UNIVERSITY of York

Bachelor of Science

in Mathematics

Introduction to Pure Mathematics

Author:

Daniel Pérez

Contents

Abstract

This lecture notes correspond to the autumn term of the modules *Algebra* MAT00010C and *Mathematical Skills I* MAT00011C. Please note that these notes have not been endorsed by the lecturers, and I have made several modifications to them (often substantial) after the lectures. They should not be considered as accurate representations of the actual lecture content, and it is highly likely that any errors present are solely my responsibility.

CHAPTER 1

Sets, functions and relations

In this chapter, our focus will be on the fundamental components of mathematics: sets, functions, and relations. These concepts serve as the building blocks for a wide range of mathematical topics, and having a solid understanding of these terms can greatly aid in navigating the realm of mathematics.

1.1 Sets

Definition 1.1.1 (Set)**.** *A set is a collection of "objects" usually of the mathematical kind, such as numbers or points in space, etc. The objects are called the elements and are only counted once. For example, if* $a = 2, b = c = 1$*, then* $A = \{a, b, c\}$ *has only two members. If* x *is an element of a set* A*, then we write* $x \in A$*.*

Proposition 1.1.1 (Russell's paradox)**.** *It is not true that every property defines a set.*

Proof. Consider the property "is not an element of itself". A set A has that property iff $A \notin A$. Imagine there were a set $B = \{A : A \notin A\}$. Then, is $B \in B$?

If $B \in B$, then B does not have the property, so $B \notin \{A : A \notin A\}$, $B \notin B$, so B does have the property, so $B \in \{A : A \notin A\}$ so $B \in B$. Contradiction. There is no such thing as the set of all sets.

You don't get into difficulties if you take an existing set X and define A inside X by some property $A = \{x \in X : \text{ the property is true for } x\}$. □

Example 1.1.1. *Common sets and the symbols used to denote them:*

i $\mathbb{N} = \{1, 2, 3, \cdots\}$ *is the natural numbers*

ii $\mathbb{N}_0 = \{0, 1, 2, \cdots\}$ *is the natural numbers with* 0

iii $\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$ *is the integers*

iv $\mathbb{Q} = \{\frac{a}{b} : a, b \in \mathbb{Z}, b \neq 0\}$ *is the rational numbers*

v $\mathbb{R}$ *is the real numbers*

vi $\mathbb{C}$ *is the complex numbers*

It is still debated whether 0 *is a natural number. Those who believe that* 0 *is a natural number usually write* $\mathbb{N}$ *for* $\{0, 1, 2, \cdots\}$, *and* $\mathbb{N}^+$ *for the positive natural numbers. However, most of the time, it doesn't matter, and when it does, you should specify it explicitly.*

Definition 1.1.2 (Equality of sets)**.** *Two sets* A *and* B *are equal if* $A \subset B$ *and* $B \subset A$, *i.e., every element of* A *is an element of* B *and vice versa. (This is what you use when you come up with two sets and need to show that they are equal.)*

Definition 1.1.3 (Empty set)**.** $\emptyset$, *or the empty set, is the set with no elements.*

Definition 1.1.4 (Subset)**.** A *is a subset of* B, *written as* $A \subseteq B$ *or* $A \subset B$, *if all elements in* A *are in* B. *i.e.*

$$(\forall x)\, x \in A \Rightarrow x \in B.$$

Theorem 1.1.1. $(A = B) \Leftrightarrow (A \subseteq B$ *and* $B \subseteq A)$

Suppose X is a set and P is the property of some elements in x, we can write a set $\{x \in X : P(x)\}$ for the subset of x comprising of the elements for which $P(x)$ is true. e.g. $\{n \in \mathbb{N} : n \text{ is prime}\}$ is the set of all primes.

Definition 1.1.5 (Intersection). *The intersection of* A *and* B*, written* $A \cap B$*, is* $\{x : x \in A$ *and* $x \in B\}$.

Definition 1.1.6 (Union). *Let* A *and* B *be sets. The union of* A *and* B *is* $A \cup B = \{x : x \in A$ *or* $x \in B\}$.

Definition 1.1.7 (Complement). *If you have a designated universal set* X *and* A *is a subset of* X*, then the complement of* A*, written* A^C*, is* $\{x : x \notin A\}$*, when it is understood that this means* $\{x \in X : x \notin A\}$.

Definition 1.1.8 (Set difference). *If* A *and* B *are sets then the difference is* $A \setminus B = \{x \in A : x \notin B\}$.

Definition 1.1.9 (Symmetric difference). *The symmetric difference of two sets* A *and* B *is* $A \Delta B = \{x : x \in A$ *xor* $x \in B\}$*, i.e. the elements in exactly one of the two sets.*

Definition 1.1.10 (Power set). *Given a set* A*, the power set of* A *is* $\mathcal{P}(A) = \{X : X \subseteq A\}$*, i.e. the set of all subsets.*

1.1.1 Notation and some basic facts about sets

Notation 1.1.1. *If* A_α *are sets for all* $\alpha \in I$*, then*

$$\bigcap_{\alpha \in I} A_\alpha = \{x : (\forall \alpha \in I) x \in A_\alpha\}$$

and

$$\bigcup_{\alpha \in I} A_\alpha = \{x : (\exists \alpha \in I) x \in A_\alpha\}.$$

We have several rules regarding how these set operations behave, which should be intuitively obvious.

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

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

(iii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Proposition 1.1.2 (De Morgan's laws)**.** *The de Morgan's laws are defined as*

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

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

More generally,

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

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

Definition 1.1.11 (Ordered pair)**.** *An ordered pair* (a, b) *is a pair of two items in which order matters. Formally, it is defined as* $\{\{a\}, \{a, b\}\}$*. We have* $(a, b) = (a', b')$ *if and only if* $a = a'$ *and* $b = b'$*.*

Definition 1.1.12 (Cartesian product)**.** *Given two sets* A, B*, the Cartesian product of* A *and* B *is* $A \times B = \{(a, b) : a \in A, b \in B\}$*. This can be extended to* n *products, e.g.* $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) : x, y, z \in \mathbb{R}\}$ *(which is officially* $\{(x, (y, z)) : x, y, z \in \mathbb{R}\}$*).*

1.2 Functions

Definition 1.2.1 (Function)**.** *Given a set A and another set B, a function f from A to B is a way of assigning to each $x \in A$ an element $y \in B$. We write $y = f(x)$ and say that y is the image of x, and x is the preimage of y.*

A is called the domain of f. B is called the range of f. We write $f : A \to B$ for the statement that f is a function with domain A and range B. If $y = f(x)$ we also write $f : x \mapsto y$ and say x maps to y. If we wish to be very formal, we can define a function to be a subset $f \subseteq A \times B$ such that for any $a \in A$, there exists a unique $b \in B$ such that $(a, b) \in f$. We then think of $(a, b) \in f$ as saying $f(a) = b$. However, while this might act as a formal definition of a function, it is a terrible way to think about functions.

Two functions f and g are equal if they have the same domain A, the same range B and $f(x) = g(x)$ for every $x \in A$.

Example 1.2.1. *$x^2 : \mathbb{R} \to \mathbb{R}$ is a function that sends x to x^2. $\frac{1}{x} : \mathbb{R} \to \mathbb{R}$ is not a function since $f(0)$ is not defined. $\pm x : \mathbb{R} \to \mathbb{R}$ is also not a function since it is multi-valued.*

If $X \subset A$ then we define

$$f(X) = \{f(x) : x \in X\} \tag{1.1}$$

This is called the image of X. (Note, this has a different meaning of "image" since ($X \subset A$ rather than $X \in A$). $f(A)$ is also called the image of f (not the same as the range). [In some books they say "codomain" instead of "range" and "range" instead of "image".]

If $Y \subset B$ then we define

$$f^{-1}(Y)\{x \in A : f(x) \in Y\} \tag{1.2}$$

This is called the inverse image of Y.

Definition 1.2.2 (Composition of functions)**.** *The composition of two functions is a function you get by applying one after another. In particular, if $f : X \to Y$ and $G : Y \to Z$, then $g \circ f : X \to Z$ is defined by $g \circ f(x) = g(f(x))$. Note that function composition is associative.*

1.2.1 Classification of functions

Definition 1.2.3 (Injective function)**.** *A function $f : X \to Y$ is injective if it hits everything at most once, i.e.*

$$(\forall x, y \in X)\, f(x) = f(y) \Rightarrow x = y.$$

Definition 1.2.4 (Surjective function)**.** *A function $f : X \to Y$ is surjective if it hits everything at least once, i.e.*

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

Example 1.2.2. *$f : \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$ with $x \mapsto x^2$ is surjective but not injective.*

Definition 1.2.5 (Bijective function)**.** *A function is bijective if it is both injective and surjective. i.e. it hits everything exactly once.*

Definition 1.2.6 (Permutation)**.** *A permutation of A is a bijection $A \to A$.*

1.2.2 Inverse function

To define the inverse function, we will first need some preliminary definitions.

Definition 1.2.7 (Identity map)**.** *If A is a set, then the identity function on A, the identity map $\iota_A : A \to A$ is defined as the map $a \mapsto a$.*

Definition 1.2.8 (Inverses)**.** *If $f : A \to B$ and there is a function $g : B \to A$ such that*

(i) $gf(x) = x$ for every $x \in A$

(ii) $fg(y) = y$ for every $y \in B$

then g is called the inverse of f and is denoted by f^{-1}.

Then in in the definition of inverses (i) says $g \circ f = \iota_A$ and (ii) says $f \circ g = \iota_B$. In the first case, we say that g is a left-inverse for f, and in the second case that it is a right-inverse.

Definition 1.2.9 (Left inverse of function)**.** *Given $f : A \to B$, a left inverse of f is a function $g : B \to A$ such that $g \circ f = \iota_A$.*

Definition 1.2.10 (Right inverse of function)**.** *Given $f : A \to B$, a right inverse of f is a function $g : B \to A$ such that $f \circ g = \iota_B$.*

Theorem 1.2.1. *The left inverse of f exists if and only if f is injective.*

Proof. $(\Rightarrow)$ If the left inverse g exists, then $\forall a, a' \in A, f(a) = f(a') \Rightarrow g(f(a)) = g(f(a')) \Rightarrow a = a'$. Therefore f is injective.

$(\Leftarrow)$ If f is injective, we can construct a g defined as

$$g : \begin{cases} g(b) = a & \text{if } b \in f(A), \text{ where } f(a) = b \\ g(b) = \text{anything} & \text{otherwise} \end{cases}.$$

Then g is a left inverse of f. □

Theorem 1.2.2. *The right inverse of f exists if and only if f is surjective.*

Proof. $(\Rightarrow)$ We have $f(g(B)) = B$ since $f \circ g$ is the identity function. Thus f must be surjective since its image is B.

$(\Leftarrow)$ If f is surjective, we can construct a g such that for each $b \in B$, pick one $a \in A$ with $f(a) = b$, and put $g(b) = a$. □

(Note that to prove the second part, for each b, we need to pick an a such that $f(a) = b$. If B is infinite, doing so involves making infinite arbitrary choices. Are we allowed to do so?

To make infinite choices, we need to use the *Axiom of choice*, which explicitly says that this is allowed. In particular, it says that given a family of sets A_i for $i \in I$, there exists a choice function $f : I \to \bigcup A_i$ such that $f(i) \in A_i$ for all i.

So can we prove the theorem without the Axiom of Choice? The answer is no. This is since if we assume surjective functions have inverses, then we can prove the Axiom of Choice.

Assume any surjective function f has a right inverse. Given a family of non-empty sets A_i for $i \in I$ (without loss of generality (wlog) assume they are disjoint), define a function $f : \bigcup A_i \to I$ that sends each element to the set that contains the element. This is surjective since each set is non-empty. Then it has a right inverse. Then the right inverse must send each set to an element in the set, i.e. is a choice function for A_i.)

Definition 1.2.11 (Inverse of function)**.** *An inverse of f is a function that is both a left inverse and a right inverse. It is written as $f^{-1} : B \to A$. It exists if and only if f is bijective, and is necessarily unique.*

1.3 Relations

Definition 1.3.1 (Relation)**.** *A relation R on a set A specifies that some elements of A are related to some others. Formally, a relation is a subset $R \subseteq A \times A$. R is a relation on A, if, when $x, y \in A$ and you write aRb, then you obtain a sentence that may be true or false.*

Example 1.3.1. *If $A = \mathbb{N}$, or $\mathbb{Z}$, or $\mathbb{Q}$, or $\mathbb{R}$ then $=$, $<$, $\geq$, $>$, $\leq$ are all relations.*

Example 1.3.2. *The following are examples of relations on natural numbers:*

(i) aRb *iff* a *and* b *have the same final digit. e.g.* $(37)R(57)$.

(ii) aRb *iff* a *divides* b. *e.g.* $2R6$ *and* $2 \not R 7$.

(iii) aRb *iff* $a \neq b$.

(iv) aRb *iff* $a = b = 1$.

(v) aRb *iff* $|a - b| \leq 3$.

(vi) aRb *iff either* $a, b \geq 5$ *or* $a, b \leq 4$.

1.3.1 Classification of relations

Definition 1.3.2 (Reflexive relation). *A relation R is reflexive if*

$$(\forall a)\, aRa.$$

Definition 1.3.3 (Symmetric relation). *A relation R is symmetric iff*

$$(\forall a, b)\, aRb \Leftrightarrow bRa.$$

Definition 1.3.4 (Transitive relation). *A relation R is transitive iff*

$$(\forall a, b, c)\, aRb \wedge bRc \Rightarrow aRc.$$

1.3.2 Equivalence relations and partitions

Definition 1.3.5 (Equivalence relation). *A relation is an equivalence relation if it is reflexive, symmetric and transitive. e.g. (i) and (vi) in the above examples are equivalence relations.*

If it is an equivalence relation, we usually write $\sim$ instead of R. As the name suggests, equivalence relations are used to describe relations that are similar to equality. For example, if we want to represent rational numbers as a pair of integers, we might have an equivalence relation defined by $(n, m) \sim (p, q)$ iff $nq = mp$, such that two pairs are equivalent if they represent the same rational number.

Example 1.3.3. *If we consider a deck of cards, define two cards to be related if they have the same suite.*

Definition 1.3.6 (Partition of set)**.** *Let A be a set. Then a partition of a set X is a collection of subsets A_α of X such that each element of X is in exactly one of A_α. The sets A_α are called the cells of the partition.*

As mentioned, we like to think of things related by $\sim$ as equal. Hence we want to identify all "equal" things together and form one new object.

Definition 1.3.7 (Equivalence class)**.** *If $\sim$ is an equivalence relation, then the equivalence class $[x]$ is the set of all elements that are related via $\sim$ to x.*

Example 1.3.4. *In the cards example, $[8\heartsuit]$ is the set of all hearts.*

Theorem 1.3.1. *If $\sim$ is an equivalence relation on A, then the equivalence classes of $\sim$ form a partition of A.*

Proof. By reflexivity, we have $a \in [a]$. Thus the equivalence classes cover the whole set. We must now show that for all $a, b \in A$, either $[a] = [b]$ or $[a] \cap [b] = \emptyset$.

Suppose $[a] \cap [b] \neq \emptyset$. Then $\exists c \in [a] \cap [b]$. So $a \sim c, b \sim c$. By symmetry, $c \sim b$. By transitivity, we have $a \sim b$. For all $b' \in [b]$, we have $b \sim b'$. Thus by transitivity, we have $a \sim b'$. Thus $[b] \subseteq [a]$. By symmetry, $[a] \subseteq [b]$ and $[a] = [b]$. □

On the other hand, given a partition $\{B_\gamma : \gamma \in \Gamma\}$ of a set A, we can define an equivalence relation $\sim$ by $x \sim y$ iff x and y lie in the same B_γ. Each partition defines an equivalence relation in which two elements are related if and only if they are in the same partition. Thus partitions and equivalence relations are "the same thing".

Definition 1.3.8 (Quotient map)**.** *The quotient map q maps each element in A to the equivalence class containing a, i.e. $a \mapsto [a]$. e.g. $q(8\heartsuit) = \{\heartsuit\}$.*

CHAPTER 2

Division

If you feel confident in your division skills, you might be correct! However, in this chapter, we'll explore the formal definitions of division and provide solid proofs for concepts we're already familiar with (and maybe discover some new ones along the way).

Definition 2.0.1 (Prime number)**.** *A natural number n is prime if the only factors of n are 1 and n, and $n \neq 1$.*

2.1 Euclid's algorithm

Definition 2.1.1 (Factor of integers)**.** *Given $a, b \in \mathbb{Z}$, we say a divides b, a is a factor of b or $a \mid b$ if $(\exists c \in \mathbb{Z})\, b = ac$. For any b, ± 1 and $\pm b$ are always factors of b. The other factors are called proper factors.*

Theorem 2.1.1. *Every natural number $n \geq 2$ can be written as a product of primes*

Proof. Suppose not. Then let n be the smallest number that cannot be written as a product of primes. Then n isn't a prime, so we can write $n = ab$ with $a, b < n$. By the minimality of n, a and b are products of primes, so $n = ab$ is also a product of primes. □

Theorem 2.1.2 (Division Algorithm)**.** *Given $a, b \in \mathbb{Z}$, $b \neq 0$, there are unique $q, r \in \mathbb{Z}$ with $a = qb + r$ and $0 \leq r < b$.*

Despite the name, the division algorithm is not an algorithm in the usual sense. Instead, it merely states that you can divide. Even the proof does not specify a (non-brute force) way of how to divide.

Proof. Choose $q = \max\{q : qb \leq a\}$. This maximum exists because the set of all q such that $qb \leq a$ is finite. Now write $r = a - qb$. We have $0 \leq r < b$ and thus q and r are found.

To show that they are unique, suppose that $a = qb + r = q'b + r'$. We have $(q - q')b = (r' - r)$. Since both r and r' are between 0 and b, we have $-b < r - r' < b$. However, $r' - r$ is a multiple of b. Thus $q - q' = r' - r = 0$. Consequently, $q = q'$ and $r = r'$. □

Definition 2.1.2 (Common factor of integers)**.** *A common factor of a and b is a number $c \in \mathbb{Z}$ such that $c \mid a$ and $c \mid b$.*

Definition 2.1.3 (Greatest common divisor)**.** *The highest common factor or greatest common divisor of two numbers $a, b \in \mathbb{N}$ is a number $d \in \mathbb{N}$ such that d is a common factor of a and b, and if c is also a common factor, $c \mid d$.*

Clearly if the GCD exists, it must be the largest common factor, since all other common factors divide it, and thus necessarily unique. It is reasonable to consider defining the greatest common divisor $\gcd(a, b)$ as the largest common factor. Under this definition, we can show that all common factors divide it. However, the aforementioned definition is superior because it doesn't rely on a predefined ordering of natural numbers. In fact, it can be extended to any ring, even if they are not ordered, as we will see in *Rings and Modules*).

Notation 2.1.1. *We write* $d = \text{hcf}(a, b) = \text{gcd}(a, b) = (a, b)$.

Here we use (a, b) *to stand for a number, and has nothing to do with an ordered pair.*

Example 2.1.1. $(25, 105) = 5$, $(34, 55) = 1$ *and* $(47, 141) = 47$.

Lemma 2.1.1. *Let* m *and* n *be positive integers and suppose that* $n = qm + r$. *Then* $(m, n) = (r, m)$.

Proof. Suppose $d \mid m$ and $d \mid n$, and write $m = ad$, $n = bd$. Then $r = n - qm = d(a - qb)$, so $d \mid r$. Hence $(d \mid m \text{ and } d \mid n) \Rightarrow (d \mid r \text{ and } d \mid m)$. Conversely, if $d \mid r$ and $d \mid m$, then $d \mid qm + r = n$. So $(d \mid r \text{ and } d \mid m) \Rightarrow (d \mid m \text{ and } d \mid n)$. Therefore, the highest common factors are the same (since the common factors of m and n are precisely the common factors of r and m). □

Proposition 2.1.1. *If* $c \mid a$ *and* $c \mid b$, $c \mid (ua + vb)$ *for all* $u, v \in \mathbb{Z}$.

Proof. By definition, we have $a = kc$ and $b = lc$. Then $ua + vb = ukc + vlc = (uk + vl)c$. So $c \mid (ua + vb)$. □

Theorem 2.1.3. *Let* $a, b \in \mathbb{N}$. *Then* (a, b) *exists.*

Proof. Let $S = \{ua + vb : u, v \in \mathbb{Z}\}$ be the set of all linear combinations of a, b. Let d be the smallest positive member of S. Say $d = xa + yb$. Hence if $c \mid a, c \mid b$, then $c \mid d$. So we need to show that $d \mid a$ and $d \mid b$, and thus $d = (a, b)$.

By the division algorithm, there exist numbers $q, r \in \mathbb{Z}$ with $a = qd + r$ with $0 \leq r < d$. Then $r = a - qd = a(1 - qx) - qyb$. Therefore r is a linear combination of a and b. Since d is the smallest positive member of S and $0 \leq r < d$, we have $r = 0$ and thus $d \mid a$. Similarly, we can show that $d \mid b$. □

Corollary 2.1.1. *Let $d = (a, b)$, then d is the smallest positive linear combination of a and b.*

Corollary 2.1.2 (Bézout's identity)**.** *Let $a, b \in \mathbb{N}$ and $c \in \mathbb{Z}$. Then there exists $u, v \in \mathbb{Z}$ with $c = ua + vb$ iff $(a, b) \mid c$.*

Proof. ($\Rightarrow$) Let $d = (a, b)$. If c is a linear combination of a and b, then $d \mid c$ because $d \mid a$ and $d \mid b$.

($\Leftarrow$) Suppose that $d \mid c$. Let $d = xa + yb$ and $c = kd$. Then $c = (kx)a + (ky)b$. Thus c is a linear combination of a and b. □

Please note that the proof of the existence of (a, b) is not a step-by-step process, but rather a confirmation that it does indeed exist. Now, let's consider how we can actually find the values of d, x, and y in the equation $d = xa + yb$.

While it may be easy to visually inspect d for small numbers, it becomes more challenging when dealing with larger numbers, such as 4931 and 3795. Additionally, using prime factorization is not feasible at this point since (a) it is a complex process, and (b) we are not working with primes yet.

However, you may notice that if a number c divides both 4931 and 3795, it also divides their difference, which is 1136. Similarly, if c divides 1136 and 3795, it also divides their sum, which is 4931. Therefore, the problem of finding common factors of 4931 and 3795 is equivalent to finding common factors of 1136 and 3795. This process can be repeated until we reach smaller numbers that are easier to work with.

Proposition 2.1.2 (Euclid's Algorithm)**.** *If we continuously break down a and*

b by the following procedure:

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

then the highest common factor is r_{n-1}.

Proof. We have (common factors of a, b) $=$ (common factors of b, r_1) $=$ (common factors of r_1, r_2) $= \cdots =$ (factors of r_{n-1}). □

This provides another way to prove the existence of greatest common divisors (GCDs). Now, let's talk about the efficiency of this algorithm. At each step, we can see that a is always greater than $b + r_1$, which means it's at least twice as large as r_1. This implies that after every two steps, the leftmost number decreases by at least half its value. Consequently, the number of digits reduces every eight steps. As a result, the required time is at most eight times the number of digits, giving it a time complexity of $O(\log b)$.

Example 2.1.2. *Suppose $a = 57$ and $b = 42$.*

$$\begin{aligned} &\textit{common factors of } 57 \textit{ and } 42 && 57 = 1 \times 42 + 15 \\ &= \textit{common factors of } 42 \textit{ and } 15 && 42 = 2 \times 15 + 12 \\ &= \textit{common factors of } 15 \textit{ and } 12 && 15 = 1 \times 12 + 3 \\ &= \textit{common factors of } 12 \textit{ and } 3 && 12 = 4 \times 3 + 0 \\ &= \textit{common factors of } 3 \textit{ and } 0 && \\ &= \textit{factors of } 3. && \end{aligned}$$

So the GCD is 3.

By reversing Euclid's Algorithm, we can find the GCD of two numbers as a linear combination of a and b.

Example 2.1.3. *Consider* 57 *and* 21.

$$\begin{aligned}
57 &= 2 \times 21 + 15 \\
21 &= 1 \times 15 + 6 \\
15 &= 2 \times 6 + 3 \\
6 &= 2 \times 3
\end{aligned}$$

In the opposite direction, we have

$$\begin{aligned}
3 &= 15 - 2 \times 6 \\
&= 15 - 2 \times (21 - 15) \\
&= 3 \times 15 - 2 \times 21 \\
&= 3 \times (57 - 2 \times 21) - 2 \times 21 \\
&= 3 \times 57 - 8 \times 21
\end{aligned}$$

This provides an alternative, constructive proof of Bézout's identity. Additionally, it offers a convenient way to express $(a, b) = ax + by$ using a quick algorithm. However, it's important to note that this algorithm requires storing the entire process of Euclid's Algorithm, which can be space-inefficient.

In order to improve space efficiency, we aim to find a recurrence relation for the coefficients A_j and B_j, where $a \times B_j - b \times A_j = (-1)^j r_j$. The inclusion of a possible factor of -1 in the equation is merely to make the recurrence relation appear more elegant. Let's assume that this relation holds for all indices less than j. Then we have

$$\begin{aligned}
(-1)^j r_j &= (-1)^j (r_{j-2} - q_j r_{j-1}) \\
&= (-1)^{j-2} r_{j-2} + q_j (-1)^{j-1} r_{j-1} \\
&= a(B_{j-2} + q_j B_{j-1}) - b(A_{j-2} + q_j A_{j-1}).
\end{aligned}$$

Hence we can obtain the following recurrence relation:

$$\begin{aligned}
A_j &= q_j A_{j-1} + A_{j-2} \\
B_j &= q_j B_{j-1} + B_{j-2}
\end{aligned}$$

with

$$a \times B_j - b \times A_j = (-1)^j r_j.$$

In particular, $a \times B_{n-1} - b \times A_{n-1} = (-1)^{n-1} r_{n-1} = (a, b)$.

