Algebra

1 Introduction and Examples

1.1 What is Algebra?

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. 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 canonical members.

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

Remark 1.1. *In linear algebra the analogous idea is $(\mathbb{R}^n,+,$ scalar multiplication $) \longrightarrow$ Vector Spaces over $\mathbb{R}$*

1.2 Sets and Functions

1.2.1 Some notations

1. If $S \subset T$ then $T\backslash S := \{x \in T \mid x \notin S\}$. $T\backslash S$ is called the **compliment** of S in T.
2. $S \times T = \{(a,b) \mid a \in S, b \in T\}$. We call this new set the **(cartesian) product** of S and T. We may naturally extend this concept to finite collections of sets.
3. We say that S and T are disjoint if $S \cap T = \emptyset$. The **union of two disjoint** sets is often written as $S \coprod T$.
4. The symbol $\exists\,!$ should be read as "there exists unique".

1.2.2 Codomain, Injective, Surjective, Bijective

Definition 1.2. ***(Function)*** *A map (or function) f from S to T is a rule which assigns to each element of S a* ***unique*** *elements of T:*

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

Example 1.3. ***(some examples)***

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

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

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

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

NTU, Singapore, June 2021

Definition 1.4. ***(Domain, Codomain, Injective, Surjective, Bijective)*** *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, $f(S)$ is the **range**, 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 $f = \mathrm{Id}_S$.*
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.*

Remark 1.5. *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 naturally contains $g(R)$.*

Lemma 1.6. ***(bijection equivalent condition)*** *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 (1) $g \circ f = Id_S$ and (2) $f \circ g = Id_T$.*

Proof. We prove it as follows:

I) "⇒": easy, since $f(x) = y$, just define $g = f^{-1}(y) = x$

II) "⇐": We seperate it into two parts:

(1) ⇒ injective: If f is not injective, for different $x \neq y$, we must have $g \circ f(x) = g \circ f(y)$, which indicates $g \circ f \neq Id_S$.

(2) ⇒ surjective: If f is not surjective, $f \circ g(T)$ is a proper subset of T, which indicates $f \circ g \neq Id_T$.

□

Remark 1.7. ***(necessity of both condition (1) and (2))*** *Both are necessary, we show by counterexamples:*

(2) ✓ (1) ✗:

$$S = [0, 2], \quad T = [0, 2]$$
$$g(x) = \frac{1}{2}x, \quad f(x) = \left\{ \begin{array}{cc} 2x & x \in [0, 1] \\ x - 1 & x \in [1, 2] \end{array} \right.$$

(1) ✓ (2) ✗:

$$S = [0, 1], \quad T = [0, 2]$$
$$f(x) = x, \quad g(x) = \left\{ \begin{array}{cc} x & x \in [0, 1] \\ x - 1 & x \in [1, 2] \end{array} \right.$$

Lemma 1.8. *A bijection exists between two **finite** sets if and only if they have the same cardinality.*

Lemma 1.9. *Let S be a finite set. Let $f : S \to S$ be an injection. Then f is also a surjection.*

Proof. Let $a \in S$. We need to show that there exists $b \in S$ such that $a = f(b)$. Consider what happens when f is applied repeatedly on S. Let f^2 denote $f \circ f$ and, generally, $f^n := f \circ f^{n-1}$. Consider the sequence of elements of S :

$$a, f(a), f^2(a), \ldots$$

Because S is a finite set, there must be repetitions. That is, there must exist $r, s \in \mathbb{N}$ such that:

$$f^r(a) = f^s(a)$$

where $r \neq s$. Without loss of generality, assume $r > s$. f is an injection. Therefore by Composite of Injections is Injection, f^s is an injection. By "Injection iff Left Cancellable", f^s is left cancellable. Thus:

$$f^r(a) = f^s(a) \Rightarrow f^s \circ f^{r-s}(a) = f^s(a) \Rightarrow f^{r-s}(a) = a$$

That is, $b = f^{r-s-1}(a)$ □

Corollary 1.10. *Let S be a finite set. Let $f : S \to S$ be an injection. Then f is a permutation.*

1.3 Equivalence Relations

Definition 1.11. ***(Equivalence Relation)*** *An equivalence relation on a set S is a subset $U \subset S \times S$ satisfying:*

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

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

Definition 1.12. ***(Equivalence Class)*** *Let $\sim$ be an equivalence relation on the set S. Let $x \in S$. The equivalence class containing x is the subset*

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

Remark 1.13. ***(Explanation)***

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 1.14. ***(Partition)*** *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.*

Lemma 1.15. ***(equivalence vs. partition)*** *An equivalence relation on a set is the same as a partition.*

Proof. We have

$\Rightarrow$: Directly from Remark 1.13.

$\Leftarrow$: For partition $\{X_i\}$, set U as $\cup_i X_i \times X_i$. □

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

Observe that $\mathbb{Z}$ is a set, and many other various properties comes from $+$ and $\times$. $\mathbb{Z}$ with operators $+$ and $\times$ is the canonical example of a **ring**.

2.1 Basic Properties of $+$ and $\times$ for $\mathbb{Z}$

We may naturally express operators $+$ and $\times$ as maps from the product space:

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

- Here are 4 elementary properties that $+$ satisfies:
    1. **Associativity:** $a + (b + c) = (a + b) + c, \forall\, a, b, c \in \mathbb{Z}$.

2. **Existence of additive identity:** $a + 0 = 0 + a = a, \forall\, a \in \mathbb{Z}$.
3. **Existence of additive inverses:** $a + (-a) = (-a) + a = 0, \forall\, a \in \mathbb{Z}$.
4. **Commutativity:** $a + b = b + a, \forall\, a, b \in \mathbb{Z}$.

- Here are 3 elementary properties that × satisfy:

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

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

1. **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 $a \times b = ab$.

Remark 2.1. ***(finitely generated)*** *$\mathbb{Z}$ is generated by 1 under addition: every integer can be attained by successively adding 1 (or -1) to itself. See Section 3.3 for* ***cyclic*** *group. Under multiplication the situation is much more complicated. There is clearly no single generator of $\mathbb{Z}$ under multiplication in the above sense.*

Remark 2.2. *We will know that $\mathbb{Z}$ with operators + and × is a* ***commutative ring****. Furthermore, $\mathbb{Z}$ is actually a commutative* ***entire*** *(no zero divisors (2.5)) ring which is called an* ***integral domain****. See the following for details.*

Remark 2.3. *Some explanation:*

1. *Because of* ***non-existence of multiplicative inverses****, $\mathbb{Z}$ is not a field.*

2. *Each of these properties is totally obvious but will form the foundations of future definitions of* ***groups*** *and* ***rings****.*

3. *All of the above hold for + and × on $\mathbb{Q}$. Additionally, $\mathbb{Q}$ will motivate the definition of a* ***field*** *with an extra property:*

   ∗ **non-zero elements have multiplicative inverses:** *Given $a \in \mathbb{Q}\backslash\{0\}, \exists\, b \in \mathbb{Q}$ such that $ab = ba = 1$.*

4. *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 in different order.*

• integral domain property:

Lemma 2.4. ***(No Zero Divisors Property of $\mathbb{Z}$)*** *In $\mathbb{Z}$ (and $\mathbb{Q}$), the product of two non-zero elements is again non-zero. More precisely:*

$$a, b \in \mathbb{Z} \text{ such that } ab = 0 \Rightarrow \text{ either } a = 0 \text{ or } b = 0.$$

Proof. We can extend $\mathbb{Z}$ to $\mathbb{Q}$ which is a field and use [1, Page 7]: Assume $x \neq 0, y \neq 0, x, y \in \mathbb{Q}$, but $xy = 0$. We then have

$$1 = \left(\frac{1}{y}\right)\left(\frac{1}{x}\right) xy = \left(\frac{1}{y}\right)\left(\frac{1}{x}\right) 0 = 0 \tag{2.5}$$

a contradiction. □

Remark 2.6. *The above proof is not that necessary, I just want to show that any field is an integral domain as shown in Corollary 4.39. The key point will be brought repeatedly that invertible element is not zero (xy is invertible, so it is not 0).*

Corollary 2.7. ***(Cancellation Law)*** *For $a, b, c \in \mathbb{Z}, ca = cb$ and $c \neq 0 \Rightarrow a = b$.*

Proof. This is proven using the distributive law together with the fact that $\mathbb{Z}$ is an integral domain. For more details see Remark 3.26, Remark 4.41 and Corollary 4.42 (cancel law for domain, not for group). □

2.2 Factorization and the Fundamental Theorem of Arithmetic

Definition 2.8. ***(Divisibility)*** *Let $a, b \in \mathbb{Z}$. Then a* ***divides*** *b iff $\exists\, c \in \mathbb{Z}$ such that $b = ca$. We denote this by $a \mid b$ and say that a is a* ***divisor*** *(or factor) of b, and b is a multiple of a.*

Remark 2.9. ***(special** 0 **and** 1**)***

1. *0 is divisible by every integer.*
2. *The only integers which divide 1 are 1 and -1.*

Definition 2.10. ***(Factorization)*** *Any way of expressing an integer as the product of a finite collection of integers is called a* ***factorization****.*

Definition 2.11. ***(Prime Number)*** *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****.*

Definition 2.12. ***(Highest Common Factor, Coprime)*** *Let $a, b \in \mathbb{Z}$. The* ***highest common factor*** *of a and b, denoted $\operatorname{HCF}(a, b)$, 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 $\operatorname{HCF}(a, b) = 1$.*

Definition 2.13. ***(Lowest Common Multiple)*** *Let $a, b \in \mathbb{Z}$. The* ***lowest common multiple*** *of a and b, denoted $\operatorname{LCM}(a, b)$, is the smallest positive integer which is a common multiple of a and b.*

Here are some important elementary properties of divisibility dating back to Euclid (300BC). We'll actually prove them later in far more generality. Here I give some basic statements and proofs.

Theorem 2.14. ***(Remainder Theorem)*** *Given $a, b \in \mathbb{Z}$, if $b > 0$ then $\exists! q, r \in \mathbb{Z}$ such that $a = bq + r$ with $0 \leq r < b$*

Proof. Assume q_1, r_1 and q_2, r_2, $0 < r_1, r_2 < b$, satisfy the condition. We will get $r_1 = r_2$ and $q_1 = q_2$. □

Theorem 2.15. ***(Bézout's identity)***

- *Given $a, b \in \mathbb{Z}, \exists\, s, t \in \mathbb{Z}$ such that $as + bt = \operatorname{HCF}(a, b) = \min\{ax + by \mid x, y \in \mathbb{Z}$ and $ax + by > 0\}$. More generally:*

  $$\forall\, z \in S = \{ax + by \mid x, y \in \mathbb{Z} \text{ and } ax + by > 0\}, \operatorname{HCF}(a, b) | z.$$

- *If c is any common divisor of a and b, we have $c| \operatorname{HCF}(a, b)$.*
- *In particular, a and b are coprime if an only if there exist $s, t \in \mathbb{Z}$ such that $as + bt = 1$.*

Proof. The set S is nonempty since it contains either a or $-a$ (with $x = \pm 1$ and $y = 0$). Since S is a nonempty set of positive integers, it has a minimum element $d = as + bt$, by the well-ordering principle. To prove that $d = \operatorname{HCF}(a, b)$, it must be proven that d is a common divisor of a and b, and that for any other common divisor c, one has $c \leq d$. The Euclidean division of a by d may be written as $a = dq + r$ with $0 \leq r < d$.

The remainder r is in $S \cup \{0\}$, because

$$\begin{aligned} r &= a - qd \\ &= a - q(as + bt) \\ &= a(1 - qs) - bqt. \end{aligned}$$

Thus r is of the form $ax + by$, and hence $r \in S \cup \{0\}$. However, $0 \leq r < d$, and d is the smallest positive integer in $S:$ the remainder r can therefore not be in S, so $r = 0$. This implies that $d|a$. Similarly $d|b$, and d is a common divisor of a and b.

Now, let c be any common divisor of a and b; that is, there exist u and v such that $a = cu$ and $b = cv$. One has thus

$$\begin{aligned} d &= as + bt \\ &= cus + cvt \\ &= c(us + vt) \end{aligned}$$

That is $c|d$, and, therefore $c \leq d$.

If $\exists\, x, y \in \mathbb{Z}$, so that $ax + by = kd + r$, and $0 < r < d$. We have $ka(s - x) + kb(t - y) = r < d$, a contradiction of d is the smallest positive integer in S. We therefore have

$$\forall\, x \in S, \mathrm{HCF}(a, b)|x.$$

The final statement follows easily if we note that 1 must be the HCF since it is the smallest positive integer. □

Remark 2.16. *If there exist $s, t \in \mathbb{Z}$ such that $as + bt = k$, we cannot get k is $\mathrm{HCF}(a, b)$, unless $k = 1$.*

Corollary 2.17. *Bézout's identity can be extended to more than two integers: if*

$$\gcd(a_1, a_2, \ldots, a_n) = d$$

then there are integers $x_1, x_2, \ldots, x_n$ such that

$$d = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

has the following properties:

- *d is the smallest positive integer of this form*
- *every number of this form is a multiple of d*

Theorem 2.18. ***(Euclid's Lemma)*** *Let p be a prime number and $a, b \in \mathbb{Z}$. Then*

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

Remark 2.19. ***(Equivalent Statements)***

1. *p is a prime number. If $p \nmid a$ and $p \nmid b$, then $p \nmid ab$.*
2. *p is a prime number. If $p \nmid a$ and $p \mid ab$, then $p \mid b$.*
3. *If $n \mid ab$, and n is coprime to a, then $n \mid b$.*

   This is a generalization because in the special case n is prime, either

   - *$n \mid a$*
   - *n is coprime to a (i.e. $n \nmid a$), so $n \mid b$.*

Proof. We prove equivalent statements 3. From Theorem 2.15, we have that $\exists\, r, s$, s.t. $rn + sa = 1$. Multiply both sides by b, we get

$$rnb + sab = b$$

Left sum is divisible by n, so $n \mid b$. □

Corollary 2.20. ***(prime vs. Euclid's Lemma)***
p is a prime $\iff$ 1). $p \neq 1$ and $p \neq 0$; 2). $\forall\, a, b \in \mathbb{N}^+, p|ab \Rightarrow p|a$ or $p \mid b$

Proof. "$\Rightarrow$" from definition and Euclid's Lemma. "$\Leftarrow$" is p is not prime, let $p = p_1p_2$, $a = p_1p_3$ and $b = p_2p_3$, we then get a contradiction of 2). □

Theorem 2.21. ***(The Fundamental Theorem of Arithmetic)*** *Every positive integer $a > 1$, can be written as a product of primes:*

$$a = p_1p_2 \ldots p_r$$

*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 $c = c_1c_2$ where $c_1, c_2 \in \mathbb{N}, c_1 < c$ 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

$$a = p_1p_2 \ldots p_r = q_1q_2 \ldots q_s$$

be two factorizations of a into a product of primes. Then $p_1 \mid q_1 q_2 \dots q_s$. By Euclid's Lemma we know that $p_1 \mid q_i$ 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

$$p_2 \dots p_r = q_2 \dots q_s$$

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 $p_i = q_i$ for all i. □

Corollary 2.22. ***(Equivalence)***

Euclid's Lemma $\Longleftrightarrow$ The Fundamental Theorem of Arithmetic

Proof.

- $\Longrightarrow$: this is just the above proof to Theorem 2.21.
- $\Longleftarrow$: $p \mid ab$ and $ab = p_1 \dots p_r$, since p is prime, from uniqueness of the factorization, we have $p \in \{p_1, \dots, p_r\}$. So $p|a$ or $p \mid b$.

□

Remark 2.23. *We will extend the results to general integer domains in Section 4.9, where we first assume the existence of Fundamental Theorem of Arithmetic, and then derive Euclid's Lemma.*

Theorem 2.24. *There are infinitely many distinct prime numbers.*

Proof. Suppose that there are finitely many distinct primes $p_1, p_2 \dots p_r$. Consider $c = p_1 p_2 \dots p_r + 1$. 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

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

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

• Fundamental Theorem and $\mathbb{Q}$:

The Fundamental Theorem of Arithmetic also tells us that:

Corollary 2.25. *Every positive element $a \in \mathbb{Q}$ can be written uniquely (up to reordering) in the form:*

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

Corollary 2.26.

two positive integers are coprime $\Longleftrightarrow$ they have no common prime divisor.

This immediately shows that every positive element $a \in \mathbb{Q}$ can be written uniquely in the form:

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

Proof. "$\Rightarrow$": if a common prime divisor, not coprime. "$\Leftarrow$": if not coprime, HCF > 1:

- If HCF is a prime, a contradiction.
- If HCF is a not prime, we have HCF $= p_1 \dots p_r$ so p_1 in both integers.

□

2.3 Congruences

Definition 2.27. ***(Modulo Division)*** *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.*

Remark 2.28. ***(equivalence relation on $\mathbb{Z}$)***

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

Definition 2.29. ***(Congruent Modulo)*** *$a, b \in \mathbb{Z}$ are **congruent modulo** m iff $m \mid (a - b)$. This can also be written:*

$$a \equiv b \bmod m$$

Definition 2.30. ***(Residue Classes)*** *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** $\mathbb{Z}/m\mathbb{Z}$.*

Remark 2.31. *Later we will know that $m\mathbb{Z} := \{ma \mid a \in \mathbb{Z}\}$ is a subgroup of $(\mathbb{Z}, +)$ is a subgroup as shown in Example 3.31 and $\mathbb{Z}/m\mathbb{Z}$ is the **quotient group**.*

There is a natural surjective map

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

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

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

Lemma 2.32. *Let $m \in \mathbb{N}$. Then, $\forall\, a, b, a', b' \in \mathbb{Z}$:*

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

Definition 2.33. ***($+$ and $\times$ on $\mathbb{Z}/m\mathbb{Z}$)***

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

Remark 2.34. ***(property of the new $+$ and $\times$)***

- *Our construction of $+$ and $\times$ on $\mathbb{Z}/m\mathbb{Z}$ is from $\mathbb{Z}$, hence they satisfy the **eight elementary properites** that $+$ and $\times$ satisfied on $\mathbb{Z}$. We have*

  1. *$[0] \in \mathbb{Z}/m\mathbb{Z}$ behaves like $0 \in \mathbb{Z}$,*

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

  2. *$[1] \in \mathbb{Z}/m\mathbb{Z}$ behaves like $1 \in \mathbb{Z}$:*

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

- *How about the "integral domain property"? We say that $[a] \in \mathbb{Z}/m\mathbb{Z}$ is non-zero if $[a] \neq [0]$. In general, $\mathbb{Z}/m\mathbb{Z}$ behave quite differently in this case. As an example, notice that*

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

  *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. $\mathbb{Z}/m\mathbb{Z}$ in general is **not an integral domain.***

  $*$ *When is $\mathbb{Z}/m\mathbb{Z}$ an integral domain? See below Corollary 2.40.*

Remark 2.35. *Why we can define operators for the residue classes from the original operators. The general reason is shown in Section 3.8.2 and Section 4.2.2, where **normal group** and **ideal** is introduced for cosets (e.g. residue classes).*

Theorem 2.36. *For every $m \in \mathbb{N}, a \in \mathbb{Z}$ the congruence*

$$ax \equiv 1 \bmod m$$

has a solution x (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 $au + mv = 1$. □

Remark 2.37. ***(muliplicative inverse)*** *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*

$[a]$ ***is invertible*** $\Leftrightarrow$ a ***is coprime to*** m***.***

Recall that $\times$ on $\mathbb{Q}$ had the extra property that all non-zero elements had multiplicative inverses. Analogously we see:

Corollary 2.38. *All non-zero elements of $\mathbb{Z}/m\mathbb{Z}$ have a multiplicative inverse $\iff m$ is prime.*

Proof. In $\mathbb{Z}/m\mathbb{Z}$, non-zero elements had multiplicative inverses $\iff \{1, 2, \cdots, m-1\}$ are all coprime to m. This can only happen if m is prime. □

Remark 2.39. ***(finite field)*** *The above corollary just means:*

$$\mathbb{Z}/m\mathbb{Z} \text{ is a field } \iff m \text{ is a prime.}$$

Corollary 2.40. ***(field is integral domain)*** *If m is prime then the product of two non-zero elements of $\mathbb{Z}/m\mathbb{Z}$ is again non-zero.*

Proof. See Corollary 4.39. Key is 0 is never invertible $\implies$ if x is invertible then $x \neq 0$ □

3 Group

3.1 Basic Definitions

Definition 3.1. ***(Binary Operator)*** *Let G be a set. A binary operation is a map of sets:*

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

For ease of notation we write $*(a, b) = a * b, \forall\, a, b \in G$.

Remark 3.2. *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 $(G, *)$.*

Definition 3.3. ***(Group)*** *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* $a * b = b * a = e$

Remark 3.4. ***(inverses)*** *Condition 3. indicates that the left and right inverses should be the same.*

Remark 3.5. ***(commutativity?)*** *Unlike $+$ in $\mathbb{Z}$, in general group does not require commutativity.*

Definition 3.6. ***(Abelian Group)*** *A group $(G, *)$ is called Abelian if it also satisfies the commutative property:*

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

Example 3.7. *We have seen different* ***Abelian group*** *examples thus far:*

- $(\mathbb{Z}, +)$
- $(\mathbb{Q}, +)$

- $(\mathbb{Q}\backslash\{0\}, \times)$
- $(\mathbb{Z}/m\mathbb{Z}, +)$
- $(\mathbb{Z}/m\mathbb{Z}\backslash\{[0]\}, \times)$
- *A real vector space under* $+$.
- ***trivial group**: a set with a **single** element admits one possible binary operation.*

$*$ *There exist non-Abelian group, see Definition 3.11 for general linear group.*

$*$ *Note that* $(\mathbb{Z}, \times)$ *is **not a group**. It is a monoid.*

Definition 3.8. ***(Monoid)*** *A monoid 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* $a * e = e * a = a, \forall\, a \in G$.

Remark 3.9. ***(monoid vs. group)*** *Monoid does not need existence of inverses for every element. A group is a monoid in which every element is invertible. Please also note monoid implies the closeness of the operator.*

Example 3.10. $(\mathbb{Z}, \times)$ *is a monoid but not a group.*

Definition 3.11. ***(General Linear Group)*** *The set of invertible* $n \times n$ *matrices with real entries, denoted* $\mathrm{GL}_n(\mathbb{R})$*, forms a group under matrix multiplication, which is not commute (i.e. not a Abelian Group).*

3.1.1 Map between Groups: Homomorphism, Isomorphism .etc.

Definition 3.12. ***(Homomorphism)*** *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* $f(x * y) = f(x) \circ f(y), \forall\, x, y \in G$.

Remark 3.13. *If* $G = H$ *and* $f = \mathrm{Id}_G$ *we call* f *the **identity homomorphism**.*

Remark 3.14. ***(explanation)*** *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.*

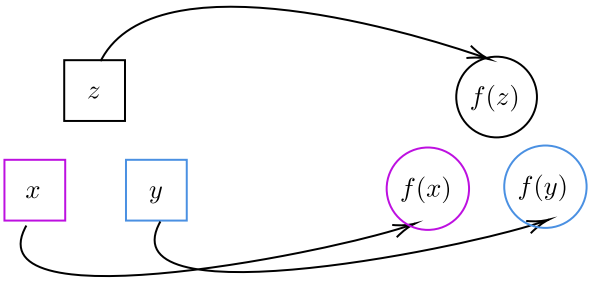


Fig. 1: Homomorphism, $z = x * y$ and $f(z) = f(x) \circ f(y)$.

Example 3.15. *Let* $H = \{e\}$ *be the trivial group, and let* $f(x) = e$ *for all* $x \in G$. f *is a homomorphism.*

Definition 3.16. ***(Isomorphism*** $\cong$***)*** *A homomorphism* $f : G \to H$ *which is **bijective** is called an **isomorphism**. Two groups* G *and* H *are said to be isomorphic if there exists an isomorphism between them, written as* $G \cong H$.

Definition 3.17. ***(Endomorphism)*** *A homomorphism from a group to itself (i.e.* $f : G \to G$*) is called an **endomorphism**.*

Definition 3.18. ***(Automorphism)*** *An endomorphism which is also an isomorphism is called an* ***automorphism****.*

Remark 3.19. ***(explanation)*** *Intuitively two groups being isomorphic means that they are the "same" group with* ***relabelled*** *elements.*

Lemma 3.20. ***(Composition of Homomorphism)*** *Let $(G, *), (H, \circ)$ and $(M, \Box)$ 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 $x, y \in G$. Then $gf(x * y) = g(f(x) \circ f(y)) = gf(x) \Box gf(y)$. □

Remark 3.21. ***(sets of endomorphisms: a monoid)***

- *The collection of endomorphisms of a group (a set of mappings) is a monoid under composition.*
- *The set of automorphisms of a group (a set of invertible mappings) is a group under composition. We denote it by $\mathrm{Aut}(G)$.*
- *$\mathrm{Aut}(G)$ is analogues to the collection of $n \times n$ invertible matrices being a group under matrix multiplication. Since invertible matrices can also be viewed as a linear invertible mapping between vector space (group).*

• ***Terminology Summary:***

- ***monomorphism*** $=$ *injective homomorphism*
- ***epimorphism*** $=$ *surjective homomorphism*
- ***isomorphism*** $=$ *bijective homomorphism*
- ***endomorphism*** $=$ ***homomorphism*** *of a group to itself*
- ***automorphism*** $=$ ***isomorphism*** *of a group with itself*

Theorem 3.22. ***(Uniqueness of** e**)*** *Let $(G, *)$ be a group. The identity element e is unique.*

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

Theorem 3.23. ***(Uniqueness of inverse)*** *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:

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

□

Remark 3.24. *Theorem 3.22 tells us that we can write $e \in G$ for the identity and it is well-defined. Similarly Theorem 3.23 tells us that for $a \in G$ we can write $a^{-1} \in G$ for the inverse in a well-defined way.*

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

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

Theorem 3.25. ***(Cancellation Law)*** *Let $a, b, c \in G$ a group. Then*

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

Remark 3.26. ***(cancel law for ring vs. cancel law for group)***

- *The cancel law for elements in rings is proven using the distributive law together with* ***not zero-divisor****. For more details see Remark 4.41 and Corollary 4.42. A special case is shown in Corollary 2.7.*
- *The cancel law for elements in groups is proven using associativity and existence of inverses.*

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

Theorem 3.27. ***(map of*** e ***and inverse in homomorphism)*** *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*

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

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 $e_H = f(e_G) = f\left(x * x^{-1}\right) = f(x) \circ f\left(x^{-1}\right)$ and $e_H = f(e_G) = f\left(x^{-1} * x\right) = f\left(x^{-1}\right) \circ f(x)$. Hence $f\left(x^{-1}\right) = (f(x))^{-1}$

□

Remark 3.28. ***(homomorphism*** $\Rightarrow$ ***subgroup)*** *For the exact definition of subgroup, see next section.*

1. f *is a homomorphism from G to H* $\Rightarrow$ *the image* $f(G)$ *is a subgroup of* H*:*

$$(f(x) \circ f(y)) \circ f(z) = f((x * y) * z) = f(x * (y * z)) = f(x) \circ (f(y) \circ f(z))$$

The other two conditions are verified in Theorem 3.27.

2. *injective homomorphism* $\Rightarrow$ *isomorphic to a subgroup: We may view G as a subgroup of H by identifying it with its image in* H *under* f*.*

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 3.29. ***(Subgroup)*** *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. $x \in H \Rightarrow x^{-1} \in H$

Remark 3.30. *A subgroup is naturally a group under the induced binary operation. It clearly has the same identity element.*

Example 3.31. ***(two important examples)***

- *If* $m \in \mathbb{N}$*, then the subset* $m\mathbb{Z} := \{ma \mid a \in \mathbb{Z}\}$ *is a subgroup of* $(\mathbb{Z}, +)$*.*
- *Vector space* V *over* $\mathbb{R}$ *is an Abelian group under* $+$*. If* W *is a subspace then it is a subgroup.*

Lemma 3.32. $H, K \subset G$ *subgroups* $\Rightarrow H \cap K \subset G$ *is a subgroup.*

Remark 3.33. *This result clearly extends to any (finite or infinite) collection of subgroups of G.*

Proof. We check the three conditions in the definition:

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

□

3.2.1 Equivalence Relation and Cosets

Definition 3.34. ***(Left Cosets: equivalence classes)*** *Let $(G, *)$ be a group and let $H \subset G$ be a **subgroup**. Let us define a **equivalence relation** on G using H as follows:*

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

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

Remark 3.35. ***(left coset vs. orbit)*** *Left coset is a special case of orbit, see Remark 3.81.*

Lemma 3.36. *This indeed is an equivalence relation on G.*

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

1. **symmetric**: $x \sim y \Rightarrow x^{-1} * y \in H \Rightarrow \left(x^{-1} * y\right)^{-1} \in H \Rightarrow y^{-1} * x \in H \Rightarrow y \sim x$
2. **reflexive**: $e \in H \Rightarrow x^{-1} * x \in H, \forall x \in G \Rightarrow x \sim x$
3. **transitive**: $x \sim y, y \sim z \Rightarrow x^{-1} * y, y^{-1} * z \in H \Rightarrow \left(x^{-1} * y\right) * \left(y^{-1} * z\right) \in H \Rightarrow x^{-1} * z \in H \Rightarrow x \sim z$

□

Lemma 3.37. *For $x \in G$ the equivalence class (or left coset) containing x equals*

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

Remark 3.38. *Some explanations:*

1. *Don't be confused with the notation $m\mathbb{Z}$. $m\mathbb{Z}$ is defined using the integer multiplication (or say using cyclic) as in Example 3.31 and $k(m\mathbb{Z})$ is then the equivalence class containing integer k.*
2. *Sometimes we use notation $x + H$ instead of xH.*
3. ***Right cosets** can be defined analogously as $Hx := \{h * x \mid h \in H\} \subset G$*

Proof. The easiest way to show that two subsets of G are equal is to prove containment in both directions:
1). $x \sim y \iff x^{-1} * y \in H \iff x^{-1} * y = h$ for some $h \in H \Rightarrow y = x * h \in xH$. Therefore $\{$equivalence class containing $x\} \subset xH$
2). $y \in xH \Rightarrow y = x * h$ 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 3.39. *Hence for $x, y \in G, xH = yH \iff x^{-1} * y \in H$.*

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

Remark 3.40. *We also have some interesting results:*

- *$y \in xH \Rightarrow yH = xH$. Hence left cosets can in general be written with different representatives at the front.*
- *The equivalence class containing $e \in G$ is just H. It is the **only equivalence class which is a subgroup**, as no other contains the identity.*

- *If $H = \{e\}$ then the left cosets are singleton sets. The collection of left cosets $= G$.*

Example 3.41. ***(vector subspaces example)*** *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 3.42. ***(Set of Left Cosets)*** *Let $(G, *)$ be a group and $H \subset G$ a subgroup. We denote*

$$G/H := \textbf{set of left cosets of } H \text{ in } G = \{xH \mid x \in G\}.$$

*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.*

Example 3.43. ***(canonical example of left cosets)***

- ***set of all residue classes** modulo m: $\mathbb{Z}/m\mathbb{Z}$. The subgroup $m\mathbb{Z} \subset Z$ has index m.*

- *The vector space in Example 3.41 is not finite index as there are infinitely many parallel planes in $\mathbb{R}^3$*

Remark 3.44. ***(Set of Left Cosets: a group?)*** *It depends. How to define the binary operator? We need H is a normal group. See Section 3.8.2. If it is, we call G/H a **quotient group.***

3.2.2 Bijection, Equal Sized Partition and Lagrange's Theorem

Theorem 3.45. *Let $x \in G$. The map (of sets)*

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

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 $g = x * h \Rightarrow g = \phi(h)$ □

Corollary 3.46. *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.** See Lemma 1.8. □

Theorem 3.47. ***(Lagrange's Theorem)***
*Let $(G, *)$ be a **finite** group and $H \subset G$ a subgroup. Then $|H|$ divides $|G|$.*

Remark 3.48. ***(inverse of Lagrange's Theorem)*** *Given any divisor of $|G|$ must there be a subgroup of that order? **No.** However if $p^n||G|$ with a prime number p, **Sylow's Theorem** shown in Theorem 3.103 and Remark 3.104 gives a **partial converse** to Lagrange's theorem.*

Proof. We can use H to define the above equivalence relation on G. We get a partition of G: set of left cosets G/H. 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.

Corollary 3.49. *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 3.50. ***(Generated Subgroup)*** *Let G be a group and $X \subset G$ be* ***any subset****. We define the subgroup generated by X to be the intersection of all subgroups of G containing X. We denote it by $\mathrm{gp}(X) \subset G$.*

Remark 3.51. ***(explanation: equivalent statement)*** *Just analogous to "closure" in [1], (Please listen to the audio in short algebra P17 too.)*

1. *$\mathrm{gp}(X)$ is the minimal subgroup containing X. By minimal we mean that if $H \subset G$ is a subgroup such that $X \subset H$ then $\mathrm{gp}(X) \subset H$.*
2. *A more constructive way of defining $\mathrm{gp}(X)$ is as all possible finite compositions of elements of X and their inverses.*

Example 3.52. *Consider the group $(\mathbb{Z}, +)$ and $X = \{1\} \subset \mathbb{Z}$. Then $\mathrm{gp}(X) = \mathbb{Z}$. This is the precise sense in which $\mathbb{Z}$ is "generated" by 1 under addition.*

Definition 3.53. ***(Finitely Generated)*** *We say a group $(G, *)$ is* ***finitely generated*** *if $\exists\, X \subset G$ that is finite such that $\mathrm{gp}(X) = G$.*

Remark 3.54.

- *Clearly all finite groups are finitely generated.*
- *$(\mathbb{Q}\backslash\{0\}, \times)$ is not finitely generated since there are infinitely many primes.*

Definition 3.55. ***(Cyclic)*** *A group $(G, *)$ is said to be* ***cyclic*** *if $\exists\, x \in G$ such that $\mathrm{gp}(\{x\}) = G$:*

$$G = \{x^n \mid n \in \mathbb{Z}\}.$$

Remark 3.56. *$(\mathbb{Z}, +)$, $(m\mathbb{Z}, +)$, and $(\mathbb{Z}/m\mathbb{Z}, +)$ are examples (m is not necessarily to be prime). Not all groups are cyclic, vector space is an example.*

Remark 3.57. *Let $(G, *)$ be a group (not necessarily cyclic) and $x \in G$. We call $\mathrm{gp}(\{x\}) \subset G$ the* ***subgroup generated by*** *x. By definition it is cyclic.*

Lemma 3.58. *Any group of prime order is cyclic generated by any* ***non-identity*** *element.*

Proof. Let G be a group of prime order p. Let x be a **non-identity** element of G. Then $\mathrm{gp}(\{x\}) \subset G$ is non-trivial and by Lagrange's theorem (Theorem 3.47) it must have order p. Hence $G = \mathrm{gp}(\{x\})$ □

Lemma 3.59. *All cyclic groups are Abelian.*

Proof. 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 $\mathrm{gp}(\{x\}) \subset G$ is Abelian. □

3.3.1 Cyclic Group: two cases

Theorem 3.60. *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. $G = \mathrm{gp}(\{x\})$, then $G = \left\{\cdots x^{-2}, x^{-1}, e, x, x^2 \cdots\right\}$. We have two cases to consider.

1. Assume all elements in this set are distinct (otherwise we turn to case 2 because of non-injective, see below.), then we can define a map of sets:

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

Then, $\forall\, a, b \in \mathbb{Z}, \phi\left(x^a * x^b\right) = \phi\left(x^{a+b}\right) = a + b = \phi\left(x^a\right) + \phi\left(x^b\right)$ so ϕ is a homomorphism. To show ϕ is a bijective: surjective is clear; injective is because if not injective, we then $x^a = x^b$ as below, and then G is finite. Thus, $(G, *)$ is isomorphic to $(\mathbb{Z}, +)$.

