

MATH 4580 LECTURE NOTES

LÉO JIMENEZ

CONTENTS

1. Sets, functions, relations	2
1.1. Sets	2
1.2. Functions	3
1.3. Relations	5
2. Induction	6
3. Number theory	9
3.1. Division, greatest common divisor	9
3.2. Primes	11
4. Groups	11
4.1. Definition and basic properties	11
4.2. Finite cyclic group and their invertibles	15
4.3. Order	17
5. Tools to study groups	18
5.1. Subgroups	18
5.2. Morphisms	20
6. Important examples	25
6.1. Cyclic groups	26
6.2. Multiplicative group of the complex numbers and the circle group	26
6.3. Permutation groups	31
6.4. Dihedral groups	36
6.5. Group presentations	39
7. Decomposing groups: cosets and quotients	41
7.1. Cosets and Lagrange's theorem	41
7.2. Normal subgroups, quotient groups	45
7.3. The first and third isomorphism theorem	46
7.4. The second isomorphism theorem	51
7.5. Simple groups	51
8. Rings	55
8.1. Definition, basic properties	55
8.2. Integral domains and fields	58
8.3. Morphisms, subrings, ideals	60
8.4. Polynomial rings	64
8.5. Prime and maximal ideals	67
8.6. The division algorithm	69
8.7. Irreducible polynomials	71
References	73

Date: March 25, 2026.

1. SETS, FUNCTIONS, RELATIONS

This section is a review of material from MATH 3345, Foundations of Higher Mathematics.

1.1. Sets.

Definition 1.1. A set is a collection of objects called its elements. If a is an object and A is a set, we write $a \in A$ for "a is an element of A ".

A common way to construct sets is the *set-builder notation*:

Definition 1.2. If P is some property an object can have, we denote $\{x : P(x)\}$ the set of objects having property P . If A is a set, then $\{x \in A, P(x)\}$ denotes the set of x in A having property P .

Example 1.3. (1) $\{m \in \mathbb{Z}, 2 \text{ divides } m\}$ is the set of even number.

A very important property of sets is that *two sets are equal if and only if they have the same elements*.

As a consequence, there is a unique set with no elements, denoted \emptyset .

Definition 1.4. Given a set A , we say that B is a subset of A , denoted $B \subseteq A$, if for any object x , if $x \in B$, then $x \in A$.

In this class, there is no difference between \subset and \subseteq .

In other words, $A \subset B$ if for all object x , we have $x \in A \Rightarrow x \in B$. We deduce from this that given two sets A and B , to prove that $B \subseteq A$, we pick an arbitrary $b \in B$, and show that $b \in A$.

Example 1.5. Let $B = \{m \in \mathbb{Z}, 4 \text{ divides } m\}$ and $A = \{m \in \mathbb{Z}, 2 \text{ divides } m\}$. Then $B \subset A$.

Proof. Let $m \in B$. So 4 divides m , which means that there is $k \in \mathbb{Z}$ such that $m = 4k$. Therefore $m = 2(2k)$, so 2 divides m , which means that $m \in A$. We conclude that $B \subseteq A$. \square

We also deduce that for any two sets A and B , $A = B$ if and only if $A \subset B$ and $B \subset A$. This means that to prove that two sets A and B are equal, we pick an arbitrary $a \in A$ and show that $a \in B$. Then we pick an arbitrary $b \in B$ and show that $b \in A$.

Example 1.6. $\{x \in \mathbb{R} : \exists y \in \mathbb{R}, y^2 = x\} = [0, +\infty)$.

Proof. Let $x \in \mathbb{R}$ such that there is $y \in \mathbb{R}$ with $x = y^2$. Then $x = y^2 \geq 0$, so $x \in [0, +\infty)$. Therefore $\{x \in \mathbb{R} : \exists y \in \mathbb{R}, y^2 = x\} \subset [0, +\infty)$.

Let $x \in [0, +\infty)$. Let $y = \sqrt{x}$, which exists because x is positive. Then $y^2 = x$, therefore $x \in \{x \in \mathbb{R} : \exists y \in \mathbb{R}, y^2 = x\}$. This implies that $[0, +\infty) \subseteq \{x \in \mathbb{R} : \exists y \in \mathbb{R}, y^2 = x\}$.

We conclude that $\{x \in \mathbb{R} : \exists y \in \mathbb{R}, y^2 = x\} = [0, +\infty)$. \square

Proposition (Properties of inclusion). For any sets A, B, C :

- (1) $A \subseteq A$,
- (2) if $A \subseteq B$ and $B \subset A$, then $A = B$,
- (3) if $A \subseteq B$ and $B \subset C$, then $A \subseteq C$.

Definition 1.7. The set of all subsets of a set A is also a set, denoted $\mathcal{P}(A)$. In other words:

$$\mathcal{P}(A) := \{X : X \subset A\} .$$

Definition 1.8 (Set operations). Let A, B be sets, define:

- $A \cup B = \{x : x \in A \text{ or } x \in B\}$, the union of A and B ,
- $A \cap B = \{x : x \in A \text{ and } x \in B\}$, the intersection of A and B ,
- $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$, the complement of B in A .

We summarize the properties of these set operations:

Theorem (Properties of set operations). Let A, B, C be sets. Then:

- (1) $(A \cap B) \cap C = A \cap (B \cap C)$
- (2) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- (3) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- (4) $C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B)$
- (5) $C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B)$

Definition 1.9. Given two sets A and B , we define the cartesian product $A \times B$ of A and B as the set $\{(a, b) : a \in A, b \in B\}$.

Theorem (Properties of the cartesian product). Let A, B, C, D be sets, then:

- (1) $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$,
- (2) $(A \cup B) \times C = (A \times C) \cup (B \times C)$,
- (3) if $A \times C = B \times C$ and $C \neq \emptyset$, then $A = B$,
- (4) $A \times \emptyset = \emptyset$.

1.2. Functions. Recall that a function from a set A to a set B is "a process which assigns to every element of A a unique element of B ". The way to make this rigorous is the following:

Definition 1.10. Given two sets A and B , a *function* f from A to B is a subset of $f \subset A \times B$ such that for any $x \in A$, there is a unique $y \in B$ such that $(x, y) \in f$.

We never use this notation. What we write instead:

- $f : A \rightarrow B$ to denote a function f from A to B
- $f(x) = y$ for $(x, y) \in f$.

Definition 1.11. Let $f : A \rightarrow B$ be a function.

- A is called the *domain* of f ,
- B is called the *codomain* of f ,
- for any $X \subset A$, we define $f(X) = \{f(x) : x \in X\} \subset B$, which we call the *image of X under f* .
- for any $Y \subset B$, we define $f^{-1}(Y) = \{x \in A : f(x) \in Y\} \subset A$, which we call the *preimage of Y under f* .

We can compose functions together:

Definition 1.12. Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be functions. We define the function $g \circ f : A \rightarrow C$, called *g composed with f* , as given by $(g \circ f)(a) = g(f(a))$ for all $a \in A$.

There is, for any set, a special function that will play a key role in composition:

Definition 1.13. Let A be a set. Define id_A to be the function such that $\text{id}_A(a) = a$ for all $a \in A$.

We say that functions are equal when they do the same thing on every object of their domain:

Definition 1.14. Two functions $f : A \rightarrow B$ and $g : A \rightarrow B$ are equal if for all $a \in A$, we have $f(a) = g(a)$.

Example 1.15. For any function $f : A \rightarrow B$, we have that:

- (1) $f \circ \text{id}_A = f$,
- (2) $\text{id}_B \circ f = f$.

Composition is *associative*:

Proposition 1.16. Let $f : A \rightarrow B$, $g : B \rightarrow C$ and $h : C \rightarrow D$ be functions. Then $h \circ (g \circ f) = (h \circ g) \circ f$.

Proof. First note that both these functions have domain A and codomain D . Let $a \in A$, then:

$$\begin{aligned} h \circ (g \circ f)(a) &= h(g \circ f(a)) \\ &= h(g(f(a))) \\ &= h \circ g(f(a)) \\ &= (h \circ g) \circ f(a) \end{aligned}$$

□

The most important definitions regarding functions are the following:

Definition 1.17. Let $f : A \rightarrow B$ be a function. We say that:

- f is injective if for all $x, y \in A$, if $f(x) = f(y)$, then $x = y$.
- f is surjective if for all $y \in B$, there is $x \in A$ such that $f(x) = y$,
- f is bijective if it is both injective and surjective.

A very good exercise to warm up:

Proposition 1.18. Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be functions. Then:

- (1) if both f and g are injective, so is $g \circ f$.
- (2) if both f and g are surjective, so is $g \circ f$.

When a function is bijective, we can "reverse it":

Theorem 1.19. Let $f : A \rightarrow B$ be a function. Then f is bijective if and only if there is a function $f^{-1} : B \rightarrow A$ such that $f^{-1} \circ f = \text{id}_A$ and $f \circ f^{-1} = \text{id}_B$. This function is unique, we call it the inverse of f .

This theorem means that to show a function $f : A \rightarrow B$ is bijective, it is enough to show that there is a function $g : B \rightarrow A$ such that $g \circ f = \text{id}_A$ and $f \circ g = \text{id}_B$.

Example 1.20. Show that the function $f : [0, +\infty] \rightarrow [0, +\infty]$ given by $f(x) = \sqrt{x}$ is bijective by finding its inverse.

Taking the inverse behaves well with respect to composition. You can prove the following as an exercise:

Proposition 1.21. Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be bijections. Then $g \circ f : A \rightarrow C$ is a bijection, and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

1.3. Relations.

Definition 1.22. A *relation* between two sets A and B is a subset $R \subset A \times B$.

If A is a set, then a relation on A is a subset $R \subset A \times A$

For $(a, b) \in A \times B$, we write aRb (or sometimes $R(a, b)$) to say that $(a, b) \in R$.

Just like we don't usually think of a function $f : A \rightarrow B$ as a special subset of $A \times B$, we don't think of a relation that way either.

More intuitive is to think of R as some information on the pair (a, b) . For example, a and b can share some properties (a and b have the same remainder modulo 7), be comparable in some way (a is larger than b)...

Here are the main properties of relations:

Definition 1.23. Let A be a set, and R a relation on A . We say that R is:

- reflexive if $\forall a \in A, aRa$,
- antireflexive if $\forall a \in A, \neg(aRa)$,
- symmetric if $\forall a, b \in A, (aRb \Rightarrow bRa)$,
- antisymmetric if $\forall a, b \in A \left((aRb \wedge bRa) \Rightarrow a = b \right)$,
- transitive if $\forall a, b, c \in A, \left((aRb \wedge bRc) \Rightarrow aRc \right)$

Example 1.24.

- \leq is a relation on \mathbb{R} . It is reflexive, antisymmetric and transitive.
- $=$ is a relation on any set. It is reflexive, symmetric, antisymmetric and transitive.
- \neq is also a relation on any set. It is symmetric and antireflexive.

The main type of relation that we will use is:

Definition 1.25. Let A be a set. A relation E on A is an *equivalence relation* if it is reflexive, symmetric and transitive.

Definition 1.26. Let A be a set and E an equivalence relation on A . For some $a \in A$, the *equivalence class of a* is the set

$$[a]_E = \{x \in A : aEx\} .$$

We will often denote it $[a]$ when there is no ambiguity possible.

The *quotient of A by E* is the set of equivalence classes, i.e.

$$A/E = \{[a] : a \in A\} .$$

We will look at some examples now.

Example 1.27 (Congruence modulo n). Let $n \in \mathbb{Z}$. We say that some integers $a, b \in \mathbb{Z}$ are *congruent modulo n* , and write $a \equiv b [n]$, if $n|b - a$ (meaning n divides $b - a$). This is an equivalence relations. Let $a, b, c \in \mathbb{Z}$, then:

- $a - a = 0$, and $n|0$, so $a \equiv a [n]$
- if $a \equiv b [n]$, then $n|b - a$. So $n|a - b$ as well, meaning $b \equiv a [n]$
- is $a \equiv b [n]$ and $b \equiv c [n]$, then $n|b - a$ and $n|c - b$. This implies that n divides $c - b + b - a = c - a$. So $a \equiv c [n]$.

The quotient \mathbb{Z}/\equiv is written $\mathbb{Z}/n\mathbb{Z}$ (and we will see why later). It can be identified with the set of remainders modulo n , i.e. $\{0, 1, \dots, n - 1\}$.

Example 1.28. The relation on \mathbb{R}^2 given by $(x, y)E(u, v)$ if and only if $x^2 + y^2 = u^2 + v^2$ is an equivalence relation.

The equivalence class of $(a, b) \in \mathbb{R}^2$ is the circle with center the origin passing through (a, b) . The quotient set can be identified with $[0, +\infty)$ by taking the radius of the circles.

Example 1.29. Two triangles are:

- congruent if they have the same side length.
- similar if they have the same angles.

Both are equivalence relations on triangles.

The previous two examples can naturally be seen as coming from a group, as we will (hopefully) see later. In fact they all come from *matrices*:

Example 1.30 (Change of basis). Let $M_n(\mathbb{R})$ be the set of $n \times n$ matrices with entries in \mathbb{R} . The relation $M \sim N$ if and only if there exists an invertible $P \in M_n(\mathbb{R})$ such that $P^{-1}MP = N$ is an equivalence relation.

Example 1.31. Let $f : A \rightarrow B$ be a function. Then xEy if and only if $f(x) = f(y)$ is an equivalence relation.

The equivalence class of $a \in A$ is $f^{-1}(\{f(a)\})$.

In fact, all equivalence relations are of the form given by this example.

Proposition 1.32. Let A be a set and E an equivalence relation on A . Then there is a surjective function $\pi_E : A \rightarrow A/E$ with $\pi(a) = [a]$.

By definition, we have $\pi(a) = \pi(b)$ if and only if aEb .

We will also sometimes look at partitions:

Definition 1.33. A *partition* of a set A is a set S of subsets of A such that every $a \in A$ is contained in exactly one set in S .

Partitions are the same as equivalence classes in the following sense:

Theorem 1.34. Let A be a set. If E is an equivalence relation on A , then the set of equivalence classes form a partition of A .

Conversely, if S is a partition of A , then the relation E given by xEy if and only if x and y belong to the same element of S , is an equivalence relation.

2. INDUCTION

In this section, we review induction from MATH3345.

The principle of mathematical induction:

Let $P(n)$ be a statement about n . Suppose that:

- (1) $P(1)$ is true,
- (2) for all $n \in \mathbb{N}$, if $P(n)$ is true, then $P(n + 1)$ is true.

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

Remark 2.1. We can "start" the induction at any $m \in \mathbb{Z}$, and we get that if:

- (1) $P(m)$ is true,
- (2) for all $n \geq m$, if $P(n)$ is true, then $P(n + 1)$ is true.

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

Here is how to structure a proof by induction (starting at 1).

- state exactly what is the property $P(n)$ you want to prove.
- Base case: prove that $P(1)$ holds.
- Inductive step: prove that if $P(n)$ is true, then $P(n+1)$ is true. To do this, assume $P(n)$ and deduce $P(n+1)$. Remember to state when you are using that $P(n)$ is true!
- Conclusion: by induction, $P(n)$ is true for all $n \in \mathbb{N}$.

If your proof does not have all these steps, it is not complete.

There are two main types of proof by induction you may be asked to do: proving a formula and proving a property.

Example 2.2. For all $n \in \mathbb{N}$, we have that $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$.

Proof. Let $P(n)$ be the property that $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$. We prove by induction that $P(n)$ holds for all $n \in \mathbb{N}$.

- Base case: We have that $\frac{1(1+1)}{2} = \frac{2}{2} = 1$ so $P(1)$ holds.
- Inductive step: Suppose that $P(n)$ holds, for some $n \in \mathbb{N}$. Then:

$$\begin{aligned} 1 + 2 + 3 + \cdots + n + (n + 1) &= \frac{n(n + 1)}{2} + n + 1 \text{ by induction hypothesis} \\ &= \frac{n(n + 1) + 2(n + 1)}{2} \\ &= \frac{(n + 1)(n + 1)}{2} \end{aligned}$$

hence $P(n + 1)$ is true.

By induction, we have that $P(n)$ is true for all $n \in \mathbb{N}$. □

Example 2.3. For all $n \geq 2$, we have $2n \leq 2^n$.

Proof. Let $P(n)$ be the property that $2n \leq 2^n$. We prove by induction that $P(n)$ holds for all $n \geq 2$.

- Base case: We have that $2 = 4 = 2^2$, so $P(2)$ holds.
- Inductive step: Suppose that $P(n)$ holds for some $n \geq 2$. Then:

$$\begin{aligned} 2(n + 1) &= 2n + 2 \\ &\leq 2^n + 2 \text{ by inductive hypothesis} \\ &= 2^n \left(1 + \frac{2}{2^n}\right) \\ &= 2^n \left(1 + \frac{1}{2^{n-1}}\right) \\ &\leq 2^n \times 2 \text{ because } n \geq 2 \\ &= 2^{n+1} \end{aligned}$$

therefore $P(n + 1)$ holds.

By induction, we have that $P(n)$ holds for all $n \geq 2$. □

We also have the principle of *strong* mathematical induction:

The principle of strong mathematical induction:

Let $P(n)$ be a statement about n . Suppose that:

- (1) $P(1)$ is true,
- (2) for all $n \in \mathbb{N}$, if $P(k)$ is true for all $k \leq n$, then $P(n + 1)$ is true.

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

Note that the two principles are equivalent: we can deduce one from the other in both directions.

We also have the *well-ordering principle*:

Let $S \subset \mathbb{Z}$ be non-empty and bounded from below, meaning that there is $a \in \mathbb{Z}$ such that $a \leq s$ for all $s \in S$. Then S has a *least element*, meaning some $s_0 \in S$ such that $s_0 \leq s$ for all $s \in S$.

Remark 2.4. *The well-ordering is a property of the order relation \leq on \mathbb{Z} . It is not true of all order relations. For example, consider the order \leq , but on the real numbers \mathbb{R} this time. The open interval $(0, 1)$ is bounded from below, by -1 for example, but it does not have a least element (you can try to prove it).*

In general, an order \leq satisfying the well-ordering principle is called a well-order.

You may not have seen this theorem in MATH 3345:

Theorem 2.5. *The well-ordering principle and the induction principle are equivalent.*

Proof. induction \Rightarrow well-ordering. Let S be a non-empty subset of \mathbb{Z} , and let $b \in \mathbb{Z}$ be such that $b \leq s$ for all $s \in S$. Assume, for a contradiction, that S has no least element. This means that for all $a \in \mathbb{Z}$, either $a \notin S$ or there is $s \in S$ such that $s < a$. We will show by strong induction that for all $a \in \mathbb{Z}$, we have $a \notin S$, which is a contradiction.

If $a < b - 1$, this is because b is a lower bound for S .

Base case: if $a = b - 1$, then again this is because b is a lower bound.

Inductive step: Let $a \in \mathbb{Z}$ and suppose that k is not in S for any $k \leq a$. By assumption, either $a + 1 \notin S$, or there is $s \in S$ such that $s < a + 1$. But the latter is impossible by induction hypothesis. So $a + 1 \notin S$.

We have proven by induction that $S = \emptyset$, which is a contradiction. So S must have a least element.

well ordering \Rightarrow induction. Assume the well-ordering principle is true. Let P be a property of numbers, and assume that:

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

We want to show that $P(n)$ is true for all $n \in \mathbb{N}$. For a contradiction, assume that $P(n)$ is false for some n . Consider the set $S = \{n \in \mathbb{N} : P(n) \text{ is false}\}$, it is non-empty by assumption. Let n be its least element, which exists by the well-ordering principle. Since $1 \notin S$, we know that $n \neq 1$. We see that $n - 1 \notin S$, as otherwise n would not be the least element of S . Therefore $P(n - 1)$ is true, and by assumption we obtain that $P(n)$ is true, a contradiction. \square

3. NUMBER THEORY

This section is also mostly MATH 3345 review.

3.1. Division, greatest common divisor. One of the source of complexity in the study of number is the *Euclidian division*:

Theorem 3.1. *Let $n, d \in \mathbb{Z}$ with $d \geq 1$. There there are unique $q, r \in \mathbb{Z}$ such that:*

- $n = dq + r$
- $0 \leq r \leq d - 1$.

The proof of this, done in MATH 3345, is a bit annoying and maybe not the most instructive. Try to do it using the well-ordering principle!

This sort of division is not exclusive to integers. We will see much later (maybe) that we can do this with polynomials.

The Euclidian division is the first step of an algorithm that allows us to find the greatest common divisor two integers:

Definition 3.2. Let $a, b \in \mathbb{Z}$. A *greatest common divisor* of a and b is some $d \in \mathbb{Z}$ such that:

- $d \geq 0$,
- $d|a$ and $d|b$,
- for all $d' \in \mathbb{Z}$, if $d'|a$ and $d'|b$, then $d'|d$.

We denote it by $\gcd(a, d)$

Remark that if the GCD of a and b exists, it must be unique (exercise: prove it!). We will therefore call it *the* greatest common divisor. But does it actually exists? The answer is yes:

Theorem 3.3. *Let $a, b \in \mathbb{Z} \setminus \{0\}$, then $\gcd(a, b)$ exists and it is the least element of the set $\{ax + by : x, y \in \mathbb{Z}\} \cap \mathbb{N}$.*

Before proving the theorem, we start with a lemma we will revisit later:

Lemma 3.4. *Let $I \subset \mathbb{Z}$ be non-empty and such that for all $r, s \in I$ and $c \in \mathbb{Z}$, we have:*

- $r + s \in I$,
- $cr \in I$.

Then there is $s_0 \in I$ such that $I = \{s_0 k : k \in \mathbb{Z}\}$. Equivalently, there is s_0 such that $s_0|s$ for all $s \in I$.

Proof. Let I be such a subset. Note that it must contain 0. Indeed, it is non-empty, so we can pick some $s \in I$. By the second condition, we have $0 \times s = 0 \in I$.

Now there are two cases. If $I = \{0\}$, then the conclusion of the lemma is true for $s_0 = 0$.

Now assume that $S \neq 0$ and consider $S = I \cap \mathbb{N}$. This is non-empty: if $s \in I$ and $s \neq 0$, then either $s \geq 1$ so $s \in S$, or $s \leq -1$ in which case $-s \in S$.

Let s_0 be its least element. We show that s_0 divides any other $s \in S$. So let $s \in S$, and divide it by s_0 , we have $s = s_0 q + r$ for some $q, r \in \mathbb{Z}$ with $0 \leq r \leq s_0 - 1$.

Because of our conditions on I , we get that $r = s - s_0 q \in I$, and if $r > 0$, this means that $r \in S$. This is a contradiction as $r < s_0$ but s_0 is the least element of S . Therefore $r = 0$, or equivalently, $s_0|s$.

This implies that $s_0|s$ for all $s \in S$. If $s = 0$, then $0 = s_0 \times 0$. If $s < 0$, then $-s > 0$ and $-s \in I$, so $s_0|-s$. This implies $s_0|s$.

We have proven that $I \subset \{s_0k : k \in \mathbb{Z}\}$. For the other inclusion, note that since $s_0 \in I$, the second condition gives $s_0k \in I$ for any $k \in \mathbb{Z}$. \square

We are now ready for:

Proof of Theorem 3.3. Let $a, b \in \mathbb{Z} \setminus \{0\}$ and let $I = \{ax + by : x, y \in \mathbb{Z}\}$. Then it is easy to check that I satisfies the conditions of Lemma 3.4. Therefore there is some $d \in I$ such that d divides all the elements of I .

Picking $x = 1$ and $y = 0$, we see that $a \in S$, so $d|a$. With $x = 0$ and $y = 1$, we get $d|b$. In particular $I \neq \{0\}$, so d is the smallest element of $I \cap \mathbb{N}$.

Now let d' be another common divisor of a and b . Then d' divides $ax + by$ for any $x, y \in \mathbb{Z}$, meaning that d' divides all elements of I . In particular d' divides d . \square

This is all great, but now we might ask how to do this effectively:

Question. *How do we find the GCD of a and b ?*

Here's a rough idea. We know that $\text{GCD}(a, b)$ is the smallest element of $S = \{ax + by : x, y \in \mathbb{Z}\} \cap \mathbb{N}$. Suppose $a, b > 0$ and $a > b$. Then we can find a smaller element in S by dividing a by b : write $a = bq + r$ with $0 \leq r < b$. But $r = a - bq$ is also in S if it is not zero! We can repeat the process by dividing b by r . Iterating, we eventually reach the smallest element of S .

Let's try an example first, pick $a = 150, b = 126$. We get:

$$\begin{aligned} 150 &= 126 \times 1 + 24 \\ 126 &= 24 \times 5 + 6 \\ 24 &= 6 \times 4 \end{aligned}$$

So $\text{GCD}(150, 126) = 6$.

The general result is:

Theorem 3.5 (Euclidian algorithm). *Let $a, b \in \mathbb{Z}$, we can find $\text{gcd}(a, b)$ by using the following algorithm. First, change a, b to $-a, -b$ if they are negative. Then*

Step 1: Set $r_0 = a$ and $r_1 = b$.

Step 2: divide r_{n-1} by r_n and write $r_{n-1} = r_n q_n + r_{n+1}$

- *if $r_{n+1} \neq 0$, repeat step 2 for r_{n+1} and r_n*
- *else, output r_n .*

Proof. Let $I = \{ax + by : x, y \in \mathbb{Z}\}$. Because $\text{gcd}(a, b) = \text{gcd}(-a, b)$, we may assume that a, b are positive.

If either one is zero, the algorithm stops at the first step. So we may assume they are both non-zero.

Let $S = I \cap \mathbb{N}$. By our assumptions, it contains both a and b . First note that we have $r_0 > r_1 > r_2 > \dots > r_{n-1} > r_n \geq 0$ for all n . Therefore the r_i form a subset of \mathbb{Z} that is bounded from below by zero, they must have a least element. This tells us the algorithm must stop at some r_n . This means that there is n such that $r_{n+1} = 0$.

Since $r_{n-1} = r_n q_n$, we have that $r_n | r_{n-1}$. We can prove by induction that for all $k \leq n$, we have $r_n | r_k$: for the next step, we would use that $r_{n-2} = r_{n-1} q_{n-1} + r_n$. Since $r_0 = a$ and $r_1 = b$, this means r_n divides a and b .

We can also prove by induction, using that $r_{m+1} = r_{m-1} - r_m q_m$, that r_m is in S for all m , and so $r_n \in S$.

