lim_{n\to\infty}(a_n+b_n)=a+b$, as desired.

(iii) Consider the limit $\lim_{n\to\infty}(a_nb_n-ab)$. We want to prove that this equals to 0. We write

$$\lim_{n\to\infty}(a_nb_n-ab)=\lim_{n\to\infty}(a_nb_n-ab_n+ab_n-ab);$$

the idea is to show that this is equal to

$$\lim_{n\to\infty}(a_nb_n-ab_n)+\lim_{n\to\infty}(ab_n-ab).$$

Note: we cannot write this yet because we have not shown that these two sequences are convergent.

For the second sequence, for a constant we can write $\lim_{n\to\infty} b=b$. By scalar multiplication, $\lim_{n\to\infty}(-b)=-\lim_{n\to\infty} b=-b$. By addition, we have

$$\begin{aligned}\lim_{n\to\infty}(b_n-b) &= \lim_{n\to\infty}[b_n+(-b)] \\ &= \lim_{n\to\infty} b_n+\lim_{n\to\infty}(-b) \\ &= b+(-b) \\ &= 0.\end{aligned}$$

Since a is a scalar, we have that

$$\lim_{n\to\infty}(ab_n-ab)=a\lim_{n\to\infty}(b_n-b)=0.$$

For the first sequence, we want to show that $\lim_{n\to\infty}(a_n-a)b_n=0$. Since b_n is convergent, b_n is bounded. Let $M>0$ be a bound of b_n, then for all $n\in\mathbf{N}$,

$$|b_n|\leqslant M.$$

Fix $\varepsilon>0$. Since $\lim_{n\to\infty} a_n=a$, there exists $N\in\mathbf{N}$ such that for all $n\geqslant N$,

$$|a_n-a|<\frac{\varepsilon}{M}.$$

Combining the two equations,

$$\begin{aligned}|a_nb_n - ab_n| &= |(a_n - a)b_n| \\ &= |a_n - a|\,|b_n| \\ &< \frac{\varepsilon}{M} \cdot M \\ &= \varepsilon.\end{aligned}$$

Thus $\lim_{n\to\infty}(a_nb_n - ab_n) = 0$.

Since $\lim_{n\to\infty}(ab_n - ab) = 0$ and $\lim_{n\to\infty}(a_nb_n - ab_n) = 0$, by addition, we have that

$$\begin{aligned}\lim_{n\to\infty}(a_nb_n - ab) &= \lim_{n\to\infty}(a_nb_n - ab_n + ab_n - ab) \\ &= \lim_{n\to\infty}(a_nb_n - ab_n) + \lim_{n\to\infty}(ab_n - ab) \\ &= 0 + 0 \\ &= 0,\end{aligned}$$

and thus $\lim_{n\to\infty} a_nb_n = ab$, as desired.

(iv) Since we have proven multiplication, it suffices to show that $\lim_{n\to\infty}\dfrac{1}{b_n} = \dfrac{1}{\lim_{n\to\infty} b_n}$. Consider the limit

$$\lim_{n\to\infty}\left(\frac{1}{b_n} - \frac{1}{b}\right) = \lim_{n\to\infty}\left(\frac{b - b_n}{b_nb}\right).$$

Let $\varepsilon > 0$ be given. Since $b = \lim_{n\to\infty} b_n$, there exists $N_1 \in \mathbf{N}$ such that for all $n \geqslant N_1$,

$$|b_n - b| < \frac{|b|}{2}.$$

Then

$$|b_nb - b^2| < \frac{b^2}{2},$$

or

$$\frac{b^2}{2} < b_nb < \frac{3b^2}{2}.$$

This shows that if $n \geqslant N_1$, b_nb would always be positive, and $\frac{1}{b_nb} < \frac{2}{b^2}$.

Let $M = \frac{2}{b^2}$, then the original statement can be rewritten as

$$\left|\frac{b - b_n}{b_nb}\right| < M|b - b_n|.$$

Pick $N_2 \in \mathbf{N}$ such that for all $n \geqslant N_2$,

$$|b_n - b| < \frac{\varepsilon}{M}.$$

Let $N := \max\{N_1, N_2\}$. Then for all $n \geqslant N$,

$$\left|\frac{b - b_n}{b_n b}\right| < M \cdot \frac{\varepsilon}{M} = \varepsilon.$$

□

Theorem 10.6 (Sandwich theorem)**.** Let $a_n \leqslant c_n \leqslant b_n$ where $(a_n), (b_n)$ are convergent sequences such that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = L$, then (c_n) is also a converging sequence and $\lim_{n\to\infty} c_n = L$.

Proof. □

Exercise

Let (x_n) be a sequence of real numbers and let $\alpha \geqslant 2$ be a constant. Define the sequence (y_n) as follows:

$$y_n = x_n + \alpha x_{n+1} \quad (n = 1, 2, \dots)$$

Show that if (y_n) is convergent, then (x_n) is also convergent.

Exercise

(i) $\lim_{n\to\infty} \dfrac{1}{n_p} = 0$ $(p > 0)$.

(ii) $\lim_{n\to\infty} \sqrt[n]{p} = 1$ $(p > 0)$.

(iii) $\lim_{n\to\infty} \sqrt[n]{n} = 1$.

(iv) $\lim_{n\to\infty} \dfrac{n^\alpha}{(1+p)^n} = 0$ $(p > 0,\ \alpha \in \mathbf{R})$.

(v) $\lim_{n\to\infty} x^n = 0$ $(|x| < 1)$.

Subsequences

Definition 10.7 (Subsequence)**.** Given a sequence (x_n), consider a sequence (n_k) of positive integers such that $n_1 < n_2 < \cdots$. Then (x_{n_i}) is called a **subsequence** of (x_n). If (x_{n_i}) converges, its limit is called a *subsequential limit* of (x_n).

Proposition 10.8. (x_n) converges to x if and only if every subsequence of (x_n) converges to x.

Proof.

$\boxed{\implies}$ Suppose (x_n) converges to x. Then $\forall \varepsilon > 0, \exists N \in \mathbf{N}, \forall n \geqslant N, d(x_n, x) < \varepsilon$. Every subsequence of (x_n) can be written in the form (x_{n_i}) where $n_1 < n_2 < \cdots$ is a strictly increasing sequence of positive integers. Pick M such that $n_M > N$, then $\forall i > M$, $d(x_{n_i}, x) < \varepsilon$. Hence every subsequence of (x_n) converges to x.

$\boxed{\Longleftarrow}$ Intuitively, if every neighbourhood of x eventually contains all x_n, then since (x_{n_i}) is a subset of (x_n) they should all be contained in the neighbourhood eventually as well. $\square$

Lemma 10.9. If (x_n) is a sequence in a compact metric space X, then there exists a convergent subsequence of (x_n).

Proof. Let E be the range of (x_n). If E is finite then there exists $x \in E$ and a sequence (n_i) with $n_1 < n_2 < \cdots$ such that

$$x_{n_1} = x_{n_2} = \cdots = x.$$

The subsequence (x_{n_i}) so obtained converges evidently to x.

If E is infinite, Theorem 2.37 shows that E has a limit point $x \in X$. Choose n_1 so that $d(x, x_{n_1}) < 1$. Having chosen $n_1, \dots, n_{i-1}$, we see from Theorem 2.20 that there exists an integer $n_i > n_{i-1}$ such that $d(x, x_{n_i}) < \frac{1}{i}$. Then $x_{n_i} \to x$. $\square$

Proposition 10.10. Every bounded sequence in $\mathbf{R}^n$ contains a convergent subsequence.

Proof. This follows from the above proposition, since Theorem 2.41 implies that every bounded subset of $\mathbf{R}^n$ lies in a compact subset of $\mathbf{R}^n$. $\square$

The following is an important corollary.

Theorem 10.11 (Bolzano–Weierstrass)**.** Every bounded sequence in $\mathbf{R}$ contains a convergent subsequence.

Proposition 10.12. The subsequential limits of a sequence (x_n) in metric space X form a closed subset of X.

Proof. Let E be the set of all subsequential limits of (x_n), let q be a limit point of E. We want to show that $q \in E$.

Choose n_1 so that $x_{n_1} \neq q$. (If no such n_1 exists, then E has only one point, and there is nothing to prove.) Put $\delta = d(q, x_{n_1})$. Suppose $n_1, \dots, n_{i-1}$ are chosen. Since q is a limit point of E, there is an $x \in E$ with $d(x, q) < 2^{-1}\delta$. Since $x \in E$, there is an $n_i > n_{i-1}$ such that $d(x, x_{n_i}) < 2^{-i}\delta$. Thus

$$d(q, x_{n_i}) < 2^{1-i}\delta$$

for $i = 1, 2, 3, \dots$. This says that (x_{n_i}) converges to q. Hence $q \in E$. $\square$

Cauchy Sequences

This is a very helpful way to determine whether a sequence is convergent or divergent, as it does not require the limit to be known. In the future you will see many instances where the convergence of all sorts of limits are compared with similar counterparts; generally we describe such properties as *Cauchy criteria.*

Definition 10.13 (Cauchy sequence)**.** A sequence (x_n) in a metric space X is said to be a **Cauchy sequence** if

$$\forall \varepsilon > 0, \quad \exists N \in \mathbf{N}, \quad \forall n, m \geqslant N, \quad d(x_n, x_m) < \varepsilon.$$

Remark. This simply means that the distances between any two terms is sufficiently small after a certain point.

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 x, then they are also becoming close to each other. The converse is not always true, however.

Proposition 10.14. A sequence (x_n) in $\mathbf{R}^n$ is convergent if and only if it is Cauchy.

Proof.

$\boxed{\Longrightarrow}$ Suppose that (x_n) converges to x, then there exists $N \in \mathbf{N}$ such that $\forall n \geqslant N$, $|x_n - x| < \dfrac{\varepsilon}{2}$. Then for $n, m > N$, by triangle inequality,

$$|x_n - x_m| \leqslant |x_n - x| + |x_m - x| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence (x_n) is a Cauchy sequence.

$\boxed{\Longleftarrow}$ First, we show that (x_n) must be bounded. Pick $N \in \mathbf{N}$ such that $\forall n, m > N$ we have $|x_n - x_m| < 1$. Centered at x_n, we show that (x_n) is bounded; to do this we pick

$$r = \max\{1, |x_n - x_1|, \dots, |x_n - x_N|\}.$$

Then the sequence x_n is in $B_r(x_n)$ and thus is bounded.

Since (x_n) is bounded, by the corollary of Bolzano–Weierstrass we know that (x_n) contains a subsequence (x_{n_i}) that converges to x.

Then $\forall \varepsilon > 0$, pick $N_1 \in \mathbf{N}$ such that for all $n, m > N$, $|x_n - x_m| < \dfrac{\varepsilon}{2}$.

Simultaneously, since $\{x_{n_i}\}$ converges to x, pick M such that for $i > M$, $|x_{n_i} - x| < \dfrac{\varepsilon}{2}$.

Now, since $n_1 < n_2 < \cdots$ is a sequence of strictly increasing natural numbers, we can pick $i > M$ such that $n_i > N$. Then $\forall n \geqslant N$, by setting $m = n_i$ we obtain

$$|x_n - x_{n_i}| < \frac{\varepsilon}{2}, \quad |x_{n_i} - x| < \frac{\varepsilon}{2}$$

and hence

$$|x_n - x| \leqslant |x_n - x_{n_i}| + |x_{n_i} - x| < \varepsilon$$

by triangle inequality. Hence (x_n) is convergent. ∎

Definition 10.15 (Diameter)**.** Let nonempty $E \subseteq X$. Then the **diameter** of E is

$$\operatorname{diam} E := \sup_{x,y\in E} d(x,y).$$

Exercise

Find the diameter of the open unit ball in $\mathbf{R}^n$ given by

$$B = \{x \in \mathbf{R}^n \mid \|x\| < 1\}.$$

Solution. First note that

$$d(x,y) = \|x-y\| \leqslant \|x\| + \|-y\| = \|x\| + \|y\| < 1+1 = 2.$$

On the other hand, for any $\varepsilon > 0$, we pick

$$x = \left(1-\frac{\varepsilon}{4}, 0, \dots, 0\right), \quad y = \left(-\left(1-\frac{\varepsilon}{4}\right), 0, \dots, 0\right).$$

Then $d(x,y) = 2 - \dfrac{\varepsilon}{2} > 2 - \varepsilon$.

Therefore $\operatorname{diam} B = 2$. □

Proposition 10.16. $E \subseteq \mathbf{R}^n$ is bounded if and only if $\operatorname{diam} E < +\infty$.

Proof.

$\boxed{\Longrightarrow}$ If E is bounded, then there exists $M > 0$ such that $\|x\| \leqslant M$ for all $x \in E$.

Thus for any $x, y \in E$,

$$d(x,y) = \|x-y\| \leqslant \|x\| + \|y\| \leqslant 2M.$$

Thus $\operatorname{diam} E = \sup d(x,y) \leqslant 2M < +\infty$.

$\boxed{\Longleftarrow}$ Suppose that $\operatorname{diam} E = r$. Pick a random point $x \in E$, suppose that $\|x\| = R$.

Then for any other $y \in E$,

$$\|y\| = \|x + (y-x)\| \leqslant \|x\| + \|y-x\| \leqslant R + r.$$

Thus, by picking $M = R + r$, we obtain $\|y\| \leqslant M$ for all $y \in E$, and we are done.

Remark. Basically we used x to confine E within a ball, which is then confined within an even bigger ball centered at the origin.

□

Definition 10.17. A metric space is said to be **complete** if every Cauchy sequence converges.

Definition 10.18. A sequence (x_n) of real number is said to be

(i) **monotonically increasing** if $x_n \leqslant x_{n+1}$ $(n = 1, 2, \ldots)$;

(ii) **monotonically decreasing** if $x_n \geqslant x_{n+1}$ $(n = 1, 2, \ldots)$.

The class of monotonic sequences consists of the increasing and decreasing sequences.

Theorem 10.19 (Monotone convergence theorem)**.** If a monotonic sequence is bounded, then it converges.

Proof. Consider the case of a monotonically increasing sequence (x_n), which is bounded above by $M \in \mathbf{R}$. By the least upper bound property of $\mathbf{R}$, the (x_n) has a supremum L.

