MA3201 + MA4203 Notes

Lou Yi

April 30, 2025

Contents

1 Introduction To Rings

1.1 Basic Definitions

Algebraic structure are non-empty sets equipped with one or more operations and are are required to satisfy certain axioms/properties.
Examples: Semi-group, Monoid, Group, Ring, Field, Vector Space, Modules, Algebra.

Definition 1.1 (Ring) *A **ring** R is a set with two binary operations, addition (denoted by $a+b$) and multiplication (denoted by $a \cdot b = ab$), such that for all $a, b, c \in \mathbb{R}$:*

1. *(Commutative under addition) $a + b = b + a$.*
2. *(Associativity under addition) $(a + b) + c = a + (b + c)$.*
3. *(Additive identity) There is an additive identity 0. That is, there is an element 0 in R such that $a + 0 = 0 + a = a$ for all $a \in R$.*
4. *(Additive inverse) There is an element $-a$ in R such that $a + (-a) = 0$.*
5. *(Associativity under multiplication) $a(bc) = (ab)c$*
6. *(Distribution) $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$.*

Remark 1.1.1 *Axiom 1,2,3,4 basically says R with the binary operations $+$ forms an abelian group.*

Definition 1.2 (Commutative Ring) *A ring is said to be a **commutative** if it is commutative under multiplication, that is $\forall a, b \in R$, we have $ab = ba$.*

Definition 1.3 (Ring with Unity and Units) *If there exists $e \in R$ such that $a \cdot e = a = e \cdot a$, $\forall a \in R$, then e is called **unity / multiplicative identity**. In this case, the ring is called a **Ring with unity / Ring with multiplicative identity**. If $e \neq 0$ is a unity of the ring, and for $b \in R$, $\exists c \in R$, such that $bc = cb = e$, then b is said to be a **unit element** of the ring R and c is called the **multiplicative inverse** of b.*

Example:

- $(\mathbb{Z}, +, \cdot)$ is a commutative ring with the unity 1 and ± 1 are units of the ring.
- $(2\mathbb{Z}+, \cdot)$ is a commutative ring with no unity.
- $(\mathbb{Z}[x], +, \cdot)$ is the ring of polynomials with integer coefficients. It is commutative, with the unity 1 and ± 1 are the units of the ring.
- $(M_n(\mathbb{Z}), +, \cdot)$ is a non-commutative ring, with the unity Id_n, and the set $\mathrm{GL}_n(\mathbb{Z})$ forms the units of the ring.
- Let X be any nonempty set and let A be any ring, then we have the ring of functions R which is the set of all functions from $f : X \to A$, under pointwise addition and multiplication. R is commutative if and only if A is commutative, and R has a unity (the constant function 1) if and only if A has a unity.

The following properties are easy to prove.

Proposition 1.4 (Properties of Rings) *Let R be a ring, and $a, b, c \in R$. Then*

1. $a0 = 0a = 0$.
2. $a(-b) = (-a)b = -(ab)$.
3. $(-a)(-b) = ab$.
4. $a(b-c) = ab - ac$ *and* $(b-c)a = ba - ca$.

Furthermore, if R has a unity element 1. Then it is unique and

5 $(-1)a = -a$.

6 $(-1)(-1) = 1$.

7 *The multiplicative inverse of an element in R is unique.*

Remark 1.4.1 *The subtraction operation $-$ is defined by $a - b = a + (-b)$, where $-b$ is the additive inverse of b.*

Definition 1.5 (Zero Ring) *A ring is called a **zero ring** if the additive identity is equal to the multiplicative identity.*

Lemma 1.6 *A zero ring consists of only one element, which is precisely the additive identity and the multiplicative identity of the ring.*

Proof: Let a be any element of the zero ring, and 1 be the multiplicative identity. Then $a = a \cdot 1 = 0$. □

Lemma 1.7 *Let R be a ring with unity and $R^\times$ (sometimes denoted $U(R)$) denote the set of unit elements of R. Then $R^\times$ is a group under multiplication, and is referred to as the **group of units of** R. In addition, if R is commutative, then $(R^\times, \cdot)$ is an abelian group.*

Proof: Note $R^\times$ has a identity 1 which is the unity of the ring R. Now suppose $a, b \in R^\times$, then $ac = ca = 1$ and $bd = db = 1$ for some $c, d \in \mathbb{R}$, then $abdc = 1$, so $ab \in R^\times$. Clearly, the multiplicative inverse of any $b \in R^\times$ is in $R^\times$. Associativity of multiplication is also clear. Hence $R^\times$ is a group. The second assertion is also clear. □

1.2 Subrings

Definition 1.8 (Subring) *A **Subring** S of the ring R is a subgroup (with respect to addition) of R that is closed under multiplication.*

Remark 1.8.1 *A subset S or the ring R is a subring if S is itself a ring with the operations of R. That is a a nonempty subset S of a ring R is a subring if closed under subtraction and multiplication, that is $a - b$ and ab are in S whenever a and b are in S.*

Example:

- $\{0\}$ and R is always a subring of R.
- $n\mathbb{Z}$ is a subring of $\mathbb{Z}$.
- $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ as subrings.
- $S = \{\bar{0}, \bar{2}, \bar{4}\}$ is a subring of $\mathbb{Z}/6\mathbb{Z}$. Note the multiplicative identity of $\mathbb{Z}/6\mathbb{Z}$ is $\bar{1}$ and the multiplicative identity of S is $\bar{2}$, the two are different.
- The integral Quaternions forms a subring of either the real or rational Quaternions.

Proposition 1.9 *A ring and its subrings may have different multiplicative identity.*

Proposition 1.10 *The arbitrary intersection of a collection of subring is a subring.*

1.3 Integral Domain

Definition 1.11 (Zero Divisor) *A nonzero element a of R is called a **zero divisor** if there is a nonzero element b in R such that either $ab = 0$ or $ba = 0$.*

Example:

- In $\mathbb{Z}/6\mathbb{Z}$, $\bar{2}$, $\bar{3}$ are zero divisors, as $\bar{2}\bar{3} = 0$.
- In $\mathbb{Z}/n\mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z} \setminus \{0\} \setminus (\mathbb{Z}/n\mathbb{Z})^*$ are the set of all zero divisors of the ring.
- The multiplicative identity 1 is never a zero divisor of any ring.

Lemma 1.12 *A zero divisor cannot be a unit and a unit cannot be a zero divisor.*

Proof: Suppose a is a zero element of R, and $\exists b \in R$ such that $b \neq 0$ and $ab = 0$ or $ba = 0$. Suppose $v \in R$ is such that $av = va = 1$. Then if $ab = 0$, then $b = 1b = (va)b = v0 = 0$ which is a contradiction; similarly, if $ba = 0$, then $b = b1 = b(av) = (ba)v = 0v = 0$.

Suppose a is a unit and let b be its multiplicative inverse, then if $ac = 0$ for some $c \in R$, we have $0 = b(ac) = (ba)c = 1c = c$ so $c = 0$; similarly, we can show if $ca = 0$ for some $c \in R$, then $c = 0$. Hence a is not a zero divisor. □

Definition 1.13 (Integral Domain) *A commutative ring with identity $1 \neq 0$ is called an **integral domain** if it has no zero divisors.*

Example:

- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{Z}[i]$ is an integral domain.
- $\mathbb{Z}/n\mathbb{Z}$ is an integral domain if and only if n is a prime.
- $M_{2\times 2}(\mathbb{Z})$ is not an integral domain.

Proposition 1.14 (Cancellation Law) *Assume a, b and c are elements of any ring with a not a zero divisor. If $ab = ac$, then either $a = 0$ or $b = c$. In particular. if a, b, c are any elements in an integral domain and $ab = ac$, then either $a = 0$ or $b = c$.*

Proof: If $ab = ac$, then $a(b-c) = 0$ so either $a = 0$ or $b-c = 0$, since a is not a zero divisor. The second assertion is clear. □

Lemma 1.15 *Let S be any subring of an integral domain R. If S contains a multiplicative identity that isn't 0_R, then S is an integral domain.*

Proof: Since S is a subring of R, then it is clear that S is commutative and and has no zero divisors. Next S has a multiplicative identity that isn't $0_R = 0_S$. Hence S is an integral domain. □

Lemma 1.16 *Every field is an integral domain.*

Proof: Since every non-zero element of the field is a unit, then it is not a zero divisor. Hence the field F has no zero divisor, and it is commutative (as it is a field), so F is an integral domain. □

Corollary 1.16.1 *Suppose R is a subring of a field F, and R contains the unity of F. Then R is an integral domain.*

Proof: Follows directly from the previous two lemmas. □

Definition 1.17 (Idempotent Element) *An element $a \in R$ is **idempotent** if $a^2 = a$.*

Corollary 1.17.1 *The only idempotent element of an integral domain are 0 and 1.*

Proof: It is clear that 0 is an idempotent element. Now suppose $a \in R$ is not zero and idempotent, we show $a = 1$. Since $a^2 = a = a \cdot 1$, then by the cancellation law, we have $a = 1$. □

Definition 1.18 (Nilpotent Element) *An element a of a ring R is called **Nilpotent** if $a \neq 0$ and $a^m = 0$ for some $m \in \mathbb{N}^+$.*

Lemma 1.19 *If $a \in R$ is idempotent, then $\forall n \in \mathbb{N}^+$, we have $a^n = a$.*

Proof: Induction. □

Corollary 1.19.1 *Any idempotent element of of a ring is not nilpotent.*

Proposition 1.20 *Let R be a commutative ring. Then the sum of a nilpotent element of R and a unit in R is a unit.*

Proof: Let a be a unit with inverse b and x is nilpotent with $x^m = 0$. Then by direct computation, one can check that

$$b - b^2x + b^3x^2 - \cdots + (-1)^{m-1}b^m x^{m-1}$$

is the multiplicative inverse of $a + x$. □

Proposition 1.21 *Any finite integral domain is a field.*

Proof: Let R be a finite integral domain and let a be a nonzero element of R. By the cancellation law, the map $x \mapsto ax$ is an injective function. Since R is finite, this map is a bijection. In particular, there is some $b \in R$ such that $ab = 1$, i.e., a is a unit in R ($a(ab) = a = (ab)a = a(ba)$, then by cancellation, we get $ba = 1$ as well). Since a is arbitrary nonzero element, and an integral domain is automatically commutative, then R is a field. □

1.4 Characteristic of a Ring

Definition 1.22 (Characteristic of a Ring) *The **characteristic of a ring** R is the least positive integer n such that $nx = 0$, $\forall x \in R$. If no such number exists, then the characteristic of the ring is said to be 0. The characteristic of a ring R is denoted* $\mathrm{Char}(R)$.

Example:

- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, $\mathbb{C}$ and $\mathbb{Z}[i]$ has zero characteristic.
- $\mathrm{Char}(\mathbb{Z}/n\mathbb{Z}) = n$.

Lemma 1.23 *Suppose R has a unity element 1. Then the characteristic of R is equal to the smallest positive number n such that $n \cdot 1 = 0$. If such number does not exists, then* $\mathrm{Char}(R) = 0$.

Proof: It is clear that if no such n exists, then $\mathrm{Char}(R) = 0$. Now if $n \neq 0$ and $n \cdot 1 = 0$, we have $\mathrm{Char}(R) \geq n$. We show $\forall x \in R$, $nx = 0$. Since

$$nx = n(1x) = (n \cdot 1)(x) = 0(x) = 0.$$

Then the claim holds. □

Theorem 1.24 (Characteristic of an Integral Domain) *Suppose R is a integral domain, then the characteristic of R is 0 or a prime.*

Proof: Let R be an integral domain, and let $n = \mathrm{Char}(R)$. Assume $n \neq 0$, we show that n is a prime. We prove this using contradiction, suppose n is not a prime, then $n = ab$ for some $1 < a, b < n$. By Lemma 1.23, we have

$$n_R = 0 \Longrightarrow a_R b_R = 0$$

Since R is an integral domain, this implies $a_R = 0$ or $b_R = 0$. However, this is a contradiction, as $a, b < n$. □

Lemma 1.25 *The characteristic of a ring R is the number n such that if the statement $ka = 0$ for all $a \in R$, then $n|k$.*

Proof: Division algorithm. □

Theorem 1.26 *If R is a ring with unity, and $\mathrm{Char}(R) = 0$, then R is infinite.*

Proof: Since $1 \in R$, then $n \cdot 1 \in R$ for any $n \in \mathbb{N}^+$ by induction. But then $n \cdot 1$ and $m \cdot 1$ must be distinct for $n \neq m$, as otherwise, the characteristic of R divides $|n - m|$. Hence R has infinitely many different elements, so it must be infinite. □

Corollary 1.26.1 *If R is finite, then R has non-zero characteristic.*

Corollary 1.26.2 *Suppose R is a ring with unity, and $\exists n \in \mathbb{N}^+$, s.t., $n \cdot 1 = 0$, then $\mathrm{Char}(R) \neq 0$ and $\mathrm{Char}(R) | n$.*

Definition 1.27 (Boolean Ring) *A ring R is called a **Boolean ring** if $a^2 = a$ for all $a \in R$.*

Lemma 1.28 *Every nontrivial Boolean ring has characteristic 2.*

Proof: Let $a \in R$, then $a + a = (a + a)^2 = a^2 + a^2 + a^2 + a^2 = a + a + a + a$. So $a + a = 0$. □

Lemma 1.29 *Every Boolean ring is commutative.*

Proof: Let $x, y \in R$ where R is a Boolean ring. Then $x + y = (x + y)^2 = x + xy + yx + y$, so $xy = yx$, as R has characteristic 2. □

1.5 Division Ring

Definition 1.30 (Division Ring) *A ring R with identity 1, where $1 \neq 0$, is called a **Division Ring (or skew field)** if every nonzero element $a \in R$ has a multiplicative inverse, i.e. every non zero element of a division ring is a unit element. A commutative division ring is called a **field**, i.e., it is a commutative ring F with identity $1 \neq 0$ in which every nonzero element is a unity: $F^\times = F \setminus \{0\}$.*

Remark 1.30.1 *A field $(F, +, \cdot)$ is a algebraic structure such that the multiplicative and additive identity are distinct, $(F, +)$ and $(F \setminus \{0\}, \cdot)$ are abelian groups, and satisfies the distribution property:*

$$\forall a, b, c \in F,\ (a + b)c = ab + bc \ \text{ and } \ a(b + c) = ab + ac.$$

Example:

- $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields, $\mathbb{Z}$ is not a division ring.
- $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ is a field.
- $R = \left\{ \begin{pmatrix} z & w \\ \bar{w} & \bar{z} \end{pmatrix} : z, w \in \mathbb{C} \right\}$ is a division ring but not a field.

- The **real Hamilton Quaternions** $\mathbb{H}$ is the collection of the form $a+bi+cj+dk$ where $a, b, c, d \in \mathbb{R}$ and ring operation is defined similar to Definition 2.3. Similarly one can define **rational Hamilton Quaternions**. Both the real and rational Hamilton Quaternions are division rings where the inverse of nonzero elements are given by
$$(a+bi+cj+dk)^{-1} = \frac{a-bi-cj-dk}{a^2+b^2+c^2+d^2}.$$

Definition 1.31 (Center) *The **center** of a ring R is defined to be the set*

$$\{z \in R \,|\, zr = rz \ \forall r \in R\},$$

i.e., it is the set of all elements which commute with every element of R. Let $a \in R$, we use $C(a)$ to denote the set of elements in R that commutes with a, that is

$$C(a) = \{r \in R \,|\, ra = ar\}.$$

One can clearly see that the center of R is the intersection of $C(a)$ over all $a \in R$.

Proposition 1.32 *The center of a ring is a subring that contains the identity (if the original ring R has an identity). Then center of a division ring is a field.*

Proof: Let Z denote the center of the ring R. Suppose $a, b \in Z$, then we show $a-b$ and $ab \in Z$. Let $\forall x \in R$, we have

$$(a-b)x = ax - bx = xa - xb = x(a-b)$$

and

$$abx = axb = xab = x(ab).$$

Consequently Z is a subring. And it is clear that if R has an identity, then Z contains that same identity.

Next, if R is a division ring. Then we show Z is a field. Suffices to show that every nonzero element of Z has a multiplicative inverse, since Z is automatically commutative. Suppose $a \in Z$, then $ab = ba = 1_R$ for some $b \in R$. We show $b \in Z$, that is the multiplicative inverse of a is in Z. Note $\forall x \in R$,

$$bx = bx(ab) = baxb = xb$$

so $b \in Z$. Thus Z is a field. □

Lemma 1.33 *$C(a)$ is a subring of R containing a and if R is a division ring, then $C(a)$ is a division ring.*

1.6 More Rings

Definition 1.34 (Direct Product of Rings) *Let I be any nonempty index set and let R_i be a ring for each $i \in I$. Then we define the **direct product** $\prod_{i \in I} R_i$ to be the set of all tuples with componentwise addition and multiplication.*

Remark 1.34.1 *Note the direct product of rings is indeed a ring. It is commutative if and only if each R_i is commutative; it has an identity if and only if every R_i has an identity.*

Definition 1.35 (Discrete Valuation Ring) *Let K be a field. A **discrete valuation** on K is a function $v : K^\times \to \mathbb{Z}$ satisfying:*

1. *$v(ab) = v(a) + v(b)$;*
2. *v is surjective;*
3. *$v(x + y) \geq \min\{v(x), v(y)\}$ for all $x, y \in K^\times$ with $x + y \neq 0$.*

*The set $R = \{x \in K^\times \mid v(x) \geq 0\} \cup \{0\}$ is called the **discrete valuation ring of** v.*

Remark 1.35.1 *Note R is always a subring of K which contains the identity. One can also show that x is a unit of R if and only if $v(x) = 0$.*

Definition 1.36 (Local Ring) *A commutative ring with $1 \neq 0$ is called a **local ring** if it contains a unique maximal ideal.*

Proposition 1.37 *Let K be a field and $\nu : K^\times \to \mathbb{Z}$ be a discrete valuation on K. Let R the be discrete valuation ring on ν, then R is a local ring.*

1.7 Ring Homomorphism and Ideals

Definition 1.38 *Let R and S be rings.*

1. *A **ring homomorphism** is a map $\varphi : R \to S$ satisfying:*
   - *$\varphi(a + b) = \varphi(a) + \varphi(b)$ for all $a, b \in R$;*
   - *$\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R$.*
2. *The **kernel** of the ring homomorphism φ, denoted $\ker \varphi$, is the set of elements of R that maps to 0 in S.*
3. *The **fiber** of the ring homomorphism above $s \in S$, is the set $\varphi^{-1}(s)$.*
4. *A bijective ring homomorphism is called an **isomorphism**, and is denoted $R \cong S$.*

Example:

1. The map $\mod : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is a ring homomorphism. The kernel of $\mod$ is $\{kn : k \in \mathbb{Z}\}$. The fiber of φ above m is $\{kn + m : k \in \mathbb{Z}\}$.
2. Let $\varphi : \mathbb{Q}[x] \to \mathbb{Q}$ be defined by $\varphi(p(x)) = p(0)$.Then φ is a ring homomorphism, with kernel being the set of all polynomials with constant term 0, and the fiber of φ above k to be the set of all polynomials with constant term k.

Definition 1.39 (Ideal Of A Ring) *Let R be a ring, let I be a subset of R and let $r \in R$.*

1. *$rI = \{ra : a \in I\}$ and $Ir = \{ar : a \in I\}$.*

2. *A subset I of R is a **left ideal of** R if I is a subring of R and I is closed under left multiplication by elements from R, i.e., $rI \subset I$ for all $r \in R$.*

3. *A subset I of R is a **right ideal of** R if I is a subring of R and I is closed under right multiplication by elements from R, i.e., $Ir \subset I$ for all $r \in R$.*

4. *I is said to be an **Ideal (two-sides)** of ring R if I is both a left ideal and right ideal. That is I is a subring of R and $\forall r \in R$ and $a \in I$, $ra \in I$ and $ar \in I$.*

Remark 1.39.1 *It is clear then a subset I of a ring R is an ideal if and only if I is non-empty, $\forall a, b \in I$ and $r \in R$, $a - b, ab, ar$ and ra are also elements of I.*

Remark 1.39.2 *For commutative rings, the notions of left, right and two-sided ideals coincide.*

Proposition 1.40 *Let R and S be rings, and $\varphi : R \to S$ be a ring homomorphism. Then*

1. *the image of φ is a subring of S. In fact, the image of any subring of R under φ is a subring of S.*

2. *the kernel of φ is a subring of R. Futhermore, if $\alpha \in \ker \varphi$ then $r\alpha$ and $\alpha r \in \ker \varphi$ for every $r \in R$. So $I = \ker \varphi$ is an ideal of R.*

3. *The fiber of φ above s is the set $r + \ker \varphi$, where $\varphi(r) = s$.*

4. *The primage of any subring of S is a subring of R containing $\ker \varphi$.*

5. *The primage of any ideal of S is an ideal of R containing $\ker \varphi$.*

6. *If φ is surjective, then the image of any ideal I of R is an ideal of S.*

Proof:

1. If $s_1, s_1 \in \operatorname{Im} \varphi$, then $s_1 = \varphi(r_1)$ and $s_2 = \varphi(r_2)$ for some $r_1, r_2 \in R$. Then $\varphi(r_1 - r_2) = s_1 - s_2$ and $\varphi(r_1 r_2) = s_1 s_2$, so $s_1 - s_2, s_1 s_2 \in \operatorname{Im} \varphi$. Hence φ is a subring of S. The second assertion is similar.

2. If $\alpha, \beta \in \ker \varphi$. Then it is clear that $\varphi(\alpha - \beta) = \varphi(\alpha\beta) =$. In addition, for any $r \in R$, we have $\varphi(ra) = \varphi(r)\varphi(a) = 0$, and $\varphi(ar) = \varphi(a)\varphi(r) = 0$. Hence the second claim holds.

3. If $\varphi(r_1) = s$, then $\varphi(r_1 - r) = 0$, so $r_1 - r \in \ker \varphi$.

4. Similar verification.

5. Similar verification.

6. Similar verification.

□

Example:

- $\{0\}$ and R are always ideals of a ring R. The ideal $\{0\}$ is known as the **trivial ideal** and is denoted 0. An ideal I is **proper** if $I \neq R$.
- $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$.
- Let R be a commutative ring with unity and $a \in R$, then the set $\langle a \rangle = \{ra,\ r \in R\}$ is an ideal of R, and is called the **principle ideal generated by** a.
  We give a quick proof of this. Suppose $x, y \in \langle a \rangle$, then $x = r_1 a$, $y = r_2 a$ for some $r_1, r_2 \in R$. Hence $x - y = (r_1 - r_2)a \in \langle a \rangle$ (by distributivity), and $xy = (r_1 r_2 a)a \in R$ (by commutativity). Lastly, for any $r \in R$, we have $rx = xr = (rr_1)a \in R$ (by commutativity). Thus $\langle a \rangle$ is an ideal.
- Consider the ring of polynomials over reals, $\mathbb{R}[x]$, then $\langle x^2+1 \rangle = \{g(x)(x^2+1) : g(x) \in \mathbb{R}[x]\}$ is the principle ideal generated by $\langle x^2 + 1 \rangle$.
- In $\mathbb{R}[x]$, the set of all polynomials with constant term being zero is an ideal of $\mathbb{R}[x]$ (in fact generated by x).
- Let $R = \mathbb{Z}[x]$, and I be the set of all polynomials whose terms are of degree at least 2 together with the zero polynomial. Then I is an ideal of R.
- Fix some $n \in \mathbb{Z}$ with $n \geq 2$ and consider the noncommutative ring $M_n(R)$. If J is any ideal of R, then $M_n(J)$ is an ideal of $M_n(R)$. This ideal is the kernel of the surjective homomorphism $M_n(R) \to M_n(R/J)$, which reduces each entry of a matrix mod J.
- Let R be a commutative ring with unity and let $G = \{g_1, \cdots, g_n\}$ be a finite group. The map from the group ring RG to R defined by $\sum_{i=1}^{n} a_i g_i \mapsto \sum_{i=1}^{n} a_i$ is a homomorphism, called the **augmentation map**. The kernel of the augmentation map, the **augmentation ideal**, is the set of elements of RG whose coefficients sum to 0.

Lemma 1.41 *Let $\varphi : R \to S$ be a surjective homomorphism. Then the image of the center of R is contained in the center of S.*

Lemma 1.42 *Let R be a ring, and $\{I_\alpha\}$ be a collection of ideals of R. Then $\bigcap_\alpha I_\alpha$ is an ideal of R. Similarly, the intersection of an arbitrary set of left ideals is a left ideal and the intersection of right ideas is a right ideal.*
If $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ are ideals of R, then

$$\bigcup_{n=1}^{\infty} I_n$$

is an ideal of R. The same applies for left and right ideals.

Proof: Direct verification. □

Definition 1.43 *Let I and J be ideals of R.*

1. *The **sum** of I and J is defined to be $I + J = \{a + b : a \in I, b \in J\}$.*
2. *The **product** of I and J, IJ is defined to be the set of all finite sums of elements of the form ab with $a \in I$ and $b \in J$.*

3. *For any $n \geq 1$, define the n^{th} **power of** I, I^n, to be the set consisting of all finite sums of elements of the form $a_1 a_2 \cdots a_n$ with $a_i \in I$ for all i. Equivalently, I^n is defined inductively defining $I^1 = I$, and $I^n = II^{n-1}$ for $n = 2, 3, \cdots$.*

Lemma 1.44 *The sum of two ideals I and J in the ring R is an ideal in R and $I + J$ is the smallest ideal I and J containing both I and J. The product IJ is an ideal contained in $I \cap J$*

Proof: It is clear that $I + J$ is nonempty if I and J are ideals. Let $i_1 + j_1, i_2 + j_2 \in I + J$, $r \in R$, then

$$\begin{aligned}(i_1 + j_1) - (i_2 + j_2) &= (i_1 - i_2) + (j_1 - j_2) \in I + J \\ (i_1 + j_1)(i_2 + j_2) &= i_1 i_2 + i_1 j_2 + j_1 i_2 + j_1 j_2 \in I + I + J + J = I + J \\ r(i_1 + j_1) &= r i_1 + r j_1 \in I + J \\ (i_1 + j_1) r &= i_1 r + j_1 r \in I + J.\end{aligned}$$

So $I + J$ is an ideal. It is also clear that if L is any ideal containing I and J in R, then it is closed under addition, hence contains $I + J$.

Next we show IJ is an ideal. Suffices to show that if $i_1 j_1, i_2 j_2 \in IJ$, $r \in R$, then $i_1 j_1 + i_2 j_2 \in IJ$, $(i_1 j_1)(i_2 j_2) \in IJ$ and $r(i_1 j_1) \in RJ, (i_1 j_1) r \in RJ$. But these are clearly true (note $(i_1 j_1)(i_2 j_2) = (i_1 j_1 i_2) j_2 \in IJ$). Hence IJ is an ideal. Lastly, the generating elements are elements of both I and J, and since ideals are closed under addition, additive inverse and multiplication by elements in R. Then IJ is contained in $I \cap J$. □

Definition 1.45 *Let A be any subset of the ring R,*

1. *Let (A) denote the smallest ideal of R containing A, called the **ideal generated by** A.*
2. *Let RA denote the set of all finite sums of elements of the form ra with $r \in R$ and $a \in A$, i.e., $RA = \{r_1 a_1 + r_2 a_2 + \cdots + r_n a_n \,|\, r_i \in R,\, a_i \in A,\, n \in \mathbb{Z}^+\}$. Similarly, $AR = \{a_1 r_1 + \cdots + a_n r_n \,|\, r_i \in R,\, a_i \in A,\, n \in \mathbb{Z}^+\}$ and $RAR = \{r_1 a_1 r_1' + \cdots + r_n a_n r_n' \,|\, r_i, r_i' \in R,\, a_i \in A,\, n \in \mathbb{Z}^+\}$. We adapt the convention $RA = \{0\}$ if $A = \emptyset$.*
3. *An ideal generated by a single element is called a **principal ideal**.*
4. *An ideal generated by a finite set is called a **finitely generated ideal**.*

Remark 1.45.1 *The ideal generated by a subset A is equal to the intersection of all ideals in R containing A. Similarly, we can define the **left ideal generated by** A and **right ideal generated by** A.*

Lemma 1.46 *If R has identity $1 \neq 0$, $A \subset R$. Then RA is the left ideal generated by A and AR is the right ideal generated by A and RAR is the two-sided ideal generated by A. In particular, if R is commutative, then*

$$RA = AR = RAR = (A).$$

Proof: Since R has identity, then it is clear that RA contains A. Next as R is a ring, then $RRA = RA$, so RA is a left ideal. Moreover, any left idea containing A must contain all the finite sums of elements of the form ra, $r \in R$ and $a \in A$, so must contain RA. Then we conclude RA is precisely the left ideal generated by A. Similarly, we have AR is the right ideal generated by A and RAR is the two-sided ideal generated by A. Lastly, if R is commutative, then all three notions of ideals coincides, hence we have

$$RA = AR = RAR = (A).$$

□

Remark 1.46.1 *If R is commutative ring, and $a \in R$, then the principal ideal (a) generated by a is just the set of all $R-$multiples of a. In this case an element $b \in R$ belongs to the ideal (a) if and only if $b = ra$ for some $r \in R$, and we say a **divides** b **in** R or b **is a multiple of** a. If R is not commutative, the set $\{ras \mid r, s \in R\}$ is not necessarily the two-sided ideal generated by a, since it need not be closed under addition (in this case the ideal generated by a is the ideal RaR, which consists of all finite sums of elements of the form ras, $r, s \in R$).*

Lemma 1.47 *Let $a, b \in R$, then $(b) \subset (a)$ if and only if $b \in (a)$.*

Proof: Forward direction is trivial. For the backward direction, recall (b) is the smallest ideal containing b and (a) is an ideal containing b, thus $(b) \subset (a)$. □

Proposition 1.48 *Let I be an ideal of R with $1 \neq 0$.*

1. *$I = R$ if and only if I contains a unit.*
2. *Assume R is commutative. Then R is a field if and only if its only ideals are 0 and R.*
3. *If the only left side ideals and the only right side ideals of D are 0 and D, then D is a division ring and the converse holds.*

Proof:

1. If $I = R$, then I contains the unit 1; conversely, if u is a unit in I with inverse v, then for any $r \in R$,

   $$r = r \cdot 1 = r(uv) = (rv)u \in I.$$

   Hence $R = I$.

2. The ring R is a field if and only if every nonzero element is a unit. If R is a field every nonzero ideal contains a unit, so by (1), R is the only nonzero ideal. Conversely, if 0 and R are the only ideals of R, let u be any nonzero element of R. By hypothesis $(u) = R$ and so $1 \in (u)$. Thus there is some $v \in R$ such that $1 = uv$.

3. Similar argument to (2), however, in this case we need to run the argument twice to obtain a left inverse and right inverse.

□

Corollary 1.48.1 *If R is a field then any nonzero ring homomorphism from R into another ring is an injection.*

Proof: The kernel of a ring homomorphism is an ideal. The kernel of a nonzero homomorphism is a proper ideal, hence is trivial by Proposition 1.48. □

Definition 1.49 (Maximal Ideal) *An ideal M in an arbitrary ring S is called a* ***maximal ideal*** *if $M \neq S$ and the only ideals containing M are M and S.*

Remark 1.49.1 *Not all rings have maximal ideals, for example, $\mathbb{Q}$ with the usual addition and multiplication defined by $ab = 0$ for all $a, b \in \mathbb{Q}$, has no maximal ideals. Also, from the definition the zero ring has no maximal ideal.*

Proposition 1.50 *In a ring with identity, every proper ideal is contained in a maximal ideal.*

Proof: Let R be a ring with identity and let I be a proper ideal (so R cannot be the zero ring). Let $\mathcal{S}$ be the set of all proper ideals of R which contain I. Then $\mathcal{S}$ is nonempty ($I \in \mathcal{S}$) and is partially ordered by inclusion. If $\mathcal{C}$ is a chain in $\mathcal{S}$, define J to be the union of all ideals in $\mathcal{C}$. Then we show J is an ideal.
It is clear that J is nonempty. If $a, b \in J$, then there are ideals $A, B \in C$ such that $a \in A$ and $b \in B$ and bydefinition of a chain either $A \subset B$ or $B \subset A$. In either case, $a - b \in J$. Similarly, we have J is closed under left and right multiplication by elements of R. So J is an ideal.
If J is not a proper ideal then $1 \in J$. In this case, by definition of J, we must have $1 \in A$ for some $A \in \mathcal{C}$, which is a contradiction because A is a proper ideal. This proves that each chain has an upper bound in $\mathcal{S}$. Then by Zorn's Lemma $\mathcal{S}$ has a maximal element which is therefore a maximal (proper) ideal containing I. □

Definition 1.51 (Annihilators) *Let a be an element of a ring R. Then we call the set $\{x \in R : ax = 0\}$ to be the* ***right annihilators*** *of a in R, and call the set $\{y \in R\ ya = R\}$ to be the* ***left annihilators*** *of a in R.*
Let L be a subset of R, then we call the set $\{x \in \mathbb{R} \,|\, xa = 0$ for all $a \in L\}$ to be the ***left annihilator*** *of L in R. Similarly, we can define the right annihilator of a set.*

Lemma 1.52 *If $a \in R$, then the left annihilators of L is a left ideal of R and the right annihilators of L is a right ideal.*

If I is a left ideal of R, then the left annihilator of I in R is a two-sided ideal; if I is a right ideal of R, then the right annihilator of I in R is a two-sided ideal.