Also, by an easy induction, $A_j B_{j-1} - B_j A_{j-1} = (-1)^j$. So $(A_j, B_j) = 1$.

These coefficients also play another role. We can put the Euclid's Algorithm's equations in the following form:

$$\begin{aligned} \frac{57}{21} &= 2 + \frac{15}{21} \\ \frac{21}{15} &= 1 + \frac{6}{15} \\ \frac{15}{6} &= 2 + \frac{3}{6} \\ \frac{6}{3} &= 2 \end{aligned}$$

Then we can write out the fraction $\frac{57}{21}$ in continued fraction form

$$\frac{57}{21} = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{2}}}$$

Expanding this continued fractions term by term, we can have the sequence $2, 2+\frac{1}{1} = 3$, $2 + \frac{1}{1+\frac{1}{2}} = \frac{8}{3}$. These are called the "convergents". The sequence happens to be $\frac{A_i}{B_i}$.

2.2 Primes

Theorem 2.2.1 (Euclid's theorem). *There are infinitely many primes.*

Proof. Assume not, and let all the primes be written in a list $p_1, p_2 \cdots p_n$. Now let $N = p_1, p_2 \cdots p_n + 1$. Then, by theorem 2.1.1, N is a product of primes. However, every p_i divides $N - 1$, so it cannot divide N. Hence, there must be primes other than $p_1, p_2 \cdots p_n$. □

Proof. (Erdös 1930) Suppose that there are finitely many primes, $p_1, p_2 \cdots p_k$. Consider all numbers that are the products of these primes, i.e. $p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k}$, where $j_i \geq 0$. Factor out all squares to obtain the form $m^2 p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}$, where $m \in \mathbb{N}$ and $i_j = 0$ or 1.

Let $N \in \mathbb{N}$. Given any number $x \leq N$, when put in the above form, we have $m \leq \sqrt{N}$. So there are at most $\sqrt{N}$ possible values of m. For each m, there are 2^k numbers of the form $m^2 p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}$. So there are only $\sqrt{N} \times 2^k$ possible values of x of this kind.

Now pick $N \geq 4^k$. Then $N > \sqrt{N} \times 2^k$. So there must be a number $\leq N$ not of this form, i.e. it has a prime factor not in this list. □

Historically, many people have came up with "new" proofs that there are infinitely many primes. However, most of these proofs were just Euclid's proof in disguise. Erdös' proof is genuinely a new proof. For example, Euclid's proof comes up with a particular number N, and says all its factors are not in the list of primes. On the other hand, Erdös' proof says that there is some number, which we don't know, with at least one factor not in the list.

Also, the proofs give different bounds on when we should expect to see the kth prime. For example, Euclid tells us that the kth prime must be less than 2^{2^k}, while Erdös tells us it is less than 4^k.

Theorem 2.2.2. *If $a \mid bc$ and $(a, b) = 1$, then $a \mid c$.*

Proof. From Euclid's algorithm, there exist integers $u, v \in \mathbb{Z}$ such that $ua + vb = 1$. So multiplying by c, we have $uac + vbc = c$. Since $a \mid bc$, a |LHS. So $a \mid c$. □

Definition 2.2.1 (Coprime numbers)**.** *We say a, b are coprime if $(a, b) = 1$.*

Corollary 2.2.1. *If p is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$. (True for all p, a, b)*

Proof. We know that $(p, a) = p$ or 1 because p is a prime. If $(p, a) = p$, then $p \mid a$. Otherwise, $(p, a) = 1$ and $p \mid b$ by the theorem above. □

Corollary 2.2.2. *If p is a prime and $p \mid n_1 n_2 \cdots n_i$, then $p \mid n_i$ for some i.*

Note that when we defined primes, we defined it in terms of factors of p. This corollary is the opposite — it is about how p behaves as a factor of other numbers.

2.2.1 Solving linear equations in integers

Suppose we are given an equation of the form $ax + by = c$, where a, b, and c are integers, and asked to find integer solutions x and y.

Let $d = (a, b)$. Then $d \mid ax + by$ for any pair x, y, so if $d \nmid c$ then there are no solutions. If $d \mid c$, then $ax + by = c \Leftrightarrow \frac{a}{d}x + \frac{b}{d}y = \frac{c}{d}$. So dividing through by d, we can concentrate on the case $(a, b) = 1$.

To solve $ax + by = c$ when $(a, b) = 1$, find h, k such that $ah + bk = 1$ (using Euclid's algorithm) and set $x = ch$ and $y = ck$. Now we'd like to find all solutions to $ax + by = c$, still assuming $(a, b) = 1$. First look at the "homogeneous equation" $ax + by = 0$. Let h, k be integers such that $ha + kb = 1$. Then, if $ax + by = 0$ we have

$$\begin{aligned} &hax + hby = 0 \\ &\Rightarrow (1 - kb)x + hby = 0 \\ &\Rightarrow x = b(kx - hy) \end{aligned}$$

So $b \mid x$. Writing $x = \lambda b$, we deduce (from $ax + by = 0$) that $y = -\lambda a$. So all solutions have the form $x = \lambda b$, $y = -\lambda a$. Conversely, all such pairs are solutions.

Now suppose we have found some solution x_0, y_0 to $ax + by = c$. If $ax_0 + by_0 = c$ as well, then $a(x_0 - x') + b(y_0 - y') = 0$, so we can find an integer λ such that $x_0 - x' = \lambda b$, $y_0 - y' = -\lambda a$. Therefore, all solutions of $ax + by = c$ have the form $x = x_0 + \lambda b$, $y = y_0 - \lambda a$ and all such pairs are solutions.

2.2.2 Fundamental theorem of arithmetic

Theorem 2.2.3 (Fundamental Theorem of Arithmetic)**.** *Every natural number is expressible as a product of primes in exactly one way. In particular, if $p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$, where p_i, q_i are primes but not necessarily distinct, then $k = l$. $q_1, \cdots q_l$ are $p_1, \cdots p_k$ in some order.*

Proof. Since we already showed that there is at least one way above, we only need to show uniqueness.

Let $p_1 \cdots p_k = q_1 \cdots q_l$. We know that $p_1 \mid q_1 \cdots q_l$. Then $p_1 \mid q_1(q_2q_3 \cdots q_l)$. Thus $p_1 \mid q_i$ for some i. wlog assume $i = 1$. Then $p_1 = q_1$ since both are primes. Thus $p_2p_3 \cdots p_k = q_2q_3 \cdots q_l$. Likewise, we have $p_2 = q_2, \cdots$ and so on. □

Corollary 2.2.3. *If $a = p_1^{i_1}p_2^{i_2} \cdots p_r^{i_r}$ and $b = p_1^{j_1}p_2^{j_2} \cdots p_r^{j_r}$, where p_i are distinct primes (exponents can be zero). Then $(a, b) = \prod p_k^{\min\{i_k,j_k\}}$. Likewise, $\operatorname{lcm}(a, b) = \prod p_k^{\max\{i_k,j_k\}}$. We have $\gcd(a, b) \times \operatorname{lcm}(a, b) = ab$.*

However, this is not an efficient way to calculate (a, b), since prime factorization is very hard.

Note that this is a property peculiar to natural numbers. There are "arithmetical systems" (permitting addition, multiplication and subtraction) where factorization is not unique, e.g. even numbers.

Example 2.2.1. *The following systems have no prime unique factorization*

(i) Even numbers. "Primes" are twice of odd numbers. So 6 is a prime (NOT divisible by 2!) while 8 is not. We have $60 = 2 \times 30 = 6 \times 10$, where $2, 6, 10, 30$ are primes. However, this example is not "proper" since there is no identity element. (i.e. not a ring)

(ii) Consider $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$. We have $6 = 2 \times 3 = (1-\sqrt{-5})(1+\sqrt{-5})$. It can be shown that these are primes (see Rings and Modules).

Exercise 2.2.1. *Where does the proof of the Fundamental Theorem of Arithmetic fail in these examples?*

CHAPTER 3

Induction and counting

Experience shows that mathematicians do not know how to count.

3.1 Principle of mathematical induction

The principle of mathematical induction is as follows.

PMI 1 Let $P(1)$, $P(2)$, … be a sequence of statements. Suppose that $P(1)$ is true, and for every k, $P(k) \Rightarrow P(k+1)$. Then $P(n)$ is true for every n.

PMI 2 Let $P(1)$, $P(2)$, … as before. Suppose that $P(1)$ is true, and that for every k, $P(1) \wedge \ldots \wedge P(k) \Rightarrow P(k+1)$. Then $P(n)$ is true for every n

WOP The well-ordering principle states that every non-empty subset of $\mathbb{N}$ has a least element.

All of the above formulations are the same.

PMI 1 ⇔ PMI 2. It is easy to see that PMI 2 ⇒ PMI 1. Indeed, if the assumptions of

PMI 1 hold then the assumptions of PMI 2 hold. Now suppose that the assumption of PMI 2 hold. For each n, let $Q(n)$ be the statement $P(1) \wedge \ldots \wedge P(n)$. Then $Q(1)$ is true, and $\forall k$, $Q(k) \Rightarrow Q(k+1)$. So, by PMI 1, $Q(n)$ is true for every n. But $Q(n) \Rightarrow P(n)$.

[PM2 $\Rightarrow$ WOP] Let $A \subset \mathbb{N}$. Suppose that A does not have a smallest element. We shall prove (by PMI 2) that $A = \emptyset$. Let $P(n)$ be the statement $n \notin A$. Then $P(1)$ is true, since otherwise 1 would be the smallest element of A. If $P(1)$, …, $P(k)$ are all true, then none of $1, \ldots, k$ belong to A. So $k+1 \notin A$ since otherwise it would be the smallest element, i.e., $P(1) \wedge \ldots \wedge P(k) \Rightarrow P(k+1)$. So by PMI 2, $\forall n\ P(n)$, which says $A = \emptyset$.

[WOP $\Rightarrow$ PMI 2] Assume $P(1)$ and that $\forall k\ P(1) \wedge \ldots \wedge P(k) \Rightarrow P(k+1)$. We would like to show that $P(n)$ is true for all n. If this is not true, then $A = \{n : P(n)$ is false$\} \neq \emptyset$, so A has a least element n, by WOP. But then $P(1)$, …, $P(n-1)$ are all true, so $P(n)$ is true, by hypothesis. This contradiction implies PMI 2. □

3.2 The Inclusion-Exclusion formula

A useful theorem is the pigeonhole principle.

Theorem 3.2.1 (Pigeonhole Principle)**.** *If we put $mn+1$ pigeons into n pigeonholes, then some pigeonhole has at least $m+1$ pigeons.*

Another useful tool for counting is the indicator function.

Definition 3.2.1 (Indicator function)**.** *Let X be a set. For each $A \subseteq X$, the indicator function or characteristic function of A is the function $i_A : X \to \{0, 1\}$ with $i_A(x) = 1$ if $x \in A$, 0 otherwise. It is sometimes written as χ_A.*

Proposition 3.2.1. *These are some properties of the indicator function:*

(i) $i_A = i_B \Leftrightarrow A = B$

(ii) $i_{A \cap B} = i_A i_B$

(iii) $i_{\overline{A}} = 1 - i_A$

(iv) $i_{A\cup B} = 1 - i_{\overline{A\cup B}} = 1 - i_{\overline{A}\cap\overline{B}} = 1 - i_{\overline{A}} i_{\overline{B}} = 1 - (1 - i_A)(1 - i_B) = i_A + i_B - i_{A\cap B}$.

(v) $i_{A\setminus B} = i_{A\cap\overline{B}} = i_A i_{\overline{B}} = i_A(1 - i_B) = i_A - i_{A\cap B}$

Example 3.2.1. *We can use the indicator function to prove certain properties about sets:*

(i) Proof that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$:

$$\begin{aligned} i_{A\cap(B\cup C)} &= i_A i_{B\cup C} \\ &= i_A(i_B + i_C - i_B i_C) \\ &= i_A i_B + i_A i_C - i_A i_B i_C \\ i_{(A\cap B)\cup(A\cap C)} &= i_{A\cap B} + i_{A\cap C} - i_{A\cap C} i_{A\cap B} \\ &= i_A i_B + i_A i_C - i_A i_C i_A i_B \\ &= i_A i_B + i_A i_C - i_A i_B i_C \end{aligned}$$

Therefore $i_{A\cap(B\cup C)} = i_{(A\cap B)\cup(A\cap C)}$ and thus $A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$. Note that $i_A = i_A^2$ since $i_A = 0$ or 1, and $0^2 = 0$ and $1^2 = 1$.

(ii) Proof that the symmetric difference is associative: Observe that $i_{A\Delta B} \equiv i_A + i_B \pmod 2$. Thus $i_{(A\Delta B)\Delta C} = i_{A\Delta(B\Delta C)} \equiv i_A + i_B + i_C \pmod 2$.

Indicator functions are handy for computing the sizes of finite sets because if $A \subseteq X$, then $|A| = \sum_{x\in X} i_A(x)$.

Proposition 3.2.2. $|A \cup B| = |A| + |B| - |A \cap B|$

Proof.

$$\begin{aligned} |A \cup B| &= \sum_{x\in X} i_{A(x)\cup B(x)} \\ &= \sum (i_A(x) + i_B(x) - i_{A\cap B}(x)) \\ &= \sum i_A(x) + \sum i_B(x) - \sum i_{A\cap B}(x) \\ &= |A| + |B| - |A \cap B| \end{aligned}$$

□

Theorem 3.2.2 (Inclusion-Exclusion Principle)**.** *Let A_1, ..., A_n be a collection of a finite sets. Then*

$$| A_1 \cup \ldots \cup A_n |= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} | A_{i_1} \cap \ldots \cap A_{i_k} |$$

Proof. Let x be an element of $A_1 \cup \ldots \cup A_n$. We must show that x contributes 1 to the right-hand side. Let $\Gamma = \{i : x \in A_i\}$. Then $x \in A_{i_1} \cap \ldots \cap A_{i_k}$ if and only if $\{i_1, \ldots, i_k\} \subset \Gamma$. So the number of $i_1 < \ldots < i_k$ such that $x \in A_{i_1} \cap \ldots \cap A_{i_k}$ is $\binom{m}{k}$, where $m =| \Gamma |$. So the contribution of x to the RHS is

$$\sum_{k=1}^{n} (-1)^{k+1} \binom{m}{k} = \binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \ldots + (-1)^{m+1} \binom{m}{m}$$

But $0 = (1-1)^m = 1 - \binom{m}{1} + \binom{m}{2} - \ldots - (-1)^{m+1}\binom{m}{m}$ so $\binom{m}{1} - \binom{m}{2} + \ldots + (-1)^{m+1}\binom{m}{m} = 1$, as required. □

Example 3.2.2. *How many numbers* ≤ 200 *are coprime to 110?*

Let $X = \{1, \cdots 200\}$, *and* $A_1 = \{x : 2 \mid x\}$, $A_2 = \{x : 5 \mid x\}$, $A_3 = \{x : 11 \mid x\}$.

We know that

$$\begin{aligned}
|A_1| &= \lfloor 200/2 \rfloor = 100 \\
|A_2| &= \lfloor 200/5 \rfloor = 40 \\
|A_3| &= \lfloor 200/11 \rfloor = 18 \\
|A_1 \cap A_2| &= \lfloor 200/10 \rfloor = 20 \\
|A_1 \cap A_3| &= \lfloor 200/22 \rfloor = 9 \\
|A_2 \cap A_3| &= \lfloor 200/55 \rfloor = 3 \\
|A_1 \cap A_2 \cap A_3| &= \lfloor 200/110 \rfloor = 1
\end{aligned}$$

Then the answer is $200 - 100 - 40 - 18 + 20 + 9 + 3 - 1 = 73$.

3.3 Combinations

Example 3.3.1. *How many subsets of* $\{1, 2, \cdots n\}$ *are there? There are* $2 \times 2 \times \cdots \times 2 = 2^n$. *Since for each subset, every element is either in or out of the subset, and there are two choices for each element. Equivalently, there are* 2^n *possible indicator functions, i.e. functions* $\{1, 2, 3, \cdots, n\} \to \{0, 1\}$.

Definition 3.3.1 (Combination $\binom{n}{r}$)**.** *The number of subsets of* $\{1, 2, 3, \cdots, n\}$ *of size* r *is denoted by* $\binom{n}{r}$. *The symbol is pronounced as "n choose r".*

This is the definition of $\binom{n}{r}$. This does not in any way specify how we can actually calculate the value of $\binom{n}{r}$.

Proposition 3.3.1. *By definition,*

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$$

Theorem 3.3.1 (Binomial theorem)**.** *For* $n \in \mathbb{N}$ *with* $a, b \in \mathbb{R}$, *we have*

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

Proof. We have $(a+b)^n = (a+b)(a+b)\cdots(a+b)$. When we expand the product, we get all terms attained by choosing b from some brackets, a from the rest. The term $a^{n-r}b^r$ comes from choosing b from r brackets, a from the rest, and there are $\binom{n}{r}$ ways to make such a choice. □

Because of this theorem, $\binom{n}{r}$ is sometimes called a "binomial coefficient".

Proposition 3.3.2. *Let's see some properties of the binomial coefficient:*

(i) $\binom{n}{r} = \binom{n}{n-r}$. *This is because choosing* r *things to keep is the same as choosing* $n - r$ *things to throw away.*

(ii) $\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r}$ *(Pascal's identity) The RHS counts the number of ways*

to choose a team of r players from $n+1$ available players, one of whom is Bob. If Bob is chosen, there are $\binom{n}{r-1}$ ways to choose the remaining players. Otherwise, there are $\binom{n}{r}$ ways. The total number of ways is thus $\binom{n}{r-1}+\binom{n}{r}$.

Now given that $\binom{n}{0}=\binom{n}{n}=1$, since there is only one way to choose nothing or everything, we can construct Pascal's triangle:

$$\begin{array}{ccccccccc} &&&&1&&&& \\ &&&1&&1&&& \\ &&1&&2&&1&& \\ &1&&3&&3&&1& \\ 1&&4&&6&&4&&1 \end{array}$$

where each number is the sum of the two numbers above it, and the rth item of the nth row is $\binom{n}{r}$ (first row is row 0).

(iii) $\binom{n}{k}\binom{k}{r}=\binom{n}{r}\binom{n-r}{k-r}$. *We are counting the number of pairs of sets (Y,Z) with $|Y|=k$ and $|Z|=r$ with $Z\subseteq Y$. In the LHS, we first choose Y then choose $Z\subseteq Y$. The RHS chooses Z first and then choose the remaining $Y\setminus Z$ from $\{1,2,\cdots n\}\setminus Z$.*

Lemma 3.3.1 (Vandermonde's theorem)**.**

$$\binom{n}{k}+\binom{n}{k+1}=\binom{n+1}{k+1} \tag{3.1}$$

Proof. (i) Check it boringly.

(ii) $\binom{n+1}{k+1}$ is the number of subsets of $\{1,\dots,n+1\}$ of size $k+1$. Of those $\binom{n}{k+1}$ do not contain the element $(n+1)$ and $\binom{n}{k}$ do include the element $(n+1)$.

□

Example 3.3.2. *A greengrocer stocks n kinds of fruit. In how many ways can we choose a bag of r fruits? If we are only allowed to choose one of each kind, then the answer is $\binom{n}{r}$. But we might have $r=4$, and we want to allow picking 2 apples, 1 plum and 1 quince. The total number of ways to choose is $\binom{n+r-1}{r}$.*

Why?

Each choice can be represented by a binary string of length $n+r-1$, with r 0's and $n-1$ 1's. The string can be constructed as follows (by example): when $n=5$ and $r=8$, a possible binary string 000100110010*. The block of zeros corresponds to the number of each fruit chosen, and the 1s separate the choices. In the string above, we have 3 of type 1, 2 of type 2, 0 of type 3, 2 of type 4 and 1 of type 5. Then clearly the number of possible strings is $\binom{n+r-1}{r}$.*

Proposition 3.3.3.

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

Proof. There are $n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$ ways to choose r elements in order. Each choice of subsets is chosen this way in $r!$ orders, so the number of subsets is $\frac{n!}{(n-r)!r!}$. □

Example 3.3.3. *A bank prepares a letter for each of its n customers, saying how much it cares. (Each of these letters costs the customer £40) There are $n!$ ways to put the letters in the envelopes. In how many ways can this be done so that no one gets the right letter (i.e. how many derangements are there of n elements)? We let X be the set of all envelopings (permutation of n). $|X| = n!$. For each i, let $A_i = \{x \in X : x$ assigns the correct letter to customer $i\}$. We want to know $|\bigcap_i \overline{A}_i|$. We know that $|A_i| = (n-1)!$ since i's letter gets in i's envelopes and all others can be placed randomly. We have $|A_i \cap A_j| = (n-2)!$ as well. Similarly, $|A_i \cap A_j \cap A_k| = (n-3)!$.*

By the inclusion-exclusion formula, we have

$$\begin{aligned}\left|\bigcap_i \overline{A}_i\right| &= |X| - \sum |A_i| + \sum |A_i \cap A_j| + \cdots \\ &= n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \cdots \\ &= n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}\right) \\ &\approx n!e^{-1}\end{aligned}$$

3.4 Well-ordering and induction

Several proofs so far involved "take the least integer such that", e.g. division algorithm; or involved a sequence of moves "and so on. . . " e.g. Euclid's algorithm, every number is a product of primes. We rely on the weak principle of induction as we have seen above.

Example 3.4.1. *All numbers are equal. Let $P(n)$ be "if $\{a_1, \cdots a_n\}$ is a set of n numbers, then $a_1 = a_2 = \cdots a_n$". $P(1)$ is trivially true. Suppose we have $\{a_1, a_2 \cdots a_{n+1}\}$. Assuming $P(n)$, apply it to $\{a_1, a_2 \cdots a_n\}$ and $\{a_2, \cdots , a_{n+1}\}$, then $a_1 = \cdots = a_n$ and $a_2 = a_3 = \cdots = a_{n+1}$. So $a_1 = a_2 = \cdots = a_{n+1}$. Hence $P(n) \Rightarrow P(n + 1)$. So $P(n)$ is true for all $n \in \mathbb{N}$.*

Theorem 3.4.1. *Inclusion-exclusion principle.*

Proof. Let $P(n)$ be the statement "for any sets $A_1 \cdots A_n$", we have $|A_1 \cup \cdots \cup A_n| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \cdots \pm |A_i \cap A_2 \cap \cdots \cap A_n|$".

$P(1)$ is trivially true. $P(2)$ is also true (see above). Now given $A_1 \cdots A_{n+1}$, Let $B_i = A_i \cap A_{n+1}$ for $1 \le i \le n$. We apply $P(n)$ both to the A_i and B_i.

Now observe that $B_i \cap B_j = A_i \cap A_j \cap A_{n+1}$. Likewise, $B_i \cap B_j \cap B_k = A_i \cap A_j \cap A_k \cap A_{n+1}$. Now

$$\begin{aligned}|A_1 \cup A_2 \cup \cdots \cup A_{n+1}| &= |A_1 \cup \cdots \cup A_n| + |A_{n+1}| - |(A_1 \cup \cdots \cup A_n) \cap A_{n+1}| \\ &= |A_1 \cup \cdots \cup A_n| + |A_{n+1}| - |B_1 \cup \cdots \cup B_n| \\ &= \sum_{i \le n} |A_i| - \sum_{i<j\le n} |A_i \cap A_j| + \cdots + |A_{n+1}| \\ &\quad - \sum_{i \le n} |B_i| + \sum_{i<j\le n} |B_i \cap B_j| - \cdots\end{aligned}$$

Note $\sum\limits_{i \le n} |B_i| = \sum\limits_{i \le n} |A_i \cap A_{n+1}|$. So $\sum\limits_{i<j\le n} |A_i \cap A_j| + \sum\limits_{i \le n} |B_i| = \sum\limits_{i<j\le n+1} |A_i \cap A_j|$, and similarly for the other terms. So

$$= \sum_{i \le n+1} |A_i| - \sum_{i<j\le n+1} |A_i \cap A_j| + \cdots$$

So $P(n) \Rightarrow P(n + 1)$ for $n \ge 2$. By WPI, $P(n)$ is true for all n. □

However, WPI is not quite what we want for "every number is a product of primes". We need a different form of induction.

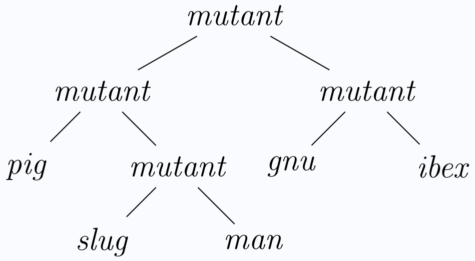
Theorem 3.4.2 (Strong principle of induction)**.** *Let $P(n)$ be a statement about $n \in \mathbb{N}$. Suppose that*

1. *$P(1)$ is true*
2. *$\forall n \in N$, if $P(k)$ is true $\forall k < n$ then $P(n)$ is true.*

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

Note that (i) is redundant as it follows from (ii), but we state it for clarity.

Example 3.4.2. *"Evolutionary trees" Imagine that we have a mutant that can produce two offsprings. Each offspring is either an animal or another mutant. A possible evolutionary tree is as follows:*



Let $P(n)$ be the statement $n - 1$ mutants produces n animals. Given some tree with n animals, remove the top mutant to get two sub-trees, with n_1 and n_2 animals, where $n_1 + n_2 = n$. If $P(k)$ is true $\forall k < n$, then $P(n_1)$ and $P(n_2)$ are true. So the total number of mutants is $1 + (n_1 - 1) + (n_2 - 1) = n - 1$. So $P(n)$ is true. Hence by strong principle of induction, $P(n)$ is true for all n.