We claim that $x_n \to L$. For any $\varepsilon > 0$, by the definition of supremum, there exists $N \in \mathbf{N}$ such that $L - \varepsilon < x_N \leqslant L$. Since the sequence is increasing, for all $n \geqslant N$, we have $x_N \leqslant x_n \leqslant L$, so $L - \varepsilon < x_n \leqslant L$, which implies $|x_n - L| < \varepsilon$.

Thus $x_n \to L$. The proof for decreasing sequences is similar. □

Upper and Lower Limits

Let (x_n) be a real sequence.

Definition 10.20 (Convergence to infinity)**.** We write $x_n \to \infty$ if

$$\forall M \in \mathbf{R}, \quad \exists N \in \mathbf{N}, \quad \forall n \geqslant N, \quad x_n \geqslant M.$$

Similarly, we write $x_n \to -\infty$ if

$$\forall M \in \mathbf{R}, \quad \exists N \in \mathbf{N}, \quad \forall n \geqslant N, \quad x_n \leqslant M.$$

Definition 10.21 (Upper and lower limits)**.** Let (x_n) be a real sequence. Let $E \subset \overline{\mathbf{R}}$ be the set of all subsequential limits of (x_n), then

$$\limsup_{n\to\infty} x_n = \sup E,$$
$$\liminf_{n\to\infty} x_n = \inf E,$$

which are called the **limit superior** (or upper limit) and **limit infimum** (or lower limit) of (x_n) respectively.

Proposition 10.22. Let (x_n) be a real sequence. Then $\limsup_{n\to\infty} x_n$ has the following two properties:

(i) $\limsup_{n\to\infty} x_n \in E$.

(ii) If $x > \limsup_{n\to\infty} x_n$, there exists $N \in \mathbf{N}$ such that $n \geqslant N$ implies $x_n < x$.

Moreover, $\limsup_{n\to\infty} x_n$ is the only number with the properties (i) and (ii).

Example

- Let (x_n) be a sequence containing all rationals. Then every real number is a subsequential limit, and $\limsup_{n\to\infty} x_n = +\infty$, $\liminf_{n\to\infty} = -\infty$.
- For a real-valued seqeunce (x_n), $\lim_{n\to\infty} x_n = x$ if and only if $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = x$.

Proposition 10.23. If $a_n \leqslant b_n$ for $n \geqslant N$ where N is fixed, then

$$\liminf_{n\to\infty} a_n \leqslant \liminf_{n\to\infty} b_n,$$
$$\limsup_{n\to\infty} a_n \leqslant \limsup_{n\to\infty} b_n.$$

Proposition 10.24 (Arithmetic properties)**.**

(i) If $k > 0$, $\limsup\limits_{n\to\infty} ka_n = k \limsup\limits_{n\to\infty} a_n$.
If $k < 0$, $\limsup\limits_{n\to\infty} ka_n = k \liminf\limits_{n\to\infty} a_n$.

(ii) $\limsup(a_n + b_n) \leqslant \limsup a_n + \limsup b_n$

Moreover, $\limsup\limits_{n\to\infty}(a_n + b_n)$ may be bounded from below as follows:

$$\limsup_{n\to\infty}(a_n + b_n) \geqslant \limsup_{n\to\infty} a_n + \liminf_{n\to\infty} b_n.$$

write down the analogous properties for liminf, and to prove (i) and (ii)

Now you should try to prove (i) for liminf as well; as for (ii), try to explain why properties (i),(ii) for limsup and property (i) for liminf would imply property (ii) for liminf

§10.2 Series

Definition 10.25 (Series)**.** Given a sequence (a_n) in X, we associate a sequence (s_n), where

$$s_n = \sum_{k=1}^{n} a_k,$$

which we call a **series**. The term s_n is called the *n-th partial sum* of the series.

If the sequence of partial sums (s_n) converges to s, we say that the (infinite) series *converges*, and write $\sum\limits_{n=1}^{\infty} a_n = s$; that is,

$$\forall \varepsilon > 0, \quad \exists N \in \mathbf{N}, \quad \forall n \geqslant N, \quad \left| \sum_{k=1}^{n} a_k - s \right| < \varepsilon.$$

The number s is called the *sum* of the series.

If (s_n) diverges, the series is said to *diverge*.

Notation. When there is no possible ambiguity, we write $\sum_{n=1}^{\infty} a_n$ simply as $\sum a_n$.

The Cauchy criterion can be restated in the following form:

Proposition 10.26 (Cauchy criterion)**.** $\sum a_n$ converges if and only if

$$\forall \varepsilon > 0, \quad \exists N \in \mathbf{N}, \quad \forall m \geqslant n \geqslant N, \quad \left| \sum_{k=m}^{n} a_k \right| \leqslant \varepsilon.$$

Corollary 10.27 (Divergence test)**.** If $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.

The name of this result stems from its restatement: if $a_n \not\to 0$ as $n \to \infty$, then $\sum a_n$ diverges.

Proof. In the above proposition, take $m = n$, then $|a_n| \leqslant \varepsilon$ for all $n \geqslant N$. □

Remark. The converse is not true; we have the very well known counterexample of the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$.

Proposition 10.28. A series of non-negative terms converges if and only if its partial sums form a bounded sequence.

Convergence Tests

Proposition 10.29 (Geometric series)**.** If $0 \leqslant x < 1$, then

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

If $x \geqslant 1$, the series diverges.

Proof. For $0 < x < 1$,

$$\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}.$$

Taking limits $n \to \infty$, the result follows.

For $x = 1$, we get $1 + 1 + 1 + \cdots$, which evidently diverges. □

Lemma 10.30 (Cauchy condensation test)**.** Suppose $a_1 \geqslant a_2 \geqslant a_3 \geqslant \cdots \geqslant 0$. Then $\sum a_n$ converges if and only if the series

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots$$

converges.

Proof. By Theorem 3.24, it suffices to consider the boundedness of the partial suums. Let

$$\begin{aligned} s_n &= a_1 + a_2 + \cdots + a_n, \\ t_k &= a_1 + 2a_2 + \cdots + 2^k a_{2^k}. \end{aligned}$$

For $n < 2^k$,

$$\begin{aligned} s_n &\leqslant a_1 + (a_2 + a_3) + \cdots + (a_{2^k} + \cdots + a_{2^{k+1}-1}) \\ &\leqslant a_1 + 2a_2 + \cdots + 2^k a_{2^k} \\ &= t_k. \end{aligned}$$

Thus if (s_n) is unbounded, then (t_k) is unbounded.

For $n > 2^k$,

$$\begin{aligned} s_n &\geqslant a_1 + a_2 + (a_3 + a_4) + \cdots + (a_{2^{k-1}+1} + \cdots + a_{2^k}) \\ &\geqslant \frac{1}{2} a_1 + a_2 + 2a_4 + \cdots + 2^{k-1} a_{2^k} \\ &= \frac{1}{2} t_k. \end{aligned}$$

Thus if (t_k) is unbounded, then (s_n) is unbounded. □

Lemma 10.31 (p-test)**.** $\sum \frac{1}{n^p}$ converges if $p > 1$, and diverges if $p \leqslant 1$.

Proof. If $p \geqslant 0$, divergence follows from Theorem 3.23.

If $p > 0$, Theorem 3.27 is applicable, and we are led to the series

$$\sum_{k=0}^{\infty} 2^k \cdot \frac{1}{2^{kp}} = \sum_{k=0}^{\infty} 2^{(1-p)k}.$$

Now $2^{1-p} < 1$ if and only if $1-p < 0$, and the result follows by comparison with the geometric series (take $x = 2^{1-p}$ in Theorem 3.26). □

We introduce the following convergence tests to as a general method to determine whether an infinite series converges or diverges:

- Comparison test (Lemma 10.32)
- Root test (Lemma 10.33)
- Ratio test (Lemma 10.34)
- Absolute convergence (Lemma 10.35)

Lemma 10.32 (Comparison test)**.** Consider two sequences (a_n) and (b_n).

(i) If $|a_n| \leqslant b_n$ for all $n \geqslant N_0$ (where N_0 is some fixed integer), and if $\sum b_n$ converges, then $\sum a_n$ converges.

(ii) If $a_n \geqslant b_n \geqslant 0$ for all $n \geqslant N_0$, and if $\sum b_n$ diverges, then $\sum a_n$ diverges.

Proof.

(i) Since $\sum b_n$ converges, by the Cauchy criterion, fix $\varepsilon > 0$, there exists $N \geqslant N_0$ such that for $m \geqslant m \geqslant N$,

$$\sum_{k=n}^{m} b_k \leqslant \varepsilon.$$

By the triangle inequality,

$$\left|\sum_{k=n}^{m} a_k\right| \leqslant \sum_{k=n}^{m} |a_k| \leqslant \sum_{k=n}^{m} c_k \leqslant \varepsilon.$$

(ii) If $\sum a_n$ converges,

□

Lemma 10.33 (Root test)**.** Given $\sum a_n$, put $\alpha = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$. Then

(i) if $\alpha < 1$, $\sum a_n$ converges;

(ii) if $\alpha > 1$, $\sum a_n$ diverges;

(iii) if $\alpha = 1$, the test gives no information.

Proof.

(i) If $\alpha > 1$, we can choose β so that $\alpha < \beta < 1$, and $n \in \mathbf{N}$ such that for all $n \geqslant N$,

$$\sqrt[n]{|a_n|} < \beta.$$

by Theorem 3.17(b). Since $0 < \beta < 1$, $\sum \beta^n$ converges. Hence by the comparison test, $\sum a_n$ converges.

(ii) If $\alpha > 1$, by Theorem 3.17, there is a sequence (n_k) such that

$$\sqrt[n_k]{|a_{n_k}|} \to \alpha.$$

Hence $|a_n| > 1$ for infinitely many values of n so that the condition $a_n \to 0$, necessary for convergence of $\sum a_n$, does not hold (Theorem 3.23).

(iii) Consider the series $\sum \frac{1}{n}$ and $\sum \frac{1}{n^2}$. For each of these series $\alpha = 1$, but the first diverges, the second converges. Hence the condition that $\alpha = 1$ does not give us information on the convergence of a series.

□

Lemma 10.34 (Ratio test)**.** The series $\sum a_n$

(i) converges if $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$;

(ii) diverges if $\left|\frac{a_{n+1}}{a_n}\right| \geqslant 1$ for all $n \geqslant n_0$, where n_0 is some fixed integer.

Proof.

(i) If $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$, there exists $\beta < 1$ and $N \in \mathbf{N}$ such tht for all $n \geqslant N$,

$$\left|\frac{a_{n+1}}{a_n}\right| < \beta.$$

In particular, from $n = N$ to $n = N + p$,

$$\begin{aligned} |a_{N+1}| &< \beta |a_N| \\ |a_{N+2}| &< \beta |a_{N+1}| < \beta^2 |a_N| \\ &\vdots \\ |a_{N+p}| &< \beta^p |a_N| \end{aligned}$$

Hence for all $n \geqslant N$,

$$|a_n| < |a_N| \beta^{-N} \cdot \beta^n.$$

Since $\sum \beta^n$ converges, by the comparison test, $\sum a_n$ converges.

(ii) Suppose $\left|\frac{a_{n+1}}{a_n}\right| \geqslant 1$ for all $n \geqslant n_0$, where n_0 is some fixed integer. Then $|a_{n+1}| \geqslant |a_n|$ for $n \geqslant n_0$, and it is easily seen that $a_n \not\to 0$, so $\sum a_n$ diverges.

□

The series $\sum a_n$ is said to *converge absolutely* if the series $\sum |a_n|$ converges.

Lemma 10.35 (Absolute convergence). If $\sum a_n$ converges absolutely, then $\sum a_n$ converges.

Proof. □

Example (The number e)

The number e is defined as follows:

$$e \coloneqq \sum_{n=0}^{\infty} \frac{1}{n!}$$

Proposition. The number e is equivalent to the following:

$$\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e.$$

Proof. Let

$$s_n = \sum_{k=0}^{n} \frac{1}{k!}, \quad t_n = \left(1 + \frac{1}{n}\right)^n.$$

By the binomial theorem,

$$t_n = 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{n-1}{n}\right).$$

Comparing term by term, we see that $t_n \leqslant s_n$. By Proposition 10.23, we have that

$$\limsup_{n\to\infty} t_n \leqslant \limsup_{n\to\infty} s_n = e.$$

Next, if $n \geqslant m$,

$$t_n \geqslant 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \cdots + \frac{1}{m!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{m-1}{n}\right).$$

Let $n \to \infty$, keeping m fixed. We get

$$\liminf_{n\to\infty} t_n \geqslant 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{m!},$$

so that

$$s_m \leqslant \liminf_{n\to\infty} t_n.$$

Letting $m \to \infty$, we finally get

$$e \leqslant \liminf_{n\to\infty} t_n.$$

□

Proposition. e is irrational.

Proof. Suppose, for a contradiction, that e is rational. Then $e = \frac{p}{q}$, where p and q are positive integers. Let

$$s_n = \sum_{k=0}^{n} \frac{1}{k!}.$$

Then

$$\begin{aligned} e - s_n &= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \frac{1}{(n+3)!} + \cdots \\ &< \frac{1}{(n+1)!}\left[1 + \frac{1}{n+1} + \frac{1}{(n+1)^2}\right] \\ &= \frac{1}{(n+1)!}\frac{n+1}{n} \\ &= \frac{1}{n!n} \end{aligned}$$

and thus
$$0 < e - s_n < \frac{1}{n!n}.$$
Taking $n = q$ gives
$$0 < q!(e - s_q) < \frac{1}{q}.$$
Since $q!e$ is an integer (by our assumption), and
$$q!s_q = q!\left(1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{q!}\right)$$
is an integer, we have that $q!(e - s_n)$ is an integer. Since $q \geqslant 1$, this implies the existence of an integer between 0 and 1. We have thus reached a contradiction. □

Example (Power series)

Given a sequence (c_n) of complex numbers, the series
$$\sum_{n=0}^{\infty} c_n z^n$$
is called a **power series**. The numbers c_n are called the *coefficients* of the series.

In general, the series will converge or diverge, depending on the choice of z. More specifically, with every power series there is associated a circle, the circle of convergence, such that $\sum c_n z^n$ converges if z is in the interior of the circle and diverges if z is in the exterior.

Proposition. Given the power series $\sum c_n z^n$, let
$$\alpha = \limsup_{n\to\infty} \sqrt[n]{|c_n|}, \quad R = \frac{1}{\alpha}.$$
(If $\alpha = 0$, $R = +\infty$; if $\alpha = +\infty$, $R = 0$.) Then $\sum c_n z^n$

(i) converges if $|z| < R$,

(ii) diverges if $|z| > R$.