Definition 1.53 (Nilradical) *Let R be a ring, then the set of nilpotent elements is called the* ***nilradical of*** *R and is denoted by $\mathfrak{N}(R)$.*

Lemma 1.54 *When R is a commutative ring, then $\mathfrak{N}(R)$ is an ideal of R.*

Proof: For the closure of additivity, let $a, b \in \mathfrak{N}(R)$, with $a^m = b^n = 0$, then consider $(a + b)^{m+n}$. □

1.8 Quotient Ring

Recall the definition of the Quotient group / Factor group. We have the similar structure for a ring. Since every ring is abelian group under addition, then every subgroup of the ring is a normal subgroup.

Definition 1.55 (Factor Ring / Quotient Ring) *Let R be a ring, and let A be an ideal of of R, then $R/A = \{r + A : r \in R\}$ is a factor group /quotient group under addition. Define the multiplication on R/A as follows:*

$$\forall s + A, t + A \in R/A,\ (s + A)(t + A) = st + A.$$

*Then the structure $(R/A, +, \cdot)$ is called the **Factor Ring / Quotient Ring**.*

Theorem 1.56 *Let R be a ring and let A be a subring of R. The set of cosets $\{r + A : r \in R\}$ is a ring (known as factor ring), under the operations:*

$$(s + A) + (t + A) = (s + t) + A \text{ and } (s + A)(t + A) = st + A$$

if and only if A is an ideal of R.

Proof: Since every subring A of R is a normal subgroup of R under addition, then addition is always well-defined on R/A. It is also clear that multiplication is closed. The distribution property follows from the distribution of the base ring. Thus we only need to check when is multiplication well-defined. Suppose $s_1 + A = s_2 + A$ and $t_1 + A = t_2 + A$, where $s_1, s_2, t_1, t_2 \in R$, then we must have $s_1 - s_2 \in A$ and $t_1 - t_2 \in A$. Also $(s_1 + A)(t_1 + A) = s_1t_1 + A$, $(s_2 + A)(t_2 + A) = s_2t_2 + A$. Now if A is an ideal then $(s_1 - s_2)t_2 \in A$, and $s_1(t_1 - t_2) \in A$, so

$$s_2t_2 + A = s_2t_2 + (s_1 - s_2)t_2 + s_1(t_1 - t_2)A = s_1t_1 + A.$$

Conversely, if A is not an ideal, then $\exists a \in A$ and $r \in R$, such that ra or ar not in A (we already assumed A is a subring). Suppose $ra \notin A$, then consider $(r + A)(0 + A)$, $(r + A)(a + A)$, the two should be equal, yet the former is equal to $0 + A$ and the later is equal equal to $ra + A$. As $ra \notin A$, then the two does not equal. Similarly, we can show that multiplication is not well-defined for the other case. Hence R/A is not a ring with the prescribed operations if A is not an ideal. □

Example:

- $\mathbb{Z}/n\mathbb{Z}$ is a factor ring of $\mathbb{Z}$. Since $n\mathbb{Z}$ is a principle idea generated by n, i.e., $n\mathbb{Z} = \langle n \rangle$.
- $2\mathbb{Z}/6\mathbb{Z}$ is a factor ring of $2\mathbb{Z}$.
- $R = \mathbb{Z}[i]/\langle 2 - i \rangle$ is a factor ring of $\mathbb{Z}[i]$. The elements of R are $\{0 + \langle 2 - i \rangle, 1 + \langle 2 - i \rangle, 2 + \langle 2 - i \rangle, 3 + \langle 2 - i \rangle, 4 + \langle 2 - i \rangle\}$. Firstly note that $5 \in \langle 2 - i \rangle$. Next, for any $a + bi \in \mathbb{Z}[i]$, we note it is equal to $c + d(2 - i)$ for some $c, d \in \mathbb{Z}$. Hence the elements in R follows consequently.
- Consider $\mathbb{R}[x]$, and $I = \langle x^2 + 1 \rangle$. Then $\mathbb{R}[x]/I = \{g(x) + \langle x^2 + 1 \rangle : g(x) \in \mathbb{R}[x]\}$ is a factor ring of $\mathbb{R}[x]$. Then by division algorithm, let $f(x) \in \mathbb{R}[x]$, we have $f(x) = (x^2 + 1)p(x) + r(x)$, where $r(x)$ is either degree 1 or degree 0. Hence, we conclude $\mathbb{R}[x]/I = \{(ax + b) + \langle x^2 + 1 \rangle : a, b \in \mathbb{R}\}$.

Theorem 1.57 (The first Isomorphism Theorem for Rings) *If $\varphi : R \to S$ is a homomorphism of rings, then the kernel of φ is an ideal of R, the image of φ is a subring of S and $R/\ker\varphi$ is isomorphic as a ring to $\varphi(R)$.*

Proof: By previous analysis, we know that $\ker\varphi$ is an ideal and $\operatorname{Im}\varphi$ is a subring of S. Let $I = \ker\varphi$, we also know R/I is well-defined as is equal to the set $\{r + I\,;\, r \in R\}$. Next, consider the map $g : R/I \to \varphi(R)$, $r + I \mapsto \varphi(r)$. One can easily check that it is a ring isomorphism. Hence the theorem follows. □

Corollary 1.57.1 *If I is an ideal of R, then the map $\psi : R \to R/I$, $r \mapsto r + I$ is a subjective ring homomorphism with kernel I. Thus every ideal is the kernel of a ring homomorphism and vice versa.*

Proof: Clear. □

Theorem 1.58 (The second Isomorphism Theorem for Rings) *Let A be a subring and let B be an ideal of R. Then $A+B = \{a+b : a \in A, b \in B\}$ is a subring of R, $A \cap B$ is an ideal of A and $(A+B)/B \cong A/(A\cap B)$.*

Proof: Let $a_1+b_1, a_2+b_2 \in A+B$, then $(a_1+b_1)-(a_2+b_2) = (a_1-a_2)+(b_1-b_2) \in A+B$ and $(a_1+b_1)(a_2+b_2) = a_1a_2 + (b_1a_2 + a_1b_2 + b_1b_2) \in A + B$. So $A + B$ is a subring of R. Similarly, we can show $A \cap B$ is an ideal of A. Lastly, Consider the map $\varphi : A \mapsto R/B$ by $a \mapsto a + B$. Note φ is the restriction of the quotient map $R \to R/B$ and A is a subring of R, hence φ is a ring homomorphism. The image of φ is the set of all cosets $a + B$ for $a \in A$. Hence $\operatorname{Im}\varphi = (A + B)/B$. Next, suppose $a \in \ker\varphi$, then $a + B = B$, that is $a \in B$. Hence $\ker\varphi = A \cap B$. Then by the First isomorphism theorem, we have

$$(A + B)/B \cong A/(A \cap B).$$

□

Theorem 1.59 (The Third Isomorphism Theorem for Rings) *Let I and J be ideals of R with $I \subset J$. Then J/I is an ideal of R/I and $(R/I)/(J/I) \cong R/J$.*

Proof: Define $\varphi : R/I \to R/J$ by $r + I \mapsto r + J$. This map is clearly additive (by group theory) and it is multiplicative since

$$(r_1 + I)(r_2 + I) = r_1r_2 + I = (r_1r_2 + I).$$

As $I \subset J$, then the image is onto. And by group theory result, we know $\ker\varphi = J/I$. Hence we conclude by First Isomorphism Theorem that $(R/I)/(J/I) \cong R/J$. □

Theorem 1.60 (The Fourth Isomorphism Theorem for Rings) *Let I be an ideal of R. Then correspondence $A \leftrightarrow A/I$ is an inclusion preserving bijection between the set of subrings A of R that contain I and the set of subrings of R/I. Furthermore, A (a subring containing I) is an ideal of R if and only if A/I is an ideal of R/I.*

Proof: Let $\varphi : A \to A/I$ be the quotient map. Let $X, Y \subset A/I$. Then it is clear that $X \subset Y \Leftrightarrow \phi^{-1}(X) \subset \phi^{-1}(Y)$ Then if $X \subset Y$ are subrings of A/I, then $\phi^{-1}(X) \subset \phi^{-1}(Y)$, and they are also subrings of A. So the map between the set of subrings A of R that contain I and the set of subrings of R/I is also bijective.
Let $I \subset A$ be any ideal of R. Since φ is surjective, then A (a subring containing I) is an ideal of R if and only if A/I is an ideal of R/I. □

Proposition 1.61 *Assume R is commutative ring with $1 \neq 0$. The ideal M is a maximal ideal if and only if the quotient ring R/M is a field.*

Proof: This follows from the Fourth Isomorphism Theorem together with Proposition 1.48 (2). the ideal M is maximal if and only if there are no ideals I with $M \subset I \subset R$. By the Fourth Isomorphism Theorem, the ideals of R containing M correspond bijectively with the ideals of R/M, so M is maximal if and only if the only ideals of R/M are 0 and R/M. And by Proposition 1.48 (2), we see that M is maximal if and only if R/M is a field. □

Remark 1.61.1 *If R is not commutative, then only one implication holds, that is I is a maximal ideal if R/I is a field.*

Example:

1. Let n be a nonnegative integer. The ideal $n\mathbb{Z}$ of $\mathbb{Z}$ is a maximal ideal if and only if $\mathbb{Z}/n\mathbb{Z}$ is a field. Hence if and only if n is a prime number.
2. The ideal $(2, x)$ is a maximal ideal in $\mathbb{Z}[x]$ because its quotient ring is the field $\mathbb{Z}/2\mathbb{Z}$.
3. Let R be the ring of all functions from $[0, 1]$ to $\mathbb{R}$ and for each $a \in [0, 1]$ let M_a be the kernel of evaluation at a. Since evaluation is a subjective homomorphism from R to $\mathbb{R}$, we see that $R/M_a \cong \mathbb{R}$ and hence M_a is a maximal ideal. Similarly, the kernel of evaluation at any fixed point is a maximal ideal in the ring of continuous real valued function on $[0, 1]$.
4. If F is a field and G is a finite group, then the augmentation ideal I is a maximal ideal of the group ring FG. The augmentation ideal is the kernel of the augmentation map which is a surjective homomorphism onto the field F, so $FG/I \cong F$, and F is a field.

Definition 1.62 (Prime Ideal) *Assume R is commutative. An ideal P is called a **prime ideal** if $P \neq R$ and whenever the product ab of two elements $a, b \in R$ is an element of P, then at least one of a and b is an element of P.*

Proposition 1.63 *Assume R is commutative with $1 \neq 0$. Then the ideal P is a prime ideal in R if and only if the quotient ring R/P is an integral domain. In particular, a commutative ring with identity is an integral domain if and only if 0 is a prime ideal.*

Proof: The ideal P is prime if and only if $P \neq R$ and whenever $ab \in P$, then either $a \in P$ or $b \in P$. Let $\bar{r}$ denote $r + P \in R/P$. Note that $r \in P$ if and only if the element $\bar{r}$ is zero in the quotient ring R/P. Thus in the terminology of quotients P is a prime ideal if and only if $\bar{r} \neq \bar{0}$ (so $\bar{1} \notin P$, otherwise $R \subset P$, so $\bar{1} \in R/P$) and whenever $\bar{a}\bar{b} = \bar{a}\bar{b} = \bar{0}$, then either $\bar{a} = \bar{0}$ or $\bar{b} = \bar{0}$, i.e., R/P is an integral domain. □

Corollary 1.63.1 *Assume R is commutative. Every maximal ideal of R is a prime ideal.*

Proof: If M is a maximal ideal, then R/M is a field and hence an integral domain. □

Proposition 1.64 *Let R be a commutative ring. Then the nilradical $\mathfrak{N}(R)$ of R equals the intersection of all prime ideals of R.*

Proof: Firstly, the nilradical is contained in every prime ideal. Indeed, if $r \in \mathcal{N}_R$, one has $r^n = 0$ for some positive integer n. Since every ideal contains 0 and every prime ideal that contains a product, here $r^n = 0$, contains one of its factors, one deduces that every prime ideal contains r.
Conversely, let $f \notin \mathcal{N}_R$; we have to prove that there is a prime ideal that does not contain f. Consider the set Σ of all ideals that do not contain any power of f. One has $(0) \in \Sigma$, by definition of the nilradical. For every chain $J_1 \subseteq J_2 \subseteq \dots$ of ideals in Σ, the union $J = \bigcup_{i \geq 1} J_i$ is an ideal that belongs to Σ, since otherwise it would contain a power of f, that must belong to some J_i, contradicting the definition of J_i.
So, Σ is a partially ordered set by inclusion such that every chain has a least upper bound. Thus, Zorn's lemma applies, and there exists a maximal element $\mathfrak{m} \in \Sigma$. We have to prove that $\mathfrak{m}$ is a prime ideal. If it were not prime there would be two elements $g \in R$ and $h \in R$ such that $g \notin \mathfrak{m}$, $h \notin \mathfrak{m}$, and $gh \in \mathfrak{m}$. By maximality of $\mathfrak{m}$, one has $\mathfrak{m} + (g) \notin \Sigma$ and $\mathfrak{m} + (h) \notin \Sigma$. So there exist positive integers r and s such that $f^r \in \mathfrak{m} + (g)$ and $f^s \in \mathfrak{m} + (h)$. It follows that $f^r f^s = f^{r+s} \in \mathfrak{m} + (gh) = \mathfrak{m}$, contradicting the fact that $\mathfrak{m}$ is in Σ. This finishes the proof, since we have proved the existence of a prime ideal that does not contain f. □

Proposition 1.65 *Let $\varphi : R \to S$ be a homomorphism of commutative rings.*

1. *If P is a prime ideal of S then either $\varphi^{-1}(P) = R$ or $\varphi^{-1}(P)$ is a prime ideal of R. In particular, if R is a subring of S, and φ is the inclusion homomorphism, then $P \cap R$ is either R or a prime ideal of R.*
2. *If M is a maximal ideal of S and φ is surjective, then $\varphi^{-1}(M)$ is a maximal ideal of R.*

Proof: Suppose P is a prime ideal and J is any ideal that strictly contains P. Then $\exists j \in J$ but $j \notin P$. Then for all $r \notin P$, $rj(r + j) = 0$ as it is a Boolean ring. So either $rj \in P$ or $r + j \in P$, but if $rj \in P$, then either $r \in P$ or $j \in P$ which is a contradiction. So it must be the case that $r + j \in P$, then $r + j \in J$, which shows $r \in J$. Hence $J = R$, so P must be maximal. □

Lemma 1.66 *In a Boolean ring, every finitely generated ideals is principal.*

Proof: We proceed by induction, it suffices to show that $(x, y) = (z)$ for some z. Note this works if $z = x + y + xy$. Then $x = xz$ and $y = yz$. □

1.9 Ring of Fractions

Theorem 1.67 *Let R be a commutative ring. Let D be any nonempty subset of R that does not contain 0, does not contain any zero divisors and is closed under multiplication, i.e., $ab \in D$ for all $a, b \in D$. Then there is a commutative ring Q with 1 such that Q contains R as a subring and every element of D is a unit in Q. The ring Q has the following additional properties:*

1. *every element of Q is of the form rd^{-1} for some $r \in R$ and $d \in D$. In particular, if $D = R - \{0\}$, then Q is a field.*

2. *(uniqueness of Q) the ring Q is the "smallest" ring containing R in which all elements of D becomes units, in the following sense. Let S be any commutative ring with identity and let $\varphi : R \to S$ be any injective ring homomorphism such that $\varphi(d)$ is a unit in S for every $d \in D$. Then there is an injective homomorphism $\Phi : Q \to S$ such that $\Phi|_R = \varphi$. In other words, any ring containing an isomorphic copy of R in which all the elements of D becomes units must also contain an isomorphic copy of Q.*

Proof: Let $\mathcal{F} = \{(r, d) \,|\, r \in R,\, d \in D\}$ and define the relation $\sim$ on $\mathcal{F}$ by $(r, d) \sim (s, e)$ if and only if $re = sd$. One can check that this is an equivalence relation, the only difficulty might be transitivity: suppose $(r, d) \sim (s, e)$ and $(s, e) \sim (t, f)$. Then $re - sd = 0$ and $sf - te = 0$. Multiplying the first of these equation by f and the second by d and adding them gives $(rf - td)e = 0$. Since $e \in D$ is neither zero nor a zero divisor, we must have $rf - td = 0$, so $(r, d) \sim (t, f)$. Next, denote the equivalence class of (r, d) by $\frac{r}{d}$:

$$\frac{r}{d} := \{(a, b) \,|\, a \in R,\, b \in D \text{ and } rb = ad\}.$$

Let Q be the set of equivalence classes under $\sim$. Note that $\frac{r}{d} = \frac{re}{de}$ in Q for all $e \in D$, since D is closed under multiplication.

We now define an additive and multiplicative structure on Q:

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

To show Q is a commutative ring with identity, one need to check:

1. the addition and multiplication are well defined;

2. Q is an abelian group under addition, where the additive identity is $\frac{0}{d}$ for any $d \in D$ and the additive inverse of $\frac{a}{d}$ is $\frac{-a}{d}$;

3. multiplication is associative, distributive and commutative;

4. Q has an identity, $\frac{d}{d}$ for any $d \in D$.

These are all routine checking which will be omitted here.

Next we embed R into Q by defining:

$$\iota : R \to Q \text{ by } \iota : r \mapsto \frac{rd}{d} \text{ where d is any element of D.}$$

One can check that ι is a ring homomorphism and it is injective, because

$$\iota(r) = 0 \Leftrightarrow \frac{rd}{d} = \frac{0}{d} \Leftrightarrow rd^2 = 0 \Leftrightarrow r = 0.$$

The subring $\iota(R)$ of Q is therefore isomorphic to R, so we can identity each $r \in R$ with $\iota(r)$ and we can consider R as a subring of Q.

Next note that each $d \in D$ has a multiplicative inverse in Q: namely, if d is represented by the fraction $\frac{de}{e}$, then its multiplicative inverse if $\frac{e}{de}$. One then sees that every element of Q may be written as $r \cdot d^{-1}$ for some $r \in R$ and some $d \in D$ ($\frac{r}{d} = rd^{-1}$). In particular, if $D = R - \{0\}$, every nonzero element of Q has a multiplicative inverse and Q is a field.

It remains to establish the uniqueness property of Q. Assume $\varphi : R \to S$ is an injective ring homomorphism such that $\varphi(d)$ is a unit in S for all $d \in D$. Extend φ to a map $\Phi : Q \to S$ by defining $\Phi(rd^{-1}) = \varphi(r)\varphi(d)^{-1}$ for all $r \in R$, $d \in D$. This map is well-defined, since $rd^{-1} = se^{-1}$ implies $re = sd$, so $\varphi(r)\varphi(e) = \varphi(s)\varphi(d)$, hence

$$\Phi(rd^{-1}) = \varphi(r)\varphi(d)^{-1} = \varphi(s)\varphi(e)^{-1} = \Phi(se^{-1}).$$

It is also straight forward to check that Φ is a ring homomorphism. Finally, Φ is injective because $rd^{-1} \in \ker \Phi$ implies $r \in \ker \Phi \cap R = \ker \varphi$ ($\Phi(d^{-1})$ is not a zero divisor); since φ is injective, this forces r and hence also rd^{-1} to be zero. □

Definition 1.68 (Ring / Field of Fractions) *Let R, D and Q be as in the Theorem 1.67.*

1. *The ring Q is called the **ring of fractions** of D with respect to R and is denoted $D^{-1}R$.*
2. *If R is an integral domain and $D = R - \{0\}$, then Q is called the **field of fractions** or **quotient field of** R.*

Corollary 1.68.1 *Let R be an integral domain and let Q be the field of fractions of R. If a field F contains a subring R' isomorphic to R, then the subfield of F generated by R' is isomorphic to Q.*

Proof: Let $\varphi : R \cong R' \subseteq F$ be a ring isomorphism of R to R'. In particular, $\varphi : R \to F$ is an injective homomorphism from R into the field F. Let $\Phi : Q \to F$ be the extension of φ to Q as in the Theorem 1.67. Then Φ is injective, so $\Phi(Q)$ is an isomorphic copy of Q in F containing $\varphi(R) = R'$. Now, any subfield of F containing $R' = \varphi(R)$ contains the elements $\varphi(r_1)\varphi(r_2)^{-1} = \varphi(r_1 r_2^{-1})$ for all $r_1, r_2 \in R$. Since every element of Q is of the form $r_1 r^{-1}$ for some $r_1, r_2 \in R$, it follows that any subfield of F containing R' contains the field $\Phi(Q)$, so that $\phi(Q)$ is the subfield of F generated by R'. □

1.10 The Chinese Remainder Theorem

Lemma 1.69 *Let $R_1, \cdots, R_n$ be rings with identities. Then every ideal of $R_1 \times \cdots \times R_n$ is of the form $I_1 \times \cdots \times I_n$ where I_i is an ideal of R_i.*

Proof: Suppose I_i is an ideal of R_n, then clearly $I_1 \times \cdots \times I_n$ is an ideal of the direct product. Conversely, suppose I is an ideal, then $I = I_1 \times \cdots \times I_n$ where I_i is a subring of R_i. As $R_1, \cdots, R_n$ have identities, then multiplication by $(1, \cdots, 1, a, \cdots, 1)$, we can show that I_i is an ideal of R_i. □

Corollary 1.69.1 *If R and S are nonzero rings, then $R \times S$ is never a field.*

Proof: Since it contains ideals that is not 0 nor $R \times S$. □

Definition 1.70 (Comaximal) *Two ideals A and B of a ring R are said to be **comaximal** if $A + B = R$.*

Lemma 1.71 *The product of the ideals $A_1, A_2, \cdots, A_k$ is the ideal of all finite sums of elements of the form $x_1 x_2 \cdots x_k$ such that $x_i \in A_i$ for all i. It is also an ideal that is contained in $\bigcap_{n=1}^{k} A_k$. In particular, if R is commutative, $A_i = (a_i)$, then $A_1 \cdots A_k = (a_1 \cdots a_k)$.*

Proof: Proof by induction. Consider the case of two, then use distributive law for the inductive step. □

Theorem 1.72 (Chinese Remainder Theorem) *Let $A_1, A_2, \cdots, A_k$ be ideals in R which is a commutative ring with identity $1 \neq 0$. The map*

$$R \to R/A_1 \times R/A_2 \times \cdots \times R/A_k \quad \text{defined by} \quad r \mapsto (r + A_1, r + A_2, \cdots, r + A_k)$$

is a ring homomorphism with kernel $A_1 \cap A_2 \cap \cdots \cap A_k$. If for each $i, j \in \{1, 2 \cdots, k\}$ with $i \neq j$, the ideals A_i and A_j are comaximal, then this map is surjective and $A_1 \cap A_2 \cap \cdots \cap A_k = A_1 A_2 \cdots A_k$ so

$$R/(A_1 A_2 \cdots A_k) = R/(A_1 \cap A_2 \cap \cdots \cap A_k) \cong R/A_1 \times R/A_2 \times \cdots \times R/A_k.$$

Proof: We first prove this for $k = 2$; the general case will follow by induction. Let $A = A_1$ and $B = A_2$. Consider the map $\varphi : R \to R/A \times R/B$ defined by $\varphi(r) = (r + A, r + B) = (r \mod A, r \mod B)$. This map is a ring homomorphism because φ is just the natural projection of R into R/A and R/B for the two components. The kernel of φ consists of all the elements $r \in R$ that are in A and in B, i.e., $A \cap B$. To complete the proof in this case, it remains to show that when A and B are comaximal, φ is surjective and $A \cap B = AB$. Since $A + B = R$, there are elements $x \in A$ and $y \in B$ such that $x + y = 1$. This equation shows that $\varphi(x) = (0, 1)$ and $\varphi(y) = (1, 0)$, as $x = 1 - y \in 1 + B$ and $y = 1 - x \in 1 + A$. If now $(r_1 \mod A, r_2 \mod B)$ is an arbitrary element in $R/A \times R/B$, then the element $r_2 x + r_1 y$ mapsto this element. This shows that φ is indeed surjective. Finally, the ideal AB is always contained in $A \cap B$. If A and B are comaximal, and x and y are as above, then for any $c \in A \cap B$, $c = c \cdot 1 = cx + cy \in AB$. This establishes the reverse inclusion $A \cap B \subset AB$ and completes the proof when $k = 2$.

The general case follows from induction by considering two ideals $A = A_1$ and $B = A_2 \cdots A_k$. However, we need to show that A_1 and $A_2 \cdots A_k$ are comaximal if we want to use the base case. By hypothesis, for each $i \in \{2, 3, \cdots, k\}$ there are elements $x_i \in A_1$ and $y_i \in A_i$ such that $x_i + y_i = 1$. Since $x_i + y_i \equiv y_i \mod A$, it follows that $1 = (x_2 + y_2) \cdots (x_k + y_k)$ is an elements in $A_1 + (A_2 \cdots A_k)$, so must be every $r \in R$. This completes the proof. □

Remark 1.72.1 *Since the isomorphism in the Chinese Remainder Theorem is an isomorphism of rings, the groups of units on both sides must be isomorphic thus. It is easy to see that the units in any direct product of rings are the elements that have units in each of the coordinates.*

Corollary 1.72.1 *Let n be a positive integer and let $p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ be prime factorization. Then*

$$\mathbb{Z}/n\mathbb{Z} \cong (\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z}) \times (\mathbb{Z}/p_2^{\alpha_2}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_k^{\alpha_k}\mathbb{Z}),$$

as rings, so in particular, we have the following isomorphism of multiplicative groups:

$$(\mathbb{Z}/n\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z})^\times \times (\mathbb{Z}/p_2^{\alpha_2}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/p_k^{\alpha_k}\mathbb{Z})^\times.$$

Corollary 1.72.2 *Let $n_1, \cdots, n_k$ be integers which are relatively prime in pairs, i.e., $(n_i, n_j) = 1$ for all $i \neq j$. Then for any $a_1, \cdots, a_k \in \mathbb{Z}$, there is a solution $x \in \mathbb{Z}$ to the simultaneous congruences*

$$\begin{aligned} x &\equiv a_1 \mod n_1, \\ x &\equiv a_2 \mod n_2, \\ &\vdots \\ x &\equiv a_k \mod n_k, \end{aligned}$$

and the solution x is unique mod $n = n_1 n_2 \cdots n_k$. In particular, let $n_i' = n/n_i$ be the quotient of n by n_i, and let t_i be the inverse of n_i' mod n_i. Then x is given by

$$x = a_1 t_1 n_1' + a_2 t_2 n_2' + \cdots + a_k t_k n_k' \mod n.$$

Corollary 1.72.3 *Let $f_1(x), f_2(x), \cdots, f_k(x)$ be polynomials with integer coefficients of the same degree d. Let $n_1, n_2, \cdots, n_k$ be integers which are relatively prime in pairs. Then there exists a polynomial $f(x)$ with integer coefficients and of degree d such that*

$$\begin{aligned} f(x) &\equiv f_1(x) \mod n_1, \\ f(x) &\equiv f_2(x) \mod n_2, \\ &\vdots \\ f(x) &\equiv f_k(x) \mod n_k. \end{aligned}$$

Further, if all the $f_i(x)$ are monic, then $f(x)$ can be chosen to be monic as well.

2 Special Rings

2.1 Quadratic Field

Definition 2.1 (Quadratic Field) *Let D be a rational number that is not a perfect square in $\mathbb{Q}$. Then the set*

$$\mathbb{Q}(\sqrt{D}) = \{a + b\sqrt{D} \,:\, a, b \in \mathbb{Q}\}$$

*as a subset of $\mathbb{C}$ forms a field, which is called the **quadratic field**. Now let D be a square free integer. Define*

$$\mathcal{O} = \mathcal{O}_{\mathbb{Q}(\sqrt{D})} = \mathbb{Z}[\omega] = \{a + b\omega \,:\, a, b \in \mathbb{Z}\},$$

where

$$\omega = \begin{cases} \sqrt{D} & \text{if } D = 2, 3 \mod 4 \\ \frac{1+\sqrt{D}}{2}, & \text{if } D = 1 \mod 4 \end{cases}$$

*called the **ring of integers** in the quadratic field $\mathbb{Q}(\sqrt{D})$.*

Remark 2.1.1 *The rational number D may be written $D = f^2 D'$ for some rational number f and a unique integer D' where D' is not divisible by the square of any integer greater than 1. If we call D' the **squarefree** part of D. Then $\sqrt{D} = f'\sqrt{D'}$, so $\mathbb{Q}(\sqrt{D}) = \mathbb{Q}(\sqrt{D'})$. Thus there is no loss in assuming that D is a squarefree integer in the definition of the quadratic field $\mathbb{Q}(\sqrt{D})$.*

Remark 2.1.2 *Note $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is indeed a subring (hence a ring) of $\mathbb{Q}(\sqrt{D})$ when D is a square free integer.*

Definition 2.2 (Ring of Gaussian Integers) *The **ring of Gaussian Integers** $\mathbb{Z}[i]$ is the set*

$$\mathbb{Z}[i] = \{a + ib \,:\, a, b \in \mathbb{Z}\}$$

with the usual addition and multiplication on complex numbers.

Remark 2.2.1 *One can verify that $\mathbb{Z}[i]$ is a commutative ring, with 1 being the unity. The units of $\mathbb{Z}[i]$ are precisely ± 1 and $\pm i$. This is really just a special case of the $\mathbb{Z}[\sqrt{D}]$ where $D = -1$.*

Definition 2.3 (Field Norm) *We define the **field norm** $\mathcal{N} : Q(\sqrt{D}) \to \mathbb{Q}$ by*

$$\mathcal{N}(a + b\sqrt{D}) = (a + b\sqrt{D})(a - b\sqrt{D}) = a^2 - Db^2 \in \mathbb{Q}.$$

Proposition 2.4 (Properties of Field Norm)

1. *$\mathcal{N}(a + b\sqrt{D}) \neq 0$ if $a + b\sqrt{D} \neq 0$.*
2. *$\mathcal{N}$ is multiplicative, i.e. $\mathcal{N}(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in \mathbb{Q}(\sqrt{D})$.*
3. *On the subring $\mathcal{O}$,*

$$N(a + b\omega) = (a + b\omega)(a + b\bar{\omega}) = \begin{cases} a^2 - Db^2, & \text{if } D = 2, 3 \mod 4 \\ a^2 + ab + \frac{1-D}{4}b^2, & \text{if } D = 1 \mod 4 \end{cases}$$

where

$$\bar{\omega} = \begin{cases} -\sqrt{D}, & \text{if } D = 2, 3 \mod 4 \\ \frac{1-\sqrt{D}}{2}, & \text{if } D = 1 \mod 4. \end{cases}.$$

In particular, $N(\alpha)$ is an integer for every $\alpha \in \mathcal{O}$.

4. *If $\alpha \in \mathcal{O}$ has field norm $N(\alpha) = \pm 1$, then α is invertible hence a unit in $\mathcal{O}$. In fact, if $\alpha = a + b\omega$, then its inverse is given by $\pm(\alpha + b\bar{\omega})$ depending on the sign of $N(\alpha)$.*

5. *An element of $\mathcal{O}$ is a unit if and only if it has field norm ± 1.*

Proof: Direct verification. □

Remark 2.4.1 *When D is negative by direct verification we have,*

- *if $D = -1$, then the units of $\mathbb{Z}[i]$ are $\{\pm 1, \pm i\}$;*
- *if $D = -3$, then the units of $\mathbb{Z}[(1 + \sqrt{-3})/2]$ are $\{\pm 1, \pm\rho, \pm\rho^2\}$, where $\rho = (-1 + \sqrt{-3})/2$;*
- *for any other $D < 0$, the only units of $\mathcal{O}$ are $\{\pm 1\}$.*

When $D > 0$, it can be shown that the group of units $\mathcal{O}^\times$ is always infinite. This is a corollary of the characterization of solutions to Pell's equation.

Theorem 2.5 *For $D < 0$, the ring of integers of the quadratic field $\mathbb{Q}(\sqrt{D})$ is a Principal Ideal Domain if and only if $D = -1, -2, -3, -7, -11, -19, -43, -67$ or -163. They are Euclidean precisely when $D = -1, -2, -3 - 7$ or -11.*

2.2 Matrix Rings

Definition 2.6 (Matrix Ring) *Fix an arbitrary ring R, and let n be a positive integer. Let $M_n(R)$ be the set of all $n \times n$ matrices with entries from R. Then the set of matrices becomes a ring under the usual rules of matrix addition and multiplication, which is called the* ***matrix ring****.*

Remark 2.6.1 *If R is any nontrivial ring and $n \geq 2$, then $M_n(R)$ is not commutative. If $ab \neq 0$ in R, let A be the matrix with a in the $1, 1$ entry and zeros elsewhere, and let B be the matrix with b in positive $1, 2$ and zeros elsewhere; then ab is the entry in position $1, 2$ of AB where BA is the zero matrix, so $AB \neq BA$. In addition, we conclude that $M_n(R)$ has zero divisors for all nonzero rings R whenever $n \geq 2$.*

Definition 2.7 *An element (a_{ij}) of $M_n(R)$ is called a* ***scalar matrix*** *if for some $a \in R$, $a_{ii} = a$ for all $i \in \{1, \cdots, n\}$ and $a_{ij} = 0$ for all $i \neq j$.*
If R has a multiplicative identity 1, then the scalar matrix with 1's down the diagonal is the ***identity matrix*** *which is also the multiplicative identity of $M_n(R)$. In this case, the units of the ring $M_n(R)$ are the invertible $n \times n$ matrices and the group of units is denoted $\mathrm{GL}_n(R)$ which is known as the* ***general linear group of degree*** *n* ***over*** *R.*
The set of ***upper triangular matrices*** *of $M_n(R)$ is the set $\{(a_{ij}) : a_{pq} = 0 \text{ whenever } p > q\}$.*

Remark 2.7.1 *The set of scalar matrix is a subring of $M_n(R)$. If R is commutative, then the subring of scalar matrix is also commutative; it fact it commutes with all elements of $M_n(R)$. It is also easy to check that the set of upper triangular matrix is a subring of $M_n(R)$.*

Lemma 2.8 *If S is a subring of R, then $M_n(S)$ is a subring of $M_n(R)$.*

2.3 Group Rings

Definition 2.9 *Fix a commutative ring R with identity $1 \neq 0$ and let $G = \{g_1, \cdots, g_n\}$ be any finite group with group operations written multiplicative. Define the **group ring**, RG, of G with coefficients in R to be the set of all formal sums*

$$a_1g_1 + a_2g_2 + \cdots + a_ng_n, \quad a_i \in R, \ 1 \leq i \leq n.$$

Addition is defined "component wise":

$$(a_1g_1 + \cdots + a_ng_n) + (b_1g_1 + \cdots + b_ng_n) = (a_1 + b_1)g_1 + \cdots + (a_n + b_n)g_n$$

Multiplication is performed by first defining $(ag_i)(bg_j) = (ab)g_k$, where $g_ig_j = g_k \in G$. The product is then extended to all formal sums by the distributive laws.