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

   Define the map:

   $$\begin{aligned} \phi : G &\to \mathbb{Z}/m\mathbb{Z} \\ x^n &\mapsto [n] \end{aligned}$$

   This is clearly a surjection, hence a bijection because $|G| = |\mathbb{Z}/m\mathbb{Z}| = m$ and Lemma 1.8. Again $\forall\, a, b \in \{0, \ldots, m-1\}$ we know $\phi\left(x^a * x^b\right) = \phi\left(x^{a+b}\right) = [a+b] = [a] + [b] = \phi\left(x^a\right) + \phi\left(x^b\right)$ is a homomorphism. Hence $(G, *)$ is isomorphic to $(\mathbb{Z}/m\mathbb{Z}, +)$.

□

Corollary 3.61. *Two finite cyclic groups of the same size are isomorphic.*

Corollary 3.62. *A subgroup H of a cyclic group G 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. $(G, *) \cong (\mathbb{Z}, +)$. Choose $m \in \mathbb{N}$ minimal such that $m \in H (m \neq 0)$. Hence $\textcircled{1}$: $\mathrm{gp}(\{m\}) = m\mathbb{Z} = \{ma \mid a \in \mathbb{Z}\} \subseteq H$. 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 $\textcircled{2}$: $H \subseteq m\mathbb{Z}$. So we get $m\mathbb{Z} = H$. Hence H is cylic.
2. $(G, *) \cong (\mathbb{Z}/m\mathbb{Z}, +)$. Again, choose $n \in \{1, ..., m-1\}$ minimal and positive such that $[n] \in H$. The same argument as above shows that the containment $\mathrm{gp}(\{[n]\}) \subseteq H$ is actually equality:

   Assume $\exists\, [x] \in H$ such that $x \notin \mathrm{gp}(\{[n]\})$. By the remainder theorem, $x = qn + r, r, q \in \mathbb{Z}$ and $0 < r < n \Rightarrow r \in H$. This is a contradiction by the minimality of n. Hence H is cyclic.

□

Theorem 3.63. *Let $(G, *)$ be a* ***finite cyclic group*** *of order d. Let $m \in \mathbb{N}$ such that m divides d. 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.

Uniqueness: Then if $n = d/m$ then $\mathrm{gp}(\{[n]\}) \subset \mathbb{Z}/d\mathbb{Z}$ 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 2) in Corollary 3.62 we know that the minimal $k \in \mathbb{N}$ such that $[k] \in H$ must be $\mathrm{gp}(\{[k]\}) = H$. $|H| = m$, and so $k = d/m$ according to the proof 2) in Corollary 3.62. We then have $k = n$, which means $\mathrm{gp}(\{[k]\}) = \mathrm{gp}(\{[n]\}$, i.e., they are the same group. □

Definition 3.64. *If $|\,\mathrm{gp}(\{x\})| < \infty$, define $\mathrm{ord}(x) := |\,\mathrm{gp}(\{x\})|$ (We say that x is of finite order). If not we say that x is of infinite order.*

Corollary 3.65.

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

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

- *If $x^n = e$, we have $\mathrm{ord}(x) | n$.*
- *$e \in G$ is the only element of G of order 1.*
- *The only element with finite order in $\mathbb{Z}$ is 0 .*

Corollary 3.66. *Let $(G, *)$ be a finite group and $x \in G$. Then $\operatorname{ord}(x)$ divides $|G|$ and $x^{|G|} = e$.*

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

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

□

3.4 Permutation Groups and Group Actions

Definition 3.67. ***(Permutation Group** $(\Sigma(S), \circ)$**)*** *Let S be a **set**. We define the **permutations group of S** to be the **set of bijections** from S to itself, denoted $(\Sigma(S), \circ)$, where the group binary operation $\circ$ is **composition of functions**.*

Remark 3.68. ***(explanation)***

1. *S is a set **not necessarily a group**. If S is a group, see Remark 3.21 for a similar definition of **group of automorphisms** (of course bijection) which is a group under composition.*
2. *By composition of functions we always mean on the left, i.e. $\forall\, f, g \in \Sigma(S)$ and $s \in S, (f \circ g)(s) = f(g(s))$*
3. *Verify $(\Sigma(S), \circ)$ is a group: 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.*
4. *Let $n \in \mathbb{N}$. We write **finite symmetric groups** $\operatorname{Sym}_n := \Sigma(\{1, 2, \ldots, n\})$. If S is any set of cardinality n then*
$$\Sigma(S) \cong \operatorname{Sym}_n$$
*The **isomorphism** is induced by writing a bijection from S to $\{1, 2, \ldots, n\}$.*

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 3.69. ***(Group Action)*** *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). $\mu(e, s) = s$

Remark 3.70. *μ is a map defined on the product space $G \times S$. However μ is quite special because the two conditions are very strict:*

- *Condition (1). has limit μ to the case that it can be write as composition of function:*

  If we define $x(s) := \mu(x, s), \forall\, x \in G, s \in S$, Condition (1). becomes:

  $$\text{and } (x * y)(s) = x(y(s)), \forall\, x, y \in G, s \in S$$

- *Condition (2). has limit that e in G must be an identity mapping.*

Remark 3.71. *G is a general group and **not necessarily the permutations group** $\Sigma(S)$. See also Section 3.4.1 example 1. for the case if $G = \Sigma(S)$.*

3.4.1 Examples of Group Action

We introduce some examples, which are very important. In example 3, 4 and 5, $S = G$.

1. There is a natural action of $\Sigma(S)$ on S :

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

This is where the definition group action comes from.

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

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

3. **left regular representation of G:** Let $(G, *)$ be a group. There is a natural action of G on itself:

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

Condition (1) holds as $*$ is associative. Condition (2) holds because e $*x = x, \forall x \in G$. Note here we can extend it to arbitrary subgroup H of G:

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

4. **right regular representation of G:** For any subgroup H of G:

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

See Remark 3.81 for more details.

5. **conjugation:** There is another natural action of G on itself:

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

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

3.4.2 Action of G on S = homomorphism from G to Permutation Group of S

Bijection φ_g:

Let $\mu : G \times S \to S$ an action of a group G on a set S. For any $g \in G$, define $\varphi_g(\cdot) := \mu(g, \cdot) = g(\cdot)$:

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

Remark 3.72. *Condition (1) of μ implies that $\varphi_{g*h} = \varphi_g(\varphi_h), \forall g, h \in G$. Here the second term is composition of functions. Similarly Condition (2) tell is that $\varphi_e = \mathrm{Id}_S$.*

Lemma 3.73. *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 in Remark 3.72 it is clear that $\varphi_{g^{-1}}$ is inverse to φ_g □

Remark 3.74. *Since for each g, φ_g is a bijection from S to S, we have $\varphi_g \in \Sigma(S)$ for any $g \in G$.*

Homomorphism φ:

Given group action μ, we define $\varphi(g) := \varphi_g$:

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

Lemma 3.75. *φ is a homomorphism.*

Proof. As we have just seen, Condition (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. □

Remark 3.76. *So an action μ 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: Given homorphism $\varphi(g) = \varphi_g$, we then construct $\mu(g, s) = \varphi_g(s)$. Both concepts are interchangeable.*

"action of G on S" = "homomorphism from G to the permutation group $(\Sigma(S), \circ)$"

Definition 3.77. ***(Faithful)*** *An action of G on S is called faithful if*

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

*is **injective**.*

Remark 3.78. *If G acts faithfully on S then G is isomorphic to a subgroup of $\Sigma(S)$. See Remark 3.28 for details.*

Theorem 3.79. ***(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 $\varphi : G \to \Sigma(G)$ 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

Similar to what we have done in Section 3.2.1, here we define an equivalence relation and then get a partition:

3.5.1 Equivalence Relation on S and Orbit

Definition 3.80. ***(Equivalence Relation using Action)*** *Let $(G, *)$ be a group, together with an action φ on a set S. We can define an equivalence relation on S by*

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

Remark 3.81. ***(explanation and comparison)***

- *In Section 3.2.1, the equivalence relation left cosets is defined on a **group** G.*
- *Please note here S is a general **set** without group structure. It however has an auxiliary group G worked as the action.*
- *However, equivalence relation left cosets can be viewed as **a sepcial case** of equivalence relation using the action, **right regular representation of** G:*

  For a subgroup H of G, define:

  $$\begin{aligned} \mu : H \times G &\to G \\ (h, x) &\mapsto x * h \end{aligned}$$

  left coset: *we then have $x \sim y$ iff $\exists\, h \in H$ such that $y = x * h \Leftrightarrow x \sim y$ iff $x^{-1} * y \in H$.*
- *Similarly, for a subgroup H of G, define the **left regular representation** of G:*

  $$\begin{aligned} \mu : H \times G &\to G \\ (h, x) &\mapsto h * x \end{aligned}$$

  right coset: *we then have $x \sim y$ iff $\exists\, h \in H$ such that $y = h * x \Leftrightarrow x \sim y$ iff $y * x^{-1} \in H$.*

Lemma 3.82. *This indeed is an equivalence relation on S.*

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

1. **symmetric**: $s \sim t \Rightarrow \exists\, g(s) = t \Rightarrow g^{-1}(t) = s \Rightarrow t \sim s$
2. **reflexive**: for $e \in G$, $e(s) = s, \forall\, s \in S \Rightarrow s \sim s$
3. **transitive**: $s \sim t, t \sim u \Rightarrow g(s) = t, h(t) = u \Rightarrow h(g(s)) = u \Rightarrow s \sim u$ since $h \circ g \in G$

□

Definition 3.83. ***(Orbit: equivalence class)*** *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**. We definite **orbit of** s as:*

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

which is the equivalence class containing $s \in S$.

Remark 3.84. $\mathrm{Orb}(s)$ *is a subset of S and hence is merely a set with no extra structure (not a group).*

Definition 3.85. ***(Transitive)*** *Let $(G, *)$ be a group, together with an action φ on a set S.*

- *We say that G acts **transitively** on S if there is **only one orbit.***
- *Equivalently, φ is **transitive** if $\forall\, s, t \in S$, $\exists\, g \in G$ such that $g(s) = t$*

Example 3.86.

1. *The natural action of $\Sigma(S)$ on S in Example 1 in Section 3.4.1 is **transitive**. This is clear because given any two points in a set S there is always a bijection which maps one to the other.*
2. *If G is not the trivial group (the group with one element) then **conjugation** in Section 3.4.1 is never transitive. To see this observe that under this action $\mathrm{Orb}(e) = \{e\}$ since $s * e * s^{-1} = e, \forall\, s \in S$.*
3. ***right regular representation of** G (left coset) is **not transitive** unless $H = G$.*

Remark 3.87. ***(equal sized equivalent class?)*** ***right or left regular representation of** G gives equal sized equivalent class (left or right cosets, partitions) as shown in Section 3.2.2. However Orb does **not** give this property on S, e.g. the above conjugation.*

3.5.2 Subgroup Stabiliser and Orbit-Stabiliser Theorem

Definition 3.88. ***(Stabiliser)*** *Let $(G, *)$ be a group, together with an action φ on a set S. Let one specific $s \in S$. We define the **stabiliser subgroup** of s to be all elements of G which **fix** s **under the action**. More precisely*

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

Lemma 3.89. $\mathrm{Stab}(s)$ *is a subgroup of G.*

Proof. Just check the 3 conditions for group:

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

□

Thus we may form the **set of left cosets of** $\mathrm{Stab}(s)$ **in** G:

$$G/\mathrm{Stab}(s) := \{x\,\mathrm{Stab}(s) \mid x \in G\}$$

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

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

hence they partition G into **disjoint subsets with equal size**.

Lemma 3.90. *Let $x, y \in G$ then $x \operatorname{Stab}(s) = y \operatorname{Stab}(s) \iff x(s) = y(s)$.*

Remark 3.91. ***(explanation: not important, just some thinking, Lemma 3.93 is the key)***

1. *What does it mean x and y are in the same left coset of stabiliser? For this s, we have the "inter state" $m \in S$: $m = x(s)(= y(s)), \forall\, x, y$ in the same left coset equivalent class. If x has map s to $m = x(s)$, for all y in the same class, we can map it back to s: $y^{-1}(x(s)) = s$.*

2. *Since for all $g \in \operatorname{Stab}(s), g(s) = s$, we then have $x(g(s)) = x(s)$.*

*Roughly speaking, the above two together tell that any element in the **same left coset** $\{x \circ g \mid g \in \operatorname{Stab}(s), x\}$ is a map that maps s to the **same element** $x(s) \in S$.*

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

Definition 3.92. *From above, we deduce that there is a well defined map (of sets):*

$$\begin{aligned} \phi : G/\operatorname{Stab}(s) &\to \operatorname{Orb}(s) \\ x \operatorname{Stab}(s) &\mapsto x(s) \end{aligned}$$

Lemma 3.93. *ϕ is a bijection.*

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

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

Therefore ϕ is injective. □

This immediately gives the following key result:

Theorem 3.94. ***(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 ($|\operatorname{Orb}(s)| < \infty$). Then $\operatorname{Stab}(s) \subset G$ is of finite index and*

$$(G : \operatorname{Stab}(s)) = |\operatorname{Orb}(s)|$$

Proof. Immediate from previous lemmas. □

We have the following corollary:

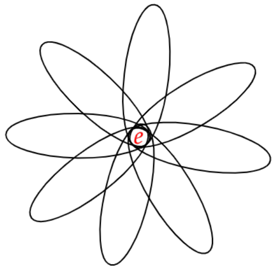
Corollary 3.95. *If $(G, *)$ is a finite group acting on a set S and $s \in S$ then*

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

Proof. In this case $(G : \operatorname{Stab}(s)) = |G|/|\operatorname{Stab}(s)|$. Applying the Orbit-Stabiliser Theorem yields the result. □

Remark 3.96. ***(more discussion)*** *For $t \in \operatorname{Orb}(s)$ and $t \neq s$, we then have $|\operatorname{Stab}(s)| = |\operatorname{Stab}(t)|$ since $\operatorname{Orb}(s) = \operatorname{Orb}(t)$. That just means if two elements are in the same orbit, the size of the stabiliser subgroup must be the same. The question is "does $\operatorname{Stab}(t)$, t over one orbit $\operatorname{Orb}(s)$, are partitions of G?" The answer is **No**. You can think of $\operatorname{Stab}(t)$ as a star shaped structure with the intersection not empty (at least e is inside). Also note in general $\cup_{t \in \operatorname{Orb}(s)} \operatorname{Stab}(t) \subsetneq G$ since **disjoint** $\{x \operatorname{Stab}(s)\}$ partition G and for each coset the size equals $\operatorname{Stab}(t)$, but the number of coset $=$ the number of $\operatorname{Stab}(t)$, $t \in \operatorname{Orb}(s)$ (**not disjoint**).*

- ⋆ *If G acts transitively on S, we have $\operatorname{Stab}(s) = \{e\}$ for all s.*
- ⋆ *If $|\operatorname{Orb}(s)| = 1$, we have $G = \operatorname{Stab}(s)$.*



3.5.3 Application of Orbit-Stabiliser Theorem

The orbit-stabiliser theorem allows us to prove non-trivial results about the structure of finite groups.

Example 3.97. ***(conjugacy classes)*** *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 $h \in G$,*

$$\operatorname{Conj}(h) := \operatorname{Orb}(h) = \left\{ g * h * g^{-1} \mid g \in G \right\}.$$

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 $|C_i| \mid |G|, \forall\, i \in \{1, \cdots, r\}$ because of Orbit-Stabiliser Theorem.

Remark 3.98. *$|C_i|$ may be different as we have mentioned in Remark 3.87. Conjugacy classes are not groups.*

Example 3.99. ***(similarity = conjugacy classes)*** *If $G = \mathrm{GL}_n(\mathbb{R})$, general linear group Definition 3.11, then two matrices are in the same conjugacy class if and only if they are similar.*

Definition 3.100. ***(Group Center)*** *Let $(G, *)$ be a group. The* ***center*** *of G is the subgroup*

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

Proof. Easy to prove it is a group. And also note that e must be in $Z(G)$. □

Lemma 3.101. *G is Abelian $\Leftrightarrow$ the center is G itself.*

Theorem 3.102. *Let G be a p-group, i.e., a finite group of order p^n, for p a prime number and $n \in \mathbb{N}$. Then the center is non-tivial: $Z(G) \neq \{e\}$.*

Proof. Let G act on itself by conjugation. We know $Z(G)$ is a subgroup. Observe that $h \in Z(G) \iff \operatorname{Conj}(h) = \{h\}$. Assume that $Z(G) = \{e\}$. Hence if $h \neq e$ (i.e. $h \notin Z(G)$) then $|\operatorname{Conj}(h)| > 1$ (note h is always inside $\operatorname{Conj}(h)$). By the Orbit-Stabiliser Theorem we know that $|\operatorname{Conj}(h)|$ must divide p^n. Hence p divides $|\operatorname{Conj}(h)|$. 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 since it means p and p are coprime, hence $Z(G)$ cannot be trivial. □

Theorem 3.103. ***(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 size p^n.*

Remark 3.104. ***(Sylow vs. Lagrange)***

- *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.*
- *Sylow's Theorem is a partial converse to Lagrange's theorem.*

Proof. Assume that $|G| = p^n m$, where $m = p^r u$ with $\mathrm{HCF}(p, u) = 1$. **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 **family of all subsets** of G of size p^n. An element of S is an unordered p^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 arbitrary ordering $\omega = \{\omega_1, \cdots, \omega_{p^n}\} \in S$, then for $g \in G$, define

$$g(\omega) := \{g * \omega_1, \cdots, g * \omega_{p^n}\}.$$

In other words, g maps a set to a set.

- We first claim that $\textcircled{1}: \forall\, \omega, |\operatorname{Stab}(\omega)| \leq p^n$. To see this define the function

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

We note that $g(\omega) = \omega, \forall\, g \in \operatorname{Stab}(\omega)$. By the cancellation property for groups f is an injective map. Hence $|\operatorname{Stab}(\omega)| \leq |\omega| = p^n$.

- 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, and that $p^n m - j$ and $p^n - j$ have the same number of p factors, namely the number of p factor of j:

Let $0 \leq k \leq n-1$ be the max such that $p^k | j$. We have $j = p^k a$, with $\operatorname{HCF}(p, a) = 1$.

$$p^k | p^n m - j \Leftrightarrow p^k | j \Leftrightarrow p^k | p^n - j$$

We know $p^k | p^n - j$ with k be the max possible integer (easy to get). We need to prove $p^k | p^n m - j$ is also the max possbile one: $p^n m - j = p^n m - p^k a = p^k(p^{n-k} m - a)$, where there is no factor p in $p^{n-k}m - a$ since $\operatorname{HCF}(p, a) = 1$.

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 $\operatorname{HCF}(p, v) = 1$.

Now recall that S is the disjoint union of the orbits of our action of G on S. Hence there must $\exists \omega \in S$ such that $|\operatorname{Orb}(\omega)| = p^s t$, where $s \leq r$ and $\operatorname{HCF}(p, t) = 1$. By the **Orbit-Stabilizer Theorem** we know that $|\operatorname{Stab}(\omega)| = p^{n+r-s} \frac{u}{t}$. Because $|\operatorname{Stab}(\omega)| \in \mathbb{N}$ and u and t are coprime to p, we deduce that $\frac{u}{t} \in \mathbb{N}$. Hence $\textcircled{2}: |\operatorname{Stab}(\omega)| \geq p^n$.

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

Lemma 3.105. *For any group $G, G/Z(G)$ is cyclic iff G is Abelian, or in otherwords: the quotient $G/Z(G)$ can never be non-trivial cyclic.*

Remark 3.106. *Note, $Z(G)$ is **normal** is proved in Remark 3.148.*

Corollary 3.107. *If Z is contained in $Z(G)$ and G/Z is cyclic, then G is Abelian.*

Proof. First step: If $G/Z(G)$ is cyclic with generator $xZ(G)$, show that every element of G can be written in the form $x^a z$ for some $a \in \mathbb{Z}$ and some element $z \in Z(G)$:

We have that $G/Z(G)$ is cyclic, and so there is an element $x \in G$ such that $G/Z(G) = \operatorname{gp}(xZ(G))$, where $xZ(G)$ is the coset with representative x. Now let $g \in G$. We know that $gZ(G) = (xZ(G))^m$ for some m, and by definition $(xZ(G))^m = x^m Z(G)$. We have that $gZ(G) = x^m Z(G)$, and this happens if and only if $g = x^m z$ for some $z \in Z(G)$.

Second step: $g, h \in G$ implies that $g = x^{a_1} z_1$ and $h = x^{a_2} z_2$, so

$$\begin{aligned} gh &= (x^{a_1} z_1)(x^{a_2} z_2) \\ &= x^{a_1} x^{a_2} z_1 z_2 \\ &= x^{a_1 + a_2} z_2 z_1 \\ &= \ldots = (x^{a_2} z_2)(x^{a_1} z_1) = hg. \end{aligned}$$

Third step: But then from Lemma 3.101), it just means $G/Z(G) = \{e\}$.

□

Theorem 3.108. *Showing any group with p or p^2 elements are both Abelian. Note, $|Z(G)| = p$ cannot happens.*

Proof. If the size is p, it just means it is cyclic following from Lemma 3.58. But then Lemma 3.59 indicates it is Abelian.

If the size if p^2, we then have $Z(G)$ is non-trivial from Theorem 3.102, so $|Z(G)| = p$ or p^2. If $Z(G) = p^2$, it just means G is Abelian following from Lemma 3.101. If $|Z(G)| = p$, we have $G/Z(G)$ is cyclic. From Lemma 3.105, we have G is Abelian. But then $Z(G) = G$, so $|Z(G)| = p$ cannot happens. □

3.6 Finite Symmetric Groups

As Theorem 3.79 (Cayley's Theorem) shows, if $(G, *)$ is a finite group of order n. Then G is isomorphic to a subgroup of Sym_n, the symmetric group on $\{1, 2 \ldots n\}$. Hence to properly understand finite groups we must understand these finite symmetric groups.

Lemma 3.109. *For $n \in \mathbb{N}$, $|\mathrm{Sym}_n| = n!$*

3.6.1 Representation of Elements in Sym_n: cycles

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

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

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

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

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

Example 3.110. *Let $n = 5$ then $\sigma = (123)(45)$ corresponds to*

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

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

Example 3.111. *Let $n = 5$ then $\sigma = (523)$ corresponds to*

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

This notation makes it clear how to compose two permutations.

Example 3.112. *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.*

Remark 3.113. *This example also shows that in general Sym_n is **not Abelian.***

Definition 3.114. ***(Cycle Structure)*** *Hence, given $\sigma \in \mathrm{Sym}_n$, 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.*

Remark 3.115. ***(explanation)*** *Cycle structure is a set of numbers (not a collection of set). There is no order since any order is okay (corresponds to reordering the cycles).*

Lemma 3.116. *Let $\sigma \in \operatorname{Sym}_n$ decompose as the disjoint product of cycles of length $n_1, ..n_m$ (so $\sum n_i = n$) . Then $\operatorname{ord}(\sigma) = \operatorname{LCM}(n_1, \ldots n_m)$.*

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** $m \in \mathbb{N}$ **must have order** m **in** Sym_n . Also recall that if G is a finite group then for any $d \in \mathbb{N}, x \in G, x^d = e \Longleftrightarrow \operatorname{ord}(x) \mid d$. 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 $\sigma^d = e \Longleftrightarrow n_i \mid d, \forall i$. The smallest value d can take with this property is $\operatorname{LCM}(n_1, \ldots n_m)$. □

Theorem 3.117. *Two permutations are* ***conjugate*** *in Sym_n if and only if they have the* ***same cycle structure****.*

Proof. Let $\sigma, \tau \in \operatorname{Sym}_n$ 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 $\alpha \in \operatorname{Sym}_n$ 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 3.118. *Conjugacy classes in Sym_n are indexed by cycle structures (i.e. partitions of n).*

Proof. Immediate from the above. □

3.6.2 Transposition, Even and Odd Permuation

Definition 3.119. ***(Transposition)*** *A transposition is a cycle of length 2 .*

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

$$(a_1, a_2, \cdots, a_n) = (a_1, a_n)(a_1, a_{n-1}) \cdots (a_1, a_3)(a_1, a_2) \tag{3.120}$$

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

Remark 3.121. *Notice that in the above example, both expressions involve an* ***even*** *number of transpositions.*

Definition 3.122. ***(Sign: even or odd)*** *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 $\operatorname{sgn}(\sigma)$, to be $+1$ or -1 depending on whether σ is even or odd.*

Remark 3.123. *This indeed is a definition because the disjoint product cycles representation is unique (up to internal shifts and reordering the cycles)*

Remark 3.124. ***(equivalent statements)***

- ***even:*** *number of product of transpositions is even*
- ***odd:*** *number of product of transpositions is odd.*

For one cycle, if it has even length, it can be rewritten as an odd number of transpositions. Two even cycle $\Rightarrow$ an even number of transpositions. So an ***odd*** *number of* ***even*** *length cycles $\Rightarrow$ an* ***odd*** *number of transpositions.*

Why this equivalent statement is also a definition? If we take (3.120) *as the canonical decomposition, of course we make the above as a definition. In the general case, $\sigma \in \operatorname{Sym}_n$ may be written as some arbitrary product of transpositions, we need the last paragraph below in the proof of Theorem 3.125.*

Theorem 3.125. *Let $\sigma \in \operatorname{Sym}_n$ 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 $2 \mid (m - n)$.*

Remark 3.126. ***(comparison of determinant and similarity of matrix)***

1. *Another proof of Theorem 3.125 use an identity matrix and its determinant, switch (i,j) means switch two rows. Determinant must be either even or odd. It cannot be even and odd at the same time.*

2. *Also note conjugacy classes plays the role of the similarity in* $\mathrm{GL}_n(\mathbb{R})$ *as shown in Example 3.99. Similarity in* $\mathrm{GL}_n(\mathbb{R})$ *will guarantees the* ***same sign of determinant****.*

3. *Here for conjugacy classes in* Sym_n*, we also have similar properties as indicated by Corollary 3.118. This is also why Item 1 can prove Theorem 3.125.*

4. ***Key****: Conjugacy classes in* Sym_n *is a special case of similarity where the* ***domain is only permutation matrix.***

Proof. First notice that a cycle of length r can be written as the product of $r-1$ transpositions by (3.120). 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.

$$(1,i)(1,2,\cdots,i,\cdots,r)=(1,2,\cdots,i-1)(i,i+1,\cdots,r)$$

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, $\operatorname{sgn}((1,i)\sigma)=-\operatorname{sgn}(\sigma)$

(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

$$(1,i)(1,2,\cdots,i-1)(i,\cdots,r)=(1,\cdots,r)$$

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 $\operatorname{sgn}((1,i)\sigma)=-\operatorname{sgn}(\sigma)$ 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 ① **(existence): 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. ② **(uniqueness)**: This is a contradiction as $\operatorname{sgn}(e)=1$. □

Remark 3.127. *We should observe that from the proof of the above we see that*

$$\forall\,\sigma,\tau\in\mathrm{Sym}_n,\operatorname{sgn}(\sigma\tau)=\operatorname{sgn}(\sigma)\operatorname{sgn}(\tau).$$

Because $\operatorname{sgn}(e)=1$ *we deduce that* $\operatorname{sgn}(\sigma)=\operatorname{sgn}\left(\sigma^{-1}\right)$ *for all* $\sigma\in\mathrm{Sym}_n$.

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 3.128. ***(Alternating Group)*** *The subgroup of* $\mathrm{Alt}_n\subset\mathrm{Sym}_n$ *consisting of* ***even*** *elements is called the* ***alternating group*** *of rank n.*

Remark 3.129. *So* Alt_n *is the permutation matrix with positive determinant.*

Observe that Alt_n contains all 3-cycles (cycles of length 3). Further more we have

Lemma 3.130. Alt_n *is generated by* 3*-cylces.*

Proof. As any element of Alt_n can be written as the product of transpositions, we only have to do it for the product of two (because of even) transpositions. There are two cases:

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

$\square$

Lemma 3.131. $|\mathrm{Alt}_n| = \frac{n!}{2}$

Proof. Recall that $|\mathrm{Sym}_n| = n!$, hence we just need to show that $(\mathrm{Sym}_n : \mathrm{Alt}_n) = 2$. Let $\sigma, \tau \in \mathrm{Sym}_n$. Recall that

$$\sigma \,\mathrm{Alt}_n = \tau \,\mathrm{Alt}_n \iff \sigma^{-1}\tau \in \mathrm{Alt}_n$$

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

$$\sigma \,\mathrm{Alt}_n = \tau \,\mathrm{Alt}_n \iff \mathrm{sgn}(\sigma) = \mathrm{sgn}(\tau)$$

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

Remark 3.132. *The alternating groups for $n \geq 5$ have a very special property called* ***simple*** *Definition 3.152.*

3.7 Symmetry of Sets with Extra Structure

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

For vector space $\mathbb{R}^n$:

- The permutation group $\Sigma\left(\mathbb{R}^n\right)$ does not take this into account and elements do not need to be a linear invertible mapping.
- $\mathrm{GL}_n(\mathbb{R}) \subset \Sigma\left(\mathbb{R}^n\right)$ only includes linear invertible mapping (permutations). These are permutations which preserve the vector space stucture under "addition" and "multiplication".

3.7.1 Symmetry in Euclidean Space: isometry, symmetry group

Definition 3.133. *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 3.134. ***(Euclidean Distance)*** *The Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ is*

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

and the norm (length) of $\boldsymbol{x}$ is:

$$||\boldsymbol{x}|| := d(\boldsymbol{x}, \mathbf{0})$$

Definition 3.135. ***(Isometry)*** *An isometry of $\mathbb{R}^n$ is a map of sets $f : \mathbb{R}^n \to \mathbb{R}^n$* ***(not necessarily linear****) such that $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, d(\boldsymbol{x}, \boldsymbol{y}) = d(f(\boldsymbol{x}), f(\boldsymbol{y}))$, where d is the Euclidean distance function. The collection of all isometries of $\mathbb{R}^n$ is denoted by* $\mathrm{Isom}\left(\mathbb{R}^n\right)$.

Remark 3.136. ***(some facts about isometry)*** *We statement the following without proof. See also [2, page 85 and 320] and the supp material.*

- *Under the* ***Euclidean distance*** *in Definition 3.134, every isometry of $\mathbb{R}^n$ is a composition of an origin fixing isometry (linear isometry) and a translation. We may call the composition as the* ***affine mapping*** *written as $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{y}$:*
  1. ***origin fixing isometry:*** *$f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ must be* ***orthogonal****.*
  2. ***translation:*** *$f(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{y}$ for some $\boldsymbol{y} \in \mathbb{R}^n$*

- *The composition of any two isometries is an isometry.*
- *All isometries are bijective and their inverses are isometries. This means* $\mathrm{Isom}\left(\mathbb{R}^n\right)$ *is a* ***subgroup*** *of* $\Sigma\left(\mathbb{R}^n\right)$ *.*
- *Give that* $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ *where* $\boldsymbol{A}$ *is* ***orthogonal****. We then have that the* ***positive scalar multiples of the Euclidean norm*** *are the only* ***unitarily invariant norms*** *on* $\mathbb{R}^n$*.*

Definition 3.137. ***(Symmetry Group)*** *Let* $X \subset \mathbb{R}^n$ *be a subset (not necessarily a subspace). We define the* ***symmetry group*** *of* X *to be the subgroup* $\mathrm{Sym}(X) \subset \mathrm{Isom}\left(\mathbb{R}^n\right)$ *with the property that* $f \in \mathrm{Sym}(X)$ *if and only if* f *permutes* X*:*

$$\mathrm{Sym}(X) = \Sigma(X) \cap \mathrm{Isom}(\mathbb{R}^n)$$

Remark 3.138. *Note in the following* $\Sigma(X)$ *in general is* ***not*** *a subset of* $\mathrm{Isom}\left(\mathbb{R}^n\right)$*.*

- *Note we have* $\mathrm{Sym}(X) \subset \mathrm{Isom}\left(\mathbb{R}^n\right) \subset \Sigma(\mathbb{R}^n)$*.*
- *Define* $\mathrm{Isom}^*\left(\mathbb{R}^n\right) := \{f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A} \text{ is orthogonal}\}$*. We have* $\mathrm{Isom}^*\left(\mathbb{R}^n\right) \subset \mathrm{GL}_n(\mathbb{R}^n) \subset \Sigma(\mathbb{R}^n)$*.*
- $\mathrm{Sym}(X) \subset \Sigma(X) \subset \Sigma(\mathbb{R}^n)$*.*
- $\mathrm{Sym}(X)$ *(Symmetry Group) vs.* Sym_k *(Permuation Group,* Σ*):* $\mathrm{Sym}(X) \subset \mathrm{Sym}_k$ *for finite* $|X| = k$*.*

There is a natural action of $\mathrm{Sym}(X)$ on the set X, coming from the fact there is a natural homomorphism $\mathrm{Sym}(X) \to \Sigma(X)$. homomorphism g :

$$\begin{aligned} \mathrm{Sym}(X) &\to \Sigma(\mathbb{R}^n) \\ f &\mapsto f \end{aligned}$$

$\mathrm{Sym}(X)$ measures how much symmetry X has. The more symmetric X, the larger $\mathrm{Sym}(X)$.

3.7.2 Dihedral Group in $\mathbb{R}^2$

Definition 3.139. ***(Dihedral Group: regular*** m***-gon)*** *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*.*

Remark 3.140. ***(***D_m***: subgroup of orthogonal matrices)*** *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.*

1. $f \in D_m$ is a group action **faithfully** on set X: see Example 1 in Section 3.4.1 with a D_m being a subgroup of $\Sigma(X)$
2. $f \in D_m$ is a group action **transitive** on set X: since only one orbit = set of verticals .

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

$$\mathrm{Rot}_m = \left\{e, \sigma, \sigma^2, \cdots, \sigma^{m-1}\right\} \subset D_m$$

Remark 3.141. Rot_m *a* ***subgroup*** *and is cyclic of order* m*.*

Given a vertex $a, \mathrm{Stab}(a) = \{e, \tau\}$, where τ is the reflection through the straight line containing a and the origin. We have $|D_m : \mathrm{Stab}(a)| = m$ since the maximum possible value is m and we can achieve it. Note this is because Lemma 3.93, and for $\mathrm{Orb}(s)$ is set of all vertex.

By the Orbit-Stabilizer theorem $|D_m| = 2m$, hence $(D_m : \mathrm{Rot}_m) = 2$. We deduce that

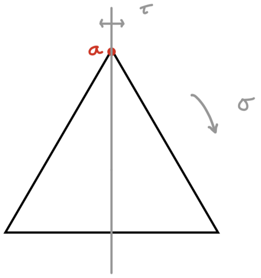
$$D_m = \mathrm{Rot}_m \coprod \tau\,\mathrm{Rot}_m$$

The left coset $\tau\,\mathrm{Rot}_m$ 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

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

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

Fig. 2: D_3

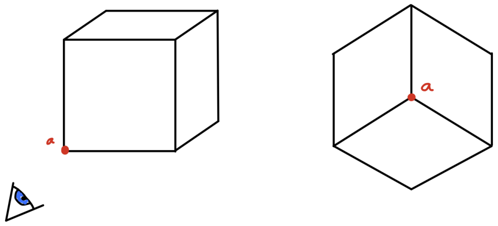


3.7.3 The Cube in $\mathbb{R}^3$

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

Again $\mathrm{Sym}(X)$ acts faithfully and transitively on the vertices. If $a \in X$ is a vertex, then $\mathrm{Stab}(a)$ can naturally be identified with D_3 (see below figure, not D_6) which has size 6 . Hence, by the Orbit-Stabilizer theorem, $|\,\mathrm{Sym}(X)| = 48$ since $\mathrm{Orb}(s) = 8$.

The same logic applies to $\mathrm{Rot}_{\square}$, **the rotational symmetries**, although the stabilizer of a now has size 3 . This tells us that $|\mathrm{Rot}_{\square}| = 24$.



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

$$\mathrm{Sym}(X) = \mathrm{Rot}_{\square} \coprod \tau\, \mathrm{Rot}_{\square}$$

It can be shown that $\tau\sigma = \sigma^{-1}\tau$ for all $\sigma \in \mathrm{Rot}_{\square}$. Thus it remains to determine the group structure of $\mathrm{Rot}_{\square}$

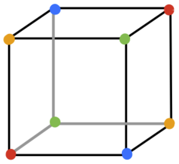
Color the vertices with four colors, making sure that opposite vertices have the same color (see below figure). Rotational symmetries act on this set of four colors, inducing a homomorphism from $\mathrm{Rot}_{\square}$ to Sym_4 . Given any two colors, it is possible to transpose them (leaving the others fixed) by a rotation. Because Sym_4 is generated by transpositions, the induced homormorphism $\mathrm{Rot}_{\square} \to \mathrm{Sym}_4$ must be surjective. However, $|\mathrm{Rot}_{\square}| = 24 = 4! = |\,\mathrm{Sym}_4\,|$. Hence it must be an isomorphism (because of bijective).