R is called the *radius of convergence* of $\sum c_n z^n$.

Proof. Put $a_n = c_n z^n$, then apply the root test:
$$\begin{aligned}\limsup_{n\to\infty} \sqrt[n]{|a_n|} &= \limsup_{n\to\infty} \sqrt[n]{|c_n z^n|} \\ &= |z| \limsup_{n\to\infty} \sqrt[n]{|c_n|} \\ &= |z|\alpha \\ &= \frac{|z|}{R}.\end{aligned}$$

(i) If $|z| < R$, then $\limsup_{n\to\infty} \sqrt[n]{|a_n|} < 1$. By the root test, $\sum c_n z^n$ converges.

(ii) If $|z| > R$, then $\limsup_{n\to\infty} \sqrt[n]{|a_n|} > 1$. By the root test, $\sum c_n z^n$ diverges.

□

Further properties of power series will be discussed in Chapter 15.

Summation by Parts

Proposition 10.36 (Partial summation formula)**.** Given two sequences (a_n) and (b_n), put

$$A_n = \sum_{k=0}^{n} a_k$$

if $n \geqslant 0$; put $A_{-1} = 0$. Then, if $0 \leqslant p \leqslant q$, we have

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p.$$

Proof. The RHS can be written as

$$\begin{aligned}
&\sum_{n=p}^{q-1} A_n b_n + A_q b_q - \sum_{n=p}^{q-1} A_n b_{n+1} - A_{p-1} b_p \\
&= \sum_{n=p}^{q} A_n b_n - \sum_{n=p-1}^{q-1} A_n b_{n+1} \\
&= \sum_{n=p}^{q} A_n b_n - \sum_{n=p}^{q} A_{n-1} b_n \\
&= \sum_{n=p}^{q} (A_n - A_{n-1})\, b_n \\
&= \sum_{n=p}^{q} a_n b_n
\end{aligned}$$

which is equal to the LHS. □

Proposition 10.37. Suppose the partial sums A_n of $\sum a_n$ form a bounded sequence, $b_0 \geqslant b_1 \geqslant b_2 \geqslant \cdots$, and $\lim_{n\to\infty} b_n = 0$. Then $\sum a_n b_n = 0$.

Proof. □

Proposition 10.38. Suppose $|c_1| \geqslant |c_2| \geqslant |c_3| \geqslant \cdots$, $c_{2m-1} \geqslant 0, c_{2m} \leqslant 0$ for $m = 1, 2, 3, \ldots$, and $\lim_{n\to\infty} c_n = 0$. Then $\sum c_n$ converges.

Addition and Multiplication of Series

Proposition 10.39. If $\sum a_n = A$ and $\sum b_n = B$, then

(i) $\sum (a_n + b_n) = A + B$,

(ii) $\sum ca_n = cA$ for any fixed c.

Proof.

(i) Let $A_n = \sum_{k=0}^n a_k$, $B_n = \sum_{k=0}^n b_k$. Then

$$A_n + B_n = \sum_{k=0}^{n} (a_k + b_k).$$

Since $\lim_{n\to\infty} A_n = A$ and $\lim_{n\to\infty} B_n = B$, we see that

$$\lim_{n\to\infty} (A_n + B_n) = A + B.$$

(ii)

□

Thus two convergent series may be added term by term, and the resulting series converges to the sum of the two series. The situation becomes more complicated when we consider multiplication of two series. To begin with, we have to define the product. This can be done in several ways; we shall consider the so-called "Cauchy product".

Definition 10.40 (Cauchy product)**.** Given $\sum a_n$ and $\sum b_n$, let

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \quad (n = 0, 1, 2, \dots)$$

We call $\sum c_n$ the *product* of the two given series.

This definition may be motivated as follows. If we take two power series $\sum a_n z^n$ and $\sum b_n z^n$, multiply them term by term, and collect terms containing the same power of z, we get

$$\begin{aligned}\sum_{n=0}^{\infty} a_n z^n \cdot \sum_{n=0}^{\infty} b_n z^n &= \left(a_0 + a_1 z + a_2 z^2 + \cdots\right)\left(b_0 + b_1 z + b_2 z^2 + \cdots\right) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0) z + (a_0 b_2 + a_1 b_1 + a_2 b_0) z^2 + \cdots \\ &= c_0 + c_1 z + c_2 z^2.\end{aligned}$$

Setting $z = 1$, we arrive at the above definition.

Theorem 10.41 (Mertens)**.** Suppose $\sum a_n = A$, $\sum b_n = B$, $\sum a_n$ converges absolutely. Then their Cauchy product converges to AB.

That is, the product of two convergent series converges, and to the right value, if at least one of the two series converges absolutely.

Proof. Let $A_n = \sum_{k=0}^{n} a_k$, $B_n = \sum_{k=0}^{n} b_k$, $C_n = \sum_{k=0}^{n} c_k$. Also let $\beta_n = B_n - B$. Then

$$\begin{aligned} C_n &= a_0b_0 + (a_0b_1 + a_1b_0) + \cdots + (a_0b_n + a_1b_{n-1} + \cdots + a_nb_0) \\ &= a_0B_n + a_1B_{n-1} + \cdots + a_nB_0 \\ &= a_0(B+\beta_n) + a_1(B+\beta_{n-1}) + \cdots + a_n(B+\beta_0) \\ &= A_nB + a_0\beta_n + a_1\beta_{n-1} + \cdots + a_n\beta_0. \end{aligned}$$

Let

$$\gamma_n = a_0\beta_n + a_1\beta_{n-1} + \cdots + a_n\beta_0.$$

We wish to show that $C_n \to AB$. Since $A_nB \to AB$, it suffices to show that $\lim_{n\to\infty} \gamma_n = 0$.

Let

$$\alpha = \sum_{n=0}^{\infty} |a_n|.$$

Let $\varepsilon > 0$. Since $B_n \to B$, $\beta_n \to 0$. Hence we can choose $N \in \mathbf{N}$ such that for all $n \geqslant N$, $|\beta_n| \leqslant \varepsilon$, in which case

$$\begin{aligned} |\gamma_n| &= |\beta_0a_n + \cdots + \beta_Na_{n-N}| + |\beta_{N+1}a_{n-N}a_{n-N-1} + \cdots + \beta_na_0| \\ &\leqslant |\beta_0a_n + \cdots + \beta_Na_{n-N}| + \varepsilon\alpha. \end{aligned}$$

Keeping N fixed, and letting $n \to \infty$, we get

$$\limsup_{n\to\infty} |\gamma_n| \leqslant \varepsilon\alpha,$$

sine $a_k \to 0$ as $k \to \infty$. Since ε is arbitrary, we have $\lim_{n\to\infty} \gamma_n = 0$, as desired. □

Theorem 10.42 (Abel)**.** Let the series $\sum a_n$, $\sum b_n$, $\sum c_n$ converge to A, B, C respectively, and $\sum c_n$ is the Cauchy product of $\sum a_n$ and $\sum b_n$. Then $C = AB$.

Rearrangements

Definition 10.43 (Rearrangement)**.** Let (k_n) be a sequence in which every positive integer appears once and only once. Putting

$$a'_n = a_{k_n} \quad (\forall n \in \mathbf{N})$$

we say that $\sum a'_n$ is a *rearrangement* of $\sum a_n$.

Proposition 10.44. Let $\sum a_n$ be a series of real numbers which converges, but not absolutely. Suppose $-\infty \leqslant \alpha \leqslant \beta \leqslant \infty$. Then there exists a rearrangement $\sum a'_n$ with partial sums s'_n such that

$$\liminf_{n\to\infty} s'_n = \alpha, \quad \limsup_{n\to\infty} s'_n = \beta.$$

Proposition 10.45. If $\sum a_n$ is a series of complex numbers which converges absolutely, then every rearrangement of $\sum a_n$ converges, and they all converge to the same sum.

11 Continuity

§11.1 Limit of Functions

Let (X, d_X) be a metric space, let $E \subseteq X$. Then the metric d_X induces a metric on E. Now consider a mapping f (or function) from E into another metric space (Y, d_Y).

In particular, if $Y = \mathbf{R}$, f is called a **real-valued function**; and if $Y = \mathbf{C}$, f is called a **complex-valued function**.

Definition 11.1 (Limit of function)**.** Consider a limit point $p \in E$. We say $\lim_{x \to p} f(x) = q$ if there exists a point $q \in Y$ such that

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall x \in E, \quad 0 < d_X(x,p) < \delta \implies d_Y\left(f(x), q\right) < \varepsilon.$$

In words, this means no matter what $B_\varepsilon(q)$ we are given, we can always find a $B_\delta(p)$ succh that $f\left(\overline{B}_\delta(p) \cap E\right) \subset B_\varepsilon(q)$.

We can recast this definition in terms of limits of sequences:

Theorem 11.2. *$\lim_{x \to p} f(x) = q$ if and only if $\lim_{n \to \infty} f(p_n) = q$ for every sequence (p_n) in E such that $p_n \neq p$, $\lim_{n \to \infty} p_n = p$.*

Proof.

$\boxed{\implies}$ Suppose $\lim_{x \to p} f(x) = q$. Choose (p_n) in E satisfying $p_n \neq p$ and $\lim_{n \to \infty} p_n = p$. We now want to show that $\lim_{n \to \infty} f(p_n) = q$.

Let $\varepsilon > 0$ be given. Since $\lim_{x \to p} f(x) = q$, there exists $\delta > 0$ such that

$$\forall x \in E, \quad 0 < d_X(x,p) < \delta \implies d_Y\left(f(x), q\right) < \varepsilon.$$

Also, since $\lim_{n \to \infty} p_n = p$, there exists $N \in \mathbf{N}$ such that

$$\forall n \geqslant N, \quad 0 < d_X(p_n, p) < \delta.$$

Thus for $n \geqslant N$, we have $d_Y\left(f(p_n), q\right) < \varepsilon$, which shows that $\lim_{n \to \infty} f(p_n) = q$.

$\boxed{\impliedby}$ We now prove the reverse direction by contrapositive. Suppose $\lim_{x \to p} f(x) \neq q$. Then

$$\exists \varepsilon > 0, \quad \forall \delta > 0, \quad \exists x \in E, \quad d_Y\left(f(x), q\right) \geqslant \varepsilon \quad \text{and} \quad 0 < d_X(x,p) < \delta.$$

Taking $\delta_n = \frac{1}{n}$ $(n = 1, 2, \dots)$, we thus find a sequence in E satisfying $p_n \neq p$ and $\lim_{n\to\infty} p_n = p$ for which $\lim_{n\to\infty} f(p_n) \neq q$. □

Corollary 11.3. If f has a limit at p, this limit is unique.

Proof. This follows from and Theorem 11.2. □

Proposition 11.4. Suppose $E \subseteq X$, limit point $p \in E$, $f, g : E \to \mathbf{R}$. Let $\lim_{x\to p} f(x) = A$ and $\lim_{x\to p} g(x) = B$. Then

(i) $\lim_{x\to p}(f+g)(x) = A + B$

(ii) $\lim_{x\to p}(fg)(x) = AB$

(iii) $\lim_{x\to p}\left(\dfrac{p}{q}\right)(x) = \dfrac{A}{B}$ $(B \neq 0)$

Proof. By the same proofs as for sequences, limits are unique, and in $\mathbf{R}$ they add/multiply/divide as expected. □

§11.2 Continuous Functions

Consider metric spaces (X, d_X) and (Y, d_Y), let $E \subseteq X$.

Definition 11.5 (Continuity)**.** We say that $f : E \to Y$ is **continuous** at $p \in E$ if

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall x \in X, \quad d_X(x,p) < \delta \implies d_Y\left(f(x), f(p)\right).$$

We say f is continuous in E if it is continuous at every point of E.

Lemma 11.6. Assume p is a limit point of E. Then f is continuous at p if and only if $\lim_{x\to p} f(x) = f(p)$.

Proof. Compare Definitions 4.1 and 4.5. □

Theorem 11.7 (Sequential criterion for continuity)**.** $f : E \subseteq X \to Y$ is continuous at $p \in E$ if and only if for every sequence (x_n) in E that converges to p, the sequence $(f(x_n))$ converges to $f(p)$.

Proof. The sequential definition of continuity follows almost directly from the sequential definition of limits. □

As for real-valued functions, the definition of continuity can be phrased in terms of limits.

Corollary 11.8. $f : X \to \mathbf{R}$ is continuous at $p \in X$ if and only if for any sequence (x_n) with $\lim_{n\to\infty} x_n = p$, we have $\lim_{n\to\infty} f(x_n) = f(p)$.

We now consider the composition of functions. The following result shows that a continuous function of a continuous function is continous.

Proposition 11.9. Suppose X, Y, Z are metric spaces, $E \subseteq X$, $f : E \to Y$, g maps the range of $f(E)$ into Z, $h : E \to Z$ defined by

$$h(x) = g \circ f(x) \quad (x \in E)$$

If f is continuous at $p \in E$, and g is continuous at $f(p)$, then h is continuous at p.

Proof. Let $\varepsilon > 0$ be given. Since g is continous at $f(p)$, there exists $\eta > 0$ such that for all $y \in f(E)$,

$$d_Y(y, f(p)) < \eta \implies d_Z\Big(g(y), g(f(p))\Big) < \varepsilon$$

Since f is continuous at p, there exists $\delta > 0$ such that for all $x \in E$,

$$d_X(x, p) < \delta \implies d_Y(f(x), f(p)) < \eta$$

It follows that for all $x \in E$,

$$d_X(x, p) < \delta \implies d_Z(h(x), h(p)) = d_Z\Big(g(f(x)), g(f(p))\Big) < \varepsilon$$

Thus h is continuous at p. □

Proposition 11.10. $f : X \to Y$ is continuous on X if and only if $f^{-1}(V)$ is open in X for every open set $V \subseteq Y$.

Proof.

$\boxed{\implies}$ Suppose f is continuous on X, $V \subseteq Y$ is open. We have to show that every point of $f^{-1}(V)$ is an interior point of $f^{-1}(V)$.