Therefore we have $r_n \in S$ and r_n divides a and b . As we've seen before, this implies that $r_n = \gcd(a, b)$. \square

3.2. Primes. Primes numbers are the building block of integers (see the fundamental theorem of arithmetic below), which is why they are important.

Definition 3.6. An integer p is *prime* if:

- $p \geq 2$,
- for all $d \in \mathbb{N}$, if $d|p$, then either $d = 1$ or $d = p$.

Recall the following important theorem:

Theorem 3.7 (Euclid's theorem). *There are infinitely many prime numbers.*

We also have the following characterization of prime numbers:

Theorem 3.8. *An integer $p \in \mathbb{Z}$ with $p \geq 2$ is prime if and only if for all $a, b \in \mathbb{Z}$, if $p|ab$, then $p|a$ or $p|b$.*

Note that the left to right direction is called *Euclid's lemma*, and is very useful.

Proof. Assume that p is prime, and let $a, b \in \mathbb{Z}$ such that $p|ab$.

If $p|a$, then we're done. So we now assume that p does not divide a . This implies that $\gcd(a, p) = 1$. Indeed, if $d|a$ and $d|p$, then $d = 1$ or p because p is prime. But $d \neq p$, as otherwise we would have $p|a$, a contradiction.

This means that there are $x, y \in \mathbb{Z}$ such that $1 = ax + py$. we can now write:

$$\begin{aligned} b &= b \times 1 \\ &= b(ax + py) \\ &= abx + pby \end{aligned}$$

and since $p|ab$ and $p|p$, we obtain $p|b$.

The right to left direction is left as an exercise. \square

We end this section by the very important:

Theorem 3.9 (Fundamental theorem of arithmetic). *Every integer greater or equal to 2 can be written in a unique way as a product of prime numbers, up to reordering the factors.*

Proof. Homework 2. \square

4. GROUPS

4.1. Definition and basic properties. In mathematics, we often want to combine objects (numbers, functions, matrices, sets ...) together into a third object of the same type. Here is the general idea behind this:

Definition 4.1. Let A be a set. A *binary operation* on A is a function

$$* : A \times A \rightarrow A .$$

We denote the image of (a, b) under $*$ by $a * b$.

Example 4.2.

- (1) $+$ and \times are binary operations on \mathbb{Z} , \mathbb{Q} , \mathbb{R} and \mathbb{N} .
- (2) $-$ is a binary operation on \mathbb{Z} , but not on \mathbb{N} , because, for example, $1 - 4 \notin \mathbb{N}$
- (3) matrix addition and multiplication are binary operations on $M_n(\mathbb{R})$
- (4) let A be any set, then \cap and \cup are binary operations on $\mathcal{P}(A)$.
- (5) let A be any set, and consider A^A , the set of all functions from A to itself. Then composition is a binary operation on A^A .
- (6) consider the set of functions $\mathcal{F} = \{f : \mathbb{R} \rightarrow \mathbb{R}\}$, and define

$$\begin{aligned} \text{ev} : \mathcal{F} \times \mathbb{R} &\rightarrow \mathbb{R} \\ (f, x) &\rightarrow f(x) \end{aligned}$$

this is not a binary operation on a single set, because $\mathcal{F} \neq \mathbb{R}$.

Here are some very important properties a binary operation can have:

Definition 4.3. A binary operation $*$ on some set A is:

- associative if for all $a, b, c \in A$, we have

$$a * (b * c) = (a * b) * c .$$

- commutative if for all $a, b \in A$, we have

$$a * b = b * a .$$

We say that e is an identity element for $*$ if for all $a \in A$, we have:

$$a * e = e * a = a .$$

If e is an identity element for $*$, we say that b is an inverse of a for $*$ if

$$a * b = b * a = e .$$

Here is a first observation:

Proposition 4.4. *Let $*$ be a binary operation on A . If $*$ has an identity, then it is unique.*

Proof. Suppose that e_1 and e_2 are both identities for A . Then we have $e_1 * e_2 = e_1$ because e_2 is an identity, and also $e_1 * e_2 = e_2$ because e_1 is an identity. So $e_1 = e_2$. \square

So from now on, I will always talk about *the* identity of a binary operation.

Let's look at our examples:

- (1) $+$ and \times are associative and commutative. For identity and inverses:
 - $+$ has no identity element on \mathbb{N} , but has 0 as identity on \mathbb{Z} , \mathbb{Q} and \mathbb{R} . The inverse of a is $-a$.
 - \times has identity 1 on all four sets. On \mathbb{Q} and \mathbb{R} , every number but 0 has an inverse. On \mathbb{N} and \mathbb{Z} , only 1 and -1 have inverses.
- (2) $-$ on \mathbb{Z} is:
 - not associative: $(4 - 3) - 2 = -1$ and $4 - (3 - 2) = 3$
 - not commutative: $1 - 2 = -1$ but $2 - 1 = 1$
 - it has no identity. For all a , we have $a - 0 = a$, but $0 - a = -a$. We say that 0 is a *right identity*.
 - Since it has no identity, it cannot have inverses in the sense of our definition. However, for any a , we have $a - a = 0$.
- (3) matrix multiplication is:

- associative
 - not commutative
 - the identity matrix $\begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$ is the identity
 - matrices with non-zero determinant have inverses.
- (4) both \cap and \cup are commutative and associative. The identity of \cap is A , the identity of \cup is \emptyset . For \cap , only A has an inverse, itself. For \cup , only \emptyset has an inverse, itself.
- (5) composition is:
- associative
 - not commutative in general. If $A = \mathbb{R}$ with $f(x) = x+1$ and $g(x) = x^2$, then $f \circ g \neq g \circ f$.
 - the identity is Id_A .
 - f has an inverse if and only if f is a bijection.

Here are some more uniqueness properties:

Proposition 4.5. *Let $*$ be a binary operation on some set A , and suppose it is associative. Let $a, b \in A$, then:*

- (1) if a has an inverse, then it is unique. We denote it a^{-1} .
- (2) if a has an inverse a^{-1} , then the inverse of a^{-1} is a , i.e. $(a^{-1})^{-1} = a$.
- (3) if both a and b have inverse, then the inverse of $a * b$ is $b^{-1} * a^{-1}$, i.e.:

$$(a * b)^{-1} = b^{-1} * a^{-1} .$$

Proof. Let $a, b \in A$, and assume they both have inverses. Let e be the identity of $*$.

- (1) assume that a has inverses c_1 and c_2 . Then we have

$$\begin{aligned} c_1 &= c_1 * e \\ &= c_1 * (a * c_2) \\ &= (c_1 * a) * c_2 \\ &= e * c_2 \\ &= c_2 \end{aligned}$$

- (2) We have $a * a^{-1} = a^{-1} * a = e$, so a^{-1} is the inverse of a

- (3) we compute

$$\begin{aligned} (a * b) * (b^{-1} * a^{-1}) &= a * (b * b^{-1}) * a^{-1} \\ &= a * e * a^{-1} \\ &= e \end{aligned}$$

and the same computation shows $(b^{-1} * a^{-1}) * (a * b) = e$. Uniqueness of inverses gives us $(a * b)^{-1} = b^{-1} * a^{-1}$.

□

We finally arrive at the definition that will keep us busy for the next few months:

Definition 4.6. A group $(G, *)$ is a set with a binary operation $*$ such that:

- $*$ is associative,
- $*$ has an identity element e ,

- every $g \in G$ has an inverse g^{-1} .

Note that by the previous theorem, identity and inverses are unique.

An important class of groups, which are much easier to understand, are:

Definition 4.7. A group $(G, *)$ is *abelian*, or *commutative*, if $*$ is commutative.

Notation. For a group $(G, *)$, we will denote the group operation multiplicatively, i.e. we write $g * h = g \cdot h$, or even often just gh . The identity can be written 1 in this case.

If G is abelian (and only in that case), we can write the group operation $+$. The identity is then written 0, and the inverse of g is $-g$.

Great, so are there groups? We know some already:

Example 4.8.

- \mathbb{Z} , \mathbb{Q} and \mathbb{R} are all abelian groups under $+$.
- $\mathbb{Q} \setminus \{0\}$ and $\mathbb{R} \setminus \{0\}$ are groups under \cdot .
- any vector space is a group under $+$. This includes $(M_n(\mathbb{R}), +)$.
- the set $\text{Gl}_n(\mathbb{R})$ of invertible matrices, i.e. with determinant non-zero, is a group under matrix multiplication.

And we also have some maybe less familiar example:

Proposition 4.9. Let A be any set, and write S_A for the set of bijections $f : A \rightarrow A$ (S stands for symmetry). Then (S_A, \circ) is a group.

Proof. All properties were proven in MATH3345. □

The following two propositions tell us how to solve equations in groups:

Proposition 4.10 (Cancellation laws). Let G be a group. For all $a, b, c \in G$, we have that:

- $ab = ac$ implies $b = c$,
- $ba = ca$ implies $b = c$.

Proof. We only prove the first one, the other being similar. Let $a, b, c \in G$, and assume that $ab = ac$. Then $a^{-1}ab = a^{-1}ac$, so simplifying we get $b = c$. □

Proposition 4.11. Let $g, h \in G$. There there is a unique $x \in G$ such that $gx = h$. Similarly, there is a unique y such that $yg = h$.

Proof. Pick $x = g^{-1}h$, then:

$$\begin{aligned} gx &= gg^{-1}h \\ &= h \end{aligned}$$

To prove uniqueness, suppose that $gx_1 = h = gx_2$, then $x_1 = x_2$ by the cancellation law. Similarly, we can pick $y = hg^{-1}$. □

Finally, what's the simplest possible group?

Definition 4.12. The trivial group is the group on the set $\{1\}$ defined by $1 \cdot 1 = 1$.

Note that the choice of $\{1\}$ as our set is arbitrary. Any group on a set with one element will be defined in this exact way.

4.2. **Finite cyclic group and their invertibles.** Let $n \in \mathbb{N}$. We could ask :

Question. *What if $n = 0$ in $(\mathbb{Z}, +)$? Do we still get a group?*

This is something we do all the time when computing with hours or days of the week:

- 60 minutes equals an hour and zero minutes, so 60 minutes is "the same" as zero minutes
- 7 days from Wednesday is Wednesday, so 7 days is "the same" as 0 days.

To make this precise, we used congruence modulo n in MATH3345.

Recall that for any two $a, b \in \mathbb{Z}$, we write $a \equiv b [n]$ is $b - a | n$. This is an equivalence relation, and we denote $[a]$ the class of some $a \in \mathbb{Z}$.

This equivalence relation has n different class, which are the classes $[0], [1], \dots, [n-1]$, and correspond to possible remainders modulo n .

We can try to define addition as follows, for any $a, b \in \mathbb{Z}$:

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

We have to check that this is *well-defined*, i.e. the result does not depend of which elements of the class we pick. More formally:

Proposition 4.13 (From 3345). *For all $a, b, c, d \in \mathbb{Z}$, if $a \equiv c [n]$ and $b \equiv d [n]$, then $a + b \equiv c + d [n]$.*

In terms of equivalence classes, this means that if $[a] = [c]$ and $[b] = [d]$, then $[a + b] = [c + d]$. We conclude that $+$ defines a binary operation on $\{[0], [1], \dots, [n-1]\}$.

Proposition 4.14. *The binary operation $(\{[0], [1], \dots, [n-1]\})$ is an abelian group.*

Proof. Associativity and commutativity come from associativity and commutativity of $(\mathbb{Z}, +)$. For associativity, let $a, b, c \in \mathbb{Z}$, then:

$$\begin{aligned} ([a] + [b]) + [c] &= [a + b] + [c] \\ &= [a + b + c] \\ &= [a] + [b + c] \text{ by associativity of } (\mathbb{Z}, +) \\ &= [a] + ([b] + [c]) \end{aligned}$$

A similar computation works for commutativity. Finally, $[0]$ is the identity, and the inverse of any $[a]$ is $[-a]$. □

We will, we the context is clear, drop the brackets and just write $\{0, 1, \dots, n-1\}$ for the set of equivalence classes.

Definition 4.15. The group $(\{0, 1, \dots, n-1\}, +)$ is called the *cyclic group of order n* , and denoted \mathbb{Z}_n , or $\mathbb{Z}/n\mathbb{Z}$.

Example 4.16. We look at the group \mathbb{Z}_4 , on the set $\{0, 1, 2, 3\}$. We can write down everything about its addition in its *Cayley table*:

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

In general, for any finite group, we can write down its Cayley table by listing every element, and computing all $g \cdot h$ for $g, h \in G$.

Remark 4.17. *Note two properties of this Cayley table:*

- *it is symmetric across the diagonal. That's because $(\mathbb{Z}_4, +)$ is abelian.*
- *every line and column contains each element of \mathbb{Z}_4 exactly once. This is true for every group, and a consequence of Proposition 4.11.*

Exercise 4.18. *Write down the Cayley table for some other $(\mathbb{Z}_n, +)$, for example \mathbb{Z}_5 .*

We could also look at classes modulo n , but with multiplication this time. Again from 3345 we have:

Proposition 4.19. *For all $a, b, c, d \in \mathbb{Z}$, if $[a] = [c]$ and $[b] = [d]$, then $[ab] = [cd]$.*

This implies that \times is a well-defined binary operation on (\mathbb{Z}_n, \times) , we write (\mathbb{Z}_n, \times) .

The class of 1 is the identity in (\mathbb{Z}_n, \times) . Observe that for all $a \in \mathbb{Z}_n$, we have $a \times 0 = 0$, so 0 has no inverse. But it's even worse than that: in \mathbb{Z}_4 , we have $2 \times 2 = 0$, so 2 has no inverse. If it had one, say a , then $0 = 2 \times 2 \times a = 2$, a contradiction. In fact:

Proposition 4.20. *For any $n \in \mathbb{N}$, the class of a has an inverse in (\mathbb{Z}_n, \times) if and only if $\gcd(n, a) = 1$.*

Proof. Suppose that $[a]$ has an inverse in (\mathbb{Z}_n, \times) . This means there is $[b] \in \mathbb{Z}_n$ with $[ab] = 1$, or, back in \mathbb{Z} :

$$ab - 1 = kn \text{ for some } k \in \mathbb{Z}$$

which implies that $\gcd(n, a) = 1$.

Conversely, suppose $\gcd(n, a) = 1$, so there is $k, b \in \mathbb{Z}$ such that $ab + kn = 1$, which means that $[a]^{-1} = [b]$ in (\mathbb{Z}_n, \times) . \square

So we need to restrict our attention to all $[a]$ such that $\gcd(n, a) = 1$ to get a group.

Definition 4.21. The set of *invertible* elements of \mathbb{Z}_n is $U(n) := \{[a] : \gcd(n, a) = 1\}$.

By the previous proposition, we have obtained:

Proposition 4.22. *The binary operation $(U(n), \times)$ is an abelian group.*

Example 4.23. We compute the Cayley table of $U(8)$. In \mathbb{Z}_8 , we see that the invertibles are $U(8) = \{1, 3, 5, 7\}$, and we get the following Cayley table:

\times	1	3	5	7
1	1	3	5	7
3	3	1	7	5
5	5	7	1	3
7	7	5	3	1

From this Cayley table, we see that for every $g \in U(8)$, we have $g \times g = 1$ (visually, the diagonal is filled with 1's). But in $(\mathbb{Z}_4, +)$, we have $1 + 1 = 2 \neq 0$. So the groups, even though they have the same size, behave differently. In fact $U(8)$ is the smallest group that is not "like"¹ $(\mathbb{Z}_n, +)$ for some n . It is called the *Klein group*.

¹Precisely, isomorphic to, which we will define later.

4.3. **Order.** First we introduce some notation.

Notation. Let G be a group. For any $n \in \mathbb{Z}$ and $g \in G$, we write:

- $g^n = \underbrace{g \cdot g \cdots g \cdot g}_{n \text{ times}}$ if $n > 0$,
- $g^n = (g^{-1})^{-n}$ if $n < 0$ (so for example $g^{-3} = g^{-1} \cdot g^{-1} \cdot g^{-1}$),
- $g^0 = 1$.

Remark that these notations are unambiguous by associativity of \cdot .
Some of the usual power laws are valid:

Proposition 4.24. *Let (G, \cdot) be a group. Then for any $g \in G$ and $n, m \in \mathbb{Z}$:*

- (1) $g^n \cdot g^m = g^{n+m}$,
- (2) $(g^n)^m = g^{nm}$

Proving this rigorously is annoying (induct on both n and m). Because this should be a fairly intuitive fact, I will not go through the proof.

Warning. Since G may not be abelian, in general $(gh)^n \neq g^n h^n$. In other words, there is no reason to be able to write something like:

$$ghgh = ghhg = gh^2g = ggh^2 = g^2h^2$$

Notation. If $(G, +)$ is an abelian group, then we can modify this notation. Let $g \in G$ and $n \in \mathbb{Z}$, then

- $ng = \underbrace{g + \cdots + g}_{n \text{ times}}$ if $n > 0$,
- $ng = (-n)(-g)$ if $n < 0$,
- $0g = 0$.

Now suppose that we keep multiplying some g by itself. Either at some point we arrive at 1, or it never happens. We make this into a definition:

Definition 4.25. Let (G, \cdot) be a group and $g \in G$. The *order* of g in G , denoted $\text{ord}(g)$, or $|g|$, is the smallest $n \in \mathbb{N}$ such that $g^n = 1$, if such an n exists, and ∞ otherwise.

Order will behave differently depending on the size of the group:

Definition 4.26. Let G be a group. If $|G| = n$ for some $n \in \mathbb{N}$, we say that G is a finite group, and we call n the *order* of G .

Otherwise, we say G is infinite, or that it is a group of infinite order.

Here is a very useful fact:

Theorem 4.27. *Let G be a finite group. Then every $g \in G$ has finite order.*

Proof. Let $g \in G$. There are two possibilities:

- (a) for all $n, m \in \mathbb{N}$, if $n \neq m$, then $g^n \neq g^m$.
- (b) there are $n, m \in \mathbb{N}$ such that $g^n = g^m$ but $n \neq m$.

First, notice that (a) is impossible. Indeed, in case (a), we have that the function $\exp_g : \mathbb{N} \rightarrow G$ defined by $\exp_g(n) = g^n$ is an injection. But \mathbb{N} is infinite and G is finite, this is a contradiction.

Therefore, there must be $n, m \in \mathbb{N}$ such that $g^n = g^m$ and $n \neq m$. We may assume that $n < m$, and then we have $g^{m-n} = g^{n-n} = g^0 = 1$. Since $n \neq m$, we get $m - n \in \mathbb{N}$. So g has finite order. \square

Let's look at some examples:

Example 4.28.

- (1) in $(\mathbb{Z}_4, +)$, we have $|1| = 4$ and $|2| = 2$
- (2) in $U(8)$, we have $|1| = 1$, and $|g| = 2$ for any other $g \in U(8)$, because $g^2 = 1$ for all g .
- (3) in $(\mathbb{Z}, +)$, every element except 0 has infinite order.
- (4) in $\text{Gl}_2(\mathbb{R})$
 - $\left| \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right| = 4$
 - $\left| \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \right| = \infty$

5. TOOLS TO STUDY GROUPS

5.1. Subgroups. Just as sets have *subsets*, groups have *subgroups*. We don't want a subgroup H of (G, \cdot) to just be a subset of G . We want it to be a group as well, *under the same operation!* In practice, this means the following:

Definition 5.1. Let (G, \cdot) be a group. A subgroup of G is a subset H of G such that:

- $1 \in H$,
- for all $g, h \in H$, we have $gh \in H$,
- for all $h \in H$, we have $h^{-1} \in H$.

We write $H < G$ to state that H is a subgroup of G (or sometimes $H \leq G$). If $H < G$ and $H \neq G$, we write $H \leq G$.

Let's look at a few examples.

Example 5.2. Consider (\mathbb{R}^*, \cdot) , where $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$. Then (\mathbb{Q}, \cdot) is a subgroup. We check the three properties:

- $1 \in \mathbb{Q}^*$,
- if $g, h \in \mathbb{Q}^*$, then $gh \in \mathbb{Q}^*$ (done in MATH3345),
- if $g \in \mathbb{Q}^*$, then $g^{-1} = \frac{1}{g} \in \mathbb{Q}^*$ (also 3345).

Example 5.3. For any $n \in \mathbb{Z}^+$, let $n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$. For all $n \in \mathbb{Z}$, the set $n\mathbb{Z}$ is a subgroup of \mathbb{Z} . Fix some $n \in \mathbb{Z}^+$, let's check the three properties.

- (1) $0 \in n\mathbb{Z}$ as $0 = 0 \times n$.
- (2) let $x, y \in n\mathbb{Z}$, then there are $a, b \in \mathbb{Z}$ such that $x = na$ and $y = nb$. So $x + y = n(a + b)$.
- (3) let $x \in n\mathbb{Z}$, so $x = na$ for some $a \in \mathbb{Z}$. Then $-x = n \times (-a)$.

In fact these are the only subgroups of \mathbb{Z} :

Theorem 5.4. *The subgroups of \mathbb{Z} are exactly the $n\mathbb{Z}$, for $n \in \mathbb{Z}^+$.*

Proof. We know that the $n\mathbb{Z}$ are subgroups by the previous proposition, so we only have to show that any subgroup of \mathbb{Z} is of that form. Let H be a subgroup of \mathbb{Z} , then $0 \in H$ so $H \neq \emptyset$. Moreover, note that for any $x, y \in H$ and $c \in \mathbb{Z}$, we have both $x + y \in H$, and $cx \in H$. We conclude, by Lemma 3.4, that there is $n \in H$ such that $H = \{nk : k \in \mathbb{Z}\} = n\mathbb{Z}$. \square

We can find the subgroups of the $(\mathbb{Z}_n, +)$ as well:

Theorem 5.5. *The non-trivial subgroups of $(\mathbb{Z}_n, +)$ are exactly those of the form $d\mathbb{Z}_n = \{dk : k \in \mathbb{Z}_n\}$ for d dividing n .*

Proof. Fix some $n \in \mathbb{N}$. First we have to show that these are subgroups. Let d be a divisor of n , and consider the subset $d\mathbb{Z}_n$. Then:

- $0 = d \times 0 \in d\mathbb{Z}_n$.
- Let $x, y \in d\mathbb{Z}_n$, so $x = da$ and $y = db$ for some $a, b \in \mathbb{Z}_n$. Then $x + y = d(a + b) \in d\mathbb{Z}_n$.
- Let $x \in d\mathbb{Z}_n$, so there is $a \in \mathbb{Z}_n$ such that $x = da$. Then $-x = d(-a) \in d\mathbb{Z}_n$.

Conversely, consider some subgroup $H < \mathbb{Z}_n$. Let $\tilde{H} = \mathbb{N} \cap \{k \in \mathbb{Z} : [k] \in H\}$.

By the well-ordering principle, the set \tilde{H} must have a least element, call it d . We will show that for any $h \in \mathbb{Z}$ with $[h] \in H$, we have $d|h$.

It is enough to show it for $h > 0$, because it is true for $h = 0$, and if $h < 0$, it is enough to show it for $-h > 0$. So let $h \in \mathbb{N}$ with $[h] \in H$. By Euclidian division, there are $r, q \in \mathbb{Z}$ with:

- $h = dq + r$
- $0 \leq r < d$.

By the first line, we get $r = h - dq$, so $[r] = [h] - q[d] \in H$. Since d is the least element of \tilde{H} , we must have $r = 0$. So $d|h$.

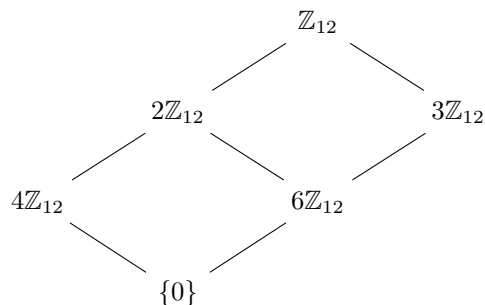
This means that $h = kd$ for some $k \in \mathbb{Z}$, and so $[h] = k[d] \in H$. So $H = d\mathbb{Z}_n$.

We still have to show that $d|n$. It is enough to show $\gcd(n, d) = d$. There are $x, y \in \mathbb{Z}$ such that $\gcd(n, d) = xd + yn$. In particular, we see that $\gcd(n, d) \in \tilde{H}$. We also have $1 \leq \gcd(n, d) \leq d$. Because d is the least element of \tilde{H} , this implies $\gcd(n, d) = d$. □

Example 5.6. Let's find all the subgroups of $(\mathbb{Z}_{12}, +)$. We have one for each divisor of 12, and the trivial subgroup $\{0\}$. In a list:

- $1\mathbb{Z}_{12} = \mathbb{Z}_{12}$,
- $2\mathbb{Z}_{12} = \{0, 2, 4, 6, 8, 10, 12\}$,
- $3\mathbb{Z}_{12} = \{0, 3, 6, 9\}$,
- $4\mathbb{Z}_{12} = \{0, 4, 8\}$,
- $6\mathbb{Z}_{12} = \{0, 6\}$

We can organize them into what's called the *subgroup lattice* of \mathbb{Z}_{12} :



Where a dash between an upper subgroup U and a lower subgroup L means $L < U$.

In general, we can draw the subgroup lattice of any finite group (and parts of the one of an infinite group). But it may get very complicated.

Let's look at a more complicated example.

Example 5.7. Recall that $\text{Gl}_2(\mathbb{R}) = \left\{ M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R}, ad - bc \neq 0 \right\}$, where $\det(M) = ad - bc$ is the *determinant* of the matrix. Consider $(\text{Gl}_2(\mathbb{R}), \cdot)$, where \cdot is matrix multiplication, it is a group (existence of the inverse matrix is equivalent to non-zero determinant). We consider $\text{Sl}_2(\mathbb{R}) = \{M \in \text{Gl}_2(\mathbb{R}) : \det(M) = 1\}$, it is a subgroup. It is easy to see that the determinant of the identity matrix $\text{id} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is 1.

The key property to check is that if M, N are matrices, then $\det(M \cdot N) = \det(M) \det(N)$. From this follows:

- if $M, N \in \text{Sl}_2(\mathbb{R})$, then $\det(M \cdot N) = \det(M) \det(N) = 1 \times 1 = 1$,
- if $M \in \text{Gl}_2(\mathbb{R})$, then $1 = \det(\text{id}) = \det(MM^{-1}) = \det(M) \det(M^{-1})$, so $\det(M^{-1}) = \frac{1}{\det(M)}$. In particular, we see that if $M \in \text{Sl}_2(\mathbb{R})$, then $\det(M^{-1}) = 1$.