Definition 3.4.1 (Partial order)**.** *A partial order on a set is a reflexive, anti-symmetric ($(aRb) \wedge (bRa) \Leftrightarrow a = b$) and transitive relation.*

Example 3.4.3. *The ordinary ordering of $\mathbb{N}$ $a \leq b$ is a partial order of $\mathbb{N}$. Also, $a \mid b$ on $\mathbb{N}$ is also a partial order.*

Definition 3.4.2 (Total order)**.** *A total order is a partial order where $\forall a \neq b$, exactly one of aRb or bRa holds. This means that every two things must be related.*

Definition 3.4.3 (Well-ordered total order)**.** *A total order is well-ordered if every non-empty subset has a minimal element, i.e. if $S \neq \emptyset$, then $\exists m \in S$ such that $x < m \Rightarrow x \notin S$.*

Example 3.4.4. $\mathbb{Z}$ *with the usual order is not well-ordered since the set of even integers has no minimum. The positive rationals are also not well-ordered under the usual order.*

Theorem 3.4.3 (Well-ordering principle)**.** $\mathbb{N}$ *is well-ordered under the usual order, i.e. every non-empty subset of* $\mathbb{N}$ *has a minimal element.*

Example 3.4.5. *Proof that every number is a product of primes by strong induction: Assume the contrary. Then there exists a minimal n that cannot be written as a product of prime (by the well-ordering principle). If n is a prime, then n is a product of primes. Otherwise, write $n = ab$, where $1 < a, b < n$. By minimality of n, both a and b are products of primes. Hence so is n. Contradiction.*

Example 3.4.6. *All numbers are interesting. Suppose that there are uninteresting numbers. Then there exists a smallest uninteresting number. Then the property of being the smallest uninteresting number is itself interesting. Contradiction.*

Example 3.4.7. *Consider a total order on $\mathbb{N} \times \mathbb{N}$ by "lexicographic" or "dictionary" order, i.e. $(a, b) \leq (c, d)$ if $a < c$ or $(a = c \land b \leq d)$.*

The Ackermann function is a function $a : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{N}$ *is defined by*

$$a(m,n) = \begin{cases} n+1 & \textit{if } m = 0 \\ a(m-1,1) & \textit{if } m > 0 \textit{ and } n = 0 \\ a(m-1, a(m, n-1)) & \textit{if } m > 0 \textit{ and } n > 0. \end{cases}$$

We want to show that this is well-defined.

Note that $a(m,n)$ *is expressed in terms of* a *at points* $(x,y) < (m,n)$*. So* a *is well-defined if lexicographic order is well-ordered, i.e. every non-empty subset has a minimal element (if* a *were not well-defined, then would be a smallest place where the definition is bad. But definition of that point is defined in terms of smaller points which are well defined).*

We can see that $\mathbb{N}_0 \times \mathbb{N}_0$ *is well-ordered: if* $S \subseteq \mathbb{N}_0 \times \mathbb{N}_0$ *is non-empty, let* S_x *be the set of* $\{x \in \mathbb{N} : (\exists y)\,(x,y) \in S\}$*, i.e. the set of all* x*-coordinates of* S*. By the well-ordering principle,* S_x *has a minimal element* m*. Then let* $S_y = \{y \in \mathbb{N}_0 : (m,y) \in S\}$*. Then* S_y *has a minimal element* n*. Then* (m,n) *is the minimal element of* S*.*

CHAPTER 4

Modular arithmetic

Loosely speaking, modular arithmetic is like the arithmetic of clocks: when you get to a certain number, you go back to the beginning again. We will study arithmetic under this number system. Like the integers, we are allowed to add and multiply numbers. However, while in $\mathbb{Z}$, we can only divide by 1 and -1, in modular arithmetic, more numbers can be divided.

4.1 Modular arithmetic

Definition 4.1.1 (Modulo)**.** *If $a, b \in \mathbb{Z}$ have the same remainder after division by m, i.e. $m \mid (a - b)$, we say a and b are congruent modulo m, and write*

$$a \equiv b \pmod{m}$$

Example 4.1.1. *The check digits of the ISBN (or Hong Kong ID Card Number) are calculated modulo 11.*

Example 4.1.2. $9 \equiv 0 \pmod 3$, $11 \equiv 6 \pmod 5$.

Proposition 4.1.1. *If $a \equiv b \pmod m$, and $d \mid m$, then $a \equiv b \pmod d$.*

Proof. $a \equiv b \pmod m$ if and only if $m \mid (a-b)$, hence $d \mid (a-b)$, i.e. $a \equiv b \pmod d$. □

Observe that with m fixed, $a \equiv b \pmod m$ is an equivalence relation. The set of equivalence classes is written as $\mathbb{Z}_m$ or $\mathbb{Z}/(m\mathbb{Z})$.

Example 4.1.3. $\mathbb{Z}_3 = \{[0], [1], [2]\}$

Proposition 4.1.2. *If $a \equiv b \pmod m$ and $u \equiv v \pmod m$, then $a + u \equiv b + v \pmod m$ and $au \equiv bv \pmod m$.*

Proof. Since $a \equiv b \pmod m$ and $u \equiv v \pmod m$, we have $m \mid (a-b) + (u-v) = (a+u) - (b+v)$. So $a + u \equiv b + v \pmod m$

Since $a \equiv b \pmod m$ and $u \equiv v \pmod m$, we have $m \mid (a-b)u + b(u-v) = au - bv$. So $au \equiv bv \pmod m$. □

This means that we can do arithmetic modulo n. Formally, we are doing arithmetic with the congruence classes, i.e $\mathbb{Z}_m$. For example, in $\mathbb{Z}_7$, $[4] + [5] = [9] = [2]$.

Modular arithmetic can sometimes be used to show that equations have no solutions.

Example 4.1.4. $2a^2 + 3b^3 = 1$ *has no solutions in $\mathbb{Z}$. If there were a solution, then $2a^2 \equiv 1 \pmod 3$. But $2 \cdot 0^2 \equiv 0$, $2 \cdot 1^2 \equiv 2$ and $2 \cdot 2^2 \equiv 2$. So there is no solution to the congruence, and hence none to the original equation.*

Observe that all odd numbers are either $\equiv 1 \pmod 4$ or $\equiv 3 \equiv -1 \pmod 4$. So we can classify primes depending on their value modulo 4.

Theorem 4.1.1. *There are infinitely many primes that are $\equiv -1 \pmod 4$.*

Proof. Suppose not. So let $p_1, \cdots p_k$ be all primes $\equiv -1 \pmod 4$. Let $N = 4p_1p_2\cdots p_k - 1$. Then $N \equiv -1 \pmod 4$. Now N is a product of primes, say $N = q_1q_2\cdots q_\ell$. But $2 \nmid N$ and $p_i \nmid N$ for all i. So $q_i \equiv 1 \pmod 4$ for all i. But then that implies $N = q_1q_2\cdots q_\ell \equiv 1 \pmod 4$, which is a contradiction. □

Example 4.1.5. *Solve* $7x \equiv 2 \pmod{10}$. *Note that* $3 \cdot 7 \equiv 1 \pmod{10}$. *If we multiply the equation by* 3, *then we get* $3 \cdot 7 \cdot x \equiv 3 \cdot 2 \pmod{10}$. *So* $x \equiv 6 \pmod{10}$. *Effectively, we divided by* 7.

"Division" doesn't always work for all numbers, e.g. you cannot divide by 2 mod 10. We give a name to numbers we can divide.

Definition 4.1.2 (Unit)**.** *u is a unit if $\exists v$ such that* $uv \equiv 1 \pmod m$*.*

Theorem 4.1.2. *u is a unit modulo m if and only if* $(u, m) = 1$*.*

Proof. ($\Rightarrow$) Suppose u is a unit. Then $\exists v$ such that $uv \equiv 1 \pmod m$. Then $uv = 1 + mn$ for some n, or $uv - mn = 1$. So 1 can be written as a linear combination of u and m. So $(u, m) = 1$.

($\Leftarrow$) Suppose that $(u, m) = 1$. Then there exists a, b with $ua + mb = 1$. Thus $ua \equiv 1 \pmod m$. □

Using the above proof, we can find the inverse of a unit efficiently by Euclid's algorithm.

Corollary 4.1.1. *If $(a, m) = 1$, then the congruence $ax \equiv b \pmod m$ has a unique solution (mod m).*

Proof. If $ax \equiv b \pmod m$, and $(a, m) = 1$, then $\exists a^{-1}$ such that $a^{-1}a \equiv 1 \pmod m$. So $a^{-1}ax \equiv a^{-1}b \pmod m$ and thus $x \equiv a^{-1}b \pmod m$. Finally we check that $x \equiv a^{-1}b \pmod m$ is indeed a solution: $ax \equiv aa^{-1}b \equiv b \pmod m$. □

Proposition 4.1.3. *There is a solution to* $ax \equiv b \pmod{m}$ *if and only if* $(a, m) \mid b$.

If $d = (a, m) \mid b$*, then the solution is the unique solution to* $\frac{a}{d}x \equiv \frac{b}{d} \pmod{\frac{m}{d}}$

Proof. Let $d = (a, m)$. If there is a solution to $ax \equiv b \pmod{m}$, then $m \mid ax - b$. So $d \mid ax - b$ and $d \mid b$.

On the contrary, if $d \mid b$, we have $ax \equiv b \pmod{m} \Leftrightarrow ax - b = km$ for some $k \in Z$. Write $a = da'$, $b = db'$ and $m = dm'$. So $ax \equiv b \pmod{m} \Leftrightarrow da'x - db' = dkm' \Leftrightarrow a'x - b' = km' \Leftrightarrow a'x \equiv b' \pmod{m'}$. Note that $(a', m') = 1$ since we divided by their greatest common factor. Then this has a unique solution modulo m'. □

Example 4.1.6. $2x \equiv 3 \pmod{4}$ *has no solution since* $(2, 4) = 2$ *which does not divide* 3.

4.2 Multiple moduli

Suppose we are given $x \equiv 2 \pmod{3}$ and $x \equiv 1 \pmod{4}$. What is the general solution to x? We work in mod 12. Since we are given that $x \equiv 2 \pmod 3$, we know that $x \equiv 2, 5, 8 \text{ or } 11 \pmod{12}$. Similarly, since $x \equiv 1 \pmod{4}$, we must have $x \equiv 1, 5 \text{ or } 9 \pmod{12}$. Combining these results, we must have $x \equiv 5 \pmod{12}$.

On the other hand, if $x \equiv 5 \pmod{12}$, then $x \equiv 5 \equiv 2 \pmod{3}$ and $x \equiv 5 \equiv 1 \pmod{4}$. So $x \equiv 5 \pmod{12}$ is indeed the most general solution.

Theorem 4.2.1 (Chinese remainder theorem)**.** *Let* $(m, n) = 1$ *and* $a, b \in \mathbb{Z}$. *Then there is a unique solution (modulo mn) to the simultaneous congruences*

$$\begin{cases} x \equiv a \pmod{m} \\ x \equiv b \pmod{n} \end{cases},$$

i.e. $\exists x$ *satisfying both and every other solution is* $\equiv x \pmod{mn}$.

Proof. Since $(m, n) = 1$, $\exists u, v \in \mathbb{Z}$ with $um + vn = 1$. Then $vn \equiv 1 \pmod{m}$ and

$um \equiv 1 \pmod{n}$. Put $x = umb + vna$. So $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$.

To show it is unique, suppose both y and x are solutions to the equation. Then

$$\begin{aligned}
&y \equiv a \pmod{m} \text{ and } y \equiv b \pmod{n} \\
\Leftrightarrow\ &y \equiv x \pmod{m} \text{ and } y \equiv x \pmod{n} \\
\Leftrightarrow\ &m \mid y - x \text{ and } n \mid y - x \\
\Leftrightarrow\ &mn \mid y - x \\
\Leftrightarrow\ &y \equiv x \pmod{mn}
\end{aligned}$$

□

This shows a congruence (mod mn) is equivalent to one (mod n) and another (mod m).

Example 4.2.1.

$$\begin{cases} x \equiv 2 \pmod{3} \\ x \equiv 5 \pmod{7} \\ x \equiv 4 \pmod{1}1 \end{cases},$$

Note that $x = 26 \pmod{3 \times 7 \times 11}$. *We first show existence: first we do the case* $r_1 = 1$, $r_2 = \ldots = r_k = 0$. *We know, by an earlier lemma (says* $(a, m) = (b, m) = 1 \Rightarrow (ab, m) = 1$*), that*

$$(a_1, a_2, \ldots, a_k) = 1$$

So we can find u*,* v *such that* $a_1u + a_2a_3 \ldots a_kv = 1$. *Then* $a_2a_3 \ldots a_kv$ *is a multiple of* a_i *when* $i \geq 2$ *and is* $\equiv 1 \pmod{a}_1$. *Call this number* x_1. *Similarly, for other* i*, we can find* x_i *such that*

$$x_i \equiv \begin{cases} 1 \pmod{a_j} & j = i \\ 0 \pmod{a_j} & j \neq i \end{cases}$$

Then $\sum_{i=1}^{k} r_ix_i \equiv r_j \pmod{a}_j$ *since* $a_j \mid x_i$ *when* $i \neq j$ *and* $x_j \equiv 1 \pmod{a}_j$.

Example 4.2.2.

$$\begin{cases} x \equiv 2 \pmod 3 \\ x \equiv 3 \pmod 7 \end{cases}$$

We have

$$7 \equiv 1 \pmod 3 \text{ and } 7 \equiv 0 \pmod 7$$

$$15 \equiv 0 \pmod 3 \text{ and } 15 \equiv 1 \pmod 7$$

and deduce that

$$x \equiv 2 \times 7 + 3 \times 15 = 59 \equiv 17 \pmod{21}$$

By uniqueness, suppose that for every i we have $x \equiv r_i \pmod{a_i}$ and $y \equiv r_i \pmod{a_i}$. Then $a_i \mid x - y$ for every i, so, by the lemma, $a_1 a_2 \dots a_k \mid x - y$. If $0 \leq x,\ y < a_1 a_2 \dots a_k$ it follows that $x = y$.

We can easily extend this to more than two moduli by repeatedly applying this theorem.

Proposition 4.2.1. *Given any $(m, n) = 1$, c is a unit mod mn iff c is a unit both mod m and mod n.*

Proof. ($\Rightarrow$) If $\exists u$ such that $cu \equiv 1 \pmod{mn}$, then $cu \equiv 1 \pmod m$ and $cu \equiv 1 \pmod n$. So c is a unit mod m and n.

($\Leftarrow$) Suppose there exists u, v such that $cu \equiv 1 \pmod m$ and $cv \equiv 1 \pmod n$. Then by CRT, $\exists w$ with $w \equiv u \pmod m$ and $w \equiv v \pmod n$. Then $cw \equiv cu \equiv 1 \pmod m$ and $cw \equiv cv \equiv 1 \pmod n$.

But we know that $1 \equiv 1 \pmod m$ and $1 \equiv 1 \pmod n$. So 1 is a solution to $cw \equiv 1 \pmod m$, $cw \equiv 1 \pmod n$. By the "uniqueness" part of the Chinese remainder theorem, we must have $cw \equiv 1 \pmod{mn}$. □

Definition 4.2.1 (Euler's totient function)**.** *We denote by $\phi(m)$ the number of integers a, $0 \leq a \leq m$, such that $(a, m) = 1$, i.e. a is a unit mod m. Note $\phi(1) = 1$.*

Proposition 4.2.2. *Here are some properties of the Euler's totient function:*

(i) $\phi(mn) = \phi(m)\phi(n)$ *if* $(m, n) = 1$, *i.e.* ϕ *is multiplicative.*

(ii) If p *is a prime,* $\phi(p) = p - 1$

(iii) If p *is a prime,* $\phi(p^k) = p^k - p^{k-1} = p^k(1 - 1/p)$

(iv) $\phi(m) = m\prod_{p|m}(1 - 1/p)$.

Proof. We will only prove (iv). In fact, we will prove it twice.

(i) Suppose $m = p_1^{k_1}p_2^{k_2}\cdots p_\ell^{k_\ell}$. Then

$$\begin{aligned}\phi(m) &= \phi(p_1^{k_1})\phi(p_2^{k_2})\cdots\phi(p_\ell^{k_\ell})\\ &= p_1^{k_1}(1 - 1/p_1)p_2^{k_2}(1 - 1/p_2)\cdots p_\ell^{k_\ell}(1 - 1/p_\ell)\\ &= m\prod_{p|m}(1 - 1/p)\end{aligned}$$

(ii) Let $m = p_1^{k_1}p_2^{k_2}\cdots p_\ell^{k_\ell}$. Let $X = \{0, \cdots m - 1\}$. Let $A_j = \{x \in X : p_j \mid x\}$. Then $|X| = m$, $|A_j| = m/p_j$, $|A_i \cap A_j| = m/(p_ip_j)$ etc. So $\phi(m) = |\overline{A}_1 \cap \overline{A}_2 \cap \cdots \overline{A}_\ell| = m\prod_{p|m}(1 - 1/p)$.

□

Example 4.2.3. $\phi(60) = 60(1 - 1/2)(1 - 1/3)(1 - 1/5) = 16$.

If a, b are both units (mod m), then so is ab, for if $au \equiv 1$ and $bv \equiv 1$, then $(ab)(uv) \equiv 1$. So the units form a multiplicative group of size $\phi(m)$.

4.3 Prime moduli

Modular arithmetic has some nice properties when the modulus is a prime number.

Theorem 4.3.1 (Wilson's theorem)**.** $(p - 1)! \equiv -1 \pmod p$ *if* p *is a prime.*

Proof. If p is a prime, then $1, 2, \cdots, p - 1$ are units. Among these, we can pair each number up with its inverse (e.g. 3 with 4 in modulo 11). The only elements that cannot be paired with a different number are 1 and -1, who are self-inverses, as show below:

$$\begin{aligned} & x^2 \equiv 1 \pmod p \\ \Leftrightarrow\ & p \mid (x^2-1) \\ \Leftrightarrow\ & p \mid (x-1)(x+1) \\ \Leftrightarrow\ & p \mid x-1 \text{ or } p \mid x+1 \\ \Leftrightarrow\ & x \equiv \pm 1 \pmod p \end{aligned}$$

Now $(p-1)!$ is a product of $(p-3)/2$ inverse pairs together with 1 and -1. So the product is -1. □

Theorem 4.3.2 (Fermat's little theorem)**.** *Let p be a prime. Then $a^p \equiv a \pmod p$ for all $a \in \mathbb{Z}$. Equivalently, $a^{p-1} \equiv 1 \pmod p$ if $a \not\equiv 0 \pmod p$.*

Proof. Two proofs are offered:

(i) The numbers $\{1, 2, \cdots p-1\}$ are units modulo p and form a group of order $p-1$. So $a^{p-1} \equiv 1$ by Lagrange's theorem.

(ii) If $a \not\equiv 0$, then a is a unit. So $ax \equiv ay$ iff $x \equiv y$. Then $a, 2a, 3a, \cdots (p-1)a$ are distinct mod p. So they are congruent to $1, 2, \cdots p-1$ in some order. Hence $a \cdot 2a \cdots 3a \cdots (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdots (p-1)$. So $a^{p-1}(p-1)! \equiv (p-1)!$. So $a^{p-1} \equiv 1 \pmod p$.

□

Neither Wilson nor Fermat's theorem hold if the modulus is non-prime. However, Fermat's theorem can be generalized:

Theorem 4.3.3 (Fermat-Euler Theorem)**.** *Let a, m be coprime. Then*

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

Proof. Lagrange's theorem: The units mod m form a group of size $\phi(m)$.

Alternatively, let $U = \{x \in \mathbb{N} : 0 < x < m, (x, m) = 1\}$. These are the $\phi(m)$ units. Since a is a unit, $ax \equiv ay \pmod m$ only if $x \equiv y \pmod m$. So if $U =$

$\{u_1, u_2, \cdots, u_{\phi(m)}\}$, then $\{au_1, au_2, \cdots au_{\phi(m)}\}$ are distinct and are units. So they must be $u_1, \cdots u_{\phi(m)}$ in some order. Then $au_1 au_2 \cdots au_{\phi(m)} \equiv u_1 u_2 \cdots u_{\phi(m)}$. So $a^{\phi(m)} z \equiv z$, where $z = u_1 u_2 \cdots u_{\phi(m)}$. Since z is a unit, we can multiply by its inverse and obtain $a^{\phi(m)} \equiv 1$. □

Definition 4.3.1 (Quadratic residues)**.** *The quadratic residues are the "squares" mod p, i.e.* $1^2, 2^2, \cdots, (p-1)^2$.

Note that if $a^2 \equiv b^2 \pmod p$, then $p \mid a^2 - b^2 = (a-b)(a+b)$. Then $p \mid a - b$ or $p \mid a + b$. So $a \equiv \pm b \pmod p$. Thus every square is a square of exactly two numbers.

Example 4.3.1. *If* $p = 7$, *then* $1^2 \equiv 6^2 \equiv 1$, $2^2 \equiv 5^2 \equiv 4$, $3^2 \equiv 4^2 \equiv 2$. *So* $1, 2, 4$ *are quadratic residues.* $3, 5, 6$ *are not.*

Proposition 4.3.1. *If* p *is an odd prime, then* -1 *is a quadratic residue if and only if* $p \equiv 1 \pmod 4$.

Proof. If $p \equiv 1 \pmod 4$, say $p = 4k+1$, then by Wilson's theorem, $-1 \equiv (p-1)! \equiv 1 \cdot 2 \cdots \left(\frac{p-1}{2}\right)\left(-\frac{p-1}{2}\right) \cdots (-2)(-1) \equiv (-1)^{\frac{p-1}{2}} \left(\frac{p-1}{2}\right)!^2 = (-1)^{2k}(2k!)^2 \equiv (2k!)^2$. So -1 is a quadratic residue.

When $p \equiv -1 \pmod 4$, i.e. $p = 4k+3$, suppose -1 is a square, i.e. $-1 \equiv z^2$. Then by Fermat's little theorem, $1 \equiv z^{p-1} \equiv z^{4k+2} \equiv (z^2)^{2k+1} \equiv (-1)^{2k+1} \equiv -1$. Contradiction. □

Proposition 4.3.2. *(Unproven) A prime p is the sum of two squares if and only if* $p \equiv 1 \pmod 4$.

Proposition 4.3.3. *There are infinitely many primes* $\equiv 1 \pmod 4$.

Proof. Suppose not, and $p_1, \cdots p_k$ are all the primes $\equiv 1 \pmod 4$. Let $N = (2p_1 \cdots p_k)^2 + 1$. Then N is not divisible by 2 or $p_1, \cdots, p_k$. Let q be a prime $q \mid N$. Then $q \equiv -1 \pmod 4$. Then $N \equiv 0 \pmod q$ and hence $(2p_1 \cdots p_k)^2 + 1 \equiv 0 \pmod q$, i.e. $(2p_1 \cdots p_k)^2 \equiv -1 \pmod{q}$. So -1 is a quadratic residue mod q, which is a contradiction since $q \equiv -1 \pmod{4}$. □

Proposition 4.3.4. *Let $p = 4k+3$ be a prime. Then if a is a quadratic residue, i.e. $a \equiv z^2 \pmod{p}$ for some z, then $z = \pm a^{k+1}$.*

Proof. By Fermat's little theorem, $a^{2k+1} \equiv z^{4k+2} \equiv z^{p-1} \equiv 1$. If we multiply by a, then $a^{2k+2} \equiv a \pmod{p}$. So $(\pm a^{k+1})^2 \equiv a \pmod{p}$. □

This allows us to take square roots efficiently. This efficiency requires an effective way of computing powers of a efficiently. This can be done by repeated squaring. For example, to find a^{37}, we can calculate this by $a^{37} = a^{32}a^4a^1 = ((((a^2)^2)^2)^2)^2 \cdot (a^2)^2 \cdot a$. Thus calculation of a^n has time complexity $O(\log n)$, as opposed to $O(n)$ if you take powers manually.

Suppose a is a square mod n, where $n = pq$ and p, q are distinct primes. Then a is a square mod p and a square mod q. So there exists some s with $(\pm s)^2 \equiv a \pmod{p}$ and some t with $(\pm t)^2 \equiv a \pmod{q}$. By the Chinese remainder theorem, we can find a unique solution of each case, so we get 4 square roots of a modulo n.

CHAPTER 5

Real numbers

Up until now, our focus has been solely on natural numbers and integers. However, in the real world, we often encounter rational numbers and even real numbers. Before we proceed, it is important to address a significant point from a philosophical standpoint. The idea is to define the "real numbers" as a set equipped with specific operations (such as addition and multiplication) that satisfy certain properties known as axioms. In doing so, we can pose two questions: first, does a set actually exist that satisfies these properties, and second, is it unique?

The first question can be answered by means of an explicit construction. Essentially, we find a concrete set that indeed satisfies these properties. However, it is crucial to note that we undertake the construction merely to demonstrate that it is reasonable to discuss such a set. For instance, we construct a real number as a pair of subsets of $\mathbb{Q}$. Yet, it would be absurd to claim that a real number "is" a pair of sets. Rather, it can be constructed as a pair of sets. If we consider each real number as a set (which is valid and true), it would be considered absurd to ask whether

$$\exists x : x \in 3 \lor x \in \pi$$

even though it is a valid question within the framework of viewing real numbers as sets.

The issue of uniqueness is more intricate. Firstly, it is evident that the constructions themselves are not unique. Instead of constructing the natural number 0 as the set $\emptyset$, as we will do later, we could define it as $\emptyset$, and the entire construction would still hold. However, we might hope that all possible constructions are somehow "isomorphic." It turns out this is true for what we present below, although the proofs are not straightforward.