So, suppose $p \in X$ and $f(p) \in V$. Since V is open, there exists $\varepsilon > 0$ such that $y \in V$ if $d_Y(f(p), y) < \varepsilon$; and since f is continuous at p, there exists $\delta > 0$ such that $d_Y(f(x), f(p)) < \varepsilon$ if $d_X(x, p) < \delta$. Thus $x \in f^{-1}(V)$ as soon as $d_X(x, p) < \delta$.

$\boxed{\impliedby}$ Conversely, suppose $f^{-1}(V)$ is open in X for every open set $V \subseteq Y$. Fix $p \in X$ and $\varepsilon > 0$, let $V = \{y \in Y \mid d_Y(y, f(p))\} < \varepsilon$. Then V is open; hence $f^{-1}(V)$ as soon as $d_X(p, x) < \delta$. But if $x \in f^{-1}(V)$, then $f(x) \in V$, so that $d_Y(f(x), f(p)) < \varepsilon$. □

Corollary 11.11. $f : X \to Y$ is continuous if and only if $f^{-1}(C)$ is closed in X for every closed set $C \subseteq Y$.

Proof. This follows from the above result, since a set is closed if and only if its complement is open, and since $f^{-1}(E^c) = [f^{-1}(E)]^c$ for every $E \subseteq Y$. □

Proposition 11.12. Let $f, g : X \to \mathbf{R}$. Then $f + g$, fg, and $\frac{f}{g}$ ($g(x) \neq 0$ for all $x \in X$) are continuous on X.

Proof. At isolated points of X there is nothing to prove. At limit points, the statement follows from Theorems 4.4 and 4.6 □

Continuity of linear functions in normed spaces

A great deal of power comes from considering the set of all functions on a space satisfying some property, such as continuity, as a metric space in its own right. In this section we consider some important examples of such spaces.

We begin with the space of bounded real-valued functions on a set X. At this stage we assume nothing about X.

Definition 11.13 (Space of bounded real-valued functions)**.** If X is any set, we define $B(X)$ to be the space of functions $f : X \to \mathbf{R}$ for which $f(X) = \{f(x) \mid x \in X\}$ is bounded. If $f \in B(X)$, define $\|f\|_\infty = \sup_{x\in X} |f(x)|$.

Lemma 11.14. For any set X, $B(X)$ is a vector space, and $\|\cdot\|_\infty$ is a norm.

Proof. □

Now we turn to the space of continuous real-valued functions, $C(X)$. To make sense of what this means we now need X to be a metric space.

Definition 11.15. Let X be a metric space. We write $C(X)$ for the space of all continuous functions $f : X \to \mathbf{R}$.

§11.3 Continuity and Compactness

Assume (X, d_X) and (Y, d_Y) are metric spaces.

Definition 11.16 (Bounded)**.** $f : E \to \mathbf{R}^n$ is said to be **bounded** if there exists $M \in \mathbf{R}$ such that $|f(x)| \leqslant M$ for all $x \in E$.

Theorem 11.17. Suppose $f : X \to Y$ is continuous. Then for any compact subset $K \subseteq X$, the image set $f(K)$ is a compact subset of Y.

Proof. We prove it by definition. Assume $\{V_i \mid i \in I\}$ is an open cover of $f(K)$. By the continuity of f and □

Theorem 11.18 (Extreme Value Theorem)**.** A continuous function on a compact set attains its maximum and minimum values.

Definition 11.19 (Uniform continuity)**.** Let (X, d_X) and (Y, d_Y) be metric spaces, let $E \subseteq X$. We say that $f : E \to Y$ is **uniformly continuous** if

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall x, y \in E, \quad d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon.$$

Let us consider the differences between the concepts of continuity and of uniform continuity. First, uniform continuity is a property of a function on a set, whereas continuity can be defined at a single point. To ask whether a given function is uniformly continuous at a certain point is meaningless. Second, if f is continuous on X, then it is possible to find, for each $\varepsilon > 0$ and for each point $p \in X$, a number $\delta > 0$ having the property specified in Definition 4.5. This δ depends ε *and* on p. If f is, however, uniformly continuous on X, then it is possible, for each $\varepsilon > 0$, to find *one* number $\delta > 0$ which will do for *all* points $p \in X$.

Evidently, every uniformly continuous function is continuous. That the two concepts are equivalent on compact sets follows from the next theorem.

Proposition 11.20. Let $f : E \subseteq X \to Y$ be continuous. Then f is uniformly continous.

Proof. □

§11.4 Continuity and Connectedness

Proposition 11.21. If $f : X \to Y$ is continous, and if $E \subseteq X$ is connected, then $f(E)$ is connected.

Proof. □

Theorem 11.22 (Intermediate Value Theorem)**.** Let $f : [a, b] \to \mathbf{R}$ be continuous. If $f(a) < f(b)$ and $f(a) < c < f(b)$, then there exists $x \in (a, b)$ such that $f(x) = c$.

Proof. □

§11.5 Discontinuities

Let $f : X \to Y$. If f is not continuous at $x \in X$, we say that f is discontinuous at x, or that f has a discontinuity at x.

If f is defined on an interval or a segment, it is customary to divide discontinuities into two types. Before giving this classification, we have to define the **right-hand** and the **left-hand limits** of f at x, denoted by $f(x+)$ and $f(x-)$ respectively.

Definition 11.23 (Right-hand and left-hand limits)**.** Let $f : (a, b) \to \mathbf{R}$. Consider any point x such that $a \leqslant x < b$.

Definition 11.24 (Discontinuities)**.** Let $f : [a, b] \to \mathbf{R}$. If f is discontinuous at x, and if $f(x+)$ and $f(x-)$ exist, then f is said to have a **discontinuity of the first kind**, or a **simple discontinuity**, at x. Otherwise the discontinuity is said to be of the **second kind**.

There are two ways in which a function can have a simple discontinuity: either

§11.6 Monotonic Functions

Proposition 11.25. Let $f : [a, b] \to \mathbf{R}$ be monotonically increasing. Then $f(x+)$ and $f(x-)$ exist for all $x \in (a, b)$; more precisely,

$$\sup_{t\in(a,x)} f(t) = f(x-) \leqslant f(x) \leqslant f(x+) = \inf_{t\in(x,b)} f(t).$$

Furthermore, if $a < x < y < b$, then

$$f(x+) \leqslant f(y-).$$

Analogous results evidently hold for monotically decreasing functions.

§11.7 Infinite Limits and Limits at Infinity

Definition 11.26. For $c \in \mathbf{R}$, the set $\{x \in \mathbf{R} \mid x > c\}$ is called a neighbourhood of $+\infty$ and is written $(c, +\infty)$. Similarly, the set $(-\infty, c)$ is a neighbourhood of $-\infty$.

Definition 11.27. Let $f : E \subset \mathbf{R} \to \mathbf{R}$. We say that $\lim\limits_{t\to x} f(t) = A$ where A and x are in the extended real number system, if for every neighbourhood of U of A there is a neighbourhood V of x such that $V \cap E$ is not empty, and such that $f(t) \in U$ for all $t \in V \cap E$, $t \neq x$.

12 Differentiation

§12.1 The Derivative of A Real Function

Definition 12.1 (Derivative)**.** Suppose $f : [a,b] \to \mathbf{R}$. For any $x \in [a,b]$, if the limit

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} \quad (a < t < b, t \neq x)$$

exists, we call it f', known as the **derivative** of f.

If f' is defined at a point x, we say that f is **differentiable** at x; If f' is defined at every point of a set $E \subseteq [a,b]$, we say that f is differentiable on E.

Lemma 12.2 (Differentiability implies continuity)**.** If $f : [a,b] \to \mathbf{R}$ is differentiable at $x \in [a,b]$, then f is continuous at x.

Proof.

$$\begin{aligned}
\lim_{t \to x}[f(t) - f(x)] &= \lim_{t \to x}\left[\frac{f(t) - f(x)}{t - x} \cdot (t - x)\right] \\
&= \lim_{t \to x} \frac{f(t) - f(x)}{t - x} \cdot \lim_{t \to x}(t - x) \\
&= f'(x) \cdot 0 = 0.
\end{aligned}$$

Since $\lim_{t \to x} f(t) = f(x)$, f is continuous at x. □

Remark. The converse of Lemma 12.2 is not true; it is easy to construct continuous functions which fail to be differentiable at isolated points.

> **Example** (Weierstrass function)
>
> Let $0 < a < 1$, let $b > 1$ be an odd integer, and $ab > 1 + \frac{3}{2}\pi$. Then the function
>
> $$W(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$
>
> is continuous and nowhere differentiable on $\mathbf{R}$.

Notation. If f has a derivative f' on an interval, and if f' is itself differentiable, we denote the derivative of f' by f'', and call f'' the **second derivative** of f. Continuing in this manner, we obtain functions

$$f, f', f'', f^{(3)}, f^{(4)}, \dots, f^{(n)},$$

each of which is the derivative of the preceding one. f' is called the n-th derivative (or the derivative or order n) of f.

Remark. In order for $f^{(n)}(x)$ to exist at a point x, $f^{(n-1)}(t)$ must exist in a neighbourhood of x (or a one-sided neighbourhood, if x is an endpoint of the interval on which f is defined), and $f^{(n-1)}(x)$ must be differentiable at x.

Notation. $C_1[a,b]$ denotes the set of differentiable functions over $[a,b]$ whose derivative is continuous. More generally, $C_n[a,b]$ denotes the set of functions whose n-th derivative is continuous. In particular, $C_0[a,b]$ is the set of continuous functions over $[a,b]$.

Later on when we talk about properties of differentiation such as the intermediate value theorems, we usually have the following requirement on the function:

> f is a continuous function on $[a,b]$ which is differentiable in (a,b).

Lemma 12.3 (Differentiation rules)**.** Suppose $f,g:[a,b]\to\mathbf{R}$ are differentiable at $x\in[a,b]$. Then

(i) Scalar multiplication: for $\alpha\in\mathbf{R}$, αf is differentiable at x, and

$$(\alpha f)'(x)=\alpha f'(x).$$

(ii) Addition: $f\pm g$ is differentiable at x, and

$$(f\pm g)'(x)=f'(x)\pm g'(x).$$

(iii) Product rule: fg is differentiable at x, and

$$(fg)'(x)=f'(x)g(x)+f(x)g'(x).$$

(iv) Quotient rule: f/g (when $g(x)\neq 0$) is differentiable at x, and

$$\left(\frac{f}{g}\right)'(x)=\frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}.$$

Proof.

(i)

(ii)

$$\begin{aligned}\frac{(f+g)(t)-(f+g)(x)}{t-x}&=\frac{f(t)+g(t)-f(x)-g(x)}{t-x}\\&=\frac{f(t)-f(x)}{t-x}+\frac{g(t)-f(x)}{t-x}\end{aligned}$$

Taking limits $t\to x$, the first term equals to $f'(x)$, and the second term equals to $g'(x)$. The case for subtraction is analogous.

(iii)

$$\begin{aligned}\frac{(fg)(t)-(fg)(x)}{t-x} &= \frac{f(t)g(t)-f(x)g(x)}{t-x}\\ &= \frac{[f(t)-f(x)]\,g(t)+f(x)\,[g(t)-g(x)]}{t-x}\\ &= \frac{f(t)-f(x)}{t-x}\cdot g(t)+f(x)\cdot\frac{g(t)-g(x)}{t-x}\\ &= f'(x)g(x)+f(x)g'(x)\end{aligned}$$

Taking limits $t\to x$, the first term equals to $f'(x)g(x)$, and the second term equals to $f(x)g'(x)$, so we are done.

(iv) Similarly,

$$\frac{\left(\frac{f}{g}\right)(t)-\left(\frac{f}{g}\right)(x)}{t-x} = \frac{1}{g(t)g(x)}\left[g(x)\cdot\frac{f(t)-f(x)}{t-x} - f(x)\cdot\frac{g(t)-g(x)}{t-x}\right]$$

Taking limits $t\to x$, the result immediately follows.

□

By induction, we can obtain the following extensions of the differentiation rules.

Corollary 12.4. Suppose $f_1,f_2,\dots,f_n:[a,b]\to\mathbf{R}$ are differentiable at $x\in[a,b]$. Then

(i) $f_1+f_2+\cdots+f_n$ is differentiable at x, and

$$(f_1+f_2+\cdots+f_n)'(x)=f_1{}'(x)+f_2{}'(x)+\cdots+f_n{}'(x).$$

(ii) $f_1f_2\cdots f_n$ is differentiable at x, and

$$\begin{aligned}(f_1f_2\cdots f_n)'(x)=&f_1{}'(x)f_2(x)\cdots f_n(x)+f_1(x)f_2{}'(x)\cdots f_n(x)\\ &+\cdots+f_1(x)f_2(x)\cdots f_n{}'(x).\end{aligned}$$

Theorem 12.5 (Chain rule)**.** Suppose f is continuous on $[a,b]$, $f'(x)$ exists at $x\in[a,b]$, g is defined on I that contains $f([a,b])$, and g is differentiable at $f(x)$. Then the composition

$$h(x)\coloneqq g\circ f(x)=g\left(f(x)\right):[a,b]\to\mathbf{R}$$

is differentiable at x, and the derivative at x can be calculated as

$$h'(x)=g'\left(f(x)\right)f'(x).$$

Proof. Let $y=f(x)$. By the definition of the derivative, we have

$$f(t)-f(x)=(t-x)[f'(x)+u(t)] \tag{1}$$

$$g(s)-g(y)=(s-y)[g'(y)+v(s)] \tag{2}$$

where $t\in[a,b]$, $s\in I$, $\lim_{t\to x}u(t)=0$, $\lim_{s\to y}v(s)=0$.

Let $s=f(t)$. Using first (2) and then (1), we obtain

$$\begin{aligned} h(t)-h(x) &= g(f(t))-g(f(x)) \\ &= [f(t)-f(x)]\cdot[g'(y)+v(s)] \\ &= (t-x)[f'(x)+u(t)][g'(y)+v(s)], \end{aligned}$$

or, if $t\neq x$,

$$\frac{h(t)-h(x)}{t-x}=[g'(y)+v(s)][f'(x)+u(t)].$$

Letting $t\to x$, we see that $s\to y$, by the continuity of f, so that the RHS of the above equation tends to $g'(y)f'(x)$, thus giving us the desired result. □

> **Example**
>
> One family of pathological examples in calculus is functions of the form
>
> $$f(x)=x^p\sin\frac{1}{x}.$$
>
> For $p=1$, the function is continuous and differentiable everywhere other than $x=0$; for $p=2$, the function is differentiable everywhere, but the derivative is discontinuous.