One of the nice things about sets is that we could combine them in various ways: union, intersection, product ... Some of these have nice group counterparts.

Proposition 5.8. *Let G be a group and H, K be subgroups of G . Then $H \cap K$ is a subgroup of G .*

Proof. Exercise. □

The union has a group counterpart, but it is a bit more complicated. We save it for later.

The product behave nicely:

Definition 5.9. Let G, H be two groups. The *cartesian product* of G and H , denoted $G \times H$, is given, as a set, by the cartesian product $G \times H$, and we define the group operation as follows:

$$(g_1, h_1)(g_2, h_2) = (g_1g_2, h_1h_2)$$

for any $g_1, g_2 \in G$ and $h_1, h_2 \in H$.

Exercise 5.10. *Check that this defines a group.*

Finally, here is a criterion to decide whether a subset is a subgroup. In practice, it does not save as much time as one might think.

Proposition 5.11. *A non-empty subset $H < G$ is a subgroup if and only if for all $g, h \in H$, we have $gh^{-1} \in H$.*

Proof. Exercise. □

Warning. If you decide to use this (which I do not recommend), you have to check that H is non-empty!

5.2. Morphisms. One overarching question in the study of groups is the following:

Question. *Can we classify all possible groups?*

Note that this is not really a well-defined question. For example, consider the groups $(\mathbb{Z}_2, +)$ and $(\{-1, 1\}, \times)$. Both groups are of order 2, and their only non-identity element has order 2 as well. So in some sense, these groups are the same. One of the goals of this section is to make this into a precise definition.

In the previous section, we saw that for groups, it is nicer to look at subgroups rather than simply subsets. What about functions?

Question. Are there special functions that are well-suited to the study of groups?

Recall our previous example: the group of 2×2 invertible $\text{GL}_2(\mathbb{R})$ has $\text{SL}_2(\mathbb{R})$, the matrices of determinant 1, as a subgroup. The key point of the proof was to check that for any $M, N \in \text{GL}_2(\mathbb{R})$, we have $\det(M \cdot N) = \det(M) \det(N)$. We make this into a definition:

Definition 5.12. Let G and H be two groups. A group *morphism* (or homomorphism) ϕ from G to H is a function $\phi : G \rightarrow H$ such that for all $g, h \in G$, we have $\phi(gh) = \phi(g)\phi(h)$.

By our previous discussion:

Example 5.13. The determinant \det is a morphism from $(\text{GL}_2(\mathbb{R}), \cdot)$ to (\mathbb{R}^*, \cdot) .

Example 5.14. Consider the map $\text{sign} : \mathbb{R}^* \rightarrow \{-1, 1\}$ defined by $\text{sign}(x) = 1$ if x is positive, and $\text{sign}(x) = -1$ if x is negative. Then sign is a morphism from (\mathbb{R}^*, \cdot) to $(\{-1, 1\}, \cdot)$.

Example 5.15. Let $\phi : (\mathbb{Z}, +) \rightarrow (\mathbb{Q}^*, \cdot)$ be defined as $\phi(n) = 2^n$ for all n . Then $\phi(n + m) = 2^{n+m} = 2^n \cdot 2^m$, so ϕ is a morphism.

A classic example from calculus:

Example 5.16. $\exp : (\mathbb{R}, +) \rightarrow (\mathbb{R}^{>0}, \cdot)$ and $\ln : (\mathbb{R}^{>0}, \cdot) \rightarrow (\mathbb{R}, +)$ are morphisms.

We also have an important general example:

Example 5.17. Let G be a group, and fix some $h \in G$. Then the function $\phi : G \rightarrow G$ such that $\phi(g) = h^{-1}gh$ for all $g \in G$ is a morphism.

Let's check it. Let $g_1, g_2 \in G$, then:

$$\begin{aligned} \phi(g_1)\phi(g_2) &= h^{-1}g_1hh^{-1}g_2h \\ &= h^{-1}g_1g_2h \\ &= \phi(g_1g_2). \end{aligned}$$

This morphism is often very useful when studying groups, and it is called *conjugation by h* .

We also have an important morphism relating $(\mathbb{Z}, +)$ and $(\mathbb{Z}_n, +)$:

Proposition 5.18. The map $\pi : (\mathbb{Z}, +) \rightarrow (\mathbb{Z}_n, +)$ defined by $\pi(x) = [x]$ (the class of x modulo n) is a surjective group morphism.

Proof. We just have to prove that for all $x, y \in \mathbb{Z}$, the equality $\pi(x + y) = [x + y] = [x] + [y] = \pi(x) + \pi(y)$ holds, which was proven in MATH 3345. \square

Here is a summary of all the nice properties of morphisms:

Proposition 5.19. Let $\phi : G_1 \rightarrow G_2$ be a morphism of group. Then:

- (1) If e_1 is the identity of G_1 , then $\phi(e_1)$ is the identity of G_2 ,
- (2) For any $g \in G$, we have $\phi(g^{-1}) = \phi(g)^{-1}$.
- (3) If H_1 is a subgroup of G_1 , then $\phi(H_1)$ is a subgroup of G_2 .
- (4) If H_2 is a subgroup of G_2 , then $\phi^{-1}(H_2)$ is a subgroup of G_1 .

Proof. Let e_2 be the identity of G_2 .

(1) We have:

$$\begin{aligned} e_2\phi(e_1) &= \phi(e_1) \\ &= \phi(e_1 \cdot e_1) \\ &= \phi(e_1)\phi(e_1) \end{aligned}$$

and by cancellation we get $e_2 = \phi(e_1)$

(2) Let $g \in G$. Then we have:

$$\begin{aligned} \phi(g)\phi(g^{-1}) &= \phi(gg^{-1}) \\ &= \phi(e_1) \end{aligned}$$

and similarly $\phi(g^{-1})\phi(g) = \phi(e_1)$. Since $\phi(e_1) = e_2$, we conclude that $\phi(g)^{-1} = \phi(g^{-1})$.

(3) Let H_1 be a subgroup of G_1 . First, since $e_1 \in H_1$, we get $e_2 = \phi(e_1) \in \phi(H_1)$. Now let $h, k \in \phi(H_1)$. This means that there are $f, g \in H_1$ such that $\phi(f) = h$ and $\phi(g) = k$. Therefore

$$\begin{aligned} hk &= \phi(f)\phi(g) \\ &= \phi(fg) \end{aligned}$$

and because H_1 is a subgroup of G , we have $fg \in H_1$, therefore $hk = \phi(fg) \in \phi(H_1)$. We can similarly prove that $h^{-1} \in \phi(H_2)$.

(4) Let H_2 be a subgroup of G_2 . We know that $\phi(e_1) = e_2$, and since $e_2 \in H_2$, this means that $e_1 \in \phi^{-1}(H_2)$. Now let $f, g \in \phi^{-1}(H_2)$. This means that $\phi(f) \in H_2$, and we have:

$$\phi(f)^{-1} = \phi(f^{-1})$$

and since H_2 is a subgroup, we get $\phi(f)^{-1} \in H_2$. This means that $\phi(f^{-1}) \in H_2$, so $f^{-1} \in \phi^{-1}(H_2)$. Similarly, we can prove that $\phi(f)\phi(g) \in \phi^{-1}(H_2)$. \square

In particular, the preimage of the identity is of crucial importance, and has a special name:

Definition 5.20. Let $\phi : G \rightarrow H$ be a morphism of group, and e_H the identity of H . Then $\phi^{-1}(e_H)$ is a subgroup of G , and we call it the *kernel* of ϕ , denoted $\ker(\phi)$.

The kernel allows us to check that a morphism is injective very quickly:

Proposition 5.21. *Let $\phi : G \rightarrow H$ be a group morphism. Then ϕ is injective if and only if $\ker(\phi) = \{e_G\}$.*

Proof. Assume first that ϕ is injective. Since $\phi(e_G) = e_H$, we have $\{e_G\} \subset \ker(\phi)$. Let $g \in \ker(\phi)$, then $\phi(g) = e_H = \phi(e_G)$, so $g = e_G$ since ϕ is injective. Therefore $\ker(\phi) = \{e_G\}$.

Now assume that $\ker(\phi) = \{e_G\}$. Let $g_1, g_2 \in G$, and assume that $\phi(g_1) = \phi(g_2)$. Then we have:

$$\begin{aligned} \phi(g_1g_2^{-1}) &= \phi(g_1)\phi(g_2)^{-1} \\ &= \phi(g_1)\phi(g_1)^{-1} \\ &= e_H \end{aligned}$$

so $g_1 g_2^{-1} \in \ker(\phi)$, which by assumption implies that $g_1 g_2^{-1} = e_G$. So $g_1 = g_2$. \square

One useful aspect of morphisms is that they allow us to compare groups. For example, they give us a way to tell when two groups are essentially the same:

Definition 5.22. A bijective group morphism $\phi : G \rightarrow H$ is called an *isomorphism*. If G and H are two groups such that there is an isomorphism $\phi : G \rightarrow H$, we say that G and H are *isomorphic*, and we write $G \simeq H$.

Example 5.23. $\exp : (\mathbb{R}, +) \rightarrow (\mathbb{R}^{>0}, \cdot)$ is an isomorphism. We know it is a morphism, and it is a bijection because \ln is its inverse function.

Example 5.24. $\det : (\text{GL}_n(\mathbb{R}), \cdot) \rightarrow (\mathbb{R}^*, \cdot)$ is not an isomorphism, because it is not injective. For example, we have $\det(\text{id}) = 1 = \det(-\text{id})$. Note that it is, however, surjective. For example $\begin{vmatrix} a & 0 \\ 0 & 1 \end{vmatrix} = a$ for all $a \in \mathbb{R}^*$.

Example 5.25. The map $\phi : (\mathbb{Z}, +) \rightarrow (\mathbb{Q}, \times)$ with $\phi(n) = 2^n$ is not an isomorphism. It is injective, but is not surjective. For example $3 \notin \phi(\mathbb{Z})$.

In the exponential and logarithm example, note that the inverse function \ln is also a morphism (and thus an isomorphism). This is a general fact:

Proposition 5.26. *If $\phi : G \rightarrow H$ is an isomorphism, then its inverse function ϕ^{-1} is also one.*

Proof. The inverse of a bijection is always a bijection, so we only have to prove that ϕ^{-1} is a morphism.

Let $x, y \in H$, then we have:

$$\begin{aligned} \phi(\phi^{-1}(x)\phi^{-1}(y)) &= \phi(\phi^{-1}(x))\phi(\phi^{-1}(y)) \\ &= xy \\ &= \phi(\phi^{-1}(xy)) \end{aligned}$$

and since ϕ is injective, this implies that $\phi^{-1}(xy) = \phi^{-1}(x)\phi^{-1}(y)$. \square

Here is a general principle about group isomorphisms:

Meta-theorem: If G and H are isomorphic, then they have the exact same group properties.

For example, the groups $(\mathbb{R}, +)$ and $(\mathbb{R}^{>0}, \cdot)$ have the same exact group properties, because they are isomorphic. Here are a few:

- abelian,
- every non-identity element is of infinite order,
- fix $n \in \mathbb{N}$. For every $x \in \mathbb{R}$, there is y such that $ny = x$. Similarly, for every $x \in \mathbb{R}^{>0}$, there is y such that $y^n = x$. We say that they are *divisible*.

In practice, this is also very useful to show that two groups G and H are not isomorphic:

Step 1. Find a group property that G has but H has not.

Step 2. Prove that this property is preserved by isomorphism.

For example, we can prove:

Proposition 5.27. *If $\phi : G \rightarrow H$ is a group isomorphism, then for all $g \in G$, we have $\text{ord}(g) = \text{ord}(\phi(g))$.*

Proof. Let $\phi : G \rightarrow H$ be an isomorphism and $g \in G$. First, assume that g has infinite order. Assume, for a contradiction, that $\phi(g)$ has finite order, then there is $n \in \mathbb{N}$ such that $\phi(g)^n = 1$, and therefore $\phi(g^n) = 1$. Because ϕ is an isomorphism, it is injective, and therefore $g^n = 1$.

Second, assume that g has finite order n , so $g^n = 1$. Then $\phi(g)^n = \phi(g^n) = 1$. If $d < n$ is another natural number such that $\phi(g)^d = 1$, then we have $1 = \phi(g)^n \phi(g)^{-d} = \phi(g^{n-d})$, and again because ϕ is injective, this implies $g^{n-d} = 1$, which is a contradiction as n is the order of g . So n is the smallest natural number such that $\phi(g)^n = 1$, i.e. $\text{ord}(\phi(g)) = n$. \square

Example 5.28. The groups $(\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$ and $(\mathbb{Z}_4, +)$ are not isomorphic.

Proof. Recall that $(\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$ is the group of pairs (a, b) with $a, b \in \mathbb{Z}_2$ and group operation coordinate-wise. In particular, notice that every element in $(\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$ is of order 2: if $(a, b) \in (\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$, then $(a, b) + (a, b) = (2a, 2b) = (0, 0)$ because all element in $(\mathbb{Z}_2, +)$ are of order 2. But in \mathbb{Z}_4 , the element 1 has order four. By the previous lemma, the two groups cannot be isomorphic. \square

However, there is another group we've seen that is isomorphic to $(\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$. Recall that $U(8) = \{1, 3, 5, 7\}$ is the multiplicative group of invertible elements of \mathbb{Z}_8 , and it has the following Cayley table:

\times	1	3	5	7
1	1	3	5	7
3	3	1	7	5
5	5	7	1	3
7	7	5	3	1

We have:

Example 5.29. The groups $(\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$ and $(U(8), \cdot)$ are isomorphic.

Proof. We define a morphism $\phi : (\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +) \rightarrow (U(8), \cdot)$ as follows. First, set $\phi((0, 0)) = 1$, there is no choice for this. Now set $\phi((0, 1)) = 3$ and $\phi((1, 0)) = 5$. Because we want ϕ to be a morphism, we have no choice but to set

$$\begin{aligned} \phi((1, 1)) &= \phi((0, 1) + (1, 0)) \\ &= \phi((0, 1)) \cdot \phi((1, 0)) \\ &= 3 \cdot 5 \\ &= 7 \end{aligned}$$

so we let $\phi((1, 1)) = 7$.

This defines all values of ϕ . I leave it as an exercise to check that it is a morphism (i.e. compatibility with the group operations). \square

Note that there were choices in the previous example, as we could have sent $(0, 1)$ and $(1, 0)$ to any two non-identity element of $U(8)$. Later, we will explain why we had choices by looking at group *automorphisms*.

To conclude this subsection, let me mention that we can make our classification question precise. We have:

Proposition 5.30. *Being isomorphic defines an equivalence relations on groups.*

Proof. We have to check three things:

- a group is isomorphic to itself. This is done via the identity morphism $\text{id}_G : G \rightarrow G$ with $\text{id}_G(g) = g$ for all $g \in G$
- if $\phi : G \rightarrow H$ is an isomorphism, then there is an isomorphism from H to G . We can pick ϕ^{-1} .
- if G is isomorphic to H and H is isomorphic to K , then G is isomorphic to K . Let $\phi : G \rightarrow H$ and $\psi : H \rightarrow K$ be isomorphism. We know from MATH 3345 that $\psi \circ \phi$ is a bijection. Therefore we just have to show it is a morphism. Let $g, f \in G$, then:

$$\begin{aligned} (\psi \circ \phi)(gf) &= \psi(\phi(gf)) \\ &= \psi(\phi(g)\phi(f)) \\ &= \psi(\phi(g))\psi(\phi(f)) \\ &= (\psi \circ \phi)(g)(\psi \circ \phi)(f) \end{aligned}$$

so $\psi \circ \phi$ is a morphism. □

In that proof, we have obtained the more general:

Proposition 5.31. *The composition of two group morphisms is a group morphism.*

Isomorphism from a group to itself are particularly important:

Definition 5.32. Let G be a group. A group isomorphism $\phi : G \rightarrow G$ is called an *automorphism*. We denote $\text{Aut}(G)$ the set of automorphisms of G .

Note that as a corollary of the proof, we have obtained:

Corollary 5.33. *Let G be a group. Then $\text{Aut}(G)$ is a group under composition.*

Going back to example 5.29, there were multiple choices for isomorphisms between $(\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$ and $(U(8), \cdot)$ because these two groups have non-trivial automorphism groups.

Knowing that being isomorphic is an equivalence relation, we can ask:

Question. *Can we classify all groups, up to isomorphisms?*

This is currently very far from achieved, even for finite groups. We know the complete answer for finite abelian groups, and finite *simple* groups (we will define simple later). In the next section, we will classify a very specific type of group.

For finite groups in general, we know all of them up to at least order 2000 (see the paper [1]). Curiously, more than 99% of them are of order 1024. In fact, if we define a 2-group to be a group with order a power of 2, we have the following:

Conjecture 5.34. *Let $\text{Gr}(n)$ be the number of groups of order less than n , and $\text{Gr}_2(n)$ the number of 2-groups of order less than n . Then $\lim_{n \rightarrow \infty} \frac{\text{Gr}_2(n)}{\text{Gr}(n)} = 1$.*

This is open and seems very difficult (so is understanding the asymptotic behavior of Gr).

6. IMPORTANT EXAMPLES

In this section, we will use our newfound knowledge to study important classes of groups.

6.1. Cyclic groups. Recall that we have defined the groups $(\mathbb{Z}_n, +)$, and we also have the well-know group $(\mathbb{Z}, +)$. What they have in common is that any element in them can be written as $m \times 1$, for some $m \in \mathbb{Z}$. This is what we will call a *cyclic* group.

Proposition 6.1. *Let G be a group and $g \in G$. The subset $\langle g \rangle = \{g^n : n \in \mathbb{Z}\}$ is a subgroup of G , called the subgroup generated by g . It is the smallest subgroup containing g .*

Proof. It contains the identity because $g^0 = 1$. Moreover, we have $g^n \cdot g^m = g^{n+m}$ and $(g^n)^{-1} = g^{-n}$ for all $n, m \in \mathbb{Z}$. Therefore $\langle g \rangle < G$.

Let H be a subgroup of G containing g . Then it must contain g^n for all $n \in \mathbb{Z}$, so it must contain $\langle g \rangle$. Therefore $\langle g \rangle$ is the smallest subgroup of G containing g . \square

Definition 6.2. A group G is called *cyclic* if there is $g \in G$ such that $\langle g \rangle = G$. In that case, we call g a generator of G .

Example 6.3. The groups $(\mathbb{Z}, +)$ and $(\mathbb{Z}_n, +)$ are cyclic, generated by 1.

Warning. The generator is not unique. For example, $(\mathbb{Z}, +)$ is generated by both 1 and -1 . And not every element of a cyclic group is a generator, for example 2 does not generate $(\mathbb{Z}, +)$.

Example 6.4. Any cyclic group is abelian, therefore any non-abelian group is not cyclic.

The abelian groups $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ and $U(8)$ are not cyclic.

Cyclic groups are the simplest groups, and we can, in fact, completely classify them:

Theorem 6.5. *A cyclic group is isomorphic to either $(\mathbb{Z}, +)$ or $(\mathbb{Z}_n, +)$ for some $n \in \mathbb{N}$.*

Proof. Let G be a cyclic group and g a generator. Suppose first that g has finite order $n \in \mathbb{N}$. Then we define $\phi : G \rightarrow \mathbb{Z}_n$ by letting $\phi(g^m) = m$. First we check that this is well-defined. If $g^m = g^k$ for some $m, k \in \mathbb{Z}$, then $g^{m-k} = 1$, which implies that $n|m - k$, and thus $m - k = 0$ in \mathbb{Z}_n . It defines a function on G since g is a generator of G . Checking that ϕ is a morphism is straightforward. Finally, it is surjective because \mathbb{Z}_n is generated by 1. For injectivity, if $\phi(g^m) = \phi(g^k)$ for some $m, k \in \mathbb{Z}$, then $m - k = \phi(g^m) - \phi(g^k) = 0$ in \mathbb{Z}_n , which implies $n|m - k$, and thus $g^m = g^k$.

Now assume that g has infinite order, we define $\phi : G \rightarrow \mathbb{Z}$ by letting $\phi(g^m) = m$ for all $m \in \mathbb{N}$. This is well-defined because $g^m = g^k$ implies $k = m$, and G is generated by g . It is injective because if $\phi(g^m) = \phi(g^k)$ for some $m, k \in \mathbb{N}$, then $m = k$ in \mathbb{Z} , and therefore $g^m = g^k$. And it is surjective because 1 generates \mathbb{Z} . \square

6.2. Multiplicative group of the complex numbers and the circle group.

Recall that the *complex numbers* are given by:

$$\{a + bi : a, b \in \mathbb{R}\}$$

where $i^2 = -1$. We can add and multiply complex numbers. As a reminder, the operations are given as follows, for $a, b, c, d \in \mathbb{R}$:

- $(a + ib) + (c + id) = a + b + (c + d)i$
- $(a + ib)(c + id) = ac + iad + ibc - bd = ac - bd + i(ad + bc)$.

We see that any element has an additive inverse, so $(\mathbb{C}, +)$ is a group, and:

Proposition 6.6. $(\mathbb{C}, +)$ is isomorphic to $(\mathbb{R}^2, +)$.

Does multiplication of complex numbers form a group? And is it isomorphic to a group we know? To prove it is, we will use the following:

Definition 6.7. Let $z = a + ib$ be a complex number. Its *complex conjugate* \bar{z} is given by $\bar{z} = a - ib$, and $z\bar{z} = a^2 + b^2$. The quantity $\sqrt{z\bar{z}}$ is called the *absolute value* of z .

We now prove:

Proposition 6.8. Every non-zero complex number is invertible. In other words, (\mathbb{C}^*, \cdot) is a group.

Proof. Let $z = a + ib$, and assume $z \neq 0$, so $a^2 + b^2 \neq 0$. We want to find c, d such that

$$(a + ib)(c + id) = 1.$$

For any $c + id \in \mathbb{C}$, we have

$$\begin{aligned} (a + ib)(c + id) = 1 &\Rightarrow (a - ib)(a + ib)(c + id) = a - ib \\ &\Rightarrow (a^2 + b^2)(c + id) = a - ib \\ &\Rightarrow c + id = \frac{a}{a^2 + b^2} + i\frac{-b}{a^2 + b^2} \text{ because } a^2 + b^2 \neq 0 \end{aligned}$$

And conversely, we can check that $(a + ib)^{-1} = \frac{a}{a^2 + b^2} + i\frac{-b}{a^2 + b^2}$.

□

Note that from the formula for multiplication, we see that:

Proposition 6.9. The map $z \rightarrow \bar{z}$ is an isomorphism from (\mathbb{C}^*, \cdot) to (\mathbb{C}^*, \cdot) .

I now want to prove that (\mathbb{C}^*, \cdot) is isomorphic to a matrix group with real coefficients.

Here is an heuristic for what we'll do: we can think of $1, i \in \mathbb{C}$ as the base vectors $(0, 1)$ and $(1, 0)$ in the real plane. Now, multiplying by a complex number gives us a linear transformation of the real plane. In fact, the transformations given by 1 and i are as follows

$$\begin{aligned} (0, 1) &\rightarrow (0, 1) \\ (1, 0) &\rightarrow (1, 0) \end{aligned}$$

and

$$\begin{aligned} (0, 1) &\rightarrow (1, 0) \\ (1, 0) &\rightarrow -(1, 0) \end{aligned}$$

So writing them as matrices, we get $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Any complex number is given by $a + ib$, which we write as $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. We prove:

Proposition 6.10. *The set of matrices*

$$G = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a, b \in \mathbb{R}, a^2 + b^2 \neq 0 \right\}$$

is a group under matrix multiplication, isomorphic to (\mathbb{C}, \cdot) .

Proof. We follow the intuition established in the previous paragraph, and define a map:

$$\begin{aligned} G &\rightarrow \mathbb{C}^* \\ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} &\rightarrow a + ib \end{aligned}$$

The image of this map is \mathbb{C}^* because of the condition that $a^2 + b^2 \neq 0$, which is equivalent to $a + ib \neq 0$. It is straightforward to see that this map is bijective, so we just show that it is a morphism. Let $a, b, c, d \in \mathbb{R}$, with $a^2 + b^2 \neq 0$ and $c^2 + d^2 \neq 0$, and let $M = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ and $N = \begin{pmatrix} c & -d \\ d & c \end{pmatrix}$. Then

$$\begin{aligned} \phi(MN) &= \phi \left(\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} c & -d \\ d & c \end{pmatrix} \right) \\ &= \phi \left(\begin{pmatrix} ac - bd & -(ad + bc) \\ ad + bc & ac - bd \end{pmatrix} \right) \\ &= ac - bd + i(ad + bc) \\ &= \phi(M)\phi(N) \end{aligned}$$

□

Recall that we know another special group of $\text{GL}_2(\mathbb{R})$: the special linear group $\text{SL}_2(\mathbb{R})$ of matrices of determinant 1.

Definition 6.11. As a group of matrices, the circle group is given by

$$\mathbb{S}^1 = G \cap \text{SL}_2(\mathbb{R}) .$$

As a group of complex numbers, it is given by complex numbers $a + ib$ such that $a^2 + b^2 = 1$, i.e. it is the complex circle.

We in fact have already seen that group on homework 6: we have

$$G \cap \text{SL}_2(\mathbb{R}) = \{a, b \in \mathbb{R}, a^2 + b^2 = 1\} = \text{SO}_2(\mathbb{R})$$

so the circle group and the special orthogonal group are isomorphic!

Again thinking of the complex number as the real plane, we see that any element of the circle group can be written as $\cos(\theta) + i \sin(\theta)$, for some $\theta \in \mathbb{R}$. In fact, we can write any complex number in *polar coordinates*:

Fact 6.12. *If z is any non-zero complex number, there are unique $r \in (0, +\infty)$ and $\theta \in [0, 2\pi)$ such that*

$$z = r(\cos(\theta) + i \sin(\theta)) .$$

Moreover, if $z_1 = r_1(\cos(\theta_1) + i \sin(\theta_1))$ and $z_2 = r_2(\cos(\theta_2) + i \sin(\theta_2))$ are complex numbers, then

$$z_1 z_2 = r_1 r_2 (\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2))$$

and

$$|z_1| = r_1 .$$

Proof. Some computation involving trigonometry, essentially done in Homework 6. \square

In particular, given two elements of the circle group associated to the angles θ_1 and θ_2 , we obtain their product by adding together the two angles.