Nevertheless, while this notion of isomorphism is intriguing, it is not particularly significant. This is because we are not concerned with how the real numbers, for example, are constructed. When we work with them, we simply assume that they satisfy the relevant defining properties. Hence, we can choose any set that fulfills the axioms and utilize it. The existence of other non-isomorphic sets is inconsequential.

Given that our main interest lies in the natural numbers and the real numbers, we will provide both an axiomatic description and an explicit construction for these. However, we will only offer explicit constructions for the integers and rationals, without delving into detailed proofs of their validity.

5.1 Construction of numbers

5.1.1 Construction of natural numbers

Our construction of natural numbers will include 0 (even though $0 \notin \mathbb{N}$).

Definition 5.1.1 (Natural numbers)**.** *The natural numbers* $\mathbb{N}$ *is defined by Peano's axioms. We call a set* $\mathbb{N}$ *“natural numbers” if it has a special element* 0 *and a map* $S : \mathbb{N} \to \mathbb{N}$ *that maps* n *to its "successor" (intuitively, it is* $+1$*) such that:*

(i) $S(n) \neq 0$ *for all* $n \in \mathbb{N}$

(ii) For all $n, m \in \mathbb{N}$*, if* $S(n) = S(m)$*, then* $n = m$*.*

(iii) For any subset $A \subseteq \mathbb{N}$*, if* $0 \in A$ *and “*$n \in A \Rightarrow S(n) \in A$*, then in fact*

$A = \mathbb{N}$.

The last axiom is the axiom of induction.

We write $1 = S(0), 2 = S(1)$, $3 = S(2)$ *etc. We can (but will not) prove that the axiom of induction allows us to define functions on* $\mathbb{N}$ *recursively (cf. IID Logic and Set Theory). Assuming this, we can define addition and multiplication recursively by*

$$\begin{aligned} n + 0 &= n & n \times 0 &= 0 \\ n + S(m) &= S(n + m) & n \times S(m) &= n \times m + n \end{aligned}$$

We can show by induction that these satisfy the usual rules (e.g. associativity, distributivity).

We can construct this explicitly by $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$ etc. In general, we define $S(n) = \{n\} \cup n$. Note that this is in some sense a circular definition, since we are defining the natural numbers recursively, but to do recursion, we need the natural numbers. Also, it is not clear how we can show this satisfies the axioms above. To actually do this properly, we will need to approach this in a slightly different way.

5.1.2 Construction of integers

Definition 5.1.2 (Integers). *$\mathbb{Z}$ is obtained from $\mathbb{N}$ by allowing subtraction. Formally, we define $\mathbb{Z}$ to be the equivalence classes of $\mathbb{N} \times \mathbb{N}$ under the equivalence relation*

$$(a, b) \sim (c, d) \quad \textit{if and only if} \quad a + d = b + c.$$

Intuitively, we think of (a, b) as $a - b$.

We write a for $[(a, 0)]$ and $-a$ for $[(0, a)]$, and define the operations by

$$\begin{aligned} (a, b) + (c, d) &= (a + c, b + d) \\ (a, b) \times (c, d) &= (ac + bd, bd + ad). \end{aligned}$$

We can check that these are well-defined and satisfy the usual properties.

5.1.3 Construction of rationals

Definition 5.1.3 (Rationals)**.** $\mathbb{Q}$ *is obtained from* $\mathbb{Z}$ *by allowing division. Formally, we define* $\mathbb{Q}$ *to be the equivalence classes of* $\mathbb{Z} \times \mathbb{N}$ *under the relation*

$$(a, b) \sim (c, d) \quad \textit{if and only if} \quad ad = bc.$$

We write $\frac{a}{b}$ *for* $[(a, b)]$*. We define*

$$\begin{aligned}(a, b) + (c, d) &= (ad + bc, bd)\\(a, b) \times (c, d) &= (ac, bd).\end{aligned}$$

We can check that these are well-defined and satisfy the usual properties.

Algebraically, we say $\mathbb{Q}$ is a "totally ordered field".

Definition 5.1.4 (Totally ordered field)**.** *A set* F *equipped with binary operations* $+, \times$ *and relation* $\leq$ *is a* totally ordered field *if*

(i) F *is an additive abelian group with identity* 0.

(ii) $F \setminus \{0\}$ *is a multiplicative abelian group with identity* 1.

(iii) Multiplication is distributed over addition: $a(b + c) = ab + ac$.

(iv) $\leq$ *is a total order.*

(v) For any $p, q, r \in F$*, if* $p \leq q$*, then* $p + r \leq q + r$.

(vi) For any $p, q, r \in F$*, if* $p \leq q$ *and* $0 \leq r$*, then* $pr \leq qr$.

Proposition 5.1.1. $\mathbb{Q}$ *is a totally ordered-field.*

Examples of non-totally-ordered fields include $\mathbb{Z}_p$, which is a field but not totally ordered.

Proposition 5.1.2. $\mathbb{Q}$ *is densely ordered, i.e. for any* $p, q \in \mathbb{Q}$*, if* $p < q$*, then there is some* $r \in \mathbb{Q}$ *such that* $p < r < q$.

Proof. Take $r = \frac{p+q}{2}$. □

However, $\mathbb{Q}$ is not enough for our purposes.

Proposition 5.1.3. *There is no rational $q \in \mathbb{Q}$ with $q^2 = 2$.*

Proof. Suppose not, and $(\frac{a}{b})^2 = 2$, where b is chosen as small as possible. We will derive a contradiction in four ways.

(i) $a^2 = 2b^2$. So a is even. Let $a = 2a'$. Then $b^2 = 2a'^2$. Then b is even as well, and $b = 2b'$. But then $\frac{a}{b} = \frac{a'}{b'}$ with a smaller b'. Contradiction.

(ii) We know that b is a product of primes if $b \neq 1$. Let $p \mid b$. Then $a^2 = 2b^2$. So $p \mid a^2$. So $p \mid a$. Contradict b minimal.

(iii) (Dirichlet) We have $\frac{a}{b} = \frac{2b}{a}$. So $a^2 = 2b^2$. For any, u, v, we have $a^2v = 2b^2v$ and thus $uab + a^2v = uab + 2b^2v$. So $\frac{a}{b} = \frac{au+2bv}{bu+av}$. Put $u = -1, v = 1$. Then $\frac{a}{b} = \frac{2b-a}{a-b}$. Since $a < 2b, a - b < b$. So we have found a rational with smaller b.

(iv) Same as 3, but pick u, v so $bu + av = 1$ since a and b are coprime. So $\frac{a}{b}$ is an integer.

□

5.1.4 Construction of real numbers

As illustrated earlier, the rational numbers do not encompass all the numbers that we require. But what precisely do we mean by stating that "numbers are missing"? One might argue that the issue lies in the fact that not all polynomial equations have solutions. However, this is not the fundamental problem. Firstly, even when working within the realm of real numbers, not all equations possess solutions. For instance, consider the equation $x^2 + 1 = 0$. Furthermore, certain real numbers, like π, do not serve as solutions to polynomial equations with integer coefficients, yet we still desire to include them in our number system.

The real problem is expressed in terms of least upper bounds, or suprema.

Definition 5.1.5 (Least upper bound and greatest lower bound)**.** *For an ordered set X, $s \in X$ is a least upper bound (or supremum) for the set $S \subseteq X$, denoted by $s = \sup S$, if*

(i) s is an upper bound for S, i.e. for every $x \in S$, we have $x \leq s$.

(ii) if t is any upper bound for S, then $s \leq t$.

Similarly, $s \in X$ is a greatest lower bound (or infimum) if s is a lower bound and any lower bound $t \leq s$.

By definition, the least upper bound for S, if exists, is unique.

The problem with $\mathbb{Q}$ is that if we let $S = \{q \in \mathbb{Q} : q^2 < 2\}$, then it has no supremum in $\mathbb{Q}$.

Recall that $\mathbb{Q}$ is a totally ordered field. We will define the real numbers axiomatically to be a totally ordered field without this problem.

Definition 5.1.6 (Real numbers)**.** *The real numbers is a totally ordered field containing $\mathbb{Q}$ that satisfies the least upper bound axiom.*

Axiom 5.1.1 (Least upper bound axiom)**.** *Every non-empty set of the real numbers that has an upper bound has a least upper bound.*

We have the requirement "non-empty" since every number is an upper bound of $\emptyset$ but it has no least upper bound.

We will leave the construction to the end of the section.

Note that $\mathbb{Q}$ is a subset of $\mathbb{R}$, in the sense that we can find a copy of $\mathbb{Q}$ inside $\mathbb{R}$. By definition of a field, there is a multiplicative identity $1 \in \mathbb{R}$. We can then define the natural numbers by

$$n = \underbrace{1 + \cdots + 1}_{n \text{ times}}.$$

We can then define the negative integers by letting $-n$ be the additive inverse of n. Then $\frac{1}{n}$ is the multiplicative inverse of n (for $n \neq 0$), and $\frac{m}{n}$ is just m copies of $\frac{1}{n}$ added together. This is our canonical copy of $\mathbb{Q}$.

Corollary 5.1.1. *Every non-empty set of the real numbers bounded below has an infimum.*

Proof. Let S be non-empty and bounded below. Then $-S = \{-x : x \in S\}$ is a non-empty set bounded above, and $\inf S = -\sup(-S)$. □

Alternatively, we can prove it just using the ordering of $\mathbb{R}$:

Proof. Let S be non-empty and bounded below. Let L be the set of all lower bounds of S. Since S is bounded below, L is non-empty. Also, L is bounded above by any member of S. So L has a least upper bound $\sup L$.

For each $x \in S$, we know x is an upper bound of L. So we have $\sup L \leq x$ by definition. So $\sup L$ is indeed a lower bound of S. Also, by definition, every lower bound of S is less than (or equal to) $\sup L$. So this is the infimum. □

Now the set $\{q \in \mathbb{Q} : q^2 < 2\}$ has a supremum in $\mathbb{R}$ (by definition).

We make some useful definitions.

Definition 5.1.7 (Closed and open intervals)**.** *A closed interval* $[a, b]$ *with* $a \leq b \in \mathbb{R}$ *is the set* $\{x \in \mathbb{R} : a \leq x \leq b\}$.
An open interval (a, b) *with* $a \leq b \in \mathbb{R}$ *is the set* $\{x \in \mathbb{R} : a < x < b\}$.
Similarly, we can have $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ *and* $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$.

Example 5.1.1. *Let* $S = [0, 1]$*. Then* $S \neq \emptyset$*. Also* S *has an upper bound, e.g.* 2*. Hence* $\sup S$ *exists.*
To find it explicitly, notice that 1 *is an upper bound for* S *by definition, and if* $t < 1$*, then* t *is not an upper bound for* S *since* $1 \in S$ *but* $1 \not\leq t$*. So every upper bound is at least* 1 *and therefore* 1 *is the supremum of* S*.*
Now let $T = (0, 1)$*. Again* T *is non-empty and has an upper bound (e.g.* 2*). So again* $\sup T$ *exists. We know that* 1 *is an upper bound. If* $t < 0$*, then* $0.5 \in S$ *but* $s \not\leq t$*. So* t *is not an upper bound. Now suppose* $0 \leq t < 1$*, then* $0 < t < \frac{1+t}{2} < 1$

and so $\frac{1+t}{2} \in S$ *but* $\frac{1+t}{2} \not\leq t$*. So* t *is not an upper bound. So* $\sup T = 1$.

Note that these cases differ by $\sup S \in S$ *but* $\sup T \notin T$*.* S *has a maximum element* 1 *and the maximum is the supremum.* T *doesn't have a maximum, but the supremum can still exist.*

The real numbers has a rather interesting property.

Theorem 5.1.1 (Axiom of Archimedes)**.** *Given* $r \in \mathbb{R}$*, there exists* $n \in \mathbb{N}$ *with* $n > r$*.*

This was considered an axiom by Archimedes but we can prove this with the least upper bound axiom.

Proof. Assume the contrary. Then r is an upper bound for $\mathbb{N}$. $\mathbb{N}$ is not empty since $1 \in \mathbb{N}$. By the least upper bound axiom, $s = \sup \mathbb{N}$ exists. Since s is the least upper bound for $\mathbb{N}$, $s - 1$ is not an upper bound for $\mathbb{N}$. So $\exists m \in \mathbb{N}$ with $m > s - 1$. Then $m+1 \in \mathbb{N}$ but $m+1 > s$, which contradicts the statement that s is an upper bound. □

Notice that every non-empty set $S \in \mathbb{R}$ which is bounded below has a greatest lower bound (or infimum). In particular, we have

Proposition 5.1.4. $\inf\{\frac{1}{n} : n \in \mathbb{N}\} = 0$.

Proof. Certainly 0 is a lower bound for S. If $t > 0$, there exists $n \in \mathbb{N}$ such that $n \geq 1/t$. So $t \geq 1/n \in S$. So t is not a lower bound for S. □

Theorem 5.1.2. $\mathbb{Q}$ *is dense in* $\mathbb{R}$*, i.e. given* $r, s \in \mathbb{R}$*, with* $r < s$*,* $\exists q \in \mathbb{Q}$ *with* $r < q < s$*.*

Proof. wlog assume first $r \geq 0$ (just multiply everything by -1 if $r < 0$ and swap r and s). Since $s - r > 0$, there is some $n \in \mathbb{N}$ such that $\frac{1}{n} < s - r$. By the Axiom of Archimedes, $\exists N \in \mathbb{N}$ such that $N > sn$.

Let $T = \{k \in \mathbb{N} : \frac{k}{n} \geq s\}$. T is not empty, since $N \in T$. Then by the well-ordering principle, T has a minimum element m. Now $m \neq 1$ since $\frac{1}{n} < s - r \leq s$. Let $q = \frac{m-1}{n}$. Since $m - 1 \notin T$, $q < s$. If $q = \frac{m-1}{n} < r$, then $\frac{m}{n} < r + \frac{1}{n} < s$, so $m \notin T$, contradiction. So $r < q < s$. □

Theorem 5.1.3. $\sqrt{2}$ *exists, i.e., if* $\mathbb{F}$ *is a complete ordered field then there exists* $x \in \mathbb{R}$ *such that* $x > 0$ *and* $x^2 = 2$.

Proof. Let $S = \{r \in \mathbb{R} : r^2 \leq 2\}$. Then $0 \in S$ so $S \neq \emptyset$. Also for every $r \in S$, we have $r \leq 3$. So S is bounded above. So $x = \sup S$ exists and $0 \leq x \leq 3$.

By trichotomy, either $x^2 < 2, x^2 > 2$ or $x^2 = 2$.

Suppose $x^2 < 2$. Let $0 < t < 1$. Then consider $(x+t)^2 = x^2 + 2xt + t^2 < x^2 + 6t + t \leq x^2 + 7t$. Pick $t < \frac{2-x^2}{7}$, then $(x+t)^2 < 2$. So $x + t \in S$. This contradicts the fact that x is an upper bound of S.

Now suppose $x^2 > 2$. Let $0 < t < 1$. Then consider $(x-t)^2 = x^2 - 2xt + t^2 \geq x^2 - 6t$. Pick $t < \frac{x^2-2}{6}$. Then $(x-t)^2 > 2$, so $x - t$ is an upper bound for S. This contradicts the fact that x is the least upper bound of S.

So by trichotomy, $x^2 = 2$. □

Now, let us embark on the construction of the real numbers using the rationals as our starting point. The concept here is that each set, such as $q \in \mathbb{Q} : q^2 < 2$, represents a "missing number," specifically the supremum $\sqrt{2}$. However, multiple sets can correspond to the same missing number. For instance, $q \in \mathbb{Q} : q^2 < 2 \cup -3$ could also "should have" the supremum $\sqrt{2}$. Consequently, we must choose a specific set that accurately represents this missing number.

To accomplish this, we select the maximal set, denoted by S, which satisfies the following condition: for any $x \in S$ and $y \in \mathbb{Q}$, if $y < x$, then $y \in S$. In cases where y is not in S, we can include it in S, and the set $S \cup y$ will still "have the same upper bound." Furthermore, each rational number $q \in \mathbb{Q}$ can also be represented by the set $x \in \mathbb{Q} : x \leq q$. These representations are commonly known as Dedekind cuts. However, for the sake of convention, a Dedekind cut is defined as the pair $(S, \mathbb{Q} \setminus S)$ and can be characterized by the following definition:

Definition 5.1.8 (Dedekind cut)**.** *A Dedekind cut of $\mathbb{Q}$ is a set of partition of $\mathbb{Q}$ into L and R such that*

$$(\forall l \in L)(\forall r \in R)\, l < r,$$

and R has no minimum.

The requirement that R has no minimum corresponds to our (arbitrary) decision that the rationals should be embedded as

$$q \mapsto \{x \in q : x \leq \mathbb{Q}\}, \{x \in \mathbb{Q} : x > q\},$$

instead of $q \mapsto \{x \in q : x < \mathbb{Q}\}, \{x \in \mathbb{Q} : x \geq q\}$,

We can then construct the set $\mathbb{R}$ from $\mathbb{Q}$ by letting $\mathbb{R}$ be the set of all Dedekind cuts. The supremum of any bounded set of real numbers is obtained by taking the union of (the left sides) of the Dedekind cuts. The definition of the arithmetic operations is left as an exercise for the reader (to actually define them is tedious but not hard).

5.2 Sequences

Here we will look at sequences, with series in the next chapter. Only a brief introduction to these topics will be provided here, as these topics will be studied later in Analysis I.

Definition 5.2.1 (Sequence)**.** *A sequence is a function $\mathbb{N} \to \mathbb{R}$. If a is a sequence, instead of $a(1), a(2), \cdots$, we usually write $a_1, a_2, \cdots$. To emphasize it is a sequence, we write the sequence as (a_n).*

We want to capture the notion of a sequence tending to a limit. For example, we want to say that $1, \frac{1}{2}, \frac{1}{3}, \cdots$ tends to 0, while $1, 2, 3, \cdots$ does not converge to a limit.

The idea is that if $a_n \to l$, then we can get as close to l as we like, as long as we are sufficiently far down the sequence. More precisely, given any "error threshold" ε, we can find a (possibly large) number N such that whenever $n \geq N$, we have $|a_n - l| < \varepsilon$.

Definition 5.2.2 (Limit of sequence)**.** *The sequence (a_n) tends to $l \in \mathbb{R}$ as n tends to infinity if and only if*

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N})(\forall n \geq N)\, |a_n - l| < \varepsilon.$$

If a_n tends to l as n tends to infinity, we write $a_n \to l$ as $n \to \infty$; $\lim\limits_{n\to\infty} a_n = l$; or a_n converges to l.

Intuitively, if $a_n \to l$, we mean given any ε, for sufficiently large n, a_n is always within $l \pm \varepsilon$.

The definition $a_n \not\to l$ is the negation of the above statement:

$$(\exists \varepsilon > 0)(\forall N \in \mathbb{N})(\exists n \geq N)\, |a_n - l| \geq \varepsilon.$$

Definition 5.2.3 (Convergence of sequence)**.** *The sequence (a_n) converges if there exists an l such that $a_n \to l$. The sequence diverges if it doesn't converge.*

Every proof of $a_n \to l$ looks like: Given $\varepsilon > 0$, (argument to show N exists, maybe depending on ε), such that $\forall n \geq N$, $|a_n - l| < \varepsilon$.

Example 5.2.1. *Show that $a_n = 1 - \frac{1}{n} \to 1$.*
Given $\varepsilon > 0$, choose $N > \frac{1}{\varepsilon}$, which exists by the Axiom of Archimedes. If $n \geq N$, then $|a_n - 1| = \frac{1}{n} \leq \varepsilon$. So $a_n \to 1$.

Example 5.2.2. *Let*

$$a_n = \begin{cases} \frac{1}{n} & n \text{ is prime} \\ \frac{1}{2n} & n \text{ is not prime} \end{cases}.$$

We will show that $a_n \to 0$. Given $\varepsilon > 0$. Choose $N > \frac{1}{\varepsilon}$. Then $\forall n \geq N$, $|a_n - 0| \leq \frac{1}{n} < \varepsilon$.

Example 5.2.3. *Prove that*

$$a_n = \begin{cases} 1 & n \text{ is prime} \\ 0 & n \text{ is not prime} \end{cases}$$

diverges.

Let $\varepsilon = \frac{1}{3}$. *Suppose* $l \in \mathbb{R}$. *If* $l < \frac{1}{2}$, *then* $|a_n - l| > \varepsilon$ *when* n *is prime. If* $l \geq \frac{1}{2}$, *then* $|a_n - l| > \varepsilon$ *when* n *is not prime. Since the primes and non-primes are unbounded,* $(\forall N)\exists n > N$ *such that* $|a_n - l| > \varepsilon$. *So* a_n *diverges.*

An important property of $\mathbb{R}$ is the following:

Theorem 5.2.1. *Every bounded monotonic sequence converges.*

In case of confusion, the terms are defined as follows: (a_n) is increasing if $m \leq n$ implies $a_m \leq a_n$. Decreasing is defined similarly. Then it is monotonic if it is increasing or decreasing. (a_n) is bounded if there is some $B \in \mathbb{R}$ such that $|a_n| \leq B$ for all n.

Proof. wlog assume (a_n) is increasing. The set $\{a_n : n \geq 1\}$ is bounded and non-empty. So it has a supremum l (least upper bound axiom). Show that l is the limit:

Given any $\varepsilon > 0$, $l - \varepsilon$ is not an upper bound of a_n. So $\exists N$ such that $a_N \geq l - \varepsilon$. Since a_n is increasing, we know that $l \geq a_m \geq a_N > l - \varepsilon$ for all $m \geq N$. So $\exists N$ such that $\forall n \geq N$, $|a_n - l| < \varepsilon$. So $a_n \to l$. □

We can show that this theorem is equivalent to the least upper bound axiom.

Definition 5.2.4 (Subsequence)**.** *A subsequence of* (a_n) *is* $(a_{g(n)})$ *where* $g : \mathbb{N} \to \mathbb{N}$ *is strictly increasing. e.g.* $a_2, a_3, a_5, a_7 \cdots$ *is a subsequence of* (a_n).

Theorem 5.2.2. *Every sequence has a monotonic subsequence.*

Proof. Call a point a_k a "peak" if $(\forall m \geq k)\, a_m \leq a_k$. If there are infinitely many peaks, then they form a decreasing subsequence. If there are only finitely many peaks, $\exists N$ such that no a_n with $n > N$ is a peak. Pick a_{N_1} with $N_1 > N$. Then pick a_{N_2} with

$N_2 > N_1$ and $a_{N_2} > a_{N_1}$. This is possible because a_{N_1} is not a peak. The pick a_{N_3} with $N_3 > N_2$ and $a_{N_3} > a_{N_2}$, ad infinitum. Then we have a monotonic subsequence. □

We will now prove the following basic properties of convergence:

Theorem 5.2.3. *(i) If $a_n \to a$ and $a_n \to b$, then $a = b$ (i.e. limits are unique)*

(ii) If $a_n \to a$ and $b_n = a_n$ for all but finitely many n, then $b_n \to a$.

(iii) If $a_n = a$ for all n, then $a_n \to a$.

(iv) If $a_n \to a$ and $b_n \to b$, then $a_n + b_n \to a + b$

(v) If $a_n \to a$ and $b_n \to b$, then $a_n b_n \to ab$

(vi) If $a_n \to a \neq 0$, and $\forall n(a_n \neq 0)$. Then $1/a_n \to 1/a$.

(vii) If $a_n \to a$ and $b_n \to a$, and $\forall n(a_n \leq c_n \leq b_n)$, then $c_n \to a$. (Sandwich theorem)

Many students are confused as to why we should prove these "obvious" properties of convergence. It seems "obvious" that if a_n converges to a and b_n converges to b, then the sum converges to $a + b$. However, it is not obvious (at least for the first-time learners) that if $(\forall \varepsilon > 0)(\exists N)(\forall n \geq N)\, |a_n - a| < \varepsilon$ and $(\forall \varepsilon > 0)(\exists N)(\forall n \geq N)\, |b_n - b| < \varepsilon$, then $(\forall \varepsilon > 0)(\exists N)(\forall n \geq N)\, |(a_n + b_n) - (a + b)| < \varepsilon$. In some sense, what we are trying to prove is that our attempt at defining convergence actually satisfies the "obvious" properties we think convergence should satisfy.

In proving this, we will make frequent use of the triangle inequality: $|x + y| \leq |x| + |y|$.

Proof. (i) Suppose instead $a < b$. Then choose $\varepsilon = \frac{b-a}{2}$. By the definition of the limit, $\exists N_1$ such that $\forall n \geq N_1$, $|a_n - a| < \varepsilon$. There also $\exists N_2$ st. $\forall n \geq N_2$, $|a_n - b| < \varepsilon$.

Let $N = \max\{N_1, N_2\}$. If $n \geq \max\{N_1, N_2\}$, then $|a - b| \leq |a - a_n| + |a_n - b| < 2\varepsilon = b - a$. Contradiction. So $a = b$.

(ii) Given $\varepsilon > 0$, there $\exists N_1$ st. $\forall n \geq N_1$, we have $|a_n - a| < \varepsilon$. Since $b_n = a_n$ for all but finitely many n, there exists N_2 such that $\forall n \geq N_2$, $a_n = b_n$.

Let $N = \max\{N_1, N_2\}$. Then $\forall n \geq N$, we have $|b_n - a| = |a_n - a| < \varepsilon$. So $b_n \to a$.

(iii) $\forall \varepsilon$, take $N = 1$. Then $|a_n - a| = 0 < \varepsilon$ for all $n \geq 1$.

(iv) Given $\varepsilon > 0$, $\exists N_1$ such that $\forall n \geq N_1$, we have $|a_n - a| < \varepsilon/2$. Similarly, $\exists N_2$ such that $\forall n \geq N_2$, we have $|b_n - b| < \varepsilon/2$.

