Introduction to Abstract Algebra (Math 113)

Alexander Paulin

Contents

1 Introduction

1.1 What is Algebra?

If you ask someone on the street this question, the most likely response will be: "Something horrible to do with x, y and z". If you're lucky enough to bump into a mathematician then you might get something along the lines of: "Algebra is the abstract encapsulation of our intuition for composition". By composition, we mean the concept of two object coming together to form a new one. For example adding two numbers, or composing real valued single variable functions. As we shall discover, the seemly simple idea of composition hides vast hidden depth.

Algebra permeates all of our mathematical intuitions. In fact the first mathematical concepts we ever encounter are the foundation of the subject. Let me summarize the first six to seven years of your mathematical education:

The concept of *Unity*. The number 1.
You probably always understood this, even as a little baby.

$\downarrow$

$\mathbb{N} := \{1, 2, 3...\}$, the natural numbers. $\mathbb{N}$ comes equipped with two natural operations $+$ and $\times$.

$\downarrow$

$\mathbb{Z} := \{... - 2, -1, 0, 1, 2, ...\}$, the integers.
We form these by using geometric intuition thinking of $\mathbb{N}$ as sitting on a line. $\mathbb{Z}$ also comes with $+$ and $\times$. Addition on $\mathbb{Z}$ has particularly good properties, e.g. additive inverses exist.

$\downarrow$

$\mathbb{Q} := \{\frac{a}{b} \,|a, b \in \mathbb{Z}, b \neq 0\}$, the rational numbers. We form these by taking $\mathbb{Z}$ and *formally* dividing through by non-negative integers. We can again use geometric insight to picture $\mathbb{Q}$ as points on a line. The rational numbers also come equipped with $+$ and $\times$. This time, multiplication is has particularly good properties, e.g non-zero elements have multiplicative inverses.

We could continue by going on to form $\mathbb{R}$, the real numbers and then $\mathbb{C}$, the complex numbers. This process is of course more complicated and steps into the realm of mathematical analysis.

Notice that at each stage the operations of $+$ and $\times$ gain additional properties. These ideas are very simple, but also profound. We spend years understanding how $+$ and $\times$ behave in $\mathbb{Q}$. For example

$$a + b = b + a \text{ for all } a, b \in \mathbb{Q},$$

or

$$a \times (b + c) = a \times b + a \times c \text{ for all } a, b, c \in \mathbb{Q}.$$

The central idea behind abstract algebra is to define a larger class of objects (sets with extra structure), of which $\mathbb{Z}$ and $\mathbb{Q}$ are definitive members.

$$\begin{aligned} (\mathbb{Z}, +) &\longrightarrow \textit{Groups} \\ (\mathbb{Z}, +, \times) &\longrightarrow \textit{Rings} \\ (\mathbb{Q}, +, \times) &\longrightarrow \textit{Fields} \end{aligned}$$

In linear algebra the analogous idea is

$$(\mathbb{R}^n, +, \text{scalar multiplication}) \longrightarrow \textit{Vector Spaces over } \mathbb{R}$$

The amazing thing is that these vague ideas mean something very precise and have far far more depth than one could ever imagine.

1.2 Sets and Functions

A set is any collection of objects. For example six dogs, all the protons on Earth, every thought you've ever had, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Observe that $\mathbb{Z}$ and $\mathbb{Q}$ are sets with extra structure coming from $+$ and $\times$. In this whole course, all we will study are sets with some carefully chosen extra structure.

Basic Logic and Set Notation

Writing mathematics is fundamentally no different than writing english. It is a language which has certain rules which must be followed to accurately express what we mean. Because mathematical arguments can be highly intricate it is necessary to use simplifying notation for frequently occurring concepts. I will try to keep these to a minimum, but it is crucial we all understand the following:

- If P and Q are two statements, then $P \Rightarrow Q$ means that if P is true then Q is true. For example: x odd $\Rightarrow x \neq 2$. We say that P implies Q.
- If $P \Rightarrow Q$ and $Q \Rightarrow P$ then we write $P \iff Q$, which should be read as P is true if and only if Q is true.
- The symbol $\forall$ should be read as "for all".
- The symbol $\exists$ should be read as "there exists". The symbol $\exists!$ should be read as "there exists unique".

Let S and T be two sets.

- If s is an object contained in S then we say that s is an *element*, or a *member* of S. In mathematical notation we write this as $s \in S$. For example $5 \in \mathbb{Z}$. Conversely $s \notin S$ means that x is not contained in S. For example $\frac{1}{2} \notin \mathbb{Z}$.
- If S has finitely many elements then we say it is a finite set. We denote its cardinality (or size) by $|S|$.
- The standard way of writing down a set S is using *curly bracket* notation.

  $$S = \{ \text{ Notation for elements in S } | \text{ Properties which specifies being in S } \}.$$

  The vertical bar should be read as "such that". For example, if S is the set of all even integer then

  $$S = \{x \in \mathbb{Z} \mid 2 \text{ divides } x\}.$$

  We can also use the curly bracket notation for finite sets without using the | symbol. For example, the set S which contains only 1,2 and 3 can be written as

  $$S = \{1, 2, 3\}.$$

- If every object in S is also an object in T, then we say that S is contained in T. In mathematical notation we write this as $S \subset T$. Note that $S \subset T$ and $T \subset S \Rightarrow S = T$. If S is *not* contained in T we write $S \not\subset T$.
- If $S \subset T$ then $T \setminus S := \{x \in T \mid x \notin S\}$. $T \setminus S$ is called the *compliment* of S in T.
- The set of objects contained in both S and T is call the intersection of S and T. In mathematical notation we denote this by $S \cap T$.
- The collection of all objects which are in either S or T is call the union on S and T. In mathematical notation we denote this by $S \cup T$.
- $S \times T = \{(a, b) | a \in S, b \in T\}$. We call this new set the (cartesian) product of S and T. We may naturally extend this concept to finite collections of sets.

- The set which contains no objects is called the empty set. We denote the empty set by $\emptyset$. We say that S and T are *disjoint* if $S \cap T = \emptyset$. The union of two disjoint sets is often written as $S \coprod T$.

Definition. *A map (or function) f from S to T is a rule which assigns to each element of S a unique elements of T. We express this information using the following notation:*

$$\begin{aligned} f : S &\to T \\ x &\mapsto f(x) \end{aligned}$$

Here are some examples of maps of sets:

1. $S = T = \mathbb{N}$,

$$\begin{aligned} f : \mathbb{N} &\to \mathbb{N} \\ a &\mapsto a^2 \end{aligned}$$

2. $S = \mathbb{Z} \times \mathbb{Z}$, $T = \mathbb{Z}$,

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \\ (a, b) &\mapsto a + b \end{aligned}$$

This very simple looking abstract concept hides enormous depth. To illustrate this, observe that calculus is just the study of certain classes of functions (continuous, differentiable or integrable) from $\mathbb{R}$ to $\mathbb{R}$.

Definition. *Let S and T be two sets,and $f : S \to T$ be a map.*

1. *We say that S is the domain of f and T is the codomain of f.*
2. *We say that f is the identity map if $S = T$ and $f(x) = x$, $\forall x \in S$. In this case we write $f = Id_S$.*
3. *f is injective if $f(x) = f(y) \Rightarrow x = y \; \forall \; x, y \in S$.*
4. *f is surjective if given $y \in T$, there exists $x \in S$ such that $f(x) = y$.*
5. *If f is both injective and surjective we say it is bijective. Intuitively this means f gives a perfect matching of elements in S and T.*

Observe that if R, S and T are sets and $g : R \to S$ and $f : S \to T$ are maps then we may compose them to give a new function: $f \circ g : R \to T$. Note that this is only possible if the domain of f is naturally contained in the codomain of g.

Important Exercise. *Let S and T be two sets. Let f be a map from S to T. Show that f is a bijection if and only if there exists a map g from T to S such that $f \circ g = Id_T$ and $g \circ f = Id_S$.*

1.3 Equivalence Relations

Within a set it is sometimes natural to talk about different elements being related in some way. For example, in $\mathbb{Z}$ we could say that $x, y \in \mathbb{Z}$ are related if $x - y$ is divisible by 2. Said another way, x and y are related is they are both odd or both even. This idea can be formalized as something called an *equivalence relation*.

Definition. *An equivalence relation on a set S is a subset $U \subset S \times S$ satisfying:*

1. $(x, y) \in U \iff (y, x) \in U$. *(This is called the symmetric property.)*
2. $\forall x \in S, (x, x) \in U$. *(This is called the reflexive property.)*
3. *Given $x, y, z \in S$, $(x, y) \in U$ and $(y, z) \in U \Rightarrow (x, z) \in U$. (This is called the transitive property.)*

If $U \subset S \times S$ is an equivalence relation then we say that $x, y \in S$ are *equivalent* if and only if $(x, y) \in U$. In more convenient notation, we write $x \sim y$ to mean that x and y are equivalent.

Definition. *Let $\sim$ be an equivalence relation on the set S. Let $x \in S$. The equivalence class containing x is the subset*

$$[x] := \{y \in S \mid y \sim x\} \subset S.$$

Remarks.
1. *Notice that the reflexive property implies that $x \in [x]$. Hence equivalence classes are non-empty and their union is S.*
2. *The symmetric and transitive properties imply that $y \in [x]$ if and only if $[y] = [x]$. Hence two equivalence classes are equal or disjoint. It should also be noted that we can represent a given equivalence class using any of its members using the $[x]$ notation.*

Definition. *Let S be a set. Let $\{X_i\}$ be a collection of subsets. We say that $\{X_i\}$ forms a partition of S if each X_i is non-empty, they are pairwise disjoint and their union is S.*

We've seen that the equivalence classes of an equivalence relation naturally form a partition of the set. Actually there is a converse: Any partition of a set naturally gives rise to an equivalence relation whose equivalence classes are the members of the partition. The conclusion of all this is that an equivalence relation on a set is the same as a partition. In the example given above, the equivalence classes are the odd integers and the even integers. **Equivalence relations and equivalence classes are incredibly important. They will the foundation of many concepts throughout the course. Take time to really internalize these ideas.**

2 The Structure of + and × on $\mathbb{Z}$

2.1 Basic Observations

We may naturally express + and × in the following set theoretic way:

$$\begin{aligned} + : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \\ (a, b) &\mapsto a + b \end{aligned}$$

$$\begin{aligned} \times : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \\ (a, b) &\mapsto a \times b \end{aligned}$$

Here are 4 elementary properties that + satisfies:

- (Associativity): $a + (b + c) = (a + b) + c \; \forall a, b, c \in \mathbb{Z}$
- (Existence of additive identity) $a + 0 = 0 + a = a \; \forall a \in \mathbb{Z}$.
- (Existence of additive inverses) $a + (-a) = (-a) + a = 0 \; \forall a \in \mathbb{Z}$
- (Commutativity) $a + b = b + a \; \forall a, b \in \mathbb{Z}$.

Here are 3 elementary properties that × satisfy:

- (Associativity): $a \times (b \times c) = (a \times b) \times c \; \forall a, b, c \in \mathbb{Z}$
- (Existence of multiplicative identity) $a \times 1 = 1 \times a = a \; \forall a \in \mathbb{Z}$.
- (Commutativity) $a \times b = b \times a \; \forall a, b \in \mathbb{Z}$.

The operations of + and × interact by the following law:

- (Distributivity) $a \times (b + c) = (a \times b) + (a \times c) \; \forall a, b, c \in \mathbb{Z}$.

From now on we'll simplify the notation for multiplication to $a \times b = ab$.

Remarks

1. Each of these properties is totally obvious but will form the foundations of future definitions: groups and rings.
2. All of the above hold for + and × on $\mathbb{Q}$. In this case there is an extra property that non-zero elements have multiplicative inverses:

   $$\text{Given } a \in \mathbb{Q} \backslash \{0\}, \; \exists b \in \mathbb{Q} \text{ such that } ab = ba = 1.$$

   This extra property will motivate the definition of a field.

3. The significance of the Associativity laws is that summing and multiplying a finite collection of integers makes sense, i.e. is independent of how we do it.

It is an important property of $\mathbb{Z}$ (and $\mathbb{Q}$) that the product of two non-zero elements is again non-zero. More precisiely: $a, b \in \mathbb{Z}$ such that $ab = 0 \Rightarrow$ either $a = 0$ or $b = 0$. Later this property will mean that $\mathbb{Z}$ is something called an *integral domain*. This has the following useful consequence:

Cancellation Law: For $a, b, c \in \mathbb{Z}$, $ca = cb$ and $c \neq 0 \Rightarrow a = b$.

This is proven using the distributive law together with the fact that $\mathbb{Z}$ is an integral domain. I leave it an exercise to the reader.

2.2 Factorization and the Fundamental Theorem of Arithmetic

Definition. *Let $a, b \in \mathbb{Z}$. Then a divides $b \iff \exists c \in \mathbb{Z}$ such that $b = ca$. We denote this by $a|b$ and say that a is a divisor (or factor) of b.*

Observe that 0 is divisible by every integer. The only integers which divide 1 are 1 and -1. Any way of expressing an integer as the product of a finite collection of integers is called a *factorization*.

Definition. *A prime number p is an integer greater than 1 whose only positive divisors are p and 1. A positive integer which is not prime is called composite.*

Remark. *$\mathbb{Z}$ is generated by 1 under addition. By this I mean that every integer can be attained by successively adding 1 (or -1) to itself. Under multiplication the situation is much more complicated. There is clearly no single generator of $\mathbb{Z}$ under multiplication in the above sense.*

Definition. *Let $a, b \in \mathbb{Z}$. The highest common factor of a and b, denoted $HCF(a, b)$, is the largest positive integer which is a common factor of a and b. Two non-zero integers $a, b \in \mathbb{Z}$ are said to be coprime if $HCF(a, b) = 1$.*

Here are some important elementary properties of divisibility dating back to Euclid (300BC), which I'll state without proof. We'll actually prove them later in far more generality.

Remainder Theorem. *Given $a, b \in \mathbb{Z}$, if $b > 0$ then $\exists!\ q, r \in \mathbb{Z}$ such that $a = bq + r$ with $0 \leq r < b$.*

Theorem. *Given $a, b \in \mathbb{Z}$, $\exists u, v \in \mathbb{Z}$ such that $au + bv = HCF(a, b)$. In particular, a and b are coprime if an only if there exist $u, v \in \mathbb{Z}$ such that $au + bv = 1$.*

Euclid's Lemma. *Let p be a prime number and $a, b \in \mathbb{Z}$. Then*

$$p|ab \Rightarrow p|a \text{ or } p|b$$

The Fundamental Theorem of Arithmetic. *Every positive integer, a, greater than* 1 *can be written as a product of primes:*

$$a = p_1 p_2 ... p_r.$$

Such a factorization is unique up to ordering.

Proof. If there is a positive integer not expressible as a product of primes, let $c \in \mathbb{N}$ be the least such element. The integer c is not 1 or a prime, hence $c = c_1 c_2$ where $c_1, c_c \in \mathbb{N}$, $c_1 < c$ and $c_2 < c$. By our choice of c we know that both c_1 and c_2 are the product of primes. Hence c much be expressible as the product of primes. This is a contradiction. Hence all positive integers can be written as the product of primes.

We must prove the uniqueness (up to ordering) of any such decomposition. Let

$$a = p_1 p_2 ... p_r = q_1 q_2 ... q_s$$

be two factorizations of a into a product of primes. Then $p_1 | q_1 q_2 ... q_s$. By Euclid's Lemma we know that $p_1 | q_i$ for some i. After renumbering we may assume $i = 1$. However q_1 is a prime, so $p_1 = q_1$. Applying the cancellation law we obtain

$$p_2 ... p_r = q_2 ... q_s.$$

Assume that $r < s$. We can continue this process until we have:

$$1 = q_{r+1} .. q_s.$$

This is a contradiction as 1 is not divisible by any prime. Hence $r = s$ and after renumbering $p_i = q_i \ \forall i$. □

Using this we can prove the following beautiful fact:

Theorem. *There are infinitely many distinct prime numbers.*

Proof. Suppose that there are finitely many distinct primes $p_1, p_2 ..... p_r$. Consider $c = p_1 p_2 ... p_r + 1$. Clearly $c > 1$. By the Fundamental Theorem of Arithmetic, c is divisible by at least one prime, say p_1. Then $c = p_1 d$ for some $d \in \mathbb{Z}$. Hence we have

$$p_1(d - p_2 ... p_r) = c - p_1 p_2 .. p_r = 1.$$

This is a contradiction as no prime divides 1. Hence there are infinitely many distinct primes. □

The Fundamental Theorem of Arithmetic also tells us that every positive element $a \in \mathbb{Q}$ can be written uniquely (up to reordering) in the form:

$$a = p_1^{\alpha_1} \cdots p_n^{\alpha_n}; \ p_i \text{ prime and } \alpha_i \in \mathbb{Z}$$

The Fundamental Theorem also tells us that two positive integers are coprime if and only if they have no common prime divisor. This immediately shows that every positive element $a \in \mathbb{Q}$ can be written uniquely in the form:

$$a = \frac{\alpha}{\beta}, \alpha, \beta \in \mathbb{N} \text{ and coprime.}$$

We have seen that both $\mathbb{Z}$ and $\mathbb{Q}$ are examples of sets with two concepts of composition ($+$ and $\times$) which satisfy a collection of abstract conditions. We have also seen that the structure of $\mathbb{Z}$ together with $\times$ is very rich. Can we think of other examples of sets with a concept of $+$ and $\times$ which satisfy the same elementary properties?

2.3 Congruences

Fix $m \in \mathbb{N}$. By the remainder theorem, if $a \in \mathbb{Z}, \exists\ !\ \ q, r \in \mathbb{Z}$ such that $a = qm + r$ and $0 \leq r < m$. We call r the *remainder* of a modulo m. This gives the natural equivalence relation on $\mathbb{Z}$:

$$a \sim b \iff a \text{ and } b \text{ have the same remainder modulo } m \iff m|(a-b)$$

Important Exercise. *Check this really is an equivalence relation!*

Definition. $a, b \in \mathbb{Z}$ *are* ***congruent modulo m*** $\iff m|(a-b)$. *This can also be written:*

$$a \equiv b \mod m.$$

Remarks. *1. The equivalence classes of $\mathbb{Z}$ under this relation are indexed by the possible remainder modulo m. Hence, there are m distinct equivalence classes which we call* ***residue classes****. We denote the set of all residue classes $\mathbb{Z}/m\mathbb{Z}$.*

2. There is a natural surjective map

$$\begin{aligned} [\,] \ : \ & \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \\ & a \mapsto [a] \end{aligned} \tag{1}$$

Note that this is clearly not injective as many integers have the same remainder modulo m. Also observe that $\mathbb{Z}/m\mathbb{Z} = \{[0], [1], ...[m-1]\}$.

The following result allows us to define $+$ and $\times$ on $\mathbb{Z}/m\mathbb{Z}$.

Propostion. *Let $m \in \mathbb{N}$. Then,$\forall a, b, a', b' \in \mathbb{Z}$:*

$$[a] = [a'] \text{ and } [b] = [b'] \Rightarrow [a+b] = [a'+b'] \text{ and } [ab] = [a'b'].$$

Proof. This is a very good exercise. □

Definition. *We **define** addition and multiplication on* $\mathbb{Z}/m\mathbb{Z}$ *by*

$$[a] \times [b] = [a \times b] \; \forall a, b \in \mathbb{Z} \qquad [a] + [b] = [a + b] \; \forall a, b \in \mathbb{Z}$$

Remark. *Note that there is ambiguity in the definition, because it seems to depend on making a choice of representative of each residue class. The proposition shows us that the resulting residue classes are independent of this choice, hence* $+$ *and* $\times$ *are well defined on* $\mathbb{Z}/m\mathbb{Z}$*.*

Our construction of $+$ and $\times$ on $\mathbb{Z}/m\mathbb{Z}$ is lifted from $\mathbb{Z}$, hence they satisfy the eight elementary properites that $+$ and $\times$ satisfied on $\mathbb{Z}$. In particular $[0] \in \mathbb{Z}/m\mathbb{Z}$ behaves like $0 \in \mathbb{Z}$:

$$[0] + [a] = [a] + [0] = [a], \; \forall [a] \in \mathbb{Z}/m\mathbb{Z};$$

and $[1] \in \mathbb{Z}/m\mathbb{Z}$ behaves like $1 \in \mathbb{Z}$:

$$[1] \times [a] = [a] \times [1] = [a], \; \forall [a] \in \mathbb{Z}/m\mathbb{Z}.$$

We say that $[a] \in \mathbb{Z}/m\mathbb{Z}$ is non-zero if $[a] \neq [0]$. Even though $+$ and $\times$ on $\mathbb{Z}/m\mathbb{Z}$ share the same elementary properties with $+$ and $\times$ on $\mathbb{Z}$, they behave quite differently in this case. As an example, notice that

$$[1] + [1] + [1] + \cdots + [1](m \text{ times}) = [m] = [0]$$

Hence we can add 1 (in $\mathbb{Z}/m\mathbb{Z}$) to itself and eventually get 0 (in $\mathbb{Z}/m\mathbb{Z}$).

Also observe that if m is composite with $m = rs$, where $r < m$ and $s < m$ then $[r]$ and $[s]$ are both non-zero ($\neq [0]$) in $\mathbb{Z}/m\mathbb{Z}$, but $[r] \times [s] = [rs] = [m] = [0] \in \mathbb{Z}/m\mathbb{Z}$. Hence we can have two non-zero elements *multiplying* together to give *zero.*

Proposition. *For every* $m \in \mathbb{N}, a \in \mathbb{Z}$ *the congruence*

$$ax \equiv 1 \mod m$$

has a solution (in $\mathbb{Z}$*) iff* a *and* m *are coprime.*

Proof. This is just a restatement of the fact that a and m coprime $\iff$ $\exists u, v \in \mathbb{Z}$ such that $au + mv = 1$. □

Observe that the congruence above can be rewritten as $[a] \times [x] = [1]$ in $\mathbb{Z}/m\mathbb{Z}$. We say that $[a] \in \mathbb{Z}/m\mathbb{Z}$ has a multiplicative inverse if $\exists [x] \in \mathbb{Z}/m\mathbb{Z}$ such that $[a] \times [x] = [1]$. Hence we deduce that the only elements of $\mathbb{Z}/m\mathbb{Z}$ with muliplicative inverse are those given by $[a]$, where a is coprime to m.

Recall that $\times$ on $\mathbb{Q}$ had the extra property that all non-zero elements had *multiplicative inverses.* When does this happen in $\mathbb{Z}/m\mathbb{Z}$?. By the above we see that this can happen $\iff$ $\{1, 2, \cdots, m-1\}$ are all coprime to m. This can only happen if m is prime. We have thus proven the following:

Corollary. *All non-zero elements of* $\mathbb{Z}/m\mathbb{Z}$ *have a multiplicative inverse* $\iff$ m *is prime.*

Later this will be restated as $\mathbb{Z}/m\mathbb{Z}$ is a *field* $\iff$ m is a prime. These are examples of things called *finite fields*.

Important Exercise. *Show that if m is prime then the product of two non-zero elements of $\mathbb{Z}/m\mathbb{Z}$ is again non-zero.*

Key Observation: There are naturally occuring sets (other than $\mathbb{Z}$ and $\mathbb{Q}$) which come equipped with a concept of $+$ and $\times$, whose most basic properties are the same as those of the usual addition and multiplication on $\mathbb{Z}$ or $\mathbb{Q}$. **Don't be fooled into thinking all other examples will come from numbers. As we'll see, there are many examples which are much more exotic**.

3 Groups

3.1 Basic Definitions

Definition. *Let G be a set. A **binary operation** is a map of sets:*

$$* : G \times G \to G.$$

For ease of notation we write $*(a, b) = a * b \ \forall a, b \in G$. Any binary operation on G gives a way of *combining* elements. As we have seen, if $G = \mathbb{Z}$ then $+$ and $\times$ are natural examples of binary operations. When we are talking about a set G, together with a fixed binary operation $*$, we often write $(G, *)$.

Fundamental Definition. *A **group** is a set G, together with a binary operation $*$, such that the following hold:*

1. *(Associativity): $(a * b) * c = a * (b * c) \ \forall a, b, c \in G$.*
2. *(Existence of identity): $\exists e \in G$ such that $a * e = e * a = a \ \forall a \in G$.*
3. *(Existence of inverses): Given $a \in G, \exists b \in G$ such that $a * b = b * a = e$.*

Remarks.

1. *We have seen five different examples thus far: $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{Q}\backslash\{0\}, \times)$, $(\mathbb{Z}/m\mathbb{Z}, +)$, and $(\mathbb{Z}/m\mathbb{Z} \setminus \{[0]\}, \times)$ if m is prime. Another example is that of a real vector space under addition. Note that $(\mathbb{Z}, \times)$ is **not** a group. Also note that this gives examples of groups which are both finite and infinite. The more mathematics you learn the more you'll see that groups are everywhere.*
2. *A set with a single element admits one possible binary operation. This makes it a group. We call this the trivial group.*
3. *A set with a binary operation is called a **monoid** if only the first two properties hold. From this point of view, a group is a monoid in which every element is invertible. $(\mathbb{Z}, \times)$ is a monoid but not a group.*

4. *Observe that in all of the examples given the binary operation is commutative, i.e.* $a * b = b * a \ \forall a, b \in G$. *We do not include this in our definition as this would be too restrictive. For example the set of invertible* $n \times n$ *matrices with real entries, denoted* $GL_n(\mathbb{R})$*, forms a group under matrix multiplication. However we know that matrix multiplication does not commute in general.*

Definition. *A group* $(G, *)$ *is called* ***Abelian*** *if it also satisfies*

$$a * b = b * a \ \forall a, b \in G.$$

This is also called the commutative property.

The fundamental Abelian group is $(\mathbb{Z}, +)$. Notice also that any vector space is an Abelian group under it's natural addition.

So a group is a set with extra structure. In set theory we have the natural concept of a map between sets (a function). The following is the analogous concept for groups:

Fundamental Definition. *Let* $(G, *)$ *and* $(H, \circ)$ *be two groups. A* ***homomorphism*** f*, from* G *to* H*, is a map of sets* $f : G \to H$*, such that* $f(x * y) = f(x) \circ f(y) \ \forall x, y \in G$*. If* $G = H$ *and* $f = Id_G$ *we call* f *the identity homomorphism.*

Remarks.