As a corollary, we obtain a morphism:

Corollary 6.13. *The map*

$$\begin{aligned} \phi : (\mathbb{R}^{>0}, \cdot) \times (\mathbb{R}, +) &\rightarrow (\mathbb{C}^*, \cdot) \\ (r, \theta) &\rightarrow r(\cos(\theta) + i \sin(\theta)) \end{aligned}$$

is a surjective group morphism.

However, note that it is not an isomorphism. In fact we can determine its kernel:

Proposition 6.14. *The kernel of the map*

$$\begin{aligned} \phi : (\mathbb{R}^{>0}, \cdot) \times (\mathbb{R}, +) &\rightarrow (\mathbb{C}^*, \cdot) \\ (r, \theta) &\rightarrow r(\cos(\theta) + i \sin(\theta)) \end{aligned}$$

is the subgroup $\{1\} \times 2\pi\mathbb{Z}$, where $2\pi\mathbb{Z} = \{2\pi k : k \in \mathbb{Z}\}$.

Proof. Let $(r, \theta) \in \ker(\phi)$, so $r(\cos(\theta) + i \sin(\theta)) = 1$. This implies $r = |r(\cos(\theta) + i \sin(\theta))| = |1| = 1$, so $r = 1$. Then we must have $1 = \cos(\theta) + i \sin(\theta)$, which implies $\theta \in 2\pi\mathbb{Z}$. \square

For another example, we have $i = \phi(1, \frac{\pi}{2}) = \phi(1, \frac{\pi}{2} + 2\pi)$. We will see later, when we introduce quotient groups, how to make it into an isomorphism.

The quantity $\cos(\theta) + i \sin(\theta)$ is actually the *complex exponential* in disguise! We will admit the following:

Fact 6.15. *The series*

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}$$

converges for any complex number z . It defines a function, the *complex exponential*, denoted $\exp(z)$ or e^z .

For any $\theta \in \mathbb{R}$, we have $e^{i\theta} = \cos(\theta) + i \sin(\theta)$.

Moreover, for any $z, w \in \mathbb{C}$, we have $e^z e^w = e^{z+w}$.

As a consequence of that fact, for any complex number z , there are $r \in \mathbb{R}^{\geq 0}$ and $\theta \in \mathbb{R}$ such that:

$$z = r(\cos(\theta) + i \sin(\theta)) = r e^{i\theta} .$$

Note that if you admit convergence of all series involved, the identities

$$e^{i\theta} = \cos(\theta) + i \sin(\theta)$$

and

$$e^z e^w = e^{z+w}$$

can be proven using direct computations.

The circle group (\mathbb{S}^1, \cdot) is of infinite order, but also has elements of finite order (we have seen plenty on HW6). For example, we have $1 = 1^4 = (-1)^4 = i^4 = (-i)^4$.

In fact, elements of finite order in \mathbb{S}^1 correspond exactly to the *roots of unity* in \mathbb{C} : solution of the equation $z^n = 1$, for some $n \in \mathbb{N}$.

Theorem 6.16. *Fix some $n \in \mathbb{N}$. The solutions of $z^n = 1$ in \mathbb{C} are contained in \mathbb{S}^1 , they are called the n -th roots of unity.*

Moreover, the n -th root of unity are a cyclic subgroup of \mathbb{S}^1 , given by

$$\{e^{i\frac{2k\pi}{n}} : 0 \leq k \leq n-1\} .$$

Proof. Let $z \in \mathbb{C}^*$ be a solution of $z^n = 1$. Writing $z = re^{i\theta}$ for some $r \in (0, +\infty)$ and $\theta \in \mathbb{R}$, we obtain $1 = z^n = r^n e^{in\theta}$, and in particular we must have $r = 1$ as $r \in (0, +\infty)$, which implies $z \in \mathbb{S}^1$.

To prove the moreover part, note that if $z^n = 1$ and $w^n = 1$, then $(zw)^n = z^n w^n = 1$, and $z^{-n} = \frac{1}{z^n} = 1$. Therefore the n -th roots of unity form a subgroup. If z is an n -th root of unity, writing $z = e^{i\theta}$ for some $\theta \in \mathbb{R}$, we obtain $1 = z^n = e^{in\theta}$, so $n\theta = 2m\pi$ for some $m \in \mathbb{Z}$, i.e. $\theta = \frac{2m\pi}{n}$. Since we may assume that $\theta \in [0, 2\pi)$, we obtain that the group of n -th roots of unity can be identified as

$$\{e^{i\frac{2k\pi}{n}} : 0 \leq k \leq n-1\} .$$

Finally, the subgroup is cyclic because it is generated by $e^{i\frac{2\pi}{n}}$. \square

We thus have identified all elements of finite order in \mathbb{S}^1 :

Corollary 6.17. *The elements of finite order in \mathbb{S}^1 are given by*

$$\{e^{i2\pi\theta} : \theta \in \mathbb{Q}\} .$$

In particular, they form a subgroup.

Of particular interest are the p^n -th roots of unity, for some prime number p :

Definition 6.18. Let p be a prime number, the Prüfer p -group is

$$Z(p^\infty) = \{z : z^{p^n} = 1 \text{ for some } n \in \mathbb{N}\} .$$

It is an infinite group, but all of its elements have finite order. They have one more interesting property, that of being *divisible*:

Definition 6.19. An abelian group $(G, +)$ is divisible if for all $g \in G$ and $n \in \mathbb{N}$, there is $h \in G$ such that $nh = g$.

The following groups are divisible:

- $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$,
- the Prüfer p -groups,
- the circle group.

The following groups are not divisible:

- any non-trivial finite group,
- $(\mathbb{Z}, +)$.

There is a general theorem that $(\mathbb{Q}, +)$ and the Prüfer p -groups are essentially the only divisible abelian groups, in the sense that any divisible group is "made up" of these two ².

Note that the Prüfer p -groups are infinite, but they also have the following somewhat strange property:

²This is called the structure theorem for divisible groups.

Theorem 6.20. *Let p be a prime number. Any subgroup $H < Z(p^\infty)$ not equal to $Z(p^\infty)$ is finite.*

Proof. Maybe homework. \square

In particular, they are infinite groups in which every element is of finite order.

6.3. Permutation groups. Recall that if A is a set, the set:

$$S_A = \{f : A \rightarrow A : f \text{ is a bijection}\}$$

is a group under composition \circ . We call it the *symmetric group on A* , its elements are called *permutations* of A . If $A = \{1, \dots, n\}$, we denote this group S_n .

Finally, if $\sigma \in S_n$, the *support* of σ is the set

$$\text{supp}(\sigma) = \{x \in \{1, \dots, n\} : \sigma(x) \neq x\}.$$

Remark 6.21. *Let A and B be two sets, and assume that there is a bijection f from A to B . Then we obtain a bijection $\hat{f} : S_A \rightarrow S_B$ given by $\hat{f}(\sigma) = f \circ \sigma \circ f^{-1}$. As an exercise, check that this is a morphism.*

By the previous remark, the symmetric group S_n is isomorphic to the symmetric group of *any* set of size n , which is why we can reduce our study of finite symmetric groups to S_n .

Example 6.22. The symmetric group S_1 is trivial, because the identity is the only permutation of $\{1\}$.

The group S_2 is isomorphic to $(\mathbb{Z}_2, +)$. Indeed, the only non-identity permutation of $\{1, 2\}$ is the one swapping 1 and 2, which has order 2.

Proposition 6.23. *If $n \geq 3$, then S_n is not abelian.*

Proof. Fix some $n \geq 3$. We consider the two permutations σ, τ of $\{1, 2, \dots, n\}$ with σ given by $\sigma(1) = 2, \sigma(2) = 1$ and $\sigma(i) = i$ for any $i \neq 1, 2$, and τ given by $\tau(2) = 3, \tau(3) = 2$, and $\tau(i) = i$ for all $i \neq 2, 3$.

Then

$$\begin{aligned} \sigma \circ \tau(2) &= \sigma(\tau(2)) \\ &= \sigma(3) \\ &= 3 \end{aligned}$$

and a similar computation shows that $\tau \circ \sigma(2) = 1$. Therefore $\sigma \circ \tau \neq \tau \circ \sigma$, so S_n is not abelian. \square

Proposition 6.24. *The group S_n has cardinality $n!$.*

Proof sketch. We can prove it by noticing that any $\sigma \in S_n$ is entirely determined by $(\sigma(1), \dots, \sigma(n)) \in \{1, \dots, n\}^n$, and that $\sigma(i) \neq \sigma(j)$ for all $i \neq j$. Therefore, we have n possibilities for $\sigma(1)$, $n - 1$ possibilities for $\sigma(2)$, ..., which we multiply into $n \times (n - 1) \times \dots \times 1$ possibilities for σ . \square

Let us do some example of composition in S_4 . Consider $\sigma \in S_4$ and $\tau \in S_4$ given as follows:

$$\begin{array}{ll}
 \sigma : 1 \rightarrow 2 & \tau : 1 \rightarrow 4 \\
 2 \rightarrow 4 & 2 \rightarrow 3 \\
 3 \rightarrow 3 & 3 \rightarrow 1 \\
 4 \rightarrow 1 & 4 \rightarrow 2
 \end{array}$$

We compute the composition $\sigma \circ \tau$:

$$\begin{array}{l}
 1 \rightarrow 4 \rightarrow 1 \\
 2 \rightarrow 3 \rightarrow 3 \\
 3 \rightarrow 1 \rightarrow 2 \\
 4 \rightarrow 2 \rightarrow 4
 \end{array}$$

where we apply τ in the first column and σ in the second.

As you can see, this notation is cumbersome, so we condense it into:

Notation. Given $\sigma \in S_n$, we denote it by:

$$\begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n-1) & \sigma(n) \end{pmatrix}$$

So the previous examples will be denoted:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

and

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}$$

In fact, we can have an even more convenient notation. Consider the two elements of S_5 :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 3 & 1 \end{pmatrix}$$

and

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 2 & 4 \end{pmatrix}$$

Notice that σ permutes the element of $\{1, \dots, 5\}$ in a cycle

$$1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 5 \rightarrow 1 .$$

This is not the case for τ . Following 1, we get

$$1 \rightarrow 3 \rightarrow 1$$

and following 2 we get

$$2 \rightarrow 5 \rightarrow 4 \rightarrow 2 .$$

So σ on one cycle of length 5, and τ can be decomposed in a cycle of length 2 and a cycle of length 3.

We make this precise:

Definition 6.25. A *k-cycle* in S_n is a permutation σ such that there are $a_1, \dots, a_k \in \{1, \dots, n\}$ with:

- $\sigma(a_i) = a_{i+1}$ for all $1 \leq i \leq k - 1$ and $\sigma(a_k) = a_1$,
- $\sigma(x) = x$ for all $x \in \{1 \cdots, n\} \setminus \{a_1, \dots, a_k\}$.

The set $\{a_1, \dots, a_k\}$ is the support of the cycle.

In other words, the cycle σ of the definition is the permutation:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_{k-1} & a_k \\ a_2 & a_3 & \cdots & a_k & a_1 \end{pmatrix}.$$

Notation. The cycle

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_{k-1} & a_k \\ a_2 & a_3 & \cdots & a_k & a_1 \end{pmatrix}$$

is denoted $(a_1 a_2 \cdots a_{k-1} a_k)$

So for example, the two first permutations we saw are $\sigma = (1 2 4 3 5)$ and $\tau = (1 3)(2 5 4)$. Note that because the support of $(1 3)$ and $(2 5 4)$ are disjoint, they commute, i.e. $(1 3)(2 5 4) = (2 5 4)(1 3)$.

This true in general:

Proposition 6.26. *If two cycles have disjoint support, then they commute.*

Proof. Fix some n , and consider two cycles $\sigma = (a_1 \cdots a_k)$ and $\tau = (b_1 \cdots b_l)$, for some $1 < k, l \leq n$.

First assume that their supports are disjoint. This implies that $\sigma(b_j) = b_j$ for all j and $\tau(a_i) = a_i$ for all i . From, there, we compute, for example, that

$$\begin{aligned} \sigma\tau(a_i) &= \sigma(a_i) \\ &= a_{i+1} \\ &= \tau(a_{i+1}) \\ &= \tau(\sigma(a_i)) \end{aligned}$$

for all i . A similar computation shows that $\sigma\tau(b_j) = \tau\sigma(b_j)$ for all j . Finally, for any x not equal to a_i or b_j for all i, j , we have $\sigma(x) = \tau(x) = x$, so also $\sigma\tau(x) = x = \tau\sigma(x)$. □

We also know the order of cycles:

Proposition 6.27. *A k -cycle has order k .*

Proof. Let $\sigma = (a_1 \cdots a_k)$ be a k -cycle. For $i < k$, we have $\sigma^i(a_1) = a_{i+1} \neq a_1$, therefore the order of σ is at least k . Moreover, we have $\sigma^k(a_i) = a_i$ for all i , and if $x \notin \{a_1, \dots, a_k\}$, then $\sigma(x) = x$, so $\sigma^k(x) = x$. Therefore $\sigma^k = \text{id}$, so σ has order k . □

One important structural theorem about permutations is the following:

Theorem 6.28. *Every permutation $\sigma \in S_n$ is equal to a product of cycles with disjoint support.*

Proof. We prove this by strong induction on the size of the support of σ . So let $\sigma \in S_n$. If $|\text{supp}(\sigma)| = 0$, then $\sigma = \text{id}$, which is a cycle of length 1.

Now assume that we have proved it for all permutations with support of size at most k , for some $k \in \mathbb{N}$. Let $\sigma \in S_n$ be a permutation with support of size $k + 1$.

Pick some $a \in \text{supp}(\sigma)$, which exists because the support has size at least 1. Consider the set $\{\sigma^i(a) : i \in \mathbb{Z}^+\}$, it is a finite set because it is contained in $\{1, \dots, n\}$. Pick some i such that $\sigma^i(a) = \sigma^j(a)$ for some $j > i$, which exists because otherwise the set $\{1, \dots, n\}$ would be infinite ($i \rightarrow \sigma^i(a)$ would define an injection from \mathbb{N} to $\{1, \dots, n\}$). Then $\sigma^{j-i}(a) = a$, and $j - i > 1$ as $a \in \text{supp}(\sigma)$. This means that the set of natural numbers $m > 1$ such that $\sigma^m(a) = a$ is non-empty, pick its least element l using the well-ordering principle. Then for all $0 < i < k < l$ we must have $\sigma^i(a) \neq \sigma^k(a)$, as otherwise $\sigma^{k-i}(a) = a$ and $0 < k - i < l$.

Let $\tau = (a \sigma(a) \dots \sigma^{l-1}(a))$, and consider the permutation $\sigma \circ \tau^{-1}$. We observe:

- (1) for all i , we have $\sigma \circ \tau^{-1}(\sigma^i(a)) = \sigma^i(a)$,
- (2) for all $x \in \{1, \dots, n\} \setminus \{a, \sigma(a), \dots, \sigma^{l-1}(a)\}$, we have $\sigma \circ \tau^{-1}(x) = \sigma(x)$ because $\tau(x) = x$.

In particular, the permutation $\sigma \circ \tau^{-1}$ has support $\text{supp}(\sigma) \setminus \{a, \sigma(a), \dots, \sigma^{l-1}(a)\}$, and as $l > 1$, we can apply our induction hypothesis and write $\sigma \circ \tau^{-1} = \tau_1 \dots \tau_m$ for some cycles τ_i with disjoint supports. Therefore $\sigma = \tau_1 \dots \tau_m \tau$, and the support of τ is disjoint from the supports of the τ_i because they are all contained in the support of τ ³. \square

Let's understand how this works in some example.

Example 6.29. Elements of S_5 comes in 7 different classes:

- the identity, which is a 1-cycle,
- 2-cycles such as (1 2),
- 3-cycles such as (1 3 5),
- 4-cycles,
- 5-cycles,
- products of 2 2-cycles with disjoint supports, such as (3 5)(2 1)
- product of a 3-cycle and a 2-cycle with disjoint supports, such as (3 5 1)(2 4).

The simplest cycles are *transpositions*, those of the form $(a b)$. They are of order two. Just like we can write every permutation as a product of cycles, we can write them as a product of transposition.

Example 6.30. Consider the cycle (1 2 3 4), we can write it as a product of transpositions as follows:

$$(1 \ 2 \ 3 \ 4) = (1 \ 4)(1 \ 3)(1 \ 2) .$$

This is in fact more general:

Proposition 6.31. *Any cycle is equal to a product of transpositions.*

Proof. The cycle $(a_1 \ a_2 \ \dots \ a_k) \in S_n$ is equal to

$$(a_1 \ a_k)(a_1 \ a_{k-1}) \dots (a_1 \ a_3)(a_1 \ a_2) .$$

\square

and as a consequence:

Corollary 6.32. *Any permutation in S_n is equal to a product of transpositions.*

³This is actually not completely immediate and may be given as homework.

But the decomposition as a product of transpositions is not unique. For example $\text{id} = (1\ 2)(1\ 2) = (2\ 3)(2\ 3)$. What is unique is the parity of the number of transpositions required. In other words, if $\sigma \in S_n$ is written as a product of transpositions in two ways, then either both involve an even number of transpositions, or both involve an odd number of transpositions.

The key observation is:

Lemma 6.33. *If we have $\text{id} = \tau_1 \cdots \tau_k$ in S_n , where the τ_i are transpositions, then k is even.*

Proof. We prove it by strong induction on the number of transpositions needed. If $k = 0$, we're done. Note that $k \neq 1$, because a single transposition is never the identity. So we can assume that $k \geq 2$. Assume that the result is true for every $l < k$: if id is the product of l transpositions, then l is even.

Let $\tau_k = (a\ b)$ for some $a, b \in \{1, \dots, n\}$. There are four cases to treat, depending on the value of τ_{k-1} .

Case 1. if $\tau_{k-1} = (a\ b)$, then $\text{id} = \tau_1 \cdots \tau_{k-2}$.

Case 2. if $\tau_{k-1} = (a\ c)$ for some $c \neq b$, then

$$\begin{aligned}\tau_{k-1}\tau_k &= (a\ c)(a\ b) \\ &= (a\ b\ c) \\ &= (a\ b)(b\ c)\end{aligned}$$

Case 3. if $\tau_{k-1} = (b\ c)$ for some $c \neq a$, then

$$\begin{aligned}\tau_{k-1}\tau_k &= (b\ c)(a\ b) \\ &= (a\ c\ b) \\ &= (a\ c)(b\ c)\end{aligned}$$

Case 4. if $\tau_{k-1} = (c\ d)$ for c, d different from a and b , then

$$\tau_{k-1}\tau_k = (a\ b)(c\ d)$$

In Cases 1., we can apply our induction hypothesis to get that $k - 2$ is even. Therefore k is even as well.

In Case 1. 2. and 3., we have transformed τ into a product of k transpositions such that a does not appear in the last transposition. We can keep applying this process to move the rightmost transposition containing a to the left. One of two possibilities must be true:

- a. we can keep going until we reach the first transposition.
- b. at some point, there is a cancellation, i.e. we are in Case 1.

If a. is true, then id is written as a product of transpositions, of which only the first contain a . But that implies $\text{id}(a) \neq a$, which is a contradiction. So b. must be true. In that case, we have written id as a product of $k - 2$ transpositions, and by induction hypothesis $k - 2$ is even, and therefore k is even. \square

As a corollary:

Corollary 6.34. *For any $\sigma \in S_n$, the parity of decompositions of σ into a product of transpositions is either always even, or always odd. If it is even, we call σ an even permutation, and odd if it is odd.*

This even gives us a morphism:

Theorem 6.35. *The map*

$$\begin{aligned} \text{sign} : S_n &\rightarrow \{-1, 1\} \\ \sigma &\rightarrow -1 \text{ if } \sigma \text{ is odd} \\ \sigma &\rightarrow 1 \text{ if } \sigma \text{ is even} \end{aligned}$$

is a group morphism. It is called the signature.

Proof. That this is a well-defined map is a consequence of the previous corollary. That it is a morphism is a consequence of the following facts:

- the product of two even permutations is even,
- the product of two odd permutations is even,
- the product of an odd and an even permutation is odd.

□

Definition 6.36. The kernel of $\text{sign} : S_n \rightarrow \{-1, 1\}$ is called the *alternating group*, denoted A_n .

In other words, the alternating group is the group of all even permutations.

6.4. Dihedral groups. In this subsection, we will study a special subgroup of S_n : the dihedral group D_n , defined for $n \geq 3$. We assume that $n \geq 3$ for the rest of this subsection.

First let's recall a few things:

Definition 6.37. A *regular n -gon* is an n -sided polygon with all sides of equal length and all angles between sides are equal.

This includes, for example, equilateral triangles, squares, ...

Definition 6.38. Let V_n be the set of vertices of a regular n -gon G . A *symmetry* of G is a bijection f from V_n to itself such that if x, y are adjacent edges, then $f(x)$ and $f(y)$ are adjacent.

Proposition 6.39. *The set of symmetries of a regular n -gon form a group under composition.*

Proof. Let $n \in \mathbb{N}$. If $n = 1$ then D_n contains only the identity, so D_1 is the trivial group. Similarly, we have that D_2 is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

Now assume that $n > 2$, let G be a regular n -gon, and let $f, g \in D_n$. The identity id preserves adjacency, so it is an element of D_n .

Now pick some adjacent vertices x, y of G . Then $g(x)$ and $g(y)$ are adjacent because $g \in D_n$. Therefore $f(g(x))$ and $f(g(y))$ are adjacent, because $f \in D_n$. So $f \circ g \in D_n$.

We also need to show that $f^{-1} \in D_n$, i.e. $f^{-1}(x)$ and $f^{-1}(y)$ are adjacent. Let v and w be the two vertices adjacent to $f^{-1}(x)$. Then $f(v)$ and $f(w)$ are adjacent to x . Since y is adjacent to x and x has exactly two adjacent vertices, we get that $y = f(v)$ or $y = f(w)$. This implies that $f^{-1}(y) = v$ or $f^{-1}(y) = w$, i.e. $f^{-1}(y)$ is adjacent to $f^{-1}(x)$. So $f^{-1} \in D_n$. □

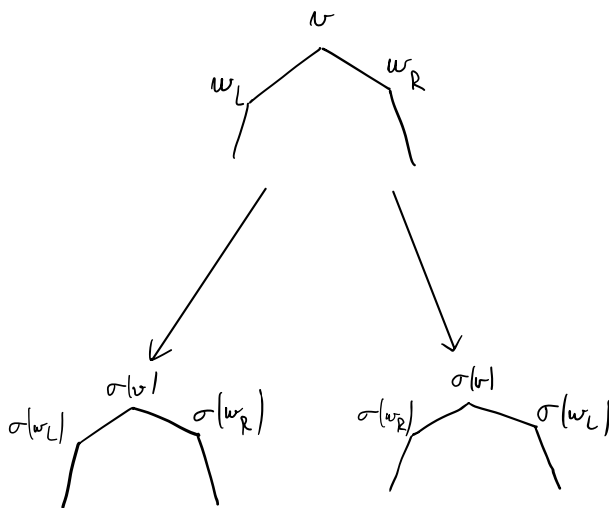
Let D_n be the group of symmetry of a regular n -gon.

Question. *Can we describe the structure of D_n ?*

First, let us compute its size:

Lemma 6.40. *The group D_n has $2n$ elements.*

Proof. Let v be any vertex of v , as well as its two neighbors w_R and w_L , with w_R coming first in the trigonometric direction. Any $\sigma \in D_n$ must send w_R and w_L to the two neighbors of $\sigma(v)$, so there are two possibilities: either $\sigma(w_R)$ comes first in the trigonometric direction, or $\sigma(w_L)$ comes first, see the following picture:



Once $\sigma(v)$ is chosen and we pick $\sigma(w_R)$ or $\sigma(w_L)$ to come first, we have identified σ , as $\sigma(w)$ is determined for any vertex w .

Reciprocally, we can always construct some element of D_n by picking $\sigma(v)$ and an orientation for $\sigma(w_R)$ and $\sigma(w_L)$. □

Ok, but what are the elements of D_n ? Let's try to find some.

- there is r , the rotation by $\frac{2\pi}{n}$ radians.
- if we fix some vertex v , there is also the reflection s fixing v .

Proposition 6.41. *The set $\{\text{id}, r, r^2, \dots, r^{n-1}\}$ is a cyclic subgroup of order n of D_n .*

Proof. This is because r has order n . □

So we have identified $\text{id}, r, r^2, \dots, r^{n-1}$ as elements of D_n . What about the other n elements?

Proposition 6.42. *The other n elements of D_n are $s, sr, sr^2, \dots, sr^{n-1}$, where r is a rotation by $\frac{2\pi}{n}$ radians and s is a reflection around a fixed vertex.*

Proof. First, we have to show that they are all distinct from $\text{id}, r, \dots, r^{n-1}$. This is because, for any vertex v , any of $s, sr, sr^2, \dots, sr^{n-1}$ changes the order of the neighbors v , but none of the $\text{id}, r, \dots, r^{n-1}$ does.

We also have to show that if $sr^i = sr^j$, then $i = j$. So assume that $sr^i = sr^j$ for some $0 \leq i, j \leq n - 1$. Then $r^i = r^j$, which implies that $i = j$ because r has order n . □

We therefore have proven:

Theorem 6.43. *The group D_n is given by*

$$D_n = \{\text{id}, r, \dots, r^{n-1}, s, sr, \dots, sr^{n-1}\}$$

where r is rotation by $\frac{2\pi}{n}$ radians and s is the reflection around some fixed vertex.

In particular, we have obtained:

Corollary 6.44. *The groups of symmetries of any two regular n -gons are isomorphic.*

This is why we can talk about *the* dihedral group D_n .

Note that we have found all elements of D_n , but there are many more products we could take: $rs, rs^{10}, rsr sr^3 s^6 \dots$ we would like to be able to simplify this sort of expression, and this is the goal of the next result.

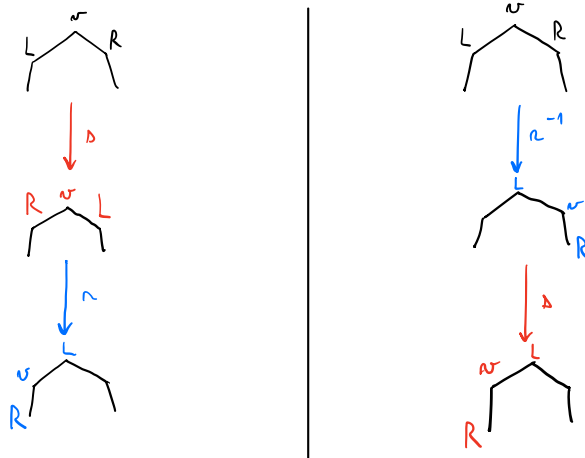
Lemma 6.45. *In D_n , we have the following equalities:*

- (1) $\text{ord}(s) = \text{ord}(sr^i) = 2$,
- (2) $r^i s = sr^{-i}$.