Lemma 3.142. *We deduce that* $\mathrm{Rot}_{\square}$ *is isomorphic to* Sym_4.

3.7.4 One 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

$$G \cong \mathrm{Sym}(X)?$$



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

Remark 3.143. *I think the answer is no with the group being $\Sigma(\mathbb{R}^n)$ in Remark 3.138.*

3.8 Normal Subgroups and Isomorphism Theorems

1. In linear algebra, we study the linear maps $\boldsymbol{A}$, and not the vector spaces themselves. The structure preserving maps between vector spaces are more interesting than the spaces themselves.
2. Philosophically it's saying that an object in isolation is uninteresting; it's how it relates to what's around it that matters.

In this section, we study **homomorphisms between abstract groups**.

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 $\forall\, x, y \in G$, $f(xy) = f(x)f(y)$

3.8.1 Kernel and Image: subgroups

Definition 3.144. ***(Kernel and Image)*** *Given $f : G \to H$ a homomorphism of groups, we define the **kernel** of f to be:*

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

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

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

Lemma 3.145. *Given a homomorphism $f : G \to H, \operatorname{Ker}(f) \subseteq G$ and $\operatorname{Im}(f) \subseteq H$ are subgroups.*

Proof. First we will show true for $\operatorname{Ker}(f)$:

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

Now we will show that $\operatorname{Im}(f)$ is a subgroup:

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

□

Lemma 3.146. *A homomorphism $f : G \to H$ is injective if and only if $\operatorname{Ker}(f) = \{e_G\}$.*

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

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

Thus f is injective. □

3.8.2 Normal Group

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 $(xy)H$. In general this is **not** well defined. It will depend on the choice of x and y. In the case $G = \mathbb{Z}$ and $m\mathbb{Z} = H$, we know for any x and y, we can define the binary operator and makes $G/m\mathbb{Z}$ a well defined new group.

What's the answer for the general case? We need H to be a **normal** group!

Definition 3.147. ***(Normal Group)*** *We call a subgroup $H \subseteq G$* ***normal*** *if it satisfies equivalently any one of the following condition*

- $gHg^{-1} := \left\{ghg^{-1} \mid g \in G, h \in H\right\} = H$*;*
- $\forall\, g \in G$*, and* $\forall\, h \in H$*,* $ghg^{-1} \in H$*;*
- *left coset equals right coset for* $\forall\, g \in G$*:* $gH = Hg$*;*
- *the union of conjugacy classes of G equals* H*, i.e., if* $H = \bigcup_{h \in H} \operatorname{Conj}(h)$*;*
- *Any two elements commute regarding the normal subgroup membership relation: for all* $g, h \in G, gh \in H$ *if and only if* $hg \in H$

We denote ***normal subgroup*** *by* $H \triangleleft G$*.*

Proof. We only show the equivalence between last one and others.

Assume the last one is true. For any $h \in H$, we have $hg^{-1}g = h \in H$ for any $g \in G$, so $ghg^{-1} \in H$.

Assume $\forall\, g \in G$, and $\forall\, h \in H$, $ghg^{-1} \in H$. If $ab \in H$ for a pair of elements $a, b \in G$. We have $b^{-1}a^{-1} \in H$, so $a^{-1}(b^{-1}a^{-1})a \in H$. That means $a^{-1}b^{-1} \in H$. So $ba \in H$. □

Remark 3.148. ***(explanation)***

1. (a) *The definition of normal group is not saying that given* $g \in G$ *and* $h \in H$*, then* $\text{ghg}^{-1} = h$*. It is merely saying that* $ghg^{-1} \in H$*.*
   (b) *The definition of group center means: if* $h \in Z(G)$*,* $\text{ghg}^{-1} = h$ *for all* $g \in G$*.*
2. *Note here* $H = \{ghg^{-1} \mid g = e_G, h \in H\}$*. So for all subgroup* H*, we have* $H \subseteq gHg^{-1}$*.*
3. *normal subgroup vs. conjugacy vs. group center:*
   - *Recall:* $\operatorname{Conj}(h) := \operatorname{Orb}(h) = \left\{g * h * g^{-1} \mid g \in G\right\}$*;* $Z(G) := \{h \in G \mid g * h = h * g, \forall\, g \in G\}$

   (a) *Observe that* $h \in Z(G) \iff \operatorname{Conj}(h) = \{h\}$*.*
   (b) *A subgroup is normal* ***if*** *the union of conjugacy classes of G equals* H*, i.e., if* $H = \bigcup_{h \in H} \operatorname{Conj}(h)$*.*
   (c) ***Group center is a normal subgroup****. And group center is more restrict than normal subgroup as we have mentioned in 1.*
4. ***If*** G ***is Abelian, every subgroup is normal as*** $ghg^{-1} = h, \forall\, g, h \in G$***.***

Example 3.149. ***(non-normal group)*** *Let $G = \mathrm{Sym}_3, H = \{e, (1,2)\}$. 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.*

Theorem 3.150. ***(Kernal of Homomorphism is Normal)***

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

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

Remark 3.151. *In general $\mathrm{Im}(f) \subset H$ is **not** normal.*

Definition 3.152. ***(Simple)*** *We say a group G is **simple** if its only normal subgroups are $\{e\}$ and G*

Example 3.153. *Groups of prime order are trivially simple by Lagrange's theorem. As we shall see later simple groups are the **core building blocks of groups theory.***

Remark 3.154. *Also recall that a finite group G with order p^n, for p a prime number and $n \in \mathbb{N}$. Then the center is non-tivial: $Z(G) \neq \{e\}$.*

1. *Additionally, if G is not Abelian, the group center, which is normal, is not G (and nontrivally). So G is then not simple.*
2. *Additionally, if G is Abelian, the center is G. Later we will see in this case G is possibly not simple as shown in Corollary 3.229.*

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

Lemma 3.155. *Let $H \subseteq G$ be a normal subgroup. Then the binary operation:*

$$G/H \times G/H \to G/H$$
$$(xH, yH) \mapsto (xy)H$$

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: $\forall\, x_1, x_2, y_1, y_2 \in G$ such that $x_1H = x_2H$ and $y_1H = y_2H$, 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 $uy_2^{-1}y_1 = y_1^{-1}\left(x_1^{-1}x_2\right)y_1$. 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.

Corollary 3.156. *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. $\forall\, x, y, z \in G, (xy)z = x(yz) \Rightarrow (xH * yH) * zH = xH * (yH * zH)$.
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.

□

Corollary 3.157. *If G is Abelian, any subgroup of H is of course Abelian, and G/H is therefore an **Abelian group.***

Remark 3.158. *The canonical example is $\mathbb{Z}/m\mathbb{Z}$.*

Theorem 3.159. *Given a* ***normal*** *subgroup H, the natural map*

$$\begin{aligned} \phi : G &\to G/H \\ x &\mapsto xH \end{aligned}$$

is a ***homomorphism*** *with* $\operatorname{Ker}(\phi) = H$.

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 $x \in \operatorname{Ker}(\phi) \iff \phi(x) = xH = H \iff x \in H$. Hence $\operatorname{Ker}(\phi) = H$. □

Observe that Theorem 3.150 and Theorem 3.159 show that

any normal subgroup can be realised as the kernel of a group homomorphism.

3.8.3 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 $\operatorname{Ker}(\phi) \subset G$ is a normal subgroup. Hence we may form the quotient group $G/\operatorname{Ker}(\phi)$.

Lemma 3.160. *The followng map:*

$$\begin{aligned} \varphi : G/\operatorname{Ker}(\phi) &\to \operatorname{Im}(\phi) \\ x\operatorname{Ker}(\phi) &\mapsto \phi(x) \end{aligned}$$

is well defined.

Proof. $x\operatorname{Ker}(\phi) = y\operatorname{Ker}(\phi) \iff x^{-1}y \in \operatorname{Ker}(\phi) \iff \phi\left(x^{-1}y\right) = e_H \iff \phi\left(x^{-1}\right)\phi(y) = e_H \iff \phi(x)^{-1}\phi(y) = e_H \iff \phi(x) = \phi(y)$. In summary, $\phi(x) = \phi(y) \iff x\operatorname{Ker}(\phi) = y\operatorname{Ker}(\phi)$. Hence ϕ is constant on each coset of $\operatorname{Ker}(\phi)$. Note here it actually prove the injective of φ. □

Theorem 3.161. ***(The First Isomorphism Theorem)*** *Let G and H be two groups. Let $\phi : G \to H$ be a homomorphism, then the induced map φ is an* ***isomorphism of groups****.*

Remark 3.162. *Since isomorphism implies injective, we know* $\operatorname{Ker}(\varphi) = \{\operatorname{Ker}(\phi)\}$ *from Lemma 3.146.*

Proof. Firstly we observe that the induced ϕ is by definition of $\operatorname{Im}(\phi)$ surjective. Note that from the proof in Section 3.8.3, we already have φ is injective: given $x, y \in G, \varphi(x\operatorname{Ker}(\phi)) = \varphi(y\operatorname{Ker}(\phi)) \iff \phi(x) = \phi(y) \iff x\operatorname{Ker}(\phi) = y\operatorname{Ker}(\phi)$.

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

3.8.4 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.

Remark 3.163. *A general normal subgroup. Not limit to the* Ker.

- H**: subgroup of** G **containing** $N \Rightarrow H/N$ **: subgroup 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 $H/N = \{hN \mid h \in H\}$. Observe that H/N is naturally a subset of G/N.

$$N \subset H \subset G$$

Lemma 3.164. $H/N \subset G/N$ *is a subgroup.*

Remark 3.165. *We know both H/N and G/N are groups because N is normal, then H/N is a subgroup. Below proof is not that necessary.*

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 $(xN)^{-1} = x^{-1}N \in H/N$.

□

• **M: subgroup of $G/N \Rightarrow H_M$: subgroup of G containing 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.

$$H_M = \bigcup_{xN \in M} xN$$

Lemma 3.166. *$N \subset H_M \subset G$, where H_M 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 with image and domain being collections of sets:

$$\alpha : \{\text{subgroups of } G \text{ containing N}\} \longrightarrow \{\text{subgroups of } G/N\}$$
$$H \mapsto H/N$$

and

$$\beta : \{\text{subgroups of } G/N\} \longrightarrow \{\text{subgroups of } G \text{ containing N}\}$$
$$M \mapsto H_M$$

Lemma 3.167. *These maps of sets are inverse to each other.*

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

- Let H be a subgroup of G containing N. Then $\beta\alpha(H) = \beta(H/N) = \cup_{x \in H} xN = H$. Thus $\beta\alpha$ is the identity map on $\{$subgroups of G containing $N\}$.
- Let M be a subgroup of G/N. then $\alpha\beta(M) = \alpha(H_M) = M$, where the last equality from $H_M = \bigcup_{xN \in M} xN$. 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:

Theorem 3.168. ***(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.*

3.8.5 The Second Isomorphism Theorem

Theorem 3.169. ***(The Second Isomorphism Theorem)*** *Let G be a group. Let S be a subgroup of G, and let N be a normal subgroup of G. Then the following hold:*

1. *The product SN is a subgroup of G,*
2. *The intersection $S \cap N$ is a normal subgroup of S, and*
3. *The quotient groups $(SN)/N$ and $S/(S \cap N)$ are isomorphic with the isomorphism*

$$\begin{gathered} f : S/(S \cap N) \to (SN)/N \\ s(S \cap N) \mapsto sN \end{gathered}$$

Proof. Note $\mathrm{Ker}(f) = \{s \in S \mid sN = N\} = \{s \in S \mid s \in N\} = S \cap N$. Surjective and injective is then obvious. Homomorphism is also obvious. The conclusion follows from the first isomorphism theorem Section 3.8.3. □

3.8.6 Factor Theorem

Assume that we have

1. a group G which contains a normal subgroup N;
2. a group H.
3. $f : G \to H$ a group homomorphism, and $N \subseteq \mathrm{Ker}(f)$.

Let π be the canonical projection as in Theorem 3.159 from G to the quotient group G/N :

$$\begin{tikzcd} G \arrow[r, "f"] \arrow[d, "\pi"'] & H \\ G/N \arrow[ur, "\bar{f}"'] & \end{tikzcd}$$

We would like to find a homomorphism $\bar{f} : G/N \to H$ that makes the diagram **commute**, namely

$$f(a) = \bar{f}(\pi(a))$$

for all $a \in G$

Theorem 3.170. ***(Factor Theorem)*** *Any homomorphism f whose kernel K contains N can be factored through G/N. In other words, there is a* ***unique homomorphism*** *$\bar{f} : G/N \to H$ such that $\bar{f} \circ \pi = f$. Furthermore*

1. *$\bar{f}$ is an* ***epimorphism*** *if and only if f is.*
2. *$\bar{f}$ is a* ***monomorphism*** *if and only if $K = N$.*
3. *$\bar{f}$ is an* ***isomorphism*** *if and only if f is an epimorphism and $K = N$.*

Proof. **Unicity.** Let us start by proving that if there exists $\bar{f}$ such that $\bar{f} \circ \pi = f$, then it is unique. Let $\tilde{f}$ be another homomorphism such that $\tilde{f} \circ \pi = f$. We thus have that

$$(\bar{f} \circ \pi)(a) = (\tilde{f} \circ \pi)(a) = f(a)$$

for all $a \in G$, that is

$$\bar{f}(aN) = \tilde{f}(aN) = f(a).$$

This tells us that for all $bN \in G/N$ for which there exists an element b in G such that $\pi(b) = bN$, then its image by either $\bar{f}$ or $\tilde{f}$ is determined by $f(b)$. This shows that $\bar{f} = \tilde{f}$ by **surjectivity** of π.

Existence. Let $aN \in G/N$ such that $\pi(a) = aN$ for $a \in G$. We define

$$\bar{f}(aN) = f(a).$$

This is the most natural way to do it, however, we need to make sure that this is indeed well-defined, in the sense that it should not depend on the choice of the representative taken in the coset. Let us thus take another representative, say $b \in aN$. Since a and b are in the same coset, they satisfy $a^{-1}b \in N \subset K$, where $K = \mathrm{Ker}(f)$ by assumption. Since $a^{-1}b \in K$, we have $f\left(a^{-1}b\right) = 1$ and thus $f(a) = f(b)$.

Now that $\bar{f}$ is well defined, let us check this is indeed a group homomorphism. First note that G/N is indeed a group since $N \lhd G$. Then, we have

$$\bar{f}(aNbN) = \bar{f}(abN) = f(ab) = f(a)f(b) = \bar{f}(aN)\bar{f}(bN)$$

and $\bar{f}$ is a homomorphism.

1. The fact that $\bar{f}$ is an epimorphism if and only if f is comes from the fact that both maps have the same image.
2. First note that the statement $\bar{f}$ is a monomorphism if and only if $K = N$ makes sense since $K = \mathrm{Ker}(f)$ is indeed a normal subgroup, as proved earlier.

   To show that $\bar{f}$ is a monomorphism is equivalent to show that $\mathrm{Ker}(\bar{f})$ is trivial. By definition, we have

   $$\begin{aligned}\mathrm{Ker}(\bar{f}) &= \{aN \in G/N, \bar{f}(aN) = 1\} \\ &= \{aN \in G/N, \bar{f}(\pi(a)) = f(a) = 1\} \\ &= \{aN \in G/N, a \in K = \mathrm{Ker}(f)\}\end{aligned}$$

   So the kernel of $\bar{f}$ is exactly those cosets of the form aN with $a \in K$, but for the kernel to be trivial, we need it to be equal to N, that is we need $K = N$.
3. This is just a combination of the first two parts.

□

3.9 Direct Products and Direct Sums

3.9.1 (External) Direct Products

Definition 3.171. ***((External) Direct Product)*** *Let G and H be two groups, with respective identities e_G and e_H. We may form the* ***(external) direct product*** *$G \times H = \{(x, g) \mid x \in G, g \in H\}$. 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 3.172. *$G \times H$ is a group under the natural binary operation.*

Proof. Check the 3 conditons:

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 $h \in H(g, h)^{-1} = \left(g^{-1}, h^{-1}\right)$.

□

Remark 3.173. *There is an obvious generalization of this concept to* ***the product of any finite collection of groups****.*

Theorem 3.174. *Let G and H be finite cyclic groups. Then $G \times H$ is cyclic if and only if $|G|$ and $|H|$ are relatively prime.*

Proof. Let $|G| = m$ and $|H| = n$, so that $|G \times H| = mn$. To prove the first half of the theorem, we assume $G \times H$ is cyclic and show that m and n are relatively prime. Suppose that $\gcd(m, n) = d$ and (g, h) is a generator of $G \times H$. Since $(g, h)^{mn/d} = \left((g^m)^{n/d}, (h^n)^{m/d}\right) = (e, e)$, we have $mn = |(g, h)| \leq mn/d$. Thus, $d = 1$.

To prove the other half of the theorem, let $G = \langle g \rangle$ and $H = \langle h \rangle$ and suppose $\gcd(m, n) = 1$. Then, $|(g, h)| = \operatorname{lcm}(m, n) = mn = |G \times H|$, so that (g, h) is a generator of $G \times H$. □

Corollary 3.175. ***(Criterion for*** $G_1 \times G_2 \times \cdots \times G_n$ ***to Be Cyclic)*** *An external direct product* $G_1 \times G_2 \times \cdots \times G_n$ *of a finite number of finite cyclic groups is cyclic if and only if* $|G_i|$ *and* $|G_j|$ *are relatively prime when* $i \neq j$.

Corollary 3.176. ***(Criterion for*** $Z_{n_1 n_2 \cdots n_k} \cong Z_{n_1} \times Z_{n_2} \times \cdots \times Z_{n_k}$***)*** *Let* $m = n_1 n_2 \cdots n_k$. *Then* Z_m *is isomorphic to* $Z_{n_1} \times Z_{n_2} \times \cdots \times Z_{n_k}$ *if and only if* n_i *and* n_j *are relatively prime when* $i \neq j$.

Proof. This is because cyclic of $Z_{n_2} \times \cdots \times Z_{n_k}$ and Theorem 3.60. □

Example 3.177. *By using the results above in an iterative fashion, one can express the same group (up to isomorphism) in many different forms. For example, we have*

$$Z_2 \times Z_2 \times Z_3 \times Z_5 \cong Z_2 \times Z_6 \times Z_5 \cong Z_2 \times Z_{30}$$

3.9.2 Direct Sum (Internal Direct Product)

Definition 3.178. ***(Direct Sum)*** *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* $G = H \oplus K$.

Remark 3.179. ***(equivalent condition and explanation)*** *We explain these two conditions:*

1. *The first condition is similar to Abelian, however we only require this* ***commutative*** *over* H *and* K. *Note in some books [3][Page 183], they use definition with* H *and* K *to be normal groups. This is* ***more strict*** *as this definition implies our definition [3][Theorem 9.6].*
2. *The second property guarantee uniqueness of the decomposition (see [2, page 2]) and it is equivalent to:*
   - $H \cap K = \{e_G\}$ *and for* $g \in G$ *there exist* $h \in H, k \in K$ *such that* $g = hk$.

   From the second property, we can define (3.184). *It may be confused why for every* h *and* k, *we can define the map* $hk \in G$. *This is because* H *and* K *are all subgroups of* G *so the operation is closed. So the second property should be read as:*
   - (3.184) *is a bijection.*
3. *The generalization of this concept to the direct sum of any finite collection of groups can be seen at [3][Page 184].*

Remark 3.180. ***(thinking)*** *What's the fundamental reason of "*f *injective if and only if* $\operatorname{Ker}(f) = \{e_G\}$ *in Lemma 3.146 " and* $H \cap K = \{e_G\}$ *guarantee uniqueness"?*

- ***The general answer is (linear) tensor over groups****. Lemma 3.146 is the special case of* 1*-tensor, while the above direct sum is* 2*-tensor over the product* $H \times K$. *If the kernel of the homomorphism* (3.184) *is* $\{(e_G, e_G)\}$. *We know it is injective.* $\{(e_G, e_G)\}$ *just means* $H \cap K = \{e_G\}$ *(hint:* $H \cap K$ *is a group, if* $H \cap K \neq \{e_G\}$, *we will get that the kernel is not* $\{(e_G, e_G)\}$*).*

Example 3.181. *For example,* $(\mathbb{Z}/15\mathbb{Z}, +)$ *is the direct sum of* $\operatorname{gp}([3])$ *and* $\operatorname{gp}([5])$

Theorem 3.182. ***(direct sum isomorphic to direct product)*** *If G is the direct sum of the subgroups* $H, K \subset G$ *then:*

$$H \oplus K \cong H \times K$$

Remark 3.183. *The concept of direct sum has a clear generalization to any finite collection of subsets of G.*

Proof. Define the map

$$\begin{aligned} \phi : H \times K &\to G \\ (h, k) &\mapsto hk \end{aligned} \tag{3.184}$$

Let $x, y \in H$ and $g, h \in K$. By property 1, we have $\phi((x, g) * (y, h)) = \phi(xy, gh) = xygh = xgyh = \phi(x, g)\phi(y, h)$. Hence ϕ is a homomorphism. Property two ensures that ϕ is bijective. □

Corollary 3.185. *If normal groups $H \cap K = \{e\}$ and are subgroups of G, we have $H \oplus K \cong H \times K$ which may not surjective over G. But note according to* (3.184)*, $H \oplus K$ must be a group as the image of the map* (3.184)*.*

3.9.3 Relation to Chinese Reminder Theorem, Unit Group

• Notation Clarify:

1. For notation simplicity, in this section, we denote $\mathbb{Z}/n\mathbb{Z}$ as $\mathbb{Z}_n$.
2. Let **unit group** $\mathbb{U}_n \subseteq \mathbb{Z}_n$ consist of those $[u]$ such that for some $[v]$, $[u] \cdot [v] = [1]$, namely, those elements of $\mathbb{Z}_n$ that have multiplicative inverses. Note the group operation now is **multiplication** of modulo.

We first recall Chinese Remainder Theorem and its corollary:

Theorem 3.186. ***(Chinese Remainder Theorem)*** *Suppose $n = ab$, with a and b relatively prime. For $x = 0, 1, \dots, n-1$, associate $[x] \in \mathbb{Z}_n$ with $([x], [x]) \in \mathbb{Z}_a \times \mathbb{Z}_b$ (note that the symbol $[x]$ means different things in $\mathbb{Z}_n, \mathbb{Z}_a$ and $\mathbb{Z}_b$). This gives a* ***one-to-one*** *correspondence between $\mathbb{Z}_n$ and $\mathbb{Z}_a \times \mathbb{Z}_b$.*

Corollary 3.187. *Suppose $n = ab$, with a and b relatively prime. For $x = 0, 1, \dots, n-1$, if $[x] \in \mathbb{U}_n$, associate $[x]$ with $([x], [x]) \in \mathbb{Z}_a \times \mathbb{Z}_b$. This gives a* ***one-to-one*** *correspondence between $\mathbb{U}_n$ and $\mathbb{U}_a \times \mathbb{U}_b$.*

• **But note $\mathbb{Z}_a$ and $\mathbb{Z}_b$ do not satisfy the conditions of intersection is $\{[0]\}$. What is happening?**

1. Corollary 3.176 is the group version (isomorphic, more than bijiection) of Chinese Remainder Theorem Theorem 3.186. Note in Corollary 3.176, cyclic is the key. We does not care the direct sum here.
2. $\text{gp}\,[a] \cap \text{gp}\,[b] = \{[0]\}$ if a and b are coprime is clear since otherwise it mean $\text{LCM}(a, b) < ab$.
3. But what's the relation between $\mathbb{Z}_{ab} \cong \mathbb{Z}_a \times \mathbb{Z}_b \cong \text{gp}\,[a] \times \text{gp}\,[b] \cong \text{gp}\,[a] \oplus \text{gp}\,[b] = \mathbb{Z}_{ab}$, where $[a]$ is in terms of $\text{mod} ab$?

   Key: this is the trick of multiple each element in $\mathbb{Z}_a$ by b, and you will get $\text{gp}\,[b]$; or we note $\text{gp}\,[b]$ is a cyclic group with order b, so it must isomorphic to $\mathbb{Z}_b$.
4. Similary, we can also get $\mathbb{U}_{ab} \cong \mathbb{U}_a \times \mathbb{U}_b \cong \mathbb{U}_{ab}^{(a)} \times \mathbb{U}_{ab}^{(b)} \cong \mathbb{U}_{ab}^{(a)} \oplus \mathbb{U}_{ab}^{(b)}$;

We next list the facts that have been stated in number theory notes or in [3]:

1.

   Definition 3.188. *Suppose that $n = de$. Let $G_e = \{x : 0 \le x < n \text{ and } (x, n) = e\}$, that is, G_e consists of all numbers whose gcd with n is e.*
2.

   Definition 3.189.

   $$R_d = \{y : 0 \le y < d \text{ and } (y, d) = 1\}$$

   Notice that this definition makes sense even for $d = 1$, namely, $R_1 = \{0\}$, and in every case, the number of elements in R_d is $\phi(d)$ (because R_d is almost $\mathbb{U}_d$, missing the $[\cdot]$ again).
3.

Definition 3.190. *If k is a divisor of n, let*

$$\mathbb{U}_n^{(k)} = \{x \in \mathbb{U}_n \mid x \bmod k = 1\}$$

This is a group. I skim the proof here.

Lemma 3.191. *From number theory notes, we have $|G_e| = |R_d|$ according to*

$$(y,d) = 1 \quad \textit{iff} \quad (ey, ed) = e \quad \textit{iff} \quad (ey, n) = e \tag{3.192}$$

Remark 3.193. *Note here is a general conclusion, no need d and e are coprime. But below we need* ***coprime*** *condition.*

Lemma 3.194. *Suppose a and b are relatively prime. Then $\mathbb{U}_{ab}$ is isomorphic to the external direct product of $\mathbb{U}_a$ and $\mathbb{U}_b$. We have*

1. $\mathbb{U}_{ab} \cong \mathbb{U}_a \times \mathbb{U}_b$;
2. $\mathbb{U}_{ab}^{(a)} \cong \mathbb{U}_b$;
3. $\mathbb{U}_{ab}^{(b)} \cong \mathbb{U}_a$.

Proof. We can construct the following.

1. An isomorphism from $\mathbb{U}_{ab}$ to $\mathbb{U}_a \times \mathbb{U}_b$ is $x \to (x \bmod a, x \bmod b)$;
2. An isomorphism from $\mathbb{U}_{ab}^{(a)}$ to $\mathbb{U}_b$ is $x \to x \bmod b$;
3. An isomorphism from $\mathbb{U}_{ab}^{(b)}$ to $\mathbb{U}_a$ is $x \to x \bmod a$.

□

Remark 3.195. *The first is obvious the isomorphic group version of Corollary 3.187. But what's the second and third means? It becomes obvious if we note the similar trick of* (3.192)*: it is just a factor trick.*

Lemma 3.196. *If a and b are relatively prime positive integers then $\mathbb{U}_{ab} = \mathbb{U}_{ab}^{(a)} \oplus \mathbb{U}_{ab}^{(b)} \cong \mathbb{U}_{ab}^{(a)} \times \mathbb{U}_{ab}^{(b)}$.*

Proof. Here, we only need to show $\mathbb{U}_{ab}^{(a)} \cap \mathbb{U}_{ab}^{(b)} = \{[1]\}$. If otherwise $[1] \neq [w] \in \mathbb{U}_{ab}^{(a)} \cap \mathbb{U}_{ab}^{(b)}$. We have $w \equiv 1 \bmod a$ and $w \equiv 1 \bmod b$. But this just means a and b are not coprime (think carefully here), a contradiction. We therefore have the conclusion $\mathbb{U}_{ab}^{(a)} \times \mathbb{U}_{ab}^{(b)} \cong \mathbb{U}_{ab}^{(a)} \oplus \mathbb{U}_{ab}^{(b)}$ according to (3.184) and $\mathbb{U}_{ab} \cong \mathbb{U}_{ab}^{(a)} \times \mathbb{U}_{ab}^{(a)}$ according to Lemma 3.194. From the finite size, we then have the map (3.184) is surjective on $\mathbb{U}_{ab}$ and hence get "=". □

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

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

Lemma 3.197. *Use Abelian properties, we have that $\forall\, a, b \in G; m, n \in \mathbb{Z}$:*

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

Remark 3.198. *Note here ma is not the multiplication of two elements in G, it is one integer and one element in G. Compare also with Lemma 4.11.*

From now in this section we assume that

G is finitely generated Abelian group

Definition 3.199. ***(Basis)*** *$\exists$ $\{a_1, \cdots, a_n\} \subset G$ such that $\text{gp}(\{a_1, \cdots, a_n\}) = G$. In other words, because G is Abelian, every $x \in G$ can be written in the form*

$$x = \lambda_1 a_1 + \cdots + \lambda_n a_n \quad \lambda_i \in \mathbb{Z} \tag{3.200}$$

*If every expression of the above form was **unique (after possibly restricting** $0 \leq \lambda_1 < \text{ord}(a_i)$) for a given $x \in G$. We call $\{a_1, \cdots, a_n\}$ a **basis** for G.*

Remark 3.201. *In general such the expression in* (3.200) *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 $ma = 0$. Observe that it is not clear that such a basis even exists at present.*

Lemma 3.202. ***(Direct Sum Decomposition)*** *If $\{a_1, \cdots, a_n\} \subset G$ were a basis then letting $A_i = \text{gp}(a_i) \subset G$ 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.*

Proof. Since G is Abelian, from Definition 3.178, we know that the two direction are both correct in the lemma. □

3.10.1 Torsion, Torsion Free, Free Abelian

Definition 3.203. ***(Torsion, Torsion Subgroup: finite order)*** *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 3.204. *$tG \subset G$ is indeed a subgroup.*

Proof. This critically requires that G be Abelian from properties list in Lemma 3.197. It is not true in general.

1. $\text{ord}(0) = 1 \Rightarrow 0 \in tG$
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 $g \in tG \Rightarrow -g \in tG$

□

Definition 3.205. ***(Torsion Group vs. Torsion Free Group)*** *If $tG = G$ we say that G is a **torsion group**. If $tG = \{0\}$ we say that G is **torsion free**.*

Lemma 3.206. *If G is torsion and finitely generated then G is finite.*

Remark 3.207. *Under condition finitely generated G, we have*

*G **is finite** $\Leftrightarrow tG = G$.*

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

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

This is a finite set. □

Lemma 3.208. *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 $n(x + tG) = nx + tG = tG$ (last equality is from $x + tG$ is torsion). Hence $nx \in tG$ so $\exists\, m \in \mathbb{N}$ such that $mnx = 0$. Hence $x \in tG \Rightarrow x + tG = tG$. □

Remark 3.209. *Here we use notation $x + tG$. This is the same as xtG but we include $+$ for better understanding.*

Definition 3.210. ***(Free Abelian)*** *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.*

Remark 3.211. ***(explanation)***

- *Non-torsion is from $\lambda_i \in \mathbb{Z}$. Compare also with* (3.200) *where the uniqueness may be there in the sense of* ***after possibly restricting*** $0 \le \lambda_1 < \operatorname{ord}(a_i)$.
  - $\operatorname{gp}(\{a_i\})$ *is infinite for each i.*
- *We still have*

$$\text{“}\sum \lambda_i a_i = 0 \Rightarrow \lambda_i = 0 \textit{ for all } i\text{”} \Leftrightarrow \text{“unique”} \Leftrightarrow \operatorname{gp}(\{a_i\}) \cap \operatorname{gp}(\{a_j\}) = \{0\} \textit{ for all } i \neq j. \tag{3.212}$$

In this case

$$G = \operatorname{gp}(a_1) \oplus \cdots \oplus \operatorname{gp}(a_n) \cong \mathbb{Z} \times \mathbb{Z} \cdots \times \mathbb{Z} = \mathbb{Z}^n \tag{3.213}$$

Lemma 3.214. *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 $2G := \{2x \mid x \in G\}$. $2G \subseteq G$ is a subgroup. Observe that $2G = \{\lambda_1 a_1 + \cdots \lambda_n a_n \mid \lambda_i \in 2\mathbb{Z}\}$. Hence $(G : 2G) = 2^n$. But the left hand side is defined independently of the basis, so 2^n (and hence n) is independently of the basis. The result follows. □

Definition 3.215. ***(Rank)*** *Let G be a finitely generated free Abelian group. The rank of G is the size of any basis.*

Theorem 3.216. *A finitely generated abelian group is free Abelian $\Longleftrightarrow$ it is torsion free.*

Proof. ($\Rightarrow$) is trivial. Assume not torsion free but free Abelian, we have a basis $\{a_1, ..., a_n\}$. $\exists\, x = \sum_i \lambda_i a_i \neq 0$ and $mx = x, m > 1$, then $\sum(\lambda_i m - \lambda_i)a_i = 0$ which is impossible from (3.212), a contradiction.

($\Leftarrow$) 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: $n = 1, G = \operatorname{gp}(a) \cong (\mathbb{Z}, +)$ from Theorem 3.60, 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 since now basis guarantees unique (after possbile restricting to 0 to $\operatorname{ord}(a_i)$), the torch free then further guarantees $\operatorname{ord}(a_i)$ is not finite, and we get (3.213). 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 $d \mid \lambda_i$ for all i, then have $d\left(\frac{\lambda_1 a_1}{d} + \frac{\lambda_2 a_2}{d} + \cdots + \ldots\right) = 0$. As G is torsion-free, $\left(\frac{\lambda_1 a_1}{d} + \frac{\lambda_2 a_2}{d} + \cdots + \ldots\right) = 0$. We can therefore **assume that the λ_i are collectively coprime.**

1). If exist a coefficient $= 1$, wlog, $\lambda_1 = 1$, 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 from the $n - 1$ case.

2). Otherwise, we will reduce to the above case as use the following additional steps: Assume all

$\lambda_i \geq 2$ (cannot $\neq 1$; if $\lambda_i = 0$, just remove a_i). 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 (because of coprime) 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. □

Theorem 3.217. *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 3.218. ***(Rank: finitely generated Abelian group)*** *Let G be a finitely generated Abelian group. We define the rank of G to be the rank of G/tG*

3.10.2 Decomposition of Finitely Generated Abelian Group

Let G be finitely generated and Abelian with G/tG being the free Abelian group with rank n. We can select $\{f_1, \cdots, f_n\}$ as a basis for G/tG, where f_i of course has infinite order.

Let $\phi : G \to G/tG$ be the natural quotient homomorphism (Theorem 3.159) which is surjective of course:

$$\begin{aligned} \phi : G &\to G/tG \\ x &\mapsto x + tG \end{aligned}$$

Choose $\{e_1, \cdots, e_n\} \subset G$ such that $\phi(e_i) = f_i, \forall\, i \in \{1, \cdots, n\}$. Since none of the f_i have finite order, we get that none of the e_i have finite order. Moreover, from homomorphism we have that

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

Then, we have $\lambda_1 e_1 + \cdots + \lambda_n e_n = 0 \iff \lambda_i = 0, \forall\, i$. We define

$$F := \text{gp}\{e_1, \cdots, e_n\} \subseteq G$$

which is is free abelian with basis $\{e_1, \cdots, e_n\}$, so 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:

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

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

Theorem 3.219. ***(Decomposition Theorem)*** *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

$$G = F \oplus tG$$

F is free Abelian by construction. We next prove tG is finite. Define the homomorphism :

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

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. □

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

$$G \cong \mathbb{Z}^n \oplus tG \tag{3.220}$$

where tG is finite. See also (3.213).

