IIT BOMBAY

Abstract Algebra

Mentor: Aryaman Maithani

Notes By: Ishan Kapnadak

Summer of Science 2021

Last Updated: July 17, 2021

Contents

§0. Preliminaries

§§0.1. Notation

1. $\mathbb{N} = \{0, 1, 2, \ldots\}$ is the set of non-negative integers
2. $\mathbb{N}^+ = \{1, 2, 3, \ldots\}$ is the set of positive integers
3. $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ is the set of integers
4. $\mathbb{Q} = \left\{\frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0\right\}$ is the set of rationals
5. $\mathbb{R}$ is the set of reals
6. $\mathbb{C}$ is the set of complex numbers
7. $\mathbb{Q}^\times$, $\mathbb{Z}^*$, $\mathbb{R}^\times$, $\mathbb{C}^\times$ will denote the set of non-zero rationals, integers, reals and complex numbers respectively.

§§0.2. Relations and Partitions

Definition 0.1. A relation on A is a subset R of $A \times A$. If $(a, b) \in R$, we say that a is related to b by R and write $a \,R\, b$ or $a \sim b$.

For example, 'equality' ($=$) is a relation on any set A. 'Less than' ($<$) is a relation on $\mathbb{R}$ or any of its subsets. Fix a positive integer n. Then, 'congruence modulo n' ($\equiv$) is a relation on $\mathbb{Z}$, defined by

$$a \equiv b \pmod{n} \iff n \mid a - b.$$

Definition 0.2. A relation $\sim$ on A is said to be an equivalence relation if it is

1. reflexive, i.e, $a \sim a$ for all $a \in A$,
2. symmetric, i.e, $a \sim b$ implies $b \sim a$ for all $a, b \in A$, and
3. transitive, i.e, $a \sim b$ and $b \sim c$ implies $a \sim c$ for all $a, b, c \in A$.

It is easy to show that $<$ is transitive but not reflexive or symmetric. $=$ is an equivalence relation on any set whereas $\equiv \pmod{n}$ is an equivalence relation on $\mathbb{Z}$.

Exercise 0.3.

1. Show that the relation $\sim$ on $\mathbb{N} \times \mathbb{N} = \{(m, n) \mid m, n \in \mathbb{N}\}$ defined by
$$(m, n) \sim (m', n') \iff m + n' = m' + n$$
is an equivalence relation.
2. Show that the relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}^* = \{(m, n) \mid m, n \in \mathbb{Z}, n \neq 0\}$ defined by
$$(m, n) \sim (m', n') \iff m \cdot n' = m' \cdot n$$
is an equivalence relation.

Remark 0.4 (Well-Ordering Property)**.** Every non-empty subset of $\mathbb{N}$ has a least element. That is, if $A \subseteq \mathbb{N}$ and $A \neq \emptyset$ then $\exists\, m \in \mathbb{N}$ such that $m \leq a$ for all $a \in A$. WOP is also true for any subset of $\mathbb{Z}$ which is bounded below. WOP also implies the principle of induction.

Definition 0.5. If $\sim$ is an equivalence relation on A, then for any $a \in A$, the set

$$[a]_\sim := \{b \in A \mid b \sim a\}$$

is called the equivalence class of a with respect to $\sim$. Elements of the equivalence class of a are said to be equivalent to a. If C is an equivalence class, then any element of C is called a representative of class C. We will denote the equivalence class of a as $[a]$ when the relation $\sim$ is clear from context.

Example 0.6. Fix some $n \in \mathbb{N}_+$ and consider the equivalence relation $\equiv \pmod n$. We have a total of n equivalence classes, defined by

$$\begin{aligned}
[0] &= \{kn \mid k \in \mathbb{Z}\} \\
[1] &= \{kn+1 \mid k \in \mathbb{Z}\} \\
&\vdots = \quad \vdots \\
[n-1] &= \{kn+n-1 \mid k \in \mathbb{Z}\}
\end{aligned}$$

These are called the *residue classes* $\pmod n$. We sometimes also denote the residue classes as $\overline{0}, \overline{1}, \ldots, \overline{n-1}$. We denote the set of residue classes of n as $\mathbb{Z}_n$, defined as

$$\mathbb{Z}_n := \left\{\overline{0}, \overline{1}, \ldots, \overline{n-1}\right\}.$$

For $\overline{a}, \overline{b} \in \mathbb{Z}_n$, we further define addition and multiplication as follows

$$\overline{a} + \overline{b} = \overline{a+b} \text{ and } \overline{a} \cdot \overline{b} = \overline{a \cdot b}.$$

One may verify that these operations are indeed well-defined.

Exercise 0.7. The following exercise builds on Exercise 0.3.

1. Show that the equivalence classes of the relation $\sim$ on $\mathbb{N} \times \mathbb{N} = \{(m,n) \mid m,n \in \mathbb{N}\}$ defined by
$$(m,n) \sim (m',n') \iff m+n' = m'+n$$
are in one-to-one correspondence with the set of integers. (We may define the set $\mathbb{Z}$ formally, using $\mathbb{N}$, as equivalence classes of this relation).
2. Show that the equivalence classes of the relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}^* = \{(m,n) \mid m,n \in \mathbb{Z}, n \neq 0\}$ defined by
$$(m,n) \sim (m',n') \iff m \cdot n' = m' \cdot n$$

are in one-to-one correspondence with the set of rationals. (We may define the set $\mathbb{Q}$ formally, using $\mathbb{Z}$, as equivalence classes of this relation).

Definition 0.8. A partition of A is a collection $\{A_i \mid i \in I\}$ of non-empty subsets of A (I is some indexing set) such that

1. $A = \bigcup_{i\in I} A_i$.
2. $A_i \cap A_j = \emptyset$ for all $i, j \in I$ with $i \neq j$.

The ideas of an equivalence relation and partitions are closely related, as we show now.

Proposition 0.9. Let A be a non-empty set.

1. If $\sim$ defines an equivalence relation on A then the set of equivalence classes of $\sim$ form a partition of A.
2. If $\{A_i \mid i \in I\}$ is a partition of A then there exists an equivalence relation $\sim$ on A whose equivalence classes are precisely the sets $A_i, i \in I$.

§§0.3. Number Theory

Definition 0.10. Given $a, b \in \mathbb{Z}$, we say that b divides a and write $b \mid a$ if $a = bc$ for some $c \in \mathbb{Z}$. In this case, a is said to be a multiple of b, or a is said to be divisible by b.

1. Divisibility is a relation on $\mathbb{Z}$, which is reflexive and transitive but not symmetric. In fact, $a \mid b$ and $b \mid a \iff b = \pm a$. This follows since $b \mid a \implies |b| \leq |a|$ (assuming $a \neq 0$).
2. If $b \mid a_1$ and $b \mid a_2$ then $b \mid (a_1 + a_2)$ and $b \mid ka_1$ for all $k \in \mathbb{Z}$. Thus, $b \mid (k_1 a_1 + k_2 a_2)$ for all $k_1, k_2 \in \mathbb{Z}$.

Proposition 0.11 (Division Algorithm)**.** Given $a, b \in \mathbb{Z}$ with $b \neq 0$, there exist unique $q, r \in \mathbb{Z}$ such that $a = bq + r$ and $0 \leq r < |b|$.

Proof. We first prove existence. Let $a, b \in \mathbb{Z}$ with $b \neq 0$. Without loss of generality, we may assume $b > 0$ (if $b < 0$, we do the following considering $-b$ and replace q by $-q$). Consider

$$S := \{a - bx \mid x \in \mathbb{Z} \text{ such that } a - bx \geq 0\}.$$

Then, S is a non-empty subset of $\mathbb{N}$ (we may take $x = -|a|$). Hence, by the Well-Ordering Property, there exists a minimal element in S. Call this r. Since $r \in S$, $r = a - bq$ for some $q \in \mathbb{Z}$ and $r \geq 0$. We now only have to show that $r < b$. Suppose $r \geq b$, then $r - b = a - b(q + 1) \geq 0$ and thus $r - b \in S$. This contradicts the minimality of r. This proves the existence of q, r satisfying the given properties.

Suppose there are $q, r, q', r' \in \mathbb{Z}$ satisfying the given conditions. We have $r - r' = b(q\prime - q)$. If $r \neq r'$, we obtain $|b| < |r - r'|$ which is a contradiction since $0 \leq r, r' < b$. Thus, we get $r = r'$. Since $b \neq 0$, we also get $q = q'$. Hence, q and r are unique. □

Corollary 0.12. Suppose H is a subset of $\mathbb{Z}$ that is non-empty and closed under inverses and addition. Then, $H = n\mathbb{Z}$ for a unique $n \in \mathbb{N}$ where

$$n\mathbb{Z} := \{nm \mid m \in \mathbb{Z}\}$$

H is called a *subgroup* of the additive group $\mathbb{Z}$, as we shall see later.

Proof. Let H be a subset of $\mathbb{Z}$ satisfying the given conditions. If $H = \{0\}$ then $H = n\mathbb{Z}$ with $n = 0$. If $H \neq \{0\}$, then $H \cap \mathbb{N}^+$ is non-empty. Suppose n is the least element in $H \cap \mathbb{N}^+$ (such an n exists by the Well-Ordering Property). Then, $n\mathbb{Z} \subseteq H$ since H is closed under addition. Further, if $m \in H$ then, by the Division Algorithm, there exist $q, r \in \mathbb{Z}$ such that $m = nq + r$ with $0 \leq r < n$. But $r = m - nq \in H$ since $m \in H$ and $n \in H$. If $r > 0$ then $r \in H \cap \mathbb{N}^+$ and the minimality of n is contradicted. Hence, $r = 0$ and $m = nq$. Thus, $H \subseteq n\mathbb{Z}$, giving us $H = n\mathbb{Z}$. □

Definition 0.13. Given $a, b \in \mathbb{Z}$, not both zero, a greatest common divisor or gcd of a and b is a positive integer d such that

1. d is a common divisor of a and b, i.e, $d \mid a$ and $d \mid b$, and
2. if e is a common divisor of a and b, i.e, $e \mid a$ and $e \mid b$, then $e \mid d$.

In the case $a = b = 0$, we define the gcd of a and b to be 0. We usually denote the gcd of a and b as $\gcd(a, b)$.

We leave it as an exercise to the reader to come up with a similar definition for the least common multiple (lcm). We denote the lcm of a and b as $\operatorname{lcm}(a, b)$.

Proposition 0.14 (Bézout's Lemma)**.** Given any $a, b \in \mathbb{Z}$, the gcd of a and b exists and is unique. Moreover, it can be expressed as a combination $ma + nb$ for some $m, n \in \mathbb{Z}$.

Proof. For $a = b = 0$, the proof is trivial. We hence assume that at least one of a and b is non-zero. Consider

$$H = \{ma + nb \mid m, n \in \mathbb{Z}\}.$$

H is a non-empty subset of $\mathbb{Z}$ that is closed under inverses and addition. Hence, by Corollary 0.12, there exists a unique $d \in \mathbb{N}^+$ such that $H = d\mathbb{Z}$. We leave it as an exercise to show that d is indeed the gcd. □

Proposition 0.15.

Let $a, b, c \in \mathbb{Z}$. If $a \mid bc$ and $\gcd(a, b) = 1$ (a and b are coprime), then $a \mid c$.

Proof. Bézout's Lemma tells us that $ma + nb = 1$ for some $m, n \in \mathbb{Z}$. Multiplying throughout by c, we get

$$mac + nbc = c.$$

Since $a \mid ac$ and $a \mid bc$, we must have $a \mid c$. □

Definition 0.16 (Prime Number)**.** An integer p is said to be a prime number if it is greater than 1 and the only positive integers that divide p are 1 and p.

Corollary 0.17 (Euclid's Lemma)**.** If p is a prime number, and $p \mid bc$ for some $b, c \in \mathbb{Z}$, then $p \mid b$ or $p \mid c$.

Proof. Suppose p is prime and $p \mid bc$ for some $\mathbb{Z}$. If $p \mid b$, we are done. If $p \nmid b$, then $\gcd(p, b) = 1$. By Proposition 0.15, $p \mid c$. □

Proposition 0.18. There are infinitely many primes.

Euclid's Proof. If there were only finitely many primes $p_1, \ldots, p_k$, then consider $n = p_1 \cdots p_k + 1$. If p is a prime that divides n then $p \neq p_i$ for all $i \in \{1, \ldots, k\}$ (since $n \equiv 1 \pmod{p_i}$ for all i), which is a contradiction since $p_1, \ldots, p_k$ are assumed to be the only primes. □

Proposition 0.19. If $b \in \mathbb{Z}$ and p is a prime number such that $p \nmid b$, then there exists $b' \in \mathbb{Z}$ such that $bb' \equiv 1 \pmod{p}$. Moreover, b' can be chosen such that $1 \leq b' < p$ and b' is unique.

Proof. Since $\gcd(b, p) = 1$, we have $pu + bv = 1$ for some $u, v \in \mathbb{Z}$ (Bézout's Lemma). Thus, $b' = v$ satisfies $bb' \equiv 1 \pmod{p}$. The uniqueness of b' satisfying $bb' \equiv 1 \pmod{p}$ and $1 \leq b' < p$ follows by replacing any $v \in \mathbb{Z}$ satisfying $bv \equiv 1 \pmod{p}$ by the unique element b' in the residue class of $v \pmod{p}$ such that $0 \leq b' < p$. Moreover, $b' = 0 \implies 0 \equiv 1 \pmod{p}$ which is a contradiction. Hence, $1 \leq b' < p$. □

Theorem 0.20 (Fundamental Theorem of Arithmetic)**.** Every positive integer n can be written as a product of primes. That is,

$$n = p_1 \cdots p_k$$

for some primes $p_1, \ldots, p_k$ (not necessarily distinct). Moreover, this factorisation is unique up to rearrangement of terms. That is, if

$$n = q_1 \cdots q_l$$

where $q_1, \ldots, q_l$ are primes then $k = l$ and $q_i = p_{\sigma(i)}$ for all $i \in \{1, \ldots, k\}$ for some permutation σ of $\{1, \ldots, k\}$.

Proof. We first prove the existence by induction. If $n = 1$, the hypothesis holds with $k = 0$ since the empty product is 1, by convention. Suppose $n > 1$ and the hypothesis holds for all positive integers strictly less than n. If n is prime, the hypothesis clearly holds with $k = 1$. If n is not prime, then $n = n_1 n_2$ for some $n_1, n_2 \in \mathbb{N}^+$ with $n_1 < n$ and $n_2 < n$. By the induction hypothesis, both n_1 and n_2 are finite products of primes and hence, so is n. Thus, existence is proved by induction.

Next, we prove uniqueness. Suppose $n \in \mathbb{N}^+$ is written as

$$n = p_1 \cdots p_k$$

and also

$$n = q_1 \cdots q_l$$

where $p_1, \ldots, p_k, q_1, \ldots, q_l$ are primes. We can induct on k. If $k = 0$, then $n = 1$ and hence $l = 0$. Hence, $k = l$ and p_i's are a permutation of q_i's (vacuously). Suppose that $k > 1$ and the result holds for $k - 1$. Then,

$$p_1 \mid n = q_1 \cdots q_l.$$

Hence, by an obvious extension of Euclid's Lemma, $p_1 \mid q_j$ for some j since q_j is a prime and $p_1 > 1$, we must have $p_1 = q_j$. Thus,

$$p_2 \cdots p_k = q_1 \cdots q_{j-1} \cdot q_{j+1} \cdots q_l.$$

By induction hypothesis, $k - 1 = l - 1$, giving us $k = l$, and $p_2, \ldots, p_k$ are a permutation of $q_1, \ldots, q_{j-1}, q_{j+1}, \ldots q_l$. Since we showed $p_1 = q_j$, it follows that $p_1, \ldots, p_k$ is a permutation of $q_1, \ldots, q_l$. Thus, uniqueness is also proved by induction. □

Following is another version of the above theorem.

Theorem 0.21 (Fundamental Theorem of Arithmetic - Version 2)**.** Every non-zero integer n can be written as

$$n = \epsilon \cdot p_1^{e_1} \cdots p_h^{e_h}$$

where $\epsilon \in \{1, -1\}$, $p_1, \ldots, p_h$ are distinct primes and $e_1, \ldots, e_h$ are positive integers and $h \geq 0$. Moreover, p_i and e_i are uniquely determined by n.

This allows us to associate with every prime an 'exponent' or 'valuation' on the set of non-zero integers. Let p be a prime. We define $\nu_p \colon \mathbb{Z}^* \to \mathbb{N}$ as follows

$$\nu_p(n) = \begin{cases} e_i & \text{if } p = p_i \text{ for some } 1 \leq i \leq h \text{ in the prime decomposition of } n, \\ 0 & \text{otherwise.} \end{cases}$$

With this notation, we can write

$$n = \epsilon \cdot \prod_p p^{\nu_p(n)}$$

where the product is over all primes. This product is well defined since $\nu_p(n) = 0$ for all but finitely many primes.

Proposition 0.22. If $m, n \in \mathbb{Z}^*$, then

$$\gcd(m,n) = \prod_p p^{\min\{\nu_p(m),\nu_p(n)\}}$$

$$\operatorname{lcm}(m,n) = \prod_p p^{\max\{\nu_p(m),\nu_p(n)\}}$$

Corollary 0.23. If $m, n \in \mathbb{Z}^*$ then $|m \cdot n| = \gcd(m,n) \cdot \operatorname{lcm}(m,n)$.

We can extend the function ν_p to non-zero rationals ($\mathbb{Q}^\times$) by defining

$$\nu_p\left(\frac{m}{n}\right) := \nu_p(m) - \nu_p(n) \text{ for } m, n \in \mathbb{Z}^*$$

We leave it as an exercise to show that the above is indeed well-defined. This allows us to write every non-zero rational number as a product of primes. Suppose $r \in \mathbb{Q}^\times$, then we have

$$r = \epsilon \cdot \prod_p p^{\nu_p(r)}$$

where $\epsilon \in \{1, -1\}$ and the product is over all primes. This product is well-defined since $\nu_p(r) = 0$ for all but finitely many primes.

Note: By convention, we often define $\nu_p(0) = \infty$. With this, the following is true.

Proposition 0.24. Let $\nu_p \colon \mathbb{Q} \to \mathbb{Z} \cup \{\infty\}$ as defined above for some prime p. Then,

1. ν_p is surjective,
2. $\nu_p(r) = \infty \iff r = 0$,
3. $\nu_p(rs) = \nu_p(r) + \nu_p(s)$ for all $r, s \in \mathbb{Q}$, and
4. $\nu_p(r+s) \geq \min\{\nu_p(r), \nu_p(s)\}$ for all $r, s \in \mathbb{Q}$.

Proof. Left as an exercise. □

Remark 0.25. The function $\nu_p \colon \mathbb{Q} \to \mathbb{Z} \cup \{\infty\}$ is called the p-*adic valuation of* $\mathbb{Q}$. One can use it to define a norm and a metric on $\mathbb{Q}$, as follows:

$$|x|_p := 2^{-\nu_p(x)} \text{ for all } x \in \mathbb{Q}.$$

with the convention $2^{-\infty} := 0$. One can see that

1. $|x|_p \geq 0$ and $|x|_p = 0 \iff x = 0$.
2. $|xy|_p = |x|_p |y|_p$.

3. $|x+y|_p \leq \max\left\{|x|_p, |y|_p\right\} \leq |x|_p + |y|_p$.

Thus, $|\cdot|_p : \mathbb{Q} \to \mathbb{R}$ defines a norm on $\mathbb{Q}$. Further, $d_p : \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$ defined by

$$d_p(x,y) := |x-y|_p \text{ for } x, y \in \mathbb{Q}$$

defines a metric on $\mathbb{Q}$. Once can consider the completion of $\mathbb{Q}$ with respect to d_p and this gives rise to a set denoted by $\mathbb{Q}_p$ called the field of p-adic numbers.[1] We will not spend much time on p-adic numbers in this document.

Definition 0.26. Given $n \in \mathbb{N}^+$, we define $\varphi(n)$ to be the number of integers $a \in \mathbb{N}^+$ with $1 \leq a \leq n$ and $\gcd(a,n) = 1$. φ is called Euler's totient function.

We can obtain an explicit formula for $\varphi(n)$ using a very basic counting principle - called the inclusion-exclusion principle.

Theorem 0.27 (Principle of Inclusion and Exclusion)**.** If $A_1, \ldots, A_r$ are finite sets, then

$$|A_1 \cup \cdots \cup A_r| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{r-1} |A_1 \cap \cdots \cap A_r|.$$

Proof. The proof is trivial using induction on r. □

For $n \in \mathbb{N}^+$, suppose $p_1, \ldots, p_r$ be the distinct primes that divide n. We have

$$\begin{aligned} n - \varphi(n) &= \left|\left\{a \in \mathbb{N}^+ \mid 1 \leq a \leq n \text{ and } \gcd(a,n) \neq 1\right\}\right| \\ &= \left|\bigcup_{i=1}^{r} A_{p_i}\right| \end{aligned}$$

where for any $m \in \mathbb{N}^+$ with $m \mid n$, $A_m := \left\{a \in \mathbb{N}^+ \mid 1 \leq a \leq n \text{ and } m \mid a\right\}$. Observe that $|A_m| = n/m$. By the Principle of Inclusion and Exclusion,

$$\begin{aligned} n - \varphi(n) &= \sum_i |A_{p_i}| - \sum_{i<j} \left|A_{p_i} \cap A_{p_j}\right| + \sum_{i<j<k} \left|A_{p_i} \cap A_{p_j} \cap A_{p_k}\right| - \cdots + (-1)^{r-1} |A_{p_1} \cap \cdots \cap A_{p_r}| \\ &= \sum_i |A_{p_i}| - \sum_{i<j} \left|A_{p_i p_j}\right| + \sum_{i<j<k} \left|A_{p_i p_j p_k}\right| - \cdots + (-1)^{r-1} |A_{p_1 \cdots p_r}| \\ &= \sum_i \frac{n}{p_i} - \sum_{i<j} \frac{n}{p_i p_j} + \sum_{i<j<k} \frac{n}{p_i p_j p_k} - \cdots + (-1)^{r-1} \frac{n}{p_1 \cdots p_r} \end{aligned}$$

[1] Within the p-adic numbers, we also have the p-adic integers, denoted by $\mathbb{Z}_p$. Note this is **not** the set of residue classes modulo p. Since we will not deal with p-adic integers in this document, we continue to use the notation $\mathbb{Z}_n$ for the set of residue classes modulo n. Another common notation for the same is $\mathbb{Z}/n\mathbb{Z}$.

$$\therefore\ \varphi(n) = n - \sum_i \frac{n}{p_i} + \sum_{i<j} \frac{n}{p_i p_j} - \sum_{i<j<k} \frac{n}{p_i p_j p_k} + \cdots + (-1)^r \frac{n}{p_1 \cdots p_r}$$
$$= n\left(1 - \sum_i \frac{1}{p_i} + \sum_{i<j} \frac{1}{p_i p_j} - \sum_{i<j<k} \frac{1}{p_i p_j p_k} + \cdots + (-1)^r \frac{1}{p_1 \cdots p_r}\right)$$

We can express the above term as a product of r factors. This gives us the famous *Euler's Product Formula*:

$$\boxed{\varphi(n) = n \cdot \prod_{p|n} \left(1 - \frac{1}{p}\right)}$$

Corollary 0.28 (Multiplicativity of φ)**.** For $m, n \in \mathbb{N}^+$, $\varphi(mn) = \varphi(m) \cdot \varphi(n)$ iff $\gcd(m, n) = 1$.

If $n = p_1^{e_1} \cdots p_h^{e_h}$ where $p_1, \ldots, p_k$ are distinct primes and $e_1, \ldots, e_h \in \mathbb{N}^+$, one may reduce the product formula to the following

$$\varphi(n) = \prod_{i=1}^{h} p_i^{e_i - 1} \cdot (p_i - 1)$$

In particular, this allows us to deduce that $\varphi(n)$ is even for all $n > 2$.

Exercise 0.29. For $n, d \in \mathbb{N}^+$, show that

$$\sum_{d|n} \varphi(d) = n.$$

Definition 0.30. Let $n \in \mathbb{N}^+$. A set $\{a_1, \ldots, a_k\}$ of integers is called a reduced system of residues (mod n) if the following hold:

1. $\gcd(a_i, n) = 1$ for all $i \in \{1, \ldots, k\}$,
2. $a_i \not\equiv a_j \pmod{n}$ for all $i, j \in \{1, \ldots, k\}, i \neq j$, and
3. $a \in \mathbb{Z}, \gcd(a, n) = 1 \implies a \equiv a_i \pmod{n}$ for some $i \in \{1, \ldots, k\}$.

For example, $\{1, 3, 5, 7\}$ is a reduced system of residues (mod 8).

Proposition 0.31. Let $n \in \mathbb{N}^+$. Then,

1. $\{a \in \mathbb{Z} \mid 1 \leq a \leq n \text{ and } \gcd(a, n) = 1\}$ is a reduced system of residues (mod n).
2. Any reduced system of residues (mod n) has cardinality $\varphi(n)$.
3. If $\{a_1, \ldots, a_k\}$ is a reduced system of residues (mod n) and $a \in \mathbb{Z}$ such that $\gcd(a, n) = 1$, then $\{aa_1, \ldots, aa_k\}$ is also a reduced system of residues (mod n).

Theorem 0.32 (Euler's Theorem)**.** Let $n \in \mathbb{N}^+$ and $a \in \mathbb{Z}$. If $\gcd(a, n) = 1$, then $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Proof. Let $\{a_1, \ldots, a_k\}$ be a reduced system of residues $\pmod{n}$. Then, we proved that $k = \varphi(n)$ and $\{aa_1, \ldots, aa_k\}$ is also a reduced system of residues $\pmod{n}$. With a little bit of effort, we can show that

$$\prod_{i=1}^{k} aa_i \equiv \prod_{i=1}^{k} a_i \pmod{n} \implies a^{\varphi(n)} \cdot \prod_{i=1}^{k} a_i \equiv \prod_{i=1}^{k} a_i \pmod{n}$$

$$\therefore\ a^{\varphi(n)} \equiv 1 \pmod{n}$$ □

Corollary 0.33 (Fermat's Little Theorem)**.** If p is a prime and $a \in \mathbb{Z}$, then $a^p \equiv a \pmod{p}$.

Theorem 0.34 (Wilson's Theorem)**.** Let $n \in \mathbb{N}^+$ with $n > 1$. n is prime iff $(n-1)! \equiv -1 \pmod{n}$.

Proof. If $p = 2$ or 3, the result clearly holds. Suppose p is a prime greater than 3. For any $a \in \mathbb{Z}$, there is a unique $a' \in \mathbb{Z}$ with $1 \leq a' \leq p-1$ and $aa' \equiv 1 \pmod{p}$ (Proposition 0.19). We see that

$$a = a' \iff a^2 \equiv 1 \pmod{p} \iff p \mid (a-1)(a+1) \iff a = 0 \text{ or } a = p-1$$

Hence the $p-3$ numbers $2, \ldots, p-2$ can be paired as (a, a') where $2 \leq a, a' \leq p-2$, $aa' \equiv 1 \pmod{p}$. Moreover, $a \neq a'$ in any of these pairs. This tells us that $2 \cdot 3 \ldots \cdot (p-2) \equiv 1 \pmod{p}$. This gives us

$$(p-1)! \equiv (p-1) \pmod{p} \implies (p-1)! \equiv -1 \pmod{p}$$

We leave the converse as an exercise. □

Theorem 0.35 (Chinese Remainder Theorem)**.** If $n_1, \ldots, n_k$ are pairwise coprime positive integers, that is, $n_1, \ldots, n_k \in \mathbb{N}^+$ and $\gcd(n_i, n_j) = 1$ for all $i, j \in \{1, \ldots, k\}$, $i \neq j$, and if $c_1, \ldots, c_k$ are any integers, then the congruences

$$x \equiv c_i \pmod{n_i} \text{ for } i \in \{1, \ldots, k\}$$

have a solution, which is unique $\pmod{n_1 \cdots n_k}$.

Proof. Let n be the product $n_1 \cdots n_k$ and let $m_i = n/n_i$ for $i \in \{1, \ldots, k\}$. Since n_i's are pairwise coprime, we get $\gcd(m_i, n_i) = 1$ for all $i \in \{1, \ldots, k\}$ Then, the congruence

$$m_i x \equiv c_i \pmod{n_i}$$

has a solution, say x_i, for each i. Consider $x = m_1x_1 + \cdots + m_kx_k$. For any $i \in \{1, \ldots, k\}$, we see that $n_i \mid m_j$ for all $j \neq i$. Hence, x as chosen above satisfies all the congruences. This proves existence. We leave the proof of uniqueness $\pmod{n}$ as an exercise. □

§1. Group Theory

§§1.1. Introduction

Definition 1.1. Suppose G is a nonempty set. A binary operation on G (or a law of composition) is a function $*\colon G \times G \to G$. For any $a, b \in G$ we denote $*(a, b)$ as $a * b$ or simply ab.

Example 1.2. Addition, subtraction and multiplication are binary operations on $\mathbb{R}$. Addition and subtraction are binary operations on $\mathbb{R}^{m\times n}$, the set of $m \times n$ real matrices. We denote the set of $n \times n$ real, invertible matrices by $GL_n(\mathbb{R})$ (this is called the general linear group, as shall be discussed later). Multiplication is a binary operation on $GL_n(\mathbb{R})$. Note however that addition is not a binary operation on $GL_n(\mathbb{R})$ since the sum of two invertible matrices may be singular. $GL_1(\mathbb{R})$ is denoted as $\mathbb{R}^\times$, the set of non-zero real numbers. Let S be any non-empty set and let $\mathcal{M}$ denote the set of all functions from S to S. Then, function composition is a binary operation on $\mathcal{M}$.

Definition 1.3. A binary operation $*$ on a set G is said to be associative if $(a * b) * c = a * (b * c)$ for all $a, b, c \in G$.

Definition 1.4. A binary operation $*$ on a set G is said to be commutative if $a * b = b * a$ for all $a, b \in G$.

Proposition 1.5 (Generalised Associative Law)**.** Let G be a set and let $*$ be an associative binary operation on G. For any $g_1, \ldots, g_n \in G$, the product $g_1 * \cdots * g_n$ is independent of how we bracket it.

Proof. This is left as an exercise. The idea is to use induction on n. First show the basis. Then, assume that for any $k < n$, any bracketing of of k elements $b_1 * \cdots * b_k$ can be reduced to $b_1 * (b_2 * (\cdots b_k)) \cdots)$. Next, argue that $a_1 * \cdots * a_n$ can be reduced to $(a_1 * \cdots * a_k) * (a_{k+1} * \cdots * a_n)$ for some $k < n$. Apply the induction condition on each subproduct to complete the proof. □

Definition 1.6. A group is an ordered pair $(G, *)$ where G is a set and $*$ is a binary operation on G such that

1. $(a * b) * c = a * (b * c)$ for all $a, b, c \in G$, that is, $*$ is associative,
2. there exists an element $e \in G$, called an identity of G, such that $a * e = e * a = a$ for all $a \in G$, and
3. for each $a \in G$, there is an element $a^{-1} \in G$, called an inverse of a, such that $a * a^{-1} = a^{-1} * a = e$.

We say that G is a group under $*$ if $(G, *)$ is a group. If $*$ is clear from context, we may simply say

that G is a group. We further say that G is a *finite group* if G is a finite set. Note that any group is nonempty by virtue of the existence of an identity element.

Definition 1.7. We say that a group $(G, *)$ is abelian if $a * b = b * a$ for all $a, b \in G$.

Definition 1.8. Let G be a group. We define the order of G as the cardinality of G, denoted by $|G|$.

Example 1.9.

1. $\mathbb{Z}, \mathbb{C}, \mathbb{Q}$ and $\mathbb{R}$ are all groups under the addition operation with $e = 0$ and $a^{-1} = -a$, for all a. $\mathbb{Q}^\times$, $\mathbb{R}^\times$, $\mathbb{C}^\times$, $\mathbb{Q}^+$, $\mathbb{R}^+$ are all groups under the multiplication operation with $e = 1$ and $a^{-1} = 1/a$, for all a. Note however that $\mathbb{Z}^*$ is not a group under the multiplication operation since the element 2 (for instance) does not have an multiplicative inverse in $\mathbb{Z}^*$. We shall take the associative laws of these sets under addition and multiplication as given.
2. Rotation matrices in 2-dimensions with multiplication also form a group. This is called the $SO_2(\mathbb{R})$ group. This is the set of special orthogonal matrices - the set of 2×2 orthogonal matrices with determinant 1. On the other hand, the set of 2×2 orthogonal matrices forms another group, called the orthogonal group, $O_2(\mathbb{R})$.
3. Consider the group of non-zero complex numbers, $\mathbb{C}^\times$ under multiplication ($GL_1(\mathbb{C})$). This also forms a group under multiplication. For any $n \in \mathbb{N}^+$, consider the n^{th} root of unity, defined as
$$\omega_n = \cos\frac{2\pi}{n} + \iota \sin\frac{2\pi}{n}$$
The set containing all powers of ω_n forms a finite group of order n under complex multiplication, whose elements are precisely the n roots of the polynomial $z^n = 1$. This group is called the *cyclic group generated by* ω_n, and is denoted as μ_n.
4. For $n \in \mathbb{N}^+$, $\mathbb{Z}_n$ is an abelian group of order n under the addition operation with $e = \overline{0}$ and the inverse of $\overline{a}$ defined as $\overline{-a}$. We denote this group as $\mathbb{Z}_n$. Notice that $\mathbb{Z}_n$ behaves similar to the cyclic group generated by ω_n and can be thought of as being generated by the equivalence class $\overline{1}$.

   However, $\mathbb{Z}_n$ does not form a group under multiplication. This is because not all numbers have a multiplicative inverse modulo n. From number theory, we know that a number a has a multiplicative inverse modulo n if and only if $(a, n) = 1$. The set of equivalence classes $\overline{a}$ which have a multiplicative inverse modulo n form an abelian group under multiplication. We denote this group as $\mathbb{Z}_n^\times$. The order of this group is equal to the number of integers between 1 and n which are co-prime with n. This is given precisely by Euler's totient function, φ. Thus, $\mathbb{Z}_n^\times$ forms an abelian group of order $\varphi(n)$ under multiplication. We sometimes also denote this group as U_n.

Theorem 1.10. Let G be a group under operation $*$. Then

1. The identity of G is unique

2. For each $g \in G$, g^{-1} is unique
3. For each $g \in G$, $(g^{-1})^{-1} = g$
4. For $a, b \in G$, $(a * b)^{-1} = b^{-1} * a^{-1}$

Proof.

1. Let f and g be two identities of G. We have $f * g = f$ and $f * g = g$. Thus $f = g$.
2. Let $a, b \in G$ be two inverses of some $g \in G$ and let e be the identity of G. We show that $b = a$.

$$\begin{aligned} b &= b * e \text{ (definition of } e) \\ &= b * (g * a) \text{ (since } a \text{ is an inverse of } g) \\ &= (b * g) * a \text{ (associativity)} \\ &= e * a \text{ (since } b \text{ is an inverse of } g) \\ &= a \text{ (definition of } e) \end{aligned}$$

3. We have $g^{-1} * g = g * g^{-1} = e$, implying that $(g^{-1})^{-1} = g$.
4. Using the generalised associative law (Proposition 1.5) on $(a * b) * (b^{-1} * a^{-1})$ and $(b^{-1} * a^{-1}) * (a * b)$ gives the required result.

□

Notation: For any group $(G, *)$, we denote $a * b$ as ab. For some group G, $g \in G$ and $n \in \mathbb{Z}^+$, we write $x \cdots x$ (n times) as x^n. For $n < 0, n \in \mathbb{Z}$, we define $x^n := (x^{-1})^{-n}$, which is the same as $(x^{-n})^{-1}$ (Prove!) We usually denote the identity element of any group as 1 and we define $x^0 := 1$.

Definition 1.11. Let G be a group and let $x \in G$. Let n be the smallest positive integer such that $x^n = 1$. n is called the order of x and is denoted by $|x|$. If no such positive power exists, we say that x is of infinite order.

Proposition 1.12. Any element of a finite group has finite order.

Proof. Let G be a group and let $x \in G$. It suffices to show that $x^n = 1$ for some $n \in \mathbb{N}$. Note that $x^0, \ldots, x^{|G|}$ are $|G| + 1$ elements of G. Since the cardinality of G is $|G|$, we may conclude that two of these must be equal (pigeonhole principle). Thus,

$$x^n = x^m$$

for some $0 \leq n < m \leq |G|$. This gives us

$$1 = x^{m-n}$$

Since $m - n \in \mathbb{N}$, the claim follows. □

§§1.2. Dihedral Groups

We now look at importance class of groups whose elements are symmetries of geometric objects. The simplest objects to consider are regular n-gons. For each $n \in \mathbb{Z}^+$, $n \geq 3$, we let D_{2n} be the set of symmetries of a regular n-gon. A symmetry is any rigid motion of the n-gon such that after this motion, the n-gon exactly covers the original n-gon. This can be thought of as first labelling n vertices as $1, 2, \ldots, n$ and then describing a symmetry uniquely by the corresponding permutation σ of $\{1, 2, \ldots, n\}$.

We make D_{2n} a group by defining st for $s, t \in D_{2n}$ to be the symmetry obtained by first applying t then s. That is, if s and t have the permutations σ and τ respectively on the vertices then st has the permutation $\sigma \circ \tau$.

We now show that $|D_{2n}| = 2n$. Observe that vertex 1 can be mapped to any one of the n vertices. Let's say that it is mapped to vertex i. Now, since vertex 2 is adjacent to vertex 1, it must be mapped to either $i+1$ or $i-1$. The position of vertex 2 fixes the entire permutation. Thus, we have $2n$ possible permutations, and so $|D_{2n}| = 2n$. We call D_{2n} the *dihedral group of order* $2n$.

These $2n$ symmetries are n rotations by $2\pi i/n$ radians about the center for $i = 1, 2, \ldots, n$ and the n reflections about the n lines of symmetry.

Let r be the clockwise rotation of the n-gon by $2\pi/n$ radians and let s be the reflection symmetry that reflects the n-gon about the axis passing through vertex 1 and the center. The following properties follow (proof is omitted):

1. $1, r, r^2, \ldots, r^{n-1}$ are all distinct and $r^n = 1$. Thus, $|r| = n$
2. $|s| = 2$
3. $s \neq r^i$ for any i
4. $sr^i \neq sr^j$ for all $0 \leq i, j \leq n-1$, $i \neq j$. Thus, we have
$$D_{2n} = \left\{1, r, r^2, \ldots, r^{n-1}, s, sr, sr^2, \ldots sr^{n-1}\right\}.$$
5. $rs = sr^{-1}$ (Since r and s do not commute, D_{2n} is non-abelian) [2]
6. $r^i s = sr^{-i}$

We conclude that all elements of D_{2n} can be expressed uniquely as $s^k r^i$ where k is 0 or 1 and $0 \leq i \leq n-1$. Moreover, identities (1), (2) and (6) will easily allow us to obtain this unique representation. Consider $n = 12$. For example, we have

$$(sr^9)(sr^6) = s(r^9 s)r^6 = s(sr^{-9})r^6 = s^2 r^{-3} = r^{-3} = r^9$$

Finally, another common way of writing the dihedral group, is as a presentation[3] as follows

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

[2] Note that to claim D_{2n} to be non-abelian, we need $r \neq r^{-1}$. This is true for $n \geq 3$.

[3] A *presentation* is another form of defining a group G. We have a set of generators, S, such that every element of G can be written as a product of these generating elements. We also have a set of relations, R, among these generators. The group is then *presented* as $\langle S \mid R \rangle$.

§§1.3. Quaternion and Heisenberg Groups

The quaternion group is a group of order 8, defined as follows

$$Q_8 = \{\pm 1, \pm \hat{\imath}, \pm \hat{\jmath}, \pm \mathbb{K}\}$$

where each element is a 2×2 complex matrix of determinant 1. Hence, this group lies within $GL_2(\mathbb{C})$. The matrices are defined as follows

$$1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \hat{\imath} = \begin{bmatrix} \imath & 0 \\ 0 & -\imath \end{bmatrix} \quad \hat{\jmath} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \hat{k} = \begin{bmatrix} 0 & \imath \\ \imath & 0 \end{bmatrix}$$

In Q_8, we have the following relations

$$\hat{\imath}^2 = \hat{\jmath}^2 = \hat{k}^2 = -1$$

$$\hat{\imath}\hat{\jmath} = \hat{k} = -\hat{\jmath}\hat{\imath}$$

$$\hat{\jmath}\hat{k} = \hat{\imath} = -\hat{k}\hat{\jmath}$$

$$\hat{k}\hat{\imath} = \hat{\jmath} = -\hat{\imath}\hat{k}$$

$$\hat{\imath}\hat{\jmath}\hat{k} = -1$$

The Heisenberg group is a group of 3×3 upper-triangular matrices, defined as follows

$$H(\mathbb{R}) = \left\{ \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \mid a, b, c \in \mathbb{R} \right\}.$$

$H(\mathbb{R})$ forms an infinite non-abelian group under matrix multiplication, where each element has determinant 1.

§§1.4. Symmetric Groups

Let Ω be any non-empty set and S_Ω be the set of all bijections from Ω to Ω (or permutations of Ω). The set S_Ω is a group under function composition, $\circ$, since function composition is associative, the identity is the identity map on Ω and every bijection has an inverse bijection. In the case where $\Omega = \{1, \ldots, n\}$, we denote S_Ω as S_n. S_n is called the *symmetric group of order n*. It is easy to show that $|S_n| = n!$. We now illustrate a convenient notation to write elements of S_n, called the *cycle decomposition*. A *cycle* is a string of integers that cyclically permutes the integers of this string, leaving all other integers fixed. For example, $(a_1\ a_2\ \ldots\ a_k)$ sends a_1 to a_2, a_2 to $a_3, \ldots, a_{k-1}$ to a_k and a_k to a_1. In general, any element σ of S_n can be rearranged and written as k (disjoint) cycles as

$$\sigma = (a_1\ \ldots\ a_{m_1})(a_{m_1+1}\ \ldots\ a_{m_2})\ldots(a_{m_{k-1}+1}\ \ldots\ a_{m_k})$$

To find where an element i is sent to by a permutation, we simply need to find the element written after i in the cycle decomposition. Any permutation σ can be easily written as its cycle decomposition using the following algorithm.

1. To start a new cycle, pick the smallest number in $\{1, \ldots, n\}$ that has not appeared in a previous cycle. Call it a. Begin the new cycle $(a$

2. Let $\sigma(a) = b$. If $b = a$, close the cycle and return to step 1. If $b \neq a$, write b next to a so that the cycle becomes $(a\, b$
3. Let $\sigma(b) = c$. If $c = a$, close the cycle and return to step 1. If $c \neq a$, write c next to b and repeat this step with c as b until the cycle closes.

The *length* of a cycle is the number of integers appearing in the cycle. A cycle of length l is called an l-cycle. (Notice that an l-cycle has order l) By convention, we omit 1-cycles. Thus, if some element i does not in a cycle decomposition of a permutation, it is understood that the permutation fixes i. The identity permutation is written as 1. The final step in the algorithm is thus to remove all 1-cycles. Note that

$$(1\,3)\,(1\,2) = (1\,2\,3) \text{ and } (1\,2)\,(1\,3) = (1\,3\,2)$$

This shows that S_n is *non-abelian* for all $n \geq 3$. Note that since disjoint cycles permute elements in disjoint sets, disjoint cycles commute. Thus, rearranging the cycles in any product of disjoint cycles does not change the permutation.

Remark 1.13. We define an equivalence relation on Ω (any general non-empty set), with $a \sim b$ if $b = \sigma^k(a)$ for some k. Here σ^k denotes the permutation σ composed k times. It is easy to verify that this is an equivalence relation. Each disjoint cycle in the cycle decomposition of σ represents an equivalence class of $\sim$. (Verify!)

Note that every symmetry transformation of an equilateral triangle can be associated with a unique permutation of the vertices. Likewise, every permutation of the vertices of an equilateral triangle can be associated with (the same) symmetry transformation. We see that D_6 and S_3 are essentially the same group. This will be made more precise when we discuss isomorphisms.

§§1.5. Conjugacy

Recall from linear algebra that an $n \times n$ real symmetric matrix, A can be diagonalised. That is, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then $A \sim \Lambda$, where Λ is a diagonal matrix containing the eigenvalues. That is, $CAC^{-1} = \Lambda$ for an *orthogonal* matrix, C, whose column vectors are the corresponding eigenvectors of A. This very idea can be made more abstract and applied to groups, in general.

Definition 1.14. Let G be a group and $g, h \in G$. We say that g is a conjugate of h if there exists an $x \in G$ such that $h = xgx^{-1}$.

Proposition 1.15. Conjugacy is an equivalence relation on the group G

Proof. We prove reflexivity, symmetry and transitivity.

1. For reflexivity, we may take x to be identity to give us that $h \sim h$ for all $h \in G$.
2. Suppose $h \sim g$. Then, $h = xgx^{-1}$ for some $x \in G$. Left-multiplying by x^{-1} and right-multiplying by x, we get $g = x^{-1}hx = x^{-1}h(x^{-1})^{-1}$. Since, $x^{-1} \in G$, we have that $g \sim h$, proving symmetry.

3. Suppose $h \sim g$ and $l \sim h$. Then, there exist $x, y \in G$ such that $h = xgx^{-1}$ and $l = yhy^{-1}$. Substituting h, we see that $l = (yx)g(x^{-1}y^{-1}) = (yx)g(yx)^{-1}$. Since $yx \in G$, we have that $l \sim g$, proving transitivity.

□

The equivalence classes of the conjugacy relation are called *conjugacy classes*. Thus, all of G is a disjoint union of conjugacy classes. We see that for any group G, the identity element is the only element in its conjugacy class. If G is abelian, then $gag^{-1} = a$ for all $a, g \in G$. Thus, each element in an abelian group is part of a unique conjugacy class. We shall typically denote the conjugacy class of $g \in G$ as $C(g)$.

Proposition 1.16. Let G be a group and let $g, h \in G$ belong to the same conjugacy class, i.e, $g \sim h$. Then, $|g| = |h|$

Proof. Since $g \sim h$, we know that $g = xhx^{-1}$ for some $x \in G$. Now

$$g^n = (xhx^{-1})^n = xh^nx^{-1}$$

Thus, $g^n = 1 \iff h^n = 1$. □

We want to understand what are the conjugacy classes in S_n. We will first begin by analysing the conjugacy classes of S_3. Using our familiar cycle decomposition notation, we may define S_3 as follows

$$S_3 = \{1, (1\,2), (1\,3), (2\,3), (1\,2\,3), (1\,3\,2)\}.$$

The conjugacy class of a 2-cycle cannot contain any 3-cycle or the identity element (since they have different orders). With a little bit of work, we can show that the number of conjugacy classes are precisely 3 - the identity, the 2-cycles and the 3-cycles. Let us relate these numbers with partitions of natural numbers. We first define what are partitions.

Definition 1.17. Let $n \in \mathbb{N}^+$. A partition of n is a tuple $\lambda = (\lambda_1, \ldots, \lambda_l)$ of positive integers $\lambda_1 \geq \ldots \geq \lambda_l$ such that $\lambda_1 + \cdots + \lambda_l = n$.

The number of partitions of 3 is equal to 3 and these are given by $(1,1,1)$, $(2,1)$ and (3). This number is exactly equal to the number of conjugacy classes of S_3. For a general symmetric group, S_n, the number of conjugacy classes is given precisely by the number of partitions of n.

Suppose $\tau = (i_1, \ldots, i_k)$ is a k-cycle. Suppose $\sigma \in S_n$. We want to show that

$$\sigma\tau\sigma^{-1} = \left(\sigma(i_1), \ldots, \sigma(i_k)\right).$$

We have

$$\begin{aligned}\sigma\tau\sigma^{-1}(\sigma(i_1)) &= \sigma\tau(\sigma^{-1}\sigma)(i_1)\\ &= \sigma\tau(i_1)\\ &= \sigma(i_2)\end{aligned}$$

Similarly, we show that the two permutations have the same effect on all k elements, $\sigma(i_1), \ldots, \sigma(i_k)$. Now, suppose $x \notin \{\sigma(i_1), \ldots, \sigma(i_k)\} \iff \sigma^{-1}(x) \neq i_s$ for any $s = 1, \ldots, k$. We have

$$\begin{aligned}\sigma\tau\sigma^{-1}(x) &= \sigma\sigma^{-1}(x) \text{ (since } \sigma^{-1}(x) \text{ is a fixed point of } \tau) \\ &= x\end{aligned}$$

Hence, the conjugate of a k-cycle is also a k-cycle. We may now explicitly construct a permutation σ to show that both 3-cycles are conjugates. Similarly, we may show that all 2-cycles form a conjugacy class.