Let $N = \max\{N_1, N_2\}$. Then $\forall n \geq N$, $|(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b| < \varepsilon$.

(v) Given $\varepsilon > 0$, Then there exists N_1, N_2, N_3 such that

$$\begin{aligned}
\forall n \geq N_1 &: |a_n - a| < \frac{\varepsilon}{2(|b| + 1)} \\
\forall n \geq N_2 &: |b_n - b| < \frac{\varepsilon}{2|a|} \\
\forall n \geq N_3 &: |b_n - b| < 1 \Rightarrow |b_n| < |b| + 1
\end{aligned}$$

Then let $N = \max\{N_1, N_2, N_3\}$. Then $\forall n \geq N$,

$$\begin{aligned}
|a_n b_n - ab| &= |b_n(a_n - a) + a(b_n - b)| \\
&\leq |b_n||a_n - a| + |a||b_n - b| \\
&< (|b| + 1)|a_n - a| + |a||b_n - b| \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon
\end{aligned}$$

(vi) Given $\varepsilon > 0$, then $\exists N_1, N_2$ such that $|a_n - a| < \frac{|a|^2}{2}\varepsilon$ and $|a_n - a| < \frac{|a|}{2}$.

Let $N = \max\{N_1, N_2\}$. The $\forall n \geq N$,

$$\begin{aligned}
\left|\frac{1}{a_n} - \frac{1}{a}\right| &= \frac{|a_n - a|}{|a_n||a|} \\
&< \frac{2}{|a|^2}|a_n - a| \\
&< \varepsilon
\end{aligned}$$

(vii) By (iii) to (v), we know that $b_n - a_n \to 0$. Let $\varepsilon > 0$. Then $\exists N$ such that $\forall n \geq N$, we have $|b_n - a_n| < \varepsilon$. So $|c_n - a_n| < \varepsilon$. So $c_n - a_n \to 0$. So $c_n = (c_n - a_n) + a_n \to a$.

$\square$

Example 5.2.4. *Let* $x_n = \frac{n^2(n+1)(2n+1)}{n^4+1}$. *Then we have*

$$x_n = \frac{(1+1/n)(2+1/n)}{1+1/n^4} \to \frac{1 \cdot 2}{1} = 2$$

by the theorem (many times).

Example 5.2.5. *Let* $y_n = \frac{100^n}{n!}$. *Since* $\frac{y_{n+1}}{y_n} = \frac{100}{n+1} < \frac{1}{2}$ *for large* $n > 200$, *we know that* $0 \leq y_n < y_{200} \cdot \frac{2^{200}}{2^n}$. *Since* $y_{200} \cdot \frac{2^{200}}{2^n} \to 0$, *we know that* $y_n \to 0$ *as well.*

5.3 Series

In a field, the sum of two numbers is defined. By induction, the sum of finitely many numbers is defined as well. However, infinite sums ("series") are not. We will define what it means to take an infinite sum. Of course, infinite sums exist only for certain nice sums. For example, $1 + 1 + 1 + \cdots$ does not exist.

Definition 5.3.1 (Series and partial sums)**.** *Let* (a_n) *be a sequence. Then* $s_m = \sum_{n=1}^{m} a_n$ *is the mth partial sum of* (a_n). *We write*

$$\sum_{n=1}^{\infty} a_n = \lim_{m \to \infty} s_m$$

if the limit exists.

Example 5.3.1. *Let* $a_n = \frac{1}{n(n-1)}$ *for* $n \geq 2$. *Then*

$$s_m = \sum_{n=2}^{m} \frac{1}{n(n-1)} = \sum_{n=2}^{m} \left(\frac{1}{n-1} - \frac{1}{n} \right) = 1 - \frac{1}{m} \to 1.$$

Then

$$\sum_{n=2}^{\infty} \frac{1}{n(n-1)} = 1.$$

Example 5.3.2. *Let* $a_n = \frac{1}{n^2}$. *Then* $s_m = \sum_{n=1}^{m} \frac{1}{n^2}$. *We know that* s_m *is increasing. We also know that* $s_m \leq 1 + \sum \frac{1}{n(n-1)} \leq 2$, *i.e. it is bounded above.*

So s_m converges and $\sum_{n=1}^{\infty} \frac{1}{n^2}$ exists (in fact it is $\pi^2/6$).

Example 5.3.3. *(Geometric series) Suppose $a_n = r^n$, where $|r| < 1$. Then $s_m = r \cdot \frac{1-r^m}{1-r} \to \frac{r}{1-r}$ since $r^n \to 0$. So*

$$\sum_{n=1}^{\infty} r^n = \frac{r}{1-r}.$$

Example 5.3.4. *(Harmonic series) Let $a_n = \frac{1}{n}$. Consider*

$$\begin{aligned} S_{2^k} &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \cdots + \frac{1}{2^k} \\ &\geq 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{2^k} \\ &\geq 1 + \frac{k}{2}. \end{aligned}$$

So $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

Decimal expansions

Definition 5.3.2 (Decimal expansion)**.** *Let (d_n) be a sequence with $d_n \in \{0, 1, \cdots 9\}$. Then $\displaystyle\sum_{n=1}^{\infty} \frac{d}{10^n}$ converges to a limit r with $0 \leq r \leq 1$ since the partial sums s_m are increasing and bounded by $\sum \frac{9}{10^n} \to 1$ (geometric series). We say $r = 0.d_1 d_2 d_3 \cdots$, the* decimal expansion *of r.*

Does every x with $0 \leq x < 1$ have a decimal expansion? Pick d_1 maximal such that $\frac{d_1}{10} \leq x < 1$. Then $0 \leq x - \frac{d_1}{10} < \frac{1}{10}$ since d_1 is maximal. Then pick d_2 maximal such that $\frac{d_2}{100} \leq x - \frac{d_1}{10}$. By maximality, $0 \leq x - \frac{d_1}{10} - \frac{d_2}{100} < \frac{1}{100}$. Repeat inductively, pick maximal d_n with

$$\frac{d_n}{10^n} \leq x - \sum_{j=1}^{n-1} \frac{d_j}{10^j}$$

so

$$0 \leq x - \sum_{j=1}^{n} \frac{d_j}{10^j} < \frac{1}{10^n}.$$

Since both LHS and RHS → 0, by sandwich, $x - \sum_{j=1}^{\infty} \frac{d_j}{10^j} = 0$, i.e. $x = 0.d_1d_2\cdots$.

Since we have shown that at least one decimal expansion, can the same number have two different decimal expansions? i.e. if $0.a_1a_2\cdots = 0.b_1b_2\cdots$, must $a_i = b_i$ for all i?

Now suppose that the a_j and b_j are equal until k, i.e. $a_j = b_j$ for $j < k$. wlog assume $a_k < b_k$. Then

$$\sum_{j=k+1}^{\infty} \frac{a_j}{10^j} \leq \sum_{j=k+1}^{\infty} \frac{9}{10^j} = \frac{9}{10^{k+1}} \cdot \frac{1}{1 - 1/10} = \frac{1}{10^k}.$$

So we must have $b_k = a_k + 1$, $a_j = 9$ for $j > k$ and $b_j = 0$ for $j > k$. For example, $0.47999\cdots = 0.48000\cdots$.

5.4 Irrational numbers

Recall $\mathbb{Q} \subseteq \mathbb{R}$.

Definition 5.4.1 (Irrational number)**.** *Numbers in $\mathbb{R} \setminus \mathbb{Q}$ are irrational.*

Definition 5.4.2 (Periodic number)**.** *A decimal is periodic if after a finite number ℓ of digits, it repeats in blocks of k for some k, i.e. $d_{n+k} = d_n$ for $n > \ell$.*

Proposition 5.4.1. *A number is periodic iff it is rational.*

Proof. Clearly a periodic decimal is rational: Say $x = 0.7413157157157\cdots$. Then

$$\begin{aligned}
10^\ell x &= 10^4 x \\
&= 7413.157157\cdots \\
&= 7413 + 157\left(\frac{1}{10^3} + \frac{1}{10^6} + \frac{1}{10^9} + \cdots\right) \\
&= 7413 + 157 \cdot \frac{1}{10^3} \cdot \frac{1}{1 - 1/10^3} \in \mathbb{Q}
\end{aligned}$$

Conversely, let $x \in \mathbb{Q}$. Then x has a periodic decimal. Suppose $x = \frac{p}{2^c 5^d q}$ with $(q, 10) = 1$. Then $10^{\max(c,d)} x = \frac{a}{q} = n + \frac{b}{q}$ for some $a, b, n \in \mathbb{Z}$ and $0 \leq b < q$. However, since $(q, 10) = 1$, by Fermat-Euler, $10^{\phi(q)} \equiv 1 \pmod{q}$, i.e. $10^{\phi(q)} - 1 = kq$ for some k. Then

$$\frac{b}{q} = \frac{kb}{kq} = \frac{kb}{999\cdots 9} = kb\left(\frac{1}{10^{\phi(q)}} + \frac{1}{10^{2\phi(q)}} + \cdots\right).$$

Since $kb < kq < 10^{\phi(q)}$, write $kb = d_1 d_2 \cdots d_{\phi(q)}$. So $\frac{b}{q} = 0.d_1 d_2 \cdots d_{\phi(q)} d_1 d_2 \cdots$ and x is periodic. □

Example 5.4.1. $x = 0.01101010001010\cdots$, *where* 1*s appear in prime positions, is irrational since the digits don't repeat.*

5.5 Euler's number

Definition 5.5.1 (Euler's number)**.**

$$e = \sum_{j=0}^{\infty} \frac{1}{j!} = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots$$

This sum exists because the partial sums are bounded by $1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \cdots = 3$ and it is increasing. So $2 < e < 3$.

Proposition 5.5.1. *e is irrational.*

Proof. Is $e \in \mathbb{Q}$? Suppose $e = \frac{p}{q}$. We know $q \geq 2$ since e is not an integer (it is between 2 and 3). Then $q!e \in \mathbb{N}$. But

$$q!e = \underbrace{q! + q! + \frac{q!}{2!} + \frac{q!}{3!} + \cdots + \frac{q!}{q!}}_{n} + \underbrace{\frac{q!}{(q+1)!} + \frac{q!}{(q+2)!} + \cdots}_{x},$$

where $n \in \mathbb{N}$. We also have

$$x = \frac{1}{q+1} + \frac{1}{(q+1)(q+2)} + \cdots .$$

We can bound it by

$$0 < x < \frac{1}{q+1} + \frac{1}{(q+1)^2} + \frac{1}{(q+1)^3} + \cdots = \frac{1}{q+1} \cdot \frac{1}{1 - 1/(q+1)} = \frac{1}{q} < 1.$$

This is a contradiction since $q!e$ must be in $\mathbb{N}$ but it is a sum of an integer n plus a non-integer x. □

5.6 Algebraic numbers

Rational numbers are “nice”, because they can be written as fractions. Irrational numbers are bad. However, some irrational numbers are worse than others. We can further classify some irrational numbers as being transcendental.

Definition 5.6.1 (Algebraic and transcendental numbers)**.** *An algebraic number is a root of a polynomial with integer coefficients (or rational coefficients). A number is transcendental if it is not algebraic.*

Proposition 5.6.1. *All rational numbers are algebraic.*

Proof. Let $x = \frac{p}{q}$, then x is a root of $qx - p = 0$. □

Example 5.6.1. *$\sqrt{2}$ is irrational but algebraic since it is a root of $x^2 - 2 = 0$.*

So do transcendental numbers exist?

Theorem 5.6.1. *(Liouville 1851; Non-examinable) L is transcendental, where*

$$L = \sum_{n=1}^{\infty} \frac{1}{10^{n!}} = 0.11000100\cdots$$

with 1s in the factorial positions.

Proof. Suppose instead that $f(L) = 0$ where $f(x) = a_k x^k + a_{k-1}x^{k-1} + \cdots + a_0$, where $a_i \in \mathbb{Z}$, $a_k \neq 0$.

For any rational p/q, we have

$$f\left(\frac{p}{q}\right) = a_k\left(\frac{p}{q}\right)^k + \cdots + a_0 = \frac{\text{integer}}{q^k}.$$

So if p/q is not a root of f, then $|f(p/q)| \geq q^{-k}$.

For any m, we can write L = first m terms + rest of the terms = $s + t$.

Now consider $|f(s)| = |f(L) - f(s)|$ (since $f(L) = 0$). We have

$$\begin{aligned}
|f(L) - f(s)| &= \left|\sum a_i(L^i - s^i)\right| \\
&\leq \sum |a_i(L^i - s^i)| \\
&= \sum |a_i|(L - s)(L^{i-1} + \cdots + s^{i-1}) \\
&\leq \sum |a_i|(L - s)i, \\
&= (L - s)\sum i|a_i| \\
&= tC
\end{aligned}$$

with $C = \sum i|a_i|$.

Writing s as a fraction, its denominator is at most $10^{m!}$. So $|f(s)| \geq 10^{-k\times m!}$. Combining with the above, we have $tC \geq 10^{-k\times m!}$.

We can bound t by

$$t = \sum_{j=m+1}^{\infty} 10^{-j!} \leq \sum_{\ell=(m+1)!}^{\infty} 10^{-\ell} = \frac{10}{9}10^{-(m+1)!}.$$

So $(10C/9)10^{-(m+1)!} \geq 10^{-k\times m!}$. Pick $m \in \mathbb{N}$ so that $m > k$ and $10^{m!} > \frac{10C}{9}$. This is always possible since both k and $10C/9$ are constants. Then the inequality gives $10^{-(m+1)} \geq 10^{-(k+1)}$, which is a contradiction since $m > k$. □

Theorem 5.6.2. *(Hermite 1873) e is transcendental.*

Theorem 5.6.3. *(Lindermann 1882) π is transcendental.*

CHAPTER 6

Countability

After messing with numbers, we finally get back to sets. Here we are concerned about the sizes of sets. We can count how big a set is by constructing bijections. Two sets have the same number of things if there is a bijection between them. In particular, a set has n things if we can bijection it with $[n] = \{1, 2, 3, \cdots, n\}$.

First first prove a few preliminary properties about bijecting with $[n]$ that should be obviously true.

Lemma 6.0.1. *If $f : [n] \to [n]$ is injective, then f is bijective.*

Proof. Perform induction on n: It is true for $n = 1$. Suppose $n > 1$. Let $j = f(n)$. Define $g : [n] \to [n]$ by

$$g(j) = n, \quad g(n) = j, \quad g(i) = i \text{ otherwise.}$$

Then g is a bijection. So the map $g \circ f$ is injective. It fixes n, i.e. $g \circ f(n) = n$. So the map $h : [n-1] \to [n-1]$ by $h(i) = g \circ f(i)$ is well-defined and injective. So h is surjective. So h is bijective. So $g \circ f$ is bijective. So is f. □

Corollary 6.0.1. *If A is a set and $f : A \to [n]$ and $g : A \to [m]$ are both bijections, then $m = n$.*

Proof. wlog assume $m \geq n$. Let $h : [n] \to [m]$ with $h(i) = i$, which is injective. Then the map $h \circ f \circ g^{-1} : [m] \to [m]$ is injective. Then by the lemma this is surjective. So h must be surjective. So $n \geq m$. Hence $n = m$. □

This shows that we cannot biject a set to two different numbers, or a set cannot have two different sizes!

Definition 6.0.1 (Finite set and cardinality of set)**.** *The set A is finite if there exists a bijection $A \to [n]$ for some $n \in \mathbb{N}_0$. The cardinality or size of A, written as $|A|$, is n. By the above corollary, this is well-defined.*

Lemma 6.0.2. *Let $S \subseteq \mathbb{N}$. Then either S is finite or there is a bijection $g : \mathbb{N} \to S$.*

Proof. If $S \neq \emptyset$, by the well-ordering principle, there is a least element $s_1 \in S$. If $S \setminus \{s_1\} \neq \emptyset$, it has a least element s_2. If $S \setminus \{s_1, s_2\}$ is not empty, there is a least element s_3. If at some point the process stops, then $S = \{s_1, s_2, \cdots, s_n\}$, which is finite. Otherwise, if it goes on forever, the map $g : \mathbb{N} \to S$ given by $g(i) = s_i$ is well-defined and is an injection. It is also a surjection because if $k \in S$, then k is a natural number and there are at most k elements of S less than k. So k will be mapped to s_i for some $i \leq k$. □

Definition 6.0.2 (Countable set)**.** *A set A is countable if A is finite or there is a bijection between A and $\mathbb{N}$. A set A is uncountable if A is not countable.*

This is one possible definition of countability, but there are some (often) more helpful definitions.

Theorem 6.0.1. *The following are equivalent:*

1. *A is countable*
2. *There is an injection from $A \to \mathbb{N}$*
3. *$A = \emptyset$ or there is a surjection from $\mathbb{N} \to A$*

Proof. (i) $\Rightarrow$ (iii): If A is finite, there is a bijection $f : A \to S$ for some $S \subseteq \mathbb{N}$. For all $x \in \mathbb{N}$, if $x \in S$, then map $x \mapsto f^{-1}(x)$. Otherwise, map x to any element of A. This is a surjection since $\forall a \in A$, we have $f(a) \mapsto a$.

(iii) $\Rightarrow$ (ii): If $A \neq \emptyset$ and $f : \mathbb{N} \to A$ is a surjection. Define a map $g : A \to \mathbb{N}$ by $g(a) = \min f^{-1}(\{a\})$, which exists by well-ordering. So g is an injection.

(ii) $\Rightarrow$ (i): If there is an injection $f : A \to \mathbb{N}$, then f gives a bijection between A and $S = f(A) \subseteq \mathbb{N}$. If S is finite, so is A. If S is infinite, there is a bijection g between S and $\mathbb{N}$. So there is a bijection $g \circ f$ between A and $\mathbb{N}$. □

Often, the injection definition is the most helpful.

Proposition 6.0.1. *The integers $\mathbb{Z}$ are countable.*

Proof. The map $f : \mathbb{Z} \to \mathbb{N}$ given by

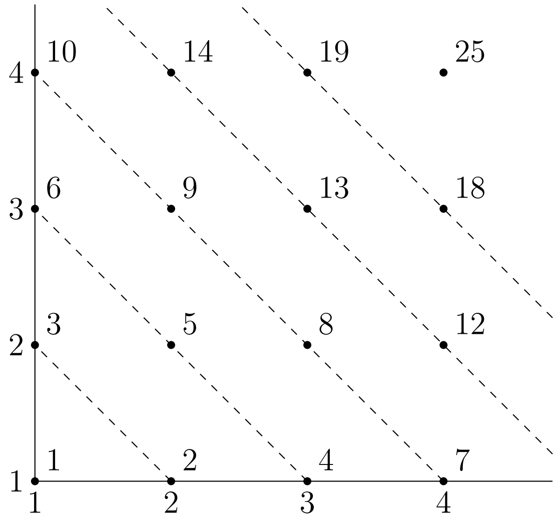
$$f(n) = \begin{cases} 2n & n > 0 \\ 2(-n) + 1 & n \leq 0 \end{cases}$$

is a bijection. □

Proposition 6.0.2. $\mathbb{N} \times \mathbb{N}$ *is countable.*

Proof. We can map $(a, b) \mapsto 2^a 3^b$ injectively by the fundamental theorem of arithmetic. So $\mathbb{N} \times \mathbb{N}$ is countable.

We can also have a bijection by counting diagonally: $(a, b) \mapsto \binom{a+b}{2} - a + 1$:



□

Since $\mathbb{Z}$ is countable, we have an injection $\mathbb{Z} \to \mathbb{N}$, so there is an injection from $\mathbb{Z} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. So $\mathbb{Z} \times \mathbb{N}$ is countable. However, the rationals are the equivalence classes of $\mathbb{Z} \times \mathbb{N}$. So $\mathbb{Q}$ is countable.

Proposition 6.0.3. *If $A \to B$ is injective and B is countable, then A is countable (since we can inject $B \to \mathbb{N}$).*

Proposition 6.0.4. *$\mathbb{Z}^k$ is countable for all $k \in \mathbb{N}$*

Proof. Proof by induction: $\mathbb{Z}$ is countable. If $\mathbb{Z}^k$ is countable, $\mathbb{Z}^{k+1} = \mathbb{Z} \times \mathbb{Z}^k$. Since we can map $\mathbb{Z}^k \to \mathbb{N}$ injectively by the induction hypothesis, we can map injectively $\mathbb{Z}^{k+1} \to \mathbb{Z} \times \mathbb{N}$, and we can map that to $\mathbb{N}$ injectively. □

Theorem 6.0.2. *A countable union of countable sets is countable.*

Proof. Let I be a countable index set, and for each $\alpha \in I$, let A_α be a countable set. We need to show that $\bigcup_{\alpha \in I} A_\alpha$ is countable. It is enough to construct an injection $h : \bigcup_{\alpha \in I} A_\alpha \to \mathbb{N} \times \mathbb{N}$ because $\mathbb{N} \times \mathbb{N}$ is countable. We know that I is countable. So there exists an injection $f : I \to \mathbb{N}$. For each $\alpha \in I$, there exists an injection $g_\alpha : A_\alpha \to \mathbb{N}$.

For $a \in \bigcup A_\alpha$, pick $m = \min\{j \in \mathbb{N} : a \in A_\alpha \text{ and } f(\alpha) = j\}$, and let α be the corresponding index such that $f(\alpha) = m$. We then set $h(a) = (m, g_\alpha(a))$, and this is an injection. □

Proposition 6.0.5. *$\mathbb{Q}$ is countable.*

Proof. It can be proved in two ways:

(i) $\mathbb{Q} = \bigcup_{n\geq 1} \frac{1}{n}\mathbb{Z} = \bigcup_{n\geq 1} \left\{\frac{m}{n} : m \in \mathbb{Z}\right\}$, which is a countable union of countable sets.

(ii) $\mathbb{Q}$ can be mapped injectively to $\mathbb{Z} \times \mathbb{N}$ by $a/b \mapsto (a, b)$, where $b > 0$ and $(a, b) = 1$. □

Theorem 6.0.3. *The set of algebraic numbers is countable.*

Proof. Let $\mathcal{P}_k$ be the set of polynomials of degree k with integer coefficients. Then $a_k x^k + a_{k-1}x^{k-1} + \cdots + a_0 \mapsto (a_k, a_{k-1}, \cdots, a_0)$ is an injection $\mathcal{P}_k \to \mathbb{Z}^{k+1}$. Since $\mathbb{Z}^{k+1}$ is countable, so is $\mathcal{P}_k$.

Let $\mathcal{P}$ be the set of all polynomials with integer coefficients. Then clearly $\mathcal{P} = \bigcup \mathcal{P}_k$. This is a countable union of countable sets. So $\mathcal{P}$ is countable.

For each polynomial $p \in \mathcal{P}$, let R_p be the set of its roots. Then R_p is finite and thus countable. Hence $\bigcup_{p\in\mathcal{P}} R_p$, the set of all algebraic numbers, is countable. □

Theorem 6.0.4. *The set of real numbers $\mathbb{R}$ is uncountable.*

Proof. (Cantor's diagonal argument) Assume $\mathbb{R}$ is countable. Then we can list the reals as $r_1, r_2, r_3 \cdots$ so that every real number is in the list. Write each r_n uniquely in decimal form (i.e. without infinite trailing '9's). List them out vertically:

$$r_1 = n_1 \,.\, d_{11}\, d_{12}\, d_{13}\, d_{14} \cdots$$

$$r_2 = n_2 \,.\, d_{21}\, d_{22}\, d_{23}\, d_{24} \cdots$$

$$r_3 = n_3 \,.\, d_{31}\, d_{32}\, d_{33}\, d_{34} \cdots$$

$$r_4 = n_4 \,.\, d_{41}\, d_{42}\, d_{43}\, d_{44} \cdots$$

Define $r = 0\,.\,d_1\,d_2\,d_3\,d_4 \cdots$ by $d_n = \begin{cases} 0 & d_{nn} \neq 0 \\ 1 & d_{nn} = 0 \end{cases}$. Then by construction, this differs from the nth number in the list by the nth digit, and is so different from every number in the list. Then r is a real number but not in the list. Contradiction. □

Corollary 6.0.2. *There are uncountable many transcendental numbers.*

Proof. If not, then the reals, being the union of the transcendentals and algebraic numbers, must be countable. But the reals is uncountable. □

This is an easy but non-constructive proof that transcendental numbers exists. "If we can't find one, find lots!" (it is debatable whether this proof is constructive or not. Some argue that we can use this to construct a transcendental number by listing all the algebraic numbers and perform the diagonal argument to obtain a number not in the list, i.e. a transcendental number. So this is in fact constructive)

Example 6.0.1. *Let $\mathcal{F}_k = \{Y \subseteq \mathbb{N} : |Y| = k\}$, i.e. the set of all subsets of $\mathbb{N}$ of size k. We can inject $\mathcal{F}_k \to \mathbb{Z}^k$ in the obvious way, e.g. $\{1,3,7\} \mapsto (1,3,7)$ etc. So it is countable. So $\mathcal{F} = \bigcup_{k \geq 0} \mathcal{F}_k$, the set of all finite subsets of $\mathbb{N}$ is countable.*

Example 6.0.2. *Recall $\mathcal{P}(X) = \{Y : Y \subseteq X\}$. Now suppose $\mathcal{P}(\mathbb{N})$ is countable. Let $S_1, S_2, S_3, \cdots$ be the list of all subsets of $\mathbb{N}$. Let $S = \{n : n \notin S_n\}$. But then S is not in the list. Contradiction. So $\mathcal{P}(\mathbb{N})$ is uncountable.*

Example 6.0.3. *Let Σ be the set of all functions $\mathbb{N} \to \mathbb{N}$ (i.e. the set of all integer sequences). If Σ were countable, we could list it as $f_1, f_2, f_3 \cdots$. But then consider f given by $f(n) = \begin{cases} 1 & f_n(n) \neq 1 \\ 2 & f_n(n) = 1 \end{cases}$. Again f is not in the list. Contradiction. So Σ is uncountable.*
Alternatively, there is a bijection between $\mathcal{P}(\mathbb{N})$ and the set of 0, 1 sequences by $S \mapsto$ the indicator function. So we can inject $\mathcal{P}(\mathbb{N}) \to \Sigma$ by $S \mapsto$ indicator function $+1$. So Σ cannot be countable (since $\mathcal{P}(\mathbb{N})$ is uncountable).