1. *Intuitively one should thing about a homomorphism as a map of sets which preserves the underlying group structure. It's the same idea as a linear map between vector spaces.*

2. *A homomorphism* $f : G \to H$ *which is bijective is called an* ***isomorphism****. Two groups are said to be* ***isomorphic*** *if there exists an isomorphism between them. Intuitively two groups being isomorphic means that they are the "same" group with relabelled elements.*

3. *A homomorphism from a group to itself (i.e.* $f : G \to G$*) is called an* ***endomorphism****. An endomorphism which is also an isomorphism is called an* ***automorphism****.*

Proposition. *Let* $(G, *), (H, \circ)$ *and* $(M, \square)$ *be three groups. Let* $f : G \to H$ *and* $g : H \to M$ *be homomorphism. Then the composition* $gf : G \to M$ *is a homomorphism.*

Proof. Let $x, y \in G$. $gf(x * y) = g(f(x) \circ f(y)) = gf(x) \square gf(y)$. □

Remark. *Composition of homomorphism gives the collection of endomorphisms of a group the structure of a monoid. The subset of automorphisms has the stucture of a group under composition. We denote it by* $Aut(G)$*. This is analogues to the collection of* $n \times n$ *invertible matrices being a group under matrix multiplication.*

Proposition. *Let* $(G, *)$ *be a group. The identity element is unique.*

Proof. Assume $e, e' \in G$ both behave like the identity. Then $e = e * e' = e'$. □

Proposition. *Let $(G, *)$ be a group. For $a \in G$ there is only one element which behaves like the inverse of a.*

Proof. Assume $a \in G$ has 2 inverses, $b, c \in G$. Then:

$$\begin{aligned} (a * b) &= e \\ c * (a * b) &= c * e \\ (c * a) * b &= c \quad \text{(associativity and identity)} \\ e * b &= c \\ b &= c \end{aligned}$$

□

The first proposition tells us that we can write $e \in G$ for the identity and it is well-defined. Similarly the second proposition tells us that for $a \in G$ we can write $a^{-1} \in G$ for the inverse in a well-defined way. The proof of the second result gives a good example of how we prove results for abstract groups. We can only use the axioms, nothing else.

Given $r \in \mathbb{Z}$ and $a \in G$, we write

$$a^r = \begin{cases} a * a * \cdots * a \ \ (r \text{ times}), & \text{if } r > 0 \\ e, & \text{if } r = 0 \\ a^{-1} * a^{-1} * \cdots * a^{-1} \ \ (-r \text{ times}), & \text{if } r < 0 \end{cases}$$

Cancellation Law for Groups. *Let $a, b, c \in G$ a group. Then*

$$a * c = a * b \Rightarrow c = b \text{ and } c * a = b * a \Rightarrow c = b$$

Proof. Compose on left or right by $a^{-1} \in G$, then apply the associativity and inverses and identity axioms. □

Proposition. *Let $(G, *)$ and $(H, \circ)$ be two groups and $f : G \to H$ a homomorphism. Let $e_G \in G$ and $e_H \in H$ be the respective identities. Then*

- $f(e_G) = e_H$.
- $f(x^{-1}) = (f(x))^{-1}, \forall x \in G$

Proof.

- $f(e_G) \circ e_H = f(e_G) = f(e_G * e_G) = f(e_G) \circ f(e_G)$. By the cancellation law we deduce that $f(e_G) = e_H$.
- Let $x \in G$. Then $e_H = f(e_G) = f(x * x^{-1}) = f(x) \circ f(x^{-1})$ and $e_H = f(e_G) = f(x^{-1} * x) = f(x^{-1}) \circ f(x)$. Hence $f(x^{-1}) = (f(x))^{-1}$.

□

3.2 Subgroups, Cosets and Lagrange's Theorem

In linear algebra, we can talk about subspaces of vector spaces. We have an analogous concept in group theory.

Definition. *Let $(G, *)$ be a group. A **subgroup** of G is a subset $H \subset G$ such that*

1. $e \in H$
2. $x, y \in H \Rightarrow x * y \in H$
3. $x \in H \Rightarrow x^{-1} \in H$

Remarks. 1. *A subgroup is naturally a group under the induced binary operation. It clearly has the same identity element.*

2. *If $m \in \mathbb{N}$, then the subset $m\mathbb{Z} := \{ma | a \in \mathbb{Z}\}$is a subgroup of $(\mathbb{Z}, +)$.*

3. *If V is a vector space over $\mathbb{R}$ then it is naturally an Abelian group under addition. If W is a subspace then it is also under a subgroup under addition.*

Proposition. *$H, K \subset G$ subgroups $\Rightarrow H \cap K \subset G$ is a subgroup.*

Proof. 1. As H, K subgroups, $e \in H$ and $e \in K \Rightarrow e \in H \cap K$.

2. $x, y \in H \cap K \Rightarrow x * y \in H$ and $x * y \in K \Rightarrow x * y \in H \cap K$.

3. $x \in H \cap K \Rightarrow x^{-1} \in H$ and $x^{-1} \in K \Rightarrow x^{-1} \in H \cap K$.

□

This result clearly extends to any collection of subgroups of G.

Let $(G, *)$ be a group and let $H \subset G$ be a subgroup. Let us define a relation on G using H as follows:

$$\text{Given } x, y \in G,\ x \sim y \iff x^{-1} * y \in H$$

Proposition. *This gives an equivalence relation on G.*

Proof. We need to check the three properties of an equivalence relation:

1. (Reflexive)$e \in H \Rightarrow x^{-1} * x \in H\ \forall x \in G \Rightarrow x \sim x$
2. (Symmetric) $x \sim y \Rightarrow x^{-1} * y \in H \Rightarrow (x^{-1} * y)^{-1} \in H \Rightarrow y^{-1} * x \in H \Rightarrow y \sim x$
3. (Transitive) $x \sim y, y \sim z \Rightarrow x^{-1} * y, y^{-1} * z \in H \Rightarrow (x^{-1} * y) * (y^{-1} * z) \in H \Rightarrow x^{-1} * z \in H \Rightarrow x \sim z$

□

Definition. *We call the equivalence classes of the above equivalence relation* ***left cosets*** *of H in G.*

Proposition. *For $x \in G$ the equivalence class (or left coset) containing x equals*

$$xH := \{x * h | h \in H\} \subset G$$

Proof. The easiest way to show that two subsets of G are equal is to prove containment in both directions.

$x \sim y \iff x^{-1} * y \in H \iff x^{-1} * y = h$ for some $h \in H \Rightarrow y = x * h \in xH$. Therefore $\{$Equivalence class containing $x\} \subset xH$.

$y \in xH \Rightarrow y = x * h$ for some $h \in H \Rightarrow x^{-1} * y \in H \Rightarrow y \sim x$. Therefore $xH \subset \{$Equivalence class containing $x\}$. □

This has the following very important consequence:

Corollary. *Hence for $x, y \in G$, $xH = yH \iff x^{-1} * y \in H$.*

Proof. By the above proposition we know that $xH = yH \iff x \sim y \iff x^{-1} * y \in H$. □

It is very important you understand and remember this fact. An immediate consequence is that $y \in xH \Rightarrow yH = xH$. Hence left cosets can in general be written with different representatives at the front. This is very important. Also observe that the equivalence class containing $e \in G$ is just H. Hence the only equivalence class which is a subgroup H, as no other contains the identity. If $H = \{e\}$ then the left cosets are singleton sets.

Remarks. *Let $G = \mathbb{R}^3$, thought of as a group under addition. Let H is a two dimensional subspace. Recall this is a subgroup under addition. Geometrically H is a plane which contains the origin. Geometrically the left cosets of H in $\mathbb{R}^3$ are the planes which are parallel to H.*

Definition. *Let $(G, *)$ be a group and $H \subset G$ a subgroup. We denote by G/H the set of left cosets of H in G. If the size of this set is finite then we say that H has* ***finite index*** *in G. In this case we write*

$$(G : H) = |G/H|,$$

and call it the index of H in G.

For $m \in \mathbb{N}$, the subgroup $m\mathbb{Z} \subset Z$ has index m. Note that $\mathbb{Z}/m\mathbb{Z}$ is naturally the set of residue classes modulo m previously introduced. The vector space example in the above remark is not finite index as there are infinitely many parallel planes in $\mathbb{R}^3$

Proposition. *Let $x \in G$. The map (of sets)*

$$\begin{aligned} \phi : H &\longrightarrow xH \\ h &\longrightarrow x * h \end{aligned}$$

is a bijection.

Proof. We need to check that ϕ is both injective and surjective. For injectivity observe that for $g, h \in H$, $\phi(h) = \phi(g) \Rightarrow x * h = x * g \Rightarrow h = g$. Hence ϕ is injective. For surjectivity observe that $g \in xH \Rightarrow \exists h \in H$ such that $g = x * h \Rightarrow g = \phi(h)$. □

Now let's restrict to the case where G is a finite group.

Proposition. *Let $(G, *)$ be a finite group and $H \subset G$ a subgroup. Then $\forall x \in G$, $|xH| = |H|$.*

Proof. We know that there is a bijection between H and xH. Both must be finite because they are contained in a finite set. A bijection exists between two finite sets if and only if they have the same cardinality. □

Lagrange's Theorem. *Let $(G, *)$ be a finite group and $H \subset G$ a subgroup. Then $|H|$ divides $|G|$.*

Proof. We can use H to define the above equivalence relation on G. Because it is an equivalence relation, its equivalence classes cover G and are all disjoint. Recall that this is called a partition of G.
We know that each equivalence class is of the form xH for some (clearly non-unique in general) $x \in G$. We know that any left coset of H has size equal to $|H|$. Hence we have partitioned G into subsets each of size $|H|$. We conclude that $|H|$ divides $|G|$. □

This is a powerful result. It tightly controls the behavior of subgroups of a finite group. For example:

Corollary. *Let $p \in \mathbb{N}$ be a prime number. Let $(G, *)$ be a finite group of order p. Then the only subgroups of G are G and $\{e\}$.*

Proof. Let H be a subgroup of G. By Lagrange $|H|$ divides p. But p is prime so either $|H| = 1$ or $|H| = p$. In the first case $H = \{e\}$. In the second case $H = G$. □

3.3 Finitely Generated Groups

Definition. *Let G be a group and $X \subset G$ be a subset. We define the **subgroup generated by X** to be the intersection of all subgroups of G containing X. We denote it by $gp(X) \subset G$.*

Remarks.

1. *$gp(X)$ is the minimal subgroup containing X. By minimal we mean that if $H \subset G$ is a subgroup such that $X \subset H$ then $gp(X) \subset H$.*

2. *A more constructive way of defining $gp(X)$ is as all possible finite compositions of elements of X and their inverses. I leave it as an exercise to check that this subset is indeed a subgroup.*

3. *Let us consider the group $(\mathbb{Z}, +)$ and $X = \{1\} \subset \mathbb{Z}$. Then $gp(X)$=$\mathbb{Z}$. This is the precise sense in which $\mathbb{Z}$ is "generated" by 1 under addition.*

Definition. *We say a group $(G, *)$ is **finitely generated** if $\exists X \subset G$ finite such that $gp(X)$=G.*

Remarks. *1. Clearly all finite groups are finitely generated.*

*2. The fact that there are infinitely many primes implies that $(\mathbb{Q}\backslash\{0\}, \times)$ is **not** finitely generated.*

Definition. *A group $(G, *)$ is said to be cyclic if $\exists x \in G$ such that $gp(\{x\}) = G$, i.e. G can be generated by a single element. In concrete terms this means that $G = \{x^n | n \in \mathbb{Z}\}$.*

By the above observations $(\mathbb{Z}, +)$ and $(\mathbb{Z}/m\mathbb{Z}, +)$ are examples.

Proposition. *Any group of prime order is cyclic.*

Proof. Let G be a group of prime order p. Let x be a non-identity element of G. Then $gp(\{x\}) \subset G$ is non-trivial and by Lagrange's theorem must have order p. Hence $G = gp(\{x\})$. □

Remarks. *It is important to understand that not all groups are cyclic. We'll see many examples throughout the course.*

Let G be a group (not necessarily cyclic). For $r, s \in \mathbb{Z}$ and $x \in G$, $x^r x^s = x^{r+s} = x^{s+r} = x^s x^r$. Hence $gp(\{x\}) \subset G$ is Abelian. We deduce that all cyclic groups are Abelian.

Theorem. *Let G be a cyclic group. Then*

1. If G is infinite, $G \cong (\mathbb{Z}, +)$

2. If $|G| = m \in \mathbb{N}$, then $G \cong (\mathbb{Z}/m\mathbb{Z}, +)$

Proof. We have two cases to consider.

1. If $G = gp(\{x\})$, then $G = \{\cdots x^{-2}, x^{-1}, e, x, x^2 \cdots\}$. Assume all elements in this set are distinct, then we can define a map of sets:

$$\begin{aligned} \phi : G &\rightarrow \mathbb{Z} \\ x^n &\rightarrow n \end{aligned}$$

   Then,$\forall a, b \in \mathbb{Z}, \phi(x^a * x^b) = \phi(x^{a+b}) = a + b = \varphi(x^a) + \varphi(x^b)$ so φ is a homomorphism which by assumption was bijective. Thus, $(G, *)$ is isomorphic to $(\mathbb{Z}, +)$.

2. Now assume $\exists a, b \in \mathbb{Z}, b > a$ such that $x^a = x^b$. Then $x^{(b-a)} = e \Rightarrow x^{-1} = x^{(b-a-1)} \Rightarrow G = \{e, \cdots, x^{b-a-1}\}$. In particular G is finite. Choose minimal $m \in \mathbb{N}$ such that $x^m = e$. Then $G = \{e, x, \cdots, x^{m-1}\}$ and all its elements are distinct by minimality of m. Hence $|G| = m$.

Define the map

$$\begin{aligned} \phi : G &\rightarrow \mathbb{Z}/m\mathbb{Z} \\ x^n &\rightarrow [n] \text{ for } n \in \{0, ...m-1\} \end{aligned}$$

This is clearly a surjection, hence a bijection because $|G| = |\mathbb{Z}/m\mathbb{Z}| = m$. Again $\forall a, b \in \{0, ..., m-1\}$ we know $\phi(x^a * x^b) = \phi(x^{a+b}) = [a+b] = [a] + [b] = \varphi(x^a) + \varphi(x^b)$ is a homomorphism. Hence $(G, *)$ is isomorphic to $(\mathbb{Z}/m\mathbb{Z}, +)$.

□

Hence two finite cyclic groups of the same size are isomorphic. What are the possible subgroups of a cyclic group?

Proposition. *A subgroup of a cyclic group is cyclic.*

Proof. If H is trivial we are done. Hence assume that H is non-trivial. By the above we need to check two cases.

1. $(G, *) \cong (\mathbb{Z}, +)$. Let $H \subset \mathbb{Z}$ be a non-trivial subgroup. Choose $m \in \mathbb{N}$ minimal such that $m \in H (m \neq 0)$. Hence $m\mathbb{Z} = \{ma | a \in \mathbb{Z}\} \subseteq H$. Assume $\exists n \in H$ such that $n \notin m\mathbb{Z}$. By the remainder theorem, $n = qm + r, r, q \in \mathbb{Z}$ and $0 < r < m \Rightarrow r \in H$. This is a contradiction by the minimality of m. Therefore $m\mathbb{Z} = H$. Observe that $gp(\{m\}) = m\mathbb{Z} \subset \mathbb{Z}$. Hence H is cylic.

2. $(G, *) \cong (\mathbb{Z}/m\mathbb{Z}, +)$. Let $H \subset \mathbb{Z}/m\mathbb{Z}$ be a non-trivial subgroup. Again, choose $n \in \mathbb{N}$ minimal and positive such that $[n] \in H$. The same argument as above shows that the containment $gp(\{[n]\}) \subseteq H$ is actually equality. Hence H is cyclic.

□

Proposition. *Let $(G, *)$ be a finite cyclic group of order d. Let $m \in \mathbb{N}$ such that m divides $|G|$. Then there is a unique cyclic subgroup of order m.*

Proof. Because $|G| = d$ we know that $G \cong (\mathbb{Z}/d\mathbb{Z}, +)$. Hence we need only answer the question for this latter group. Let m be a divisor of d. Then if $n = d/m$ then $gp(\{[n]\}) \subset \mathbb{Z}/d\mathbb{Z}$ is cyclic of order m by construction. If $H \subset \mathbb{Z}/d\mathbb{Z}$ is a second subgroup of order m then by the above proof we know that the minimal $n \in \mathbb{N}$ such that $[n] \in H$ must be $n = d/m$. Hence $H = gp(\{[n]\})$.

□

Let $(G, *)$ be a group (not necessarily cyclic) and $x \in G$. We call $gp(\{x\}) \subset G$ the **subgroup generated by x**. By definition it is cyclic.

Definition. *If $|gp(\{x\})| < \infty$ we say that x is of finite order and its order, written $ord(x)$ equals $|gp(\{x\})|$. If not we say that x is of infinite order.*

Remarks. *1. Observe that by the above we know that if $x \in G$ is of finite order, then*

$$ord(x) = \text{minimal } m \in \mathbb{N} \text{ such that } x^m = e$$

2. $e \in G$ is the only element of G of order 1.

3. The only element with finite order in $\mathbb{Z}$ is 0.

Proposition. *Let $(G, *)$ be a finite group and $x \in G$. Then $ord(x)$ divides $|G|$ and $x^{|G|} = e$.*

Proof. By definition $ord(x) = |gp(\{x\})|$. Therefore, by Lagrange's theorem, $ord(x)$ must divide $|G|$. Also note that by definition $x^{ord(x)} = e$. Hence

$$x^{|G|} = x^{(ord(x) \times \frac{|G|}{ord(x)})} = e^{\frac{|G|}{ord(x)}} = e.$$

□

3.4 Permutation Groups and Group Actions

Definition. *Let S be a set. We define **the group of permutations of S** to be the set of **bijections** from S to itself, denoted $\Sigma(S)$, where the group binary operation is **composition of functions**.*

Remarks. *1. By composition of functions we always mean on the left, i.e. $\forall f, g \in \Sigma(S)$ and $s \in S$ $(f * g)(s) = f(g(s))$.*

2. Associativity clearly has to hold. The identity element e of this group is the identity function on S, i.e. $\forall x \in S$, $e(s) = s$. Inverses exist because any bijective map from a set to itself has an inverse map.

3. Let $n \in \mathbb{N}$. We write $Sym_n := \Sigma(\{1, 2, ..., n\})$. If S is any set of cardinality n then $\Sigma(S)$ is isomorphic to Sym_n, the isomorphism being induced by writing a bijection from S to $\{1, 2, ..., n\}$. We call these groups the finite symmetric groups.

4. Observe that given $\sigma \in \Sigma(S)$ we can think about σ as "moving" S around. In this sense the group $\Sigma(S)$ naturally "acts" on S. Let's make this precise.

Definition. *Let $(G, *)$ be a group and S a set. By a group action of $(G, *)$ on S we mean a map:*

$$\mu : G \times S \to S$$

such that

1. $\forall x, y \in G, s \in S, \mu(x * y, s) = \mu(x, \mu(y, s))$

2. $\mu(e, s) = s$

If the action of the group is understood we will write $x(s) = \mu(x, s) \forall x \in G, s \in S$. This notation makes the axioms clearer: (1) becomes $(x * y)(s) = x(y(s)) \ \forall x, y \in G, s \in S$ and (2) becomes $e(s) = s \ \forall s \in S$.

Remarks. *1. Notice that there is a natural action of $\Sigma(S)$ on S:*

$$\begin{aligned} \mu : \Sigma(S) \times S &\rightarrow S \\ (f, s) &\rightarrow f(s) \end{aligned}$$

This is where the definition comes from.

2. Let $(G, *)$ be a group. There is a natural action of G on itself:

$$\begin{aligned} \mu : G \times G &\rightarrow G \\ (x, y) &\rightarrow x * y \end{aligned}$$

*Property (1) holds as $*$ is associative. Property (2) holds because $e * x = x \ \forall x \in G$. This is called the **left regular representation** of G.*

3. We define the trivial action of G on S by

$$\begin{aligned} \mu : G \times S &\rightarrow S \\ (g, s) &\rightarrow s \ \ \forall s \in S, \ g \in G \end{aligned}$$

4. There is another natural action of G on itself:

$$\begin{aligned} \mu : G \times G &\rightarrow G \\ (x, y) &\rightarrow x^{-1} * y * x \end{aligned}$$

Property (1) holds because of associativity of $*$ and that $(g * h)^{-1} = h^{-1} * g^{-1}$. Property (2) is obvious. This action is called conjugation.

Let $\mu : G \times S \rightarrow S$ an action of a group G on a set S. Any $g \in G$ naturally gives rise to a map:

$$\begin{aligned} \varphi_g : S &\rightarrow S \\ s &\mapsto g(s) \end{aligned}$$

Observe that property (1) of μ being an action implies that $\varphi_{g*h} = \varphi_g \varphi_h \ \forall g, h \in G$. Here the second term is composition of functions. Similarly property (2) tell is that $\varphi_e = Id_S$..

Proposition. *φ_g is a bijection.*

Proof. Given φ_g, if we can find an inverse function, then we will have shown bijectivity. By the above two observations it is clear that $\varphi_{g^{-1}}$ is inverse to φ_g.

□

Hence μ gives rise to a map of sets:

$$\begin{aligned} \varphi : G &\rightarrow \Sigma(S) \\ g &\rightarrow \varphi_g \end{aligned}$$

Proposition. *φ is a homomorphism.*

Proof. As we have just seen, property (1) of μ being an action $\Rightarrow \varphi_h \circ \varphi_g = \varphi_{h*g} \forall h, g \in G$. This is precisely the statement that φ is a homomorphism. □

So an action of a group G on a set S gives a homorphism $\varphi : G \longrightarrow \Sigma(S)$. It is in fact true that any such homorphism comes from a unique group action. Hence an action of G on S is the same thing as homomorphism from G to the permutation group of S. Both concepts are interchangeable.

Definition. *An action of G on S is called* ***faithful*** *if*

$$\begin{aligned} \varphi : G &\rightarrow \Sigma(S) \\ g &\mapsto \varphi_g \end{aligned}$$

is injective.

Notice that if G and H are two groups and $f : G \rightarrow H$ is an injective homomorphism then we may view G as a subgroup of H by identifying it with its image in H under f. Hence if G acts faithfully on S then G is isomorphic to a subgroup of $\Sigma(S)$.

Cayley's Theorem. *Let G be a group. Then G is isomorphic to a subgroup of $\Sigma(G)$. In particular if $|G| = n \in \mathbb{N}$, then G is isomorphic to a subgroup of Sym_n.*

Proof. The result will follow if we can show that the left regular representation is faithful. Let $\varphi : G \rightarrow \Sigma(G)$ be the homomorphism given by the left regular representation. Hence for $g, s \in G$, $\varphi_g(s) = g * s$. For$h, g \in G$, suppose $\varphi_h = \varphi_g$. Then $h * s = g * s \ \forall s \in G \Rightarrow h = g$. Hence φ is injective.

□

3.5 The Orbit-Stabiliser Theorem and Sylow's Theorem

Definition. *Let $(G, *)$ be a group, together with an action φ on a set S. We can define an equivalence relation on S by*

$$s \sim t \iff \exists g \in G \text{ such that } g(s) = t$$

Remarks. *This is an equivalence relation as a consequence of the group axioms, together with the definition of an action. I leave it as an exercise to check this.*

Definition. *Let $(G, *)$ be a group, together with an action φ on a set S. Under the above equivalence relation we call the equivalence classes orbits, and we write*

$$Orb(s) := \{t \in S | \exists g \in G \text{ such that } g(s) = t\} \subset S$$

*for the equivalence class containing $s \in S$. We call it the **orbit** of s.*

It is important to observe that $Orb(s)$ is a subset of S and hence is merely a set with no extra structure.

Definition. *Let $(G, *)$ be a group, together with an action φ on a set S. We say that G acts **transitively** on S is there is only one orbit. Equivalently, φ is transitive if given $s, t \in S$, $\exists g \in G$ such that $g(s) = t$.*

An example of a transitive action is the natural action of $\Sigma(S)$ on S. This is clear because given any two points in a set S there is always a bijection which maps one to the other. If G is not the trivial group (the group with one element) then conjugation is never transitive. To see this observe that under this action $Orb(e) = \{e\}$.

Definition. *Let $(G, *)$ be a group, together with an action φ on a set S. Let $s \in S$. We define the stabiliser subgroup of s to be all elements of G which **fix** s under the action. More precisely*

$$Stab(s) = \{g \in G | g(s) = s\} \subset G$$

For this definition to make sense we must prove that $Stab(s)$ is genuinely a subgroup.

Proposition. *$Stab(s)$ is a subgroup of G.*

Proof.
1. $e(s) = s \Rightarrow e \in Stab(s)$
2. $x, y \in Stab(s) \Rightarrow (x * y)(s) = x(y(s)) = x(s) = s \Rightarrow x * y \in Stab(s)$.
3. $x \in Stab(s) \Rightarrow x^{-1}(s) = x^{-1}(x(s)) = (x^{-1} * x)(s) = e(s) = s \Rightarrow x^{-1} \in Stab(s)$

□

Thus we may form the left cosets of $Stab(s)$ in G:

$$G/Stab(s) := \{xStab(s) | x \in G\}.$$

Recall that these subsets of G are the equivalence classes for the equivalence relation:

$$\text{Given } x, y \in G,\ x \sim y \iff x^{-1} * y \in Stab(s),$$

hence they partition G into disjoint subsets.

Proposition. *Let $x, y \in G$ then $xStab(s) = yStab(s) \iff x(s) = y(s)$.*

Proof. Recall that x and y are in the same left coset $\iff x^{-1}y \in Stab(s)$. Hence $x^{-1}y(s) = s$. Composing both sides with x and simplifying by the axioms for a group action implies that $x(s) = y(s)$. □

We deduce that there is a *well defined* map (of sets):

$$\begin{aligned} \phi : G/Stab(s) &\longrightarrow Orb(s) \\ xStab(s) &\longrightarrow x(s) \end{aligned}$$

Proposition. *ϕ is a bijection.*

Proof. By definition, $Orb(s) := \{x(s) \in S | x \in G\}$. Hence ϕ is trivially surjective. Assume $\phi(xStab(s)) = \phi(yStab(s))$ for some $x, y \in G$. This implies the following:

$$\begin{aligned} x(s) = y(s) &\Rightarrow x^{-1}(y(s)) = s \\ &\Rightarrow (x^{-1} * y)(s) = s \\ &\Rightarrow x^{-1} * y \in Stab(s) \\ &\Rightarrow xStab(s) = yStab(s) \end{aligned}$$

Therefore ϕ is injective.