Proof. We have $s \neq \text{id}$. However, we know that $s^2(v) = v$, and it also fixes its neighbors. This implies it is the identity. We will compute the order of sr^i at the end of the proof.

We prove that $r^i s = sr^{-i}$ by induction on i .

For the base case, we need to prove that $rs = sr^{-1}$ (we could start at zero, but in fact we will need the $i = 1$ case in the proof later). Again, to show equality, we just have to show that they both send v and its neighbors to the same vertices. This is showed in the following picture:



Now assume that we have proven that $r^i s = sr^{-i}$ for some $i \leq 1$. Then we compute

$$\begin{aligned} r^{i+1}s &= rr^i s \\ &= rsr^{-i} \text{ by induction hypothesis} \\ &= sr^{-1}r^{-i} \text{ by the } i = 1 \text{ case} \\ &= sr^{-(i+1)}. \end{aligned}$$

Finally, we prove that $\text{ord}(sr^i) = 2$. We have

$$\begin{aligned} (sr^i)^2 &= sr^i sr^i \\ &= s sr^{-i} r^i \\ &= s^2 \\ &= \text{id} . \end{aligned}$$

□

As an aside, we can talk about what happens when $n = 1$ or $n = 2$. If $n = 1$, a regular 1-gon is just a point, and its group of symmetry is trivial. If $n = 2$, a regular 2-gon is a line segment. Its group of symmetry consists of the identity and the symmetry swapping the two extremities of the segment. In particular, we see that $|D_1| = 1$ and $|D_2| = 2$, and thus our general formula for the size of D_n is not valid for $n = 1$ or $n = 2$.

6.5. Group presentations. Recall that we have proven, for the dihedral group D_n , that any element can be written as a product of r and s of the form $s^j r^i$ for $j = 0, 1$ and $i \in \{0, \dots, n - 1\}$. Moreover we have the equalities:

- $s^2 = r^n = \text{id}$
- $rs = sr^{-1}$.

If I give you a word of the form

$$s^{n_1} r^{n_2} \dots s^{n_l} r^{n_l}$$

i.e. a product of power of r and s , you can simplify it, using these rules, into a word of the form $s^j r^i$, for $j = 0, 1$ and $i \in \{0, \dots, n - 1\}$. For example, in D_7 :

$$\begin{aligned} s^6 r^{10} sr^2 &= r^3 sr^2 \\ &= sr^{-3} r^2 \\ &= sr^{-1} \\ &= sr^6 \end{aligned}$$

So we can describe D_n using generator and relations:

$$D_n = \left\langle \underbrace{r, s}_{\text{generators}} \mid \underbrace{s^2 = r^n = \text{id}, rs = sr^{-1}}_{\text{relations}} \right\rangle$$

What this means:

- (1) every element of D_n can be written as a product of r and s ,
- (2) every equality between elements of D_n is a consequence of the equalities in the brackets.

In general, consider any non empty subset S of a group G . The set of all products of elements of S and $S^{-1} = \{s^{-1} : s \in S\}$ is a subgroup of G : it is stable under multiplication, and contains the identity.

Definition 6.46. Let G be a group and S a non-empty subset of G . The *subgroup generated by S* consists of all products of elements of S and $S^{-1} = \{s^{-1} : s \in S\}$. We say that S *generates* G if the subgroup generated by S is G .

Remark 6.47. *One can show that the subgroup generated by S is the smallest subgroup of G containing S in the following sense: any subgroup of G containing S must contain the subgroup generated by S .*

Using this, we can introduce *groups presentations*. If S is a subset of G , we call a *relation* in S an equality involving products of elements of S and their inverses.

Definition 6.48. Let G be a group, and S a subset of G , as well as R a set of relations in S . We write

$$G = \langle S | R \rangle$$

to mean that:

- (1) G is generated by S ,
- (2) every relation in G is a consequence of the set of relations R .

Warning. This definition is enough to get an intuitive idea of what's going on, but it is not rigorous: what do we mean precisely by "is a consequence of"? Maybe we will clarify this later.

Our previous work on the dihedral group amounts to finding a presentation for D_n .

Here is an example of a presentation of a familiar group:

$$\mathbb{Z}^2 = \langle a, b | aba^{-1}b^{-1} \rangle$$

which means that \mathbb{Z}^2 is generated by elements a and b such that a and b commute, i.e. $a = (0, 1)$ and $b = (1, 0)$.

Here is another one:

$$\begin{aligned} S_n = \langle \tau_i, i = 1 \cdots n-1 | \tau_i^2 = \text{id} \\ \tau_i \tau_j = \tau_j \tau_i \text{ if } |i-j| > 1 \\ \tau_i \tau_{i+1} \tau_i = \tau_{i+1} \tau_i \tau_{i+1} \rangle \end{aligned}$$

where $\tau_i = (i \ i+1)$. This one takes some work to obtain! In fact, we can ask:

Question. *Given a group presentation $G = \langle S | R \rangle$, what can we deduce about the group G ?*

Maybe it is reasonable to restrict ourselves to presentations containing finite amount of informations:

Definition 6.49. A group G is *finitely generated* if there is a finite $S \subset G$ such that G is generated by S .

It is *finitely presented* if it is generated by a finite subset S , and has presentation

$$G = \langle S | R \rangle$$

where R is finite.

Example 6.50.

- any finite group is finitely presented.
- $(\mathbb{Z}^n, +)$ is finitely presented for all n .
- $(\mathbb{Q}, +)$ is not finitely generated (and so not finitely presented either).

It is difficult to find finitely generated groups that are not finitely presented. The problem is that we need to show that a given finitely generated group does not have *any* finite presentation.

Example 6.51. The lamplighter group:

$$L = \langle a, t \mid a^2 = (at^n at^{-n})^2 = 1, n \in \mathbb{N} \rangle$$

is finitely generated, but not finitely presented.

Even for finitely presented groups, simple questions are difficult in general. For example, what is the group given by the presentation:

$$\langle a, b \mid a^{-1}ba = b^2, b^{-1}ab = a^2 \rangle ?$$

It is actually the trivial group! More precisely, these relations imply $a = b = 1$.

Exercise 6.52. Prove that this presentation gives the trivial group.

It gets worse: in general, we cannot deduce anything of a group from its presentation, not even if it is trivial or not:

Theorem 6.53 (Undecidability of the group isomorphism problem). *There is no algorithm that, given a group presentation $G = \langle R \mid S \rangle$, decides whether the group G is trivial or not.*

Group presentations are a very active area of research in mathematics, full of open problems. I'll just give one of them.

Question. *Is there a finitely presented infinite group in which every element has finite order?*

Some examples with infinite presentations are known, and are already very difficult to construct.

7. DECOMPOSING GROUPS: COSETS AND QUOTIENTS

In this section, we will introduce some of the most important tools to decompose groups:

- given $H < G$, we obtain a partition of G into the *cosets* of H , which can be used to understand the structure of G ,
- if H is a so-called *normal* subgroup of G , we obtain a group G/H and a morphism $\pi : G \rightarrow G/H$ with kernel H . This can be used to study G while "ignoring" H .

7.1. Cosets and Lagrange's theorem.

Definition 7.1. Let G be a group, H a subgroup and $g \in G$. We define the *left coset* of H with representative g to be the set:

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

Similarly, we define the right coset of H with representative g to be:

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

Remark 7.2. *If $h \in H$, then $hH = H = Hh$.*

Example 7.3. In \mathbb{Z} , the subgroup $3\mathbb{Z}$ has three left cosets:

- $3\mathbb{Z}$,
- $1 + 3\mathbb{Z} = \{1 + 3k : k \in \mathbb{Z}\}$,
- $2 + 3\mathbb{Z} = \{2 + 3k : k \in \mathbb{Z}\}$.

Because \mathbb{Z} is abelian, the right cosets are the same.

In general we have:

Remark 7.4. *In an abelian group, the left and right cosets of a subgroup coincide.*

It may not be the case in a non-abelian subgroup:

Example 7.5. Consider the subgroup $H = \{\text{id}, (1\ 2)\}$ of S_3 . Its left cosets are:

- H ,
- $(1\ 3)H = (1\ 2\ 3)H = \{(1\ 3), (1\ 2\ 3)\}$,
- $(2\ 3)H = (1\ 3\ 2)H = \{(2\ 3), (1\ 3\ 2)\}$.

and its right cosets are:

- H ,
- $H(1\ 3) = H(1\ 3\ 2) = \{(1\ 3), (1\ 3\ 2)\}$,
- $H(2\ 3) = H(1\ 2\ 3) = \{(2\ 3), (1\ 2\ 3)\}$.

In particular, no left coset is equal to a right coset, except H itself.

Note that there is ambiguity in how we write a cosets: two elements $g_1, g_2 \in G$ can have the same coset, i.e. $g_1H = g_2H$. In practice, it is often crucial to figure out when this happens. The following is very useful:

Proposition 7.6. *Let $H < G$ and $g_1, g_2 \in G$. The following are equivalent:*

- (1) $g_1H = g_2H$,
- (2) $Hg_1^{-1} = Hg_2^{-1}$,
- (3) $g_1H \subset g_2H$,
- (4) $g_2 \in g_1H$,
- (5) $g_1^{-1}g_2 \in H$.

Proof. We prove some of it, the rest will be left as an exercise. We use the following chain of implications: (1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (1).

(1) \Rightarrow (2). Assume (1) is true, and let $g \in Hg_1^{-1}$, so $g = hg_1^{-1}$ for some $h \in H$. Then $g^{-1} = g_1h^{-1} \in g_1H$. By assumption, there is $k \in H$ such that $g^{-1} = g_2k$, and therefore $g = k^{-1}g_2^{-1} \in Hg_2^{-1}$.

(3) \Rightarrow (4). Assume that $g_1H \subset g_2H$. In particular, this means that $g_1 \in g_2H$, so there is $h \in H$ such that $g_2 = g_1h$. From this we deduce that $g_1^{-1}g_2 = h$, and thus $g_1^{-1}g_2 \in H$.

(5) \Rightarrow (1). Assume that $g_1^{-1}g_2 \in H$. Then

$$\begin{aligned} g_1H &= g_1(g_1^{-1}g_2H) \\ &= g_1g_1^{-1}g_2H \\ &= g_2H . \end{aligned}$$

The implications (2) \Rightarrow (3) and (4) \Rightarrow (5) are left as exercise. \square

The following is a very useful consequence:

Theorem 7.7. *Let G be a group and $H < G$. The left cosets of H form a partition of G . The associated equivalence relation is $g_1 \sim g_2$ if and only if $g_1^{-1}g_2 \in H$.*

The same is true for right cosets, the associated equivalence relation is $g_1 \sim g_2$ if and only if $g_1 g_2^{-1} \in H$.

Proof. We only do the proof for left cosets, the one for right cosets being similar. We have to prove that for any two cosets $g_1 H$ and $g_2 H$, either $g_1 H \cap g_2 H = \emptyset$, or $g_1 H = g_2 H$. So pick two cosets $g_1 H$ and $g_2 H$, and assume that $g_1 H \cap g_2 H \neq \emptyset$. Let $a \in g_1 H \cap g_2 H$. What this means is that there are $h_1, h_2 \in H$ such that both $a = g_1 h_1$ and $a = g_2 h_2$. So $g_1 h_1 = g_2 h_2$, and in particular $g_1^{-1} g_2 = h_1 h_2^{-1} \in H$. By Proposition 7.6, we conclude that $g_1 H = g_2 H$.

The associated equivalence relation is given by $g_1 \sim g_2$ if and only if g_1 and g_2 are in the same coset of H , i.e. there is $g \in G$ such that $g_1, g_2 \in gH$. Again by Proposition 7.6, this is equivalent to $g_1 H = gH = g_2 H$. \square

Note that even though the proof for right cosets would be the same, the equivalence relation we obtained is not be the same. This is because if we try to prove the equivalent of Proposition 7.6, but for right cosets, we would end up proving that for all $g_1, g_2 \in G$, $Hg_1 = Hg_2$ if and only if $g_1 g_2^{-1} \in H$. *You can prove this as an exercise.*

Notation. The set of left cosets is denoted G/H , and the set of right cosets is denoted $H \backslash G$. These sets are the quotients of G under the equivalence relations defined in the previous theorem.

Note that even though left and right cosets may be different, there is the same quantity of each. More precisely

Proposition 7.8. *Let G be a group, and H a subgroup. There is a bijection between G/H and $H \backslash G$.*

Proof. We define a map $\text{inv} : G/H \rightarrow H \backslash G$ by $\text{inv}(gH) = Hg^{-1}$.

We first have to show that inv is well defined. Assume that $g_1 H = g_2 H$, then by Proposition 7.6, we have $Hg_1^{-1} = Hg_2^{-1}$, so inv is well defined.

Surjectivity is straightforward: let $g \in G$, then $\text{inv}(g^{-1}H) = Hg$.

For injectivity, assume that $\text{inv}(g_1 H) = \text{inv}(g_2 H)$, which means $Hg_1^{-1} = Hg_2^{-1}$. Again by Proposition 7.6, we obtain $g_1 H = g_2 H$, so inv is injective. \square

Definition 7.9. Let G be a group and H a subgroup. We denote $[G : H]$ the cardinality of G/H (or equivalently, of $H \backslash G$). We call it *the index of H in G* .

We will now prove *Lagrange's theorem* that if G is finite, then $|G| = |H| \times [G : H]$. A key fact is that all cosets have the same cardinality:

Lemma 7.10. *Let G be a group and H a subgroup. For all $g \in G$, there is a bijection between H and gH . In particular, all cosets of H have the same cardinality.*

Proof. We define $\mu_g : H \rightarrow gH$ by letting $\mu_g(h) = gh$. This is a bijection because it has inverse $\mu_{g^{-1}} : gH \rightarrow H$ given by $\mu_{g^{-1}}(k) = g^{-1}k$. \square

We deduce:

Theorem 7.11 (Lagrange's theorem). *Let G be a finite group and H a subgroup. Then $|G| = |H| \times [G : H]$. In particular, the order of the subgroup H divides the order of G .*

Proof. We know by Theorem 7.7 that the cosets of H form a partition of G . By Lemma 7.10, all cosets have cardinality $|H|$. Since there are $[G : H]$ cosets, we obtain $|G| = |H|[G : H]$. \square

Remark 7.12. *There is a version of this for infinite groups: for any group G and $H < G$, there is a bijection between G and $H \times G/H$. However, this turns out to depend on some mathematical principle called the axiom of choice, which is not provable from basic set theory.*

As an aside, here is a formulation of the axiom of choice: any surjective function $f : A \rightarrow B$ has a right inverse, i.e. there is a function $g : B \rightarrow A$ such that $f \circ g = \text{id}_B$.

It is impossible to prove this from just basic set theory, but also impossible to disprove it.

This theorem is one of the many sources of interaction between groups and number theory. For example we deduce from it:

Corollary 7.13. *Let G be a finite group. The order of any element of G divides the order of G .*

Proof. Let $g \in G$, and consider $\langle g \rangle$, the subgroup generated by g . Then $|\langle g \rangle| = \text{ord}(g)$. But by Lagrange's theorem, the order of $\langle g \rangle$ divides $|G|$. \square

As a result, we can classify groups of prime order:

Corollary 7.14. *Let p be a prime number. Any group of order p must be isomorphic to $(\mathbb{Z}_p, +)$.*

Proof. By Theorem 6.5, we just need to show that it is cyclic. Let $g \in G \setminus \{1\}$. By Corollary 7.13, the order of g divides p . Since $g \neq 1$, the order is not 1, so it must be p . Therefore the subgroup $\langle g \rangle$ has order p , and thus we must have $G = \langle g \rangle$, i.e. G is cyclic. \square

7.1.1. *An application: Euler's theorem.* Recall that for any $n \in \mathbb{N}$, the group $(U(n), \cdot)$ is the multiplicative group of the invertible elements of \mathbb{Z}_n , i.e. elements $a \in \mathbb{Z}_n$ such that there is $x \in \mathbb{Z}_n$ with $ax = 1$. The following was already proved:

Proposition 7.15. *Let $n \in \mathbb{N}$. Then some $a \in \mathbb{Z}_n$ is invertible if and only if $\text{gcd}(n, a) = 1$.*

Proof. Let $a \in \mathbb{Z}_n$. Assume that a is invertible. Then there is $x \in \mathbb{Z}_n$ such that $ax = 1$. Seeing a, x as elements of \mathbb{Z} , this means that $n | ax - 1$, i.e. there is $y \in \mathbb{Z}$ such that $1 = ax + ny$. This implies that $1 = \text{gcd}(n, a)$.

Conversely, if $\text{gcd}(n, a) = 1$, then there are $x, y \in \mathbb{Z}$ such that $ax + ny = 1$. In \mathbb{Z}_n , this gives $ax = 1$, i.e. a is invertible. \square

This means that the group $U(n)$ has order the number of $1 \leq a \leq n$ such that $\text{gcd}(n, a) = 1$.

Definition 7.16. The number of $1 \leq a \leq n$ such that $\text{gcd}(n, a) = 1$ is a function of n called Euler's totient function and denoted ϕ .

We can now prove:

Theorem 7.17. (*Euler's theorem*)

Let $a \in \mathbb{Z}$ and $n \in \mathbb{N}$, and assume that $\text{gcd}(n, a) = 1$. Then $a^{\phi(n)} \equiv 1 [n]$.

Proof. Since $(U(n), \cdot)$ has order $\phi(n)$, Corollary 7.13 gives $a^{\phi(n)} = 1$ in $U(n)$, which gives the congruence in \mathbb{Z} . \square

In particular, if p is a prime number, we have $\phi(p) = p - 1$, and we obtain the following:

Theorem 7.18 (Fermat's little theorem). *Let p be a prime number and $a \in \mathbb{Z}$, not divisible by p . Then $a^p \equiv a \pmod{p}$.*

It is difficult to see how to prove this without using groups!

7.2. Normal subgroups, quotient groups. We begin by asking the following:

Question. *Let G be a group and H a subgroup. Does the binary operation $(g_1H, g_2H) \rightarrow g_1g_2H$ make G/H into a group?*

Given a function defined on cosets, the first question we should ask is: is it well-defined?

Assume that $g_1H = f_1H$ and $g_2H = f_2H$, is it the case that $g_1g_2H = f_1f_2H$? Using Proposition 7.6, this can be reformulated as follows: assume that $g_1f_1^{-1} \in H$ and $g_2f_2^{-1} \in H$, is it the case that $(g_1g_2)(f_1f_2)^{-1} \in H$? We can attempt the computation:

$$\begin{aligned} (g_1g_2)(f_1f_2)^{-1} &= g_1 \underbrace{g_2f_2^{-1}}_{\in H} f_1^{-1} \\ &= g_1hf_1^{-1} \text{ where } h = g_2f_2^{-1} \in H \end{aligned}$$

If we knew that $hf_1^{-1} = f_1^{-1}k$ for some $k \in H$, we could conclude that $g_1f_1^{-1} \in H$. In other words, what we need is the following equality of cosets:

$$Hf_1^{-1} = f_1^{-1}H.$$

We make it into a definition:

Definition 7.19. Let G be a group. A subgroup H of G is *normal* if for all $g \in G$, we have:

$$gH = Hg.$$

We write $H \triangleleft G$ to say that H is a normal subgroup of G .

Here is a useful characterization of normal subgroups:

Proposition 7.20. *Let G be a group and H a subgroup of G . The following are equivalent:*

- (1) H is normal in G ,
- (2) for all $g \in G$, we have $g^{-1}Hg \subset H$,
- (3) for all $g \in G$, we have $g^{-1}Hg = H$.

Proof. (1) \Rightarrow (2). Assume that H is normal, and pick $g \in G$. Let $g^{-1}hg \in g^{-1}Hg$. Because H is normal and $hg \in Hg$, we get that $hg \in gH$. In other words, there is $k \in H$ such that $hg = gk$. This gives $g^{-1}hg = g^{-1}gk = k \in H$.

(2) \Rightarrow (3). Assume that $g^{-1}Hg \subset H$ for all $g \in G$. Pick some $g \in G$. Then we have, by using the assumption, but for g^{-1} , that $gHg^{-1} \subset H$. Let $h \in H$, then $ghg^{-1} \in gHg^{-1}$, which is a subset of H . Therefore there is $k \in H$ such that $ghg^{-1} = k$, which gives $h = g^{-1}hg$, so $h \in g^{-1}Hg$. This gives $H = g^{-1}Hg$.

(3) \Rightarrow (1). Exercise. \square

As it turns out, normality is the only property needed to make the quotient G/H into a group:

Theorem 7.21. *Let G be a group and H a normal subgroup. Then G/H equipped with the binary operation $(g_1H, g_2H) \rightarrow g_1g_2H$ is a group. We call it the quotient group of G by H .*

Moreover, the map $\pi : G \rightarrow G/H$ given by $\pi(g) = gH$ is a surjective group morphism with kernel H .

Proof. The proof was essentially given in the introductory discussion, but we give it again. Let G be a group and H a normal subgroup. We first show that this operation is well-defined. Let $g_1, f_1, g_2, f_2 \in G$ and assume that $g_1H = f_1H$ and $g_2H = f_2H$. Then:

$$\begin{aligned} g_1g_2H &= g_1Hg_2 \text{ as } H \text{ is normal} \\ &= f_1Hg_2 \\ &= f_1g_2H \text{ again as } H \text{ is normal} \\ &= f_1f_2H . \end{aligned}$$

Therefore this binary operation is well-defined.

For any $g \in G$, we have $(gH)(H) = g \cdot 1H = H$, so H is the identity. Straightforward computations show that $(gH)^{-1} = g^{-1}H$ and associativity.

Checking that the map π is a surjective morphism is left as an exercise. \square

Remark 7.22. *If G is abelian, every subgroup is normal. So every subgroup gives us a quotient group!*

Example 7.23. Consider the subgroup $n\mathbb{Z}$ of \mathbb{Z} , for some $n \in \mathbb{N}$. Because \mathbb{Z} is abelian, the subgroup $n\mathbb{Z}$ is normal. The cosets are $\{k + n\mathbb{Z} : 0 \leq k \leq n - 1\}$, and the group operation is given by

$$(k + n\mathbb{Z})(l + n\mathbb{Z}) = (k + l) + n\mathbb{Z}$$

In other word, we recover the group \mathbb{Z}_n . This is why this group is denoted $\mathbb{Z}/n\mathbb{Z}$: it is the quotient of \mathbb{Z} by its normal subgroup $n\mathbb{Z}$.

Example 7.24. Consider the subgroup $\text{Sl}_n(\mathbb{R}) < \text{Gl}_n(\mathbb{R})$, the subgroup of matrices of determinant 1. Because \det is a morphism, we have, for any $S \in \text{Sl}_n$ and $M \in \text{Gl}_n$, that $\det(M^{-1}SM) = \det(M)^{-1}\det(S)\det(M) = 1$. So Sl_n is a normal subgroup.

Example 7.25. Consider the symmetric group S_n and its subgroup A_n . It is easy to see that if $\tau \in A_n$ and $\sigma \in S_n$, then $\sigma^{-1}\tau\sigma$ is an even permutation (just do cases according to the sign of σ). Therefore $A_n \triangleleft S_n$.

7.3. The first and third isomorphism theorem. Note that all three examples of the previous subsection have the following property: the normal subgroup is given by the kernel of a group morphism. This is a general fact.

Proposition 7.26. *Let $\phi : G \rightarrow H$ be a group morphism. Then $\ker(\phi) \triangleleft G$.*

Proof. Let $g \in G$ and $f \in \ker(\phi)$. Then

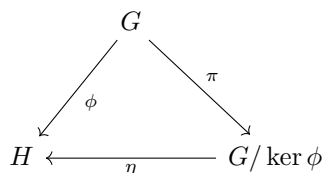
$$\begin{aligned} \phi(g^{-1}fg) &= \phi(g^{-1})\phi(f)\phi(g) \\ &= \phi(g)^{-1}1\phi(g) \\ &= 1 \end{aligned}$$

so $g^{-1}fg \in \ker(\phi)$. Therefore $\ker(\phi) \triangleleft G$. □

In fact, given a surjective group morphism $\phi : G \rightarrow H$, the quotient map $\pi : G \rightarrow G/\ker(\phi)$ is essentially the same map as ϕ . The following makes this precise:

Theorem 7.27 (First isomorphism theorem). *Let $\phi : G \rightarrow H$ be a surjective group morphism and consider the quotient morphism $\pi : G \rightarrow G/\ker(\phi)$. Then there is a unique isomorphism $\eta : G/\ker(\phi) \rightarrow H$ such that $\eta \circ \pi = \phi$.*

Before starting with the proof, we give a graphical representation of the result. We can put all these maps into the following representation, called a *diagram*:



The information that $\eta \circ \pi = \phi$ means that no matter what path of functions we follow in that diagram, the end result is the same. We say that the diagram *commutes*.

Proof of Theorem 7.27. We define $\eta : G/\ker(\phi) \rightarrow H$ as $\eta(g\ker(\phi)) = \phi(g)$. We first have to show that this is well-defined. So assume that $g_1\ker(\phi) = g_2\ker(\phi)$ for some $g_1, g_2 \in G$. In particular, by Proposition 7.6, we have $g_1^{-1}g_2 \in \ker(\phi)$. We compute

$$\begin{aligned} \phi(g_1) &= \phi(g_1)\phi(g_1^{-1}g_2) \text{ as } g_1^{-1}g_2 \in \ker(\phi) \\ &= \phi(g_1g_1^{-1}g_2) \\ &= \phi(g_2) \end{aligned}$$

so η is well-defined.

We show it is a morphism. First, we have $\eta(\ker(\phi)) = \eta(1\ker(\phi)) = \phi(1) = 1$, so the image of the identity is the identity. Now let $g_1, g_2 \in G$, we compute:

$$\begin{aligned} \eta(g_1\ker(\phi) \cdot g_2\ker(\phi)) &= \eta(g_1g_2\ker(\phi)) \\ &= \phi(g_1g_2) \\ &= \phi(g_1)\phi(g_2) \\ &= \eta(g_1\ker(\phi))\eta(g_2\ker(\phi)) \end{aligned}$$

so η is a morphism.