Or, we can let $\Sigma^* \subseteq \Sigma$ be the set of bijections from $\mathbb{N} \to \mathbb{N}$. Let $\Sigma^{**} \subseteq \Sigma^*$ be the bijections of the special form: for every n,

$$\textit{either} \begin{cases} f(2n-1) = 2n-1 \\ f(2n) = 2n \end{cases}, \textit{ or } \begin{cases} f(2n-1) = 2n \\ f(2n) = 2n-1 \end{cases},$$

i.e. for every odd-even pair, we either flip them or keep them the same.

But there is a bijection between Σ^{**} and $0, 1$ sequences: if the nth term in the sequence $= 0$, don't flip the nth pair in the function, vice versa. Hence Σ^{**} is uncountable.

Theorem 6.0.5. *Let A be a set. Then there is no surjection from $A \to \mathcal{P}(A)$.*

Proof. Suppose $f : A \to \mathcal{P}(A)$ is surjective. Let $S = \{a \in A : a \notin f(a)\}$. Since f is surjective, there must exist $s \in A$ such that $f(s) = S$. If $s \in S$, then $s \notin S$ by the definition of S. Conversely, if $s \notin S$, then $s \in S$. Contradiction. So f cannot exist. □

This shows that there are infinitely many different possible "infinite sizes" of sets.

We conclude by two theorems that we will not prove.

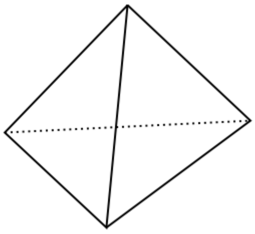
Theorem 6.0.6 (Cantor-Schröder-Bernstein theorem)**.** *Suppose there are injections $A \to B$ and $B \to A$. Then there's a bijection $A \leftrightarrow B$.*

Continuum hypothesis. There is no set whose size lies between $\mathbb{N}$ and $\mathbb{R}$. In 1963, Paul Cohen proved that it is impossible to prove this or disprove this statement.

CHAPTER 7

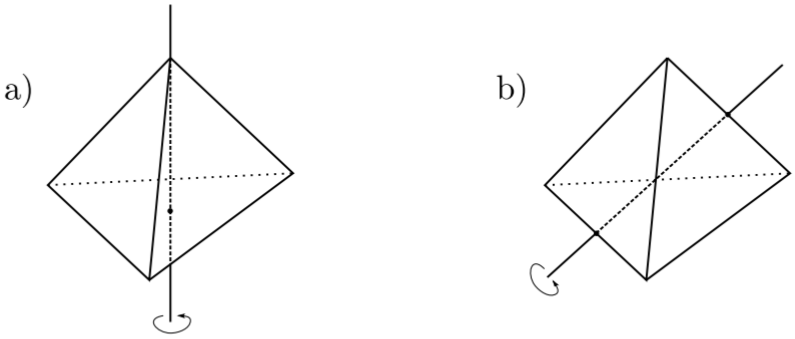
Groups and homomorphisms

Group theory is the study of symmetry, and is best begun by considering some symmetries. First lets consider those simmetries of the following shape, a regular tetrahedron, which are given by rotations of 3-dimensional space. The most trivial kind of rotational symmetry is to do nothing at all: that is, rotate by 0°.

$T =$ 

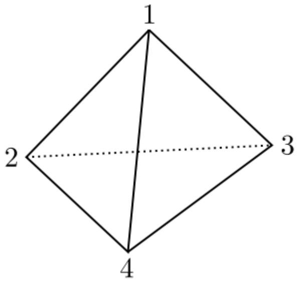
Firstly we can rotate around an axis passing through a vertex and the middle of a face, as in Figure a), and here we can rotate either by 120° or 240°. We can also rotate by 360°, but this leaves every point on the tetrahedron back where it started: it is the same as the "do nothing" rotation.

Secondly we can rotate around an axis passing through the midpoints of two opposite



edges, as in Figure *b*), and here we can rotate by 180○. We can convince ourselves that there are no more rotations; how many have we found? For rotations of type a) there are 4 possible axes, which have 2 rotations (by 120° and 240°) each, so 8 rotations. For rotations of type *b*) there are 3 possible axes, which have 1 rotation each, so 3 rotations. Adding to these the "do nothing" rotation, and we find that T has 12 rotational symmetries

It is not obvious but if we do one rotation and then do another, maybe about a different axis, the result is again a rotation (we will prove this later in the course). Lets see an example of this, for which it is useful to number the vertices of the tetrahedron. Let



us write r for the rotation by 120° about the axis in Figure *a*), and s for the rotation by 180° about the axis in Figure *b*). The rotation r moves the vertices as follows

$$1 \mapsto 1$$

$$2 \mapsto 4$$

$$3 \mapsto 2$$

$$4 \mapsto 3$$

and the rotation s moves the vertices as follows

$$1 \mapsto 3$$
$$2 \mapsto 4$$
$$3 \mapsto 1$$
$$4 \mapsto 2$$

If we do r first and then do s, the vertices are moved as

$$1 \mapsto 1 \mapsto 3$$
$$2 \mapsto 4 \mapsto 2$$
$$3 \mapsto 2 \mapsto 4$$
$$4 \mapsto 3 \mapsto 1$$

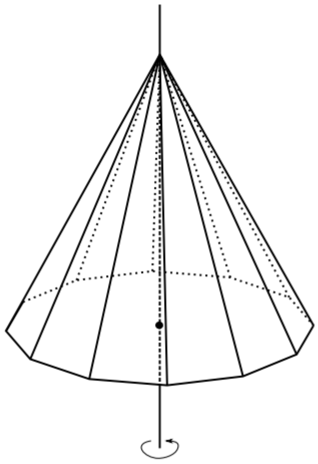
and some thinking shows that this is a rotation about the axis going through the vertex called “2” and the middle of the opposite face. On the other hand if we do s first and then do r, the vertices are moved as

$$1 \mapsto 3 \mapsto 2$$
$$2 \mapsto 4 \mapsto 3$$
$$3 \mapsto 1 \mapsto 1$$
$$4 \mapsto 2 \mapsto 4$$

and this is a rotation about the axis going through the vertex called “4” and the middle of the opposite face. So it is also a rotation, but is a different rotation to the one we got by doing r first and then s!

Now lets consider a new geometrical shape, given by the cone on a dodecagon. The symmetries of C are a bit easier to classify: we may rotate by any multiple of $\frac{360^\circ}{12} = 30^\circ$ around the vertical axis, including the “do nothing” rotation, so this shape also has 12 symmetries about the vertical axis.

Both T and C then have the same number of symmetries, 12, but I am sure that you agree that they do not have the same kinds of symmetries. There are several ways of describing the differences; here are two:

$C =$ 

(i) All symmetries of T have the property that if we repeat them twice—for those as in Figure b), or three times, for those as in Figure a), then we get the "do nothing" symmetry. But if we repeat the "rotate by 30°" symmetry of C two or three times we get rotation by 60° or 90°, neither of which do nothing.

(ii) If we rotate C by $30a°$ and then by $30b°$ then that means rotating it by $30(a+b)°$. This is the same as rotating it by $30(b+a)°$, so is the same as rotating by $30b°$ and then by $30a°$. So the order in which we apply symmetries of C does not matter, whereas we saw that it does matter for T (doing r then s is not the same as doing s then r).

Group theory will allow us to understand these kinds of phenomena. In fact group theory will give us a language, and tools, to describe symmetries of any mathematical object, not just geometrical shapes as we have discussed here.

7.1 Groups

Definition 7.1.1 (Binary operation). *A binary operation is a way of combining two elements to get a new element. Formally, it is a map* $* : A \times A \to A$*.*

Definition 7.1.2 (Group). *A group is a set* G *with a binary operation* $*$ *satisfying the following axioms:*

(i) There is some $e \in G$ such that for all a, we have

$$a * e = e * a = a. \tag{identity}$$

(ii) For all $a \in G$, there is some $a^{-1} \in G$ such that

$$a * a^{-1} = a^{-1} * a = e. \tag{inverse}$$

(iii) For all $a, b, c \in G$, we have

$$(a * b) * c = a * (b * c). \tag{associativity}$$

Definition 7.1.3 (Order of group)**.** *The order of the group, denoted by $|G|$, is the number of elements in G. A group is a finite group if the order is finite.*

Note that technically, the inverse axiom makes no sense, since we have not specified what e is. Even if we take it to be the e given by the identity axiom, the identity axiom only states there is some e that satisfies that property, but there could be many! We don't know which one $a * a^{-1}$ is supposed to be equal to! So we should technically take that to mean there is some a^{-1} such that $a * a^{-1}$ and $a^{-1} * a$ satisfy the identity axiom. Of course, we will soon show that identities are indeed unique, and we will happily talk about "the" identity.

Some people put a zeroth axiom called "closure":

() For all $a, b \in G$, we have $a * b \in G$. (closure)

Technically speaking, this axiom also makes no sense — when we say $*$ is a binary operation, by definition, $a * b$ must be a member of G. However, in practice, we often have to check that this axiom actually holds. For example, if we let G be the set of all matrices of the form

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}$$

under matrix multiplication, we will have to check that the product of two such matrices is indeed a matrix of this form. Officially, we are checking that the binary operation is a well-defined operation on G.

It is important to know that it is generally not true that $a*b = b*a$. There is no a priori reason why this should be true. For example, if we are considering the symmetries of a triangle, rotating and then reflecting is different from reflecting and then rotating.

However, for some groups, this happens to be true. We call such groups abelian groups.

Definition 7.1.4 (Abelian group)**.** *A group is abelian if it satisfies*

(iv) $(\forall a, b \in G)\, a * b = b * a$. *(commutativity)*

If it is clear from context, we are lazy and leave out the operation $*$, and write $a * b$ as ab. We also write $a^2 = aa$, $a^n = \underbrace{aaa \cdots a}_{n \text{ copies}}$, $a^0 = e$, $a^{-n} = (a^{-1})^n$ etc.

Example 7.1.1. *The following are abelian groups:*

(i) $\mathbb{Z}$ *with* $+$

(ii) $\mathbb{Q}$ *with* $+$

(iii) $\mathbb{Z}_n$ *(integers mod n) with* $+_n$

(iv) $\mathbb{Q}^*$ *with* $\times$

(v) $\{-1, 1\}$ *with* $\times$

The following are non-abelian groups:

(i) Symmetries of an equilateral triangle (or any n-gon) with composition. (D_{2n})

(ii) 2×2 *invertible matrices with matrix multiplication* $(\mathrm{GL}_2(\mathbb{R}))$

(iii) Symmetry groups of 3D objects

Recall that the first group axiom requires that there exists an identity element, which we shall call e. Then the second requires that for each a, there is an inverse a^{-1} such that $a^{-1}a = e$. This only makes sense if there is only one identity e, or else which identity should $a^{-1}a$ be equal to?

We shall now show that there can only be one identity. It turns out that the inverses are also unique. So we will talk about the identity and the inverse.

Proposition 7.1.1. *Let $(G, *)$ be a group. Then*

(i) The identity is unique.

(ii) Inverses are unique.

Proof. (i) Suppose e and e' are identities. Then we have $ee' = e'$, treating e as an inverse, and $ee' = e$, treating e' as an inverse. Thus $e = e'$.

(ii) Suppose a^{-1} and b both satisfy the inverse axiom for some $a \in G$. Then $b = be = b(aa^{-1}) = (ba)a^{-1} = ea^{-1} = a^{-1}$. Thus $b = a^{-1}$.

□

Proposition 7.1.1. *Let $(G, *)$ be a group and $a, b \in G$. Then*

(i) $(a^{-1})^{-1} = a$

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

Proof. 1. Given a^{-1}, both a and $(a^{-1})^{-1}$ satisfy

$$xa^{-1} = a^{-1}x = e.$$

By uniqueness of inverses, $(a^{-1})^{-1} = a$.

2. We have

$$\begin{aligned}(ab)(b^{-1}a^{-1}) &= a(bb^{-1})a^{-1} \\ &= aea^{-1} \\ &= aa^{-1} \\ &= e\end{aligned}$$

Similarly, $(b^{-1}a^{-1})ab = e$. So $b^{-1}a^{-1}$ is an inverse of ab. By the uniqueness of inverses, $(ab)^{-1} = b^{-1}a^{-1}$.

□

Sometimes if we have a group G, we might want to discard some of the elements. For example if G is the group of all symmetries of a triangle, we might one day decide that we hate reflections because they reverse orientation. So we only pick the rotations in G and form a new, smaller group. We call this a *subgroup* of G.

Definition 7.1.1 (Subgroup). *A H is a* subgroup *of G, written $H \leq G$, if $H \subseteq G$ and H with the restricted operation $*$ from G is also a group.*

Exercise 7.1.1.

- $(\mathbb{Z}, +) \leq (\mathbb{Q}, +) \leq (\mathbb{R}, +) \leq (\mathbb{C}, +)$
- $(e, *) \leq (G, *)$ *(trivial subgroup)*
- $G \leq G$
- $(\{\pm 1\}, \times) \leq (\mathbb{Q}^*, \times)$

According to the definition, to prove that H is a subgroup of G, we need to make sure H satisfies all group axioms. However, this is often tedious. Instead, there are some simplified criteria to decide whether H is a subgroup.

Lemma 7.1.1 (Subgroup criteria I). *Let $(G, *)$ be a group and $H \subseteq G$. $H \leq G$ iff*

1. $e \in H$
2. $(\forall a, b \in H)\, ab \in H$
3. $(\forall a \in H)\, a^{-1} \in H$

Proof. The group axioms are satisfied as follows:

0. Closure: (ii)
1. Identity: (i). Note that H and G must have the same identity. Suppose that e_H and e_G are the identities of H and G respectively. Then $e_H e_H = e_H$. Now e_H has an inverse in G. Thus we have $e_H e_H e_H^{-1} = e_H e_H^{-1}$. So $e_H e_G = e_G$. Thus $e_H = e_G$.
2. Inverse: (iii)
3. Associativity: inherited from G. □

Humans are lazy, and the test above is still too complicated. We thus come up with an even simpler test:

Lemma 7.1.2 (Subgroup criteria II)**.** *A subset $H \subseteq G$ is a subgroup of G iff:*

(I) H is non-empty

(II) $(\forall a, b \in H)\, ab^{-1} \in H$

Proof. (I) and (II) follow trivially from (i), (ii) and (iii).

To prove that (I) and (II) imply (i), (ii) and (iii), we have

1. H must contain at least one element a. Then $aa^{-1} = e \in H$.

3. $ea^{-1} = a^{-1} \in H$.

2. $a(b^{-1})^{-1} = ab \in H$.

□

Proposition 7.1.2. *The subgroups of $(\mathbb{Z}, +)$ are exactly $n\mathbb{Z}$, for $n \in \mathbb{N}$ ($n\mathbb{Z}$ is the integer multiples of n).*

Proof. Firstly, it is trivial to show that for any $n \in \mathbb{N}$, $n\mathbb{Z}$ is a subgroup. Now show that any subgroup must be in the form $n\mathbb{Z}$.

Let $H \leq \mathbb{Z}$. We know $0 \in H$. If there are no other elements in H, then $H = 0\mathbb{Z}$. Otherwise, pick the smallest positive integer n in H. Then $H = n\mathbb{Z}$.

Otherwise, suppose $(\exists a \in H)\, n \nmid a$. Let $a = pn + q$, where $0 < q < n$. Since $a - pn \in H$, $q \in H$. Yet $q < n$ but n is the smallest member of H. Contradiction. So every $a \in H$ is divisible by n. Also, by closure, all multiples of n must be in H. So $H = n\mathbb{Z}$. □

7.2 Homomorphisms

It is often helpful to study functions between different groups. First, we need to define what a function is. These definitions should be familiar from IA Numbers and Sets.

Definition 7.2.1 (Function)**.** *Given two sets X, Y, a* function *$f : X \to Y$ sends each $x \in X$ to a particular $f(x) \in Y$. X is called the domain and Y is the* *co-domain.*

Exercise 7.2.1.

- *Identity function: for any set X, $1_X : X \to X$ with $1_X(x) = x$ is a function. This is also written as* id_X.
- *Inclusion map: $\iota : \mathbb{Z} \to \mathbb{Q}$: $\iota(n) = n$. Note that this differs from the identity function as the domain and codomain are different in the inclusion map.*
- $f_1 : \mathbb{Z} \to \mathbb{Z}$: $f_1(x) = x + 1$.
- $f_2 : \mathbb{Z} \to \mathbb{Z}$: $f_2(x) = 2x$.
- $f_3 : \mathbb{Z} \to \mathbb{Z}$: $f_3(x) = x^2$.
- *For $g : \{0, 1, 2, 3, 4\} \to \{0, 1, 2, 3, 4\}$, we have:*
  - $g_1(x) = x + 1$ *if* $x < 4$; $g_1(4) = 4$.
  - $g_2(x) = x + 1$ *if* $x < 4$; $g_1(4) = 0$.

Definition 7.2.2 (Composition of functions)**.** *The* composition *of two functions is a function you get by applying one after another. In particular, if $f : X \to Y$ and $G : Y \to Z$, then $g \circ f : X \to Z$ with $g \circ f(x) = g(f(x))$.*

Exercise 7.2.2. $f_2 \circ f_1(x) = 2x + 2$. $f_1 \circ f_2(x) = 2x + 1$. *Note that function composition is not commutative.*

Definition 7.2.3 (Injective functions)**.** *A function f is* injective *if it hits everything at most once, i.e.*

$$(\forall x, y \in X)\, f(x) = f(y) \Rightarrow x = y.$$

Definition 7.2.4 (Surjective functions)**.** *A function is* surjective *if it hits everything at least once, i.e.*

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

Definition 7.2.5 (Bijective functions)**.** *A function is* bijective *if it is both injective and surjective. i.e. it hits everything exactly once. Note that a function has an inverse iff it is bijective.*

Exercise 7.2.3. *ι and f_2 are injective but not subjective. f_3 and g_1 are neither. 1_X, f_1 and g_2 are bijective.*

Lemma 7.2.1. *The composition of two bijective functions is bijective*

When considering sets, functions are allowed to do all sorts of crazy things, and can send any element to any element without any restrictions. However, we are currently studying groups, and groups have additional structure on top of the set of elements. Hence we are not interested in arbitrary functions. Instead, we are interested in functions that "respect" the group structure. We call these *homomorphisms.*

Definition 7.2.6 (Group homomorphism)**.** *Let $(G, *)$ and $(H, \times)$ be groups. A function $f : G \to H$ is a* group homomorphism *iff*

$$(\forall g_1, g_2 \in G)\, f(g_1) \times f(g_2) = f(g_1 * g_2),$$

Definition 7.2.7 (Group isomorphism)**.** Isomorphisms *are bijective homomorphisms. Two groups are* isomorphic *if there exists an isomorphism between them. We write $G \cong H$.*

We will consider two isomorphic groups to be "the same". For example, when we say that there is only one group of order 2, it means that any two groups of order 2 must be isomorphic.

Exercise 7.2.4.

- *$f : G \to H$ defined by $f(g) = e$, where e is the identity of H, is a homomorphism.*
- *$1_G : G \to G$ and $f_2 : \mathbb{Z} \to 2\mathbb{Z}$ are isomorphisms. $\iota : \mathbb{Z} \to \mathbb{Q}$ and $f_2 : \mathbb{Z} \to \mathbb{Z}$ are homomorphisms.*

- $\exp : (\mathbb{R}, +) \to (\mathbb{R}^+, \times)$ *with* $\exp(x) = e^x$ *is an isomorphism.*
- *Take* $(\mathbb{Z}_4, +)$ *and* $H : (\{e^{ik\pi/2} : k = 0, 1, 2, 3\}, \times)$. *Then* $f : \mathbb{Z}_4 \to H$ *by* $f(a) = e^{i\pi a/2}$ *is an isomorphism.*
- $f : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^*$ *with* $f(A) = \det(A)$ *is a homomorphism, where* $\mathrm{GL}_2(\mathbb{R})$ *is the set of* 2×2 *invertible matrices.*

Proposition 7.2.1. *Suppose that* $f : G \to H$ *is a homomorphism. Then*

1. *Homomorphisms send the identity to the identity, i.e.*

$$f(e_G) = e_H$$

2. *Homomorphisms send inverses to inverses, i.e.*

$$f(a^{-1}) = f(a)^{-1}$$

3. *The composite of 2 group homomorphisms is a group homomorphism.*
4. *The inverse of an isomorphism is an isomorphism.*

Proof.

1.

$$\begin{aligned} f(e_G) &= f(e_G^2) = f(e_G)^2 \\ f(e_G)^{-1} f(e_G) &= f(e_G)^{-1} f(e_G)^2 \\ f(e_G) &= e_H \end{aligned}$$

2.

$$\begin{aligned} e_H &= f(e_G) \\ &= f(aa^{-1}) \\ &= f(a)f(a^{-1}) \end{aligned}$$

   Since inverses are unique, $f(a^{-1}) = f(a)^{-1}$.

3. Let $f : G_1 \to G_2$ and $g : G_2 \to G_3$. Then $g(f(ab)) = g(f(a)f(b)) = g(f(a))g(f(b))$.

4. Let $f : G \to H$ be an isomorphism. Then

$$\begin{aligned} f^{-1}(ab) &= f^{-1}\Big\{f\left[f^{-1}(a)\right]f\left[f^{-1}(b)\right]\Big\} \\ &= f^{-1}\Big\{f\left[f^{-1}(a)f^{-1}(b)\right]\Big\} \\ &= f^{-1}(a)f^{-1}(b) \end{aligned}$$

So f^{-1} is a homomorphism. Since it is bijective, f^{-1} is an isomorphism. □

Definition 7.2.8 (Image of homomorphism)**.** *If $f : G \to H$ is a homomorphism, then the* image *of f is*

$$\operatorname{im} f = f(G) = \{f(g) : g \in G\}.$$

Definition 7.2.9 (Kernel of homomorphism)**.** *The* kernel *of f, written as*

$$\ker f = f^{-1}(\{e_H\}) = \{g \in G : f(g) = e_H\}.$$

Proposition 7.2.2. *Both the image and the kernel are subgroups of the respective groups, i.e.* $\operatorname{im} f \leq H$ *and* $\ker f \leq G$.

Proof. Since $e_H \in \operatorname{im} f$ and $e_G \in \ker f$, $\operatorname{im} f$ and $\ker f$ are non-empty. Moreover, suppose $b_1, b_2 \in \operatorname{im} f$. Now $\exists a_1, a_2 \in G$ such that $f(a_i) = b_i$. Then $b_1 b_2^{-1} = f(a_1)f(a_2^{-1}) = f(a_1 a_2^{-1}) \in \operatorname{im} f$.

Then consider $b_1, b_2 \in \ker f$. We have $f(b_1 b_2^{-1}) = f(b_1)f(b_2)^{-1} = e^2 = e$. So $b_1 b_2^{-1} \in \ker f$. □

Proposition 7.2.3. *Given any homomorphism $f : G \to H$ and any $a \in G$, for all $k \in \ker f$, $aka^{-1} \in \ker f$.*

This proposition seems rather pointless. However, it is not. All subgroups that satisfy this property are known as *normal subgroups*, and normal subgroups have very important properties. We will postpone the discussion of normal subgroups to later lectures.

Proof. $f(aka^{-1}) = f(a)f(k)f(a)^{-1} = f(a)ef(a)^{-1} = e$. So $aka^{-1} \in \ker f$. □

Exercise 7.2.5. *Images and kernels for previously defined functions:*

1. *For the function that sends everything to* e*,* $\operatorname{im} f = \{e\}$ *and* $\ker f = G$*.*
2. *For the identity function,* $\operatorname{im} 1_G = G$ *and* $\ker 1_G = \{e\}$*.*
3. *For the inclusion map* $\iota : \mathbb{Z} \to \mathbb{Q}$*, we have* $\operatorname{im} \iota = \mathbb{Z}$ *and* $\ker \iota = \{0\}$
4. *For* $f_2 : \mathbb{Z} \to \mathbb{Z}$ *and* $f_2(x) = 2x$*, we have* $\operatorname{im} f_2 = 2\mathbb{Z}$ *and* $\ker f_2 = \{0\}$*.*
5. *For* $\det : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^*$*, we have* $\operatorname{im} \det = \mathbb{R}^*$ *and* $\ker \det = \{A : \det A = 1\} = \mathrm{SL}_2(\mathbb{R})$

Proposition 7.2.4. *For all homomorphisms* $f : G \to H$*,* f *is*

1. *surjective iff* $\operatorname{im} f = H$
2. *injective iff* $\ker f = \{e\}$

Proof.

1. By definition.
2. We know that $f(e) = e$. So if f is injective, then by definition $\ker f = \{e\}$. If $\ker f = \{e\}$, then given a, b such that $f(a) = f(b)$, $f(ab^{-1}) = f(a)f(b)^{-1} = e$. Thus $ab^{-1} \in \ker f = \{e\}$. Then $ab^{-1} = e$ and $a = b$. □

So far, the definitions of images and kernels seem to be just convenient terminology to refer to things. However, we will later prove an important theorem, the *first isomorphism theorem*, that relates these two objects and provides deep insights (hopefully).