§12.2 Mean Value Theorems

Let (X,d) be a metric space.

Definition 12.6 (Local maximum and minimum)**.** We say that $f:X\to\mathbf{R}$ has

(i) a **local maximum** at $x_0\in X$ if there exists $\delta>0$ such that $f(x_0)\geqslant f(x)$ for all $x\in B_\delta(x_0)$;

(ii) a **local minimum** at $x_0\in X$ if there exists $\delta>0$ such that $f(x_0)\leqslant f(x)$ for all $x\in B_\delta(x_0)$.

Lemma 12.7 (Fermat's theorem)**.** Suppose $f:[a,b]\to\mathbf{R}$. If f has a local maximum or minimum at $x_0\in(a,b)$, and if $f'(x_0)$ exists, then $f'(x_0)=0$.

Proof. If f is not differentiable at x_0, we are done. Assume now f is differentiable at x_0 and x_0 is a local maximum. By definition, there exists $\delta>0$ such that $f(x_0)\leqslant f(x)$, for all $x\in B_\delta(x_0)$. Then

$$\frac{f(x)-f(x_0)}{x-x_0}\begin{cases}\geqslant 0 & x_0-\delta<x<x+\delta\\ \leqslant 0 & x_0<x<x_0+\delta\end{cases}$$

Since $f'(x_0)$ exists, we have

$$f'(x_0-)\geqslant 0,\quad f'(x_0+)\leqslant 0,$$

but we know that $f'(x_0-) = f'(x_0+) = f'(x_0)$ since f is differentiable at x_0. Hence $f'(x_0) = 0$. □

Theorem 12.8 (Rolle's theorem)**.** If f is continuous on $[a, b]$, differentiable in (a, b) and $f(a) = f(b)$, then there exists $c \in (a, b)$ such that

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

Proof. Let $h(x)$ be a function defined on $[a, b]$ where $h(a) = h(b)$.

The idea is to show that h has a local maximum/minimum, then by Fermat's Theorem this will then be the stationary point that we're trying to find.

First note that h is continuous on $[a, b]$, so h must have a maximum M and a minimum m.

If M and m were both equal to $h(a) = h(b)$, then h is just a constant function and so $h'(x) = 0$ everywhere.

Otherwise, h has a maximum/minimum that is not $h(a) = h(b)$, so this extremal point lies in (a, b).

In particular, this extremal point is also a local extremum. Since h is differentiable on (a, b), by Fermat's theorem this extremum point is stationary, thus Rolle's Theorem is proven. □

Theorem 12.9 (Generalised mean value theorem)**.** If f and g are continuous on $[a, b]$ and differentiable in (a, b), then there exists $c \in (a, b)$ such that

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

Proof. For $t \in [a, b]$, put

$$h(t) = [f(b) - f(a)]g(t) - [g(b) - g(a)]f(t).$$

Then h is continuous on $[a, b]$, and h is differentiable on (a, b). Moreover,

$$h(a) = f(b)g(a) - f(a)g(b) = h(b)$$

thus by Rolle's Theorem, there exists $c \in (a, b)$ such that $h'(c) = 0$, i.e. $[g(b) - g(a)]f'(c) = [f(b) - f(a)]g'(c)$ □

Theorem 12.10 (Mean value theorem)**.** If f is continuous on $[a, b]$ and differentiable in (a, b), then there exists $c \in (a, b)$ such that

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

Proof. Take $g(x) = x$ in Theorem 12.9. □

Proposition 12.11. Suppose f is differentiable in (a, b).

(i) If $f'(x) \geqslant 0$ for all $x \in (a, b)$, then f is monotonically increasing.

(ii) If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.

(iii) If $f'(x) \leqslant 0$ for all $x \in (a,b)$, then f is monotonically decreasing.

Proof. All conclusions can be read off from the equation

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

which is valid, for each pair of numbers x_1, x_2 in (a,b), for some x between x_1 and x_2. □

Exercise

Let f and g be continuous on $[a,b]$ and differentiable on (a,b). If $f'(x) = g'(x)$, then $f(x) = g(x) + C$.

Exercise

Given that $f(x) = x^\alpha$ where $0 < \alpha < 1$. Prove that f is uniformly continuous on $[0,+\infty)$.

Exercise

Let f be a function continuous on $[0,1]$ and differentiable on $(0,1)$ where $f(0) = f(1) = 0$. Prove that there exists $c \in (0,1)$ such that

$$f(x) + f'(x) = 0.$$

§12.3 Darboux's Theorem

The following result implies some sort of a "intermediate value" property of derivatives that is similar to continuous functions.

Theorem 12.12 (Darboux's Theorem)**.** *Suppose f is differentiable on $[a,b]$, and suppose $f'(a) < c < f'(b)$. Then there exists $x \in (a,b)$ such that $f'(x) = c$.*

Proof. Put $g(t) = f(t) - ct$. Then $g'(a) < 0$, so that $g(t_1) < g(a)$ for some $t_1 \in (a,b)$, and $g'(b) > 0$, so that $g(t_2) < g(b)$ for some $t_2 \in (a,b)$.

Hence g attains its minimum on $[a,b]$ (Theorem 4.16) at some point x such that $a < x < b$. By Theorem 5.8, $g'(x) = 0$. Hence $f'(x) = c$. □

Corollary 12.13. If f is differentiable on $[a,b]$, then f' cannot have any simple discontinuities on $[a,b]$.

Remark. But f' may very well have discontinuities of the second kind.

§12.4 L'Hopital's Rule

The following theorem is frequently used in the evaluation of limits.

Theorem 12.14 (L'Hopital's Rule)**.** Suppose f and g are differentiable over (a, b) with $g'(x) \neq 0$ for all $x \in (a, b)$, where $-\infty \leqslant a < b \leqslant +\infty$. If either

(i) $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = 0$; or

(ii) $\lim_{x \to a} |g(x)| = +\infty$,

and

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

then

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

Proof. The entire proof is rather tedious because we have to many cases.

We first consider the case in which $-\infty \leqslant A < +\infty$. Choose $q \in \mathbf{R}$ such that $A < q$, and choose $r \in \mathbf{R}$ such that $A < r < q$.

1. $\frac{0}{0}$ or $\frac{\infty}{\infty}$ 2. a is normal or $a = -\infty$ 3. A is normal or $A = \pm\infty$

We'll only prove the most basic one here: 0/0, a and A are normal This is the case which will be required for Taylor series

First we define f(a)=g(a)=0, so that f and g are continuous at $x = a$

Now let $x \in (a, b)$, then f and g are continuous on $[a, x]$ and differentiable in (a, x) : Thus by Cauchy's Mean Value Theorem, there exists $\xi \in (a, x)$ such that

$$\frac{f'(\xi)}{g'(\xi)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x)}{g(x)}$$

For each x, we pick ξ which satisfies the above, so that ξ may be seen as a function of x satisfying $a < \xi(x) < x$

Then by squeezing we have $\lim_{x \to a^+} \xi(x) = a$.

Since $\frac{f'}{g'}$ is continuous near a, the theorem regarding the limit of composite functions give

$$\lim_{x \to a^+} \frac{f(x)}{g(x)} = \lim_{x \to a^+} \frac{f'(\xi)}{g'(\xi)} = \lim_{x \to a^+} \left(\frac{f'}{g'}\right)(\xi(x)) = A$$

Now the same reasoning can be used for b where we will use lim(x→b-) to replace all the $\lim_{x \to a^+}$, and ξ will be a function which maps to (x, b). $\square$

§12.5 Taylor Expansion

Theorem 12.15 (Taylor's Theorem)**.** Suppose $f : [a,b] \to \mathbf{R}$, $f^{(n-1)}$ is continuous on $[a,b]$, $f^{(n)}(t)$ exists for every $t \in (a,b)$. Let α and β be distinct points of $[a,b]$, and define

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(t-\alpha)^k.$$

Then there exists $x \in [\alpha, \beta]$ such that

$$f(\beta) = P(\beta) + \frac{f^{(n)}(x)}{n!}(\beta - \alpha)^n.$$

13 Riemann–Stieltjes Integral

§13.1 Definition of Riemann–Stieltjes Integral

A **partition** P of a closed interval $[a,b] \subset \mathbf{R}$ is a finite set of points $x_0, x_1, \ldots, x_n$ where

$$a = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_{n-1} \leqslant x_n = b.$$

Let $f : [a,b] \to \mathbf{R}$ be bounded, and α be an increasing function over $[a,b]$. Denote by

$$M_i = \sup_{[x_{i-1},x_i]} f(x),$$
$$m_i = \inf_{[x_{i-1},x_i]} f(x),$$

and by

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}).$$

The **upper sum** of f with respect to the partition P and α is

$$U(f,\alpha;P) = \sum_{i=1}^{n} M_i \Delta\alpha_i$$

and the **lower sum** of f with respect to the partition P and α is

$$L(f,\alpha;P) = \sum_{i=1}^{n} m_i \Delta\alpha_i.$$

Define the upper Riemann–Stieltjes integral as

$$\overline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) := \inf_P U(f,\alpha;P)$$

and the lower Riemann–Stieltjes integral as

$$\underline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) := \sup_P L(f,\alpha;P).$$

It is easy to see from definition that

$$\underline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) \leqslant \overline{\int_a^b} f(x)\,\mathrm{d}\alpha(x)\,.$$

Definition 13.1 (Riemann–Stieltjes integrability)**.** A function f is **Riemann–Stieltjes integrable** with respect to α over $[a, b]$, if

$$\underline{\int_a^b} f(x)\,\mathrm{d}\alpha(x) = \overline{\int_a^b} f(x)\,\mathrm{d}\alpha(x)\,.$$

Notation. $\int_a^b f(x)\,\mathrm{d}\alpha(x)$ denotes the common value, which is called the Riemann–Stieltjes of f with respect to α over $[a, b]$.

Notation. $\mathcal{R}_\alpha[a, b]$ denotes the set of Riemann–Stieljes integrable functions with respect to α over $[a, b]$.

In particular, when $\alpha(x) = x$, we call the corresponding Riemann–Stieljes integration the *Riemann integration*, and use $\mathcal{R}[a, b]$ to denote the set of Riemann integrable functions.

Definition 13.2 (Refinement)**.** The partition P' is a **refinement** of P if $P' \supset P$. Given two partitions P_1 and P_2, we say that P' is their **common refinement** if $P' = P_1 \cup P_2$.

Intuitively, a refinement will give a better estimation than the original partition, so the upper and lower sums of a refinement should be more restrictive. We will now show this.

Lemma 13.3. If P' is a refinement of P, then

(i) $L(f, \alpha; P) \leqslant L(f, \alpha; P')$

(ii) $U(f, \alpha; P') \leqslant U(f, \alpha; P)$

Proof.

(i) Suppose first that P' contains just one point more than P Let this extra point be x', and suppose $x_{i-1} < x' < x_i$ for some i, where $x_{i-1}, x_i \in P$. Put

$$w_1 = \inf_{x \in [x_{i-1}, x']} f(x)$$

and

$$w_2 = \inf_{x \in [x', x_i]} f(x).$$

Let, as before,

$$m_i = \inf_{x \in [x_{i-1}, x_i]} f(x).$$

Clearly $w_1 \geqslant m_i$ and $w_2 \geqslant m_i$. Hence

$$\begin{aligned}
&L(f, \alpha; P') - L(f, \alpha; P) \\
&= w_1[\alpha(x') - \alpha(x_{i-1})] + w_2[\alpha(x_i) - \alpha(x')] - m_i[\alpha(x_i) - \alpha(x_{i-1})] \\
&= (w_1 - m_i)[\alpha(x') - \alpha(x_{i-1})] + (w_2 - m_i)[\alpha(x_i) - \alpha(x')] \geqslant 0.
\end{aligned}$$

If P' contains k more points than P, we repeat this reasoning k times.

(ii) Analogous to the proof of (i).

□

One would expect the lower RS integral to be less than or equal to the upper RS integral. We now show this.

Lemma 13.4.
$$\underline{\int_a^b} f \,\mathrm{d}\alpha \leqslant \overline{\int_a^b} f \,\mathrm{d}\alpha\,.$$

Proof. Let P' be the common refinement of partitions P_1 and P_2. By Lemma 13.3,

$$L(f,\alpha;P_1) \leqslant L(f,\alpha;P') \leqslant U(f,\alpha;P') \leqslant U(f,\alpha;P_2)$$

and so
$$L(f,\alpha;P_1) \leqslant U(f,\alpha;P_2).$$

Fix P_2 and take sup over all P_1 gives

$$\underline{\int} f \,\mathrm{d}\alpha \leqslant U(f,\alpha;P_2).$$

Then take inf over all P_2, which gives the desired result. □

Now we discuss integrability conditions for f.

Theorem 13.5. $f \in \mathcal{R}_\alpha[a,b]$ if and only if

$$\forall \varepsilon > 0, \quad \exists P, \quad U(f,\alpha;P) - L(f,\alpha;P) < \varepsilon.$$

Proof.

$\boxed{\implies}$ Suppose $f \in \mathcal{R}_\alpha[a,b]$. Let $\varepsilon > 0$ be given. Then there exists partitions P_1 and P_2 such that

$\boxed{\impliedby}$ For every P, from Lemma 13.4 we have

$$L(f,\alpha;P) \leqslant \underline{\int} f \,\mathrm{d}\alpha \leqslant \overline{\int} f \,\mathrm{d}\alpha \leqslant U(f,\alpha;P).$$

□

Example (Dirichlet function)

The Dirichlet function is defined over $\mathbf{R}$ by

$$f(x) = \begin{cases} 1 & x \in \mathbf{Q} \\ 0 & x \in \mathbf{R} \setminus \mathbf{Q} \end{cases}$$

We try to calculate the two on the interval $[0,1]$.

The Dirichlet function is pathological because for each subinterval $[x_{i-1},x_i]$, the supremum is always 1 and the infimum is always 0.

So no matter what partition we use, $U(f,P)$ is always 1 whereas $L(f,P)$ is always 0. This means that $U(f) = 1$ and $L(f) = 0$, so there are two different values for "the integral of f".

This is like the case where we try to find the limit of the Dirichlet function where x is approaching any given real number r, there exists two sequences approaching r whose image approaches two different values.