Now, we look at conjugates of any permutation in S_n. Suppose $\sigma \in S_n$ and $\sigma = \tau_1 \cdots \tau_k$ where $\tau_1, \ldots, \tau_k$ are disjoint cycles of length at least 2. Let $\gamma \in S_n$. We have

$$\begin{aligned}\gamma\sigma\gamma^{-1} &= \gamma(\tau_1 \cdots \tau_k)\gamma^{-1} \\ &= (\gamma\tau_1\gamma^{-1}) \cdots (\gamma\tau_k\gamma^{-1})\end{aligned}$$

Here $\gamma\tau_i\gamma^{-1}$ is a conjugate of τ_i. Hence, the cycle structure of the conjugate of the permutation remains the same as the original permutation. Moreover, the conjugates of disjoint cycles are also disjoint.

For example, consider S_5 and the permutation

$$\tau = (1)(2)(3\,4\,5)$$

This corresponds to the partition $(3, 1, 1)$ of 5. Moreover, given any $\sigma \in S_5$, the conjugate of τ will be

$$(\sigma(1))(\sigma(2))\left(\sigma(3)\,\sigma(4)\,\sigma(5)\right)$$

Thus, the conjugacy class of τ is precisely the set

$$\left\{(\sigma(1))(\sigma(2))\left(\sigma(3)\,\sigma(4)\,\sigma(5)\right) \mid \sigma \in S_5\right\}$$

Thus, corresponding to every partition of 5, we have a unique conjugacy class of S_5. In general, corresponding to every partition of n, we have a unique conjugacy class of S_n. Thus, the number of conjugacy classes in S_n is $p(n)$, the number of partitions of n.

§§1.6. Odd and Even Permutations

Proposition 1.18. Every permutation can be written as a product of transpositions (2-cycles)

Proof. We first express a k-cycle as a product of 2-cycles. Consider the k-cycle $(a_1 \ \ldots \ a_k)$. Verify that we have

$$(a_1 \ \ldots \ a_k) = (a_1\,a_k)(a_1\,a_{k-1}) \ldots (a_1\,a_3)\,(a_1\,a_2)$$

We have thus shown that every k-cycle is a product of transpositions. We also know that any permutation can be written as a product of disjoint cycles. Hence, the claim follows. □

Definition 1.19. A permutation $\sigma \in S_n$ is called even (odd) if σ is a product of an even (odd) number of transpositions

We must prove that this is indeed a well-defined notion (that is, every permutation must either be an even permutation or an odd permutation). Consider the *Vandermonde* polynomial, defined as

$$P(x_1, \ldots, x_n) = \prod_{1 \le j < i \le n} (x_i - x_j)$$

Consider a permutation $\sigma \in S_n$. We define

$$\sigma P := \prod_{1 \le j < i \le n} (x_{\sigma(i)} - x_{\sigma(j)})$$

If σ is a transposition, then $\sigma P = -P$. Thus, if σ is an even permutation, we see that $\sigma P = P$ whereas if σ is an odd permutation, we have $\sigma P = -P$. Thus, σ must either be even or odd, as determined by its effect on P. Using this idea, we also trivially see that the product of two even permutations is even, the product of two odd permutations is even and the product of an even permutation and an odd permutation is odd. Moreover, the inverse of an even permutation is also an even permutation. The identity permutation is also an even permutation. We thus see that the set of even permutations forms a group by itself! We call this group A_n, the *alternating group of degree* n.

Proposition 1.20. The order of A_n is $n!/2$.[4]

Proof. To show this, we set up a bijection between the set of even permutations and the set of odd permutations. Let $Z = S_n \setminus A_n$. We define a map $\varphi \colon A_n \to Z$, defined by $\varphi(\sigma) = (1\,2)\,\sigma$. This is a one-to-one map from A_n to Z. This is also an onto map since given any $\tau \in Z$, we have $\varphi\left((1\,2)\,\tau\right) = \tau$. Thus, φ is a bijection and hence $|A_n| = |Z|$. But we know that $|S_n| = n! = |A_n| + |Z|$ (since A_n and Z are disjoint). This gives us that $|A_n| = n!/2$. □

Remark 1.21. The group A_5 is fundamental in proving that there exists a quintic polynomial which is not solvable by radicals.

§§1.7. Subgroups and Cyclic Groups

Definition 1.22. Let G be a group. A subset H of G is a subgroup of G if H is non-empty and closed under products and inverses. That is, $x, y \in H$ implies that $x^{-1} \in H$ and $xy \in H$. If H is a subgroup of G, we write $H \le G$.

Example 1.23.

1. A_3 is a subgroup of S_3.
2. $SL_n(\mathbb{R})$, the group of $n \times n$ real matrices with determinant 1, is a subgroup of $GL_n(\mathbb{R})$ (under matrix multiplication).

[4] This assumes $n > 1$. For $n = 1$, the group S_n is the trivial group, and so is A_n. For this case, we have $|A_n| = |S_n| = 1$.

3. $O_n(\mathbb{R})$ is a subgroup of $GL_n(\mathbb{R})$ while $SO_n(\mathbb{R})$ is a subgroup of $O_n(\mathbb{R})$.
4. The set of complex numbers of unity magnitude, denoted as S^1, forms a group under multiplication and is in fact a subgroup of $\mathbb{C}^\times$. The cyclic group generated by ω_n, which we denote as μ_n, is a (finite) subgroup of S^1.

Theorem 1.24 (Subgroup Criterion)**.** A subset H of a group G is a subgroup of G if and only if

1. $H \neq \emptyset$.
2. for all $x, y \in H$, $xy^{-1} \in H$.

Proof. We only prove the converse. Let x be any element of H (such an element exists since $H \neq \emptyset$). We have $xx^{-1} \in H \implies 1 \in H$. For any $h \in H$, we have $1h^{-1} \in H \implies h^{-1} \in H$. Thus, H is closed under inverses. For any $x, y \in H$, we know that $y^{-1} \in H$, and thus, $x(y^{-1})^{-1} \in H \implies xy \in H$. Hence, H is also closed under multiplication. □

Let G be any group and $g \in G$. We define the subgroup *generated* by g to be the smallest subgroup of G containing g. We leave it as an exercise to verify that this is the group $\langle g \rangle := \{1, g^{\pm 1}, g^{\pm 2}, \ldots\}$. Groups generated by a single element are called *cyclic groups*.

Proposition 1.25. Suppose H is a cyclic group generated by x, $H = \langle x \rangle$. If the order of H is infinite, then $H = \left\{1, x^{\pm 1}, x^{\pm 2}, \ldots\right\}$, all of which are distinct elements. If H is of order n, then the order of x is also n and $H = \{1, x, \ldots, x^{n-1}\}$.

Proof. Suppose the order of H is infinite and H is generated by x. All we need to show is that every power of x is distinct. Suppose that $x^m = x^n$ for some $m > n$. This gives us $x^{m-n} = 1$. Let $d = m - n$, giving us $x^d = 1$. If $l \in \mathbb{Z}$, then by the Division Algorithm, $l = dq + r$ where $q \in \mathbb{Z}$ and $0 \leq r < d$. Now

$$x^l = x^{dq+r} = (x^d)^q x^r = x^r$$

Hence, every integral power of x is x^r for some $0 \leq r < d$. Thus, H is finite, which is a contradiction. Hence, $x^m \neq x^n$ for $m \neq n$.

Suppose $H = \langle x \rangle$ and $|H| = n$. Since H is finite, $\{1, x^{\pm 1}, x^{\pm 2}, \ldots\}$ is a finite list. As proved before, $x^d = 1$ for some $d \in \mathbb{N}^+$. Let m be the smallest positive integer such that $x^m = 1$ (such a number exists because of Well-Ordering Property). Thus $|x| = m$. This means that $\{1, x, \ldots, x^{m-1}\}$ is a group, which is precisely the same as H. Equating the number of elements, we get $m = n$. Thus, $|x| = n$. □

Proposition 1.26. Let $H = \langle x \rangle$.

1. If $|x|$ is infinite, then $|x^a|$ is also infinite for any $a \neq 0$.

2. If $|x| = n$, then
$$|x^a| = \frac{n}{\gcd(a, n)}$$

Proof. The first part is rather trivial to prove and is left as an exercise. We now prove the second part. Let $d = (a, n)$. We have
$$(x^a)^{n/d} = (x^n)^{a/d}$$
Since the order of x is n and a/d is an integer, we see that $(x^a)^{n/d} = 1$. It is also not too difficult to show that no $m < n/d$ satisfies $(x^a)^m = 1$. □

Recall Corollary 0.12, which states that any subset of $\mathbb{Z}$ that is closed under inverses and addition must be of the form $n\mathbb{Z}$ for some $n \in \mathbb{Z}$. In other words, any subgroup of $\mathbb{Z}$ is of the form $n\mathbb{Z}$. Notice that $\mathbb{Z}$ is a cyclic group, generated by 1 or -1, and all its subgroups are also cyclic subgroups. This idea extends to all cyclic groups, as stated next.

Proposition 1.27. Suppose H is a cyclic group and $K \leq H$. Then, K is cyclic.

Proof. If K is the trivial subgroup (containing only the identity), it is clearly cyclic. Suppose K is non-trivial. Then, there is a non-zero integer n for which $x^n \in K$. Since K is closed under inverses, $x^{-n} \in K$. Hence, there exists a $d \geq 1$ such that $x^d \in K$. Let d be the smallest such integer (such a d exists because of Well-Ordering Property). We claim that $K = \langle x^d \rangle$. This is easily proven using the division algorithm, and is left as an exercise. □

Theorem 1.28. Let $H = \langle x \rangle$ be a finite cyclic subgroup of order n and $m \in \mathbb{N}^+$. Then, H has a subgroup of order m if and only if $m \mid n$. Moreover, for each divisor m of n, there is exactly one subgroup of order m in H. (Alternatively, there is a one-one correspondence between subgroups of H and divisors of n).

Proof. We know that $|x| = n$ and that every subgroup of H is cyclic. Suppose $K = \langle x^a \rangle$ with $a \geq 1$. Proposition 1.26 tells us that $|x^a| = n/d$ where $d = \gcd(a, n)$. Thus, the subgroup $\langle x^a \rangle$ is mapped to the divisor $\gcd(a, n)$. Conversely, assume that d is a divisor of n. Then,
$$\left|x^{n/d}\right| = \frac{n}{\gcd\left(\frac{n}{d}, n\right)} = \frac{n}{n/d} = d$$
Hence, we map the divisor d to the subgroup generated by $x^{n/d}$. It remains to show that for each divisor d of n, there exists a unique subgroup of order d.

Suppose $K \leq H$ and $|K| = d$. We must show that $K = \langle x^{n/d} \rangle$. Since K is cyclic, there is a $b \in \mathbb{N}^+$ such that $K = \langle x^b \rangle$. This implies that $|K| = n/\gcd(b, n)$. Thus, $\gcd(b, n) = n/d \implies n/d \mid b$. Thus $b = (n/d) \cdot c$ for some $c \in \mathbb{Z}$. Now,
$$x^b = x^{c(n/d)} \in \langle x^{n/d} \rangle$$

Thus, $K \leq \langle x^{n/d} \rangle$. However, order of both these groups are equal to d, which concludes that that the two groups are equal, or $K = \langle x^{n/d} \rangle$. □

Using this theorem, we can in fact find the total number of subgroups of H (total number of divisors of n) and also construct each of these subgroups with the help of the prime-factorisation of n.

Proposition 1.29. Suppose $H = \langle x \rangle$.

1. If H is infinite, then H has only one other generator, x^{-1}.
2. If $|H| = n$ then x^a generates H if and only if $\gcd(a, n) = 1$.

Proof. First consider that H is infinite. We have shown that $H = \{1, x^{\pm 1}, x^{\pm 2}, \ldots\}$, all of which are distinct elements. Suppose x^a also generates H. Then, $H = \{1, x^{\pm a}, x^{\pm 2a}, \ldots\}$. Comparing the two forms, $x = x^{na}$ for some $n \in \mathbb{Z}^*$. This gives us $x^{na-1} = 1$. If $na - 1$ is non-zero then x generates a finite cyclic group, which is a contradiction. Hence, $na = 1$. This gives us precisely the two solutions $a = \pm 1$. Thus, any infinite cyclic group has only two possible generators x or x^{-1}. We leave it to the reader to verify that x^{-1} indeed generates H.

Now suppose H is finite. $|H| = n \implies |x| = n$. We know from Proposition 1.26 that $|x^a| = n/d$ where $d = \gcd(a, n)$. If x^a also generates H then $|x^a| = n$. This gives us $d = 1$. Conversely, suppose that $\gcd(a, n) = 1$. Then, $|x^a| = n$ and hence, x^a generates a subgroup of H of order n. However, the only subgroup of H of order n is H itself. Hence, x^a generates H. This also gives us that the number of generators of H are $\varphi(n)$. □

Example 1.30.

1. Consider the multiplicative group $\mathbb{Z}_2^\times = \left\{\overline{1}, \overline{5}, \overline{7}, \overline{11}\right\}$. This is a group of order 4. However, all its elements have order either 1 or 2. Hence, this group is not cyclic since a cyclic group of order 4 must have an element of order 4.
2. Consider the real Heisenberg group, $H(\mathbb{R})$ which is a subgroup of $SL_3(\mathbb{R})$. Consider a matrix
$$M = \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$
where M is not an identity matrix. We may decompose M as
$$M = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{I} + \underbrace{\begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}}_{N}$$
One may verify that $N^3 = 0$ and
$$N^2 = \begin{bmatrix} 0 & 0 & ac \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Now,

$$M^n = (I + N)^n = I + nN + \binom{n}{2} N^2$$

For this matrix to generate a finite cyclic group, we need $M^n = I$ for some n. This gives us $a = b = c = 0$, which is a contradiction since $M \neq I$. Thus, every non-identity matrix in $H(\mathbb{R})$ generates an infinite cyclic group.

§2. Group Homomorphisms

Definition 2.1. Let $(G, \star)$ and $(H, \diamond)$ be two groups. A map $\varphi\colon G \to H$ is a (group) homomorphism if φ satisfies

$$\varphi(a \star b) = \varphi(a) \diamond \varphi(b) \text{ for all } a, b \in G.$$

This is more compactly written as

$$\varphi(ab) = \varphi(a)\varphi(b)$$

where the product is the "appropriate" group operation.

Definition 2.2. Let G, H be two groups. A map $\varphi\colon G \to H$ is called an isomorphism if φ is a homomorphism and φ is a bijection. In this case, we say that G and H are isomorphic and write $G \cong H$.

Definition 2.3. Let G be a group. An automorphism is an isomorphism φ from G to itself.

Example 2.4. Following are some examples of homomorphisms

1. Let $\mathbb{F}$ be any field[5] and $\det\colon GL_n(\mathbb{F}) \to \mathbb{F}^\times$ be the determinant function. This is an example of a homomorphism since $\det(AB) = \det(A)\det(B)$. $\mathbb{F}^\times$ is the multiplicative group associated with the field $\mathbb{F}$.
2. Consider the symmetric group S_n and define $f\colon S_n \to \{1, -1\}$ by taking $f(\sigma) = \operatorname{sign}(\sigma)$ where $\operatorname{sign}(\sigma)$ is -1 if σ is an odd permutation and 1 if it is an even permutation. f is a homomorphism since $\operatorname{sign}(\sigma\tau) = \operatorname{sign}(\sigma)\operatorname{sign}(\tau)$.
3. Suppose $H \leq G$ where G is any group. The identity map $i\colon H \to G$ is a homomorphism, trivially.
4. Consider $\varphi\colon \mathbb{C}^\times \to \mathbb{R}^\times$ defined as $\varphi(z) = |z|$ for all $z \in \mathbb{C}^\times$. φ is a homomorphism.
5. Consider $\phi\colon \mathbb{R} \to \mathbb{R}^\times$ defined by $\phi(x) = e^x$ for all $x \in \mathbb{R}$. This is also a homomorphism, since $\phi(x+y) = e^{x+y} = e^x e^y = \phi(x)\phi(y)$.
6. Consider the group
$$G = \left\{ \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \mid x \in \mathbb{R} \right\} \leq GL_2(\mathbb{R}).$$
One can verify that the map $\varphi\colon G \to \mathbb{R}$ defined as
$$\varphi\left(\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \right) = x$$
is an isomorphism. Hence $G \cong \mathbb{R}$.

[5] We will define a field later while discussing rings. For now, the reader may assume the field to be $\mathbb{R}$ or $\mathbb{C}$.

Definition 2.5. Let G be a group and $a \in G$. The map $T_a \colon G \to G$ defined as $T_a(g) = ag$ for all $g \in G$, is called translation by a.

T_a is a bijection from G to G but it is *not* a homomorphism, in general. Notice however that every element $a \in G$ gives rise to a permutation of G (assuming G is finite). Let S_G denote the group of permutations of G and define $\varphi \colon G \to S_G$ defined as $\varphi(a) = T_a$. We claim that φ is an injective homomorphism. To show that φ is a homomorphism, we need to show that $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in G$. It is trivial to verify that $T_{ab} = T_a T_b$. We will prove the injectivity of φ soon while proving Cayley's Theorem.

Proposition 2.6. Let G, H be groups and let $\varphi \colon G \to H$ be a homomorphism. Then,

1. $\varphi(1) = 1$,
2. $(\varphi(a))^{-1} = \varphi(a^{-1})$ for all $a \in G$, and
3. the image of G under φ is a subgroup of H, that is,

$$\operatorname{im} \varphi = \{\varphi(a) \mid a \in G\} \leq H$$

Proof. We have

$$\varphi(1 \cdot 1) = \varphi(1) \cdot \varphi(1)$$

Also, since $1 \cdot 1 = 1$, we have

$$\varphi(1) \cdot \varphi(1) = \varphi(1) \implies \varphi(1) = 1$$

For any $a \in G$, we have

$$\varphi(aa^{-1}) = 1 = \varphi(a) \cdot \varphi(a^{-1})$$

This gives us

$$(\varphi(a))^{-1} = \varphi(a^{-1})$$

The proof of the third part is left as an exercise and follows directly from the first two parts. □

Proposition 2.7. Suppose $\varphi \colon G \to H$ is an isomorphism. Then, $\varphi^{-1} \colon H \to G$ is also an isomorphism.

Proof. Suppose $x, y \in H$. We show that $\varphi^{-1}(xy) = \varphi^{-1}(x)\varphi^{-1}(y)$. Let $\varphi(g) = x$ and $\varphi(h) = y$ where $g, h \in G$. Since φ is a homomorphism, we have

$$\varphi(gh) = \varphi(g)\varphi(h) = xy \implies \varphi^{-1}(xy) = gh = \varphi^{-1}(x)\varphi^{-1}(y)$$ □

Definition 2.8. Let G, H be groups and let $\varphi \colon G \to H$ be a homomorphism. The kernel of φ is defined as

$$\ker \varphi := \{g \in G \mid \varphi(g) = 1\}$$

Proposition 2.9. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. Then, $\ker \varphi \leq G$.

Proof. Left as an exercise. □

Proposition 2.10. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. φ is injective if and only if $\ker \varphi = \{1\}$, the trivial subgroup.

Proof. Observe that for all $a, b \in G$

$$\varphi(a) = \varphi(b) \iff \varphi(a) \cdot (\varphi(b))^{-1} = 1 \iff \varphi(ab^{-1}) = 1.$$

If φ is injective, then $\varphi(a) = \varphi(b) \iff a = b$. Hence, $\varphi(ab^{-1}) = 1 \iff a = b$ and $\ker \varphi = \{1\}$. Conversely, suppose $\ker \varphi = \{1\}$. Then, $\varphi(ab^{-1}) = 1 \iff ab^{-1} = 1 \iff a = b$. Thus, $\varphi(a) = \varphi(b) \iff a = b$ and φ is injective. □

Proposition 2.11. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. If φ is injective then $G \cong \operatorname{im} \varphi$.

Proof. Note that $\varphi\colon G \to \operatorname{im} \varphi$ is surjective by definition. If φ is also injective, then φ is a bijection. Moreover, φ is also a homomorphism. Thus, φ is an isomorphism and $G \cong \operatorname{im} \varphi$. □

Theorem 2.12 (Cayley's Theorem)**.** Every group is isomorphic to a subgroup of a permutation group.

Proof. We showed that $\varphi\colon G \to S_G$ defined as $\varphi(a) = T_a$ is a homomorphism. From Proposition 2.6, we have that $\operatorname{im} \varphi \leq S_G$, a permutation group. Now, we have

$$a \in \ker \varphi \implies \varphi(a) = 1 \implies T_a(1) = 1 \implies a = 1$$

Hence, $\ker \varphi = \{1\}$. By Proposition 2.10, φ is injective and hence $G \cong \operatorname{im} \varphi$. Hence, G is isomorphic to a subgroup of a permutation group. □

Example 2.13. Suppose $G = \mathbb{Z}_3$ and define $\varphi\colon G \to S_3$ as above. One can show that $\operatorname{im} \varphi$ contains the identity permutations and both the 3-cycles and hence forms a subgroup of S_3. In fact, $\operatorname{im} \varphi \cong A_3$.

Suppose $\varphi\colon G \to H$ is a homomorphism. We saw that elements of G that map to the identity in H form a subgroup of G, its kernel. We now generalise this idea by looking at what elements in G map to a particular element $h \in H$.

Definition 2.14. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. The fiber of an element $h \in H$ is defined as

$$\varphi^{-1}(h) := \{g \in G \mid \varphi(g) = h\}.$$

Remark 2.15. $\ker \varphi = \varphi^{-1}(1)$.

Definition 2.16. Let $H \leq G$. For $g \in G$, we define the left coset of H by g as

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

Similarly, we define the right coset of H by g as

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

Proposition 2.17. Let $H \leq G$. Then, the number of left and right cosets of H in G are equal. This common value is called the *index of* H *in* G and is denoted as $[G : H]$.

Proposition 2.18. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. Suppose $g \in G$ and $\varphi(g) = h$. Then, the fiber of h is the left coset of $\ker \varphi$ by g.

Proof. Suppose $x \in G$ is such that $\varphi(x) = h = \varphi(g)$. We have

$$\varphi(x) = \varphi(g) \iff \varphi(g^{-1}x) = 1 \iff g^{-1}x \in \ker \varphi \iff x \in g \ker \varphi. \qquad \square$$

Proposition 2.19. Let G be a group and let $H \leq G$. Then, any left or right coset of H in G has the same cardinality as H itself.

Corollary 2.20. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. Given any two $h, h' \in H$, the cardinality of $\varphi^{-1}(h)$ and $\varphi^{-1}(h')$ is the same, and is equal to the cardinality of $\ker \varphi$.

Proposition 2.21. Let $H \leq G$ and $a, b \in G$. Then,

1. $aH = bH \iff b^{-1}a \in H$,
2. either $aH = bH$ or $aH \cap bH = \emptyset$, and
3. in particular, if $b \in aH$, then $aH = bH$.

Proposition 2.22. Suppose $H \leq G$. Then the following two equivalent statements are true.

1. For $a, b \in G$, the relation $a \sim b \iff a \in bH$ is an equivalence relation on G.
2. The left cosets of H, namely gH for $g \in G$, form a partition of G. In other words, G is a disjoint union of left cosets of H.

Proposition 2.23. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. Then, G is a disjoint union of fibers of elements in $\operatorname{im}\varphi$. That is,

$$G = \bigsqcup_{h \in \operatorname{im}\varphi} \varphi^{-1}(h)$$

Theorem 2.24 (Counting Principle)**.** Let G be a group and let $H \leq G$. Then, $|G| = |H| \cdot [G : H]$.

Proof. The proof is straightforward since G can be written as a disjoint union of distinct left cosets of H. □

Corollary 2.25. Let $\varphi\colon G \to H$ be a group homomorphism. Then, $|G| = |\ker\varphi| \cdot |\operatorname{im}\varphi|$.

Another corollary of the counting principle is the following.

Theorem 2.26 (Lagrange's Theorem)**.** Let G be a group and let $H \leq G$. Then, $|H|$ divides $|G|$.

Corollary 2.27. Let G be a finite group of order p where p is a prime number. Then, the only subgroups of G are the trivial subgroup and G itself.

We now provide a second proof of Euler's Theorem using Lagrange's Theorem.

Theorem 2.28 (Euler's Theorem)**.** Let $a, n \in \mathbb{N}^+$ and $(a, n) = 1$. Then, $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Proof. Consider the multiplicative group $\mathbb{Z}_n^\times$ of order $\varphi(n)$. Since $\gcd(a, n) = 1$, $\overline{a} \in \mathbb{Z}_n^\times$. Consider the cyclic group H generated by $\overline{a}$. H is a subgroup of $\mathbb{Z}_n^\times$. By Lagrange's Theorem, $|H|$ divides $\varphi(n)$. But $|H| = |\overline{a}|$. We know that

$$(\overline{a})^{|\overline{a}|} = \overline{1}$$

Now, since $\varphi(n) = |\overline{a}| \cdot m$ for some $m \in \mathbb{Z}$, we have

$$(\overline{a})^{\varphi(n)} = \overline{1} \implies a^{\varphi(n)} \equiv 1 \pmod{n}$$

□

Note that the converse of Lagrange's Theorem is not true. That is, if G is a group of order n and d is a divisor of n, then there need not exist a subgroup $H \leq G$ of order d. Consider the group A_4, of order 12.

Claim: There exists no subgroup of A_4 of order 6.

Proof. Suppose there is a subgroup $H \leq A_4$ of order 6. By the counting principle, H will have 2 left cosets. Then, $A_4 = H \cup \sigma H$ where $\sigma \in A_4$ and $\sigma \notin H$. A_4 consists of the identity permutation, 8 3-cycles and 3 elements which are products of 2 disjoint 2-cycles. If τ is a 3-cycle in A_4 then $\tau^3 = 1$ and hence $\tau = \tau^4 = (\tau^2)^2$. Thus, if $\tau \in A_4$, then $\tau^2 \in H$, since the square of every permutation in A_4 is a 3-cycle. If $\tau \in H$ then clearly $\tau^2 \in H$. Pick an element $\tau \in \sigma H$. Then, $\tau = \sigma h$ for some $h \in H$. We then have $\tau^2 = \sigma h \sigma h$. Notice that G can also be written as a disjoint union of right cosets. We then get

$$G = H \cup \sigma H = H \cup H\sigma \implies \sigma H = H\sigma.$$

Now, $h\sigma = \sigma h'$ for some $h' \in H$. We thus get $\tau^2 = \sigma^2 h' h \in \sigma H$. Thus, $\sigma^2 h' h \sigma \tilde{h}$ for some $\tilde{h} \in H$. However, this gives us $\sigma h' h = \tilde{h} \implies \sigma \in H$ which is a contradiction. □

Theorem 2.29. Suppose $K \leq H \leq G$ and G is finite. Then, $[G : K] = [G : H] \cdot [H : K]$.

Proof. Suppose $[G : H] = r$. Then,

$$G = \bigsqcup_{i=1}^{r} g_i H \text{ for some } g_1, \ldots, g_r \in G.$$

Suppose $[H : K] = s$. Then,

$$H = \bigsqcup_{j=1}^{s} h_j K \text{ for some } h_1, \ldots, h_s \in H.$$

It is left as an exercise to show that cosets of the form $g_i h_j K$ are disjoint and to show that these cosets exhaust G. Thus,

$$G = \bigsqcup_{i=1}^{r} \bigsqcup_{j=1}^{s} g_i h_j K \implies [G : K] = r \cdot s$$ □

Definition 2.30. Let G be a group $a \in G$. The map $\gamma_a : G \to G$ with $\gamma_a(g) = aga^{-1}$ for all $g \in G$ is an isomorphism and is called conjugation by a or an inner automorphism of G.

Definition 2.31. Let G be a group. Then, the set

$$\operatorname{Aut} G := \{\varphi : G \to G \mid \varphi \text{ is an automorphism}\}$$

forms a group under composition of maps and is called the automorphism group of G.

Proposition 2.32. Suppose G is *the*[6] cyclic group of order n, that is, $G = \mathbb{Z}_n$. Suppose $\varphi\colon G \to G$ is an automorphism. Then, the following are true.

1. $\varphi(\overline{1}) = \overline{m}$ where $\gcd(m, n) = 1$ and $1 \leq m \leq n-1$.
2. The map $\psi\colon \operatorname{Aut} G \to \mathbb{Z}_n^\times$ defined by $\psi(\varphi) = \varphi(\overline{1})$ is an isomorphism.

Definition 2.33. Suppose G is a group and $N \leq G$. For $g \in G$, we define $gNg^{-1} := \{gng^{-1} \mid n \in N\}$. N is said to be a normal subgroup of G if $gNg^{-1} = N$ for all $g \in G$. We denote this as $N \trianglelefteq G$.

Proposition 2.34. Let G, H be groups and let $\varphi\colon G \to H$ be a homomorphism. Then, $\ker \varphi \trianglelefteq G$.

Proposition 2.35. Suppose $H \leq G$ and $[G : H] = 2$. Then, $H \trianglelefteq G$.

Proof. Since $[G : H] = 2$, G is the disjoint union of two distinct left (or right) cosets of H. That is, for some $g \in G \setminus H$, we have $G = H \cup gH$ and $G = H \cup Hg$. Let $gh \in gH$. Since $gh \in G$, we must have either $gh \in H$ or $gh \in Hg$. Note that $gh \in H \implies g \in H$. Since we assumed $g \notin H$, we get that $gh \in Hg \implies gH \subseteq Hg$. A similar argument also shows that $Hg \subseteq gH$, giving us $gH = Hg$. We leave it as an exercise to show that $gH = Hg \implies H \trianglelefteq G$. □

Corollary 2.36. For all $n \in \mathbb{N}^+$, $n \geq 3$, $A_n \trianglelefteq S_n$.

Definition 2.37 (Center)**.** Let G be a group. The center of G is defined as

$$Z(G) := \{z \in G \mid zg = gz \ \forall g \in G\}$$

Proposition 2.38. Let G be a group and let $Z(G)$ denote its center. Then, $Z(G) \trianglelefteq G$.

Proof. Let $x \in G$. Then, $xZ(G)x^{-1} = \{xyx^{-1} \mid y \in Z(G)\}$. Since $y \in Z(G)$, $xyx^{-1} = yxx^{-1} = y$. Thus, $xZ(G)x^{-1} = Z(G)$ for any $x \in G$. □

Proposition 2.39. Let $H \leq G$. Then, the following statements are equivalent.

1. $H \trianglelefteq G$.

[6] It turns out that up to isomorphism, there is only one cyclic group of order n and only one infinite cyclic group. This is listed as an exercise.

2. $gH = Hg$.
3. Each left coset of H is some right coset of H.

Proof. We only prove that (2) $\iff$ (3) as the rest of the implications are easy to verify. Suppose gH is some left coset of H and $gH = Ha$ for some $a \in G$. Now, $g \in gH \implies g \in Ha$. However, we know that $g \in Hg$. Since any two right cosets of H are either equal or disjoint, we conclude that $Ha = Hg$ and hence, $gH = Hg$. □

Corollary 2.40. If H is the only subgroup of order d in a group G, then $H \trianglelefteq G$.

Proof. Fix a $g \in G$. We know that $\gamma_g \colon G \to G$ is an isomorphism. Hence, $\left|gHg^{-1}\right| = |H| = d$. But since H is the only subgroup of order d, we conclude that $gHg^{-1} = H$ for all $g \in G$. □

Theorem 2.41 (Correspondence Theorem)**.** Let $\varphi \colon G \to G'$ be a homomorphism. Then, the following are true.

1. $\varphi^{-1}(H') \leq G$ for all subgroups $H' \leq G'$. Here, $\varphi^{-1}(H') := \{g \in G \mid \varphi(g) \in H'\}$.
2. If $H' \trianglelefteq G'$ then $\varphi^{-1}(H') \trianglelefteq G$.
3. If φ is surjective and $\varphi^{-1}(H') \trianglelefteq G$, then $H' \trianglelefteq G'$.
4. If φ is surjective, then there is a one-to-one correspondence between the following sets.

$$\{H \leq G \mid \ker \varphi \leq H\} \longleftrightarrow \{H' \leq G'\}$$

Under this correspondence, $H' \trianglelefteq G' \iff \varphi^{-1}(H') \trianglelefteq G'$.

Proof. Let H be a subgroup of G containing the kernel and let H' be a subgroup of G'. To show a bijection, we need to show that $\varphi^{-1}(\varphi(H)) = H$ and $\varphi(\varphi^{-1}(H')) = H'$. The second statement is trivial, hence we only look at the first statement. It is easy to show that $H \subseteq \varphi^{-1}(\varphi(H))$. Let $x \in \varphi^{-1}(\varphi(H)) \implies \varphi(x) \in \varphi(H)$. Thus, $\varphi(x) = \varphi(h)$ for some $h \in H$. Thus, $\varphi(xh^{-1}) = 1 \implies xh^{-1} \in \ker \varphi \subseteq H$. Thus, $x \in H \implies \varphi^{-1}(\varphi(H)) \subseteq H$, completing the proof.

Let $g \in G$. We need to show that $g\varphi^{-1}(H')g^{-1} \subseteq H'$. Let $x \in \varphi^{-1}(H')$. Then, $\varphi(x) \in H'$. Now,

$$\varphi(gxg^{-1}) = \varphi(g)\varphi(x)\varphi(g^{-1}) \in H'$$

Thus, $gxg^{-1} \in \varphi^{-1}(H')$ for all $x \in \varphi^{-1}(H')$. Thus, $\varphi^{-1}(H') \trianglelefteq G$.

We now show that if φ is surjective and $H \trianglelefteq G$ containing $\ker \varphi$ then $\varphi(H) \trianglelefteq G'$. We look at the conjugate $g'\varphi(x)(g')^{-1}$ where $x \in H$ and $g' \in G'$. Since φ is surjective, $g' = \varphi(g)$ for some $g \in G$. Now,

$$g'\varphi(x)(g')^{-1} = \varphi(g)\varphi(x)(\varphi(g))^{-1} = \varphi(gxg^{-1}) \in \varphi(H)$$ □

Proposition 2.42. Let $\varphi\colon G \to G'$ be a surjective homomorphism and let $H' \leq G'$ and $\varphi^{-1}(H')$ be its inverse image. Let $\delta\colon \varphi^{-1}(H') \to H'$ be the map φ restricted to $\varphi^{-1}(H')$. Then,

1. δ is a surjective homomorphism.
2. $\ker \varphi = \ker \delta$.
3. $\left|\varphi^{-1}(H')\right| = |\ker \varphi| \cdot \left|H'\right|$.

We know the subgroups of S_3 - the trivial subgroup, 3 subgroups of order 3, the alternating group A_3, and S_3 itself. We now use this knowledge to understand a certain class of subgroups of S_4. Let $\mathcal{A} = \{T_1, T_2, T_3\}$ where

$$T_1 = \{(1\,2), (3\,4)\}$$
$$T_2 = \{(1\,3), (2\,4)\}$$
$$T_3 = \{(1\,4), (2\,3)\}$$

Let $\sigma = (1\,2\,3\,4)$. Then, σ gives rise to a permutation φ_σ of $\mathcal{A}$. We have

$$\varphi(\sigma)(T_1) = T_3 \quad \varphi(\sigma)(T_2) = T_2 \quad \varphi(\sigma)(T_3) = T_1$$

Similarly, any permutation in S_4 is mapped to a permutation in S_3. We leave it as an exercise to show that $\varphi\colon S_4 \to S_3$ as defined above forms a surjective homomorphism. One may also verify that

$$V := \{1, (1\,2)(3\,4), (1\,3)(2\,4), (1\,4)(2\,3)\} \subseteq \ker \varphi$$

V forms a subgroup of S_4, called the *Klein-four group* which corresponds to reflections of a square. Note also that $|S_4| = |\ker \varphi| \cdot |S_3| \implies |\ker \varphi| = 4$. Thus, the Klein-four group is in fact the kernel of this homomorphism. Using the correspondence theorem, we can deduce a lot about the subgroups of S_4 containing the Klein-four group. There are exactly 6 such subgroups, each having order equal to a multiple of 4 that divides 24. This gives us $4, 8, 12$ and 24 as the possible orders. The subgroup of order 4 is the Klein-four itself while the subgroup of order 24 is S_4 itself. One can further show that all subgroups of order 8 that contain the Klein-four group arise from subgroups of order 2 in S_3, which are exactly three in number. We hence conclude that there are exactly 3 subgroups of order 8 in S_4 which contain the Klein-four group. Moreover, the three subgroups of order 2 in S_3 are not normal, hence the subgroups of order 8 in S_4 containing the Klein-four group are also not normal. Finally, there exists a unique subgroup of order 12 in S_4 containing the Klein-four group, which turns out to be the alternating group A_4. Note that characterising these subgroups of S_4 was a herculean task to carry out by only looking at S_4. However, the correspondence theorem allows us to map these subgroups to simpler subgroups of a group which we understand.

§3. Direct Products and Quotient Groups

Definition 3.1 (Direct Product)**.** Suppose $G_1, \ldots, G_n$ are groups. We define the direct product of these groups as

$$G_1 \times \ldots \times G_n = \{(g_1, \ldots, g_n) \mid g_i \in G_i \text{ for all } i \in \{1, \ldots, n\}\}$$

with the associated binary operation as

$$(g_1, \ldots, g_n)(h_1, \ldots, h_n) = (g_1 h_1, \ldots, g_n h_n)$$

where each $g_i, h_i \in G_i$ for all $i \in \{1, \ldots, n\}$. The direct product, along with this binary operation, forms a group.

Example 3.2. Let C_2 be the[7] cyclic group of order 2 and let C_3 be the cyclic group of order 3. Suppose $C_2 = \langle x \rangle$ and $C_3 = \langle y \rangle$. Then,

$$C_2 \times C_3 = \left\{(1,1), (1,y), (1,y^2), (x,1), (x,y), (x,y^2)\right\}$$

Notice that $|(x,y)| = 6$. Since $|C_2 \times C_3| = 6$, we see that (x,y) in fact generates $C_2 \times C_3$. That is, $C_2 \times C_3 = \langle (x,y) \rangle$, a cyclic group of order 6. Since there is only one cyclic group of order 6, we conclude that $C_2 \times C_3 \cong C_6$. We leave it as an exercise to explicitly define the isomorphism between these two groups. This is generalised by the following proposition.

Proposition 3.3. Suppose C_m and C_n are cyclic groups of order m and n respectively. Then,

1. $|C_m \times C_n| = mn$.
2. $C_m \times C_n$ is cyclic if and only if $\gcd(m,n) = 1$.
3. if $\gcd(m,n) = 1$, then $C_m \times C_n \cong C_{mn}$.
4. if $x \in C_m$ and $y \in C_n$, then $\left|\langle x,y \rangle\right| = \operatorname{lcm}(|x|, |y|)$.

Definition 3.4 (Inclusion Maps)**.** Let A and B be two groups and let $A \times B$ denote their direct product. Define $i_A : A \to A \times B$ and $i_B : B \to A \times B$ as

$$i_A(a) = (a, 1) \text{ for all } a \in A$$

$$i_B(b) = (1, b) \text{ for all } b \in B$$

We call i_A the inclusion map of A in $A \times B$ and i_B the inclusion map of B in $A \times B$.

[7] Since, up to isomorphism, there is only a single cyclic group of order n, we denote this group as C_n.

Proposition 3.5. Let A, B be two groups and let i_A, i_B be their respective inclusion maps in $A \times B$. Then,

1. i_A and i_B are group homomorphisms.
2. $\operatorname{im} i_A \leq A \times B$ and $\operatorname{im} i_B \leq A \times B$.
3. i_A and i_B are injective.
4. $\operatorname{im} i_A \cong A$ and $\operatorname{im} i_B \cong B$.

Definition 3.6. Let A and B be two groups and let $A \times B$ denote their direct product. Define $P_A : A \times B \to A$ and $P_B : A \times B \to B$ as

$$P_A(a, b) = a$$

$$P_B(a, b) = b$$

We call P_A the projection map of $A \times B$ onto A and P_B the projection map of $A \times B$ onto B.

Proposition 3.7. Let A, B be two groups and let P_A, P_B be their respective projection maps. Then,

1. P_A and P_B are group homomorphisms.
2. $\ker P_A = \operatorname{im} i_B$ and $\ker P_B = \operatorname{im} i_A$.

Corollary 3.8. Let A, B be two groups and let i_A, i_B be their respective inclusion maps in $A \times B$. Then, $\operatorname{im} i_A \trianglelefteq A \times B$ and $\operatorname{im} i_B \trianglelefteq A \times B$.

Proof. We leave the proof as an exercise to the reader. We will later show that The proof follows quite trivially since $\operatorname{im} i_A$ and $\operatorname{im} i_B$ are both the kernels of some group homomorphisms (namely, the projection maps described above). □

Proposition 3.9. Let A, B be two groups and let i_A, i_B be their respective inclusion maps in $A \times B$. Then,

1. $\operatorname{im} i_A \cap \operatorname{im} i_B = \{(1, 1)\}$.
2. $\operatorname{im} i_A \operatorname{im} i_B = A \times B$ where

$$\operatorname{im} i_A \operatorname{im} i_B := \left\{(a, 1)(1, b) \mid (a, 1) \in \operatorname{im} i_A, (1, b) \in \operatorname{im} i_B\right\}$$

3. for any $(a, b) \in A \times B$, the decomposition of (a, b) into a product of two elements in $\operatorname{im} i_A$ and $\operatorname{im} i_B$ is unique.

Definition 3.10 (Internal Direct Product)**.** Let G be a group and $N_1, \ldots, N_t$ be normal subgroups of G. We say that G is an internal direct product of $N_1, \ldots, N_t$ if

1. $G = N_1 \cdots N_t$.
2. Every $g \in G$ has a unique decomposition $g = n_1 \cdots n_t$ where $n_i \in N_i$ for all $i \in \{1, \ldots, t\}$.

Hence, we just showed that $A \times B$ is an internal direct product of $\operatorname{im} i_A$ and $\operatorname{im} i_B$. Note that the definition does not require that these normal subgroups intersect trivially. In fact, this follows from the definition itself.

Proposition 3.11. Let G be a group and $N_1, \ldots, N_t$ be normal subgroups of G. If G is an internal direct product of $N_1, \ldots, N_t$, then $N_i \cap N_j = \{1\}$ for all $i, j \in \{1, \ldots, t\}$ with $i \neq j$.

Proof. We will consider the case that G is an internal direct product of 2 normal subgroups, N_1 and N_2. The idea used next generalises very well to any internal direct product. Suppose that N_1 and N_2 both contain $g \neq 1$. Then, they also contain g^{-1}. Now consider the identity $1 \in G$. We can write

$$1 = 1 \cdot 1 = g \cdot g^{-1}$$

Thus, giving us two distinct ways of writing the same element in G as a product of elements in N_1, N_2. Hence, N_1 and N_2 must intersect trivially. □

Lemma 3.12. Let G be an internal direct product of normal subgroups $N_1, \ldots, N_t$. Then, for all $a \in N_i, b \in N_j$ with $i, j \in \{1, \ldots, t\}$ and $i \neq j$, we have $ab = ba$. That is, elements in distinct N_i's commute.

Proof. Let $a \in N_i$ and $b \in N_j$ and $i \neq j$. Define $h = aba^{-1}b^{-1}$. We can write h as $(aba^{-1})b^{-1}$. The bracketed term is a conjugate of b. Since $b \in N_j$ and N_j is normal, we have $aba^{-1} \in N_j$. Since $b^{-1} \in N_j$, we have $h \in N_j$. Similarly, we can write $h = a(ba^{-1}b)$ and conclude that $h \in N_i$. However, since $i \neq j$, we have $N_i \cap N_j = \{1\}$. Thus, $h = 1$. This gives us $aba^{-1}b^{-1} = 1 \implies ab = ba$. □

Theorem 3.13. Let G be a group and $N_1, \ldots, N_t$ be normal subgroups of G. If G is an internal direct product of $N_1, \ldots, N_t$, then the map $\varphi\colon N_1 \times \ldots \times N_t \to G$ defined by

$$\varphi(n_1, \ldots, n_t) = n_1 \cdots n_t$$

is an isomorphism, and hence $G \cong N_1 \times \ldots \times N_t$.

Proof. We first prove that φ is a homomorphism. Let $(n_1, \ldots, n_t), (m_1, \ldots, m_t) \in N_1 \times \ldots \times N_t$ be two n-tuples. Then,

$$\varphi\left((n_1, \ldots, n_t)(m_1, \ldots, m_t)\right) = \varphi\left(n_1 m_1, \ldots, n_t m_t\right) = n_1 m_1 \cdots n_t m_t$$

We also have

$$\varphi(n_1,\dots,n_t)\varphi(m_1,\dots,m_t) = n_1\cdots n_t m_1\cdots m_t$$

Now, since n_i, m_i all belong to distinct N_i, the above lemma allows us to conclude commutativity of these elements. Thus,

$$\varphi\left((n_1,\dots,n_t)(m_1,\dots,m_t)\right) = \varphi(n_1,\dots,n_t)\varphi(m_1,\dots,m_t)$$

and hence φ is a homomorphism. It is trivial to check that φ is surjective. Also observe that $\ker\varphi = \{(1,\dots,1)\}$ since $1 \in G$ has the unique representation $1\cdots 1$. Thus, φ is also injective and hence a bijection. This proves isomorphism. □

Proposition 3.14. Let G be a group and let H, K be finite subgroups of G. Define

$$HK := \{hk \mid h \in H, k \in K\} = \bigcup_{h\in H} hK$$

Then, the following hold true.

1. $|HK| = \dfrac{|H||K|}{|H\cap K|}$.
2. $HK \leq G \iff HK = KH$.
3. If $K \trianglelefteq G$ then $HK \leq G$.

Proof. We see that HK is a union of left cosets of K taken over elements in H. Hence, to count the number of elements in HK, we only need to count the number of distinct left cosets of K by elements in H. To this end, observe that

$$h_1K = h_2K \iff h_2^{-1}h_1 \in K$$

Since $h_2^{-1}h_1$ must also lie in H, we have

$$h_1K = h_2K \iff h_2^{-1}h_1 \in K\cap H \iff h_1(K\cap H) = h_2(K\cap H)$$

Hence, the number of distinct left cosets of K by elements in H is equal to the number of distinct left cosets of $K\cap H$ in H. However, this is precisely equal to $[H : K\cap H]$. If H, K are finite, this is equal to $\frac{|H|}{|K\cap H|}$. Each distinct left coset has precisely $|K|$ elements. Thus,

$$|HK| = |K|\cdot[H : K\cap H] = \frac{|H||K|}{|H\cap K|}$$

Now, suppose that HK is a subgroup and let $h \in H$ and $k \in K$. We then have $h = h1 \in HK$ and $k = 1k \in HK$. Since HK is closed under products, we have $kh \in HK$ and thus $KH \subseteq HK$. We also have $(hk)^{-1} \in HK$, so $(hk)^{-1} = xy$ for some $x \in H$, $y \in K$. Thus, $hk = (xy)^{-1} = y^{-1}x^{-1} \in KH$ since $y^{-1} \in K$ and $x^{-1} \in H$. Thus, $HK \subseteq KH$. Hence, if $HK \leq G$ then $HK = KH$.

Conversely, suppose that $HK = KH$. We trivially see that $1 \in HK$. Suppose $a, b \in HK$. Thus, $a = h_1k_1$ and $b = h_2k_2$ for some $h_1, h_2 \in H$ and $k_1, k_2 \in K$. Now, $k_1h_2 \in KH = HK$ and thus $k_1h_2 = hk$ for some $h \in H$ and $k \in K$. We then have

$$ab = h_1k_1h_2k_2 = h_1hkk_2$$

Since H and K are closed under products, $h_1h \in H$ and $kk_2 \in K$, giving us $ab \in HK$. Also,

$$a^{-1} = (h_1k_1)^{-1} = k_1^{-1}h_1^{-1} \in KH = HK$$

Thus, HK is closed under products and inverses and also contains the identity. Thus, $HK \leq G$.