Before we get to that, we will first study some interesting classes of groups and develop some necessary theory.

7.3 Cyclic groups

The simplest class of groups is *cyclic groups*. A cyclic group is a group of the form $\{e, a, a^2, a^2, \cdots, a^{n-1}\}$, where $a^n = e$. For example, if we consider the group of all rotations of a triangle, and write $r =$ rotation by $120°$, the elements will be $\{e, r, r^2\}$ with $r^3 = e$.

Officially, we define a cyclic group as follows:

Definition 7.3.1 (Cyclic group C_n)**.** *A group G is* cyclic *if*

$$(\exists a)(\forall b)(\exists n \in \mathbb{Z})\, b = a^n,$$

i.e. every element is some power of a. Such an a is called a generator of G. We write C_n for the cyclic group of order n.

Exercise 7.3.1.

1. *$\mathbb{Z}$ is cyclic with generator* 1 *or* -1*. It is* the *infinite cyclic group.*
2. $(\{+1, -1\}, \times)$ *is cyclic with generator* -1*.*
3. $(\mathbb{Z}_n, +)$ *is cyclic with all numbers coprime with n as generators.*

Notation 7.3.1. *Given a group G and $a \in G$, we write $\langle a \rangle$ for the cyclic group generated by a, i.e. the subgroup of all powers of a. It is the smallest subgroup containing a.*

Definition 7.3.2 (Order of element)**.** *The* order *of an element a is the smallest integer n such that $a^n = e$. If n doesn't exist, a has infinite order. Write* $\text{ord}(a)$ *for the order of a.*

We have given two different meanings to the word "order". One is the order of a group and the other is the order of an element. Since mathematicians are usually (but not always) sensible, the name wouldn't be used twice if they weren't related. In fact, we have

Lemma 7.3.1. *For a in g,* $\text{ord}(a) = |\langle a \rangle|$.

Proof. If $\text{ord}(a) = \infty$, $a^n \neq a^m$ for all $n \neq m$. Otherwise $a^{m-n} = e$. Thus $|\langle a \rangle| = \infty = \text{ord}(a)$.

Otherwise, suppose $\text{ord}(a) = k$. Thus $a^k = e$. We now claim that $\langle a \rangle = \{e, a^1, a^2, \cdots a^{k-1}\}$. Note that $\langle a \rangle$ does not contain higher powers of a as $a^k = e$ and higher powers will loop back to existing elements. There are also no repeating elements in the list provided since $a^m = a^n \Rightarrow a^{m-n} = e$. So done. □

It is trivial to show that

Proposition 7.3.1. *Cyclic groups are abelian.*

Definition 7.3.3 (Exponent of group)**.** *The* exponent *of a group G is the smallest integer n such that $a^n = e$ for all $a \in G$.*

7.4 Dihedral groups

Definition 7.4.1 (Dihedral groups D_{2n})**.** *Dihedral groups are the symmetries of a regular n-gon. It contains n rotations (including the identity symmetry, i.e. rotation by $0°$) and n reflections.*
We write the group as D_{2n}. Note that the subscript refers to the order of the group, not the number of sides of the polygon.

The dihedral group is not hard to define. However, we need to come up with a presentation of D_{2n} that is easy to work with.

We first look at the rotations. The set of all rotations is generated by $r = \frac{360°}{n}$. This r has order n.

How about the reflections? We know that each reflection has order 2. Let s be our favorite reflection. Then using some geometric arguments, we can show that any reflection can be written as a product of r^m and s for some m. We also have $srs = r^{-1}$.

Hence we can define D_{2n} as follows: D_{2n} is a group generated by r and s, and every element can be written as a product of r's and s's. Whenever we see r^n and s^2, we replace it by e. When we see srs, we replace it by r^{-1}.

It then follows that every element can be written in the form $r^m s$.

Formally, we can write D_{2n} as follows:

$$\begin{aligned} D_{2n} &= \langle r, s \mid r^n = s^2 = e, srs^{-1} = r^{-1} \rangle \\ &= \{e, r, r^2, \cdots r^{n-1}, s, rs, r^2 s, \cdots r^{n-1} s\} \end{aligned}$$

This is a notation we will commonly use to represent groups. For example, a cyclic group of order n can be written as

$$C_n = \langle a \mid a^n = e \rangle.$$

7.5 Direct products of groups

Recall that if we have to sets X, Y, then we can obtain the product $X \times Y = \{(x, y) : x \in X, y \in Y\}$. We can do the same if X and Y are groups.

Definition 7.5.1 (Direct product of groups)**.** *Given two groups $(G, \circ)$ and $(H, \bullet)$, we can define a set $G \times H = \{(g, h) : g \in G, h \in H\}$ and an operation $(a_1, a_2) * (b_1, b_2) = (a_1 \circ b_1, a_2 \bullet b_2)$. This forms a group.*

Why would we want to take the product of two groups? Suppose we have two independent triangles. Then the symmetries of this system include, say rotating the first triangle, rotating the second, or rotating both. The symmetry group of this combined system would then be $D_6 \times D_6$.

Exercise 7.5.1.

$$\begin{aligned} C_2 \times C_2 &= \{(0,0), (0,1), (1,0), (1,1)\} \\ &= \{e, x, y, xy\} \textit{ with everything order 2} \\ &= \langle x, y \mid x^2 = y^2 = e, xy = yx \rangle \end{aligned}$$

Proposition 7.5.1. *$C_n \times C_m \cong C_{nm}$ iff $\mathrm{hcf}(m, n) = 1$.*

Proof. Suppose that $\mathrm{hcf}(m, n) = 1$. Let $C_n = \langle a \rangle$ and $C_m = \langle b \rangle$. Let k be the order of (a, b). Then $(a, b)^k = (a^k, b^k) = e$. This is possible only if $n \mid k$ and $m \mid k$, i.e. k is a common multiple n and m. Since the order is the minimum value of k that satisfies the above equation, $k = \mathrm{lcm}(n, m) = \frac{nm}{\mathrm{hcf}(n,m)} = nm$.

Now consider $\langle (a, b) \rangle \leq C_n \times C_m$. Since (a, b) has order nm, $\langle (a, b) \rangle$ has nm elements. Since $C_n \times C_m$ also has nm elements, $\langle (a, b) \rangle$ must be the whole of $C_n \times C_m$. And we know that $\langle(a, b)\rangle \cong C_{nm}$. So $C_n \times C_m \cong C_{nm}$.

On the other hand, suppose $\text{hcf}(m, n) \neq 1$. Then $k = \text{lcm}(m, n) \neq mn$. Then for any $(a, b) \in C_n \times C_m$,we have $(a, b)^k = (a^k, b^k) = e$. So the order of any (a, b) is at most $k < mn$. So there is no element of order mn. So $C_n \times C_m$ is not a cyclic group of order nm. □

Given a complicated group G, it is sometimes helpful to write it as a product $H \times K$, which could make things a bit simpler. We can do so by the following theorem:

Proposition 7.5.2 (Direct product theorem)**.** *Let $H_1, H_2 \leq G$. Suppose the following are true:*

1. *$H_1 \cap H_2 = \{e\}$.*
2. *$(\forall a_i \in H_i)\, a_1a_2 = a_2a_1$.*
3. *$(\forall a \in G)(\exists a_i \in H_i)\, a = a_1a_2$. We also write this as $G = H_1H_2$.*

Then $G \cong H_1 \times H_2$.

Proof. Define $f : H_1 \times H_2 \to G$ by $f(a_1, a_2) = a_1a_2$. Then it is a homomorphism since

$$\begin{aligned} f((a_1, a_2) * (b_1, b_2)) &= f(a_1b_1, a_2b_2) \\ &= a_1b_1a_2b_2 \\ &= a_1a_2b_1b_2 \\ &= f(a_1, a_2)f(b_1, b_2). \end{aligned}$$

Surjectivity follows from (iii). We'll show injectivity by showing that the kernel is $\{e\}$. If $f(a_1, a_2) = e$, then we know that $a_1a_2 = e$. Then $a_1 = a_2^{-1}$. Since $a_1 \in H_1$ and $a_2^{-1} \in H_2$, we have $a_1 = a_2^{-1} \in H_1 \cap H_2 = \{e\}$. Thus $a_1 = a_2 = e$ and $\ker f = \{e\}$. □

CHAPTER 8

Symmetric groups

We will devote two full chapters to the study of symmetric groups, because it is really important. Recall that we defined a symmetry to be an operation that leaves some important property of the object intact. We can treat each such operation as a bijection. For example, a symmetry of $\mathbb{R}^2$ is a bijection $f : \mathbb{R}^2 \to \mathbb{R}^2$ that preserves distances. Note that we must require it to be a bijection, instead of a mere function, since we require each symmetry to be an inverse.

We can consider the case where we don't care about anything at all. So a "symmetry" would be any arbitrary bijection $X \to X$, and the set of all bijections will form a group, known as the *symmetric group*. Of course, we will no longer think of these as "symmetries" anymore, but just bijections.

In some sense, the symmetric group is the most general case of a symmetry group. In fact, we will, in Group Theory (MAT00032I), show that every group can be written as a subgroup of some symmetric group.

8.1 Symmetric groups

Definition 8.1.1 (Permutation). *A* permutation *of X is a bijection from a set X to X itself. The set of all permutations on X is* $\operatorname{Sym} X$.

When composing permutations, we treat them as functions. So if σ and ρ are permutations, $\sigma \circ \rho$ is given by first applying ρ, then applying σ.

Theorem 8.1.1. $\operatorname{Sym} X$ *with composition forms a group.*

Proof. The groups axioms are satisfied as follows:

0. If $\sigma : X \to X$ and $\tau : X \to X$, then $\sigma \circ \tau : X \to X$. If they are both bijections, then the composite is also bijective. So if $\sigma, \tau \in \operatorname{Sym} X$, then $\sigma \circ \tau \in \operatorname{Sym} X$.
1. The identity $1_X : X \to X$ is clearly a permutation, and gives the identity of the group.
2. Every bijective function has a bijective inverse. So if $\sigma \in \operatorname{Sym} X$, then $\sigma^{-1} \in \operatorname{Sym} X$.
3. Composition of functions is associative. □

Definition 8.1.2 (Symmetric group S_n). *If X is finite, say $|X| = n$ (usually use $X = \{1, 2, \cdots, n\}$), we write* $\operatorname{Sym} X = S_n$. *The is the* symmetric group *of degree n.*

It is important to note that the *degree* of the symmetric group is different from the *order* of the symmetric group. For example, S_3 has degree 3 but order 6. In general, the order of S_n is $n!$.

There are two ways to write out an element of the symmetric group. The first is the *two row notation.*

Notation 8.1.1. *(Two row notation) We write $1, 2, 3, \cdots n$ on the top line and their* *images below, e.g.*

$$\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix} \in S_3 \text{ and } \begin{pmatrix}1 & 2 & 3 & 4 & 5\\ 2 & 1 & 3 & 4 & 5\end{pmatrix} \in S_5$$

In general, if $\sigma : X \to X$*, we write*

$$\begin{pmatrix}1 & 2 & 3 & \cdots & n\\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n)\end{pmatrix}$$

Exercise 8.1.1. *For small* n*, we have*

1. *When* $n = 1$*,* $S_n = \left\{\begin{pmatrix}1\\1\end{pmatrix}\right\} = \{e\} \cong C_1$.
2. *When* $n = 2$*,* $S_n = \left\{\begin{pmatrix}1 & 2\\1 & 2\end{pmatrix}, \begin{pmatrix}1 & 2\\2 & 1\end{pmatrix}\right\} \cong C_2$
3. *When* $n = 3$*,*

$$S_n = \left\{\begin{matrix}\begin{pmatrix}1 & 2 & 3\\1 & 2 & 3\end{pmatrix}, & \begin{pmatrix}1 & 2 & 3\\2 & 3 & 1\end{pmatrix}, & \begin{pmatrix}1 & 2 & 3\\3 & 1 & 2\end{pmatrix}\\ \\ \begin{pmatrix}1 & 2 & 3\\2 & 1 & 3\end{pmatrix}, & \begin{pmatrix}1 & 2 & 3\\3 & 2 & 1\end{pmatrix}, & \begin{pmatrix}1 & 2 & 3\\1 & 3 & 2\end{pmatrix}\end{matrix}\right\} \cong D_6.$$

Note that S_3 *is not abelian. Thus* S_n *is not abelian for* $n \geq 3$ *since we can always view* S_3 *as a subgroup of* S_n *by fixing* $4, 5, 6, \cdots n$.

In general, we can view D_{2n} as a subgroup of S_n because each symmetry is a permutation of the corners.

While the two row notation is fully general and can represent any (finite) permutation, it is clumsy to write and wastes a lot of space. It is also very annoying to type using LaTeX. Hence, most of the time, we actually use the cycle notation.

Notation 8.1.2 (Cycle notation)**.** *If a map sends* $1 \mapsto 2$*,* $2 \mapsto 3$*,* $3 \mapsto 1$*, then we write it as a cycle* $(1\ 2\ 3)$*. Alternatively, we can write* $(2\ 3\ 1)$ *or* $(3\ 1\ 2)$*, but by convention, we usually write the smallest number first. We leave out numbers that don't move. So we write* $(1\ 2)$ *instead of* $(1\ 2)(3)$*.*

For more complicated maps, we can write them as products of cycles. For example, in S_4, we can have things like $(1\ 2)(3\ 4)$.

The order of each cycle is the length of the cycle, and the inverse is the cycle written the other way round, e.g. $(1\ 2\ 3)^{-1} = (3\ 2\ 1) = (1\ 3\ 2)$.

Exercise 8.1.2.

1. *Suppose we want to simplify* $(1\ 2\ 3)(1\ 2)$*. Recall that composition is from right to left. So* 1 *gets mapped to* 3 *(*$(1\ 2)$ *maps* 1 *to* 2*, and* $(1\ 2\ 3)$ *further maps it to* 3*). Then* 3 *gets mapped to* 1*.* 2 *is mapped to* 2 *itself. So* $(1\ 2\ 3)(1\ 2) = (1\ 3)(2)$
2. $(1\ 2\ 3\ 4)(1\ 4) = (1)(2\ 3\ 4) = (2\ 3\ 4)$*.*

Definition 8.1.3 (k-cycles and transpositions)**.** *We call* $(a_1\ a_2\ a_3 \cdots a_k)$ *a* k-cycle*.* 2*-cycles are called* transpositions*. Two cycles are* disjoint *if no number appears in both cycles.*

Exercise 8.1.3. $(1\ 2)$ *and* $(3\ 4)$ *are disjoint but* $(1\ 2\ 3)$ *and* $(1\ 2)$ *are not.*

Lemma 8.1.1. *Disjoint cycles commute.*

Proof. If $\sigma, \tau \in S_n$ are disjoint cycles. Consider any n. Show that: $\sigma(\tau(a)) = \tau(\sigma(a))$. If a is in neither of σ and τ, then $\sigma(\tau(a)) = \tau(\sigma(a)) = a$. Otherwise, wlog assume that a is in τ but not in σ. Then $\tau(a) \in \tau$ and thus $\tau(a) \notin \sigma$. Thus $\sigma(a) = a$ and $\sigma(\tau(a)) = \tau(a)$. Therefore we have $\sigma(\tau(a)) = \tau(\sigma(a)) = \tau(a)$. Therefore τ and σ commute. □

In general, non-disjoint cycles may not commute. For example, $(1\ 3)(2\ 3) = (1\ 3\ 2)$ while $(2\ 3)(1\ 3) = (1\ 2\ 3)$.

Theorem 8.1.2. *Any permutation in* S_n *can be written (essentially) uniquely as a product of disjoint cycles. (Essentially unique means unique up to re-ordering of cycles and rotation within cycles, e.g.* $(1\ 2)$ *and* $(2\ 1)$*)*

Proof. Let $\sigma \in S_n$. Start with $(1\ \sigma(1)\ \sigma^2(1)\ \sigma^3(1)\ \cdots)$. As the set $\{1, 2, 3 \cdots n\}$ is finite, for some k, we must have $\sigma^k(1)$ already in the list. If $\sigma^k(1) = \sigma^l(1)$, with $l < k$, then $\sigma^{k-l}(1) = 1$. So all $\sigma^i(1)$ are distinct until we get back to 1. Thus we have the first cycle $(1\ \sigma(1)\ \sigma^2(1)\ \sigma^3(1)\ \cdots\ \sigma^{k-1}(1))$.

Now choose the smallest number that is not yet in a cycle, say j. Repeat to obtain a cycle $(j\ \sigma(j)\ \sigma^2(j)\ \cdots\ \sigma^{l-1}(j))$. Since σ is a bijection, nothing in this cycle can be in previous cycles as well.

Repeat until all $\{1, 2, 3 \cdots n\}$ are exhausted. This is essentially unique because every number j completely determines the whole cycle it belongs to, and whichever number we start with, we'll end up with the same cycle. □

Definition 8.1.4 (Cycle type). *Write a permutation $\sigma \in S_n$ in disjoint cycle notation. The* cycle type *is the list of cycle lengths. This is unique up to re-ordering. We often (but not always) leave out singleton cycles.*

Exercise 8.1.4. $(1\ 2)$ *has cycle type* 2 *(transposition).* $(1\ 2)(3\ 4)$ *has cycle type* $2, 2$ *(double transposition).* $(1\ 2\ 3)(4\ 5)$ *has cycle type* $3, 2$.

Lemma 8.1.2. *For $\sigma \in S_n$, the order of σ is the least common multiple of cycle lengths in the disjoint cycle notation. In particular, a k-cycle has order k.*

Proof. As disjoint cycles commute, we can group together each cycle when we take powers. i.e. if $\sigma = \tau_1\tau_2 \cdots \tau_l$ with τ_i all disjoint cycles, then $\sigma^m = \tau_1^m\tau_2^m \cdots \tau_l^m$.

Now if cycle τ_i has length k_i, then $\tau_i^{k_i} = e$, and $\tau_i^m = e$ iff $k_i \mid m$. To get an m such that $\sigma^m = e$, we need all k_i to divide m. i.e. m is a common multiple of k_i. Since the order is the least possible m such that $\sigma^m = e$, the order is the least common multiple of k_i. □

Exercise 8.1.5. *Any transpositions and double transpositions have order 2.* $(1\ 2\ 3)(4\ 5)$ *has order 6.*

8.2 Sign of permutations

To classify different permutations, we can group different permutations according to their cycle type. While this is a very useful thing to do, it is a rather fine division. In this section, we will assign a "sign" to each permutation, and each permutation can either be odd or even. This high-level classification allows us to separate permutations into two sets, which is also a useful notion.

To define the sign, we first need to write permutations as products of transpositions.

Proposition 8.2.1. *Every permutation is a product of transpositions.*

This is not a deep or mysterious fact. All it says is that you can rearrange things however you want just by swapping two objects at a time.

Proof. As each permutation is a product of disjoint cycles, it suffices to prove that each cycle is a product of transpositions. Consider a cycle $(a_1\ a_2\ a_3\ \cdots\ a_k)$. This is in fact equal to $(a_1\ a_2)(a_2\ a_3)\cdots(a_{k-1}\ a_k)$. Thus a k-cycle can be written as a product of $k-1$ transpositions. □

Note that the product is not unique. For example,

$$(1\ 2\ 3\ 4\ 5) = (1\ 2)(2\ 3)(3\ 4)(4\ 5) = (1\ 2)(2\ 3)(1\ 2)(3\ 4)(1\ 2)(4\ 5).$$

However, the number of terms in the product, mod 2, is always the same.

Theorem 8.2.1. *Writing $\sigma \in S_n$ as a product of transpositions in different ways, σ is either always composed of an even number of transpositions, or always an odd number of transpositions.*

The proof is rather magical.

Proof. Write $\#(\sigma)$ for the number of cycles in disjoint cycle notation, including singleton cycles. So $\#(e) = n$ and $\#((1\ 2)) = n-1$. When we multiply σ by a transposition $\tau = (c\ d)$ (wlog assume $c < d$),

- If c, d are in the same σ-cycle, say, $(c\ a_2\ \cdots\ a_{k-1}\ d\ a_{k+1}\ \cdots a_{k+l})(c\ d) = (c\ a_{k+1}\ a_{k+2}\ \cdots a_{k+l})(d\ a_2\ a_3\ \cdots\ a_{k-1})$. So $\#(\sigma\tau) = \#(\sigma) + 1$.
- If c, d are in different σ-cycles, say

  $(d\ a_2\ a_3\ \cdots\ a_{k-1})(c\ a_{k+1}\ a_{k+2}\ \cdots\ a_{k+l})(c\ d)$

  $= (c\ a_2\ \cdots\ a_{k-1}\ d\ a_{k+1}\ \cdots a_{k+l})(c\ d)(c\ d)$

  $= (c\ a_2\ \cdots\ a_{k-1}\ d\ a_{k+1}\ \cdots a_{k+l})$ and $\#(\sigma\tau) = \#(\sigma) - 1$.

Therefore for any transposition τ, $\#(\sigma\tau) \equiv \#(\sigma) + 1 \pmod 2$.

Now suppose $\sigma = \tau_1 \cdots \tau_l = \tau_1' \cdots \tau_k'$. Since disjoint cycle notation is unique, $\#(\sigma)$ is uniquely determined by σ.

Now we can construct σ by starting with e and multiplying the transpositions one by one. Each time we add a transposition, we increase $\#(\sigma)$ by $1 \pmod 2$. So $\#(\sigma) \equiv \#(e) + l \pmod 2$. Similarly, $\#(\sigma) \equiv \#(e) + k \pmod 2$. So $l \equiv k \pmod 2$. □

Definition 8.2.1 (Sign of permutation)**.** *Viewing $\sigma \in S_n$ as a product of transpositions, $\sigma = \tau_1 \cdots \tau_l$, we call $\operatorname{sgn}(\sigma) = (-1)^l$. If $\operatorname{sgn}(\sigma) = 1$, we call σ an even permutation. If $\operatorname{sgn}(\sigma) = -1$, we call σ an odd permutation.*

While l itself is not well-defined, it is either always odd or always even, and $(-1)^l$ is well-defined.

Theorem 8.2.2. *For $n \geq 2$, $\operatorname{sgn} : S_n \to \{\pm 1\}$ is a surjective group homomorphism.*

Proof. Suppose $\sigma_1 = \tau_1 \cdots \tau_{l_1}$ and $\sigma_2 = \tau_1' \cdots \tau_{l_2}$. Then $\operatorname{sgn}(\sigma_1\sigma_2) = (-1)^{l_1 + l_2} = (-1)^{l_1}(-1)^{l_2} = \operatorname{sgn}(\sigma_1)\operatorname{sgn}(\sigma_2)$. So it is a homomorphism.

It is surjective since $\operatorname{sgn}(e) = 1$ and $\operatorname{sgn}((1\ 2)) = -1$. □

It is this was rather trivial to prove. The hard bit is showing that sgn is well defined. If a question asks you to show that sgn is a well-defined group homomorphism, you *have* to show that it is well-defined.

Lemma 8.2.1. *σ is an even permutation iff the number of cycles of even length is even.*

Proof. A k-cycle can be written as $k-1$ transpositions. Thus an even-length cycle is odd, vice versa.

Since sgn is a group homomorphism, writing σ in disjoint cycle notation, $\sigma = \sigma_1\sigma_2\cdots\sigma_l$, we get $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma_1)\cdots\operatorname{sgn}(\sigma_l)$. Suppose there are m even-length cycles and n odd-length cycles, then $\operatorname{sgn}(\sigma) = (-1)^m 1^n$. This is equal to 1 iff $(-1)^m = 1$, i.e. m is even. □

Rather confusingly, odd length cycles are even, and even length cycles are odd.

Definition 8.2.2 (Alternating group A_n)**.** *The* alternating group A_n *is the kernel of* sgn*, i.e. the even permutations. Since A_n is a kernel of a group homomorphism, $A_n \leq S_n$.*

Among the many uses of the sgn homomorphism, it is used in the definition of the determinant of a matrix: if $A_{n\times n}$ is a square matrix, then

$$\det A = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)}\cdots a_{n\sigma(n)}.$$

Proposition 8.2.2. *Any subgroup of S_n contains either no odd permutations or exactly half.*

Proof. If S_n has at least one odd permutation τ, then there exists a bijection between the odd and even permutations by $\sigma \mapsto \sigma\tau$ (bijection since $\sigma \mapsto \sigma\tau^{-1}$ is a well-defined inverse). So there are as many odd permutations as even permutations. □

After we prove the isomorphism theorem later, we can provide an even shorter proof of this.

CHAPTER 9

Lagrange's theorem

One can model a Rubik's cube with a group, with each possible move corresponding to a group element. Of course, Rubik's cubes of different sizes correspond to different groups.

Suppose I have a $4 \times 4 \times 4$ Rubik's cube, but I want to practice solving a $2 \times 2 \times 2$ Rubik's cube. It is easy. I just have to make sure every time I make a move, I move two layers together. Then I can pretend I am solving a $2 \times 2 \times 2$ cube. This corresponds to picking a particular subgroup of the $4 \times 4 \times 4$ group.

Now what if I have a $3 \times 3 \times 3$ cube? I can still practice solving a $2 \times 2 \times 2$ one. This time, I just look at the corners and pretend that the edges and centers do not exist. Then I am satisfied when the corners are in the right positions, while the centers and edges can be completely scrambled. In this case, we are not taking a subgroup. Instead, we are identifying certain moves together. In particular, we are treating two moves as the same as long as their difference is confined to the centers and edges.

Let G be the $3 \times 3 \times 3$ cube group, and H be the subgroup of G that only permutes the edges and centers. Then for any $a, b \in G$, we think a and b are "the same" if $a^{-1}b \in H$. Then the set of things equivalent to a is $aH = \{ah : h \in H\}$. We call this a *coset*, and the set of cosets form a group.