Finally, let's check injectivity and surjectivity.

Let $g \in G$, and assume that $\eta(g\ker(\phi)) = e_H$. So $\phi(g) = e_H$, which means that $g \in \ker(\phi)$. Therefore $g\ker(\phi) = \ker(\phi)$. So η is injective.

For surjectivity, let $g \in G$, then $\eta(g\ker(\phi)) = \phi(g)$. This shows η is surjective.

We also have to show that $\eta \circ \pi = \phi$ and that η is the unique isomorphism satisfying this equality. Let $g \in G$, then $\eta \circ \pi(g) = \eta(g\ker(\phi)) = \phi(g)$. So $\eta \circ \pi = \phi$.

Now assume that there is another isomorphism η' satisfying $\eta' \circ \pi = \phi$. Let $g \in G$, then $\phi(g) = \eta'(\pi(g)) = \eta'(g \ker(\phi))$. So $\eta = \eta'$. \square

To see why this theorem is useful, let's go back to our examples of determinant and signature.

Example 7.28. We have that $\text{Sl}_n(\mathbb{R}) = \ker(\det)$, and \det is a surjective morphism from $\text{Gl}_n(\mathbb{R})$ to \mathbb{R}^* . Therefore \mathbb{R}^* is isomorphic to $\text{Gl}_n(\mathbb{R})/\text{Sl}_n(\mathbb{R})$.

Example 7.29. $\text{sign} : S_n \rightarrow \{-1, 1\}$ is a surjective group morphism, with kernel A_n . Therefore $\{-1, 1\}$ is isomorphic to S_n/A_n . In particular by Lagrange's theorem, we see that $|A_n| = \frac{n!}{2}$.

Here's one last familiar example:

Example 7.30. We go back to the multiplicative group of the complex numbers and the circle group \mathbb{S}^1 . Recall that we have a surjective morphism:

$$\begin{aligned} \phi : (\mathbb{R}, +) &\rightarrow (\mathbb{S}^1, \cdot) \\ \theta &\rightarrow e^{i\theta} \end{aligned}$$

If we identify its kernel, then we would have identified (\mathbb{S}^1, \cdot) , up to isomorphism. We have, for any $\theta \in \mathbb{R}$

$$\begin{aligned} \theta \in \ker(\phi) &\Leftrightarrow e^{i\theta} = 1 \\ &\Leftrightarrow \cos(\theta) + i \sin(\theta) = 1 \\ &\Leftrightarrow \cos(\theta) = 1 \text{ and } \sin(\theta) = 0 \\ &\Leftrightarrow \theta \in \{2k\pi : k \in \mathbb{Z}\} = 2\pi\mathbb{Z} \end{aligned}$$

Therefore \mathbb{S}^1 is isomorphic to $\mathbb{R}/2\pi\mathbb{Z}$.

We can also apply a similar technique to (\mathbb{C}, \cdot) . Here, we have a surjective morphism:

$$\begin{aligned} ((0, +\infty), \cdot) \times (\mathbb{R}, +) &\rightarrow (\mathbb{C}^*, \cdot) \\ (r, \theta) &\rightarrow re^{i\theta} \end{aligned}$$

Using a similar reasoning, we obtain that (\mathbb{C}, \cdot) is isomorphic to $((0, +\infty), \cdot) \times (\mathbb{R}, +)/\{1\} \times 2\pi\mathbb{Z}$.

One could wonder how to relate subgroups of G and subgroups of its quotients:

Question. *Given a group G and a normal subgroup N , is there any connection between subgroups of G and subgroups of G/N ?*

Before giving the answer, we prove a useful lemma:

Lemma 7.31. *Let $\pi : G \rightarrow F$ be a group morphism. Then for any subgroup $\ker(\pi) < H < G$, we have $\pi^{-1}(\pi(H)) = H$.*

Proof. The inclusion $H \subset \pi^{-1}(\pi(H))$ is true in general for functions and subsets. We need to prove the other one. So let $g \in \pi^{-1}(\pi(H))$. This means that $\pi(g) \in \pi(H)$, so there is $h \in H$ such that $\pi(g) = \pi(h)$. This implies $\pi(gh^{-1}) = 1$, i.e. $gh^{-1} \in \ker(\pi)$. Since $\ker(\pi) < H$, we also have $gh^{-1} \in H$. So both $h \in H$ and $gh^{-1} \in H$, from which we deduce that $g \in H$. \square

We will also need a lemma showing that morphisms preserve normality:

Lemma 7.32. *Let $\phi : G \rightarrow F$ be a surjective group morphism, and H a subgroup of G with $\ker(\phi) < H$. Then H is normal in G if and only if $\phi(H)$ is normal in F .*

Proof. Assume first that $H \triangleleft G$. Let $\phi(h) \in \phi(H)$, and pick some $f \in F$. Since ϕ is surjective, there is $g \in G$ such that $\phi(g) = f$. Then:

$$\begin{aligned} f^{-1}\phi(h)f &= \phi(g^{-1})\phi(h)\phi(g) \\ &= \phi(g^{-1}hg) \\ &\in \phi(H) \text{ as } g^{-1}hg \in H \end{aligned}$$

which implies $\phi(H) \triangleleft F$.

Now assume that $\phi(H)$ is normal in F . Let $h \in H$, and $g \in G$. Then $\phi(g)^{-1}\phi(h)\phi(g) \in \phi(H)$ by assumption. Therefore $\phi(g^{-1}hg) \in \phi(H)$. In other words, we have $g^{-1}hg \in \phi^{-1}(\phi(H))$. By Lemma 7.31, this implies that $g^{-1}hg \in H$. Thus $H \triangleleft G$. \square

Remark 7.33. *The left to right direction is true without the assumption on $\ker(\phi)$.*

The answer to our question is then given by:

Theorem 7.34 (Correspondence theorem). *Let G be a group, N a normal subgroup and $\pi : G \rightarrow G/N$ the quotient map. Consider the sets:*

$$\begin{aligned} \text{Sub}(G)_N &= \{H : N < H < G\} \\ \text{Sub}(G/N) &= \{H : H < G/N\} . \end{aligned}$$

Then the map $\tilde{\pi}$ defined by $\tilde{\pi}(H) = \pi(H) = H/N$ is a bijection from $\text{Sub}(G)_N$ to $\text{Sub}(G/N)$. Moreover, it restricts to a bijection between the subsets of normal subgroups in $\text{Sub}(G)_N$ and $\text{Sub}(G/N)$.

Proof. We first have to prove that it is a well-defined map, i.e. that if $H \in \text{Sub}(G)_N$, then $\tilde{\pi}(H) \in \text{Sub}(G/N)$. This is simply because the image of a subgroup under a morphism is a subgroup.

We will show that the map $\widetilde{\pi^{-1}} : \text{Sub}(G/N) \rightarrow \text{Sub}(G)_N$ given by $\widetilde{\pi^{-1}}(K) = \pi^{-1}(K)$, is its inverse. Again, it is well-defined because the inverse image of a subgroup is a subgroup.

To show these two maps are inverse of each other, we need to show that $\pi^{-1}(\pi(H)) = H$ for all $N < H < G$, and $\pi(\pi^{-1}(K)) = K$ for all $K < G/N$. The first equality is given by Lemma 7.31. The second is because π is surjective.

Finally, we have to show the moreover part, i.e. that if H is a subgroup of G , then $H \triangleleft G$ if and only if $H/N \triangleleft G/N$. This is a direct consequence of Lemma 7.32, applied to the morphism $\pi : G \rightarrow G/N$. \square

Here is an application:

Example 7.35. The subgroups of the circle group \mathbb{S}^1 are in bijection with subgroups of $(\mathbb{R}, +)$ containing $2\pi\mathbb{Z}$. We recover the subgroups of roots of unity: it corresponds to $2\pi\mathbb{Q}$.

The subgroup $\{1, i, -1, -i\}$ corresponds to the subgroup $\frac{\pi}{2}\mathbb{Z}$.

So we know that subgroups of G/N are in bijection with subgroups of G containing N . If we have another normal subgroup K of G with $N \triangleleft K \triangleleft G$, then we can form the quotient group $G/N / K/N$, as K/N is a normal subgroup of G/N . Can we write this in a simpler way?

Theorem 7.36 (Third isomorphism theorem). *Let G be a group and consider normal subgroups N and K of G , with $N \triangleleft K$. Then the groups $G/N / K/N$ and G/K are isomorphic.*

Proof. We will use the first isomorphism theorem here. We have two quotient morphisms:

$$\begin{aligned}\pi : G &\rightarrow G/N \\ \rho : G/N &\rightarrow G/N / K/N\end{aligned}$$

and by composing them, we get a morphism $\rho \circ \pi : G \rightarrow G/N / K/N$. By the first isomorphism theorem, it is enough to prove that $\ker(\rho \circ \pi) = K$.

Note that we have:

$$\begin{aligned}\ker(\rho \circ \pi) &= \pi^{-1}(\ker(\rho)) \\ &= \pi^{-1}(K/N) \\ &= \pi^{-1}(\pi(K)) \\ &= K \text{ by Lemma 7.31}\end{aligned}$$

which gives the result. □

The content of this theorem can also be summarized in a commutative diagram:

$$\begin{array}{ccc} G & \longrightarrow & G/K \\ \downarrow \pi & & \updownarrow \\ G/N & \xrightarrow{\rho} & G/N / K/N \end{array}$$

where the rightmost arrow is the isomorphism given by the first isomorphism theorem, and the top arrow is the quotient map.

Example 7.37. Let $n, m \in \mathbb{N}$, we have the following group inclusions $nm\mathbb{Z} \triangleleft m\mathbb{Z} \triangleleft \mathbb{Z}$, and so we have

$$\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z} \simeq \mathbb{Z}/nm\mathbb{Z} / m\mathbb{Z}/nm\mathbb{Z}$$

We can unpack this further. What is $m\mathbb{Z}/nm\mathbb{Z}$? On the face of it, it is the quotient group of $m\mathbb{Z}$ by its subgroup $nm\mathbb{Z}$. But that's confusing (at least to me). Notice that there is a group isomorphism

$$\begin{aligned}\phi : \mathbb{Z} &\rightarrow m\mathbb{Z} \\ x &\rightarrow mx\end{aligned}$$

(check that it is one as an exercise). Moreover, it takes $n\mathbb{Z} < \mathbb{Z}$ to $nm\mathbb{Z} < m\mathbb{Z}$. Therefore $m\mathbb{Z}/nm\mathbb{Z} \simeq \mathbb{Z}/n\mathbb{Z}$.

Another way to see this: $m\mathbb{Z}/nm\mathbb{Z}$ is the subgroup of $\mathbb{Z}/nm\mathbb{Z}$ given by $\{0, m, 2m, \dots, (n-1)m\}$. From this description, one also sees that it is isomorphic to $\mathbb{Z}/n\mathbb{Z}$.

One can sometime sees this written as

$$\mathbb{Z}/m\mathbb{Z} \simeq \mathbb{Z}/nm\mathbb{Z}/\mathbb{Z}/n\mathbb{Z}$$

which maybe makes the “factorization” clearer.

7.4. The second isomorphism theorem. There is one last isomorphism theorem that is also very useful. It is different enough from the previous two to get its own section. We start by a motivating lemma:

Lemma 7.38. *Let G be a group, N a normal subgroup of G and H a subgroup of G . Then*

- (1) $HN = \{hg : h \in H, g \in N\}$ is a subgroup of G ,
- (2) $H \cap N$ is a normal subgroup of H .

Proof. We start by proving (1). Since $1 \in H$ and $a \in N$, we see that $1 \in HN$. Now let $h_1, h_2 \in H$ and $g_1g_2 \in N$.

First we check that HN is stable under product. We have $h_1g_1h_2g_2 = h_1h_2gg_2$ for some $g_3 \in N$, because $N \triangleleft G$. Since $h_1h_2 \in H$ and $gg_2 \in N$, we get that $h_1g_1h_2g_2 \in HN$.

For inverses, we have that $(h_1g_1)^{-1} = g_1^{-1}h_1^{-1} = h_1^{-1}g$ for some $g \in N$, because $N \triangleleft G$. Therefore $(h_1g_1)^{-1} \in HN$. So $HN < G$.

We now prove (2). We have already seen that intersection of subgroups are subgroups, and since $H \cap N \subset H$, it must be a subgroup of H . Let $g \in H \cap N$ and $h \in H$, then $h^{-1}gh \in N$ because $g \in N$ and N is normal in G . But we also have $h^{-1}gh \in H$ because $h \in H$ and $g \in H$ and H is a subgroup. Therefore $h^{-1}gh \in H \cap N$. \square

We now state and prove the:

Theorem 7.39 (Second isomorphism theorem). *Let G be a group, N a normal subgroup of G and H a subgroup of G . Then*

$$HN/N \simeq H/H \cap N .$$

Proof. We use the first isomorphism theorem. Define a map $\phi : H \rightarrow HN/N$ by $\phi(h) = hN$, it is enough to show that $\ker(\phi) = H \cap N$. We have that, for any $h \in H$:

$$\begin{aligned} h \in \ker(\phi) &\Leftrightarrow hN = N \\ &\Leftrightarrow h \in N \\ &\Leftrightarrow h \in H \cap N \end{aligned}$$

so $\ker(\phi) = H \cap N$. \square

7.5. Simple groups. Recall that an integer is prime if its only positive divisors are 1 and itself. Through the fundamental theorem of arithmetic, we know that any integer can be written as a product of primes.

Question. *Is there any similar decomposition for groups?*

We will see that the answer is yes, at least for finite groups, although it is quite a bit more complex than for integers.

If G is a group and H is a normal subgroup, we have a surjective morphism $\pi : G \rightarrow G/H$, and we can hope to capture the structure of G through these of H and G/H . Therefore, our "building blocks" should be those group for which such a decomposition is impossible:

Definition 7.40. A group G is simple if its only normal subgroups are $\{1\}$ and G .

Remark 7.41. This is equivalent to saying that the only surjective group morphisms $\phi : G \rightarrow K$ are for $K = G$ and $K = \{1\}$.

Example 7.42. If p is prime, then $(\mathbb{Z}_p, +)$ is simple. Indeed, the order of any element of $(\mathbb{Z}_p, +)$ must be 1 or p . So any non-identity element has order p , which means that \mathbb{Z}_p has no proper subgroups other than $\{0\}$.

Other examples of simple groups are not so easy to find. We have encountered at least one before:

Theorem 7.43. The alternating group A_n is simple for $n \geq 5$.

One key lemma, of independent interest, is the following:

Lemma 7.44. Let $\sigma = (a_1 a_2 \cdots a_k)$ be a cycle in S_n , and let $\tau \in S_n$. Then

$$\tau\sigma\tau^{-1} = (\tau(a_1) \tau(a_2) \cdots \tau(a_k)) .$$

Proof. Fix some σ, τ as in the theorem, and pick $x \in \{1, \dots, n\}$. Then

$$\begin{aligned} (\tau \circ \sigma \circ \tau^{-1})(\tau(x)) &= (\tau \circ \sigma)(\tau(\tau^{-1}(x))) \\ &= (\tau \circ \sigma)(x) \\ &= \tau(\sigma(x)) . \end{aligned}$$

In other words, for all $x \in \{1, \dots, n\}$, we have that $\tau\sigma\tau^{-1}(\tau(x)) = \tau(\sigma(x))$.

If $x = a_i$ for some i , this gives us $\tau\sigma\tau^{-1}(\tau(a_i)) = \tau(\sigma(a_i)) = \tau(a_{i+1})$ (or $\tau(a_1)$ if $i = k$). Else, we have $\sigma(x) = x$, and therefore $\tau\sigma\tau^{-1}(\tau(x)) = \tau(\sigma(x)) = \tau(x)$. This yields the lemma. \square

To prove that any non-trivial normal subgroup of A_n is equal to A_n , we will use the following:

Lemma 7.45. The group A_n is generated by 3-cycles.

Proof. We know it is generated by products of two transpositions. Any product of two transposition is one of the following

$$\begin{aligned} (a b)(a b) &= \text{id} \\ (a b)(c d) &= (a c b)(a c d) \\ (a b)(a c) &= (a c b) \end{aligned}$$

for some distinct a, b, c, d . \square

And one last lemma:

Lemma 7.46. If $n \geq 5$, then any two 3-cycles are conjugate in A_n .

Proof. Consider two 3-cycles (abc) and (def) , and let $\tau \in S_n$ such that $\tau(a) = d$, $\tau(b) = e$ and $\tau(c) = f$. Then $\tau(abc)\tau^{-1} = (def)$ by Lemma 7.44. If τ is even, we are done. If not, because $n \geq 5$, there are $g, h \in \{1, \dots, n\}$ such that $g, h \notin \{d, e, f\}$. Then we compute that:

$$\begin{aligned} (gh)\tau(abc)\tau^{-1}(gh) &= (gh)(def)(gh) \\ &= (def) \end{aligned}$$

and $(gh)\tau$ is even. □

Putting these previous two lemmas together, here is the strategy: pick some $H \triangleleft A_n$ non trivial. If we show that H contains a 3-cycle, then it contains all 3-cycles by Lemma 7.46. Therefore it must be equal to A_n by Lemma 7.45.

We can now prove the theorem:

Theorem 7.47. *The alternating group A_n is simple for $n \geq 5$.*

Proof. Let $H \triangleleft A_n$ be non trivial. As discussed before, it is enough to show that H contains a 3-cycle.

Let $\sigma \in H \setminus \{\text{id}\}$ with $|\text{supp}(\sigma)|$ of minimal size. In other words, the number of i with $\sigma(i) = i$ is maximal.

We can decompose σ into a product of cycles with disjoint support. Suppose first that σ is a product of 2-cycles with disjoint support. There is at least two 2-cycles (ab) and (cd) , because σ is even. As $n \geq 5$, we can find $e \in \{1, \dots, n\} \setminus \{a, b, c, d\}$. Let $\tau = (cde) \in A_n$. Consider $\tilde{\sigma} = \tau\sigma\tau^{-1}\sigma^{-1}$, this is an element of H because H is normal in A_n . Observe that $\tilde{\sigma}$ fixes a and b . Moreover, any element in $\{1, \dots, n\} \setminus \{a, b, c, d, e\}$ which is fixed by σ is also fixed by $\tilde{\sigma}$. Therefore $|\text{supp}(\tilde{\sigma})| \leq |\text{supp}(\sigma)| - 1$, which is a contradiction.

So we can assume that σ has at least one cycle of length ≥ 3 in its decomposition, denote it $(a_1 \dots a_l)$. If $\sigma \neq (a_1 a_2 a_3)$, then $\text{supp}(\sigma) \geq 4$, as otherwise $\sigma = (a_1 a_2 a_3 a_4)$, which is odd. Let $b, c \in \text{supp}(\sigma) \setminus \{a_1, a_2, a_3\}$. Let $\tau = (a_3 bc)$, and again consider $\tilde{\sigma} = \tau\sigma\tau^{-1}\sigma^{-1}$. Again, any element fixed by σ is fixed by $\tilde{\sigma}$, but we also have that $\tilde{\sigma}(a_1) = a_1$. Therefore $|\text{supp}(\tilde{\sigma})| \leq |\text{supp}(\sigma)| - 1$, a contradiction.

We conclude that σ must be a 3-cycle, and as explained previously, this implies that $H = A_n$. □

There is another group we know that is not simple, but quite close: $\text{Sl}_n(\mathbb{R})$, for $n \geq 2$.

Definition 7.48. Let G be a group. The *center* of G is

$$Z(G) = \{g \in G : h^{-1}gh = g \text{ for all } h \in G\} .$$

Proposition 7.49. *The center of G is a normal subgroup of G .*

The proof is left as an exercise.

We can determine the centers of $\text{Gl}_n(\mathbb{R})$ and $\text{Sl}_r(\mathbb{R})$:

Proposition 7.50. *The center of $\text{Gl}_n(\mathbb{R})$ is the group $\{\lambda \text{id} : \lambda \in \mathbb{R}^*\}$, and the center of $\text{Sl}_n(\mathbb{R})$ is $Z(\text{Gl}_n(\mathbb{R})) \cap \text{Sl}_n(\mathbb{R})$.*

In other words, the center of $\text{Gl}_n(\mathbb{R})$ is made of matrices of the form $\begin{pmatrix} \lambda & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda \end{pmatrix}$ and the center of Sl_n of the same matrices, but with determinant 1, i.e. $\lambda^n = 1$.

We can define two groups. The projective linear group is $\mathrm{PGL}_n(\mathbb{R}) = \mathrm{GL}_n(\mathbb{R})/Z(\mathrm{GL}_n(\mathbb{R}))$, and the projective special linear group $\mathrm{PSL}_n(\mathbb{R}) = \mathrm{SL}_n(\mathbb{R})/Z(\mathrm{SL}_n(\mathbb{R}))$. Note that they are sometimes isomorphic, but not always. We have

Theorem 7.51. *The group $\mathrm{PSL}_n(\mathbb{R})$ is simple if $n \geq 2$.*

We skip the proof of this. Before we move on, note the following interesting phenomenon:

- if n is even, then $Z(\mathrm{SL}_n(\mathbb{R}))$ contains two matrices id and $-\mathrm{id}$,
- if n is odd, then the only n -th root of 1 is 1 itself, so in fact $Z(\mathrm{SL}_n(\mathbb{R})) = \{\mathrm{id}\}$, and $\mathrm{PSL}_n(\mathbb{R}) = \mathrm{SL}_n(\mathbb{R})$.

To use simple groups to decompose any group, we can make use of the following:

Definition 7.52. Let G be a group. A *composition series* for G is a finite sequence of groups:

$$\{1\} = G_0 \triangleleft G_1 \triangleleft \cdots \triangleleft G_{l-1} \triangleleft G_l = G$$

such that G_i/G_{i-1} is simple for all $1 \leq i \leq l$.

We have:

Theorem 7.53. *If G is a finite group, then it has a composition series.*

Recall our big question:

Question. *Can we classify all finite groups, up to isomorphism?*

Using composition series, we can decompose this into two goals:

- (1) classify finite simple groups,
- (2) understand how they can fit together into a composition series.

The good news is that we have a complete classification of finite simple groups! The bad news is that the full proof, due to many people, is more than 10.000 pages long. A new, simpler proof is currently being written, with the hope of reducing it to less than 5000 pages. Here's a very rough explanation of the classification. A finite simple group is either:

- A. $(\mathbb{Z}_p, +)$ for some prime p ,
 - B. A_n for some $n \geq 5$,
 - C. a *group of Lie type*: these are constructed from groups of matrices,
 - D. one of 26 *sporadic groups*, which do not seem to fit into a global pattern.
- The largest is called the *monster group*, of size approximately 10^{53} . The second largest is the baby monster, how cute!

As far as I know, part (2) of the goal has not been achieved. Here is an example showing why knowing all the G_i/G_{i+1} is not enough. Consider the two groups $(\mathbb{Z}_4, +)$ and $(\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)$.

The group \mathbb{Z}_4 has $\{0, 2\}$ as a subgroup, isomorphic to \mathbb{Z}_2 . Moreover, the quotient $\mathbb{Z}_4/\mathbb{Z}_2$ is also isomorphic to \mathbb{Z}_2 (we are using that any group of order 2 is isomorphic to \mathbb{Z}_2). Let $\rho : \mathbb{Z}_4 \rightarrow \mathbb{Z}_2$ be the quotient map.

For the group $\mathbb{Z}_2 \times \mathbb{Z}_2$, we define a map $\pi : \mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow \mathbb{Z}_2$ by letting $\pi(a, b) = a$ for any a, b . Then its kernel is $\{(0, b) : b \in \mathbb{Z}_2\}$, which is isomorphic to \mathbb{Z}_2 .

To summarize: the kernel and images of π and ρ are all isomorphic to \mathbb{Z}_2 . However, the groups \mathbb{Z}_4 and $\mathbb{Z}_2 \times \mathbb{Z}_2$ are not isomorphic.

8. RINGS

So far, we have only looked at sets with one binary operations, and mostly groups. But there are very interesting structures with *two* binary operations, the integers with addition and multiplication being an important example. In this section, we will study *rings*, one of the most important class of structures with two binary operations. It is essentially an abstraction of the properties of integers.

8.1. Definition, basic properties. We start right away with:

Definition 8.1. A *ring* is a set R equipped with two binary operations $+$ and \cdot such that:

- (1) $(R, +)$ is an abelian group with identity 0 .
- (2) \cdot is associative, i.e. $(ab)c = a(bc)$ for all $a, b, c \in R$,
- (3) \cdot distributes over $+$ on both sides, i.e. for all $a, b, c \in R$, we have:

$$a(b + c) = ab + ac$$

$$(a + b)c = ac + bc$$

We denote a ring and its operations by $(R, +, \cdot)$.

If there is some $1 \in R$ such that

- (4) $1 \cdot a = a \cdot 1$ for all $a \in R$,

R is called a *ring with identity*.

If \cdot is commutative, we call R a *commutative ring*.

Some important class of rings are those satisfying a very useful dichotomy:

Definition 8.2. An *integral domain* is a commutative ring R with identity such that for any $a, b \in R$, if $ab = 0$, either $a = 0$ or $b = 0$.

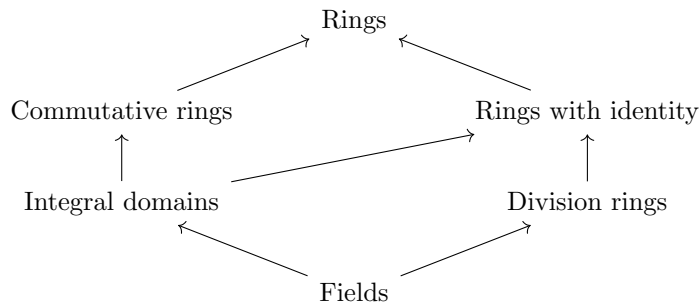
Finally, the existence of inverses for multiplication is important:

Definition 8.3. A *division ring* is a ring R with identity such that every $a \in R \setminus \{0\}$ has a multiplicative inverse, meaning there is a^{-1} such that $aa^{-1} = a^{-1}a$. Note that this inverse must be unique.

A commutative division ring is called a *field*.

One may ask why we insist that $a \neq 0$ in the previous definition. This is because one can prove (and we will), that for any element a in a ring, $a \times 0 = 0 \times a = 0$.