□

This immediately gives the following key result:

Orbit-Stabiliser Theorem. *Let $(G, *)$ be a group together with an action, φ, on a set S. Let $s \in S$ such that the orbit of s is finite ($|Orb(s)| < \infty$). Then $stab(s) \subset G$ is of finite index and*

$$(G : Stab(s)) = |Orb(s)|$$

Proof. Immediate from previous proposition. □

We have the following corollary:

Corollary. *If $(G, *)$ is a finite group acting on a set S and $s \in S$ then*

$$|G| = |Stab(s)| \cdot |Orb(s)|.$$

Proof. In this case $(G : Stab(s)) = |G|/|Stab(s)|$. Applying the orbit-stabiliser theorem yields the result. □

The orbit-stabiliser theorem allows us to prove non-trivial results about the structure of finite groups. As an example let us consider the action of G (a finite group) on itself by conjugation. The orbits under this action are called conjugacy classes.

Concretely, for $h \in G, Conj(h) := Orb(h) = \{g^{-1} * h * g | g \in G\}$. If $C_1, \cdots, C_r \subset G$ are the distinct conjugacy classes then we deduce that $|G| = \sum_{i=1}^{r} |C_i|$ and $|C_i||G| \ \forall i \in \{1, \cdots, r\}$.

If $G = GL_n(\mathbb{R})$, the group of invertible $n \times n$ matrices with real entries (under matrix multiplication), then two matrices are in the same conjugacy class if and only if they are similar.

Definition. *Let $(G, *)$ be a group. The center of G is the subset*

$$Z(G) := \{h \in G | g * h = h * g, \ \forall g \in G\}.$$

We leave it as an exercise to check that the center is a subgroup.

Theorem. *Let G be a finite group of order p^n, for p a prime number and $n \in \mathbb{N}$. Then the center is non-tivial.*

Proof. Let G act on itself by conjugation. We know $Z(G)$ is a subgroup. Observe that $h \in Z(G) \iff Conj(h) = \{h\}$. Recall that $Conj(e) = \{e\}$, hence $|Conj(e)| = 1$. Assume that $Z(G) = \{e\}$. Hence if $h \neq e$ then $|Conj(h)| > 1$. By the orbit-stabiliser theorem we know that $|Conj(h)|$ must divide p^n. Hence p divides $|Conj(h)|$. Because the conjugacy classes form a partition of G we deduce that $\exists m \in \mathbb{N}$ such that $p^n = 1 + pm$. This is not possible, hence $Z(G)$ cannot be trivial. □

Recall that Lagrange's theorem says that if G is a finite group and H is a subgroup then $|H|$ divides $|G|$. It is not true, in general, that given any divisor of $|G|$ there is a subgroup of that order. We shall see an example of such a group later. There are, however, partial converses to Lagrange's theorem.

Sylow's Theorem. *Let $(G, *)$ be a finite group such that p^n divides $|G|$, where p is prime. Then there exists a subgroup of order p^n.*

Proof. Assume that $|G| = p^n m$, where $m = p^r u$ with $HCF(p, u) = 1$. Our central strategy is to consider a cleverly chosen group action of G and prove one of the stabilizer subgroups has size p^n. We'll need to heavily exploit the orbit-stabilizer theorem.

Let S be the set of all *subsets* of G of size p^n. An element of S is an unordered n-tuple of distinct elements in G. There is a natural action of G on S by term-by-term composition on the left.

Let $\omega \in S$. If we fix an ordering $\omega = \{\omega_1, \cdots, \omega_{p^n}\} \in S$, then $g(\omega) := \{g * \omega_1, \cdots, g * \omega_{p^n}\}$.

- We first claim that $|Stab(\omega)| \leq p^n$. To see this define the function

$$\begin{aligned} f : Stab(\omega) &\to \omega \\ g &\to g * \omega_1 \end{aligned}$$

  By the cancellation property for groups this is an injective map. Hence $|Stab(\omega)| \leq |\omega| = p^n$.

- Observe that

$$|S| = \binom{p^n m}{p^n} = \frac{p^n m!}{p^n!(p^n m - p^n)!} = \prod_{j=0}^{p^n-1} \frac{p^n m - j}{p^n - j} = m \prod_{j=1}^{p^n-1} \frac{p^n m - j}{p^n - j}.$$

  Observe that if $1 \leq j \leq p^n - 1$ then j is divisible by p at most $n-1$ times. This means that $p^n m - j$ and $p^n - j$ have the same number of p factors, namely the number of p factor of j. This means that

$$\prod_{j=1}^{p^n-1} \frac{p^n m - j}{p^n - j}$$

  has no p factors. Hence $|S| = p^r v$, where $HCF(p, v) = 1$.

  Now recall that S is the disjoint union of the orbits of our action of G on S. Hence there must be an $\omega \in S$ such that $|Orb(\omega)| = p^s t$, where $s \leq r$ and $HCF(p, t) = 1$. By the orbit-stabilizer theorem we know that $|Stab(\omega)| = p^{n+r-s}\frac{u}{t}$. Because $|Stab(\omega)| \in \mathbb{N}$ and u and t are coprime to p, we deduce that $\frac{u}{t} \in \mathbb{N}$. Hence $|Stab(\omega)| \geq p^n$.

For this choice of $\omega \in S$, $Stab(\omega)$ is thus a subgroup of size p^n. □

Historically this is a slight extension of what is called Sylow's First Theorem. There are two more which describe the properties of such subgroups in greater depth.

3.6 Finite Symmetric Groups

As we have proven, if $(G, *)$ is a finite group of order n. Then G is isomorphic to a subgroup of Sym_n, the symmetric group on $\{1, 2...n\}$. Hence to properly understand finite groups we must understand these finite symmetric groups.

Proposition. *For $n \in \mathbb{N}$, $|Sym_n| = n!$.*

Proof. Any permutation σ of $\{1, 2...n\}$ is totally determined by a choice of $\sigma(1)$, then $\sigma(2)$ and so on. At each stage the possibilities drop by one. Hence the number of permutations is $n!$. □

We need to think of a way of elegantly representing elements of Sym_n. For $a \in \{1, 2...n\}$ and $\sigma \in Sym_n$ we represent the action of σ on a by a cycle:

$$(abc...f) \text{ where } b = \sigma(a),\ c = \sigma(b)...\sigma(f) = a.$$

We know that eventually we get back to a because σ has finite order. In this way every $\sigma \in Sym_n$ can be written as a product of disjoint cycles:

$$\sigma = (a_1...a_r)(a_{r+1}...a_s)...(a_{t+1}...a_n).$$

This representation is unique up to internal shifts and reordering the cycles.

E.g. Let $n = 5$ then $\sigma = (123)(45)$ corresponds to

$$\begin{array}{rcl} & 1 & \longrightarrow & 2 \\ & 2 & \longrightarrow & 3 \\ \sigma: & 3 & \longrightarrow & 1 \\ & 4 & \longrightarrow & 5 \\ & 5 & \longrightarrow & 4 \end{array}$$

If an element is fixed by σ we omit it from the notation.

E.g. Let $n = 5$ then $\sigma = (523)$ corresponds to

$$\begin{array}{rcl} & 1 & \longrightarrow & 1 \\ & 2 & \longrightarrow & 3 \\ \sigma: & 3 & \longrightarrow & 5 \\ & 4 & \longrightarrow & 4 \\ & 5 & \longrightarrow & 2 \end{array}$$

This notation makes it clear how to compose two permutations. For example, let $n = 5$ and $\sigma = (23), \tau = (241)$, then $\tau\sigma = (241)(23) = (1234)$ and $\sigma\tau = (23)(241) = (1324)$. Observe that composition is on the left when composing permutations. This example also shows that in general Sym_n is not Abelian.

Hence, given $\sigma \in Sym_n$, we naturally get a well-defined partition of n, taking the lengths of the disjoint cycles appearing in σ. This is call the **cycle structure** of σ.

Proposition. *Let $\sigma \in Sym_n$ decompose as the disjoint product of cycles of length $n_1, ..n_m$ (so $\sum n_i = n$). Then $ord(\sigma) = LCM(n_1, ...n_m)$, where LCM denotes the lowest common multiple.*

Proof. Let $\sigma = (a_1, \cdots, a_r)(a_{r+1}, \cdots, a_s)\cdots(a_{t+1}, \cdots, a_n)$, be a representation of σ as the disjoint product of cycles. We may assume that $r = n_1$, etc, without any loss of generality. Observe that a cycle of length $d \in \mathbb{N}$ must have order d in Sym_n. Also recall that if G is a finite group then for any $d \in \mathbb{N}$, $x \in G$, $x^d = e \iff ord(x)|d$. Also observe that for all $d \in \mathbb{N}$, $\sigma^d = (a_1, \cdots, a_r)^d(a_{r+1}, \cdots, a_s)^d \cdots (a_{t+1}, \cdots, a_n)^d$. Thus we know that $\sigma^d = e \iff n_i|d\ \forall i$. The smallest value d can take with this property is $LCM(n_1, ...n_m)$. □

Theorem. *Two permutations are conjugate in Sym_n if and only if they have the same cycle structure.*

Proof. Let $\sigma, \tau \in Sym_n$ have the same cycle structure. Hence we may represent both in the form:

$$\sigma = (a_1, \cdots, a_r)(a_{r+1}, \cdots, a_s)\cdots(a_{t+1}, \cdots, a_n),$$

$$\tau = (b_1, \cdots, b_r)(b_{r+1}, \cdots, b_s) \cdots (b_{t+1}, \cdots, b_n).$$

Define $\alpha \in Sym_n$ such that $\alpha(a_i) = b_i \ \forall i$. By construction $\alpha^{-1}\tau\alpha = \sigma$. Going through the above process in reverse, the converse is clear. □

Corollary. *Conjugacy classes in Sym_n are indexed by cycle structures (i.e. partitions of n).*

Proof. Immediate from the above. □

Definition. *A **transposition** is a cycle of length 2.*

Observe that we can write any cycle as a product of transpositions:

k

Hence any permutation $\sigma \in \mathrm{Sym}_n$ may be written as the (not necessarily disjoint) product of transpositions. This representation is non-unique as the following shows:
e.g. n=6, σ=(1 2 3)=(1 3) (1 2)=(4 5)(1 3)(4 5)(1 2)
Notice that both expressions involve an even number of transpositions.

Theorem. *Let $\sigma \in Sym_n$ be expressed as the product of transpositions in two potentially different ways. If the first has m transpositions and the second has n transpositions then $2|(m-n)$.*

Proof. First notice that a cycle of length r can be written as the product of $r-1$ transpositions by the above. Let us call σ even if there are an even number of even length cycles (once expressed as a disjoint product); let us call σ odd if there are an odd number of even length cycles. We also define the sign of σ, denoted $sgn(\sigma)$, to be $+1$ or -1 depending on whether σ is even or odd.
Consider how sign changes when we multiply by a transposition (1 i). We have two cases:

1. 1 and i occur in the same cycle in σ. Without loss of generality we consider (1 2 $\cdots$ i $\cdots$ r) as being in σ.

   $$(1\ \mathrm{i})(1\ 2 \cdots\ i \cdots\ r) = (1\ 2 \cdots\ i-1)(i\ i+1 \cdots\ r)$$

   If r is even then either we get two odd length cycles or two even length cycles. If r is odd then exactly one of the cycles on the right is even length. In either case, $sgn((1\ i)\sigma) = -sgn(\sigma)$.

2. 1 and i occur in distinct cycles. Again, without loss of generality we may assume that $(1 \cdots i-1)(i \cdots r)$ occurs in σ. In this case

   $$(1\ \mathrm{i})(1\ 2 \cdots\ i-1)(i \cdots\ r) = (1 \cdots\ r).$$

   In either of the cases r even or odd, we see that the number of even length cycles must drop or go up by one. Hence $sgn((1\ i)\sigma) = -sgn(\sigma)$ as in case 1.

We deduce that multiplying on the left by a transposition changes the sign of our permutation. The identity must have sign 1, hence by induction we see that the product of an odd number of transpositions has sign -1, and the product of an even number of transpositions has sign 1.

Note that if we write any product of transpositions then we can immediately write down an inverse by reversing their order. Let us assume that we can express σ as the product of transpositions in two different ways, one with an odd number and one with an even number. Hence we can write down σ as the product of evenly many transpositions and σ^{-1} as a product of an odd number of transpositions. Thus we can write $e = \sigma * \sigma^{-1}$ as a product of an odd number of transpositions. This is a contradiction as $sgn(e) = 1$. □

We should observe that from the proof of the above we see that $\forall \sigma, \tau \in Sym_n$, $sgn(\sigma\tau) = sgn(\sigma)sgn(\tau)$. Because $sgn(e) = 1$ we deduce that $sgn(\sigma) = sgn(\sigma^{-1})$ for all $\sigma \in Sym_n$.

In particular this shows that the set of even elements of Sym_n contains the identity and is closed under composition and taking inverse. Hence we have the following:

Definition. *The subgroup of $Alt_n \subset Sym_n$ consisting of even elements is called the Alternating group of rank n.*

Observe that Alt_n contains all 3-cycles (cycles of length 3).

Proposition. *Alt_n is generated by 3-cylces.*

Proof. By generate we mean that any element of Alt_n can be expressed as the product of three cycles. As any element of Alt_n can be written as the product of three cycles we only have to do it for the product of two transpositions. There are two cases:

1. $(i\ j)(k\ l) = (k\ i\ l)(i\ j\ k)$.
2. $(i\ j)(i\ k) = (i\ k\ j)$.

□

Proposition. $|Alt_n| = \frac{n!}{2}$.

Proof. Recall that $|Sym_n| = n!$, hence we just need to show that $(Sym_n : Alt_n) = 2$. Let $\sigma, \tau \in Sym_n$. Recall that

$$\sigma Alt_n = \tau Alt_n \iff \sigma^{-1}\tau \in Alt_n.$$

But $sgn(\sigma^{-1}\tau) = sgn(\sigma)sgn(\tau)$, hence

$$\sigma Alt_n = \tau Alt_n \iff sgn(\sigma) = sgn(\tau).$$

Hence Alt_n has two left cosets in Sym_n, one containing even permutations and one odd permutations. □

Later we shall see that the alternating groups for $n \geq 5$ have a very special property.

3.7 Symmetry of Sets with Extra Structure

Let S be a set and $\Sigma(S)$ its permutation group. The permutation group completely ignores the fact that there may be extra structure on S.

As an example, $\mathbb{R}^n$ naturally has the structure of a vector space. The permutation group $\Sigma(\mathbb{R}^n)$ does not take this into account. However within the full permutation group there are linear permutations, namely $GL_n(\mathbb{R})$. These are permutations which preserve the vector space stucture.

Symmetry in Euclidean Space

Definition. *Given $n \in \mathbb{N}$, n-dimensional Euclidean space is the vector space $\mathbb{R}^n$ equipped with the standard inner product (the dot product).*

Concretely, if $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$ then $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := x_1 y_1 + \cdots + x_n y_n$.

Definition. *The distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ is*

$$d(\boldsymbol{x}, \boldsymbol{y}) := \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle}.$$

Definition. *An **isometry** of $\mathbb{R}^n$ is a map of sets $f : \mathbb{R}^n \to \mathbb{R}^n$ (not necessarily linear) such that $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $d(\boldsymbol{x}, \boldsymbol{y}) = d(f(\boldsymbol{x}), f(\boldsymbol{y}))$. The collection of all isometries of $\mathbb{R}^n$ is denoted by $Isom(\mathbb{R}^n)$.*

Remarks.

- *The identity function is an isometry and the composition of any two isometries is an isometry.*
- *We say an isometry, f, fixes the origin if $f(\mathbf{0}) = \mathbf{0}$. It is a fact that f fixes the origin if and only if $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^n$, where $\boldsymbol{A}$ is an orthogonal matrix.*
- *We say an isometry, f, is a translation if*

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ \boldsymbol{x} &\longrightarrow \boldsymbol{x} + \boldsymbol{y}. \end{aligned}$$

  for some $\boldsymbol{y} \in \mathbb{R}^n$.
- *Every isometry of $\mathbb{R}^n$ is a composition of an origin fixing isometry and a translation. As a consequence, all isometries are bijective and their inverses are isometries. This means $Isom(\mathbb{R}^n)$ is a subgroup of $\Sigma(\mathbb{R}^n)$.*

Let $X \subset \mathbb{R}^n$ be a subset (not necessarily a subspace).

Definition. *We define the symmetry group of X to be the subgroup $Sym(X) \subset Isom(\mathbb{R}^n)$ with the property that $f \in Sym(X)$ if and only if f permutes X.*

There is a natural action of $Sym(X)$ on the set X, coming from the fact there is a natural homomorphism $Sym(X) \to \Sigma(X)$. $Sym(X)$ measures how much symmetry X has. The more symmetric X, the larger $Sym(X)$.

The Dihedral Group

Let $m \in \mathbb{N}$ and $X \subset \mathbb{R}^2$ be a regular m-gon centered at the origin. We call the symmetry group of X the dihedral group of rank m, and we denote it by D_m.

First observe that every element of D_m must fix the center of X (the origin). Thus we may view D_m as a subgroup of the group of 2×2 orthogonal matrices. We shall not take this approach here.

Also observe that $f \in D_m$ acts faithfully and transitive on the set of vertices of X. Hence D_m can naturally by identified with a subgroup of Sym_m. Let σ be the rotation by $\frac{2\pi}{m}$ clockwise about the origin. All possible rotational symmetries are generated by σ, namely

$$Rot_m = \{e, \sigma, \sigma^2, \cdots, \sigma^{m-1}\} \subset D_m.$$

Hence Rot_m is cyclic of order m.

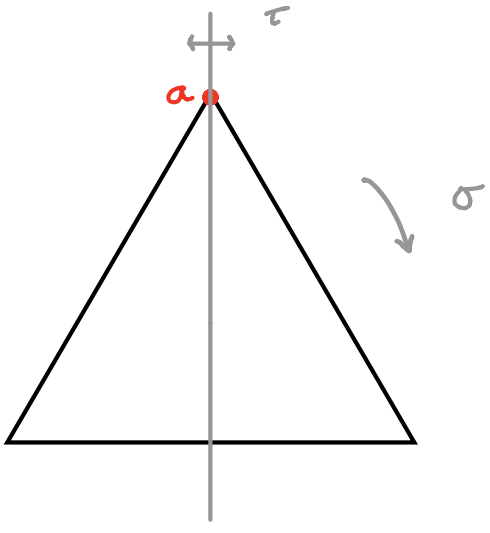
Given a vertex a, $Stab(a) = \{e, \tau\}$, where τ is the reflection through the straight line containing a and the origin. By the orbit-stabilizer theorem $|D_m| = 2m$, hence $(D_m : Rot_m) = 2$. We deduce that

$$D_m = Rot_m \coprod \tau Rot_m.$$

The left coset τRot_m is precisely the set of reflective symmetries. Hence every element of D_m can be written in the form σ^k (if a rotation) or $\tau\sigma^k$ (if a reflection). The group structure is completely determined by the following properties

- $ord(\sigma) = m$
- $ord(\tau) = 2$
- $\tau\sigma = \sigma^{-1}\tau$ (consider the action on the vertices)

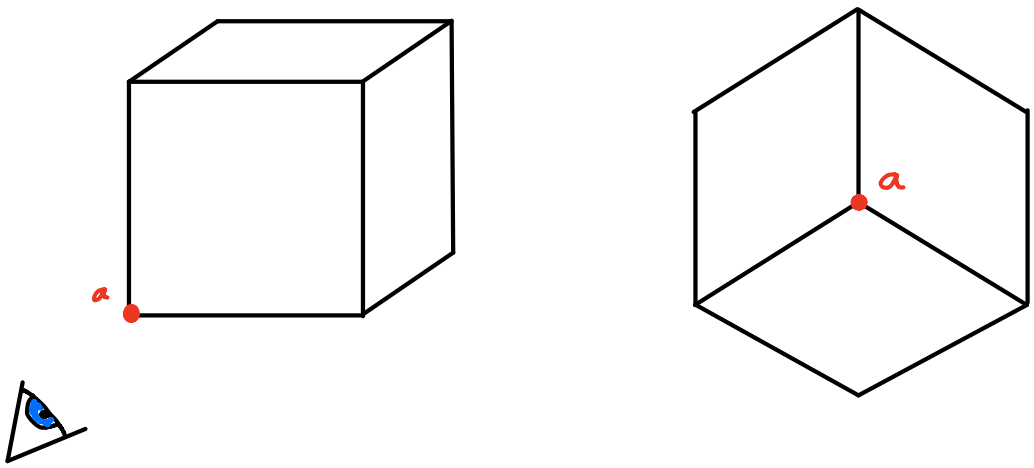
Observe that the third property implies that D_m is not Abelian. Here is a picture for $n = 3$.



The Cube in $\mathbb{R}^3$

Let $X \subset \mathbb{R}^3$ be a solid cube centered at the origin. Again, elements of $Sym(X)$ must fix the origin, hence, if we wished, we could identify $Sym(X)$ with a subgroup of the group of 3×3 orthogonal matrices.

Again $Sym(X)$ acts faithfully and transitively on the vertices. If $a \in X$ is a vertex, then $Stab(a)$ can naturally be identified with D_3 (see below figure) which has size 6. Hence, by the orbit-stabilizer theorem, $|Sym(X)| = 48$. The same logic applies to $Rot_\square$, the rotational symmetries, although the stabilizer of a now has size 3. This tells us that $|Rot_\square| = 24$.

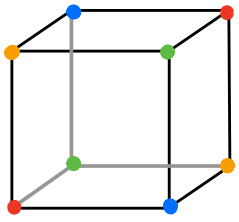


If $\tau \in Sym(X)$ is the symmetry sending $\boldsymbol{x}$ to $-\boldsymbol{x}$ (this is not a rotation), then again

$$Sym(X) = Rot_\square \coprod \tau Rot_\square.$$

It can be shown that $\tau\sigma = \sigma\tau$ for all $\sigma \in Rot_\square$. Thus it remains to determine the group structure of $Rot_\square$.

Color the vertices with four colors, making sure that opposite vertices have the same color (see below figure). Rotational symmetries act on this set of four colors, inducing a homomorphism from $Rot_\square$ to Sym_4. Given any two colors, it is possible to transpose them (leaving the others fixed) by a rotation. Because Sym_4 is generated by transpositions, the induced homormorphism $Rot_\square \to Sym_4$ must be surjective. However, $|Rot_\square| = 24 = 4! = |Sym_4|$. Hence it must be an isomorphism. We deduce that $Rot_\square$ is isomorphic to Sym_4.



Interesting Question:

Let $(G, *)$ be an abstract group. When is it true that we can find $X \subset \mathbb{R}^n$, for some $n \in \mathbb{N}$ such that

$$G \cong Sym(X)?$$

Less formally, when can an abstract group be realised in geometry?

3.8 Normal Subgroups and Isomorphism Theorems

In linear algebra the predominant objects we study are the maps between vector spaces, and not the vector spaces themselves. The structure preserving maps between vector spaces are more interesting than the spaces themselves. This a deep observation and it is true far beyond the confines of linear algebra. Philosophically it's saying that an object in isolation is uninteresting; it's how it relates to what's around it that matters. The world of group theory is no different. Here the objects are groups and the maps between them are homomorphisms. Now we'll study homomorphisms between abstract groups in more detail.

Let G and H be two groups. We'll suppress the $*$ notation as it will always be obvious where composition is taking place. Let e_G and e_H be the respective identity elements. Recall that a homomorphism from G to H is a map of sets $f : G \to H$ such that $\forall x, y \in G$, $f(xy) = f(x)f(y)$.

Definition. *Given $f : G \to H$ a homomorphism of groups, we define the **kernel** of f to be:*

$$Ker(f) := \{x \in G | f(x) = e_H\}$$

*We define the **image** of f to be:*

$$Im(f) := \{y \in H | \exists x \in G \text{ such that } f(x) = y\}$$

Proposition. *Given a homomorphism $f : G \to H$, $Ker(f) \subseteq G$ and $Im(f) \subseteq H$ are subgroups.*

Proof. First we will show true for Ker(f):

1. $f(e_G) = e_H \Rightarrow e_G \in Ker(f)$.
2. Suppose $x, y \in Ker(f)$. Then $f(xy) = f(x)f(y) = e_H \Rightarrow xy \in Ker(f)$.
3. Given $x \in Ker(f)$, $f(x^{-1}) = e_H^{-1} = e_H \Rightarrow x^{-1} \in Ker(f)$.

Now we will show that Im(f) is a subgroup:

1. $f(e_G) = e_H$ so $e_H \in$ Im(f).
2. $f(xy) = f(x)f(y) \forall x, y \in G$ so Im(f) is closed under composition.
3. Note that $f(x)^{-1} = f(x^{-1}) \Rightarrow y \in Im(f) \Rightarrow y^{-1} \in Im(f)$.

□

Proposition. *A homomorphism $f : G \to H$ is injective if and only if $ker(f)$ is trivial.*

Proof. f injective $\Rightarrow Ker(f) = \{e_G\}$ trivially. Now assume $ker(f) = \{e_G\}$. Suppose $x, y \in G$ such that $f(x) = f(y)$.

$$\begin{aligned} f(x) = f(y) &\Rightarrow f(x)f(y)^{-1} = e_H \\ &\Rightarrow f(x)f(y^{-1}) = e_H \\ &\Rightarrow f(xy^{-1}) = e_H \\ &\Rightarrow xy^{-1} = e_G \\ &\Rightarrow x = y \end{aligned}$$

Thus f is injective.

□

Recall that for $m \in \mathbb{N}$ the set of left cosets of $m\mathbb{Z}$ in $\mathbb{Z}$, denoted $\mathbb{Z}/m\mathbb{Z}$ naturally inherited the structure of a group from $+$ on $\mathbb{Z}$. It would be reasonable to expect that this was true in the general case, i.e. given G a group and H, a subgroup, the set G/H naturally inherits the structure of a group from G. To make this a bit more precise let's think about what *naturally* means. Let $xH, yH \in G/H$ be two left cosets. Recall that x and y are not necessarily unique. The only obvious way for combining xH and yH would be to form $(xy)H$.

Warning: in general this is not well defined. It will depend on the choice of x and y.

Something very special happens in the case $G = \mathbb{Z}$ and $m\mathbb{Z} = H$.

Fundamental Definition. *We call a subgroup $H \subseteq G$ **normal** if, for all $g \in G$, $gHg^{-1} = \{ghg^{-1} | g \in G, h \in H\} = H$. We denote normal subgroup by $H \triangleleft G$.*

Remarks.