3.11 Finite Abelian Groups

Definition 3.221. *(*p***-group)** 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.*

Lemma 3.222. *The order of a finite* p*-group must be a power of p.*

Proof. We prove by contradiction. Assume $|G| = p^n u_1 \dots u_n$, where $\mathrm{HCF}(p, u_i) = 1$ for all i. By Sylow's Theorem Theorem 3.103, we have that there exist a subgroup M of G with size u_i. Let $x \in M$ but $x \neq e_G$. We then have $\mathrm{gp}(x)$, a subgroup of M. But note x need to be a power of p according to the definition of p-group. So we have $p|u_i$ which is impossible. □

From now in this section we assume that

G **is finite Abelian group**

Let $p \in \mathbb{N}$ be a prime. We define $G_p := \{g \in G \mid \text{ord }(p) \text{ is a power of } p\} \subset G$

Theorem 3.223. $G_p \subset G$ *is a subgroup.*

Proof.

1. $\mathrm{ord}(0) = 1 = p^0 \Rightarrow 0 \in G_p$
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 $p^r g = 0 \Rightarrow -p^r g = p^r(-g) = 0 \Rightarrow -g \in G_p$.

□

Remark 3.224. *Note the above critically relies on G being Abelian.*

Lemma 3.225. *We have the following facts:*

1. *By definition* G_p *is a* ***maximal*** p***-group** contained in G.*
2. $G_p = \{0\}$ *or p divides* $|G|$*.*
3. *By Sylow's Theorem Theorem 3.103, we deduce that if* $|G| = p^n u$*, where* $\mathrm{HCF}(p, u) = 1$*, then* $|G_p| = p^n$*. (Note, from Lemma 3.222, we already know it has order* p^r *with* $r \leq n$*. Now, we further know* $r = n$*. Again, confirm it is the* ***maximal*** p*-subgroup)*

Proof. 1). By definition.

2). Two ways: one is from Lemma 3.222 directly. The other is from that if $x \in G_p$ but $x \neq 0$, then we have $\mathrm{gp}(x)$ will divides G. Note $\mathrm{ord}(x)$ is a positive power of p. So $p \mid |G|$. Recall that $\forall\, g \in G$, $\mathrm{ord}(g) \mid |G|$ by Lagrange's Theorem.

3). If $|G_p| = p^r$ where $r < n$, we have $|G/G_p| = p^{n-r}u$. By Sylow's Theorem Theorem 3.103, we then can select a subgroup M of G/G_p s.t. $|M| = p$. We then have $G_p \subseteq H_M$ by the Third Isomorphism Theorem Theorem 3.168. Note $|H_M| = p^{r+1}$. We select one $x \in H_M \backslash G_p$, then we have $\mathrm{ord}(x)|p^{r+1}$. But this just means $x \in G_p$, a contradiction. □

The importance of the maximal p-subgroups is the following theorem.

Theorem 3.226. *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 from Corollary 2.17 we have

$$P_1 Q_1 + \cdots + P_r Q_r = 1$$

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 uniqueness of this sum. Assume we had

$$g = g_1' + \cdots + g_r', \quad g_i' \in G_{p_i} \tag{3.227}$$

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. This is because each $g_i' - g_i'$, $i \geq 2$, belongs to G_{p_i} which means that the order of $g_i' - g_i'$ is a power of p_i. We then have $P_1 \times$right hand size$= 0$. Similarly, the left hand side has order dividing P_i, $i \geq 2$. That means all $P_i x = 0$, $i = 1, ..., r$.

Now, from the above discussion, we then have $x = \sum_{i=1}^{n} Q_i P_i x = 0 \Rightarrow g_1 = g_1'$. Similarly we find $g_i = g_i'$ for all $i \in \{1, \cdots, r\}$, hence the sum is unique and we deduce that

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

Let $\{q_1, \cdots, q_s\}$ 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. From the The Fundamental Theorem of Arithmetic Theorem 2.21, clearly $\{p_1, \cdots, p_r\} = \{q_1, \cdots, q_s\}$ and any such representation is unique. □

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

Theorem 3.228. *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 $B = \mathrm{gp}(b)$ of **maximal order**, say p^n.

Strategy: show that there is a p-subgroup $D \subset G$ such that $G = B \oplus D$. Because then we can apply this recursively to D get conclusion.

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 $c \in G \backslash B$ 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 because $p^n c = 0$ (recall **maximal order** is n for any element in G).

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 $p^n \mid sp^{n-r}$. Therefore we have $p \mid s$, say $s = ps'$. Hence $c_1 = p^{r-1}c - s'b$ has order p and is not in B. Therefore $C = \mathrm{gp}\,(c_1)$ is the required subgroup.

Let $BC = \{ab \mid a \in B, b \in C\}$. 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 $(a_1 b_1)(a_2 b_2) = (a_1 a_2)(b_1 b_2) \in BC$. Hence BC is closed under composition.
3. Let $a_1 \in B, b_1 \in C$. Then $(a_1 b_1)^{-1} = b_1^{-1} a_1^{-1} = a_1^{-1} b_1^{-1} \in BC$. Hence BC is closed under taking inverses.

First observe that $|G/C| < |G|$. (note the strictly $<$ here) 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. In other words, we get that

$$G/C = BC/C \oplus D/C$$

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 $g = a + d + c$ 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 $G = B \oplus D$. □

Corollary 3.229. *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 $pG = \{pg \mid g \in G\}$. It is a straightforward to prove that pG is a subgroup of G. Note that for $r > 1, \mathbb{Z}/p^{r-1}\mathbb{Z} \cong p\left(\mathbb{Z}/p^r\mathbb{Z}\right)$, 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. □

Lemma 3.230. *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

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

□

Theorem 3.231. ***(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 can also stated this theorem as direct product.

- **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:

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

Definition 3.232. *let G be a finite group. A composition series for G is a nested collection of subgroups*

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

such that

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

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

Remark 3.233. *By the third isomorphism theorem a composition series cannot be extended, meaning we cannot add any intermediate normal subgroups. It means if $G_1 \lhd G_2$, we cannot insert one G' s.t. $G_1 \lhd G' \lhd G_2$. This is because that if could then G'/G_1 is a normal subgroup in G_2/G_1 which is a contradiction with G_2/G_1 is simple.*

Theorem 3.234. *Any finite group G has a composition series.*

Example 3.235. *Observe that if G is simple that $\{e\} = G_0 \lhd G_1 = G$ is a composition series. If $G = \mathrm{Sym}_3$ then*

$$\{e\} \lhd \mathrm{gp}((123)) \lhd \mathrm{Sym}_3$$

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 simple.

Theorem 3.236. ***(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 3.237. ***(simple components)*** *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.

Example 3.238. *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 3.239. ***(solvable)*** *A finite group is called **solvable** (or soluble) if its simple components are **Abelian**. Note that solvable groups need not be Abelian themselves.*

Example 3.240. *Note that Sym_3 is solvable (please visually think three or two balls at a triangle or a line switching order or rotate, that's also why its simple components is isomorphism to $\mathbb{Z}/3\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z}$), while Alt_5 **(being simple and non-Abelian) is non-solvable**.*

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

  1. The simple groups are like atoms
  2. Finite groups have simple components, like molecules have constituent atoms.
  3. 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

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

Remark 3.241. *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**:

  1. Cyclic groups of prime order. **These are the only Abelian simple groups.**
  2. Alt_n for $n > 4$.
  3. 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.**
  4. 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 1970 s. 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 Ring and Field

4.1 Basic Definitions

Definition 4.1. ***(ring)*** *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).

Remark 4.2. *Please note*

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:*

$$\left(\sum_i x_i\right)\left(\sum_j y_j\right)=\sum_{i, j} x_i y_j$$

in a well-defined way.

Example 4.3. *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.*
3. *Let S be a set and $\mathbb{P}(S)$ be the power set of S. On $\mathbb{P}(S)$ define $+$ and $\times$ by*

$$X+Y=\left(X \cap Y^{\prime}\right) \cup\left(X^{\prime} \cap Y\right), XY=X \cap Y,$$

*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 $\mathbb{M}_{n\times n}\left(\mathbb{R}^n\right)$. This has the structure of a ring under the usual addition and multiplication of matrices **(no multiplication inverse and not multiplication commutative)**.*

Definition 4.4. ***(commutative ring)*** *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 4.5. ***(ring homomorphism)*** *Let R and S be two rings. A **ring homomorphism** ϕ from R to S is a map of sets $\phi: R \rightarrow S$ such that $\forall\, x, y \in R$*

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

Remark 4.6. *Once again, if $R=S$ and $\phi=\mathrm{Id}_R$ then we call it the **identity homomorphism.***

Remark 4.7. ***(explanation)***

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

   $\star$ However, if the ring is entire, we then have $\phi\left(1_R\right)=1_S$ because of cancel law.

Definition 4.8. *Some analogous definition for ring to group in Section 3.1.1:*

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

Theorem 4.9. *In any ring R we have the following elementary consequences of the axioms:*

1. $x0 = x(0+0) = x0 + x0 \Rightarrow x0 = 0$
2. *Similarly,* $0x = 0$ *for all* $x \in R$*.*

Theorem 4.10. ***(trivial ring)*** *The ring with one element is called the* ***trivial ring.***

$$R \textbf{ consists of one element} \Leftrightarrow 1 = 0.$$

Proof. If $1 = 0$ then $\forall\, x \in R, x = x1 = x0 = 0$, hence R consists of one element. The other direction is also trivial. □

• Some notations analogous to group Lemma 3.197:

1. In a ring we abbreviate expressions like
$$a + a + a + \cdots + a(n \text{ times }) = na(n \in \mathbb{N})$$
It is clear that we may naturally extend this to all $n \in \mathbb{Z}$.
2. Similarly,
$$a \times a \times \cdots \times a(n \text{ times }) = a^n \text{ for } n \in \mathbb{N}.$$

Lemma 4.11. *We have the identities:* $\forall\, a, b \in R$ *and* $m, n \in \mathbb{Z}$*,*

1. $m(a + b) = ma + mb$
2. $(m + n)a = ma + na$
3. $(mn)a = m(na)$
4. $m(a \times b) = (ma) \times b = a \times (mb)$
5. $a^{m+n} = a^m a^n$
6. $a^{mn} = (a^m)^n$
7. $(a + b)^n = \sum_{2^n} ...$ *(all possible ordering with n items with mixed a and b. Note $ab \neq ba$)*

If further R is commutative, we have

1. $(a \times b)^n = a^n \times b^n$
2. $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

Remark 4.12. *Compare with Lemma 3.197.*

Definition 4.13. ***(subring)*** *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*). More precisely,*

1. *R is a* ***subgroup*** *of S under addition.*
2. *R is closed under multiplication (i.e. still a monoid).*
3. $1_S \in R$*.*

Lemma 4.14. *Analogous to subgroups,* ***an arbitrary intersection of subrings is again a subring.***

Example 4.15. $(\mathbb{Z}, +, \times) \subset (\mathbb{Q}, +, \times)$ *is a subring.*

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

• Recall:

1. Let G and H be groups and $\phi : G \to H$ a group homomorphism.
2. Recall that $\operatorname{Ker}(\phi) \subset G$ is a normal subgroup, thus the set of right coset $G/\operatorname{Ker}(\phi)$ naturally forms a group (the quotient group).
3. Recall that all normal subgroups arise in this manner (See Theorem 3.159).
4. The First Isomorphism Theorem in Section 3.8.3 states that there is a natural isomorphism

$$G/\operatorname{Ker}(\phi) \cong \operatorname{Im}(\phi).$$

• **Question:** Does something analogous hold for rings? **Ans:** Yes. We next introduce it.

4.2.1 Kernel and Image

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

Definition 4.16. ***(Kernel and Image)*** *The **kernel** of ϕ is the subset*

$$\operatorname{Ker}(\phi) := \{r \in R \mid \phi(r) = 0_S\} \subset R.$$

*The **image** of ϕ is the subset*

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

Remark 4.17. *Remember that ring homomorphism ϕ need to be a **group homomorphism** with respect to the additive Abelian group structures on R and S. From group theory, we know*

1. $\operatorname{Ker}(\phi) = \{0_R\} \iff \phi$ *is injective*
2. $\operatorname{Ker}(\phi) \subset R$ *and* $\operatorname{Im}(\phi) \subset S$ *are subgroups under addition.* $\operatorname{Ker}(\phi)$ *is normal.*

Lemma 4.18. $\operatorname{Im}(\phi) \subset S$ *is a subring.*

Proof. We need to check that $\operatorname{Im}(\phi)$ is closed under multiplication and contains 1_S. Let $s_1, s_2 \in \operatorname{Im}(\phi)$. 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 $s_1 s_2 \in \operatorname{Im}(\phi)$. Hence $\operatorname{Im}(\phi)$ is closed under multiplication. □

4.2.2 Ideal

We already know $\phi(1_R) = 1_S$, $1_S \in \operatorname{Im}(\phi)$ and $\operatorname{Im}(\phi)$ is a **subring.** Furthermore, for $\operatorname{Ker}(\phi)$, we have

1. If S is non trivial then because $\phi(1_R) = 1_S$ we know that $1_R \notin \operatorname{Ker}(\phi)$. Hence in this case $\operatorname{Ker}(\phi) \subset R$ **is not a subring.**
2. $\operatorname{Ker}(\phi) \subset R$ is a Abelian subgroup under $+$.
3. Let $a \in \operatorname{Ker}(\phi)$ and $r \in R$. Observe that $\phi(ra) = \phi(r)\phi(a) = \phi(r)0_S = 0_S$. Hence $ra \in \operatorname{Ker}(\phi)$. Similarly $ar \in \operatorname{Ker}(\phi)$. Hence $\operatorname{Ker}(\phi)$ **is closed under both left and right multiplication by all of R.**

Definition 4.19. ***(ideal)*** *Let R be a ring. An **ideal** $I \subset R$ is a subset which is a subgroup (**not subring**) under addition and is **closed under both left and right multiplication** by **all** of R. More precisely,*

$$\text{if } x \in I \text{ then } xr, rx \in I \text{ for all } r \in R.$$

Remark 4.20. *Analogy to Remark 3.148:*

1. *Note, for any subset* $I \subseteq \{ra, ar \mid r \in R, a \in I\}$.
2. *If I is ideal, we have* $I = \{ra, ar \mid r \in R, a \in I\}$

Lemma 4.21. *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.

Lemma 4.22. *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. Let us define a multiplication for $\forall\, a,b \in R$:*

$$\begin{aligned}R/I \times R/I &\to R/I\\(a+I, b+I) &\mapsto (a+I)\times(b+I) := (ab)+I\end{aligned}$$

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

$$a_1b_1 - a_2b_2 = a_1(b_1 - b_2) + (a_1 - a_2)b_2 \in I$$

because I is an ideal. Thus

$$a_1b_1 + I = a_2b_2 + I$$

Lemma 4.23. *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

1. $0 + I$ is the **additive identity** and
2. $1 + I$ is the **ultiplicative identity.**

□

Analogous to the case of groups shown in Theorem 3.159.

Theorem 4.24. *For ideal I in ring R, the natural map*

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

is ***surjective quotient ring homomorphism***. *We also see that* $\mathrm{Ker}(\phi) = I$.

This is totally analogous to the group theory situation:
any ideal can be realised as the kernel of a ring homomorphism.

4.2.3 The First Isomorphism Theorem

Theorem 4.25. ***(The First Isomorphism Theorem)***

Let $\phi : R \to S$ be a ring homomorphism. Then the induced map

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

is a ***ring isomorphism.***

Proof. The first isomorphism theorem Theorem 3.161 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 $a, b \in R$. $\varphi((a + \mathrm{Ker}(\phi))(b + \mathrm{Ker}(\phi))) = \varphi(ab + \mathrm{Ker}(\phi)) = \phi(ab) = \phi(a)\phi(b) = \varphi(a + \mathrm{Ker}(\phi))\varphi(b + \mathrm{Ker}(\phi))$. Also $\varphi(1 + I) = \phi(1) = 1$. Hence φ is a ring homomorphism and we are done. □

Definition 4.26. ***(Embedding)*** *An injective (i.e.* $\mathrm{Ker}(\varphi) = \{0\}$*) ring homomorphims $\phi : R \to S$ is called an* ***embedding***. *By the first isomorphism theorem, R is* ***isomorphic to the subring*** $\mathrm{Im}(\phi) \subset S$.

4.2.4 Factor Theorem for Rings

We are now ready to state a factor theorem and a 1st isomorphism theorem for rings, the same way we did for groups in Section 3.8.6. It may help to keep in mind the analogy between two-sided ideals and normal subgroups mentioned above.

Assume that we have a ring R which contains a proper two-sided ideal $\mathcal{I}$, another ring S, and $f : R \to S$ a ring homomorphism. Let π be the canonical projection as in Theorem 4.24 from R to the quotient group $R/\mathcal{I}$:

$$\begin{tikzcd}
R \arrow[r, "f"] \arrow[d, "\pi"'] & S \\
R/\mathcal{I} \arrow[ur, "\bar{f}"'] &
\end{tikzcd}$$

We would like to find a ring homomorphism $\bar{f} : R/\mathcal{I} \to S$ that makes the diagram **commute**, namely

$$f(a) = \bar{f}(\pi(a))$$

for all $a \in R$.

Theorem 4.2.A. ***(Factor Theorem for Rings)*** *Any ring homomorphism f whose kernel K contains $\mathcal{I}$ can be factored through $R/\mathcal{I}$. In other words, there is a* ***unique ring homomorphism*** *$\bar{f} : R/\mathcal{I} \to S$ such that $\bar{f} \circ \pi = f$. Furthermore*

1. *$\bar{f}$ is an* ***epimorphism*** *if and only if f is.*
2. *$\bar{f}$ is a* ***monomorphism*** *if and only if $K = \mathcal{I}$.*
3. *$\bar{f}$ is an* ***isomorphism*** *if and only if f is an epimorphism and $K = \mathcal{I}$.*

Proof. Since we have already done the proof for groups in Theorem 3.170 with many details, here we will just mention a few important points in the proof. Let $a + \mathcal{I} \in R/\mathcal{I}$ such that $\pi(a) = a + \mathcal{I}$ for $a \in R$. We define

$$\bar{f}(a + \mathcal{I}) = f(a)$$

This is the most natural way to do it, however, we need to make sure that this is indeed **well-defined**, in the sense that it should not depend on the choice of the representative taken in the coset. Let us thus take another representative, say $b \in a + \mathcal{I}$. Since a and b are in the same coset, they satisfy $a - b \in \mathcal{I} \subset K$, where $K = \mathrm{Ker}(f)$ by assumption. Since $a - b \in K$, we have $f(a - b) = 0$ and thus $f(a) = f(b)$.

Now that $\bar{f}$ is well defined, it is an easy computation to check that $\bar{f}$ inherits the property of ring homomorphism from f. The rest of the proof works exactly the same as for groups as in Theorem 3.170. □

4.3 Integral Domain, Entire, (skew) Field: Properties of Elements of Rings

Definition 4.27. ***(invertible; unit)*** *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 $a'a = aa' = 1$.*

1. *We know that such an inverse is unique if it exists, hence we shall write it as a^{-1}.*
2. ***Note that when*** $1 \neq 0$***,*** 0 ***is never invertible. Again, when*** $1 \neq 0$ ***(i.e. ring is non-trivial), if*** x ***is invertible then*** $x \neq 0$ *(See Theorem 4.10, Theorem 4.9).*
3. *We denote the set of units in R by R^*.*

Remark 4.28. ***(clarify)***

1. *From group definition we need then to the left and right additive inverse to be the same. However, in* ***any*** *ring R, if a has* ***both*** *a left and a right inverse, then* ***the left and right inverses are the same and the element is a unit.***

   If $ca = ab = 1$, we have

   $$b = 1 \cdot b = (ca)b = c(ab) = c \cdot 1 = c$$

   Since $b = c$, then $ca = ba = ab = 1$, so $b = c = a^{-1}$, as desired.

2. *However,* ***a ring element with a left inverse may have but no right inverse****. Example: Take the ring of linear operators on the space of polynomials. Then consider (formal) integration and differentiation. Integration is injective but not surjective. Differentiation is surjective but not injective. See also the proof of Lemma 1.6.*

Lemma 4.29. *For any ring R, $(R^*, \times)$ is a group.*

Definition 4.30. ***(division ring; skew field; field)*** *A* ***non-trivial*** *ring R in which every non-zero element is invertible (i.e $R\backslash\{0\} = R^*$) is called a* ***division ring*** *(or* ***skew field****). If R is a* ***commutative*** *division ring then R is called a* ***field****.*

Example 4.31. $\mathrm{GL}_n(\mathbb{R})$ *is* ***skew field*** *but not commutative.*

Remark 4.32.

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.*

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 $\mathbb{C}$.*

Example 4.33. *The collection of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ is the set $\mathbb{M}_{n\times n}(\mathbb{R}^n)$ is a ring but not a skew field (division ring).*

Definition 4.34. ***(zero-divisor)*** *Let R be a* ***non-trivial*** *ring. Given $a \in R\backslash\{0\}$, if there exists $b \in R\backslash\{0\}$ such that $ab = 0$ or $ba = 0$, then a is said to be a* ***zero-divisor****. Note that* ***0 is not a zero-divisor.***

Example 4.35. *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 4.36. ***(entire (domain); integral domain)*** *A* ***non-trivial*** *ring R with* ***no zero divisors*** *is said to be* ***entire*** *or* ***domain****; a* ***commutative*** *domain is called an* ***integral domain****. More concretely: R is* ***domain*** *if and only if $1 \neq 0$ and $\forall\, x, y \in R, xy = 0 \Rightarrow x = 0$ or $y = 0$.*

Example 4.37. *$(\mathbb{Z}, +, \times), (\mathbb{Q}, +, \times)$ are integral domains. $(\mathbb{Z}/m, +, \times)$ is an integral domain $\Longleftrightarrow$ m prime. The above example shows that $\mathbb{M}_2(\mathbb{R})$ is not integral domain.*

Theorem 4.38. *A ring R is domain $\Longleftrightarrow$ its set of non-zero elements forms a monoid under multiplication. In other words, R is domain $\Longleftrightarrow$ $R\backslash\{0\}$ is* ***closed under multiplication****, i.e., a monoid.*

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 4.39. *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. □

Corollary 4.40. *Any **skew field** F is an **entire** ring.*

Remark 4.41. *Note* entire *and* invertible *are two different properties. In rudin it directly proves field has the entire property. The key is that roughly speaking, we have invertile $\Rightarrow$ entire.*

Corollary 4.42. ***(Cancel Law of Ring)*** *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.*

Remark 4.43. *See also Remark 3.26 for comparison of cancel law for ring and cancel law for group.*

We next have the other direction of Corollary 4.39:

Theorem 4.44. *Every **finite integral domain** R is a **field**.*

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

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

ψ is well define because R is an integral domain. By the cancellation law for integral domains, we know that given $r_1, r_2 \in R, r_1a = r_2a \Rightarrow r_1 = r_2 \Rightarrow \psi$ injective. Since $R\backslash\{0\}$ is finite, ψ is surjective $\Rightarrow \exists\, b \in R\backslash\{0\}$ such that $ba = ab = 1$. Hence a has a multiplicative inverse. Therefore, $R^* = R\backslash\{0\}$. □

Corollary 4.45. ***(Wedderburn's Little Theorem)*** *We have*

1. *Every **finite domain** R is a **skew field**.*
2. *More generally, **Wedderburn's little theorem** states that every finite domain is a field. In other words, for **finite** rings, **there is no distinction between domains, division rings and fields.***

We next show several applications using Cancellation Law of Ring Corollary 4.42

Lemma 4.46. *If matrix $AB = 1$, then $BA = 1$*

Proof. Since $AB = I$ then $B = B(AB) = (BA)B$. Note from $AB = I$ that $1 = \det(AB) = \det(A)\det(B)$ so $\det(B) \neq 0$. So by $(BA)B = B$ we have: $(BA - I)B = 0$. Since $\det(B) \neq 0 \Leftrightarrow B$ is invertible (i.e. $Bx = 0 \Leftrightarrow x = 0$) $\Leftrightarrow B$ is not a 0 divisor. So $BA = I$. □

4.4 Polynomial Rings

4.4.1 One Variable Polynomial Ring

We consider **ring** R**.**

Definition 4.47. ***(polynomial ring)*** *The polynomial ring in X with **cofficients in a ring** R consists of formal expressions of the form:*

$$g(X) = b_0 + b_1X + b_2X^2 + \cdots + b_mX^m, b_i \in R, m \in \mathbb{N}$$

If $f(X) = a_0 + a_1X + \cdots + a_nX^n$ is another polynomial then we define that

$$f(X) = g(X) \Longleftrightarrow a_i = b_i, \forall\, i$$

*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** $\times$ 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. $f(X) \times g(X) = (a_0b_0) + (a_0b_1 + a_1b_0)X + (a_0b_2 + a_1b_1 + a_2b_0)X^2 + \cdots + a_nb_mX^{n+m}$

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

Remark 4.48. ***(explanation)***

1. *In general $f \times g \neq g \times f$*

2. *Note that here we may view a polynomial as **a sequence with only finite manly nonzero elements**.*

Lemma 4.49. *Check the above $R[X]$, the set of polynomials in X with coefficients in R, has the ring structure with*

1. *0 and 1 in R are the zero and one in $R[X]$, i.e. $m = 0$*

2. *If further R commutative $\Rightarrow$ $R[X]$ commutative ring.*

Proof. The Abelian group properties of $R[X]$ follows easily from the Abelian group properties of R and the coefficients from all from R. We next prove the the associated property in the monoid of the multiplication: $(f \times g) \times h = f \times (g \times h)$, where $f = \sum a_i X^i$, $g = \sum b_i X^i$ and $h = \sum c_i X^i$.

$$\begin{aligned}(f \times g) \times h &= (\sum_k \sum_{i+j=k} a_i b_j X^k)(\sum c_l X^l)\\ &= \sum_w \sum_{i+j+l=w} a_i b_j c_l X^w\end{aligned}$$

We also get

$$f \times (g \times h) = \sum_w \sum_{i+j+l=w} a_i b_j c_l X^w.$$

Compare the w-th item, we get they are same. Finally the distributive law $(f+g) \times h = f \times h + g \times h$ is also obvious. Commutative case is easy. □

• There is a natural embedding:

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

• 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.

Definition 4.50. ***(root (zero))*** *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 4.51. ***(degree, monic, constant)*** *Let R be a ring and $f \in R[X]$ be a non-zero polynomial. Hence we may write $f = c_n X^n + c_{n-1} X^{n-1} + \cdots + c_0, c_i \in R, c_n \neq 0$.*

1. *We call n the **degree** of f and write $\deg(f) = n$.*

2. *If in addition $c_n = 1$, we say that f is **monic**.*

3. *Elements of degree 0 are called **constant polynomials**.*

Theorem 4.52. *If R is **entire** then $R[X]$ satisfies:*

1. *$\forall f, g \in R[X] \backslash \{0\}, \deg(f+g) \leq \max\{\deg(f), \deg(g)\}$*

2. *$\forall f, g \in R[X] \backslash \{0\} \Rightarrow fg \neq 0$ 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, $a_n b_m \neq 0 \Rightarrow$ "$\deg(fg) = n + m = \deg(f) + \deg(g)$, and $fg \neq 0$." □

Corollary 4.53. *We have that*

1. *R **entire** $\Rightarrow R[X]$ **entire.***
2. *R **an integral domain** $\Rightarrow R[X]$ **an integral domain.***

Proof. From the last sentence of the above proof. □

Theorem 4.4.B. *Let A and B be rings and let $\sigma : A \to B$ be a **ring homomorphism** with kernel K. Define (with the abuse of notation, better to use for example $\bar{\sigma}$) $\sigma : A[x] \to B[x]$ by*

$$\sigma\left(a_0 + a_1 x + \cdots a_n x^n\right) = \sigma\left(a_0\right) + \sigma\left(a_1\right) x + \cdots + \sigma\left(a_n\right) x^n$$

*We have σ is a **ring homomorphism** from $A[x]$ to $B[x]$.*

Proof. Easy, just check the properties of ring homomorphism Definition 4.5 are satisfied, all from the ring homomorphism of h. □

Corollary 4.4.A. *For $\alpha \in A$, and $f \in A[X]$ we have*

$$\begin{aligned}\sigma(f(\alpha)) &= \sigma\left(c_n \alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0\right) \\ &= \sigma\left(c_n\right)\sigma(\alpha)^n + \sigma\left(c_{n-1}\right)\sigma(\alpha)^{n-1} + \cdots + \sigma\left(c_1\right)\sigma(\alpha) + \sigma\left(c_0\right) \\ &= (\sigma f)(\sigma(\alpha)).\end{aligned}$$

*So if $f(\alpha) = 0$ then $(\sigma f)(\sigma(\alpha)) = \sigma(f(\alpha)) = \sigma(0) = 0$, so σ sends any **root** α of $f(X)$ in A to a **root** $\sigma(\alpha)$ of $(\sigma f)(X)$ in B.*

4.4.2 Multi Variable Polynomial Ring:

Definition 4.54. ***(multi-variable polynomial ring)*** *Since $R[X]$ is still a ring. Inductively, we can define*

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

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

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

where the sum is finite.

Definition 4.55. ***(monomials)*** *Expressions of the form $m_{(i)} = X_1^{i_1} \cdots X_n^{i_n}$ are called **monomials**.*

Corollary 4.56. *Analogous to Corollary 4.53, We have that*

1. *R **entire** $\Rightarrow R[X_1, X_2, \cdots, X_n]$ **entire.***
2. *R **an integral domain** $\Rightarrow R[X_1, X_2, \cdots, X_n]$ **an integral domain.***

4.5 Field of Fractions

In this section, we **construct a field from any integral domain** and contains the integral domain as a subset (i.e. embedding the integral domain to a field), analogous to the **construct of $\mathbb{Q}$ from $\mathbb{Z}$**:

4.5.1 Recall the $\mathbb{Q}$

1. Elements in $\mathbb{Q}$ often are in the form of $\frac{a}{b}$ with $a, b \in \mathbb{Z}, b \neq 0$. This is not unique. $\frac{a}{b} = \frac{c}{d} \Longleftrightarrow ad - bc = 0$.
2. In $\mathbb{Q}$, we define $+$ and $\times$ by the following rules:
   (a) $\frac{a}{b} + \frac{c}{d} = \frac{ad+cb}{bd}$
   (b) $\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$
3. 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:**

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

   Hence, $\mathbb{Q}$ can be thought of as $(\mathbb{Z} \times \mathbb{Z} \backslash \{0\} / \sim)$.
4. 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.

4.5.2 Generalise the Construction on Integral Domain

Definition 4.57. ***(Elements in the Field)*** *Let R be an* ***integral domain****. We define the* ***equivalence relation*** *(See Lemma 4.58) on $R \times R\backslash\{0\}$ by:*

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

Let us denote the equivalence classes by $(R \times (R\backslash\{0\}))/\sim$. For $(a, b) \in R \times (R\backslash\{0\})$ we often denote the equivalence class containing (a, b) by $\frac{a}{b}$.

Lemma 4.58. *$\sim$ is indeed an equivalence relation.*

Remark 4.59. *Note, properties (e.g.commutative, cancel law) of integral domain are quite crucial for the equivalence.*

Proof.

1. $(a, b) \sim (a, b)$ as $ab - ab = 0$ since R is commutative.
2. $(a, b) \sim (c, d) \Rightarrow ad - bc = 0 \Rightarrow cb - da = 0 \Rightarrow (c, d) \sim (a, b)$ since R is commutative.
3. Let $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. Then $ad - bc = 0, cf - de = 0$. Consider

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

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

□

Definition 4.60. ***($+$ and $\times$ in the Field)*** *Let us define multiplication and addition on $R \times R\backslash\{0\}/\sim$ by*

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

Lemma 4.61. *$+$ and $\times$ are well-defined on $(R \times (R\backslash\{0\}))/\sim$.*

Proof. The first thing to note is that if $b, d \in R\backslash\{0\} \Rightarrow bd \in R\backslash\{0\}$ as R is an integral domain. We just need to check that choosing different representatives gives the same answer. □

Corollary 4.62. ***(0 and 1 in the Field)*** *We have*

1. *The zero in $(R \times (R\backslash\{0\}))/\sim$ is given by the equivalence class containing $(0, 1)$.*
2. *The one in $R \times (R\backslash\{0\})/\sim$ is given by the equivalence class containing $(1, 1)$.*

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

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

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

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

Theorem 4.63. *$(R \times (R\backslash\{0\}))/\sim$ is a field.*

Remark 4.64. *It is a straight forward exercise to check that under these operations $(R \times (R\backslash\{0\}))/\sim$ is a commutative ring. Also observe that*

1. *$(a, b) \in (R \times (R\backslash\{0\}))$ is in the* ***zero class*** *if and only if $a = 0$.*
2. *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.

Proof. We just need to check non-zero elements have multiplicative inverses. Let $\frac{a}{b} \in (R \times (R \backslash \{0\}))/\sim$ be non-zero. By Remark 4.64 this implies that $a \neq 0$. Hence $\frac{b}{a} \in (R \times (R \backslash \{0\}))/\sim$. But

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

Hence we are done. □

Definition 4.65. ***(Field of Fraction)*** *From now on we denote the constructed field as:*

1. *Let R be an integral domain. The* ***field of fractions*** *of R is the field*

$$\operatorname{Frac}(R) \coloneqq (R \times (R \backslash \{0\}))/\sim .$$

2. *Given an integral domain R and indeterminants $\{X_1, \cdots, X_n\}$, we know from Corollary 4.56 that $R[X_1, \cdots, X_n]$ is an integral domain. We define*

$$R(X_1, \cdots, X_n) := \operatorname{Frac}(R[X_1, \cdots, X_n])$$

Example 4.66. *The canonical example is* $\operatorname{Frac}(\mathbb{Z}) = \mathbb{Q}$.

Theorem 4.67. *The map*

$$\begin{aligned} \phi : R &\to \operatorname{Frac}(R) \\ a &\mapsto \frac{a}{1} \end{aligned} \tag{4.68}$$

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

Corollary 4.70. *Let R be a field. The natural embedding* (4.68) $R \to \operatorname{Frac}(R)$ *is an* ***isomorphism***.

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

4.6 Characteristic: Additive Order of Elements

Let R be **entire** (non-trivial with no zero-divisors).

Theorem 4.71. *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 $a \in R\backslash\{0\}$ 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 $b \in R\backslash\{0\}$ 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 $k1_R = (rs)1_R = (r1_R)\,(s1_R) = 0$ (See Lemma 4.11). 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 4.72. ***(Characteristic)*** *Suppose R an **entire** ring.*

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

Remark 4.73. ***(explanation)***

1. *Recall that in Remark 2.39 we have mentioned that:*

   $$\mathbb{Z}/m\mathbb{Z} \text{ is a field } \iff m \text{ is a prime.}$$

   *Theorem 4.71 can be viewed as the **generalization** of Remark 2.39 where prime is from Remark 2.39, and same order of all elements follows from Lemma 3.58.*
2. *Under finite assumption, in general we have field ⇔ entire (See Corollary 4.45), we therefore have that:*
   The element in any finite field or entire must have prime order.
3. *Note this **does not** means any finite field have prime order. Instead, we have p^n as shown in Theorem 4.74.*

Theorem 4.74. *Let F be a finite field. The order of F is always of order p^n where p is prime.*

Proof. Directly follows from Lemma 3.222. Here we give a detailed proof again, but the proof idea is the same:

Let p be the characteristic of a finite field F. Then since 1 has order p in $(F, +)$, we know that p divides $|F|$. Now let $q \neq p$ be any other prime dividing $|F|$. Then by Theorem 3.103 and Lemma 3.58, there is an element $x \in F$ whose order in $(F, +)$ is q.

Then $qx = 0$. But we also have $px = 0$. Now since p and q are relatively prime, we can find integers a and b such that $ap + bq = 1$.

Thus $(ap + bq)x = x$. But $(ap + bq)x = a(px) + b(qx) = 0$, giving $x = 0$, which is not possible since x has order at least 2 in $(F, +)$. So there is no prime other than p which divides $|F|$. □

Example 4.75. *When studying abstract fields, the characteristic is very important. Eg. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all fields (hence entire) of characteristic zero.*

Definition 4.76. *($\mathbb{F}_p$) 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 4.77. *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.

Step 1: Because $\operatorname{char}(F) = 0$, the natural map homomorphism:

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

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.**

Step 2: Now we will extend this notion to $\mathbb{Q}$.

We define the following map:

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

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

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

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

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

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

By definition $\psi\left(\frac{1}{1}\right) = 1_F$. 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 $\psi\left(\frac{n}{m}\right) = 0$. 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 4.78. *There is an embedding of $\mathbb{F}_p$ in any field F of characteristic p, where p is prime.*

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

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

□

4.7 Ring Extensions

Let R be a subring of **commutative ring** S.