Now, suppose $K \trianglelefteq G$. It suffices to show that $HK = KH$. Let $x \in HK$. Thus, $x = hk$ for some $h \in H$, $k \in K$. We have

$$xh^{-1} = hkh^{-1} \in K \text{ (since } K \text{ is normal)}$$

Thus, $x \in HK \implies x \in KH$ and hence $HK \subseteq KH$. One can similarly show that $KH \subseteq HK$. Thus $HK = KH$ and $HK \leq G$. □

Proposition 3.15. Let G be a group and let $N \trianglelefteq G$. Define $G/N := \{gN \mid g \in G\}$, the set of all left cosets of N. Then, G/N forms a group with binary operation defined as

$$(gN)(hN) = (gh)N \text{ for all } gN, hN \in G/N$$

Proof. We first prove that this binary operation is well-defined. That is, if $gN = g_1N$ and $hN = h_1N$, then $(gh)N = (g_1h_1)N$. $gN = g_1N \implies g_1^{-1}g \in N$ and $hN = h_1N \implies h_1^{-1}h \in N$. Now,

$$(gh)N = (g_1h_1)N \iff (g_1h_1)^{-1}gh \in N \iff h_1^{-1}g_1^{-1}gh \in N$$

Now, $g^{-1}g = n_1$ and $h_1^{-1}h = n_2$ for some $n_1, n_2 \in N$, giving us $g = g_1n_1$ and $h = h_1n_2$. Thus,

$$(gh)N = (g_1h_1)N \iff h_1^{-1}g_1^{-1}g_1n_1h_1n_2 \in N \iff h_1^{-1}n_1h_1n_2 \in N$$

Now $h_1^{-1}n_1h_1 \in N$ since N is normal. Thus, $h_1^{-1}n_1h_1n_2 \in N \implies (gh)N = (g_1h_1)N$.

Now, we prove associativity. Let $g_1N, g_2N, g_3N \in G/N$. We have

$$(g_1Ng_2N)\, g_3N = (g_1g_2N)g_3N = \big((g_1g_2)g_3\big)\, N$$

Since G is associative, $(g_1g_2)g_3 = g_1(g_2g_3)$. Thus,

$$(g_1Ng_2N)\, g_3N = \big(g_1(g_2g_3)\big)\, N = g_1N(g_2g_3N) = g_1N\,(g_2Ng_3N)$$

Since $(gN)N = N(gN) = gN$, N is the element. It is also trivial to check that $(gN)^{-1} = g^{-1}N$. Hence, G/N forms a group. □

Remark 3.16. The group G/N, as defined above, is called the *quotient group of* N *in* G.

Proposition 3.17. Let G be a group and let $N \trianglelefteq G$. Define $\varphi\colon G \to G/N$ with $\varphi(g) = gN$ for all $g \in G$. Then, φ is a surjective group homomorphism.

Proof. Surjectivity is evident. We hence only show that φ is a homomorphism. Given any $g_1, g_2 \in G$, we have

$$\varphi(g_1 g_2) = g_1 g_2 N = (g_1 N)(g_2 N) = \varphi(g_1)\varphi(g_2)$$ □

Proposition 3.18. Every normal subgroup of a group G is the kernel of some group homomorphism.

Proof. Let G be a group and let $N \trianglelefteq G$ be a normal subgroup. We define $\varphi\colon G \to G/N$ with $\varphi(g) = gN$ for all $g \in G$. The identity element of G/N is N. Hence,

$$\ker \varphi = \{g \in G \mid \varphi(g) = N\} = \{g \in G \mid gN = N\} = N$$ □

Proposition 3.19. Let G be a group and let $N \trianglelefteq G$. Then, every subgroup of the quotient group G/N is of the form $H/N = \{hN \mid h \in H\}$ where $N \leq H \leq G$. Conversely, if $N \leq H \leq G$, then $H/N \leq G/N$. Moreover, if $N \leq N' \trianglelefteq G$, then $N'/N \trianglelefteq G/N$.

Proof. This is a direct application of the Correspondence Theorem. □

There is a reason why we are interested in only quotient groups generated by normal subgroups. Such a nice structure will not exist if the subgroup isn't normal. In fact, the binary operation defined above is indeed a well-defined binary operation then the subgroup must be normal, as we now show.

Proposition 3.20. Let G be a group and let $N \leq G$. The operation $\cdot\colon G/N \times G/N \to G/N$ defined by

$$(gN) \cdot (hN) = (gh)N \text{ for all } gN, hN \in G/N$$

is well-defined if and only if $N \trianglelefteq G$.

Proof. We have already shown that if $N \trianglelefteq G$ then the binary operation on G/N is well defined. Now suppose the operation is well-defined. That is, if $gN = g_1N$ and $hN = h_1N$ then $(gh)N = (g_1h_1)N$. Let $n \in N$ and let $g \in G$. We have $nN = 1N$. Thus, $(ng)N = (1g)N = gN$, which gives us $gN = ngN$. Thus, $N = g^{-1}ngN$ and hence $g^{-1}ng \in N$ for all $g \in G$, giving us $N \trianglelefteq G$. □

§4. Semidirect Products

We now study the "semidirect product" of two groups H and K, which is a generalisation of the direct product which relaxes the requirement that both H and K be normal. Suppose G is a group and H, K are subgroups of G such that

1. $H \trianglelefteq G$ (but K is not necessarily normal in G), and
2. $H \cap K = 1$.

By Proposition 3.14, $HK \leq G$. Moreover, every element in HK can be written uniquely as hk for some $h \in H$, $k \in K$. That is, there is a bijection between HK and $H \times K$, given by $hk \mapsto (h, k)$. Here, the group H appears as elements of the form $(h, 1)$, while the group K appears as elements of the form $(1, k)$. Given $h_1k_1, h_2k_2 \in HK$, we have

$$\begin{aligned}(h_1k_1)(h_2k_2) &= h_1k_1h_2(k_1^{-1}k_1)k_2) \\ &= h_1(k_1h_2k_1^{-1})k_1k_2 \\ &= h_3k_3\end{aligned}$$

where $h_3 = h_1(k_1h_2k_1^{-1}) \in H$ since H is normal, and $k_3 \in K$. Since H is normal in G, K acts on H via conjugation, with action defined as $(k, h) \mapsto khk^{-1}$. With this, the product of two elements of HK can be written as

$$(h_1k_1)(h_2k_2) = (h_1\, k_1 \cdot h_2)(k_1k_2)$$

The action of K on H gives rise to a homomorphism of K into $\mathrm{Aut}(H)$. We now use this interpretation to define a group given two groups H and K, and a homomorphism from K to $\mathrm{Aut}(H)$.

Theorem 4.1. Let H and K be groups and let $\varphi\colon K \to \mathrm{Aut}(H)$ be a homomorphism. Let $\cdot$ be the (left) action of K on H determined by φ. Let G be the set of ordered pairs (h, k) with $h \in H$, $k \in K$, and define operation on G as

$$(h_1, k_1)(h_2, k_2) = (h_1\, k_1 \cdot h_2, k_1k_2).$$

1. G is a group under this operation with $|G| = |H| \cdot |K|$.
2. The sets $\{(h, 1) \mid h \in H\}$ and $\{(1, k) \mid k \in K\}$ are subgroups of G and the maps $h \mapsto (h, 1)$ for $h \in H$, and $k \mapsto (1, k)$ for $k \in K$, are isomorphisms of these subgroups with the groups H and K respectively. That is,

$$H \cong \{(h, 1) \mid h \in H\} \text{ and } K \cong \{(1, k) \mid k \in K\}.$$

Identifying H and G with their isomorphic copies as above, the following are true.

3. $H \trianglelefteq G$.
4. $H \cap K = 1$.
5. For all $h \in H, k \in K$, $khk^{-1} = k \cdot h = \varphi(k)(h)$.

Proof. Since we have discussed the motivation for the above, the proof becomes easy.

1. We leave it as a simple exercise to prove that G is a group, with identity $(1,1)$ and $(h,k)^{-1} = (k^{-1}\cdot h^{-1}, k^{-1})$. Moreover, we clearly have $|G| = |H|\cdot|K|$.
2. Let $\tilde{H} := \{(h,1) \mid h \in H\}$ and $\tilde{K} := \{(1,k) \mid k \in K\}$. For all $a,b \in H$ and all $x,y \in K$, we clearly have
$$(a,1)(b,1) = (ab,1) \text{ and } (1,x)(1,y) = (1,xy)$$
which shows that $\tilde{H}$ and $\tilde{K}$ are subgroups of G, and that the maps as defined are isomorphisms.

4 It is clear by definition that $\tilde{H}\cap\tilde{K} = 1$.

5 We have
$$\begin{aligned}(1,k)(h,1)(1,k)^{-1} &= (k\cdot h, k)(1,k^{-1})\\ &= (k\cdot h\, k\cdot 1, kk^{-1})\\ &= (k\cdot h, 1)\end{aligned}$$
Identifying $(h,1)$ with h and $(1,k)$ with k, we get $khk^{-1} = k\cdot h$.

3 We have shown above that $K \leq N_G(H)$. Since $H \leq N_G(H)$ and $G = HK$, it follows that $G = N_G(H)$. Hence, $H \trianglelefteq G$.

□

Definition 4.2. Let H and K be groups and let $\varphi\colon K \to \mathrm{Aut}(H)$ be a homomorphism. The group G described in Theorem 4.1 is called the semidirect product of H and K with respect to φ. We denote this group as $H \rtimes_\varphi K$. When there is no danger of confusion, we simply write $H \rtimes K$.

Proposition 4.3. Let H and K be groups and let $\varphi\colon K \to \mathrm{Aut}(H)$ be a homomorphism. Then, the following are equivalent.

1. The identity map between $H \rtimes K$ and $H \times K$ is an isomorphism.
2. φ is the trivial homomorphism.
3. $K \trianglelefteq H \rtimes K$.

Proof.

1 $\implies$ 2. By the definition of the group operation on $H \rtimes K$, we have
$$(h_1,k_1)(h_2,k_2) = (h_1\, k_1\cdot h_2, k_1k_2)$$
for all $h_1,h_2 \in H$ and $k_1,k_2 \in K$. If the identity map is an isomorphism, then $(h_1,k_1)(h_2,k_2) = (h_1h_2, k_1k_2)$. Thus, we get $k_1\cdot h_2 = h_2$ for all $h_2 \in H, k_1 \in K$, so that φ is the trivial homomorphism.

2 $\implies$ 3. If φ is trivial, then the action of K on H is trivial. Thus, the elements of H commute with K by Theorem 4.1, and H normalises K. Since K normalises itself, we get that $G = HK$ normalises K, so that $K \trianglelefteq H \rtimes K$.

3 $\implies$ 1. If $K \trianglelefteq H \rtimes K$, then both H and K are normal subgroups of $H \rtimes K$. Now, for any $h \in H$, $k \in K$, we have

$$h^{-1}k^{-1}hk = h^{-1}(k^{-1}hk) \in H \text{ since } H \text{ is normal, and}$$
$$h^{-1}k^{-1}hk = (h^{-1}k^{-1}h)k \in K \text{ since } K \text{ is normal.}$$

Since $H \cap K = 1$, it follows that $hk = kh$ for all $h \in H, k \in K$. Thus, the action of K on H is trivial and we get $(h_1, k_1)(h_2, k_2) = (h_1h_2, k_1k_2)$, which completes the proof. □

Example 4.4.

1. The dihedral group D_{2n} can be expressed as the semidirect product of two cyclic groups. In fact, we have $D_{2n} \cong \mathbb{Z}_n \rtimes \mathbb{Z}_2$. Recall that any element in D_{2n} can be written as $r^i s^j$ where i is unique modulo n, and j is unique modulo 2. We leave it to the reader to verify that $(i, j) \mapsto r^i s^j$ is an isomorphism.
2. With the above, we may generalise the dihedral group to infinite order, by considering the semidirect product $\mathbb{Z} \cong \mathbb{Z}_2$. We denote this infinite-order group as D_∞.

Theorem 4.5. Let G be a group and let H, K be subgroups such that

1. $H \trianglelefteq G$, and
2. $H \cap K$.

Let $\varphi\colon K \to \mathrm{Aut}(H)$ be the homomorphism defined by mapping $k \in K$ to the automorphism induced by conjugation by k on H. Then, $HK \cong H \rtimes K$. In particular, if $G = HK$ with H, K satisfying the above two, then G is the semidirect product of H and K.

§5. Isomorphism Theorems

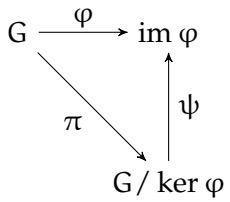
Theorem 5.1 (First Isomorphism Theorem)**.** If $\varphi\colon G \to H$ is a group homomorphism, then $G/\ker\varphi \cong \operatorname{im}\varphi$.

Proof. Let N be $\ker\varphi$ and let $G' = \operatorname{im}\varphi$. We have shown that $N \trianglelefteq G$. Let $\pi\colon G \to G/N$ be the map defined as $\pi(g) = gN$ for all $g \in G$. Now, we define $\psi\colon G/N \to G'$ as $\psi(gN) = \varphi(g)$ for all $gN \in G/N$. We first show that this map is well-defined. For $g, h \in G$, we have

$$gN = hN \iff h^{-1}g \in N \iff \varphi(h^{-1}g) = 1 \iff (\varphi(h))^{-1}\varphi(g) = 1 \iff \varphi(g) = \varphi(h)$$

This not only proves that the map is well-defined but also that it is injective. Moreover, surjectivity of ψ is easy to verify. Hence, ψ is an isomorphism between G/N and G'. Recall that we had defined $N = \ker\varphi$ and $G' = \operatorname{im}\varphi$. Thus, $G/\ker\varphi \cong \operatorname{im}\varphi$. □

Pictorially, we can visualise the first isomorphism theorem as follows.



For example, consider $\varphi\colon S_n \to \{1, -1\}$ defined by $\varphi(\sigma) = \operatorname{sign}\sigma$ for all $\sigma \in S_n$. The kernel of this map is the group of all even permutations, or the alternating group. Thus, $\ker\varphi = A_n$. By the first isomorphism theorem, $S_n/A_n \cong \{1, -1\}$ Note that $\{1, -1\}$ is just the cyclic group of order 2, denoted as C_2. Hence, $S_n/A_n \cong C_2$.

Let V be the Klein-four group in S_4, that is, $V = \{1, (1\,2)(3\,4), (1\,3)(2\,4), (1\,4)(2\,3)\}$. We have seen that $V \trianglelefteq S_4$. After a laborious argument earlier, we were able to show that S_4/V was in fact S_3. But this is achieved rather simply using the first isomorphism theorem. We leave it as an exercise to show that $S_4/V \cong S_3$.

Let $S^1 \leq \mathbb{C}^\times$ be the subgroup of complex numbers with magnitude unity. We define a map $f\colon \mathbb{R} \to S^1$ defined as $f(x) = e^{2\pi i x}$. We leave it as an exercise to show that f is a homomorphism. Notice that $\ker f = \mathbb{Z}$. This shows us that $\mathbb{Z} \trianglelefteq \mathbb{R}$. Moreover, $\mathbb{R}/\mathbb{Z} \cong S^1$.

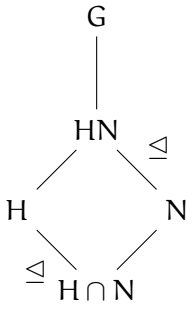
Theorem 5.2 (Second Isomorphism Theorem)**.** Let G be a group. Let $N \trianglelefteq G$ and $H \leq G$. Then, $N \trianglelefteq HN$, $H \cap N \trianglelefteq H$ and $HN/N \cong H/H \cap N$.

Proof. Since N forms a normal subgroup of G, it is also clear that it forms a normal subgroup HN. Let $\pi\colon G \to G/N$ be the 'natural' homomorphism, that is, $\pi(g) = gN$ for all $g \in G$. Using π, we define $\pi_H\colon H \to G/N$ which is the restriction of π to H. It is easy to verify that this restriction is a homomorphism too. We will now show that $\operatorname{im}\pi_H$ is precisely HN/N. We have

$$HN/N = \{hnN \mid h \in H, n \in N\} = \{(hN)(nN) \mid h \in H, n \in N\} = \{hN \mid h \in H\} = \operatorname{im}\pi_H$$

Also, $h \in \ker \pi_H \iff hN = N \iff h \in N$. But, we already know that $h \in H$. Hence, $\ker \pi_H$ is given precisely by $H \cap N$. This shows us that $H \cap N \trianglelefteq H$ since it is the kernel of some group homomorphism. By the First Isomorphism Theorem, we have $H/\ker \pi_H \cong \operatorname{im} \pi_H$ and thus $HN/N \cong H/H \cap N$. □

The second isomorphism theorem is sometimes also called the *Diamond Isomorphism Theorem*, the reason for which should be clear from the following diagram.



Theorem 5.3 (Third Isomorphism Theorem)**.** Let G be a group and let H, N be normal subgroups of G such that $N \leq H$. Then, $(H/N) \trianglelefteq (G/N)$ and $G/H \cong \dfrac{G/N}{H/N}$.

$$\begin{array}{ccccc} G & \xrightarrow{\varphi} & G/N & \xrightarrow{\pi} & \dfrac{G/N}{H/N} \\ \trianglelefteq \big| & & \big| \trianglelefteq & & \\ H & \xrightarrow{\varphi_H} & H/N & & \end{array}$$

Proof. We will use the above diagrammatic representation to prove this theorem. $\varphi\colon G \to G/N$ represents the natural homomorphism $\varphi(g) = gN$. φ_H is the restriction of φ to H. Let $X \in H/N$ and $Y \in G/N$. We have $X = hN$ for some $h \in H$ and $Y = gN$ for some $g \in G$. Now,

$$YXY^{-1} = (gN)(hN)(g^{-1}N) = (ghg^{-1})N$$

Since $H \trianglelefteq G$, we have that $ghg^{-1} = h'$ for some $h' \in H$. Thus, $YXY^{-1} = h'N \in H/N$ and hence, $H/N \trianglelefteq G/N$. Consider the map $\pi \circ \varphi\colon G \to \dfrac{G/N}{H/N}$. Since π and φ are surjective homomorphisms, $\pi \circ \varphi$ is also a surjective homomorphism. We know that $\ker \pi = H/N$. Thus, $\ker(\pi \circ \varphi) = \{g \in G \mid \varphi(g) = H/N\}$. Thus, $\ker(\pi \circ \varphi) = H$. By the First Isomorphism Theorem, we conclude that

$$G/H \cong \frac{G/N}{H/N}$$

□

Theorem 5.4. Let G be a finite abelian group of order n and let $d \in \mathbb{N}^+$ be such that $d \mid n$. Then, G has a subgroup of order d.

Proof. We will apply induction on the order of G. If $n = 1$, the theorem is true. Suppose the theorem is true for all abelian groups of order strictly less than n. We first prove that for any prime p with $p \mid n$, G has an element of order p.

We may assume $|G| > 1$. Suppose $a \in G$ such that $|a| = m \geq 2$. Suppose $p \mid m$. Then, $a^{m/p}$ has order p. Suppose $p \nmid m$. Consider $N = \langle a \rangle$ with $|N| = m$. Since G is abelian, every subgroup of G is normal. Hence, $N \trianglelefteq G$. Now, consider the quotient group G/N. We know that $|G/N| = n/m$. Since $p \mid n$ and $p \nmid m$, we conclude that p divides $|G/N|$. By the induction hypothesis, there exists an element of order p in G/N, since G/N is abelian and of order strictly less than n. Thus, $|bN| = p$ for some $b \in G$. Suppose $|b| = k$. This gives us $b^k = 1 \implies (bN)^k = N$. Thus, $p \mid k$ and we are back to the first case. Hence, there is an element of order p and hence a subgroup of order p.

Fix a prime p such that $p \mid d$. Let $a \in G$ be an element of order p in G and let $N = \langle a \rangle$. Thus, $|N| = p$. We have $|G/N| = n/p < n$. Now, by the induction hypothesis, there exists a subgroup H/N of G/N with $|H/N| = d/p$. Thus, $|H| / |N| = d/p$, which gives us $|H| = d$ and we are done. □

§6. Group Actions

§§6.1. Definitions

Definition 6.1 (Group Action)**.** Let G be a group and let S be a set. A group action of G on S is a map from $G \times S$ to S (denoted as $g \cdot s$ for all $g \in G$, $s \in S$) satisfying

1. $g \cdot (h \cdot s) = (gh) \cdot s$ for all $g, h \in G$, $s \in S$.
2. $1 \cdot s = s$ for all $s \in S$.

In this case, we say that G acts on S or that S is a G-set.

Definition 6.2 (Permutation Representation)**.** Let G be a group and let S be a set. A homomorphism $\varphi \colon G \to S_S$ is called a permutation representation of G on S.

Theorem 6.3. Let G be a group and let S be a set. Define $\sigma_g \colon S \to S$ with $\sigma_g(s) = g \cdot s$ for all $s \in S$. Then, the following is true.

1. σ_g is a bijection, and hence a permutation of S.
2. The map $\varphi \colon G \to S_S$ defined by $\varphi(g) = \sigma_g$ is a permutation representation of G on S.

Proof. To prove the first part, we show that $\sigma_{g^{-1}}$ is the inverse of σ_g. Indeed, we have

$$\sigma_{g^{-1}} \circ \sigma_g(s) = g^{-1} \cdot (g \cdot s) = (g^{-1}g) \cdot s = 1 \cdot s = s \text{ for all } s \in S$$

Similarly, σ_g is the inverse of $\sigma_{g^{-1}}$. This shows that σ_g is a bijection, and hence a permutation of S.

We already know that $1 \cdot s = s$ for all $s \in S$. To prove that φ as defined above is a homomorphism, we see that

$$\sigma_g \circ \sigma_h(s) = g \cdot (h \cdot s) = (gh) \cdot s = \sigma_{gh}(s) \text{ for all } g, h \in G, s \in S$$

Hence, $\varphi(g)\varphi(h) = \varphi(gh)$ for all $g, h \in G$. □

It turns out that the converse of the above theorem is also true.

Proposition 6.4. Let G be a group and let S be a set. Let $\psi \colon G \to S_S$ be a permutation representation of G on S. Then, the map from $G \times S$ to S, defined by

$$g \cdot s = \psi(g)(s) \text{ for all } g \in G, s \in S$$

is a group action of G on S.

Proof. Clearly, $\psi(1)$ is the identity permutation and hence $1 \cdot s = s$ for all $s \in S$. We have

$$g \cdot (h \cdot s) = \psi(g)\big(\psi(h)(s)\big)$$

Since ψ is a group homomorphism, we get

$$g \cdot (h \cdot s) = \psi(gh)(s) = (gh) \cdot s \qquad \square$$

Exercise 6.5.

1. Consider $S = G$. The map $\psi\colon G \times G \to G$ defined by $(g, h) \mapsto gh$, is a group action. The permutation representation induced by this group action is the map from G to S_G defined by $g \mapsto T_g$, where T_g is the translation map, defined by $T_g(h) = gh$. Moreover, the kernel of this permutation representation is trivial, and hence is an injective group homomorphism from G to S_G. Hence, G is isomorphic to a subgroup of S_G, which is precisely Cayley's Theorem (Theorem 2.12).

2. Again, consider $S = G$. The map $\psi\colon G \times G \to G$ defined by $(g, h) \mapsto ghg^{-1}$ is also a group action. The permutation induced by this group action is the map from G to S_G defined by $g \mapsto \gamma_g$, where γ_g is the conjugation map, defined by $\gamma_g(h) = ghg^{-1}$. The kernel of this permutation is the set
$$\left\{ g \in G \mid ghg^{-1} = h \text{ for all } h \in G \right\}$$
which is precisely the center of G, $Z(G)$. By the First Isomorphism Theorem (Theorem 5.1), we get
$$G/Z(G) \cong \left\{ \gamma_g \mid g \in G \right\}.$$
The group on the right is the group of inner automorphisms of G.

3. Let $\mathbb{F}$ be a field. The group $GL_n(\mathbb{F})$ acts on the n-dimensional vector space $V := \mathbb{F}^n$ with the group action naturally defined from $GL_n(\mathbb{F}) \times V \to V$ as
$$(A, u) \mapsto Au \text{ (the matrix product).}$$
In the case that the field is $\mathbb{F}_2$, the vector space $V = \mathbb{F}^2$ has precisely 4 vectors, given by
$$V = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{e_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{e_2}, \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{e_1+e_2} \right\}$$
Now, consider $S = \{e_1, e_2, e_1 + e_2\}$ and consider $G = GL_2(\mathbb{F}_)$ with the group action defined as above. This group action gives rise to a permutation representation from G to S_3 (since S has three elements) defined by $A \mapsto L_A$, where L_A is the *linear map* induced by A, defined by $L_A(u) = Au$. The kernel of this permutation representation is trivial and hence, is an injective group homomorphism. Moreover, since $|GL_2(\mathbb{F}_2)| = |S_3| = 6$, we must have that this homomorphism is also onto, and hence an isomorphism. Thus, $GL_2(\mathbb{F}_2) \cong S_3$.

§§6.2. Orbits and Stabilisers

Definition 6.6. Let G be a group and let S be a set. Let $\cdot\colon G \times S \to S$ be a group action. For a fixed

$s \in S$, we define the orbit of s under this group action as

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

Definition 6.7. Let G be a group and let S be a set. Let $\cdot\colon G \times S \to S$ be a group action. For a fixed $s \in S$, we define the stabiliser of s under this group action as

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

Note that $O_s \subseteq S$ and $G_s \leq G$.

Definition 6.8. Let G be a group and let S be a set. Let $\cdot\colon G \times S \to S$ be a group action. The group is said to act transitively on S (via the action $\cdot$) if $O_s = S$ for all $s \in S$. That is, for every pair $s, t \in S \times S$, there exists $g \in G$ such that $g \cdot s = t$.

Proposition 6.9. Let G be a group and let S be a set. Let $\cdot\colon G \times S \to S$ be a group action. Define a relation $\sim$ on S defined by $s \sim s'$ if $s' = g \cdot s$ for some $g \in G$. Then, $\sim$ is an equivalence relation. In other words, $s' \sim s$ if $s' \in O_s$.

Proof. Note that $s \sim s$ since $s = 1 \cdot s$, hence $\sim$ is reflexive. If $s \sim s'$, then $s' = g \cdot s$ for some $g \in G$. We then have

$$g^{-1} \cdot s' = g^{-1} \cdot g \cdot s = (g^{-1}g) \cdot s = 1 \cdot s = s \implies s \sim s'.$$

Hence, $\sim$ is symmetric. If $s \sim s'$ and $s' \sim s''$, we have that $s' = g \cdot s$ and $s'' = h \cdot s'$ for some $g, h \in G$. Now,

$$s'' = h \cdot s' = h \cdot (g \cdot s) = (hg) \cdot s \implies s \sim s'' \text{ since } hg \in G.$$

Hence, $\sim$ is also transitive, and thus an equivalence relation. □

Corollary 6.10. Let G be a group and let S be a set. Let $\cdot\colon G \times S \to S$ be a group action. Then, S can be written as a disjoint union of orbits.

Proof. This follows trivially from Proposition 6.9 and Proposition 0.9. □

Theorem 6.11 (Orbit-Stabiliser Formula)**.** Let G be a group and let S be a set. Let $\cdot\colon G \times S \to S$ be a group action. Let G/G_s denote the set[8] of cosets of the stabiliser of an element $s \in S$. Then, $\varphi\colon G/G_s \to O_s$ defined by

$$\varphi(gG_s) := g \cdot s$$

is a bijection. In particular, $|O_s| = [G\colon G_s]$, where $[G\colon G_s] := |G/G_s|$ is the *index* of G_s in G. In the case that G is finite, we have $[G\colon G_s] = |G| / |G_s|$.

Proof. We first show that this map is well-defined by noting that

$$gG_s = hG_s \iff h^{-1}g \in G_s \iff (h^{-1}g)\cdot s = s \iff g\cdot s = h\cdot s.$$

Note that this also proves that φ is injective. φ is also trivially surjective, and hence a bijection. □

Proposition 6.12. Let G be a group and let S be a set. Let $\cdot\colon G\times S\to S$ be a group action. Let $s\in S$ and $g\in G$. Then, $G_{g\cdot s} = gG_sg^{-1}$.

Proof. We have

$$h\in G_{g\cdot s} \iff h\cdot(g\cdot s) = g\cdot s \iff (g^{-1}hg)\cdot s = s \iff g^{-1}hg\in G_s \iff h\in gG_sg^{-1}.$$

□

Corollary 6.13. Let G be a group and let S be a set. Let $\cdot\colon G\times S\to S$ be a group action. Then,

$$|S| = \sum_{s_i} \left[G\colon G_{s_i}\right]$$

where the sum runs over one element s_i from each orbit of S.

Definition 6.14. Let G be a group and let $g\in G$. We define the centraliser of g as

$$Z(g) := \left\{h\in G \mid gh = hg\right\}.$$

Moreover, the centraliser $Z(g)$ is a subgroup of G.

Definition 6.15. Let G be a group and let H be a subgroup of G. We define the normaliser of H as

$$N(H) := \left\{g\in G \mid gH = Hg\right\}.$$

Proposition 6.16. Let G be a group and let H be a subgroup of G. Then, $H\trianglelefteq N(H)$ and $N(H)$ is the largest subgroup of G in which H is a normal subgroup.

Theorem 6.17. Let G be a finite group and let $g\in G$. Let $C(g)$ be the conjugacy class of g, and let $Z(g)$ be the centraliser of g. Then,

$$|G| = |C(g)|\cdot|Z(g)|$$

[8] Note that in general G_s need not be a normal subgroup of G. However, we can still talk about its set of cosets.

Proof. Consider G acting on G via conjugation. That is, consider the group action $\cdot : G \times G \to G$, defined by $(h, g) \mapsto hgh^{-1}$. In this case, O_g is clearly the conjugacy class, $C(g)$, of g, and G_g is the centraliser, $Z(g)$, of g. Applying the Orbit-Stabiliser Formula gives us the required result. □

Corollary 6.18. Let G be a finite group. Then,

$$|G| = \sum_g |C(g)|$$

where the sum runs over one element from each conjugacy class of G.

Proof. This follows directly by applying Theorem 6.17 to Corollary 6.13. □

Corollary 6.19 (Class Equation)**.** Let G be a group and let $Z(G)$ be its center. Then,

$$|G| = |Z(G)| + \sum_g \left[G : Z(g)\right]$$

where the sum runs over one element from each conjugacy class that is not in the center.

Proof. The proof trivially follows from Corollary 6.18 once we note that each element of the center $Z(G)$ forms a conjugacy class containing only itself. □

Definition 6.20. Let G be a finite group and let p be a prime. G is called a p-group if $|G| = p^n$ for some $n \in \mathbb{N}^+$.

Theorem 6.21. If G is a p-group, then $|Z(G)| \geq p$. In particular, every p-group has a non-trivial center.

Proof. By the Class Equation of G, we have

$$p^n = |Z(G)| + \sum_g \left[G : Z(g)\right]$$

Note that in the sum to the right, each index is at least 2 (since the sum varies over only non-trivial conjugacy classes). However, $\left[G : Z(g)\right]$ must divide $|G| = p^n$. It follows that each index is a power of p, and hence, $|Z(G)|$ is also a power of p. Hence, $|Z(G)| \geq p$. □

In the case that G is abelian, $Z(G)$ is the entire group G. However, the above theorem is truly powerful for non-abelian groups as it states that every non-abelian p-group has a non-trivial proper normal subgroup, namely, its center.

Theorem 6.22. Let G be a p-group and let $|G| = p^n$. Then, there is a sequence of subgroups H_i such that $|H_i| = p^i$ for $i = 1, \ldots, n$. Moreover,

1. $1 \trianglelefteq H_1 \trianglelefteq \ldots \trianglelefteq H_n$, and
2. $(H_1 \cong) H_1/1, H_2/H_1, \ldots, H_n/H_{n-1}$ are all cyclic groups of order p.

Here, 1 represents the trivial subgroup of G.

Proof. We apply induction on n. If $n = 1$, the theorem is trivially true. Suppose the theorem holds for groups of order p^{n-1} ($n > 1$). Let $x \in Z(G)$ and $x \neq 1$ (such an x exists since the center is non-trivial, by Theorem 6.21). Moreover, $|x| = p^r$ where $r < n$ (why?). We also have $\left|x^{p^{r-1}}\right| = p$. Define $y := x^{p^{r-1}}$ and let $H = \langle y \rangle$. We have $H \trianglelefteq G$. Now, consider the quotient group G/H, of order p^{n-1}. By the induction hypothesis, G/H has a sequence of subgroups as follows.

$$1 \trianglelefteq H_2/H \trianglelefteq H_3/H \trianglelefteq \ldots \trianglelefteq H_n/H = G/H.$$

By the Correspondence Theorem, we may conclude the result. □

Definition 6.23. A simple group is a non-trivial group that has no non-trivial normal subgroups.

Proposition 6.24. The alternating group A_5 is simple.

Proof. We have $|A_5| = 60$. Let the elements

$$(1), \underbrace{(1\,2\,3)}_{\sigma}, \underbrace{(1\,2\,3\,4\,5)}_{\tau}, \underbrace{(1\,2)(3\,4)}_{\alpha}$$

be representative elements of the conjugacy classes in A_5. Clearly, $|C(1)| = 1$. By the Orbit-Stabiliser Formula,

$$|C(\sigma)| = \frac{60}{|Z(\sigma)|}.$$

Now, the elements in the centraliser of σ in A_5 are the powers[9] of σ, of which there are precisely three - (1), σ, and σ^2. Thus, $|C(\sigma)| = 20$. Similarly, $|C(\tau)| = 12$. Since the number of 5-cycles in A_5 is 24, there is another 5-cycle, say γ, that lies outside of $C(\tau)$ and has its own conjugacy class of 12 elements. That is, $|C(\gamma)| = 12$. Elementary combinatorics shows that there are 15 permutations having structure $(1\,2)(3\,4)$. Moreover, all such elements lie in the same conjugacy class (why?). We then have $|C(\alpha)| = 15$, so that the class equation of A_5 becomes

$$60 = 1 + 12 + 12 + 15 + 20.$$

Now, suppose A_5 had a non-trivial normal subgroup, say H. Then, H must be a disjoint union of conjugacy classes of A_5 and must contain the identity (the center). Moreover, $|H|$ must divide 60. However, from the class equation, we can verify that no possible combination of conjugacy classes along with the center has order that is a divisor of 60. Hence, A_5 is a simple group. □

Remark 6.25. The simplicity of A_5 is crucial in proving that there is a quintic polynomial that is not solvable by radicals.

Lemma 6.26. Let $n \geq 3$. Any even permutation in S_n can be written as a product of 3-cycles.

Proof. We leave the proof as an exercise to the reader. The proof should be fairly trivial after noting the following identities.

$$(a\ b)(b\ c) = (a\ b\ c) \text{ and } (a\ b)(c\ d) = (a\ b\ c)(b\ c\ d).$$

□

Theorem 6.27 (Galois)**.** A_n is simple for all $n \geq 5$.

Proof. Let $n \geq 5$. Suppose that there is a non-trivial normal subgroup of A_n, that is, there is a normal subgroup $N \trianglelefteq A_n$ such that $N \neq (1)$ and $N \neq A_n$. Recall that a fixed point of a permutation is a point that is not 'moved' by the permutation. Among all the permutations in $A_n \setminus (1)$, pick a permutation σ that has the maximum number of fixed points. We will show that σ must be a 3-cycle. First, we write σ as a product of disjoint cycles.

$$\sigma = (a_1 \ldots a_k)(b_1 \ldots b_m)\cdots$$

Suppose $k < m$. Then, observe that σ^k is a nontrivial permutation in A_n that has strictly more fixed points than σ, which is a contradiction. Similarly, k cannot be strictly greater than m, and hence, by trichotomy, we conclude that $k = m$. Proceeding this way, we see that σ has to decompose as a product of cycles of equal length, say m. Suppose $m = 2$. Then, σ is a product of transpositions. Since σ is an even permutation, there must be an even number of transpositions in the decomposition. Moreover, since σ is non-trivial, we must have at least one (and hence, at least two) transpositions. Thus,

$$\sigma = (a_1\ a_2)(a_3\ a_4)\cdots(a_{2r-1}\ a_{2r})$$

with $r \geq 2$. Since $n \geq 5$, there exists a $b \neq a_1, a_2, a_3, a_4$. Let $\tau = (a_3\ a_4\ b)$. Define the *commutator* of τ with σ as $\gamma := \tau\sigma\tau^{-1}\sigma^{-1}$. Since $\sigma \in N$ and N is normal, we conclude that $\gamma \in N$. The fixed points of σ are carried over to γ. That is, $\sigma(j) = j \implies \gamma(j) = j$. Moreover, a_1 and a_2, which were not fixed points of σ, have become fixed under γ. Thus, γ has strictly more fixed points than σ, which is a contradiction. Hence, $m \neq 2$ and $m \geq 3$.

We again consider the decomposition

$$(a_1 \ldots a_m)(b_1 \ldots b_m)\cdots$$

where m is now at least 3. Suppose σ is not a 3-cycle. Then, choose distinct $r, s \neq a_1, a_2, a_3$ (this is again possible since $n \geq 5$). Now, consider $\tau = (a_3\ r\ s) \in A_n$ and the commutator $\gamma = \tau\sigma\tau^{-1}\sigma^{-1}$. As before, we have $\gamma \in N$. We may again verify that γ preserves the fixed points of σ, and that

[9] Note that the two cycle $(4\,5)$ also permutes with σ but it is an odd permutation, hence not an element of A_5.

$\gamma(a_2) = a_2$. Hence, γ has strictly more fixed points than σ, which is a contradiction. Hence, σ must be a 3-cycle in which case it generates the entire group A_n by Lemma 6.26, and $N = A_n$. Thus, a nontrivial subgroup N cannot exist, and hence A_n is simple for $n \geq 5$. □

§7. Sylow Theorems

Theorem 7.1 (Cauchy's Theorem)**.** Let G be a finite group and let p be a prime. If p divides the order of G, then G has a subgroup of order p.

Proof. Consider the set

$$S = \left\{(x_1, \ldots, x_p) \in G^p \mid x_1 \cdots x_p = 1\right\}.$$

$$\left(\sigma, (x_1, \ldots, x_p)\right) \longmapsto \left(x_{\sigma(1)}, \ldots, x_{\sigma(p)}\right)$$

for all $\sigma \in H$ and $(x_1, \ldots, x_p) \in S$. Notice that the orbit of the $(1, \ldots, 1)$ is itself. Since S can be written as a disjoint union of orbits, it follows that there is at least another orbit that has only one element (this follows from divisibility considerations, since p divides $|S|$). Thus, there exists a p-tuple $(x_1, \ldots, x_p) \neq (1, \ldots, 1)$ such that $O((x_1, \ldots, x_p)) = \left\{(x_1, \ldots, x_p)\right\}$. Since the orbit of this tuple contains only itself, and since permutations in H cyclically permute the elements of the tuple, it follows that each element of the tuple. That is, such an element of the form $(x, \ldots, x)$ for some $x \in G$ with $x \neq 1$. Since $(x, \ldots, x) \in S$, it follows that $x^p = 1$. It then follows that $|x| = p$, and the cyclic subgroup $\langle x \rangle$ is a subgroup of G of order p. □

Proposition 7.2. Let G be a finite group of order p^n where p is a prime and $n \in \mathbb{N}^+$. Then, there are normal subgroups

$$\{1\} = G_0 < G_1 < \ldots < G_n = G$$

such that $|G_i| = p^i$ and $G_i \trianglelefteq G$ for $i = 0, \ldots, n$.

Proof. We prove this by induction on n. In the case that $n = 1$, we trivially have $G_0 = \{1\}$ and $G_1 = G$. Now, assume that $n \geq 2$ and that the result holds for $n - 1$. By Theorem 6.21, G has a non-trivial center, and p divides $|Z(G)|$. By Cauchy's Theorem, $Z(G)$ has an element of order p, say z. We define $G_1 = \langle z \rangle$. Clearly, $|G_1| = p$. Moreover, since $G_1 \leq Z(G)$ and $Z(G) \trianglelefteq G$, we conclude that $G_1 \trianglelefteq G$. Now, define $H = G/G_1$. We have $|H| = p^n/p = p^{n-1}$. By the induction hypothesis, H has normal subgroups

$$\{1\} = H_0 < H_1 < \ldots < H_{n-1} = H$$

such that $|H_i| = p^i$ and $H_i \trianglelefteq H$ for $i = 0, \ldots, n-1$. By the Correspondence Theorem, there is a normal subgroup G_{i+1} of G such that $G_{i+1}/G_1 = H_i$ for $i = 0, \ldots, n-1$. Moreover, $|G_{i+1}| = |G_1| \cdot |H_i| = p^{i+1}$ for $i = 0, \ldots, n-1$. We also have that $G_i \leq G_{i+1}$ by the Correspondence Theorem. This concludes the proof. □

Definition 7.3. Let G be a group and let $H \leq G$ be a subgroup. If $|H| = p^i$ for a prime p and some positive integer i, then H is called a p-subgroup of G.

Definition 7.4. Let G be a finite group and $|G| = p^n m$ where p is a prime, n is a positive integer, and $\gcd(p, m) = 1$. A subgroup of G having order p^n is called a Sylow p-subgroup of G. The set of all Sylow p-subgroups of G is denoted as $\mathrm{Syl}_p(G)$. The number of Sylow p-subgroups of G is denoted as $n_p := \left|\mathrm{Syl}_p(G)\right|$.

Theorem 7.5 (Sylow Theorems)**.** Let G be a finite group and $|G| = p^n m$ where p is a prime, n is a positive integer, and $\gcd(p, m) = 1$. Then, the following are true.

1. G has subgroups of order p^i for $i = 1, \ldots, n$. In particular, G has a Sylow p-subgroup, that is, $n_p \geq 1$.
2. Any p-subgroup of G is contained in a Sylow p-subgroup of G.
3. Any two Sylow p-subgroups of G are conjugates of each other.
4. $n_p \equiv 1 \pmod{p}$ and $n_p \mid m$.

Proof.

1. When $|G| = 1$, the statement is trivially true. Now, assume that the statement is true for all finite groups of order less than $|G|$. The class equation of G is

$$|G| = |Z(G)| + \sum_g \left[G : Z(g)\right]$$

where the sum runs over one representative from each conjugacy class that is not the center. Suppose that $p \nmid |Z(G)|$. Since p divides $|G|$, there exists a g in the second sum such that $p \nmid |G| / |Z(g)|$. Since p^n divides the order of G, it follows that p^n divides $|Z(g)|$. Note that $Z(g)$ is not the whole group since g is not a central element. Hence, $|Z(g)| < |G|$. From the induction hypothesis, it follows that $Z(g)$ has subgroups of order p^i for $i = 1, \ldots, n$ and in particular, a Sylow p-subgroup. Since $Z(g) \leq G$, this result extends to G as well. Now, if p divides $Z(G)$, then by Cauchy's Theorem, there exists an element $z \in Z(G)$ of order p. Let $H = \langle z \rangle$. Then, $H \trianglelefteq G$. We leave the rest of the proof as an exercise to the reader. The proof follows along similar lines as Proposition 7.2, by considering the quotient group G/H (which has order strictly less than $|G|$), proving the result there and pulling it back to G by the Correspondence Theorem.

2. Let $H' \leq G$ and $\left|H'\right| = p^i$ for some $0 \leq i \leq n$. We must show that H' is contained in a Sylow p-subgroup of G. Suppose H is a Sylow p-subgroup of G. Consider the set

$$S = G/H = \left\{gH \mid g \in G\right\}.$$

We have $|S| = |G| / |H| = m$. Let H' act on S by translation, with action defined from $H \times S$ to S as $(h, gH) \mapsto hgH$. We leave it to the reader to verify that this is indeed a group action. Now, S is a disjoint union of H'-orbits and thus

$$|S| = \sum_h |O_h|$$

where the sum runs over one representative from each orbit. Note that each orbit must have cardinality of the form p^k for $k = 0, \ldots, i$, since $|O|_h$ must divide $|H|'$ which is p^i. However, $|S| = m$ and $p \nmid m$. We hence conclude that there is at least one orbit that has only one element, that is, an orbit consisting of a single left coset, say gH. We thus have

$$\begin{aligned} & hgH = gH && \text{for all } h \in H' \\ \implies & g^{-1}hg \in H && \text{for all } h \in H' \\ \implies & g^{-1}H'g \subseteq H \\ \implies & H' \subseteq gHg^{-1} \end{aligned}$$

Since $\left|gHg^{-1}\right| = p^n$, gHg^{-1} is also a Sylow p-subgroup of G. Hence, H' is contained in a Sylow p-subgroup of G.

3. In the above, if $|H|' = p^n$, then we clearly have $H' = gHg^{-1}$. Hence, any two Sylow p-subgroups are conjugates of each other.

4. As defined earlier, $\mathrm{Syl}_p(G)$ denotes the set of all Sylow p-subgroups of G. G acts on $\mathrm{Syl}_p(G)$ by conjugation, with action from $G \times \mathrm{Syl}_p(G)$ to $\mathrm{Syl}_p(G)$ defined as $(g, H) \mapsto gHg^{-1}$. Since every Sylow p-subgroup is a conjugate of a Sylow p-subgroup, there is only one orbit with respect to this group action, and hence, the group G acts transitively on $\mathrm{Syl}_p(G)$. If P is a Sylow p-subgroup, then $\mathrm{Syl}_p(G) = \left\{gPg^{-1} \mid g \in G\right\}$. By the Orbit-Stabiliser Formula,

$$\left|\mathrm{Syl}_p(G)\right| = [G : G_P] = \frac{|G|}{|G_P|}.$$

Now,

$$G_P = \left\{g \in G \mid gPg^{-1} = P\right\} = N(P).$$

Since $N(P)$ contains P, $|N(P)| \geq p^n$. Now, the orbit-stabiliser formula immediately tells us that $\left|\mathrm{Syl}_p(G)\right| = n_p \mid m$. Now, it remains to show that $n_p \equiv 1 \pmod p$.

Now, we consider the Sylow p-subgroup, P, act on $\mathrm{Syl}_p(G)$ with the same action as defined as above. Now, $P \in \mathrm{Syl}_p(G)$ and $O_P = \{P\}$. Suppose $Q \in \mathrm{Syl}_p(G)$, $Q \neq P$ with $O_Q = \{Q\}$. Now,

$$O_Q = \left\{gQg^{-1} \mid g \in P\right\} = \{O_Q\} \iff gQg^{-1} = Q \text{ for all } g \in P$$

Thus, $P \subseteq N(Q)$, and thus $P \leq N(Q)$. Moreover, $Q \trianglelefteq N(Q)$ by Proposition 6.16. By Proposition 3.14, $PQ \leq N(Q)$. Since $Q \trianglelefteq N(Q)$, we also have $Q \trianglelefteq PQ$. By the Third Isomorphism Theorem,

$$\frac{PQ}{Q} \cong \frac{P}{P \cap Q} \implies \left|\frac{PQ}{Q}\right| = \left|\frac{P}{P \cap Q}\right| = p^i \text{ for some } i.$$

Thus, $|PQ| = p^{i+n}$, which implies that $i = 0$ and $|PQ| = p^n$. Since $P \leq PQ$ and $Q \trianglelefteq PQ$, and $|P| = |Q| = |PQ|$, it follows that $P = Q = PQ$. The main conclusion is that when P acts on $\mathrm{Syl}_p(G)$, there is only one singleton orbit. All other orbits have cardinality p^i for some $i \geq 1$. Since $\mathrm{Syl}_p(G)$ is a disjoint union of these orbits, we may write out the cardinality of $\mathrm{Syl}_p(G)$ (which is n_p) as the sum of cardinalities of all disjoint orbits. Now, 'modding' out by p gives us $n_p \equiv 1 \pmod p$. □

§8. Classification of Groups

§§8.1. Isomorphism Classes of Groups

We now classify all groups of order at most 13 using the Sylow theorems, along with some other specific orders.

Proposition 8.1. Let p be a prime. Up to isomorphism, there is exactly one group of order p, namely the cyclic group C_p.

Proof. We leave the proof as an exercise to the reader. (Hint: Lagrange's Theorem). □

Theorem 8.2. Let p be a prime. Up to isomorphism, there are only two groups of order p^2, namely, the cyclic group C_{p^2}, and the group $C_p \times C_p$.