1. *Observe that this is **not** saying that given $g \in G$ and $h \in H$, then $ghg^{-1} = h$. It is merely saying that $ghg^{-1} \in H$. A normal subgroup is the union of conjugacy classes of G.*
2. *If G is Abelian, every subgroup is normal as $ghg^{-1} = h \forall g, h \in G$.*
3. *Let G=Sym_3, $H = \{e, (12)\}$. Then $(13)(12)(13) = (23) \notin H$*

   *Hence H is **not** normal in Sym_3, so in general not all subgroups of a group are normal.*

Proposition. *Let G and H be two groups. Let $f : G \to H$ a homomorphism. Then $Ker(f) \subset G$ is a normal subgroup.*

Proof. Let $h \in Ker(f)$ and $g \in G$. Then $f(ghg^{-1}) = f(g)f(h)f(g^{-1}) = f(g)e_H f(g)^{-1} = e_H \Rightarrow ghg^{-1} \in Ker(f)$. □

In general $Im(f) \subset H$ is **not** normal.

Fundamental Definition. *We say a group G is **simple** if its only normal subgroups are $\{e\}$ and G.*

Cyclic groups of prime order are trivially simple by Lagrange's theorem. It is in fact true that for $n \geq 5$, Alt_n is simple, although proving this will take us too far afield. As we shall see later simple groups are the core building blocks of groups theory.

The importance of normal subgroups can be seen in the following:

Proposition. *Let $H \subseteq G$ be a normal subgroup. Then the binary operation:*

$$\begin{gathered} G/H \times G/H \to G/H \\ (xH, yH) \mapsto (xy)H \end{gathered}$$

is well defined.

Proof. As usual the problem is that that coset representatives are not unique and thus we could have two representatives giving different maps. Thus our goal is to show:

$$\forall x_1, x_2, y_1, y_2 \in G \text{ such that } x_1H = x_2H \text{ and } y_1H = y_2H, \text{ then } (x_1y_1)H = (x_2y_2)H$$

By assumption we know $x_1^{-1}x_2,\ y_1^{-1}y_2 \in H$. Consider

$$u = (x_1y_1)^{-1}(x_2y_2) = y_1^{-1}x_1^{-1}x_2y_2$$

Hence $uy_2^{-1}y_1 = y_1^{-1}(x_1^{-1}x_2)y_1$. Therefore, by the normality of $H, uy_2^{-1}y_1 \in H \Rightarrow u \in H \Rightarrow (x_1y_1)H = (x_2y_2)H$.

□

This shows that if $H \subset G$ normal, G/H can be endowed with a natural binary operation.

Proposition. *Proposition Let G be a group; $H \subset G$ a normal subgroup. Then G/H is a group under the above binary operation. We call it the quotient group.*

Proof. Simple check of three axioms of being a group.

1. $\forall x, y, z \in G, (xy)z = x(yz) \Rightarrow (xH * yH) * zH \Rightarrow xH * (yH * zH)$.
2. $xH * H = xH = H * xH \Rightarrow H \in G/H$ is the identity.
3. $xH * x^{-1}H = xx^{-1}H = eH = H = x^{-1}xH = x^{-1}H * xH \Rightarrow$ inverses exist.

□

Proposition. *The natural map*

$$\begin{array}{rcl} \phi : G & \longrightarrow & G/H \\ x & \longrightarrow & xH \end{array}$$

is a homomorphism with $Ker(\phi) = H$.

Proof. Observe that $\forall x, y \in G$, $\phi(xy) = xyH = xHyH = \phi(x)\phi(y) \Rightarrow \phi$ is a homomorphism. Recall that the identity element in G/H is the coset H. Hence for $x \in Ker(\phi) \iff \phi(x) = xH = H \iff x \in H$. Hence $Ker(\phi) = H$.
Observe that this shows that any normal subgroup can be realised as the kernel of a group homomorphism.

□

The First Isomorphism Theorem

Let G and H be groups, with respective identities e_G and e_H. Let $\phi : G \to H$ be a homomorphism. Recall that $Ker(\phi) \subset G$ is a normal subgroup. Hence we may form the *quotient* group $G/Ker(\phi)$. Let $x, y \in G$ such that they are in the same left coset of $Ker(\phi)$. Recall that $xKer(\phi) = yKer(\phi) \iff x^{-1}y \in Ker(\phi) \iff \phi(x^{-1}y) = e_H \iff \phi(x^{-1})\phi(y) = e_H \iff \phi(x)^{-1}\phi(y) = e_H \iff \phi(x) = \phi(y)$. In summary, $\phi(x) = \phi(y) \iff xKer(\phi) = yKer(\phi)$ Hence ϕ is **constant** on each coset of $Ker(\phi)$.

Hence we get a map of sets :

$$\begin{aligned} \varphi : G/Ker(\phi) &\longrightarrow Im(\phi) \\ xKer(\phi) &\longrightarrow \phi(x) \end{aligned}$$

This is well define precisely because of the above observations.

The First Isomorphism Theorem. *Let G and H be two groups. Let $\phi : G \to H$ be a homomorphism, then the induced map*

$$\begin{aligned} \varphi : G/Ker(\phi) &\longrightarrow Im(\phi) \\ xKer(\phi) &\longrightarrow \phi(x) \end{aligned}$$

is an isomorphism of groups.

Proof. Firstly we observe that the induced ϕ is by definition of $Im(\phi)$ surjective. Note that given $x, y \in G$, $\varphi(xKer(\phi)) = \varphi(yKer(\phi)) \iff \phi(x) = \phi(y) \iff xKer(\phi) = yKer(\phi)$, hence φ is injective.

It is left for us to show that φ is a homomorphism. Given $x, y \in G$, $\varphi(xKer(\phi)yKer(\phi)) = \varphi(xyKer(\phi)) = \phi(xy) = \phi(x)\phi(y) = \varphi(xKer(\phi))\varphi(yKer(\phi))$.
Therefore $\phi : G/Ker(\phi) \to Im(\phi)$ is a homomorphism, and thus an isomorphism.

□

The Third Isomorphism Theorem

Let G be a group and N a normal subgroup. The third isomorphism theorem concerns the connection between certain subgroups of G and subgroups of G/N.

Let H be a subgroup of G containing N. Observe that N is automatically normal in H. Hence we may form the quotient group $H/N = \{hN | h \in H\}$. Observe that H/N is naturally a subset of G/N.

Lemma. *$H/N \subset G/N$ is a subgroup.*

Proof. We need to check the three properties.

1. Recall that $N \in G/N$ is the identity in the quotient group. Observe that $N \subset H \Rightarrow N \in H/N$.
2. Let $x, y \in H$. By definition $xy \in H$. Thus $xNyN = (xy)N \in H/N$.
3. Let $x \in H$. By definition $x^{-1} \in H$. Thus $(xN)^{-1} = x^{-1}N \in H/N$.

□

Conversely, let $M \subset G/N$ be a subgroup. Let $H_M \subset G$ be the union of the left cosets contained in M .

Lemma. *$H_M \subset G$ is a subgroup.*

Proof. We need to check the three properties.

1. Recall that $N \in G/N$ is the identity in the quotient group. Hence $N \in M \Rightarrow N \subset H_M$. N is a subgroup hence $e_G \in N \Rightarrow e_G \in H_M$.
2. Let $x, y \in H_M$. This implies that $xN, yN \in M$. M is a subgroup, hence $xNyN = xyN \in M$. This implies that $xy \in H_M$.
3. Let $x \in H_M$. Hence $xN \in M$. M is a subgroup, hence $(xN)^{-1} = x^{-1}N \in M$. This implies that $x^{-1} \in H_M$.

□

Hence we have two maps of sets:

$$\begin{array}{rcl} \alpha : \{\text{Subgroups of } G \text{ containing } N\} & \longrightarrow & \{\text{Subgroups of } G/N\} \\ H & \longrightarrow & H/N \end{array}$$

and

$$\begin{array}{rcl} \beta : \{\text{Subgroups of } G/N\} & \longrightarrow & \{\text{Subgroups of } G \text{ containing } N\} \\ M & \longrightarrow & H_M \end{array}$$

Proposition. *These maps of sets are inverse to each other.*

Proof. We need to show that composition in both directions gives the identity function.

1. Let H be a subgroup of G containing N. Then $\beta\alpha(H) = \beta(H/N) = H$. Thus $\beta\alpha$ is the identity map on {Subgroups of G containing N}.

2. Let M be a subgroup of G/N. then $\alpha\beta(M) = \alpha(H_M) = M$. Thus $\alpha\beta$ is the identity map on {Subgroups of G/N}.

□

We deduce that both α and β are bijections and we have the following:

The Third Isomorphism Theorem. *Let G be a group and $N \subset G$ a normal subgroup. There is a natural bijection between the subgroups of G containing N and subgroups of G/N.*

Proof. Either map α or β exhibits the desired bijection. □

3.9 Direct Products and Direct Sums

Definition. *Let G and H be two groups, with respective identities e_G and e_H. We may form the **direct product** $G \times H = \{(x, g) | x \in G \; g \in H\}$. Let $x, y \in G$ and $g, h \in H$. Observe that there is a natural binary operation on $G \times H$ given by:*

$$(x, g) * (y, h) := (xy, gh).$$

Lemma. *$G \times H$ is a group under the natural binary operation.*

Proof. 1. Associativity holds for both G and $H \Rightarrow$ associativity hold for $G \times H$.

2. (e_G, e_H) is the identity.

3. For $g \in G$ and $h \in H$ $(g, h)^{-1} = (g^{-1}, h^{-1})$.

□

There is an obvious generalization of this concept to any finite collection of groups.

Definition. *Let G be a group and $H, K \subset G$ two subgroups. Let us furthermore assume that*

1. $\forall h \in H$ and $\forall k \in K$, $hk = kh$.

2. Given $g \in G$ there exist unique $h \in H$, $k \in K$ such that $g = hk$.

*Under these circumstances we say that G is the **direct sum** of H and K and we write $G = H \oplus K$. Observe that the second property is equivalent to:*

3. $H \cap K = \{e_G\}$ and for $g \in G$ there exist $h \in H$, $k \in K$ such that $g = hk$.

For example, $(\mathbb{Z}/15\mathbb{Z}, +)$ is the direct sum of $gp([3])$ and $gp([5])$.

Proposition. *If G is the direct sum of the subgroups $H, K \subset G$ then $G \cong H \times K$.*

Proof. Define the map

$$\begin{aligned} \phi : H \times K &\longrightarrow G \\ (h, k) &\longrightarrow hk \end{aligned}$$

Let $x, y \in H$ and $g, h \in K$. By property one $\phi((x, g)(y, h)) = \phi(xy, gh) = xygh = xgyh = \phi(x, g)\phi(y, h)$. Hence ϕ is a homomorphism. Property two ensures that ϕ is bijective. □

The concept of direct sum has a clear generalization to any finite collection of subsets of G.

3.10 Finitely Generated Abelian Groups

Let G be an Abelian group. We shall now use additive notation to express composition within G. In particular we will denote the identity by 0 (not to be confused with $0 \in \mathbb{Z}$). We do this because we are very familiar with addition on $\mathbb{Z}$ being commutative. Given $m \in \mathbb{Z}$ and $a \in G$, we write

$$ma = \begin{cases} a * a * \cdots * a \ \ (m \text{ times}), & \text{if } m > 0 \\ 0, & \text{if } m = 0 \\ a^{-1} * a^{-1} * \cdots * a^{-1} \ \ (-m \text{ times}), & \text{if } m < 0 \end{cases}$$

We have the identities:

1. $m(a + b) = ma + mb$
2. $(m + n)a = ma + na$
3. $(mn)a = m(na)$

$\forall a, b \in G; m, n \in \mathbb{Z}$

Now assume that G is finitely generated. Hence $\exists\{a_1, \cdots, a_n\} \subset G$ such that $gp(\{a_1, \cdots, a_n\}) = G$. In other words, because G is Abelian, every $x \in G$ can be written in the form

$$x = \lambda_1 a_1 + \cdots + \lambda_n a_n \ \ \lambda_i \in \mathbb{Z}.$$

In general such an expression is not unique. For example is G is of order $m \in \mathbb{N}$ then $(m + 1)a = a$ for all $a \in G$. This is because $ma = 0$ A reasonable goal would be to find a generating set such that every expression of the above form was unique (after possibly restricting $0 \leq \lambda_1 < ord(a_i)$) for a given $x \in G$. Such a generating set is called a basis for

G. Observe that it is not clear that such a basis even exists at present. If $\{a_1, \cdots, a_n\} \subset G$ were a basis then letting $A_i = gp(a_i) \subset G$ we have the direct sum decomposition:

$$G = A_1 \oplus \cdots \oplus A_n.$$

Conversely, if G can be represented as the direct sum of cyclic subgroups then choosing a generator for each gives a basis for G.

Definition. *Let G be an Abelian group. $x \in G$ is **torsion** is it is of finite order. We denote the subgroup of torsion elements by $tG \subset G$, called the torsion subgroup.*

Lemma. *$tG \subset G$ is a subgroup.*

Proof. This critically requires that G be Abelian. It is not true in general.

1. $ord(0) = 1 \Rightarrow 0 \in tG$
2. Let $g, h \in tG \Rightarrow \exists n, m \in \mathbb{N}$ such that $ng = mg = 0 \Rightarrow nm(g+h) = (mng + nmh) = m0 + n0 = 0 \Rightarrow g + h \in tG$.
3. $ng = 0 \Rightarrow -(ng) = n(-g) = 0$. Hence $g \in tG \Rightarrow -g \in tG$.

□

Clearly if G is finite then $tG = G$.

Definition. *If $tG = G$ we say that G is a torsion group. If $tG = \{0\}$ we say that G is torsion free.*

Proposition. *If G is torsion and finitely generated then G is finite.*

Proof. Let $\{a_1, \cdots, a_n\} \subset G$ be a generating set. Each element is of finite order hence every element $x \in G$ can be written in the form

$$x = \lambda_1 a_1 + \cdots + \lambda_n a_n, \quad \lambda_i \in \mathbb{Z},\ 0 \leq \lambda_1 < ord(a_i).$$

This is a finite set. □

Proposition. *$G/tG$is a torsion free Abelian group.*

Proof. Firstly note that $tG \subset G$ is normal as G is Abelian, hence G/tG is naturally an abelian group. Let $x \in G$. Assume that $x + tG \in G/tG$ is torsion. Hence $\exists n \in \mathbb{N}$ such that $n(x + tG) = nx + tG = tG$. Hence $nx \in tG$ so $\exists m \in \mathbb{N}$ such that $mnx = 0$. Hence $x \in tG \Rightarrow xtG = tG$. □

Definition. *An finitely generated Abelian group G is said to be **free Abelian** if there exists a finite generating set $\{a_1, \cdots, a_n\} \subset G$ such that every element of G can be uniquely expressed as*

$$\lambda_1 a_1 + \cdots \lambda_n a_n \text{ where } \lambda_i \in \mathbb{Z}.$$

*In other words, if we can find a **basis** for G consisting of non-torsion elements.*

In this case

$$G = gp(a_1) \oplus \cdots \oplus gp(a_n) \cong \mathbb{Z} \times \mathbb{Z} \cdots \times \mathbb{Z} = \mathbb{Z}^n.$$

Proposition. *Let G be a finitely generated free abelian group. Any two bases must have the same cardinality.*

Proof. Let $\{a_1, \cdots, a_n\} \subset G$ be a basis. Let $2G := \{2x | x \in G\}$. $2G \subseteq G$ is a subgroup. Observe that $2G = \{\lambda_1 a_1 + \cdots \lambda_n a_n | \lambda \in 2]z\}$. Hence $(G : 2G) = 2^n$. But the left hand side is defined independently of the basis. The result follows. □

Definition. *Let G be a finitely generated free Abelian group. The **rank** of G is the size of a any basis.*

Theorem. *A finitely generated abelian group is free Abelan $\iff$ it is torsion free.*

Proof. ($\Rightarrow$) is trivial.
($\Leftarrow$)
Assume G is torsion-free, let $\{a_1, \cdots, a_n\} \subset G$ generate G. We will prove the result by induction on n.
Base Case: $n = 1$. $G = gp(a) \cong (\mathbb{Z}, +)$ which is free abelian. Therefore result is true for $n = 1$.
If $\{a_1, \cdots, a_n\} \subset G$ is a basis we have nothing to prove. Suppose that it is not a basis. then we have a non-trivial relation:

$$\lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_n a_n = 0$$

If $\exists d \in \mathbb{Z}$ such that $d | \lambda_i$ for all i, then have $d(\frac{\lambda_1 a_1}{d} + \frac{\lambda_2 a_2}{d} + \cdots + ...) = 0$. As G is torsion-free, $(\frac{\lambda_1 a_1}{d} + \frac{\lambda_2 a_2}{d} + \cdots + ...) = 0$. We can therefore assume that the λ_i are collectively coprime. If $\lambda_1 = 1$, then we can shift terms to get $a_1 = -(\lambda_2 a_2 + \lambda_3 a_3 + \cdots + \lambda_n a_n)$. Therefore, G is generated by the $\{a_2, \cdots, a_n\} \subset G$ and the result follows by induction. We will reduce to this cases as follows: Assume $|\lambda_1| \geq |\lambda_2| > 0$. By the remainder theorem we may choose $\alpha \in \mathbb{Z}$ such that $|\lambda_1 - \alpha\lambda_2| < |\lambda_2|$. Let $a_2' = a_2 + \alpha a_1$ and $\lambda_1' = \lambda_1 - \alpha\lambda_2$, then

$$\lambda_1' a_1 + \lambda_2 a_2' + \cdots + \lambda_n a_n = 0.$$

Also observe that $\{a_1, a_2', \cdots, a_n\} \subset G$ is still a generating set and $\{\lambda_1', \cdots, \lambda_n\}$ are still collectively coprime. This process must must eventually terminate with one of the coefficients equal either 1 or -1. In this case we can apply the inductive step as above to conclude that G is free abelian. □

Proposition. *Let G be finitely generated and Abelian. Then G/tG is a finitely generated free Abelian group.*

Proof. G/tG is torsion free. We must show that G/tG is finitely generated. Let $\{a_1, \cdots, a_n\} \subset G$ generate G. Then $\{a_1 + tG, \cdots, a_n + tG\} \subset G/tG$ forms a generating set. By the above theorem G/tG is free Abelian. □

Definition. *Let G be a finitely generated Abelian group. We define the rank of G to be the rank of G/tG.*

Let G be finitely generated and Abelian. Let G/tG be of rank $n \in \mathbb{N}$ and let $f_1, \cdots, f_n$ be a basis for G/tG. Let $\phi : G \to G/tG$ be the natural quotient homomorphism. Clearly ϕ is surjective. Choose $\{e_1, \cdots, e_n\} \subset G$ such that $\phi(e_i) = f_i \ \forall i \in \{1, \cdots, n\}$. None of the f_i have finite order $\Rightarrow$ none of the e_i have finite order. Moreover

$$\phi(\lambda_1 e_1 + \cdots + \lambda_n e_n) = \lambda_1 f_1 + \cdots + \lambda_n f_n \in G/tG.$$

Because $\{f_1, \cdots, f_n\}$ is a free basis for G/tG we deduce that $\lambda_1 e_1 + \cdots + \lambda_n e_n = 0 \iff \lambda_i = 0 \forall i \Rightarrow F := gp\{e_1, \cdots, e_n\} \subseteq G$ is free abelian with basis $\{e_1, \cdots, e_n\} \Rightarrow F$ is torsion free. Therefore $F \cap tG = \{0\}$.
Let $g \in G$. By definition, $\exists \lambda_1, \cdots, \lambda_n \in \mathbb{Z}$ such that $\phi(g) = \lambda_1 f_1 + \cdots + \lambda_n f_n$. Then we have:

$$\begin{aligned} \phi(g) = \lambda_1 f_1 + \cdots + \lambda_n f_n \quad &\Rightarrow \quad \phi(g) = \phi(\lambda_1 e_1 + \cdots + \lambda_n e_n) \\ &\Rightarrow \quad \phi(g - (\lambda_1 e_1 + \cdots + \lambda_n e_n)) = 0 \\ &\Rightarrow \quad g - (\lambda_1 e_1 + \cdots + \lambda_n e_n) \in ker\phi = tG \\ &\Rightarrow \quad \exists h \in tG \ s.t. \ g = (\lambda_1 e_1 + \cdots + \lambda_n e_n) + h \end{aligned}$$

Hence every x may be written uniquely in the form $x = f + g$ where $f \in F$ and $g \in tG$.

Proposition. *Every finitely generated Abelian group can be written as a direct sum of a free Abelian group and a finite group.*

Proof. By the above, we may write

$$G = F \oplus tG$$

Define the homomorphism :

$$\begin{aligned} G = F \oplus tG \quad &\longrightarrow \quad tG \\ f + h \quad &\longrightarrow \quad h \end{aligned}$$

This is surjective with kernel F, hence by the first isomorphism theorem tG is isomorphic to G/F. The image of any generating set of G is a generating set for G/F under the quotient homomorphism. Hence tG is finitely generated and torsion, hence finite. F is free Abelian by construction.

□

Hence we have reduced the study of finitely generated Abelian groups to understanding finite Abelian groups.

3.11 Finite Abelian Groups

Definition. *A finite group G (not necessarily Abelian) is a* ***p-group****, with $p \in \mathbb{N}$ a prime, if every element of G has order a power of p.*

By Sylow's Theorem the order of a finite p-group must be a power of p. From now on let G be a finite Abelian group. Let $p \in \mathbb{N}$ be a prime. We define $G_p := \{g \in G | ord(p) \text{ is a power of } p\} \subset G$.

Theorem 1. *$G_p \subset G$ is a subgroup.*

Proof. 1. $ord(0) = 1 = p^0 \Rightarrow 0 \in G_p$.

2. Let $g, h \in G_p \Rightarrow \exists r, s \in \mathbb{N}$ such that $p^r g = p^s h = 0 \Rightarrow p^{r+s}(g+h) = p^s(p^r g) + p^r(p^s h) = 0 + 0 = 0 \Rightarrow g + h \in G_p$.

3. Let $g \in G_p \Rightarrow \exists r \in \mathbb{N}$ such that $p^r g = 0 \Rightarrow -p^r g = p^r(-g) = 0 \Rightarrow -g \in G_p$

□

This critically relies on G being Abelian. By definition G_p is a p-group. Recall that $\forall g \in G$, $\text{ord}(g) || G|$ by Lagrange's Theorem. Therefore $G_p = 0$ unless **possibly** if p divides $|G|$. By Sylow's Theorem we deduce that if $|G| = p^n u$, where $HCF(p, u) = 1$, then $|G_p| = p^n$. Thus $G_p \subseteq G$ is the maximal p-subgroup contained in G. The importance of the maximal p-subgroups is the following theorem.

Theorem. *Let G is a finite Abelian group. Let $\{p_1, \cdots, p_r\}$ be the primes dividing $|G|$. Then*

$$G = G_{p_1} \oplus \cdots \oplus G_{p_r}$$

Moreover this is the unique way to express as the direct sum of p-subgroups for distinct primes.

Proof. Let $|G| = n = a_1 a_2 \cdots a_r$ where $a_i = p_i^{\alpha_i}$. Let $P_i = n/a_i$. $\{P_1, \cdots, P_r\} \subset \mathbb{Z}$ are collectively coprime $\Rightarrow$ $\exists\, Q_1, \cdots, Q_r \in \mathbb{Z}$ such that

$$P_1 Q_1 + \cdots + P_r Q_r = 1 \text{ (Extension of Euclid)}$$

Let $g \in G$ and $g_i = P_i Q_i g$. Clearly $g = g_1 + g_2 + \cdots + g_r$ and $p_i^{\alpha_i} g_i = Q_i(ng) = 0$. Hence $g_i \in G_{p_i}$.
We must prove the uniquness of this sum. Assume we had

$$g = g_1' + \cdots + g_r', \quad g_i' \in G_{p_i}.$$

Therefore $x = g_1 - g_1' = (g_2' - g_2) + (g_3' - g_3) + \cdots + (g_r' - g_r)$. The right hand size has order dividing P_1, the left hand side has order dividing Q_1. P_1 and Q_1 are coprime $\Rightarrow$ $\exists\, u, v \in \mathbb{Z}$ such that $up_1 + vq_1 = 1 \Rightarrow x = u(p_1 x) + v(q_1 x) = 0 + 0 = 0 \Rightarrow g_1 = g_1'$. Similarly we find $g_i = g_i'$ for all $i \in \{1, \cdots, r\}$, hence the sum is unique and we deduce

$$G = G_{p_1} \oplus \cdots \oplus G_{p_r}.$$

Let $\{q_i, \cdots, q_s\}$ be a finite collection of distinct primes. Assume that G can be expressed as the direct sum

$$G = H_1 \oplus \cdots \oplus H_s \cong H_1 \times \cdots \times H_s$$

where H_i is a finite q_i-subgroup. Clearly $G_{q_i} = H_i$ and if p is a prime not in $\{q_1, \cdots, q_s\}$ $G_p = \{0\}$. Thus $\{p_1, \cdots, p_r\} = \{q_1, \cdots, q_s\}$ and any such representation is unique. □

We have however reduced the study of finite abelian groups to finite abelian p-groups.

Theorem 1. *Every finite Abelian p-group is a direct sum of* ***cyclic groups****.*

Proof. Let G be a finite Abelian p-group. If G is cyclic, we are done, otherwise take a cyclic subgroup $B = gp(b)$ of maximal order, say p^n. Our strategy is to show that there is a p-subgroup $D \subset G$ such that $G = B \oplus D$. We apply the following inductive hypothesis: For any finite Abelian p-group F of size less than $|G|$, if $M \subset F$ is a maximal cyclic subgroup then there exists $N \subset F$ such that $M \oplus N = F$. This is clearly true for F trivial.

We claim that there is a subgroup C of order p such that $B \cap C = \{0\}$. Recall that because G is Abelian G/B is naturally an Abelian p-group. Let $c \in G \setminus B$ and suppose $cB \in G/B$ has order p^r for $r > 0$. Observe that the maximal order of any element in G/B is less than or equal to p^n. Thus we know $n \geq r$. By definition $p^r(cB) = B \Rightarrow p^r c \in B$. Thus there exists $s \in \mathbb{N}$ such that $p^r c = sb$. By maximality of the order of b we know $0 = p^n c = sp^{n-r}b$. But $ord(b) = p^n$, hence $p^n | sp^{n-r}$. Therefore we have $p|s$, say $s = ps'$. Hence $c_1 = p^{r-1}c - s'b$ has order p and is not in B. Therefore $C = gp(c_1)$ is the required subgroup.

Let $BC = \{ab | a \in B, b \in C\}$. We claim that $BC \subset G$ is a subgroup.