Here is a diagram summarizing the properties and implications between them:



All of these implications follow by the definitions. Less immediate is the fact that they are all strict. For this, we will need to look at examples.

Example 8.4. The integers $(\mathbb{Z}, +, \times)$ form a commutative ring, with identity 1. In fact, it is an integral domain. It is not a division ring however, as for example, there is not integer a such that $a \times 2 = 1$.

Example 8.5. \mathbb{Q} , \mathbb{R} and \mathbb{C} are fields with their usual addition and multiplication.

Example 8.6. The 2×2 matrices with real coefficients $M_2(\mathbb{R})$ forms a ring with matrix addition and multiplication. It is commutative with identity $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

We have $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 0$ therefore it is not an integral domain. It is not a division ring either.

Example 8.7. For any $n \in \mathbb{N}$, $(\mathbb{Z}_n, +, \cdot)$ is a commutative ring with identity. It is in general not an integral domain nor a division ring. For example, we have $2 \cdot 3 = 0$ in \mathbb{Z}_6 .

Recall that we defined $U(n)$ to be the group of invertible elements of \mathbb{Z}_n for multiplication. Therefore, we know that $U(n)$ is a field if and only if $U(n) = \{1, \dots, n-1\}$. Recall that we proved:

Proposition 8.8. *Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}_n$. Then a is invertible for \cdot if and only if $\gcd(n, a) = 1$.*

From this we deduce:

Theorem 8.9. *$(\mathbb{Z}_n, +, \cdot)$ is a field if and only if n is a prime number.*

In particular, we have constructed finite fields. These are essential in many applications of mathematics, in particular to cryptography and error-correcting codes. And have plenty of applications in pure math as well! Note that not all finite fields are of the form $(\mathbb{Z}_p, +, \times)$: we will see other examples.

For now, I will set up some common notation:

Notation. The finite field $(\mathbb{Z}_p, +, \times)$, for some prime p , is denoted \mathbb{F}_p .

Example 8.10. The ring $(2\mathbb{Z}, +, \times)$ is a commutative ring without an identity.

Just as we had subgroups, we have:

Definition 8.11. Let R be a ring. A non-empty subset $S \subset R$ is a *subring* of R if

- (1) S is a subgroup of $(R, +)$,
- (2) for all $a, b \in S$, we have $ab \in S$.

Let us look at an interesting example:

Example 8.12. Consider the ring $M_2(\mathbb{C})$, and the four matrices

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \mathbf{j} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \quad \mathbf{k} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

they satisfy the identities

$$\begin{aligned} i^2 = j^2 = k^2 &= -1 \\ ij &= k \\ jk &= i \\ ki &= j \\ ji &= -k \\ kj &= -i \\ ik &= -j \end{aligned}$$

From all these identities, we deduce that $\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\}$ forms a subring of $M_2(\mathbb{C})$. It is called the ring of *quaternions*. It has a unit, and is not commutative. However, it is a division ring.

Indeed, consider some $a + bi + cj + dk \in \mathbb{H}$, then we can compute

$$(a + bi + cj + dk)(a - bi - cj - dk) = a^2 + b^2 + c^2 + d^2$$

and thus the inverse of $a + bi + cj + dk$ is given by

$$\frac{a - bi - cj - dk}{a^2 + b^2 + c^2 + d^2}.$$

Non-zero elements a, b such that $ab = 0$ have a special name:

Definition 8.13. Let R be a ring. A *zero-divisor* of R is some $a \in R \setminus \{0\}$ such that there is some non-zero $b \in R$ with $ab = 0$.

Therefore, a commutative ring with identity is an integral domain if and only if it has no zero divisor.

One way we can make new rings out of old ones is, just like for groups, cartesian products:

Definition 8.14. Let R, S be rings. Then the cartesian product $R \times S$ is the ring obtained by defining addition and multiplication coordinatewise.

Example 8.15. We can consider the cartesian product $\mathbb{Z}_2 \times \mathbb{Z}_3$, which is defined as:

$$\{(a, b) : a \in \mathbb{Z}_2, b \in \mathbb{Z}_3\}$$

with $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b)(c, d) = (ac, bd)$.

Note that this shows that the cartesian product of two integral domains need not be an integral domain.

Finally, some useful identities in rings:

Proposition 8.16. Let R be a ring, then for all $a, b \in R$:

- (1) $a0 = 0a = 0$,
- (2) $a(-b) = (-a)b = -ab$,
- (3) $(-a)(-b) = ab$.

Proof. Let $a, b \in R$. We have

$$\begin{aligned} a0 &= a(0 + 0) \\ &= a0 + a0 \end{aligned}$$

which implies $a0 = 0$. Similarly, we can prove $0a = 0$.

For (2), notice that

$$\begin{aligned} a(-b) + ab &= a(-b + b) \\ &= a0 \\ &= 0 \end{aligned}$$

which implies that $a(-b) = -ab$. The rest of the identities are proven in a similar way. \square

8.2. Integral domains and fields. Integral domains and fields are important throughout mathematics, for examples in the areas of number theory and geometry.

Let us look at more examples.

Example 8.17. Let $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}$. This forms a subring of the complex numbers, and even an integral domain (exercise: prove it). It is called the ring of *Gaussian integers*.

The ring of Gaussian integers is not a field. Recall that complex conjugation, defined by

$$a + ib \rightarrow a - ib = \overline{a + ib}$$

is a morphism from (\mathbb{C}^*, \cdot) to itself, i.e. $\overline{zw} = \bar{z}\bar{w}$ for any $z, w \in \mathbb{C}$. In particular, for any $z \in \mathbb{C}^*$, we have $\overline{z^{-1}} = \overline{z}^{-1}$.

Assume that $z = a + ib \in \mathbb{Z}[i]$ is invertible. Its conjugate \bar{z} is also in $\mathbb{Z}[i]$, and is also invertible, as $\bar{z}^{-1} = \overline{z^{-1}} \in \mathbb{Z}[i]$. So $z\bar{z}$ is invertible in $\mathbb{Z}[i]$, its inverse is $z^{-1}\bar{z}^{-1}$. But $z\bar{z} = a^2 + b^2$, so $a^2 + b^2$ is invertible in $\mathbb{Z}[i]$.

Now, if $w \in \mathbb{Z}[i]$ is such that $(a^2 + b^2)z = 1$, then we must have (in \mathbb{C}) that $w = \frac{1}{a^2 + b^2}$. But $\frac{1}{a^2 + b^2} \in (0, +\infty)$, so $w \in \mathbb{N}$, and the only possibility is $w = 1$. Therefore $z \in \{1, i, -1, -i\}$.

Example 8.18. Let $\mathbb{Q}[\sqrt{5}] = \{a + \sqrt{5}b : a, b \in \mathbb{Q}\}$. This is a subring of \mathbb{R} . Moreover, we can show that it is a field, in the same way we proved that \mathbb{C} was a field. If $z = a + \sqrt{5}b \in \mathbb{Q}[\sqrt{5}]$, define its conjugate by $\bar{z} = a - \sqrt{5}b$. Assume that a, b are not both zero. To find the inverse, we write

$$(a + \sqrt{5}b)(c + \sqrt{5}d) = 1$$

Multiplying by the conjugate, we find

$$(a^2 - 5b^2)(c + \sqrt{5}d) = a - \sqrt{5}b$$

Note that because $a, b \in \mathbb{Q}$, we have $a^2 - 5b^2 \neq 0$. Indeed, otherwise we could conclude that $\sqrt{5} \in \mathbb{Q}$, which is not true.

Finally we obtain

$$c + \sqrt{5}d = \frac{a}{a^2 - 5b^2} - \sqrt{5} \frac{b}{a^2 - 5b^2}.$$

This field is called the *Golden field*, because of its connection with the golden ratio $\varphi = \frac{1 + \sqrt{5}}{2}$.

Remark 8.19. We could do essentially the same construction for any integer d that is not a square, and obtain a field $\mathbb{Q}[\sqrt{d}]$ (for example $\mathbb{Q}[\sqrt{-1}] = \mathbb{Q}[i]$). A crucial feature of these fields is the existence of the conjugation. More about this when you learn Galois theory.

Finally, I want to show an example of a finite field that is not of the form $(\mathbb{Z}_p, +, \cdot)$:

Example 8.20. Let $\mathbb{F}_4 = \{a + bX : a, b \in \mathbb{F}_2\}$. We equip \mathbb{F}_4 with coordinate wise addition, i.e. $(a + bX) + (c + dX) = a + c + (b + d)X$. For multiplication, coordinate wise multiplication does not give a field. We will fix $X^2 = X + 1$, and the rest is usual polynomial multiplication. This gives the formula:

$$(a + bX)(c + dX) = (ac + bd) + (bc + ad + bd)X$$

This forms a field with 4 elements.

For example, the inverse of X is $X - 1$, as we have $X^2 = X + 1$, so $X(X - 1) = 1$.

The following theorem gives us some useful tool to study integral domains:

Theorem 8.21 (Cancellation laws). *Let R be a commutative ring with identity. Then R is an integral domain if and only if for all $a, b, c \in R$ with $a \neq 0$, if $ab = ac$ then $b = c$.*

Proof. Suppose first that R is an integral domain, and let $a, b, c \in R$ with $a \neq 0$ and $ab = ac$. Then $a(b - c) = 0$, so $a = 0$ or $b - c = 0$ as R is an integral domain. Therefore $b = c$.

Conversely, suppose that R is not an integral domain. Then there are $a, b \in R$ with $a \neq 0$, $b \neq 0$ and $ab = 0$. But we also have $a0 = 0$. So $ab = 0 = a0$, $a \neq 0$ and $b \neq 0$. \square

Here's a fun theorem of Wedderburn:

Theorem 8.22. *Every finite integral domain is a field.*

Proof. Let R be a finite integral domain. The key observation is the following.

Claim. *For any $a \in R \setminus \{0\}$, we have an injective map*

$$\begin{aligned} \mu_a : R \setminus \{0\} &\rightarrow R \setminus \{0\} \\ x &\rightarrow ax \end{aligned}$$

Proof of claim. First, note that this indeed defines a map from $R \setminus \{0\}$ to itself, because if $x \neq 0$, then $ax \neq 0$ as R is an integral domain. Now we show that the map is injective. Let $x, y \in R \setminus \{0\}$, and assume that $ax = ay$. Then $x = y$ by the previous theorem. \square

But $R \setminus \{0\}$ is finite, so μ_a must also be surjective. In particular, there must be some $x \in R$ such that $ax = \mu_a(x) = 1$, i.e. x is the inverse of a . We can do this for any $a \in R \setminus \{0\}$, which shows that R is a field. \square

An important number associated to a ring is its *characteristic*:

Definition 8.23. Let R be a ring, we define, for any natural number n and $r \in R$

$$nr := \underbrace{r + r + \cdots + r}_{n \text{ times}} .$$

The characteristic of a ring R is the smallest natural number n such that $nr = 0$ for all $r \in R$, if it exists. If such a natural number does not exist, the characteristic of R is defined to be zero. We denote it $\text{Char}(R)$.

So for example:

- \mathbb{Z}_n has characteristic n ,
- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and \mathbb{C} all have characteristic zero.

- $\mathbb{Z}_2 \times \mathbb{Z}_3$, with coordinatewise addition and multiplication, has characteristic 6. But $\mathbb{Z}_2 \times \mathbb{Z}_2$ has characteristic 2.

In R has an identity, we only have to look at its order to compute the characteristic:

Lemma 8.24. *Let R be a ring with identity. If 1 has finite order n in $(R, +)$, then $\text{Char}(R) = n$. Otherwise $\text{Char}(R) = 0$.*

Proof. Assume first that 1 has order n . Pick any $a \in R$, then

$$\begin{aligned} na &= \underbrace{a + a + \cdots + a}_{n \text{ times}} \\ &= \underbrace{a \cdot 1 + a \cdot 1 + \cdots + a \cdot 1}_{n \text{ times}} \\ &= a \underbrace{(1 + 1 + \cdots + 1)}_{n \text{ times}} \\ &= a \cdot n1 \\ &= 0 \end{aligned}$$

If 1 has infinite order, then the characteristic is zero. \square

For integral domain, we can say even more:

Theorem 8.25. *Let R be an integral domain. Then $\text{Char}(R)$ is either 0 or prime.*

Proof. Suppose that $\text{Char}(R) \neq 0$. Then 1 has finite order n in $(R, +)$. If $n = pq$ for some $p, q \in \mathbb{N}$, then we have

$$\begin{aligned} 0 &= n1 \\ &= p(q1) \\ &= (p1)(q1) \end{aligned}$$

so either $p1 = 0$ or $q1 = 0$. Assume that $p1 = 0$. As n is the order of 1, we must have $p = n$, and thus $q = 1$. If $q1 = 0$, we obtain similarly that $q = n$ and $p = 1$. So n is prime. \square

8.3. Morphisms, subrings, ideals. Recall that for groups, there are special maps, group morphisms, that are particularly useful. The same goes for rings:

Definition 8.26. Let R and S be ring. A *ring morphism* $\phi : R \rightarrow S$ is a map from R to S such that for all $a, b \in R$, we have:

- $\phi(a + b) = \phi(a) + \phi(b)$,
- $\phi(ab) = \phi(a)\phi(b)$.

A bijective ring morphism is called a *ring isomorphism*. When two rings R and S have an isomorphism between them, we say that they are isomorphic, denoted $R \simeq S$.

Let's look at some examples.

Example 8.27. Fix some $n \in \mathbb{N}$, there is a ring morphism $\pi : \mathbb{Z} \rightarrow \mathbb{Z}_n$ given by $\pi(a) = [a]_n$.

Example 8.28. Consider the ring of complex numbers \mathbb{C} . Then the conjugate map \bar{z} , taking $z = a + ib$ to $\bar{z} = a - ib$ is an isomorphism from $(\mathbb{C}, +, \cdot)$ to itself.

In general, an isomorphism from a ring to itself is called an *automorphism*.

For groups, we also introduced the special class of subgroups, *normal subgroups*, that are exactly the kernels of group morphisms. Similarly, kernel of ring morphism have special properties.

Definition 8.29. Let $\phi : R \rightarrow S$ be a ring morphism. The *kernel* of ϕ , denoted $\ker(\phi)$, is the subring $\phi^{-1}(\{0\})$.

As an exercise, you can check that it is indeed a subring.

But it has one more property. Indeed, suppose that $a \in \ker(\phi)$ and x is any element of R . Then $\phi(xa) = \phi(x)\phi(a) = \phi(x)0 = 0$. So $xa \in \ker(\phi)$ for all $a \in \ker(\phi)$, and the same goes for ax . In other words, $\ker(\phi)$ is stable under exterior multiplication. We make this into a definition:

Definition 8.30. Let R be a ring. A subring I of R is a left *ideal* of R if for all $a \in I$ and $x \in R$, we have $xa \in I$, and a right ideal if $ax \in I$ instead. It is an ideal if it is both a left and right ideal.

We have:

Proposition 8.31. *The kernel of a ring morphism $\phi : R \rightarrow S$ is an ideal.*

Note that we will mostly work with commutative rings, in which there is no distinction between these three notions. In fact, we will now adopt the

Convention: All rings are commutative.

Any ring has at least two ideals, itself and $\{0\}$. They are called the trivial ideals. Sometimes that's all there is:

Proposition 8.32. *A field has no non-trivial ideal.*

Proof. Let K be a field and $I \neq \{0\}$ an ideal. Let $a \in I \setminus \{0\}$. Then $1 = a^{-1}a \in I$. Now consider any $x \in K$, we have $x = x1 \in I$. So $I = K$. \square

Remark that from the proof, we can extract the following:

Fact 8.33. *Any ideal of a ring R , if it contains 1, must be equal to R .*

Let's give some examples.

Example 8.34. In \mathbb{Z} , for any $n \in \mathbb{N}$, we have that $n\mathbb{Z}$ is an ideal. It is non-trivial, unless $n = 1$ in which case it is \mathbb{Z} .

Example 8.35. Let $\mathcal{C}[a, b]$ be the ring of continuous functions $[a, b] \rightarrow \mathbb{R}$ for some $a < b$. For any $c \in (a, b)$, the set of functions $f \in \mathcal{C}[a, b]$ such that $f(c) = 0$ is an ideal. In fact, it is the kernel of the *evaluation morphism*:

$$\begin{aligned} \text{ev}_c : \mathcal{C}[a, b] &\rightarrow \mathbb{R} \\ f &\rightarrow f(c) \end{aligned}$$

Note that there is an important difference between the first example and the second. The ideal $n\mathbb{Z}$ is of the form $\{nk : k \in \mathbb{Z}\}$. However, there is no function f such that $\ker(\phi) = \{fg : g \in \mathcal{C}[a, b]\}$.

We call ideals of the first form *principal*.

Proposition 8.36. *Let R be a ring, and $a \in R$. The set $I = \{ra : r \in R\}$ is an ideal.*

Proof. First, we have $0 = a0 \in I$. Now pick $ra, sa \in I$, for some $r, s \in R$. Then $ra + sa = (r + s)a \in I$. Finally, if we pick any $t \in R$, we have $tra = (tr)a \in I$. So I is an ideal. \square

Definition 8.37. An ideal I of a ring R is principal if there is $a \in R$ such that $I = \{ra : r \in R\}$. In that case, we say that I is *generated* by a , and denote it $\langle a \rangle$.

Sometimes, all ideal are principal:

Theorem 8.38. *Every ideal of $(\mathbb{Z}, +, \cdot)$ is principal, of the form $n\mathbb{Z}$ for some $n \in \mathbb{Z}$.*

Proof. We already proved this, it is exactly Lemma 3.4. To summarize, any non-zero ideal I of \mathbb{Z} is generated by the smallest element n of $I \cap \mathbb{N}$, and then $I = n\mathbb{Z}$. \square

We go back to continuous functions for an example of a non-principal ideal:

Proposition 8.39. *Let $a < b$, some $c \in (a, b)$ and $\mathcal{C}[a, b]$ be the ring of continuous functions $[a, b] \rightarrow \mathbb{R}$. Consider the evaluation morphism:*

$$\begin{aligned} \text{ev}_c : \mathcal{C}[a, b] &\rightarrow \mathbb{R} \\ f &\rightarrow f(c) \end{aligned}$$

Then $\ker(\text{ev}_c)$ is not principal.

Proof. If it was, then there would be a function $f \in \mathcal{C}[a, b]$ such that for all $g \in \mathcal{C}[a, b]$, if $g(c) = 0$, then there is $h \in \mathcal{C}[a, b]$ such that $g = hf$.

First note that this implies that for all $x \in [a, b] \setminus \{c\}$, we have $f(x) \neq 0$. Otherwise, there would be $d \in [a, b] \setminus \{c\}$ such that $f(d) = 0$. This would imply that for all $g \in \mathcal{C}[a, b]$, if $g(c) = 0$, then $g(d) = 0$. But if we look at $g(x) = x - c$, we see that $g(d) \neq 0$.

Now consider $g = \sqrt[3]{f}$. Since $f(c) = 0$, we get $g(c) = 0$, and thus $g \in \ker(\text{ev}_c)$. Consider the function $h \in \mathcal{C}[a, b]$ with $g = hf$. Since $f(x) \neq 0$ for all $x \in [a, b] \setminus \{c\}$, we have $h = \frac{g}{f}$ on $[a, b] \setminus \{c\}$. So outside of $x = c$, we have $h = \frac{f^{\frac{1}{3}}}{f} = \frac{1}{f^{\frac{2}{3}}}$. As $f(c) = 0$ and f is continuous, this implies that $\lim_{x \rightarrow c} h(x)$ does not exist, so h is not continuous at c . \square

Just as for groups, we have a correspondence between ideals and morphisms.

Theorem 8.40. *Let R be a ring and I an ideal. The quotient group R/I is a ring when equipped with the multiplication*

$$(r + I)(s + I) = rs + I$$

for all $r, s \in R$. Moreover, the map

$$\begin{aligned} \pi : R &\rightarrow R/I \\ r &\rightarrow r + I \end{aligned}$$

is a ring morphism with kernel I .

Proof. The main point is to show that this multiplication is well defined. Let $r, r', s, s' \in R$ such that $r + I = r' + I$ and $s + I = s' + I$. This means that there are $a, b \in I$ such that $r = r' + a$ and $s = s' + b$. Then

$$\begin{aligned} rs &= (r' + a)(s' + b) \\ &= r's' + as' + br' + ab \end{aligned}$$

Since $as' + br' + ab \in I$, we obtain $rs + I = r's' + I$. This shows that the multiplication is well-defined. The associative and distributive laws follow from associativity and distributivity in \mathbb{R} .

We already know that this map π is a surjective morphism for addition. It remains to show it respects multiplication. Let $r, s \in R$, then

$$\begin{aligned}\pi(rs) &= rs + I \\ &= (r + I)(s + I) \text{ by definition} \\ &= \pi(r)\pi(s) .\end{aligned}$$

Finally, we need to show that the kernel of π is I . Let $r \in \ker(\pi)$, then $r + I = I$, which means that there is $a \in I$ such that $r + a = 0$, i.e. $r = -a \in I$. \square

We call π the *canonical* morphism associated to I .

Example 8.41. A example of a quotient ring that we have already seen is $\mathbb{Z}/n\mathbb{Z}$: it is the quotient of \mathbb{Z} by its ideal $n\mathbb{Z}$.

Finally, just as we had isomorphism theorems for groups, we have them for rings! Let me state them without proof (they are very similar to the proofs for groups).

Theorem 8.42 (First isomorphism theorem). *Let $\phi : R \rightarrow S$ be surjective ring homomorphism and $\pi : R \rightarrow R/\ker(\phi)$ the canonical morphism. Then there is a unique isomorphism $\eta : R/\ker(\phi) \rightarrow S$ such that $\eta \circ \pi = \phi$.*

As an application, we have:

Example 8.43. Let $\mathcal{C}[a, b]$ be the set of continuous functions from $[a, b]$ to \mathbb{R} , some $c \in (a, b)$, and consider the morphism

$$\begin{aligned}\text{ev}_c : \mathcal{C}[a, b] &\rightarrow \mathbb{R} \\ f &\rightarrow f(c)\end{aligned}$$

It is surjective. We have seen that its kernel is the set of functions g such that $g(c) = 0$. By the first isomorphism theorem, we get $\mathcal{C}[a, b]/\ker(\text{ev}_c) \simeq \mathbb{R}$.

We have the correspondence theorem too:

Theorem 8.44 (Correspondence theorem). *Let R be a ring and I an ideal of R . Then $S \rightarrow S/I$ is a bijection between subrings of R containing I and subrings of R/I . Moreover, it restricts to a bijection between ideals of R containing I and ideals of R/I .*

To state the second isomorphism theorem, we need:

Proposition 8.45. *Let I be a subring of R and J be an ideal of R . Then*

$$I + J = \{a + b : a \in I, b \in J\}$$

is a subring of R . If I is an ideal of R , then it is an ideal of R .

And we now state:

Theorem 8.46 (Second isomorphism theorem). *Let R be a ring, I a subring of R and J an ideal of R . Then $I \cap J$ is an ideal of R and*

$$I/I \cap J \simeq (I + J)/J .$$

Finally:

Theorem 8.47 (Third isomorphism theorem). *Let R be a ring and I, J ideals of R , with $J \subset I$. Then*

$$R/I \simeq \frac{R/J}{I/J} .$$

In this class, I will not expect you to apply the second isomorphism theorem for rings.

8.4. Polynomial rings. In calculus, we learned about polynomial functions, for example $P(x) = 4x^3 + 2x - 6$ or $Q(x) = 13x^{10} - 2x^2$. We know how to add and multiply these functions together to obtain expressions for $P(x) + Q(x)$ and $P(x)Q(x)$.

In fact, it really is not relevant for computations that polynomials are functions: to add and multiply them, we really just apply some abstract set of rules. This idea is at the core of the study of *polynomial rings*.

In this subsection, R is a ring with identity.

Definition 8.48. The set of polynomials with indeterminate x over R is the set of expressions of the form

$$P(x) = \sum_{i=0}^n a_i x^i = a_0 + a_1 x + \cdots + a_n x^n$$

for $i \in \mathbb{Z}^+$, with $a_0, a_1, \dots, a_n \in R$, and $a_n \neq 0$ if $n \neq 0$. We denote this set $R[x]$.

The a_i are called the *coefficients* of P , and a_n its called is *leading coefficient*. The term $a_n x^n$ is called its *leading term*.

If $n > 0$, it is called the *degree* of P , denoted $\deg(P)$. If $n = 0$ and $a_0 \neq 0$, then we have $P = a_0$, a *constant* polynomial, and $\deg(P) = 0$. By convention, if $P = 0$, we have $\deg(P) = 0$. This is actually a very useful convention, as we will see later.

Notation. We will often write P, Q, \dots instead of $P(x), Q(x)$, when it is clear what the variable is.

We have to equip it with addition and multiplication to make it into a ring. We define them exactly as they were defined for functions. Consider two polynomials:

$$P = \sum_{i=0}^n a_i x^i$$

$$Q = \sum_{j=0}^m b_j x^j .$$

We define

$$P + Q = \sum_{k=0}^{\max(n,m)} c_k x^k$$

where $c_k = a_k + b_k$.

We define multiplication as

$$PQ = \sum_{k=0}^{n+m} d_k x^k$$

where

$$d_k = \sum_{i=0}^k a_i b_{k-i} = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0 = \sum_{i+j=k} a_i b_j .$$

Note that in these formulas, we use the convention that if $k > n$, then $a_k = 0$, and if $k > m$, then $b_k = 0$.

Maybe these look a bit unusual, but the bottom line is that *you add and multiply abstract polynomials just as you did when they were functions.*

Proposition 8.49. *The set $R[x]$ is a ring with identity, equipped this addition and multiplication.*

Proof. We have to check that all the properties of a ring hold. That addition makes it into an abelian group is a consequence of the fact that $(R, +)$ is an abelian group. The zero polynomial is the additive identity. For multiplication, the polynomial 1 is the identity.