Proof. Let $|G| = p^2$ and let $Z(G)$ be the center of G. We first show that G must be abelian. By Theorem 6.21, $Z(G)$ is non-trivial. By Lagrange's Theorem, $Z(G)$ must have order p or p^2. If $Z(G)$ has order p^2, then $Z(G) = G$, so that G is abelian. If $Z(G)$ has order p, then $G/Z(G)$ has order p. Hence, $G/Z(G)$ is cyclic and hence abelian. However, this forces G to be abelian, in which case $Z(G) = G$ and $|Z(G)| = p^2$, which is a contradiction. Hence, G is abelian.

Now, suppose there was an element of order p^2 in G. Then, $G \cong C_{p^2}$. If not, then all elements (except identity) have order p. Let x be one such element. Now, choose $y \notin \langle x \rangle$. Both $\langle x \rangle$ and $\langle y \rangle$ are normal subgroups of G, intersecting trivially. By Proposition 3.14, the cardinality of their product matches the cardinality of G. Hence, G is the internal direct product of $\langle x \rangle$ and $\langle y \rangle$, both of which are isomorphic to C_p. Theorem 3.13 now tells us that $G \cong C_p \times C_p$. □

Corollary 8.3. Up to isomorphism, there are only two groups of order 4, namely, the cyclic group C_4, and the Klein-four group V.

Proof. By Theorem 8.2, C_4 and $C_2 \times C_2$ are the only two groups of order 4. Since V is a group of order 4, and V is not isomorphic to C_4 (why?), we conclude that $V \cong C_2 \times C_2$. □

Proposition 8.4. Let G be a group of order mp^n where $m > 1$ and p is prime. If $n_p = 1$, then G is not simple.

Proof. Let H be the *unique* Sylow p-subgroup of G. H is clearly a proper non-trivial subgroup of G since $|H| = p^n$ and $1 < p^n < mp^n$. Moreover, for any $g \in G$, gHg^{-1} is a Sylow p-subgroup of G by the third Sylow theorem. By uniqueness, $gHg^{-1} = H$ for all $g \in G$, so that H is normal. Hence, G is not simple. □

Theorem 8.5. Up to isomorphism, there are only two groups of order 6, namely the cyclic group of order 6, C_6, and the group of permutations S_3.

Proof. Let $|G| = 6 = 2 \cdot 3$. By the first Sylow theorem, G has a Sylow 2-subgroup, say H, and a Sylow 3-subgroup, say K. We will let n_2 denote the number of Sylow 2-subgroups of G, and n_3 denote the number of Sylow 3-subgroups of G. By the fourth Sylow theorem,

$$n_2 \equiv 1 \pmod 2 \text{ and } n_2 \mid 3 \implies n_2 = 1 \text{ or } 3.$$
$$n_3 \equiv 1 \pmod 3 \text{ and } n_3 \mid 2 \implies n_3 = 1.$$

Hence, there is a unique Sylow 3-subgroup of G, which is K, that is a normal subgroup of G by Proposition 8.4. If $n_2 = 1$, then H is also normal in G. Moreover, $|H| = 2$ and $H = C_2$, and $|K| = 3$ and $K = C_3$. Note that H and K are two normal subgroups of G that intersect only in identity. By Proposition 3.14, we have

$$|HK| = \frac{|H| \cdot |K|}{|H \cap K|} = \frac{2 \cdot 3}{1} = 6.$$

Again, from Proposition 3.14, we know that $HK \leq G$. However, $|HK| = |G|$ and hence $HK = G$. Thus, G is an internal direct product of H and K. But H and K are the unique cyclic groups C_2 and C_3. Hence, $G = C_2C_3$. By Theorem 3.13, $G \cong C_2 \times C_3$. Now, since 2 and 3 are coprime, by Proposition 3.3 and Example 3.2 in particular, we get that $G \cong C_6$.

Now, consider that $n_2 = 3$, that is, there are 3 Sylow 2-subgroups of G, say H_1, H_2 and H_3. Let $\mathrm{Syl}_2(G) = \{H_1, H_2, H_3\}$ and let G act on $\mathrm{Syl}_2(G)$ by conjugation, with action defined from $G \times \mathrm{Syl}_2(G)$ to $\mathrm{Syl}_2(G)$ as

$$(g, H) \longmapsto gHg^{-1} \text{ for all } g \in G, H \in \mathrm{Syl}_2(G).$$

For a fixed $g \in G$, $\gamma_g \colon \mathrm{Syl}_2(G) \to \mathrm{Syl}_2(G)$ defined by $\gamma_g(H) = gHg^{-1}$ defines a permutation of $\mathrm{Syl}_2(G)$, a set with 3 elements. Now, construct the natural homomorphism $\varphi \colon G \to S_3$ with $\varphi(g) = \gamma_g$. Now,

$$\ker \varphi = \left\{ g \in G \mid gH_ig^{-1} = H_i \text{ for all } H_i \in \mathrm{Syl}_2(G) \right\} = N(H_1) \cap N(H_2) \cap N(H_3).$$

We leave it as an exercise to show that $\ker \varphi = \{1\}$. Hence, φ is injective. Moreover, $|G| = |S_3| = 6$. Thus, φ is a bijection, and $G \cong S_3$. □

Theorem 8.6. Let p, q be distinct primes with $p < q$ and let $p \nmid q - 1$. Up to isomorphism, there's only one group of order pq, namely the cyclic group C_{pq}.

Proof. Let $|G| = pq$. By the first Sylow theorem, G has a Sylow p-subgroup, say H, and a Sylow q-subgroup, say K. By the fourth Sylow theorem,

$$n_p \equiv 1 \pmod p \text{ and } n_p \mid q \implies n_p = 1 \text{ (since } p \nmid q - 1).$$
$$n_q \equiv 1 \pmod q \text{ and } n_q \mid p \implies n_q = 1.$$

Thus there is a unique Sylow p-subgroup, H, and a unique Sylow q-subgroup, K. By Proposition 8.4, $H \trianglelefteq G$ and $K \trianglelefteq G$. Since they also intersect only in identity, by Proposition 3.14, we again have

$$|HK| = \frac{|H| \cdot |K|}{|H \cap K|} = pq$$

and $HK \leq G$. Since $|HK| = |G|$, we conclude that $HK = G$. Thus, G is an internal direct product of H and K. But H and K are the unique cyclic groups C_p and C_q. Hence, $G = C_p C_q$. By Theorem 3.13, $G \cong C_p \times C_q$. Since q and p are coprime, by Proposition 3.3, $G \cong C_{pq}$. □

Theorem 8.7. Up to isomorphism, there are only two groups of order 21, namely, the cyclic group C_{21} and the group presented as $\left\langle x, y \mid x^7 = y^3 = 1; yx = x^2y \right\rangle$.

Proof. Let $|G| = 21 = 3 \cdot 7$. By the first Sylow theorem, G has a Sylow 3-subgroup, say H, and a Sylow 7-subgroup, say K. By the fourth Sylow theorem,

$$n_3 \equiv 1 \pmod 3 \text{ and } n_3 \mid 7 \implies n_3 = 1 \text{ or } 7.$$
$$n_7 \equiv 1 \pmod 7 \text{ and } n_7 \mid 3 \implies n_7 = 1.$$

Thus, there is a unique Sylow 7-subgroup, K, of G. By similar reasoning as before, $K \trianglelefteq G$. Again. if $n_3 = 1$, then there is only one Sylow 3-subgroup of G. By similar reasoning as before, we get $G \cong C_{21}$ in this case.

Now let us assume $n_3 = 7$ and let H be a Sylow 3-subgroup of G. Since $K \trianglelefteq G$, it follows from Proposition 3.14 that $HK \leq G$. Since H and K intersect only in identity, we get $HK = G$ by similar reasoning. Now, since $|H| = 3$ and $|K| = 7$ are both prime, these are both isomorphic to cyclic groups of corresponding orders. Thus, $H = \langle y \rangle$ and $K = \langle x \rangle$ where $|y| = 3$ and $|x| = 7$. Now, since $K \trianglelefteq G$, we have

$$yxy^{-1} \in K \implies yxy^{-1} = x^i \text{ for some } i.$$

If $i = 1$, then G becomes abelian. If G were abelian then for any two Sylow 3-subgroups H and H′, we have $gHg^{-1} = H' \implies H = H'$, by commutativity. Hence, if G is abelian, there is a unique Sylow 3-subgroup, but we have assumed $n_3 = 7$. Hence, $i \neq 1$. Now,

$$\begin{aligned} yxy^{-1} = x^i \implies y^2xy^{-2} &= y(yxy^{-1})y^{-1} = yx^iy^{-1} \\ &= (yxy^{-1})^i = x^{i^2}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} y^3xy^{-3} &= y(y^2xy^{-2})y^{-1} = yx^{i^2}y^{-1} \\ &= (yxy^{-1})^{i^2} = x^{i^3}. \end{aligned}$$

Since $y^3 = 1$, we get $x = x^{i^3}$. Since $|x| = 7$, we get $i^3 \equiv 1 \pmod 7$, which has as its solutions $i = 1, 2, 4 \pmod 7$. Since $i \neq 1$, we conclude that $i = 2$ or 4. When $i = 2$, we get the presentation we desired. We now show that the case $i = 4$ boils down to the same case as $i = 2$. In the case that $i = 4$, we have

$$yxy^{-1} = x^4 \implies y^2xy^{-2} = x^{16} = x^2.$$

Note that y^2 is also a generator of H. Hence, replacing y^2 by y reduces the case $i = 4$ to the case $i = 2$.

Note that we are not done with the proof since we have not yet proved that such a group exists! Consider the group $GL_2(\mathbb{F}_7)$, where $\mathbb{F}_7$ is the finite field of 7 elements, namely $\mathbb{Z}_7$. Now consider the elements

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

It is easy to show that $|A| = 7$ and $|B| = 3$ in $GL_2(\mathbb{F}_7)$. We leave it as a simple computational exercise to show that $BA = A^2B$. □

Theorem 8.8. Up to isomorphism, there are exactly five groups of order 12, namely

1. C_{12},
2. $C_2 \times C_6$,
3. A_4,
4. D_{12}, and
5. $\left\langle x, y \mid x^4 = y^3 = 1; xy = y^2x \right\rangle$.

Proof. Let $|G| = 12 = 2^2 \cdot 3$. By the first Sylow theorem, G has a Sylow 2-subgroup, say H, and a Sylow 3-subgroup, say K. Now, H has order 4, and hence, by Corollary 8.3, H is either the cyclic group C_4, or the Klein-four group V. Of course, $K \cong C_3$. Now, by the fourth Sylow theorem,

$$\begin{aligned} n_2 \equiv 1 \pmod 2 \text{ and } n_2 \mid 3 &\implies n_2 = 1 \text{ or } 3. \\ n_3 \equiv 1 \pmod 3 \text{ and } n_3 \mid 4 &\implies n_3 = 1 \text{ or } 4. \end{aligned}$$

We claim that one of H and K has to be normal in G. Suppose K is not normal in G. Then, there are 4 Sylow-3 subgroups, say K_1, K_2, K_3, and K_4. Moreover, each pair intersects only in identity. Hence, we have

$$\left| \bigcup_{i=1}^{4} K_i \right| = 9.$$

Since any of the Sylow 2-subgroups intersect with Sylow 3-subgroups only in identity, and since $|G| = 12$, it follows that there must be exactly one Sylow 2-subgroup of G, which is normal by reasoning as before. Hence, one of H and K will always be normal.

Case 1: Both H and K are normal in G. Of course, H and K intersect only in identity.

In this case, Proposition 3.14 tells us that $G = HK$ since $|HK| = |G|$ and $HK \leq G$. Since both H and K are normal, G is their internal direct product. By Theorem 3.13, $G \cong H \times K$. Thus, we have the following two possibilities.

1. $G \cong C_4 \times C_4 \cong C_{12}$.
2. $G \cong V \times C_3 \cong C_2 \times C_2 \times C_3 \cong C_2 \times C_6$.

Case 2: H is normal in G, but K is not normal.

In this case, there are 4 Sylow 3-subgroups, all of which are conjugate to each other. Let $\mathrm{Syl}_3(G) = \{K_1, K_2, K_3, K_4\}$. Suppose G acts on $\mathrm{Syl}_3(G)$ by conjugation, with action defined as

$$(g, K_i) \longmapsto gK_ig^{-1} \text{ for all } g \in G, K_i \in \mathrm{Syl}_3(G).$$

This gives rise to a permutation representation $\varphi\colon G \to S_4$, with $\varphi(g) = \gamma_g$, where $\gamma_g\colon \mathrm{Syl}_3(G) \to \mathrm{Syl}_3(G)$ is defined as

$$\gamma_g(K_i) = gK_ig^{-1} \text{ for all } g \in G, K_i \in \mathrm{Syl}_3(G).$$

The kernel is given by

$$\ker\varphi = \left\{g \in G \mid gK_ig^{-1} = K_i \text{ for all } i\right\} = \bigcap_{i=1}^{4} N(K_i).$$

Note that since every K_i is conjugate to every K_j, the orbit of each K_i is the entire set $\mathrm{Syl}_3(G)$, which has cardinality 4. The Orbit-Stabiliser Formula now gives us that $|N(K_i)| = 3$ for all i. But, $K_i \subseteq N(K_i)$ and $|K_i| = 3$ for all i. Hence, $N(K_i) = K_i$ for all i. Since K_i's intersect in identity. so do $N(K_i)$'s. Thus, $\ker\varphi$ is identity and φ is injective. Since G has 8 elements of order 3 (2 from each Sylow 3-subgroup), $\operatorname{im}\varphi$ has 8 3-cycles. However, there are exactly 8 3-cycles in the group S_4. Hence, $\operatorname{im}\varphi$ is a subgroup of S_4 that contains all 3-cycles. Moreover, $|\operatorname{im}\varphi| = |A_4| = 12$. Since A_4 is generated by 3-cycles, it follows that $\operatorname{im}\varphi = A_4$. Now, since φ is injective, $G \cong \operatorname{im}\varphi$, by Proposition 2.11. Hence, $G \cong A_4$.

Case 3: K is normal in G, H is not normal in G, and $H \cong C_4$.

Let H act on K via conjugation, with action defined as $(h, k) \mapsto hkh^{-1}$ for all $h \in H, k \in K$. Now, define $\gamma_h\colon K \to K$ with $\gamma_h(k) = hkh^{-1}$ for all $h \in H, k \in K$. Notice that γ_h cannot be the identity map since that would force G to be abelian, which is a contradiction since H is not a normal subgroup of G. Hence, there is only one other possibility for γ_h (why?). Suppose $H = \langle x\rangle$ and $K = \langle y\rangle$. Now, $\gamma_x(y) = y^2$ since γ_x is not the identity map. Hence, $y^2 = xyx^{-1} \implies xy = y^2x$. As before, we must show that there is indeed such a group. That is, we must find a group G which is presented as

$$G = \left\langle x, y \mid x^4 = y^3 = 1;\ xy = y^2x \right\rangle.$$

Define

$$X = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \text{ and } Y = \begin{bmatrix} \omega & 0 \\ 0 & \omega^2 \end{bmatrix}$$

where $\omega = \exp\left(\frac{2\pi i}{3}\right)$. We leave it as a simple computational exercise to show that these two elements satisfy the give requirements.

Case 4: K is normal in G, H is not normal in G, and $H \cong V$.

Suppose $K = \langle y\rangle$. That is, $K = \{1, y, y^2\}$. Consider the set $S = \{y, y^2\}$ and let H act on S via conjugation. We can restrict the action of H to the set S since the conjugate of y must be y or y^2, and likewise, the conjugate of y^2 must be y or y^2. This is because conjugation preserves order (Proposition 1.16). Now, the stabiliser of y, given by

$$G_y = \left\{h \in H \mid hyh^{-1} = y\right\}$$

can be easily shown to have order 2. Thus, there is a $z \in H$ such that $zyz^{-1} = y$ and $z \neq 1$, and there is an $x \in H$ such that $xyx^{-1} = y^2$. Since H is abelian, we have $xz = zx$. Hence, we have the following presentation for the group.

$$G = \left\langle x, y, z \mid x^2 = y^3 = z^2 = 1;\ xz = zx, yz = zy, xy = y^2x \right\rangle.$$

We leave it as an exercise to show that the dihedral group D_{12} satisfies the above presentations. □

§§8.2. Simplicity of Groups

We now state some important results that allow us to classify several groups on the basis of their simplicity.

Theorem 8.9. Any group with prime order is simple.

Proof. Let p be a prime number and let G be a group of order p. Let H be any subgroup of G. By Lagrange's Theorem, we have either $|H| = 1$ or $|H| = p$. In either case, H is a trivial subgroup of G. Thus, G is simple and has no non-trivial normal subgroups. □

Proposition 8.10. A group G is simple abelian if and only if it is of prime order.

Theorem 8.11. Let p be a prime number and let $n \geq 2$ be an integer. Any group with order p^n is not simple.

Proof. As G is a p-group, it has a non-trivial center, $Z(G)$, by Theorem 6.21. If $Z(G) \neq G$, then $Z(G)$ is a proper non-trivial normal subgroup of G, and hence G is not simple. Now, assume $Z(G) = G$, so that G is abelian. Let $x \in G$ with $x \neq 1$. We have $|x| = p^m$ for some $1 \leq m \leq n$. Define $y := x^{p^{m-1}}$, so that $|y| = p$. Now, $\langle y \rangle$ is a proper non-trivial subgroup of G which is normal since G is abelian. Hence, G is not simple. □

Theorem 8.12. Let p be a prime number and let m be an integer with $1 < m < p$. Any group with order mp^n is not simple.

Proof. By the fourth Sylow theorem $n_p \equiv 1 \pmod{p}$ so that $n_p = 1 + kp$ for some $k \in \mathbb{N}$. However, $n_p \mid m$ and since $m < p$, this forces $k = 0$. Thus, $n_p = 1$ and there is exactly one Sylow p-subgroup of G, say H. By similar reasoning as before, H is a normal subgroup of G. It is also non-trivial since $|H| = p^n$ and $1 < p^n < mp^n$. Hence, G is not simple. □

Theorem 8.13. Let p and q be distinct prime numbers. Any group with order p^2q is not simple.

Proof. Let G be a group of order p^2q. We show that G is not simple.

Case 1: $p > q$.

By the fourth Sylow theorem, $n_p \mid q$ and $n_p \equiv 1 \pmod{p}$. The first condition gives us $n_p = 1$ or $n_p = q$. Since $q < p$, $q \not\equiv 1 \pmod{p}$. Thus, $n_p = 1$, and G is not simple.

Case 2: $p < q$.

Again, by the fourth Sylow theorem, we have $n_q \in \{1, p, p^2\}$. If $n_q = 1$, we are done. As before, $n_q \neq p$ since $p < q$. Now, assume that $n_q = p^2$. Thus, there are exactly p^2 Sylow q-subgroups of G. Moreover, each pair of Sylow q-subgroups intersects only in identity since each has order q, a prime. Hence, these p^2 Sylow q-subgroups capture exactly $p^2(q-1)$ non-identity elements of G. Since the remaining p^2 elements (barring identity) cannot be part of any Sylow q-subgroup, and since $n_p \geq 1$, we conclude that these remaining p^2 elements form a unique Sylow p-subgroup of G, giving us $n_p = 1$. Thus, G is not simple. □

Theorem 8.14. Let p, q, r be distinct prime numbers. Any group with order pqr is not simple.

Proof. We may assume without loss of generality that $p < q < r$. Let G be a group of order pqr. If any of n_p, n_q or n_r are 1, we know that G is not simple. Assume now that each of the above is strictly greater than 1. Now, $n_r \mid pq$. Since we have assumed $n_r > 1$, and since $p, q < r$, we conclude that $n_r = pq$. Thus, we have pq Sylow r-subgroups, that intersect pairwise in identity (since each has prime order). Thus, the number of elements having order r is $o_r = pq(r-1)$. Now, $n_q > 1$ and $n_q \mid pr$ gives us $n_q \in \{p, r, pr\}$. Since $p < q$, we conclude that $n_q \neq p$ and hence $n_q \geq r$. Thus, $o_q \geq r(q-1)$. Similarly, $o_p \geq p(q-1)$.

Since o_r, o_q, and o_p are counting distinct non-identity elements of G, we have

$$\begin{aligned} &|G| \geq o_r + o_q + o_p + 1 \geq pq(r-1) + r(q-1) + q(p-1) \\ &\implies |G| \geq pqr + \underbrace{(r-1)(q-1)}_{>0} > pqr. \end{aligned}$$

Thus, we arrive at a contradiction since $|G| = pqr$, which concludes the proof. □

Theorem 8.15. Let p be a prime and let n be an integer with $n > 1$. Any group with order $(p+1) \cdot p^n$ is not simple.

Proof. Let G be a group with the given order. By the fourth Sylow theorem, we have $n_p \mid (p+1)$ and $n_p \equiv 1 \pmod{p}$. This gives us $n_p = 1$ or $n_p = p+1$. If $n_p = 1$, we are done. Now, assume $n_p = p+1$. Thus, G has $p+1$ Sylow p-subgroups, so that $\mathrm{Syl}_p(G) = \{P_1, \ldots, P_{p+1}\}$. Let $\varphi \colon G \to S_{p+1}$ be the natural homomorphism induced by the group action. By the third Sylow theorem, G acts on $\mathrm{Syl}_p(G)$ transitively, and hence $\ker \varphi \neq G$. Assume now that the kernel is trivial, in which case φ is injective. Thus, $|\operatorname{im} \varphi| = (p+1) \cdot p^n$. Now, since $\operatorname{im} \varphi \leq S_{p+1}$ and $|S_{p+1}| = (p+1)!$, by Lagrange's Theorem, we must have $(p+1) \cdot p^n \mid (p+1)! \iff p^n \mid p!$ which is a contradiction (why?), since $n > 1$. Hence, the kernel is a proper non-trivial subgroup of G. Since kernels of all homomorphisms are normal, it follows that G is not simple. □

The above theorems are able to classify most groups of order at most 200 on the basis of their simplicity. For an exhaustive list, I urge the reader to refer to Aryaman's website.

§9. Rings and Fields

§§9.1. Definitions

Definition 9.1 (Ring)**.** A ring is a set R together with two binary operations $+$ and $\cdot$ (called addition and multiplication) satisfying the following properties.

1. $(R, +)$ is an abelian group with the identity element with respect to $+$ denoted by 0.
2. Multiplication is associative, that is, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in R$.
3. There is a multiplicative identity in R, denoted by 1, that is, $\exists\, 1 \in R$ such that $1 \cdot a = a \cdot 1 = a$ for all $a \in R$.
4. The distributive laws hold. That is,

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

$$(b + c) \cdot a = b \cdot a + c \cdot a$$

for all $a, b, c \in R$.

We henceforth forgo the use of $\cdot$ and denote multiplication simply by juxtaposition. Moreover, we denote the additive inverse of a as $-a$.

Definition 9.2 (Pseudo-Ring)**.** Let R be a set together with two binary operations $+$ and $\cdot$ (called addition and multiplication). If R satisfies only the ring axioms $1, 2$ and 4, we call R a pseudo-ring or a non-unital ring or a rng ("ring" without "i", the multiplicative identity).

Proposition 9.3. Let R be a ring. Then, the following is true.

1. $0a = a0 = 0$ for all $a \in R$.
2. $(-a)b = a(-b) = -(ab)$ for all $a, b \in R$.
3. $(-a)(-b) = ab$ for all $a, b \in R$.
4. The multiplicative identity, 1, is unique and $-a = (-1)a$.

Example 9.4. Some examples of rings:

1. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are all rings with respect to the usual addition and multiplication.
2. For any $n \in \mathbb{N}^+$, $\mathbb{Z}_n$ is a ring with respect to addition and multiplication modulo n.
3. Consider $n \in \mathbb{N}^+$ and let $M_n(\mathbb{R})$ be the set of all $n \times n$ matrices with entries in $\mathbb{R}$. $M_n(\mathbb{R})$ is a ring with respect to the usual addition and matrices.

Definition 9.5 (Commutative Ring)**.** A ring R is said to be commutative if $ab = ba$ for all $a, b \in R$.

Definition 9.6 (Zero Divisor)**.** Let R be a ring. An element $a \in R$ is called a zero divisor if there is a non-zero $b \in R$ such that $ab = 0$ or $ba = 0$.

Definition 9.7 (Unit)**.** Let R be a ring. An element $a \in R$ is called a unit if there is some $b \in R$ such that $ab = ba = 1$. The set of units in R is denoted as $R^\times$. We call b the[10] multiplicative inverse of a and denote it as a^{-1} or $1/a$.

Definition 9.8 (Irreducible Element)**.** Let R be a commutative ring. An element $f \in R$ is said to be irreducible if f is non-zero, non-unit in R and whenever $f = gh$ for some $g, h \in R$, either g is a unit or h is a unit.

Definition 9.9 (Division Ring)**.** Let R be a ring. R is called a division ring or a skew field if $R^\times = R \setminus \{0\}$.

Remark 9.10. A division ring necessarily requires $1 \neq 0$ since $1 = 0 \implies R = \{0\}$. In this case, 0 is indeed a unit and $R^\times = \{0\} \neq \varnothing = R \setminus \{0\}$.

Definition 9.11 (Field)**.** A commutative division ring is called a field.

Proposition 9.12. A field has zero as the only zero divisor.

Definition 9.13 (Domain)**.** A ring R with $1 \neq 0$ is called a domain if it has no non-zero zero divisors.

Definition 9.14 (Integral Domain)**.** A commutative domain is called an integral domain.

Proposition 9.15. If R is an integral domain, then $ab = ac \implies a = 0$ or $b = c$ for all $a, b, c \in R$.

Proof. If $ab = ac$, then $a(b - c) = 0$. If $a = 0$, the result follows trivially. If $a \neq 0$, then a is also not a zero divisor, since R is an integral domain. Hence, $b - c = 0$, giving us $b = c$. □

[10] prove that it is unique.

Proposition 9.16.

1. Every field is an integral domain.
2. Every finite integral domain is a field.

Proof.

1. This follows trivially from Proposition 9.12.
2. We give an outline of the proof. Suppose that the elements of the finite integral domain are $0, a_1, \ldots, a_n$. Fix some non-zero a_i. Now, consider the set

$$\{0a_i, a_1a_i, \ldots, a_na_i\}$$

From Proposition 9.15, it follows that each element of the above set is distinct. Hence, one of them must be equal to 1, the multiplicative identity. Hence, for every non-zero a_i, we have a multiplicative inverse.

□

§§9.2. Polynomial Rings

Definition 9.17 (Polynomial)**.** Let R be a commutative ring and let x be an indeterminate. The formal sum

$$a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0$$

with $n \geq 0$ and each $a_i \in R$ is called a polynomial in x with coefficients in R.

Definition 9.18 (Degree)**.** Let $f(x)$ be a polynomial in x with coefficients in R. The degree of $f(x)$ is defined as

$$\deg f(x) := \max\left\{i \in \mathbb{N} \mid a_i \neq 0\right\}$$

By convention, we define the degree of the *zero* polynomial (one which has all coefficients equal to 0) as $-\infty$.

We denote the set of all polynomials in x with coefficients in R as $R[x]$. We define the addition or sum of two polynomials "componentwise". That is,

$$\begin{aligned}(a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0) + (b_nx^n + b_{n-1}x^{n-1} + \ldots + b_1x + b_0)\\ = (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \ldots + (a_1 + b_1)x + (a_0 + b_0)\end{aligned}$$

For multiplication, we first define $(ax^i)(bx^j) = abx^{i+j}$ for polynomials with only one non-zero term. We then extend this to all polynomials using the distributive laws. That is,

$$\begin{aligned}&(a_0 + a_1x + a_2x^2 + \ldots) \cdot (b_0 + b_1x + b_2x^2 + \ldots)\\ &= (a_0b_0) + (a_1b_0 + a_0b_1)x + (a_2b_0 + a_1b_1 + a_0b_2)x^2 + \ldots\end{aligned}$$

In general, the coefficient of x^k in the product will be $\sum_{i=0}^{k} a_i b_{k-i}$.

Proposition 9.19. Let R be a commutative ring and let $R[x]$ denote the set of all polynomials in x with coefficients in R. Then, under the above defined addition and multiplication, $R[x]$ forms a commutative ring, called the *ring of polynomials in* x *with coefficients in* R.

The ring R itself appears in $R[x]$ as the *constant polynomials*. The multiplicative identity in $R[x]$ is the constant polynomial 1 where 1 is the multiplicative identity in R. For example, $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$ are examples of such rings. The ring $\mathbb{Z}_3[x]$ consists of polynomials in x where coefficients are either $0, 1$ or 2 and addition, multiplication is carried out modulo p. For example, if

$$p(x) = x^2 + 2x + 1 \text{ and } q(x) = x^3 + x + 2$$

then

$$\begin{aligned} p(x) + q(x) &= x^3 + x^2 \\ p(x) \cdot q(x) &= x^5 + 2x^4 + 2x^3 + x^2 + 2x + 2 \end{aligned}$$

Definition 9.20 (Power Series)**.** Let R be a commutative ring and let x be an indeterminate. The formal sum

$$a_0 + a_1 x + a_2 x^2 + \dots$$

with each $a_i \in R$ is called a (formal)[11] power series in x with coefficients in R.

Note that unlike a polynomial, a power series may have infinitely many terms. We can think of both polynomials and power series as a sequence of coefficients. The sequence of coefficients in a polynomial would have to be an eventually zero sequence (or a sequence with finite support), whereas there is no such restriction on the sequence of coefficients for a power series. Moreover, addition and multiplication of two formal power series follow the same pattern as the polynomials.

Proposition 9.21. Let R be a commutative ring and let $R[\![x]\!]$ denote the set of all formal power series in x with coefficients in R. Then, with addition and multiplication as in $R[x]$, $R[\![x]\!]$ forms a commutative ring, called the *ring of formal power series in* x *with coefficients in* R.

Unlike the polynomials, there are non-trivial (non-constant) units in the ring of power series. For example, consider the ring $\mathbb{Z}[\![x]\!]$ and consider $f(x) = 1 - x$. One may verify that the power series $g(x) = 1 + x + x^2 + \dots$ is a multiplicative inverse of $f(x)$, i.e, $f(x)g(x) = 1$. In fact, the following is true.

Proposition 9.22. Let R be a commutative ring and let $f(x) = \sum_{i=0}^{\infty} a_i x^i$ be a formal power series in $R[\![x]\!]$. Then $f(x)$ is a unit in $R[\![x]\!]$ if and only if a_0 is a unit in R.

[11] the term 'formal' signifies that we are only dealing with the 'expression' $a_0 + a_1 x + \dots$ but not actually evaluating it, so we need not worry about convergence.

Proof. Let $f(x)$ be the formal power series $\sum_{i=0}^{\infty} a_i x^i$ and suppose $g(x) = \sum_{j=0}^{\infty} b_j x^j$ be a formal power series that is a multiplicative inverse of $f(x)$. We then have

$$f(x)g(x) = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) x^k = 1$$

On comparing coefficients, we get $a_0 b_0 = 1$. If a_0 is not a unit in R then there does not exist any b_0 in R satisfying the above equation, and hence, such a $g(x)$ does not exist and $f(x)$ is not a unit in $R[\![x]\!]$. If a_0 is invertible in R, then we define $b_0 := a_0^{-1}$. For $k \geq 1$, we have

$$\sum_{i=0}^{k} a_i b_{k-i} = 0 \implies a_0 b_k = -\sum_{i=1}^{k} a_i b_{k-i}$$

Multiplying throughout by b_0, we get

$$b_k = -b_0 \sum_{i=1}^{k} a_i b_{k-i}$$

This allows us to solve for each coefficient b_i by substituting $k = 1, 2, \ldots$ sequentially. Thus, a multiplicative inverse of $f(x)$ exists in $R[\![x]\!]$ and hence $f(x)$ is a unit in this ring. □

We now restrict ourselves to polynomials over fields.

Proposition 9.23 (Division Algorithm)**.** Let $\mathbb{F}$ be a field and let $f(x), g(x) \in \mathbb{F}[x]$ with $g(x) \neq 0$. Then, there are unique polynomials $q(x), r(x) \in \mathbb{F}[x]$ such that

$$f(x) = g(x)q(x) + r(x)$$

and $\deg r(x) < \deg g(x)$.

Proof. We first prove existence. If $\deg f(x) < \deg g(x)$, then taking $q(x) = 0$ and $r(x) = f(x)$ works. Assume that $\deg f(x) \geq \deg g(x)$. We can induct on $n = \deg f(x)$. If $n = 0$, then $\deg g(x) = 0$, since $g(x)$ is non-zero. In this case, $f(x), g(x)$ are constant polynomials. We take $r(x) = 0$ and $q(x) = f(x)/g(x)$. Here $1/g(x)$ represents the multiplicative inverse of $g(x)$, which must exist since $g(x)$ is a non-zero constant polynomial and $\mathbb{F}$ is a field.

Now, suppose $n > 0$ and the result holds for polynomials of degree less than n. Suppose $f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$ and $g(x) = b_m x^m + b_{m-1} x^{m-1} + \ldots + b_1 x + b_0$ where $a_i, b_j \in \mathbb{F}$ with $a_n \neq 0$ and $b_m \neq 0$. Since $b_m \neq 0$, it has a multiplicative inverse. Moreover, $n \geq m$ by assumption. Now, consider

$$f_1(x) = f(x) - \frac{a_n}{b_m} x^{n-m} \cdot g(x)$$

Then, $f_1(x) \in \mathbb{F}[x]$ and $\deg f_1(x) < n$. By the induction hypothesis, there exist $q_1(x), r_1(x) \in \mathbb{F}[x]$ such that $f_1(x) = g(x)q_1(x) + r_1(x)$ and $\deg r_1(x) < m$. Now, define

$$q(x) := \frac{a_n}{b_m} x^{n-m} + q_1(x) \text{ and } r(x) := r_1(x)$$

Verify that these two polynomials satisfy the conditions of the proposition, proving existence.

To prove uniqueness, suppose there are $\tilde{q}(x), \tilde{r}(x) \in \mathbb{F}[x]$ also satisfying the above conditions. We have

$$q(x)g(x) + r(x) = \tilde{q}(x)g(x) + \tilde{r}(x) \implies r(x) - \tilde{r}(x) = g(x)\left(q(x) - \tilde{q}(x)\right)$$

Observe that if $q - \tilde{q}$ is non-zero, then the degree of the RHS is at least $\deg g(x)$ and, if $r - \tilde{r}$ is non-zero, then the degree of the LHS is strictly less than $\deg g(x)$. Hence, equality holds only if $\tilde{r}(x) = r(x)$ and $\tilde{q}(x) = q(x)$, proving uniqueness. □

Remark 9.24. The above result is also valid for an integral domain, provided the leading coefficient of $g(x)$ is a unit.

Definition 9.25 (Root)**.** Let R be a commutative ring and let $f(x) \in R[X]$. An element $\alpha \in R$ is said to be a root of $f(x)$ if $f(\alpha) = 0$.

Theorem 9.26 (Remainder Theorem)**.** Let K be an integral domain and let $\alpha \in K$, $f(x) \in K[x]$. Then, the remainder upon dividing $f(x)$ by $(x - \alpha)$ is $f(\alpha)$.

Proof. Since the polynomial $g(x) = (x - \alpha)$ is monic, the Division Algorithm applies. Thus, there exist unique $q(x), r(x) \in K[x]$ such that

$$f(x) = q(x)(x - \alpha) + r(x)$$

and $\deg r(x) < 1$. Hence, $r(x)$ is a constant polynomial, say r. Substituting $x = \alpha$, we get

$$f(\alpha) = q(\alpha)(\alpha - \alpha) + r \implies f(\alpha) = r$$ □

Theorem 9.27 (Factor Theorem)**.** Let K be an integral domain. Let $\alpha \in K$ and let $f(x) \in K[x]$. Then, α is a root of $f(x)$ if and only if $(x - \alpha)$ is a factor of $f(x)$, that is, $f(x) = (x - \alpha)h(x)$ for some $h(x) \in K[x]$.

Proof. If α is a root of $f(x)$, then $f(\alpha) = 0$. Hence, by the Remainder Theorem, the remainder upon dividing $f(x)$ by $(x - \alpha)$ is 0. Hence, $f(x) = h(x)(x - \alpha)$ for some $h(x) \in K[x]$.

Conversely, suppose that $f(x) = (x - \alpha)h(x)$. Then, the remainder upon diving $f(x)$ by $(x - \alpha)$ is 0. Hence, by the Remainder Theorem, $f(\alpha) = 0$ and α is a root of $f(x)$. □

Corollary 9.28. Let K be an integral domain and let $f(x) \in K[x]$ be a non-zero polynomial of degree n. Then, $f(x)$ has at most n roots in K.

Proof. We leave the proof as an exercise to the reader. (Hint: induction). □

Theorem 9.29 (Fundamental Theorem of Algebra - Version 1)**.** Every non-zero polynomial $f(x) \in \mathbb{C}[x]$ has a root in $\mathbb{C}$.

Theorem 9.30 (Fundamental Theorem of Algebra - Version 2)**.** For $f(x) \in \mathbb{C}[x]$ of degree $n \neq 0$, we can write

$$f(x) = a\prod_{i=1}^{h}(x-\alpha_i)^{e_i}$$

for some $a \in \mathbb{C}$, $a \neq 0$, $h \geq 0$, distinct $\alpha_1, \ldots, \alpha_h \in \mathbb{C}$ and $e_1, \ldots, e_h \in \mathbb{N}^+$. In particular, e_i is the multiplicity of α_i as a root of $f(x)$ and

$$\sum_{i=1}^{h} e_i = n$$

Definition 9.31. Let R be a commutative ring. Given $f(x), g(x) \in R[x]$, we say that $g(x)$ divides $f(x)$ and write $g(x) \mid f(x)$ if $f(x) = g(x)h(x)$ for some $h(x) \in R[x]$.

Definition 9.32 (Irreducible Polynomial)**.** Let R be a commutative ring. A polynomial $f(x) \in R[x]$ is said to be irreducible if $f(x)$ is an irreducible element in $R[x]$.

Proposition 9.33. Let R be an integral domain. $(x-\alpha)$ is irreducible in $R[x]$ for any $\alpha \in R$.

Definition 9.34 (Nilpotent Element)**.** Let R be a commutative ring. An element $a \in R$ is said to be nilpotent if there exists $n \in \mathbb{N}^+$ such that $a^n = 0$.

Proposition 9.35. Let R be a commutative ring and let a, b be nilpotent elements in R. Then,

1. $a + b$ is nilpotent.
2. $ar = ra$ is nilpotent for all $r \in R$. In particular, $-a$ is nilpotent.
3. If $u \in R$ is a unit, then $u - a$ is also a unit.

Proof. We leave the first part as an exercise to the reader (Hint: Binomial theorem). The second part is also trivial. We now prove the third part.

Let $n \in \mathbb{N}^+$ be such that $a^n = 0$. Suppose that $u = 1$. In this case, we have

$$1 = 1 - a^n = (1-a)(a^{n-1} + a^{n-2} + \ldots + a + 1)$$

Thus $1 - a = u - a$ is a unit with multiplicative inverse $(a^{n-1} + a^{n-2} + \ldots + a + 1)$. Now, in

general, suppose u is a unit with $uw = 1$ and suppose a is nilpotent. We then have

$$(u - a) = u(1 - wa)$$

Since a is nilpotent, wa is also nilpotent, by the second part. Thus, by the above argument, $(1 - wa)$ is a unit. Thus, $(u - a)$ is a product of two units and is a unit itself. □

Theorem 9.36. Let R be a commutative ring and let $f(x) = a_n x^n + \ldots + a_0 \in R[x]$. $f(x)$ is a unit in $R[x]$ if and only if a_0 is a unit in R and a_i is nilpotent for each $i > 0$.

Proof. One direction is easy to show. We leave it as an exercise to show that if $a_0 \in R$ is a unit then the constant polynomial a_0 is a unit in $R[x]$. It is also trivial to show that $a_i x^i$ is nilpotent in $R[x]$ if and only if a_i is nilpotent. Thus, if a_0 is a unit and a_i is nilpotent for each $i > 0$, the polynomial $f(x) = a_0 + \ldots + a_n x^n$ is a unit in $R[x]$ since it is the sum of a unit and nilpotent elements. □

Proposition 9.37. An integral domain has no nonzero nilpotent elements.

Corollary 9.38. Let R be an integral domain. Then, the group of units of $R[x]$ is precisely the group of constant polynomials in $R[x]$ which are units in R.

Proof. The proof follows trivially from Theorem 9.36 and Proposition 9.37. □

Proposition 9.39. Let $\mathbb{F}$ be a field and $f(x) \in \mathbb{F}[x]$ have degree d such that $1 \leq d \leq 3$. If $f(x)$ has no root in $\mathbb{F}[x]$, then $f(x)$ is irreducible in $\mathbb{F}[x]$.

Proof. Since $\deg f(x) \geq 1$, we see that $f(x)$ is non-zero and non-unit in $\mathbb{F}[x]$ (Corollary 9.38). Further, if $f(x) = g(x)h(x)$ for some $g(x), h(x) \in \mathbb{F}[x]$ and if both $g(x), h(x)$ are non-units, then $\deg g(x) \geq 1$ and $\deg h(x) \geq 1$. Since $\deg f(x) \leq 3$ and $\deg g(x) + \deg h(x) = \deg f(x)$, we conclude that at least one of $g(x)$ and $h(x)$ must have degree 1. Hence, at least one of them has a root in $\mathbb{F}$. However, this implies that $f(x)$ has a root in $\mathbb{F}$, which is a contradiction. Hence, $f(x)$ is irreducible. □

This allows us to easily conclude that $x^2 + 1$ is irreducible in $\mathbb{R}[x]$. Note that Proposition 9.39 breaks down for degree-4 polynomials. For example, consider the polynomial $f(x) = x^4 + 3x^2 + 2 \in \mathbb{R}[x]$. $f(x)$ clearly does not have a root in $\mathbb{R}$ since $a^4 + 3a^2 + 2 \geq 2 > 0$ for all $a \in \mathbb{R}$. However, we may factorise $f(x)$ as $(x^2 + 1)(x^2 + 2)$, both of which are non-units.

Proposition 9.40. An odd-degree polynomial of degree greater than 1 in $\mathbb{R}[x]$ is reducible.

Proof. This follows from an elementary result in calculus which states that an odd-degree polynomial has a root in $\mathbb{R}$. Once we know that the polynomial has a root in $\mathbb{R}$, we may appeal to the Factor Theorem, to conclude. We leave it to the reader to work out the details. □

Does the same work for $\mathbb{Q}[x]$? That is, is an odd-degree polynomial in $\mathbb{Q}[x]$ of degree greater than 1 always reducible in $\mathbb{Q}[x]$? (Hint: No).

Theorem 9.41 (Fundamental Theorem of Algebra - Version 3)**.** Every non-zero polynomial in $\mathbb{R}[x]$ can be factored as

$$f(x) = a \cdot (x - \alpha_1) \cdots (x - \alpha_r) \cdot q_1(x) \cdots q_s(x)$$

where $a \in \mathbb{R}^\times$, $\alpha_1, \ldots, \alpha_r \in \mathbb{R}$, not necessarily distinct, and $q_1(x), \ldots, q_s(x)$ are monic, quadratic polynomials in $\mathbb{R}[x]$ with negative discriminants. That is, $q_i(x)$ is of the form $x^2 + bx + c$ where $b, c \in \mathbb{R}$ with $b^2 - 4c < 0$, for all $i \in \{1, \ldots, s\}$.

Theorem 9.42 (Fundamental Theorem of Algebra - Version 4)**.** The only monic, irreducible polynomials in $\mathbb{C}[x]$ are $(x - \alpha)$ where $\alpha \in \mathbb{C}$.

Theorem 9.43 (Fundamental Theorem of Algebra - Version 5)**.** The only monic, irreducible polynomials in $\mathbb{R}[x]$ are of the form $(x - \alpha)$ where $\alpha \in \mathbb{R}$, or of the form $x^2 + bx + c$ where $b, c \in \mathbb{R}$ with $b^2 - 4c < 0$.

Theorem 9.44. Let $\mathbb{F}$ be a field and $f(x)$ be a nonzero polynomial in $\mathbb{F}[x]$. Then, $f(x)$ can be factored as

$$f(x) = a \cdot p_1(x) \cdots p_h(x)$$

where $a \in \mathbb{F}^\times$, $h \in \mathbb{N}$ and $p_1(x), \ldots, p_h(x)$ are monic irreducible polynomials in $\mathbb{F}[x]$.

Proof. We leave the proof as an exercise to the reader. The proof follows along similar lines as the proof for Theorem 0.20. □

Definition 9.45 (Greatest Common Divisor)**.** Let R be an integral domain and let $f(x), g(x) \in R[x]$ be such that $f(x)$ and $g(x)$ are not both zero. A polynomial $h(x) \in R[x]$ is said to be a greatest common divisor or gcd of $f(x)$ and $g(x)$ if the following hold.

1. $h(x) \mid f(x)$ and $h(x) \mid g(x)$.
2. If $\tilde{h}(x) \in R[x]$ is such that $\tilde{h}(x) \mid f(x)$ and $\tilde{h}(x) \mid g(x)$, then $\tilde{h}(x) \mid h(x)$.

If $f(x), g(x)$ are both zero, we define the zero polynomial as the gcd of $f(x)$ and $g(x)$. We denote the gcd of $f(x)$ and $g(x)$ as $\big(f(x), g(x)\big)$.

Remark 9.46. Suppose R is an integral domain. If the gcd of $f(x)$ and $g(x)$ exists, then it is unique up to multiplication by a unit in $R[x]$.

Lemma 9.47. If $\mathbb{F}$ is a field, then for any $f(x), g(x) \in \mathbb{F}[x]$, $\big(f(x), g(x)\big)$ exists and moreover, it can be expressed as $u(x)f(x) + v(x)g(x)$ for some $u(x), v(x) \in \mathbb{F}[x]$.

Proof. We leave this as an exercise too. The proof follows along similar lines as Proposition 0.14. □

Corollary 9.48. Let $\mathbb{F}$ be a field and $p(x) \in \mathbb{F}[x]$ be irreducible. If $p(x) \mid f(x)g(x)$ for some $f(x), g(x) \in \mathbb{F}[x]$, then $p(x) \mid f(x)$ or $p(x) \mid g(x)$.

Proof. Again, the proof follows along similar lines as Proposition 0.15 and Corollary 0.17. □

§§9.3. Subrings and Ideals

Definition 9.49 (Subring)**.** Let S be a ring. A subset R of S is said to be a subring of S if R is a ring with respect to addition and multiplication induced from S and R contains 1, the multiplicative identity of S. In this case, we say that S is an overring or a ring extension of R.

Proposition 9.50. Let S be a ring and let $R \subseteq R$. R is a subring of S iff the following properties hold.

1. $1 \in R$.
2. R is closed under addition and subtraction. That is, $a, b \in R \implies a + b \in R$ and $a - b \in R$.
3. R is closed under multiplication. That is, $a, b \in R \implies ab \in R$.

Definition 9.51 (Subfield)**.** If $\mathbb{K}$ is a field and $\mathbb{F}$ is a subring of $\mathbb{K}$ such that $\mathbb{F}$ is also a field, then $\mathbb{F}$ is called a subfield of $\mathbb{K}$. In this case, we say that $\mathbb{K}$ is an overfield or a field extension of $\mathbb{F}$.

Example 9.52.

1. $\mathbb{Z}$ is a subring of $\mathbb{Q}$. $\mathbb{Q}$ is a subring of $\mathbb{R}$. This also tells us that $\mathbb{Z}$ is a subring of $\mathbb{R}$. In general, if S is a subring of R and T is a subring of S, then T is also a subring of R. In other words, the relation "is a subring of", is transitive (we leave the proof as an exercise). In fact, the above examples are also fields themselves. Hence, $\mathbb{Z}$ is a subfield of $\mathbb{Q}$, $\mathbb{Q}$ is a subfield of $\mathbb{R}$ and so on. Equivalently, $\mathbb{Q}$ is a field extension of $\mathbb{Z}$, $\mathbb{R}$ is a field extension of $\mathbb{Q}$ and so on. Naturally, the relation "is a subfield of", is also transitive.
2. $2\mathbb{Z}$ is closed under addition and multiplication, however it has no multiplicative identity. Hence, it is not a subring of $\mathbb{Z}$.

3. $\mathbb{Z}[\iota]$ is a subring of $\mathbb{C}$ where $\mathbb{Z}[\iota]$ is defined as

$$\mathbb{Z}[\iota] := \{m + n\iota \mid m, n \in \mathbb{Z}\}.$$

   $\mathbb{Z}[\iota]$ is called the *ring of Gaussian integers*.