Definition 4.79. ***(Ring Extension)*** *The **ring extension** of R by $\{\alpha_1, \cdots, \alpha_n\} \subset S$ is the subring*

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

Lemma 4.80. *$R[\alpha_1, \cdots, \alpha_n]$ is a subring. Further more it is the intersection of all subrings containing R and the subset $\{\alpha_1, \cdots, \alpha_n\}$.*

Remark 4.81. *So $R[\alpha_1, \cdots, \alpha_n]$ is the smallest subring containing R and $\{\alpha_1, \cdots, \alpha_n\}$.*

4.8 Principal, Prime and Maximal Ideals

Definition 4.82. ***(proper ideal)*** *An ideal $I \subset R$ is **proper** if $I \neq R$.*

Lemma 4.83. *I does not contain a **unit** $\Longleftrightarrow I \subset R$ is proper $\Longleftrightarrow R/I$ is a non-trivial ring.*

Proof. This is because ideal I contains a **unit** $\Longleftrightarrow$ I contains **one**. □

Definition 4.84. ***(principal ideal)*** *Let R be a* ***commutative*** *ring. We say an ideal $I \subset R$ is* ***principal*** *if there exist $a \in R$ such that $I = \{ra \mid r \in R\}$. In this case we write $I = (a) := \{ra \mid r \in R\}$.*

Lemma 4.85. *Let R be a* ***commutative*** *ring. $(a) := \{ra \mid r \in R\}$ is an ideal and it is the smallest ideal containing a.*

Proof. From definition, it is easy to check and here **commutative** is import. □

Lemma 4.86. *Let $I = (a)$, we have*

I ***is proper*** $\iff$ a ***is not unit*** $\iff$ I ***does not contain unit***

Proof. Others are easy. Here we only point out if a is not a unit, $1 \in (a)$. □

Definition 4.87. ***(prime ideal)*** *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 $b \in I$. In other words, the set $R \backslash I$ (not the quotient ring R/I) is closed under $\times$.*

Remark 4.88. *Later, we will see in* ***integral domain****,*

principle ideal (a) is a ***prime ideal*** $\iff$ *a is a* ***prime element****.*

Lemma 4.89. *Let R be a* ***commutative ring****. Let $I \subset R$ be an ideal. Then*

I is ***prime*** $\iff$ *R/I is an* ***integral domain****.*

Remark 4.90. *Analogously, if we remove the commutative condition in the definition of prime Definition 4.87, we have under R is ring, prime $\iff$ entire.*

Proof. $\Rightarrow$: 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.

$\Leftarrow$: 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 4.91. ***(maximal ideal)*** *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 $J = R$*

Remark 4.92. *This is a partial order, so the maximal ideal may be not unique.*

Lemma 4.93. *Let R be a* ***commutative*** *ring. Let $I \subset R$ be an ideal. Then*

I is ***maximal*** $\iff$ *R/I is a* ***field****.*

Proof. $\Rightarrow$ 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:

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

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.

$\Leftarrow$ 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 4.94. *Let R be a* ***commutative*** *ring. Let $I \subset R$ be an ideal. Then*

I ***maximal*** $\Rightarrow$ *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, Irreducible and Prime Elements

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.

If R is not commutative or has zero-divisors the factorisation is not well defined or too complicated. If c is a zero-divisors, $(a-b)c = 0$, but we have $ac = bc$ with $a \neq b$.

• We assume R is an **integral domain**.

4.9.1 Unique Factorization Domain, Irreducible Elements

We definite the analogy concept of **basic (i.e. irreducible)** divisors in R that will works similar to prime integer in $\mathbb{Z}$. Later we will define general **prime element** in R and show the relation between **irreducible element** and **prime element**.

Definition 4.95. *Let $a, b \in R$. $a \mid b$ will mean that $\exists\, c \in R$ such that $b = ac$.*

Definition 4.96. ***(associated)*** *Two non-zero elements a, b in an integral domain R are* ***associated*** *if $a \mid b$ and $b \mid a$, i.e. $\exists\, c, d \in R$ such that $b = ac$ and $a = bd$.*

Theorem 4.97. *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 $u \in R^*$*

Remark 4.98. *From Theorem 4.97, we now have an* ***equivalent*** *class: a and b are in the same class $\iff$ a and b are associates. Recall Lemma 4.29, for any ring R, $(R^*, \times)$ is a group. Here it looks like the coset generated by subgroup $(R^*, \times)$ but note now R is just a monoid under $\times$.*

Proof. Association of a and $b \Rightarrow a \mid b$ and $b \mid a \Rightarrow \exists\, c, d \in R$ 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. □

Lemma 4.99. *Let R be an integral domain with $a, b \in R$. Then*

1. *$(a) \subset (b) \iff b \mid a$.*
2. *If $a = bc$ where c is not a unit, the inclusion is strict.*
3. *a and b are* ***associated*** *$\iff (a) = (b)$.*

Proof. We only show 2. Others are obvious. If $(a) = (b)$, we have $a \mid b$ which means a and b are associated. But this is not possible by Theorem 4.97. □

Example 4.100. *In $\mathbb{Z}$, m and n are associated if and only if $n = \pm m$.*

Definition 4.101. ***(irreducible element)*** *We call $a \in R \backslash \{0\}$ an* ***irreducible element*** *if it satisfies the two conditions:*

1. *It is a non-unit;*
2. *It is* ***NOT*** *the product of two non-units.*

Lemma 4.102. *If a is* ***irreducible*** *then so are* ***all its associates****.*

Remark 4.103. *From Remark 4.98, we then can state that whether* ***equivalent*** *class is irreducible.*

Proof. Obvious p is irreducible $\iff pu$ is irreducible where $u \in R^*$. The key is the closedness of the unit set R^* under the multiplication according to Lemma 4.29: au is not unit $\iff a$ is not unit. □

Example 4.104. *Recall $\mathbb{Z}$, we have*

1. *In $\mathbb{Z}$, the units are ± 1, m is* ***irreducible*** *if and only if it is ± 1 times a prime.*
2. *The Fundamental Theorem of Artithmetic 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. $10 = 2 \times 5 = (-2) \times (-5)$.*

Definition 4.105. ***(Unique Factorization Domain)*** *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 two 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 $i \in \{1, \cdots, n\}$.

Remark 4.106. *Note*

1. *Clearly $\mathbb{Z}$ is a UFD by the* ***Fundamental Theorem of Artithmetic****.*
2. *But* ***not all integral domains are UFDs.***

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 $a, b \in R\backslash\{0\}$. We first give the definition of them.

Definition 4.107. ***(Highest Common Factor)*** *Given $a, b \in R\backslash\{0\}$ a* ***highest common factor*** *of a and b is element $d \in R$ such that*

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

Definition 4.108. ***(Lowest Common Multiplier)*** *Given $a, b \in R\backslash\{0\}$ a* ***lowest common multiplier*** *of $a, b \in R$ is an element $c \in R$ such that*

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

Remark 4.109. ***(explanation)***

1. *$a \mid b$ can be viewed as a* ***partial ordering.***
2. *The common factors set F of a and b can deployed this partial ordering, and* HCF *can be viewed as the* ***upper bound*** *of the set where each element in F need to be comparable with* HCF*. That why viewed it as a maximal elements it not good.*
3. *Similar for the common multiplier set,* LCM *can be viewed as the* ***lower bound*** *of the set.*
4. *It is* ***not true*** *that* HCF *and* LCM *exist in an arbitrary integral domain.*
5. *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.***

Lemma 4.110. *In a* ***UFD****,*

1. HCF *and* LCM *exist for two $a, b \in R\backslash\{0\}$.*
2. *Furthermore, 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 $\mathrm{HCF}(a, b) = p_1^{\gamma_1} \cdots p_r^{\gamma_r}$ where $\gamma_i = \min(\alpha_i, \beta_i)$.*
3. *Similarly, $\mathrm{LCM}(a, b) = p_1^{e_1} \cdots p_r^{e_r}$ where $e_i = \max(\alpha_i, \beta_i)$.*

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 $\delta_i \leq \beta_i, \forall\, i$. (If you don't see why please see Sec. 3.5 in note "logic, proof and number theory" for why Fundamental Theorem of Arithmetic implies LCM and HCF). Therefore, $\delta_i \leq \gamma_i \Rightarrow \mathrm{HCF}(a, b) = p_1^{\gamma_1} \cdots p_r^{\gamma_r}$. The proof for LCM is similar. □

Remark 4.111. *From definition, we have*

1. *If $a \in R$ a unit then*

$$\mathrm{HCF}(a,b) = 1, \mathrm{LCM}(a,b) = b, \forall\, b \in R \backslash \{0\}$$

2. *If $a \mid b$, $\mathrm{HCF}(a,b) = a$ and $\mathrm{LCM}(a,b) = b$.*

4.9.2 Prime Elements

In Section 4.9.1, we introduce irreducible element in integral domain which plays the role of prime integers in $\mathbb{Z}$. Let us now introduce another natural generalisation of prime number to an arbitrary integral domain.

Definition 4.112. ***(prime element)*** *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 $p \mid b$*

Remark 4.113. ***(explanation)*** *The about is the analogy to Corollary 2.20, we know*
p is a prime $\iff$ 1). $p \neq 1$ and $p \neq 0$; 2). $\forall\, a, b \in \mathbb{N}^+, p|ab \Rightarrow p|a$ or $p \mid b$

Here in Definition 4.112, we instead take the form of Euclid's Lemma as definition.

Example 4.114. *In $\mathbb{Z}$ prime elements are the prime numbers and their negatives.*

Lemma 4.115. *All elements associated to a prime are themselves prime.*

Proof. Easy to prove: p is prime $\iff pu$ is prime for unit $u \in R^*$. □

Theorem 4.116. ***(prime vs. irreducible)*** *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 $p \mid a$ or $p \mid b$. Say $p \mid a \Rightarrow a = pc = abc$ 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. □

Remark 4.117. *We shall see that for a general integral domain the converse does not always hold. But later in the special case principal ideal domain or UFD, as shown in Theorem 4.134 and Corollary 4.119, we have the converse holds.*

Theorem 4.118. *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. $\Rightarrow$: suppose R is a UFD. Then, by definition, (1). holds. Suppose $p_1 \in R$ irreducible. Then suppose $a, b \in R$ such that $p_1 \mid ab$. If $a = 0, p_1 \mid a$ 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 $p_1 \mid ab$, then by the uniqueness of factorization present in a UFD, this forces $(\alpha_1 + \beta_1) > 0 \Rightarrow \alpha_1$ or $\beta_1 > 0 \Rightarrow p \mid a$ or $p \mid b$. Therefore p_1 is prime.

$\Leftarrow$: 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 (directly from (1).), 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.

For $r = 1$: $a_1 = b_1 \cdots b_s$. Since irreducible a_1 is prime from the assumption, $a_1 \mid b_j$ for some j. WLOG, let $j = 1$, so $a_1 \mid b_1$. We then have $a_1 = b_1 u, u \in R^*$ according to the assumption that a_1 and b_1 both are irreducible.

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 $r > 1$. Use the same way as above: By hypothesis, a_1 is prime and $a_1 \mid b_1 \cdots b_s \Rightarrow a_1 \mid b_j$ for some j. WLOG assume $j = 1$. b_1 is irreducible and $b_1 = a_1 u \Rightarrow u \in R^* \Rightarrow a_1$ 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 $b_i, \forall\, i \in \{2, \cdots, r\}$. Hence R is a UFD. □

Corollary 4.119. *Let R be a **unique factorization domain**. Consider the following properties of an nonzero, **non-unit** element p of R :*

1. *(p) is a **maximal ideal**.*
2. *(p) is a **prime ideal**.*
3. *p is **prime**.*
4. *p is **irreducible**.*

The following implications hold:

$$(p) \textbf{\textit{ maximal}} \Longrightarrow (p) \textbf{\textit{ prime}} \Longleftrightarrow p \textbf{\textit{ prime}} \Longleftrightarrow p \textbf{\textit{ irreducible}}$$

Remark 4.120.

- *Note here UFD condition is only used to get "irreducible ⇒ prime".*
- ***Other conclusions are correct in general integral domains.***

Proof. We only need to show (p) prime $\Longleftrightarrow p$ prime. Others are obvious from Corollary 4.94, Theorem 4.116 and Theorem 4.118.

$\Rightarrow$: If p is not prime, we have $p|ab$ but $p \nmid a$ and $p \nmid b$. But this means $a \notin (p)$ and $b \notin (p)$, but $ab \in (p)$, which is a contradiction of (p) is prime.

$\Leftarrow$: If same reverse the deduction above. □

4.10 Euclidean Integral Domain

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 **Euclidean rings** there is a faster way to determine the HCF of two elements.

Definition 4.121. ***(Euclidean)*** *If R is an **integral domain**, R is **Euclidean** if it admits a function $\varphi : R\backslash\{0\} \to \mathbb{N} \cup \{0\}$ such that*

(1) $\varphi(ab) \geq \varphi(a), \forall\, a, b \in R\backslash\{0\}$

(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 $\varphi(r) < \varphi(b)$.

Remark 4.122. ***(explanation)***

1. *This is intended to model the behavior of the function*

$$\varphi : \mathbb{Z}\backslash\{0\} \to \mathbb{N} \cup \{0\} : a \to |a|$$

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.*

3. *The well order property (any nonempty set has a minimal element) of $\mathbb{N} \cup \{0\}$ is quite important here.*

Lemma 4.123. *The second axiom of a Euclidean Ring is equivalent to the following:*
(2'): $\forall\, a, b \in R \backslash \{0\}$, if $\varphi(a) \geq \varphi(b)$ then $\exists\, c \in R$ such that either $a = bc$ or $\varphi(a - bc) < \varphi(a)$

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 $a, b \in R \backslash \{0\}$ if $b \mid a$ we are done. Therefore, assume $b \nmid a$. Hence $a - bq \neq 0, \forall\, q \in R$. Choose $q \in R$ 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 4.124. *F **field** $\Rightarrow F[X]$ **Euclidean** with the function being the* $\deg$ *function.*

Proof.

$$\begin{aligned} \varphi : F[X] \backslash \{0\} &\longrightarrow \mathbb{N} \cup \{0\} \\ f &\longrightarrow \deg(f) \end{aligned}$$

Check (1): As F is a field, $F[X]$ is an integral domain. From Theorem 4.52, $\Rightarrow \deg(fg) = \deg(f) + \deg(g) \geq \deg(f), \forall\, g, f \in F[X] \backslash \{0\} \Rightarrow \varphi(fg) \geq \varphi(f) \forall\, f, g \in F[X] \backslash \{0\}$.

Check (2'): Let $f = a_0 + a_1 X + \cdots + a_n X^n$, $g = b_0 + b_1 X + \cdots + b_m X^m$ where $a_i, b_j \in F, n, m \in \mathbb{N} \cup \{0\}$ and $a_n \neq 0, b_m \neq 0$.

Assume $\varphi(f) \geq \varphi(g) \Rightarrow n \geq m \Rightarrow n - m \geq 0 \Rightarrow X^{n-m} \in F[X] \Rightarrow X^{n-m} b_m^{-1} a_n g$ has leading term $a_n X^n \Rightarrow \deg\left(f - X^{n-m} b_m^{-1} a_n g\right) < \deg(f)$.

Hence setting $c = a_n b_m^{-1} X^{n-m}$ we have $\varphi(f - cg) = \deg(f - cg) < \deg(f) = \varphi(f)$. Therefore, (2') is satisfied. □

Remark 4.125. *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.*

Theorem 4.126. ***(The Euclidean Algorithm, HCF)*** *Let R be **Euclidean**, with Euclidean function φ. We have*

1. *Any two $a, b \in R$ have an* $\mathrm{HCF}(a, b)$.

2. *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 $\mathrm{HCF}(a, b) = b$ and we are done setting $u = 0$ and $v = 1$. If not then applying property (2) again we get

$$b = q_2 r_1 + r_2,$$

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 (well ordering of $\mathbb{N} \cup \{0\}$ is used). 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.

$$r_{n-2} = q_n r_{n-1} + r_n$$
$$r_{n-1} = q_{n+1} r_n \text{ (with } r_{n+1} = 0)$$

Assume it is true for r_{i-1} and r_{i-2}. By definition $r_i = -q_i r_{i-1} + r_{i-2}$. 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 $r_n \mid r_{n-1} \to r_n \mid r_{n-2}$. Inductively $r_n \mid r_i$ for all i. In particular $r_n \mid b$ and $r_n \mid r_1 \Rightarrow r_n \mid a$. Hence r_n is a common divisor of both a and b. Let $d \in R$ such that $d \mid a$ and $d \mid b$. Hence $d|(ua + vb) \Rightarrow d|r_n$. Hence $\text{HCF}(a, b) = r_n = ua + vb$ □

Corollary 4.127. ***(LCM)*** *Let R be* ***Euclidean*** *ring. Then for any $a, b \in R\backslash\{0\}$ have*

$$\text{LCM} = \frac{ab}{\text{HCF}(a, b)}.$$

Proof. By the above $\text{HCF}(a, b) = au + bv$ for $u, v \in R$. We will define $m = \frac{ab}{\text{HCF}(a,b)}$. Note that this makes sense as $\text{HCF}(a, b) \mid a$. It is clear that $a \mid m$ and $b \mid m$. Let m' be a common multiple, i.e. $a\,|m', b|\,m'$. Then $ab \mid bm'$ and $ab\,|am' \Rightarrow ab|\,aum' + bvm' \Rightarrow ab \mid (au + bv)m' \Rightarrow ab\,|(a, b)m' \Rightarrow \text{HCF}(a, b)m|\,\text{HCF}(a, b)m'$. Because a and b are non-zero $\text{HCF}(a, b)$ is nonzero. Because R is an integral domain we can cancel resulting in $m \mid m'$. Therefore m is an LCM of a, b □

Remark 4.128. *It is worth mentioning that as of yet we have only shown Euclidean rings admit* HCF *and* LCM. *We do not yet know if they are UFDs.*

4.11 Principal Ideal Domains

Definition 4.129. ***(Principal Ideal Domain (PID))*** *Let R be an integral domain. We say that a R is a* ***principal ideal domain (PID)*** *if* ***every ideal of*** *R* ***is principal****. More precisely, if $I \subset R$ is an ideal then there exists $a \in I$ such that $I = (a)$.*

Theorem 4.130. *R* ***Euclidean*** *$\Rightarrow$ R* ***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 4.131. ***(ascending chain, stationary)*** *Let R be an integral domain and $I_1, I_2, I_3, \cdots$ be a sequence of ideals* ***(not necessarily principle)****. 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 4.132. *R* ***PID*** *$\Rightarrow$ every* ***ascending chain*** *of ideals is* ***stationary****.*

Proof.

$$I_1 \subset I_2 \subset I_3 \subset \cdots$$

be an ascending chain of ideals in R. Let I be the union of all the I_i.

We claim that I **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 4.133. *R **PID** $\Rightarrow$ 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 (c_1). 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 $a = p_1 p_2 \cdots p_r c_r$. □

Theorem 4.134. ***(prime vs. irreducible)*** *Let R be a **PID** and $p \in R$. Then*

1. *p **irreducible** $\Longrightarrow$ (p) **is maximal.***
2. *p **irreducible** $\Longleftrightarrow$ p **prime.***
3. *(p) **prime** $\Longleftrightarrow$ (p) **maximal.***

Proof. 1. 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 not 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.

2. Observe now that $R/(p)$ is a field. Hence $R/(p)$ is an integral domain implying that (p) is a prime ideal. p is prime follows from Remark 4.120. (This is just a repeat of Corollary 4.94 proof.)

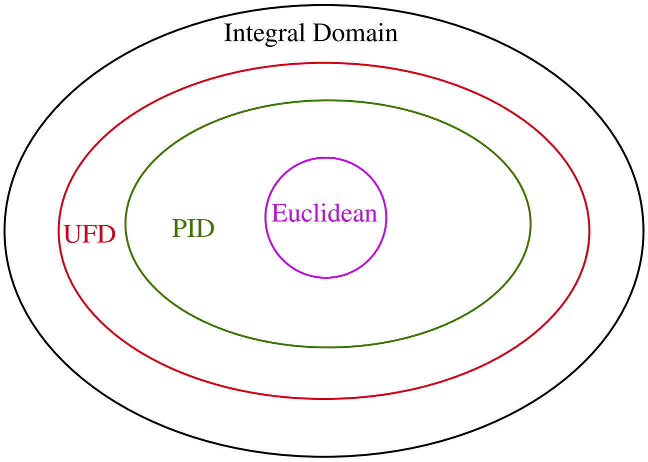
3. "$\Longleftarrow$" is from Corollary 4.94. "$\Longrightarrow$": from 1. and 2. □

Theorem 4.135. *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. So a PID is a UFD according to Theorem 4.118. □

Theorem 4.136. *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. □



4.11.1 Summary: maximal, prime, irreducible

According to Corollary 4.119, Remark 4.120 and Theorem 4.134, we have

1. **integer domain**:
$$(p) \textbf{ maximal} \Longrightarrow (p) \textbf{ prime} \Longleftrightarrow p \textbf{ prime} \Longrightarrow p \textbf{ irreducible}$$
2. **unique factorization domain (UFD)**:
$$(p) \textbf{ maximal} \Longrightarrow (p) \textbf{ prime} \Longleftrightarrow p \textbf{ prime} \Longleftrightarrow p \textbf{ irreducible}$$
3. **principal ideal domain (PID)**:
$$\begin{tikzcd}(p) \textbf{ prime} \arrow[r, Leftrightarrow] \arrow[d, Leftrightarrow] & p \textbf{ prime} \arrow[d, Leftrightarrow] \\ (p) \textbf{ maximal} \arrow[r, Leftrightarrow] & p \textbf{ irreducible}\end{tikzcd}$$
4. **Euclidean**:
$$\begin{tikzcd}(p) \textbf{ prime} \arrow[r, Leftrightarrow] \arrow[d, Leftrightarrow] & p \textbf{ prime} \arrow[d, Leftrightarrow] \\ (p) \textbf{ maximal} \arrow[r, Leftrightarrow] & p \textbf{ irreducible}\end{tikzcd}$$

4.12 Factorization in Polynomial Rings

Theorem 4.137. *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.** We want to study the **factorization in the polynomial ring $F[X]$.**

Lemma 4.138. *$F[X]^* = F^*$, where we view $F \subset F[X]$ as the degree zero polynomials (the constant polynomials).*

Proof. The "one" element in $F[X]$ is $1 \in F \subset F[X]$ which is a degree zero polynomials. If $f \in F[X]$ and $\deg(f) > 0$ then $\deg(fg) > 0, \forall\, g \in F[X] \backslash \{0\}$ according to Theorem 4.52. Thus all invertible elements of $F[X]$ must be degree zero, so they must be constant polynomials. Because F is a field we deduce that $F[X]^* = F^* = F \backslash \{0\}$. □

• We need determine the **irreducible elements** in $F[X]$

Definition 4.139. ***(linear polynomial)*** *We call $f \in F[X]$ such that $\deg(f) = 1$ **linear polynomial***

Lemma 4.140. *Every **linear polynomial** must be **irreducible** for reasons of degree.*

Please see Section 4.12.3 for the summary. Roughly speaking, we next show two cases

1. Algebraically Closed Field: the **irreducible elements** are the **linear polynomials**. **No** others polynomials are linear.
2. If F is not algebraically closed, then the result may be complicated. For example, **irreducible elements** in $F[X]$ at least contain all **primitive irreducible polynomials** in $R[X]$ where F is the field of fraction. See Corollary 4.163 for details.

4.12.1 Algebraically Closed Field

Definition 4.141. ***(algebraically closed)*** *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.*

Theorem 4.142. *Given **field** F,*

$$\textbf{\textit{the only irreducible}}\textit{ elements of } F[X] \textit{ are linear}$$
$$\Updownarrow$$
$$F \textbf{\textit{ is algebraically closed}}$$

Proof. $\Downarrow$: Assume every irreducible in $F[X]$ is linear. Then take $f \in F[X]$ with $\deg(f) > 0$. 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.

$\Uparrow$: 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.

We claim that if $p(\alpha) = 0 \Rightarrow (x - \alpha) \mid p$**.**

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 $(x - \alpha) \mid p$.

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. □

Corollary 4.143. *Let F be a field, we know that $F[X]$ is Euclidean. If $p(\alpha) = 0 \Rightarrow (x - \alpha) \mid p$.*

Remark 4.144. ***(explanation)***

1. *F is **field** is very crucial in Theorem 4.142. It is not always true without this assumption.*
2. *Since $\mathbb{Q}[X]$ does not satisfy every positive degree poly has a root in $\mathbb{Q}$, i.e. $X^2 + 1$, the irreducible elements can be not linear, and the reducible polynomial like $(X^2 + 1)(X^2 + 1)$ can have no root.*

Corollary 4.145. *F **algebraically closed** $\Longleftrightarrow$ **Any** $f \in F[X]$ such that $f \notin F[X]^*, f \neq 0$ can be factored into **linear terms**.*

Theorem 4.146. ***(The Fundamental Theorem of Algebra)*** *$\mathbb{C}$ is **algebraically closed.***

Remark 4.147. *$\mathbb{R}$ and $\mathbb{Q}$ are not algebraically closed. 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.*

Theorem 4.148. *Every field can be embedded in an **algebraically closed field**.*

Example 4.149. *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 $\mathbb{F}_p$.*

Theorem 4.150. *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. □

Corollary 4.151. ***two monic polynomials are associated if and only if they are equal.***

4.12.2 Gauss' Lemma and UFD

• However, for $F = \mathbb{Q}$, it is much complicated and is the starting of algebraic number theory.

Recall that $\mathbb{Q} = \mathrm{Frac}(\mathbb{Z})$. 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**. We next study a general **UFD** R.

Definition 4.152. ***(primitive)*** *$f \in R[X]\backslash\{0\}$ is **primitive** if $\deg(f) > 0$ and its coefficients **do not have an irreducible common factor.***

Example 4.153. *$R = \mathbb{Z}, f = 5x^3 + 3x^2 + 10$ is **primitive**.*

Lemma 4.154. *Let R be a **UFD**. Any element $g(x) \in R[X]$ can be written as*

$$g(x) = dg_1(x)$$

*where $d \in R$ and $g_1(x)$ is **primitive**. Moreover, this **decomposition is unique up to units of R.***

Proof. In fact, let d be a greatest common divisor of the (nonzero) coefficients of g, and let $g_1(x) = (1/d)g(x)$. Then $g_1(x)$ is primitive and $g(x) = dg_1(x)$. Conversely, if $g(x) = dg_1(x)$, where $d \in R$ and $g_1(x)$ is primitive, then d is a greatest common divisor of the coefficients of $g(x)$. Since the **greatest common divisor is unique up to units in** R (See Remark 4.109), it follows that the decomposition is also unique up to units in R (i.e. up to association). □

Definition 4.155. ***(content)*** *Let R be a **UFD** and $f \in R[X]\backslash\{0\}$. 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. Note, according to Lemma 4.154, we have that the **content is unique up to associated.***

Example 4.156. *$R = \mathbb{Z}, f = 9x^3 + 3x + 18$, the content of f is 3.*

Corollary 4.157. *Let F be the **field of fraction** of **UFD** R, Any element $\varphi(x) \in F[X]$ can be written as*

$$\varphi(x) = (d_1/b_1)\, f_1(x),$$

*where d_1 and b_1 are nonzero elements of R and $f_1(x) \in R[X]$ is primitive. Moreover, this **decomposition is unique up to units of R.***

Proof. Just take b to be the product of the denominators of the coefficients of $\varphi(x)$. Factoring $g(x)$ as in Lemma 4.154 gives

$$\varphi(x) = (d/b)f(x),$$

where $f(x)$ is primitive in $R[X]$. This decomposition is unique up to units in R. In fact, if

$$(d_1/b_1)\, f_1(x) = (d_2/b_2)\, f_2(x)$$

where f_1 and f_2 are primitive in $R[X]$, then $d_1 b_2 f_1(x) = d_2 b_1 f_2(x)$. By the uniqueness of the decomposition Lemma 4.154 for $R[X]$, there exists a **unit** u in R^* such that $d_1 b_2 = u d_2 b_1$. Thus $d_1/b_1 = u d_2/b_2$. □

Example 4.158. *Take $R = \mathbb{Z}$.*

$$7/10 + 14/5x + 21/20x^3 = (7/20)\left(2 + 8x + 3x^3\right)$$

where $2 + 8x + 3x^3$ is primitive in $Z[x]$.

Lemma 4.159. ***(Gauss' Lemma)*** *Let R be a **UFD** with field of fractions F.*

1. *The **product of two primitive elements of** $R[X]$ **is primitive.***
2. *Suppose $f(x) \in R[X]$. Then $f(x)$ has a **factorization** $f(x) = \varphi(x)\psi(x)$ in $F[X]$ with $\deg(\varphi), \deg(\psi) \geq 1 \iff f(x)$ has the **same factorization** in $R[X]$. (Note it does not say $\varphi(x) \in R[X]$. It just means "factorization" is the same. See Corollary 4.161 2.)*

Remark 4.160. *Note, here we only mentioned **factorization** but whether it is unique of the factorization of $R[X]$ is not studied. The uniqueness is studied in Corollary 4.163 and Theorem 4.165.*

Proof. 1). 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

$$\pi c_{r+s} = a_0 b_{r+s} + \cdots + a_r b_s + \cdots + a_{r+s} b_0 \Rightarrow a_r b_s = \pi c_{r+s} - a_0 b_{r+s} - \cdots - a_{r+s} b_0$$

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.

2). $\Leftarrow$ is obvious. We need to prove $\Rightarrow$. Suppose that $f(x)$ has the factorization $f(x) = \varphi(x)\psi(x)$ in $F[X]$ with $\deg(\varphi), \deg(\psi) \geq 1$. According to Lemma 4.154 and Corollary 4.157, write $f(x) = ef_1(x), \varphi(x) = (a/b)\varphi_1(x)$ and $\psi(x) = (c/d)\psi_1(x)$, where $f_1(x), \varphi_1(x)$, and $\psi_1(x)$ are primitive in $R[X]$. Then $f(x) = ef_1(x) = (ac/bd)\varphi_1(x)\psi_1(x)$. By part 1), the product $\varphi_1(x)\psi_1(x)$ is primitive in $R[X]$. By the uniqueness of such decompositions in Lemma 4.154 , it follows that $(ac/bd) = eu$, where u is a unit in R, so $f(x)$ factors as $f(x) = ue\varphi_1(x)\psi_1(x)$ in $R[X]$. □

Corollary 4.161. *Let R be a **UFD**. Suppose $f, g \in R[X]\backslash\{0\}$ 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. □

Corollary 4.162. *We have*

1. *If a polynomial $f(x)$ in $\mathbb{Z}[x]$ has a proper factorization in $\mathbb{Q}[x]$, then it has a proper factorization in $\mathbb{Z}[x]$.*
2. *Furthermore, if $f(x)$ has the factorization $f(x) = \varphi(x)\psi(x)$ in $F[X]$ with $\deg(\varphi) \geq 1$ $\deg(\psi) \geq 1$, and $\varphi(x) = (a/b)\varphi_1(x)$ where $\varphi_1(x)$ is primitive. We have $\varphi_1(x) \mid f(x)$*

Proof. It is from Lemma 4.159 and its proof. □

Corollary 4.163. *A **primitive polynomial is irreducible** in $R[X] \iff$ it is **irreducible** in $F[X]$.*

*So we have the conclusion that the **irreducible elements** of $R[X]$ are of two types:*

1. ***irreducible elements** of R, and*
2. ***primitive elements** of $R[X]$ that are **irreducible** in $F[X]$.*

Remark 4.164. *Note, without the assumption **primitive** in $R[X]$, $\Leftarrow$ is **not correct** anymore. 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 $3 \in \mathbb{Q}[X]^*$.*

Proof. Suppose that $f(x) \in R[X]$ is primitive in $R[X]$ and irreducible in $F[X]$. If $f(x) = a(x)b(x)$ in $R[X]$, then one of $a(x)$ and $b(x)$ must be a unit in $F[X]$, so of degree 0 . Suppose without loss of generality that $a(x) = a_0 \in R$. Then a_0 divides all coefficients of $f(x)$, and, because $f(x)$ is primitive, a_0 is a unit in R. This shows that $f(x)$ is irreducible in $R[X]$.

Conversely, suppose that $f(x)$ is irreducible in $R[X]$ and of degree ≥ 1. Then $f(x)$ is necessarily primitive. Moreover, by Gauss's lemma, $f(x)$ has no factorization $f(x) = a(x)b(x)$ in $F[X]$ with $\deg(a(x)) \geq 1$ and $\deg(b(x)) \geq 1$, so $f(x)$ is irreducible in $F[X]$. □

Theorem 4.165. *R is **UFD** $\Longleftrightarrow$ $R[X]$ is **UFD**.*

Proof. Let $g(x)$ be a nonzero, nonunit element of $R[X]$. First, $g(x)$ can be written as $df(x)$, where $f(x)$ is primitive and $d \in R$; furthermore, this decomposition is unique up to units in R. The element d has a unique factorization in R, by assumption, so it remains to show that $f(x)$ has a unique factorization into irreducibles in $R[X]$. But using the factorization (See Theorem 4.137) of $f(x)$ in $F[X]$ and Gauss's Lemma Lemma 4.159 2., we can write

$$f(x) = p_1(x)p_2(x) \cdots p_s(x)$$

where the $p_i(x)$ are **elements of** $R[X]$ **that are irreducible in** $F[X]$**.** Since $f(x)$ is **primitive**, it follows that $p_i(x)$ are primitive as well, and hence irreducible in $R[X]$, by Corollary 4.163.

The **uniqueness** of this factorization follows from the **uniqueness** of irreducible factorization in $F[X]$ together with the uniqueness of the factorization in Corollary 4.157. In fact, suppose that

$$f(x) = p_1(x)p_2(x) \cdots p_s(x) = q_1(x)q_2(x) \cdots q_r(x),$$

where the $p_i(x)$ and $q_i(x)$ are irreducible in $R[X]$. Since $f(x)$ is primitive, each $p_i(x)$ and $q_i(x)$ is primitive, and in particular of degree ≥ 1. By Corollary 4.163, each $p_i(x)$ and $q_i(x)$ is irreducible in $F[X]$. By the uniqueness of the irreducible factorization in $F[X]$, after possibly renumbering the $q_i(x)$, we have $p_i(x) = c_i q_i(x)$ for each i for some $c_i \in F$. But then, by the uniqueness of the decompostion of Corollary 4.157, each c_i is actually a unit in R. □

Corollary 4.166. *Let $f = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[X]$ have a **rational zero** $\frac{\alpha}{\beta}$ where α and β are coprime integers. Then $\beta \mid a_n$, and if $\alpha \neq 0, \alpha \mid a_0$. In particular, if $a_n = 1$, all **rational zeros are integral**.*

Proof. By Corollary 4.143, $f\left(\frac{\alpha}{\beta}\right) = 0 \Rightarrow \left(X - \frac{\alpha}{\beta}\right) \mid f$ in $\mathbb{Q}[X] \Rightarrow \exists\, g \in \mathbb{Q}[X]$ such that $f = \left(X - \frac{\alpha}{\beta}\right) g$. Observe that $\beta X - \alpha$ is primitive, as shown in Corollary 4.162, we deduce that $(\beta X - \alpha) | f \Rightarrow \beta | a_n$ and if $\alpha \neq 0, \alpha \mid a_0$. 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. □

In Corollary 4.163, we have one way to get **irreducible** elements in $F[X]$ (**primitive and irreducible** elements in $R[X]$). Next in Theorem 4.167, we show another interesting criterion for $\mathbb{Q}[X]$:

Theorem 4.167. ***(Eisenstein's Criterion)*** *Let $f = a_0 + a_1x + a_2x^2 + \cdots a_nx^n \in \mathbb{Z}[X] \backslash \{0\}$. If there is a **prime** number $p \in \mathbb{N}$ such that*

1) $p \nmid a_n$

2) $p \mid a_i, \forall\, 0 \leq i < n$

3) $p^2 \nmid a_0$

*then f is **irreducible** over $\mathbb{Q}[X]$.*

Remark 4.168. *Eisenstein's Criterion works (with same proof) for **any UFD and its field of fractions.** Here for simplification, we only show $\mathbb{Z}$ and $\mathbb{Q}$.*

Proof. By Lemma 4.159, we know if f reducible over $\mathbb{Q} \Rightarrow f$ reducible over $\mathbb{Z}$. Suppose that f satisfies the conditions 1), 2), and 3) but is reducible over $\mathbb{Q}$ and hence over $\mathbb{Z}$. 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 $p \mid a_0$ and $p^2 \nmid a_0 \Rightarrow p \nmid b_0$ or $p \nmid c_0$. Without loss of generality assume that $p \mid b_0$ and $p \nmid c_0$. Furthermore, $b_rc_s=a_n$ is not divisible by $p \Rightarrow p \nmid b_r$ and $p \nmid c_s$. Hence the **first coefficient of g is divisible by p but not the last.** While both the first and last coefficient of h is not divisible by p.