1. $e_G \in B$ and $e_G \in C \Rightarrow e_G \in BC$.
2. Let $a_1, a_2 \in B$, $b_1, b_2 \in C$. Then $(a_1b_1)(a_2b_2) = (a_1a_2)(b_1b_2) \in BC$. Hence BC is closed under composition.
3. Let $a_1 \in B$, $b_1 \in C$. Then $(a_1b_1)^{-1} = b_1^{-1}a_1^{-1} = a_1^{-1}b_1^{-1} \in BC$. Hence BC is closed under taking inverses.

First observe that $|G/C| < |G|$. Hence the inductive hypothesis applies to G/C. Observe that $BC \subset G$ is a subgroup containing C. Observe that BC/C is cyclic, generated by $bC \in BC/C$. Because $B \cap C = \{0\}$ we also know that $|BC/C| = p^n$. Note that the size of the maximal cyclic subgroup of G must be larger than or equal to the size of the maximal cyclic subgroup of G/C. However we have constructed a cyclic subgroup $BC/C \subset G/C$ whose order equals that of a B. Hence $BC/C \subset G/C$ is a maximal cyclic subgroup. Thus by our inductive hypothesis $\exists N \subset G/C$ such that $BC/C \oplus N = G/C$. By the third isomorphism

theorem we know that $N = D/C$ for a unique subgroup $D \subset G$ containing C. We claim that G is the direct sum of B and D.

Let $g \in G$. Then $gC \in G/C$ is uniquely expressible in form $g + C = (a + C) + (d + C) = (a+d)+C$, where $a \in B$ and $d \in D$. Hence $g = a+d+e$ for some $c \in C$. However $C \subset D$ so this expresses g as a sum of elements of B and D. Let $x \in B \cap D$. Hence $xC \in BC/C \cap D/C$. Assume that $x \neq 0$. Note that $x \notin C$. Hence xC is non-zero on BC/C and D/C. However by construction $BC/C \cap D/C = \{C\}$. This is a contraction. Hence $B \cap D = \{0\}$ and we deduce that $G = B \oplus D$.

Thus we have shown that given any finite Abelian p-group G and a maximal cyclic subgroup $B \subset G$, there exists a subgroup $D \subset G$ such that $G = B \oplus D$. Observe that D is a finite Abelian p-group, thus we can continue this process until eventually it must terminate. The end result will be an expression of G as a direct sum of cyclic p-groups.

□

Corollary. *For any finite Abelian p-group G , there exist a unique decreasing sequence of natural numbers $\{r_1, \cdots, r_n\} \subset \mathbb{N}$ such that*

$$G \cong \mathbb{Z}/p^{r_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{r_n}\mathbb{Z}.$$

Proof. By the previous theorem we know that G is the direct sum of cyclic groups each of p-power order. Thus we know that such integers exist . We will prove uniqueness by induction on $|G|$. Assume that there is are isomorphisms

$$G \cong \mathbb{Z}/p^{r_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{r_n}\mathbb{Z} \cong \mathbb{Z}/p^{s_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{s_m}\mathbb{Z},$$

where the r_i and s_j are a decreasing sequence of natural numbers. We therefore see that $|G| = p^{\sum_{i=1}^n r_i} = p^{\sum_{j=1}^m s_j}$. Hence $\sum_{i=1}^n r_i = \sum_{j=1}^m s_j$.

Let $pG = \{pg | g \in G\}$. It is a straightforward exercise (which we leave to the reader) to prove that pG is a subgroup of G. Note that for $r > 1$, $\mathbb{Z}/p^{r-1}\mathbb{Z} \cong p(\mathbb{Z}/p^r\mathbb{Z})$, where the isomorphism is given by sending $a + p^{r-1}\mathbb{Z}$ to $pa + p^r\mathbb{Z}$. We deduce therefore that there are isomorphisms

$$pG \cong \mathbb{Z}/p^{r_1-1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{r_n-1}\mathbb{Z} \cong \mathbb{Z}/p^{s_1-1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{s_m-1}\mathbb{Z}.$$

Observe now that $|pG| < |G|$, thus by induction we deduce that the r_i and s_j agree when restricted to entries strictly greater than 1. This, together with the fact that $\sum_{i=1}^n r_i = \sum_{j=1}^m s_j$, implies that the two sets are the same and thus uniqueness is proven. □

Proposition. *Let G is an Abelian group such that $p \in \mathbb{N}$ is a prime dividing $|G|$. Then G_p is non-trivial.*

Proof. Recall that if $\{p_1, \cdots, p_r\}$ are the primes dividing $|G|$ then

$$G \cong G_{p_1} \times \cdots \times G_{p_r}.$$

Hence $|G| = |G_{p_1}| \cdots |G_{p_r}|$. By the above corollary p_i divides $|G|$ if and only if G_{p_i} is non-trivial. □

Basis Theorem for Finitely Generated Abelain Groups. *Every finitely generated Abelian group G can be written as a direct sum of cyclic groups:*

$$G = \beta_1 \oplus \cdots \oplus \beta_r$$

where each β_i is either infinite or of prime power order, and the orders which occurs are uniquely determined.

Proof. $G{=}F \oplus tG$. F is free and finitely generated, hence the direct sum of infinite cyclic groups $(\mathbb{Z}, +)$. The number equals the rank of G. tG is finite Abelian, hence the is the unique direct sum of p-groups for distinct primes p. Each p-group is the unique direct sum (up to order) of p-power cyclic groups. □

Note that we could have stated this theorem with direct product in place of direct sum. Thus we have classified all finitely generate Abelian groups up to isomorphism.

3.12 The Classification of Finite Groups (Proofs Omitted)

In the last section we classified all finite Abelian groups up to isomorphism. Is it possible to do the same for all finite groups? It turns out that the situation is far more complicated in the non-Abelian case.

Here is the basic strategy:

- Show that any finite group G can be broken down into simple pieces.
- Classify these simple pieces.
- Understand how these simple pieces can fit together.

Definition. *let G be a finite group. A **composition series** for G is a nested collection of subgroups*

$$\{e\} = G_0 \triangleleft G_1 \triangleleft \cdots \triangleleft G_{r-1} \triangleleft G_r = G.$$

such that

- $G_{i-1} \neq G_i$ *for all* $0 < i \leq r$.
- G_i/G_{i-1} *is simple for all* $0 < i \leq r$.

Remarks. *By the third isomorphism theorem a composition series cannot be extended, meaning we cannot add any intermediate normal subgroups.*

Theorem. *Any finite group G has a composition series.*

Observe that if G is simple that $\{e\} = G_0 \lhd G_1 = G$ is a composition series.

If $G = Sym_3$ then

$$\{e\} \lhd gp((123)) \lhd Sym_3$$

gives a composition series. To see why, observe that each quotient group has size 3 or 2 and are therefore isomorphism to $\mathbb{Z}/3\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z}$ which are both.

Jordan-Holder Theorem. *Let G be a finite group. Suppose we have two composition series for G*

$$\{e\} = G_0 \lhd G_1 \lhd \cdots \lhd G_{r-1} \lhd G_r = G.$$
$$\{e\} = H_0 \lhd H_1 \lhd \cdots \lhd H_{s-1} \lhd H_s = G.$$

Then $r = s$ and the quotient groups

$$\{G_1/G_0, \cdots, G_r/G_{r-1}\}, \quad \{H_1/H_0, \cdots, H_s/H_{s-1}\}$$

are pairwise isomorphic (perhaps after reordering).

Definition. *If G has composition series*

$$\{e\} = G_0 \lhd G_1 \lhd \cdots \lhd G_{r-1} \lhd G_r = G.$$

we call the quotient groups

$$\{G_1/G_0, \cdots, G_r/G_{r-1}\}$$

*the **simple components** of G.*

By the Jordan-Holder Theorem the simple components are well-defined up to isomorphism. It is possible that two non-isomorphic groups have the same (up to isomorphism) simple components. As an example Sym_3 and $\mathbb{Z}/6\mathbb{Z}$ both have simple components $\{\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}\}$.

Definition. *A finite group is called solvable (or soluble) if its simple components are Abelian. Note that Solvable groups need not be Abelian themselves*

Note thatSym_3 is solvable, while Alt_5 (being simple and non-Abelian) is non-solvable.

To summarize our study: Finite group theory if much like the theory of chemical molecules.

- The simple groups are like atoms
- Finite groups have simple components, like molecules have constituent atoms.
- Non-isomorphic finite groups with the same simple components are like molecules with the same atoms but different structure (isomers).

We now have two goals

- Classify all finite simple groups up to isomorphism.
- Classify all finite simple groups with given simple components.

The theory of groups was initiated by Galois in 1832. Galois was the first to discover the first known simple groups, namely $\mathbb{Z}/p\mathbb{Z}$ for p prime and Alt_n for $n > 4$. Amazingly it took until 2004 until a complete classification was known. The proof stretches across over 10000 pages and is the combined work of thousands of mathematicians. Here's a very rough breakdown the the different four distinct classes of finite simple group:

- Cyclic groups of prime order. These are the only Abelian simple groups.
- Alt_n for $n > 4$
- Finite groups of Lie type. These groups are very complicated to describe in general. The basic idea is that they can be realized as subgroups and quotients of matrix groups. There are 16 infinite families of finite simple groups of Lie type.
- There are 26 sporadic groups. Very strangely these do not fall into any fixed pattern. The first were discovered in 1852 by Mathieu, while he was thinking about subgroups of finite permutation groups with extremely strong transitivity properties. The largest sporadic group was discovered in the 1970s. It's called the monster group and has size

$$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$$

  The monster contains all but six of the other sporadic groups as quotients of subgroups.

The theory of finite simple groups is one of the crown jewels of mathematics. It's demonstrates how profound the definiton of a group really is. All of this complexity is contained in those three innocent axioms.

The next question, of course, is to classify all finite groups with given simple components. This is still a wide open problem. As such a complete classification of all finite groups is still unknown.

One may ask about classifying infinite groups. Unsurprisingly the situation is even more complicated, although much progress has been made if specific extra structure (topological, analytic or geometric) is imposed.

4 Rings and Fields

4.1 Basic Definitions

A group $(G, *)$ is a set with a binary operation satisfying three properties. The motivation for the definition reflected the behavior of $(\mathbb{Z}, +)$. Observe that $\mathbb{Z}$ also comes naturally equipped with multiplication $\times$. In the first lectures we collected some of the properties of $(\mathbb{Z}, +, \times)$. Motivated by this we make the following fundamental definition:

Definition. *A **ring** is a set R with two binary operations, $+$, called addition, and $\times$, called multiplication, such that:*

1. *R is an **Abelian group** under addition.*
2. *R is a **monoid** under multiplication (inverses do not necessarily exist).*
3. *$+$ and $\times$ are related by the distributive law:*

$$(x + y) \times z = x \times z + y \times z \text{ and } x \times (y + z) = x \times y + x \times z \; \forall x, y, z \in R$$

The identity for $+$ is "zero", denoted 0_R (often just written as 0), and the identity for $\times$ is "one", denoted 1_R (often just written as 1).

Remarks. 1. *To simplify the notation we will write $x \times y = xy$ for all $x, y \in R$.*

2. *Distributivity implies that we can "multiply" together finite sums:*

$$(\sum x_i)(\sum y_j) = \sum x_i y_j$$

in a well-defined way.

Here are some examples of rings:

1. The integers under the usual addition and multiplication.
2. $\mathbb{Z}/m\mathbb{Z}$ under the addition and multiplication described in 2.3.
3. Let S be a set and $\mathbb{P}(S)$ be the set of all subsets. This is called the power set of S. On $\mathbb{P}(S)$ define $+$ and $\times$ by

$$X + Y = (X \cap Y') \cup (X' \cap Y),\; XY = X \cap Y$$

   Where X' denotes the complement of X in S. Then $\mathbb{P}(S)$ is a ring with $\emptyset = 0$ and $S = 1$. This strange looking ring has applications to mathematical logic.

4. In linear algebra the collection of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ is the set $\mathbb{M}_{n\times n}(\mathbb{R}^n)$. This has the structure of a ring under the usual addition and multiplication of matrices.

Note that matrix multiplication is not commutative in general. So it is perfectly possible for a multiplication not to be commutative in a ring.

Definition. *Let R be a ring with multiplication $\times$. If $\times$ is commutative, i.e. $xy = yx \; \forall \; x, y \in R$ then we say that R is a* ***commutative ring****.*

Definition. *Let R and S be two rings. A* ***homomorphism*** *ϕ from R to S is a map of sets $\phi : R \to S$ such that $\forall x, y \in R$*

1. $\phi(x + y) = \phi(x) + \phi(y)$
2. $\phi(xy) = \phi(x)\phi(y)$
3. $\phi(1_R) = 1_S$

Once again, if $R = S$ and $\phi = Id_R$ then we call it the identity homomorphism.

Note that R and S are abelian groups under $+$ so ϕ is a group homomorphism with respect to $+$ so $\phi(0_R) = 0_S$. We have to include (3) as $(R, \times)$ is only a monoid so it does not follow from (2) alone that $\phi(1_R) = 1_S$.

Remarks.
1. *As for groups, the composition of two ring homomorphisms is again a ring homomorphism.*
2. *As before, an* ***isomorphism*** *is a bijective homomorphism, or equivalently one with an inverse homomorphism. A homomorphism from R to itself is called an* ***endomorphism****. An endomorphism which is also an isomorphism is called an* ***automorphism****. This is exactly the same terminology as for groups.*

In any ring R we have the following elementary consequences of the axioms:

$$x0 = x(0 + 0) = x0 + x0 \Rightarrow x0 = 0$$

Similarly, $0x = 0$ for all $x \in R$.

If R consists of one element, then $1 = 0$, conversely if $1 = 0$ then $\forall x \in R, x = x1 = x0 = 0$, hence R consists of one element. The ring with one element is called the **trivial ring**.

In a ring we abbreviate expressions like

$$a + a + a + \cdots + a \; (n \; times) = na (n \in \mathbb{N})$$

It is clear that we may naturally extend this to all $n \in \mathbb{Z}$.
Similarly,

$$a \times a \times \cdots \times a \; (n \; times) = a^n \; for \; n \in \mathbb{N}.$$

By the Distributive Law, we have the identities

1. $m(a+b) = ma + mb$
2. $(m+n)a = ma + na$
3. $(mn)a = m(na)$

$\forall a, b \in R$ and $m, n \in \mathbb{Z}$.

Definition. *Given R and S two rings we say that R is a **subring** of S if it is a subset and is a ring under the induced operations (with same 0 and 1). Eg. $(\mathbb{Z}, +, \times) \subset (\mathbb{Q}, +, \times)$. More precisely,*

1. *R is a subgroup of S under addition.*
2. *R is closed under multiplication.*
3. *$1_S \in R$.*

Remarks. *As with subgroups, an arbitrary intersection of subrings is again a subring.*

4.2 Ideals, Quotient Rings and the First Isomorphism Theorem for Rings

Let G and H be groups and $\phi : G \to H$ a group homomorphism. Recall that $ker(\phi) \subset G$ is a normal subgroup, thus the set of right coset $G/ker(\phi)$ naturally forms a group (the quotient group). Recall that all normal subgroups arise in this manner. The 1^{st} Isomorphism theorem states that there is a natural isomorphism

$$G/ker(\phi) \cong Im(\phi).$$

Does something analogous hold for rings?

Let R and S be two rings. Let $\phi : R \to S$ be a ring homomorphism.

Definition. *The kernel of ϕ is the subset*

$$ker(\phi) := \{r \in R | \phi(r) = 0_S\} \subset R.$$

The image of ϕ is the subset

$$Im(\phi) := \{s \in S | \exists r \in R \text{ s.t. } \phi(r) = s\} \subset S.$$

Remember that ϕ is a group homomorphism with respect to the additive Abelian group structures on R and S. With respect to this structure these definitions are exactly the same as in group theory. In particular we know that

$$ker(\phi) = \{0_R\} \iff \phi \text{ is injective .}$$

We also know that $ker(\phi) \subset R$ and $Im(\phi) \subset S$ are subgroups under addition.

Proposition. *$Im(\phi) \subset S$ is a subring.*

Proof. We need to check that $Im(\phi)$ is closed under multiplication and contains 1_S. Let $s_1, s_2 \in Im(\phi)$. Hence $\exists r_1, r_2 \in R$ such that $\phi(r_1) = s_1$ and $\phi(r_2) = s_2$. But $s_1 s_2 = \phi(r_1)\phi(r_2) = \phi(r_1 r_2)$. Hence $s_1 s_2 \in Im(\phi)$. Hence $Im(\phi)$ is closed under multiplication.

By definition $\phi(1_R) = 1_S$. Hence $1_S \in Im(\phi)$. Thus $Im(\phi)$ is a subring.

□

If S is non trivial then because $\phi(1_R) = 1_S$ we know that $1_R \notin ker(\phi)$. Hence in this case $\ker(\phi) \subset R$ is not a subring. What properties does it satisfy?

1. $ker(\phi) \subset R$ is a subgroup under $+$.
2. Let $a \in ker(\phi)$ and $r \in R$. Observe that $\phi(ra) = \phi(r)\phi(a) = \phi(r)0_S = 0_S$. Hence $ra \in ker(\phi)$. Similarly $ar \in ker(\phi)$. Hence $ker(\phi)$ is closed under both left and right multiplication by **all** of R.

Definition. *Let R be a ring. An ideal $I \subset R$ is a subset which is a subgroup under addition and is closed under both left and right multiplication by all of R. More precisely, if $x \in I$ then $xr, rx \in I$ for all $r \in R$.*

We have just shown that the kernel of a homomorphism is always an ideal. An ideal is the ring theoretic analogue of normal subgroup in group theory.

Let $I \subset R$ be an ideal. Recall that $(R, +)$ is an abelian group, Hence $(I, +) \subset (R, +)$ is a normal subgroup. Hence the right cosets R/I naturally have a group structure under addition. We have completely ignored the multiplicative structure on R. Let us define a multiplication by:

$$(a + I) \times (b + I) := (ab) + I, \ \ \forall a, b \in R.$$

Lemma. *This binary operation is well defined.*

Proof. Let $a_1 + I = a_2 + I$ and $b_1 + I = b_2 + I$ where $a_1, a_2, b_1, b_2 \in R$. Observe that

$$a_1 b_1 - a_2 b_2 = a_1(b_1 - b_2) + (a_1 - a_2)b_2$$

is contained in I because I is an ideal. Thus

$$a_1 b_1 + I = a_2 b_2 + I.$$

□

Proposition. *R/I is a ring under the natural operations. We call it the **quotient ring**.*

Proof. This is just a long and tedious exercise to check the axioms which all follow because they hold on R. Unsurprisingly $0 + I$ is the additive identity and $1 + I$ is the multiplicative identity. □

As in the case of groups there is a natural surjective quotient ring homomorphism

$$\phi : R \to R/I.$$

From the definitions we see that $ker(\phi) = I$. We deduce that ideals of a ring are precisely the kernels of ring homomorphisms. This is totally analogous to the group theory situation.

The First Isomorphism Theorem. *Let $\phi : R \to S$ be a ring homomorphism. Then the induced map*

$$\begin{aligned} \varphi : R/ker(\phi) &\longrightarrow Im(\phi) \\ a + ker(\phi) &\longrightarrow \phi(a) \end{aligned}$$

is a ring isomorphism.

Proof. The first isomorphism theorem for groups tells us that it is an isomorphism of additive group. Hence we merely need to check that it is a ring homomorphism.

Let $a, b \in R$. $\varphi((a + ker(\phi))(b + ker(\phi))) = \varphi(ab + ker(\phi)) = \phi(ab) = \phi(a)\phi(b) = \varphi(a + ker(\phi))\varphi(b + ker(\phi))$. Also $\varphi(1 + I) = \phi(1) = 1$.

Hence φ is a ring homomorphism and we are done.

□

Definition. *An injective ring homomorphims $\phi : R \to S$ is called and **embedding**. By the first isomorphism theorem, R is isomorphic to the subring $Im(\phi) \subset S$.*

4.3 Properties of Elements of Rings

Definition. *Let R be a ring. An element $a \in R$ is said to be **invertible**, or a **unit**, if it has a multiplicative inverse, i.e. $\exists a' \in R$ such that $a'a = aa' = 1$. We know that such an inverse is unique if it exists, hence we shall write it as a^{-1}. Note that if $1 \neq 0$ then 0 is never invertible. We denote the set of units in R by R^*.*

It is clear that for any ring R, $(R^*, \times)$ is a group.

Definition. *A non-trivial ring R in which every non-zero element is invertible (i.e $R \setminus \{0\} = R^*$) is called a **division ring** (or **skew field**). If R is a commutative division ring then R is called a **field**.*

Remarks.

1. *$(\mathbb{Q}, +, \times)$ is the canonical example of a field. Other natural examples include $(\mathbb{R}, +\times)$, $(\mathbb{C}, +, \times)$ and $(\mathbb{Z}/p\mathbb{Z}, +, \times)$, where p is a prime number. There are examples of division rings which are not fields (i.e. not commutative) but we will not encounter them in this course.*

2. *All of linear algebra (except the issue of eigenvalues existing) can be set up over an arbitrary field. All proofs are exactly the same, we never used anything else about $\mathbb{R}$ or $\mathbb{C}$.*

In an arbitrary ring it is possible that two non-zero elements can multiply to give zero. For example, in $\mathbb{M}_{2\times 2}(R)$, the non-zero matrices

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$$

multiply to give the zero matrix.

Definition. *Let R be a non-trivial ring. Given $a \in R \setminus \{0\}$, if there exists $b \in R \setminus \{0\}$ such that $ab = 0$ or $ba = 0$, then a is said to be a **zero-divisor**. Note that 0 is not a zero-divsor.*

Definition. *A non-trivial ring R with **no zero divisors** is said to be **entire**; a commutative entire ring is called an **integral domain**. More concretely: R is entire if and only if $1 \neq 0$ and $\forall x, y \in R, xy = 0 \Rightarrow x = 0$ or $y = 0$.*

$(\mathbb{Z}, +, \times), (\mathbb{Q}, +, \times)$ are integral domains. $(\mathbb{Z}/m, +, \times)$ is an integral domain $\iff$ m prime. The above example shows that $\mathbb{M}_2(\mathbb{R})$ is not entire.

Theorem. *A ring R is **entire** $\iff$ its set of non-zero elements forms a **monoid** under multiplication. Another way to state this is that R entire $\iff$ $R \setminus \{0\}$ is closed under multiplication.*

Proof. In any ring R observe that if $x, y \in R$ are two non-zero divisors then by definition $xy \in R$ must be a non-zero divisor. Hence, If R is non-trivial the non-zero divisors of R are a monoid under multiplication. If R is entire the set of non-zero divisors is precisely $R\backslash\{0\}$, which implies it is a monoid under multiplication. Conversly if $R\backslash\{0\}$ is a monoid then firstly it is non-empty so R is non-tivial. But if $x, y \in R\backslash\{0\}$ then $xy \in R\backslash\{0\}$. Hence R is entire by definition.

□

Corollary. *Any field F is an integral domain.*

Proof. If $x, y \in F$, $x \neq 0 \neq y$ then $\exists\, x^{-1}, y^{-1} \in F$ such that $xx^{-1} = x^{-1}x = 1 = yy^{-1} = y^{-1}y$, therefore xy is invertible so is non-zero.
Hence, non-zero elements are closed under multiplication, so F is entire. F is a field so F is commutative, so it is an integral domain. □

Cancellation Law:

Let R be a ring. If $c \in R$ is not a zero-divisor, then for any $a, b \in R$ such that $ca = cb$ or $ac = bc$, then $a = b$.

This is because $ca - cb = c(a - b)$ and $ac - bc = (a - b)c$. In particular, if R is entire, then we can "cancel" any non-zero element. It is important to note that we cannot do this in an arbitrary ring.

Theorem. *Every **finite** integral domain R is a field.*

Proof. We need to show that $R^* = R \setminus \{0\}$. Let $a \in R \setminus \{0\}$. Define the following map of sets:

$$\begin{aligned} \psi : R \setminus \{0\} &\to R \setminus \{0\} \\ r &\mapsto ra. \end{aligned}$$

ψ is well define because R is an integral domain. By the cancellation law for integral domains, we know thatgiven $r_1, r_2 \in R$ $r_1 a = r_2 a \Rightarrow r_1 = r_2 \Rightarrow \psi$ injective. Since $R \setminus \{0\}$ is finite, ψ is surjective $\Rightarrow \exists\ b \in R \setminus \{0\}$ such that $ba = ab = 1$. Hence a has a multiplicative inverse. Therefore, $R^* = R \setminus \{0\}$. □

4.4 Polynomial Rings

Let R be a ring.

Definition. *The **polynomial ring** in X with cofficients in a ring R consists of formal expressions of the form:*

$$g(X) = b_0 + b_1 X + b_2 X^2 + \cdots + b_m X^m, b_i \in R, m \in \mathbb{N}$$

If $f(X) = a_0 + a_1 X + \cdots + a_n X^n$ is another polynomial then we decree that $f(X) = g(X) \iff a_i = b_i\ \forall i$. Note that we set $a_i = 0$ if $i > n$ and $b_j = 0$ if $j > m$. We refer to X as the indeterminant.

Addition and multiplication are defined by the rules

1. $f(X) + g(X) = (a_0 + b_0) + (a_1 + b_1)X + \cdots + (a_n + b_n)X^n$ *(if $m \leq n$)*
2. $f(X) \times g(X) = (a_0 b_0) + (a_0 b_1 + a_1 b_0)X + (a_0 b_2 + a_1 b_1 + a_2 b_2)X^2 + \cdots + a_n b_m X^{n+m}$

We will denote this ring by $R[X]$.

Important Exercise. *Check this genuinely gives a ring structure on the set of polynomials in X with coefficients in R.*

Note that there is a natural embedding:

$$\begin{aligned} \phi : R &\longrightarrow R[X] \\ a &\longrightarrow a \text{ (polynomial with } m = 0 \text{ and } a = a_0) \end{aligned}$$

Remarks. 1. *The zero and one elements in $R[X]$ are the image of the zero and one element in R under ϕ.*

2. *R commutative $\Rightarrow R[X]$ commutative.*

3. *Given $f(X) \in R[X]$ we can construct a map (of sets):*

$$\begin{aligned} \varphi_f : R &\longrightarrow R \\ a &\mapsto f(a), \end{aligned}$$

where $f(a) \in R$ is the element of R given be replacing X by a. For a general ring R this process can be quite subtle as we shall see.