Example (Heaviside step function)

The Heaviside step function H is a real-valued function defined by

$$H(x) = \begin{cases} 0 & x \leqslant 0 \\ 1 & x > 0 \end{cases}$$

Proposition. f bounded on $[a,b]$, f continuous at $s \in (a,b)$. Let $\alpha(x) = H(x-s)$, then

$$\int_a^b f \,\mathrm{d}\alpha = f(s).$$

Proposition. Suppose $c_n \geqslant 0$ for $n = 1, 2, \dots$, $\sum c_n$ converges, (s_n) is a sequence of distinct points in (a,b), and

$$\alpha(x) = \sum_{n=1}^{\infty} c_n I(x - s_n).$$

Let f be continuous on $[a,b]$. Then

$$\int_a^b f \,\mathrm{d}\alpha = \sum_{n=1}^{\infty} c_n f(s_n).$$

Proposition 13.6.

(i) For all $\varepsilon > 0$, if there exists P such that $U(f,\alpha;P) - L(f,\alpha;P) < \varepsilon$, then $U(f,\alpha;P) - L(f,\alpha;P') < \varepsilon$ where P' is a refinement of P.

(ii)

(iii)

Proof. □

Lemma 13.7 (Continuity implies integrability)**.** If f is continuous on $[a,b]$, then $f \in \mathcal{R}_\alpha[a,b]$.

Proof. Let $\varepsilon > 0$ be given. Choose $\eta > 0$ such that

$$[\alpha(b) - \alpha(a)]\eta < \varepsilon.$$

Since f is uniformly continuous on $[a,b]$ (Theorem 4.19), there exists $\delta > 0$ such that

$$|f(x) - f(t)| < \eta$$

if $x \in [a,b]$, $t \in [a,b]$, $|x-t| < \delta$.

If P is any partition of $[a,b]$ such that $\Delta x_i < \delta$ for all i, then (16) implies that

$$M_i - m_i \leqslant \eta$$

and therefore

$$\begin{aligned} U(f,\alpha;P) - L(f,\alpha;P) &= \sum_{i-1}^{n} (M_i - m_i)\Delta\alpha_i \\ &\leqslant \eta \sum_{i=1}^{n} \Delta\alpha_i \\ &= \eta[\alpha(b) - \alpha(a)] \\ &< \varepsilon. \end{aligned}$$

By Theorem 6.6, $f \in \mathcal{R}_\alpha[a,b]$. □

Proposition 13.8. If f is monotonic on $[a,b]$, and if α is continuous on $[a,b]$, then $f \in \mathcal{R}_\alpha[a,b]$.

Proposition 13.9. Suppose f is bounded on $[a,b]$, f has only finitely many points of discontinuity on $[a,b]$, and α is continuous at every point at which f is discontinuous. Then $f \in \mathcal{R}_\alpha[a,b]$.

Proposition 13.10. $f \in \mathcal{R}_\alpha[a,b]$, $m \leqslant f \leqslant M$, and ϕ is uniformly continuous on $[m,M]$. Then

$$\phi \circ f \in \mathcal{R}_\alpha[a,b].$$

Proof. Choose $\varepsilon > 0$. Since ϕ is uniformly continuous on $[m,M]$, there exists $\delta > 0$ such that $\delta < \varepsilon$ and $|\phi(s) - \phi(t)| < \varepsilon$ if $|s-t| <\leqslant \delta$ and $s,t \in [m,M]$.

Since $f \in \mathcal{R}_\alpha[a,b]$, there exists a partition $P = \{x_0, x_1, \dots, x_n\}$ of $[a,b]$ such that

$$U(f,\alpha;P) - L(f,\alpha;P) < \delta^2.$$

□

§13.2 Properties of the Integral

Theorem 13.11.

(i) If $f_1, f_2 \in \mathcal{R}_\alpha[a,b]$, then

$$f_1 + f_2 \in \mathcal{R}_\alpha[a,b];$$

$cf \in \mathcal{R}_\alpha[a,b]$ for every $c \in \mathbf{R}$, and

$$\int_a^b (f_1 + f_2)\,\mathrm{d}\alpha = \int_a^b f_1\,\mathrm{d}\alpha + \int_a^b f_2\,\mathrm{d}\alpha\,,$$

$$\int_a^b (cf)\,\mathrm{d}\alpha = c\int_a^b f\,\mathrm{d}\alpha\,.$$

(ii) If $f_1, f_2 \in \mathcal{R}_\alpha[a,b]$ and $f_1 \leqslant f_2$, then

$$\int_a^b f_1 \, \mathrm{d}\alpha \leqslant \int_a^b f_2 \, \mathrm{d}\alpha \,.$$

(iii) If $f \in \mathcal{R}_\alpha[a,b]$ and $c \in [a,b]$, then $f \in \mathcal{R}_\alpha[a,c]$ and $f \in \mathcal{R}_\alpha[c,b]$, and

$$\int_a^b f \, \mathrm{d}\alpha = \int_a^c \mathrm{d}\alpha + \int_c^b \mathrm{d}\alpha \,.$$

(iv) If $f \in \mathcal{R}_\alpha[a,b]$ and $|f| \leqslant M$, then

$$\left| \int_a^b f \, \mathrm{d}\alpha \right| \leqslant M \left[\alpha(b) - \alpha(a) \right] .$$

(v) If $f \in R_{\alpha_1}[a,b]$ and $f \in R_{\alpha_2}[a,b]$, then $f \in R_{\alpha_1+\alpha_2}[a,b]$ and

$$\int_a^b f \, \mathrm{d}(\alpha_1 + \alpha_2) = \int_a^b f \, \mathrm{d}\alpha_1 + \int_a^b f \, \mathrm{d}\alpha_2 \,;$$

if $f \in \mathcal{R}_\alpha[a,b]$ and c is a positive constant, then $f \in R_{c\alpha}[a,b]$ and

$$\int_a^b f \, \mathrm{d}(c\alpha) = c \int_a^b f \, \mathrm{d}\alpha \,.$$

(vi) If $f \in \mathcal{R}_\alpha[a,b]$ and $g \in \mathcal{R}_\alpha[a,b]$, then $fg \in \mathcal{R}_\alpha[a,b]$.

Proof.

(i) If $f = f_1 + f_2$ and P is any partition of $[a,b]$, we have

$$\begin{aligned} L(f_1, \alpha; P) + L(f_2, \alpha; P) &\leqslant L(f, \alpha; P) \\ &\leqslant U(f, \alpha; P) \\ &\leqslant U(f_1, \alpha; P) + U(f_2, \alpha; P). \end{aligned}$$

If $f_1 \in \mathcal{R}_\alpha[a,b]$ and $f_2 \in \mathcal{R}_\alpha[a,b]$, let $\varepsilon > 0$ be given. There are partitions P_1 and P_2 such that

(ii)

(iii)

(iv)

(v)

(vi)

□

Theorem 13.12 (Triangle inequality)**.** $f \in \mathcal{R}_\alpha[a,b]$, then $|f| \in \mathcal{R}_\alpha[a,b]$,

$$\left| \int_a^b f \, \mathrm{d}\alpha \right| \leqslant \int_a^b |f| \, \mathrm{d}\alpha \,.$$

Proof. □

6.14 6.15 Heaviside step function

6.16 corollary for intinite sum, need $\sum c_n$ to converge (23) comparison test

Proposition 13.13 (Integration by substitution)**.** Assume α increases monotonically, $\alpha' \in R[a,b]$. Let f be a bounded real function on $[a,b]$, then

$$f \in \mathcal{R}_\alpha[a,b] \iff f\alpha' \in R[a,b].$$

Proposition 13.14 (Change of variables)**.** Suppose $\phi : [A,B] \to [a,b]$ is a strictly increasing continuous function. Suppose α is monotonically increasing on $[a,b]$, $f \in \mathcal{R}_\alpha[a,b]$. Define β and g on $[A,B]$ by

$$\beta(y) = \alpha(\phi(y)), \quad g(y) = f(\phi(y)).$$

Then $g \in R(\beta)$ and

$$\int_A^B g \, \mathrm{d}\beta = \int_a^b f \, \mathrm{d}\alpha \, .$$

§13.3 Integration and Differentiation

We shall show that integration and differentiation are, in a certain sense, inverse operations.

Lemma 13.15. $f \in \mathcal{R}_\alpha[a,b]$. For $x \in [a,b]$, put

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

Then F is continuous on $[a,b]$; furthermore, if f is continuous at $x_0 \in [a,b]$, then F is differentiable at x_0, and

$$F'(x_0) = f(x_0).$$

Theorem 13.16 (Fundamental Theorem of Calculus)**.** $f \in \mathcal{R}_\alpha[a,b]$, there is a differentiable function F on $[a,b]$ such that $F' = f$, then

$$\int_a^b f(x) \, \mathrm{d}x = F(b) - F(a). \tag{13.1}$$

Theorem 13.17 (Integration by parts)**.** Suppose F and G are differentiable functions on $[a,b]$, $F' = f \in R$, $G' = g \in R$. Then

$$\int_a^b F(x)g(x) \, \mathrm{d}x = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x) \, \mathrm{d}x \, . \tag{13.2}$$

14 Sequence and Series of Functions

§14.1 Uniform Convergence

Definition 14.1 (Pointwise convergence)**.** Suppose (f_n) is a sequence of functions defined on a set E, and suppose that $(f_n(x))$ converges for every $x \in E$. We can then define a function f by

$$f(x) = \lim_{n\to\infty} f_n(x) \quad (\forall x \in E)$$

We say that (f_n) **converges pointwise** to f on E, denoted by $f_n \to f$, if

$$\forall \varepsilon > 0, \quad \forall x \in E, \quad \exists N \in \mathbf{N}, \quad \forall n > N, \quad |f_n(x) - f(x)| < \varepsilon.$$

f is called the **limit** or limit function of (f_n).

Similarly, if $\sum f_n(x)$ converges for every $x \in E$, and if we define

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad (\forall x \in E)$$

the function f is called the **sum of the series** $\sum f_n$.

Most properties are not preserved by pointwise continuity; that is, f does not inherit most properties of f_n.

Example (f_n continuous, f discontinuous)

Let $f_n(x) = x^n$ for $x \in [0, 1]$. Then

$$f(x) = \lim_{n\to\infty} f_n(x) = \begin{cases} 0 & \text{if } x \in (0, 1] \\ 1 & \text{if } x = 1 \end{cases}$$

and so the limit function $f(x)$ is discontinuous.

Example (f_n integrable, f not integrable)

Recall that the Dirichlet function

$$D(x) = \begin{cases} 1 & \text{if } x \in \mathbf{Q} \\ 0 & \text{if } x \in \mathbf{R} \setminus \mathbf{Q} \end{cases}$$

is not integrable.

Proof. Consider the interval $[0,1]$. We partition $P: 0 = x_0 < x_1 < \cdots < x_n = 1$. The sum is given by $\sum_{i=1}^{n} D(t_i)\Delta x_i$. Then

$$M_i = \max_{t \in [x_{i-1}, x_i]} D(t) = 1 \implies U(D;P) = 1 \quad \forall P$$

and

$$m_i = \min_{t \in [x_{i-1}, x_i]} D(t) = 0 \implies L(D;P) = 0 \quad \forall P.$$

Hence

$$\overline{\int_0^1} D(x)\,\mathrm{d}x = 1, \qquad \underline{\int_0^1} D(x)\,\mathrm{d}x = 0$$

so $\overline{\int_0^1} D(x)\,\mathrm{d}x \neq \underline{\int_0^1} D(x)\,\mathrm{d}x$, and thus $D(x)$ is not integrable. □

We define a sequence of functions as follows:

$$D_n(x) = \begin{cases} 1 & \text{if } x = \frac{p}{q}, p \in \mathbf{Z}, q \in \mathbf{Z} \setminus \{0\}, |q| \leqslant n \\ 0 & \text{if otherwise} \end{cases}$$

Definition 14.2 (Uniform convergence)**.** We say that (f_n) **uniformly converges** to f over E, denoted by $f_n \rightrightarrows f$, if

$$\forall \varepsilon > 0, \quad \exists N \in \mathbf{N}, \quad \forall x \in E, \quad \forall n > N, \quad |f_n(x) - f(x)| < \varepsilon.$$

For series, we say that the series $\sum f_n(x)$ converges uniformly on E if the sequence of partial sums (S_n) defined by

$$S_n(x) = \sum_{i=1}^{n} f_i(x)$$

converges uniformly on E.

Uniform convergence is stronger than pointwise convergence, since N is uniform (or "fixed") for all $x \in E$; for pointwise convergence, the choice of N is determined by x.

The Cauchy criterion for uniform convergence is as follows.

Lemma 14.3 (Cauchy criterion)**.** $(f_n) \rightrightarrows f$ if and only if

$$\forall \varepsilon > 0, \quad \exists N \in \mathbf{N}, \quad \forall x \in E, \quad \forall n, m \geqslant N, \quad |f_n(x) - f_m(x)| \leqslant \varepsilon.$$

Proof.

$\boxed{\implies}$ Suppose $f_n \rightrightarrows f$ on E. Let $\varepsilon > 0$ be given. Then there exists $N \in \mathbf{N}$ such that for all $x \in E$, for all $n > N$,

$$|f_n(x) - f(x)| < \frac{\varepsilon}{2}.$$

Then for all $n, m > N$,

$$\begin{aligned} |f_n(x) - f_m(x)| &= \Big|(f_n(x) - f(x)) - (f_m(x) - f(x))\Big| \\ &\leqslant |f_n(x) - f(x)| + |f_m(x) - f(x)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

by triangle inequality.

$\boxed{\Longleftarrow}$ Conversely, suppose the Cauchy condition holds. By Theorem 3.11, the sequence $(f_n(x))$ converges, for every x, to a limit which we may call $f(x)$. Thus $(f_n) \to f$ on E. We have to prove that the convergence is uniform.

Let $\varepsilon > 0$ be given. Choose $N \in \mathbf{N}$ such that (13) holds. Fix n, and let $m \to \infty$ in (13). Since $f_m(x) \to f(x)$ as $m \to \infty$, thus for all $n \geqslant N$ and for all $x \in E$,

$$|f_n(x) - f(x)| \leqslant \varepsilon,$$

which completes the proof. □

The following criterion is sometimes useful.

Proposition 14.4. Suppose $f_n \to f$ on E. Let

$$M_n = \sup_{x \in E} |f_n(x) - f(x)|.$$

Then $f_n \rightrightarrows f$ on E if and only if $M_n \to 0$ as $n \to \infty$.