4. $\mathbb{Q}[\sqrt{2}]$ is a subring of $\mathbb{R}$ where $\mathbb{Q}[\sqrt{2}]$ is defined as

$$\mathbb{Q}[\sqrt{2}] := \left\{r + s\sqrt{2} \mid r, s \in \mathbb{Q}\right\}.$$

   It is, in fact, a field. Note that for $r, s \in \mathbb{Q}$, $r + s\sqrt{2} \neq 0 \iff (r, s) \neq (0, 0)$. For $(r, s) \neq (0, 0)$, we leave it as an exercise to verify that

$$\frac{r - s\sqrt{2}}{r^2 - 2s^2} \in \mathbb{Q}[\sqrt{2}]$$

   is a multiplicative inverse of $r + s\sqrt{2}$. Thus, $\mathbb{Q}[\sqrt{2}]$ is a subfield of $\mathbb{R}$. One may also verify that $\mathbb{Q}[\sqrt{2}]$ is a field extension of $\mathbb{Q}$.

Definition 9.53. Given rings $R \subseteq S$, and $\alpha \in S$, we define $R[\alpha]$ to be the smallest subring of S containing α and R.

Given fields $\mathbb{F} \subseteq \mathbb{K}$, and $\alpha \in \mathbb{K}$, we define $\mathbb{F}(\alpha)$ to be the smallest subfield of $\mathbb{K}$ containing α and $\mathbb{F}$.

Similarly, given a set $A \subseteq R$ (or $A \subseteq \mathbb{F}$), we can talk about $R[A]$ (or $\mathbb{F}(A)$) to be the smallest ring (or field) generated by A over R (or $\mathbb{F}$).

Proposition 9.54. Let $\mathbb{F} \subseteq \mathbb{K}$ be field and let $A \subseteq \mathbb{K}$ be a set. If $A = \emptyset$, then $\mathbb{F}(A) = \mathbb{F}$. Assume $A \neq \emptyset$.

Let

$$M := \{a_1 \cdots a_n \mid n \in \mathbb{N}, a_1, \ldots, a_n \in A\}$$

be the set of all finite products of elements of A. Let

$$S := \{b_0 + b_1 m + 1 + \cdots + b_n m_n \mid n \in \mathbb{N}, m_1, \ldots, m_n \in M, b_0, b_1, \ldots, b_n \in \mathbb{F}\}$$

be the set of all finite sums of elements of M. (These are polynomials in A with coefficients in $\mathbb{F}$). Then,

$$\mathbb{F}(A) = \left\{\frac{s_1}{s_2} \mid s_1, s_2 \in S \text{ and } s_2 \neq 0\right\}.$$

Proof. The case $A = \emptyset$ is trivial. Assume $A \neq \emptyset$. Let the set on the RHS of the last equation be Q. Note that M is closed under products, and S is closed under sums and products both. Moreover, S contains $\mathbb{F}$ as the constant polynomials. It is hence clear that Q is a subfield of $\mathbb{K}$. Taking denominator to be 1, we also see that $S \subseteq Q$, and thus Q contains $\mathbb{F}$ as well. Since $A \subseteq M \subseteq S$, Q

also contains A. Thus, $\mathbb{F}(A) \subseteq Q$.

On the other hand, note that $M \subseteq \mathbb{F}(A)$ since $A \subseteq \mathbb{F}(A)$. Since $\mathbb{F} \subseteq \mathbb{F}(A)$, we get $S \subseteq \mathbb{F}(A)$, so that $Q \subseteq \mathbb{F}(A)$. (All these assertions follow from the relevant closure properties of $\mathbb{F}(A)$, a field). Thus, $Q = \mathbb{F}(A)$. □

Corollary 9.55. Let $\mathbb{F} \subseteq \mathbb{K}$ be fields and let $A \subseteq \mathbb{K}$ be a set. If $a \in \mathbb{F}(A)$, then there exists a finite set $B \subseteq A$, such that $a \in \mathbb{F}(B)$.

Proof. Let $a \in \mathbb{F}(A)$ and let M, S be as in Proposition 9.54. Then, $a = s_1/s_2$ for some $s_1, s_2 \in S$. Then, both s_1 and s_2 are polynomials in finitely many $a_i \in A$ with coefficients in $\mathbb{F}$. Let B be the set of those finitely many a_i. Then, $a \in \mathbb{F}(B)$. □

Definition 9.56 (Ideal)**.** A subset I of a ring R is called an ideal[12] of R if

1. I is an additive subgroup of R, or equivalently, $0 \in I$ and $a, b \in I \implies a - b \in I$ (Theorem 1.24).
2. $a \in I$ and $r \in R \implies ra \in I$ and $ar \in I$.

Proposition 9.57. Let R be a commutative ring and let I be an ideal of R. Show that $I \neq R \iff 1 \notin I \iff I$ does not contain any unit.

Proof. This follows trivially from the second condition and is left as an exercise. □

Motivated by Proposition 9.57, we sometimes call R the *unit ideal* of R and any ideal I different from R is called a *non-unit ideal* of R.

Corollary 9.58. If $\mathbb{F}$ is a field, then the only ideals of $\mathbb{F}$ are the trivial or zero ideal $\{0\}$, and $\mathbb{F}$ itself.

Proof. This follows directly from Proposition 9.57 since every non-zero element in a field is a unit. □

Definition 9.59. Let R be a commutative ring. Given any $a_1, \ldots, a_n \in R$, the set

$$\langle a_1, \ldots, a_n \rangle := \{r_1 a_1 + \ldots + r_n a_n \mid r_1, \ldots, r_n \in R\}$$

is an ideal of R called the ideal generated by $a_1, \ldots, a_n$. More generally, if $A \subseteq R$, then the set of all finite R-linear combinations of elements in A, defined as

$$\langle A \rangle := \{r_1 a_1 + \ldots + r_n a_n \mid r_1, \ldots, r_n \in R, a_1, \ldots, a_n \in A, n \in \mathbb{N}\}$$

[12] We sometimes call this a *two-sided* ideal since we require the set to be closed under both right multiplication and left multiplication by elements in the ring. We may also define a *right* and *left* ideal, similarly.

is an ideal of R called the ideal generated by A.[13]

Definition 9.60 (Principal Ideal)**.** An ideal I of a commutative ring R is called a principal ideal of R if it can be generated by a single element in R, i.e, $I = \langle a \rangle$ for some $a \in R$.

Example 9.61.

1. $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$ for any $n \in \mathbb{Z}$. For $n = 1$, we get the unit ideal and for $n = 0$, we get the trivial or zero ideal. Moreover, $n\mathbb{Z}$ and $-n\mathbb{Z}$ are the same ideal for any $n \in \mathbb{Z}$. Additionally, one can prove that these are the only ideals of $\mathbb{Z}$ (Hint: Corollary 0.12). Note also that $n\mathbb{Z}$ is the ideal generated by n. We hence conclude that all ideals of $\mathbb{Z}$ are principal.

2. Let $\mathbb{F}$ be a field. For any $f(x) \in \mathbb{F}[x]$, the set

$$\langle f(x) \rangle := \{ f(x)g(x) \mid g(x) \in \mathbb{F}[x] \}$$

is an ideal of $\mathbb{F}[x]$. Are these the only ideals of $\mathbb{F}[x]$? That is, if I is an ideal of $\mathbb{F}[x]$ then is $I = \langle f(x) \rangle$ for some $f(x) \in \mathbb{F}[x]$? The answer is again **yes**. In fact, the proof is also eerily similar to the one for ideals of $\mathbb{Z}$. This is not surprising since they both are a consequence of the division algorithm, which holds true for both integers as well as polynomials. We again leave the exact details of the proof as an exercise to the reader. If I was the zero ideal, then taking $f(x)$ to be the zero polynomial suffices. If I is a nonzero ideal, then one may consider $f(x)$ to be a nonzero polynomial in I such that $\deg f(x)$ is the least among the degrees of all nonzero polynomials in I. One may then appeal to the division algorithm (Proposition 9.23) to conclude that any polynomial $h(x)$ will be a multiple of $f(x)$. Thus, any ideal of $\mathbb{F}[x]$ is principal, just like $\mathbb{Z}$.

3. Consider the ring $\mathbb{Z}[x]$ and the ideal $I = \langle 2, x \rangle$. We have

$$I = \{ 2f(x) + xg(x) \mid f(x), g(x) \in \mathbb{Z}[x] \}$$

We would like to show that the ideal I is *not* principal. Suppose I is principal and $I = \langle f(x) \rangle$ for some $f(x) \in \mathbb{Z}[x]$. Since $2 \in I$, we have $2 = f(x)g(x)$ for some $g(x) \in \mathbb{Z}[x]$. Comparing degrees, we get $\deg f(x) + \deg g(x) = 0$, which tells us that $f(x), g(x)$ are both non-zero constant polynomials whose product is 2. We hence conclude that $f(x) = \pm 1$ or $f(x) = \pm 2$. If $f(x) = \pm 1$, then $I = \mathbb{Z}[x]$, which is a contradiction since $1 \notin I$ (since 2 is not a unit in $\mathbb{Z}[x]$) but $1 \in \mathbb{Z}[x]$. Thus, $f(x) = \pm 2$ and I is the ideal generated by 2 (which is the same as the ideal generated by -2). Now, $x \in I$ and hence $x = 2h(x)$ for some $h(x) \in \mathbb{Z}[x]$. On comparing degrees, we conclude that $h(x)$ must be a linear polynomial, of the form $ax + b$ for some $a, b \in \mathbb{Z}$. Comparing the leading coefficients, we get $2a = 1$ which is not possible since, again, 2 is not a unit in $\mathbb{Z}$. Thus, I is not principal.

4. Consider the set

$$R = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$$

[13] Instead of $(a_1, \ldots, a_n)$ or (A), it is also common to denote these ideals as $\langle a_1, \ldots, a_n \rangle$ or $\langle A \rangle$.

R forms a commutative ring under the usual addition and multiplication of matrices. The set

$$S = \left\{ \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \mid a \in \mathbb{R} \right\}$$

forms a ring and is a subset of R. However, it does **not** form a subring of R since the multiplicative identity of R ($I_{2\times 2}$) is not present in the set S. The ring S has its own multiplicative identity (put $a = 1$ in the definition above), which is different from the identity in R.

§10. Ring Homomorphisms

Definition 10.1 (Ring Homomorphism)**.** Let R and S be rings. A map $\varphi\colon R \to S$ is called a ring homomorphism if

1. $\varphi(a+b) = \varphi(a) + \varphi(b)$ for all $a, b \in R$.
2. $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R$.
3. $\varphi(1) = 1$.[14]

A field homomorphism is a ring homomorphism between fields.

Definition 10.2. Since any field homomorphism is injective (why?), we also call them embeddings.

Example 10.3.

1. Of course, for any ring R, the identity map of R is a ring homomorphism.
2. Let R be a commutative ring and let R' be any overring of R, that is, R is a subring of R'. For some $\alpha \in R'$, we define the map $\pi_\alpha\colon R[x] \to R'$ as follows
$$\pi_\alpha\big(f(x)\big) = f(\alpha) \text{ for all } f(x) \in R[x].$$
We leave it as an easy exercise to the reader to show that π_α is a homomorphism, called the *substitution homomorphism* or the *substitution map*.
3. If R is a subring of R', then the *inclusion map* $\varphi\colon R \to R'$ defined as
$$\varphi(a) = a \text{ for all } a \in R$$
is a homomorphism.
4. Let R be a commutative ring and let $M_n(R)$ be the ring of $n \times n$ matrices with entries in R. For a fixed invertible matrix $P \in M_n(R)$, the map, $\varphi_P\colon M_n(R) \to M_n(R)$ defined by
$$\varphi_P(A) = PAP^{-1}$$
is a homomorphism.

Proposition 10.4. Let $\varphi\colon R \to S$ be a ring homomorphism. Then,

1. $\varphi(0) = 0$.
2. $\varphi(-r) = -\varphi(r)$ for all $r \in R$.
3. If $r \in R^\times$ then $s \in S^\times$ and $\varphi(r^{-1}) = (\varphi(r))^{-1}$.

[14] Note that a more suggestive way of writing this is $\varphi(1_R) = 1_S$. For brevity, we drop the subscript and 1 will be assumed to be the multiplicative identity of the appropriate ring.

4. The image of R under φ is a subring of S, where the image is defined as

$$\operatorname{im}\varphi := \{\varphi(r) \mid r \in R\}$$

Proof. We have

$$\varphi(0+0) = \varphi(0) + \varphi(0)$$

Since $0+0=0$, we have

$$\varphi(0) + \varphi(0) = \varphi(0) \implies \varphi(0) = 0$$

For any $a \in R$, we have

$$\varphi(a-a) = 0 = \varphi(a) + \varphi(-a)$$

This gives us

$$\varphi(-a) = -\varphi(a)$$

Let $r \in R^\times$. Then, r^{-1} exists. We have

$$\varphi(r)\varphi(r^{-1}) = \varphi(rr^{-1}) = \varphi(1) = 1$$

from which, the third part clearly follows. The proof of the fourth part is left as an exercise and follows directly from the first three parts. □

Definition 10.5 (Kernel)**.** Let $\varphi\colon R \to S$ be a ring homomorphism. The kernel of φ, denoted as $\ker\varphi$ is defined as

$$\ker\varphi := \{a \in R \mid \varphi(a) = 0\}$$

Note that the definition straightaway implies that $0 \in \ker\varphi$. Moreover, we have the following.

Proposition 10.6. If $\varphi\colon R \to S$ is a ring homomorphism, then $\ker\varphi$ is an ideal of R.

Proof. Left as an exercise. □

Definition 10.7 (Quotient Ring)**.** Let R be a ring and let I be an ideal of R. Then, I is an additive subgroup of R and the quotient group

$$R/I := \{r + I \mid r \in R\}$$

is an abelian group with addition defined as

$$(a+I) + (b+I) := (a+b) + I \text{ for all } a, b \in R.$$

Moreover, R/I is a ring where multiplication is defined[15] as

$$(a+I)(b+I) := ab + I \text{ for all } a, b \in R.$$

In fact, R/I is a ring with respect to addition and multiplication as defined above, with $1+I$ as the multiplicative identity in R/I.[16] We call R/I the quotient ring of I in R.

Proposition 10.8. Let R be a ring and let I be an ideal of R. Let $\varphi\colon R \to R/I$ be a map defined by

$$\varphi(r) = r + I \text{ for all } r \in R.$$

Then,

1. φ is a homomorphism,
2. $\ker \varphi = I$.

From Proposition 10.6 and Proposition 10.8, we see that any ideal is the kernel of some ring homomorphism and that the kernel of any homomorphism is an ideal.

Definition 10.9 (Isomorphism)**.** Let R, S be rings. A ring homomorphism $\varphi\colon R \to S$ is called an isomorphism if φ is a bijection. In this case, R and S are said to be isomorphic and we denote this as $R \cong S$.

Exercise 10.10. Let R, S be rings and let $\varphi\colon R \to S$ be an isomorphism. Then,

1. φ^{-1} is an isomorphism,
2. $r \in R$ is a unit if and only if $\varphi(r)$ is a unit in S,
3. $r \in R$ is a zero divisor if and only if $\varphi(r)$ is a zero divisor in S,
4. R is commutative if and only if S is commutative,
5. R is an integral domain if and only if S is an integral domain, and
6. R is a field if and only if S is a field.

Theorem 10.11 (Isomorphism Theorem for Rings)**.** Let R, S be rings. If $\varphi\colon R \to S$ is a homomorphism, then $\operatorname{im} \varphi \cong R/\ker \varphi$.

Proof. Verify that the map $r + \ker \varphi \mapsto \varphi(r)$ defines an isomorphism. □

Definition 10.12 (Prime Ideal)**.** Let R be a ring and let P be an ideal of R. P is called a prime ideal of R if $P \neq R$ and for all $a, b \in R$,

$$ab \in P \implies a \in P \text{ or } b \in P$$

[16] One should check that this is indeed well-defined. That is, if $a + I = a' + I$ and $b + I = b' + I$, then $ab + I = a'b' + I$.
[16] We leave it as an exercise to verify that R/I is indeed a ring.

Definition 10.13 (Maximal Ideal)**.** Let R be a ring and let M be an ideal of R. M is called a maximal ideal of R if $M \neq R$ and whenever J is an ideal of R such that $M \subseteq J$, we have either $J = M$ or $J = R$.

Example 10.14.

1. In the ring $\mathbb{Z}$, the ideals $\langle 0 \rangle = \{0\}$ and $p\mathbb{Z}$, where p is a prime, are prime ideals of $\mathbb{Z}$, as is proved trivially by Euclid's Lemma (Corollary 0.17). In fact, these are the only prime ideals of $\mathbb{Z}$. Moreover, the ideals $p\mathbb{Z}$ are the only maximal ideals of $\mathbb{Z}$.
2. Similarly, in the ring $\mathbb{F}[x]$ where $\mathbb{F}$ is a field, the only maximal ideals are ideals of the form $\langle f(x) \rangle$ where $f(x)$ is irreducible in $\mathbb{F}[x]$. Further, these ideals, together with the zero ideal, are the only prime ideals of $\mathbb{F}[x]$. In particular, $\mathbb{F} = \mathbb{C}$, we have two more versions of the Fundamental Theorem of Algebra, as stated ahead,

Theorem 10.15 (Fundamental Theorem of Algebra - Version 6)**.** The only maximal ideals in $\mathbb{C}[x]$ are $\langle x - \alpha \rangle$ where $\alpha \in \mathbb{C}$.

Theorem 10.16 (Fundamental Theorem of Algebra - Version 7)**.** The only maximal ideals in $\mathbb{R}[x]$ are of the form $\langle x - a \rangle$ or $\langle x^2 + bx + c \rangle$ where $a, b, c \in \mathbb{R}$ and b, c are such that $b^2 - 4c < 0$.

Proposition 10.17. Let R be a commutative ring and let I be an ideal of R. Then,

1. I is a prime ideal if and only if R/I is an integral domain;
2. I is a maximal ideal if and only if R/I is a field.

Proof.

1. Suppose I is a prime ideal. Since $I \neq R$, R/I is not the trivial ring. Thus, $1 \neq 0$ in the ring R/I. We now show that if the product of two cosets of I is equal to I (the additive identity of the ring R/I), then at least one of them must be equal to I. For $a, b \in R$, if we have $(a + I)(b + I) = I$, then $ab + I = I \implies ab \in I$. Since I is a prime ideal, this means that either $a \in I$ or $b \in I$. Thus, either $a + I = I$ or $b + I = I$, proving that R/I is an integral domain. The converse is straightforward as well and is left as an exercise.
2. Suppose I is a maximal ideal of R. Then, $I \neq R$ and thus $1 \neq 0$ in R/I. Now suppose that $a + I$ is a non-zero element of R/I for some $a \in R$. Then, $a \notin I$. Let J be the ideal generated by a and I. We have
$$J = \{ra + u \mid r \in R \text{ and } u \in I\}$$
We leave it as an exercise to show that J is indeed an ideal. Notice that J contains I (take r to be 0). Moreover, $J \neq I$ since $a \in J$ and $a \notin I$. Thus, by the maximality of I, we must have

that $J = R$. In particular, we have that

$$1 = ba + u \text{ for some } b \in R \text{ and } u \in I.$$

This implies that

$$(a + I)(b + I) = ab + I = ba + I = 1 - u + I$$

Since $u \in I$, we get that

$$(a + I)(b + I) = 1 + I$$

and hence, $b + I$ is a multiplicative inverse of $a + I$ in R/I. Hence, R/I is a field. We again leave the converse as an exercise to the reader.

□

Corollary 10.18. Let R be a commutative ring and let I be an ideal of R. If I is maximal, then I is prime.

Proof. This follows from Proposition 10.17 since every field is an integral domain. □

Proposition 10.19. In a finite commutative ring, every prime ideal is maximal.

Proof. We leave the proof as an exercise. (Hint: Proposition 9.16 and Proposition 10.17) □

Exercise 10.20. Let R be a commutative ring and let I be an ideal of R. Show that there is an inclusion-preserving[17] bijection between ideals of R containing I and the set of ideals of R/I, which is given by $J \mapsto J/I := \{a + I \mid a \in J\}$ where J is an ideal of R containing I. Moreover, this correspondence preserves primality and maximality.

Proposition 10.21. Let R be a ring and let I be an ideal of R such that $I \neq R$. Then, there exists a maximal ideal M of R such that $I \subseteq M$.

Proof. Consider the set

$$\mathcal{F} = \{J \mid J \text{ is an ideal of } R \text{ with } J \neq R \text{ and } I \subseteq J\}.$$

$\mathcal{F}$ is clearly non-empty since $I \in \mathcal{F}$. Further, suppose that $\{I_\alpha\}_{\alpha \in \Lambda}$ is a chain in $\mathcal{F}$, that is, a subset of $\mathcal{F}$ such that for any $\alpha, \beta \in \Lambda$, either $I_\alpha \subseteq I_\beta$ or $I_\beta \subseteq I_\alpha$. Now, we define

$$J := \bigcup_{\alpha \in \Lambda} I_\alpha.$$

[17] Let φ be the bijection. By 'inclusion-preserving', we mean that if J_1 and J_2 are ideals containing I and if $J_1 \subset J_2$, then $\varphi(J_1) \subset \varphi(J_2)$.

We claim that J is an ideal of R (the proof is left as an exercise) and $I \subseteq J$. Moreover, $J \neq R$, since $1 \in J \implies 1 \in I_\alpha \implies I_\alpha = R$ for some $\alpha \in \Lambda$, which is a contradiction. Thus $J \in \mathcal{F}$ and clearly, $I_\alpha \subseteq J$ for all $\alpha \in \Lambda$. So, J is an upper bound in $\mathcal{F}$ on the chain $\{I_\alpha\}_{\alpha \in \Lambda}$. Hence, by Zorn's Lemma, $\mathcal{F}$ has a maximal element with respect to inclusion, say M. Clearly, M has the desired properties.

□

Proposition 10.22. Every subring of a field is an integral domain.

Proof. Left as an exercise. □

Interestingly, the converse of Proposition 10.22 is also true.

Proposition 10.23. Every integral domain is a subring of some field.

The simplest way to prove Proposition 10.23 is to construct the so-called field of fractions of the integral domain. This is similar to the construction of $\mathbb{Q}$ from $\mathbb{Z}$.

§§10.1. Construction of Field of Fractions of an Integral Domain

For the remainder, we let R denote an integral domain and let $S := R \times R^\times = \{(a, b) \mid a, b \in R, b \neq 0\}$

Definition 10.24. We define a relation $\sim$ on S as follows:

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

Lemma 10.25. The relation $\sim$ on S is an equivalence relation.

Proof. Given any $(a, b) \in S$, we clearly have $ab = ab \implies (a, b) \sim (a, b)$. Hence $\sim$ is reflexive.

Suppose $(a, b) \sim (c, d)$. Then, $ad = bc \implies cb = ad \implies (c, d) \sim (a, b)$ and hence $\sim$ is symmetric.

Suppose $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. Then, $ad = bc$ and $cf = de$. Now,

$$\begin{aligned} ad = bc &\implies adf = bcf \\ &\implies adf = bde \\ &\implies af = be && \text{(since } d \neq 0 \text{ and } R \text{ is an integral domain)} \\ &\implies (a, b) \sim (e, f) \end{aligned}$$

Hence, $\sim$ is transitive. □

Definition 10.26 (Field of Fractions)**.** The field of fractions of an integral domain R, denoted as $\mathrm{Frac}(R)$, is the collection of equivalence classes of the relation $\sim$ on S. If $(a, b) \in S$, we denote the equivalence class of (a, b) with respect to $\sim$ as $\frac{a}{b}$. Thus,

$$\mathrm{Frac}(R) := \left\{ \frac{a}{b} \mid (a, b) \in S \right\}$$

We are yet to prove that the above construction is a field. First, we define the field operations on this set and show that these are indeed well-defined.

Definition 10.27. We define addition and multiplication as follows. Given, $\frac{a}{b}, \frac{c}{d} \in \mathrm{Frac}(R)$

$$\frac{a}{b} + \frac{c}{d} := \frac{ad + bc}{bd}$$
$$\frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}$$

Proposition 10.28. Addition in $\mathrm{Frac}(R)$ is well-defined. That is, if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$, then

$$(ad + bc, bd) \sim (a'd' + b'c', b'd').$$

Proposition 10.29. Multiplication in $\mathrm{Frac}(R)$ is well-defined. That is, if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$, then

$$(ac, bd) \sim (a'c', b'd').$$

Theorem 10.30. The set $\mathrm{Frac}(R)$ along with addition and multiplication as defined above, forms a field where

1. the additive identity is $\frac{0}{1}$,
2. the additive inverse of $\frac{a}{b}$ is $\frac{-a}{b}$,
3. the multiplicative identity is $\frac{1}{1}$, and
4. for $\frac{a}{b} \neq \frac{0}{1}$, the multiplicative inverse of $\frac{a}{b}$ is $\frac{b}{a}$.

We leave the proofs of the above results as an instructive exercise.

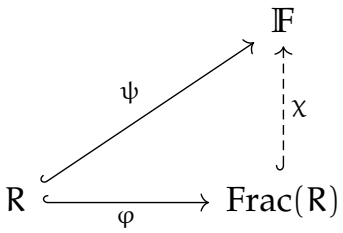
The map $\varphi \colon R \to \mathrm{Frac}(R)$ defined by

$$\varphi(a) = \frac{a}{1} \text{ for } a \in R$$

is an injective homomorphism, which we refer to as the *natural inclusion map*. Thus, $R \cong \operatorname{im} \varphi$,

which is a subring of $\mathrm{Frac}(R)$. Thus, identifying R with $\operatorname{im}\varphi$, we may regard R to be a subring of $\mathrm{Frac}(R)$, which is what we had wanted to show all along.

Theorem 10.31 (Universal Property)**.** Let R be an integral domain and let $R \overset{\varphi}{\hookrightarrow} \mathrm{Frac}(R)$ be the natural inclusion map. If $\mathbb{F}$ is a field such that there is an injective homomorphism $R \overset{\psi}{\hookrightarrow} \mathbb{F}$, then there exists an injective homomorphism $\mathrm{Frac}(R) \overset{\chi}{\hookrightarrow} \mathbb{F}$ such that $\psi = \chi \circ \varphi$.



Intuitively, the universal property states that if $\mathbb{F}$ is any field that contains R as a subring, then the field $\mathbb{F}$ also contains $\mathrm{Frac}(R)$. Thus, $\mathrm{Frac}(R)$ is the smallest field containing R as a subring.

Proof. We define $\chi\colon \mathrm{Frac}(R) \to \mathbb{F}$ as

$$\chi\left(\frac{a}{b}\right) := \psi(a)\psi(b)^{-1} \text{ for } a, b \in R, b \neq 0.$$

Since ψ is injective and $\mathbb{F}$ is a field, the above map is well-defined.[18] It is also straightforward to check that χ is a homomorphism. Now, it remains to show that χ is injective. We have

$$\begin{aligned}
\chi\left(\frac{a}{b}\right) = 0 &\implies \psi(a)\psi(b)^{-1} = 0 \\
&\implies \psi(a) = 0 && \text{(since } b \neq 0 \text{ and hence } \psi(b)^{-1} \neq 0\text{)} \\
&\implies a = 0 && \text{(since } \psi \text{ is injective)} \\
&\implies \frac{a}{b} = 0.
\end{aligned}$$

Hence, $\ker\chi$ is trivial and χ is indeed injective. Now, for any $a \in R$, we have

$$\chi \circ \varphi(a) = \chi\left(\frac{a}{1}\right) = \psi(a)\psi(1)^{-1} = \psi(a) \implies \chi \circ \varphi = \psi. \qquad \square$$

Corollary 10.32. If $\mathbb{F}$ is any field such that there is an injective homomorphism $R \overset{\psi}{\hookrightarrow} \mathbb{F}$ satisfying the universal property, then $\mathbb{F}$ and $\mathrm{Frac}(R)$ are isomorphic. Moreover, there exists an isomorphism $\chi\colon \mathrm{Frac}(R) \to \mathbb{F}$ such that $\psi = \chi \circ \varphi$.

Proof. Left as an exercise. This indicates that $\mathrm{Frac}(R)$ is unique up to isomorphism. $\square$

[18] One must also check that if $\frac{a}{b} = \frac{a'}{b'}$, then $\chi\left(\frac{a}{b}\right) = \chi\left(\frac{a'}{b'}\right)$.

Definition 10.33. Let R be an integral domain. The field $\mathrm{Frac}(R[x])$ is called the field of rational functions with coefficients in R and is denoted as $R(x)$.

The field of rational functions $R(x)$ consists of elements of the form $\frac{p(x)}{q(x)}$ where $p(x), q(x) \in R[x]$ and $q(x) \neq 0$.

§11. Domains

For the remainder, we assume all rings to be commutative.

§§11.1. Euclidean Domains

Definition 11.1. Let R be an integral domain. Any function $N\colon R \to \mathbb{N}$ with $N(0) = 0$ is called a norm on R. If $N(a) > 0$ for $a \neq 0$, we call N a positive norm.

Observe that this definition of a norm is fairly weak, and hence any integral domain R possesses several different norms.

Definition 11.2. An integral domain R is said to be a Euclidean domain (or possess a division algorithm) if there is a norm N on R such that for any two elements $a, b \in R$ with $b \neq 0$, there exist elements $q, r \in R$ such that

$$a = qb + r \text{ with } r = 0 \text{ or } N(r) < N(b).$$

q is called the quotient and r is called the remainder of the division. Such a norm N is called a Euclidean function.

In a Euclidean domain, we have the Euclidean algorithm, which allows us to write the following by successive "divisions".

$$\begin{aligned} a &= q_0 b + r_0 \\ b &= q_1 r_0 + r_1 \\ r_0 &= q_2 r_1 + r_2 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n \\ r_{n-1} &= q_{n+1} r_n \end{aligned}$$

where r_n is the last nonzero remainder. Such an r_n exists since $N(b) > N(r_0) > N(r_1) \ldots > N(r_n)$ is a decreasing sequence of non-negative integers if the remainders are nonzero, and such a sequence cannot continue indefinitely. Note that there is no guarantee that these elements are unique.

Example 11.3.

1. Fields are trivial examples of Euclidean domains, that satisfy the defining conditions with any norm. This is because for all elements a, b with $b \neq 0$, we have

   $$a = qb + 0$$

   where $q = ab^{-1}$.

2. The integers, $\mathbb{Z}$, form a Euclidean domain with norm given by $N(a) = a$, the usual absolute value, for all $a \in \mathbb{Z}$. Of course, a division algorithm does exist for the integers, as we have

seen before (Proposition 0.11). Note however that if a is not a multiple of b, we have two possible pairs of quotient and remainder. For example, we have

$$5 = 2 \cdot 2 + 1 \text{ and } 5 = 3 \cdot 2 - 1.$$

If however we restrict the remainder to be non-negative, this factorisation is unique.

Definition 11.4. Let $\mathbb{F}$ be a field. A discrete valuation on $\mathbb{F}$ is a function $\nu \colon \mathbb{F}^\times \to \mathbb{Z}$ satisfying the following.

1. $\nu(ab) = \nu(a) + \nu(b)$, i.e, ν is a homomorphism from the multiplicative group of units of $\mathbb{F}$ to the additive group $\mathbb{Z}$.
2. ν is surjective.
3. $\nu(x+y) \geq \min\{\nu(x), \nu(y)\}$ for all $x, y \in \mathbb{F}^\times$ with $x + y \neq 0$.

Definition 11.5. Let ν be a discrete valuation on a field $\mathbb{F}$. The set $\{x \in \mathbb{F}^\times \mid \nu(x) \geq 0\} \cup \{0\}$ forms a subring of $\mathbb{F}$ called the valuation ring of ν.

Proposition 11.6. Let ν be a discrete valuation on a field $\mathbb{F}$. Then,

1. $\nu(1) = 0$, and
2. $\nu(b^{-1}) = -\nu(b)$ for all $b \in \mathbb{F}^\times$.

Proof. We leave the proof as an exercise to the reader. These properties follow directly from the definition. □

Definition 11.7. An integral domain R is called a discrete valuation ring if there exists a discrete valuation ν on $\mathrm{Frac}(R)$ such that R is the valuation ring of ν.

Proposition 11.8. A discrete valuation ring is a Euclidean domain.

Proof. Let R be a discrete valuation ring. We define the norm on R to be $N(0) = 0$ and $N = \nu$ on non-zero elements of R. Now, for $a, b \in R$ with $b \neq 0$, we have that

1. if $N(a) < N(b)$, then $a = 0b + a$, and
2. if $N(a) \geq N(b)$, then we have $q = ab^{-1} \in R$, since

$$\nu(q) = \nu(ab^{-1}) = \nu(a) - \nu(b)$$

$$\therefore \nu(q) \geq 0 \implies q \in R.$$

We then have $a = qb + 0$. □

Proposition 11.9. Every ideal in a Euclidean domain is principal. More precisely, if I is a nonzero ideal in a Euclidean domain R, then $I = \langle d \rangle$, where d is any nonzero element of I of minimum norm.

Proof. If I is the zero ideal, we are done. Else, let d be an element of I of minimum norm. Such an element exists since the set $\{N(a) \mid a \in I\}$ has a minimum element by the well-ordering property (Remark 0.4). Clearly, $\langle d \rangle \subseteq I$ since d is an element of I. Suppose $a \in I$. Since R is a Euclidean domain applying the division algorithm allows us to write $a = qd + r$ with $r = 0$ or $N(r) < N(d)$. Then, $r = a - qd$. Since $d \in I$, we have $qd \in I$. Now, since $a \in I$ and $qd \in I$, we have that $r \in I$. By the minimality of norm of d, we must have $r = 0$, giving us $a = qd \implies a \in \langle d \rangle$. Hence, $I = \langle d \rangle$. □

Example 11.10. Let $R = \mathbb{Z}[x]$. Since $\langle 2, x \rangle$ is not a principal ideal of $\mathbb{Z}[x]$ (why?), it follows that $\mathbb{Z}[x]$ is not a Euclidean domain.

Proposition 11.11. Let R be a commutative ring and let $a, b \in R$. If $I = \langle a, b \rangle$, then d is a gcd of a and b if

1. I is contained in the principal ideal $\langle d \rangle$, and
2. if $\langle d' \rangle$ is any principal ideal containing I, then $\langle d \rangle \subseteq \langle d' \rangle$.

Proposition 11.12. If a and b are nonzero elements of a commutative ring R such that the ideal generated by a and b is a principal ideal $\langle d \rangle$, then d is a gcd of a and b.

Definition 11.13. An integral domain in which every ideal generated by two elements is principal is called a Bézout domain.

Exercise 11.14. In a Bézout domain, every finitely generated ideal is principal.

Proposition 11.15. Let R be an integral domain. If for two elements $d, d' \in R$, we have $\langle d \rangle = \langle d' \rangle$, then $d' = ud$ for some unit u. In particular, if d and d' are both greatest common divisors of two elements, then $d' = ud$ for some unit u.

Proof. The proof is trivial if either of d or d' is zero, so we may assume that both d and d' are nonzero. Since $d \in \langle d' \rangle$, we have $d = xd'$ for some $x \in R$. Similarly, since $d' \in \langle d \rangle$, we have $d' = yd$ for some $y \in R$. Thus, $d = xyd$ and $d(1 - xy) = 0$. Since $d \neq 0$ and R is an integral domain, it follows that $xy = 1$. That is, both x and y are units in R. This proves the first part. The second

part follows trivially since any two greatest common divisors of two elements generate the same principal ideal. □

Theorem 11.16. Let R be an integral domain and let a and b be two non-zero elements of R. Let $d = r_n$ be the last non-zero remainder in the Euclidean algorithm for a and b. Then,

1. d is a greatest common divisor of a and b, and
2. the principal ideal $\langle d \rangle$ is the ideal generated by a and b. In particular, d can be written as an R-linear combination of a and b, that is, there exist elements $x, y \in R$ such that

$$d = ax + by.$$

One we may regard the last statement of the above theorem as an extension of Bézout's Lemma.

Proof. Since the ideal generated by a and b is principal, a and b do have a gcd, namely, any element which generates the principal ideal $\langle a, b \rangle$. Both parts of the theorem follow once we show that $d = r_n$ generates the said ideal. To do so, we will show that

1. $d \mid a$ and $d \mid b$, so that $\langle a, b \rangle \subseteq \langle d \rangle$, and
2. d is an R-linear combination of a and b, so that $\langle d \rangle \subseteq \langle a, b \rangle$.

It is easy to show via induction that d indeed divides a and b (keep track of the divisibilities in the Euclidean algorithm). To prove the second part, we may again proceed inductively to show that $r_n \in \langle a, b \rangle$. More specifically, we have that $r_0 \in \langle a, b \rangle$. Assuming $r_{k-1}, r_k \in \langle a, b \rangle$, we have

$$r_{k+1} = r_{k-1} - q_{k+1} r_k \in \langle r_{k-1}, r_k \rangle \subseteq \langle a, b \rangle.$$

Hence, by induction, $r_n \in \langle a, b \rangle$, which completes the proof. □

Aside. The above discussion gives an interesting perspective on the existence of a solution in the integers to the first-order Diophantine equation, given by

$$ax + by = N$$

where $a, b, N \in \mathbb{Z}$. Observe that the existence of a solution (x, y) to the above equation is just another way of saying that $N \in \langle a, b \rangle$. By the above theorem, this is the same as saying that $N \in \langle d \rangle$ where $d = \gcd(a, b)$. Hence, the equation $ax + by = N$ is solvable in integers x and y if and only if N is divisible by $\gcd(a, b)$. Moreover, let (x_0, y_0) be a solution of the above equation. Then, the complete set of solutions to the equation is given by

$$x = x_0 + m \cdot \frac{a}{\gcd(a, b)}$$
$$y = y_0 - m \cdot \frac{b}{\gcd(a, b)}$$

where m varies over $\mathbb{Z}$.

We end this discussion with another useful notion that is sometimes used to prove that a given integral domain is not a Euclidean domain. For any integral domain R, we define $\tilde{R} := R^\times \cup \{0\}$.

Definition 11.17. Let R be an integral domain. An element $u \in R \setminus \tilde{R}$ is called a universal side divisor if for every $x \in R$, there is some $z \in \tilde{R}$ such that u divides $x - z$. That is, every $x \in R$ can be written as $x = qu + z$ where z is either zero or a unit.

Proposition 11.18. Let R be an integral domain that is not a field. If R is a Euclidean domain, then there are universal side divisors in R.

Proof. Suppose R is a Euclidean domain with some norm N, and let u be an element of $R \setminus \tilde{R}$ (this is nonempty since R is not a field) of minimal norm. For any $x \in R$, write $x = qu + r$, where either $r = 0$, or $N(r) < N(u)$. In either case, the minimality of $N(u)$ implies that $r \in \tilde{R}$. Hence, u is a universal side divisor of R. □

§§11.2. Principal Ideal Domains

Definition 11.19. A principal ideal domain (PID) is an integral domain in which every ideal is principal.

An immediate consequence of the above definition is the following.

Corollary 11.20. Every Euclidean domain is a principal ideal domain.

Example 11.21.

1. We have seen in Example 9.61, that $\mathbb{Z}$ is a principal ideal domain.
2. Example 11.10 tells us that $\mathbb{Z}[x]$ is not a principal ideal domain.

Proposition 11.22. Every nonzero prime ideal in a principal ideal domain is a maximal ideal.

Proof. Let $\langle p \rangle$ be a prime ideal in a principal ideal domain R, and let $I = \langle m \rangle$ be an ideal containing $\langle p \rangle$. We must show that $I = \langle p \rangle$ or $I = R$. Since $p \in \langle m \rangle$, we must have $p = rm$ for some $r \in R$. Since $\langle p \rangle$ is a prime ideal and $rm \in \langle p \rangle$, we must have $r \in \langle p \rangle$ or $m \in \langle p \rangle$. If $m \in \langle p \rangle$, then $\langle p \rangle = \langle m \rangle = I$, and we are done. Suppose otherwise that $r \in \langle p \rangle$. Then, $r = ps$ for some $s \in R$. Now, we have $rm = psm = p$ and hence $sm = 1$, since R is an integral domain. Thus, m is a unit and hence $I = R$. □

Corollary 11.23. If the polynomial ring $R[x]$ is a principal ideal domain, then R is a field.

Proof. Assume that $R[x]$ is a principal ideal domain. Since R is a subring of $R[x]$, R is an integral domain. Define $\psi\colon R[x] \to R$ by

$$\psi\big(f(x)\big) = f(0) \text{ for all } f(x) \in R[x].$$

We leave it as an exercise to verify that ψ is a homomorphism. By the Isomorphism Theorem for Rings, $R[x]/\ker\psi$ is isomorphic to R. We may show that $\ker\psi = \langle x\rangle$ and hence, $R[x]/\langle x\rangle \cong R$. Thus, $R[x]/\langle x\rangle$ is an integral domain. By Proposition 10.17, $\langle x\rangle$ is a non-zero prime ideal. By Proposition 11.22, $\langle x\rangle$ is also maximal. Again, Proposition 10.17 tells us that $R[x]/\langle x\rangle$ is a field, and hence R is a field. □

Definition 11.24. A norm N on an integral domain R is called a Dedekind-Hasse norm if N is a positive norm, and for every non-zero $a, b \in R$, either $a \in \langle b\rangle$, or there is a nonzero element in $\langle a, b\rangle$ with norm strictly lesser than $N(b)$. That is, either b divides a in R, or there exist $s, t \in R$ with $0 < N(sa - tb) < N(b)$.

Notice that the R is a Euclidean domain with respect to a positive norm N if it is always possible to satisfy the Dedekind-Hasse condition with $s = 1$. Thus, the Dedekind-Hasse condition is a weakening of the Euclidean condition, and we may think of a Dedekind-Hasse norm as a generalisation of a Euclidean function.

Proposition 11.25. An integral domain R is a principal ideal domain if and only if R has a Dedekind-Hasse norm.

Proof. Let I be a nonzero ideal in R and let b be a nonzero element of I, chosen such that $N(b)$ is minimal (again, such a b exists by the Well-Ordering Property). Suppose a is a nonzero element in I, so that $\langle a, b\rangle \subseteq I$. Let N be a Dedekind-Hasse norm on R. We must have that either b divides a, or that there exist $s, t \in R$ such that $0 < N(sa - tb) < N(b)$. Since $N(b)$ is minimal, we conclude that $a \in \langle b\rangle$. Thus, I is principal and R is a principal ideal domain. We shall prove the converse in a later section. □

§§11.3. Unique Factorisation Domains

So far, we have computed the gcd of two elements algorithmically. However, Proposition 0.22 shows us that for elements in $\mathbb{Z}$, we may calculate the gcd of two numbers using their prime factorisations. This idea generalises to a large class of rings, as we now show. We first recall the definition of an irreducible element. We restrict ourselves to integral domains for the following discussion.

Definition 11.26. Let R be an integral domain. An element $f \in R$ is said to be irreducible if f is non-zero, non-unit in R, and whenever $f = gh$ for some $g, h \in R$, at least one of g and h is a unit.

Definition 11.27. Let R be an integral domain. An element $p \in R$ is said to be prime if p is non-zero, and the ideal $\langle p\rangle$ is a prime ideal. In other words, a non-zero element p is a prime if it is not a unit, and whenever $p \mid ab$ for $a, b \in R$, then either $p \mid a$ or $p \mid b$.

Definition 11.28. Let R be an integral domain. Two elements $a, b \in R$ are said to be associate if $a = ub$ for some unit $u \in R^\times$.

Proposition 11.29. In an integral domain, every prime element is irreducible.

Proof. Let p be a prime, so that $\langle p \rangle$ is a nonzero prime ideal. Suppose $p = ab$ for some $a, b \in R$, so that $ab = p \in \langle p \rangle$. Since $\langle p \rangle$ is a prime ideal, one of a or b must be in $\langle p \rangle$. Assume without loss of generality that $a \in \langle p \rangle$, so that $a = pr$ for some $r \in R$. Now, $p = ab = prb \implies rb = 1$. Thus, b is a unit and p is irreducible. □

Proposition 11.30. In a principal ideal domain, an element is prime iff it is irreducible.

Proof. Proposition 11.29 already shows us that prime implies irreducible. Hence it suffices to show that in a principal ideal domain, an irreducible element is prime. Let M be any ideal containing $\langle p \rangle$. Since R is a principal ideal domain, $M = \langle m \rangle$ for some $m \in R$. Since $p \in \langle m \rangle$, $p = rm$ for some $r \in R$. Now, since p is irreducible, one of r or m must be a unit. Thus, either $\langle p \rangle = \langle m \rangle$ or $\langle m \rangle = R$. Thus, the only ideals containing $\langle p \rangle$ are $\langle p \rangle$ itself, and R. Hence, $\langle p \rangle$ is maximal ideal. By Corollary 10.18, $\langle p \rangle$ is a prime ideal, and hence p is prime. □

Definition 11.31. A unique factorisation domain (UFD) is an integral domain R in which every non-zero, non-unit element $r \in R$ has the following properties.

1. r can be written as a finite product of irreducibles (not necessarily distinct): $r = p_1 \cdots p_n$.
2. The decomposition is above is unique up to associates. That is, if $r = q_1 \cdots q_m$ is another factorisation of r into irreducibles, then $n = m$, and there is some permutation σ such that p_i is associate to $\sigma(q_i)$ for $i = 1, \ldots, n$.

As a trivial example, observe that any field is vacuously a unique factorisation domain since there are no non-zero non-unit elements.

Proposition 11.32. In a unique factorisation domain, an element is prime iff it is irreducible.

Proof. As before, Proposition 11.29 already shows us that prime implies irreducible. We now show that in a unique factorisation domain, an irreducible element is prime. Let p be an irreducible element and suppose $p \mid ab$ for some $a, b \in R$. Thus, $ab = pc$ for some $c \in R$. Writing a, b, c as their irreducible decompositions, and from the uniqueness of this decomposition, we see that p must be associate to one of the irreducibles on the LHS. Without loss of generality, assume that p is associate to one of the irreducibles in the decomposition of a, so that $a = (up)p_2 \cdots p_n$ for some unit $u \in R^\times$ and some (possibly empty) set of irreducibles $p_2, \ldots, p_n$. Thus, $p \mid a$ since $a = pd$ with $d = up_2 \cdots p_n$. This completes the proof. □

We may now use the terms prime and irreducible interchangeably. This allows us to talk about 'prime factorisations', which is just the decomposition of nonzero elements into irreducibles or primes.

Proposition 11.33. Let a and b be two nonzero elements of a unique factorisation domain R and suppose

$$a = up_1^{e_1} \cdots p_n^{e_n}$$
$$b = vp_1^{f_1} \cdots p_n^{f_n}$$

where u and v are units, the primes $p_1, \ldots, p_n$ are distinct, and the exponents e_i and f_i are non-negative. Then, the element

$$d = p_1^{\min(e_1,f_1)} \cdots p_n^{\min(e_n,f_n)}$$

is a gcd of a and b.

Proof. Since the exponents of each of the primes occurring in d are no larger than the exponents occurring in the prime factorisations of a and b, it follows that d is a common divisor of a and b. We leave it as an exercise to show that if c is a common divisor of a and b, then $c \mid d$. □

Theorem 11.34. Every principal ideal domain is a unique factorisation domain.

Proof. Let R be a principal ideal domain and let $r \in R$ be non-zero and non-unit. We must show that r can be written as a finite product of irreducibles in R, and that this decomposition is unique up to units. We first prove that such a decomposition indeed exists.

If r is itself reducible, then we are done. If not, then $r = r_1 r_2$ where r_1, r_2 are both non-unit. If both these elements are irreducible then again we are done. If not, at least one element, say r_1, can be written as a product of two non-unit elements, $r_1 = r_{11} r_{12}$, and so forth. We must verify that this process terminates, that is, we must necessarily reach a point where all factors of r are irreducible. If this is not the case, we obtain an *infinite ascending chain* of ideals, as follows.

$$\langle r \rangle \subset \langle r_1 \rangle \subset \langle r_{11} \rangle \subset \ldots \subset R$$

where all inclusions are proper. Moreover, such an infinite chain exists thanks to the Axiom of Choice. We now show that such an ascending chain of ideals $I_1 \subseteq I_2 \subseteq \ldots \subseteq R$ eventually becomes stationary. That is, there is some positive integer n such that $I_k = I_n$ for all $k \geq n$. This means that it is not possible to have an infinite ascending chain of ideals where all containments are proper. Let $I := \cup_{i=1}^{\infty} I_i$. It is easy to show that I is an ideal of R. Since R is a principal ideal domain, $I = \langle a \rangle$ for some $a \in R$. Since $a \in I$, we must have $a \in I_n$ for some n. We then have $I_n \subseteq I = \langle a \rangle \subseteq I_n$, so that $I = I_n$ and the chain becomes stationary at I_n. This proves that every non-zero, non-unit element in R has a finite decomposition into irreducibles.