Definition. *Let R be a ring and $f \in R[X]$ be a non-zero polynomial. We say that $a \in R$ is a **root**, or zero, of f if $f(a) = 0$.*

Definition. *Let R be a ring and $f \in R[X]$ be a non-zero polynomial. Hence we may write $f = c_n X^n + c_{n-1}X^{n-1} + \cdots + c_0,\ c_i \in R, c_n \neq 0$. We call n the degree of f and write deg(f)=n. If in addition $c_n = 1$, we say that f is monic. Elements of degree 0 are called constant polynomials.*

Theorem. *If R is entire then $R[X]$ satisfies:*

1. $\forall f, g \in R[X] \setminus \{0\}, deg(f+g) \leq max\{deg(f), deg(g)\}$
2. $\forall f, g \in R[X] \setminus \{0\} \Rightarrow fg \neq 0$ *and* $deg(fg) = deg(f) + deg(g)$.

Proof. By the definition of degree, (1) is clear. For (2):
Let $deg(f) = n, deg(g) = m$. Then suppose a_n, b_m the leading coefficients of f and g respectively. Hence fg has maximal power of X given by $a_n b_m X^{n+m}$. As R is entire, $a_n b_m \neq 0 \Rightarrow fg \neq 0$ and $deg(fg) = n + m = deg(f) + deg(g)$. □

Corollary. *R entire $\Rightarrow R[X]$ entire.*

Proof. Immediate from above.

□

Corollary. *R an integral domain $\Rightarrow R[X]$ and integral domain.*

Proof. Immediate from above.

□

The process of adjoining indeterminants to a ring R can be iterated to form polynomials in more than one variable with coefficients in R. We of course use another symbol for the indeterminants, ie. $R[X][Y]$, polynomials in X and Y with coefficients in R, e.g. $X^2 + Y^2X + X^3Y^6$.
We simplify this notation to $R[X][Y] = R[X, Y]$. Inductively, we define

$$R[X_1, \cdots, X_n] = R[X_1, \cdots, X_{n-1}][X_n]$$

$f \in R[X_1, \cdots, X_n]$ has a unique expression of the form

$$f = \sum a_{i_1 \cdots i_n} X^{i_1} \cdots X_n^{i_n} \ (a_{i_1 \cdots i_n} \in R)$$

where the sum is finite.
Expressions of the form $m_{(i)} = X_1^{i_1} \cdots X_n^{i_n}$ are called **monomials**. The example we'll study most deeply is when R is a field.

4.5 Field of Fractions

What is the process by which we go from $(\mathbb{Z}, +, \times)$ to $(\mathbb{Q}, +, \times)$? Intuitively, we are "dividing" through by all non-zero elements. Let us think more carefully about what is actually happening and try to generalize the construction to R an integral domain. What is an element of $\mathbb{Q}$? We usually write it in the form $\frac{a}{b}$ with $a, b \in \mathbb{Z}$, $b \neq 0$. This is **not** unique. $\frac{a}{b} = \frac{c}{d} \iff ad - bc = 0$.
As we are all aware, we define $+$ and $\times$ by the following rules:

1. $\frac{a}{b} + \frac{c}{d} = \frac{ad+cb}{bd}$
2. $\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$

We should therefore think of elements of $\mathbb{Q}$ as pairs of integers (a, b) such that $b \neq 0$, up to an equivalence relation.

$$(a, b) \sim (c, d) \iff ad - cb = 0$$

Hence, $\mathbb{Q}$ can be thought of as $(\mathbb{Z} \times \mathbb{Z} \setminus \{0\}/\sim)$. The well-definedness of $+$ and $\times$ is not obvious and needs checking, i.e. choosing different elements of the same equivalence class should give the same results.

Let us now generalise this construction. Let R be an integral domain. We define the relation on $R \times R \backslash \{0\}$ by:

$$(a, b) \sim (c, d) \iff ad - bc = 0.$$

Proposition. $\sim$ *is an equivalence relation.*

Proof. 1. $(a, b) \sim (a, b)$ as $ab - ab = 0$ since R is commutative.

2. $(a, b) \sim (c, d) \Rightarrow ad - bc = 0 \Rightarrow bc - ad = 0 \Rightarrow (c, d) \sim (a, b)$

3. Let $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. Then $ad - bc = 0, cf - de = 0$. Consider

$$\begin{aligned} (af - be)d &= adf - bed \\ &= f(ad - bc) + b(cf - de) \\ &= f0 + b0 = 0 \end{aligned}$$

$d \neq 0 \Rightarrow af - be = 0 \Rightarrow (a, b) \sim (e, f)$

□

Let us denote the equivalence classes by $(R \times (R \setminus \{0\}))/\sim$. It is convenient to use the usual notation: for $(a, b) \in R \times (R \setminus \{0\})$ we denote the equivalence class containing (a, b) by $\frac{a}{b}$. Let us define multiplication and addition on $R \times R \setminus \{0\}/\sim$ by

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

Proposition. *$+$ and $\times$ are well-defined on $(R \times (R \setminus \{0\}))/\sim$.*

Proof. The first thing to note is that if $b, d \in R \setminus \{0\} \Rightarrow bd \in R \setminus \{0\}$ as R is an integral domain. We just need to check that choosing different representatives gives the same answer. It's just an exercise in keeping the notation in order - you can do it. □

Proposition. *$0 \in (R \times (R \setminus \{0\}))/\sim$ is given by the equivalence class containing $(0, 1)$. $1 \in R \times (R \setminus \{0\})/\sim$ is given by the equivalence class containing $(1, 1)$.*

Proof. For all $(a, b) \in (R \times (R \setminus \{0\}))$,

$$\frac{a}{b} + \frac{0}{1} = \frac{a \times 1 + b \times 0}{b \times 1)} = \frac{a}{b}.$$

$$\frac{a}{b} \times \frac{1}{1} = \frac{a1}{b1} = \frac{a}{b}$$

Both operations are clearly commutative because R is commutative. Hence we are done.

□

It is a straight forward exercise to check that under these operations $(R \times (R \setminus \{0\}))/\sim$ is a commutative ring. Also observe that $(a, b) \in (R \times (R \setminus \{0\}))$ is in the zero class if and only if $a = 0$. Similarly (a, b) give the one class if and only in $a = b$. This is good. It's the same as in $\mathbb{Q}$, so we've done something right.

Theorem. *$(R \times (R \setminus \{0\}))/\sim$ is a field.*

Proof. We just need to check non-zero elements have multiplicative inverses. Let $\frac{a}{b} \in (R \times (R\setminus\{0\}))/\sim$ be non-zero. By the above this implies that $a \neq 0$. Hence $\frac{b}{a} \in (R\times(R\setminus\{0\}))/\sim$. But

$$\frac{a}{b} \times \frac{b}{a} = \frac{ab}{ab} = \frac{1}{1}.$$

Hence we are done. Multiplication:
Let $(a_1, b_1) \sim (a_2, b_2)$ and $(c_1, d_1) \sim (c_2, d_2)$.

□

Definition. *Let R be an integral domain. The field of fractions of R is the field $Frac(R) := (R \times (R \setminus \{0\}))/\sim$.*

The canonical example is $Frac(\mathbb{Z}) = \mathbb{Q}$.

Definition. *Given an integral domain R and indeterminants $\{X_1, \cdots, X_n\}$ we know that $R[X_1, \cdots, X_n]$ is an integral domain. We define*

$$R(X_1, \cdots, X_n) := Frac(R[X_1, \cdots, X_n]).$$

Theorem. *The map*

$$\begin{aligned} \phi : R &\to Frac(R) \\ a &\mapsto \tfrac{a}{1} \end{aligned}$$

is an embedding.

Proof. We need to check that ϕ is a homomorphism first.

1. Given $a, b \in R$, $\phi(a+b) = \frac{a+b}{1} = \frac{a}{1} + \frac{b}{1} = \phi(a) + \phi(b)$.
2. Given $a, b \in R$, $\phi(ab) = \frac{ab}{1} = \frac{a}{1} \times \frac{b}{1} = \phi(a)\phi(b)$.
3. $\phi(1) = \frac{1}{1}$.

To check it is injective we just need to show that the kernel (as a homomorphism of Abelain groups) is trivial.
$\phi(a) = \frac{a}{1} = \frac{0}{1} \iff a = 0$. Thus the kernel is trivial and so ϕ is injective. □

Corollary. *Every integral domain may be embedded in a field.*

Proposition. *Let R be a field. The natural embedding $R \subset Frac(R)$ is an isomorphism.*

Proof. We must show ϕ is surjective. Let ϕ denote the natural embedding $R \subset Frac(R)$. Let $\frac{a}{b} \in Frac(R)$. R is a field so there exist b^{-1}, a multiplicative inverse to b. But $\frac{a}{b} = \frac{ab^{-1}}{1} = \phi(ab^{-1})$. Hence ϕ is surjective. Therefore ϕ is an isomorphism. □

This is backed up by our intuition. Clearly taking fractions of rationals just gives the rationals again.

4.6 Characteristic

Let R be entire (non-tvial with no zero-divisors). Recall that $(R, +)$ is an abelian group, hence given $a \in R$ we may talk about its additive order. Recall that if $a \in R$ does not have finite order, then we say it has infinite order.

Theorem. *In an entire ring R, the additive order of every non-zero element is the same. In addition, if this order is **finite** then it is **prime**.*

Proof. Let $a \in R \setminus \{0\}$ be of finite (additive) order $k > 1$, i.e. k is minimal such that $ka = 0$. This implies $(k \times 1_R)a = 0 \Rightarrow k \times 1_R = 0$ as R is entire and contains no zero-divisors. Therefore if we choose $b \in R \setminus \{0\}$ then $kb = (k \times 1_R)b = 0 \times b = 0 \Rightarrow$ every element has order dividing k. Choosing a with minimal order $k > 1$ ensures that every nonzero element must have order k. If no element has finite order, all elements must have infinite order.
Now assume that $1_R \in R$ has finite order $k > 1$ and that we have factored $k = rs$ in $\mathbb{N}$. Then $k1_R = (rs)1_R = (r1_R)(s1_R) = 0$. Since R entire, either $r1_R = 0$ or $s1_R = 0$. However, since k is the minimal order of 1_R, $r = k$ or $s = k$. Therefore, k must be prime.

□

Definition. *Suppose R an entire ring. R has* ***characteristic zero*** *if all of its non-zero elements have infinite additive order, denoted char(R)=0. If all non-zero elements of R are of additive order $p \in \mathbb{N}$, then R is* ***characteristic p****, or char(R)=p. In this case, R is* ***finite characteristic****.*

When studying abstract fields, the characteristic is very important.
Eg. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all fields (hence entire) of characteristic zero. If p is a prime number $\mathbb{Z}/p\mathbb{Z}$ is a field of characteristic p. We denote this later field by $\mathbb{F}_p$.

Theorem. *There is an embedding of $\mathbb{Q}$ in any field F of characteristic 0.*

Proof. Let 1_F denote the multiplicative identity in F. Let 0_F denote the additive identity in F. We must find a suitable **embedding** of $\mathbb{Q}$ in F. Because $char(F) = 0$ the natural map homomorphism:

$$\begin{gathered}\phi : \mathbb{Z} \to F \\ n \mapsto n1_F\end{gathered}$$

is injective. We claim that it is a homomorphism (of rings). Let $a, b \in \mathbb{Z}$, then $\phi(ab) = ab1_F = ab1_F1_F = a1_Fb1_F = \phi(a)\phi(b)$; $\phi(a + b) = (a + b)1_F = a1_F + b1_F = \phi(a) + \phi(b)$. $\phi(1) = 1_F$. Thus ϕ is an injective homomorphism.
Now we will extend this notion to $\mathbb{Q}$. We define the following map:

$$\begin{gathered}\psi : \mathbb{Q} \to F \\ \tfrac{n}{m} \mapsto \phi(n)\phi(m)^{-1}\end{gathered}$$

We must check that ψ is well defined and is an embedding.
For $a, b, n, m \in \mathbb{Z}$, $\frac{n}{m} = \frac{a}{b} \Rightarrow nb - am = 0$. Therefore

$$\begin{aligned}\phi(nb - am) = \phi(0) = 0_F = \phi(nb) - \phi(am) \quad &\Rightarrow \quad \phi(nb) = \phi(am) \\ &\Rightarrow \quad \phi(n)\phi(b) = \phi(a)(m) \\ &\Rightarrow \quad \phi(n)\phi(m)^{-1} = \phi(a)\phi(b)^{-1} \\ &\Rightarrow \quad \psi(\frac{n}{m}) = \psi(\frac{a}{b})\end{aligned}$$

This shows that ψ is well defined.
Next: ψ is a homomorphism.

$$\begin{aligned}\psi(\frac{a}{b} + \frac{n}{m}) &= \psi(\frac{am + bn}{bm}) \\ &= (\phi(a)\phi(m) + \phi(b)\phi(n))\phi(bm)^{-1} \\ &= \phi(a)\phi(b)^{-1} + \phi(n)\phi(m)^{-1} \\ &= \psi(\frac{a}{b}) + \psi(\frac{n}{m})\end{aligned}$$

$$\begin{aligned}\psi(\frac{a}{b}\frac{n}{m}) &= \psi(\frac{an}{bm}) \\ &= \phi(an)\phi(bm)^{-1} \\ &= \phi(a)\phi(n)\phi(b)^{-1}\phi(m)^{-1} \\ &= \phi(a)\phi(b)^{-1}\phi(n)\phi(m)^{-1} \\ &= \psi(\frac{a}{b})\psi(\frac{n}{m})\end{aligned}$$

By definition $\psi(\frac{1}{1}) = 1_F$. Thus we have a homomorphism. We claim that it is injective. We must show that the kernel (as a homomorphism of Abelian groups) is trivial. Let $\frac{n}{m} \in \mathbb{Q}$ such that $\psi(\frac{n}{m}) = 0$. Then $\phi(n)\phi(m)^{-1} = 0 \Rightarrow \phi(n) = 0 \Rightarrow n = 0$ as ϕ was already shown to be injective. Therefore the kernel is trivial, so ψ is an embedding. □

Theorem. *Let p be a prime number. There is an embedding of $\mathbb{F}_p$ in ant field F of characteristic p.*

Proof. Note that $\{0_F, 1_F, \cdots, (p-1)1_F\} \subseteq F$ is closed under $+$ and $\times$, hence forms a subring. Clearly $\mathbb{F}_p$ is isomorphic to this subring under the embedding

$$\begin{aligned}\psi : \mathbb{F}_p &\longrightarrow F \\ [\mathrm{a}] &\longrightarrow a1_F\end{aligned}$$

□

4.7 Ring Extensions

let R be a subring of S. Recall that this means R is a subgroup under addition, is closed under multiplication and contains 1_S.

Definition. *The ring extension of R by $\{\alpha_1, \cdots, \alpha_n\} \subset S$ is the subring*

$$R[\alpha_1, \cdots, \alpha_n] = \{f(\alpha_1, \cdots, \alpha_n) | f \in R[X_1, \cdots, X_n]\}$$

This is the intersection of all subrings containing R and the subset $\{\alpha_1, \cdots, \alpha_n\}$.

4.8 Principal, Prime and Maximal Ideals

Definition. *An ideal $I \subset R$ is proper if $I \neq R$.*

Note that $I \subset R$ is proper if and only if R/I is a non-trivial ring.

Definition. *Let R be a commutative ring. We say an ideal $I \subset R$ is principal if there exist $a \in R$ such that $I = \{ra | r \in R\}$. In this case we write $I = (a)$.*

Definition. *Let R be a commutative ring. We say an ideal $I \subset R$ is prime if it is proper and given $a, b \in R$ such that $ab \in I$ then either $a \in I$ or $b \in I$.*

Proposition. *Let R be a commutative ring. Let $I \subset R$ be an ideal. Then I is prime if and only if R/I is an integral domain.*

Proof. I is a proper ideal hence R/I is non-trivial.

Observe that R commutative trivially implies that R/I is commutative. Let $I \subset R$ be prime and assume that R/I has zero divisors. Then there exists $a, b \in R$ such that $a, b \notin I$ but $(a + I)(b + I) = 0 + I$. But this trivially implies that $ab \in I$. But this contradicts the fact that I is prime.

Assume that R/I is an integral domain but I is not prime. Hence we can find $a, b \in R$ such that $ab \in I$ but $a, b \notin I$. But then $(a + I)$ and $(b + I)$ are zero divisors, which is a contradiction.

□

Definition. *Let R be a commutative ring. We say that an ideal is maximal if it is maximal among the set of proper ideals. More precisely $I \subset R$ is a maximal ideal if given an ideal $J \subset R$ such that $I \subset J$, then either $I = J$ or $J = R$.*

Proposition. *Let R be a commutative ring. Let $I \subset R$ be an ideal. Then I is maximal if and only if R/I is a field.*

Proof. First observe that R commutative trivially implies that R/I is commutative.

Assume that $I \subset R$ is maximal. Take a non-zero element of R/I, i.e. $a + I$ for $a \notin I$. Consider the ideal $(a) \subset R$. Consider the following new ideal:

$$(a) + I = \{ra + b | r \in R, \ \ b \in I\}.$$

Note that this is certainly an ideal because it is closed under addition and scalar multiplication by all R. Note that by construction $I \subset (a) + I$ and $a \in (a) + I$. Hence I is strictly contained in $(a) + I$. But I is maximal. Hence $(a) + I = R$. Thus there exist $r \in R$ and $b \in I$ such that $ra + b = 1$. Hence $(r + I)(a + I) = ra + I = 1 + I$. Thus $(a + I)$ has a multiplicative inverse. Hence R/I is a field.

Assume that R/I is a field. Assume that J is a proper ideal of R which strictly contains I, i.e. I is not maximal. Let $a \in J$ and $a \notin I$. Thus $(a + I)$ is non-zero in R/I. Thus it has a multiplicative inverse. Hence there exists $b \in R$ such that $ab + I = 1 + I$. This implies that $ab - 1 \in I$, which in turn implies that $ab - 1 \in J$. But $a \in J$, hence $1 \in J$, which implies that $J = R$. This is a contradiction. Hence I is maximal. □

Corollary. *Let R be a commutative ring. Let $I \subset R$ be an ideal. Then I maximal implies that I is prime.*

Proof. I maximal $\Rightarrow$ R/I is a field $\Rightarrow$ R/I is an integral domain $\Rightarrow$ I prime. □

4.9 Factorisation in Integral Domains

Let R be a ring. In $\mathbb{Z}$ we have the "Fundamental Theorem of Arithmetic" - every non-zero element of $\mathbb{Z}$ is ± 1 times a unique product of prime numbers. Does something analogous hold for R? Clearly, if R is **not** commutative or has zero-divisors the issue is very subtle. Hence we will resrict to the case when R is an integral domain.
The first issue to address is what does a prime element of R mean? The problem, as we will see, is that we can easily come up with several different natural definitions which are equivalent in $\mathbb{Z}$, but in R may not be.

Let $a, b \in R$. As in $\mathbb{Z}$, $a|b$ will mean that $\exists c \in R$ such that $b = ac$.

Definition. *Two non-zero elements a, b in an integral domain R are associated if $a|b$ and $b|a$, i.e. $\exists c, d \in R$ such that $b = ac$ and $a = bd$.*

Theorem. *In R an integral domain, and $a, b \in R$ be two non-zero elements. Then, a and b are associated $\iff$ $a = bu$ for $u \in R^*$*

Proof. Association of a and $b \Rightarrow a|b$ and $b|a \Rightarrow \exists c, d \in R$ such that $a = bd$ and $b = ac \Rightarrow a = acd \Rightarrow a = 0$ or $cd = 1$. If $a = 0 \Rightarrow b = 0$, which is not true by assumption. Thus we have $cd = 1 \Rightarrow c, d$ are inverses of each other and thus units. □

Theorem. *Let R be an integral domain with $a, b \in R$. Then $(a) \subset (b) \iff b|a$. Hence a and b are associated if and only if $(a) = (b)$.*

In $\mathbb{Z}$, m and n are associated if and only if $n = \pm m$.

Definition. *We call $a \in R \backslash \{0\}$ an **irreducible element** if it is a non-unit and is NOT the product of two non-units.*

If a is irreducible then so are all its associates. In $\mathbb{Z}$, m is irreducible if and only if it is ± 1 times a prime. The FTOA says that every $m \in \mathbb{Z}$ can be factored into irreducible elements in "essentially" one way. Here, essentially means up to switching irreducibles for *associated* irreducibles, i.e. $10 = 2 \times 5 = (-2) \times (-5)$. This motivates the important definition:

Definition. *A **unique factorization domain** (UFD)is an integral domain in which every element NOT zero or a unit can be written as the product of **irreducibles**. Moreover, given 2 complete factorizations of the same element*

$$X = a_1 \cdots a_n = b_1 \cdots b_m,$$

into irreducibles, $n = m$ and after renumbering a_i is associated to b_i for all $i \in \{1, \cdots, n\}$.

Clearly $\mathbb{Z}$ is a UFD by the Fundamental Theorem of Artithmetic. A natural question to ask is whether all integral domains are UFDs. The answer, rather surprisingly, is no.
Let R be a UFD. Many of the properties of $\mathbb{Z}$ carry over to R. For example we can talk about highest common factor (HCF) and least common multiple (LCM) for two $a, b \in R \setminus \{0\}$.

Definition. *Given $a, b \in R \setminus \{0\}$ a highest common factor of a and b is element $d \in R$ such that*

1. *$d|a$ and $d|b$*
2. *Given $d' \in R$ such that $d'|a$ and $d'|b$, then $d'|d$.*

Definition. *Given $a, b \in R \setminus \{0\}$ a lowest common multiplie of $a, b \in R$ is an element $c \in R$ such that*

1. *$a|c$ and $b|c$*
2. *Given $c' \in R$ such that $a|c'$ and $b|c'$, then $c|c'$.*

Remarks.
1. *It should be observed that there is no reason to believe that HCFs and LCMs exist in an arbitrary integral domain. Indeed it is not true in general.*
2. *Clearly a HCF (if it exists) is NOT unique: If d is an HCF of a and b then so is d' for d' associated to d. Similarly for LCM. Hence when we talk about the HCF or LCM of two elements we must understand they are well defined only up to association.*

Theorem. *In a UFD any two non-zero elements have a HCF. Moreover, if $a = up_1^{\alpha_1} \cdots p_r^{\alpha_r}$ and $b = vp_1^{\beta_1} \cdots p_r^{\beta_r}$ where u, v are units, and the p_i are pairwise non-associated irreducible elements, then $HCF(a, b) = p_1^{\gamma_1} \cdots p_r^{\gamma_r}$ where $\gamma_i = min(\alpha_i, \beta_i)$.*

Proof. Let d be a common factor of a and b. By the uniqueness of complete factorisation we know that (up to association) d is a product of p_i for $i \in \{1, \cdots p_r\}$. Without loss of generality we may therefore assume that $d = \prod_{i=1}^{r} p_i^{\delta_i}$. Again by the uniqueness of complete factorisation d is a common factor of a and $b \iff \delta_i \leq \alpha_i$ and $\delta_i \leq \beta_i \forall i$. Therefore, $\delta_i \leq \gamma_i \Rightarrow HCF(a, b) = p_1^{\gamma_1} \cdots p_r^{\gamma_r}$. □

Proposition. *In a UFD any two non-zero elements have a LCM. Moreover, if $a = up_1^{\alpha_1} \cdots p_r^{\alpha_r}$ and $b = vp_1^{\beta_1} \cdots p_r^{\beta_r}$ where u, v are units, and the p_i are pairwise non-associated irreducible elements, then $LCM(a, b) = p_1^{\gamma_1} \cdots p_r^{\gamma_r}$ where $\gamma_i = max(\alpha_i, \beta_i)$.*

Proof. Exactly the same argument as above works in this case observing that $d = \prod_{i=1}^{r} p_i^{\delta_i}$ is a common multiple of a and b if and only if $\delta_i \geq \alpha_i$ and $\delta_i \geq \beta_i$ for all $i \in \{1, \cdots p_r\}$. □

Remarks. *If $a \in R$ a unit then*

$$HCF(a, b) = 1,\ LCM(a, b) = b\ \forall b \in R \setminus \{0\}$$

Even if we know that R is a UFD, there is no easy way to completely factor any element. This is clearly apparent in $\mathbb{Z}$. Fortunately for certain rings there is a faster way to determine the HCF of two elements.

Definition. *If R is an integral domain, R is **Euclidean** if it admits a function $\varphi : R \setminus \{0\} \to \mathbb{N} \cup \{0\}$ such that*

1. $\varphi(ab) \geq \varphi(a) \forall a, b \in R \setminus \{0\}$
2. *For any* $a, b \in R$, *if* $b \neq 0$, *then* $\exists q, r \in R$ *such that* $a = bq + r$ *where either* $r = 0$ *or* $\varphi(r) < \varphi(b)$.

Remarks. 1. *This is intended to model the behavior of the function*

$$\varphi : \mathbb{Z} \setminus \{0\} \to \mathbb{N} \cup \{0\} \; a \to |a|$$

The second property is just a generalization of the remainder theorem on $\mathbb{Z}$. *Hence we see that* $\mathbb{Z}$ *is Euclidean.*

2. *We include* 0 *in the codomain as this enlarges the collection of rings under consideration.*

Lemma. *The second axiom of a Euclidean Ring is equivalent to the following:*

(2'): $\forall a, b \in R \setminus \{0\}$; *if* $\varphi(a) \geq \varphi(b)$ *then* $\exists c \in R$ *such that either* $a = bc$ *or* $\varphi(a - bc) < \varphi(a)$ *or* $a = bc$.

Proof. $(2) \Rightarrow (2')$.
Suppose (2) holds. Then we have $\varphi(a) \geq \varphi(b) \Rightarrow \exists q, r$ such that $a = qb + r$ where where either $r = 0$ or $\varphi(r) < \varphi(b)$. If $r = 0 \Rightarrow q = c$ and we are done . Otherwise $\varphi(r) = \varphi(a - qb) < \varphi(b) \leq \varphi(a)$ so we are done with $c = q$.
$(2') \Rightarrow (2)$
Given $a, b \in R \setminus \{0\}$ if $b|a$ we are done. Therefore, assume $b \nmid a$. Hence $a - bq \neq 0 \; \forall q \in R$. Choose $c \in R$ such that $\varphi(a - bq)$ is minimal. Note that by assumption $b \nmid (a - bq)$. If $\varphi(a - bq) \geq \varphi(b) \Rightarrow \exists c \in R$ such that $\varphi(a - bq - bc) < \varphi(a - bq)$. This is a contradiction by the minimality condition. Therefore $\varphi(a - bq) < \varphi(b)$, i.e. setting $r = a - bq$ we have

$$a = bq + r \text{ with } \varphi(r) < \varphi(b)$$

hence (2) holds.