For series, there is a very convenient test for uniform convergence, due to Weierstrass.

Lemma 14.5 (Weierstrass M-test)**.** Suppose (f_n) is a sequence of functions defined on E, and suppose there exists $(M_n) \in \mathbf{R}^+$ such that $|f_n(x)| \leqslant M_n$ for all $n \geqslant 1$ and for all $x \in E$.

Then $\sum f_n$ converges uniformly on E if $\sum M_n$ converges.

§14.2 Uniform Convergence and Continuity

We now consider properties preserved by uniform convergence.

Proposition 14.6. Suppose $f_n \rightrightarrows f$ on E. Let $x \in E$ be a limit point, let

$$\lim_{t \to x} f_n(t) = A_n.$$

Then (A_n) converges, and $\lim\limits_{t \to x} f(t) = \lim\limits_{n \to \infty} A_n$.

Proposition 14.7. Let (f_n) be a sequence of continuous functions on E, $f_n \rightrightarrows f$. Then f is continuous in E.

Definition 14.8 (Supremum norm)**.** If X is a metric space, we denote the set of all complex-valued, continuous, bounded functions with domain X by $C(X)$.

If $f \in C(X)$, we define

$$\|f\| := \sup_{x \in X} |f(x)|,$$

known as the **suprenum norm** of f.

Lemma 14.9. $\|f\|$ gives a norm on $C(X)$.

Proof. Check that $\|f\|$ satisfies the conditions for a norm:

(i)

□

Proposition 14.10. $(C(X), \|\cdot\|)$ is a metric space.

§14.3 Uniform Convergence and Integration

Theorem 14.11. Assume (f_n) is a sequence of functions defined over $[a,b]$ and each $f_n \in R_\alpha[a,b]$. If $f_n \to f$, then $f \in R_\alpha[a,b]$, and

$$\lim_{n\to\infty} \int_a^b f_n \,\mathrm{d}\alpha = \int_a^b f \,\mathrm{d}\alpha \,.$$

Proof. Define □

Corollary 14.12. Assume $a_n \in R_\alpha[a,b]$ and

$$f(x) := \sum_{n=0}^{\infty} a_n(x)$$

converges uniformly. Then it follows

$$\int_a^b f \,\mathrm{d}\alpha = \sum_{n=0}^{\infty} a_n \,\mathrm{d}\alpha \,.$$

Proof. Consider the sequence of partial sums

$$f_n(x) := \sum_{k=0}^{n} a_k(x), \quad n = 0, 1, \dots$$

It follows $f_n \in R_\alpha[a,b]$ and $f_n \rightrightarrows f$. Apply above theorem to (f_n) and the conclusion follows. □

§14.4 Uniform Convergence and Differentiation

Theorem 14.13. (f_n) differentiable on $[a,b]$, $\exists x_0 \in [a,b]$ s.t. $f_n(x_0) \to y_0 = f(x_0)$ and $f_n' \rightrightarrows f'$. Then $f_n \rightrightarrows f$ on $[a,b]$, and f is differentiable, $f'(x) = \lim_{n\to\infty} f_n'(x)$ for any $x \in [a,b]$.

Proof. $f_n(x_0) \to y_0$ thus □

§14.5 Stone–Weierstrass Approximation Theorem

Theorem 14.14 (Weierstrass approximation theorem)**.** If f is a continuous complex function on $[a,b]$, there exists a sequence of polynomials P_n such that $P_n \rightrightarrows f$ on $[a,b]$.

If f is real, then P_n may be taken real.

15 Some Special Functions

§15.1 Power Series

Definition 15.1 (Analytic function)**.** An **analytic function** is a function that can be represented by a power series, i.e., functions of the form

$$f(x)=\sum_{n=0}^{\infty}c_nx^n$$

or, more generally,

$$f(x)=\sum_{n=0}^{\infty}c_n(x-a)^n.$$

As a matter of convenience, we shall often take $a=0$ without any loss of generality.

We shall restrict ourselves to real values of x. The **radius of convergence** is the maximum R such that $f(x)$ converges in $(-R,R)$. If $f(x)$ converges for all $x\in(-R,R)$, for some $R>0$, we say that f is expanded in a power series about the point $x=0$.

Proposition 15.2. Suppose the series $\sum_{n=0}^{\infty}c_nx^n$ converges for $x\in(-R,R)$. Then

(i) $\sum_{n=0}^{\infty}c_nx^n$ converges uniformly on $[-R+\varepsilon,R-\varepsilon]$ for all $\varepsilon>0$;

(ii) $f(x)$ is continuous and differentiable on $(-R,R)$, and

$$f'(x)=\sum_{n=1}^{\infty}nc_nx^{n-1}.$$

Proof.

(i) Let $\varepsilon>0$ be given. For $|x|\leqslant R-\varepsilon$, we have

$$|c_nx^n|\leqslant|c_n(R-\varepsilon)^n|$$

and since

$$\sum c_n(R-\varepsilon)^n$$

converges absolutely (every power series converges absolutely in the interior of its internal of convergence, by the root test), Theorem 7.10 show that $\sum_{n=0}^{\infty}c_nx^n$ uniformly converges on $[-R+\varepsilon,R-\varepsilon]$.

(ii)

□

Corollary 15.3. f has derivatives of all orders in $(-R, R)$, which are given by

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)c_n x^{n-k}.$$

In particular,

$$f^{(k)}(0) = k!c_k, \quad k = 0, 1, 2, \ldots$$

(Here $f^{(0)}$ means f, and $f^{(k)}$ is the k-th derivative of f, for $k = 1, 2, 3, \ldots$)

Proof. Apply theorem successively to $f, f', f'', \ldots$. Put $x = 0$. □

Proposition 15.4. Suppose $\sum c_n$ converges. Put

$$f(x) = \sum_{n=0}^{\infty} c_n x^n$$

for $x \in (-R, R)$

§15.2 Exponential and Logarithmic Functions

Definition 15.5 (Exponential function)**.**

$$\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}. \tag{15.1}$$

Proposition 15.6. $\exp(z)$ converges for every $z \in \mathbf{C}$.

Proof. Ratio test. □

Proposition 15.7. For $z, w \in \mathbf{C}$,

$$\exp(z + w) = \exp(z) + \exp(w).$$

Corollary 15.8. For $z \in \mathbf{C}$,

$$\exp(z)\exp(-z) = 1.$$

Proof. Take $z = z$, $w = -z$ in the previous result. □

Proposition 15.9. exp is strictly increasing in $\mathbf{R}$.

Proposition 15.10. For $z \in \mathbf{C}$,

$$\exp'(z) = \exp(z)$$

Further,

$$\exp'(z) = \lim_{h\to 0} \frac{\exp(z+h) - \exp(z)}{h} = \lim_{h\to 0} \frac{\exp(z+h) - 1}{h} \exp(z).$$

Let $\exp(1) = e$. So $\exp(n) = \exp(1 + \cdots + 1) = \exp(1) \cdots \exp(1) = e^n$. This holds for any $n \in \mathbf{Q}$.

§15.3 Trigonometric Functions

Define

$$C(x) = \frac{\exp(ix) + \exp(-ix)}{2}$$
$$S(x) = \frac{\exp(ix) - \exp(-ix)}{2i}$$

Our goal here is to show that $C(x)$ and $S(x)$ coincide with the functions $\cos x$ and $\sin x$, whose definition is usually based on geometric considerations.

Proposition 15.11 (Euler's identity)**.**

$$\exp(ix) = C(x) + iS(x).$$

Proof. □

From definition, it is easy to see that $C(0) = 1$, $S(0) = 0$, and

$$C'(x) = S(x)$$
$$S'(x) = C(x)$$

Proposition 15.12.

(i) exp is periodic, with period $2\pi i$.

(ii) C and S are periodic, with period 2π.

(iii) If $0 < t < 2\pi$, then $\exp(it) \neq 1$.

(iv) If $z \in \mathbf{C}$, $|z| = 1$, there exists a unique $t \in [0, 2\pi)$ such that $\exp(it) = z$.

§15.4 Algebraic Completeness of the Complex Field

We now prove that the complex field is *algebraically complete*; that is, every non-constant polynomial with complex coefficients has a complex root.

Theorem 15.13 (Fundamental Theorem of Algebra)**.** Suppose $a_0, \dots, a_n$ are complex numbers, $n \geqslant 1$, $a_n \neq 0$,

$$P(z) = \sum_{k=0}^{n} a_k z^k.$$

Then $P(z) = 0$ for some complex number z.

Proof. Without loss of generality, assume $a_n = 1$. Let $\mu = \inf |P(z)|$. □

§15.5 Fourier Series

Definition 15.14 (Trigonometric polynomial)**.** A **trigonometric polynomial** is a finite sum of the form

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \quad (x \in \mathbf{R})$$

where $a_0, \dots, a_N, b_1, \dots, b_N \in \mathbf{C}$.

We can write the above in the form

$$f(x) = \sum_{n=-N}^{N} c_n e^{inx}.$$

It is clear that every trigonometric polynomial is periodic, with period 2π.

For non-zero integer n, e^{inx} is the derivative of $\frac{1}{in} e^{inx}$, which also has period 2π. Hence

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{inx} \,\mathrm{d}x = \begin{cases} 1 & (n = 0) \\ 0 & (n = \pm 1, \pm 2, \dots) \end{cases}$$

Definition 15.15 (Fourier coefficients)**.** If f is an integrable function on $[-\pi, \pi]$, the numbers c_m defined by

$$c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{inx} \,\mathrm{d}x$$

for all integers m are called the **Fourier coefficients** of f.

Definition 15.16 (Fourier series)**.** The series

$$\sum_{n=-\infty}^{\infty} c_n e^{inx}$$

formed with the Fourier coefficients is called the **Fourier series** of f.

Definition 15.17 (Orthogonal system of functions)**.** Let (ϕ_n) be a sequence of complex functionns on $[a, b]$ such that

$$\int_a^b \phi_n(x) \overline{\phi_m(x)} \,\mathrm{d}x = 0 \quad (n \neq m)$$

Then (ϕ_n) is said to be an **orthogonal system of functions** on $[a, b]$. If in addition

$$\int_a^b |\phi_b(x)|^2 \,\mathrm{d}x = 1$$

for all n, (ϕ_n) is said to be **orthonormal**.

§15.6 Gamma Function

Definition 15.18 (Gamma function)**.** For $0 < x < \infty$, the **Gamma function** is defined as

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} \,\mathrm{d}t\,. \tag{15.2}$$

The integral converges for these x. (When $x < 1$, both 0 and ∞ have to be looked at.)

Lemma 15.19.

(i) The functional equation

$$\Gamma(x + 1) = x\Gamma(x)$$

holds for $0 < x < \infty$.

(ii) $\Gamma(n + 1) = n!$ for $n = 1, 2, 3, \ldots$

(iii) $\log \Gamma$ is convex on $(0, \infty)$.

Proof.

(i) Integrate by parts.

(ii) Since $\Gamma(1) = 1$, (1) implies (2) by induction.

(iii)

□

In fact, these three properties characterise Γ completely.

Lemma 15.20 (Characteristic properties of Γ)**.** If f is a positive function on $(0, \infty)$ such that

(i) $f(x + 1) = xf(x)$,

(ii) $f(1) = 1$,

(iii) $\log f$ is convex,

then $f(x) = \Gamma(x)$.

Proof. □

Definition 15.21 (Beta function)**.** For $x > 0$ and $y > 0$, the **beta function** is defined as

$$B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t\,.$$

Lemma 15.22.

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Proof. Let $f(x) = \dfrac{\Gamma(x+y)}{\Gamma(y)}B(x,y)$. We want to prove that $f(x) = \Gamma(x)$, using Lemma 15.20.

(i)

$$B(x+1,y) = \int_0^1 t^{x}(1-t)^{y-1}\,\mathrm{d}t\,.$$

Integrating by parts gives

$$\begin{aligned} B(x+1,y) &= \underbrace{\left[t^x \cdot \frac{(1-t)^y}{y}(-1)\right]_0^1}_{0} + \int_0^1 xt^{x-1}\frac{(1-t)^y}{y}\,\mathrm{d}t \\ &= \frac{x}{y}\int_0^1 t^{x-1}(1-t)^{y-1}(1-t)\,\mathrm{d}t \\ &= \frac{x}{y}\left(\int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t - \int_0^1 t^{x}(1-t)^{y-1}\,\mathrm{d}t\right) \\ &= \frac{x}{y}\left(B(x,y) - B(x+1,y)\right) \end{aligned}$$

which gives $B(x+1,y) = \dfrac{x}{x+y}B(x,y)$. Thus

$$\begin{aligned} f(x+1) &= \frac{\Gamma(x+1+y)}{\Gamma(y)}B(x+1,y) \\ &= \frac{(x+y)B(x+y)}{\Gamma(y)} \cdot \frac{x}{x+y}B(x,y) \\ &= xf(x). \end{aligned}$$

(ii)

$$B(1,y) = \int_0^1 (1-t)^{y-1}\,\mathrm{d}t = \left[-\frac{(1-t)^y}{y}\right]_0^1 = \frac{1}{y}$$

and thus

$$f(1) = \frac{\Gamma(1+y)}{\Gamma(y)}B(1,y) = \frac{y\Gamma(y)}{\Gamma(y)}\frac{1}{y} = 1.$$

(iii) We now show that $\log B(x,y)$ is convex, so that

$$\log f(x) = \underbrace{\log\Gamma(x+y)}_{\text{convex}} + \log B(x,y) - \underbrace{\log\Gamma(y)}_{\text{constant}}$$

is convex with respect to x.

$$B(x_1,y)^{\frac{1}{p}}B(x_2,y)^{\frac{1}{q}} = \left(\int_0^1 t^{x_1-1}(1-t)^{y-1}\,\mathrm{d}t\right)^{\frac{1}{p}}\left(\int_0^1 t^{x_2-1}(1-t)^{y-1}\,\mathrm{d}t\right)^{\frac{1}{q}}$$

By Hölder's inequality,

$$\begin{aligned} B(x_1,y)^{\frac{1}{p}}B(x_2,y)^{\frac{1}{q}} &= \int_0^1 \left[t^{x_1-1}(1-t)^{y-1}\right]^{\frac{1}{p}}\left[t^{x_2-1}(1-t)^{y-1}\right]^{\frac{1}{q}}\mathrm{d}t \\ &= \int_0^1 t^{\frac{x_1}{p}+\frac{x_2}{q}-1}(1-t)^{y-1}\,\mathrm{d}t \\ &= B\left(\frac{x_1}{p}+\frac{x_2}{q},y\right). \end{aligned}$$

Taking log on both sides gives

$$\log B(x,y)^{\frac{1}{p}}B(x_2,y)^{\frac{1}{q}} \geqslant \log B\left(\frac{x_1}{p}+\frac{x_2}{q},y\right)$$

or

$$\frac{1}{p}\log B(x,y)+\frac{1}{q}\log B(x_2,y) \geqslant \log B\left(\frac{x_1}{p}+\frac{x_2}{q},y\right).$$

Hence $\log B(x,y)$ is convex, so $\log f(x)$ is convex.