Let $i \in \{1,\cdots,r\}$ be minimal such that $p \nmid b_i$. 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 $p \mid a_i$ by 2) and $p \mid b_{i-j}c_j, \forall j \in \{1,\cdots,i\}$ by minimality $\Rightarrow p\,|b_ic_0 \Rightarrow p|\, b_i$ or $p \mid c_0$ which is a contradiction. Hence f is irreducible in $\mathbb{Q}[X]$. □

Corollary 4.169. *There are **irreducible polynomials of arbitrary degree** in $\mathbb{Q}[X]$.*

Proof. If we try to use Corollary 4.163, we need to construct **primitive and irreducible** elements in $\mathbb{Z}[X]$. Primitive property is easy to satisfied, but irreducible is not that easy to check. So we use Theorem 4.167 to construct.

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. □

Corollary 4.170. *Prove x^n+1 is irreducible over $\mathbb{Q}[X]$ iff $n=2^k$ for $k \in \mathbb{N}$*

Proof. $\Rightarrow$: If n is divisible by an odd positive integer d, then x^n+1 is divisible by $x^{n/d}+1$: $x^n+1=(x^{n/d}+1)(x^d-x^{d-1}+\cdots+1)$

$\Leftarrow$: If $f(x)=x^{2^k}+1$, apply Eisenstein to $f(x+1)$. □

4.12.3 Summary and Explanation

Here's a useful analogy from chemistry:

1. Let F be a field. One should think about $f \in F[X]\backslash\{0\}, f \notin F[X]^*$ (up to association) as a **molecule**.
2. One should think about the **irreducible** f (up to association) as **atoms**.
3. The fact that $F[X]$ is a **UFD** says that **every molecule is constructed from a unique finite collection of atoms.**
4. 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.
   (a) 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.
   (b) 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!

4.12.4 Zeros of Polynomials Over a Field

Let's now study the zeros of polynomials over a field. We next show a extended version of Corollary 4.143.

Theorem 4.171. *Let F be a field and $f \in F[X]\backslash\{0\}$ have **distinct roots** $\alpha_1,\cdots,\alpha_n \in F$. Then $(x-\alpha_1)\cdots(x-\alpha_n) \mid f$.*

Proof. We have already proven in Corollary 4.143 that $f(\alpha_i)=0 \Rightarrow (x-\alpha_i) \mid f$. 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 $\Rightarrow (x-\alpha_1)\cdots(x-\alpha_n) \mid f$. □

Corollary 4.172. *Let F be a field and $f \in F[X]\backslash\{0\}$ 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 roots 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 4.173. *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 4.174. *Let F be an **infinite field**. Let $f, g \in F[X]$ such that $f(a) = g(a)$ for all $a \in F$ then $f = g$*

Remark 4.175. *This is **not true if** F **is finite!** For example over $\mathbb{F}_p$ the polynomial $x^p - x$ is zero for every value of $\mathbb{F}_p$.*

Theorem 4.176. *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 $f = 0$. □

4.13 The Unit Group and More

Theorem 4.177. *The unit group of a **finite field** is a **cyclic group**.*

Proof. Let G be the unit group of a finite field, n its order. Let d be a divisor of n, $\psi(d)$ the number of elements order d in G. Suppose there exists an element a of G whose order is d. Let H be the subgroup of G generated by a. Then every element of H satisfies the equation $x^d = 1$ (See Corollary 3.66). Since the number of the solutions of $x^d = 1$ is less than or equal to d and the order of H is d, $H = \left\{x \in G \mid x^d = 1_G\right\}$. Therefore $\psi(d) = 0$ or $\phi(d)$, where $\phi(d)$ is the Euler's function, i.e. **the number of elements of order d in a cyclic group of order d**. Note, here the set containing those elements is only a subset of H and $|H| = d \geq \psi(d)$. Also note finite cyclic group is isomorphism to the $(\mathbb{Z}/d\mathbb{Z}, +)$ (See Theorem 3.60). We would like to find elements in $(\mathbb{Z}/d\mathbb{Z}, +)$ s.t. the order of then is d, this just mean we want to find some element $a \equiv 1 (\mathrm{mod} d)$ (note here 1 is the element in $(\mathbb{Z}/d\mathbb{Z}, +)$, not the 1_G in G). This is because if a satisfies this, we have $\mathrm{gp}(a) = H$. If a does not satisfy this, this just means $1 \notin \mathrm{gp}(a)$ so $H \neq \mathrm{gp}(a)$.

Since $\sum_{d \mid n} \psi(d) = n = \sum_{d \mid n} \phi(d)$, $\psi(d) = \phi(d)$ for all $d \mid n$. In particular $\psi(n) = \phi(n)$, which means there exists an element of order n in G. This completes the proof. □

Corollary 4.178. $\mathbb{U}_p := \mathbb{F}_p^\times$ *($= \mathbb{F}_p \backslash \{0\}$), p is a **prime**, is a **cyclic group** under ×.*

Corollary 4.179. *See supp for details:*

1. *Let p be a **prime**, $n \in \mathbb{N}$. Then $\left(\mathbb{Z}/p^2\mathbb{Z}\right)^\times$, the unit group of $\left(\mathbb{Z}/p^2\mathbb{Z}\right)$, is cyclic.*

2. *Let p be an **odd prime**, $n \in \mathbb{N}$. Then $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic.*

5 Field Theory

5.1 Field Extension and Minimal Polynomial

5.1.1 Field Extension

Definition 5.1. ***(field extension)*** *If F and E are fields, and $F \subseteq E$, we say that E is an extension of F, and we write either $F \leq E$ or E/F.*

Remark 5.2. *Here means isomorphism subset, i.e.* $\mathrm{F} \cong G \subset E$.

Example 5.3. *Here are some classical examples:*

1. $\mathbb{C} = \{a + bi, a, b \in \mathbb{R}\}$ *is a field extension of* $\mathbb{R}$.
2. $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2}, a, b \in \mathbb{Q}\}$ *is a field extension of* $\mathbb{Q}$.
3. $\mathbb{Q}(i) = \{a + bi, a, b \in \mathbb{Q}\}$ *is a field extension of* $\mathbb{Q}$.

Definition 5.4. ***(vector field structure of field extension)*** *If E is an extension of F, then in particular E is an* ***abelian group under addition****, and we may* ***multiply*** *$x \in E$ by $\lambda \in F$. We can see that this* ***endows*** ***E with a structure of F-vector space****:*

1. *elements of E are seen as vectors,*
2. *elements of field F are seen as scalars.*

It then makes sense to speak of the ***dimension*** *of E over F.*

Definition 5.5. ***(degree of the extension)*** *Let E/F be a field extension. The* ***dimension*** *of E as F-vector space is called the* ***degree of the extension****, written $[E : F]$. If $[E : F] < \infty$, we say that E is a* ***finite extension*** *of F, or that the extension E/F is finite.*

Example 5.6. *Let us get back to our examples:*

1. *Consider the field extension $\mathbb{C}/\mathbb{R}$. We have that $\mathbb{C}$ is a vector space of dimension 2 over $\mathbb{R}$. It is thus an extension of degree 2 (with basis $\{1, i\}$).*
2. *The field extension $\mathbb{Q}(\sqrt(2))/\mathbb{Q}$ is of degree 2 , it is called a* ***quadratic extension*** *of $\mathbb{Q}$.*
3. *The field extension $\mathbb{Q}(i)/\mathbb{Q}$ is a also a* ***quadratic field extension*** *of $\mathbb{Q}$.*

Definition 5.7. ***(number field) Finite extensions of $\mathbb{Q}$ are called number fields.***

Example 5.8. *Both $\mathbb{Q}(\sqrt(2))/\mathbb{Q}$ and $\mathbb{Q}(i)/\mathbb{Q}$ are finite field extensions of $\mathbb{Q}$ and therefore are number fields.*

Remark 5.9. *If we look at $\mathbb{C}$, we see it is obtained by adding i to $\mathbb{R}$, and i is a root of the polynomial X^2+1. Similarly, $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is obtained by adding a root of the polynomial X^2-2. See Example 5.24.*

In what follows, we will make formal the **connection between roots of polynomials and field extensions.**

Definition 5.10. ***(field homomorphism)*** *If we have two fields E, F, a* ***field homomorphism*** *between them is a* ***ring homomorphism*** *between two fields.*

Lemma 5.11. ***Field homomorphism*** *f must be a* ***monomorphism****.*

Proof. The kernel of a ring homomorphism is an ideal, and a field has only trivial ideals, namely $\{0\}$ and itself since every non-zeros element is unit. It cannot be that the whole field is the kernel since from ring homomorphism we know $f(1_E) = 1_F$. □

Theorem 5.12. *Let f be a non-constant polynomial over a field F. Then there is an extension E/F and an element $\alpha \in E$, such that $f(\alpha) = 0$.*

Proof. Recall that $F[X]$ is a unique factorization domain, thus f can be factored into a product of irreducible polynomials, and we may assume without loss of generality that f is itself **irreducible**. Consider now the ideal

$$\mathcal{I} = (f(X))$$

in $F[X]$, the ring of polynomials with indeterminate X and coefficients in F. Note, $F[X]$ is Euclidean according to Theorem 4.124, and then from Section 4.11.1, $(f(X)) = \mathcal{I}$ is **maximal**. Thus by the characterization of maximal ideals with respect to their quotient ring Lemma 4.93, we have that

$$E = F[X]/\mathcal{I}$$

is a **field**. We now place an isomorphic copy of F inside E via the **monomorphism**

$$h : F \to E, a \mapsto a + \mathcal{I}.$$

This thus gives a **field extension** E/F. Now let

$$\alpha = X + \mathcal{I} \in E.$$

We are left to prove that α is a **root** of $f(X)$. If $f(X) = a_0 + a_1X + \ldots + a_nX^n$, then

$$\begin{aligned} f(\alpha) &= (a_0 + \mathcal{I}) + a_1(X + \mathcal{I}) + \ldots + a_n(X + \mathcal{I})^n \\ &= a_0 + \mathcal{I} + a_1X + a_1\mathcal{I} + \ldots + a_nX^n + \ldots + a_n\mathcal{I}^n \\ &= (a_0 + a_1X + \ldots + a_nX^n) + \mathcal{I} \\ &= f(X) + \mathcal{I} \end{aligned}$$

which is zero in E. □

Corollary 5.13. ***(extension from adjoining root)*** *Let the extension of F to be $E = F[X]/(f(X))$, where $f(X)$ is an **irreducible polynomial** in $F[X]$. Then $X + (f(X)) \in E$ is a **root** of f. The extension E/F is finite.*

Proof. We only need to show 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 $g(X) \mid f(X)$ or $\exists\, q(X), r(X) \in F[X]$ 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. □

Remark 5.14. *The extension is said to be obtained from F by **adjoining a root of** f.*

*Later in Theorem 5.23, we will know the phenomena shown in Corollary 5.13, i.e. finite dimension, is a general fact, with f (set it to be monic as in Lemma 5.16) is called the **minimal polynomial**. Note this is because the irreducible polynomial with the root $X + (f(X))$ is the **unique monic polynomial** (See Lemma 5.17).*

5.1.2 Minimal Polynomial

Definition 5.15. ***(algebraic, transcendental, algebraic extension)***

1. ***algebraic**: If E is an extension of F, an element $\alpha \in E$ is said to be **algebraic** over F if there is a **non-constant** polynomial $f \in F[X]$ such that $f(\alpha) = 0$.*
2. ***transcendental**: If α is **not algebraic** over F, it is said to be **transcendental** over F.*

*If every element of E is algebraic over F, then E is said to be an **algebraic extension** of F.*

Lemma 5.16. *For a field every non-zero polynomial $p(X)$ has **exactly one unique associated monic polynomial** $q(X)$: p divided by its leading coefficient.*

Suppose that $\alpha \in E$ is algebraic over F. Thus there exists by definition a polynomial $f \in F[X]$ with $f(\alpha) = 0$. It thus makes sense to consider the set $\mathcal{I}$ of all polynomials $g \in F[X]$ such that $g(\alpha) = 0$.

Clearly we have

- if g_1, g_2 are in $\mathcal{I}$, so does $g_1 \pm g_2$,
- if $g \in \mathcal{I}$ and $h \in F[X]$, then $gh \in \mathcal{I}$.

This tells us that

Lemma 5.17. *$\mathcal{I} = \{g \in F[X], g(\alpha) = 0\}$ is an ideal of $F[X]$. Since $F[X]$ is a principal ideal domain, we have*

$$\mathcal{I} = (m(X))$$

for some ***monic*** *$m(X)$ in $F[X]$. Any two generators of $\mathcal{I}$ are thus multiple of each others, so they must be of same degree, and since $m(X)$ is monic, it has to be* ***unique****. This* ***unique*** *polynomial $m(X)$ has the following properties:*

1. *If $g \in F[X]$, then $g(\alpha) = 0$ if and only if $m(X)$ divides $g(X)$. This is clear from the definition of $\mathcal{I}$.*
2. *$m(X)$ is the* ***monic polynomial of least degree*** *such that $m(\alpha) = 0$, which follows from the above property.*
3. *$m(X)$ is the* ***unique monic irreducible polynomial*** *such that $m(\alpha) = 0$.*

Remark 5.1.A. *Note the first uniqueness is from the generators are monic and are multiple of each others. While the uniqueness in 3. is a different story emphasizing that we find a monic irreducible polynomial with $m(\alpha) = 0$, it should be unique and should be* ***the*** *generator.*

Proof. All are clear. We only need to show 3.:

irreducible: If $m(X) = h(X)k(X)$ with $\deg h < \deg m, \deg k < \deg m$, then either $h(\alpha) = 0$ or $k(\alpha) = 0$, so that either $h(X)$ or $k(X)$ is a multiple of $m(X)$ by the first property, which is impossible. Thus $m(X)$ is irreducible.

unicity: We are left to prove the unicity of $m(X)$. More generally speaking if there were two irreducible monic polynomials $m(X)$ and $m'(X)$ such that $m(\alpha) = m'(\alpha) = 0$, we have that $m(X)$ and $m'(X)$ cannot be distinct (see Lemma 5.21 below). □

Definition 5.18. ***(minimal polynomial)*** *The polynomial $m(X)$ is called the* ***minimal polynomial*** *of α over F. It may be denoted by $\min(\alpha, F)$ or $\mu_{\alpha,F}$.*

Corollary 5.19. ***minimal polynomial*** *is* ***unique, monic, irreducible polynomial****. A monic irreducible polynomial with a root $= \alpha$ must be* ***the*** *unique* ***minimal polynomial*** *$\min(\alpha, F)$.*

Example 5.20. *The polynomial $X^2 + 1$ is the minimal polynomial of i over $\mathbb{Q}$. It also the minimal polynomial of i over $\mathbb{R}$.*

Lemma 5.21. *We have that*

1. *Let f and g be polynomials over the field F. Then f and g are* ***relatively prime*** *if and only if f and g have* ***no common root*** *in* ***any*** *extension of F.*
2. *If f and g are* ***distinct monic irreducible polynomials*** *over F, then f and g have* ***no common*** *roots in* ***any*** *extension of F.*

Proof. 1. "⇒": If f and g are relatively prime, their greatest common divisor is 1 , so there are polynomials $a(X)$ and $b(X)$ over F such that

$$a(X)f(X) + b(X)g(X) = 1$$

If there is a common root say α, then we get that $0 = 1$, a contradiction.

"⇐": Conversely, let us assume that the greatest common divisor $d(X)$ of $f(X)$ and $g(X)$ is non-constant and show that then $f(X)$ and $g(X)$ have a common root. By Theorem 5.12, there exists E an extension of F in which $d(X)$ has a root α. Since $d(X)$ divides both $f(X)$ and $g(X), \alpha$ is a common root of f and g in E.

2. By the first part, it is enough to show that f and g are **relatively prime**. Let h is a non-constant divisor of the polynomials f and g which are **irreducible**. Then $f = f'h$ and $g = g'h$ with f', g' **nonzero** constant, and $h = \frac{f}{f'} = \frac{g}{g'}$, that is, $f = \frac{f'}{g'}g$. According to Corollary 4.151, it is impossible for f to be a constant multiple of g, because f and g are monic and distinct. □

Definition 5.22. ***(generated smallest field*** $F(\alpha)$***)*** *If E is an extension of F and $\alpha \in E$ is a **root** of a polynomial $f \in F[X]$ (i.e. α is algebraic over F), one may consider the **field** $F(\alpha)$ **generated** by F and α, which is the **smallest subfield** of E containing both F and α.*

Alternatively, $F(\alpha)$ can be described as

1. *the **intersection of all subfields of** E containing F and α, or*
2. *the **set of all rational functions***

$$\frac{a_0 + a_1\alpha + \cdots + a_m\alpha^m}{b_0 + b_1\alpha + \ldots + b_n\alpha^n}$$

with $a_i, b_j \in F, m, n = 0, 1, \ldots$ and the denominator is different from 0.

Theorem 5.23. *Let $\alpha \in E$ be **algebraic over** F, with **minimal polynomial** $m(X)$ over F of degree n.*

1. *We have $F(\alpha) = F[\alpha] = F_{n-1}[\alpha]$ where $F_{n-1}[\alpha]$ denotes the set of all polynomials of degree at most $n-1$ with coefficients in F. Also, recall $F[\alpha]$ is the **ring extension** as defined in Definition 4.79.*
2. *$\{1, \alpha, \ldots, \alpha^{n-1}\}$ forms a basis for the vector space $F(\alpha)$ over the field F. Consequently $[F(\alpha) : F] = n$.*

Proof. Let us first prove that $F_{n-1}[\alpha]$ is a **field**. Let $f(X)$ be any non-zero polynomial over F of degree at most $n-1$. Since $m(X)$ is irreducible with $\deg f < \deg m$, $f(X)$ and $m(X)$ are relatively prime, and there exist polynomials $a(X)$ and $b(X)$ over F such

$$a(X)f(X) + b(X)m(X) = 1.$$

Using that α is a root of m, we get

$$a(\alpha)f(\alpha) = 1$$

so that any non-zero element of $F_{n-1}[\alpha]$ **has an inverse in** $F[\alpha]$. However from that the degree function on $F[X]$ is Euclidean, we know every $g(X)$ can be written as $g(X) = m(X)k(X) + r(X)$ with $\deg(r) < n$ (including $r = 0$). That just means every element in $F[\alpha]$ must be inside $F_{n-1}[\alpha]$.

1. Thus

$$F_{n-1}[\alpha] = F[\alpha] \subset F(\alpha).$$

But $F(\alpha)$ is the smallest field containing F and α, so

$$F(\alpha) \subset F_{n-1}[\alpha]$$

and we conclude that

$$F(\alpha) = F[\alpha] = F_{n-1}[\alpha].$$

2. Now $1, \alpha, \ldots, \alpha^{n-1}$ certainly **span** $F_{n-1}[\alpha]$, and they are linearly independent because if a non-trivial linear combination of them were zero, this would yield a non-zero polynomial of degree less than that of $m(X)$ with α as a root, a contradiction. □

Example 5.24. *We give two example:*

1. *Let ζ_5 denote a primitive 5th root of unity (that is, $\zeta_5^5 = 1$ and $\zeta_5^k \neq 1$ for $1 \leq k \leq 4$). We have that $\zeta_5 \in \mathbb{Q}(\zeta_5)$ is algebraic over $\mathbb{Q}$, with **minimal polynomial** $X^4 + X^3 + X^2 + X + 1 = 0$ of degree 4 over $\mathbb{Q}$ (by factoring $X^5 - 1$). A $\mathbb{Q}$-basis is given by $\{1, \zeta_5, \zeta_5^2, \zeta_5^3\}$ and $[\mathbb{Q}(\zeta_5) : \mathbb{Q}] = 4$*
2. *For the root $X^2 + 1 = 0$, denoted by i, we have the field extension $\mathbb{Q}(i)$. $X^2 + 1 = 0$ is the minimal polynomial. So we have the $[\mathbb{Q}(i) : \mathbb{Q}] = 2$. The basis is given by $\{1, i\}$. And the field is therefore $\mathbb{Q}(i) = \{a + bi, a, b \in \mathbb{Q}\}$ is a field extension of $\mathbb{Q}$.*

In Corollary 5.13, we have prove $F[X]/(f(X))$ where $f(X)$ is an **irreducible polynomial** in $F[X]$ is a finite extension. Here is a more detailed fact:

Corollary 5.1.B. ***(extension from adjoining root)*** *Let the extension of F to be $E = F[X]/(f(X))$, where $f(X)$ is an* ***irreducible polynomial*** *in $F[X]$ with* ***degree*** *n. Then $\alpha = X + (f(X)) \in E$ is a* ***root*** *of f. The extension E/F is finite with* ***degree*** *n and*

$$F[X]/(f(X)) = F(\alpha),$$

where $\alpha = X + (f(X)) \in E$.

Proof. From definition and Corollary 5.13, we know $f(X)$ is the minimal polynomial over F for the root $\alpha = X + (f(X)) \in E$, we then have the $F(\alpha) = F[\alpha] = F_{n-1}[\alpha]$ with degree n. Note $F_{n-1}[\alpha]$ is just the $\{1 + (f(X)), X + (f(X)), \cdots, X^{n-1} + (f(X))\} \subset E$ (forms a spanning set for E over F) used in the proof of Corollary 5.13 (But we can only state the degree $\leq n$ there). Now, we know, this spanning set must form as the basis as shown in Theorem 5.23. $F[X]/(f(X))$ is a field in the spanning set. So $F[X]/(f(X)) = F_{n-1}[\alpha]$ which means it should be the smallest generated field.. □

5.1.3 Finite and Tower Field Extension

Recall that an algebraic extension is a field extension where every element is algebraic. The result below describes families of algebraic extensions.

Theorem 5.25. *If E is a* ***finite extension*** *of F, then E is an* ***algebraic extension*** *of F.*

Remark 5.26. *The converse is not true. There are infinite algebraic extensions, for example, the field of all* ***algebraic numbers*** *over the rationals is algebraic and of infinite degree. See Example 5.28.*

$\star$ ***algebraic numbers****: a number that is a* ***root of a non-zero polynomial*** *in one variable with* ***integer*** *(or, equivalently, rational) coefficients*

Proof. Let $\alpha \in E$ with degree $[E : F] = n$. Then $1, \alpha, \ldots, \alpha^n$ are $n + 1$ elements while the dimension is n, so they must be linearly dependent, say

$$a_0 + a_1\alpha + \ldots + a_n\alpha^n = 0, a_i \in F.$$

Take $p(X) = a_0 + a_1X + \ldots + a_nX^n \in F[X], \alpha$ is a root of $p(X)$ and by definition α is algebraic over F. □

Corollary 5.27. *By definition, a* ***number field*** *is a finite extension of $\mathbb{Q}$. Thus a number field is an algebraic extension of $\mathbb{Q}$.*

Example 5.28. *Consider the algebraic numbers $\sqrt[n]{2}$, for which one has*

$$|\mathbb{Q}[\sqrt[n]{2}] : \mathbb{Q}| = n,$$

since the minimal polynomial of $\sqrt[n]{2}$ over $\mathbb{Q}$ is $x^n - 2$, as the latter is irreducible over $\mathbb{Q}$ by Eisenstein's criterion Theorem 4.167.

Since $n \geq 1$ is arbitrary, this shows that the degree of the field of the algebraic numbers over the rationals cannot be finite.

Once we have a field extension K/F, we can take again K as base field and get another field extension E/K, yielding a **tower of extensions** $E/K/F$.

Lemma 5.29. *Consider the field extensions $E/K/F$.*

1. *If $\alpha_i, i \in I$, form a* ***basis*** *for E over K, and $\beta_j, j \in J$ form a* ***basis*** *for K over F, then $\alpha_i\beta_j, i \in I, j \in J$, form a* ***basis*** *for E over F.*

2. *The degree is multiplicative, namely*

$$[E : F] = [E : K][K : F].$$

In particular, $[E : F]$ is ***finite*** *if and only if $[E : K]$ and $[K : F]$ are* ***finite****.*

Proof. 1. Take $\gamma \in E$. Then

$$\begin{aligned}\gamma &= \sum_{i \in I} a_i \alpha_i, a_i \in K \\ &= \sum_{i \in I} \left(\sum_{j \in J} b_{ij} \beta_j \right) \alpha_i, b_{ij} \in F\end{aligned}$$

Thus $\alpha_i \beta_j$ span E over F. We now check the **linear independence**.

$$\sum_{i,j} \lambda_{ij} \alpha_i \beta_j = 0 \Rightarrow \sum_i \lambda_{ij} \alpha_i = 0$$

for all j and consequently $\lambda_{ij} = 0$ for all i, j which concludes the proof.

2. It is enough to use the first part, with

$$[E : K] = |I|, \quad [K : F] = |J|, \quad [E : F] = |I||J|.$$

□

Example 5.30. *Consider the field extension $\mathbb{Q}(\zeta_8)/\mathbb{Q}$ where ζ_8 is a primitive 8th root of unity. We have that*

$$\zeta_8 = \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}$$

and $\mathbb{Q}(\zeta_8)/\mathbb{Q}$ is the same field extension as $\mathbb{Q}(i, \sqrt{2})/\mathbb{Q}$. We have

$$[\mathbb{Q}(i, \sqrt{2}) : \mathbb{Q}] = [\mathbb{Q}(i, \sqrt{2}) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2 \cdot 2 = 4$$

5.2 Splitting Fields and Algebraic Closures

5.2.1 Splitting Fields

Definition 5.31. ***(generated smallest field** $F(\alpha_1, \ldots, \alpha_k)$**)***

1. *For $\alpha \in E$, an extension of F, $F(\alpha)$ in defined as the **intersection of all the subfields** of E containing F and α.*
2. *This can be of course generalized if we pick $\alpha_1, \ldots, \alpha_k \in E$, and $F(\alpha_1, \ldots, \alpha_k)$ is the **intersection of all the subfields** of E containing F and $\alpha_1, \ldots, \alpha_k$.*

Remark 5.32. *Note, Definition 5.22 has the assumption that $\alpha \in E$ being **algebraic over** F, this is because with this assumption, in Theorem 5.23 has good properties. In this general definiton, we do **NOT** assume this. But in many cases, we use Theorem 5.12 to get the **algebraic over**. See section 5.2.1.2 for more discussion.*

Definition 5.33. ***(split polynomial** f**)** If E is an extension of F and polynomial $f \in F[X]$, we say that f **splits over** E if f can be written as $\lambda (X - \alpha_1) \cdots (X - \alpha_k)$ for some $\alpha_1, \ldots, \alpha_k \in E$ and $\lambda \in F$.*

Definition 5.34. ***(splitting field for a polynomial)** If K is an extension of F and polynomial $f \in F[X]$, we say that K is a **splitting field for** f **over** F is f **splits over** K **but not over any proper subfield of** K **containing** F.*

Example 5.35. *Consider the polynomial $f(X) = X^3 - 2$ over $\mathbb{Q}$. Its roots are*

$$\sqrt[3]{2}, \sqrt[3]{2}\left(-\frac{1}{2} + i\frac{1}{2}\sqrt{3}\right), \sqrt[3]{2}\left(-\frac{1}{2} - i\frac{1}{2}\sqrt{3}\right).$$

Alternatively, if ζ_3 denotes a primitive 3-rd root of unity, we can write the roots as

$$\sqrt[3]{2}, \zeta_3 \sqrt[3]{2}, \zeta_3^2 \sqrt[3]{2}$$

*The polynomial f is **irreducible** (of course over $\mathbb{Q}[X]$ or $\mathbb{Z}[X]$, for example using **Eisenstein's criterion**). Since it is also **monic**, according to Corollary 5.19 it is the **minimal polynomial** of $\sqrt[3]{2}$, and*

$$[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$$

Now since $\sqrt[3]{2}$ and $i\sqrt{3}$ (or ζ_3) generate all the roots of f, the splitting field of f is

$$K = \mathbb{Q}(\sqrt[3]{2}, i\sqrt{3}) = \mathbb{Q}\left(\sqrt[3]{2}, \zeta_3\right)$$

*We next **compute the degree of K over $\mathbb{Q}$.***

Clearly $i\sqrt{3}$ cannot belong to $\mathbb{Q}(\sqrt[3]{2})$ which is a subfield of $\mathbb{R}$, thus $[\mathbb{Q}(\sqrt[3]{2}, i\sqrt{3}) : \mathbb{Q}(\sqrt[3]{2})]$ is at least 2. Since $i\sqrt{3}$ is a root of $X^2 + 3 \in \mathbb{Q}(\sqrt[3]{2})[X]$, this degree is exactly 2, according to Theorem 5.23.

By multiplicativity of the degrees, we get that

$$[K : \mathbb{Q}] = 6.$$

(Using that ζ_3 is a root of $X^2 + X + 1$ stays irreducible over $\mathbb{Q}(\sqrt{2})$ gives the same result.)

Lemma 5.36. *Let $\alpha_1, \ldots, \alpha_k$ be all the **roots** of $f \in F[X]$ which **splits** over some extended field E.*

1. *K is a **splitting field** for f over F if "f splits over K and K is generated over F by all the roots $\alpha_1, \ldots, \alpha_k$, that is $K = F(\alpha_1, \ldots, \alpha_k)$".*
2. *E contains a **unique splitting field** for f, namely $F(\alpha_1, \ldots, \alpha_k)$.*
3. *$F(\alpha_1, \ldots, \alpha_k) = F[\alpha_1, \ldots, \alpha_k]$.*

Remark 5.37. ***(compare with Theorem 5.23)*** *In Theorem 5.23, we have shown $F[\alpha] = F(\alpha)$ where $F[\alpha]$ is the ring extension and $F(\alpha)$ is the smallest generated field of F and α. $\alpha \in E$ is algebraic over F.*

Proof. 1. is directly from the definition Definition 5.34 as explained above. 2. is also from the definition. Note if there exists another subfield W of E that f splitting, the intersection of two field $W \cap F(\alpha_1, \ldots, \alpha_k)$ is still a field. Because $F(\alpha_1, \ldots, \alpha_k)$ is the smallest one, we then have $W = F(\alpha_1, \ldots, \alpha_k)$. 3. follows from Remark 5.37 by iteratively applying Theorem 5.23. □

Theorem 5.38. *If $f \in F[X]$ and $\deg f = n$, then f has a **splitting field** K over F with $[K : F] \leq n!$*

Remark 5.39. *Note that Example 5.35 shows that this bound is **tight.** A general version is shown in Theorem 5.43. It indicates that **the degree of the splitting field for a polynomial may much larger that the degree of the polynomial.** Need to compare with Theorem 5.23 and Theorem 5.43.*

Proof. First we may assume that $n \geq 1$, for if $n = 0$, then f is constant, and we take $K = F$ with $[K : F] = 1$.

By Theorem 5.12, f has at least one root, saying α_1, and there is an extension E_1 of F containing α_1. Since $f(\alpha_1) = 0$, the minimal polynomial $m_1(X)$ of α_1 divides $f(X)$, that is $f(X) = m_1(X)h(X)$ for some $h(X)$, and since $\deg f = n$, $\deg m_1(X) \leq n$, implying that $F(\alpha_1)/F$ has degree at most n.

We may then further write $f(X) = (X - \alpha_1)^{r_1} g(X)$ where $g(\alpha_1) \neq 0$ according to **??**ccording to Corollary 4.143, and $\deg g \leq n - 1$ (note $(X - \alpha_1)^{r_1} \in F(\alpha)[X]$ is different from the $m_1[X] \in F[X]$, don't be confused). If g is constant, then $f(X)$ has no other root than α_1, and its splitting field is $F(\alpha_1)/F$ whose degree is at most n which is indeed smaller than $n!$

Now if g is non-constant, we can iterate on g the reasoning we did on f. Namely, we have that g has degree at least 1 , and thus it has at least one root α_2. Invoking again Theorem 5.12, there is an extension of $F(\alpha_1)$ containing α_2 and the extension $F(\alpha_1, \alpha_2)$ has degree at most $n - 1$ over $F(\alpha_1)$ (corresponding to the case where $r_1 = 1$). Thus we have

$$\begin{aligned}[F(\alpha_1, \alpha_2) : F] &= [F(\alpha_1, \alpha_2) : F(\alpha_1)][F(\alpha_1) : F] \\ &\leq (n - 1)n.\end{aligned}$$

We can now continue inductively to reach that if $\alpha_1, \ldots, \alpha_n$ are all the roots of f, then

$$[F(\alpha_1, \alpha_2, \ldots, \alpha_n) : F] \leq n!$$

□

If $f \in F[X]$ and f splits over E, then we may take any root α of f and adjoin it to F to get the extension $F(\alpha)$. More precisely:

5.2.1.1 F-isomorphism and Algebraic Extension Transitivity

Theorem 5.40. *If α and β are roots of the irreducible polynomial $f \in F[X]$ in an extension E of F, then $F(\alpha)$ is **isomorphic** to $F(\beta)$.*

$$\begin{tikzcd}
\alpha, \beta \in E \\
F \arrow[u]
\end{tikzcd}$$

Proof. If f is not monic, start by dividing f by its leading coefficient, so that we can assume that f is monic (Lemma 5.16). Since f is monic, irreducible and $f(\alpha) = f(\beta) = 0$, f is the **minimal polynomial** of α and β, say of degree n. According to Theorem 5.23, if $a \in F(\alpha)$, then a can be uniquely written as

$$a = a_0 + a_1\alpha + \ldots + a_{n-1}\alpha^{n-1}.$$

Note $\{1, \alpha, \ldots, \alpha^{n-1}\}$ and $\{1, \beta, \ldots, \beta^{n-1}\}$ are the two basis in $F(\alpha)$ and $F(\beta)$ respectively. The map

$$a_0 + a_1\alpha + \ldots + a_{n-1}\alpha^{n-1} \mapsto a_0 + a_1\beta + \ldots + a_{n-1}\beta^{n-1}$$

defines a field isomorphism between $F(\alpha)$ and $F(\beta)$. □

When discussing field isomorphisms, one may want to **emphasize the base field.**

Definition 5.41. ***(F-isomorphism)*** *If E and E' are extensions of F, and $\iota : E \to E'$ is an isomorphism, we say that ι is an **F-isomorphism** if ι fixes F, that is, if*

$$\iota(a) = a, a \in F.$$

Remark 5.42. ***(explanation)***

1. *So there exists a **F-isomorphism** from $F(\alpha)$ to $F(\beta)$ as constructed in the proof.*
2. *For a more general version, see Lemma 5.57 where it does not require **F-isomorphism**.*

Corollary 5.2.C. *According to Corollary 5.1.B, let the extension of F to be $E = F[X]/(f(X))$, where $f(X)$ is an **irreducible polynomial** in $F[X]$ with **degree** n. We have*

$$F[X]/(f(X)) = F(\alpha),$$

where $\alpha = X + (f(X)) \in E$. Furthermore, for any other possible extension E' of F with with $f(\beta) = 0$. We have

$$F[X]/(f(X)) \cong F(\beta).$$

Proof. hint: According to Theorem 5.23, use the unique representation in two basis for the elements.

$$a_0 + a_1\alpha + \ldots + a_{n-1}\alpha^{n-1} \mapsto a_0 + a_1\beta + \ldots + a_{n-1}\beta^{n-1}$$

□

Remark 5.2.B. *Compare with Theorem 5.40, Corollary 5.2.C can be viewed as a generalization where α and β are not in the same fields. But both contain F as a subfield, and so there still exists a **F-isomorphism** from $F(\alpha)$ to $F(\beta)$ as constructed in the proof.*

Theorem 5.43.