Remark 2.9.1 *It is straightforward to check that RG with the prescribed definition of addition and multiplication is a ring. RG is commutative if and only if G is a commutative group.*

Remark 2.9.2 *If g_1 is the identity of G, we write a_1g_1 simply as a_1. Similarly, we shall write element $1g$ for $g \in G$ simply as g. Hence the ring R appears in RG as the "constant" formal sums and commutes with all elements of RG. The identity of R is the identity of RG. The group G also appears in RG as $g_i = 1g_i$ for $g_i \in G$. Each element of G has a multiplicative inverse in the ring RG (namely, its inverse in G). So G is a subgroup of the group of units of RG.*

Lemma 2.10 *If $|G| > 1$, then RG is not an integral domain and always have zero divisors.*

Proof: We show RG always have zero divisors if $|G| > 1$. Let g be any element of G of order $m > 1$ (G is finite). Then

$$(1 - g)(1 + g + \cdots + g^{m-1}) = 1 - g^m = 1 - 1 = 0.$$

Hence $1 - g$ is a zero divisor. □

Lemma 2.11 *If S is a subring of R then SG is a subring of RG. If H is a subgroup of G, then RH is a subring of RG.*

Example:

1. $\mathbb{Z}G$ (called the **integral group ring of** G) is a subring of $\mathbb{Q}G$ (the **rational group ring of** G).
2. $\mathbb{R}Q_8$ and $\mathbb{H}$ is not the same ring, as $\mathbb{R}Q_8$ has zero divisors (by Lemma 2.10) and $\mathbb{H}$ has no zero divisors (recall $\mathbb{H}$ is a division ring).

Proposition 2.12 *Let $G = \{g_1, g_2, \cdots, g_n\}$ be a finite group, then the element $N = g_1 + g_2 + \cdots + g_n$ is in the center of RG. More generally, if $\mathcal{K} = \{k_1, \cdots, k_m\}$ is a conjugacy class of G, then $K = k_1 + \cdots + k_m$ is in the center of RG. Moreover, let $\{\mathcal{K}_1, \cdots, \mathcal{K}_r\}$ be the set of all conjugacy classes of G, and for each $\mathcal{K}_i$, let K_i be the formal sum of the members of $\mathcal{K}_i$. Then α in RG is in the center of RG if and only if*

$$\alpha = a_1 K_1 + \cdots + a_r K_r$$

for some $a_1, \cdots, a_r \in R$.

Proof: By distributivity, suffices to show that ag_i commutes with N for any $i \in \{1, \cdots, n\}$. However, this follows from the fact at the collection $\{g_i g_1, \cdots, g_i g_n\}$ is the same as the collection $\{g_1 g_i, \cdots, g_n g_i\}$. The next two statement follows similarly. □

Lemma 2.13 *The augmentation ideal in the group ring RG is generated by $\{g - 1 \,|\, g \in G\}$.*

Proof: Firstly, since for any $g \in G$, $g - 1$ is in the augmentation ideal, then we have one direction. On the other hand, if

$$\sum_{i=1}^{n} a_i g_i$$

is in the augmentation ideal, then it can be written as the sum

$$\sum_{g_i \neq 1} a_i (g_i - 1)$$

which gives the other inclusion. □

3 Euclidean Domains, P.I.D., Unique Factorization Domain

In this section, all rings are commutative.

3.1 Euclidean Domains

Definition 3.1 (Norm) *Any function $N : R \to \mathbb{Z}_{\geq 0}$ with $N(0) = 0$ is called a **norm** on the integral domain R. If $N(a) > 0$ for $a \neq 0$, then N is called a **positive norm**.*

Definition 3.2 (Euclidean Domain) *The 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 and b of R with $b \neq 0$, there exists elements q and r in R with*

$$a = qb + r, \quad r = 0 \text{ or } N(r) < N(b).$$

*The element q is called the **quotient** and the element r is called the **remainder** of the division.*

Example:

1. Fields are always Euclidean Domains, because any norm on the field satisfy the defining condition. This is because for every a, b with $b \neq 0$, we have $a = qb + 0$, where $q = ab^{-1}$.
2. The integers $\mathbb{Z}$ are a Euclidean Domain with norm given by $N(a) = |a|$.
3. If F is a field, then the polynomial ring $F[x]$ is a Euclidean Domain with norm given by $N(p(x)) = \deg p(x)$. We will see that $R[x]$ is not a Euclidean Domain if R is not a field.
4. The quadratic integer ring $\mathcal{O}$ is an integral domain with norm defined to be the absolute value of the field norm. But in general, $\mathcal{O}$ is not Euclidean with respect to this norm (or any other norm). However, the Gaussian integers $\mathbb{Z}[i]$ is a Euclidean Domain with respect to the norm $N(a + bi) = a^2 + b^2$.

Proposition 3.3 *Every ideal in a Euclidean Domain is principal. More precisely, if I is any nonzero ideal in the Euclidean Domain R, then $I = (d)$, where d is any nonzero element of I of minimum norm.*

Proof: If I is the zero ideal, there is nothing to prove. Otherwise let d be any nonzero element of I of minimum norm (Exists by WOP). Clearly $(d) \subset I$ since $d \in I$. To show the reverse inclusion, let $a \in I$ be arbitrary, then $a = qd + r$ with $r = 0$ or $N(r) < N(d)$. Then $r = a - qd \in I$, then by the minimality of the norm of d, we must have $r = 0$. Hence $a = qd$, and $a \in (d)$. □

Corollary 3.3.1 *Every ideal of $\mathbb{Z}$ is principle.*

Example:

1. Let $R = \mathbb{Z}[x]$. Since the ideal $(2, x)$ is not principle, then $\mathbb{Z}[x]$ is not a Euclidean Domain.
2. Let R be the quadratic integer ring $\mathbb{Z}[\sqrt{-5}]$, let N be the associated field norm $N(a + b\sqrt{-5}) = a^2 + 5b^2$ and consider the ideal $I = (3, 2 + \sqrt{-5})$ generated by 3 and $2 + \sqrt{-5}$. Suppose $I = (a + b\sqrt{-5})$, $a, b \in \mathbb{Z}$, that is I is principle, then $3 = \alpha(a + b\sqrt{-5})$ and $2 + \sqrt{-5} = \beta(a + b\sqrt{-5})$ for some $\alpha, \beta \in R$. Taking norm in the first

equation gives $9 = N(\alpha)(a^2 + 5b^2)$ and since $a^2 + 5b^2$ is a positive integer, it must be $1, 3$ or 9. If the value is 9, then $N(\alpha) = 1$ and $\alpha = \pm 1$, so $a + b\sqrt{-5} = \pm 3$, which is impossible because the coefficients of $2 + \sqrt{-5}$ are not divisible by 3. The value cannot be 3 since there are no integer solutions to $a^2 + 5b^2 = 3$. If the value is 1, then $a + b\sqrt{-5} = \pm 1$ and the ideal I would be the entire ring R, but clearly $1 \notin (3, 2 + \sqrt{-5})$. Hence I is not a principal ideal and so R is not a Euclidean Domain (with respect to any norm).

Definition 3.4 *Let R be a commutative ring and let $a, b \in R$ with $b \neq 0$.*

1. *a is said to be a **multiple** of b if there exists an element $x \in R$ with $a = bx$. In this case b is said to **divide** a or be a **divisor** of a, written $b|a$.*

2. *A **greatest common divisor** of a and b is a nonzero element d such that*

   - *$d|a$ and $d|b$, and*
   - *if $d'|a$ and $d'|b$, then $d'|d$.*

   A greatest common divisor of a and b will be denoted by $\gcd(a, b)$ or simply (a, b).

Remark 3.4.1 *$b|a$ in a ring if and only if $a \in (b)$ if and only if $(a) \subset (b)$. In particular, if d is any divisor of both a and b, then (d) must contain both a and b, hence (a, b). Then the definition of gcd of a and b is equivalent to: if I is the ideal of R generated by a and b, then d is a greatest common divisor of a and b if*

- *I is contained in the principal ideal (d), and*
- *if (d') is any principal ideal containing I, then $(d) \subset (d')$.*

The following proposition is immediate from the remark:

Definition 3.5 (Sufficient Condition for GCD) *If a and b are nonzero elements in the commutative ring R such that the ideal generated by a and b is a principal ideal (d), then d is a greatest common divisor of a and b.*

Remark 3.5.1 *This explains why the symbol (a, b) is often used to denote both the ideal generated by a and b and a greatest common divisor of a and b.*

Remark 3.5.2 *This is not a necessary condition, for example in $\mathbb{Z}[x]$, the gcd of 2 and x is 1, yet $(2, x) \neq (1)$.*

Definition 3.6 (Bezout Domain) *An integral domain in which every ideal (a, b) generated by two elements is principal is called a **Bezout Domain**.*

Remark 3.6.1 *It is possible for Bezout Domain to have nonprincipal ideals (necessarily infinitely generated).*

Lemma 3.7 *An integral domain R is a Bezout Domain if and only if every pair of elements a, b of R has a g.c.d. d in R that can be written as an R-linear combination of a and b, i.e., $d = ax + by$ for some $x, y \in R$.*

Proof: Clear since $(a, b) = (a) + (b)$. □

Proposition 3.8 (Uniqueness of GCD in Integral Domains) *Let R be an integral domain. If two elements d and d' of R generate the same principal idea, i.e., $(d) = (d')$, then $d' = ud$ for some unit u in R. In particular, if d and d' are both greatest common divisor of a and b, then $d' = ud$ for some unit u.*

Proof: The assertion is clear if either d or d' is zero, so assume d, d' are nonzero. Since $d \in (d')$, then there is some $x \in R$ such that $d = xd'$; since $d' \in (d)$, then there is some $y \in R$ such that $d' = yd$. So $d = xyd$ and so $d(1 - xy) = 0$, and since $d \neq 0$, $xy = 1$, which imply x and y are units (integral domains are commutative). This proves the first assertion. The second assertion is a direct consequence of the first assertion. □

Theorem 3.9 (Existence of GCD in Euclidean Domains) *Let R be a Euclidean Domain and let a and b be nonzero element of R. Let $d = r_n$ be the last nonzero remainder in the Euclidean Algorithm for a and b. Then*

1. *d is the greatest common divisor of a and b, and*
2. *the principal ideal (d) is the ideal generated by a and b. In particular, there are elements x and y in R such that*
$$d = ax + by.$$

Note the Euclidean Algorithm is ran by applying the division algorithm repeatedly

$$\begin{aligned} a &= q_0 b + r_0 \\ b &= q_1 r_0 + r_1 \\ r_0 &= q_2 r_1 + r_2 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n \\ r_{n-1} &= q_{n+1} r_n \end{aligned}$$

where r_n is the last nonzero remainder.

Proof: Firstly, r_n exists, because $N(b) > N(r_0) > \cdots > N(r_n)$ is a decreasing sequence of nonnegative integers.

Next since R is a Euclidean Domain, then (a, b) is principle, thus $\gcd(a, b)$ exists and is equal to any element that generate ideal (a, b). Hence it suffices to show that $d = r_n$ generates this ideal.

By backtracking the steps of the Euclidean Algorithm, we can see that $d = ax + by$ for some $x, y \in R$ and d divides a and b. Hence $(d) \subset (a, b)$ and $(a, b) \subset (d)$. □

Definition 3.10 (Universal Side Divisor) *Let R be an integral domain, we denote $\tilde{R} = R^\times \cup \{0\}$. An element $u \in R - \tilde{R}$ is called a **universal side divisor** if for every $x \in R$, there is some $z \in \tilde{R}$ such that u divides $x - z$ in R, i.e., there is a type of "division algorithm" for u: every x may be written $x = qu + z$ where z is either zero or a unit.*

Proposition 3.11 *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 Euclidean with respect to some norm N and let u be an element of $R-\tilde{R}$ (which is nonempty since R is not a field) of minimal norm. For any $x \in R$, write $x = qu + r$, where r is either 0 or $N(r) < N(u)$. In either case the minimality of u implies $r \in \tilde{R}$ (any element with norm smaller than u should be in $\tilde{R}$). Hence u is a universal side divisor in R. □

Example:
The quadratic integer ring $R = \mathbb{Z}[(1+\sqrt{-19})/2] = \mathbb{Z}[\omega]$ is not a Euclidean domain with respect to any norm because R contains no universal side divisor, however, any ideals in R are principle.

One can easily show that ± 1 are the only units in R, so $\tilde{R} = \{0, \pm 1\}$. Suppose $u \in R$ is a universal side divisor and let $N(a + b\omega) = a^2 + ab + 5b^2$ denote the field norm on R. Note that if $a, b \in \mathbb{Z}$ and $b \neq 0$, then $a^2 + ab + 5b^2 = (a + b/2)^2 + 19/4b^2 \geq 5$ and so the smallest nonzero values of N on R are 1 and 4. Taking $x = 2$ in the definition of a universal side divisor, it follows that u must divide one of $2 - 0$ or 2 ± 1 in R, i.e., u is a nonunit divisor of 2 or 3 in R. If $2 = \alpha\beta$ then $4 = N(\alpha)N(\beta)$, then the norm of one of α and β must be 1, so one of them equal to ± 1. Hence the only divisors of 2 is in R are $\{\pm 1, \pm 2\}$. Similarly, the only divisors of 3 in R are $\{\pm 1, \pm 3\}$, so the only possible value for u are ± 2 or ± 3. But taking $x = \omega$, it is easy to check that none of $x, x \pm 1$ are divisible by ± 2 or ± 3 in R, so none of these is a universal side divisor. Hence R is not a Euclidean Domain.

Lemma 3.12 *Let R be a Euclidean Domain. If $(a, b) = 1$ and a divides bc, then a divides c. More generally, if a divides bc with nonzero a, b, then $\frac{a}{(a,b)}$ divides c.*

Proof: Similar to the case where $R = \mathbb{Z}$. □

Corollary 3.12.1 *Suppose a, b, N are integers, and a, b are nonzero. If (x_0, y_0) is a solution to the Diophantine equation*

$$ax_0 + by_0 = N,$$

then the set of all solutions of the equation is given by

$$x = x_0 + m\frac{b}{(a,b)}, \quad y = y_0 - m\frac{a}{(a,b)}$$

where m ranges over the integers.

Proof: Note that if x, y are solutions to $ax + by = N$, then $a(x - x_0) = b(y_0 - y)$. Then use Lemma 3.12. □

3.2 Principal Ideal Domains (P.I.D.s)

Definition 3.13 (Principal Ideal Domain) *A **principal ideal domain(P.I.D.)** is an integral domain in which every ideal is principal.*

Remark 3.13.1 *By Proposition 3.3, every Euclidean domain is a P.I.D. Hence any result that holds for P.I.D. automatically holds for Euclidan Domains.*

The following lemma is immediate from the previous results:

Lemma 3.14 *Let R be a principal ideal domain, and let a, b be nonzero elements of R. Let d be a generator for the principal ideal generated by a and b. Then*

1. *d is a greatest common divisor of a and b;*
2. *d can be written as an R-linear combination of a and b, i.e., there are elements x and y in R with*

$$d = ax + by$$

3. *d is unique up to multiplication by a unit of R.*

Lemma 3.15 *The product of two nonprincipal ideals in R can be principal.*

Proof: We provide an example. Let $R = \mathbb{Z}[\sqrt{-5}]$, and define $I = (2, 1 + \sqrt{-5})$. Note I is not principle, however, I^2 is the principal ideal generated by 2. □

Proposition 3.16 *Every nonzero prime ideal in a principal ideal domain is a maximal ideal.*

Proof: Let (p) be a nonzero prime ideal in the P.I.D. R and let $I = (m)$ be any ideal containing (p). We must show that $I = (p)$ or $I = R$. Now $p \in (m)$ so $p = rm$ for some $r \in R$. Since (p) is a prime ideal and $rm \in (p)$, then either r or must m lie in (p). If $m \in (p)$, then $(p) = (m) = I$; if $r \in (p)$, then write $r = ps$, so $p = rm = psm$, so $sm = 1$ and m is a unit, hence $I = R$. □

Corollary 3.16.1 *If R is any commutative ring such that the polynomial ring $R[x]$ is a P.I.D. (or a Euclidean Domain), then R must be a field.*

Proof: Assume $R[x]$ is a P.I.D. Since R is a subring of $R[x]$, then R must be an integral domain. The ideal (x) is a nonzero prime ideal in $R[x]$ because $R[x]/(x)$ is isomorphic to the integral domain R. Hence by Proposition 3.16, (x) is a maximal ideal, hence the quotient R is a field. □

Definition 3.17 (Dedekind-Hass Norm) *Define N to be a **Dedekind-Hass Norm** if N is a positive norm and for every nonzero $a, b \in R$ either a is an element of the ideal (b) or there is nonzero element in the ideal (a, b) of norm sctrictly smaller than the norm of b (i.e., either b divides a in R, or there exist $s, t \in R$ with $0 < N(sa - tb) < N(b)$).*

Remark 3.17.1 *If R is Euclidean with respect to a positive norm N, then it is always possible to satisfy the Dedekind-Hass condition with $s = 1$.*

Proposition 3.18 *The integral domain R is a P.I.D. if and only if R has a Dedekind-Hass norm.*

Proof: Suppose R has a Dedekind-Hass norm N. Let I be any nonzero ideal in R and let b be a nonzero element of I with $N(b)$ minimal. Suppose a is any nonzero element in I, so that the ideal (a, b) is contained in I. Then the Dedekind-Hass condition on N and the minimality of b implies that $a \in (b)$ (as there is no nonzero element in the ideal (a, b) with norm less than $N(b)$), so $I = (b)$ is principle.

The converse is proved in Corollary 3.28.2. □

Example:
Let $R = \mathbb{Z}[(1 + \sqrt{-19})/2] = \mathbb{Z}[\omega]$, then the field norm $N(a + b\omega) = a^2 + ab + 5b^2$ is a Dedekind-Hass norm on R, so R is a principle ideal domain.

Lemma 3.19 *In a P.I.D., two ideals (a) and (b) are comaximal if and only if a greatest common divisor of a and b is 1.*

Proof: If $(a, b) = 1$, then $\exists x, y \in \mathbb{R}$ such that $xa + yb = 1$. Since $xa \in (a)$ and $yb \in (b)$, then $(a) + (b) = R$. Conversely, if (a) and (b) are comaximal, then $(a) + (b) = R$, then $xa + yb = 1$ for some $x, y \in R$. □

Theorem 3.20 *Let R be an integral domain and suppose that every prime ideal in R is principal, then every ideal of R is principal, i.e., R is a P.I.D.*

Proof: Suppose towards a contradiction, we assume that the set $\mathcal{S}$ of ideals of R that are not principal is nonempty. Then this set is partially ordered by $\subsetneq$. We show that for any chain $\mathcal{I}$ under this partial order, it has an upper bounded. Take $I = \bigcup \mathcal{I}$, then I is clearly an ideal. We show it is not principal. Suppose $I = (a)$, then exists some ideal $L \in \mathcal{I}$ such that $a \in L$. However, this is a contradiction. Hence I is not principal. So by Zorn's Lemma, there is an maximal element for $\mathcal{S}$.

Denote any maximal ideal of $\mathcal{S}$ by I. And let $a, b \in R$, with $ab \in I$ but $a \notin I$ and $b \notin I$. Let $I_a = (I, a)$ be the ideal generated by I and a, and let $I_b = (I, b)$ be the ideal generated by I and b. Define $J = \{r \in R \,|\, rI_a \subseteq I\}$. Then I_a must be principal, so let $I_a = (\alpha)$. We can also easily verify J is an ideal and $I_b \subseteq J$. In particular, J must also be principal, with $J = (\beta)$, and $I \subsetneq I_b \subseteq J$. Then $I_a J = (\alpha\beta) \subseteq I$.

Now if $x \in I$, then $x \in I_a J$, because $I \subset I_a$ and $I \subset J$. So $x = s\alpha$ for some $s \in J$. So $I = I_a J$ is principal, which is a contradiction. Hence R must be an P.I.D. □

Proposition 3.21 *If R is a P.I.D. and D is a multiplicatively closed subset of R than $D^{-1}R$ (the ring of fractions) is also a P.I.D.*

Proof: Let $I = D^{-1}R$ be a nontrivial ideal. Then for every $x \in I$, $x = \frac{r}{d}$ for some $r \in R$ and $d \in D$. Then consider the set

$$L := \{r \in R \,|\, \frac{r}{d} \in I\}$$

We claim this is an ideal in R. If $a, b \in L$, then we have $\frac{a}{d} \in I$ for some $d \in D$ and $\frac{b}{e} \in I$ since I is an ideal. Then

$$\frac{a}{d} - \frac{e}{d}\frac{b}{e} = \frac{a-b}{d} \in I$$

so $a - b \in L$. Similarly, we can show closure by multiplication. Thus $L = (\alpha)$ as every ideal in R is principal. Now let d be such that $\frac{\alpha}{p} \in I$. Then $(\frac{\alpha}{p}) \subseteq I$. Now let $\frac{a}{d} \in I$ with $a = c\alpha$. Then we have $\frac{a}{d} = \frac{cp}{d}\frac{\alpha}{p}$, hence $I = (\frac{\alpha}{p})$. □

3.3 Unique Factorizaiton Domains (U.F.D.s)

Definition 3.22 *Let R be an integral domain.*

1. *Suppose $r \in R$ is nonzero and is not a unit. Then r is called **irreducible** in R if whenever $r = ab$ with $a, b \in R$. at least one of a or b must be a unit in R. Otherwise r is said to be **reducible**.*
2. *The nonzero element $p \in R$ is called **prime** in R if the ideal (p) generated by p is a prime ideal. In other words, a nonzero element p is a prime if it is not a unit and whenever $p|ab$ for any $a, b \in R$, then either $p|a$ or $p|b$.*
3. *Two elements a and b of R differing by a unit are said to be **associate** in R, i.e., $a = ub$ for some unit u in R.*

Lemma 3.23 *In an integral domain, a prime element is always irreducible.*

Proof: Suppose (p) is a nonzero prime ideal and $p = ab$. Then $ab = p \in (p)$, so by definition of prime ideal, one of a or b, say a, is in (p). Thus $a = pr$ for some r, then $p = ab = prb$, so $rb = 1$ and b is a unit. Thus this shows that p is irreducible. □

Remark 3.23.1 *The converse of this statement is not true. An irreducible element may not be prime.*

Proposition 3.24 *In a P.I.D. a nonzero element is a prime if and only if it is irreducible.*

Proof: By lemma 3.23, primes are irreducible. So we just need to show the converse. Suppose p is irreducible in a P.I.D. R, we show (p) is a prime ideal. If M is any ideal containing (p) then by hypothesis $M = (m)$ is a principal ideal. Sine $p \in (m)$, $p = rm$ for some r. But p is irreducible so by the definition, either r or m is a unit. This means either $(p) = (m)$ or $(m) = (1)$, respectively. Thus the only ideals containing (p) are (p) or $(1) = R$, i.e., (p) is a maximal ideal. Since maximal ideals are prime ideals, then the proof is complete. □

Definition 3.25 (Unique Factorization Domain) *A **Unique Factorization Domain (U.F.D.)** is an integral domain R in which every nonzero element $r \in R$ which is not a unit has the following two properties:*

1. *r can be written as a finite product of irreducible p_i of R (not necessarily distinct): $r = p_1 p_2 \cdots p_n$ and*

2. *the decomposition in* (1) *is unique up to associates: namely if* $r = q_1 q_2 \cdots q_m$ *is another factorization of* r *into irreducible, then* $m = n$ *and there is some renumbering of the factors so that* p_i *is associate to* q_i *for* $i = 1, 2, \cdots, n$.

Example:

1. A field F is trivially a U.F.D., since every nonzero element is a unit.

2. Every Principle Ideal Domain is a Unique Factorization Domain, which will be shown below. So every result for U.F.D. holds for P.I.D., hence Euclidean Domains.

3. The ring of polynomials $R[x]$ is a unique factorization domain if R is a unique factorization domain. In particular, $\mathbb{Z}[x]$ is a unique factorization domain.

4. $\mathbb{Z}[\sqrt{-5}]$ is an integral domain that is not a U.F.D. As $6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. And 2 is not an associate to $1 \pm \sqrt{-5}$ because they have different norm. The principal ideal (6) in $\mathbb{Z}[\sqrt{-5}]$ can be written as a product of 4 nonprinicpal prime ideals:$(6) = P_2^2 P_3 P_3'$ and the two distinct factorization of the elements 6 in $\mathbb{Z}[\sqrt{-5}]$ can be interpret as arising from two rearrangement of this product of ideals into product of principal ideals: the product of $P_2^2 = (2)$ with $P_3 P_3' = (3)$ and the product of $P_2 P_3 = (1+\sqrt{-5})$ with $P_2 P_3' = (1-\sqrt{-5})$.

Proposition 3.26 *In a Unique Factorization Domain, a nonzero element is a prime if and only if it is irreducible.*

Proof: Let R be a U.F.D. We only need to prove irreducible elements are prime, as the other direction is always true. Let p be irreducible in R and assume $p|ab$ for some $a, b \in R$. Writing a and b as a product of irreducibles, then from the uniqueness of the decompositions into irreducibles of ab, the irreducible element p must be associate to one of the irreducibles occurring either in the factorization of a in the factorization of b. We may assume that p is associate to one of the factorization of a, then $a = (up)p_2 \cdots p_n$ for u a unit and some (possibly empty set of) irreducibles $p_2, \cdots, p_n$. But then p divides a, so p is a rime. □

Proposition 3.27 *Let* a *and* b *be two nonzero elements of the Unique Factorization Domain* R *and suppose*

$$a = u p_1^{e_1} p_2^{e_2} \cdots p_n^{e_n} \quad \text{and} \quad b = v p_1^{f_1} p_2^{f_2} \cdots p_n^{f_n}$$

are prime factorizations for a *and* b*, where* u *and* v *are units, the primes* $p_1, \cdots, p_n$ *are distinct and the exponents* e_i *and* f_i *are* ≥ 0*. Then the element*

$$d = p_1^{\min(e_1, f_1)} p_2^{\min(e_2, f_2)} \cdots p_n^{\min(e_n, f_n)}$$

(where $d = 1$ *if all the exponents are* 0*) is the greatest common divisor 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 factorizations of both a and b, d divides both a and b. To show that d is a greatest common divisor, let c be any common divisor of a and b and let $c = q_1^{g_1} q_2^{g_2} \cdots q_m^{g_m}$ be the prime factorization of c. Since each q_i divides c, hence divides a and b, then q_i must divides one of the primes p_j (q_i's are irreducible hence primes). In particular, up to associates, the prime occurring in c must be a subset of the primes occurring in a and b:

$\{q_1, q_2, \cdots, q_n\} \subset \{p_1, p_2, \cdots, p_n\}$. Similarly, the exponents for the primes occurring in c must be no larger than those occurring in d. This implies that c divides d. $\square$

Theorem 3.28 *Every Principal Ideal Domain is a Unique Factorization Domain. In particular, every Euclidean Domain is a Unique Factorization Domain.*

Proof: Let R be a principal ideal domain, and let r be a nonzero element of R which is not a unit. We must show that r can be written as a finite product of irreducible elements of R and then we must verify that this decomposition is unique up to units.

If r is itself irreducible, then we are done. If not, then by definition, r can be written as a product $r = r_1 r_2$, where neither r_1 nor r_2 is a unit. If both these elements are irreducible, then we are done. Otherwise, at least on of the two elements, say r_1 is reducible, hence can be written as a product of two nonunit elements $r_1 = r_{11} r_{12}$, and so forth. We need to show this process terminates. Suppose this is not the case, from the factorization $r = r_1 r_2$, we obtain a proper inclusion of ideals (since r_2 is not a unit):

$$(r) \subset (r_1) \subset R.$$

Similarly, we have a proper inclusion

$$(r) \subset (r_1) \subset (r_{11}) \subset \cdots \subset R.$$

If this process do not terminate after a finite number of steps, then we would obtain an infinite ascending chain of ideals.

We show that any ascending chain $I_1 \subset I_2 \subset \cdots \subset R$ of ideals in a P.I.D. eventually becomes stationary, i.e., there is some positive integer n such that $I_k = I_n$ for all $k \geq n$. Let $I = \bigcup_{i=1}^{\infty} I_i$. It follows easily that I is an ideal. Since R is a principal ideal domain, it is generated by one element a. Then this a must be an element of one of the ideals in the chain, say $a \in I_n$. Then it must be the case that $I_n \subseteq I = (a) \subseteq I_n$. So the chain become stationary at I_n. This proves that every nonzero element of R which is not a unit has some factorization into irreducibles in R.

It remains to prove that the above decomposition is unique up to associates. We proceed by induction on the number, n, of irreducible factors in some factorization of the element r. If $n = 0$, then r is a unit. If we had $r = qc$ (some other factorization) for some irreducible q, then q would divide a unit, hence would itself be a unit, which is a contradiction (units are not irreducibles). So the case when $n = 0$ clearly holds.

Suppose now that n is at least 1 and that we have two products

$$r = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_m, \;\; m \geq n$$

for r where p_i and q_j are (not necessarily distinct) irreducibles. Since then p_1 divides the product on the right, then p_1 must divide one of the factors. Renumbering if necessary, we may assume p_1 divides q_1. But then $q_1 = p_1 u$ for some element u of R which must in fact be a unit since q_1 is irreducible. Thus p_1 and q_1 are associates. Now

we can cancel p_1 (this is an integral domain), and obtain

$$p_2 \cdots p_n = uq_2 q_3 \cdots q_m = q_2' q_3 \cdots q_m, \;\; m \geq n.$$

where $q_2' = uq_2$is again an irreducible. By induction on n, we conclude that each of the factors on the left matches bijectively (up to associates) with the factors on the right. Hence we complete the proof of the theorem. □

Corollary 3.28.1 (Fundamental Theorem of Arithmetic) *The integers $\mathbb{Z}$ is a unique Factorization Domain.*

Corollary 3.28.2 *Let R be a Principal ideal domain. Then there exists a multiplicative Dedekind-Hass norm on R.*

Proof: Let R be a P.I.D., then R is a U.F.D. Define the norm N by setting $N(0) = 0$, and $N(u) = 1$ if u is a unit, and $N(a) = 2^n$ if $a = p_1 p_2 \cdots p_n$, where the p_i are irreducibles in R. Then this norm is well-defined by the unique factorization of a. It is also clear that $N(ab) = N(a)N(b)$, so N is positive and multiplicative.

To show that N is a Dedekind-Hass norm, suppose that a, b are nonzero elements of R. Then the ideal generated by a and b is principal by assumption, say $(a, b) = (r)$. If a is not contained in the ideal (b), then also r is not contained in (b) (otherwise we would have $(r) = (b)$, a contradiction), i.e., r is not divisible by $b = xr$ for some $\in R$, it follows that x is not a unit in R and so $N(b) = N(x)N(r) > N(r)$. Hence (a, b) contains a nonzero element with norm strictly smaller than the norm of b, completing the proof. □

Proposition 3.29 *R is a P.I.D. if and only if R is a U.F.D. that is also a Bezout Domain.*

Proof: The forward direction has been shown in the previous results. We show the converse. Let I be any ideal of a U.F.D. R that is also a Bezout Domain, we show I is principle. Firstly, suppose I contain a unit, then we are done. So assume I does not contain a unit, and let a be the nonzero element of the ideal I with the minimal number of irreducible factors. Now we show $I = (a)$. If there exists $b \in I$ that is not in (a), then $(a, b) = (d)$ for some d, since R is a Bezout Domain. But then d has strictly less irreducible factors than a, which is a contradiction. □

In summary, we have the following inclusions among classes of commutative rings with identity:

$$\textbf{fields} \subset \textbf{Euclidean Domains} \subset \textbf{P.I.D.s} \subset \textbf{U.F.D.s} \subset \textbf{integral domains}.$$

3.3.1 Factorization in the Gaussian Integers

We now study the irreducible elements in $\mathbb{Z}[i]$, as we know $\mathbb{Z}[i]$ is a principal ideal domain, hence is a unique factorization domain.