□

Theorem. F *field* $\Rightarrow F[X]$ *Euclidean*

Proof. Define:

$$\begin{array}{rcl} \varphi : F[X] \setminus \{0\} & \longrightarrow & \mathbb{N} \cup \{0\} \\ f & \longrightarrow & deg(f) \end{array}$$

Property (1):
As F is a field, $F[X]$ is an integral domain $\Rightarrow deg(fg) = deg(f) + deg(g) \geq deg(f) \forall g, f \in F[X] \setminus \{0\} \Rightarrow \varphi(fg) \geq \varphi(f) \forall f, g \in F[X] \setminus \{0\}$.
Property $(2')$

Let $f = a_0 + a_1X + \cdots + a_nX^n$, $g = b_0 + b_1X + \cdots + b_mX^m$ where $a_i, b_j \in F, n, m \in \mathbb{N} \cup \{0\}$, and $a_n \neq 0, b_m \neq 0$.
Assume $\varphi(f) \geq \varphi(g) \Rightarrow n \geq m \Rightarrow n - m \geq 0 \Rightarrow X^{n-m} \in F[X] \Rightarrow X^{n-m}b_m^{-1}a_ng$ has leading term $a_nX^n \Rightarrow \deg(f - X^{n-m}b_m^{-1}a_ng) < \deg(f)$.
Hence setting $c = a_nb_m^{-1}X^{n-m}$ we have $\varphi(f - cg) = \deg(f - cg) < \deg(f) = \varphi(f)$. Therefore, Property $(2')$ is satisfied. □

Remarks. *Note that to get this proof to work we need $b_m \neq 0$ to have an inverse. This critically relied on F being a field. If we relax this condition we will not necessarily get a Euclidean Domain.*

This shows that despite the fact that $\mathbb{Z}$ and $F[X]$ (F a field) are very different rings they share an important property. Euclidean domains have many pleasant properties.

Theorem. *Let R be Euclidean, with Euclidean function φ. Any two $a, b \in R$ have an HCF, (a, b). Moreover, it can be expressed in the form $(a, b) = au + bv$ where $u, v \in R$.*

Proof. Without loss of generality assume that $\varphi(a) \geq \varphi(b)$. Apply property (2) to get

$$a = bq_1 + r_1,$$

where either $r_1 = 0$ or $\varphi(r_1) < \varphi(b)$. If $r_1 = 0$ then we know that $HCF(a, b) = b$ and we are done setting $u = 0$ and $v = 1$. If not then applying property (2) again we get

$$b = q_2r_1 + r_2,$$

where either $r_2 = 0$ or $\varphi(r_2) < \varphi(r_1)$. If $r_2 = 0$ stop. If not continue the algorithm. We claim that after a finite number of steps this process must terminate with the remainder reaching zero. To see this observe that we have a strictly decreasing sequence

$$\varphi(b) > \varphi(r_1) > \varphi(r_2) \cdots$$

in $\mathbb{N} \cup \{0\}$. Hence it must have finite length so the algorithm must terminate. Assume it terminates at the n^{th} stage, i.e. $r_{n+1} = 0$. We claim that r_n can be written in form $ua + vb$ for some $u, v \in R$. We do it by induction on n. If we set $r_0 = b$ then the result is true for $r_0 and r_1$. Assume it is true for r_{i-1} and r_{i-2}. By definition $r_i = -q_ir_{i-1} + r_{i-2}$. hence the result must be true for r_i. Hence by induction we know that we may write r_n in the form $ua + vb$.

Now we claim that r_n must divide both a and b. By construction $r_n | r_{n-1} \rightarrow r_n\ r_{n-2}$. Inductively $r_n | r_i$ for all i. In particular $r_n | b$ and $r_n | r_1 \Rightarrow r_n | a$ Hence r_n is a common divisor of both a and b. Let $d \in R$ such that $d | a$ and $d | b$. Hence $d | (ua + vb) \Rightarrow d | r_n$. Hence $HCF(a, b) = r_n = ua + vb$. □

Remarks. *This procedure is known as the Euclidean Algorithm.*

Corollary. *Let R be Euclidean ring. Then any $a, b \in R \setminus \{0\}$ have a LCM.*

Proof. By the above $HCF(a,b) = au + bv$ for $u, v \in R$. We will define $m = \frac{ab}{HCF(a,b)}$. Note that this makes sense as $HCF(a,b)|a$. It is clear that $a|m$ and $b|m$. Let m' be a common multiple, i.e. $a|m'$, $b|m'$. Then $ab|bm'$ and $ab|am' \Rightarrow ab|aum' + bvm' \Rightarrow ab|(au+bv)m' \Rightarrow ab|(a,b)m' \Rightarrow HCF(a,b)m|HCF(a,b)m'$. Because a and b are non-zero $HCF(a,b)$ is non-zero. Because R is an integral domain we can cancel resulting in $m|m'$. Therefore m is an LCM of a, b. □

It is worth mentioning that as of yet we have only shown Euclidean rings admit HCFs and LCMs. We do not yet know if they are UFDs.

4.10 Principal Ideal Domains

Definition. *Let R be an integral domain. We say that a R is a **principal ideal domain** is every ideal of R is principal. More precisely, if $I \subset R$ is an ideal then there exists $a \in I$ such that $I = (a)$. We write PID for short.*

Theorem. *R Euclidean $\Rightarrow$ R a PID.*

Proof. . Let $I \subset R$ be an ideal. If I is the zero ideal then $I = (0)$. Assume that I is not the zero ideal. Choose $a \in I$ such that $\phi(a) \leq \phi(b)$ for all $b \in I$. We aim to prove that $I = (a)$. Assume this is not the case. Hence there exists $r \in I$ such that $r \notin (a)$. This means that a does not divide r. Hence by the Euclidean property there exist $q, s \in R$ such that $r = qa + s$ where $\phi(s) < \phi(a)$. However, $s = r - qa \in I$. This contradicts the minimality of $\phi(a)$. Thus no such r exists and $I = (a)$. □

Definition. *Let R be an integral domain and $I_1, I_2, I_3, \cdots$ be a sequence of ideals. Assume that*

$$I_1 \subset I_2 \subset I_3 \subset \cdots$$

*We call this an **ascending chain of ideals**. We say that it is **stationary** if there exists some $n \in \mathbb{N}$ such that $I_n = I_m$ for all $m \geq n$.*

Theorem. *If R is a PID then every ascending chain of ideals is stationary.*

Proof. Let

$$I_1 \subset I_2 \subset I_3 \subset \cdots$$

be an ascending chain of ideals in R. Let I be the union of all the I_i. We claim that this is an ideal. Observe that $0 \in I$ as it is contained in each I_i. Similarly $r \in I \Rightarrow r \in I_i$ for some $i \Rightarrow -r \in I_i \Rightarrow -r \in I$. Let $r, s \in I$. Hence $r \in I_i$ and $s \in I_j$ for some i and j. Without loss of generality assume that $i \leq j$. Hence $r, s \in I_j \Rightarrow r + s \in I_j \Rightarrow r + s \in I$. Hence I is a subgroup under addition.

If $r \in I$ then $r \in I_i$ for some i. Thus given any $a \in R$, $ar \in I_i \subset I$. We deduce that I is an ideal.

Because R is a PID there exists $b \in I$ such that $I = (b)$. This means that $b \in I_n$ for some n. Hence $(b) \subset I_n$. Hence we have $I \subset I_n$ and $I_n \subset I$ implying that $I_n = I$. This implies that $I_m = I_n$ for all $m \geq n$.

□

Theorem. *If R is a PID then every non-zero non-units can be factored into irreducible elements.*

Proof. We will begin by showing that every non-zero, non-unit admits an irreducible factor.

Let $a \in R$ be a non-zero, non-unit. If a is irreducible we are done. Assume, therefore that $a = b_1 a_1$, where b_1 and a_1 are non-units. This implies that

$$(a) \subset (a_1)$$

Note that because b_1 is a non-unit a and a_1 are not associated by the cancellation law. Hence this is a strict inclusion. If a_1 is irreducible we are done. If not then we can repeat this process with a_1. This would give a factorization $a_1 = b_2 a_2$, where b_2 and a_2 are non-units. Thus we again get a strict inclusion

$$(a_1) \subset (a_2).$$

If a_2 is irreducible we are done. If not we can repeat the process. This builds an ascending chain of ideals. Because R is a PID we know that this ascending chain must be stationary. This can only happen if we eventually get an irreducible factor. We deduce that a must admit an irreducible factor.

Now we show that a is the product of a finite number of irreducible elements of R. If a is not irreducible then by the above we can write $a = p_1 c_1$ where p_1 is irreducible and c_1 is not a unit. Thus (a) is strictly contained in the ideal $(c1)$. If c_1 is irreducible we are done. If c_1 is not irreducible then $c_1 = p_2 c_2$ where p_2 is irreducible and c_2 is not a unit. We can build a strictly ascending chain of ideals :

$$(a) \subset (c_1) \subset (c_2) \subset \cdots$$

Because R is a PID we know that this chain is stationary, which means eventually c_r must be an irreducible. Hence $a = p_1 p_2 \cdots p_r c_r$. □

Let us now introduce another natural generalisation of prime number to an arbitrary integral domain.

Definition. *Let R be an integral domain. We say that $p \in R$ is a **prime element** if:*

1. *$p \notin R^*$ and $p \neq 0$*
2. *$\forall a, b \in R, p|ab \Rightarrow p|a$ or $p|b$*

Remarks. *1. In $\mathbb{Z}$ prime elements are the prime numbers and their negatives.*

2. *All elements associated to a prime are themselves prime.*

Proposition. *Let R be an integral domain, $p \in R$. Then p prime $\Rightarrow$ p irreducible.*

Proof. Let $p \in R$ be prime and $p = ab$ for some $a, b \in R$. Then $p|a$ or $p|b$. Say $p|a \Rightarrow a = pc = abc$ for some $c \in R$. Note that $a \neq 0 (p \neq 0)$, therefore by the cancellation law, $1 = bc \Rightarrow b$ is a unit. Hence p is irreducible.

□

We shall see that for a general integral domain the converse does not always hold. However in the case of PIDs we have the following:

Theorem. *Let R be a PID and $p \in R$. Then p irreducible $\iff$ p prime.*

Proof. By the previous proposition we only need to show that p irreducible $\Rightarrow$ p prime. We will begin by showing that if p is irreducible then $(p) \subset R$ is maximal.

First observe that p is not a unit. Hence (p) is a proper ideal of R. Assume now that there exists $I \subset R$ a proper ideal such that $(p) \subset I$. Because R is a PID, there exists $a \in I$ such that $(a) = I$. Note that a is no a unit. Hence $(p) \subset (a)$ and we deduce that $p = ab$ for some $b \in R$. Observe that because p is irreducible b must be a unit. Hence p and a are associated implying that $I = (a) = (p)$. We deduce that (p) is maximal.

Observe now that $R/(p)$ is a field. Hence $R/(p)$ is an integral domain implying that (p) is a prime ideal.

Now assume that $p|ab$ for some $a, b \in R$. Hence $ab \in (p)$. Because (p) is prime either $a \in (p)$ or $b \in (p)$. These imply that $p|a$ or $p|b$ respectively. Hence p is prime. □

Theorem. *An integral domain is a UFD iff*

1. *Every $a \in R$ such that $a \neq 0$ and $a \notin R^*$ can be factored into irreducibles (has a complete factorization)*

2. *Every irreducible element is prime*

Proof. First suppose R is a UFD. Then, by definition, (1) holds. Suppose $p_1 \in R$ irreducible. Then suppose $a, b \in R$ such that $p_1|ab$. If $a = 0$, $p_1|a$ trivially, so we will assume $a, b \neq 0$. R UFD means we can uniquely factor a, b

$$a = up_1^{\alpha_1} \cdots p_r^{\alpha_r}, b = vp_1^{\beta_1} p_2^{\beta_2} \cdots p_r^{\beta_r},$$

where u, v are units, $\alpha_i, \beta_i \in \mathbb{N} \cup \{0\}$. and the p_i are pairwise non-associated irreducible elements. It follows that ab can be factored into $uvp_1^{\alpha_1+\beta_1} \cdots p_r^{\alpha_r+\beta_r}$. Suppose $p_1|ab$, then by the uniqueness of factorization present in a UFD, this forces $(\alpha_1 + \beta_1) > 0 \Rightarrow \alpha_1$ or $\beta_1 > 0 \Rightarrow p|a$ or $p|b$. Therefore p_1 is prime.

Conversely, suppose R is an integral domain and (1) and (2) hold. Then we need to show that every non-zero, non-unit has a unique factorization into irreducibles, and the factorization is unique up to association. Let $c \in R$ such that $c \neq 0$ and $c \notin R^*$. By (1) we know we can factor into irreducibles. So let us consider two factorizations of c.

$$c = a_1 \cdots a_r, c = b_1 \cdots b_s$$

We must show $r = s$ and each b_i associated to a_i after renumbering. Let us use induction on r. $r = 1 \Rightarrow a_1 = b_1 \cdots b_s \Rightarrow b_1 | a_1 \Rightarrow a_1 = b_1 u, u \in R^*$. Then if $s > 1$, we cancel to get $u = b_2 \cdots b_s \Rightarrow b_2 \in R^*$ which is a contradiction since b_2 is an irreducible by assumption. Therefore $s = 1$ and we are done.
Let $r > 1$. By hypothesis (ii) a_1 is prime and $a_1 | b_1 \cdots b_s \Rightarrow a_1 | b_j$ for some j. WLOG assume $j = 1$. b_1 is irreducible and $b_1 = a_1 u \Rightarrow u \in R^* \Rightarrow b_1 u^{-1} = a_1 \Rightarrow b_1 | a_1 \Rightarrow$ a_1 and b_1 are associated.
By the cancellation property, we have

$$u^{-1} a_2 \cdots a_r = b_2 \cdots b_s$$

$u^{-1} a_2$ is irreducible and hence this gives a complete factorization of the same element. By induction, $r - 1 = s - 1 \Rightarrow r = s$ and we can renumber such that a_i is associated to $b_i \forall i \in \{2, \cdots, r\}$. We've just seen this holds for $i = 1$, hence R is a UFD. □

Theorem. *Every PID is a UFD.*

Proof. In a PID every non-zero non-unit can be factored into irreducibles. In addition every irreducible is prime. This a PID is a UFD.

□

Theorem. *Every Euclidean ring is a UFD*

Proof. . Every Euclidean ring is a PID. Every PID is a UFD. Hence every Euclidean ring is a UFD. □

4.11 Factorization in Polynomial Rings

Theorem. *For any field F, the polynomial ring $F[X]$ is a UFD.*

Proof. F field $\Rightarrow F[X]$ Euclidean $\Rightarrow F[X]$ is a UFD. □

From now on fix F a field. Let us return to trying to understand factorization in the polynomial ring $F[X]$.

Our first task is to determine the irreducible elements in $F[X]$.

Proposition. *$F[X]^* = F^*$, where we view $F \subset F[X]$ as the degree zero polynomials (the constant polynomials).*

Proof. The unit element in $F[X]$ is $1 \in F \subset F[X]$. If $f \in F[X]$ and $def(f) > 0$ then $deg(fg) > 0 \ \forall g \in F[X] \setminus \{0\}$. Thus all invertible elements of $F[X]$ must be degree zero, so constant polynomials. Because F is a field we deduce that $F[X]^* = F^*$. □

Definition. *We call $f \in F[X]$ such that deg(f)=1* ***linear****.*

Clearly every linear polynomial must be irreducible for reasons of degree. Here is a partial converse:

Theorem. *Given F field, the only irreducible elements of $F[X]$ are linear iff every positive degree polynomial has a zero (or root) in F*

Proof. $\Rightarrow$
Assume every irreducible in $F[X]$ is linear. Then take $f \in F[X]; deg(f) > 0$. As $F[X]$ is a UFD (since F is a field), we can factor f into linear factors. Choose $ax + b \in F[X]$ to be one such factor, $a \neq 0$. Choose $x = \frac{-b}{a}$ to be a root of f.
$\Leftarrow$
Suppose every positive degree polynomial has a root in F. Then take $p \in F[X]$ to be irreducible, $deg(p) > 0$. By our assumption, there must exist $\alpha \in F$ such that $p(\alpha) = 0$. Since F is a field, we know that $F[X]$ is Euclidean. Hence we know that $(x - \alpha)|p$. To see why let us apply property (2) of the Euclidean degree function. If $(x - \alpha)$ did not divide p then we know that there exists $q, r \in F[X]$ such that $p = q(x - a) + r$ where $r \neq 0$ and $deg(r) < deg(x - \alpha) \Rightarrow deg(r) < 1 \Rightarrow r$ is a constant. If $r \neq 0$, then $p(\alpha) \neq 0$, so $(x - \alpha)|p$. We deduce that $\exists c \in F[X]$ such that $p = (x - \alpha)c$ but since p is irreducible, c must be a unit, i.e. $c \in F^*$. Thus p is linear. □

Note that this is a property of the field F. It is not always true. For example if $F = \mathbb{Q}$, then $X^2 + 1$ does not have a root in $\mathbb{Q}[X]$ and consequently cannot be reducible. Don't let this example mislead you: there are reducible polynominals in $\mathbb{Q}[X]$ which do not have a root in $\mathbb{Q}$. For example $(X^2 + 1)(X^2 + 1)$.

Definition. *Given F a field, we call F* ***algebraically closed*** *if every $f \in F[X]$ such that $deg(f) > 0$ has a root in F.*

Remarks. *1. By the above theorem, F algebraically closed $\iff$ Any $f \in F[X]$ such that $f \notin F[X]^*, f \neq 0$ can be factored into linear terms.*

2. The ***Fundamental Theorem of Algebra*** *says that $\mathbb{C}$ is algebraically closed. $\mathbb{C}$ is defined analytically so it is unsurprising that all proofs rely on some form of analysis. $\mathbb{R}$ and $\mathbb{Q}$ are not algebraically closed as the above example demonstrates. Gauss gave about four proofs of this fact in his PhD thesis at age 21!*

It is important to realize how miraculous it is that $\mathbb{C}$ is algebraically closed. $\mathbb{C}$ is formed from $\mathbb{R}$ by jumping only one dimension. We'll see later that this almost never occurs in general.

Fact: Every field can be embedded in an algebraically closed field.

For example both $\mathbb{Q}$ and $\mathbb{R}$ naturally embed in $\mathbb{C}$. This tells us that something analogous is true even for more exotic fields like $\mathbb{F}_p$.

Proposition. *If $f \in \mathbb{R}[X]$ is irreducible then it is either linear or quadratic (degree 2).*

Proof. Let $f \in \mathbb{R}[X]$ be irreducible. Note that we may naturally consider f as being in $\mathbb{C}[X]$. Hence we may factor f as follows.

$$f = a\prod_i (x - \alpha_i),$$

where $a \in \mathbb{C}^*$ and $\alpha_i \in \mathbb{C}\ \forall i$. By the uniqueness of this factorisation we know that a is unique and the α_i are unique up to reordering. Because $f \in \mathbb{R}[X]$ we also know that $a \in \mathbb{R}$. Because $f \in \mathbb{R}[X]$, taking complex conjugation gives two linear factorisations :

$$f = a\prod_i (x - \alpha_i) = a\prod_i (x - \bar{\alpha_i}),$$

where $\bar{\alpha_i}$ denotes complex conjugation. Observe that two monic linear polynomials in $\mathbb{C}[X]$ are associated if and only if they are equal. Therefore, by uniqueness of irreducible factorisation we know that either $\alpha_i \in \mathbb{R}$ or they occur in complex conjugate pairs. Note that for any $\alpha \in \mathbb{C}$, $(x - \alpha)(x - \bar{\alpha}) \in \mathbb{R}[X]$. Hence f be written as the product of linear and quadratic real polynomials. Hence either f is linear or quadratic. □

What about other fields? The most natural place to start is $F = \mathbb{Q}$. A naive belief would be that because $\mathbb{Q}$ is relatively simple, $\mathbb{Q}[X]$ is easy to understand. You could not be further from the truth. To see this for $\mathbb{Q}$, observe that we have linked the issue of factorisation in $\mathbb{Q}[X]$ to finding rational roots of positive degree polynomials. As you are no doubt aware this second problem can be very difficult and subtle to understand. The point of departure for algebraic number theory (the algebraic study of $\mathbb{Q}$) is trying to determine the structure of $\mathbb{Q}[X]$.

Recall that $\mathbb{Q} = Frac(\mathbb{Z})$. Hence there is a natural inclusion $\mathbb{Z}[X] \subset \mathbb{Q}[X]$. Let us address the problem of factorisation in $\mathbb{Z}[X]$ first. The fundamental theorem of arithmetic says that $\mathbb{Z}$ is a UFD. Thus let R be a UFD and consider $R[X]$. It is a fact that $R[X]$ is again a UFD. I'll get you to prove this in the homework.

Definition. *$f \in R[X] \setminus \{0\}$ is **primitive** if $deg(f) > 0$ and its coefficients do not have an irreducible common factor.*
e.g. $R = \mathbb{Z}, f = 5x^3 + 3x^2 + 10$

Gauss' Lemma. *Let R be a UFD. The product of two primitive polynomials in $R[X]$ is again primitive.*

Proof. Let $f, g \in R[X]$ be primitive. Thus $f = \sum a_i x^i, g = \sum b_j x^j$ for $a_i, b_j \in R$. Because R is an integral domain, so is $R[X]$. Thus $fg \neq 0$. Assume that fg is not primitive. Thus $\exists \pi \in R$ irreducible and $h \in R[X]$ such that $fg = \pi h$. Because f and g are primitive π does not divide all the a_i and b_j. Choose r and s minimal such that π does not divide a_r and b_s. Let $h = \sum c_k x^k$.

Thus

$$\pi c_{r+s} = a_0 b_{r+s} + \cdots + a_r b_s + \cdots + a_{r+s} b_0 \quad \Rightarrow \quad a_r b_s = \lambda c_{r+s} - a_0 b_{r+s} - \cdots - a_{r+s} b_0$$

By the minimality of r and s we deduce that π divides every term in the sum on the right. Hence π divides $a_r b_s$. But R is a UFD, which implies that π is prime. Thus π must divide either a_r or b_s. This is a contradiction. Hence fg is primitive. □

This is a fantastic proof - it's got Gauss written all over it! It has some profound consequences as we'll see in a moment.

Definition. *Let R be a UFD and $f \in R[X] \setminus \{0\}$. The **content** of f is the HCF of its coefficients, i.e. If $f = \sigma g$ where $\sigma \in R$ and g primitive, σ is the content of f. e.g. $R = \mathbb{Z}, f = 9x^3 + 3x + 18$, the content of f is 3.*

Observe that because R is a UFD the content of $f \in R[X] \setminus \{0\}$ always exists. Also observe that the content is only unique up to association.

Proposition. *Let R be a UFD. Suppose $f, g \in R[X] \setminus \{0\}$ with contents $\alpha, \beta \in R$ respectively. Then the content of fg is $\alpha\beta$.*

Proof. $f = \alpha f_1, g = \beta g_1 \Rightarrow fg = (\alpha\beta) f_1 g_1$. By Gauss' Lemma, $f_1 g_1$ is also primitive so $\alpha\beta$ is the content of fg. □

The following theorem illustrates the real meaning of Gauss' Lemma.

Theorem. *Let R be a UFD, and $F = Frac(R)$. Choose $f \in R[X] \subset F[X]$. Then f is irreducible in $R[X] \Rightarrow f$ is irreducible in $F[X]$.*

Proof. Assume $f \in R[X]$ can be factored into non-units in $F[X]$. This implies that $f = gh$ for some $g, h \in F[X]$, where $deg(g), deg(h) > 0$. Clearing denominators and pulling out the content, we can obtain

$$\alpha f = \beta g_1 h_1,$$

where $\alpha, \beta \in R$ and $g_1, h_1 \in R[X]$ primitive.
Let γ be the content of f, i.e. $f = \gamma f_1$, where f_1 is primitive. Because the content is well defined deduce that $\alpha\gamma = \beta$ (perhaps after changing γ by a unit). Therefore $\alpha f = \alpha\gamma g_1 h_1 \Rightarrow f = \gamma g_1 h_1$. Observe that $deg(g) = deg(g_1)$ and $deg(h) = deg(h_1)$. Also observe that just as for a field $R[X]^* = R^*$. Thus $g_1, h_1 \in R[X]$ are not units. Thus f is reducible in $R[X]$.
□

We should note that in general the converse is not true. For example $3(x-2)$ is reducible in $\mathbb{Z}[X]$, but irreducible in $\mathbb{Q}[X]$. This is because $3 \notin \mathbb{Z}[X]^*$, but $3 \in \mathbb{Q}[X]^*$. This theorem has the surprising consequence:

Corollary. *Let $f = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{Z}[X]$ have a ratonal zero $\frac{\alpha}{\beta}$ where α and β are coprime integers. Then $\beta | a_n$ and if $\alpha \neq 0, \alpha | a_0$. In particular, if $a_n = 1$, all rational zeros are integral.*

Proof. $f(\frac{\alpha}{\beta}) = 0 \Rightarrow (X - \frac{\alpha}{\beta})|f$ in $\mathbb{Q}[X] \Rightarrow \exists g \in \mathbb{Q}[X]$ such that $f = (X - \frac{\alpha}{\beta})g$. Observe that $\beta X - \alpha$ is primitive, hence by the proof of the theorem we deduce that $(\beta X - \alpha)|f \Rightarrow \beta|a_n$ and if $\alpha \neq 0, \alpha|a_0$. Hence if $a_n = 1 \Rightarrow \beta \pm 1 \Rightarrow \frac{\alpha}{\beta} \in \mathbb{Z}$ □

Hence all rational zeroes of a monic polynomial with integer coefficients are integers.

This is kind of amazing. It's not at all obvious from the definitions.

Theorem (Eisenstein's Criterion). *Let $f = a_0 + a_1x + a_2x^2 + \cdots a_nx^n \in \mathbb{Z}[X] \setminus \{0\}$. If there is a prime number $p \in \mathbb{N}$ such that*

1. *(i) $p \not| a_n$*
2. *(ii) $p|a_i \ \ \forall 0 \geq i < n$*
3. *(iii) $p^2 \not| a_0$*

then f is irreducible over $\mathbb{Q}$.