We now show that this decomposition is unique up to units. We induct on the number, n, of irreducible factors in some factorisation of the element $r \in R$. If $n = 0$, then r is a unit and the factorisation is trivially unique since if $r = qc$ for some irreducible q, then q divides a unit, which

is a contradiction. Suppose now that $n \geq 1$ and we have that

$$r = p_1 \cdots p_n = q_1 \cdots q_m \quad m \geq n$$

where p_i and q_j are (not necessarily distinct) irreducibles. Since p_1 divides the product on the right, it must divide one of the factors. Without loss of generality, suppose p_1 divides q_1 so that $q_1 = p_1 u$ for some $u \in R$. Since q_1 is irreducible, u is a unit and thus p_1 and q_1 are associates. Since we are operating in an integral domain, we may 'cancel' p_1 to get

$$p_2 \cdots p_n = u q_2 \cdots q_m = q_2' \cdots q_m \quad m \geq n$$

where $q_2' = u q_2$ is also irreducible. By induction on n, we may conclude that up to associates, each factor on the left matches bijectively with factor on the right. Since we have already shown that p_1 and q_1 are associates, we are done. □

Corollary 11.35. Every Euclidean domain is a unique factorisation domain.

Proof. This follows straight from Theorem 11.34 and Corollary 11.20. □

Corollary 11.36 (Fundamental Theorem of Arithmetic)**.** The integers $\mathbb{Z}$ form a unique factorisation domain.

Proof. This is trivial since $\mathbb{Z}$ is a Euclidean domain. □

Proposition 11.37. Let R be a principal ideal domain. Then, there exists a multiplicative Dedekind-Hasse norm on R.

Proof. If R is a principal ideal domain, then R is a unique factorisation domain. Define the norm N by setting $N(0) = 0$, $N(u) = 1$ if u is a unit, and $N(a) = 2^n$ if $a = p_1 \cdots p_n$ where p_i's are irreducibles in R. This is well-defined since the number of irreducible factors of a is unique. Clearly, $N(ab) = N(a)N(b)$, so that N is positive and multiplicative. Suppose a, b are nonzero elements in R. Since R is a principal ideal domain, we have $\langle a, b \rangle = \langle r \rangle$ for some $r \in R$. If b divides a in R then we are done. If b does not divide a, that is, $a \notin \langle b \rangle$, and hence $r \notin \langle b \rangle$. However, $b = xr$ for some $x \in R$, and thus x is not a unit in R. We then have that $N(b) = N(x)N(r) > N(r)$, proving that there is an element in $\langle a, b \rangle$ with norm strictly smaller than $N(b)$. Hence, N is a multiplicative Dedekind-Hasse norm on R. □

§12. Polynomial Rings

§§12.1. Definitions

For this section, whenever we talk about rings, we assume commutative rings. Recall that the polynomial ring $R[x]$ in the indeterminate x with coefficients in the ring R is defined as the set of all formal sums $a_n x^n + \dots a_0$ with $n \geq 0$ and each $a_i \in R$. If $a_n \neq 0$, then $a_n x^n$ is the leading term, a_n is the leading coefficient, and the degree of the polynomial is n. We define the leading coefficient of the zero polynomial as zero. If $a_n = 1$, we call the polynomial monic. With addition and multiplication of polynomials defined the usual way, $R[x]$ is a commutative ring that borrows its identity from R itself. Moreover, we identify R with the subring of constant polynomials. We now state, without proof, a proposition that summarises a bunch of results from Section 9.2.

Proposition 12.1. Let R be an integral domain. Then,

1. $\deg(p(x)q(x)) = \deg p(x) + \deg q(x)$ for all $p(x), q(x) \in R[x]$,
2. the units of $R[x]$ are just the units of R, and
3. $R[x]$ is an integral domain.

Recall also that if R is an integral domain, then we denote by $R(x)$, the field of fractions of $R[x]$ (the field of rational functions in x with coefficients in R), which consists of all quotients of the form $\frac{p(x)}{q(x)}$ where $p(x), q(x) \in R[x]$ and $q(x)$ is non-zero.

Proposition 12.2. Let I be an ideal of the ring R and let $\langle I \rangle = I[x]$ denote the ideal of $R[x]$ generated by I (the set of polynomials with coefficients in I). Then,

$$R[x]/\langle I \rangle \cong (R/I)[x].$$

In particular, if I is a prime ideal of R, then $\langle I \rangle$ is a prime ideal of $R[x]$.

Proof. We have a natural map $\varphi \colon R[x] \to (R/I)[x]$ obtained by reducing each coefficient of a polynomial in $R[x]$ modulo I. Moreover, φ is a ring homomorphism. Observe that $\ker \varphi = I[x] = \langle I \rangle$ and thus, $R[x]/\langle I \rangle \cong (R/I)[x]$ by Theorem 10.11, proving the first part. If I is a prime ideal of R, then R/I is an integral domain by Proposition 10.17, and hence, by Proposition 12.1, $(R/I)[x]$ is an integral domain. Once again, by Proposition 10.17, we conclude that $\langle I \rangle$ is a prime ideal of $R[x]$. □

Definition 12.3. The polynomial ring in variables $x_1, \dots, x_n$ with coefficients in R, denoted as $R[x_1, \dots, x_n]$ is defined inductively as

$$R[x_1, \dots, x_n] := R[x_1, \dots, x_{n-1}][x_n].$$

A more concrete formulation of this idea is as follows.

Definition 12.4. A polynomial in n variables $x_1, \ldots, x_n$ with coefficients in a commutative ring R is an expression of the form

$$\sum a_{d_1,\ldots,d_n} x_1^{d_1} \cdots x_n^{d_n}$$

where the summation is over a finite set of n-tuples $(d_1, \ldots, d_n)$ in $\mathbb{N}^n$ where $a_{d_1,\ldots,d_n} \in R$ for every such n-tuple.

Definition 12.5. Let R be a commutative ring and let Λ be a finite subset of $\mathbb{N}^n$. Let $f(x_1, \ldots, x_n)$ be a polynomial in $x_1, \ldots, x_n$ of the form

$$f(x_1, \ldots, x_n) = \sum_{(d_1,\ldots,d_n) \in \Lambda} a_{d_1,\ldots,d_n} x_1^{d_1} \cdots x_n^{d_n}.$$

Then,

$$a_{d_1,\ldots,d_n} x_1^{d_1} \cdots x_n^{d_n}$$

is called a term of the polynomial $f(x_1, \ldots, x_n)$ provided $a_{d_1,\ldots,d_n} \neq 0$. Moreover, d_i is called the degree of x^i in the above term, and $d := d_1 + \ldots + d_n$ is called the degree of this term. We call the n-tuple $(d_1, \ldots, d_n)$ the multidegree of the term.

For brevity, we represent a polynomial in n variables, $f(x_1, \ldots, x_n)$ as simply f.

Two polynomials are equal if and only if they have the same terms. The zero polynomial is defined as the polynomial having no terms. Since any non-zero polynomial must have at least one term, we can define the degree of such polynomials as follows.

Definition 12.6. Let f be a non-zero polynomial in $x_1, \ldots, x_n$. The degree or total degree of f is defined as

$$\deg f := \max \left\{ d_1 + \ldots + d_n \mid (d_1, \ldots, d_n) \in \Lambda \text{ and } a_{d_1,\ldots,d_n} \neq 0 \right\}.$$

As in the case of single variable polynomials, we define the degree of the zero polynomial as $-\infty$.

Definition 12.7. Let f be a non-zero polynomial in $x_1, \ldots, x_n$. If every term of the polynomial has the same degree d, then f is said to be a homogeneous polynomial of degree d.

Definition 12.8. A polynomial that has a single term with coefficient 1, of the form $x_1^{i_1} \cdots x_n^{i_n}$ is called a monomial.

With this, we may think of a polynomial as a finite R-linear combination of monomials. That is, a finite linear combination of monomials with coefficients in the commutative ring R.

Definition 12.9. Let f be a non-zero polynomial in $x_1, \ldots, x_n$. The sum of all monomial terms in f of degree k is called the homogeneous component of degree k in f.

If f is a nonzero polynomial having degree d, then we may write f uniquely as $f_0 + \dots f_d$ where f_k is the homogeneous component of degree k in f, for $0 \leq k \leq d$.

§§12.2. Polynomial Rings over Fields

We now consider the special case when the coefficient ring is itself a field, say $\mathbb{F}$. We define a norm N on $\mathbb{F}[x]$ with $N(p(x)) = \deg p(x)$ and $N(0) = 0$. Recall from Proposition 9.23 that the division algorithm holds. We restate this proposition for sake of completeness.

Proposition 12.10 (Division Algorithm)**.** Let $\mathbb{F}$ be a field and let $f(x), g(x) \in \mathbb{F}[x]$ with $g(x) \neq 0$. Then, there are unique polynomials $q(x), r(x) \in \mathbb{F}[x]$ such that

$$f(x) = g(x)q(x) + r(x)$$

with $r(x) = 0$ or $\deg r(x) < \deg g(x)$.

Notice that $\deg r(x) = N(r(x))$ and $\deg g(x) = N(g(x))$. Thus, the above proposition states that $\mathbb{F}[x]$ is a Euclidean domain. An immediate corollary is the following.

Corollary 12.11. If $\mathbb{F}$ is a field, then $\mathbb{F}[x]$ is a principal ideal domain and a unique factorisation domain.

Proof. This is immediate since every Euclidean domain is a principal ideal domain and every principal ideal domain is a unique factorisation domain. □

Proposition 12.12 (Gauss' Lemma)**.** Let R be a unique factorisation domain and let $\mathbb{F}$ be its field of fractions. If $p(x) \in R[x]$ is reducible in $\mathbb{F}[x]$, then $p(x)$ is reducible in $R[x]$. More precisely, if $p(x) = A(x)B(x)$ for some non-constant polynomials $A(x), B(x) \in \mathbb{F}[x]$, then there exist nonzero elements $r, s \in \mathbb{F}$ such that $a(x) := rA(x)$ and $b(x) := sB(x)$ both lie in $R[x]$ and $p(x) = a(x)b(x)$ is a factorisation in $R[x]$.

Proof. The coefficients of polynomials $A(x), B(x)$ are elements of $\mathbb{F}$ and hence quotients of elements in R. Multiplying throughout by a common denominator, we get $dp(x) = a'(x)b'(x)$, where $a'(x), b'(x) \in R[x]$, and d is a nonzero element of R. If d is a unit in R, then the proposition holds with $a(x) = d^{-1}a'(x)$ and $b(x) = b'(x)$. If not, we can write d as a product of irreducibles in R (since R is a unique factorisation domain), say $d = p_1 \cdots p_n$. Since p_1 is irreducible and R is a unique factorisation domain, p_1 is also prime by Proposition 11.32. Hence, the ideal $\langle p_1 \rangle$ is prime. Now, by Proposition 12.2, $p_1R[x]$ is a prime ideal of $R[x]$ and $(R/p_1R)[x]$ is an integral domain. Reducing the equation $dp(x) = a(x)b(x)$ modulo p_1, we obtain $0 = \overline{a'(x)}\,\overline{b'(x)}$, where the bar indicates the images of these polynomials in the quotient ring. Since $(R/p_1R)[x]$ is an integral domain, one of the factors, say $\overline{a'(x)}$ must be zero. This means that all coefficients of $a'(x)$ are divisible by p_1, so that $\frac{1}{p_1}a'(x) \in R[x]$. Thus, from the equation $dp(x) = a(x)b(x)$, we can cancel a factor of p_1 from both sides while still having an equation in $R[x]$. Proceeding the same way with all

remaining factors of d, we obtain $p(x) = a(x)b(x)$ as a factorisation in $R[x]$ with $a(x), b(x)$ being $\mathbb{F}$-multiples of $A(x), B(x)$ respectively. □

Corollary 12.13. Let R be a unique factorisation domain, let $\mathbb{F}$ be its field of fractions, and let $p(x) \in R[x]$. If the greatest common divisor of coefficients of $p(x)$ is 1, then $p(x)$ is irreducible in $R[x]$ if and only if $p(x)$ is irreducible in $\mathbb{F}[x]$. In particular, if $p(x)$ is a monic polynomial that is irreducible in $R[x]$, then $p(x)$ is irreducible in $\mathbb{F}[x]$.

Proof. By Gauss' Lemma (Proposition 12.12), if $p(x)$ is reducible in $\mathbb{F}[x]$, then it is irreducible in $R[x]$. Conversely, since the greatest common divisor of coefficients of $p(x)$ is 1, we have that if $p(x)$ is reducible in $R[x]$, then $p(x) = a(x)b(x)$ where $a(x), b(x) \in R[x]$ are both non-constant. This same factorisation also shows that $p(x)$ is reducible in $\mathbb{F}[x]$, completing the proof. □

Theorem 12.14. R is a unique factorisation domain if and only if $R[x]$ is a unique factorisation domain.

Proof. If $R[x]$ is a unique factorisation domain, then R is trivially a unique factorisation domain (since the factorisation of any element of R in $R[x]$ must be a factorisation in R itself, due to degree considerations). Suppose conversely that R is a unique factorisation domain and let $\mathbb{F}$ be its field of fractions. Let $p(x)$ be a nonzero polynomial in $R[x]$. Let d be the greatest common divisor of the coefficients of $p(x)$, so that $p(x) = dp'(x)$ where the greatest common divisor of coefficients of $p'(x)$ is 1. Notice that d is unique up to units in R (which are also units in $R[x]$) and d can be factored into irreducibles in R (which are also irreducibles in $R[x]$). It suffices to show that $p'(x)$ can be uniquely (up to units) factored into irreducibles in $R[x]$. We may hence assume that the greatest common divisor of coefficients of $p(x)$ is 1 and that $p(x)$ is not a unit in $R[x]$, that is, $\deg p(x) > 0$. By Corollary 12.11, $\mathbb{F}[x]$ is a unique factorisation domain, and hence $p(x)$ can be factored uniquely as a finite product of irreducibles in $\mathbb{F}[x]$. Using Gauss' Lemma (Proposition 12.12) and Corollary 12.13, we can show that $p(x)$ can be written as a finite product of irreducibles in $R[x]$. We leave the proof of uniqueness as an exercise to the reader. It follows from the uniqueness of decomposition in $\mathbb{F}[x]$. □

§§12.3. Irreducibility Criteria

Proposition 12.15. Let $\mathbb{F}$ be a field and let $p(x) \in \mathbb{F}[x]$. Then $p(x)$ has a factor of degree one if and only if $p(x)$ has a root in $\mathbb{F}$, that is, there is an $\alpha \in \mathbb{F}$ with $p(\alpha) = 0$.

Proof. Since $\mathbb{F}$ is a field, if $p(x)$ has a factor of degree one, we may assume it to be monic, i.e, of the form $(x - \alpha)$ for some $\alpha \in \mathbb{F}$. Then clearly $p(\alpha) = 0$. Conversely, suppose that $p(\alpha) = 0$ for some $\alpha \in \mathbb{F}$. By the division algorithm in $\mathbb{F}[x]$ (Proposition 9.23), we may write

$$p(x) = q(x)(x - \alpha) + r$$

where r is a constant (since $\deg r < \deg(x - \alpha)$). Since $p(\alpha) = 0$, r must be zero, and hence $p(x)$ has $(x - \alpha)$ as a factor. □

Proposition 12.16. A polynomial of degree two or three over a field $\mathbb{F}$ is reducible in $\mathbb{F}[x]$ if and only if it has a root in $\mathbb{F}$.

Proof. We leave this as an exercise. □

Proposition 12.17 (Rational Root Theorem)**.** Let $p(x) = a_n x^n + \ldots + a_0$ be a polynomial of degree n in $\mathbb{Z}[x]$. If $\frac{r}{s} \in \mathbb{Q}$ is in lowest terms (i.e, $\gcd(r, s) = 1$), and $\frac{r}{s}$ is a root of $p(x)$, then $r \mid a_0$ and $s \mid a_n$.

Proof. We have

$$p\left(\frac{r}{s}\right) = a_n\left(\frac{r}{s}\right)^n + \ldots + a_0 = 0 \implies a_n r^n + a_{n-1} r^{n-1} s + \ldots + a_0 s^n = 0$$

Thus, we have $a_n r^n = s(-a_{n-1} r^{n-1} - \ldots - a_0 s^{n-1})$ so that $s \mid a_n r^n$. Since $\gcd(r, s) = 1$, it follows that $s \mid a_n$. The proof for $r \mid a_0$ follows along similar lines. □

Corollary 12.18. Let $p(x) \in \mathbb{Z}[x]$ be monic. If $p(d) \neq 0$ for all integers d dividing the constant term of $p(x)$, then $p(x)$ has no root in $\mathbb{Q}$.

Proof. This follows trivially from Proposition 12.17. □

Proposition 12.19. Let I be a proper ideal[19] in the integral domain R. Let $p(x)$ be a non-constant monic polynomial in $R[x]$. If the image of $p(x)$ in $(R/I)[x]$ cannot be factored in $(R/I)[x]$ into two polynomials of smaller degree, then $p(x)$ is irreducible in $R[x]$.

Proof. Suppose that $p(x)$ cannot be factored in $(R/I)[x]$ but is reducible in $R[x]$. Then, there exist non-constant polynomials $a(x), b(x) \in R[x]$ such that $p(x) = a(x)b(x)$. Moreover, $a(x), b(x)$ are both monic since $p(x)$ is monic. By Proposition 12.2, reducing the coefficients modulo I gives us a non-constant factorisation in $(R/I)[x]$, which is a contradiction. □

Proposition 12.20 (Eisenstein-Schönemann Criterion)**.** Let P be a prime ideal of the integral domain R. Let $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$ be a polynomial in $R[x]$ ($n \geq 1$). If $a_{n-1}, \ldots, a_0 \in P$ and $a_0 \notin P^2$, then $f(x)$ is irreducible in $R[x]$.

Proof. Suppose $f(x)$ were reducible in $R[x]$, say $f(x) = a(x)b(x)$ where $a(x), b(x) \in R[x]$ are non-constant polynomials. Reducing this equation modulo P, we obtain $x^n = \overline{a(x)}\,\overline{b(x)}$ in $(R/P)[x]$, where the bar indicates polynomials whose coefficients are reduced modulo P. Since P is a prime

[19] That is, I is an ideal of R and $I \neq R$.

ideal, R/P is an integral domain, it follows that the constant terms of both $\overline{a(x)}$ and $\overline{b(x)}$ are 0, that is, the constant terms of both $a(x)$ and $b(x)$ are elements of P. However, this implies that the constant term a_0 of $f(x)$ is an element of P^2, which is a contradiction. □

Proposition 12.20 is most frequently used in the case of $\mathbb{Z}[x]$, so we state this result explicitly as a corollary.

Corollary 12.21 (Eisenstein-Schönemann Criterion for Integers)**.** Let p be a prime in $\mathbb{Z}$ and let $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$ be a polynomial in $\mathbb{Z}[x]$, with $n \geq 1$. If $p \mid a_i$ for all $i \in \{0, \ldots, n-1\}$ but $p^2 \nmid a_0$, then $f(x)$ is irreducible in both $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

Proof. Trivial. □

§13. Algebraic Extensions

§§13.1. The Prime Subfield

Definition 13.1. A number field is a subfield of $\mathbb{C}$.

Proposition 13.2. Any number field contains the field $\mathbb{Q}$.

Proof. We leave this as a simple exercise to the reader. □

Definition 13.3. The characteristic of a field $\mathbb{F}$ is the smallest positive integer p such that $p \cdot 1 = 0$ if such a p exists, and is defined to be 0 otherwise. Here, $p \cdot 1$ is defined as

$$p \cdot 1 := \underbrace{1 + \cdots + 1}_{p \text{ times}}.$$

We denote the characteristic of $\mathbb{F}$ as $\mathrm{ch}(\mathbb{F})$.

Proposition 13.4. Let $\mathbb{F}$ be a field. Then, the following are true.

1. $\mathrm{ch}(\mathbb{F})$ is either 0 or a prime.
2. If $\mathrm{ch}(\mathbb{F}) = p$, a prime and if $n \cdot 1 = 0$ for some $n \in \mathbb{Z}$[20], then $p \mid n$.
3. If $\mathrm{ch}(\mathbb{F}) = p$, a prime, then for any $\alpha \in \mathbb{F}$,
$$p \cdot \alpha = \underbrace{\alpha + \cdots + \alpha}_{p \text{ times}} = 0.$$
4. If $\mathrm{ch}(\mathbb{F}) = p$, a prime, then for any $x, y \in \mathbb{F}$,
$$(x + y)^p = x^p + y^p.$$

Proof.

1. Observe that
$$m \cdot 1 + n \cdot 1 = (m + n) \cdot 1 \text{ and}$$
$$(m \cdot 1)(n \cdot 1) = (mn) \cdot 1$$
for $m, n \in \mathbb{N}^+$. It follows that $\mathrm{ch}(\mathbb{F})$ is either 0 or prime. Suppose that $\mathrm{ch}(\mathbb{F})$ is some composite number $n = ab$ ($a, b \in \mathbb{N}^+$ and $a, b < n$). We then have $n \cdot 1 = 0 \implies (ab) \cdot 1 = 0 \implies (a \cdot 1)(b \cdot 1) = 0$. Since $\mathbb{F}$ is a field, it follows that one of $a \cdot 1$ or $b \cdot 1$ must be zero, which is a contradiction since $a, b < n$.

[20] We define $(-n) \cdot 1 := -(n \cdot 1)$ for positive n, and define $0 \cdot 1 := 0$.

2. This follows trivially from the first part itself, along with an elementary application of the Division Algorithm.
3. This is trivial as well, since $p \cdot \alpha = p \cdot (1\alpha) = (p \cdot 1)\alpha = 0$.
4. This is again trivial and follows from the the third part. (Hint: Binomial Theorem)

□

Corollary 13.5. Let $\mathbb{F}$ be a field of characteristic $p > 0$. Then, for any $a_1, \ldots, a_n \in \mathbb{F}$, we have

$$(a_1 + \cdots + a_n)^p = a_1^p + \cdots + a_n^p.$$

Proof. Apply induction on part 4 of Proposition 13.4. □

Definition 13.6. Let $\mathbb{F}$ be a field. The prime subfield of $\mathbb{F}$ is the subfield of $\mathbb{F}$ generated by the multiplicative identity, $1 \in \mathbb{F}$.

Proposition 13.7. Let $\mathbb{F}$ be a field. If $\mathrm{ch}(\mathbb{F}) = 0$, then the prime subfield of $\mathbb{F}$ is isomorphic to $\mathbb{Q}$. If $\mathrm{ch}(\mathbb{F}) = p$ for some prime p, then the prime subfield of $\mathbb{F}$ is isomorphic to $\mathbb{Z}_p =: \mathbb{F}_p$.

Proof. We have the natural ring homomorphism $\varphi\colon \mathbb{Z} \to \mathbb{F}$ defined by

$$\varphi(n) = n \cdot 1.$$

Note that $\ker \varphi = (\mathrm{ch}(\mathbb{F})\mathbb{Z}$. Depending on the characteristic, quotienting by the kernel gives us an injection of either $\mathbb{Z}$ or $\mathbb{Z}_p$ into $\mathbb{F}$. Since $\mathbb{F}$ is a field, the Universal Property tells us that $\mathbb{F}$ must either contain an isomorphic copy of $\mathbb{Q}$, the field of fractions of $\mathbb{Z}$ (when $\mathrm{ch}(\mathbb{F}) = 0$), or an isomorphic copy of $\mathbb{F}_p := \mathbb{Z}_p$, the field of fractions of $\mathbb{Z}_p$ (when $\mathrm{ch}(\mathbb{F}) = p$). □

§§13.2. Extensions and Degrees

Definition 13.8. Let $\mathbb{F}$ be a subfield of $\mathbb{K}$. We say that $\mathbb{K}$ is an extension field of $\mathbb{F}$ and we call $\mathbb{F}$ the base field. We denote this as $\mathbb{K}/\mathbb{F}$.

Remark 13.9. Note that $\mathbb{K}/\mathbb{F}$ is not a quotient. In fact, since the only ideals of $\mathbb{K}$ are 0 and $\mathbb{K}$, quotienting does not make sense in the first place.

Definition 13.10. Let $\mathbb{K}/\mathbb{F}$ be a field extension. We may regard $\mathbb{K}$ as a vector space over $\mathbb{F}$. We denote $\dim_{\mathbb{F}} \mathbb{K}$ as $[\mathbb{K}\colon \mathbb{F}]$ and call it the degree of the field extension $\mathbb{K}/\mathbb{F}$.

Definition 13.11. A field extension $\mathbb{K}/\mathbb{F}$ is said to be a finite extension if $[\mathbb{K}\colon \mathbb{F}]$ is finite.

Definition 13.12. A field extension $\mathbb{K}/\mathbb{F}$ is said to be a simple extension if there exists $\alpha \in \mathbb{K}$ such that $\mathbb{K} = \mathbb{F}(\alpha)$.

Definition 13.13. Let $\mathbb{K}/\mathbb{F}$ be a field extension and let $\alpha \in \mathbb{K}$. α is said to be algebraic over $\mathbb{F}$ if there exists a non-zero polynomial $f(x) \in \mathbb{F}[x]$ such that $f(\alpha) = 0$.

α is called transcendental over $\mathbb{F}$ if it is not algebraic over $\mathbb{F}$.

If every element of $\mathbb{K}$ is algebraic over $\mathbb{F}$, then $\mathbb{K}/\mathbb{F}$ is called an algebraic extension.

Proposition 13.14. Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be fields and let $\alpha \in \mathbb{K}$. If α is algebraic over $\mathbb{F}$, then α is algebraic over $\mathbb{E}$.

Proof. We leave this as an exercise to the reader. □

Corollary 13.15. Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be fields. If $\mathbb{K}/\mathbb{F}$ is algebraic, then so are $\mathbb{K}/\mathbb{E}$ and $\mathbb{E}/\mathbb{F}$.

Proposition 13.16. Every finite extension is an algebraic extension.

Proof. Let $\mathbb{K}/\mathbb{F}$ be a finite extension and let $n := \dim_{\mathbb{F}} \mathbb{K}$. Let $a \in \mathbb{K}$ be an arbitrary element. We show that α is algebraic over $\mathbb{F}$. Since the set $\{1, \alpha, \ldots, \alpha^n\}$ has $n+1$ elements, it is linearly dependent over $\mathbb{F}$. Thus, there $b_0, \ldots, b_n \in \mathbb{F}$ not all 0 such that

$$b_0 + b_1 a + \cdots + b_n a^n = 0.$$

Thus, $f(x) := b_0 + b_1 x + \cdots + b_n x^n \in \mathbb{F}[x]$ is a non-zero polynomial satisfying $f(\alpha) = 0$. □

Example 13.17.

1. Consider the extensions $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. It is known that $\pi \in \mathbb{R}$ is transcendental over $\mathbb{Q}$. A consequence of this is that $\pi\imath \in \mathbb{C}$ is also transcendental over $\mathbb{Q}$. However, $\pi\imath$ is algebraic over $\mathbb{R}$ since it satisfies the non-zero polynomial $x^2 + \pi^2 \in \mathbb{R}[x]$.

   Thus, the property of being algebraic depends on the base field. In particular, we have shown that $\mathbb{C}/\mathbb{Q}$ is not an algebraic extension, whereas $\mathbb{C}/\mathbb{R}$ is, by Proposition 13.16, since $[\mathbb{C}\colon \mathbb{R}] = 2$.

2. Let $\mathbb{K}$ be a finite field and let $\mathbb{F}$ be its prime subfield. Then, $\mathbb{K}$ is a finite dimensional vector space over $\mathbb{F}$ (since $\mathbb{K}$ is finite) and thus, $\mathbb{K}/\mathbb{F}$ is an algebraic extension by Proposition 13.16.

Proposition 13.18. Let $\mathbb{K}/\mathbb{F}$ be a field extension and let $\alpha \in \mathbb{K}$ be algebraic over $\mathbb{F}$. Then, the following are true (with "irreducible" meaning "irreducible" over $\mathbb{F}[x]$).

1. There exists a unique monic irreducible polynomial $f(x) \in \mathbb{F}[x]$ such that $f(\alpha) = 0$.
2. $f(x)$ generates the kernel of the map $\mathbb{F}[x] \to \mathbb{F}[\alpha] \subseteq \mathbb{K}$ defined by $p(x) \mapsto p(\alpha)$.
3. If $g(x) \in \mathbb{F}[x]$ is such that $g(\alpha) = 0$, then $f(x) \mid g(x)$.
4. $f(x)$ has the least positive degree among all polynomials in $\mathbb{F}[x]$ that are satisfied by α.

Proof. Define $\psi\colon \mathbb{F}[x] \to \mathbb{K}$ by $p(x) \mapsto p(\alpha)$. Since α is algebraic over $\mathbb{F}[x]$, $I := \ker\psi$ is non-zero. By Corollary 12.11, $\mathbb{F}[x]$ is a principal ideal domain, and hence $I = \langle f(x)\rangle$ for some $f(x) \in \mathbb{F}[x]$. Moreover, $f(x)$ is non-zero since I is non-zero. By the Isomorphism Theorem for Rings, $\mathbb{F}[x]/I$ is isomorphic to a subring of $\mathbb{K}$, and hence is an integral domain. By Proposition 10.17, I is a prime ideal, and hence $f(x)$ is prime. By Proposition 11.30, $f(x)$ is irreducible. Scaling by an appropriate factor, we may assume that $f(x)$ is monic. Clearly, any other $g(x)$ that is satisfied by α must be an element of I, and thus $f(x) \mid g(x)$. In particular, if $g(x)$ is irreducible and monic, then $f(x) \mid g(x) \implies g(x) = af(x)$ for some $a \in \mathbb{F}^\times$. Since $g(x)$ is also monic, we have that $a = 1$, giving us $f(x) = g(x)$. Thus, such an $f(x)$ is unique.

This proves the first three parts. The fourth part follows from the third via a simple application of the Division Algorithm. □

Definition 13.19. Let $\mathbb{K}/\mathbb{F}$ be a field extension, and let $\alpha \in \mathbb{K}$ be algebraic over $\mathbb{F}$. The unique irreducible monic polynomial in $\mathbb{F}[x]$ that is satisfied by α is called the minimal polynomial of α over $\mathbb{F}$. We denote this as $\operatorname{irr}(\alpha, \mathbb{F})$.

The degree of $\operatorname{irr}(\alpha, \mathbb{F})$ is called the degree of α over $\mathbb{F}$ and is denoted as $\deg_{\mathbb{F}} \alpha$.

Example 13.20. Let $\alpha \in \mathbb{C}$ be a square root of ι. Then, α satisfies $f(x) := x^4 + 1$. We may show that $f(x) = \operatorname{irr}(\alpha, \mathbb{Q})$, and hence $\deg_{\mathbb{Q}} \alpha = 4$. However, α also satisfies $x^2 - \iota$, so that $\operatorname{irr}(\alpha, \mathbb{Q}(\iota)) = x^2 - \iota$, and $\deg_{\mathbb{Q}(\iota)} \alpha = 2$. Hence, the degree also depends on the base field.

Proposition 13.21. Let $\mathbb{K}/\mathbb{F}$ be a field extension and $\alpha \in \mathbb{K}$ be algebraic over $\mathbb{F}$. Let $f(x) := \operatorname{irr}(\alpha, \mathbb{F})$ and let $n := \deg f(x)$. Then,

1. $\mathbb{F}[\alpha] = \mathbb{F}(\alpha) \cong \mathbb{F}[x]/\langle f(x)\rangle$.
2. $\dim_{\mathbb{F}}(\mathbb{F}(\alpha)) = n$ and $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is an $\mathbb{F}$-basis of $\mathbb{F}(\alpha)$.

Proof. Consider the substitution homomorphism $\psi\colon \mathbb{F}[x] \to \mathbb{F}[\alpha]$ defined by $p(x) \mapsto p(\alpha)$. By Proposition 13.18, $\ker\psi = \langle f(x)\rangle$. By Corollary 12.11, $\mathbb{F}[x]$ is a principal ideal domain. Hence, $f(x)$ is a prime element by Proposition 11.30, and hence $\langle f(x)\rangle$ is a prime ideal. Since $f(x) \neq 0$, we get that $\langle f(x)\rangle$ is also a maximal ideal, by Proposition 11.22. Moreover, ψ is clearly surjective so that

$\operatorname{im}\psi = \mathbb{F}[\alpha]$. Hence, by the Isomorphism Theorem for Rings, $\mathbb{F}[x]/\langle f(x)\rangle \cong \mathbb{F}[\alpha]$. Since $\langle f(x)\rangle$ is maximal, $\mathbb{F}[x]/\langle f(x)\rangle$ is a field by Proposition 10.17. Thus, $\mathbb{F}[\alpha] = \mathbb{F}(\alpha)$.

Consider the set $B = \{1, \alpha, \ldots, \alpha^{n-1}\}$. Using $f(x)$, we may write all higher powers of α as an $\mathbb{F}$-linear combination of elements of B. Hence, B spans $\mathbb{F}[\alpha]$. Now, suppose $a_0, \ldots, a_{n-1} \in \mathbb{F}$ satisfy

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = 0.$$

Then, $g(x) := a_0 + a_1 x + \ldots + a_{n-1}x^{n-1} \in \mathbb{F}[x]$ is a polynomial that is satisfied by α. However since $\deg g(x) < \deg f(x)$, by Proposition 13.18, we get that $g(x) = 0$. This proves linear independence. □

Corollary 13.22. Let $\mathbb{K}/\mathbb{F}$ be a field extension and let $\alpha \in \mathbb{K}$ be algebraic over $\mathbb{F}$. Then, $\mathbb{F}(\alpha)/\mathbb{F}$ is a finite, and hence, algebraic extension.

Definition 13.23. Let $\mathbb{F} \subseteq \mathbb{E}_1, \mathbb{E}_2$ be fields. A $\mathbb{F}$-homomorphism from $\mathbb{E}_1$ to $\mathbb{E}_2$ is a field homomorphism $\varphi\colon \mathbb{E}_1 \to \mathbb{E}_2$ that fixes $\mathbb{F}$.

If φ is an isomorphism, we call it an $\mathbb{F}$-isomorphism.

Proposition 13.24. Let $\mathbb{K}/\mathbb{F}$ be a field extension and let $\alpha, \beta \in \mathbb{K}$ be algebraic over $\mathbb{F}$. Then, the following two statements are equivalent.

1. There exists an $\mathbb{F}$-isomorphism $\psi\colon \mathbb{F}(\alpha) \to \mathbb{F}(\beta)$ such that $\psi(\alpha) = \beta$.
2. $\operatorname{irr}(\alpha, \mathbb{F}) = \operatorname{irr}(\beta, \mathbb{F})$.

Proof. (1 $\implies$ 2) Let $\psi\colon \mathbb{F}(\alpha) \to \mathbb{F}(\beta)$ be as mentioned. Let $f(x) := \operatorname{irr}(\alpha, \mathbb{F})$ and $g(x) := \operatorname{irr}(\beta, \mathbb{F})$. Then,

$$\begin{aligned} 0 &= \psi(0) \\ &= \psi(f(\alpha)) \\ &= f(\psi(\alpha)) && \text{(since } \psi \text{ is an } \mathbb{F}\text{-isomorphism)} \\ &= f(\beta). \end{aligned}$$

Thus, by Proposition 13.18, $g(x) \mid f(x)$. Since both are irreducible and monic, $g(x) = f(x)$.

(2 $\implies$ 1) Let $f(x) := \operatorname{irr}(\alpha, \mathbb{F}) = \operatorname{irr}(\beta, \mathbb{F})$. The isomorphisms $\mathbb{F}(\alpha) \cong \mathbb{F}[x]/\langle f(x)\rangle \cong \mathbb{F}(\beta)$ are $\mathbb{F}$-isomorphisms, and thus, so is their composition. □

Definition 13.25. A field extension $\mathbb{K}/\mathbb{F}$ is called a quadratic extension if $[\mathbb{K}\colon \mathbb{F}] = 2$.

Remark 13.26. Every quadratic extension is simple. If $\mathbb{K}/\mathbb{F}$ is a quadratic extension and $\alpha \in K \setminus \mathbb{F}$, then $[\mathbb{F}(\alpha)\colon \mathbb{F}] > 1$, and thus $[\mathbb{F}(\alpha)\colon \mathbb{F}] = 2$. Thus, $\mathbb{F}(\alpha) = \mathbb{K}$ and $\mathbb{K}/\mathbb{F}$ is simple.

Definition 13.27. A chain of fields $\mathbb{F}_1 \subseteq \ldots \subseteq \mathbb{F}_n$ is called a tower of fields if $\mathbb{F}_i$ is a subfield of $\mathbb{F}_{i+1}$ for all $i = 1, \ldots, n-1$.

Proposition 13.28 (Tower Law)**.** Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be a tower of fields. Then,

$$[\mathbb{K} : \mathbb{F}] = [\mathbb{K} : \mathbb{E}] \cdot [\mathbb{E} : \mathbb{F}].$$

In particular, the left side is ∞ iff the right side is.

Proof. If $\mathbb{K}/\mathbb{F}$ is a finite extension, then so are $\mathbb{K}/\mathbb{E}$ (any finite basis for $\mathbb{K}/\mathbb{F}$ is a spanning set for $\mathbb{K}/\mathbb{E}$) and $\mathbb{E}/\mathbb{F}$ ($\mathbb{E}$ is an $\mathbb{F}$-subspace of $\mathbb{K}$). Thus, if either of $\mathbb{K}/\mathbb{E}$ or $\mathbb{E}/\mathbb{F}$ is not a finite extension, then neither is $\mathbb{K}/\mathbb{F}$.

Now, suppose $[\mathbb{K} : \mathbb{E}] =: n$ and $[\mathbb{E} : \mathbb{F}] =: m$ are both finite. Let $\{\alpha_i\}_{i=1}^n \subseteq \mathbb{K}$ be an $\mathbb{E}$-basis of $\mathbb{K}$, and let $\{\beta_j\}_{j=1}^m \subseteq \mathbb{E}$ be an $\mathbb{F}$-basis of $\mathbb{E}$. Now, put $B := \{\alpha_i\beta_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$. We show that B is an $\mathbb{F}$-basis for $\mathbb{K}$.

Let $a \in \mathbb{K}$ be arbitrary. Then,

$$a = \sum_{i=1}^{n} a_i \alpha_i$$

for $a_i \in \mathbb{E}$. For each $i = 1, \ldots, n$, we may write

$$a_i = \sum_{j=1}^{m} b_{ij} \beta_j$$

for $b_{ij} \in \mathbb{F}$. Now,

$$a = \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij}(\alpha_i \beta_j)$$

is an $\mathbb{F}$-linear combination of elements of B. Hence, B spans $\mathbb{K}$.

Now, suppose $\{b_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq m\} \subseteq \mathbb{F}$ satisfies

$$\sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij}(\alpha_i \beta_j) = 0.$$

We may group the terms to get

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{m} b_{ij} \alpha_i \right] \beta_j = 0.$$

Linear independence of $\{\beta_j\}_{j=1}^m$ forces $\sum_{j=1}^m b_{ij}\alpha_i = 0$ for all i, Now, linear independence of $\{\alpha_i\}_{i=1}^n$ forces $b_{ij} = 0$ for all i, j, which proves linear independence. It remains to show that $|B| = nm$, which we leave as an exercise. □

Corollary 13.29. Let $\mathbb{K}/\mathbb{F}$ be a finite extension and let $\alpha \in \mathbb{K}$. Then, $\deg_{\mathbb{F}} \alpha$ divides $[\mathbb{K}\colon \mathbb{F}]$.

Proof. This follows from the Tower Law by considering the tower $\mathbb{F} \subseteq \mathbb{F}(\alpha) \subseteq \mathbb{K}$. Note that since $\mathbb{K}/\mathbb{F}$ is a finite extension, α is algebraic over $\mathbb{F}$. □

Proposition 13.30. Let $\mathbb{K}/\mathbb{F}$ be a field extension and let $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$ be algebraic over $\mathbb{F}$. Then, $\mathbb{F}(\alpha_1, \ldots, \alpha_n)$ is a finite (and hence, algebraic) extension of $\mathbb{F}$.

Proof. We leave the details of the proof to the reader. The following tower might be helpful.

$$\mathbb{F} \subseteq \mathbb{F}(\alpha_1) \subseteq \mathbb{F}(\alpha_1, \alpha_2) \cdots \subseteq \mathbb{F}(\alpha_1, \ldots, \alpha_n)$$ □

Corollary 13.31. Let $\mathbb{E}/\mathbb{F}$ and $\mathbb{K}/\mathbb{E}$ be algebraic extensions. Then, $\mathbb{K}/\mathbb{F}$ is an algebraic extension.

Proof. Let $\alpha \in \mathbb{K}$ and let $\operatorname{irr}(\alpha, \mathbb{E}) = a_0 + a_1 x + \cdots a_{n-1}x^{n-1} + x^n =: f(x)$. Let $\mathbb{L} = \mathbb{F}(a_0, \ldots, a_{n-1})$. Then, $\mathbb{L}/\mathbb{F}$ is a finite extension since each $a_i \in \mathbb{E}$ is algebraic over $\mathbb{F}$. Moreover, $0 \neq f(x) \in \mathbb{L}[x]$. Thus, α is algebraic over $\mathbb{L}$ and $\mathbb{L}(\alpha)/\mathbb{L}$ is a finite extension. By the Tower Law, $\mathbb{L}(\alpha)/\mathbb{F}$ is a finite extension. Hence, α is algebraic over $\mathbb{F}$ by Proposition 13.16. □

Corollary 13.32. Let $\mathbb{K}/\mathbb{F}$ be a field extension. Then,

$$\mathbb{A} := \{\alpha \in \mathbb{K} \mid \alpha \text{ is algebraic over } \mathbb{F}\}$$

is a subfield of $\mathbb{K}$ containing $\mathbb{F}$. Moreover, $\mathbb{A}/\mathbb{F}$ is an algebraic extension.

Proof. It is clear that $\mathbb{A}$ contains $\mathbb{F}$. We now show that $\mathbb{A}$ is a subfield. Let $\alpha, \beta \in \mathbb{A}$ with $\beta \neq 0$. Then, $\mathbb{L} := \mathbb{F}(\alpha, \beta)$ is a finite extension of $\mathbb{F}$. Thus, all elements of $\mathbb{L}$ are algebraic over $\mathbb{F}$. In particular, so are $\alpha \pm \beta$, $\alpha\beta$ and $\alpha\beta^{-1}$. □

§§13.3. Compositum of Fields

Definition 13.33. Let $\mathbb{E}_1, \mathbb{E}_2 \subseteq \mathbb{K}$ be fields. The compositum of $\mathbb{E}_1$ and $\mathbb{E}_2$ is the smallest subfield of $\mathbb{K}$ containing $\mathbb{E}_1$ and $\mathbb{E}_2$. We denote this by $\mathbb{E}_1\mathbb{E}_2$.

Example 13.34.

1. Suppose $\mathbb{F} \subseteq \mathbb{E}_1, \mathbb{E}_2 \subseteq \mathbb{K}$ and $\mathbb{E}_1 = \mathbb{F}(\alpha_1, \ldots, \alpha_n)$. Then,

$$\mathbb{E}_1\mathbb{E}_2 = \mathbb{E}_2(\alpha_1, \ldots, \alpha_n).$$

2. Let m and n be coprime positive integers. Consider the subfields $\mathbb{F} := \mathbb{Q}(\zeta_m)$ and $\mathbb{E} := \mathbb{Q}(\zeta_n)$ of $\mathbb{C}$. Then,
$$\mathbb{E}\mathbb{F} = \mathbb{Q}(\zeta_{mn}).$$
It is clear that $\mathbb{E}\mathbb{F} \subseteq \mathbb{Q}(\zeta_{mn})$, since $\zeta_n = \zeta_{mn}^m$ and $\zeta_m = \zeta_{mn}^n$. On the other hand, since $\gcd(m,n) = 1$, by Bézout's Lemma, there exist integers $a, b \in \mathbb{Z}$ such that $am + bn = 1$. Thus,
$$\frac{a}{n} + \frac{b}{m} = \frac{1}{mn}$$
which gives us $\zeta_{mn} = \zeta_n^a \zeta_m^b$ and thus $\mathbb{Q}(\zeta_{mn}) \subseteq \mathbb{E}\mathbb{F}$.

Proposition 13.35. Let $\mathbb{F}$ be a field that is a subring of an integral domain R. If R is finite dimensional as an $\mathbb{F}$ vector space, then R is a field.

Proof. It suffices to show that every non-zero element in R has a multiplicative inverse in R. Let $a \in R$ be arbitrary with $a \neq 0$. Since $\dim_{\mathbb{F}} R < \infty$, there is a smallest $n \geq 1$ such that the set $\{1, a, \ldots, a^n\}$ is linearly dependent. Then, let $b_0, \ldots, b_n \in \mathbb{F}$ be not all zero such that
$$b_0 + b_1 a + \cdots + b_n a^n = 0.$$
If $b_n = 0$, then the minimality of n is contradicted. Hence, $b_n \neq 0$. If $b_0 = 0$, we may cancel a (since R is an integral domain and $a \neq 0$) to again contradict the minimality of n. Thus, $b_0 \neq 0$. Now, we have
$$a(b_1 + \cdots b_n a^{n-1}) = -b_0$$
which shows that
$$-\frac{1}{b_0}(b_1 + \cdots + b_n a^{n-1}) \in R$$
is a multiplicative inverse of a. □

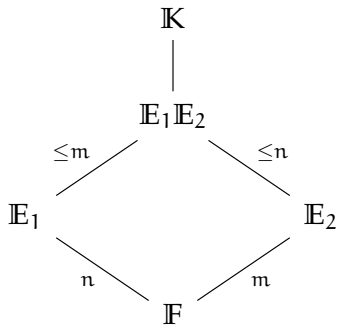
Proposition 13.36. Let $\mathbb{F} \subseteq \mathbb{E}_1, \mathbb{E}_2 \subseteq \mathbb{K}$ be fields. Consider
$$\mathbb{L} := \left\{ \sum_{i=1}^{n} \alpha_i \beta_i \mid n \in \mathbb{N}, \alpha_i \in \mathbb{E}_1, \beta_i \in \mathbb{E}_2 \right\}.$$
That is, $\mathbb{L}$ is the set of all finite sums of products of elements of $\mathbb{E}_1$ and $\mathbb{E}_2$.

Suppose $d := [\mathbb{E}_1 : \mathbb{F}] \cdot [\mathbb{E}_2 : \mathbb{F}] < \infty$. Then, $\mathbb{L} = \mathbb{E}_1\mathbb{E}_2$ and $[\mathbb{L} : \mathbb{F}] \leq d$. If $[\mathbb{E}_1 : \mathbb{F}]$ and $[\mathbb{E}_2 : \mathbb{F}]$ are coprime, then equality holds.

Proof. We leave it as an exercise to the reader to show that $\mathbb{L}$ is a subring of $\mathbb{K}$. Thus, $\mathbb{L}$ is an integral domain. Let $n := [\mathbb{E}_1 : \mathbb{F}]$ and $m := [\mathbb{E}_2 : \mathbb{F}]$. If $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_m\}$ are $\mathbb{F}$-bases for $\mathbb{E}_1$ and $\mathbb{E}_2$, then the set $\{\alpha_i \beta_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ clearly spans $\mathbb{L}$ over $\mathbb{F}$. Hence $\dim_{\mathbb{F}} \mathbb{L} \leq mn = d$. In particular, $\dim_{\mathbb{F}} \mathbb{L}$ is finite, so that $\mathbb{L}$ is a field by Proposition 13.35.

Lastly, note that $[\mathbb{E}_i : \mathbb{F}]$ divides $[\mathbb{L} : \mathbb{F}]$ by the Tower Law. In particular, if $\gcd(m, n) = 1$, then $mn \mid [\mathbb{L} : \mathbb{F}]$. Since $[\mathbb{L} : \mathbb{F}] \leq mn$, we are done.