1. *If E is generated over F by finitely many elements $\alpha_1, \ldots, \alpha_n$ **algebraic over** (See section 5.2.1.2) F, then E is a finite extension of F.*

$$\begin{tikzcd}
\alpha \in E & & \beta \in E' \\
& F \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

2. ***(Transitivity of algebraic extensions)*** *If E is **algebraic over** K, and K is **algebraic over** F, then E is **algebraic over** F.*

Proof. 1. Set $E_0 = F, E_k = F(\alpha_1, \ldots, \alpha_k), 1 \leq k \leq n$, in particular $E_n = F(\alpha_1, \ldots, \alpha_n) = E$ by definition of E. Then $E_k = E_{k-1}(\alpha_k)$, where α_k is algebraic over F, and hence over E_{k-1}. Now $[E_k : E_{k-1}]$ is the degree of the minimal polynomial of α_k over E_{k-1}, which is finite. By multiplicativity of the degrees, we conclude that

$$[E : F] = \prod_{k=1}^{n} [E_k : E_{k-1}] < \infty.$$

2. Let $\alpha \in E$ with minimal polynomial

$$m(X) = b_0 + b_1 X + \ldots + b_{n-1} X^{n-1} + X^n$$

over K since by assumption α is algebraic over K. The coefficients b_i are in K and thus are algebraic over F. Set $L = F(b_0, b_1, \ldots, b_{n-1})$, by the first part, L is a finite extension of F. We have $m(X) \in L[X], \alpha$ is algebraic over L, and $L(\alpha) \subseteq E$ is a finite extension of L. This gives us the following tower of field extensions:

$$L(\alpha)/\left(L = F(b_0, b_1, \ldots, b_{n-1})\right)/F.$$

By transitivity of the degrees, since $[L : F] < \infty$ and $[L(\alpha) : L] < \infty$, we get that $[L(\alpha) : F] < \infty$. We conclude α is algebraic over F since we know that all finite extensions are algebraic (Theorem 5.25). □

Corollary 5.44. *If E is an extension of F and A is the set of all elements in E that are **algebraic over** F, then A is a **subfield** of E.*

Proof. If $\alpha, \beta \in A$, then the sum, difference, product and quotient (if $\beta \neq 0$) of α and β belong to $F(\alpha, \beta)$, which is a finite extension of F by the first part of Theorem 5.43. This is thus an algebraic extension since all finite extensions are algebraic (Theorem 5.25). We have that $\alpha + \beta, \alpha - \beta, \alpha\beta$ and α/β in $F(\alpha, \beta)$ (since it is a field) are algebraic and so they are in A, proving that A is a field. □

Theorem 5.2.C. *If K/F is finite, say $n = [K : F] < \infty$, we have $K = F[k_1, \ldots, k_n] = F(k_1, \ldots, k_n)$ is finitely generated for some $k_1, \ldots, k_n$ from K.*

Remark 5.2.C. *Here it may be interesting to notice that $F(\alpha)$ in Theorem 5.23 may be at the same time write as $F[k_1, \ldots, k_n]$. The converse of Theorem 5.2.C is shown in Theorem 5.43.*

Proof. By definition $K = \operatorname{Span}_F \{k_1, \ldots, k_n\}$ where $k_i \in K$ for each i. Hence every element of K can be written as a finite linear combination of the k_i over F, and so we know

$$F(k_1, \ldots, k_n) \subseteq K \subseteq F[k_1, \ldots, k_n].$$

At the same time

$$F[k_1, \ldots, k_n] \subseteq F(k_1, \ldots, k_n) \subseteq K$$

is clear since K is closed under addition and multiplication. So $K = F[k_1, \ldots, k_n] = F(k_1, \ldots, k_n)$ is finitely generated. □

5.2.1.2 Summary of Generated Field

1. In Definition 5.22, we only need to assume that all α_i are inside some extension E of F. This assumption is quite weak and it is only to ensure the operations between elements from F and α_i make sense.
2. The assumption **algebraic over** or **roots** or **splitting** in theorems like Theorem 5.23 and Lemma 5.36 are the same story: adjoining a root to get the extension. This assumption is quite important, it ensures the generated field $F(\alpha_1, \dots, \alpha_k)$ (or $F(\alpha)$) is a finite extension.
3. If we do not assume **algebraic over**, the extension may be infinite. For example $\mathbb{Q}[\pi]$ is an infinite extension of $\mathbb{Q}$ since π is **transcendental** over $\mathbb{Q}$. (From Theorem 5.25, we know **transcendental** extension must be infinite extensions.)
4. Sometimes, we can have $F(\alpha_1, \dots, \alpha_k) = F(\alpha_1, \dots, \alpha_n)$ for different number of adjoining elements (not necessarily roots). For example, in Theorem 5.38 if $f \in F[X]$ and $\deg f = n$, then f has a **splitting field** K over F with $[K : F] \leq n!$, a **tight** bound. But in Theorem 5.2.C, we have we can select another set of $[K : F]$ to be the generating set.
   - Later we will see, in Theorem of the Primitive Element Theorem 5.92, if E/F is a **finite separable** extension, then
   $$E = F(\gamma),$$
   which means **only one other element is enough to generate the field**! (Of course here we can also use another set of $[E : F]$ to be the generating set according to Theorem 5.2.C.)

5.2.2 Algebraic Closure

Factoring in linear polynomials is equivalent to the splitting in Section 5.2.1. **Algebraically closed** in then introduced in Definition 4.141. We show that it is equivalent to that the only irreducible elements are linear (so f must be factored in linear polynomials) in Theorem 4.142 and Corollary 4.145. We now study algebraically closed more deeply.

Example 5.45. *If F is $\mathbb{Q}, \mathbb{R}$ or more generally $\mathbb{C}$, we know that there is a field C with the property that any polynomial in $\mathbb{C}[X]$ splits over C, namely $C = \mathbb{C}$ itself.*

Theorem 5.46. *If C is a field, the following conditions or definitions are equivalent.*

1. *Every non-constant polynomial $f \in C[X]$ has at least one **root** in C (Definition 4.141).*
2. *Every non-constant polynomial $f \in C[X]$ **splits** over C (Corollary 4.145).*
3. *Every **irreducible** polynomial $f \in C[X]$ is **linear** (Theorem 4.142).*
4. *C has **no proper algebraic extension.** (C is the **largest algebraic extension**)*

Corollary 5.47. *From Theorem 5.46 4)., we have*

Algebraically closed** if and only if **the algebraic extension must be itself.

Proof. Note, we only need to prove 4. $\iff$ 1. since others are shown in Section 4.12.1. However, here let's prove 1. $\Rightarrow$ 2. $\Rightarrow$ 3. $\Rightarrow$ 4. $\Rightarrow$ 1.

1. $\Rightarrow$ 2. Take $f \in C[X]$ a non-constant polynomial. Since f has at least one root, we write $f = (X - \alpha_1)\, g$ for g some polynomial in $C[X]$. If g is constant, we are done since f splits. If g is non-constant, then again by assumption it has one root and $g = (X - \alpha_2)\, h$ for some h. We conclude by repeating inductively.

2. $\Rightarrow$ 3. Take $f \in C[X]$ which is irreducible, thus non-constant. By assumption it is a product of linear factors. But f is irreducible, so there can be only one such factor.

3. $\Rightarrow$ 4. Let E be an algebraic extension of C. Take $\alpha \in E$ with minimal polynomial f over C. Then f is irreducible and of the form $X - \alpha \in C[X]$ by assumption. Thus $\alpha \in C$ and $E = C$.

4. $\Rightarrow$ 1. Let f be a non-constant polynomial in $C[X]$, with root α (using for example Corollary 5.13). We can adjoin α to C to obtain $C(\alpha)$. But by assumption, there is **no proper algebraic extension** of C, so $C(\alpha) = C$ and $\alpha \in C$. Thus f has at least one root in C and we are done. Note here $C(\alpha)$ is a

algebraic extension is because any finite extension is algebraic extension Theorem 5.25 and $C(\alpha)$ is finite extension (Theorem 5.23).

□

Remark 5.48. ***(algebraic extension vs. algebraically closed extension)***

1. *Note Corollary 5.47 is* ***NOT*** *saying "****algebraically closed extension is unique (i.e. must be algebraic closure)*** *".*
2. ***Algebraically closed extension*** *and* ***algebraic extension*** *are two different thing:*
   (a) ***algebraic extension*** *E/F require any element in E is a root for one $f \in F[X]$.*
   (b) ***algebraically closed extension*** *E/F require any $f \in E[X]$ can be splitted over E.*

   Does ***algebraically closed extension*** *implies* ***algebraic extension****? **No. For example** $\mathbb{C}$ **is not algebraic extension of** $\mathbb{Q}$.*

   But *the* ***largest algebraic extension*** *is the algebraic closure and of course* ***algebraically closed****.*
3. ***Algebraic closure*** *require both* ***algebraic extension*** *and* ***algebraically closed extension*** *as in the definition Definition 5.50.*
4. ***Algebraic closure (i.e. both algebraic extension and algebraically closed extension)*** *is a* ***minimal algebraically closed extension*** *(Lemma 5.53) and a* ***largest algebraic extension*** *(Theorem 5.46 4.).*
5. *You may then ask why there exists such a field satisfies these two conditions. The* ***existence of algebraic closure*** *for* ***any*** *field in given in Theorem 5.55.*

Example 5.49. 1. *The field $\mathbb{R}$ is not algebraically closed, since $X^2 + 1 = 0$ has not root in $\mathbb{R}$.*

2. ***No finite field*** *$\mathbb{F}$* ***is algebraically closed****, since if $a_1, \ldots, a_n$ are all the elements of F, then the polynomial $(X - a_1) \ldots (X - a_n) + 1$ has no zero in $\mathbb{F}$. (Recall that but any finite extension field is an algebraic entension)*

   *Note but we can embed it in an algebraically closed field as follows (****algebraic closure****).*
3. *The field $\mathbb{C}$ is algebraically closed, this is the* ***fundamental theorem of algebra*** *Theorem 4.146.*
4. ***The field of all algebraic numbers is algebraically closed.*** *(We will not prove this here, but for a proof that algebraic numbers in a field extension indeed form a field, see Corollary 5.44.)*

We can embed an arbitrary field F in an algebraically closed field as follows as we have mentioned in Theorem 4.148.

Definition 5.50. ***(algebraic closure)*** *An extension C of F is called an* ***algebraic closure*** *if C is* ***algebraic extension*** *over F and C is* ***algebraically closed****.*

Remark 5.51. *Note* ***algebraically closed extension (not*** *require to be algebraic extension) may be not unique as $\mathbb{C}$ and all algebraic numbers are both algebraically closed extension of $\mathbb{Q}$. However,* ***algebraically closed algebraic extension*** *is* ***unique up to*** *F****-isomorphic*** *as shown in Theorem 5.55.*

Example 5.52. *Two examples:*

1. *The field $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$.*
2. *The field of all algebraic numbers is the algebraic closure of $\mathbb{Q}$.*

Lemma 5.53. *C is* ***minimal*** *among algebraically closed extensions of F.*

Remark 5.54. *Note, here* ***minimal*** *is the partial ordering, there may exist other algebraic closure. But as stated in Theorem 5.55, they are unique up to F****-isomorphic.***

Proof. let us assume that there is an algebraically closed field K such that $C/K/F$. Let $\alpha \in C$ but $\alpha \notin K$ (it exists if we assume that $C \neq K$). Then α is algebraic over F, and consequently algebraic over K. But since $\alpha \notin K$, the minimal polynomial of α over K cannot contain the factor $X - \alpha$, which contradicts that K is an algebraically closed field. □

We can prove the following theorems (we will omit the proof, but we will prove the uniqueness of the splittig field in Section 5.2.3).

Theorem 5.55. *We have*

1. *Every field F has an* ***algebraic closure.***
2. *Any two algebraic closures C and C' of F are F**-isomorphic.***
3. *If E is an* ***algebraic extension*** *of F, C is an* ***algebraic closure*** *of F, and ι is an embedding of F into C. Then ι can be extended to an embedding of E into C.*

5.2.3 Uniqueness of Splitting Fields

In Definition 5.34 of a splitting field, it is not clear how many splitting fields there are for some fixed $f \in F[X]$. In this section, we prove that any two such splitting fields are **isomorphic**. So every **splitting field of** f is **unique up to isomorphism.**

Lemma 5.56. *We can observe the following*

1. *If $\varphi : F \to F'$ is a* ***field homomorphism*** *(so* ***monomorphism*** Lemma 5.11*), then*
$$\begin{gathered}\tilde{\varphi} : F[X] \to F'[X] \\ a_0 + a_1 x + \cdots + a_n x^n \mapsto \varphi(a_0) + \varphi(a_1)x + \cdots + \varphi(a_n)x^n\end{gathered}$$
is a ***ring homomorphism*** *(See Theorem 4.4.B).*
2. *If $\varphi : F \to F'$ is an* ***isomorphism****, then $\tilde{\varphi}$ is also an* ***isomorphism****, and for any $f \in F[X]$, it sends the* ***ideal*** *$\langle f \rangle$ to the* ***ideal*** *$\langle \tilde{\varphi}(f) \rangle$.*

Lemma 5.57. *Let $\varphi : F \to F'$ be an* ***isomorphism****, and $f \in F[X]$ be* ***irreducible*** *in $F[X]$. If*

1. *α is a root of f in some extension of F,*
2. *β is a root of $\tilde{\varphi}(f)$ in some extension of F',*

then there is an ***isomorphism*** *$\psi : F(\alpha) \to F'(\beta)$ such that $\psi(\alpha) = \beta$, and $\psi|_F = \varphi$.*

$$\begin{tikzcd}
\alpha \in E & \beta \in E' \\
F \arrow[u] \arrow[r, "\varphi"] & F' \arrow[u]
\end{tikzcd}$$

Remark 5.58. ***(compare with Theorem 5.40)*** *We then see Lemma 5.57 is a generalization of Theorem 5.40 or Corollary 5.2.C where ψ is the identity map in Theorem 5.40 and Corollary 5.2.C.*

Proof. Because f is irreducible in $F[X], \varphi(f)$ is irreducible in $F'[X]$. Because $\tilde{\varphi}$ sends $\langle f \rangle$ to $\langle \tilde{\varphi}(f) \rangle$, the map $F[X] \mapsto F'[X]/\langle \tilde{\varphi}(f) \rangle$ will have kernel equal to $\langle f \rangle$, so we have an **isomorphism** (Theorem 4.25)
$$\frac{F[X]}{\langle f \rangle} \to \frac{F'[X]}{\langle \varphi(f) \rangle}$$
that sends a coset to another coset. The composition of **isomorphisms** (See Corollary 5.2.C)
$$F(\alpha) \to \frac{F[X]}{\langle f \rangle} \to \frac{F'[X]}{\langle \tilde{\varphi}(f) \rangle} \to F'(\beta)$$
gives the desired isomorphism. □

Example 5.59. *We can apply the above proposition to the identity isomorphism $\mathbb{Q} \to \mathbb{Q}$, the polynomial $f(x) = x^3 - 2 \in \mathbb{Q}[x]$, and the roots $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and $\beta = \xi\sqrt[3]{2} \in \mathbb{C}$, where $\xi = e^{2\pi i/3}$. The isomorphism ψ from*

$$\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt[3]{2}) = \left\{a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 \mid a_i \in \mathbb{Q}\right\}$$

to

$$\mathbb{Q}(\beta) = \mathbb{Q}(\xi\sqrt[3]{2}) = \left\{b_0 + b_1\xi\sqrt[3]{2} + b_2(\xi\sqrt[3]{2})^2 \mid b_i \in \mathbb{Q}\right\}$$

*is the **identity** on $\mathbb{Q}$ and sends $\sqrt[3]{2}$ to $\xi\sqrt[3]{2}$, so by the properties of homomorphisms it sends*

$$a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 \mapsto a_0 + a_1\xi\sqrt[3]{2} + a_2(\xi\sqrt[3]{2})^2$$

Notice also that the irreducible factorization of f in $\mathbb{Q}(\sqrt[3]{2})$ is given by

$$f = (x - \sqrt[3]{2})\left(x^2 - \sqrt[3]{2}x + (\sqrt[3]{2})^2\right)$$

while the irreducible factorization of f in $\mathbb{Q}(\xi\sqrt[3]{2})$ is given by

$$f = (x - \xi\sqrt[3]{2})\left(x^2 - \xi\sqrt[3]{2}x + (\xi\sqrt[3]{2})^2\right)$$

and notice that $\tilde{\psi}$ sends $\left(x^2 - \sqrt[3]{2}x + (\sqrt[3]{2})^2\right)$ to $\left(x^2 - \xi\sqrt[3]{2}x + (\xi\sqrt[3]{2})^2\right)$.

*If we wanted to, we could apply Lemma 5.57 again, applied to the isomorphism $\psi : \mathbb{Q}(\sqrt[3]{2}) \to \mathbb{Q}(\xi\sqrt[3]{2})$ and the polynomial $\left(x^2 - \sqrt[3]{2}x + (\sqrt[3]{2})^2\right)$. Lemma 5.57 would give us **new field extensions of** $\mathbb{Q}(\sqrt[3]{2})$ **and** $\mathbb{Q}(\xi\sqrt[3]{2})$**, and an isomorphism between these two extensions that agrees with** ψ **(and therefore equals the identity on** $\mathbb{Q}$ **).***

This logic is the key idea in the proof that splitting fields are unique.

Theorem 5.60. *Let $\varphi : F \to F'$ be an isomorphism, and f be **any** polynomial in $F[X]$. If*

1. *E is a **splitting field** for f in F, and*
2. *E' is a **splitting field** for $\tilde{\varphi}(f)$ in F',*

*then there is an **isomorphism** $\psi : E \to E'$ such that $\psi|_F = \varphi$.*

Proof. Let p be an irreducible factor of f of degree ≥ 2. Let $\alpha_1 \in E$ be a root of p and $\beta_1 \in E'$ be root of $\tilde{\varphi}(p)$. By Lemma 5.57, there is an **isomorphism** $F(\alpha_1) \to F'(\beta_1)$ that restricts to φ on F. If we repeat this process (until f no longer has any irreducible factors of degree ≥ 2, then we have a isomorphism $F(\alpha_1, \ldots, \alpha_k) \to F'(\beta_1, \ldots, \beta_k)$ that restricts to φ on F. Because f splits in $F(\alpha_1, \ldots, \alpha_k)$, and $\varphi(f)$ splits in $F'(\beta_1, \ldots, \beta_k)$, it follows that $E = F(\alpha_1, \ldots, \alpha_k)$ and $E' = F'(\beta_1, \ldots, \beta_k)$, which completes the proof.

Note in the above proof, we just care about irreducible factor of f. So no need to consider whether the f has repeated roots and whether β_i is the repeated root. If f splits over some field M, and the map is isomorphism, $\tilde{\varphi}(f)$ will also splits over the image field. Also note, in the above process, $\psi|_F = \varphi$ always, so we can apply Lemma 5.57 repeatedly. □

Corollary 5.61. *The number of such extensions is **at most** $[E : F]$ with the equality achieved when the splitting is **separated.***

Proof. The proof is quite long, please see the supp splittingfields.pdf. Roughly speaking, One root α in f will be sent to exactly one root α' in $\tilde{\varphi}(f)$. So the number of extensions depend on the number of different roots. When the roots are separated (see Section 5.3).

We use induction by a intermediate $F(\alpha)$ field similar to the proof in Lemma 5.86. □

Corollary 5.62. *Let F be a field, and $f \in F[X]$. Any two **splitting fields** for f over F are **isomorphic**. Moreover, this **isomorphism restricts to the identity isomorphism on** F.*

Proof. Apply the previous theorem to the case when $F' = F$ and φ is the identity map. □

Corollary 5.63. *Let F be a field, and $f \in F[X]$.* ***Any splitting field for** f **over** F **is algebraic.***

Proof. This is because the degree of the total extension will also have the same finite degree, and therefore be algebraic. □

Example 5.64. *The polynomial $x^2 - 2$ has two splitting fields over the rationals: $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}[X]/\left(x^2 - 2\right)$. They aren't the same. But they are isomorphic!*

This example is also an example for Remark 5.2.B (because the splitting fields are in the form of $F(\alpha)$).

5.3 Separability

If f is a polynomial in $F[X]$, we have seen above that we can construct a **splitting field** K **for** f **over** F, and K is such that all roots of f lie in it. The term "separable" comes from distinctness of the roots: they are separate in the sense that there are no multiple roots.

5.3.1 Definitions

Definition 5.65. ***(separable, inseparable)***

1. *An irreducible polynomial $f \in F[X]$ is **separable** if f has **no repeated roots** in a **splitting field.***
2. *An irreducible polynomial $f \in F[X]$ is **inseparable** if f has **repeated roots** in a **splitting field.***

*Note that if f is not necessarily irreducible, then we call f **separable** if each of its **irreducible factors is separable**.*

Definition 5.66. ***(separable over field, separable extension)***

1. *If E is an extension of F and $\alpha \in E$, then α is said to be **separable** over F if α is **algebraic over** F and its **minimal polynomial** $\mu_{\alpha,F}$ is a **separable polynomial** (in a splitting field).*
2. *If every element of E is **separable over** F, we say that E is a **separable extension** of F or that E/F is **separable.***

Lemma 5.67. ***Algebraic closure** C of F is a **separable extension** of C (not F).*

Proof. **Algebraic closure** C admit linear factors which each is irreducible and has an unique root. □

Example 5.68. *For example*

1. *$f(X) = (X-1)^2(X-2) \in \mathbb{Q}[X]$ is separable, since its irreducible factors $X-1$ and $X-2$ are separable.*
2. *In $\mathbf{R}[X]$, the polynomial $X^2 - X$ is separable since its roots are 0 and 1 and $X^3 - 2$ is separable since there are 3 different cube roots of 2 in the complex numbers.*
3. *In $\mathbb{F}_3[X]$ the polynomial $X^3 - 2$ is inseparable because $X^3 - 2 = (X+1)^3$ in $\mathbb{F}_3[X]$ so it has a triple root.*
4. *The real numbers $\sqrt{2}$ and $\sqrt{3}$ are both separable over $\mathbf{Q}$, as they have minimal polynomials $X^2 - 2$ and $X^2 - 3$ in $\mathbf{Q}[X]$, which are both separable.*

Definition 5.69. ***(separable closure)*** *If F is a field with **algebraic closure** C, then C contains a **smallest field containing all finite separable extensions** of F, called the **separable closure** of F.*

Corollary 5.70. ***Separable closure** of F is a separable extension of F and is a **subfield of the algebraic closure.***

Proof. From definition, it is correct. □

5.3.2 Criteria of Separable

In the definition, checking a polynomial is separable requires building a splitting field to check the roots are distinct. But in the following, we introduce criteria to test if a polynomial has multiple roots without having to work in a splitting field.

Definition 5.71. ***(derivative of polynomial)*** *The* ***derivative*** *of a polynomial*

$$f(X) = a_0 + a_1X + \cdots + a_nX^n \in F[X]$$

is defined to be

$$f'(X) = a_1 + 2a_2X + \cdots + na_nX^{n-1}.$$

Lemma 5.72. ***(criterion of multiple roots)*** *Consider*

$$f(X) = a_0 + a_1X + \cdots + a_nX^n \in F[X]$$

and its formal derivative

$$f'(X) = a_1 + 2a_2X + \cdots + na_nX^{n-1}.$$

Then f has a ***repeated root*** *(i.e.* ***inseparable****) in* ***a splitting field*** $\Longleftrightarrow$ *the degree of the* ***greatest common divisor*** *of f and f' is at least* 1.

Corollary 5.73. *For every field F, an* ***irreducible*** *polynomial in $F[X]$ is separable if and only if its* ***derivative is not 0*** *in $F[X]$.*

Proof. "$\Rightarrow$": Let us assume that f has a repeated root in its splitting field, say α. Then we can write

$$f(X) = (X-\alpha)^r h(X)$$

where $r \geq 2$ since we consider a repeated root. Now we compute the derivative of f :

$$f'(X) = r(X-\alpha)^{r-1}h(X) + (X-\alpha)^r h'(X)$$

and since $r-1 \geq 1$, we have that $(X-\alpha)$ is a factor of both f and f'.

"$\Leftarrow$": Conversely, let us assume that the greatest common divisor g of f and f' has degree at least 1, and let α be a root of g (in a splitting field). By definition of $g, X-\alpha$ is then a factor of both f and f'. We are left to prove that α is a repeated root of f. Indeed, if it were not the case, then $f(X)$ would be of the form $f(X) = (X-\alpha)h(X)$ where $h(\alpha) \neq 0$ and by computing the derivative, we would get (put $r=1$ in the above expression for f') $f'(\alpha) = h(\alpha) \neq 0$ which contradicts the fact that $X-\alpha$ is a factor of f'. □

Example 5.74. *For example*

1. *In $\mathbb{F}_3[X]$, let $f(X) = X^6 + X^5 + X^4 + 2X^3 + 2X^2 + X + 2$. Using Euclid's algorithm in $\mathbb{F}_3[X]$ on $f(X)$ and $f'(X)$*

$$\begin{aligned} f(X) &= f'(X)\left(2X^2+X\right) + \left(2X^2+2\right) \\ f'(X) &= \left(2X^2+2\right)\left(X^2+2X+2\right), \end{aligned}$$

   so $(f(X), f'(X)) = 2X^2+2$ (which is the same as X^2+1 up to scaling). The greatest common divisor is nonconstant, so $f(X)$ is inseparable. In fact, $f(X) = \left(X^2+1\right)^2\left(X^2+X+2\right)$. ***Notice we were able to detect that*** $f(X)$ ***has a repeated root before we gave its factorization.***

2. *In $\mathbf{Q}[X]$, let $f(X) = X^4 - 3X - 2$. Using Euclid's algorithm in $\mathbf{Q}[X]$ on $f(X)$ and $f'(X)$,*

$$\begin{aligned} f(X) &= f'(X)\cdot\left(\frac{1}{4}X\right) - \frac{9}{4}X - 2 \\ f'(X) &= \left(-\frac{9}{4}X - 2\right)\left(-\frac{16}{9}X^2 + \frac{128}{81}X - \frac{1024}{729}\right) - \frac{4235}{729}. \end{aligned}$$

   We have reached a remainder that is a nonzero constant, so $f(X)$ and $f'(X)$ are relatively prime. Therefore $f(X)$ is separable over $\mathbf{Q}$.

As a corollary of Corollary 5.73, we can exhibit two classes of separable polynomials.

Corollary 5.75. ***(two classes of separable polynomials)***

1. *Over a field of **characteristic zero**, every polynomial is **separable**.*

   *(Note c**haracteristic 0** field implies **infinite order** of the field.)*

2. *Over a field F of **characteristic** p (must be **prime**),*

$$\text{an } \textbf{irreducible} \text{ polynomial } f \text{ is } \textbf{inseparable}$$
$$\Updownarrow$$
$$f' \text{ is the } \textbf{zero polynomial} \text{ (equivalently } f \text{ is in } F\left[X^p\right]).$$

   *(Note **finite characteristic** field **does not** imply infinite order of the field nor finite, both are possible.)*

Proof. We can directly use Corollary 5.73, but we write it here more details since I omit the proof for Corollary 5.73. The idea behind them is clear.

1. Without loss of generality, consider f an irreducible polynomial in $F[X]$, where F is of characteristic zero. If f is a polynomial of degree n, then its derivative f' is of degree less than n, and it **cannot possibly be the zero polynomial**. Since f is **irreducible**, the greatest common divisor of f and f' is either 1 or f, but it cannot be f since f' is of **smaller** degree. Thus it is 1 , and f is separable by the above proposition.

2. We now consider the case where F is of characteristic p. As above, we take f an irreducible polynomial of degree n in $F[X]$ and compute its derivative f'.

$\Downarrow$: If f' is **non-zero**, we can use the same argument.

$\Uparrow$: But f' could also be zero, in which case the **greatest common divisor** of f and f' is actually f, and by Lemma 5.72, f has a multiple root and is then not separable. Also note that $f' = 0$ means that $f \in F\left[X^p\right]$ since we work in characteristic p. □

Example 5.76. ***(infinite field with characteristic zero)*** *Polynomials over $\mathbb{R}[X]$ and $\mathbb{Q}[X]$ are separable because $\mathbb{R}[X]$ and $\mathbb{Q}[X]$ have characteristic 0.*

Corollary 5.77. ***Finite field extensions** (i.e. **number fields**) of $\mathbb{Q}$ are **separable extensions**.*

Proof. Because all (including minimal) polynomials are separable in $\mathbb{Q}$. The condition finite field extension is to guarantee the existence of minimal polynomials Theorem 5.25. □

Another class of separable polynomials are polynomials over **finite fields**, but this asks a little bit more work.

Lemma 5.78. *Let F be a **finite field of characteristic** p. Consider the map*

$$f : F \to F, f(\alpha) = \alpha^p.$$

*Then f is an **automorphism** (called the **Frobenius Automorphism**). In particular, we have for all $\alpha \in F$ that*

$$\alpha = \beta^p$$

for some $\beta \in F$.

Proof. We have that f is a ring automorphism since

$$\begin{aligned} f(1) &= 1 \\ f(\alpha+\beta) &= (\alpha+\beta)^p = \alpha^p + \beta^p = f(\alpha) + f(\beta) \\ f(\alpha\beta) &= (\alpha\beta)^p = \alpha^p\beta^p = f(\alpha)f(\beta) \end{aligned}$$

The second set of equalities uses the binomial expansion modulo p. Now f is a **monomorphism** since F is a field (Lemma 5.11), and an injective map from a finite set to itself is necessarily surjective Lemma 1.9. □

Theorem 5.79. *Every polynomial is* ***separable*** *over a* ***finite field*** *F (of prime characteristic).*

Proof. Suppose that f is an **irreducible polynomial** which, by contradiction, has **multiple roots** in a splitting field. Using the criterion Corollary 5.75, $f(X)$ must be in $F\left[X^{p}\right]$, namely

$$f(X)=a_{0}+a_{1} X^{p}+\cdots+a_{n} X^{n p}, a_{i} \in F .$$

Using the bijectivity of the **Frobenius automorphism** in Lemma 5.78, we can write $a_{i}=b_{i}^{p}$, yielding

$$\left(b_{0}+b_{1} X+\cdots+b_{n} X^{n}\right)^{p}=b_{0}^{p}+b_{1}^{p} X^{p}+\cdots+b_{n}^{p} X^{n p}=f(X)$$

which contradicts the **irreducibility** of f. □

Corollary 5.80. *Let F and L be fields. If $\sigma: F \rightarrow L$ is a* ***field embedding****, then a polynomial $f(X) \in F[X]$ is* ***separable*** *if and only if $(\sigma f)[X] \in L[X]$ is* ***separable****.*

Remark 5.81. *So we have* ***field embedding does not affect separable*** *.*

Proof. $\Rightarrow$: Assume $f(X)$ is separable, according to Lemma 5.72, we can write

$$f(X) u(X)+f^{\prime}(X) v(X)=1$$

for some $u(X)$ and $v(X)$ in $F[X]$. Applying σ to coefficients is a ring embedding $F[X] \rightarrow L[X]$ and $\sigma\left(f^{\prime}\right)=(\sigma f)^{\prime}$, so

$$(\sigma f)(X)(\sigma u)(X)+(\sigma f)^{\prime}(X)(\sigma v)(X)=1$$

Therefore $(\sigma f)(X)$ and its derivative are relatively prime in $L[X]$, so $(\sigma f)(X)$ is separable.

$\Leftarrow$: Now assume $f(X)$ is inseparable, so some nonconstant $d(X)$ in $F[X]$ divides $f(X)$ and $f^{\prime}(X)$ in $F[X]$. Then $(\sigma d)(X)$ is nonconstant and divides $(\sigma f)(X)$ and $\sigma\left(f^{\prime}(X)\right)=(\sigma f)^{\prime}(X)$ in $L[X]$, so $(\sigma f)(X)$ and its derivative are not relatively prime and thus $(\sigma f)(X)$ is inseparable. □

Corollary 5.82. *We have*

1. *If $f(X) \in F[X]$ is* ***separable*** *and $F \subseteq L$ then every factor of $f(X)$ in $L[X]$ is also* ***separable****.*

2. *An element in an extension of L/F that is* ***separable*** *over F is also* ***separable*** *over L.*

Remark 5.3.D. *Note but for L/F,* ***separable*** *over L does* ***not*** *implies* ***separable*** *over F. For example in Lemma 5.67 the* ***Algebraic closure*** *C of F is a* ***separable extension*** *of C (not F).*

Proof. Let $g(X)$ be a factor of $f(X)$, say $f(X)=g(X) h(X)$ in $L[X]$. Since $f(X)$ is separable we can write $1=f(X) u(X)+f^{\prime}(X) v(X)$ for some polynomials $u(X)$ and $v(X)$ in $F[X]$. Then

$$\begin{aligned} 1 & =(g(X) h(X)) u(X)+\left(g(X) h^{\prime}(X)+g^{\prime}(X) h(X)\right) v(X) \\ & =g(X)\left(h(X) u(X)+h^{\prime}(X) v(X)\right)+g^{\prime}(X)(h(X) v(X))\end{aligned}$$

The last expression shows a polynomial-linear combination of $g(X)$ and $g^{\prime}(X)$ equals 1 , so $g(X)$ is separable.

Suppose α is in an extension of L and it is separable over F. Since its minimal polynomial $\mu_{\alpha, L}$ in $L[X]$ divides its minimal polynomial $\mu_{\alpha, F}$ in $F[X]$ (Lemma 5.17), separability of α over F then implies separability of α over L because repeated roots of $\mu_{\alpha, L}$ would lead to repeated roots of $\mu_{\alpha, F}$. □

Here is how separability behaves in a **tower of extensions** analog to Theorem 5.43:

Lemma 5.83. *We have*

1. *If $E/K/F$ and E is* ***separable*** *over F, then K is* ***separable*** *over F and E is* ***separable*** *over K.*

2. ***(transitivity of separable extensions):*** *If K/F and E/K are* ***separable****, then E/F is* ***separable****.*

Proof. 1. "K/F is separable": Since K is a subfield of E, every element $\beta \in K$ belongs to E, and every element of E is separable over F by assumption.

"E/K is separable": directly from 2. in Corollary 5.82.

2. I omit the proof here. Just need to check the **separable** since transitivity of algebraic extensions has been shown in Theorem 5.43. The proof need theorem of the Primitive Element (see Theorem 5.92). Please see supp separable1.pdf for details. We only need prove the number of embedding exactly gets the maximum so the separability is guaranteed. □

It is less easy to construct inseparable extensions, but here is a classical example.