In general, let $\mathcal{O}$ be a quadratic integer ring and let N be the associated field norm, suppose $\alpha \in \mathcal{O}$ is an element whose norm is a prime p in $\mathbb{Z}$. If $\alpha = \beta\gamma$ for some $\beta, \gamma \in \mathcal{O}$, then $p = N(\alpha) = N(\beta)N(\gamma)$ so that one of $N(\beta)$ or $N(\gamma)$ is ± 1 and the other is $\pm p$. That is one of the factors of α is a unit. Then we have the following lemma:

Lemma 3.30 *If $N(\alpha)$ is $\pm$ a prime in $(\mathbb{Z})$, then α is irreducible in $\mathcal{O}$.*

Suppose that π is a prime element in $\mathcal{O}$ and let (π) be the ideal generated by π in $\mathcal{O}$. Then we claim that $(\pi) \cap \mathbb{Z}$ is a prime ideal in $\mathbb{Z}$.

Lemma 3.31 *Suppose π is a prime element in $\mathcal{O}$, then $(\pi) \cap \mathbb{Z}$ is a prime ideal in $\mathbb{Z}$.*

Proof: It is clear that $(\pi) \cap \mathbb{Z}$ is an ideal in $\mathbb{Z}$, since (π) is an ideal in a large ring $\mathcal{O}$. Next suppose $a, b \in \mathbb{Z}$ are such that $ab \in (\pi) \cap \mathbb{Z}$, then $ab \in (\pi)$, so one of a or b in (π), hence one of a or b in $(\pi) \cap \mathbb{Z}$. □

By the lemma, we have $(\pi) \cap \mathbb{Z} = p\mathbb{Z}$ for some integer prime p. It follows from $p \in (\pi)$ that π is a divisor in $\mathcal{O}$ of the integer prime p, and so the prime elements in $\mathcal{O}$ can be found by determining how the primes in $\mathbb{Z}$ factor in the larger ring $\mathcal{O}$. Suppose π divides the prime p in $\mathcal{O}$, say $p = \pi\pi'$, then $N(\pi)N(\pi') = N(p) = p^2$. Then either $N(\pi) = \pm p^2$ or $N(\pi) = \pm p$. In the former case, $N(\pi') = \pm 1$, so $p = \pi$ up to associates is irreducible in $\mathcal{O}$. In the latter case, $N(\pi) = N(\pi') = \pm p$, hence π' is also irreducible and $p = \pi\pi'$ is the product of two irreducibles.

Now we come back to the case where $D = -1$, that is we come back to the Gaussian Integers $\mathbb{Z}[i]$. The units in $\mathbb{Z}[i]$ are ± 1 and $\pm i$. In this case, $\alpha = a + bi$ has $N(\alpha) = a\bar{\alpha} = a^2 + b^2$. It follows that p factors in $\mathbb{Z}[i]$ into precisely two irreducibles if and only if $p = a^2 + b^2$ is the sum of two integer squares, otherwise p remains irreducible in $\mathbb{Z}[i]$. We have the following number theory result:

Lemma 3.32 *The prime number $p \in \mathbb{Z}$ divides an integer of the form $n^2 + 1$ if and only if p is either 2 or is an odd prime congruent to 1 modulo 4.*

Proof: The case for $p = 2$ is trivial. Now if p is an odd prime, note that $p | n^2 + 1$ is equivalent to $n^2 = -1$ in $p/p\mathbb{Z}$. That is -1 is a quadratic residue modulo p, and we know this holds if and only if $p = 1 \mod 4$. □

Now we introduce the following definition:

Definition 3.33 (Gauss Primes) *A number $a + bi \in \mathbb{Z}[i]$ is a **Gauss Prime** if it is an irreducible / prime element in $\mathbb{Z}[i]$.*

In order to establish all Gauss primes in the ring of $\mathbb{Z}[i]$, we need one more lemma:

Lemma 3.34 *An integer d divides a Gaussian integer $a + bi$ in $\mathbb{Z}[i]$ if and only if d divides both a and b in $\mathbb{Z}$.*

Proof: This follows directly from complex division. □

Now we finally have the following theorem:

Theorem 3.35

1. *Let p be an integer prime. Then p is either a Gauss prime or the product $\pi\bar{\pi}$ of a Gauss prime and its complex conjugate.*
2. *The integer primes p that are Gauss primes are those congruent to 3 modulo 4. This is equivalent to saying $p \in \mathbb{Z}$ and $\mathbb{Z}[i]/(p)$ is a field.*
3. *(Fermat's Theorem on sums of squares) The prime p is the sum of two integer squares, $p = a^2 + b^2$,*

$a, b \in \mathbb{Z}$, if and only if $p = 2$ or $p = 1 \mod 4$. Except for interchanging a and b or changing the signs of a and b, the representation of p as a sum of two squares is unique.

4. *Let p be an integer prime, then the following statement are equivalent:*
   (a) *p is the product of complex conjugate Gauss primes.*
   (b) *p is congruent to 1 modulo 4 or $p = 2$.*
   (c) *p is the sum of two integer squares.*
   (d) *-1 is a Quadratic residue modulo p.*
5. *The irreducible elements in the Gaussian integers $\mathbb{Z}[i]$ are as follows:*
   (a) *$1 \pm i$ (which has norm 2).*
   (b) *Integer primes that is 3 modulo 4 (which has norm p^2).*
   (c) *$a + bi$, $a - bi$, the distinct irreducible factors of $p = a^2 + b^2 = (a + bi)(a - bi)$ for the prime $p \in \mathbb{Z}$ with $p = 1 \mod 4$ (both of which have norm p).*

Proof: Most of the results have been established previously, the only thing we need to show is Fermat's Theorem on sums of squares.

It is clear that if $p = 3 \mod 4$, then it cannot be written as the sum of two primes by simple mod 4 argument.

The case $p = 2$ is trivial, take $2 = 1^2 + 1^2$. Now if p is a prime that is 1 mod 4. Then by Lemma 3.32, we have $p|n^2 + 1$ for some integer n. Suppose p is irreducible, i.e., p is a prime in $\mathbb{Z}[i]$, then $p|n - i$ or $p|n + i$, but by Lemma 3.34, this is not possible, as $p \nmid 1$. This implies that p is not irreducible in $\mathbb{Z}[i]$. So by previous analysis, we must have $p = N(a + bi)$ for some $a + bi \in \mathbb{Z}$, that is $p = a^2 + b^2$ for some $a, b \in \mathbb{Z}$. Now in this case, since $p = (a + bi)(a - bi)$ and $a + bi, a - bi$ are irreducible because of their field norms. Then by the unique factorization of $\mathbb{Z}[i]$, and $\mathbb{Z}[i]$ only have units ± 1 and $\pm i$, then the representation $p = a^2 + b^2$ is unique up to sign and ordering of a and b. □

Corollary 3.35.1 *Let n be a positive integer and write*

$$n = 2^k p_1^{a_1} \cdots p_r^{a_r} q_1^{b_1} \cdots q_s^{b_s}$$

where $p_1, \cdots, p_r$ are distinct primes congruent to 1 mod 4 and $q_1, \cdots, q_s$ are distinct primes congruent to 3 mod 4. Then n can be written as a sum of two squares in $\mathbb{Z}$, if and only if each b_i is even. Further, if this condition on n is satisfied, then the number of representation of n as a sum of two squares is

$$4(a_1 + 1)(a_2 + 1) \cdots (a_r + 1).$$

Proof: If n can be written as the sum of two squares, iff $n = (x + yi)(x - yi)$ for some $x, y \in \mathbb{Z}$. Then if q_i divides n, as q_i is irreducible, q_i divides one of $x + y_i$ or $x - y_i$, but this implies $q_i|x$ and $q_i|y$, so q_i divides both factors, so exponents of q_i must be even. The backward direction is trivial, and follows from the fact that field norm is

multiplicative.

Next if $b_1, \cdots, b_s$ are all even. For each prime p_i congruent to 1 mod 4, with $p_i = \pi_i \bar{\pi}_i$ for $i = 1, \cdots, r$, where π_i and $\bar{\pi}_i$ are irreducibles. If $N(A + B_i) = n$, then examining norms, we see that, up to units, the factorization of $A + Bi$ into irreducible in $\mathbb{Z}[i]$ is given by

$$A + Bi = (1 + i)^k (\pi_1^{a_{1,1}} \bar{\pi_1}^{a_{1,2}}) \cdots (\pi_r^{a_{r,1}} \bar{\pi_r}^{a_{r,2}}) q_1^{b_1/2} \cdots q_s^{b_s/2}$$

with nonnegative integers $a_{i,1}, a_{i,2}$ satisfying $a_{i,1} + a_{i,2} = a_i$ for $i = 1, 2, \cdots, r$. Since $a_{i,1}$ can have the values $0, 1, \cdots, a_i$, and there are total of $(a_1 + 1)(a_2 + 1) \cdots (a_r + 1)$ distinct elements $A + Bi$ in $\mathbb{Z}[i]$ of norm n, up to units. Finally, since there are four units in $\mathbb{Z}[i]$, the second statement in the corollary follows. □

Corollary 3.35.2 *If an integer is the sum of two rational squares, then it is the sum of two integer squares.*

Proof: Let $n = \frac{a^2}{c^2} + \frac{b^2}{c^2}$, where $c \neq 0$. Then $nc^2 = a^2 + b^2$. Now as nc^2 can be written as the sum of two squares, then its prime factorization must be of the form given in Corollary 3.35.1. Then it follows n is of the form that can be written as the sum of two squares. □

Proposition 3.36 *Let $a, b \in \mathbb{Z}$ with $(a, b) = 1$, then $\mathbb{Z}[i]/(a + bi) \cong \mathbb{Z}/(a^2 + b^2)\mathbb{Z}$.*

Proof: Since $(a, b) = 1$, then there exists $k, j \in \mathbb{Z}$ such that $ak + bj = 1$. Then in the ideal $(a + bi)$, we have $(a+bi)(j+ki) = aj-bk+i$. Let $e = aj-bk$, then $i \equiv -e \mod a+bi$. Consequently, it foolows that the projection ring homomorphism $\pi : \mathbb{Z} \to \mathbb{Z}[i]/(a + bi)$ is onto.

Let $n = N(a + bi)$. Then if $m \in \ker \pi$, $a + bi | m$, which is equivalent to $\frac{m}{a+bi} = \frac{m(a-bi)}{n} = \frac{ma-mbi}{n} \in \mathbb{Z}[i]$. This can happen iff $n|ma$ and $n|mb$, so $n|(ma, mb) = m$. Thus by the First isomorphism Theorem, we conclude that $\mathbb{Z}/n\mathbb{Z} \cong \pi(\mathbb{Z}) = \mathbb{Z}[i]/(a + bi)$. □

Now, using the Chinese Remainder Theorem, we can determine $\mathbb{Z}[i]/(a)$, for any $a \in \mathbb{Z}[i]$. It suffices to consider the case when $\alpha = q^n$ where q is a prime element of $\mathbb{Z}[i]$.

4 Polynomial Rings

In this section, the ring R will always be a commutative ring with identity $1 \neq 0$.

4.1 Definitions and Basic Properties

Definition 4.1 (Polynomial Rings) *Fix a commutative ring R with $1 \neq 0$, we define the **Ring of polynomials in the variable x with coefficients in** R, $R[x]$, in the familiar way.*

*If R is an integral domain, then the quotient field of $R[x]$, or the **field of rational functions in x with coefficients in** R consists of all quotients $\frac{p(x)}{q(x)}$ where $q(x)$ is not the zero polynomial.*

Remark 4.1.1 *Note the ring R appears in $R[x]$ as the constant polynomials. Hence $R[x]$ is a commutative ring with identity 1_R.*

Lemma 4.2 *Let R be a commutative ring with 1. Then the principal ideal generated by x in the polynomial ring $R[x]$ is a prime ideal if and only if R is an integral domain; (x) is a maximal ideal if and only if R is a field.*

Definition 4.3 *Let E be a field containing the field F, and let $f(x) \in F[x]$. An element $\alpha \in E$ is a **root / zero** of $f(x)$ if $f(\alpha) = 0$.*
*A polynomial $f(x) \in \mathbb{Z}[x]$ is called **primitive**, if the greatest common divisor of the coefficients of $f(x) = 1$.*
*A polynomial $a_0 + a_1x + \cdots + a_nx^n$ over a ring R is called **monic** if $a_n = 1$.*
*A polynomial $f(x) \in \mathbb{F}[x]$ is called **irreducible** if the degree of $f(x) \geq 1$ and whenever $f(x) = g(x)h(x)$, where $g(x), h(x) \in F[x]$, then either $g(x) \in F$ or $h(x) \in F$.*
*Any polynomial that is not irreducible is called **reducible**.*
We use $\mathbb{R}[x]^x$ to denote the constant polynomials in $\mathbb{R}[x]$.

Remark 4.3.1 *Every monic polynomial is primitive.*

Proposition 4.4 *Suppose R is an integral domain, and $f, g \in \mathbb{R}[x]$, then*

1. $\deg(f \cdot g) = \deg f + \deg g$.
2. *The units of $R[x]$ are just the units of R.*
3. *$R[x]$ is an integral domain.*

Proof: If R has no zero divisors, then neither does $R[x]$: if $f(x)$ and $g(x)$ are polynomials with leading terms a_nx^n and b_mx^m repsectively, then the leading term of $p(x)q(x)$ is $a_nb_mx^{n+m}$ and $a_nb_m \neq 0$. This proves both 3 and 1. Next if $p(x) \in R[x]$ is a unit, say $p(x)q(x) = 1$ in $R[x]$. Then $\deg p(x) + \deg q(x) = 0$, so $p(x)$ and $q(x)$ must be the constant polynomials, i.e., elements of R. Hence the units of $R[x]$ are just units of R. □

Remark 4.4.1 *If the ring R has zero divisors then so does $R[x]$, because $R \subset R[x]$.*

Lemma 4.5 (McCoy's Theorem) *If $f(x)$ is a zero divisor of $R[x]$, then $cf(x) = 0$ for some nonzero $c \in R$.*

Proof: Suppose such c does not exists. Let $g \in R[x]$ be the polynomial with minimal degree such that $g(x)f(x) = 0$. Let

$$f(x) = \sum_{i=0}^{n} a_i x_i$$
$$g(x) = \sum_{i=0}^{m} b_i x_i$$

Then $b_m f(x)$ is either 0 or a polynomial with degree less than n with $(b_m f)g = 0$. Then we can proceed by induction on n. Then $(b_m^2 f)g = 0$ and $b_m^2 f$ is either 0 or a polynomial with degree less than $b_m f$ and so on. Eventually, there is a $k \in \mathbb{Z}_{\geq 0}$ such that $b_m^k f = 0$. □

Lemma 4.6 *If S is a subring of R, then $S[x]$ is a subring of $R[x]$.*

Proof: Clear from the definition of polynomial ring. □

Proposition 4.7 *Let I be an ideal of the ring R and let $(I) = I[x]$ denote the ideal of $R[x]$ generated by I. Then*

$$R[x]/(I) \cong (R/I)[x].$$

In particular, if I is a prime ideal of R, then (I) is a prime ideal of $R[x]$.

Proof: there is a natural map $\varphi : R[x] \mapsto (R/I)[x]$ given by reducing each of the coefficients of a polynomial modulo I. The definition of addition and multiplication in these two rings shows that φ is a ring homomorphism . The kernel is precisely the set of polynomials each of whose coefficients is an element of I, so $\ker \varphi = I[x] = (I)$.

Next if I is a prime ideal in R, then R/I is an integral domain, hence $(R/I)[x]$ is also an integral domain. This shows that if I is a prime ideal of R, then (I) is a prime ideal of $R[x]$. □

Remark 4.7.1 *It is not true that if I is a maximal ideal of R, then (I) is a maximal ideal of $R[x]$. However, if I is a maximal in R, then the ideal of $R[x]$ generated by I and x is maximal in $R[x]$.*

Example:

1. Let $R = \mathbb{Z}$ and consider the ideal $n\mathbb{Z}$ of $\mathbb{Z}$. Then the isomorphism above can be written

   $$\mathbb{Z}[x]/n\mathbb{Z}[x] \cong \mathbb{Z}/n\mathbb{Z}[x]$$

   and the natural projection map of $\mathbb{Z}[x]$ to $\mathbb{Z}/n\mathbb{Z}[x]$ by reducing the coefficients modulo n is a ring homomorphism. If n is composite, then the quotient ring is not an integral domain; if however, n is a prime p, then $\mathbb{Z}/pZ$ is a field and so $\mathbb{Z}/p\mathbb{Z}[x]$ is an integral domain. We also see that the set of polynomials whose coefficients are divisible by p is a prime ideal in $\mathbb{Z}[x]$.

Definition 4.8 (Polynomial Ring in Several Variables) *The **polynomial ring in the variables** $x_1, x_2, \cdots, x_n$ **with coefficients in** R, denoted $R[x_1, x_2, \cdots, x_n]$, is defined inductively by*

$$R[x_1, x_2, \cdots, x_n] = R[x_1, x_2, \cdots, x_{n-1}][x_n].$$

Remark 4.8.1 *This definition means that we can consider polynomials in n variables with coefficients in R simply as polynomials in one variable (say x_n) but now with coefficients that are themselves polynomials in $n-1$ variables. Alternatively, a nonzero polynomial in $x_1, x_2, \cdots, x_n$ with coefficients in R is a finite sum of nonzero monomial terms, i.e., a finite sum of elements of the form*

$$ax_1^{d_1}x_2^{d_2}\cdots x_n^{d_n}$$

where $a \in R$ and d_i are nonnegative integers. A monic term $x_1^{d_1}x_2^{d_2}\cdots x_n^{d_n}$ is called simply a ***monomial*** *and is the* ***monomial part*** *of the term $ax_1^{d_1}x_2^{d_2}\cdots x_n^{d_n}$. The exponent d_i is called the* ***degree*** *in x_i of the term and the sum*

$$d = d_1 + d_2 + \cdots + d_n$$

is called the ***degree*** *of the term. The order $n-$tuple $(d_1, \cdots, d_n)$ is the* ***multidegree*** *of the term. The* ***degree*** *of the nonzero polynomial is the largest degree of any of its monomial terms. A polynomial is called* ***homogeneous / form*** *if all its terms have the same degree. If f is a nonzero polynomial in n variables, the sum of all the monomial terms in f of degree k is called the* ***homogeneous component of f of degree*** *k. If f has degree d then f may be written uniquely as the sum $f_0 + f_1 + \cdots + f_d$ where f_k is the homogeneous component of f of degree k.*

Lemma 4.9 *Let $p(x_1, x_2, \cdots, x_n)$ be a homogeneous polynomial of degree k in $R[x_1, x_2, \cdots, x_n]$, then for all λ, $p(\lambda x_1, \cdots, \lambda x_n) = \lambda^k p(x_1, x_2, \cdots, x_n)$. Conversely, if p is homogeneous in degree k, then for any λ, we have $p(\lambda x_1, \cdots, \lambda x_n) = \lambda^k p(x_1, x_2, \cdots, x_n)$.*
The product of two homogeneous polynomials is again homogeneous.

Definition 4.10 (Polynomial Ring with Arbitrary Variables) *We define a* ***polynomial ring in an arbitrary number of variables with coefficients in*** *R by take finite sums of monomial terms of the type*

$$ax_1^{d_1}x_2^{d_2}\cdots x_n^{d_n}$$

but the variables are not restricted to just $x_1, \cdots, x_n$. And impose the natural addition and multiplication. Alternatively, we could define this ring as the union of all the polynomial rings in a finite number of the variables being considers.

4.2 Polynomial Rings over Fields

Let us consider the case when the coefficient ring is a field $\mathbb{F}$, and define the norm on $\mathbb{F}[x]$ by $N(p(x)) = \deg p(x)$ and $N(0) = 0$.

Theorem 4.11 (Division Algorithm) *Let $\mathbb{F}$ be a field. The polynomial ring $\mathbb{F}[x]$ is a Euclidean Domain. Specifically, if $f(x)$ and $s(x)$ are two polynomials in $\mathbb{F}[x]$ with $b(x)$ nonzero, then there are unique $q(x)$ and $r(x)$ in $\mathbb{F}[x]$ such that*

$$f(x) = q(x)s(x) + r(x) \quad \text{with } r(x) = 0 \text{ or } \deg r(x) < \deg s(x).$$

Proof: Let $n = \deg f$ and $m = \deg s$. If $n < m$, then take $q = 0$ and $r = f$, then we have the desired q and r, and it is clearly unique. Hence assume $n \geq m$.

Uniqueness:
Define $T : \mathbb{F}_{n-m}[x] \times \mathbb{F}_{m-1}[x] \to \mathbb{F}_n[x]$ by

$$T(q, r) = sq + r.$$

We show that T is a linear map:

$$T((q, r) + (q', r')) = s(q + q') + (r + r') = T(q, r) + T(q', r')$$

$$T(\lambda(q, r)) = \lambda sq + \lambda r = \lambda(T(q, r))$$

Hence T is indeed a linear map.
Suppose $(q, r) \in N(T)$, then $sq + r = 0$, and from the degree of the polynomials we get that $q = 0$ and $r = 0$. Hence $\dim N(T) = 0$. Hence proving the uniqueness part of the theorem.

Existence:
$\dim(\mathbb{F}_{n-m}[x] \times \mathbb{F}_{m-1}[x]) = (n - m + 1) + (m - 1 + 1) = n + 1 = \dim \mathbb{F}_n[x]$ Since T is injective as $N(T) = \{0\}$, then T is surjective. Thus proving the existence part of the theorem. □

Remark 4.11.1 *We note that the quotient and remainder in the Division Algorithm applied to $f(x)$ and $s(x)$ are independent of field extension. Suppose the field $\mathbb{F}$ is contained in the field $\mathbb{E}$, and $f(x) = Q(x)s(x) + R(x)$ for some $Q(x)$, $R(x)$ satisfying the condition in the theorem in $\mathbb{E}[x]$, then write $f(x) = q(x)s(x) + r(x)$ for some $q(x), r(x) \in \mathbb{F}[x]$ and apply the uniqueness condition in the ring $\mathbb{E}[x]$, we can deduce that $Q(x) = q(x)$ and $R(x) = r(x)$. In particular, if $s(x)$ divides $f(x)$ in the ring $\mathbb{E}[x]$ if and only if $s(x)|f(x)$ in $\mathbb{F}[x]$.*

Corollary 4.11.1 *If $\mathbb{F}$ is a field, then $\mathbb{F}[x]$ is a Principal Ideal Domain and a Unique Factorization Domain.*

Corollary 4.11.2 (Criterion for divisibility) *For any $f(x), g(x) \in \mathbb{F}[x]$ with $g(x) \neq 0$, one has $g(x)|f(x)$ in $\mathbb{F}[x]$ if and only if the unique remainder $r(x) \in \mathbb{F}[x]$ with $\deg r(x) < \deg g(x)$ in $\mathbb{N}$ s.t., $f(x) = q(x)g(x) + r(x)$ in $\mathbb{F}[x]$, is equal to 0.*

Proof: This follows from the uniqueness of the division algorithm. □

Definition 4.12 (relative prime) *Two nonzero polynomials $f_1(x), f_2(x) \in \mathbb{F}[x]_{\neq 0}$ are **relatively prime** if and only if there does not exists a polynomial $g(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ of positive degree such that $g(x)|f_1(x)$ and $g(x)|f_2(x)$ in $\mathbb{F}[x]$.*

Theorem 4.13 (Bézout's Theorem) *Let $f_1(x), f_2(x) \in F[x]_{\neq 0}$ be nonzero polynomial. Then $f_1(x)$ and $f_2(x)$ are relatively prime if and only if there exists a polynomial $q_1(x)$ and $q_2(x) \in \mathbb{F}[x]$ such that*

$$q_1(x)f_1(x) + q_2(x)f_2(x) = 1 \text{ in } \mathbb{F}[x]$$

Proof: $\Leftarrow$: suppose $q_1(x), q_2(x) \in \mathbb{F}[x]$ are such that

$$q_1(x)f_1(x) + q_2(x)f_2(x) = 1 \text{ in } \mathbb{F}[x],$$

but $f_1(x)$ and $f_2(x)$ are not relatively prime. Then there exists $g(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ of positive degree such that $g(x)|f_1(x)$ and $g(x)|f_2(x)$ in $\mathbb{F}[x]$, then

$$g(x)|(q_1(x)f_1(x) + q_2(x)f_2(x) \Longrightarrow g(x)|1$$

but then $\deg(g) \leq \deg(1) = 0$ in $\mathbb{N}$, which is a contradiction.

$\Rightarrow$: given any $f_1(x), f_2(x) \in \mathbb{F}[x]$, consider

$$I = \{q_1(x)f_1(x) + q_2(x)f_2(x) \in \mathbb{F}[x] \,:\, q_1(x), q_2(x) \in \mathbb{F}[x]\}$$

Then I is an ideal of the polynomial ring $\mathbb{F}[x]$, but as $\mathbb{F}[x]$ is a principal ideal domain, $I = (g(x))$ for some $g \in \mathbb{F}[x]$. But since $\gcd(f(x), g(x)) = c$ for some $c \in \mathbb{F}$, then $g(x)$ must be the constant polynomial, WLOG we can take $g(x) = 1 \in \mathbb{F}$. □

Theorem 4.14 (Euclid's Lemma for $\mathbb{F}(x)$) *Let $\phi(x) \in \mathbb{F}[x]_{\neq 0}$ be a nonzero polynomial. For any nonzero polynoimals $f(x), g(x) \in \mathbb{F}[x]_{\neq 0}$, if $\phi(x)|f(x)g(x)$ in $\mathbb{F}[x]$and $\phi(x), f(x)$ are relatively prime then $\phi(x)|g(x)$ in $\mathbb{F}[x]$.*

Proof: Since $\phi(x), f(x)$ are relatively prime, by Bézout's Theorem, there exists polynomials $q_1(x)$ and $q_2(x) \in \mathbb{F}[x]$ such that

$$q_1(x)\phi(x) + q_2(x)f(x) = 1 \text{ in } \mathbb{F}[x]$$

Multiply by $g(x)$, get

$$q_1(x)\phi(x)g(x) + q_2(x)f(x)g(x) = g(x) \text{ in } \mathbb{F}[x].$$

By hypothesis, we have $\phi(x)|f(x)g(x)$, so $\phi(x)|$LHS, and hence $\phi(x)|$RHS$= g(x)$ in $\mathbb{F}[x]$. □

Lemma 4.15 *Let $\phi(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ be an irreducible polynomial. For any nonzero polynoimal $f(x) \in \mathbb{F}[x]_{\neq 0}$, we have $\phi(x) \nmid f(x)$ in $\mathbb{F}[x]$ if and only if $\phi(x)$ and $f(x)$ are relatively prime.*

Proof: We prove the contrapositive of the statement:
If $\phi(x)|f(x)$ in $\mathbb{F}[x]$, then $\phi(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ is a polynomial of positive degree dividing both $\phi(x), f(x)$. So $\phi(x), f(x)$ are not relatively prime.
Conversely, if $\phi(x), f(x)$ are not relatively prime, then there exists $g(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ of positive degree such that $g(x)|\phi(x)$ and $g(x)|f(x)$ in $\mathbb{F}[x]$.
But since $\phi(x)$ is irreduible, $g(x)|\phi(x)$ implies $g(x) = c \cdot \phi(x)$ with $c \in \mathbb{F}[x]^x$ a nonzero constant. So $g(x)|f(x)$ implies $\phi(x)|f(x)$ in $\mathbb{F}[x]$. □

Theorem 4.16 (Euclid's Theorem for $\mathbb{F}(x)$) *Let $\phi(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ be an irreducible polynomial. For any nonzero polynomials $f(x), g(x) \in \mathbb{F}[x]_{\neq 0}$, if $\phi(x)|f(x)g(x)$ in $\mathbb{F}[x]$, then $\phi(x)|f(x)$ or $\phi(x)|g(x)$ in $\mathbb{F}[x]$.*

Proof: $\mathbb{F}[x]$ is a Euclidean Domain, hence the irreducible elements are primes elements. □

Corollary 4.16.1 *Let $\phi(x)$ and $\phi_1(x), \phi_2(x), \cdots, \phi_n(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ be irreducible monic polynomials. If $\phi(x)|\phi_1(x)\phi_2(x)\cdots\phi_n(x)$ in $\mathbb{F}[x]$, then there exists $i \in \{1, 2, \cdots, n\}$ such that $\phi(x) = \phi_i(x)$.*

Proof: By induction on n and Euclid's Theorem, we must have $\phi(x)|\phi_i(x)$ for some $i \in \{1, 2, \cdots, n\}$. But since $\phi_i(x)$ is monic irreducible, this implies $\phi_i(x) = c \cdot \phi(x)$ with $c \in F[x]^x$ a nonzero constant. Since $\phi(x)$, $\phi_i(x)$ are both monic, then it must be the case that $c = 1$. □

Theorem 4.17 (Unique Factorization Theorem for Polynomials) *For any polynomial $f(x) \in \mathbb{F}[x]_{\neq 0} \setminus \mathbb{F}[x]^x$ of positive degree, there exists a nonzero constant $c \in \mathbb{F}[x]^x$, and distinct irreducible monic polynomials $\phi_1(x)$, $\phi_2(x)$, $\cdots$, $\phi_k(x)$ and positive integers $n_1, n_2, \cdots, n_k$ such that*

$$f(x) = c \cdot [\phi_1(x)]^{n_1}[\phi_2(x)]^{n_2}\cdots[\phi_k(x)]^{n_k} \text{ in } \mathbb{F}[x].$$

The above $c \in \mathbb{F}[x]^x$ and $\phi_1(x), \phi_2(x), \cdots, \phi_k(x)$, $n_1, n_2, \cdots, n_k$ (up to permutation) are uniquely determined by $f(x)$.

Proof: $\mathbb{F}[x]$ is a unique factorization domain. □

Lemma 4.18 *Let $f(x)$ be a polynomial in $F[x]$, then $F[x]/(f(x))$ is a field if and only if $f(x)$ is irreducible.*

Proof: Since F is a field, then $F[x]$ is a principle ideal domain, then $(f(x))$ is maximal iff $(f(x))$ is a prime ideal iff $f(x)$ is irreducible. □

Lemma 4.19 *Let F be any field, then the ring $F[x]$ has infinitely many prime elements.*

Proof: Suppose not, consider $p_1 \cdots p_n + 1$. □

Let $f(x)$ be a polynomial in $F[x]$, since $F[x]$ is a principle ideal domain, then any ideal of $F[x]$ is of the form $(p(x))$. Now by the fourth isomorphism theorem, all the ideals of $F[x]/(f(x))$ is in bijection to the ideals of $F[x]$ containing $f(x)$, hence is given by $\overline{(p(x))}$ where $p(x)|f(x)$.

Lemma 4.20 *Let F be a finite field of order q and let $f(x)$ be a polynomial in $F[x]$ of degree $n \geq 1$, then $F[x]/(f(x))$ has q^n elements.*

Proof: In fact, $\bar{1}, \bar{x}, \cdots, \overline{x^{n-1}}$ is a basis of the vector space $F[x]/(f(x))$ over F. □

4.3 Polynomials Rings and U.F.D.

Proposition 4.21 (Gauss' Lemma) *Let R be a unique factorization domain with field of fractions F and let $p(x) \in R[x]$. If $p(x)$ is reducible in $F[x]$ then $p(x)$ is reducible in $R[x]$. More precisely, if $p(x) = A(x)B(x)$ for some nonconstant polynomials $A(x), B(x) \in F[x]$, then there are nonzero element $r, s \in F$ such that $rA(x) = a(x)$ and $sB(x) = b(x)$ both lie in $R[x]$ and $p(x) = a(x)b(x)$ is a factorization in $R[x]$.*

Proof: The coefficients of the polynomial on the right hand side of equation $p(x) = A(x)B(x)$ are elements in the field F, hence are quotients of elements from the Unique Factorization Domain R. Multiplying through by a common denominator for all these coefficients, we obtain an equation $dp(x) = a'(x)b'(x)$ where now $a'(x)$ and $b'(x)$ are elements of $R[x]$ and d is a nonzero element of R. If d is a unit in R, then the propositon is true with $a(x) = d^{-1}a'(x)$ and $b(x) = b'(x)$.