Diagrammatically, this is depicted as



□

§§13.4. Splitting Fields

Definition 13.37. Let $\mathbb{F}$ be a field and $f(x) \in \mathbb{F}[x]$ be a non-constant polynomial of degree n with leading coefficient $a \in \mathbb{F}^\times$. A field $\mathbb{K} \supseteq \mathbb{F}$ is called a splitting field of $f(x)$ over $\mathbb{F}$ if there exist (not necessarily distinct) $r_1, \ldots, r_n \in \mathbb{K}$ such that $f(x) = a(x - r_1) \cdots (x - r_n)$ and $\mathbb{K} = \mathbb{F}(r_1, \ldots, r_n)$.

Example 13.38. Consider $\mathbb{F} = \mathbb{Q}$, $f(x) = x^2 + 1 \in \mathbb{Q}[x]$, and $\mathbb{K} = \mathbb{C}$. Although $f(x)$ does factor linearly over $\mathbb{C}$ as $(x + \iota)(x - \iota)$, $\mathbb{C}$ is **not** a splitting field of $f(x)$ over $\mathbb{Q}$ since $\mathbb{C} \neq \mathbb{Q}(\iota, -\iota)$. On the other hand, $\mathbb{C}$ *is* a splitting field of $f(x)$ over $\mathbb{R}$ since $\mathbb{C} = \mathbb{R}(\iota, -\iota)$.

Corollary 13.39. Let $f(x) \in \mathbb{F}[x]$ be non-constant and let $\mathbb{K}$ be a splitting field of $f(x)$ over $\mathbb{F}$. Then, $\mathbb{K}/\mathbb{F}$ is an algebraic extension.

Proof. This follows trivially from Proposition 13.30. □

Theorem 13.40. Let $\mathbb{F}$ be a field and let $f(x) \in \mathbb{F}[x]$ be non-constant. Then, there exists a field $\mathbb{K} \supseteq \mathbb{F}$ such that $f(x)$ has a root in $\mathbb{K}$.

Proof. Let $g(x)$ be an irreducible factor of $f(x)$. Since $\mathbb{F}[x]$ is a principal ideal domain (Corollary 12.11), $g(x)$ is also a prime element (Proposition 11.30), so that $\langle g(x) \rangle$ is a prime ideal. Since $g(x)$ is non-zero, $\langle g(x) \rangle$ is also a maximal ideal (Proposition 11.22). Now, put $\mathbb{K} = \mathbb{F}[x]/\langle g(x) \rangle$. $\mathbb{K}$ is clearly a field by Proposition 10.17. $\mathbb{K}$ clearly contains $\mathbb{F}$ as a subfield via the identification $a \mapsto \overline{a}$, where the bar indicates the image in the quotient. Moreover, $\overline{x}$ is a root of $g(x)$ since $g(\overline{x}) = \overline{g(x)} = 0$ in the quotient. □

Theorem 13.41 (Existence of Splitting Field)**.** Let $\mathbb{F}$ be a field. Any polynomial $f(x) \in \mathbb{F}[x]$ of positive degree has a splitting field.

Proof. Let $n := \deg f(x)$. By Theorem 13.40, there exists a field $\mathbb{F}_1 \supseteq \mathbb{F}$ such that $f(x)$ has a root, say a_1, in $\mathbb{F}_1$. Now,

$$f(x) = (x - a_1) \cdot f_1(x)$$

where $\deg f_1(x) = n - 1$. Continuing inductively, we get fields

$$\mathbb{F}_n \supseteq \cdots \supseteq \mathbb{F}_1 \supseteq \mathbb{F}$$

with $a_i \in \mathbb{F}_i$ such that

$$f(x) = a(x - a_1) \cdots (x - a_n).$$

Then, $\mathbb{K} = \mathbb{F}(a_1, \ldots, a_n) \subseteq \mathbb{F}_n$ is a splitting field. □

§14. Symmetric Polynomials

Definition 14.1. Let R be a ring and let $S = R[x_1, \dots, x_n]$. Let S_n denote the symmetric group. Then, any $\tau \in S_n$ induces an automorphism $g_\tau \colon S \to S$, defined by

$$g_t\left(f(x_1, \dots, x_n)\right) := f\left(x_{\tau(1)}, \dots, x_{\tau(n)}\right).$$

Example 14.2. Suppose $n = 3$ and $\tau = (1\,2\,3)$. Then, the polynomial $x_1 + x_2^2 + x_3^3$ in $\mathbb{Z}[x]$ is mapped to $x_2 + x_3^2 + x_1^3$ via g_τ.

Definition 14.3. A polynomial $f \in R[x_1, \dots, x_n]$ is said to be a symmetric polynomial (in n variables) if $g_\tau(f) = f$ for all $\tau \in S_n$.

Definition 14.4. Let $S = R[x_1, \dots, x_n]$ and consider $f(T) \in S[T]$ given by

$$f(T) = (T - x_1) \cdots (T - x_n).$$

We may write $f(T)$ as

$$f(T) = T^n - \sigma_1 T^{n-1} + \cdots + (-1)^n \sigma_n,$$

for $\sigma_1, \dots, \sigma_n \in S$. Then, $\sigma_1, \dots, \sigma_n$ are symmetric polynomials, called the elementary symmetric polynomials (in n variables).

Remark 14.5. Note that one can explicitly write down the elementary symmetric polynomials, as follows.

$$\begin{aligned}
\sigma_1 &= \sum_{i=1}^{n} x_i \\
\sigma_2 &= \sum_{1 \le i_1 < i_2 \le n} x_{i_1} x_{i_2} \\
&\vdots \\
\sigma_n &= x_1 \cdots x_n.
\end{aligned}$$

It is now easy to verify that these are all indeed symmetric polynomials.

Definition 14.6. Given an elementary symmetric polynomial $\sigma_i \in R[x_1, \dots, x_n]$ (for $n \ge 2$), we define the elementary symmetric polynomial σ_i^0 in $(n-1)$ variables as

$$\sigma_i^0 := \sigma_i(x_1, \dots, x_{n-1}, 0).$$

Example 14.7. Consider $n = 3$ and $\sigma_2 = x_1x_2 + x_1x_3 + x_2x_3$. Then, $\sigma_2^0 = x_1x_2$ is the second symmetric polynomial in 2 variables. In fact, any elementary symmetric polynomial in $(n-1)$ variables is of the form σ_i^0 for the corresponding elementary symmetric polynomial σ_i in n variables.

Theorem 14.8 (Fundamental Theorem of Symmetric Polynomials)**.** Let R be a commutative ring. Then, every symmetric polynomial in $S := R[x_1, \ldots, x_n]$ is a polynomial in the elementary symmetric polynomials in a unique way. More precisely, if $f(x_1, \ldots, x_n) \in S$ is symmetric, then there exists a unique $g \in R[u_1, \ldots, u_n]$ such that

$$g(\sigma_1, \ldots, \sigma_n) = f(x_1, \ldots, x_n)$$

where the above equality is in S.

Proof. **Existence.** We apply induction on n, the number of variables. The case $n = 1$ is clear since every polynomial is symmetric and $\sigma_1 = x_1$. Thus, we may choose g to have the same coefficients as f.

Suppose the theorem is true for $n-1$ variables. We now apply induction on $\deg f$. If f is constant, then again choosing g to have the same coefficients as f works. Suppose $\deg f \geq 1$. Now, we define

$$f^0 := f(x_1, \ldots, x_{n-1}, 0) \in R[x_1, \ldots, x_{n-1}].$$

Then, f^0 is a symmetric polynomial in $n-1$ variables. By the induction hypothesis on the number of variables, there exists $g \in R[u_1, \ldots, u_{n-1}]$ such that

$$f^0(x_1, \ldots, x_{n-1}) = g(\sigma_1^0, \ldots, \sigma_{n-1}^0).$$

Now, define $f_1(x_1, \ldots, x_n) \in R[x_1, \ldots, x_n]$ as

$$f_1(x_1, \ldots, x_n) := f(x_1, \ldots, x_n) - g(\sigma_1, \ldots, \sigma_{n-1}).$$

Then, we have $f_1(x_1, \ldots, x_{n-1}, 0) = 0$. Thus, $x_n \mid f_1$. Since f_1 is symmetric, each x_i divides f_1, so that $\sigma_n \mid f_1$. Thus, we can write

$$f_1(x_1, \ldots, x_n) = \sigma_n \cdot h(x_1, \ldots, x_n)$$

for some $h \in R[x_1, \ldots, x_n]$. Since σ_n is not a zero-divisor in $R[x_1, \ldots, x_n]$, we get that h is also a symmetric polynomial and $\deg h < \deg f$. Thus, h is a polynomial in $\sigma_1, \ldots, \sigma_n$ and hence, so is f.

Uniqueness. For uniqueness, it suffices to show that the elementary symmetric polynomials are algebraically independent. That is, the map $\varphi\colon R[z_1, \ldots, z_n] \to R[x_1, \ldots, x_n]$ defined by

$$z_i \mapsto \sigma_i \text{ and } \varphi|_R = \mathrm{id}_R$$

is an injection. We use induction on n to prove this. The case $n = 1$ is clear since $\sigma_1 = x_1$. Assume now that $n \geq 2$, and that the result holds for $n-1$. If φ is not an injection, then pick a nonzero polynomial $f(z_1, \ldots, z_n) \in \ker \varphi$ of least degree. We may write f as a polynomial in z_n as follows.

$$f(z_1, \ldots, z_n) = f_0(z_1, \ldots, z_{n-1}) + \cdots + f_d(z_1, \ldots, z_{n-1})z^n$$

with $f_d \neq 0$. Since d is minimal, and σ_n is not a zero-divisor, we also get $f_0 \neq 0$. Since $f \in \ker \varphi$, we have

$$f_0(\sigma_1, \ldots, \sigma_{n-1}) + \cdots + f_d(\sigma_1, \ldots, \sigma_{n-1})\sigma_n^d = 0.$$

The above equality is in $R[x_1, \ldots, x_n]$. Putting $x_n = 0$, we get

$$f_0(\sigma_1^0, \ldots, \sigma_{n-1}^0) = 0$$

which shows that the corresponding φ for $n-1$ variables is not an injection, a contradiction. □

Definition 14.9. Let $S = R[x_1, \ldots, x_n]$. For $k \geq 1$, we define

$$w_k = x_1^k + \cdots + x_n^k.$$

Theorem 14.10 (Newton's Identities)**.** With w_k as defined above, we have

$$w_k = \begin{cases} \sigma_1 w_{k-1} - \sigma_2 w_{k-2} + \cdots + (-1)^k \sigma_{k-1} w_1 + (-1)^{k+1} \sigma_k k^{21} & k \leq n, \\ \sigma_1 w_{k-1} - \sigma_2 w_{k-2} + \cdots + (-1)^{n+1} \sigma_n w_{k-n} & k > n. \end{cases}$$

Proof. Let z be an indeterminate over $S := R[x_1, \ldots, x_n]$. Observe that

$$(1 - x_1 z) \cdots (1 - x_n z) = 1 - \sigma_1 z + \cdots + (-1)^n \sigma_n z^n =: \sigma(z).$$

Now, define $w(z) \in S[\![z]\!]$ as

$$\begin{aligned} w(z) &= \sum_{k=1}^{\infty} w_k z^k \\ &= \sum_{k=1}^{\infty} \left(\sum_{i=1}^{n} x_i^k \right) z^k \\ &= \sum_{i=1}^{n} \left(\sum_{k=1}^{\infty} (x_i z)^k \right) \\ &= \sum_{i=1}^{n} \frac{x_i z}{1 - x_i z}. \end{aligned}$$

Now, since $\sigma(z) = (1 - x_1 z) \cdots (1 - x_n z)$, we have

$$\sigma'(z) = -\sum_{i=1}^{n} \frac{x_i \sigma(z)}{1 - x_i z}$$

[21] Note that the last term is $\sigma_k k$ and not $\sigma_k n$, as one might have expected.

where we have taken the formal derivative in $S[\![z]\!]$. (We shall define this derivative more formally later on in Definition 16.1, but for now, one may think of it as the 'usual' derivative). On rearranging, we get

$$-\frac{z\sigma'(z)}{\sigma(z)} = \sum_{i=1}^{n} \frac{x_i z}{1 - x_i z} = w(z)$$

which further gives us

$$w(z)\sigma(z) = -z\sigma'(z).$$

Moreover, one may compute $\sigma'(z)$ independently from the first equation. Using that expression, we get

$$w(z)\sigma(z) = \sigma_1 z - 2\sigma_2 z^2 + \cdots + (-1)^n n\sigma_n z^n.$$

Comparing the coefficients of z^k on both sides gives us the desired result. □

Definition 14.11. Let $\mathbb{F}$ be a field and let $f(x) \in \mathbb{F}[x]$ be a non-constant monic polynomial. Let $\mathbb{K}$ be a splitting field of $f(x)$ over $\mathbb{F}$, so that

$$f(x) = (x - r_1) \cdots (x - r_n)$$

for $r_1, \ldots, r_n \in \mathbb{K}$. Then, the discriminant of $f(x)$ is defined as

$$\operatorname{disc}_{\mathbb{K}}(f(x)) := \prod_{1 \le i < j \le n} (r_i - r_j)^2.$$

Remark 14.12. Note that $\operatorname{disc}_{\mathbb{K}}(f(x)) = 0 \iff f(x)$ has repeated roots in $\mathbb{K}$. Moreover, by construction, $\operatorname{disc}_{\mathbb{K}}(f(x))$ has a square root in $\mathbb{K}$, given by

$$\prod_{1 \le i < j \le n} (r_i - r_j) \in \mathbb{K}.$$

Proposition 14.13. Let $\mathbb{F}$ be a field and let $f(x) \in \mathbb{F}[x]$ be non-constant and monic. Suppose $\mathbb{K}$ and $\mathbb{K}'$ are two splitting fields of $f(x)$ over $\mathbb{F}$. Then,

$$\operatorname{disc}_{\mathbb{K}}(f(x)) = \operatorname{disc}_{\mathbb{K}'}(f(x)) \in \mathbb{F}.$$

In other words, the discriminant takes value in $\mathbb{F}$ and is independent of the splitting field chosen.

Proof. Let $r_1, \ldots, r_n \in \mathbb{K}$ be such that $f(x) = (x - r_1) \cdots (x - r_n)$. Consider the Vandermonde

matrix,

$$M = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_n \\ r_1^2 & r_2^2 & \cdots & r_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_1^{n-1} & r_2^{n-1} & \cdots & r_n^{n-1} \end{bmatrix}.$$

Then, $\operatorname{disc}_{\mathbb{K}}(f(x)) = \big(\det(M)\big)^2 = \det(MM^\top)$. Let $\sigma_1, \ldots, \sigma_n \in \mathbb{F}[x_1, \ldots, x_n]$ be the elementary symmetric polynomials and define

$$s_i := \sigma_i(r_1, \ldots, r_n).$$

Now, note that

$$f(x) = x^n - s_1 x^{n-1} + \cdots + (-1)^n s_n$$

and since $f(x) \in \mathbb{F}[x]$, $s_i \in \mathbb{F}$ for all i. Also, define

$$\nu_k := r_1^k + \cdots + r_n^k$$

for all $k \geq 1$. By Newton's Identities, each ν_k can be written as a combination of s_i's, so that each $\nu_k \in \mathbb{F}$. Moreover, we have

$$MM^\top = \begin{bmatrix} n & \nu_1 & \cdots & \nu_{n-1} \\ \nu_1 & \nu_2 & \cdots & \nu_n \\ \nu_2 & \nu_3 & \cdots & \nu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{n-1} & \nu_n & \cdots & \nu_{2n-2} \end{bmatrix}.$$

Thus, $\operatorname{disc}_{\mathbb{K}}(f(x)) = \det(MM^\top) \in \mathbb{F}$.[22] Note that since ν_k can be calculated directly in terms of s_i, the coefficients of $f(x)$ itself, the discriminant does not depend on the choice of the splitting field. □

In view of the above proof, we have the following alternate definition of the discriminant, that is independent of the splitting field.

Definition 14.14. Let $\mathbb{F}$ be a field and let $f(x) = x^n - \sigma_1 x^{n-1} + \cdots + (-1)^n \sigma_n \in \mathbb{F}[x]$ be a monic polynomial. With w_k for $k = 1, \ldots, 2n-2$ as defined in Newton's Identities, we have

$$\operatorname{disc}(f(x)) := \det \begin{bmatrix} n & w_1 & \cdots & w_{n-1} \\ w_1 & w_2 & \cdots & w_n \\ w_2 & w_3 & \cdots & w_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n-1} & w_n & \cdots & w_{2n-2} \end{bmatrix}.$$

[22] Note that the n in the matrix is defined as $1 + \cdots + 1$ (n times), where 1 is the identity in $\mathbb{F}$. We may also regard n to represent the image of $n \in \mathbb{Z}$ under the homomorphism that sends $1_{\mathbb{Z}}$ to $1_{\mathbb{F}}$.

Remark 14.15. In the above definition, σ_i are **not** the elementary symmetric polynomials. They are arbitrary elements of $\mathbb{F}$. We are *defining* w_k recursively in terms of σ_i, using Newton's Identities, which is what motivates the use of the same notation.

Proposition 14.16 (Discriminant in terms of derivative)**.** Suppose $f(x) = \prod_{i=1}^{n}(x - r_i)$, then $f'(x) = (-1)^{\binom{n}{2}} \prod_{i=1}^{n} f'(r_i)$.

Proof. Observe that

$$f'(x) = \sum_{i=1}^{n} \frac{f(x)}{x - r_i} = \sum_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} (x - r_j).$$

Thus, we have

$$f'(r_i) = \prod_{\substack{j=1 \\ j \neq i}}^{n} (r_i - r_j)$$

from which the result follows. □

Example 14.17 (Discriminant of a Quadratic)**.**

Let $x^2 + bx + c \in \mathbb{F}[x]$ be a quadratic. We have $\sigma_1 = -b, \sigma_2 = c$. Thus, Newton's Identities give us

$$\begin{aligned} w_1 &= -b, \\ w_2 &= b^2 - 2c. \end{aligned}$$

Thus, we have

$$\operatorname{disc}(f(x)) = \det \begin{bmatrix} 2 & -b \\ -b & b^2 - 2c \end{bmatrix} = b^2 - 4c,$$

which is the usual determinant of a quadratic.

We now (finally) prove the Fundamental Theorem of Algebra.

Lemma 14.18.

1. Every real polynomial of odd degree has a real root.
2. Every complex number has a square root. Thus, every complex quadratic polynomial has a root in $\mathbb{C}$.

Proof. The first part follows trivially from the intermediate value property. For the second, for any $a + b\iota \in \mathbb{C}$, with $a, b \in \mathbb{R}$, we define $c, d \in \mathbb{R}$ as follows.

$$c := \sqrt{\frac{1}{2}[a + \sqrt{a^2 + b^2}]} \text{ and } d := \sqrt{\frac{1}{2}[-a + \sqrt{a^2 + b^2}]}.$$

Then, we have $(c + d\iota)^2 = a + b\iota$. □

Theorem 14.19 (Fundamental Theorem of Algebra)**.** Every non-constant complex polynomial has a root in $\mathbb{C}$.

Proof. Let $g(x) \in \mathbb{C}[x]$ be a non-constant complex polynomial. Then, $f(x) := g(x)\overline{g}(x)$ is a non-constant real polynomial, where $\overline{g}(x)$ denotes the polynomial whose coefficients are complex conjugates of the coefficients of $g(x)$. Note that if $f(z) = 0$ for some $z \in \mathbb{C}$, then $g(z) = 0$ or $\overline{g}(z) = 0$. If $\overline{g}(z) = 0$, then $g(\overline{z}) = 0$, so that $g(x)$ has a complex root in either case. Thus, it suffices to show that every non-constant real polynomial has a complex root.

Given any $f(x) \in \mathbb{R}[x]$, we can write $\deg f(x) = 2^n q$ for unique $n \geq 0$, and odd $q \in \mathbb{N}$. We prove the statement by induction on n. In the case that $n = 0$, $f(x)$ has odd degree, and hence has a real (and consequently, a complex) root. Now, assume that $n \geq 1$ and that the statement is true for $n-1$. Let $d := \deg f(x)$ and let $\mathbb{K} := \mathbb{C}(\alpha_1, \ldots, \alpha_d)$ be a splitting field of $f(x)$ over $\mathbb{C}$, where α_i are the roots of $f(x)$. For $r \in \mathbb{R}$, define

$$y_{ij}(r) = \alpha_i + \alpha_j + r\alpha_i\alpha_j$$

for $1 \leq i \leq j \leq d$. There are $\binom{d+1}{2}$ such pairs (i, j), so that the polynomial

$$h_r(x) := \prod_{1 \leq i \leq j \leq d} (x - y_{ij}(r))$$

has degree

$$\deg h_r(x) = \binom{d+1}{2} = \frac{1}{2}d(d+1) = 2^{n-1}\underbrace{q(d+1)}_{\text{odd}}.$$

Note that the coefficients of $h_r(x)$ are elementary symmetric polynomials in y_{ij}'s, and hence, are symmetric polynomials in $\alpha_i, \ldots, \alpha_d$. Thus, the coefficients of $h_r(x)$ are polynomials in coefficients of $f(x)$, so that $h_r(x) \in \mathbb{R}[x]$. By the inductive hypothesis on n, $h_r(x)$ has a root $z_r \in \mathbb{C} \subseteq \mathbb{K}$. Thus, $z_r = y_{i(r),j(r)}(r)$ for some $1 \leq i(r) \leq j(r) \leq d$.

Let $P := \{(i, j) \mid 1 \leq i \leq j \leq d\}$ and define $\varphi \colon \mathbb{R} \to P$ by $r \mapsto (i(r), j(r))$. Since P is finite and $\mathbb{R}$ is not, φ is not injective. Thus, there exist $c, d \in \mathbb{R}$, with $c \neq d$, such that

$$(i(c), j(c)) = (i(d), j(d)) =:= (a, b) \in P.$$

Thus,

$$z_c = \alpha_a + \alpha_b + c\alpha_a\alpha_b \text{ and } z_d = \alpha_a + \alpha_b + d\alpha_a\alpha_b.$$

Although apriori we only know that $\alpha_a, \alpha_b \in \mathbb{K}$, we now have

$$\alpha_a\alpha_b = \frac{z_c - z_d}{d - c} \in \mathbb{C}$$

and consequently,

$$\alpha_a + \alpha_b = z_c - c\alpha_a\alpha_b \in \mathbb{C}.$$

Thus, $\alpha_a\alpha_b$ and $\alpha_a + \alpha_b \in \mathbb{C}$. However, these are the coefficients of the quadratic

$$x^2 - (\alpha_a + \alpha_b)x + \alpha_a\alpha_b \in \mathbb{C}[x],$$

of which α_a and α_b are both roots. By Lemma 14.18, one of α_a and α_b must be in $\mathbb{C}$. Since α_a and α_b are both roots of $f(x)$, we are done. □

§15. Algebraic Closure of a Field

Definition 15.1. A field $\mathbb{K}$ is called an algebraically closed field if every non-constant polynomial $f(x) \in \mathbb{K}[x]$ has a root in $\mathbb{K}$.

Definition 15.2. Let $\mathbb{K}/\mathbb{F}$ be a field extension. We say that $\mathbb{K}$ is an algebraic closure of $\mathbb{F}$ if $\mathbb{K}$ is algebraically closed and $\mathbb{K}/\mathbb{F}$ is an algebraic extension.

We have the following simple proposition.

Proposition 15.3.

1. $\mathbb{K}$ is algebraically closed iff every non-constant polynomial in $\mathbb{K}[x]$ factors as a product of linear factors.
2. $\mathbb{C}$ is algebraically closed.
3. If $\mathbb{K}$ is algebraically closed and $\mathbb{L}/\mathbb{K}$ is an algebraic extension, then $\mathbb{L} = \mathbb{K}$.

Proposition 15.4. Let $\mathbb{K}/\mathbb{F}$ be an extension where $\mathbb{K}$ is algebraically closed. Define,

$$\mathbb{A} := \{\alpha \in \mathbb{K} \mid \alpha \text{ is algebraic over } \mathbb{F}\}.$$

Then, $\mathbb{A}$ is an algebraic closure of $\mathbb{F}$.

Proof. By Corollary 13.32, we already know that $\mathbb{A}/\mathbb{F}$ is an algebraic extension. Hence, it remains to show that $\mathbb{A}$ is algebraically closed. Let $f(x) \in \mathbb{A}[x]$ be non-constant. Then, $f(x)$ has a root $\alpha \in \mathbb{K}$ since $\mathbb{K}$ is algebraically closed. Thus. α is algebraic over $\mathbb{A}$, and hence over $\mathbb{F}$, by Corollary 13.31. Thus, $\alpha \in \mathbb{A}$. □

Lemma 15.5. Let $\{\mathbb{F}_i\}_{i\geq 1}$ be a sequence of fields with

$$\mathbb{F}_1 \subseteq \mathbb{F}_2 \subseteq \cdots$$

and let $\mathbb{F} := \bigcup_{i\geq 1} \mathbb{F}_i$. Then, $\mathbb{F}$ is a field with the following operations: Given $a, b \in \mathbb{F}$, there exist smallest $i, j \in \mathbb{N}$ such that $a \in \mathbb{F}_i$ and $b \in \mathbb{F}_j$. Then, $a, b \in \mathbb{F}_{i+j}$ and we define $a + b$ and ab to be the corresponding elements from $\mathbb{F}_{i+j}$.

Moreover, each $\mathbb{F}_i$ is a subfield of $\mathbb{F}$.

Proof. We leave this as an exercise to the reader. Note that we have used "smallest" just to ensure that the operations are well-defined. Of course, since $\mathbb{F}_i \subseteq \mathbb{F}_j$ (by which we always mean that $\mathbb{F}_i$ is a subfield of $\mathbb{F}_j$) for any $i \leq j$, we may pick any i and j. □

Theorem 15.6 (Existence of Algebraically Closed Extension)**.** Let $\mathbb{F}$ be a field. Then, there exists an algebraically closed field containing $\mathbb{F}$.

Proof. (Artin). We first show that given any field $\mathbb{F}$, we can construct a field $\mathbb{F}_1 \supseteq \mathbb{F}$ containing roots of any non-constant polynomial in $\mathbb{F}[x]$. Let S be a set of indeterminates which are in one-to-one correspondence with the set of non-constant polynomials in $\mathbb{F}[x]$. Let $x_f \in S$ denote the indeterminate corresponding to f.

Consider the polynomial ring $\mathbb{F}[S]$. Let

$$I := \left\langle f(x_f) \mid f \in \mathbb{F}[x], \deg f \geq 1 \right\rangle$$

be the ideal generated by the polynomials $f(x_f) \in S$. We now show that $1 \notin I$, so that I is a proper ideal of $\mathbb{F}[S]$. Suppose that $1 \in I$. Then,

$$1 = g_1 f_1(x_{f_1}) + \cdots + g_n f_n(x_{f_n})$$

for some $g_1, \ldots, g_n \in \mathbb{F}[S]$. Note that the polynomials g_j involve only finitely many variables. Let $x_i := x_{f_i}$ for $i = 1, \ldots, n$ and let $x_{n+1}, \ldots, x_m$ be the remaining variables in $g_1, \ldots, g_n$. Then, we have

$$\sum_{i=1}^{n} g_i\left(x_1, \ldots, x_n, x_{n+1}, \ldots, x_m\right) f_i(x_i) = 1.$$

Now, let $\mathbb{E} \supseteq \mathbb{F}$ be an extension containing roots α_i of f_i ($\mathbb{E}$ exists thanks to Theorem 13.40). Then, putting $x_i = \alpha_i$ for $i = 1, \ldots, n$ and $x_{n+1} = \cdots = x_m = 0$, we arrive at a contradiction.

Hence, I is a proper ideal of $\mathbb{F}[S]$, and is thus contained in some maximal ideal $\mathfrak{m} \subseteq \mathbb{F}[S]$. Put $\mathbb{F}_1 := \mathbb{F}[S]/\mathfrak{m}$. Then, $\mathbb{F}_1$ is a field extension of $\mathbb{F}$. Moreover, $\overline{x_f} := x_f + \mathfrak{m} \in \mathbb{F}_1$ is a root of $f(x) \in \mathbb{F}[x]$. Thus, $\mathbb{F}_1$ is an extension of $\mathbb{F}$ in which every non-constant polynomial of $\mathbb{F}[x]$ has a root.

Repeating this procedure, we get a sequence of fields

$$\mathbb{F} =: \mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \mathbb{F}_2 \cdots$$

such that every non-constant polynomial in $\mathbb{F}_i$ has a root in $\mathbb{F}_{i+1}$.

Now, put $\mathbb{K} = \bigcup_{i \geq 0} \mathbb{F}_i$. $\mathbb{K}$ is a field by Lemma 15.5 that has each $\mathbb{F}_i$ as a subfield. Now, if $f(x) \in \mathbb{K}[x]$, then $f(x) \in \mathbb{F}_n[x]$ for some n. Thus, $f(x)$ has a root in $\mathbb{F}_{n+1}[x] \subseteq \mathbb{K}$, as desired. □

Corollary 15.7 (Existence of Algebraic Closure)**.** Every field $\mathbb{F}$ has an algebraic closure.

Proof. By Theorem 15.6, there exists an algebraically closed field $\mathbb{L} \supseteq \mathbb{F}$. Now, use Proposition 15.4. □

Proposition 15.8. Let $\sigma\colon \mathbb{F} \to \mathbb{L}$ be an embedding of fields and let $\mathbb{L}$ be algebraically closed. Let $\alpha \in \mathbb{K} \supseteq \mathbb{F}$ be algebraic over $\mathbb{F}$ and let $p(x) = \operatorname{irr}(\alpha, \mathbb{F})$. Suppose $p(x) = \sum a_i x^i$ and define $p^\sigma(x) := \sum \sigma(a_i)x^i$. Then, $\tau \mapsto \tau(\alpha)$ is a bijection between the following sets.

$$\left\{\tau\colon \mathbb{F}(\alpha) \to \mathbb{L} \mid \tau \text{ is an embedding and } \tau|_{\mathbb{F}} = \sigma\right\} \leftrightarrow \left\{\beta \in \mathbb{L} \mid p^\sigma(\beta) = 0\right\}.$$

Proof. First, we note that the map is indeed well-defined. Let τ be an embedding that extends σ. Then,

$$\tau(p(\alpha)) = p^\sigma(\tau(\alpha)) = 0.$$

Thus, $\tau(\alpha)$ is indeed a root of p^σ.

Now, let $\beta \in \mathbb{L}$ be such that $p^\sigma(\beta) = 0$. Define $\tau_\beta\colon \mathbb{F}(\alpha) \to \mathbb{L}$ by $\tau_\beta(f(\alpha)) = f^\sigma(\beta)$ for $f(x) \in \mathbb{F}[x]$. We show that τ_β is well-defined. Suppose $f(\alpha) = g(\alpha)$. Then, $(f-g)(\alpha) = 0$, so that $p(x) \mid f(x) - g(x)$ by Proposition 13.18. Hence, $p^\sigma(x) \mid f^\sigma(x) - g^\sigma(x)$, giving us $f^\sigma(\beta) = g^\sigma(\beta)$. Hence, τ_β is well-defined. Moreover, it is clearly a homomorphism that extends σ. It is easily seen that $\beta \mapsto \tau_\beta$ is a two-sided inverse of the map $\tau \mapsto \tau_\alpha$. □

Remark 15.9. The above proposition essentially says that the number of ways to extend from $\mathbb{F}$ to $\mathbb{F}(\alpha)$ is precisely the number of roots that $p^\sigma(x)$ has in $\mathbb{L}$. In particular, this set is non-empty since $\mathbb{L}$ is algebraically closed.

Theorem 15.10. Let $\sigma\colon \mathbb{F} \to \mathbb{L}$ be an embedding where $\mathbb{L}$ is algebraically closed. If $\mathbb{K}/\mathbb{F}$ is an algebraic extension, then there exists an embedding $\tau\colon \mathbb{K} \to \mathbb{L}$ that extends σ.

Moreover, if $\mathbb{K}$ is an algebraic closure of $\mathbb{F}$, and $\mathbb{L}$ is an algebraic closure of $\sigma(\mathbb{F})$, then τ is an isomorphism extending σ.

Proof. Consider the set

$$\Sigma := \left\{(\mathbb{E}, \tau) \mid \mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K} \text{ and } \tau\colon \mathbb{E} \to \mathbb{L} \text{ such that } \tau|_{\mathbb{F}} = \sigma\right\}.$$

Note that Σ is non-empty since $(\mathbb{F}, \sigma) \in \Sigma$. Define the relation $\leq$ on Σ by

$$(\mathbb{E}, \tau) \leq (\mathbb{E}', \tau') \iff \mathbb{E} \subseteq \mathbb{E}' \text{ and } \tau'|_{\mathbb{E}} = \tau.$$

Then, $(\Sigma, \leq)$ is a partially ordered set. Moreover, if $\Lambda = \left\{(\mathbb{E}_i, \tau_i)\right\}_{i\in I}$ is a chain in Σ, then $\mathbb{E} := \bigcup_{i\in I} \mathbb{E}_i$ is a subfield of $\mathbb{K}$, and $\tau\colon \mathbb{E} \to \mathbb{L}$ defined as $\tau(x) := \tau_i(x)$ for $x \in \mathbb{E}_i$ is well-defined. Furthermore, $(\mathbb{E}, \tau)$ is an upper bound on Λ. By Zorn's Lemma, there exists a maximal element $(\mathbb{E}, \tau) \in \Sigma$. We show that $\mathbb{E} = \mathbb{K}$. If not, pick $\alpha \in \mathbb{K} \setminus \mathbb{E}$. By Proposition 15.8, we can extend τ to an embedding $\tau'\colon \mathbb{E}(\alpha)\colon \mathbb{L}$, which contradicts the maximality of $(\mathbb{E}, \tau)$. Thus, $\tau\colon \mathbb{K} \to \mathbb{L}$ is the desired embedding that extends σ.

Suppose now that $\mathbb{K}$ is an algebraic closure of $\mathbb{F}$ and $\mathbb{L}$ of $\sigma(\mathbb{F})$. We have

$$\sigma(\mathbb{F}) \subseteq \tau(\mathbb{K}) \subseteq \mathbb{L}.$$

Thus, $\mathbb{L}/\tau(\mathbb{K})$ is also algebraic. Since $\tau(\mathbb{K})$ is algebraically closed, we get $\mathbb{L} = \tau(\mathbb{K})$, which proves that τ is surjective, and hence an isomorphism. (Since τ is an embedding, it is injective to begin with). □

Corollary 15.11 (Isomorphism of Algebraic Closures)**.** If $\mathbb{K}_1$ and $\mathbb{K}_2$ are two algebraic closures of $\mathbb{F}$, then they are $\mathbb{F}$-isomorphic.

Proof. Consider the inclusion map $i\colon \mathbb{F} \hookrightarrow \mathbb{K}_2$. Theorem 15.10 allows us to extend it to an $\mathbb{F}$-isomorphism $\tau\colon \mathbb{K}_1 \to \mathbb{K}_2$. □

Definition 15.12. Given a field $\mathbb{F}$, we use $\overline{\mathbb{F}}$ to denote an algebraic closure of $\mathbb{F}$.

Theorem 15.13 (Isomorphism of Splitting Fields)**.** Any two splitting fields $\mathbb{E}$ and $\mathbb{E}'$ of a non-constant polynomial $f(x) \in \mathbb{F}[x]$ over $\mathbb{F}$ are $\mathbb{F}$-isomorphic.

Proof. Let $\overline{\mathbb{E}}$ be an algebraic closure of $\mathbb{E}$. Then, it is also one of $\mathbb{F}$. Thus, there exists an embedding $\tau\colon \mathbb{E}' \to \overline{\mathbb{E}}$ that extends the inclusion $i\colon \mathbb{F} \hookrightarrow \overline{\mathbb{E}}$, by Theorem 15.10.

Let $f(x) = a(x-\alpha_1)\cdots(x-\alpha_n)$ be a factorisation of $f(x)$ in $\mathbb{E}'[x]$. Then,

$$f^\tau(x) = a(x-\tau(\alpha_1))\cdots(x-\tau(\alpha_n)) \in \overline{\mathbb{E}}[x].$$

We have $\mathbb{E}' = \mathbb{F}(\alpha_1,\ldots,\alpha_n)$ and so $\tau(\mathbb{E}') = \mathbb{F}(\tau(\alpha_1),\ldots,\tau(\alpha_n))$. Thus, $\tau(\mathbb{E}')$ is a splitting field of f^τ over $\mathbb{F}$. But $f^\tau = f$ since τ extends the inclusion map. Thus, $\tau(\mathbb{E}') = \mathbb{E}$ since any algebraic closure contains a unique splitting field. □

§16. Separable Extensions

§§16.1. Definitions

Definition 16.1. Let $\mathbb{F}$ be a field. Define the $\mathbb{F}$-linear map $D_{\mathbb{F}} : \mathbb{F}[x] \to \mathbb{F}[x]$ by

$$D_{\mathbb{F}}\left(\sum_{i=0}^{n} a_i x^i\right) := \sum_{i=1}^{n} i a_i x^{i-1}.$$

Given any $f(x) \in \mathbb{F}[x]$, we call $D_{\mathbb{F}}(f(x))$ the (formal) derivative of $f(x)$, and denote it by $f'(x)$.

Remark 16.2. Note that the above definition requires no notion of limits. In the case that $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, it coincides with the usual derivative if we identify a polynomial by the function it represents.

We leave the proofs of the following two easy propositions as exercises.

Proposition 16.3. Let $f(x), g(x) \in \mathbb{F}[x]$ and let $a \in \mathbb{F}$ be arbitrary. Then,

1. $(f \pm ag)'(x) = f'(x) \pm ag'(x)$,
2. $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$.

Proposition 16.4. Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension. Then, $D_{\mathbb{E}}|_{\mathbb{F}} = D_{\mathbb{F}}$. Thus, the notation $f'(x)$ is unambiguous.

Definition 16.5. Let $f(x) \in \mathbb{F}[x]$ be a non-constant monic polynomial and let $\mathbb{E}$ be a splitting field of $f(x)$ over $\mathbb{F}$. In $\mathbb{E}[x]$, factorise $f(x)$ uniquely as

$$f(x) = (x - r_1)^{e_1} \cdots (x - r_g)^{e_g},$$

where $r_1, \ldots, r_g \in \mathbb{E}$ are distinct and each $e_i \in \mathbb{N}^+$.

The numbers $e_1, \ldots, e_g$ are called the multiplicities of the roots $r_1, \ldots, r_g$. If $e_i = 1$ for some i, then r_i is called a simple root and a repeated root otherwise.

If each root is a simple root, then $f(x)$ is said to be a separable polynomial.

If $f(x)$ is not monic, we have the same definitions upon division by the leading coefficient.

Remark 16.6. Note that as stated above, the separability of a polynomial depends on the splitting field chosen. However, in view of Remark 14.12, we see that separability depends only on $\operatorname{disc}(f(x))$, which we have seen to be independent of the splitting field chosen. (Proposition 14.13). The following proposition shows something stronger.

Proposition 16.7. The number of roots and their multiplicities are independent of the splitting field chosen for $f(x)$ over $\mathbb{F}$.

Proof. Let $\mathbb{E}$ and $\mathbb{K}$ be splitting fields of $f(x)$ over $\mathbb{F}$. By Theorem 15.13, there exists an $\mathbb{F}$-isomorphism $\tau\colon \mathbb{E} \to \mathbb{K}$. In turn, we get an isomorphism $\varphi_\tau\colon \mathbb{E}[x] \to \mathbb{K}[x]$, defined by

$$\sum a_i x^i \mapsto \sum \tau(a_i) x^i.$$

Now, let $\prod_{i=1}^{g} (x - r_i)^{e_i}$ be the unique factorisation of $f(x)$ in $\mathbb{E}[x]$. Then, the above isomorphism shows that

$$f(x) = \prod_{i=1}^{g} (x - \tau(r_i))^{e_i}$$

is the unique factorisation of $f(x)$ in $\mathbb{K}[x]$, from which the result follows. □

Proposition 16.8. Let $f(x) \in \mathbb{F}[x]$ be monic and let $r \in \mathbb{E} \supseteq \mathbb{F}$ be a root of $f(x)$. Then, r is a repeated root iff $f'(r) = 0$.

Proof. ($\implies$) If r is a repeated root, then write $f(x) = (x-r)^2 g(x)$ for $g(x) \in \mathbb{E}[x]$. Then, taking the derivative gives us

$$f'(x) = 2(x-r)g(x) + (x-r)^2 g'(x).$$

Thus, $f'(r) = 0$.

($\impliedby$) Suppose $f(x) = (x-r)g(x)$. Then,

$$0 = f'(r) = (r-r)g'(r) + g(r) = g(r).$$

Thus, $(x-r) \mid g(x)$ and hence, $(x-r)^2 \mid f(x)$. □

Theorem 16.9 (Derivative Criterion for Separability)**.** Let $f(x) \in \mathbb{F}[x]$ be monic.

1. If $f'(x) = 0$, then every root of $f(x)$ is a repeated root.
2. If $f'(x) \neq 0$, then $f(x)$ has all roots simple iff $\gcd(f(x), f'(x)) = 1$.

Proof. Let $\mathbb{E}$ be a splitting field of $f(x)$ over $\mathbb{F}$.

1. Let $r \in \mathbb{E}$ be a root of $f(x)$. Then, $f'(r) = 0$ by hypothesis, and hence r is a repeated root by Proposition 16.8.
2. Assume $f'(x) \neq 0$.

   ($\implies$) Suppose $f(x)$ has all roots simple. We need to show that $f(x)$ and $f\prime(x)$ has no common root. Let r be a root of $f(x)$. Then, $f'(r) \neq 0$, by Proposition 16.8, and we are done.

   ($\impliedby$) Suppose $\gcd(f(x), f'(x)) = 1$ and $r \in \mathbb{E}$ is an arbitrary root of $f(x)$. Then, $f'(r) \neq 0$, and r is a simple root by Proposition 16.8. □

Proposition 16.10. Let $f(x) \in \mathbb{F}[x]$ be irreducible and non-constant.

1. $f(x)$ is separable iff $f'(x) \neq 0$.
2. If $\mathrm{ch}(\mathbb{F}) = 0$, then $f(x)$ is separable.

In other words, irreducible polynomials over fields of characteristic zero are separable.

Proof. Let $\mathbb{E}$ be a splitting field of $f(x)$ over $\mathbb{F}$.

1. ($\implies$) $f(x)$ has no repeated roots, thus $f'(x) \neq 0$ by Derivative Criterion for Separability.

   ($\impliedby$) Suppose $f'(x) \neq 0$ and $r \in \mathbb{E}$ is a repeated root of $f(x)$. Then, by Proposition 16.8, $f'(r) = 0$. Thus, $g(x) := \gcd(f(x), f'(x)) \neq 1$. Irreducibility of $f(x)$ forces $f(x) = g(x)$, which implies $f(x) \mid f'(x)$, a contradiction since $\deg f'(x) < \deg f(x)$.
2. In fields of characteristic zero, only the constant polynomials have derivative zero (since $i \cdot a_i \neq 0$ if $a_i \neq 0$). Since $f(x)$ is non-constant, $f'(x) \neq 0$, and thus the previous part applies. □

Definition 16.11. Let $\mathbb{F}$ be a field of prime characteristic p. Define

$$\mathbb{F}^p := \left\{ \alpha^p \in \mathbb{F} \mid \alpha \in \mathbb{F} \right\}$$

to be the set of all p^{th} powers of elements of $\mathbb{F}$.

Proposition 16.12. $\mathbb{F}^p$ is a subfield of $\mathbb{F}$.

Proof. Only closure under addition is not obvious. For this, recall that $(x+y)^p = x^p + y^p$. (Proposition 13.4). □

Proposition 16.13. Let $\mathbb{F}$ be a field with $\mathrm{ch}(\mathbb{F}) = p > 0$. Then, $f(x) := x^p - a \in \mathbb{F}[x]$ is either irreducible in $\mathbb{F}[x]$, or $a \in \mathbb{F}^p$. In other words, either $f(x)$ is irreducible or it has a root.

Proof. Suppose $f(x)$ is not irreducible. Write $f(x) = g(x)h(x)$ with $1 \leq \deg g(x) =: m < p$. Let $b \in \mathbb{E}$ be a root of $f(x)$ in a splitting field $\mathbb{E}$ of $f(x)$ over $\mathbb{F}$. Then, $b^p = a$. Thus, $f(x)$ factorises in $\mathbb{E}[x]$ as

$$f(x) = x^p - b^p = (x-b)^p.$$

Since $\mathbb{E}[x]$ is a unique factorisation domain (Corollary 12.11), we see that $g(x) = (x-b)^m$ (we may assume that $g(x)$ is monic). Note that the coefficient of x^{m-1} is mb. By assumption, $mb \in \mathbb{F}$. Since $1 \leq m < p$, we see that $b \in \mathbb{F}$. Thus, $a = b^p \in \mathbb{F}^p$. □

Proposition 16.14. Let $f(x) \in \mathbb{F}[x]$ be an irreducible polynomial and let $p := \mathrm{ch}(\mathbb{F}) > 0$. If $f(x)$ is not separable, then there exists $g(x) \in \mathbb{F}[x]$ such that $f(x) = g(x^p)$.

Proof. Since $f(x)$ is irreducible and not separable, we must have $f'(x) = 0$. Write

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

and note that

$$f'(x) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1} = 0.$$

Thus, $k a_k = 0$ for $k = 1, \ldots, n$. When $p \nmid k$, we clearly have $a_k = 0$, since we may cancel the k. Thus, $f(x)$ is of the form

$$f(x) = a_0 + a_p x^p + \cdots + a_{mp} x^m p.$$

for some $m \in \mathbb{N}^+$. Thus, $g(x) = a_0 + a_p x + \cdots + a_{mp} x^m$ works. □

Definition 16.15. Let $\mathbb{K}/\mathbb{F}$ be a field extension. An algebraic element $\alpha \in \mathbb{K}$ over $\mathbb{F}$ is called a separable element over $\mathbb{F}$ if $\operatorname{irr}(\alpha, \mathbb{F})$ is separable over $\mathbb{F}$.

We say that $\mathbb{K}/\mathbb{F}$ is a separable extension if every $\alpha \in \mathbb{K}$ is separable.

We say that $\mathbb{F}$ is a perfect field if every algebraic extension of $\mathbb{F}$ is separable. Equivalently, every irreducible polynomial in $\mathbb{F}[x]$ is separable.

Corollary 16.16. Every field of characteristic zero is perfect.

Proof. Proposition 16.10. □

Proposition 16.17. Let $\mathbb{F}$ be a field with characteristic $p > 0$. Then, $\mathbb{F}$ is perfect iff $\mathbb{F} = \mathbb{F}^p$.

Proof. ($\Longrightarrow$) Suppose $\mathbb{F}$ is perfect and suppose $\mathbb{F} \neq \mathbb{F}^p$. Pick $\alpha \in \mathbb{F} \setminus \mathbb{F}^p$. Consider the polynomial $f(x) = x^p - \alpha \in \mathbb{F}[x]$. By Proposition 16.13, $f(x)$ is irreducible. However, $f'(x) = p x^{p-1} = 0$ (since $\mathbb{F}$ has characteristic p). Thus, by Proposition 16.10, $f(x)$ is not separable, which is a contradiction since $\mathbb{F}$ was assumed to be perfect.

($\Longleftarrow$) Suppose $\mathbb{F} = \mathbb{F}^p$ and $f(x) \in \mathbb{F}[x]$ is irreducible and not separable. By Proposition 16.14, we may write

$$f(x) = \sum_{i=0}^{m} a_i x^{ip}.$$

Let $b_i \in \mathbb{F}$ be such that $a_i = b_i^p$. (Such a b_i exists since $\mathbb{F} = \mathbb{F}^p$). Now, we have

$$f(x) = \sum_{i=0}^{m} a_i x^{ip} = \sum_{i=0}^{m} b_i^p x^{ip} = \left(\underbrace{\sum_{i=0}^{m} b_i x^i}_{\in \mathbb{F}[x]} \right)^p, \qquad \text{(By Corollary 13.5)}$$

which contradicts the irreducibility of $f(x)$. Thus, $\mathbb{F}$ is perfect. □

Corollary 16.18. Every finite field is perfect.