An immediate question one can ask is: why not $Ha = \{ha : h \in H\}$? In this particular case, the two happen to be the same for all possible a. However, for a general subgroup H, they need not be. We can still define the coset $aH = \{ah : h \in H\}$, but these are less interesting. For example, the set of all $\{aH\}$ will no longer form a group. We will look into these more in-depth in the next chapter. In this chapter, we will first look at results for general cosets. In particular, we will, step by step, prove the things we casually claimed above.

Definition 9.0.1 (Cosets). *Let $H \leq G$ and $a \in G$. Then the set $aH = \{ah : h \in H\}$ is a* left coset *of H and $Ha = \{ha : h \in H\}$ is a* right coset *of H.*

Exercise 9.0.1.

1. *Take $2\mathbb{Z} \leq Z$. Then $6 + 2\mathbb{Z} = \{\text{all even numbers}\} = 0 + 2\mathbb{Z}$. $1 + 2\mathbb{Z} = \{\text{all odd numbers}\} = 17 + 2\mathbb{Z}$.*
2. *Take $G = S_3$, let $H = \langle(1\ 2)\rangle = \{e, (1\ 2)\}$. The left cosets are*

$$\begin{aligned} eH = (1\ 2)H &= \{e, (1\ 2)\} \\ (1\ 3)H = (1\ 2\ 3)H &= \{(1\ 3), (1\ 2\ 3)\} \\ (2\ 3)H = (1\ 3\ 2)H &= \{(2\ 3), (1\ 3\ 2)\} \end{aligned}$$

3. *Take $G = D_6$ (which is isomorphic to S_3). Recall $D_6 = \langle r, s \mid r^3 e = s^2, rs = sr^{-1}\rangle$.Take $H = \langle s\rangle = \{e, s\}$. We have left coset $rH = \{r, rs = sr^{-1}\}$ and the right coset $Hr = \{r, sr\}$. Thus $rH \neq Hr$.*

Proposition 9.0.1. *$aH = bH \Leftrightarrow b^{-1}a \in H$.*

Proof. ($\Rightarrow$) Since $a \in aH$, $a \in bH$. Then $a = bh$ for some $h \in H$. So $b^{-1}a = h \in H$.

($\Leftarrow$). Let $b^{-1}a = h_0$. Then $a = bh_0$. Then $\forall ah \in aH$, we have $ah = b(h_0 h) \in bH$. So $aH \subseteq bH$. Similarly, $bH \subseteq aH$. So $aH = bH$. □

Definition 9.0.2 (Partition)**.** *Let X be a set, and $X_1, \cdots X_n$ be subsets of X. The X_i are called a* partition *of X if $\bigcup X_i = X$ and $X_i \cap X_j = \emptyset$ for $i \neq j$. i.e. every element is in exactly one of X_i.*

Lemma 9.0.1. *The left cosets of a subgroup $H \leq G$ partition G, and every coset has the same size.*

Proof. For each $a \in G$, $a \in aH$. Thus the union of all cosets gives all of G. Now we have to show that for all $a, b \in G$, the cosets aH and bH are either the same or disjoint.

Suppose that aH and bH are not disjoint. Let $ah_1 = bh_2 \in aH \cap bH$. Then $b^{-1}a = h_2h_1^{-1} \in H$. So $aH = bH$.

To show that they each coset has the same size, note that $f : H \to aH$ with $f(h) = ah$ is invertible with inverse $f^{-1}(h) = a^{-1}h$. Thus there exists a bijection between them and they have the same size. □

Definition 9.0.3 (Index of a subgroup)**.** *The* index *of H in G, written $|G : H|$, is the number of left cosets of H in G.*

Theorem 9.0.1 (Lagrange's theorem)**.** *If G is a finite group and H is a subgroup of G, then $|H|$ divides $|G|$. In particular,*

$$|H||G : H| = |G|.$$

Note that the converse is not true. If k divides $|G|$, there is not necessarily a subgroup of order k, e.g. $|A_4| = 12$ but there is no subgroup of order 6. However, we will later see that this is true if k is a prime (cf. Cauchy's theorem).

Proof. Suppose that there are $|G : H|$ left cosets in total. Since the left cosets partition G, and each coset has size $|H|$, we have

$$|H||G : H| = |G|.$$

□

Again, the hard part of this proof is to prove that the left cosets partition G and have the same size. If you are asked to prove Lagrange's theorem in exams, that is what you actually have to prove.

Corollary 9.0.1. *The order of an element divides the order of the group, i.e. for any finite group G and $a \in G$, $\operatorname{ord}(a)$ divides $|G|$.*

Proof. Consider the subgroup generated by a, which has order $\operatorname{ord}(a)$. Then by Lagrange's theorem, $\operatorname{ord}(a)$ divides $|G|$. □

Corollary 9.0.2. *The exponent of a group divides the order of the group, i.e. for any finite group G and $a \in G$, $a^{|G|} = e$.*

Proof. We know that $|G| = k \operatorname{ord}(a)$ for some $k \in \mathbb{N}$. Then $a^{|G|} = (a^{\operatorname{ord}(a)})^k = e^k = e$. □

Corollary 9.0.3. *Groups of prime order are cyclic and are generated by every non-identity element.*

Proof. Say $|G| = p$. If $a \in G$ is not the identity, the subgroup generated by a must have order p since it has to divide p. Thus the subgroup generated by a has the same size as G and they must be equal. Then G must be cyclic since it is equal to the subgroup generated by a. □

A useful way to think about cosets is to view them as equivalence classes. To do so, we need to first define what an equivalence class is.

Definition 9.0.4 (Equivalence relation)**.** *An* equivalence relation $\sim$ *is a relation that is reflexive, symmetric and transitive. i.e.*

1. $(\forall x)\, x \sim x$ *(reflexivity)*
2. $(\forall x, y)\, x \sim y \Rightarrow y \sim x$ *(symmetry)*
3. $(\forall x, y, z)\, [(x \sim y) \wedge (y \sim z) \Rightarrow x \sim z]$ *(transitivity)*

Exercise 9.0.2. *The following relations are equivalence relations:*

1. *Consider $\mathbb{Z}$. The relation $\equiv_n$ defined as $a \equiv_n b \Leftrightarrow n \mid (a - b)$.*
2. *Consider the set (formally: class) of all finite groups. Then "is isomorphic to" is an equivalence relation.*

Definition 9.0.5 (Equivalence class)**.** *Given an equivalence relation $\sim$ on A, the* equivalence class *of a is*

$$[a]_\sim = [a] = \{b \in A : a \sim b\}$$

Proposition 9.0.2. *The equivalence classes form a partition of A.*

Proof. By reflexivity, we have $a \in [a]$. Thus the equivalence classes cover the whole set. We must now show that for all $a, b \in A$, either $[a] = [b]$ or $[a] \cap [b] = \emptyset$.

Suppose $[a] \cap [b] \neq \emptyset$. Then $\exists c \in [a] \cap [b]$. So $a \sim c, b \sim c$. By symmetry, $c \sim b$. By transitivity, we have $a \sim b$. Now for all $b' \in [b]$, we have $b \sim b'$. Thus by transitivity, we have $a \sim b'$. Thus $[b] \subseteq [a]$. Similarly, $[a] \subseteq [b]$ and $[a] = [b]$. □

Lemma 9.0.2. *Given a group G and a subgroup H, define the equivalence relation on G with $a \sim b$ iff $b^{-1}a \in H$. The equivalence classes are the left cosets of H.*

Proof. First show that it is an equivalence relation.

1. Reflexivity: Since $aa^{-1} = e \in H$, $a \sim a$.
2. Symmetry: $a \sim b \Rightarrow b^{-1}a \in H \Rightarrow (b^{-1}a)^{-1} = a^{-1}b \in H \Rightarrow b \sim a$.
3. Transitivity: If $a \sim b$ and $b \sim c$, we have $b^{-1}a, c^{-1}b \in H$. So $c^{-1}bb^{-1}a = c^{-1}a \in H$. So $a \sim c$.

To show that the equivalence classes are the cosets, we have $a \sim b \Leftrightarrow b^{-1}a \in H \Leftrightarrow aH = bH$. □

Exercise 9.0.3. *Consider* $(\mathbb{Z}, +)$*, and for fixed* n*, take the subgroup* $n\mathbb{Z}$*. The cosets are* $0 + H, 1 + H, \cdots (n-1) + H$*. We can write these as* $[0], [1], [2] \cdots [n]$*. To perform arithmetic "mod n", define* $[a] + [b] = [a + b]$*, and* $[a][b] = [ab]$*. We need to check that it is well-defined, i.e. it doesn't depend on the choice of the representative of* $[a]$*.*
If $[a_1] = [a_2]$ *and* $[b_1] = [b_2]$*, then* $a_1 = a_2 + kn$ *and* $b_1 = b_2 + kn$*, then* $a_1 + b_1 = a_2 + b_2 + n(k + l)$ *and* $a_1 b_1 = a_2 b_2 + n(kb_2 + la_2 + kln)$*. So* $[a_1 + b_1] = [a_2 + b_2]$ *and* $[a_1 b_1] = [a_2 b_2]$*.*

We have seen that $(\mathbb{Z}_n, +_n)$ is a group. What happens with multiplication? We can only take elements which have inverses (these are called units, cf. IB Groups, Rings and Modules). Call the set of them $U_n = \{[a] : (a, n) = 1\}$. We'll see these are the units.

Definition 9.0.6 (Euler totient function)**.** *(Euler totient function)* $\phi(n) = |U_n|$*.*

Exercise 9.0.4. *If* p *is a prime,* $\phi(n) = p - 1$*.* $\phi(4) = 2$*.*

Proposition 9.0.3. U_n *is a group under multiplication mod* n*.*

Proof. The operation is well-defined as shown above. To check the axioms:

0. Closure: if a, b are coprime to n, then $a \cdot b$ is also coprime to n. So $[a], [b] \in U_n \Rightarrow [a] \cdot [b] = [a \cdot b] \in U_n$
1. Identity: $[1]$
2. Let $[a] \in U_n$. Consider the map $U_n \to U_n$ with $[c] \mapsto [ac]$. This is injective: if $[ac_1] = [ac_2]$, then n divides $a(c_1 - c_2)$. Since a is coprime to n, n divides $c_1 - c_2$, so $[c_1] = [c_2]$. Since U_n is finite, any injection $(U_n \to U_n)$ is also a surjection. So there exists a c such that $[ac] = [a][c] = 1$. So $[c] = [a]^{-1}$.
3. Associativity (and also commutativity): inherited from $\mathbb{Z}$. □

Theorem 9.0.2 (Fermat-Euler theorem)**.** *Let $n \in N$ and $a \in \mathbb{Z}$ coprime to n. Then*

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

In particular, (Fermat's Little Theorem) if $n = p$ is a prime, then for any a not a multiple of p.

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

Proof. As a is coprime with n, $[a] \in U_n$. Then $[a]^{|U_n|} = [1]$, i.e. $a^{\phi(n)} \equiv 1 \pmod{n}$. □

9.1 Small groups

We will study the structures of certain small groups.

Exercise 9.1.1 (Using Lagrange theorem to find subgroups)**.** *To find subgroups of D_{10}, we know that the subgroups must have size 1, 2, 5 or 10:*

1: $\{e\}$

2: The groups generated by the 5 reflections of order 2

5: The group must be cyclic since it has prime order 5. It is then generated by an element of order 5, i.e. r, r^2, r^3 and r^4. They generate the same group $\langle r \rangle$.

10: D_{10}

As for D_8, subgroups must have order 1, 2, 4 or 8.

1: $\{e\}$

2: 5 elements of order 2*, namely 4 reflections and r^2.*

4: First consider the subgroup isomorphic to C_4, which is $\langle r \rangle$. There are two other non-cyclic group.

8: D_8

Proposition 9.1.1. *Any group of order 4 is either isomorphic to C_4 or $C_2 \times C_2$.*

Proof. Let $|G| = 4$. By Lagrange theorem, possible element orders are 1 (e only), 2 and 4. If there is an element $a \in G$ of order 4, then $G = \langle a \rangle \cong C_4$.

Otherwise all non-identity elements have order 2. Then G must be abelian (For any a, b, $(ab)^2 = 1 \Rightarrow ab = (ab)^{-1} \Rightarrow ab = b^{-1}a^{-1} \Rightarrow ab = ba$). Pick 2 elements of order 2, say $b, c \in G$, then $\langle b \rangle = \{e, b\}$ and $\langle c \rangle = \{e, c\}$. So $\langle b \rangle \cap \langle c \rangle = \{e\}$. As G is abelian, $\langle b \rangle$ and $\langle c \rangle$ commute. We know that $bc = cb$ has order 2 as well, and is the only element of G left. So $G \cong \langle b \rangle \times \langle c \rangle \cong C_2 \times C_2$ by the direct product theorem. □

Proposition 9.1.2. *A group of order* 6 *is either cyclic or dihedral (i.e. is isomorphic to C_6 or D_6). (See proof in next section)*

9.2 Left and right cosets

As $|aH| = |H|$ and similarly $|H| = |Ha|$, left and right cosets have the same size. Are they necessarily the same? We've previously shown that they might *not* be the same. In some other cases, they are.

Exercise 9.2.1.

1. *Take $G = (\mathbb{Z}, +)$ and $H = 2\mathbb{Z}$. We have $0 + 2\mathbb{Z} = 2\mathbb{Z} + 0 =$ even numbers and $1 + 2\mathbb{Z} = 2\mathbb{Z} + 1 =$ odd numbers. Since G is abelian, $aH = Ha$ for all $a, \in G$, $H \leq G$.*
2. *Let $G = D_6 = \langle r, s \mid r^3 = e = s^2, rs = sr^{-1} \rangle$. Let $U = \langle r \rangle$. Since the cosets partition G, so one must be U and the other $sU = \{s, sr = r^2s, sr^2 = rs\} = Us$. So for all $a \in G, aU = Ua$.*
3. *Let $G = D_6$ and take $H = \langle s \rangle$. We have $H = \{e, s\}$, $rH = \{r, rs = sr^{-1}\}$ and $r^2H = \{r^2, r^s\}$; while $H = \{e, s\}, Hr = \{r, sr\}$ and $Hr^2 = \{r^2, sr^2\}$. So the left and right subgroups do not coincide.*

This distinction will become useful in the next chapter.

CHAPTER 10

Quotient groups

Definition 10.0.1 (Normal subgroup)**.** *A subgroup K of G is a* normal subgroup *if*

$$(\forall a \in G)(\forall k \in K)\, aka^{-1} \in K.$$

We write $K \lhd G$. This is equivalent to:

1. *$(\forall a \in G)\, aK = Ka$, i.e. left coset = right coset*
2. *$(\forall a \in G)\, aKa^{-1} = K$ (cf. conjugacy classes)*

From the example last time, $H = \langle s \rangle \leq D_6$ is not a normal subgroup, but $K = \langle r \rangle \lhd D_6$. We know that every group G has at least two normal subgroups $\{e\}$ and G.

Lemma 10.0.1.

1. *Every subgroup of index 2 is normal.*
2. *Any subgroup of an abelian group is normal.*

Proof.

1. If $K \leq G$ has index 2, then there are only two possible cosets K and $G \setminus K$. As $eK = Ke$ and cosets partition G, the other left coset and right coset must be $G \setminus K$. So all left cosets and right cosets are the same.
2. For all $a \in G$ and $k \in K$, we have $aka^{-1} = aa^{-1}k = k \in K$. □

Proposition 10.0.1. *Every kernel is a normal subgroup.*

Proof. Given homomorphism $f : G \to H$ and some $a \in G$, for all $k \in \ker f$, we have $f(aka^{-1}) = f(a)f(k)f(a)^{-1} = f(a)ef(a)^{-1} = e$. Therefore $aka^{-1} \in \ker f$ by definition of the kernel. □

In fact, we will see in the next section that all normal subgroups are kernels of some homomorphism.

Exercise 10.0.1. *Consider $G = D_8$. Let $K = \langle r^2 \rangle$ is normal. Check: Any element of G is either sr^ℓ or r^ℓ for some ℓ. Clearly e satisfies $aka^{-1} \in K$. Now check r^2: For the case of sr^ℓ, we have $sr^\ell r^2 (sr^\ell)^{-1} = sr^\ell r^2 r^{-\ell} s^{-1} = sr^2 s = ssr^{-2} = r^2$. For the case of r^ℓ, $r^\ell r^2 r^{-\ell} = r^2$.*

Proposition 10.0.2. *A group of order 6 is either cyclic or dihedral (i.e. $\cong C_6$ or D_6).*

Proof. Let $|G| = 6$. By Lagrange theorem, possible element orders are $1, 2, 3$ and 6. If there is an $a \in G$ of order 6, then $G = \langle a \rangle \cong C_6$. Otherwise, we can only have elements of orders 2 and 3 other than the identity. If G only has elements of order 2, the order must be a power of 2 by Sheet 1 Q. 8, which is not the case. So there must be an element r of order 3. So $\langle r \rangle \lhd G$ as it has index 2. Now G must also have an element s of order 2 by Sheet 1 Q. 9.

Since $\langle r \rangle$ is normal, we know that $srs^{-1} \in \langle r \rangle$. If $srs^{-1} = e$, then $r = e$, which is not true. If $srs^{-1} = r$, then $sr = rs$ and sr has order 6 (lcm of the orders of s and r), which was ruled out above. Otherwise if $srs^{-1} = r^2 = r^{-1}$, then G is dihedral by definition of the dihedral group. □

10.1 Quotient groups

Proposition 10.1.1. *Let $K \lhd G$. Then the set of (left) cosets of K in G is a group under the operation $aK * bK = (ab)K$.*

Proof. First show that the operation is well-defined. If $aK = a'K$ and $bK = b'K$, we want to show that $aK * bK = a'K * b'K$. We know that $a' = ak_1$ and $b' = bk_2$ for some $k_1, k_2 \in K$. Then $a'b' = ak_1bk_2$. We know that $b^{-1}k_1b \in K$. Let $b^{-1}k_1b = k_3$. Then $k_1b = bk_3$. So $a'b' = abk_3k_2 \in (ab)K$. So picking a different representative of the coset gives the same product.

1. Closure: If aK, bK are cosets, then $(ab)K$ is also a coset
2. Identity: The identity is $eK = K$ (clear from definition)
3. Inverse: The inverse of aK is $a^{-1}K$ (clear from definition)
4. Associativity: Follows from the associativity of G. □

Definition 10.1.1 (Quotient group)**.** *Given a group G and a normal subgroup K, the* quotient group *or* factor group *of G by K, written as G/K, is the set of (left) cosets of K in G under the operation $aK * bK = (ab)K$.*

Note that the *set* of left cosets also exists for non-normal subgroups (abnormal subgroups?), but the group operation above is not well defined.

Exercise 10.1.1.

1. *Take $G = \mathbb{Z}$ and $n\mathbb{Z}$ (which must be normal since G is abelian), the cosets are $k + n\mathbb{Z}$ for $0 \leq k < n$. The quotient group is $\mathbb{Z}_n$. So we can write $\mathbb{Z}/(n\mathbb{Z}) = \mathbb{Z}_n$. In fact these are the only quotient groups of $\mathbb{Z}$ since $n\mathbb{Z}$ are the only subgroups.*

   Note that if G is abelian, G/K is also abelian.
2. *Take $K = \langle r \rangle \lhd D_6$. We have two cosets K and sK. So D_6/K has order 2 and is isomorphic to C_2.*

3. *Take $K = \langle r^2\rangle \lhd D_8$. We know that G/K should have $\frac{8}{2} = 4$ elements. We have $G/K = \{K, rK = r^3K, sK = sr^2K, srK = sr^3K\}$. We see that all elements (except K) has order 2, so $G/K \cong C_2 \times C_2$.*

Note that quotient groups are *not* subgroups of G. They contain different kinds of elements. For example, $\mathbb{Z}/n\mathbb{Z} \cong C_n$ are finite, but all subgroups of $\mathbb{Z}$ infinite.

Exercise 10.1.2. *(Non-example) Consider D_6 with $H = \langle s\rangle$. H is not a normal subgroup. We have $rH * r^2H = r^3H = H$, but $rH = rsH$ and $r^2H = srH$ (by considering the individual elements). So we have $rsH * srH = r^2H \neq H$, and the operation is not well-defined.*

Lemma 10.1.1. *Given $K \lhd G$, the* quotient map *$q : G \to G/K$ with $g \mapsto gK$ is a surjective group homomorphism.*

Proof. $q(ab) = (ab)K = aKbK = q(a)q(b)$. So q is a group homomorphism. Also for all $aK \in G/K$, $q(a) = aK$. So it is surjective. □

Note that the kernel of the quotient map is K itself. So any normal subgroup is a kernel of some homomorphism.

Proposition 10.1.2. *The quotient of a cyclic group is cyclic.*

Proof. Let $G = C_n$ with $H \leq C_n$. We know that H is also cyclic. Say $C_n = \langle c\rangle$ and $H = \langle c^k\rangle \cong C_\ell$, where $k\ell = n$. We have $C_n/H = \{H, cH, c^2H, \cdots c^{k-1}H\} = \langle cH\rangle \cong C_k$. □

10.2 The Isomorphism Theorem

Now we come to the Really Important Theorem$^{\text{TM}}$.

Theorem 10.2.1 (The Isomorphism Theorem)**.** *Let $f : G \to H$ be a group homomorphism with kernel K. Then $K \lhd G$ and $G/K \cong \operatorname{im} f$.*

Proof. We have proved that $K \lhd G$ before. We define a group homomorphism $\theta : G/K \to \operatorname{im} f$ by $\theta(aK) = f(a)$.

First check that this is well-defined: If $a_1K = a_2K$, then $a_2^{-1}a_1 \in K$. So

$$f(a_2)^{-1}f(a_1) = f(a_2^{-1}a_1) = e.$$

So $f(a_1) = f(a_2)$ and $\theta(a_1K) = \theta(a_2K)$.

Now we check that it is a group homomorphism:

$$\theta(aKbK) = \theta(abK) = f(ab) = f(a)f(b) = \theta(aK)\theta(bK).$$

To show that it is injective, suppose $\theta(aK) = \theta(bK)$. Then $f(a) = f(b)$. Hence $f(b)^{-1}f(a) = e$. Hence $b^{-1}a \in K$. So $aK = bK$.

By definition, θ is surjective since $\operatorname{im}\theta = \operatorname{im} f$. So θ gives an isomorphism $G/K \cong \operatorname{im} f \leq H$. □

If f is injective, then the kernel is $\{e\}$, so $G/K \cong G$ and G is isomorphic to a subgroup of H. We can think of f as an inclusion map. If f is surjective, then $\operatorname{im} f = H$. In this case, $G/K \cong H$.

Exercise 10.2.1.

1. *Take* $f : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^*$ *with* $A \mapsto \det A$, $\ker f = \mathrm{SL}_N(\mathbb{R})$. $\operatorname{im} f = \mathbb{R}^*$ *as for all* $\lambda \in \mathbb{R}^*$, $\det \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix} = \lambda$. *So we know that* $\mathrm{GL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{R}) \cong \mathbb{R}^*$.
2. *Define* $\theta : (\mathbb{R}, +) \to (\mathbb{C}*, \times)$ *with* $r \mapsto \exp(2\pi i r)$. *This is a group homomorphism since* $\theta(r+s) = \exp(2\pi i(r+s)) = \exp(2\pi i r)\exp(2\pi i s) = \theta(r)\theta(s)$. *We know that the kernel is* $\mathbb{Z} \lhd \mathbb{R}$. *Clearly the image is the unit circle* $(S_1, \times)$. *So* $\mathbb{R}/\mathbb{Z} \cong (S_1, \times)$.
3. $G = (\mathbb{Z}_p^*, \times)$ *for prime* $p \neq 2$. *We have* $f : G \to G$ *with* $a \mapsto a^2$. *This is a homomorphism since* $(ab)^2 = a^2b^2$ *(*$\mathbb{Z}_p^*$ *is abelian). The kernel is* $\{\pm 1\} = \{1, p-1\}$. *We know that* $\operatorname{im} f \cong G/\ker f$ *with order* $\frac{p-1}{2}$. *These are known*

as quadratic residues.

Lemma 10.2.1. *Any cyclic group is isomorphic to either $\mathbb{Z}$ or $\mathbb{Z}/(n\mathbb{Z})$ for some $n \in \mathbb{N}$.*

Proof. Let $G = \langle c \rangle$. Define $f : \mathbb{Z} \to G$ with $m \mapsto c^m$. This is a group homomorphism since $c^{m_1+m_2} = c^{m_1}c^{m_2}$. f is surjective since G is by definition all c^m for all m. We know that $\ker f \lhd \mathbb{Z}$. We have three possibilities. Either

1. $\ker f = \{e\}$, so F is an isomorphism and $G \cong \mathbb{Z}$; or
2. $\ker f = \mathbb{Z}$, then $G \cong \mathbb{Z}/\mathbb{Z} = \{e\} = C_1$; or
3. $\ker f = n\mathbb{Z}$ (since these are the only proper subgroups of $\mathbb{Z}$), then $G \cong \mathbb{Z}/(n\mathbb{Z})$.

□

Definition 10.2.1 (Simple group)**.** *A group is* simple *if it has no non-trivial proper normal subgroup, i.e. only $\{e\}$ and G are normal subgroups.*

Exercise 10.2.2. *C_p for prime p are simple groups since it has no proper subgroups at all, let alone normal ones. A_5 is simple (non-examinable).*

The finite simple groups are the building blocks of all finite groups. All finite simple groups have been classified (The Atlas of Finite Groups). If we have $K \lhd G$ with $K \neq G$ or $\{e\}$, then we can "quotient out" G into G/K. If G/K is not simple, repeat. Then we can write G as an "inverse quotient" of simple groups.