The book has the proof that multiplication is associative. I will write the proof of distributivity, and leave commutativity as an exercise. Let $P = \sum_{i=0}^n a_i x^i$, $Q = \sum_{j=0}^m b_j x^j$ and $R = \sum_{k=0}^p c_k x^k$ be polynomials. Then

$$\begin{aligned} P(Q + R) &= P \sum_{j=0}^{\max(m,p)} (b_j + c_j) x^j \\ &= \sum_{k=0}^{n+\max(m,p)} \left(\sum_{i+j=k} a_i (b_j + c_j) \right) x^k \\ &= \sum_{k=0}^{\max(n+m,n+p)} \left(\sum_{i+j=k} a_i b_j \right) x^k + \left(\sum_{i+j=k} a_i c_j \right) x^k \\ &= \sum_{k=0}^{\max(n+m,n+p)} \left(\sum_{i+j=k} a_i b_j \right) x^k + \sum_{k=0}^{\max(n+m,n+p)} \left(\sum_{i+j=k} a_i c_j \right) x^k \\ &= \sum_{k=0}^{n+m} \left(\sum_{i+j=k} a_i b_j \right) x^k + \sum_{k=0}^{n+p} \left(\sum_{i+j=k} a_i c_j \right) x^k \\ &= PQ + PR . \end{aligned}$$

□

Definition 8.50. The ring $R[x]$ is called the ring of polynomials over R with indeterminate x .

From the formulas, we see that:

Proposition 8.51. *Let P, Q be two polynomials. Then*

$$\begin{aligned} \deg(P + Q) &\leq \max(\deg(P), \deg(Q)) \\ \deg(PQ) &\leq \deg(P) + \deg(Q) . \end{aligned}$$

Proof. This follows from the formulas for addition and multiplication, as long as P and Q are non-zero. If one of them is zero, the formulas still work because we picked $\deg(0) = -\infty$. \square

Here are examples that show that we do not have equality in general:

- in $\mathbb{R}[x]$, let $P = x^2 - 3$ and $Q = -x^2 + x$, then $\deg(P + Q) = \deg(x - 3) = 1 < 2$,
- in $\mathbb{Z}_6[x]$, let $P = 2x$ and $Q = 3x + 1$, then:

$$\begin{aligned} PQ &= 2x(3x + 1) \\ &= 6x^2 + 2x \\ &= 2x \end{aligned}$$

$$\text{so } \deg(PQ) = 1 < \deg(P) + \deg(Q) = 2.$$

Note that to get strict inequality for the degree of the product, we have to work over a ring that is not an integral domain. This is true in general:

Theorem 8.52. *If R is an integral domain, then $R[x]$ is also an integral domain, and for any $P, Q \in R[x]$, we have*

$$\deg(PQ) = \deg(P) + \deg(Q)$$

Proof. Let P, Q be two polynomials. As before, if one of them is zero, the equality of the theorem is verified, so assume they are both non-zero, of degree $n, m \in \mathbb{N}$ respectively, and dominant coefficients a_n and b_m .

The leading term of PQ is $a_n b_m x^{n+m}$. Because the dominant coefficient of a non-zero polynomial is always non-zero, we have $a_n \neq 0$ and $b_m \neq 0$, which implies, as R is an integral domain, that $a_n b_m \neq 0$. This gives us $\deg(PQ) = n + m = \deg(P) + \deg(Q)$. This also proves that the product of any two non-zero polynomials is non-zero, so $R[x]$ is an integral domain. \square

We can connect this abstract definition of polynomials with actual functions:

Proposition 8.53. *Let $\mathcal{C}(\mathbb{R})$ be the ring of continuous functions from \mathbb{R} to \mathbb{R} . The map:*

$$\begin{aligned} \phi : \mathbb{R}[x] &\rightarrow \mathcal{C}(\mathbb{R}) \\ P &\rightarrow (x \rightarrow P(x)) \end{aligned}$$

is an injective ring morphism. It is not surjective.

Proof. Exercise/homework \square

We can also evaluate polynomials over a ring. Let R be any ring, some $a \in R$, and consider the map:

$$\begin{aligned} \text{ev}_a : \mathbb{R}[x] &\rightarrow R \\ P &\rightarrow P(a) \end{aligned}$$

It is a morphism, which is called the *evaluation morphism at a* . It is not injective in general. For example $\text{ev}_a(x - a) = 0$ for all $a \in R$.

You probably saw multivariable polynomials in calculus, for example $x^2y - 2x + 2$. We can also obtain these as rings of polynomials, by iterating the construction. For example, to obtain polynomials in x and y over R , we can form $(R[x])[y]$, i.e. polynomials with indeterminate y over $R[x]$.

But wait... What if we construct $(R[y])[x]$ instead? It should not come as a surprise that:

Proposition 8.54. *For any ring R , there is a unique isomorphism between the polynomial rings $(R[x])[y]$ and $(R[y])[x]$ that fixes R , x and y .*

We therefore identify these two polynomial rings and call them the *polynomial ring with indeterminates x and y over R* , which we denote $R[x, y]$. We can do this construction for any number of variables and obtain polynomial rings $R[x_1, \dots, x_n]$ for any $n \in \mathbb{N}$.

Before moving on, I want to revisit our finite field example \mathbb{F}_4 .

Example 8.55. Consider the polynomial ring $\mathbb{F}_2[x]$, and the ideal $\langle x^2 - x - 1 \rangle$.

The field \mathbb{F}_4 can be defined as $\mathbb{F}_2[x]/\langle x^2 - x - 1 \rangle$.

One issue with this construction is that we still have to check by hand that the quotient is a field. What we'd like is a way to tell, just from the polynomial $x^2 - x - 1$, that the quotient $\mathbb{F}_2[x]/\langle x^2 - x - 1 \rangle$ is a field. Finding a way to do this is the point of the next three subsections.

8.5. Prime and maximal ideals. Fields and integral domains are of central importance in many areas of mathematics, from number theory to geometry. Essentially, they have enough rigidity to be easier to manipulate than arbitrary rings, but are still able to encode a lot of information about equations/numbers/shapes.

This is why we would like to be able to produce fields and integral domains from any ring. This is the point of *maximal ideals*.

Definition 8.56. An ideal I of a ring R is *maximal* if $R \neq I$ and for any other ideal J or R , if $I \subset J$, then $J = I$ or $J = R$.

We will prove that the quotient of any ring by a maximal ideal is a field, but we start with a lemma:

Lemma 8.57. *A ring with identity R is a field if and only if R has no ideal besides 0 and R .*

Proof. The left to right direction has been proved before. Now suppose that R has no ideal besides 0 and R .

Let $a \in R \setminus \{0\}$. The ideal generated by R is non-zero, and therefore equal to R . In particular 1 is in $\langle a \rangle$, so there is $b \in R$ such that $ba = 1$. This implies that b is the multiplicative inverse of a . Therefore R is a field. \square

We can now prove:

Theorem 8.58. *Let R be a ring with identity. An ideal I of R is maximal if and only if R/I is a field.*

Proof. Let I be an ideal of R and $\pi : R \rightarrow R/I$ the quotient map. Assume that I is maximal. Let J be an ideal of R/I , then $\pi^{-1}(J)$ is an ideal of R containing I . Therefore it must be either I or R . In the first case, we get that $J = 0$, in the second that $J = R$. By Lemma 8.57, we conclude that R/I is a field.

Conversely, assume that R/I is a field. Then R/I has at least two elements, 0 and 1 , therefore $I \neq R$. Let J be an ideal of R containing I , with $J \neq I$. Let $a \in J \setminus I$. Then $\pi(a) \neq 0$, so it has a multiplicative inverse in R/I , which, because π is surjective, can be written $\pi(b)$ for some $b \in R$. Then $1 = \pi(a)\pi(b) = \pi(ab)$,

and in particular $ab \in 1 + I$. So there is $r \in I$ such that $ab = 1 + r$. This implies that $1 = ab - r$. As $ab \in J$ and $r \in I$, we get $1 \in J$ and thus $J = R$. \square

Example 8.59. In $(\mathbb{Z}, +, \times)$, all ideals are of the form $n\mathbb{Z}$, and $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime. Therefore $n\mathbb{Z}$ is maximal if and only if n is prime.

If n is not prime, it is contained in $d\mathbb{Z}$, for any $d|n$ with $d \neq 1, n$.

Example 8.60. Consider again an interval $[a, b]$, $a \neq b$, some $c \in (a, b)$, and the kernel of the evaluation map:

$$\begin{aligned} \text{ev}_c : \mathcal{C}[a, b] &\rightarrow \mathbb{R} \\ f &\rightarrow f(c) \end{aligned}$$

Then $\ker(\text{ev}_c)$ is maximal. This can be seen by the first isomorphism theorem: the image \mathbb{R} is a field, and we know that $\mathbb{R} \simeq \mathcal{C}[a, b]/\ker(\pi)$.

Integral domains are the next best things compared to fields, and in fact we can always make a field from an integral domain in a canonical way, called its *quotient field*. I will not detail this construction here.

So we may also ask whether there is a type of ideal corresponding to integral domains.

Definition 8.61. Let R be a ring. An ideal I of R is *prime* if for all $a, b \in R$, if $ab \in I$, then either $a \in I$ or $b \in I$.

We have:

Theorem 8.62. Let R be a ring with identity and I an ideal of R . Then I is a prime ideal in R if and only if R/I is an integral domain.

Proof. Assume first that I is prime, let $\pi : R \rightarrow R/I$ be the quotient map. Pick any two non-zero elements of R/I , which can be written as $\pi(a), \pi(b)$ for some $a, b \in R$, as π is surjective. Assume that $\pi(a)\pi(b) = 0$, then $\pi(ab) = 0$, so $ab \in I$. Because I is prime, either $a \in I$, in which case $\pi(a) = 0$, or $b \in I$, in which case $\pi(b) = 0$. So R/I is an integral domain.

Now assume that R/I is an integral domain. Let $a, b \in R$ with $ab \in I$. Then $\pi(ab) = 0$, so $\pi(a)\pi(b) = 0$. This implies that either $\pi(a) = 0$ or $\pi(b) = 0$. Therefore either $a \in I$ or $b \in I$. Thus I is prime. \square

Since any field is an integral domain, we have obtained:

Corollary 8.63. Any maximal ideal of a ring with identity is prime.

The converse is not true. To find prime ideals that are not maximal, it is enough to obtain an integral domain which is not a field as a quotient ring. Here is an example with \mathbb{Z} :

Example 8.64. Consider $(\mathbb{Z}, +, \times)^2$ with coordinate-wise addition and multiplication. It is a commutative ring with identity $(1, 1)$.

The subset $\{0\} \times \mathbb{Z}$ is an ideal of \mathbb{Z}^2 . It is easy to check directly that it is prime. Alternatively, we can see that the map

$$\begin{aligned} \pi : \mathbb{Z}^2 &\rightarrow \mathbb{Z} \\ (a, b) &\rightarrow b \end{aligned}$$

has kernel $\{0\} \times \mathbb{Z}$, which shows that $\{0\} \times \mathbb{Z}$ is prime by the first isomorphism theorem.

However, it is not maximal as \mathbb{Z} is not a field. To see this directly, notice that it is contained in the ideal $p\mathbb{Z} \times \mathbb{Z}$. This one is maximal: the quotient is \mathbb{F}_p .

Finally, an example of an ideal that is not prime:

Example 8.65. Consider the two variable polynomial ring $F[x, y]$, for some field F , and the ideal $\langle xy \rangle$. This ideal is not prime. Indeed, we have $x, y \notin \langle xy \rangle$, but $xy \in \langle xy \rangle$.

The quotient $F[x, y]/\langle xy \rangle$ is therefore not an integral domain. If $\pi : F[x, y] \rightarrow F[x, y]/\langle xy \rangle$ is the quotient map, then we have $\pi(x)\pi(y) = 0$, but $\pi(x) \neq 0$ and $\pi(y) \neq 0$.

Going back once again to the field $\mathbb{F}_2[x]/\langle x^2 - x - 1 \rangle$, we have now reduced the problem of showing this is a field to showing that the ideal $\langle x^2 - x - 1 \rangle$ is maximal. But how do we show this? This will be revealed in the next two sections.

8.6. The division algorithm. Recall the division algorithm for integers: for any integers a, b with $b > 0$, there are unique integers q, r with

- $0 \leq r < b$,
- $a = bq + r$.

We now state and prove the polynomial version:

Theorem 8.66 (The division algorithm). *Let F be a field and $f, g \in F[x]$, with $g \neq 0$. There are unique polynomials $q, r \in F[x]$ such that:*

- $-\infty \leq \deg(r) < \deg(g)$,
- $f = gq + r$.

Before proving the theorem, let me show how it works in practice: consider $f = 2x^3 - x + 1$ and $g = x^2 - x$.

We multiply g by $2x$ and subtract from f to get:

$$\begin{aligned} f - 2xg &= 2x^3 - x + 1 - 2x(x^2 - x) \\ &= 2x^3 - x + 1 - 2x^3 + 2x^2 \\ &= 2x^2 - x + 1 \end{aligned}$$

This still has degree greater or equal to $\deg(g)$, so we repeat the process:

$$\begin{aligned} 2x^2 - x + 1 - 2(x^2 - x) &= 2x^2 - x + 1 - 2x^2 + 2x \\ &= x + 1 \end{aligned}$$

We now insert this back into the first equation:

$$\begin{aligned} f - 2xg &= 2x^2 + 1 \\ &= 2g + x + 1 \end{aligned}$$

And therefore:

$$f = (2x + 2)g + x + 1$$

So just like for euclidian division of numbers, the goal is simply to make the rest as small as possible. If you want to do that by hand, it is just like long division of integers. To prove that the algorithm works in general, we need the well-ordering principle.

Proof of the division algorithm. Consider the set $S = \{f - gq : q \in F[x]\}$, and the subset $\deg(S)$ of $\{-\infty\} \cup \mathbb{N}$ given by $\deg(S) = \{\deg(P) : P \in S\}$.

The set $\deg(S)$ is non-empty, and bounded from below by $-\infty$. By the well-ordering principle ⁴, the set $\deg(S)$ must have a smallest element.

Suppose that $s \in S$ with $\deg(s) \geq \deg(g)$, and write $s = f - gq_0$ for some $q_0 \in F[x]$. We can also write $s = \sum_{i=1}^n a_i x^i$ and $g = \sum_{j=1}^m b_j x^j$. Since we assume that $\deg(s) \geq \deg(g)$, we have $n \geq m$. Consider the polynomial $s - \frac{a_n}{b_m} x^{n-m} g$.

We have

$$\begin{aligned} s - \frac{a_n}{b_m} x^{n-m} g &= f - gq_0 - \frac{a_n}{b_m} x^{n-m} g \\ &= f - g\left(q_0 + \frac{a_n}{b_m} x^{n-m}\right) \end{aligned}$$

so this polynomial is in S . Its degree n coefficient is $a_n - \frac{a_n}{b_m} b_m = 0$, so it has degree strictly less than $\deg(s)$.

This implies that the least element of $\deg(S)$ must be strictly less than $\deg(g)$. Therefore, there is $r \in S$ such that $\deg(r) < \deg(g)$. This gives the existence part of the theorem.

For the uniqueness part, assume that there are two pairs of polynomials r_1, q_1 and r_2, q_2 satisfying the conditions of the theorem. Then we have $gq_1 + r_1 = gq_2 + r_2$, so $g(q_1 - q_2) = r_2 - r_1$. In terms of degree, this gives us

$$\begin{aligned} \deg(g) + \deg(q_1 - q_2) &= \deg(r_2 - r_1) \\ &\leq \deg(r_1) \\ &< \deg(g) \end{aligned}$$

Therefore $\deg(q_1 - q_2)$ cannot be positive, and thus must be $-\infty$, which implies $q_1 = q_2$. \square

We now define zeroes of polynomials:

Definition 8.67. Let F be a field, some element $a \in F$ and $P \in F[x]$. We say that a is a *zero*, or *root*, of P if $P(a) = 0$. More precisely, we have the evaluation morphism:

$$\begin{aligned} \text{ev}_a : F[x] &\rightarrow F \\ P &\rightarrow P(a) \end{aligned}$$

and a zero is an element of $\ker(\text{ev}_a)$.

We also can define divisibility for polynomials:

Definition 8.68. Let f, g be polynomials in $F[x]$, for some field F . We say that f *divides* g , and write $f|g$, if there is $q \in F[x]$ such that $g = fq$.

⁴We are applying the well-ordering principle to a set that is not a subset of \mathbb{Z} here. However, recall that the well-ordering principle only depends of the property of \mathbb{Z} as an ordered set. Therefore we can identify $\{-\infty\} \cup \mathbb{N}$ with $\{-1\} \cup \mathbb{N}$ and be fine.

Remark that we could, in fact, make this definition for any ring, not just a polynomial ring.

The following characterization of zeroes is very useful:

Lemma 8.69. *Let F be a field and $p \in F[x]$. Then $a \in F$ is a zero of p if and only if $x - a$ divides p .*

Proof. Assume that a is a zero of p . By the division algorithm, there is $q, r \in F[x]$ such that $p = q(x - a) + r$ and $\deg(r) < \deg(x - a) = 1$. Therefore r is a constant polynomial, i.e. $r \in F$. Evaluating at a , we obtain:

$$\begin{aligned} 0 &= p(a) \\ &= q(a) \times 0 + r \\ &= r \end{aligned}$$

so $p = q(x - a)$, i.e. $x - a$ divides p .

Now assume that $x - a$ divides p , so there is $q \in F[x]$ with $p = q(x - a)$. Then $p(a) = q(a) \times 0 = 0$, so a is a zero of p . \square

Finally, we can bound the number of roots of a polynomial using its degree:

Theorem 8.70. *Let F be a field and $p \in F[x]$ a non-zero polynomial. Then p has at most $\deg(p)$ distinct roots in F .*

Proof. We use induction on $\deg(p)$. If $\deg(p) = 0$, then p has no roots because $p \neq 0$.

Let p be of degree $n > 0$. If p has no roots, we are done. Now assume that p has a root a . By the previous lemma, we know that $x - a$ divides p , so let $q \in F[x]$ such that $p = (x - a)q$.

Let b be a root of p with $a \neq b$. Then $0 = p(b) = (b - a)q(b)$. Since $b \neq a$, this implies that $q(b) = 0$, so b is a root of q .

Therefore, all roots of p distinct from a are roots of q . By induction, the polynomial q has less than $\deg(q)$ roots, and we can compute that $\deg(q) = n - 1$. Therefore p has less than $n = \deg(p)$ roots. \square

Remark 8.71. *It is not the case that a polynomial of degree n must have n distinct roots, or even any root at all!*

For example, the polynomial $x^2 + 1 \in \mathbb{R}[x]$ has no root in \mathbb{R} : a root would give a number with a negative square. However, note that as a polynomial in $\mathbb{C}[x]$, it has i and $-i$ as its two roots.

8.7. Irreducible polynomials. Irreducible polynomials will play for polynomial rings $F[x]$ the same role as prime numbers for integers. In particular, they will give rise to fields when quotienting.

Definition 8.72. Let F be a field. A polynomial $f \in F[x]$ is *irreducible* if it is not in F , and there does not exist $g, h \in F[x]$ with both

- $\deg(g) < \deg(f)$ and $\deg(h) < \deg(f)$,
- $f = gh$.

For example, any polynomial f of degree 1 is irreducible: if $f = hg$, we cannot have $\deg(h), \deg(g) < 1$. Here is a more interesting example:

Example 8.73. The polynomial $x^2 + 1 \in \mathbb{R}[x]$ is irreducible.

Indeed, suppose it was not. Then there is $g \in \mathbb{R}[x]$ with $\deg(g) = 1$ and $g|x^2 + 1$. We can write $g = ax + b$ for some $a, b \in \mathbb{R}$ with $a \neq 0$. Then $\frac{b}{a}$ is a root of g , and therefore also of $x^2 + 1$, which is a contradiction.

We want to show that just like for prime numbers, irreducible polynomials allow use to construct fields. We will need the following result, of independent interest:

Theorem 8.74. *Let F be a field. Then every ideal in $F[x]$ is principal.*

Proof. Let I be an ideal of $F[x]$. If I is the zero ideal, the theorem is true. So assume that $I \neq \{0\}$.

Pick some $g \in I \setminus \{0\}$ of minimal degree. If $\deg(g) = 0$, then $g \in F$, and thus $gg^{-1} = 1 \in I$. Therefore $I = F[x] = \langle 1 \rangle$.

Now assume that $\deg(g) > 0$. Let $f \in I$. By the division algorithm, there are $q, r \in F[x]$ such that $f = qg + r$ and $\deg(r) < \deg(g)$. This implies that $r = f - qg \in I$. Because $\deg(g)$ is minimal, we obtain $r = 0$, and thus $f = qg$. This gives us $I = \langle g \rangle$. \square

Warning. This fails for more than one variable. For example, in $F[x, y]$, the ideal $\langle x, y \rangle = \{xp + yq : p, q \in F[x, y]\}$ is not principal.

Indeed, assume that is was, and that $\langle x, y \rangle = \langle q \rangle$ for some $q \in F[x, y]$. Then $q|x$ and $q|y$. One can show that the only possibility is $q \in F$, in which case $\langle x, y \rangle = F[x, y]$. But this is not true, as $\langle x, y \rangle$ does not contain any element of F .

This theorem means that if we want to understand ideals in $F[x]$, it is enough to consider principal ideals. In particular, we can characterize maximal ideals:

Theorem 8.75. *Let F be a field and $p \in F[x]$. Then the ideal $\langle p \rangle$ of $F[x]$ is maximal if and only if p is irreducible.*

Proof. Assume first that $\langle p \rangle$ is maximal. Note that this implies that $p \neq 0$. Assume that $p = gh$ for some $g, h \in F[x]$ of degree strictly less than $\deg(p)$. Since $\langle p \rangle$ is maximal, it must be prime, and therefore $g \in \langle p \rangle$ or $h \in \langle p \rangle$.

Assume that $g \in \langle p \rangle$, then there is $q \in F[x]$ with $g = qp$. Taking degrees, we obtain that $\deg(g) = \deg(q) + \deg(p)$. But $\deg(g) < \deg(p)$, which implies that $\deg(q) = -\infty$, or equivalently $q = 0$. This is a contradiction as then $g = 0$, and thus $p = 0$. Similarly, we get a contradiction if $h \in \langle p \rangle$. So p is irreducible.

Conversely, assume that p is irreducible. This implies that $p \notin F$, and in particular $\langle p \rangle \neq F[x]$. Let I be an ideal of $F[x]$ containing $\langle p \rangle$. By Theorem 8.74, we know that I is principal. Therefore there is $f \in F[x]$ with $I = \langle f \rangle$. This implies $p \in \langle f \rangle$, and thus $f|p$. So there is $q \in F[x]$ with $p = fq$. This implies that $\deg(p) = \deg(f) + \deg(q)$, and in particular that $0 \leq \deg(f) \leq \deg(p)$ and $0 \leq \deg(q) \leq \deg(p)$.

Because f is irreducible, we obtain that $\deg(p) = \deg(f)$ or $\deg(q) = \deg(f)$. Therefore either $\deg(q) = 0$ or $\deg(f) = 0$.

If $\deg(q) = 0$, then $q \in F \setminus \{0\}$, and $p = q^{-1}f$, so $I = \langle p \rangle = \langle f \rangle = J$. If $\deg(f) = 0$, then $f \in F \setminus \{0\}$, so $I = F[x]$. \square

We have obtain the following corollary:

Corollary 8.76. *Let F be a field, and $p \in F[x]$ an irreducible polynomial. Then $F[x]/\langle p \rangle$ is a field.*

We can use this to recover the complex numbers:

Proposition 8.77. *The quotient $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ is a field, isomorphic to \mathbb{C} .*

Proof. We already know that $x^2 + 1$ is irreducible. So we just have to prove that the quotient is isomorphic to \mathbb{C} . To see this, consider the morphism:

$$\begin{aligned} \text{ev}_i : \mathbb{C}[x] &\rightarrow \mathbb{C} \\ P &\rightarrow P(i) \end{aligned}$$

It restricts to a morphism:

$$\begin{aligned} \text{ev}_i : \mathbb{R}[x] &\rightarrow \mathbb{C} \\ P &\rightarrow P(i) \end{aligned}$$

The kernel of this morphism is an ideal of $\mathbb{R}[x]$, which must be principal. It contains $x^2 + 1$, but no degree one polynomial (otherwise we would have $i \in \mathbb{R}$). Therefore $\ker(\text{ev}_i) = \langle x^2 + 1 \rangle$, and we get the result by the first isomorphism theorem. \square

Note that this allows us to reconstruct \mathbb{C} from \mathbb{R} alone.

For another application, let us revisit, once again, the field with four elements.

Proposition 8.78. *In the ring $\mathbb{F}_2[x]$, the polynomial $x^2 - x - 1$ is irreducible. Therefore the quotient $\mathbb{F}_2[x]/\langle x^2 - x - 1 \rangle$ is a field, and it has four elements.*

Proof. Suppose, for a contradiction, that $x^2 - x - 1$ is not irreducible. Therefore there must be two other polynomials $p, q \in \mathbb{F}_2[x]$ such that $pq = x^2 - x - 1$ and degree strictly less than 2. Neither of them can have degree 0 or $-\infty$, therefore they must be both of degree 1. This implies that they are of the form $ax + b$, for some $a, b \in \mathbb{F}_2$ with $a \neq 0$. In particular, write $p = ax + b$ for such a, b . Then $\frac{-b}{a}$ is a root of p , and therefore a root of $x^2 - x - 1$.

But $x^2 - x - 1$ has no root: evaluating at 0 or 1 gives 1. This is a contradiction, so $x^2 - x - 1$ is irreducible, which implies, by Theorem 8.75, that $\mathbb{F}_2[x]/\langle x^2 - x - 1 \rangle$ is a field.

We still have to determine its number of elements. Note that it has size at least four, because the classes of $0, 1, x$ and $x + 1$ are distinct: the difference of any two of these polynomials has degree strictly less than 2, thus cannot belong to $\langle x^2 - x - 1 \rangle$. Moreover, since $x^2 = x + 1$ in $\mathbb{F}_2[x]/\langle x^2 - x - 1 \rangle$, we see that any class has an element of degree less or equal to 1, which must be one of $0, 1, x$ and $x + 1$. This gives the result. \square

This may seem like it's a bit too much for constructing a field with four elements. The advantage of this method is that it will generalize to any prime number p : we construct a field with p^2 elements by finding a degree 2 irreducible polynomial in \mathbb{F}_p .

In general, for any $n \in \mathbb{N}$ and prime p , we can construct a field of size p^n using similar techniques. But that's a story for another course.

REFERENCES

- [1] Hans Ulrich Besche, Bettina Eick, and Eamonn A O'Brien. The groups of order at most 2000. *Electron. Res. Announc. Amer. Math. Soc.*, 7(1-4):3, 2001.