Suppose d is not a unit, then we can write d as a product of irreducibles in R, say $d = p_1 \cdots p_n$. Since p_1 is irreducible in R, the ideal (p_1) is prime, so by Proposition 4.7, we have that $p_1R[x]$ is prime in $R[x]$ (as $(p_1) = p_1R$), and $(R/p_1R)[x]$ is an integral domain. Reducing the equation $dp(x) = a'(x)b'(x)$ modulo p_1, we obtain the equation $0 = \overline{a'(x)b'(x)}$ in this integral domain, which implies one of the two factors, say $\overline{a'(x)}$ must be 0 But this means all the coefficients of $a'(x)$ are divisible by p_1, so that $\frac{1}{p_1}a'(x)$ also has coefficients in R. In other words, in the equation $dp(x) = a'(x)b'(x)$ we can cancel a factor of p_1 from d and from either $a'(x)$ or $b'(x)$ and still have an equation in $\mathbb{R}[x]$. Proceed with this inductively, we eventually will get $p(x) = a(x)b(x)$ with $a(x), b(x) \in R[x]$ and $a(x), b(x)$ being $F-$multiplies of $A(x), B(x)$ respectively. □

Remark 4.21.1 *The elements of the ring R becomes units in the Unique Factorization Domain $F[x]$. So the constant polynomials are always units in $F[x]$, while it may be irreducible in $R[x]$.*

Corollary 4.21.1 *Let R be a unique factorization domain, let F be its field of fractions and let $p(x) \in R[x]$. Suppose the greatest common divisor of the coefficients of $p(x)$ is 1. Then $p(x)$ is irreducible in $R[x]$ if and only if it is irreducible in $F[x]$. In particular, if $p(x)$ is monic polynomial that is irreducible in $R[x]$, then $p(x)$ is irreducible in $F[x]$.*

Proof: By Gauss' Lemma, if $p(x)$ is reducible in $F[x]$, then it is reducible in $R[x]$, so if $p(x)$ is irreducible in $R[x]$, then it cannot be reducible in $F[x]$. Conversely, the assumption on the greatest common divisor of the coefficient of $p(x)$ implies that if it is reducible in $R[x]$, then $p(x) = a(x)b(x)$ where neither $a(x)$ nor $b(x)$ are constant polynomials in $R[x]$, so they cannot be units in $\mathbb{F}[x]$. This same factorization shows that $p(x)$ is reducible in $F[x]$, hence completing the proof. □

Remark 4.21.2 *The corollary does not hold if R is not a U.F.D. That is if R is only an integral domain, and let F be the field of fractions of R, then $p(x)$ being an irreducible polynomial in $R[x]$ doesn't imply it being an irreducible polynomial in $F[x]$ (the other direction is true however). Thus if there is a ring R and a field of fraction F, with a polynomial $p(x) \in R[x]$, such that $p(x)$ is irreducible in $F[x]$ but not in $R[x]$, then R is not a U.F.D.*

Corollary 4.21.2 *Let R be an integral domain with quotient field F and let $p(x)$ be a monic polynomial in $R[x]$. Assume that $p(x) = a(x)b(x)$ where $a(x)$ and $b(x)$ are monic polynomials in $F[x]$ of smaller degree than $p(x)$. Then if $a(x) \notin R[x]$, R is not a Unique Factorization Domain.*

Proof: Suppose R is a unique factorization domain, then by Gauss Lemma $p(x) = a'(x)b'(x)$, where $a'(x) = ra(x)$ and $b'(x) = sb(x)$, where $r, s \in F^\times$. Now since p, a, b are monic, then $rs = 1$ and $ra \in R[x]$ implies $r \in R$, similarly, we have $s \in R$. Hence $r, s \in R^\times$ and $a(x) = r^{-1}a'(x) \in R[x]$ which is a contradiction. □

Corollary 4.21.3 *If $f(x)$ and $g(x)$ are polynomials with rational coefficients whose product $f(x)g(x)$ has integer coefficients, then the product of any coefficient of $g(x)$ with any coefficient of $f(x)$ is an integer.*

Proof: By Gauss Lemma, $\exists r \in Q$ such that $rf(x)$ and $\frac{1}{r}g(x)$ have integer coefficients. Then for any coefficients a, b of f and g respectively, we have $ra, \frac{b}{r} \in \mathbb{Z}$, hence $ab \in \mathbb{Z}$. □

Theorem 4.22 *R is a Unique Factorization Domain if and only $R[x]$ is a Unique Factorization Domain.*

Proof: It is clear that $R[x]$ being a U.F.D. forces R to be a U.F.D. Suppose conversely that R is a unique factorization domain, F is its field of fractions and $p(x)$ is a nonzero element of $R[x]$. Let d be the greatest common divisor of the coefficients of $p(x)$, so that $p(x) = dp'(x)$, where the *g.c.d* of the coefficients of $p'(x)$ is 1. Such a factorization of $p(x)$ is unique to a change in d, so up to a unit in R, and since d can be factored uniquely into irreducibles in R, it suffices to prove that $p'(x)$ can be factored uniquely into irreducibles in $R[x]$.

Hence we only need to consider the case where the g.c.d. of the coefficients of $p(x)$ is 1 and $p(x)$ is not a unit in $R[x]$, i.e., $\deg p(x) > 0$. Since $\mathbb{F}[x]$ is a U.F.D., then $p(x)$ can be factored uniquely into irreducibles in $F[x]$. By Gauss' Lemma, such a factorization implies there is a factorization of $p(x)$ in $R[x]$ whose factors are $F-$multiplies of the factors in $F[x]$. Since the greatest common divisor of the coefficients of $p(x)$ is 1, the *g.c.d* of the coefficients in each of these factors in $R[x]$ must be 1. Then by Corollary 4.21.1, each of these factors is an irreducible in $R[x]$. This shows that $p(x)$ can always be written as a finite product of irreducibles in $R[x]$.

The uniqueness of the factorization of $p(x)$ follows from the uniqueness in $F[x]$. Suppose

$$p(x) = q_1(x) \cdots q_r(x) = q_1'(x) \cdots q_s'(x)$$

are two factorization of $p(x)$ into irreducibles in $R[x]$. Since the g.c.d. of the coefficients of $p(x)$ is 1, the same is true for each of the irreducible factors, in particular, each has positive degree. Then again by Corollary 4.21.1, each $q_i(x)$ and $q_j'(x)$ is an irreducible in $F[x]$. By unique factorization in $F[x]$, $r = s$, and possibly after rearrangement, $q_i(x)$ and $q_i'(x)$ are associates in $F[x]$ for all $i = 1, 2, \cdots, r$. Since the units of $F[x]$ are precisely the elements of $F^\times = F \setminus \{0\}$ we need to consider when $q(x) = \frac{a}{b}q'(x)$ for some $q(x), q'(x) \in R[x]$ and nonzero element of a, b of R, where the greatest common divisor of the coefficients of each of $q(x)$ and $q'(x)$ is 1. In this case $bq(x) = aq'(x)$, then by comparing g.c.d., we must have $a = ub$ for some unit $u \in R$. Thus $q(x) = uq'(x)$ and so $q(x)$ and $q'(x)$ are associates in R as well. □

Corollary 4.22.1 *If R is a Unique Factorization Domain, then a polynomial ring in an arbitrary number of variables with coefficients in R is also a Unique Factorization Domain.*

Proof: For finitely many variables, this follows from induction using Theorem 4.22. The general case follows from the definition of a polynomial ring in an arbitrary number of variables as the infinite union of polynomial rings in n variables. □

4.4 Irreducibility Criteria

In this section we want to determine the irreducible elements in a Unique Factorization Domain $R[x]$. In the one-variable case, a nonconstant monic polynomial is irreducible in $R[x]$ if it cannot be factored as the product of two other polynomials of smaller degrees.

Proposition 4.23 *Let F be a field and let $p(x) \in F[x]$. Then $p(x)$ has a factor of degree one if and and only if $p(x)$ has a root in F, i.e., there is an $\alpha \in F$ with $p(\alpha) = 0$.*

Proof: If $p(x)$ has a factor of degree one, then since F is a field, we may assume the factor is monic, i.e., is of the form $(x - \alpha)$ for some $\alpha \in F$. But then $p(\alpha) = 0$. Conversely, suppose $p(\alpha) = 0$, by the division algorithm in $F[x]$, we can write

$$p(x) = q(x)(x - a) + r$$

where $r \in F$ is a constant. Since $p(\alpha) = 0$, r must be 0, hence $p(x)$ has $(x - \alpha)$ as a factor. □

Corollary 4.23.1 *A polynomial of degree two or three over a field F is reducible if and only if it has a root in F.*

Proof: A polynomial of degree two or three is reducible if and only if it has at least one linear factor. □

Proposition 4.24 *Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be a polynomial of degree n with integer coefficients. If $r/s \in \mathbb{Q}$ is in lowest terms and r/s is a root of $p(x)$, then r divides the constant term and s divides the leading coefficient of $p(x)$, that is $r|a_0$ and $s|a_n$. In particular, if $p(x)$ is a monic polynomial with integer coefficients and $p(d) \neq 0$ for all integers dividing the constant term of $p(x)$, then $p(x)$ has no roots in $\mathbb{Q}$.*

Proof: By hypothesis, $p(r/s) = 0 = a_n(r/s)^n + a_{n-1}(r/s)^{n-1} + \cdots + a_0$. Multiplying through by s^n gives

$$0 = a_n r^n + a_{n-1}r^{n-1}s + \cdots + a_0 s^n.$$

Thus

$$a_n r^n = s(-a_{n-1}r^{n-1} - \cdots - a_0 s^{n-1}),$$

so s divides $a_n r^n$. By assumption, s is relatively prime to r and it follows $s|a_n$. Similarly, we can show $r|a_0$. The second assertion is clear. □

Example:

1. The polynomial $x^3 - 3x - 1$ is irreducible in $\mathbb{Z}[x]$, as it has no rational roots. Since only ± 1 divides -1, and ± 1 are not roots of the polynomial, then this polynomial has no rational roots.
2. For any prime p, the polynomials $x^2 - p$ and $x^3 - p$ are irreducible in $\mathbb{Q}[x]$.

3. The polynomial x^2+1 is reducible in $\mathbb{Z}/2\mathbb{Z}[x]$, since it has 1 as a root, and it factors as $x^2+1=(x+1)(x+1)$.

4. The polynomial x^2+x+1 is irreducible in $\mathbb{Z}/2\mathbb{Z}$ since it does not have a root in $\mathbb{Z}/2\mathbb{Z}$ (check 0 and 1 are not roots).

5. The polynomial x^3+x+1 is irreducible in $\mathbb{Z}/2\mathbb{Z}[x]$.

Proposition 4.25 *Let I be a proper ideal in the integral domain R and let $p(x)$ be a nonconstant 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 $p(x)$ cannot be factored in $(R/I)[x]$ but that $p(x)$ is reducible in $R[x]$. Then $p(x)=a(x)b(x)$, for some nonconstant monic polynomial $a,b\in R[x]$. But then by reducing the coefficients modulo I, it gives a factorization in $(R/I)[x]$ with nonstant factors, a contradiction. □

Proposition 4.26 (Eisenstein's Criterion) *Let P be a prime ideal of the integral domain R and let $f(x)=x^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ be a polynomial in $R[x]$ ($n\geq 1$). Suppose $a_{n-1},\cdots,a_1,a_0$ are all elements of P and suppose a_0 is not an element of P^2, then $f(x)$ is irreducible in $R[x]$. Moreover, $f(x)$ is irreducible in $F[x]$ where F is the quotient field of R.*

Proof: Suppose $f(x)$ were reducible, say $f(x)=a(x)b(x)$ in $R[x]$, where $a(x)$ and $b(x)$ are nonconstant polynomials. Reducing this requation modulo P and using the assumption on the coefficients of $f(x)$, we obtain the equation $x^n=\overline{a(x)b(x)}$ in $(R/P)[x]$. Since P is a prime ideal, R/P is an integral domain, and it follows that both $\overline{a(x)}$ and $\overline{b(x)}$ have 0 constant term, i.e., the constant term of both $a(x)$ and $b(x)$ are elements of P. But then the constant term a_0 of $f(x)$ as the product of these two would be an element of P^2, a contradiction.

The second statement follows from the Gauss' Lemma. □

Corollary 4.26.1 (Eisenstein's Criterion for $\mathbb{Z}(x)$) *Let p be a prime in $\mathbb{Z}$ and let $f(x)=x^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0\in\mathbb{Z}[x]$, $n\geq 1$. Suppose p divides a_i for all $i\in\{0,1,\cdots,n-1\}$ but that p^2 does not divide a_0. Then $f(x)$ is irreducible in both $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.*

Corollary 4.26.2 *Let P be a prime ideal in the U.F.D. D and let $f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ be a polynomial in $R[x]$, $n\geq 1$. Suppose $a_n\notin P$, $a_{n-1},\cdots,a_0\in P$ and $a_0\notin P^2$. Then $f(x)$ is irreducible in $F[x]$, where F is the quotient field of R.*

Example:

1. x^n-a is irreducible in $\mathbb{Z}[x]$ if $p|a$ but $p^2\nmid a$ for some prime p.

2. Consider the polynomial $f(x)=x^4+1$, Eisenstein's Criterion cannot be applied directly. However, consider

$$g(x)=f(x+1)=(x+1)^4+1=x^4+4x^3+6x^2+4x+2$$

which is irreducible by Eisenstein's Criterion. Hence f is irreducible.

3. Similarly, consider the cyclotomic polynomial, let p be a prime

$$\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1.$$

Consider

$$\Phi_p(x+1) = \frac{(x+1)^p - 1}{x} = x^{p-1} + px^{p-2} + \cdots + \frac{p(p-1)}{2}x + p \in \mathbb{Z}[x]$$

which is irreducible in $\mathbb{Z}[x]$. Hence $\Phi_p(x)$ is irreducible in $\mathbb{Z}[x]$.

4. Let S be any integral domain, and let $R = S[x]$ and let n be any positive integer. Consider the polynomial $y^n - x$ in the ring $R[y]$. The ideal (x) is prime in the coefficient ring R, since $R/(x) = S[x]/(x)$ is the integral domain S. Eisenstein's Criterion for the ideal (x) of R applies directly to show that $y^n - x$ is irreducible in $R[y]$.

Proposition 4.27 *$x^{n-1} + x^{n-2} + \cdots + x + 1$ is irreducible over $\mathbb{Z}$ if and only if n is a prime.*

Proof: Suppose $n = ab$ is not a prime, then notice that

$$1 + x + \cdots + x^{ab-1} = (1 + x^a + \cdots + x^{a-1})(1 + x + \cdots + x^{a(b-1)}).$$

The converse is shown in the example above. □

4.5 More on Polynomial Rings over Fields

Proposition 4.28 *The maximal ideals in $F[x]$ are the ideals $(f(x))$ generated by irreducible polynomials $f(x)$. In particular, $F[x]/(f(x))$ is a field if and only if $f(x)$ is irreducible.*

Proof: $F[x]$ is a principle ideal domain. Then $(f(x))$ is an maximal ideal in $F[x]$ if and only if $f(x)$ is a prime element, which happens if and only if $f(x)$ is irreducible. □

Proposition 4.29 *Let $g(x)$ be a nonconstant element of $F[x]$ and let*

$$g(x) = f_1(x)^{n_1} f_2(x)^{n_2} \cdots f_k(x)^{n_k}$$

be its factorization into irreducibles, where the $f_i(x)$ are distinct. Then we have the following isomorphism of rings:

$$F[x]/(g(x)) \cong F[x]/(f_1(x)^{n_1}) \times F/[x]/(f_2(x)^{n_2}) \times \cdots \times F[x]/(f_k(x)^{n_k})$$

Proof: This follows from the Chinese Remainder Theorem, since the ideals $(f_i(x)^{n_i})$ and $(f_j(x)^{n_j})$ are comaximal if $f_i(x)$ and $f_j(x)$ are distinct (they are relatively prime in the Euclidean Domain $F[x]$). □

Lemma 4.30 *If the polynomial $f(x)$ has roots $\alpha_1, \alpha_2, \cdots, \alpha_k$ in F (not necessarily distinct), then $f(x)$ has $(x - \alpha_1) \cdots (x - \alpha_k)$ as a factor. In particular, a polynomial of degree n in one variable over a field F has at most n roots in F, even counted with multiplicity.*

Proof: Induction using Proposition 4.23 gives the first statement. For the second statement, a prove by contradiction can yields the result. □

Proposition 4.31 *Suppose m is a non-negative integer, $z_1, \cdots .z_{m+1}$ are distinct elements of F, and $w_1, \cdots, w_{m+1} \in F$. Then there is a unique polynomial $p \in F_m[x]$ such that*

$$p(z_j) = w_j$$

for $j = 1, \cdots, m+1$.

Proof: Define $T : F_m[x] \to F^{m+1}$ by

$$T(p) = (p(z_1), \cdots, p(z_{m+1})).$$

Then T is clearly a linear map. The null space of T is clearly $\{0\}$ as a polynomial of degree less than m cannot have more than m roots, hence T is injective. Since $\dim F_m[x] = m + 1 = \dim F^{m+1}$, then it follows that T is surjective as well. Therefore it must be the case that there exists a unique polynomial with the desired property. □

Proposition 4.32 *A finite subgroup of the multiplicative group of a field is cyclic. In particular, if F is a finite field, then the multiplicative group $F^\times$ of nonzero element of F is a cyclic group.*

Proof: We give a proof of this result using the Fundamental Theorem of Finitely Generated Abelian Groups. By the Fundamental Theorem, the finite subgroup can be written as the direct product of cyclic groups

$$\mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z},$$

where $n_k | n_{k-1} | \cdots | n_2 | n_1$. In general, if G is a cyclic group and $d||G|$ then G contains precisely d elements of order dividing d. Since n_k divides the order of each of the cyclic groups in the direct product, it follows that each direct factor contains n_k elements of order dividing n_k. If k were greater than 1, there would therefore be a total of more than n_k such elements, But then there would be more than n_k roots of the polynomial $x^{n_k} - 1$ in the field F, contradicting Lemma 4.30. Hence $k = 1$ and the group is cyclic. □

Corollary 4.32.1 *Let p be a prime. The multiplicative group $(\mathbb{Z}/p\mathbb{Z})^\times$ of nonzero residue class mod p is cyclic.*

Corollary 4.32.2 *Let $n \geq 2$ be an integer with factorization $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ in $\mathbb{Z}$, where $p_1, \cdots, p_r$ are distinct primes. Then we have the following isomorphisms of (multiplicative) groups:*

1. $(\mathbb{Z}/n\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z})^\times \times (\mathbb{Z}/p_2^{\alpha_2}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/p_r^{\alpha_r}\mathbb{Z})^\times$
2. *$(\mathbb{Z}/2^\alpha\mathbb{Z})^\times$ is the direct product of a cyclic group of order 2 and a cyclic group of order $2^{\alpha-2}$, for all $\alpha \geq 2$.*
3. *$(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ is a cyclic group of order $p^{\alpha-1}(p-1)$, for all odd primes p.*

Proof: The isomorphism in (1) follows from the Chinese Remainder Theorem. The isomorphism in (2) follows directly from the fact that $1+p$ is an element of order p^{n-1} in the multiplicative group $(\mathbb{Z}/p^n\mathbb{Z})^\times$ and $(\mathbb{Z}/2^n\mathbb{Z})^\times$ is not cyclic for any $n \geq 3$ (there are two distinct subgroups of order 2).

For p an odd prime, $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ is an abelian group of order $p^{\alpha-1}(p-1)$. Then the Sylow $p-$subgroup of this group is cyclic, as there is an element of order $p^{\alpha-1}$. The map

$$\mathbb{Z}/p^\alpha\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \text{ defined by } a+(p^\alpha) \mapsto a+(p)$$

is a ring homomorphism (reduction mod p) which gives a surjective group homomorphism from $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ onto $(\mathbb{Z}/p\mathbb{Z})^\times$. The latter group is cyclic of order $p-1$. The kernel of this map is of order $p^{\alpha-1}$, hence for all primes $q \neq p$, the Sylow q-subgroup of $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ maps isomorphically into the cyclic group $(\mathbb{Z}/p\mathbb{Z})^\times$. All Sylow subgroups of $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ are therefore cyclic, thus completing the proof. □

Remark 4.32.1 *These isomorphism describe the group-theoretic structure of the automorphism group of the cyclic group, Z_n, of order n. Since* $\mathrm{Aut}(Z_n) \cong (\mathbb{Z}/n\mathbb{Z})^\times$.

Theorem 4.33 *Let φ denote the Euler's φ-function, then*

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

Let $\psi(d)$ denote the number of elements of G of order d. Then $\varphi(d) = \psi(d)$ for every divisor d of n.

Proof: If G is a finite subgroup of order n of the multiplicative group $F^\times$ of nonzero elements of the field F. Then for each d dividing n, G contains a unique subgroup of order d and the number of elements of G of order d is thus $\varphi(d)$ (consider this subgroup of order d). Then $|G|$ is the sum of $\varphi(d)$ as d runs over all divisors of n. □

Lemma 4.34 *The additive and multiplicative groups of a field are never isomorphic.*

Proof: If $|F|$ is finite, then this is clear as $|F| \neq |F^\times|$. So we consider the case that $|F|$ is infinite.

Let $\varphi : F^\times \to F$ be a group isomorphism.
If $-1 \neq 1$ in F, then $(-1)^2 = 1$ and $\varphi(-1) + \varphi(-1) = \varphi(1) = 0$. Then $2\varphi(-1) = 0$ which implies $\varphi(-1) = 0$ as $2 \neq 0$ in F. This gives a contradiction.

Now if $-1 = 1$ in F, then F is of characteristic 2, so every element of $F^\times$ is of order 2. Let a, b be any two non-identity element of $F^\times$. Then $\langle a, b\rangle$ is a cyclic group of order 4 (consider $\langle \varphi(a), \varphi(b)\rangle$), but this contradicts to the fact that a, b are of order 2. □

4.6 Polynomials in Several Variables over A Field and Gröbner Bases

We know that a polynomial ring $F[x]$ in a variable x over a field F is a Unique Factorization Domain and hence the polynomial ring $F[x_1, \cdots, x_n]$ is a Unique factorization Domain. However the ring $F[x_1, \cdots, x_n]$ is not a Principal Ideal Domain unless $n = 1$. Our first result is to show that ideals in such polynomial rings, although not necessarily principal, are always finitely generated. General rings with this property are given a special name:

Definition 4.35 (Noetherian) *A commutative ring R with 1 is called **noetherian** if every ideal of R is finitely generated.*

Theorem 4.36 (Hilbert's Basis Theorem) *If R is a Noetherian ring then so is the polynomial ring $R[x]$.*

Proof: Let I be an ideal in $R[x]$ and let L be the set of all leading coefficients of the elements in I. We first show that L is an ideal of R, as follows. Since I contains the zero polynomial, $0 \in L$. Let $f = ax^d + \cdots$ and $g = bx^e + \cdots$ be polynomials in I of degree d, e and leading coefficients $a, b \in R$. Then for any $r \in R$, either $ra - b$ is zero or it is the leading coefficient of the polynomial $rx^e f - x^d g$. Since the latter polynomial is in I, we have $ra - b \in L$, which shows L is an ideal of R. Since R is assumed Noetherian, the ideal L in R is finitely generated, say by $a_1, a_2, \cdots, a_n \in R$. For each $i = 1, \cdots, n$, let f_i be an element of I whose leading coefficients is a_i. Let e_i denote the degree of f_i, and let N be the maximum of $e_1, e_2, \cdots, e_n$.

For each $d \in \{0, 1, \cdots, N-1\}$, let L_d be the set of all leading coefficients of polynomials in I of degree d together with 0. A similar argument as that for L shows that each L_d is also an ideal of R, again finitely generated since R is Noetherian. For each nonzero ideal L_d, let $b_{d,1}, b_{d,2}, \cdots, b_{d,n_d} \in R$ be a set of generators for L_d, and let f_{d_i} be a polynomial in I of degree d with leading coefficients b_{d_i}.

We show that the polynomials $f_1, \cdots, f_n$ together with all the polynomials f_{d_i} for all the nonzero ideals L_d are a set of generators for I, i.e., that

$$I = (\{f_1, \cdots, f_n\} \cup \{f_{d_i} \,|\, 0 \leq d \leq N,\, 1 \leq i \leq n_d\}).$$

By construction, the ideal I' on the right above is contained in I since all the generators were chosen in I. If $I' \neq I$, there exists a nonzero polynomial $f \in I$ of minimum degree with $f \notin I'$. Let $d = \deg f$ and let a be the leading coefficient of f.

Suppose first that $d \geq N$, since $a \in L$ we may write a as an $R-$linear combination of the generators of L : $a = r_1 a_1 + \cdots + r_n a_n$, then $g = r_1 x^{d-e_1} f_1 + \cdots + r_n x^{d-e_n} f_n$ is an element of I' with the same degree d and the same leading coefficient a as f. Then $f - g \in I$ is a polynomial in I of smaller degree than f. By the minimality of f, we must have $f - g = 0$, so $f = g \in I'$ a contradiction.

Suppose next that $d < N$. In this case $a \in L_d$ for some $d < N$, and so we may write $a = r_1 b_{d,1} + \cdots + r_{n_d} b_{n_d}$ for some $r_i \in R$. Then $g = r_1 f_{d,1} + \cdots + r_{n_d} f_{n_d}$ is a polynomial in I' with the same degree d and the same leading coefficient a as f, and we have a contradiction as before.

It follows that $I = I'$ is finitely generated, and since I was arbitrary, this completes the proof that $R[x]$ is Noetherian. □

Corollary 4.36.1 *Every ideal in the polynomial ring $F[x_1, x_2, \cdots, x_n]$ with coefficients from a field F is finitely generated.*

Definition 4.37 (Monomial Ordering) *A **monomial ordering** is a well ordering $\geq$ on the set of monomials that satisfies $mm_1 \geq mm_2$ whenever $m_1 \geq m_2$ for monomials m, m_1, m_2. Equivalently, a monomial ordering may be specified by defining a well ordering on the n-tuples $\alpha = (\alpha_1, \cdots, \alpha_n) \in \mathbb{Z}^n$ of multidegrees of monomials $Ax_1^{\alpha_1} \cdots x_n^{\alpha_n}$ that satisfies $\alpha + \gamma \geq \beta + \gamma$ if $\alpha \geq \beta$.*

Remark 4.37.1 *Any total ordering on monomials which for every monomial m satisfies $m \geq 1$ and $mm_1 \geq mm_2$ whenever $m_1 \geq m_2$ is necessarily a well ordering (hence a monomial ordering) by Hilbert's Basis Theorem.*

Definition 4.38 *Fix a monomial ordering on the polynomial ring $F[x_1, x_2, \cdots, x_n]$.*

1. *The **leading term** of a nonzero polynomial f in $F[x_1, x_2, \cdots, x_n]$, denoted $\mathrm{LT}(f)$, is the monomial term of maximal order in f and the leading term of $f = 0$ is 0. Define the **multidegree of** f, denoted $\partial(f)$, to be the multidegree of the leading term of f.*
2. *If I is an ideal in $F[x_1, x_2, \cdots, x_n]$, the **ideal of leading terms**, denoted $\mathrm{LT}(I)$, is the ideal generated by the leading terms of all the elements in the ideal, i.e., $\mathrm{LT}(I) = (LT(f) \mid f \in I)$.*

Remark 4.38.1 *The leading term and the multidegree of a polynomial depends on the choice of the ordering.*

Lemma 4.39 *Fix a monomial ordering, when f and g are nonzero,*

1. *$\partial(fg) = \partial f + \partial g$;*
2. *$\mathrm{LT}(fg) = \mathrm{LT}(f) + \mathrm{LT}(g)$.*

Definition 4.40 (Gröbner Basis) *A **Gröbner basis** for an ideal I in the polynomial ring $F[x_1, \cdots, x_n]$ is a finite set of generators $\{g_1, \cdots, g_n\}$ for I whose leading terms generate the ideal of all leading terms in I, i.e.,*

$$I = (g_1, \cdots, g_m) \quad \textit{and} \quad \mathrm{LT}(I) = (\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m)).$$

4.6.1 General Polynomial Division

Fix a monomial ordering on $F[x_1, \cdots, x_n]$, and suppose $g_1, \cdots, g_m$ is a set of nonzero polynomials in $F[x_1, \cdots, x_n]$. If f is any polynomial in $F[x_1, \cdots, x_n]$, start with a set of quotients $q_1, \cdots, q_m$ and a remainder r initially all equal to 0 and successively test whether the leading term of the dividend f is divisible by the leading term of the divisors $g_1, \cdots, g_m$ in that order. Then

1. If $\mathrm{LT}(f)$ is divisible by $\mathrm{LT}(g_i)$, say, $\mathrm{LT}(f) = a_i \mathrm{LT}(g_i)$, add a_i to the quotient q_i, replace f by the dividend $f - a_i g_i$, and reiterate the entire process.
2. If the leading term of the dividend f is not divisible by any of the leading terms $\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m)$, add the leading term of f to the remainder r, replace f by the dividend $f - LT(f)$, and reiterate the entire process.

3. The process terminates when the dividend is 0 and results in a set of quotients $q_1, \cdots, q_m$ and a remainder r with

$$f = q_1 g_1 + \cdots + q_m g_m + r.$$

Each $q_i g_i$ has multidegree less than or equal to the multidegree of f and the remainder r has the property that no zero term in r is divisble by any of the leading terms $\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m)$.

Example:
Fix the lexicographic ordering $x > y$ on $F[x, y]$.

1. Suppose $f = x^3y^3 + 3x^2y^4$ and $g = xy^4$. The leading term of f is x^3y^4, which is not divisible by the leading term of g, so x^3y^3 is added to the remainder r, and replace f by $f - \mathrm{LT}(f) = 3x^2y^4$ and we start over. $3x^2y^4$ is divisible by $\mathrm{LT}(g) = xy^4$, with quotient $a = 3x$, we add $3x$ to the quotient q, and the next f be comes 0, at which point the process terminates. Thus

$$x^3y^3 + 3x^2y^4 = (3x)(xy^4) + x^3y^3.$$

2. Similarly, if $f = x^2 + x - y^2 + y$, with $g_1 = xy + 1$ and $g_2 = x + y$, then

$$f = x^2 + x - y^2 + y = (-1)(xy + 1) + (x + 1)(x + y) + (-y^2 + 1).$$

If $g_1 = x + y$ and $g_2 = xy + 1$, then

$$f = x^2 + x - y^2 + y = (x - y + 1)(x + y).$$

This shows that the polynomial $f = x^2 + x - y^2 + y$ is an element of the ideal $I = (x + y, xy + 1)$.

Theorem 4.41 *Fix a monomial ordering on $R = F[x_1, \cdots, x_n]$ and suppose $\{g_1, \cdots, g_m\}$ is a Gröbner basis for the nonzero ideal I in R. Then*

1. *Every polynomial $f \in R$ can be written uniquely in the form*

$$f = f_I + r$$

where $f_I \in I$ and no nonzero monomial term of the remainder r is divisible by any of the leading terms $\mathrm{LT}(g_1) < \cdots, \mathrm{LT}(g_m)$.

2. *Both f_1 and r can be computed by general polynomial division by $g_1, \cdots, g_m$ and are independent of the order in which these polynomials are used in the division.*

3. *The remainder r provides a unique representative for the coset of f in the quotient ring $F[x_1, \cdots, x_n]/I$. In particular, $f \in I$ if and only if $r = 0$.*

Proof: Letting $f_I = \sum_{i=1}^{m} q_i g_i \in I$ in the general polynomial division of f by $g_1, \cdots, g_m$ immediately gives a decomposition $f = f_I + r$ for any generators $g_1, \cdots, g_m$. Suppose now that $\{g_1, \cdots, g_m\}$ is a Gröbner basis, and $f = f_I + r = f_I' + r'$. Then $r - r' = f_I' - f_I \in I$, so its leading term $\mathrm{LT}(r - r')$ is an element of $\mathrm{LT}(I)$, which is the ideal $(\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m))$ since $\{g_1, \cdots, g_m\}$ is a Gröbner basis for I. Every element in this ideal is a

sum of multiples of the monomial terms $\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m)$, so is a sum of terms each of which is divisible by one of the $\mathrm{LT}(g_i)$. But both r and r', hence also $r - r'$, are sums of monomial terms none of which is divisible by $\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m)$, which is a contradiction unless $r - r' = 0$. IT follows that $r = r'$ is unique, hence so is $f_I = f - r$, which proves (1).

We know f_I and r can be computed algorithmically by polynomial division, and the uniqueness in (1) implies that r is independent of the order in which the polynomials $g_1, \cdots, g_m$ are used in the division, so (2) holds.

The first assertion in (3) is immediate from the uniqueness in (1). If $r = 0$¡,then $f = f_I \in I$; conversely, if $f \in I$, then $f = f + 0$ together with the uniqueness of r implies that $r = 0$. □

Lemma 4.42 *Suppose I is an ideal of $F[x_1, \cdots, x_n]$ generated by a (possibly infinite) set S of polynomials. Then a finite subset of polynomials in S is sufficient to generate I.*

Proof: Suppose $I = (S)$. By Hilbert's Basis Theorem, we know I is finitely generated, so $I = (f_1, \cdots, f_m)$ for some $f_1, \cdots, f_m \in I$. But then each f_i is in the ideal generated by S, so it is the finite $F[x_1, \cdots, x_n]$-combination of elements in S, that is for each f_i, there exists $s_{i1}, s_{i2}, \cdots, s_{in_i} \in S$ such that

$$f_i = q_{i1}s_{i1} + q_{i2}s_{i2} + \cdots + q_{in_i}s_{in_i}$$

where $q_{ij} \in F[x_1, \cdots, x_n]$. Then the collection of all such s_{ij} is finite and certainly generate $(f_1, \cdots, f_m)$, hence generates I. □

Proposition 4.43 *Fix a monomial ordering on $R = F[x_1, \cdots, x_n]$ and let I be a nonzero ideal in R.*

1. *If $g_1, \cdots, g_m$ are any elements of I such that $\mathrm{LT}(I) = (\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m))$, then $\{g_1, \cdots, g_m\}$ is a Gröbner basis for I.*
2. *The ideal I has a Gröbner basis.*

Proof:

1. Suppose $g_1, \cdots, g_m \in I$ with $\mathrm{LT}(I) = (\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m))$. We need to see that $g_1, \cdots, g_m$ generate the ideal I. If $f \in I$, use general polynomial division to write $f = \sum_{i=1}^{m} q_i g_i + r$ where no nonzero term in the remainder r is divisible by any $\mathrm{LT}(g_i)$. Since $f \in I$, also $r \in I$, which means $\mathrm{LT}(r)$ is in $\mathrm{LT}(I)$. But then $\mathrm{LT}(r)$ would be divisible by one of $\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m)$, which is a contradiction unless $r = 0$. Hence $f = \sum_{i=1}^{m} q_i g_i$ and $g_1, \cdots, g_m$ generate I. So $g_1, \cdots, g_m$ is a Gröbner basis for I.
2. Note that the ideal $\mathrm{LT}(I)$ of leading terms of any ideal I is a monomial ideal generated by all the leading terms of the polynomials in I, and by Lemma 4.42, there is a finite number of these monomial $h_1, \cdots, h_k \in I$ such that $\mathrm{LT}(I) = (\mathrm{LT}(h_1), \cdots, \mathrm{LT}(h_k))$. Hence $h_1, \cdots, h_k$ forms a Gröbner basis of I by part (1).