Example 5.84. ***(classical example)*** *Let $\mathbb{F}_p$ denote the **finite field of integers modulo** p. Consider the field $F = \mathbb{F}_p(t)$ of **rational functions** in t with coefficients in the finite field with p elements $\mathbb{F}_p$. We get a field extension of E/F by adjoining to F a root α (sometimes, we denote it as $\sqrt[p]{t}$) of the polynomial $X^p - t$. We can check that $X^p - t$ is irreducible over $F = \mathbb{F}_p[t]$ using (note t is prime in F).*

*The extension E/F is **inseparable** since*

$$X^p - t = X^p - (\alpha)^p = (X - \alpha)^p$$

*which has **multiple roots.***

5.3.3 Primitive Element Theorem

$$\begin{tikzcd}
 & \mathbf{C} \arrow[dd, dash] \\
\mathbf{Q}(\sqrt{2}) \arrow[ur, "2 \text{ ways}"] \arrow[dr, "2 \text{ ways}"] \arrow[dd, dash, "2"'] & \\
 & \mathbf{R} \\
\mathbf{Q} \arrow[ur, dash] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \mathbf{C} \arrow[dd, dash] \\
\mathbf{Q}(\sqrt[3]{2}) \arrow[ur, "3 \text{ ways}"] \arrow[dr, "1 \text{ way}"] \arrow[dd, dash, "3"'] & \\
 & \mathbf{R} \\
\mathbf{Q} \arrow[ur, dash] &
\end{tikzcd}$$

Example 5.85. *To embed the field $\mathbf{Q}(\sqrt{2})$ into $\mathbf{R}$, there are **two ways** this can be done: send $\sqrt{2}$ to itself or send it to $-\sqrt{2}$. That there are two embeddings is related to the fact that $X^2 - 2$ has **two different roots** in $\mathbf{R}$. Similarly, there are **two embeddings** of $\mathbf{Q}(\sqrt{2})$ into $\mathbf{C}$.*

*If we try to embed $\mathbf{Q}(\sqrt[3]{2})$ into $\mathbf{R}$, there is **only one way** to do this since there is only one real cube root of 2. Enlarging our target field to $\mathbf{C}$ provides us with **3 different cube roots of 2** (one is real, two are non-real), so $\mathbf{Q}(\sqrt[3]{2})$ has **3 different embeddings** into the complex numbers (determined by sending $\sqrt[3]{2}$ to each of the 3 cube roots of 2 in $\mathbf{C}$).*

*The number of embeddings $\mathbf{Q}(\sqrt[3]{2}) \to \mathbf{C}$ is **3**, but we had to make the **target field large enough** (target field $\mathbf{R}$ was too small). The number of embeddings of $\mathbf{Q}(\sqrt{2})$ and $\mathbf{Q}(\sqrt[3]{2})$ into $\mathbf{R}$ and $\mathbf{C}$ is related to the number of different roots of $X^2 - 2$ and $X^3 - 2$ in $\mathbf{R}$ and $\mathbf{C}$. That the number of roots equals the degree of each polynomial when they **split completely** is related to the **polynomials being separable**.*

We start with a lemma.

Lemma 5.86. *We have*

1. *Let $\sigma : E \to E$ be an F**-monomorphism** and assume that $f \in F[X]$ **splits over** E. Then σ **permutes** the roots of f, namely, **if** α **is a root of** f **in** E **then so is** $\sigma(\alpha)$.*
2. *Similarly, Let $\sigma : E \to E'$ be an **field monomorphism**, **if** α **is a root of** f **in** E **then so is** $\sigma(\alpha)$ **in** E'.*

Proof. Write $f(X)$ as

$$f(X) = b_0 + b_1 X + \cdots + b_n X^n, b_i \in F.$$

If α is a root of f in E, then

$$f(\alpha) = b_0 + b_1 \alpha + \cdots + b_n \alpha^n = 0.$$

Apply σ to the above equation, and use that σ is a field homomorphism that fixes F to get

$$b_0 + b_1 \sigma(\alpha) + \cdots + b_n \sigma(\alpha)^n = 0,$$

showing that $\sigma(\alpha)$ is a root.

□

Let E/F be a separable extension of F and let C be an **algebraic closure** of E. We next count the number of F**-monomorphisms** of E into C (**number of embeddings** of E in C that **fix** F).

Theorem 5.87. *Let E/F be a* ***finite separable extension*** *of degree n, and let σ be an embedding of F into an* ***algebraic closure*** *C.*

Then

1. *σ extends to exactly n* ***embeddings*** *of E in C.*
2. *Namely, there are exactly n* ***embeddings*** *τ of E into C, such that the restriction $\tau|_F$ of τ to F coincides with σ.*
3. *In particular, taking σ to be the* ***identity*** *on F, there are exactly n F-monomorphisms of E into C.*

$$\begin{tikzcd}
& C \\
E \arrow[ur, dotted, "\tau"] & \\
F \arrow[u] \arrow[uur, "\sigma"'] &
\end{tikzcd}$$

Proof. We do a proof by induction. If $n = 1$, then $E = F$ and σ extends to exactly 1 embedding, namely itself.

We now assume that $n > 1$ and choose $\alpha \in E, \alpha \notin F$. Let $f = \mu_{\alpha,F}$ be the **minimal polynomial** of α over F of degree say r. It is **irreducible** and **separable** (E/F is separable by assumption).

In order to use the induction hypothesis, we need to split the field extension E/F, which we do by considering the field extension $F(\alpha)$, which satisfies

$$E/F(\alpha)/F, [E : F(\alpha)] = n/r, [F(\alpha) : F] = r.$$

Note in the proof of Theorem 5.46, we have mentioned $F(\alpha)$ is an **algebraic extension** because any finite extension is algebraic extension Theorem 5.25 and $F(\alpha)$ is finite extension (Theorem 5.23). From Lemma 5.83, we further know $E/F(\alpha)$ and $F(\alpha)/F$ are **separable**.

Let σ be an embedding of F into C, and define the polynomial $g = \sigma(f) \in \sigma(F)[X]$, where σ is applied on all the coefficients of f. The polynomial g inherits the property of being **irreducible** and **separable** from f (in the space of $\sigma(F)[X]$ not $C[X]$).

Note, we must map α to one root for g as mentioned in Corollary 4.4.A. Let β denotes a root of g. We can thus define a **unique isomorphism** π that maps α to β.

$$\begin{gathered} F(\alpha) \to (\sigma(F))(\beta) \\ b_0 + b_1\alpha + \ldots + b_r\alpha^r \mapsto \sigma(b_0) + \sigma(b_1)\beta + \ldots + \sigma(b_r)\beta^r \end{gathered}$$

and restricted to F it **indeed coincides with σ.** (uniqueness is easy to see, but why there exist such a isomorphism, or say why the above is a isomorphism. Please recall the basis in Theorem 5.23 and recall the same technique used in Corollary 4.4.A)

This isomorphism is defined by the choice of β, and **there are exactly r choices for it, corresponding to the r roots of** g (note that the **separability** of g is crucial).

For each of these r isomorphisms, using the induction hypothesis on $[E : F(\alpha)] = n/r < n$, we can extend them to exactly n/r embeddings of E into C. This gives us a total of $n/r \cdot r$ distinct embeddings of E into C extending σ. □

Corollary 5.88. *If E/F is not separable, we have **less than n embeddings of E in C.***

Proof. We go through the same processing as in the prove. Then we can select one $\alpha \in E$ whose minimal polynomial has multiple roots and then the number of possible embedding is less that r. Consequently, we have the strictly less than n embedding. □

Corollary 5.89. *Let L/F be a finite extension and write $L = F(\alpha_1, \ldots, \alpha_r)$. Then*

$$L/F \textbf{ is separable} \iff \textbf{each } \alpha_i \textbf{ is separable over } F.$$

Remark 5.90. *The usefulness of Corollary 5.89 is that it gives a practical way to check a finite extension L/F is separable: rather than show every element of L is separable over F **it suffices to show there is a set of field generators for L/F that are each separable over F.***

Proof. See the supp separable1.pdf. □

Corollary 5.91. *If $f(X) \in F[X]$ is **separable** then a splitting field for f over F is separable over F.*

Proof. Let L/F be a splitting field for f over F. Then $L = F(\gamma_1, \ldots, \gamma_n)$ where the γ_i 's are all roots of $f(X)$. Therefore the γ_i 's are separable over F, so L/F is a separable extension. □

Theorem 5.92. ***(Theorem of the Primitive Element)** If E/F is a **finite separable** extension, then*

$$E = F(\gamma)$$

*for some $\gamma \in E$. We say that γ is a **primitive element** of E over F.*

Proof. If F is a finite field then every finite extension L is a finite field. Therefore $L^\times$is cyclic by Theorem 4.177. Letting γ be a generator of $L^\times$, we have $L^\times = \langle\gamma\rangle$, so $L = F(\gamma)$.

Now consider the case when F is infinite. A finite separable extension of F has the form $F(\alpha_1, \ldots, \alpha_r)$ where each α_i is separable over F. It suffices by induction on the number of field generators to show when $F(\alpha, \beta)/F$ is separable that $F(\alpha, \beta) = F(\gamma)$ for some γ.

Let $L = F(\alpha, \beta)$ and $n = [L : F]$. Recall that a F-homomorphism is a homomorphism of extensions of F that fixes F pointwise. Since L/F is separable, Theorem 5.87 tells us there is a field extension C/F such that the number of F-homomorphisms $L \to C$ is n. Pick $c \in F$. If $F(\alpha + c\beta) \neq L$ then $[F(\alpha + c\beta) : F] < [L : F]$. We will show **there are only finitely many such c, so on account of F being infinite there is a $c \in F$ such that $F(\alpha + c\beta) = L$.**

The degree $[F(\alpha+c\beta) : F]$ is an upper bound on the number of F-homomorphisms $F(\alpha+c\beta) \to C$. Since there are n F-homomorphisms $L \to C$, if $[F(\alpha + c\beta) : F] < [L : F] = n$ then there are different F-homomorphisms $L \to C$, say σ and τ, which are equal on $F(\alpha + c\beta)$. Therefore $\sigma \neq \tau$ on L but $\sigma(\alpha + c\beta) = \tau(\alpha + c\beta)$. Then $\sigma(\alpha) + c\sigma(\beta) = \tau(\alpha) + c\tau(\beta)$. If $\sigma(\beta) = \tau(\beta)$ then we get $\sigma(\alpha) = \tau(\alpha)$, so $\sigma = \tau$ as functions on $F(\alpha, \beta) = L$, which isn't true. Hence $\sigma(\beta) \neq \tau(\beta)$, so we can solve for c :

$$c = \frac{\tau(\alpha) - \sigma(\alpha)}{\sigma(\beta) - \tau(\beta)}.$$

There are only finitely many σ and τ, so only finitely many such c. □

Corollary 5.93. *When F has **characteristic** 0, all of its **finite extensions are separable** according to Corollary 5.75, so the primitive element theorem says **every finite extension field of F has the form $F(\gamma)$ for some γ.***

Example 5.94. *The field $\mathbf{Q}(\sqrt{2}, \sqrt{3})$ is separable over $\mathbf{Q}$ and it equals $\mathbf{Q}(\sqrt{2} + \sqrt{3})$.*

5.3.4 Simple Extension

Definition 5.95. ***(simple extension)*** *A* ***simple extension*** *is a field extension which is generated by the* ***adjoining of a single element.***

Lemma 5.96. *Thus the primitive element Theorem above provides a characterization of the finite extensions which are simple.*

Corollary 5.97. *Number fields are simple extensions.*

5.4 Normality

1. Algebraic and separable field extensions are **transitive**
2. We now introduce a third property, which is **not transitive,** the one of being normal.

Definition 5.98. ***(normal, conjugate root)***

1. *An* ***algebraic extension*** *E/F is* ***normal*** *if every* ***irreducible polynomial*** *over F that has at least one root in E* ***splits*** *over E.*
2. *If we call the other roots of this polynomial the* ***conjugates*** *of α, we can rephrase the definition by saying that* ***if*** *$\alpha \in E$****, then all conjugates of*** *α* ***over*** *F* ***are in*** *E.*

Example 5.99. *Consider the field extension $E = \mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$. The roots of the irreducible polynomial $f(X) = X^3 - 2$ are*

$$\sqrt[3]{2}, \zeta_3 \sqrt[3]{2}, \zeta_3^2 \sqrt[3]{2}$$

where ζ_3 is a primitive 3 rd root of unity (for example $\zeta_3 = e^{2\pi i/3}$). Thus E is ***not*** *a normal extension.*

• We can give another characterization in terms of **monomorphisms of E.**

Theorem 5.100. *The* ***finite extension*** *E/F is* ***normal*** *if and only if every F****-monomorphism*** *of E into an* ***algebraic closure*** *C is actually an F****-automorphism*** *of E. (****Finite*** *could be replaced by* ***algebraic****, which we will not prove).*

Proof. $\Rightarrow$: If E/F is normal, then an F-monomorphism τ of E into C must map each element of E to one of its conjugates (as is the case in the proof of Lemma 5.86). Note all elements have conjugates for some polynomial in $F[X]$. Thus $\tau(E) \subseteq E$, but $\tau(E)$ is an isomorphic copy of E and thus has the **same degree (i.e. dimension)** as E and $E = \tau(E)$, showing that τ is indeed an F-automorphism of E.

$\Leftarrow$: Consider $\alpha \in E$ and let β be a conjugate of α over F. There exists an F-monomorphism of E into C that carries α to β (the construction is given in the proof of Theorem 5.87). If all such embeddings are F-automorphisms of E, that means β must be in E, and we conclude that E/F is normal. □

• Here is another characterization of normal extensions in terms of **splitting fields.**

Theorem 5.101. *The finite extension E/F is* ***normal*** *if and only if E is a* ***splitting field*** *for some polynomial f in $F[X]$.*

Proof. Let E/F be a finite normal extension of degree n, and let $\alpha_1, \dots, \alpha_n$ be a **basis** for E over F. Consider for each α_i its minimal polynomial f_i over F. By definition of normal extension, since f_i has a root in E, then f_i splits over E, and so does the polynomial

$$f = f_1 \cdots f_n.$$

To prove that E is a splitting field, we are left to prove it is the smallest field over which f splits. This is here that we understand why we take such an f. If f were to split over a subfield K, that is K such that

$$F \subset K \subset E$$

then each $\alpha_i \in K$, and $K = E$ (this is a conclusion we cannot reach if we take for f only one f_i or a subset of them). This proves that E is a splitting field for f over F.

Conversely, let E be a splitting field for some f over F, whose roots are denoted by $\alpha_1, \ldots, \alpha_n$. Let τ be an F-**monomorphism** of E into an algebraic closure, that is τ takes each α_i into another root of f.

Since E is a splitting field for f, we have

$$F\left(\alpha_1, \ldots, \alpha_n\right) = E$$

and $\tau(E) \subset E$ (to see this, note $\alpha_1, \ldots, \alpha_r$ are the basis, so every elements can be linear represented using the roots). Thus since E and $\tau(E)$ have same dimension, we get that

$$\tau(E) = E$$

and τ is actually an **automorphism** of E, and by the above theorem, we conclude the E/F is normal. □

- As a corollary, we see how a subextension inherits the property of normality.

Corollary 5.102. *Let $E/K/F$ be a finite extension ($[E : F] < \infty$). If E/F is normal, so is E/K.*

Proof. Since E/F is normal, E is a splitting field for some polynomial $f \in F[X]$, that is E is generated over F by the roots of f. Since $f \in F[X] \subset K[X]$, f can also be seen as a polynomial in $K[X]$ and E is generated over K by the roots of f, and again by the above theorem, E/K is normal. □

There is no reason for an arbitrary field extension E/F to be normal. However, if E/F is finite (or more generally algebraic) one can always embed it in normal extension.

Definition 5.103. ***(normal closure)*** *Let E/F be an* ***algebraic extension****. The* ***normal closure*** *of E/F is an* ***extension field*** *N* ***of*** *E such that N/E is* ***normal*** *and N is minimal with this property.*

Example 5.104. *If E/F is finite, we can see it as follows: E is finitely generated over F, so it can be written as $E = F\left(\alpha_1, \ldots, \alpha_n\right)$. What's the normal closure of E/F?*

Let now K be a normal extension of F that contains E :

$$K/E/F.$$

Since K is normal, it must contain not only all the α_i but also all their conjugates. Let f_i be the minimal polynomial of $\alpha_i, i = 1, \ldots, n$. Then we can rephrase the last statement and say that K must contain all the roots of f_i, $i = 1, \ldots, n$. Consider the polynomial

$$f = f_1 \cdots f_n$$

Let N be ***the splitting field for*** *f* ***over*** *F (which is unique up to isomorphism Section 5.2.3). Then K must contain N. But N/F is normal from Theorem 5.101, so N must be the smallest normal extension of F that contains E. Thus N is a* ***normal closure of*** *E* ***over*** *F.*

5.5 Summary of Separable and Normal

Remark 5.105. *Note, in Lemma 5.67 (note not F in the lemma), we know*

1. Let us consider a field K and let us say we fix an algebraic closure K^{alg} of it.
   (a) Now, K^{alg} might contain elements that are not separable (over K), that is their **minimal polynomial is not a separable polynomial** (i.e., it has roots of multiplicity greater 1 in K^{alg} or equivalently it is not co-prime with its derivative). This is what Lemma 5.67 (note not F in the lemma) and Lemma 5.72 are saying.
   (b) Non-separable extensions and elements are not so nice in some ways (in particular sometimes we only want **Galois** i.e., normal and separable). So one might consider **only considering all separable (over K) elements** in K^{alg}. The collection of all these forms again a field and is called the **separable closure of K.**

- This extension is a **Galois extension** (See Corollary 5.107; In fact it is the **maximal Galois extension of K.**)

(c) From Corollary 5.75 and Theorem 5.79, we have for some fields such as **any field of characteristic 0** or for **finite fields**, **the separable closure equals the algebraic closure**. The point being that this fields are perfect fields and thus every **algebraic extension is separable.**

2. **Explanation of normal:** An extension E/K is called normal **if each irreducible polynomial in $K[X]$ that has a zero in E can be decomposed into linear factors in E.** Or put differently, if E contains one of the zeros of a polynomial P it contains all the zeros of P.

Lemma 5.106. *Let E be an algebraic extension of a field F. Show that the set of all elements in E that are separable over F forms a subfield of E, the **separable closure of** F **in** E.*

Proof. From Corollary 5.89, we know if α, β are both separable over F, then $\alpha \pm \beta$, $\alpha\beta$, and α/β. □

Corollary 5.107. *The **separable closure** of F in its **algebraic closure** is normal.*

Proof. We only need to prove the conjugate roots are all in the separable closure F^{sep}. For element α in F^{sep}, its minimal polynomial P is separable over F. But then we have all the conjugate roots are in F^{sep} since they all have a minimal polynomial (factors of P). □

6 Galois Theory

6.1 Galois Group and Fixed Fields

6.1.1 Galois Group

Definition 6.1. ***(Galois Extension)*** *If E/F is **normal** and **separable**, it is said to be a **Galois extension**, or alternatively, we say that E is **Galois over** F.*

Remark 6.2. ***(Explanation)***

1. ***Normal** tell us the minimal polynomial of any element in E, say f_α for $\alpha \in E$, splitting over E.*

2. ***Separable** tell us the minimal polynomial of any element in E, say f_α for $\alpha \in E$, does not have repeated root over a splitting field.*

3. *So together, we need **Galois extension** to have the minimal polynomial of any element in E, say f_α for $\alpha \in E$, splitting over E **with no repeat linear factors** $X - a_i$**,** $a_i \in E$.*

Lemma 6.3. *Take E/F a **Galois extension** of degree n.*

1. *Since it is **separable** of degree n, we know that there are exactly n F**-monomorphisms** of E into an algebraic closure C (See Theorem 5.87).*

2. *But E/F being also **normal**, every F**-monomorphisms** into C is actually an F**-automorphism** of E (See Theorem 5.100).*

*Thus **there are exactly** $n = [E : F]$ F**-automorphisms of** E.*

We can define the notion of a Galois group for an **arbitrary field extension**.

Definition 6.4. ***(Galois group)*** *If E/F is a field extension, the Galois group of E/F, denoted by $\text{Gal}(E/F)$, is the set of F**-automorphisms** of E. It forms a group under the composition of functions.*

Corollary 6.5. *In a **Galois extension**, the **order of the Galois group** is actually the **degree of the field extension.***

Example 6.6. *If $E = \mathbb{Q}(\sqrt[3]{2})$, then $\text{Gal}(E/\mathbb{Q}) = \{1\}$, that is the identity on E. But note $[E : F] = 3$ since $\sqrt[3]{2}$ is the root of $X^3 - 2 = 0$.*

Remark 6.7. *The above example illustrates the fact that though one can always define a Galois group, we need the **extension to be actually Galois** to say that the **order of the Galois group is actually the degree of the field extension.***

6.1.2 Fixed Fields

Definition 6.8. ***(fixed field; fixing group)***

1. *Let $G = \mathrm{Gal}(E/F)$ be the **Galois group** of the extension E/F. If H is a **subgroup of** G, the **fixed field** $\mathcal{F}(H)$ of H, is the set of elements fixed by every **automorphism in** H, that is*
$$F \subseteq \mathcal{F}(H) = \{x \in E, \sigma(x) = x \textit{ for all } \sigma \in H\} \subseteq E$$
2. *If K is an **intermediate field**: $E/K/F$. Define*
$$\mathcal{G}(K) = \mathrm{Gal}(E/K) = \{\sigma \in G, \sigma(x) = x \textit{ for all } x \in K\} \subseteq \mathrm{Gal}(E/F)$$
*It is the **group fixing** intermediate K.*

Galois theory has much to do with studying the relations between **fixed fields and fixing groups.**

Lemma 6.9. *Let E/F be a **finite Galois extension** with Galois group $G = \mathrm{Gal}(E/F)$. Then*

1. *The fixed field of G is F.*
2. *If H is a **proper subgroup** of G, then the fixed field $\mathcal{F}(H)$ of H **properly** contains F.*

Proof. 1. Let F_0 be the fixed field of G (and we have the field extensions $E/F_0/F$). We want to prove that $F_0 = F$.

We first note that if σ is an F**-automorphism** of E (that is σ is in G), then by definition of F_0, σ fixes everything in F_0, meaning that σ is an F_0**-automorphism**. Thus the F-automorphisms in the group G coincide with the F_0-automorphisms in the group G.

Now we further have that E/F_0 **is Galois**: indeed, we have $E/F_0/F$ with E/F Galois thus normal and separable, and E/F_0 inherits both properties (See Corollary 5.102 and Lemma 5.83).

We now look at the degrees of the extensions considered:
$$|\mathrm{Gal}\,(E/F_0)| = [E : F_0]\,, |\,\mathrm{Gal}(E/F)| = [E : F],$$
since both are Galois. Furthermore from the above, the number of F- and F_0-automorphisms in G coincide:
$$|\mathrm{Gal}\,(E/F_0)| = |\,\mathrm{Gal}(E/F)|$$
showing that
$$[E : F_0] = [E : F]$$
and by multiplicativity of the degrees
$$[E : F] = [E : F_0]\,[F_0 : F] \Rightarrow [F_0 : F] = 1$$
and $F = F_0$

2. In order to prove that $F \subsetneq \mathcal{F}(H)$, let us assume by contradiction that $F = \mathcal{F}(H)$

Since we consider a finite Galois extension, we can invoke the **Theorem of the Primitive Element Theorem 5.92** and claim that
$$E = F(\alpha), \alpha \in E.$$
Consider the polynomial
$$f(X) = \prod_{\sigma \in H} (X - \sigma(\alpha)) \in E[X].$$
It is a priori in $E[X]$ (because of F-automorphism of E), but we will prove now that it is actually in $F[X]$. Since by contradiction we are assuming that $F = \mathcal{F}(H)$, it is enough to proof that $f(X)$ is fixed by H. Indeed, take $\tau \in H$, then
$$\prod_{\sigma \in H} (X - \tau\sigma(\alpha)) = \prod_{\sigma \in H} (X - \sigma(\alpha))$$

since $\tau\sigma$ ranges over all H as does σ. Thus $f(X) \in F[X]$ and $f(\alpha) = 0$ (σ must be the identity once while ranging through H). Now on the one hand, we have

$$\deg f = |H| < |G| = [E : F]$$

since we assume that H is proper and E/F is Galois. On the other hand,

$$\deg f \geq [F(\alpha) : F] = [E : F]$$

since f is a multiple of the minimal polynomial of α over F (equality holds if f is the minimal polynomial of α over F), and $E = F(\alpha)$. We cannot possibly have $\deg f < [E : F]$ and $\deg f \geq [E : F]$ at the same time, which is a contradiction and concludes the proof. □

6.2 The Fundamental Theorem of Galois theory

The most significant discovery of Galois is that under some hypotheses, there is a **one-to-one correspondence** between

1. **subgroups** of the Galois group $\mathrm{Gal}(E/F)$
2. **subfields** M of E such that $F \subseteq M \subseteq E$.

The correspondence goes as follows:

- To each **intermediate subfield** M, associate the group $\mathrm{Gal}(E/M)$ of all M-automorphisms of E :

$$\begin{aligned} \mathcal{G} : \{ \text{ intermediate fields } \} &\to \{ \text{ subgroups of } \mathrm{Gal}(E/F)\} \\ M &\mapsto \mathcal{G}(M) = \mathrm{Gal}(E/M) \end{aligned}$$

- To each **subgroup** H of $\mathrm{Gal}(E/F)$, associate the **fixed subfield** $\mathcal{F}(H)$:

$$\begin{aligned} \mathcal{F} : \{ \text{ subgroups of } \mathrm{Gal}(E/F)\} &\to \{ \text{ intermediate fields } \} \\ H &\mapsto \mathcal{F}(H). \end{aligned}$$

We will prove below that, under the right hypotheses, we actually have a **bijection** (namely $\mathcal{G}$ is the inverse of $\mathcal{F}$). Let us start with an example.

Example 6.10. *Consider the field extension $E = \mathbb{Q}(i, \sqrt{5})/\mathbb{Q}$. It has four $\mathbb{Q}$ automorphisms, given by (it is enough to describe their actions on i and $\sqrt{5}$):*

$$\begin{array}{rcc} \sigma_1 : & i \mapsto i, & \sqrt{5} \mapsto \sqrt{5} \\ \sigma_2 : & i \mapsto -i, & \sqrt{5} \mapsto \sqrt{5} \\ \sigma_3 : & i \mapsto i, & \sqrt{5} \mapsto -\sqrt{5} \\ \sigma_4 : & i \mapsto -i, & \sqrt{5} \mapsto -\sqrt{5} \end{array}$$

thus

$$\mathrm{Gal}(E/\mathbb{Q}) = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$$

The proper subgroups of $\mathrm{Gal}(E/\mathbb{Q})$ *are*

$$\{\sigma_1\}, \{\sigma_1, \sigma_2\}, \{\sigma_1, \sigma_3\}, \{\sigma_1, \sigma_4\}$$

and their corresponding subfields are

$$E, \mathbb{Q}(\sqrt{5}), \mathbb{Q}(i), \mathbb{Q}(i\sqrt{5})$$

Theorem 6.11. *Let E/F be a **finite Galois extension** with Galois group G.*

1. *The map $\mathcal{F}$ is a **bijection** from **subgroups** to **intermediate fields**, with inverse $\mathcal{G}$.*
2. *Consider the intermediate field $K = \mathcal{F}(H)$ which is fixed by H, and $\sigma \in G$. Then the intermediate field*

$$\sigma K = \{\sigma(x), x \in K\}$$

is fixed by $\sigma H \sigma^{-1}$, namely $\sigma K = \mathcal{F}\left(\sigma H \sigma^{-1}\right)$.

Proof. 1. We first consider the composition of maps

$$H \to \mathcal{F}(H) \to \mathcal{G}\mathcal{F}(H)$$

We need to prove that $\mathcal{G}\mathcal{F}(H) = H$. Take σ in H, then σ fixes $\mathcal{F}(H)$ by definition and $\sigma \in \operatorname{Gal}(E/\mathcal{F}(H)) = \mathcal{G}(\mathcal{F}(H))$, showing that

$$H \subseteq \mathcal{G}\mathcal{F}(H)$$

To prove equality, we need to rule out the strict inclusion. If H **were a proper subgroup** of $\mathcal{G}(\mathcal{F}(H))$, by Lemma 6.9 the fixed field $\mathcal{F}(H)$ of H should properly contain the fixed field of $\mathcal{G}\mathcal{F}(H)$ which is $\mathcal{F}(H)$ itself, a contradiction, showing that

$$H = \mathcal{G}\mathcal{F}(H).$$

Now consider the reverse composition of maps

$$K \to \mathcal{G}(K) \to \mathcal{F}\mathcal{G}(K).$$

This time we need to prove that $K = \mathcal{F}\mathcal{G}(K)$. But

$$\mathcal{F}\mathcal{G}(K) = \text{ fixed field by } \operatorname{Gal}(E/K)$$

which is exactly K by Lemma 6.9.

2. It is enough to compute $\mathcal{F}\left(\sigma H \sigma^{-1}\right)$ and show that it is actually equal to $\sigma K = \sigma \mathcal{F}(H)$.

$$\begin{aligned}
\mathcal{F}\left(\sigma H \sigma^{-1}\right) &= \left\{x \in E, \sigma\tau\sigma^{-1}(x) = x \text{ for all } \tau \in H\right\} \\
&= \left\{x \in E, \tau\sigma^{-1}(x) = \sigma^{-1}(x) \text{ for all } \tau \in H\right\} \\
&= \left\{x \in E, \sigma^{-1}(x) \in \mathcal{F}(H)\right\} \\
&= \{x \in E, x \in \sigma(\mathcal{F}(H))\} = \sigma(\mathcal{F}(H))
\end{aligned}$$

□

We now look at **subextensions of the finite Galois extension** E/F and ask about their respective Galois group.

Theorem 6.12. *Let E/F be a* ***finite Galois extension*** *with* ***Galois group*** *G. Let K be an* ***intermediate subfield, fixed by a subgroup*** *H according to Theorem 6.11. We have*

1. *The extension E/K is* ***Galois.***

2. *The extension K/F is* ***normal*** *if and only if H is a* ***normal subgroup*** *of G.*

3. *If H is a* ***normal subgroup*** *of G, then*

$$\operatorname{Gal}(K/F) \simeq G/H = \operatorname{Gal}(E/F)/\operatorname{Gal}(E/K)$$

4. *Whether K/F is normal or not, we have*

$$[K : F] = [G : H].$$

Proof. 1. That E/K is Galois is immediate from the fact that a subextension $E/K/F$ inherits **normality and separability** from E/F.

2. First note that σ is an F-monomorphism of K into E (or C) if and only if σ is the restriction to K of an element of G :

a). if σ is an F-monomorphism of K into E, it can be extended to an F-monomorphism of E into itself thanks to the normality of E (so it is the the restriction to K of this F-monomorphism of E into itself).

b). Conversely, if τ is an F-automorphism of E, then $\sigma = \tau|_K$ is surely a F-monomorphism of K into E.

Now, by Theorem 5.100, we have

$$K/F \text{ normal } \iff \sigma(K) = K \text{ for all } \sigma \in G$$

Note here "for all $\sigma \in G$" is from the above discussion. Since $K = \mathcal{F}(H)$, we just rewrite

$$K/F \text{ normal } \iff \sigma(\mathcal{F}(H)) = \mathcal{F}(H) \text{ for all } \sigma \in G.$$

Now by Theorem 6.11, we know that $\sigma(\mathcal{F}(H)) = \mathcal{F}\left(\sigma H \sigma^{-1}\right)$, and we have

$$K/F \text{ normal } \iff \mathcal{F}\left(\sigma H \sigma^{-1}\right) = \mathcal{F}(H) \text{ for all } \sigma \in G.$$

We are almost there, we now use again the above theorem that tells us that $\mathcal{F}$ is invertible, with inverse $\mathcal{G}$, to get the conclusion:

$$K/F \text{ normal } \iff \sigma H \sigma^{-1} = H \text{ for all } \sigma \in G.$$

3. To prove this isomorphism, we will use the First Isomorphism Theorem for groups Section 3.8.3. Consider the group homomorphism

$$\operatorname{Gal}(E/F) \to \operatorname{Gal}(K/F),\, \sigma \mapsto \sigma|_K\,.$$

This map is surjective (we showed it above, when we mentioned that we can extend $\sigma|_K$ to σ. Its kernel is given by

$$\operatorname{Ker} = \{\sigma,\, \sigma|_K = 1\} = H = \operatorname{Gal}(E/K).$$

Applying the First Isomorphism Theorem for groups Section 3.8.3, we get

$$\operatorname{Gal}(K/F) \simeq \operatorname{Gal}(E/F)/\operatorname{Gal}(E/K)$$

4. Finally, by multiplicativity of the degrees:

$$[E:F] = [E:K][K:F].$$

Since E/F and E/K are Galois, we can rewrite

$$|G| = |H|[K:F].$$

We conclude by Lagrange Theorem:

$$[G:H] = |G|/|H| = [K:F].$$

□

6.3 Finite Fields

We will provide a precise classification of finite fields. Theorem 7.4. Let E be a finite field of characteristic p. 1. The cardinality of E is

$$|E| = p^n,$$

for some $n \geq 1$. It is denoted $E = \mathbb{F}_{p^n}$. 2. Furthermore, E is the splitting field for the separable polynomial

$$f(X) = X^{p^n} - X$$

over $\mathbb{F}_p$, so that any finite field with p^n elements is isomorphic to E. In fact, E coincides with the set of roots of f.

Proof. 1. Let $\mathbb{F}_p$ be the finite field with p elements, given by the integers modulo p. Since E has characteristic p, it contains a copy of $\mathbb{F}_p$. Thus E is a field extension of $\mathbb{F}_p$, and we may see E as a vector space over $\mathbb{F}_p$. If the dimension is n, then let $\alpha_1, \ldots, \alpha_n$ be a basis. Every x in E can be written as

$$x = x_1\alpha_1 + \cdots + x_n\alpha_n$$

and there are p choices for each x_i, thus a total of p^n different elements in E. 2. Let $E^\times$be the multiplicative group of non-zero elements of E. If $\alpha \in E^\times$, then

$$\alpha^{p^n-1} = 1$$

by Lagrange's Theorem, so that

$$\alpha^{p^n} = \alpha$$

for all α in E (including $\alpha = 0$). Thus each element of E is a root of f, and f is separable.

Now f has at most p^n distinct roots, and we have already identified the p^n elements of E as roots of f.

Corollary 7.5. If E is a finite field of characteristic p, then $E/\mathbb{F}_p$ is a Galois extension, with cyclic Galois group, generated by the Frobenius automorphism

$$\sigma : x \mapsto \sigma(x) = x^p, x \in E.$$

Proof. By the above proposition, we know that E is a splitting field for a separable polynomial over $\mathbb{F}_p$, thus $E/\mathbb{F}_p$ is Galois. Since $x^p = x$ for all x in $\mathbb{F}_p$, we have that

$$\mathbb{F}_p \subset \mathcal{F}(\langle\sigma\rangle)$$

that is $\mathbb{F}_p$ is contained in the fixed field of the cyclic subgroup generated by the Frobenius automorphism σ. But conversely, each element fixed by σ is a root of $X^p - X$ so $\mathcal{F}(\langle\sigma\rangle)$ has at most p elements. Consequently

$$\mathbb{F}_p = \mathcal{F}(\langle\sigma\rangle)$$

and

$$\operatorname{Gal}(E/\mathbb{F}_p) = \langle\sigma\rangle$$

This can be generalized when the base field is larger than $\mathbb{F}_p$. Corollary 7.6. Let E/F be a finite field extension with $|E| = p^n$ and $|F| = p^m$. Then E/F is a Galois extension and $m \mid n$. Furthermore, the Galois group is cyclic, generated by the automorphism

$$\tau : x \mapsto \tau(x) = x^{p^m}, x \in E$$

Proof. If the degree $[E : F] = d$, then every x in E can be written as

$$x = x_1\alpha_1 + \cdots + x_d\alpha_d$$

and there are p^m choices for each x_i, thus a total of

$$(p^m)^d = p^n$$

different elements in E, so that

$$d = m/n \text{ and } m \mid n.$$

The same proof as for the above corollary holds for the rest. Thus a way to construct a finite field E is, given p and n, to construct $E = \mathbb{F}_{p^n}$ as a splitting field for $X^{p^n} - X$ over $\mathbb{F}_p$

Theorem 7.7. If G is a finite subgroup of the multiplicative group of an arbitrary field, then G is cyclic. Thus in particular, the multiplicative group $E^\times$of a finite field E is cyclic.

Proof. The proof relies on the following fact: if G is a finite abelian group, it contains an element g whose order r is the exponent of G, that is, the least common multiple of the orders of all elements of G.

Assuming this fact, we proceed as follows: if $x \in G$, then its order divides r and thus

$$x^r = 1.$$

Therefore each element of G is a root of $X^r - 1$ and

$$|G| \leq r.$$

Conversely, $|G|$ is a multiple of the order of every element, so $|G|$ is at least as big as their least common multiple, that is

$$|G| \geq r$$

and

$$|G| = r.$$

Since the order of $|G|$ is r, and it coincides with the order of the element g whose order is the exponent, we have that G is generated by g, that is $G = \langle g\rangle$ is cyclic.

7 Notations

Some notations need to be unified, but in this version, for simplicity of written, I use several notation for one concept. The declare is as follows:

1. Unit group: R^*, $\mathbb{U}$.

References

[1] W. Rudin *et al.*, *Principles of mathematical analysis*, 1976, vol. 3.

[2] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[3] J. A. Gallian, *Contemporary abstract algebra*. Chapman and Hall/CRC, 2021.