Therefore $f(x)=\Gamma(x)$ which implies $B(x,y)=\dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. □

An alternative form of Γ is as follows:

$$\Gamma(x) = 2\int_0^{+\infty} t^{2x-1}e^{-t^2}\,\mathrm{d}t\,.$$

Using this form of Γ, we present an alternative proof.

Proof.

$$\begin{aligned} \Gamma(x)\Gamma(y) &= \left(2\int_0^{+\infty} t^{2x-1}e^{-t^2}\,\mathrm{d}t\right)\left(2\int_0^{+\infty} s^{2y-1}e^{-s^2}\,\mathrm{d}s\right) \\ &= 4\iint_{[0,+\infty)\times[0,+\infty)} t^{2x-1}s^{2y-1}e^{-\left(t^2+s^2\right)}\,\mathrm{d}t\,\mathrm{d}s \end{aligned}$$

Using polar coordinates transformation, let $t=r\cos\theta$, $s=r\sin\theta$. Then $\mathrm{d}t\,\mathrm{d}s = r\,\mathrm{d}r\,\mathrm{d}\theta$. Thus

$$\begin{aligned} \Gamma(x)\Gamma(y) &= 4\int_0^{\frac{\pi}{2}}\left[\int_0^{+\infty} r^{2x-1}\cos^{2x-1}\theta\cdot r^{2y-1}\sin^{2y-1}\theta\cdot e^{-r^2}\cdot r\,\mathrm{d}r\right]\mathrm{d}\theta \\ &= \underbrace{2\int_0^{\frac{\pi}{2}}\cos^{2x-1}\theta\sin^{2y-1}\theta\,\mathrm{d}\theta}_{B(x,y)}\cdot\underbrace{2\int_0^{+\infty} r^{2(x+y)-1}e^{-r^2}\,\mathrm{d}r}_{\Gamma(x+y)} \end{aligned}$$

since

$$\begin{aligned} B(x,y) &= \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t \qquad t=\cos^2\theta \\ &= \int_{\frac{\pi}{2}}^0 \cos^{2(x-1)}\theta \sin^{2(y-1)}\theta \cdot 2\cos\theta(-\sin\theta)\,\mathrm{d}\theta \\ &= 2\int_0^{\frac{\pi}{2}} \cos^{2x-1}\theta \sin^{2y-1}\theta\,\mathrm{d}\theta\,. \end{aligned}$$

Hence $B(x,y) = \dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. □

More on polar coordinates:

$$I = \int_{-\infty}^{+\infty} e^{-x^2}\,\mathrm{d}x \tag{15.3}$$

Proof.

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} e^{-x^2}\,\mathrm{d}x \int_{-\infty}^{+\infty} e^{-y^2}\,\mathrm{d}y \\ &= \iint_{\mathbf{R}^2} e^{-\left(x^2+y^2\right)}\,\mathrm{d}x\,\mathrm{d}y \qquad x=r\cos\theta, y=r\sin\theta \\ &= \int_0^{2\pi} \underbrace{\int_0^{+\infty} e^{-r^2} r\,\mathrm{d}r}_{\text{constant w.r.t. }\theta}\,\mathrm{d}\theta \qquad s=r^2, \mathrm{d}s=2r\,\mathrm{d}r \\ &= 2\pi\int_0^{+\infty} e^{-s}\cdot\frac{1}{2}\,\mathrm{d}s \\ &= 2\pi\left[\frac{1}{2}e^{-s}(-1)\right]_0^{\infty} = \pi \end{aligned}$$

and thus

$$I = \int_{-\infty}^{+\infty} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi}.$$

□

From this, we have

$$\Gamma\left(\frac{1}{2}\right) = 2\int_0^{\infty} e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}.$$

Lemma 15.23.

$$\Gamma(x) = \frac{2^{x-1}}{\sqrt{\pi}}\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right).$$

Proof. Let $f(x) = \dfrac{2^{x-1}}{\sqrt{\pi}}\Gamma\left(\dfrac{x}{2}\right)\Gamma\left(\dfrac{x+1}{2}\right)$. We want to prove that $f(x) = \Gamma(x)$.

(i)

$$\begin{aligned} f(x+1) &= \frac{2^x}{\sqrt{\pi}}\Gamma\left(\frac{x+1}{2}\right)\Gamma\left(\frac{x}{2}+1\right) \\ &= \frac{2^x}{\sqrt{\pi}}\Gamma\left(\frac{x+1}{2}\right)\frac{x}{2}\Gamma\left(\frac{x}{2}\right) \\ &= xf(x) \end{aligned}$$

(ii) $f(1) = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2}\right)\Gamma(1) = 1$ since $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

(iii)
$$\log f(x) = \underbrace{(x-1)\log 2}_{\text{linear}} + \underbrace{\log\Gamma\left(\frac{x}{2}\right)}_{\text{convex}} + \underbrace{\log\Gamma\left(\frac{x+1}{2}\right)}_{\text{convex}} - \underbrace{\log\sqrt{\pi}}_{\text{constant}}$$

and hence $\log f(x)$ is convex.

Therefore $f(x) = \Gamma(x)$. □

Theorem 15.24 (Stirling's formula)**.** This provides a simple approximate expression for $\Gamma(x+1)$ when x is large (hence for $n!$ when n is large). The formula is

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{(x/e)^x\sqrt{2\pi x}} = 1. \tag{15.4}$$

Proof. □

Lemma 15.25.
$$B(p, 1-p) = \Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin p\pi}.$$

Proof. □

IV

Topology

You have already studied metric spaces in some detail. These are objects where one has a notion of distance between points, satisfying some simple axioms. They have a rich and interesting theory, which leads to such concepts as connectedness, completeness and compactness.

Two metric spaces are viewed as "the same" if there is an isometry between them, which is a bijection that preserves distances. But there is a much more flexible notion of equivalence: two spaces are homeomorphic if there is a continuous bijection between them with continuous inverse. Many properties of metric spaces are preserved by a homeomorphism (for example, connectedness and compactness). Thus homeomorphic metric spaces may have very different metrics, but nevertheless have many properties in common. The conclusion to draw from this is that a metric is, frequently, a somewhat artificial and rigid piece of structure. So, one is led naturally to the study of Topology. The fundamental objects in Topology are topological spaces. Here, there is no metric in general. But one still has a notion of open sets, and so concepts such as connectedness and compactness continue to make sense.

Why study Topology? The reason is that it simultaneously simplifies and generalises the theory of metric spaces. By discarding the metric, and focusing solely on the more basic and fundamental notion of an open set, many arguments and proofs are simplified. And many constructions (such as the important concept of a quotient space) cannot be carried out in the setting of metric spaces: they need the more general framework of topological spaces. But perhaps the most important reason is that the spaces that arise naturally in Topology have a particularly beautiful theory.

16 Topological Spaces

§16.1 Definitions and Examples

Definition 16.1 (Topological space)**.** A **topological space** $(X, \mathcal{T})$ consists of a non-mepty set X together with a family $\mathcal{T}$ of subsets of X satisfying:

(i) $X, \emptyset \in \mathcal{T}$;

(ii) if $U_i \in \mathcal{T}$ for all $i \in I$, then $\bigcup_{i \in I} U_i \in \mathcal{T}$ (preserved under arbitrary unions);

(iii) if $U_1, \ldots, U_n \in \mathcal{T}$, then $\bigcap_{i=1}^{n} U_i \in \mathcal{T}$ (preserved under finite intersections).

The family $\mathcal{T}$ is called a **topology** for X. The sets in $\mathcal{T}$ are called the **open sets** of X.

Notation. When $\mathcal{T}$ is understood we talk about the topological space X.

Remark. A consequence of (ii) is that if $U_1, \ldots, U_n$ is a collection of open sets, then $U_1 \cap \cdots \cap U_n$ is open. But the intersection of infinitely many open sets need not be open!

On the other hand, in (iii), the indexing set I is allowed to be infinite. It may even be uncountable.

Example

The following are some examples of topological spaces. Let X be any non-empty set.

- The **discrete topology** on X is the set of all subsets of X; that is,

$$\mathcal{T} = \mathcal{P}(X).$$

- The **indiscrete topology** (or trivial topology) on X is

$$\mathcal{T} = \{X, \emptyset\}.$$

- The **co-finite topology** on X consists of the empty set together with every subset U of X such that $X \setminus U$ is finite.

Definition 16.2 (Basis)**.** A set $\mathcal{B}$ of subsets of X is a basis if

(i) $\bigcup_{B \in \mathcal{B}} B = X$;

(ii) for all $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \cap B_2$, there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.

Theorem 16.3. A basis $\mathcal{B}$ generates a topology $\mathcal{T}$ via

$$U \in \mathcal{T} \iff \forall x \in U \exists B \in \mathcal{B} \text{ such that } x \in B \subset U.$$

Proposition 16.4. Let (X, d) be a metric space. Then the open subsets of X form a topology, denoted by $\mathcal{T}_d$.

Proof. Check through the conditions in the definition for a topological space:

(i) Trivial.

(ii) Let U and V be open subsets of X. Consider an arbitrary point $x \in U \cap V$.

As U is open, there exists $r_1 > 0$ such that $B_{r_1}(x) \subset U$. Likewise, as $x \in V$ and V is open, there exists $r_2 > 0$ such that $B_{r_2}(x) \subset V$.

Take $r := \min\{r_1, r_2\}$. Then $B_r(x) \subset B_{r_1}(x) \subset U$ and $B_r(x) \subset B_{r_2}(x) \subset V$. Hence $B_r(x) \subset U \cap V$.

(iii) For every $x \in \bigcup_{i \in I} U_i$ there exists $k \in I$ such that $x \in U_k$. Since U_k is open, there exists $r > 0$ such that $B_r(x) \subset U_k \subset \bigcup_{i \in I} U_i$.

□

Definition 16.5. A topological space $(X, \mathcal{T})$ is **metrisable** if it arises from (at least oe) metric space (X, d), i.e. there is at least one metric d on X such that $\mathcal{T} = \mathcal{T}_d$.

Definition 16.6. Two metrics on a set are **topologically equivalent** if they give rise to the same topology.

Example

- The metrics d_1, d_2, d_∞ on $\mathbf{R}^n$ are all topologically equivalent. (Recall that d_1, d_2, d_∞ are the metrics arising from the norms $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$, respectively.) We shall call the topology defined by the above metrics the **standard** (or canonical) topology on $\mathbf{R}^n$.

- The discrete topology on a non-empty set X is metrisable, using the metric

$$d(x, y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

It is easy to check that this is a metric. To see that is gives the discrete topology, consider any subset $U \subset X$. Then for every $x \in U$, $B_{\frac{1}{2}}(x) \subset U$.

Definition 16.7. Given two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on the same set, we say $\mathcal{T}_1$ is **coarser** than $\mathcal{T}_2$ if $\mathcal{T}_1 \subset \mathcal{T}2$.

Remark. For any space $(X, \mathcal{T})$, the indiscrete topology on X is coarser than $\mathcal{T}$ which in turn is coarser than the discrete topology on X.

Definition 16.8. Let $(X, \mathcal{T})$ be a topological space. A subset V of X is **closed** in X if $X \setminus V$ is open in X (i.e. $X \setminus V \in \mathcal{T}$).

> **Example**
>
> - In the space $[0, 1)$ with the usual topology coming from the Euclidean metric, $[1/2, 1)$ is closed.
> - In a discrete space, all subsets are closed since their complements are open.
> - In the co-finite topology on a set X, a subset is closed if and only if it is finite or all of X.

Proposition 16.9. Let X be a topological space. Then

(i) X, $\emptyset$ are closed in X;

(ii) if V_1, V_2 are closed in X then $V_1 \cup V_2$ is closed in X;

(iii) if V_i is closed in X for all $i \in I$ then $\bigcap_{i \in I} Vi$ is closed in X.

Proof. These properties follow from (i), (ii), (iii) of definition of topological space, and from the De Morgan laws. □

Definition 16.10 (Convergent sequence)**.** A sequence $\{x_n\}_{n \in \mathbf{N}}$ in a topological space X converges to a point $x \in X$ if given any open set U containing x there exists $N \in \mathbf{N}$ such that $x_n \in U$ for all $n > N$.

> **Example**
>
> - In a metric space this is equivalent to the metric definition of convergence.
> - In an indiscrete topological space X any sequence converges to any point $x \in X$.
> - In an infinite space X with the co-finite topology any sequence $\{x_n\}$ of pairwise distinct elements (i.e. such that $x_n \neq x_m$ when $n \neq m$) converges to any point $x \in X$.

Bibliography

[Alc14] L. Alcock. *How to Think About Analysis*. Oxford University Press, 2014.

[Apo57] T. M. Apostol. *Mathematical Analysis*. Addison-Wesley, 1957.

[Art11] M. Artin. *Algebra*. Pearson Education, 2011.

[Axl24] S. Axler. *Linear Algebra Done Right, 4th edition*. Springer, 2024.

[BS11] R. G. Bartle and D. R. Sherbert. *Introduction to Real Analysis*. John Wiley & Sons, Inc., 2011.

[DF04] D. S. Dummit and R. M. Foote. *Abstract Algebra*. John Wiley & Sons, 2004.

[Mun18] J. R. Munkres. *Topology*. Pearson Education Limited, 2018.

[Pin10] C. C. Pinter. *A Book of Abstract Algebra*. Dover Publications, Inc., 2010.

[Pól45] G. Pólya. *How to Solve It*. Princeton University Press, 1945.

[Rud76] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[Sch92] A. H. Schoenfeld. "Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics". In: *Handbook for Research on Mathematics Teaching and Learning*. Macmillan, 1992, pp. 334–370.

Index