□

Lemma 4.44 *Suppose $f_1, \cdots, f_m \in F[x_1, \cdots, x_n]$ are polynomials with the same multidegree α and that the linear combination $h = a_1 f_1 + \cdots + a_m f_m$ with constant $a_i \in F$ has strictly smaller multidegree. Then*

$$h = \sum_{i=2}^{m} b_i S(f_{i-1}, f_i), \quad \text{for some constants } b_i \in F$$

where

$$S(f_s, f_t) = \frac{M}{\mathrm{LT}(f_s)} f_s - \frac{M}{\mathrm{LT}(f_t)} f_t$$

and M is the monic least common multiple of the monomial terms $\mathrm{LT}(f_1)$ and $\mathrm{LT}(f_2)$.

Proof: Write $f_1 = c_i f_i'$ where $c_i \in F$ and f_i' is a monic polynomial of multidegree α. We have

$$\begin{aligned} h &= \sum_{a_i c_i f_i'} \\ &= a_1 c_1 (f_1' - f_2') + (a_1 c_1 + a_2 c_2)(f_2' - f_3') + \cdots \\ &\quad + (a_1 c_1 + \cdots + a_{m-1} c_{m-1})(f_{m-1}' - f_m') + (a_1 c_1 + \cdots + a_m c_m) f_m' \end{aligned}$$

Note that $f_{i-1}' - f_i' = S(f_{i-1}, f_i)$. And since

$$S(f_{i-1}, f_i) = \frac{M}{\mathrm{LT}(f_{i-1})} f_{i-1} - \frac{M}{\mathrm{LT}(f_i)} f_i$$

cancels the leading term of f_1 and f_2, hence have multidegree strictly smaller than α. Then since h and each $f_{i-1}' - f_i'$ has multidegree strictly smaller than α, we have $a_1 c_1 + \cdots + a_m c_m = 0$, so the last term on the right hand side is 0 and the lemma follows. □

Using this lemma, we can check whether a set of generators $g_1, \cdots, g_m$ is a Gröbner basis. This happens if there are no new leading terms among the differences $S(g_i, g_j)$ not already accounted for by g_i. This result provides the principal ingredient in an algorithm to construct a Gröbner basis.

For a fixed monomial ordering on $R = F[x_1, \cdots, x_n]$ and ordered set of polynomials $G = \{g_1, \cdots, g_m\}$ in R, write $f \equiv r \mod G$ if r is the remainder obtained by general polynomial division of $f \in R$ by $g_1, \cdots, g_m$.

Proposition 4.45 (Buchberger's Criterion) *Let $R = F[x_1, \cdots, x_n]$ and fix a monomial ordering on R. If $I = (g_1, \cdots, g_m)$ is a nonzero ideal in R, then $G = \{g_1, \cdots, g_m\}$ is a Gröbner basis for I if and only if $S(g_i, g_j) \equiv 0 \mod G$ for $1 \leq i < j \leq m$.*

Proof: If $\{g_1, \cdots, g_m\}$ is a Gröbner basis for I, then $S(g_i, g_j) \equiv 0 \mod G$ since each $S(g_i, g_j)$ is an element of I.

Conversely, if $S(g_i, g_j) \equiv 0 \mod G$ for $1 \leq i < j \leq m$ and take any element $f \in I$. To see that G is a Gröbner basis, we need ot see that $(\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m))$ contains $\mathrm{LT}(f)$. Since $f \in I$, we can write $f = \sum_{i=1}^{m} h_i g_i$ for some polynomials $h_1, \cdots, h_m$. Such a representation may not be unique. Among all such representations, choose one for which the largest multidegree of any summand (*i.e.*, $\max_{i=1,c\ldots,m} \partial(h_i g_i)$) is minimal, say α. It is clear that the

multidegree of f is no worse than the largest multidegree of all the summands $h_i g_i$, so $\partial(f) \leq \alpha$. Write

$$f = \sum_{i=1}^{m} h_i g_i = \sum_{\partial(h_i g_i)=\alpha} h_i g_i + \sum_{\partial(h_i g_i)<\alpha} h_i g_i \tag{4.1}$$

$$= \sum_{\partial(h_i g_i)=\alpha} \mathrm{LT}(h_i) g_i + \sum_{\partial(h_i g_i)=\alpha} (h_i - \mathrm{LT}(h_i)) g_i + \sum_{\partial(h_i g_i)<\alpha} h_i g_i. \tag{4.2}$$

Suppose that $\partial(f) < \alpha$. Then since the multidegree of the second two sums is also strictly smaller than α it follows that the multidegree of the first sum is strictly smaller than α. If $\alpha_i \in F$ denotes the constant coefficients of the monomial term $\mathrm{LT}(h_i)$, then $\mathrm{LT}(h_i) = a_i h_i'$ where h_i' is a monomial. We can apply Lemma 4.44 to $\sum a_i(h_i' g_i)$ to write the first sum above as $\sum b_i S(h_{i-1}' g_{i-1}, h_i' g_i)$ with $|partial(h_{i-1}' g_{i-1}) = \partial(h_i g_i) = \alpha$. Let $\beta_{i-1,i}$ be the multidegree of the monic least common multiple of $\mathrm{LT}(g_{i-1})$ and $\mathrm{LT}(g_i)$. Then an easy computation shows that $S(h_{i-1}' g_{i-1}, h_i' g_i)$ is just $S(g_{i-1}, g_i)$ multiplied by the monomial of multidegree $\alpha - \beta_{i-1,i}$. The polynomial $S_(g_{i-1}, g_i)$ has multidegree less than $\beta_{i-1,i}$ and, by assumption, $S(g_{i-1}, g_i) \equiv 0 \mod G$. This means that after general polynomial division of $S(g_{i-1}, g_i)$ by $g_1, \cdots, g_m$, each $S(g_{i-1}, g_i)$ can be written as a sum $\sum q_j g_j$ with $\partial(q_j g_j) < \beta_{i-1,i}$. It follows that each $S(h_{i-1}' g_{i-1}, h_i' g_i)$ isa sum $\sum q_j' g_j$ with $\partial(q_j' g_j) < \alpha$. But then all the sums on the right hand side of equation 4.2 can be written as a sume of terms of the form $p_i g_i$ with polynomials p_i satisfying $\partial(p_i g_i) < \alpha$. This contradicts the minimality of α and shows that in fact $\partial(f) = \alpha$, i.e., the leading term of f has multidegree α.

If we now take the terms in 4.2 of multidegree α, we see that

$$\mathrm{LT}(f) = \sum_{\partial(h_i g_i)=\alpha} LT(h_i) LT(g_i),$$

so $\mathrm{LT}(f) \in (\mathrm{LT}(g_1), \cdots, \mathrm{LT}(g_m))$. It follows that $G = \{g_1, \cdots, g_m\}$ is a Gröbner basis. □

4.6.2 Buchberger's Algorithm

Buchberger's Criterion can be used to provide an algorithm to find a Gröbner basis for an ideal I. If $I = (g_1, \cdots, g_m)$ and each $S(g_i, g_j)$ leaves a remainder of 0 when divided by $G = \{g_1, \cdots, g_m\}$ using general polynomial division then G is a Gröbner basis. Otherwise $S(g_i, g_j)$ has a nonzero remainder r. Increase G by appending the polynomial $g_{m+1} = r$, and repeat the algorithm with the set of generators $G' = \{g_1, \cdots, g_m, g_{m+1}\}$. Note that once an $S(g_i, g_j)$ yields a remainder of 0 after division by the polynomials in G it also yields a remainder of 0 when additional polynomials are appended to G.

If $\{g_1, \cdots, g_m\}$ is a Gröbner basis for the ideal I and $\mathrm{LT}(g_j)$ is divisible by $\mathrm{LT}(g_i)$ for some $j \neq i$, then $\mathrm{LT}(g_i)$ is not needed as generator for $\mathrm{LT}(I)$. By Proposition 4.43, we may delete g_j and still retain a Gröbner basis for I. We may also assume without loss of generality that the leading term of each g_i is monic.

Definition 4.46 *A Gröbner basis $\{g_1, \cdots, g_m\}$ for I where each $\mathrm{LT}(g_i)$ is monic and where $\mathrm{LT}(g_i)$ is not divisible by $\mathrm{LT}(g_i)$ for $i \neq j$ is called a* ***minimal Gröbner basis****.*

Fix a monomial order on $R = F[x_1, \cdots, x_n]$. A Gröbner basis $\{g_1, \cdots, g_m\}$ for the nonzero ideal I in R is called a ***reduced Gröbner basis*** *if*

1. *each g_i has monic leading term, i.e., $\mathrm{LT}(g_i)$ is monic, $i = 1, \cdots, m$, and*
2. *no term in g_j is divisible by $\mathrm{LT}(g_i)$ for $J \neq i$.*

Remark 4.46.1 *While a minimal Gröbner basis is not unique even with a fixed monomial ordering on $F[x_1, \cdots, x_n]$, the number of elements and their leading terms are unique.*

Remark 4.46.2 *Any reduced Gröbner basis is a minimal Gröbner basis. If $G = \{g_1, \cdots, g_m\}$ is a minimal Gröbner basis for I, then the leading term $\mathrm{LT}(g_j)$ is not divisible by $\mathrm{LT}(g_i)$ for $i \neq j$. As a result, if we use polynomial division to divide g_j by other polynomials in G, we obtain a remainder g_j' in the ideal I with the same leading term as g_j (the remainder does not depend on the order of the polynomials used in the division). Then replacing g_j by g_j' in G again gives a minimal Gröbner basis for I by Proposition 4.43. And in this basis no term of g_j' is divisible by $\mathrm{LT}(g_i)$ for any $I \neq j$. Replacing each element in G by its remainder after division by the other elements in G therefore results in a reduced Gröbner basis for I. In this case, the reduced Gröbner basis is unique for a given monomial ordering.*

Theorem 4.47 *Fix a monomial ordering on $R = F[x_1, \cdots, x_n]$. Then there is a unique reduced Gröbner basis for every nonzero ideal I in R.*

Proof: We know two reduced bases have the same number of elements and the same leading terms since they are also minimal basis. If $G = \{g_1, \cdots, g_m\}$ and $G' = \{g_1', \cdots, g_m'\}$ are two reduced bases for the same nonzero ideal I, then after a possible rearrangement we may assume $\mathrm{LT}(g_i) = \mathrm{LT}(g_i') = h_i$ for $i = 1, \cdots, m$. For any fixed i, consider the polynomial $f_i = g_i - g_i'$. If g_i is nonzero, then since $g_i \in I$, its leading term must be divisible by some h_j, by definition of a reduced basis,h_j for $j \neq i$ does not divide any of the terms in either g_i or g_i', hence does not divide $\mathrm{LT}(f_i)$ (suppose not, $\exists j$, s.t., $h_j | \mathrm{LT}(f_i)$, then at least one of g_i or g_i' contains a term with the same multidegree as $\mathrm{LT}(f_i)$, which is also divisible by h_j). But h_i also does not divide $\mathrm{LT}(f_i)$ since all terms in f_i has strictly smaller multidegree. This forces $f_i = 0$, i.e., $g_i = g_i'$ for every i, so $G = G'$. □

Corollary 4.47.1 *Let I and J be two ideals in $F[x_1, \cdots, x_n]$. Then $I = J$ if and only if I and J have the same reduced Gröbner basis with respect to any fixed monomial ordering on $F[x_1, \cdots, x_n]$.*

Proof: By the uniqueness of reduced Gröbner basis, if $I = J$, then they have the same reduced Gröbner basis with respect to any fixed monomial ordering of $F[x_1, \cdots, x_n]$. On the other hand, if they have the same generator, then the two ideal must be equal. □

Example:

1. Choose the lexicographic ordering $x > y$ on $F[x, y]$ and consider the ideal I generated by $f_1 = x^3y - xy^2 + 1$ and $f_2 = x^2y^2 - y^3 - 1$. To test whether $G = \{f_1, f_2\}$ is a Gröbner basis, we compute $S(f_1, f_2) = yf_1 - xf_2 = x + y$, which is its own remainder when divided by $\{f_1, f_2\}$, so G is not a Gröbner basis for I. Set $f_3 = x + y$, and increase the generating set $G' = \{f_1, f_2, f_3\}$. Now $S(f_1, f_2) \equiv 0 \mod G'$, and a brief computation yields

$$S(f_1, f_3) = f_1 - x^2yf_3 = -x^2y^2 - xy^2 + 1 \equiv 0 \mod G'$$

$$S(f_2, f_3) = f_2 - xy^2 f_3 = -xy^3 - y^3 - 1 \equiv y^4 - y^3 - 1 \mod G'$$

Let $f_4 = y^4 - y^3 - 1$ and increase the generating set to $G'' = \{f_1, f_2, f_3, f_4\}$. The previous 0 remainder is still 0, and now $S(f_2, f_3) \equiv 0 \mod G''$ and $S(f_1, f_4) \equiv S(f_2, f_4) \equiv S(f_3, f_4) \equiv 0 \mod G''$. So $\{f_1, f_2, f_3, f_4\}$ is a Gröbner basis for I. In particular, $\mathrm{LT}(I)$ is generated by the leading terms of these four polynomials so $\mathrm{LT}(I) = (x^3y, x^2y^2, x, y^4) = (x, y^4)$. Then $x + y$ and $y^4 - y^3 - 1$ in I have leading terms generating $\mathrm{LT}(I)$, so $\{x + y, y^4 - y^3 - 1\}$ gives a minimal Gröbner basis for I.

2. Choose the lexicographic ordering $y > x$ on $F[x, y]$ and consider the same ideal I as in the previous example. In this case, the Gröbner basis generated is

$$\{x^3y - xy^2 + 1, x^2y^2 - y^3 - 1, -x - y, -x^4 - x^3 + 1\}.$$

Here $\mathrm{LT}(I) = (-xy^2, -y^3, -y, -x^4) = (y, x^4)$. So $\{x + y, x^4 + x^3 - 1\}$ gives a minimal Gröbner basis for I with respect to this ordering.

4.6.3 Gröbner Basis and Solving Algebraic Equations

Suppose $S = \{f_1, \cdots, f_m\}$ is a collection of polynomials in n variables $x_1, \cdots, x_n$ and we are trying to find the solutions of the system of equations $f_1 = 0$, $f_2 = 0$, $\cdots$, $f_m = 0$ (i.e., the common set of zeros of the polynomials in S). If $(a_1, \cdots, a_n)$ is any solution to this system, then every element f of the ideal I generated by S also satisfies $f(a_1, \cdots, a_n) = 0$. On the other hand, if $S' = \{g_1, \cdots, g_s\}$ is any set of generator for the ideal I, then the set of solutions to the system $g_1 = 0, \cdots, g_s = 0$ is the same as the original solution set.

In the case where $f_1, \cdots, f_m$ are linear polynomials, a solution to the system to equations can be obtained by successively eliminating the variables $x_1, x_2, \cdots$ by elementary means (using Gaussian Elimination). It is much more complicated to solve nonlinear polynomial equitations, but the basic principle is the same. If there is a nonzero polynomial in the ideal I involving only one of the variables, say $p(x_n)$, then the last coordinate a_n is a solution of $p(x_n) = 0$. If now there is a polynomial in I involving only x_{n-1} and x_n, say $q(x_{n-1}, x_n)$ then the coordinate a_{n-1} would be a solution of $q(x_{n-1}, a_n) = 0$, etc. If we can successively find polynomials in I that eliminate the variables $x_1, x_2, \cdots$, then we will be able to determine all the solutions $(a_1, \cdots, a_n)$ to our original system of equations explicitly. Finding equations that follow from the system of equations in S, i.e., finding elements of the ideal I that do not involve some of the variables, is referred to as *elimination theory.*

Definition 4.48 (Elimination Ideal) *If I is an ideal in $F[x_1, \cdots, x_n]$, then $I_i = I \cap F[x_{i+1}, \cdots, x_n]$ is called the i^{th} **elimination ideal of** I with respect to the ordering $x_1 > \cdots > x_n$.*

Remark 4.48.1 *It is clear that I_i is an ideal of $F[x_{i+1}, \cdots, x_n]$ when considered as a subset of $F[x_{i+1}, \cdots, x_n]$.*

Proposition 4.49 (Elimination) *Suppose $G = \{g_1, \cdots, g_m\}$ is a Gröbner basis for the nonzero ideal I in $F[x_1, \cdots, x_n]$ with respect to the lexicographic monomial ordering $x_1 > \cdots > x_n$. Then $G \cap F[x_{i+1}, \cdots, x_n]$ is a Gröbner basis of the i^{th} elimination ideal $I_i = I \cap F[x_{i+1}, \cdots, x_n]$ of I. In particular, $I \cap F[x_{i+1}, \cdots, x_n] = 0$*

if and only if $G \cap F[x_{i+1}, \cdots, x_n] = \emptyset$.

Proof: Denote $G_i = G \cap F[x_{i+1}, \cdots, x_n]$. Then $G_i \subset I_i$, so by Proposition 4.43, we show that G_i is a Gröbner basis of I_i, it suffices to see that $\mathrm{LT}(G_i)$ generate $\mathrm{LT}(I_i)$ as an ideal in $\mathbb{F}[x_{i+1}, \cdots, x_n]$. Certainly $(\mathrm{LT}(G_i)) \subset \mathrm{LT}(I_i)$ as ideals in $F[x_{i+1}, \cdots, x_n]$. To show the reverse containment, let f be any element in I_i. Then $f \in I$ and since G is Gröbner basis for I we have

$$\mathrm{LT}(f) = a_1(x_1, \cdots, x_n)\,\mathrm{LT}(g_1) + \cdots + a_m(x_1, \cdots, x_n)\,\mathrm{LT}(g_m)$$

for some polynomials $a_1, \cdots, a_m \in \mathbb{F}[x_1, \cdots, x_n]$. Writing each polynomial a_i as a sum of monomial terms, we see that $\mathrm{LT}(f)$ is a sum of monomial terms of the form $ax_1^{s_1} \cdots x_n^{s_n}\,\mathrm{LT}(g_i)$. Since $\mathrm{LT}(f)$ involves only the variables $x_{i+1}, \cdots, x_n$, the sum of all such terms containing any of the variables $x_1, \cdots, x_i$ must be 0. It follows that $\mathrm{LT}(f)$ can be written as a $F[x_{i+1}, \cdots, x_n]$-linear combination of some monomial terms $\mathrm{LT}(g_t)$ where $\mathrm{LT}(g_t)$ does not involve the variables $x_1, \cdots, x_i$. But by the choice of the ordering, if $\mathrm{LT}(g_t)$ does not involve $x_1, \cdots, x_i$, then neither do any of the other terms ing_t, i.e., $g_t \in G_i$. Hence $\mathrm{LT}(f)$ can be written as a $F[x_{i+1}, \cdots, x_n]$-linear combination of elements $\mathrm{LT}(g_i)$, completing the proof. □

Example:

1. The ellipse $2x^2 + 2xy + y^2 - 2x - 2y = 0$ intersects the circle $x^2 + y^2 = 1$ in two points. To find them we compute a Gröbner basis for the ideal $I = (2x^2 + 2xy + y^2 - 2x - 2y, x^2 + y^2 - 1) \subset \mathbb{R}[x, y]$, using the lexicographic monomial order $x > y$ to eliminate x. Such basis is $g_1 = 2x + y^2 + 5y^3 - 2$ and $g_2 = 5y^4 - 4y^3$. Hence $ry^4 = 4y^3$ and $y = 0$ or $y = 4/5$. Substituting these values into $g_1 = 0$, and solving for x we find the two intersection points are $(1, 0)$ and $(-3/5, 4/5)$.
   Instead using the lexicographic monomial order $y > x$ to eliminate y results in the Gröbner basis $\{y^2 + x^2 - 1, 2yx - 2y + x^2 - 2x + 1, 5x^3 - 7x^2 - x + 3\}$. Then we get $x = 1$ or $-3/5$. Hence we obtain the same solutions as before.

2. Consider the solutions in $\mathbb{C}$ to the system of two equations

$$f_1 : x^3 - 2xy + y^3 = 0 \text{ and } f_2 : x^5 - 2x^2y^2 + y^5 = 0.$$

   Computing a Gröbner basis for the ideal generated by f_1 and f_2 with respect to the lexicographic monomial order $x > y$, we obtain the basis

$$\begin{aligned} g_1 &= x^3 - 2xy + y^3 \\ g_2 &= 200xy^2 + 193y^9 + 158y^8 - 45y^7 - 456y^6 + 50y^5 - 100y^4 \\ g_3 &= y^10 - y^8 - 2y^7 + 2y^6 \end{aligned}$$

   Any solution to our original equations would satisfy $g_1 = g_2 = g_3 = 0$. Since $g_3 = y^6(y-1)^2(y^2 + 2y + 2)$, we have $y = 0, 1$ or $-1 \pm i$. Since $g_1(x, 0) = x^3$ and $g_2(x, 0) = 0$, we see that $(0, 0)$ is the only solution with $y = 0$. Similarly, for the other solutions of y, we have all the solutions are given by

$$(x, y) = (0, 0),\ (1, 1),\ (-1 + i, -1 - i),\ \text{or } (-1 - i, -1 + i).$$

Lemma 4.50 *If I and J are two ideals of ring R, such that $I = (f_1, \cdots, f_s)$ and $J = (h_1, \cdots, h_t)$, then*

$$I + J = (f_1, \cdots, f_s, h_1, \cdots, h_t) IJ \qquad = (f_1 h_1, \cdots, f_j h_j, \cdots, f_s h_t).$$

Proof: Clear. □

Proposition 4.51 *If I and J are two ideals in $F[x_1, \cdots, x_n]$, then $(t)I + ((1-t))J$ is an ideal in $F[t, x_1, \cdots, x_n]$ and $I \cap J = ((t)I + ((1-t))J) \cap F[x_1, \cdots, x_n]$. In particular, $I \cap J$ is the first elimination ideal of $(t)I + ((1-t))J$ with respect to the ordering $t > x_1 > \cdots > x_n$.*

Proof: Firstly, $(t)I$ and $((1-t))J$ are clearly ideals in $F[x_1, \cdots, x_n, t]$, so must their also sum $((t)I + ((1-t))J)$.

If $f \in I \cap J$, then $f = tf + (1-t)f$ shows $I \cap J \subset (((t)I + ((1-t))J)) \cap F[x_1, \cdots, x_n]$.

Conversely, the generator of $((t)I + ((1-t))J)$ are $\{tf_1 + (1-t)f_2 \mid f_1 \in I,\ f_2 \in J\}$, so suffices to check the generators. Suffices to check that all the generators are of $((t)I + ((1-t))J)$ that is also in $F[x_1, \cdots, x_n]$ is in $I \cap J$. Suppose $f = tf_1 + (1-t)f_2$ is an element of $F[x_1, \cdots, x_n]$, where $f_1 \in I$ and $f_2 \in J$. Then $t(f_1 - f_2) = f - f_2 \in F[x_1, \cdots, x_n]$ shows that $f_1 - f_2 = 0$ and $f = f_2$. So $f = f_1 = f_2 \in I \cap J$. Since $I \cap J = (tI + (1-t)J) \cap F[x_1, \cdots, x_n]$, $I \cap J$ is the first elimination ideal of $tI + (1-t)J$ with respect to the ordering $t > x_1 > \cdots > x_n$. □

Example:
Let $I = (x, y)^2 = (x^2, xy, y^2)$ and let $J = (x)$. For the lexicographic monomial ordering $t > x > y$, the reduced Gröbner basis for $tI + (1-t)J$ in $F[t, x, y]$ is $\{tx - x, ty^2, x^2, xy\}$ and so $I \cap J = (x^2, xy)$.

5 Introduction To Module Theory

5.1 Basic Definition

Definition 5.1 (Modules) *Let R be a ring (not necessarily commutative nor with 1). A **left R-module or left module over** R is a set M together with*

1. *a binary operation $+$ on M under which M is an abelian group, and*
2. *an operation of R on M (that is, a map $R \times M \to M$) denoted by rm (also called scalar multiplication), for all $r \in R$ and for all $m \in M$, which satisfies*
   - *$(r+s)m = rm + sm$, for all $r, s \in R$, $m \in M$;*
   - *$(rs)m = r(sm)$, for all $r, s \in R$, $m \in M$, and*
   - *$r(m+n) = rm + rn$, for all $r \in R$, $m, n \in M$;*
   - *if the ring R has a 1, we impose the additional axiom*

$$1m = m, \ \ \forall m \in M.$$

   *If this is the case, the M is called a **unital module**. In this and coming section, we will always work with unital modules.*

*We define the **right R-module** in the analogous way.*

Remark 5.1.1 *If the ring R is commutative and M is a left R-module, we can make M into a right R-module by defining $mr = rm$, for all $m \in M$ and $r \in R$. In this and the coming sections, we will always work with left modules.*

Remark 5.1.2 *If R is a field F, then the axioms for an R-module are precisely the same as those for a vector space over F, so that modules over a field F and vector spaces over F are the same.*

Definition 5.2 (Submodules) *Let R be a ring and let M be an R-module. An **R-submodule** of M is a subgroup N of M which is closed under the action of ring elements, i.e., $rn \in N$, for all $r \in R$ and $n \in N$, i.e., submodules of M are subsets of M which are themselves modules under the restricted operations.*

Remark 5.2.1 *If R is a field, then submodules are the same as subspaces.*

Example:

1. Let R be any ring. Then $M = R$ is a left R-module, where the action of a ring element on a module element is just the usual multiplication in the ring. In particular, every field can be considered as a 1-dimensional vector space over itself. When R is considered as a left module over itself in this fasion, the submodules of R are precisely the left ideals of R.
2. Suppose M is any R-module, then $\{0\}$ and M are always R-submodules of M.

3. Let R be a ring with 1 and let $n \in \mathbb{Z}^+$. Define

$$R^n = \{(a_1, a_2, \cdots, a_n) \,|\, a_i \in R, 1 \leq i \leq n\}$$

Define scalar multiplication and addition in the natural way, then R^n is an R-module. The module R^n is called the *free module of rank n over R*. The set of n-tuples with arbitrary ring elements in the i^{th} component and zeros in the j^{th} component for all $j \neq i$ forms a submodule of R^n.

4. If M is an R-module, and S is a subring of R with $1_S = 1_R$, then M is an S-module.

5. If M is an R-module and for some ideal I of R, $am = 0$, $\forall a \in I$ and $m \in M$. We say M is *annihilated by I*. In this situation, we can make M into an (R/I)-module by defining an action of the quotient ring R/I on M by $(r + I)m = rm$.

6. Let $R = \mathbb{Z}$, let A be any abelian group and write the operation of A as $+$. Make A into a $\mathbb{Z}$-module as follows: for any $n \in \mathbb{Z}$ and $a \in A$ define

$$na = \begin{cases} a + \cdots + a \ (n \text{ times}) & \text{if } n > 0 \\ 0_A & \text{if } n = 0 \\ -a - \cdots - a \ (-n \text{ times}) & \text{if } n < 0 \end{cases}.$$

This action makes A into a $\mathbb{Z}$-module; if fact this is the only possible action of $\mathbb{Z}$ on A making it a $\mathbb{Z}$-module. Thus every abelian group is a $\mathbb{Z}$-module. Conversely, by definition of a module, if M is any $\mathbb{Z}$-module, then M must be an abelian group. So $\mathbb{Z}$-modules are the same as abelian groups. Furthermore, it is immediate that $\mathbb{Z}$-submodules are the same as subgroups. If A has order m, then by Lagrange's Theorem, $mx = 0$ for all $x \in A$, then A is a module over $\mathbb{Z}/m\mathbb{Z}$. In particular, if p is a prime and A is an abelian group such that $px = 0$ for all $x \in A$, then A is a $\mathbb{Z}/p\mathbb{Z}$-module. These groups are the *elementary abelian p-groups*.

7. Let $\mathbf{V}$ be a vector space over F and let T be a linear transformation from $\mathbf{V}$ to $\mathbf{V}$. Then we can define $\mathbf{V}$ to be an $F[x]$-module: if

$$p(x) = \sum_{i=0}^{n} a_n x^n$$

where $a_0, \cdots, a_n \in F$. Then for each $v \in V$, define an action of the ring element $p(x)$ in the module element v by

$$p(x)v = \left(\sum_{i=0}^{n} a_n T^n \right)(v)$$

This in fact describes all $F[x]$-modules on a F-vector space $\mathbf{V}$; this is because if $\mathbf{V}$ is any $F[x]$-module, then $\mathbf{V}$ is an F-module and the action of the ring element x on $\mathbf{V}$ is a linear transformation from $\mathbf{V}$ to $\mathbf{V}$. Then the axioms for a module enforces the module structure on $\mathbf{V}$ to be defined in the above manner. Thus there is a bijection between the collection of $F[x]$-modules and the collection of pairs $\mathbf{V}$, T. With some extra work, one can show that the $F[x]$-submodules of $\mathbf{V}$ are precisely the T-invariant subspaces of $\mathbf{V}$.

Proposition 5.3 (The Submodule Criterion) *Let R be a ring with unity and M be an R-module. A subset N of M is a submodule of M if and only if*

- $N \neq \emptyset$, *and*
- $x + ry \in N$ *for all* $r \in R$ *and for all* $x, y \in N$.

Proof: $\Rightarrow$: this direction is clear. $\Leftarrow$: Suppose the two conditions holds, let $r = -1$ and apply the subgroup criterion, we get N is a subgroup of M. In particular $0 \in N$. Now let $x = 0$, then we have that $ry \in N$ for every $r \in R$ and $y \in N$, so N is closed under scalar multiplication. □

Corollary 5.3.1 *The arbitrary intersection of any nonempty collection of submodules is a submodule. The union of an ascending chain of submodules is a submodule.*

Definition 5.4 (Torsion) *An element m of the R-module M is called a* ***torsion element*** *if $rm = 0$ for some nonzero element $r \in R$. The set of torsion element is denoted* $\mathrm{Tor}(M)$.

Lemma 5.5 *If R is an integral domain, then* $\mathrm{Tor}(M)$ *is a submodule of M.*

Remark 5.5.1 *The condition that R is an integral domain is necessary, as for otherwise, note that R is an R-module.*

Proof: Firstly $\mathrm{Tor}(M)$ is non-empty, since $0_M \in \mathrm{Tor}(M)$. Now let $x, y \in \mathrm{Tor}(M)$ and $r \in R$. One can easily see that $x + ry \in \mathrm{Tor}(M)$. □

Definition 5.6 (Algebra) *Let R be an commutative ring with identity. An* ***R-algebra*** *is a ring A with identity together with a ring homomorphism $f : R \to A$ mapping 1_R to 1_A such that the subring $f(R)$ of A is contained in the center of of A.*

Remark 5.6.1 *If A is an R-algebra, then it is easy to check that A has a natural left an right (unital) R-module structure defined by $r \cdot a = a \cdot r = f(r)a$. It is possible for an R-algebra A to have other left or right R-module structures, but unless otherwise stated, this natural module structure on an algebra will be assumed.*

Definition 5.7 (Algebra Homomorphism And Isomorphism) *If A and B are two R-algebras, an* ***R-algebra homomorphism (isomorphism, resp.)*** *is a ring homomorphism (isomorphism, resp) $\varphi : A \to B$ mapping 1_A to 1_B such that $\varphi(r \cdot a) = r \cdot \varphi(a)$ for all $r \in R$ and $a \in A$.*

Example:
Let R be a commutative ring with 1.

1. Any ring with identity is a $\mathbb{Z}$-algebra.
2. For any ring A with identity, if R is a subring of the center of A containing the identity of A then A is an R-algebra. In particular, a commutative ring A containing 1 is an R-algebra for any subring R of A containing 1. For example, the polynomial ring $R[x]$ is an R-algebra, the polynomial ring over R in any number of variables is an R-algebra, and the group ring RG for a finite group G is an R-algebra.

3. If A is an R-algebra then the R-module structure of A depends only on the subring $f(R)$ contained in the center of A. If we replace R by its image $f(R)$, we see that up to a ring homomorphism, every algebra A arises from a subring of the center of A that contains 1_A.

4. A special case of the previous example occurs when $R = F$ is a field. In this case F is isomorphic to its image under f, so we can identify F itself as a subring of A. Hence, saying that A is an algebra over a field F is the same as saying that the ring A contains the field F in its center and the identity of A and of F are the same.