Proof. By the above f reducible over $\mathbb{Q} \Rightarrow f$ reducible over $\mathbb{Z}$. Suppose that f satisfies the conditions (i), (ii), and (iii) but is reducible over $\mathbb{Q}$ and hence over $\mathbb{Z}$. By the proof of the above theorem we know that there exist $g, h \in \mathbb{Z}[X]$ such that $deg(g), deg(h) > 0$ and $f = gh$. Let us write

$$g = b_0 + b_1x + \cdots + b_rx^r,\ h = c_0 + c_1x + \cdots + c_sx^s$$

when $r + s = n = \deg(f)$, $r, s > 0$. We have $a_0 = b_0c_0$. Because $p|a_0$ and $p^2 \not| a_0 \Rightarrow p \not| b_0$ or $p \not| c_0$. Without loss of generality assume that $p|b_0$ and $p \not| c_0$. Furthermore, $b_rc_s = a_n$ is not divisible by $p \Rightarrow \ p \not| b_r$ and $p \not| c_s$. Hence the first coefficient of g is divisible by p but not the last.
Let $i \in \{1, \cdots, r\}$ be minimal such that $p \not| b_i$. Observe that $i \leq r < n$. Note that $a_i = b_ic_0 + b_{i-1}c_1 + \cdots + b_0c_i \Rightarrow \ b_ic_0 = a_i - b_{i-1}c_1 - \cdots - b_0c_i$. But $p|a_i$ by (ii) and $p|b_{i-j}c_j \ \forall \ j \in \{1, \cdots, i\}$ by minimality $\Rightarrow \ p|b_ic_0 \Rightarrow p|b_i$ or $p|c_0$ which is a contradiction. Hence f is irreducible in $\mathbb{Q}[X]$.

□

Corollary. *There are irreducible polynomials of arbitrary degree in $\mathbb{Q}[X]$.*

Proof. Let $p \in \mathbb{N}$ be a prime. Let $n \in \mathbb{N}$ and define $f = p + px + px^2 + \cdots + x^n \in \mathbb{Q}[X]$. By Eisenstein's Criterion, f is irreducible of degree n.

□

Remarks.
1. *Eisenstein's Criterion works (with same proof) for any UFD and its field of fractions.*
2. *Remember irreducible polynomials in $\mathbb{R}[X]$ or $\mathbb{C}[X]$ are of degree 2 or less. $\mathbb{Q}[X]$ is **very** different.*

3. *Here's a useful analogy from chemistry: Let F be a field. One should think about $f \in F[X] \setminus \{0\}$, $f \notin F[X]^*$ (up to association) as a molecule. One should think about the irreducible such f (up to association) as atoms. The fact that $F[X]$ is a UFD says that every molecule is constructed from a unique finite collection of atoms. Trying to determine the irreducible elements of $F[X]$ is the same as trying to construct the period table. So for every F we have an equivalent of a period table. How complicated this periodic table is depends on F. F being algebraically closed says that the atoms are indexed by elements of F, i.e. every irreducible is associated to one of the form $(x-\alpha)$ for a unique $\alpha \in F$. Hence for algebraically closed fields the period table is very easy. The further from being algebraically closed F is the more complicated it becomes. For $\mathbb{Q}$ the periodic table is bewilderingly complicated. The atoms can have a enormous internal complexity. There is* ***far*** *more depth to $\mathbb{Q}$ than meets the eye!*

Let's now study the zeros of polynomials over a field.

Theorem. *Let F be a field and $f \in F[X] \setminus \{0\}$ have distinct roots $\alpha_1, \cdots, \alpha_n \in F$. Then $(x-\alpha_1)\cdots(x-\alpha_n)|f$.*

Proof. We have already proven that $f(\alpha_i) = 0 \Rightarrow (x-\alpha_i)|f$. Recall that for $\alpha, \beta \in F$, $(x-\alpha)$ and $(x-\beta)$ are associated if and only if $\alpha = \beta$. As $\alpha_i \neq \alpha_j \ \forall \ i \neq j \Rightarrow \ x-\alpha_i$ and $x-\alpha_j$ non-associated irreducible factors of $f \ \forall \ i, j$. $F[X]$ is a UFD $\Rightarrow (x-\alpha_1)\cdots(x-\alpha_n)|f$. □

Corollary. *Let F be a field and $f \in F[X]$ be a polynomial of degree $n \in \mathbb{N}$. The number of distinct roots of f in F is at most n.*

Proof. Assume that $deg(f) = n$ and $\{\alpha_1, \cdots \alpha_{n+1}\} \subset F$ are $n+1$ distinct reoots of f in F. By the theorem $g = (x-\alpha_1)\cdots(x-\alpha_{n+1})$ divides f. By the first Euclidean property of the degree function this implies that $deg(f) \geq deg(g) = n+1$. This is a contradiction. Hence the number of distinct zeros of f in F cannot exceed n. □

Corollary. *If F is a field and $f, g \in F[X]$ such that $deg(f), deg(g) \leq n$ and f and g agree on at least $n+1$ values of F then $f = g$.*

Proof. $f - g \in F[X]$ is a polynomial of degree less than or equal to n. By assumption it has $n+1$ roots in F. Hence it is the zero polynomial. □

Corollary. *Let F be an infinite field. Let $f, g \in F[X]$ such that $f(a) = g(a)$ for all $a \in F$ then $f = g$*

Proof. Immediate from the preceding corollary. □

Remarks. *This is not true if F is finite!. For example I'll get you to show that over $\mathbb{F}_p$ the polynomial $x^p - x$ is zero for every value of $\mathbb{F}_p$. This is why thinking about polynomials as functions is a bad plan.*

Theorem. *Let F be an infinite field. Let $f \in F[X_1, \cdots, X_n]$. If $f(\alpha_1, \cdots, \alpha_n) = 0$ for all $\alpha_i \in F$, then $f = 0$.*

Proof. We'll use induction on n. The previous corollary says that the result is true for $n = 1$. Let $n > 1$ and write f as a polynomial in X_1 with coefficients in $F[X_2, \cdots, X_n]$.

$$f(x_1, \cdots, x_n) = a_0 + \cdots + a_k x_1^k,$$

where $a_i = a_i(x_2, \cdots, x_n)$. Fix $\alpha_2, \cdots, \alpha_n \in F$. Then $f(x_1, \alpha_2, \cdots \alpha_n)$ vanishes for all values of F.By the preceding corollary we deduce that

$$a_i(\alpha_2, \cdots, \alpha_n) = 0 \ \forall i.$$

But the α_j were arbitrary. Hence by the induction hypothesis $a_i = 0$ for all i. Hence $f = 0$. □

5 Field Extensions and Galois Theory

5.1 Field Extensions and Minimal Polynomials

Definition. *Let E be a field and $F \subset E$ a subfield, i.e. a subring which is a field. Then we call E an extension of F and we write E/F.*

Let F be a field. Recall that a vector space V over F is an Abelian group with a good concept of scalar multiplication by F. If we have an extension of fields E/F then we may naturally regard E as a vector space over F. This is because there is a natural concept of scalar multiplication on E by F. The properties of a vector space are automatically satisfied by the ring axioms for E. If you've only ever seen vector spaces over $\mathbb{R}$ or $\mathbb{C}$, don't worry, all of the theory is identical. A trivial observation is that E is a vector space over itself.

Definition. *Let E/F be a field extension. We say that E/F is finite if E is a finite dimensional vector space over F, i.e. there is a finite spanning set for E over F. If E/F is finite then we call the dimension of E over F the degree of the extension, written $[E : F]$.*

Concretely this means that we may find a finite subset $\{x_1, \cdots, x_n\} \subset E$ such that

$$E = \{\lambda_1 x_1 + \cdots \lambda_n x_n | \lambda_i \in F\}.$$

Hence if $[E : F] = n$, then as an F-vector space $E \cong F^n$. We should be careful of this definition for the following reason:

If E/F is a finite extension then $(F, +) \subset (E.+)$ is a subgroup as an Abelian group. It is not necessarily true that $(F, +)$ is of finite index in $(E, +)$. I'll get you to prove this in the homework.

Let E/F be en extension of finite fields. Trivially we can see that the extension is finite. Hence if $[E : F] = n$, then as an F-vector space $E \cong F^n \Rightarrow |E| = |F|^n$. Hence $|E| = |F|^{[E:F]}$. From this observation we deduce that

Theorem. *Let E be a finite field of characteristic $p \in \mathbb{N}$. Then $|E| = p^n$ for some $n \in \mathbb{N}$.*

Proof. $char(E) = p \Rightarrow \mathbb{F}_p \subset E$. Hence $E/\mathbb{F}_p$ is a finite extension. Hence $|E| = |\mathbb{F}_p|^{[E:\mathbb{F}_p]} = p^{[E:\mathbb{F}_p]}$. □

Definition. *Let E/F be a field extension. Let $\alpha \in E$. We say that α is **algebraic** over F is $\exists f \in F[X]$ such that $f(\alpha) = 0$. If every $\alpha \in E$ is algebraic over F we say that the extension E/F is algebraic. If α is not algebraic then we say that it is transcendental. e.g. over $\mathbb{Q}$, $\sqrt{2}$ is algebraic, whereas π is transcendental.*

Proposition. *Let E/F be a finite field extension. Then E/F is algebraic.*

Proof. Let $\alpha \in E$. Assume that $[E : F] = n$. Thus any set subset of E of cardinality greater than n must be linearly dependent (over F). Thus $\{1, \alpha, \cdots, \alpha^n\} \subset E$ must be linearly dependent over F. Hence $\exists b_0, \cdots, b_n \in F$ such that

$$b_0 + b_1\alpha + \cdots + b_n\alpha^n = 0.$$

Let $f = b_0 + b_1x + \cdots + b_nx^n \in F[X]$. By construction $f(\alpha) = 0$. Thus α is algebraic over F. □

The converse is not true. I'll give you an example in the homework.

Definition. *Let E/F be a field extension. Let $\alpha \in E$ be algebraic (over F). Then the monic polynomial $f \in F[X]$ of minimal degree such that $f(\alpha) = 0$ is called the minimal polynomial of α (over F).*

Proposition. *Let E/F be a field extension. Let $\alpha \in E$ be algebraic (over F). The minimal polynomial of α (over F) is irreducible (in $F[X]$).*

Proof. Let $F \in F[X]$ be the minimal polynomial of α. Recall that f is reducible if and only if we can find $g, h \in F[X]$ such that $f = gh$ and $deg(g), deg(h) < deg(f)$. However, if such a factorisation exists, we know that $f(\alpha) = g(\alpha)h(\alpha) = 0$. But E is a field and is thus an integral domain. Consequently either $g(\alpha) = 0$ or $h(\alpha) = 0$. But this contradicts the minimality of $deg(f)$. □

Corollary. *Let E/F be a field extension. Let $\alpha \in E$ be algebraic (over F). The minimal polynomial of α (over F) is unique.*

Proof. Let $g, f \in F[X]$ both be monic of minimal degree such that $f(\alpha) = g(\alpha) = 0$. Recall that monic polynomials in $F[X]$ are associated if and only if they are equal. Thus if $f \neq g$, then by the unique factorisation property of $F[X]$, we know they are coprime (HCF(f,g) =1). If this were the case then y the Euclidean property of $F[X]$ $\exists u, v \in F[X]$ such that $fu + gv = 1$. But this would imply that $f(\alpha)u(\alpha) + g(\alpha)v(\alpha) = 1$. But the left hand side equals 0, which is a contradiction because E is a field so is by definition non-trivial. Thus $f = g$. □

Corollary. *Let E/F be a field extension. Let $\alpha \in E$ be algebraic (over F). Then α is the root of a unique irreducible monic polynomial in $F[X]$*

Proof. The above two results shows that the minimal polynomial of α (over F). is irreducible and necessarily unique. The proof of the corollary shows that it must be the only monic irreducible polynomial with α as a root. □

Definition. *Let E/F be a field extension. Let $\alpha \in E$ (not necessarily algebraic over F). We define the subfield generated by α to be the minimal subfield of E containing F and α. We denote this subfield by $F(\alpha)$.*

Proposition. *Let E/F be a field extension. Let $\alpha \in E$ be algebraic (over F). Let $F[\alpha] := \{f(\alpha) | f \in F[X]\} \subset E$. Then $F[\alpha] = F(\alpha)$. Moreover the the degree of $F(\alpha)$ over F equals the degree of the minimal polynomial of α over F*

Proof. We should first observe that $F[\alpha] \subset E$ is the minimal subring of E containing F and α: it is clearly closed under addition and multiplication because $g(\alpha)h(\alpha) = (gh)(\alpha)$ and $g(\alpha) + h(\alpha) = (g + h)(\alpha)$ for all $g, h \in F[X]$. We need to show therefore that it is actually a subfield. Note that $F[\alpha]$ is an F-vector space. Let $f = x^n + \sum_{i=0}^{n-1} b_i x^i \in F[X]$ be the minimal polynomial of α. We claim that the subset $\{1, \alpha, \cdots, \alpha^{n-1}\} \subset F[\alpha]$ is an F-basis.

Spanning

Let $Sp_F(1, \alpha, \cdots, \alpha^{n-1}) \subset F[\alpha]$ be the F-linear span of $\{1, \alpha, \cdots, \alpha^{n-1}\}$. We will show that all positive powers of α are in $Sp_F(1, \alpha, \cdots, \alpha^{n-1})$ by induction. Let $k \in \mathbb{N}$. If $k < n$ then α^k is trivially in the span. Observe that because $f(\alpha) = \alpha^n + \sum_{i=0}^{n-1} b_i \alpha^i = 0$ we see that α^n is in $Sp_F(1, \alpha, \cdots, \alpha^{n-1})$ Hence $Sp_F(1, \alpha, \cdots, \alpha^{n-1}, \alpha^n) = Sp_F(1, \alpha, \cdots, \alpha^{n-1})$. Finally assume that $k > n$. Inductively we may assume that $\alpha^{k-1} \in Sp_F(1, \alpha, \cdots, \alpha^{n-1})$. But then $\alpha^k \in Sp_F(1, \alpha, \cdots, \alpha^{n-1}, \alpha^n) = Sp_F(1, \alpha, \cdots, \alpha^{n-1})$. Thus all positive powers of α are contained in $Sp_F(1, \alpha, \cdots, \alpha^{n-1})$. Ever element of $F[\alpha]$ is an F-linear combination of such terms, hence $Sp_F(1, \alpha, \cdots, \alpha^{n-1}) = F[\alpha]$.

Linear Independence

If $\{1, \alpha, \cdots, \alpha^{n-1}\}$ were linearly dependent over F, then the minimal polynomial of α over F would have degree strictly less than n. This is a contradiction.

Now we must show that $F[\alpha]$ is a subfield of E. Let $f \in F[X]$ and $\beta = f(\alpha) \neq 0$. By the above we know that the set $\{1, \beta, \cdots, \beta^n\} \subset F[\alpha]$ is linearly dependent over F. Hence we may know that $\exists \{a_0, \cdots, a_n\} \subset F$ such that

$$a_0 + a_1\beta + \cdots + a_n\beta^n = 0.$$

Because $\beta \neq 0$ we conclude that there exists $k \in \mathbb{N}$ and $g \in F[X]$ such that $g(\beta) = 0$ and

$$g = 1 + b_1 x + \cdots + b_k x^k, \;\; b_i \in F.$$

But then

$$1 = \beta(-b_1 - \cdots - b_k \beta^{k-1}).$$

Thus $-b_1 - \cdots - b_k \beta^{k-1} \in F[\alpha]$ is the multiplicative inverse of β in E. We conclude that $F[\alpha]$ is a field and thus $F[\alpha] = F(\alpha)$.

□

Proposition. *Let R be a Euclidean domain. Let $a \in R$ be an irreducible element. Then the principal ideal $(a) \subset R$ is maximal.*

Proof. Recall that $(a) = \{ar | r \in R\}$. First observe that a is a non-unit so (a) is proper. Now observe that if $I \subset R$ is an ideal and $\exists\ b \in I$ such that $b \in R^*$ then $I = R$. This is clear because I is closed under multiplication by all R. Hence any proper ideal of R cannot contain any units. Assume that (a) is not maximal and that $I \subset R$ is a proper ideal of R strictly containing (a), i.e. $(a) \neq I$. Let $b \in I$ such that $b \notin (a)$. Hence a does not divide b, which is a non-zero, non-unit. But R is Euclidean, which in particular implies that it is a UFD. Hence HCFs exist and by construction $HCF(a, b) = 1$. We also know that by the Euclidean property $\exists u, v \in R\ ua + vb = 1$. But by construction $a, b \in I$. This implies therefore that $1 \in I$. This is a contradiction as I is proper. Hence (a) is maximal. □

5.2 Splitting Fields

Theorem. *Let F be a field. Let $f \in F[X]$ be a non-constant polynomial. Then there exists a field extension E/F such that f has a root in E. Moreover E can be chosen to be finite over F.*

Proof. The key observation is that $F[X]$ is Euclidean. As a result $F[X]$ is a UFD, hence we may factor f into irreducible polynomials in $F[X]$. Without loss of generality we may therefore assume that f itself is irreducible. By the above we know that the ideal $(f(X)) \subset F[X]$ is maximal. This implies that the quotient $E := F[X]/(f(X))$ is a field. There is a natural ring homomorphism:

$$\begin{array}{ccc} F & \longrightarrow & E \\ \lambda & \longrightarrow & \lambda + (f(X)) \end{array}$$

E is injective because it has trivial kernel. Hence we may naturally think of E as a field extension of F. Let $g \in F[X]$. Let $a(X) + (f(X)) \in E$. By definition $g(a(X) + (f(X))) = g(a(X)) + (f(X))$. consider $X + (f(X)) \in E$. $f(X + (f(X))) = f(X) + (f(X)) = (f(X))$. But $(f(X)) \in E$ is the additive identity. Thus $X + (f(X))$ is a root of f in E.

Finally we need to show that E/F is finite. Assume that $deg(f) = n$. We claim that $\{1 + (f(X)), X + (f(X)), \cdots, X^{n-1} + (f(X))\} \subset E$ forms a spanning set for E over F.

Given any $g \in F[X]$ we have the element $g(X)+(f(X)) \in E$. Remember that the degree function on $F[X]$ is Euclidean. Hence we have a version of the remainder theorem: either $g(X)|f(X)$ of $\exists q(X), r(X) \in F[X]$ such that $g(X) = q(X)f(X)+r(X)$ where $deg(r(X)) < n$. In the first case $g(X) \in (f(X))$ which implies that $g(X)+(f(X))$ is zero in E. In the second case we have $g(X) + (f(X)) = r(X) + (f(X))$. But $r(X) + (f(X))$ is clearly in the F-span of $\{1 + (f(X)), X + (f(X)), \cdots, X^{n-1} + (f(X))\}$. Thus E/F is finite. □

Corollary. *Let F be a field. Let $f \in F[X]$ be a non-constant polynomial. Then there exists a finite field extension E/F such that f splits into linear factors in $E[X]$.*

Proof. We'll use induction on the degree of f. Clearly if f is linear the result is true. Assume therefore that $deg(f) > 1$. By the above theorem we may find a finite field extension K/F such that $\exists\ \alpha \in K$ such that $f(\alpha) = 0$. This implies that $f = (X - \alpha)g$ for some $g \in K[X]$. By construction $deg(g) < deg(f)$. By induction we know that there is a finite field extension E/K in which g, and thus f, splits into linear factors. Because both E/K and K/F are finite E/F is finite. □

This is beautiful result. In particular it facilitates the following fundamental definition:

Definition. *Let F be a field. Let $f \in F[X]$. A splitting field for f is a finite extension E/F of minimal degree over F such that f splits into linear factors in $E[X]$.*

Theorem. *Let F be a field and $f \in F[X]$. Let E and E' be two splitting fields of f. Then E is isomorphic to E'*

Proof. We don't quite have enough time to prove this. It isn't too hard though. Intuitively it is unsurprising because a splitting field is some kind of minimal field generated by F and the roots of f. You will prove it in a second course in abstract algebra.

□

When we are thinking about $\mathbb{Q}$ we are lucky enough to have a natural embedding of $\mathbb{Q}$ in $\mathbb{C}$ which is algebraically closed. This means the splitting field of any polynomial $f \in \mathbb{Q}[X]$ can naturally be considered a subfield of $\mathbb{C}$. Concretely, if $\{\alpha_1, \cdots, \alpha_n\} \subset \mathbb{C}$ are the roots of f in $\mathbb{C}$ then the minimal subfield of $\mathbb{C}$ containing $\mathbb{Q}$ and $\{\alpha_1, \cdots, \alpha_n\}$, denoted $\mathbb{Q}(\alpha_1, \cdots, \alpha_n) \subset \mathbb{C}$, is a splitting field for f.

5.3 Galois Theory (Proofs Omitted)

Definition. *Let E/F be a finite extension. We say that E/F is **normal** if E is a splitting field for some $f \in F[X]$.*

Remarks. *An extension E/F being normal is is equivalent to the condition that if $f \in F[X]$ admits a root in E then it must split into linear factors in $E[X]$.*

From now on assume that all fields are of characteristic zero.

Definition. *Let E/F be a finite extension. We say that E/F is **Galois** if is normal.*

Remarks. *For characteristic p fields, being Galois requires an extra condition called separability. Separability is automatically satisfied by characteristic zero field extension.*

If $F = \mathbb{Q}$ then $E = \mathbb{Q}(\sqrt[3]{2}, e^{\frac{2\pi i}{3}})$ is a Galois extension as it is the splitting field of $x^3 - 2$. Note that $\mathbb{Q}(\sqrt[3]{2})$ is not a Galois extension of $\mathbb{Q}$ as $x^3 - 2$ does not split into linear factors in $\mathbb{Q}(\sqrt[3]{2})[X]$.

Definition. *Let E/F be a finite Galois extension. The Galois group of E/F, denoted $Gal(E/F)$, is the group of field automorphisms of E which fix F,i.e. $\sigma \in Gal(E/F)$ is a field automorphism $\sigma : E \to E$ such that $\sigma(\alpha) = \alpha \ \forall \alpha \in F$. Composition is given by composition of functions.*

This concept was first introduced by Evariste Galois (1811-1832). In fact this is where the term group comes from. Galois was the first to use it. Here are some nice facts about Galois extensions and Galois groups:

- Galois groups are finite. Moreover $|Gal(E/F)| = [E : F]$.
- If E/F is Galois with E the splitting field of a degree n polynomial $f \in F[X]$, then $Gal(E/F)$ acts faithfully on the roots of f in E. In particular we can identify $Gal(E/F)$ with a subgroup of Sym_n. It is important to realize that in general $Gal(E/F)$ will not be isomorphic to Sym_n. Because elements of $Gal(E/F)$ fix F they must preserve algebraic relations (over F) among the roots.

We should therefore think about $Gal(E/F)$ as permutations of the roots of a splitting polynomial which preserve all algebraic relationships between them.

This makes Galois groups extremely subtle. In some instances there may be no relationships (so $Gal(E/F) \cong Sym_n$), whereas in others there may be many (so $Gal(E/F)$ is much smaller than Sym_n).

Fundamental Theorem of Galois Theory. *Let E/F be a finite Galois extension. Then there is a natural bijection*

$$\begin{aligned} \{\textit{Intermediate Subfields } F \subset K \subset E\} &\longrightarrow \{\textit{Subgroups } H \subset Gal(E/F)\} \\ K &\longrightarrow \{\sigma \in Gal(E/F) | \sigma(k) = k, \forall k \in K\} = Gal(E/K) \end{aligned}$$

In addition K/F is Galois if and only if $Gal(E/K) \triangleleft Gal(E/F)$. In this case

$$Gal(K/F) \cong Gal(E/F)/Gal(E/K).$$

5.4 Solving Polynomials By Radicals

Suppose we wish to find the root of a polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{Q}[X]$. If $n = 2$ we have the quadratic formula. It is natural to ask if there is a similar formula (in terms of the coefficients) for $n > 2$. It turns out that for $n = 3$ and $n = 4$ there are formulae,

although they are extremely complicated. For many centuries mathematicians search for a formula in the case $n = 5$. Eventually it was proven (first by Abel and then later by Galois) that no such formula exists if $n \geq 5$. This is a very surprising result. What is so special about $n = 5$?

What would it mean for there to be an analogue of the quadratic formula for higher degree polynomials? In simpler terms, it would mean that all the zeroes could be constructed by repeatedly taking roots and applying basic algebraic operations. This would mean that the splitting field would have to have the following very specific property.

Definition. *Let $f \in \mathbb{Q}[X]$ with splitting field K_f. We say that f is solvable by radicals if there is a chain of fields*

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_m \subset \mathbb{C}$$

such that

- $K_f \subset K_m$
- $K_{i+1} = K_i(\alpha_i)$, *where α_i is a root of a polynomial of the form $x^{n_i} - b_i \in K_i[X]$, for all $0 < i < m$.*
- $e^{2\pi i/n} \in K_1$ *where $n = \prod n_i$. (This last condition is non-standard. It's included to simplify the exposition.)*

It is a fact that K_i/K_{i-1} is Galois and $Gal(K_i/K_{i-1})$ is Abelian By the fundamental theorem of Galois theory the chain

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_m$$

gives rise to the nested collection of subgroups

$$\{e\} \triangleleft Gal(K_m/K_{m-1}) \triangleleft \cdots Gal(K_m/K_2) \triangleleft Gal(K_m/K_1) \triangleleft Gal(K_m/K_0)$$

where

$$Gal(K_m/K_{i-1})/Gal(K_m/K_i) \cong Gal(K_i/K_{i-1}).$$

Note that these quotients are Abelian, hence $Gal(K_m/K_0)$ is a solvable group. Now observe that $Gal(K_f/\mathbb{Q}) \cong Gal(K_m/\mathbb{Q})/Gal(K_m/K_f)$. It is a fact that a quotient of a finite solvable group is solvable. Hence we deduce the following result

Theorem. *$f \in \mathbb{Q}[X]$ solvable by radicals $\Rightarrow$ $Gal(K_f/\mathbb{Q})$ a solvable group.*

This was Galois' key insight. He realized that if there was a version of the quadratic formula then the corresponding Galois group would have Abelian simple components. What's this got to do with degree five polynomials?

If $f \in \mathbb{Q}[X]$ is an irreducible, degree five polynomial with exactly three real roots (for example $x^5 - 9x + 3$) then it's possible to show that $Gal(K_f/\mathbb{Q}) \cong Sym_5$. Galois showed

that the simple components of Sym_5 are $\{Alt_5, \mathbb{Z}/2\mathbb{Z}\}$, hence $Gal(K_f/\mathbb{Q})$ is not solvable. Hence, in this case, f is not solvable by radicals, so there can be no version of the quadratic formula. Why isn't there a problem for degree 2,3 or 4? It's because Sym_2, Sym_3 and Sym_4 are solvable.

This proof is one of the great achievements in mathematics. Galois was a true genius. He died at 21 in a duel over a woman. He wrote all this down the night before he died, running out of time in the end. Hermann Weyl, one of the greatest mathematicians of the 20th century, said of this testament,

This letter, if judged by the novelty and profundity of ideas it contains, is perhaps the most substantial piece of writing in the whole literature of mankind.