Proof. Let $\mathbb{F}$ be a finite field of characteristic $p > 0$ (a finite cannot have characteristic zero since $\mathbb{Z}$ is infinite). We show that $\mathbb{F} = \mathbb{F}^p$.

Recall that the prime subfield of $\mathbb{F}$ is the field $\mathbb{F}_p$ with p elements. Since $\mathbb{F}$ is a vector space over $\mathbb{F}_p$, we have $|\mathbb{F}| = p^n$ for some $n \in \mathbb{N}^+$, where $n = \left[\mathbb{F}\colon \mathbb{F}_p\right]$ (Note that $\left[\mathbb{F}\colon \mathbb{F}_p\right] < \infty$ since $\mathbb{F}$ is finite). By Lagrange's Theorem, $\alpha^{p^n-1} = 1$ for all $\alpha \in \mathbb{F}^\times$ (consider the multiplicative group $\mathbb{F}^\times$ of order $p^n - 1$). Thus, $\alpha^{p^n} = \alpha$ for all $\alpha \in \mathbb{F}$ (including 0).

Thus, given any arbitrary $\alpha \in \mathbb{F}$, choosing $\beta = \alpha^{p^{n-1}}$ gives us $\alpha = \beta^p \in \mathbb{F}^p$. The result follows from Proposition 16.17. □

§§16.2. Extensions of Embeddings

Proposition 16.19. Let $f(x) \in \mathbb{F}[x]$ be an irreducible monic polynomial. Then, all roots of $f(x)$ have equal multiplicity (in any splitting field). If $\mathrm{ch}(\mathbb{F}) = 0$, then all roots are simple. If $\mathrm{ch}(\mathbb{F}) =: p > 0$, then all roots have multiplicity p^n for some $n \in \mathbb{N}$.

Proof. Let $\overline{\mathbb{F}} \supseteq \mathbb{F}$ be an algebraic closure of $\mathbb{F}$. Let $\alpha, \beta \in \mathbb{F}$ be roots of f. We have an $\mathbb{F}$-isomorphism $\sigma\colon \mathbb{F}(\alpha) \to \mathbb{F}(\beta)$ determined by $\alpha \mapsto \beta$. Thus, σ can be extended to an automorphism τ of $\mathbb{F}$. Suppose $f(x) = (x-\alpha)^m h(x)$ where m is the multiplicity of α and $h(x) \in \overline{\mathbb{F}}[x]$. Since τ fixes $\mathbb{F}$, it also fixes $f(x) \in \mathbb{F}[x]$. Thus, applying τ, we get

$$f(x) = f^\tau(x) = (x-\beta)^m h^\tau(x).$$

Thus, the multiplicity of β is at least m. By symmetry, we have equality.

If $\mathrm{ch}(\mathbb{F}) = 0$, then $f(x)$ is separable by Proposition 16.10. Thus, all roots are simple.

Now, suppose $\mathrm{ch}(\mathbb{F}) =: p > 0$. Let $n \in \mathbb{N}$ be the largest such that there exists a polynomial $g(x) \in \mathbb{F}[x]$ such that $f(x) = g(x^{p^n})$. Note that if no such positive n exists, we can take $g = f$ and $n = 0$. Then, g is irreducible since f is so. Moreover, g must be separable. If not, by Proposition 16.14, we must have $g(x) = h(x^p)$ for some $h(x) \in \mathbb{F}[x]$. But then, $f(x) = h(x^{p^{n+1}})$, contradicting the maximality of n. Thus, $g(x)$ factors as $(x-r_1)\cdots(x-r_g)$ in $\overline{\mathbb{F}}$ where each factor is distinct. Since $\overline{\mathbb{F}}$ is algebraically closed, we can find $s_1, \ldots, s_g$, necessarily distinct, such that $s_i^{p^n} = r_i$ for all i. We then have

$$f(x) = g(x^{p^n}) = (x-s_1)^{p^n} \cdots (x-s_g)^{p^n},$$

as desired. □

Theorem 16.20. Let $\sigma\colon \mathbb{F} \to \mathbb{L}$ be an embedding of fields where $\mathbb{L}$ is an algebraic closure of $\sigma(\mathbb{F})$. Similarly, let $\tau\colon \mathbb{F} \to \mathbb{L}'$ be an embedding of fields where $\mathbb{L}'$ is an algebraic closure of $\tau(\mathbb{F})$. Let $\mathbb{E}$ be an algebraic extension of $\mathbb{F}$.

Let S_σ (resp. S_τ) denote the set of extensions of σ (resp. τ) to embeddings of $\mathbb{E}$ into $\mathbb{L}$ (resp. $\mathbb{L}'$). Let $\lambda\colon \mathbb{L} \to \mathbb{L}'$ be an isomorphism extending $\tau \circ \sigma^{-1}\colon \sigma(\mathbb{F}) \to \tau(\mathbb{F})$.

The map $\psi\colon S_\sigma \to S_\tau$ given by $\psi(\widetilde{\sigma}) = \lambda \circ \widetilde{\sigma}$ is a bijection.

$$\begin{array}{ccccc}
\mathbb{L}' & \xleftarrow{\quad\lambda\quad} & & & \mathbb{L} \\
\Big| & & & & \Big| \\
\widetilde{\tau}(\mathbb{E}) & \xleftarrow{\widetilde{\tau}\in S_\tau} & \mathbb{E} & \xrightarrow{\widetilde{\sigma}\in S_\sigma} & \widetilde{\sigma}(\mathbb{E}) \\
\Big| & & & & \Big| \\
\tau(\mathbb{F}) & \xleftarrow{\quad\tau\quad} & \mathbb{F} & \xrightarrow{\quad\sigma\quad} & \sigma(\mathbb{F})
\end{array}$$

Proof. If $\widetilde{\sigma} \in S_\sigma$, then for any $x \in \mathbb{F}$, we have

$$(\lambda \circ \widetilde{\sigma})(x) = \lambda(\sigma(x)) = (\tau \circ \sigma^{-1})(\sigma(x)) = \tau(x).$$

Thus, ψ actually maps into S_τ. Since λ is an isomorphism, ψ is easily seen to be a bijection. Explicitly, the inverse of ψ is the map $\widetilde{\tau} \mapsto \lambda^{-1} \circ \tau$. □

Remark 16.21. The above proposition says that the "number" (cardinality) of extensions does not depend on $\mathbb{L}$ or on the embedding σ. Since $\mathbb{E}$ is an arbitrary algebraic extension of $\mathbb{F}$, the set S_σ need not be finite.

Thus, we may assume $\mathbb{L} \supseteq \mathbb{F}$ to be an algebraic closure of $\mathbb{F}$ and σ to be the inclusion map.

Definition 16.22. If $\mathbb{E}/\mathbb{F}$ is an algebraic extension, then the cardinality of S_σ (as in Theorem 16.20) is called the separable degree of $\mathbb{E}/\mathbb{F}$ and is denoted as $[\mathbb{E}\colon \mathbb{F}]_s$.

Proposition 16.23. Let $\alpha \in \mathbb{E} \supseteq \mathbb{F}$ be algebraic over $\mathbb{F}$ and $n := \deg\big(\operatorname{irr}(\alpha, \mathbb{F})\big)$. Then, $[\mathbb{F}(\alpha)\colon \mathbb{F}]_s \leq n = [\mathbb{F}(\alpha)\colon \mathbb{F}]$ with equality iff α is separable over $\mathbb{F}$.

Proof. By Proposition 15.8, $[\mathbb{F}(\alpha)\colon \mathbb{F}]_s$ is exactly the number of roots of $p(x) := \operatorname{irr}(\alpha, \mathbb{F})$ in $\overline{\mathbb{F}}$. This is at most $n = \deg p(x)$. Moreover, equality occurs implies that all roots are distinct, and thus α is separable over $\mathbb{F}$. □

Theorem 16.24 (Tower Law for separable degree)**.** Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be a tower of finite algebraic extensions. Then, $[\mathbb{E}\colon \mathbb{F}]_s \leq [\mathbb{E}\colon \mathbb{F}]$, and

$$[\mathbb{K}\colon \mathbb{F}]_s = [\mathbb{K}\colon \mathbb{E}]_s \cdot [\mathbb{E}\colon \mathbb{F}]_s.$$

Proof. We first show that the separable degree is multiplicative. Let $n := [\mathbb{K}\colon \mathbb{E}]_s$ and $m := [\mathbb{E}\colon \mathbb{F}]_s$, and let $\sigma\colon \mathbb{F} \to \mathbb{L}$ be an embedding into an algebraically closed field $\mathbb{L}$.

Let $\sigma_1, \dots, \sigma_m\colon \mathbb{E} \to \mathbb{L}$ be extensions of σ. Then, each σ_i has extensions $\sigma_i^{(1)}, \dots, \sigma_i^{(n)}\colon \mathbb{K} \to \mathbb{L}$. Note that the set $\left\{\sigma_i^{(j)}\colon 1 \le i \le m, 1 \le j \le n\right\}$ has cardinality mn since all extensions obtained are distinct. Clearly, any embedding $\tau\colon \mathbb{K} \to \mathbb{L}$ extending σ is obtained this way. ($\tau|_{\mathbb{E}} = \sigma_i$ for some i, and thus, $\tau = \sigma_i^{(j)}$ for some j). Thus, $[\mathbb{K}\colon \mathbb{F}]_s = mn$ as desired.

Now, since $\mathbb{E}/\mathbb{F}$ is finite, we can construct $\alpha_1, \dots, \alpha_g$ such that $\mathbb{E} = \mathbb{F}(\alpha_1, \dots, \alpha_g)$. We have the chain

$$\mathbb{F} \subseteq \mathbb{F}(\alpha_1) \subseteq \mathbb{F}(\alpha_1, \alpha_2) \subseteq \cdots \subseteq \mathbb{F}(\alpha_1, \dots, \alpha_g).$$

By Proposition 16.23, we have

$$\left[\mathbb{F}(\alpha_1, \dots, \alpha_{i+1})\colon \mathbb{F}(\alpha_1, \dots, \alpha_i)\right]_s \le \left[\mathbb{F}(\alpha_1, \dots, \alpha_{i+1})\colon \mathbb{F}(\alpha_1, \dots, \alpha_i)\right]$$

for all $i = 0, \dots, g-1$. Since both degrees are multiplicative, we are done. □

Corollary 16.25. Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be a tower of finite algebraic extensions. Then, $[\mathbb{K}\colon \mathbb{F}] = [\mathbb{K}\colon \mathbb{F}]_s$ iff the equality holds at each stage.

Theorem 16.26. Let $\mathbb{E}/\mathbb{F}$ be a finite extension. Then, $\mathbb{E}/\mathbb{F}$ is separable iff $[\mathbb{E}\colon \mathbb{F}] = [\mathbb{E}\colon \mathbb{F}]_s$.

Proof. Write $\mathbb{E} = \mathbb{F}(\alpha_1, \dots, \alpha_n)$ for $\alpha_i \in \mathbb{E}$. (Since $\mathbb{E}/\mathbb{F}$ is finite, it is also algebraic by Proposition 13.16). Now, put

$$\mathbb{F}_0 := \mathbb{F} \text{ and } \mathbb{F}_i := \mathbb{F}(\alpha_1, \dots, \alpha_i),$$

for $i = 1, \dots, n$.

($\implies$) Assume $\mathbb{E}/\mathbb{F}$ is separable. Then, since each α_i is separable over $\mathbb{F}$, it follows that α_i is separable over $\mathbb{F}_i$ for $i = 1, \dots, n$. (Note that $\operatorname{irr}(\alpha_i, \mathbb{F}_i) \mid \operatorname{irr}(\alpha_i, \mathbb{F})$). Thus, we see that

$$[\mathbb{F}_i\colon \mathbb{F}_{i-1}]_s = [\mathbb{F}_i\colon \mathbb{F}_{i-1}]$$

for all $i = 1, \dots, n$. Multiplying gives us $[\mathbb{E}\colon \mathbb{F}]_s = [\mathbb{E}\colon \mathbb{F}]$.

($\impliedby$) Let $\alpha \in \mathbb{E}$ be arbitrary. Consider the tower

$$\mathbb{F} \subseteq \mathbb{F}(\alpha) \subseteq \mathbb{E}.$$

Since $[\mathbb{E}\colon \mathbb{F}]_s = [\mathbb{E}\colon \mathbb{F}]$, we must also have $[\mathbb{F}(\alpha)\colon \mathbb{F}]_s = [\mathbb{F}(\alpha)\colon \mathbb{F}]$, by Corollary 16.25. Thus, α is separable over $\mathbb{F}$ by Proposition 16.23. □

Corollary 16.27. Let $\alpha \in \mathbb{E} \supseteq \mathbb{F}$ be separable over $\mathbb{F}$. Then, $\mathbb{F}(\alpha)/\mathbb{F}$ is separable.

Proof. By Proposition 16.23, $[\mathbb{F}(\alpha)\colon \mathbb{F}]_s = [\mathbb{F}(\alpha)\colon \mathbb{F}]$. The result follows by Theorem 16.26. □

Proposition 16.28. Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be a tower of fields. Then, $\mathbb{K}/\mathbb{F}$ is separable iff $\mathbb{K}/\mathbb{E}$ and $\mathbb{E}/\mathbb{F}$ are.

Proof. For both parts, note that if $\alpha \in \mathbb{K}$ is algebraic over $\mathbb{F}$, then it is also algebraic over $\mathbb{E}$. Moreover, $\operatorname{irr}(\alpha, \mathbb{E}) \mid \operatorname{irr}(\alpha, \mathbb{F})$. (The divisibility is in $\mathbb{E}[x]$).

($\implies$) Let $\alpha \in \mathbb{K}$ be arbitrary. Then, α is algebraic over $\mathbb{F}$ and thus, over $\mathbb{E}$. Since $\operatorname{irr}(\alpha, \mathbb{F})$ has no repeated roots, neither does its factor $\operatorname{irr}(\alpha, \mathbb{E})$. Thus, $\mathbb{K}/\mathbb{E}$ is separable. Now, let $\beta \in \mathbb{E}$ be arbitrary. Then, $\beta \in \mathbb{K}$ and thus, $\operatorname{irr}(\alpha, \mathbb{F})$ is separable. Thus, $\mathbb{E}/\mathbb{F}$ is separable.

($\impliedby$) Let $\alpha \in \mathbb{K}$ be arbitrary. Note that α is algebraic over $\mathbb{E}$ since it is separable over $\mathbb{E}$. Let $\operatorname{irr}(\alpha, \mathbb{E}) = a_1 + \cdots + a_n x^{n-1} + x^n \in \mathbb{E}[x]$. Put

$$\mathbb{F}_0 = \mathbb{F} \text{ and } \mathbb{F}_i = \mathbb{F}(a_1, \ldots, a_i),$$

for $i = 1, \ldots, n$. By ($\implies$), we see that a_i is separable over $\mathbb{F}_{i-1}$ and hence,

$$[\mathbb{F}_i : \mathbb{F}_{i-1}]_s = [\mathbb{F}_i : \mathbb{F}_{i-1}]$$

for all $i = 1, \ldots, n$. Finally, put $\mathbb{F}_{n+1} = \mathbb{F}_n(\alpha)$. Then, the above equality also holds for $i = n+1$, since α is separable over $\mathbb{F}_n$. (Note that by construction, $\operatorname{irr}(\alpha, \mathbb{F}_n) = \operatorname{irr}(\alpha, \mathbb{E})$, and the latter is separable by assumption). Upon multiplying, we get $[\mathbb{F}_{n+1} : \mathbb{F}]_s = [\mathbb{F}_{n+1} : \mathbb{F}]$, and thus $\mathbb{F}_{n+1}/\mathbb{F}$ is separable. Since $\alpha \in \mathbb{F}_{n+1}$, α is separable over $\mathbb{F}$, and thus $\mathbb{K}/\mathbb{F}$ is separable. □

Corollary 16.29. Let $f(x) \in \mathbb{F}[x]$ be a separable polynomial and let $\mathbb{E} \supseteq \mathbb{F}$ be a splitting field of $f(x)$ over $\mathbb{F}$. Then, $\mathbb{E}/\mathbb{F}$ is separable.

Proof. Write $\mathbb{E} = \mathbb{F}(r_1, \ldots, r_n)$ where $f(x) = a(x - r_1) \cdots (x - r_n)$, and apply the previous proposition and corollary repeatedly. □

Proposition 16.30. Let $\mathbb{E}/\mathbb{F}$ be a finite extension. Then, $[\mathbb{E} : \mathbb{F}]_s$ divides $[\mathbb{E} : \mathbb{F}]$. If $\operatorname{ch}(\mathbb{F}) =: p > 0$, then the quotient $\dfrac{[\mathbb{E} : \mathbb{F}]}{[\mathbb{E} : \mathbb{F}]_s}$ is a power of p.

Proof. If $\operatorname{ch}(\mathbb{F}) = 0$, then $\mathbb{F}$ is a perfect field by Corollary 16.16. Since $\mathbb{E}/\mathbb{F}$ is a finite extension, it is also algebraic by Proposition 13.16. Thus. $\mathbb{E}/\mathbb{F}$ is separable and the two degrees are equal. Suppose now that $\operatorname{ch}(\mathbb{F}) =: p > 0$.

First suppose that $\mathbb{E} = \mathbb{F}(\alpha)$ for some $\alpha \in \mathbb{E}$. Let $p(x) := \operatorname{irr}(\alpha, \mathbb{E})$ and let $d := \deg p(x)$. By Proposition 16.19, $p(x)$ factors in $\overline{\mathbb{F}}[x]$ as

$$p(x) = (x - \alpha)^{p^n} (x - \alpha_2)^{p^n} \cdots (x - \alpha_g)^{p^n}$$

for some $n \in \mathbb{N}$, where $\alpha_2, \ldots, \alpha_g \in \overline{\mathbb{F}} \setminus \{\alpha\}$ are distinct. Note that $gp^n = d$. By Proposition 15.8, $[\mathbb{F}(\alpha) : \mathbb{F}]_s = g$. Thus, the statement is true.

For a general finite extension $\mathbb{E}/\mathbb{F}$, write $\mathbb{E} = \mathbb{F}(\beta_1, \ldots, \beta_k)$ and use the fact that degrees are multiplicative. □

§17. Finite Fields

§§17.1. Existence and Uniqueness

We let p denote an arbitrary prime number.

Theorem 17.1 (Uniqueness of finite fields)**.** Let $\mathbb{K}$ and $\mathbb{L}$ be two finite fields with the same cardinality. Then, $\mathbb{K}$ and $\mathbb{L}$ are isomorphic.

Proof. Let $q := |\mathbb{K}|$ and $p := \mathrm{ch}(\mathbb{K})$. Then, $q = p^n$ for some $n \in \mathbb{N}^+$. Note that $\mathbb{K}^\times$ is a (multiplicative) group of order $q-1$. By Lagrange's Theorem, we have that $a^{q-1} = 1$ for all $a \in \mathbb{K}^\times$. We thus get that $a^q - a = 0$ for *all* $a \in \mathbb{K}$ (including 0). Hence, $\mathbb{K}$ is a splitting field of $x^q - x$ over $\mathbb{F}_p$, and so is $\mathbb{L}$. By the Isomorphism of Splitting Fields, $\mathbb{K}$ and $\mathbb{L}$ are isomorphic. □

Definition 17.2. We shall denote *the* finite field with p^n elements as $\mathbb{F}_{p^n}$.

Remark 17.3. Note that we have not yet shown that the finite field $\mathbb{F}_{p^n}$ exists for all prime p, and all $n \in \mathbb{N}^+$. We have only shown that if such a field exists, then it is unique up to isomorphism.

Theorem 17.4 (Existence of finite fields)**.** Fix a prime p and an algebraic closure $\overline{\mathbb{F}}_p$. For every $n \in \mathbb{N}^+$, there exists a unique subfield of $\overline{\mathbb{F}}_p$ of size p^n, denoted by $\mathbb{F}_{p^n}$. Furthermore,

$$\overline{\mathbb{F}}_p = \bigcup_{n \in \mathbb{N}^+} \mathbb{F}_{p^n}.$$

Proof. Fix $n \in \mathbb{N}^+$ and let $q := p^n$. $\overline{\mathbb{F}}_p$ contains a unique splitting field of $x^q - x =: f(x)$ over $\mathbb{F}_p$. We show that this splitting field has q elements. Consider

$$\mathbb{K} = \left\{ \alpha \in \overline{\mathbb{F}}_p \mid f(\alpha) = 0 \right\}.$$

Then, $|\mathbb{K}| = q$, since $f(x)$ is separable by the Derivative Criterion for Separability. Thus, $\mathbb{K}$ is the desired splitting field. Conversely, any field with q elements would be the set of roots of $x^q - x$, hence we have uniqueness.

We now show that $\overline{\mathbb{F}}_p = \bigcup_{k \geq 1} \mathbb{F}_{p^k}$. Let $\alpha \in \overline{\mathbb{F}}_p$ and let $d := \deg_{\mathbb{F}_p}(\alpha)$. Then, $[\mathbb{F}_p(\alpha) : \mathbb{F}_p] = d$ and hence, $\alpha \in \mathbb{F}(\alpha) = \mathbb{F}_{p^d}$. □

Proposition 17.5. The polynomial $f(x) := x^4 + 1$ is irreducible in $\mathbb{Z}[x]$ but it is reducible in $\mathbb{F}_p$ for every prime p.

Proof. For irreducibility in $\mathbb{Z}[x]$, note that

$$f(x+1) = x^4 + 4x^3 + 6x^2 + 4x + 2$$

is irreducible by applying the Eisenstein-Schönemann Criterion at the prime 2.

Now, let p be a prime. If $p = 2$, then $x^4 + 1 = (x+1)^4$. Let $p > 2$ be an odd prime. Then, $p^2 \equiv 1 \pmod 8$. We then have

$$x^4 + 1 \mid x^8 - 1 \mid x^{p^2-1} - 1 \mid x^{p^2} - x.$$

Now, suppose $x^4 + 1$ is irreducible and let $\alpha \in \overline{\mathbb{F}}_p$ be a root. Then, $[\mathbb{F}_p(\alpha) : \mathbb{F}_p] = \deg(x^4 + 1) = 4$. But, α is clearly contained in the splitting field of $x^{p^2} - x$ over $\mathbb{F}_p$, which is $\mathbb{F}_{p^2} \subseteq \overline{\mathbb{F}}_p$ and so, is contained in a degree 2 extension, a contradiction. □

§§17.2. Gauss' Necklace Formula

We recall (without proof) the Möbius inversion formula.

Definition 17.6. The Möbius function $\mu : \mathbb{N}^+ \to \mathbb{N}^+$ is defined as

$$\mu(n) := \begin{cases} 1 & n = 1, \\ (-1)^r & n \text{ is a product of } r \text{ distinct primes, and} \\ 0 & p^2 \mid n \text{ for some prime } p. \end{cases}$$

Theorem 17.7 (Möbius inversion formula)**.** Let $f, g : \mathbb{N}^+ \to \mathbb{N}^+$ be functions satisfying

$$f(n) = \sum_{d \mid n} g(d).$$

Then, they also satisfy

$$g(n) = \sum_{d \mid n} f\left(\frac{n}{d}\right) \mu(d).$$

For the remainder, p denotes an odd prime and q denotes a positive integral power of p.

Lemma 17.8. If $m \mid n$, then $x^{q^m} - x \mid x^{q^n} - x$ in $\mathbb{F}_q[x]$.

Proof. Fix an algebraic closure $\overline{\mathbb{F}}_q$. Since $f(x) := x^{q^m} - x$ is separable, by the Derivative Criterion for Separability, it suffices to show that every root of $f(x)$ is also a root of $x^{q^n} - x =: g(x)$. Let α be a root of $f(x)$. We have

$$\alpha^{q^m} = \alpha.$$

Raising both sides to the power q^m, we obtain

$$\alpha^{q^{2m}} = \alpha^{q^m} = \alpha.$$

Continuing repeatedly, we see that

$$\alpha^{km} = \alpha$$

for all $k \in \mathbb{N}^+$, and for $k = n/m$ in particular. This gives us $g(\alpha) = 0$, as desired. □

Lemma 17.9. Let $f(x) \in \mathbb{F}_q[x]$ be a monic irreducible polynomial. Then, $f(x) \mid x^{q^n} - x$ iff $\deg f \mid n$.

Proof. ($\implies$) Suppose $f(x) \mid x^{q^n} - x$. Then, $\mathbb{F}_{q^n}$ contains all roots of $f(x)$. Let $\alpha \in \mathbb{F}_{q^n}$ be a root of $f(x)$. Considering the tower $\mathbb{F}_q \subseteq \mathbb{F}_q(\alpha) \subseteq \mathbb{F}_{q^n}$, shows that $\deg f(x) = [\mathbb{F}_q(\alpha) : \mathbb{F}_q]$ divides $[\mathbb{F}_{q^n} : \mathbb{F}_q] = n$.

($\impliedby$) Let $d := \deg f(x)$ and suppose $d \mid n$. Fix an algebraic closure $\overline{\mathbb{F}}_q$ of $\mathbb{F}_q$. Let $\alpha \in \overline{\mathbb{F}}_q$ be a root of $f(x)$. Then, $[\mathbb{F}(\alpha) : \mathbb{F}] = d$ and thus, by Theorem 17.4, we have that

$$\mathbb{F}(\alpha) = \mathbb{F}_{q^d} = \left\{ \beta^{q^d} - \beta = 0 \mid \beta \in \overline{\mathbb{F}}_q \right\}.$$

Thus, every root of $f(x)$ satisfies $x^{q^d} - x$. Since this divides $x^{q^n} - x$ by Lemma 17.8, we are done. □

Remark 17.10. Lemma 17.9 essentially shows that the monic factorisation of $x^{q^n} - x$ in $\mathbb{F}_q[x]$ consists of every (monic) irreducible polynomial of degree d where d runs over all the divisors of n. Moreover, no factor can be repeated since $x^{q^n} - x$ is separable.

Theorem 17.11 (Gauss' Necklace Formula)**.** The number of irreducible polynomials of degree n over $\mathbb{F}_q$ is given by

$$N_q(n) := \frac{1}{n} \sum_{d \mid n} \mu(d) q^{n/d}.$$

Proof. By Lemma 17.9, we note that

$$x^{q^n} - x = \prod_{d \mid n} f_1^{(d)}(x) \cdots f_{N_q(d)}^{(d)}(x),$$

where $f_1^{(d)}(x), \ldots, f_{N_q(d)}^{(d)}(x)$ are all the monic irreducible polynomials of degree d. Equating the degree on both sides gives us

$$q^n = \sum_{d \mid n} d N_q(d).$$

Defining $f(n) := q^n$ and $g(n) := n N_q(n)$ allows us to use the Möbius inversion formula to obtain the result. □

§§17.3. Primitive Element Theorem

Definition 17.12. Let $\mathbb{E}/\mathbb{F}$ be a field extension. An element $\alpha \in \mathbb{E}$ is called a primitive element for $\mathbb{E}$ over $\mathbb{F}$ if $\mathbb{E} = \mathbb{F}(\alpha)$.

We say that $\mathbb{E}$ is primitive over $\mathbb{F}$ if there exists a primitive element for $\mathbb{E}$ over $\mathbb{F}$.

Theorem 17.13 (Primitive Element Theorem)**.** Let $\mathbb{K}/\mathbb{F}$ be a finite extension.

1. There is a primitive element for $\mathbb{K}/\mathbb{F}$ iff the number of intermediate subfields $\mathbb{E}$ such that $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ is finite.
2. If $\mathbb{K}/\mathbb{F}$ is separable, then $\mathbb{K}/\mathbb{F}$ is primitive.

Proof. Omitted due to time constraints. □

§18. Normal Extensions

Definition 18.1. An algebraic extension $\mathbb{E}/\mathbb{F}$ is called a normal extension if whenever $f(x) \in \mathbb{F}[x]$ is irreducible and has a root in $\mathbb{E}$, then $f(x)$ splits into linear factors in $\mathbb{E}[x]$.

Definition 18.2. Let $\mathbb{E}/\mathbb{F}$ be an extension and let $\mathcal{F} = \{f_i(x)\}_{i\in\mathcal{I}}$ be a (possibly infinite) family of non-constant polynomials in $\mathbb{F}[x]$. Then, $\mathbb{E}$ is said to be a splitting field for the family $\mathcal{F}$ over $\mathbb{F}$ if each $f_i(x) \in \mathcal{F}$ splits as a product of linear factors in $\mathbb{E}[x]$ and is generated by the roots of the polynomials.

Remark 18.3. A splitting field for any family always exists since an algebraic closure exists. So, we consider $A \subseteq \overline{\mathbb{F}}$ to be the set of roots of all polynomials in the family, and put $\mathbb{E} := \mathbb{F}(A) \subseteq \overline{\mathbb{F}}$.

Proposition 18.4. Let $\mathbb{F}$ be a field and let $\mathcal{F} \subseteq \mathbb{F}[x]$ be a family of separable polynomials. Then, $\mathbb{E}/\mathbb{F}$ is separable where $\mathbb{E} \subseteq \overline{\mathbb{F}}$ is the splitting field of $\mathcal{F}$ over $\mathbb{F}$.

Proof. Let $a \in \mathbb{E} = \mathbb{F}(A)$ where A is as in Remark 18.3. By Corollary 9.55, there is a finite set $\{a_1, \ldots, a_n\} \subseteq A$ such that $a \in \mathbb{F}(a_1, \ldots, a_n)$. Since each a_i is a root of a separable polynomial, it is separable. Applying Corollary 16.27 repeatedly, we see that $\mathbb{F}(a_1, \ldots, a_n)/\mathbb{F}$ is a separable extension, hence a is separable over $\mathbb{F}$. □

Lemma 18.5. Let $\mathbb{E}/\mathbb{F}$ be an algebraic extension. Let $\sigma\colon \mathbb{E} \to \mathbb{E}$ be an $\mathbb{F}$-embedding. Then, σ is an automorphism of $\mathbb{E}$.

Proof. We only need to prove that σ is onto. Let $\alpha \in \mathbb{E}$ be arbitrary. Put $p(x) := \operatorname{irr}(\alpha, \mathbb{F})$. Let $\mathbb{K} \subseteq \mathbb{E}$ be the subfield of $\mathbb{E}$ generated by the roots of $p(x)$ in $\mathbb{E}$. Then, $\mathbb{K}$ is a finite dimensional vector space over $\mathbb{F}$ and $\alpha \in \mathbb{K}$. Since σ is an $\mathbb{F}$-embedding, it maps roots of $p(x)$ to roots of $p(x)$. Thus, $\sigma(\mathbb{K}) \subseteq \mathbb{K}$. Since σ is an $\mathbb{F}$-linear map and $\mathbb{K}$ is a finite dimensional vector space over $\mathbb{F}$, $\sigma|_{\mathbb{K}}$ is onto and contains α in its image. Since $\alpha \in \mathbb{E}$ was arbitrary, we are done. □

Theorem 18.6. Let $\mathbb{F}$ be a field and fix an algebraic closure $\overline{\mathbb{F}}$ of $\mathbb{F}$. Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \overline{\mathbb{F}}$ be fields. Then, the following are equivalent.

1. Every $\mathbb{F}$-embedding $\sigma\colon \mathbb{E} \to \overline{\mathbb{F}}$ is an automorphism of $\mathbb{E}$.
2. $\mathbb{E}$ is a splitting field of a family of polynomials in $\mathbb{F}[x]$.
3. $\mathbb{E}/\mathbb{F}$ is a normal extension.

Proof. (1 $\implies$ 2) Let $a \in \mathbb{E}$ and $p_a(x) = \operatorname{irr}(a, \mathbb{F})$. If $b \in \overline{\mathbb{F}}$ is a root of $p_a(x)$, then there exists an $\mathbb{F}$-isomorphism $\mathbb{F}(a) \text{to} \overline{\mathbb{F}}$ with $a \mapsto b$. Extend this to a map $\sigma\colon \mathbb{E} \to \overline{\mathbb{F}}$. By hypothesis, we have

$\mathbb{E} = \sigma(\mathbb{E}) \ni b$. Thus, $\mathbb{E}$ is a splitting field of the family $\{p_a(x)\}_{a\in\mathbb{E}}$.

(2 $\implies$ 3) Let $\mathbb{E}$ be a splitting field of the family $\{p_i(x)\}_{i\in I} \subseteq \mathbb{F}[x]$ over $\mathbb{F}$. Let $f(x) \in \mathbb{F}[x]$ be an irreducible having a root $a \in \mathbb{E}$. Let $b \in \overline{\mathbb{F}}$ be any root of $f(x)$. There exists an $\mathbb{F}$-embedding $\mathbb{F}(a) \to \overline{\mathbb{F}}$ with $a \mapsto b$. Extend this to an $\mathbb{F}$-embedding $\sigma\colon \mathbb{E} \to \overline{\mathbb{F}}$. Since σ fixes $\mathbb{F}$, it maps roots of $p_i(x)$ to its roots for all $i \in I$. Since $\mathbb{E}$ is generated by these roots, we see that $\sigma(\mathbb{E}) \subseteq \mathbb{E}$ and hence $b \in \mathbb{E}$. Thus, all roots of $f(x)$ lie in $\mathbb{E}$, and hence, $f(x)$ splits linearly over $\mathbb{E}$. Since $f(x)$ was an arbitrary irreducible polynomial, we are done.

(3 $\implies$ 1) Let $\sigma\colon \mathbb{E} \to \overline{\mathbb{F}}$ be an $\mathbb{F}$-embedding. Let $a \in \mathbb{E}$. Then, $p(x) := \operatorname{irr}(a, \mathbb{F})$ splits linearly over $\mathbb{E}$. Since σ fixes $\mathbb{F}$, $\sigma(a)$ is a root of $p(x)$, and thus $\sigma(a) \in \mathbb{E}$. Thus, $\sigma(\mathbb{E}) \subseteq \mathbb{E}$. By Lemma 18.5, we get that σ is an automorphism of $\mathbb{E}$. (Note that $\mathbb{E}/\mathbb{F}$ is indeed algebraic since $\mathbb{E} \subseteq \overline{\mathbb{F}}$.) □

Proposition 18.7. Let $\mathbb{F} \subseteq \mathbb{E}_1, \mathbb{E}_2 \subseteq \mathbb{K}$ be fields. Suppose that $\mathbb{E}_i/\mathbb{F}$ are normal. Then, so are $\mathbb{E}_1\mathbb{E}_2/\mathbb{F}$ and $(\mathbb{E}_1 \cap \mathbb{E}_2)/\mathbb{F}$.

Proof. Fix an algebraic closure $\overline{\mathbb{F}} \supseteq \mathbb{K}$. Let $\sigma\colon \mathbb{E}_1\mathbb{E}_2 \to \overline{\mathbb{F}}$ be an $\mathbb{F}$-embedding. Then, $\sigma(\mathbb{E}_1\mathbb{E}_2) = \sigma(\mathbb{E}_1)\sigma(\mathbb{E}_2) = \mathbb{E}_1\mathbb{E}_2$. Since this is true for any $\mathbb{F}$-embedding, $\mathbb{E}_1\mathbb{E}_2/\mathbb{F}$ is normal by Theorem 18.6. Similar calculations also hold for the intersection. □

Example 18.8. Quadratic extensions are always normal. Pick $\alpha \in \mathbb{E} \setminus \mathbb{F}$, then $\mathbb{E} = \mathbb{F}(\alpha)$ is a splitting field of $\operatorname{irr}(\alpha, \mathbb{F})$ over $\mathbb{F}$.

Remark 18.9. Unlike the "tower laws" for algebraic and separable extensions, the "composition" of normal extensions need not be normal. For example, consider the chain

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[4]{2}).$$

Each successive extension is quadratic and hence normal. However, $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not normal since the irreducible (via Eisenstein-Schönemann Criterion) polynomial $x^4 - 2 \in \mathbb{Q}[x]$ has a root in $\mathbb{Q}(\sqrt[4]{2})$ but does not factor completely. However, one part of the "tower law" does hold, as can be easily verified.

Proposition 18.10. Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be fields. If $\mathbb{K}/\mathbb{F}$ is normal, then so is $\mathbb{K}/\mathbb{E}$.

§19. Galois Extensions

§§19.1. Introduction

Definition 19.1. A field extension $\mathbb{E}/\mathbb{F}$ is called a Galois extension if it is normal and separable. The Galois group of a Galois extension $\mathbb{E}/\mathbb{F}$ is the group of all $\mathbb{F}$-automorphisms of $\mathbb{E}$ under composition. It is denoted as $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$.

If $f(x) \in \mathbb{F}[x]$ is a separable polynomial and $\mathbb{E}$ is a splitting field of $f(x)$ over $\mathbb{F}$, then $\mathbb{E}/\mathbb{F}$ is a Galois extension and the Galois group of $f(x)$ over $\mathbb{F}$ is defined to be $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ and is denoted as $\mathrm{Gal}(f(x), \mathbb{F})$ or simply G_f if $\mathbb{F}$ is clear from context.

Remark 19.2. The definition of $\mathrm{Gal}(f(x), \mathbb{F})$ does not depend on the splitting field chosen, up to isomorphism. Let $\mathbb{E}$ and $\mathbb{E}'$ be two splitting fields of $f(x)$ over $\mathbb{F}$. By the Isomorphism of Splitting Fields, there is an $\mathbb{F}$-isomorphism $\tau\colon \mathbb{E} \to \mathbb{E}'$. Then, $\sigma \mapsto \tau \circ \sigma \circ \tau^{-1}$ is an isomorphism from $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ to $\mathrm{Gal}(\mathbb{E}'/\mathbb{F})$.

Example 19.3.

1. Let $\mathbb{E}/\mathbb{F}$ be an extension of finite fields. Then, $|\mathbb{F}| = q$ and $|\mathbb{E}| = q^n$ for some prime power q and some $n \in \mathbb{N}^+$. Then, $\mathbb{E}$ is a splitting field for $x^{q^n} - x \in \mathbb{F}[x]$ over $\mathbb{F}$. Thus, the extension is normal. Since the fields are finite, it is also separable. Thus, $\mathbb{E}/\mathbb{F}$ is Galois.
2. The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is **not** Galois. Since $\mathrm{ch}(\mathbb{Q}) = 0$, it is separable by Corollary 16.16. However, it is not normal since the irreducible (via Eisenstein-Schönemann Criterion) polynomial $x^3 - 2 \in \mathbb{Q}[x]$ has a root in $\mathbb{Q}(\sqrt[3]{2})$ but does not split as a product of linear factors.

Proposition 19.4. Let $\mathbb{E}/\mathbb{F}$ be a finite Galois extension. Then, $|\mathrm{Gal}(\mathbb{E}/\mathbb{F})| = [\mathbb{E}\colon \mathbb{F}]_s = [\mathbb{E}\colon \mathbb{F}]$.

Proof. Fix an algebraic closure $\overline{\mathbb{F}} \supseteq \mathbb{E}$. Let $n : [\mathbb{E}\colon \mathbb{F}]_s$. Let $\sigma_1, \ldots, \sigma_n\colon \mathbb{E} \to \overline{\mathbb{F}}$ be $\mathbb{F}$-embeddings. Since $\mathbb{E}/\mathbb{F}$ is normal, $\sigma_i \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. Thus, $|\mathrm{Gal}(\mathbb{E}/\mathbb{F})| \geq n$. On the other hand, if $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$, then σ is an $\mathbb{F}$-embedding of $\mathbb{E}$ into $\overline{\mathbb{F}}$ when composed with the inclusion. Thus, $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = \{\sigma_1, \ldots, \sigma_n\}$. The last equality follows by Theorem 16.26. □

Remark 19.5. The above proposition shows why both normality and separability are needed. If the extension is normal but not separable, then the order of the group would be the separable degree, which would not be equal to the degree by Theorem 16.26 again.

On the other hand, if the extension was separable but not normal, then there would be an extension $\sigma\colon \mathbb{E} \to \overline{\mathbb{F}}$ that would map $\mathbb{E}$ outside $\mathbb{E}$ and so, not all extensions will belong to the Galois group.

For example, consider $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$. Since there is only one root of $x^3 - 2 \in \mathbb{Q}[x]$ in $\mathbb{Q}(\sqrt[3]{2})$, there is only one $\mathbb{Q}$-automorphism of $\mathbb{Q}(\sqrt[3]{2})$.

Proposition 19.6. Let q be a prime power. The Galois group of the Galois extension $\mathbb{F}_{q^n}/\mathbb{F}_q$ is the cyclic group of order n generated by the Frobenius automorphism $\varphi\colon \mathbb{F}_{q^n} \to \mathbb{F}_{q^n}$ defined by $a \mapsto a^q$.

Proof. φ is an $\mathbb{F}_q$-automorphism since any $a \in \mathbb{F}_q$ satisfies $x^q - x$. Thus, $\varphi \in \mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$. By Proposition 19.4, we have $\left|\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)\right| = n$. Thus, it suffices to show that φ has order at least n. Let the order of φ be d. Then, $d \leq n$. Note that

$$\varphi^d(a) = a^{q^d}.$$

If $\varphi^d = \mathrm{id}_{\mathbb{F}_{q^n}}$, then every $a \in \mathbb{F}_{q^n}$ satisfies $x^{q^d} - x$. Thus, $q^d \geq q^n$ and thus $d \geq n$. Hence, $d = n$. □

Example 19.7. An extension $\mathbb{K}/\mathbb{F}$ is called biquadratic if $[\mathbb{K}\colon \mathbb{F}] = 4$ and $\mathbb{K}$ is generated over $\mathbb{F}$ by roots of two irreducible separable quadratic polynomials. In particular, $\mathbb{K}/\mathbb{F}$ is Galois. Write $\mathbb{K} = \mathbb{F}(\alpha, \beta)$ and let $p(x) := \mathrm{irr}(\alpha, \mathbb{F})$ and $q(x) := \mathrm{irr}(\beta, \mathbb{F})$. Let $\overline{\alpha}, \overline{\beta} \in \mathbb{K}$ denote the other root of $p(x)$ and $q(x)$ respectively. By separability, $\alpha \neq \overline{\alpha}$ and $\beta \neq \overline{\beta}$.

Since $\left[\mathbb{F}(\alpha, \beta)\colon \mathbb{F}\right] = 4$, $p(x)$ is irreducible over $\mathbb{F}(\beta)$ and $q(x)$ over $\mathbb{F}(\alpha)$. The four automorphisms are thus determined by sending α to α or $\overline{\alpha}$ and β to β or $\overline{\beta}$.

Define the automorphisms $\tau, \sigma\colon \mathbb{K} \to \mathbb{K}$ by

$$\tau(\alpha) = \overline{\alpha}, \tau(\beta) = \beta,$$
$$\sigma(\alpha) = \alpha, \tau(\beta) = \overline{\beta}.$$

Then, $\tau^2 = \sigma^2 = \mathrm{id}_{\mathbb{K}}$. Thus, $\mathrm{Gal}(\mathbb{K}/\mathbb{F}) = \mathbb{Z}_2 \times \mathbb{Z}_2$, the Klein-four group, V_4.

Example 19.8 (Galois group of a separable cubic)**.** Let $\mathbb{F}$ be a field with $\mathrm{ch}(\mathbb{F}) \neq 2, 3$. Let $f(x) = x^2 + px + q \in \mathbb{F}[x]$ be an irreducible cubic. In particular, $f(x)$ has no roots in $\mathbb{F}[x]$. Note that

$$f'(x) = 3x^2 + p \neq 0$$

since $\mathrm{ch}(\mathbb{F}) \neq 3$. Thus, $f(x)$ is separable by Proposition 16.10. Thus, a splitting field $\mathbb{E}$ of $f(x)$ over $\mathbb{F}$ must have degree either 3 or 6. Thus, by Proposition 19.4, $|\mathrm{Gal}(\mathbb{E}/\mathbb{F})| = 3$ or 6. We now show how the discriminant determines this.

Let $\mathbb{E} = \mathbb{F}(\alpha_1, \alpha_2, \alpha_3)$, where $f(x) = \prod_{i=1}^{3}(x - \alpha_i)$. Any $\sigma \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ permutes these roots. Let $p_\sigma \in S_3$ denote the corresponding permutation. Then, $\sigma \mapsto p_\sigma$ is injective. Under this, we identify $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ with a subgroup of S_3. Thus, $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = A_3$ or S_3.

Let

$$\delta = (\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_3 - \alpha_1).$$

Then, $\delta^2 = \mathrm{disc}(f(x)) = -(4p^3 + 7q^2) \in \mathbb{F}$. Thus, $\left[\mathbb{F}(\delta)\colon \mathbb{F}\right] \leq 2$. Now, if $\delta \in \mathbb{F}$, then $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ cannot have any odd permutations since they do not fix δ, and thus $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = A_3$. On the other hand, if $\delta \notin \mathbb{F}$, then $2 = \left[\mathbb{F}(\delta)\colon \mathbb{F}\right] \mid [\mathbb{E}\colon \mathbb{F}]$, and thus $\mathrm{Gal}(\mathbb{E}/\mathbb{F}) = S_3$.

Note that $\delta \in \mathbb{F} \iff \operatorname{disc}(f(x))$ is a perfect square in $\mathbb{F}$, so the above is completely determined by the discriminant being a perfect square. For example, if $f(x) = x^3 + x + 1 \in \mathbb{Q}[x]$, then $\operatorname{disc}(f(x)) = -31$ and $\operatorname{Gal}(\mathbb{E}/\mathbb{Q}) \cong S_3$. On the other hand, if $f(x) = x^3 + 3x + 1$, then $\operatorname{disc}(f(x)) = 81 = 9^2$, and thus, $\operatorname{Gal}(\mathbb{E}/\mathbb{Q}) \cong A_3$.

§§19.2. The Fundamental Theorem of Galois Theory

Definition 19.9. Let $\mathbb{E}$ be a field and let G be *a* group of automorphisms of $\mathbb{E}$. Then,

$$\mathbb{E}^G := \{a \in \mathbb{E} \mid \sigma(a) = a \text{ for all } \sigma \in G\}$$

is called the fixed field of G acting on $\mathbb{E}$.

Theorem 19.10 (Fundamental Theorem of Galois Theory (FTGT))**.** Let $\mathbb{K}/\mathbb{F}$ be a *finite* Galois extension. Consider the sets

$$\mathcal{I} = \{\mathbb{E} \mid \mathbb{E} \text{ is an intermediate field of } \mathbb{K}/\mathbb{F}\} \text{ and } \mathcal{G} = \{H \mid H \leq \operatorname{Gal}(\mathbb{K}/\mathbb{F})\}.$$

1. The maps
$$\mathbb{E} \mapsto \operatorname{Gal}(\mathbb{K}/\mathbb{E}) \text{ and } H \mapsto \mathbb{K}^H$$
gives a one-to-one correspondence between $\mathcal{I}$ and $\mathcal{G}$, called the Galois correspondence. Moreover, this correspondence is inclusion reversing.
2. $\mathbb{E}/\mathbb{F}$ is Galois iff $\operatorname{Gal}(\mathbb{K}/\mathbb{E}) \trianglelefteq \operatorname{Gal}(\mathbb{K}/\mathbb{F})$, and in this case
$$\operatorname{Gal}(\mathbb{E}/\mathbb{F}) \cong \frac{\operatorname{Gal}(\mathbb{K}/\mathbb{F})}{\operatorname{Gal}(\mathbb{K}/\mathbb{E})}.$$
3. $\mathbb{K}/\mathbb{E}$ is always Galois and $|\operatorname{Gal}(\mathbb{K}/\mathbb{E})| = [\mathbb{K}\colon \mathbb{E}] = \dfrac{[\mathbb{K}\colon \mathbb{E}]}{[\mathbb{E}\colon \mathbb{F}]}$.
4. If $\mathbb{E}_1, \mathbb{E}_2 \in \mathcal{I}$ correspond to $H_1, H_2 \in \mathcal{G}$, then $\mathbb{E}_1 \cap \mathbb{E}_2$ corresponds to $\langle H_1, H_2 \rangle$ and $\mathbb{E}_1\mathbb{E}_2$ corresponds to $H_1 \cap H_2$.

Proof. Omitted due to time constraints. □

References

[1] Lectures on MA419: Basic Algebra, *Sudhir R. Ghorpade and Jugal K. Verma*.

[2] Abstract Algebra, *David S. Dummit and Richard M. Foote*.

[3] Notes on Galois Theory, *Aryaman Maithani*.

[4] Lectures on MA414: Galois Theory, *Jugal K. Verma*.