5.2 Quotient Modules and Module Homomorphisms

Definition 5.8 *Let R be a ring and let M and N be R-modules.*

- *A map $\varphi : M \to N$ is an **R-module homomorphism** if it respects the R-module structures of M and N, i.e.,*
  1. $\varphi(x + y) = \varphi(x) + \varphi(y)$, $\forall x, y \in M$ *and*
  2. $\varphi(rx) = r\varphi(x)$, $\forall r \in R$, $x \in M$.
- *An R-module homomorphism is an **isomorphism** if it is both injective and surjective. The modules M and N said to be **isomorphic** denoted $M \cong N$, if there is some R-module isomorphism $\varphi : M \to N$.*
- *If $\varphi : M \to N$ is an R-module homomorphism, let $\ker \varphi = \{m \in M \mid \varphi(m) = 0\}$ and let $\varphi(M) = \{\varphi(m) \mid m \in M\}$.*
- *Let M and N be R-modules and define $\mathrm{Hom}_R(M, N)$ to be the set of all R-module homomorphism from M into N.*

Remark 5.8.1 *If R is a field, then R-module homomorphism are linear transformations.*

Lemma 5.9 *Let $\varphi : M \to N$ be an R-module homomorphism, then*

1. *$\ker \varphi$ is an submodule of M. Moreover, $\varphi^{-1}(n) = m + \ker \varphi$, where $\varphi(m) = n$.*
2. *$\mathrm{Im}\, \varphi$ is an submodule of N. In fact, the image of any submodule of M under φ is a submodule of N.*

Proposition 5.10 *Let M, N and L be R-modules.*

1. *A map $\varphi : M \to N$ is an R-module homomorphism if and only if*

$$\varphi(rx + y) = r\varphi(x) + \varphi(y), \forall x, y \in M, r \in R.$$

2. *Let φ, ψ be elements of $\mathrm{Hom}_R(M, N)$. Define $\varphi + \psi$ by*

$$(\varphi + \psi)(m) = \varphi(m) + \psi(m), \quad \forall m \in M$$

Then $\varphi + \psi \in \mathrm{Hom}_R(M, N)$ and with this operation $\mathrm{Hom}_R(M, N)$ is an abelian group. If R is a commu-

tative ring then for $r \in R$ define $r\varphi$ by

$$(r\varphi)(m) = r(\varphi(m)), \ \ \forall m \in M.$$

Then $r\varphi \in \mathrm{Hom}_R(M,N)$ and with this action of the commutative ring R the abelian group $\mathrm{Hom}_R(M,N)$ is an R-module.

3. *If $\varphi \in \mathrm{Hom}_R(L,M)$ and $\psi \in \mathrm{Hom}_R(M,N)$, then $\psi \circ \varphi \in \mathrm{Hom}_R(L,N)$.*

4. *With addition as above and multiplication defined as function composition, $\mathrm{Hom}_R(M,M)$ is a ring with 1. When R is commutative $\mathrm{Hom}_R(M,M)$ is an R-algebra.*

Proof:

1. This is clear.

2. It is clear that $\mathrm{Hom}_R(M,N)$ is an abelian group as N must be an abelian group. If R is commutative, then it is also clear that $\mathrm{Hom}_R(M,N)$ is an R-module.

3. Let φ and ψ be as given and let $r \in R$, $x, y \in L$. Then

$$\begin{aligned}(\psi \circ \varphi)(rx+y) &= \psi(\varphi(rx+y)) \\ &= \psi(r\varphi(x) + \varphi(y)) \\ &= r\psi(\varphi(x)) + \psi(\varphi(y)) \\ &= r(\psi \circ \varphi)(x) + (\psi \circ \varphi)(y)\end{aligned}$$

   so $\psi \circ \varphi$ is an R-module homomorphism.

4. Note that since the domain and codomain of the elements of $\mathrm{Hom}_R(M,M)$ are the same, function composition is well-defined. By (3), it is a binary operation on $\mathrm{Hom}_R(M,M)$ and it is associative. The identity function is the multiplicative identity of the ring. When R is commutative, then (2) shows that the ring $\mathrm{Hom}_R(M,M)$ is a left R-module and defining $\varphi r = r\varphi$ (abusing notation, denoting r to be the map rId) for all $\varphi \in \mathrm{Hom}_R(M,M)$ and $r \in R$ makes $\mathrm{Hom}_R(M,M)$ into an R-algebra.

□

Definition 5.11 (Endomorphism Ring) *The ring $\mathrm{Hom}_R(M,M)$ is called the **endomorphism ring of** M and will often be denoted by $\mathrm{End}_R(M)$, or just $\mathrm{End}(M)$. Elements of $\mathrm{End}(M)$ are called **endomorphism**.*

Remark 5.11.1 *If R is a field, then the map $r \mapsto rI$ is injective and the copy of R in $\mathrm{End}_R(M)$ is called the **scalar transformations**.*

Lemma 5.12 *Let A be any $\mathbb{Z}$-module, then $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n, A) \cong A_n$, where $A_n = \{a \in A : na = 0\}$, i.e., A_n is the annihilator of A of the ideal (n) of $\mathbb{Z}$.*

Proof: Firstly, note that if $a \in A_n$, then $\varphi(\bar{k}) = ka$ defines a $\mathbb{Z}$-module homomorphism from $\mathbb{Z}/n\mathbb{Z} \to A$. In fact, this is a $\mathbb{Z}$-module homomorphism if and only if $na = 0$. Next, if φ and ψ are elements of $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n, A)$ corresponding to a and b in A_n respectively, then we can see that $\varphi + \psi$ corresponds to $a + b \in A_n$. □

Proposition 5.13 $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}/m\mathbb{Z}) \cong \mathbb{Z}/(n,m)\mathbb{Z}$ *and* $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}) \cong \{0\}$.

Proof: Using the previous lemma, the second half of the statement follows immediately. For the first half, it suffices to find all the elements of $\mathbb{Z}/m\mathbb{Z}$ that is annihilated by (n). Note if k is such an element, then $m|nk$, so $\frac{m}{(n,m)}|k$. Then the multiples of $\frac{m}{(n,m)}$ are all the possible elements, and with brief verification, they are indeed annihilated by (n). Notice there are exactly (n, m) distinct such multiplies in $\mathbb{Z}/m\mathbb{Z}$, written additively, we conclude that this submodule is isomorphic to $\mathbb{Z}/n\mathbb{Z}$. □

Lemma 5.14 *Let R be a commutative ring, and let A, B and M be R-modules. Then*

- $\mathrm{Hom}_R(A \times B, M) \cong \mathrm{Hom}_R(A, M) \times \mathrm{Hom}_R(B, M)$*;*
- $\mathrm{Hom}_R(M, A \times B) \cong \mathrm{Hom}_R(M, A) \times \mathrm{Hom}_R(M, B)$*.*

Proposition 5.15 *Let R be a ring, M be an R-module and N be a submodule of M. The (additive, abelian) quotient group M/N can be made into an R-module by defining an action of elements of R by*

$$r(x + N) = (rx) + N, \ \ \forall r \in R, \ x + N \in M/N.$$

The natural projection map $\pi : M \to M/N$ defined by $\pi(x) = x + N$ is an R-module homomorphism with kernel N.

Proof: Since M is an abelian group under $+$, the quotient group M/N is defined and is an abelian group. To see the action of the ring element r on the coset $x + N$ is well defined, suppose $x + N = y + N$, i.e., $x - y \in N$. Since N is an R-submodule, $r(x - y) \in N$. Thus $rx - ry \in N$ and $rx + N = ry + N$, as desired. Now since the operation in M/N is "compatible" with those of M, the axioms for an R-module holds for M/N.

Lastly, the natural projection map π is a group homomorphism with kernel N. It remains only to show π is a module homomorphism, i.e., $\pi(rm) = r\pi(m)$. But it is clear that

$$\begin{aligned} \pi(rm) &= rm + N \\ &= r(m + N) \\ &= r\pi(m). \end{aligned}$$

which completes the proof. □

Definition 5.16 (Sum Of Submodules) *Let A, B be submodules of the R-module M. The **sum** of A and B is the set*

$$A + B = \{a + b \,|\, a \in A,\, b \in B\}.$$

Remark 5.16.1 *One can check that the sum of two submodules A and B is a submodule and is the smallest submodule which contains both A and B.*

Theorem 5.17 (Isomorphism Theorems)

1. *(The First Isomorphism Theorem for Modules) Let M, N be R-modules and let $\varphi : M \to N$ be an R-module homomorphism. Then $\ker\varphi$ is a submodule of M and $M/\ker\varphi \cong \varphi(M)$.*
2. *(The Second Isomorphism Theorem for Modules) Let A, B be submodules for the R-module M. Then $(A+B)/B \cong A/(A\cap B)$.*
3. *(The Third Isomorphism Theorem for Modules) Let M be an R-module, and let A and B be submodules of M with $A \subset B$. Then $(M/A)/(B/A) \cong M/B$.*
4. *(The Fourth Isomorphism Theorem for Modules) Let N be a submodule of the R-module M. There is a bijection between the submodules of M which contain N and the submodules of M/N. The correspondence is given by $A \leftrightarrow A/N$, for all $A \subseteq N$. This correspondence commutes with the processes of taking sums and intersections.*

Lemma 5.18 *Let $A_1, A_2, \cdots, A_n$ be R-modules and let B_i be a submodule of A_i for each $i = 1, 2, \cdots$. Then*

$$(A_1 \times \cdots \times A_n)/(B_1 \times \cdots \times B_n) \cong (A_1/B_1) \times \cdots \times (A_n/B_n).$$

5.3 Generation of Modules, Direct Sums, Free Modules

Let R be a ring with 1 in this section.

Definition 5.19 (Sum of Modules) *Let M be an R-module and let $N_1, \cdots, N_n$ be submodules of M, then*

- *the **sum** of $N_1, \cdots, N_n$ is the set of all finite sums of elements from the set N_i: $\{a_1 + a_2 + \cdots + a_n \,|\, a_i \in N_i, 1 \le i \le n\}$. Denote this sum by $N_1 + \cdots + N_n$.*
- *For any subset A of M let*

$$RA = \{r_1a_1 + r_2a_2 + \cdots + r_ma_m \,|\, r_1, \cdots, r_m \in R,\ a_1, \cdots, a_m \in A,\ m \in \mathbb{Z}^+\}.$$

  *If A is the finite set $\{a_1, \cdots, a_n\}$, we write $Ra_1 + \cdots + Ra_n$ for RA (in particular, if $A = \emptyset$, then $RA = \{0\}$). We call RA the **submodule of M generated by A**. If N is a submodule of M, then $N = RA$ for some subset A of M, we call A a **set of generators or generating set for N**, and we say N is **generated** by A.*
- *A submodule N of M is **finitely generated** if there is some finite subset A of M such that $N = RA$, that is, if N is generated by some finite subset. If this is the case, then there is a smallest nonnegative integer d such that N is generated by d elements. Any generating set consisting of d elements will be called a **minimal set of generators for N**. If N is not finitely generated, it need not have a minimal generating set.*

- *A submodule N of M is **cyclic** if there exists an element $a \in M$ such that $N = Ra$, that is, if N is generated by one element.*

Remark 5.19.1 *One can verify that RA is indeed a submodule of M and is the smallest submodule that contains A. In particular, for submodules $N_1, \cdots, N_n$ of M, $N_1 + \cdots + N_n$ is just the submodule generated by the set $N_1 \cup \cdots \cup N_n$, and is the smallest submodule of M containing N_i for all i.*

Definition 5.20 (Direct Product) *Let $M_1, \cdots, M_k$ be a collection of R-modules. The collection of k-tuples $(m_1, \cdots, m_k)$ where $m_i \in M_i$ with addition and action of R defined componentwise is called the **direct product/ (external) direct sum** of $M_1, \cdots, M_k$, denoted $M_1 \times \cdots \times M_k$ or $M_1 \oplus \cdots \oplus M_k$.*

Remark 5.20.1 *The direct product of a collection of R-modules is again an R-module. Direct product and direct sums are canonically isomorphic when the collection is finite. However when the collection is infinite, then direct products and direct sums are different in general.*

Proposition 5.21 *Let $N_1, \cdots, N_k$ be submodules of the R-module M. Then the following are equivalent:*

1. *The map $\pi : N_1 \times N_2 \times \cdots \times N_k \to N_1 + N_2 + \cdots + N_k$ defined by*

$$\pi(a_1, \cdots, a_k) = a_1 + \cdots + a_k$$

   is an isomorphism (of R-modules): $N_1 + \cdots + N_k \cong N_1 \times \cdots \times N_k$.

2. *$N_j \cap (N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = \{0\}$ for all $j \in \{1, \cdots, k\}$.*

3. *Every $x \in N_1 + \cdots + N_k$ can be written uniquely in the form $a_1 + \cdots + a_k$ with $a_i \in N_i$.*

Proof: Analogous to the case of direct sums of vector subspaces. □

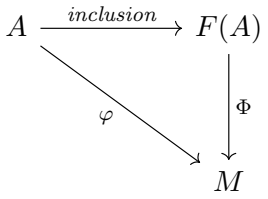
Definition 5.22 ((Internal) Direct Sum) *If an R-module $M = N_1 + N_2 + \cdots + N_k$ is the sum of submodules $N_1, \cdots, N_k$ of M satisfying the equivalent conditions of Proposition 5.21, then M is said to be the **(internal) direct sum** of $N_1, \cdots, N_k$, written*

$$M = N_1 \oplus \cdots \oplus N_k.$$

Definition 5.23 (Free Modules) *An R-module F is said to be **free** on the subset A of F if for every nonzero element x of F, there exist unique nonzero elements $r_1, r_2, \cdots, r_n$ of R and unique $a_1, a_2, \cdots, a_n$ in A such that $x = r_1a_1 + r_2a_2 + \cdots + r_na_n$ for some $n \in \mathbb{Z}^+$. In this situation we say A is a **basis** or set of **free generators** for F. If R is a commutative ring, the cardinality of A is called the **rank of** F.*

Theorem 5.24 (Universal Property of Free Modules) *5.33 For any set A there is a free R-module $F(A)$ on the set A and $F(A)$ satisfies the following universal property: if M is any R-module and $\varphi : A \to M$ is any map of sets, then there is a unique R-module homomorphism $\Phi : F(A) \to M$ such that $\Phi(a) = \varphi(a)$, for all*

$a \in A$, that is, the following diagram commute



When A is the finite set $\{a_1, \cdots, a_n\}$, $F(A) = Ra_1 \oplus \cdots \oplus Ra_n \cong \mathbb{R}^n$.

Proof: Let $F(A) = \{0\}$ if $A = \emptyset$. If A is nonempty, let $F(A)$ be the collection of all set functions $f : A \to R$ such that $f(a) = 0$ for all but finitely many $a \in A$. Make $F(A)$ into an R-module by pointwise addition of functions and pointwise multiplication of a ring element times a function, i.e.,

$$(f + g)(a) = f(a) + g(a)$$
$$(rf)(a) = r(f(a))$$

for all $a \in A$, $r \in R$ and $f, g \in F(A)$. It is easy to check that all the R-module axioms hold. Identify A as a subset of $F(A)$ by $a \mapsto f_a$, where f_a is the function which is 1 at a and zero elsewhere. We can, in this way, think of $F(A)$ as all finite R-linear combinations of elements of A by identifying each function f with the sum $r_1a_1 + \cdots + r_na_n$, where f takes on the value r_i at a_i and is zero at all other elements of A. Moreover, each element of $F(A)$ has a unique expression as such a formal sum. To establish the universal property of $F(A)$, suppose $\varphi : A \to M$ is a map of the set A into the R-module M, define $\Phi : F(A) \to M$ by

$$\Phi : \sum_{i=1}^{n} r_i a_i \mapsto \sum_{i=1}^{n} r_i \varphi(a_i)$$

By the uniqueness of the expression for the elements of $F(A)$ as a linear combinations of the a_i, we see easily that Φ is a well defined R-module homomorphism. By definition, the restriction of Φ to A equals φ. Finally since $F(A)$ is generated by A, once we know the values of an R-module homomorphism on A, its values on every element of $F(A)$ are uniquely determined. So Φ is the unique extension of φ to all of $F(A)$.
Lastly, when A is the finite set $\{a_1, \cdots, a_n\}$, $F(A) = Ra_1 \oplus \cdots \oplus Ra_n$ by Proposition 5.21 (3); since $R \cong Ra_i$ for all i (under the map $r \mapsto ra_i$), then the direct sum is isomorphic to R^n by Proposition 5.21 (1). □

Remark 5.24.1 *When $R = \mathbb{Z}$, the free module on a set A is called the **free abelian group on** A. If $|A| = n$, $F(A)$ is called the free abelian group of rank n and is isomorphic to $\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$ (n times).*

Corollary 5.24.1

1. *If F_1 and F_2 are free modules on the same set A, there is a unique isomorphism between F_1 and F_2 which is the identity map on A.*
2. *If F is any free R-module with basis A, then $F \cong F(A)$. In particular, F enjoys the same universal property with respect to A as $F(A)$ does in Theorem 5.33.*
3. *If F is a free R-module with basis A, then to define an R-module homomorphism from F into other R-modules, it suffices to specify their values on the elements of A; that is, the homomorphism is uniquely determined by the function value of the generators.*

Theorem 5.25 *Suppose R is commutative, then $R^n \cong R^m$ if and only if $n = m$.*

Lemma 5.26 *If M is a finitely generated R-module that is generated by n elements then every quotient of M may be generated by n (or fewer) elements.*

Proof: Suppose $a_1, \cdots, a_n$ generate M, then clearly $a_1 + N, \cdots, a_n + N$ generate M/N, for any submodule N of M. □

Corollary 5.26.1 *The quotients of cyclic modules are cyclic.*

Lemma 5.27 *Let N be a submodule of M, then if both M/N and N are finitely generated then so is M.*

Proof: Let $a_1, \cdots, a_n$ generate N and $b_1 + N, \cdots, b_m + N$ generate M/N, then the set $a_1, \cdots, a_n, b_1, \cdots, b_m$ generates M. □

Definition 5.28 (Torsion Module) *An R-module M is called a **torsion module** if for each $m \in M$ there is a nonzero $r \in \mathbb{R}$ such that $rm = 0$, where r may depend on m.*

Definition 5.29 (Irreducible Modules) *An R-module M is called **irreducible** if $M \neq 0$ and if 0 and M are the only submodules of M.*

Lemma 5.30 *M is irreducible if and only if $M \neq 0$ and M is a cyclic module with any nonzero element as generator.*

Proof: Suppose M is irreducible, then if exists an nonzero element $m \in M$ such that m does not generate M, then Rm is a proper submodule of M. Conversely, if M is a cyclic module with any nonzero element as a generator. Suppose towards a contradiction M is irreducible, then let N be any of its proper submodules. If $n \in N$ is nonzero, then as $Rm \subset N$, we must have that $M \subset N$ which is a contradiction. Hence N must be the zero submodule. □

Lemma 5.31 *If M_1 and M_2 are irreducible R-modules, then any nonzero R-module homomorphism from M_1 to M_2 is an isomorphism.*

Corollary 5.31.1 (Schur's Lemma) *If M is irreducible , then $\mathrm{End}_R(M)$ is a division ring.*

5.4 Tensor Products of Modules

In this section, the ring S will always have a 1, and if R is is a subring of the ring S, then we always assume that $1_R = 1_S$ (this ensures that S is a unital R-module).

If N is a left S-module, then N is naturally a left R-module. If $f : R \to S$ is a ring homomorphism from R into S with $f(1_R) = 1_S$, then it is easy to see that N can be considered as an R-module with $rn = f(r)n$ for $r \in R$ and $n \in N$. In this situation S can be considered as an extension of the ring R and the resulting R-module is said to

be obtained from N by restriction of scalars from S to R. Now we try to reverse this process by considering an arbitrary R-module N, how can we embedded into an S-module M by module homomorphism.

If the R-module N were already an S-module, then there is no difficulty "extending" the scalars from R to S, so we consider whether it is possible to define "products" of the form sn for $s \in S$ and $n \in N$. That is, we want to define a map $S \times N$ to N such that it satisfies the S-module axioms, where the image of the pairs (s, n) is denoted by sn. It is therefore natural to consider the free $\mathbb{Z}$-module on the set $S \times N$, i.e., the collection of all finite commuting sums of elements of the form (s_i, n_i), where $s_i \in S$ and $n_i \in N$. Now since N is an R-module, and in order for the defined action of S on N to satisfies the axioms of an S-module, we need

$$(s_1 + s_2)n = s_1 n + s_2 n \tag{5.1}$$
$$s(n_1 + n_2) = sn_1 + sn_2 \tag{5.2}$$
$$(sr)n = s(rn) \tag{5.3}$$
$$\tag{5.4}$$

for all $s_1, s_2, s \in S$, $r \in R$ and $n_1, n_2, n \in N$. So we must take the quotient of the abelian group by the subgroup H generated by all elements of the form

$$(s_1 + s_2, n) - (s_1, n) - (s_2, n)$$
$$(s, n_1 + n_2) - (s, n_1) - s(n_2)$$
$$(sr, n) - (s, rn)$$

for all $s_1, s_2, s \in S$, $r \in R$ and $n_1, n_2, n \in N$, where rn is the last element refers to R-module structures already defined on N. The resulting quotient group is denoted by $S \otimes_R N$ or $S \otimes N$.

Definition 5.32 (Tensor Product) *Let R be a submodule of S with $1_S = 1_R$. If N is an R-module that is also an S module. Then the quotient group $S \otimes_R N$ is called the **tensor product of S and N over r**. The elements of $S \otimes_R N$ are called **tensors** and can be written (non-uniquely in general) as finite sums of the form $s \otimes n$, with $s \in S$, $n \in N$, where $s \otimes n$ denotes the cosets containing (s, n).*

Remark 5.32.1 *By definition of the quotient, we have forced the relations*

$$(s_1 + s_2) \otimes n = s_1 \otimes n + s_2 \otimes n,$$
$$s \otimes (n_1 + n_2) = s \otimes n_1 + s \otimes n_2,$$
$$sr \otimes n = s \otimes rn.$$

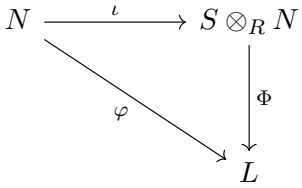
Hence we can define the tensor product $S \otimes_R N$ to be a left S-module in the natural way:

$$s\left(\sum_{finite} s_i \otimes n_i\right) = \sum_{finite} (ss_i) \otimes n_i.$$

*One can check this operation is well defined and indeed makes $S \otimes_R N$ into a left S-module, so it is called the **S-module obtained by extension of scalards from the R-module** N. There is a natural map $\iota : N \to S \otimes_R N$*

defined by $n \mapsto 1 \otimes n$. Since $1 \otimes rn = r \otimes n = r(1 \otimes n)$, then it is easy to check that ι is an R-module homomorphism from N to $S \otimes_R N$.

Theorem 5.33 (Universal Property of Tensor Produces) *Let R be a subring of S, let N be a left R-module and let $\iota : N \to S \otimes_R N$ be the R-module homomorphism defined by $\iota(n) = 1 \otimes n$. Suppose that L is any left S-module (hence also an R-module) and that $\varphi : N \to L$ is an R-module homomorphism from N to L. Then there is a unique S-module homomorphism $\Phi : S \otimes_R N \to L$ such that φ factors through Φ, i.e., $\varphi = \Phi \circ \iota$ and the diagram*



commutes. Conversely, if $\Phi : S \otimes_R N \to L$ is an S-module homomorphism then $\varphi = \Phi \circ \iota$ is an R-module homomorphism from N to L.

Proof: Suppose $\varphi : N \to L$ is an R-module homomorphism to the S-module L. By the universal property of free modules there is a $\mathbb{Z}$-module homomorphism from the free $\mathbb{Z}$-module F on the set $S \times N$ to L that sends each generator (s, n) to $s\varphi(n)$. Since φ is an R-module homomorphism, the generators of the subgroup H given by the relations 5.1, 5.2, 5.3 all map to zero in L. Hence this $\mathbb{Z}$-module homomorphism factors through H, i.e., there is a well defined $\mathbb{Z}$-module homomorphism Φ from $F/H = S \otimes_R H$ to L satisfying $\Phi(s \otimes n) = s\varphi(n)$. Moreover, on simple tensors we have

$$s'\Phi(s \otimes n) = s'(s\varphi(n)) = (s's)\varphi(n) = \Phi((s's) \otimes n) = \Phi(s'(s \otimes n))$$

for any $s' \in S$. Since Φ is additive, it follows that Φ is an S-module homomorphism, which proves the existence statement of the theorem. The module $S \otimes_R N$ is generated as an S-module by elements of the form $1 \otimes n$, so any S-module homomorphism is uniquely determined by its values on these elements. Since $\Phi(1 \otimes n) = \varphi(n)$, it follows that the S-module homomorphism Φ is uniquely determined by φ, which proves the uniqueness statement of the theorem.

The converse of the statement is clear. □

Corollary 5.33.1 *Let $\iota : N \to S \otimes_R N$ be the R-module homomorphism defined by $\iota(n) = 1 \otimes n$. Then $N/\ker \iota$ is the unique largest quotient of N that can be embedded in any S-module. In particular, N can be embedded as an R-submodule of some left S-module if and only if ι is injective, in which case N is isomorphic to the R-submodule $\iota(N)$ of the S-module $S \otimes_R N$.*

Proof: The quotient $N/\ker \iota$ is mapped injectively into the S-module $S \otimes_R N$. Suppose now that φ is an R-module homomorphism injecting the quotient $N/\ker \varphi$ of N into an S-module L. Then by the universal property, $\ker \iota$ is mapped to 0 by φ, i.e., $\ker \iota \subset \ker \varphi$. Hence $N/\ker \varphi$ is a quotient of $N/\ker \iota$. It follows that $N/\ker \iota$ is the unique largest quotient of N that can be embedded in any S-module. The second statement is immediate. □

Example:

1. $R \otimes_R N \cong N$.

2. Extension of scalars for free modules: If $N \cong \mathbb{R}^n$ is a free module of rank n over R, then $S \otimes_R N \cong S^n$ is free module of rank n over S.

3. Extension of scalars for vector spaces: let F be a subfield of the field K and let $\mathbf{V}$ be an n-dimensional vector space over F. Then $K \otimes_F \mathbf{V} \cong K^n$ is a vector space over the larger field K of the same dimension, and the original vector space $\mathbf{V}$ is contained in $K \otimes_F \mathbf{V}$ as an F-vector subspace.

4. Induced modules for finite groups: let R be a commutative ring with 1, let G be a finite group and let H be a subgroup of G. For any RH-module N, we define the *induced module* $RG \otimes_{RH} N$.

Next we study the general tensor product construction. We can only construct such tensor product $M \otimes_R N$ when M is right R-module and N is a left R-module. Suppose this is the case, the quotient of the free $\mathbb{Z}$-module on the set $M \times N$ by the subgroup generated by all elements of the form

$$\begin{gathered}(m_1 + m_2, n) - (m_1, n) - (m_2, n) \\ (m, n_1 + n_2) - (m, n_1) - (m, n_2) \\ (mr, n) - (m, rn)\end{gathered}$$

for $m, m_1, m_2 \in M$, $n, n_1, n_2 \in N$ and $r \in R$ is an abelian group, denoted by $M \otimes_R N$ or $M \otimes N$ and is called the *tensor product* of M and N over R, its elements are called *tensors* and the coset $m \otimes n$, of (m, n) is called a *simple tensor*. We have the relations

$$\begin{aligned}(m_1 + m_2) \otimes n &= m_1 \otimes n + m_2 \otimes n \\ m \otimes (n_1 + n_2) &= m \otimes n_1 + \otimes n_2 \\ mr \otimes n &= m \otimes rn.\end{aligned}$$

Every tensor can be written (non-uniquely in general) as a finite sum of simple tensors.

Definition 5.34 *Let M be a right R-module and N be a left R-module. Let L be an abelian group (written additively). A map $\varphi : M \times N \to L$ is called* ***R-balanced*** *or* ***middle linear with respect to*** *R if*

$$\begin{aligned}\varphi(m_1 + m_2, n) &= \varphi(m_1, n) + \varphi(m_2, n) \\ \varphi(m, n_1 + n_2) &= \varphi(m, n_1) + \varphi(m, n_2) \\ \varphi(m, rn) &= \varphi(mr, n)\end{aligned}$$

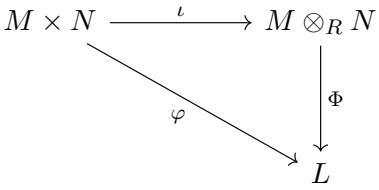
for all $m, m_1, m_2 \in M$, $n, n_1, n_2 \in N$ and $r \in R$.

Remark 5.34.1 *The Mapping $M \times N$ to the free $\mathbb{Z}$-modle on $M \times N$ and then passing to the quotient defines a map $\iota : M \times N \to M \otimes_R N$ with $\iota(m, n) = m \otimes n$, which is R-balanced.*

Theorem 5.35 (Universal Property of R-balanced Maps) *Suppose R is a ring with 1, M is a right R-module, and N is a left R-module. Let $M \otimes_R N$ be the tensor product of M and N over R, and let $\iota : M \times N \to M \otimes_R N$ defined by $\iota(m, n) = m \otimes n$.*

1. *If $\Phi : M \otimes_R N \to L$ is any group homomorphism from $M \otimes_R N$ to an abelian group L then the composite map $\varphi = \Phi \circ \iota$ is an R-balanced map from $M \times N$ to L.*
2. *Conversely, suppose L is an abelian group and $\varphi : M \times N \to L$ is any R-balanced map. Then there is a unique group homomorphism $\Phi : M \otimes_R N \to L$ such that φ factors through ι, i.e., $\varphi = \Phi \circ \iota$.*

Equivalently, the correspondence $\varphi \Leftrightarrow \Phi$ in the commutative diagram



establishes a bijection between the set of R-balanced maps $\varphi : M \times N \to L$ and the set of group homomorphism $\Phi : M \otimes_R N \to L$.

Proof: Clear from the definition of tensor products. □

Corollary 5.35.1 *Suppose D is an abelian group and $\iota' : M \times N \to D$ is an R-balanced map such that*

1. *the image of ι' generates D as an abelian group and*
2. *every R-balanced map defined on $M \times N$ factors through ι' as stated in the universal property.*

Then there is an isomorphism $f : M \otimes_R N \cong D$ of the abelian groups with $\iota' = f \circ \iota$.

Proof: Since $\iota' : M \times N \to D$ is a balanced map, the universal property implies there is a unique group homomorphism $f : M \otimes_R N \to D$ with $\iota' = f \circ \iota$. In particular $\iota'(m, n) = f(m \otimes n)$ for every $m \in M$, $n \in N$. By the first assumption on ι', these elements generated D as an abelian group, so f is surjective map. Now the balanced map $\iota : M \times N \to M \otimes_R N$ together with the second assumption ι' implies there is a unique gropu homomorphism $g : D \to M \otimes_R N$ with $\iota = g \circ \iota'$. Then $m \otimes n = (g \circ f)(m \otimes n)$. Since the simple tensors $m \otimes n$ generate $M \otimes N$, it follows that $g \circ f$ is the identity map, hence an isomorphism. □

Definition 5.36 (Bimodule) *Let R and S be any rings with 1. An abelian group M is called an (S, R)-**bimodule** if M is a left S-module, a right R-module and $s(mr) = (sm)r$ for all $s \in S$, $r \in R$ and $m \in M$.*

*Suppose M is a left or right R-module over the commutative ring R. Then the (R,R)-bimodule structure on M defined by letting the left and right R-actions coincide, i.e., $mr = rm$ for all $m \in M$ and $r \in R$, will be called the **standard R-module** structure on M.*

Now suppose N is a left R-module and M is an (S,R)-bimodule. Then

$$s\left(\sum_{\text{finite}} m_i \otimes n_i\right) = \sum_{\text{finite}} (sm_i) \otimes n_i$$

gives a well defined action of S under which $M \otimes_R N$ is a left S-module. In this case the abelian group $M \otimes_R N$ has an S-module structure, and enjoys the analogous universal property as that of Theorem 5.33. A special case of this is when M and N are two left modules over a commutative ring R and $S = R$. Then the standard R-module structure on M defined previously gives M the structure of an (R,R)-bimodule, so in this case the tensor product $M \otimes_R N$ always has the structure of a left R-module, which is closely linked to the study of bilinear maps.

Definition 5.37 (Bilinear Maps) *Let R be a commutative ring with 1 and let M, N and L be left R-modules. The map $\varphi : M \times N \to L$ is called **R-bilinear** if it is R-linear in each factor, i.e., if*

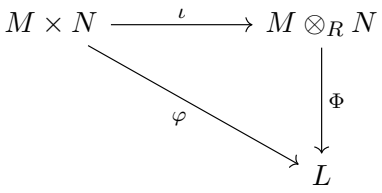
$$\varphi(r_1 m_1 + r_2 m_2, n) = r_1 \varphi(m_1, n) + r_2 \varphi(m_2, n)$$
$$\varphi(m, r_1 n_1 + r_2 n_2) = r_1 \varphi(m, n_1) + r_2 \varphi(m, n_2)$$

for all $m, m_1, m_2 \in M$, $n, n_1, n_2 \in N$ and $r_1, r_2 \in R$.

Proposition 5.38 (Universal Property of Bilinear Maps) *Suppose R is a commutative ring. Let M and N be two left R-modules and let $M \otimes_R N$ be the tensor product of M and N over R, where M is given the standard R-module structure. Then $M \otimes_R N$ is a left R-module with*

$$r(m \otimes n) = (rm) \otimes n = (mr) \otimes n = m \otimes (rn),$$

and the map $\iota : M \times N \to M \otimes_R N$ with $\iota(m, n) = m \otimes n$ is an R-bilinear map. If L is any left R-module then there is a bijection between the set of R-bilinear maps $\varphi : M \times N \to L$ and the set of R-module homomorphism $\Phi : M \otimes_R N \to L$ where the correspondence between φ and Φ is given by the commutative diagram



Proof: It is clear that $M \otimes_R N$ is an R-module and that ι is bilinear. It remains only to check that in the bijective correspondence in the Universal Property of R-balanced Maps correspond with the R-module homomorphism. If $\varphi : M \times N \to L$ is bilinear then it is a R-balanced map, so the corresponding $\Phi : M \otimes_R N$ is a group homomorphism. Moreover, on simple tensors $\Phi((rm) \otimes n) = \varphi(rm, n) = r\varphi(m, n) = r\Phi(m \otimes n)$. Since Φ is additive this extends to sums of simple tensors to show Φ is an R-module homomorphism. Conversely, if Φ is an R-module homomorphism, then one can show that the corresponding balanced map φ is bilinear. □

5.5 Basic Properties of Tensor Products

Theorem 5.39 (The "Tensor Product" of Two Homomorphisms) *Let M, M' be right R-modules, let N, N' be left R-modules, and suppose $\varphi : M \to M'$ and $\psi : N \to N'$ are R-module homomorphism.*

1. *There is a unique group homomorphism, denoted by $\varphi \otimes \psi$, mapping $M \otimes_R N$ into $M' \otimes_R N'$ such that $(\varphi \otimes \psi)(m \otimes n) = \varphi(m) \otimes \psi(n)$ for all $m \in M$ and $n \in N$.*
2. *If M, M' are also (S, R)-bimodules for some ring S and φ is also an S-module homomorphism, then $\varphi \otimes \psi$ is a homomorphism of left S-modules. In particular, if R is commutative then $\varphi \otimes \psi$ is always an R-module homomorphism for the standard R-module structures.*
3. *If $\lambda : M' \to M''$ and $\mu : N' \to N''$ are R-module homomorphism then $(\lambda \otimes \mu) \circ (\varphi \otimes \psi) = (\lambda \circ \varphi) \otimes (\mu \circ \psi)$.*

Proof: The map $(m, n) \mapsto \varphi(m) \otimes \psi(n)$ from $M \times N$ to $M' \otimes_R N'$ is clearly R-balanced, as (1) follows from the Universal Property.
In (2) the definition of the action of S on M together with the assumption that φ is an S-module homomorphism imply that on simple tensors

$$(\varphi \otimes \psi)(s(m \otimes n)) = (\varphi \otimes \psi)(sm \otimes n) = \varphi(sm) \otimes \psi(n) = s\varphi(m) \otimes \psi(n).$$

Since $\varphi \otimes \psi$ is additive, this extends to sums of simple tensors, hence $\varphi \otimes \psi$ is an S-module homomorphism.
The uniqueness of the Universal Property gives (3), which completes the proof. □

Theorem 5.40 (Associativity of the Tensor Product) *Suppose M is a right R-module, N is an (R, T)-bimodule, and L is a left T-module. Then there is a unique isomorphism*

$$(M \otimes_R N) \otimes_T L \cong M \otimes_R (N \otimes_T L)$$

of abelian groups such that $(m \otimes n) \otimes l \mapsto m \otimes (n \otimes l)$. If M is an (S,R)-bimoudle, then there is an isomorphism of S-modules.

Proof: Firstly, the (R, T)-bimodule structure on N makes $M \otimes_R N$ into a right T-module and $N \otimes_T L$ into a left R-module, so both sides of the isomorphism are well-defined. For each fixed $l \in L$, the mapping $(m, n) \mapsto m \otimes (n \otimes l)$ is R-balanced, so by the universal property, there is a homomorphism $M \otimes_R N \to M_{\otimes} R(N \otimes_T L)$ with $m \otimes n \; mapstom \otimes (n \otimes l)$. This shows that the map from $(M \otimes_R N) \times L$ to $M \otimes_R (N \otimes_T L)$ is well defined. Since it is easily seen to be T balanced, then using the universal property again would give a homomorphism $(M \otimes_R N) \otimes_T L \to M \otimes_R (N \otimes_T L)$ such that $(m \otimes n) \otimes l \mapsto m \otimes (n \otimes l)$. In a similar way we can construct a homomorphism in the opposite direct that is the inverse to this one. This proves the isomorphism.

Assume in addition M is an (S,R)-bimodule. Then for $s \in S$ and $t \in T$ we have

$$s((m \otimes n)t) = s(m \otimes nt) = sm \otimes nt = (sm \otimes n)t = (s(m \otimes n))t$$

so that $M \otimes_R N$ is an (S, T)-bimodule. Hence $(M \otimes_R N) \otimes_T L$ is a left S-module, similarly, $M|otimes_R(N \otimes_T L)$ is a left S-module. The group isomorphism just established is then a homomorphism of left S-modules. □

Remark 5.40.1 *With this theorem, we can write $M \otimes N \otimes L$ or more generally, an* ***n-fold tensor product*** *$M_1 \otimes M_2 \otimes \cdots \otimes M_n$ unambiguously whenever it is defined.*

Corollary 5.40.1 *Suppose R is commutative and M, N and L are left R-modules. Then*

$$(M \otimes N) \otimes L \cong M \otimes (N \otimes L)$$

as R-modules for the standard R-module structures on M, N and L.

Definition 5.41 (Multilinear Maps) *Let R be a commutative ring with 1 and let $M_1, M_2, \cdots, M_n$ and L be R-modules with the standard R-module structures. A map $\varphi : M_1 \times \cdots \times M_n \to L$ is called* ***n-multilinear over*** *R if it is an R-module homomorphism in each component when the other component entries are kept constant.*

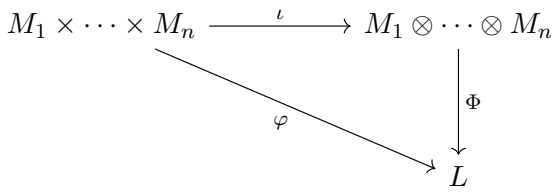
Corollary 5.41.1 *Let R be a commutative ring and let $M_1, \cdots, M_n$, L be the R-modules. Let $M_1 \otimes M_2 \otimes \cdots \otimes M_n$ denote any bracketing of the tensor product of these modules and let*

$$\iota : M_1 \times \cdots \times M_n \to M_1 \otimes \cdots \otimes M_n$$

be the map defined by $\iota(m_1, \cdots, m_n) = m_1 \otimes \cdots \otimes m_n$. Then

1. *for every R-module homomorphism $\Phi : M_1 \otimes \cdots \otimes M_n \to L$ the map $\varphi = \Phi \circ \iota$ is n-multilinear from $M_1 \times \cdots \times M_n$ to L, and*
2. *if $\varphi : M_1 \times \cdots \times M_n \to L$ is an n-multilinear map then there is a unique R-module homomorphism $\Phi : M_1 \otimes \cdots \otimes M_n \to L$ such that $\varphi = \Phi \circ \iota$.*

Hence there is a bijection between the set of n-multilinear maps and R-module homomorphism with respect to which the following diagram commutes:



Theorem 5.42 (Tensor Products of Direct Sums) *Let M, M' be right R-modules and let N, N' be left R-modules. Then there are unique group isomorphism*

$$(M \oplus M') \otimes_R N \cong (M \otimes_R N) \oplus (M' \otimes_R N)$$
$$M \otimes_R (N \oplus N') \cong (M \otimes_R N) \oplus (M \otimes_R N')$$

such that $(m, m') \otimes n \mapsto (m \otimes n, m' \otimes n)$ and $m \otimes (n, n') \mapsto (m \otimes n, m \otimes n')$ respectively. If M, M' are also (S, R)-bimodules, then there are also isomorphisms of left S-modules. In particular, if R is commutative, these are isomorphism of R-modules.

Theorem 5.43 *The map $(M\oplus M')\times N \to (M\otimes_R N)\oplus(M'\otimes_R N)$ defined by $((m,m'),m)\mapsto(m\otimes n, m'\otimes n)$ is well defined since m and m' in $M\oplus M'$ are uniquely defined in the direct sum. The maps is clearly R-balanced, so induces a homomorphism f from $(M\oplus M')\otimes N$ to $(M\otimes_R N)\oplus(M'\otimes_R N)$ with*

$$f((m,m')\otimes n) = (m\otimes n, m'\otimes n).$$

In the other direction, the R-balanced maps $M\times N\to(M\oplus M')\otimes_R N$ and $M'\times N\to(M\oplus M')\otimes_R N$ given by $(m,m)\mapsto(m,0)\otimes n$ and $(m',n)\mapsto(0,m')\times n$, respectively define homomorphism from $M\otimes_R N$ and $M'\otimes_R N$ to $(M\oplus M')\otimes_R N$. These in turn give a homomorphism g from the direct sum $(M\otimes_R N)\oplus(M'\otimes_R N)$ to $(M\oplus M')\otimes_R N$ with

$$g((m\otimes n_1, m'\otimes n_2)) = (m,0)\otimes n_1 + (0,m')\otimes n_2.$$

It is clear that f and g are inverse homomorphism and are S-module isomorphism when M and M' are (S,R)-bimodules.

Remark 5.43.1 *Using induction, we can extend the result to any finite direct sum of R-modules, that is tensor products commute with direct sums.*

Corollary 5.43.1 (Extension of Scalars for Free Modules) *The module obtained from the free R-module $N\cong R^n$ by extension of scalars is the free S-module S^n, i.e.,*

$$S\otimes_R R^n \cong S^n$$

as left S-modules.

Proof: Since $S\otimes_R R\cong S$ and $R^n = R\oplus\cdots\oplus R$ and $S^n = S\oplus\cdots\oplus S$. □

Corollary 5.43.2 *Let R be a commutative ring and let $M\cong R^s$ and $N\cong R^t$ be free R-modules with bases $m_1,\cdots,m_s$ and $n_1,\cdots,n_t$, respectively. Then $M\otimes_R N$ is a free R-module of rank st, with basis $m_i\otimes n_j$, $1\le i\le s$ and $1\le j\le t$, i.e.,*

$$R^s\otimes_R R^t \cong R^{st}.$$

Proposition 5.44 *Suppose R is a commutative ring and M,N are left R-modules, considered with the standard R-module structures. Then there is a unique R-module isomorphism*

$$M\otimes_R N\cong N\otimes_R M$$

mapping $m\otimes n$ to $n\otimes m$.

Proof: The map $M\times N\mapsto N\otimes M$ defined by $(m,n)\mapsto n\otimes m$ is R-balanced. Hence it induces a unique homomorphism f from $M\otimes N$ to $N\otimes M$ with $f(m\otimes n) = n\otimes m$. Similarly, we have a unique homomorphism g from $N\otimes M$ to $M\otimes N$ with $g(n\otimes m) = m\otimes n$ giving the inverse of f, and both maps are easily seen to be R-module isomorphism. □

Remark 5.44.1 *When $M=N$, it is not in general true that $a\otimes b = b\otimes a$ for $a,b\in M$.*

Proposition 5.45 *Let R be a commutative ring and let A and B be R-algebras. Then the multiplication $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$ is well defined and makes $A \otimes_R B$ into an R-algebra.*

Proof: Note first that the definition of an R-algebra shows that

$$r(a \otimes b) = ra \otimes b = ar \otimes b = a \otimes rb = a \otimes br = (a \otimes b)r$$

for every $r \in R$, $a \in A$ and $b \in B$. To show that $A \otimes B$ is an R-algebra, we need to show that the specified multiplication is well-defined. One way to proceed is that consider the multilinear map $\varphi : A \times B \times A \times B \Rightarrow A \otimes B$ induced by $f(a, b, a', b') = aa' \otimes bb'$ using Universal Property twice. As there is a corresponding R-module homomorphism Φ from $A \otimes B \otimes A \otimes B$ to $A \otimes B$ with $\Phi(a \otimes b \otimes a' \otimes b') = aa' \otimes bb'$. Viewing $A \otimes B \otimes A \otimes B$ as $(A \otimes B) \otimes (A \otimes B)$, then by universal property, we obtain a well defined R-bilinear mapping φ' from $(A \otimes B) \times (A \otimes B)$ to $A \otimes B$ with $\varphi'(a \otimes b, a' \otimes b') = aa' \otimes bb'$. This shows that the multiplication is indeed well-defined. And one can check that his multiplication $A \otimes B$ is an R-algebra. $\square$

6 Exact Sequences

Throughout this section, the ring will always contain a 1.
In this section, we study how one could extend a module. That is given two modules A and C, does there exists a module b containing an isomorphic copy of A such that the resulting quotient module B/A is isomorphic to C, in which case B is said to be an *extension* of C by A.

6.1 Exact Sequences

Definition 6.1 (Exact Sequence) *The pair of homomorphism $X \xrightarrow{\alpha} Y \xrightarrow{\beta} Z$ is said to be **exact** (at Y) if image $\alpha = \ker \beta$. A sequence $\cdots \to X_{n-1} \to X_n \to X_{n+1} \to \cdots$ of homomorphism is said to be an **exact sequence** if it is exact at every X_n between a pair of homomorphism.*

If we want to find a (right) module B such that $B/A \cong C$, it is equivalent to finding an exact sequence $A \xrightarrow{\psi} B \xrightarrow{\varphi} C$ that is exact at B and ψ is injective, φ is surjective. However, we can also use the language of exact sequence such that to describe injectivity and surjectivity.

Proposition 6.2 *Let A, B and C be R-modules over some ring R. Then*

- *The sequence $\{0\} \longrightarrow A \xrightarrow{\psi}$ is exact at A if and only if ψ is injective.*
- *The sequence $B \xrightarrow{\varphi} C \longrightarrow \{0\}$ is exact at C if and only if φ is surjective.*

Proof: This is clear from definition. □

Corollary 6.2.1 *The sequence $0 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 0$ is exact if and only if ψ is injective, φ is surjective, and the image $\psi = \ker \varphi$, i.e., B is an extension of C by A.*

Definition 6.3 (Short Exact Sequence) *The exact sequence $0 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 0$ is called a **short exact sequence**.*

Remark 6.3.1 *Similarly, we can consider the context of groups. The we have the analogous definition and Proposition 6.2 and Corollary 6.2.1.*

Remark 6.3.2 *Note every exact sequence can be written as a succession of short exact sequences, since to say $X \xrightarrow{\alpha} Y \xrightarrow{\beta} Z$ is exact at Y is the same as saying the sequence $0 \longrightarrow \alpha(X) \longrightarrow Y \longrightarrow Y/\ker \beta \longrightarrow 0$ is a short exact sequence.*

Definition 6.4 *Let $0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$ and $0 \longrightarrow A' \longrightarrow B' \longrightarrow C' \longrightarrow 0$ be two short exact sequences of modules.*

1. *A **homomorphism** of short exact sequences is a triple α, β, γ of module homomorphism such that the following diagram commutes:*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\ 0 & \longrightarrow & A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & 0 \end{array}$$

*The homomorphism is an **isomorphism** of short exact sequences if α, β, γ are all isomorphism, in which case the extension B and B' are said to be **isomorphic extensions**.*

2. *Two exact sequences are called **equivalent** if $A = A'$, $C = C'$, and there is an isomorphism between them. In this case, the corresponding extensions B and B' are said to be **equivalent extensions**, in particular they are also isomorphic extensions.*

3. *Homomorphisms and isomorphisms between short exact sequences of multiplicative groups are defined similarly.*

Remark 6.4.1 *It is an easy exercise to see that the composition of homomorphisms of short exact sequences is also a homomorphism. Likewise, if the triple α, β, γ is an isomorphism (or equivalence) then $\alpha^{-1}, \beta^{-1}, \gamma^{-1}$ is an isomorphism (equivalence, respectively) in the reverse direction. It follows that "isomorphism" (or equivalence) is an equivalence relation on any set of short exact sequences.*

Proposition 6.5 (The short five Lemma) *Let α, β, γ be a homomorphism of short exact sequences*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\ 0 & \longrightarrow & A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & 0 \end{array}$$

1. *If α and γ are injective then so is β.*

2. *If α and γ are surjective, then so is β.*

3. *If α and γ are isomorphisms, then so is β.*

Proof: Note (3) follows from (1) and (2).
For (1), suppose that α and γ are injective and suppose $b \in B$ with $\beta(b) = 0$. Let $\psi : A \to B$ and $\varphi : B \to C$ denote the homomorphisms in the first short exact sequences. Since $\beta(b) = 0$, it follows in particular that the image of $\beta(b)$ is in the quotient C' is also zero By the commutativity of the diagram, this implies that $\gamma(\varphi(b)) = 0$, since γ is assumed injective, then $\varphi(b) = 0$, i.e., b is in the kernel of φ. By the exactness of the first sequence, this means that b is in the image of ψ, i.e., $b = \psi(a)$ for some $a \in A$. Then, again by the commutativity of the diagram, the image of $\alpha(a)$ in B' is the same as $\beta(\psi(a)) = \beta(b) = 0$. But α and the map from A' to B' are injective by assumption, and it follows that $a = 0$. Finally, $b = \psi(a) = \psi(0) = 0$ and we see that β is injective.
The proof for (2) is similar. □

Definition 6.6 (Split Extension)

1. *Let R be a ring and let $0 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 0$ be a short exact sequence of R-modules. The sequence is said to be **split** if there is an R-module complement to $\psi(A)$ in B. In this case, up to isomorphism, $B = A \oplus C$ or more precisely, $B = \psi(A) \oplus C'$, for some submodule C' and C' is mapped isomorphically onto C by φ, i.e., $\varphi(C') \cong C$.*

2. *If* $1 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 1$ *is a short exact sequence of groups, then the sequence is said to be* ***split*** *if there is a subgroup complement to* $\psi(A)$ *in* B*. In this case, up to isomorphism,* $B = A \rtimes C$*, or more precisely* $\psi(A) \rtimes C'$ *for some subgroup* C'*, and* C *is mapped isomorphically onto* C *by* φ*, i.e.,* $\varphi(C') \cong C$*.*

In either case, the extension B *is said to be a* ***split extension*** *of* C *by* A*.*

Proposition 6.7 *The short exact sequence* $0 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 0$ *of R-modules is split if and only if there is an R-module homomorphism* $\mu : C \to B$ *such that* $\varphi \circ \mu$ *is the identity map on* C*. Similarly, the short exact sequence* $1 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 1$ *of groups is split if and only if there is a group homomorphism* $\mu : C \to B$ *such that* $\varphi \circ \mu$ *is the identity map on* C*. If this is the case, any set map* $\eta : C \to B$ *such that* $\varphi \circ \eta = id$ *is called a* ***section of*** φ *and* $\mu : C \to B$ *is called the* ***splitting homomorphism*** *for the sequence.*

Proof: If μ given defines $C' = \mu(C) \subset B$ and if C' is given define $\mu = \varphi^{-1} : C \cong C' \subset B$. □

Remark 6.7.1 *Note that a section of* φ *is nothing more than a choice of coset representatives in* B *for the quotient* $B/\ker\varphi \cong C$*. A section is a (splitting) homomorphism if this set of coset representatives forms a submodule (respectively, subgroup) in* B*, in which case this submodule (respectively, subgroup) gives a complement to* $\psi(A)$ *in* B*.*

Proposition 6.8 *Let* $0 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 0$ *be a short exact sequence of modules (respectively,* $1 \longrightarrow A \xrightarrow{\psi} B \xrightarrow{\varphi} C \longrightarrow 1$ *a short exact sequence of groups). Then* $B = \psi(A) \oplus C'$ *for some submodule* C' *of* B *with* $\varphi(C') \cong C$ *(respectively,* $B = \psi(A) \times C'$ *for some subgroup* C' *of* B *with* $\varphi(C') \cong C$*) if and only if there is a homomorphism* $\lambda : B \to A$ *such that* $\lambda \circ \psi$ *is the identity map on* A*.*

Proof: If λ is given, define $C' = \ker\lambda \subseteq B$; if C' is given, define $\lambda : B = \psi(A) \oplus C' \to A$ by $\lambda((\psi(a), c') = a$. Note that in this case $C' = \ker\lambda$ is normal in B, so that C' is a normal complement to $\psi(A)$ in B, which in turn implies that B is the direct sum of $\psi(A)$ and C'. □

6.2 Projective Modules and $\mathrm{Hom}_R(D, \_)$

Let R be a ring with 1 and suppose the R-module M is an extension of N by L, with

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

the corresponding short exact sequence of R-modules. It is natural to ask whether properties for L and N imply related properties for the extensions M.

Proposition 6.9 *Let* D, L *and* M *be R-modules, and let* $\psi : L \to M$ *be an R-module homomorphism. Then the map*

$$\begin{aligned} \psi' : \mathrm{Hom}_R(D, L) &\longrightarrow \mathrm{Hom}_R(D, M) \\ f &\longmapsto f' = \psi \circ f \end{aligned}$$

is a homomorphism of abelian groups. If ψ is injective, then ψ' is also injective, i.e., if

$$0 \longrightarrow L \xrightarrow{\psi} M$$

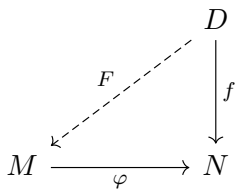
is exact, then

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi} \mathrm{Hom}_R(D, M)$$

is also exact.

Proof: $\mathrm{Hom}_R(D, L)$ and $\mathrm{Hom}_R(D, M)$ are abelian groups under addition. It is clear that ψ' is a homomorphism. If ψ is injective, then distinct homomorphism f and g from D into L give distinct homomorphism $\psi \circ f$ and $\psi \circ g$ from D into M, which is to say that ψ' is also injective. □

While obtaining homomorphisms into M from homomorphisms into the submodule L is straightforward, the situation for homomorphisms into the quotient N is much less evident. More precisely, given an R-module homomorphism $f : D \to N$, the question is whether there exists an R-module homomorphism $F : D \to M$ that extends or lifts f to M, i.e., that makes the following diagram commute:



If this is possible, then φ induces a homomorphism of abelian groups $\varphi' : \mathrm{Hom}_R(D, M) \to \mathrm{Hom}_R(D, N)$, $F \mapsto F' = \varphi \circ F$. However, in general it may not be possible to lift a homomorphism f from D to N to a homomorphism from D to M. That is if $M \xrightarrow{\varphi} N \longrightarrow 0$ is exact, then

$$\mathrm{Hom}_R(D, M) \xrightarrow{\varphi} \mathrm{Hom}_R(D, N) \longrightarrow 0$$

is not necessarily exact. However, we can shrink this sequence to obtain the following theorem.

Theorem 6.10 *Let D, L, M and N be R-modules. If*

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is exact, then the associated sequence

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{\varphi'} \mathrm{Hom}_R(D, N) \tag{6.1}$$

is exact.
A homomorphism $f : D \to N$ lifts to a homomorphism $F : D \to M$ if and only if $f \in \mathrm{Hom}_R(D, N)$ is the image of φ'. In general $\varphi' : \mathrm{Hom}_R(D, M) \to \mathrm{Hom}_R(D, N)$ need not be surjective; the map φ' is surjective if and only if every homomorphism from D to N lifts to a homomorphism from D to M, in which case the sequence (6.1) can be extended to a short exact sequence.

The sequence (6.1) is exact for all R-modules D if and only if the sequence

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N$$

is exact.

Proof: The only thing that have not been proved in the first statement is the exactness of the sequence (6.1) at $\mathrm{Hom}_R(D, M)$, i.e., $\ker \varphi' = \text{image}\, \psi'$. Suppose $F : D \to M$ is an element of $\mathrm{Hom}_R(D, M)$ lying in the kernel of φ', i.e., with $\varphi \circ F = 0$ as homomorphism from D to N. If $d \in D$ is any element of D, this implies that $\varphi(F(d)) = 0$ and $F(d) \in \ker \varphi$. By the exactness of the sequence defining the extension M, we have $\ker \varphi = \text{image}\, \psi$, so there is some element $l \in L$ with $F(d) = \psi(l)$. Since ψ is injective, the element l is unique, so this gives a well defined map $F' : D \to L$ given by $F'(d) = l$. One can verify F' is a homomorphism from D to L. Since $\psi \circ F'(d) = \psi(l) = F(d)$, we have $F = \psi'(F')$ which shows that F is in the image of ψ', showing $\ker \varphi' \subset \text{image}\, \psi'$.
Conversely, if F is in the image of ψ', then $F = \psi'(F')$ for some $F' \in \mathrm{Hom}_R(D, L)$ and so $\varphi(F(d)) = \varphi(\psi(F'(d)))$ for any $d \in D$. Since $\ker \varphi = \text{image}\, \psi$ we have $\varphi \circ \psi = 0$, and it follows that $\varphi(F(d)) = 0$ for any $d \in D$, i.e., $\varphi'(F) = 0$. Hence F is in the kernel of φ, proving the reverse containment.

For the last statement in the theorem, not first that the surjectivity of φ is not need to show that sequence (6.1) is exact. So the "if" direction has been proven. For the converse, suppose that the sequence (6.1) is exact for all R-modules D. In general, $\mathrm{Hom}_R(R, X) \cong X$ for any left R-module X, the isomorphism being given by mapping a homomorphism to its value on the element $1 \in R$. Taking $D = R$ in the sequence (6.1), the exactness of the sequence

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N$$

follows easily. □

Proposition 6.11 *Let D, L and N be R-modules. Then*

1. $\mathrm{Hom}_R(D, L \oplus N) \cong \mathrm{Hom}_R(D, L) \oplus \mathrm{Hom}(D, N)$ *and*
2. $\mathrm{Hom}_R(L \oplus N, D) \cong \mathrm{Hom}_R(L, D) \oplus \mathrm{Hom}_R(N, D)$.

Proof: We only prove 1 as the proof for 2 is similar.
Let $\pi_1 : L \oplus N \to L$ be the natural projection from $L \oplus N$ to L and similarly let π_2 be the projection to N. If $f \in \mathrm{Hom}_R(D, L \oplus N)$ then the compositions $\pi_1 \circ f$ and $\pi_2 \circ f$ give elements in $\mathrm{Hom}_R(D, L)$ and $\mathrm{Hom}_R(D, N)$, respectively. This defines a map from $\mathrm{Hom}_R(D, L \oplus N)$ to $\mathrm{Hom}_R(D, L) \oplus \mathrm{Hom}_R(D, N)$, and one can show that this is a homomorphism. Conversely, given $f_1 \in \mathrm{Hom}_R(D, L)$ and $f_2 \in \mathrm{Hom}_R(D, N)$, define the map $f \in \mathrm{Hom}_R(D, L \oplus N)$ by $f(d) = (f_1(d), f_2(d))$. This defines a map from $\mathrm{Hom}_R(D, L) \oplus \mathrm{Hom}_R(D, N)$ to $\mathrm{Hom}_R(D, L \oplus N)$ that is easily checked to be a homomorphism inverse to the map above, showing the desired statement. □

Remark 6.11.1 *The result can be extended easily to any finite direct sum of R-modules using induction. These results are referred to by saying that* Hom *commutes with finite direct sums in either variable. For infinite direct sums the situation is more complicated. Statement* (1) *remains true if $L \oplus N$ is replaced by an arbitrary direct sum and the direct sum on the right hand side is replaced by a direct product. Part* (2) *remains true if the direct sum on both sides are replaced by direct products.*

Remark 6.11.2 *This proposition shows that if the sequence*

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N$$

is a split short exact sequence of R-modules, then

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{\varphi'} \mathrm{Hom}_R(D, N) \longrightarrow 0$$

is also a split short exact sequence of abelian groups for every R-module D. In fact, the converse also holds, that is if

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{\varphi'} \mathrm{Hom}_R(D, N) \longrightarrow 0$$

is exact for every R-module D, then

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N$$

is a split short exact sequence.

Proposition 6.12 *Let P be an R-module. Then the following are equivalent:*

1. *For any R-modules L, M and N, if*

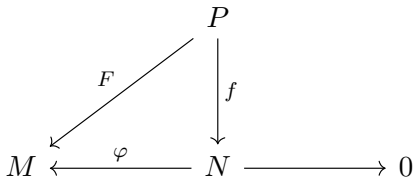
$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is a short exact sequence, then

$$0 \longrightarrow \mathrm{Hom}_R(P, L) \xrightarrow{\psi'} \mathrm{Hom}_R(P, M) \xrightarrow{\varphi'} \mathrm{Hom}_R(P, N) \longrightarrow 0$$

is also a short exact sequence.

2. *For any R-modules M and N, if $M \xrightarrow{\varphi} N \longrightarrow 0$ is exact, then every R-module homomorphism from P into N lifts to an R-module homomorphism into M, i.e., given $f \in \mathrm{Hom}_R(P, N)$, there is a lift $F \in \mathrm{Hom}_R(P, M)$ making the following diagram commute:*



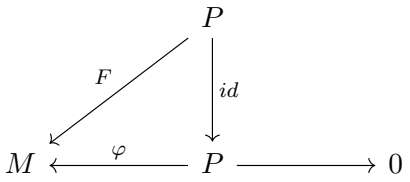
3. *If P is a quotient of the R-module M then P is isomorphic to a direct summand of M, i.e., every short exact sequence*

$$0 \longrightarrow L \longrightarrow M \longrightarrow P \longrightarrow 0$$

splits.

4. *P is a direct summand of a free R-module.*

Proof: (1) and (2) are equivalent by Theorem 6.10. We show (2) implies (3). Let $M \xrightarrow{\varphi} N \longrightarrow 0$ be exact, then the identity map from P to P lifts to a homomorphism μ making the following diagram commute



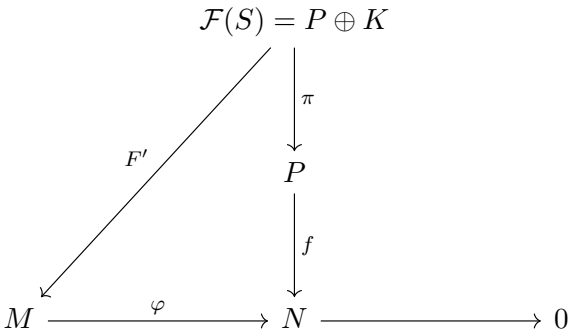
Then $\varphi \circ \mu = id$ so μ is a splitting homomorphism for the sequence which proves (3).

For (4), we know every module P is the quotient of a free module, so there is always an exact sequence

$$0 \longrightarrow \ker \varphi \longrightarrow \mathcal{F} \xrightarrow{\varphi} P \longrightarrow 0$$

where $\mathcal{F}$ is a free R-module. If (3) is satisfies, then this sequence splits, so $\mathcal{F}$ is isomorphic to the direct sum of $\ker \varphi$ and P, which proves (4).

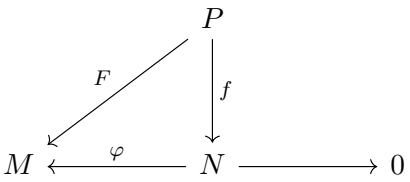
Finally, we show (4) implies (2). Suppose that P is a direct summand of a free R-module on some set S, say $\mathcal{F}(S) = P \oplus K$, and that we are given a homomorphism f from P to N as in (2). Let π denote the natural projection from $\mathcal{F}(S)$ to P, so that $f \circ \pi$ is a homomorphism from $\mathcal{F}(S)$ to N. For any $s \in S$ define $n_s = f \circ \pi(s) \in N$ and let $m_s \in M$ be any element of M with $\varphi(m_s) = n_s$ (which exists because φ is surjective). By the Universal property of free modules, there is a unique R-module homomorphism F' from $\mathcal{F}(S)$ to M with $F'(s) = m_s$. The diagram is the following:



By definition of homomorphism F' we have $\varphi \circ F'(s) = \varphi(m_s) = n_s = f \circ \pi(s)$, from which it follows that $\varphi \circ F' = f \circ \pi$ on $\mathcal{F}(S)$, i.e., the diagram commutes. Now define a map $F : P \to M$ by $F(d) = F'((d, 0))$. Since F is the composite of the injection $P \to \mathcal{F}(S)$ with the homomorphism F', it follows that F is an R-module homomorphism. Then

$$\varphi \circ F(d) = \varphi \circ F'((d, 0)) = f \circ \pi((d, 0)) = f(d)$$

i.e., $\varphi \circ F = f$, so the diagram



commutes, which shows that (4) implies (2). □

Definition 6.13 (Projective Modules) *An R-module P is called **projective** if it satisfies any of the equivalent conditions stated in Proposition 6.12.*

Corollary 6.13.1 *Free modules are projective. A finitely generated module is projective if and only if it is a direct summand of a finitely generated free modules. Every module is a quotient of a projective module.*

The map $\mathrm{Hom}_R(D, \_)$ is a covariant functor from the category of R-modules to the category of abelian groups, where D is a fixed R-module. Applying this functor to the terms in the exact sequence

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

produces an exact sequence

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{\varphi'} \mathrm{Hom}_R(D, N)$$

This is referred to by saying that $\mathrm{Hom}_R(D, \_)$ is a *left exact (covariant) functor.*

Corollary 6.13.2 *If D is an R-module, then the functor $\mathrm{Hom}_R(D, \_)$ from the category of R-modules to the category of abelian group is left exact. It is exact if and only if D is a projective R-module.*

Remark 6.13.1 *Note that if $\mathrm{Hom}_R(D, \_)$ takes short exact sequences to short exact sequences, then it takes exact sequences of any length to exact sequences since any exact sequence can be broken up into a succession of short exact sequences.*

6.3 Injective Modules and $\mathrm{Hom}_R(\_, D)$

If $0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$ is a short exact sequence of R-module, then instead of considering maps from an R-module D into L or N and the extent to which these determine maps from D into M, we can consider the "dual" question of maps from L or N to D. In this case, it is easy to dispose of the situation of a map from N to D: an R-module map from N to D immediately gives a map from M to D simply by composing with φ. It is easy to check that this defines an injective homomorphism of abelian groups.

$$\begin{aligned} \varphi' : \mathrm{Hom}_R(N, D) &\longrightarrow \mathrm{Hom}_R(M, D) \\ f &\longmapsto f' = f \circ \varphi \end{aligned}$$

or put another way, if $M \xrightarrow{\varphi} N \longrightarrow 0$ is exact, then so is

$$0 \longrightarrow \mathrm{Hom}_R(N, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(M, D).$$

On the other hand, given an R-module homomorphism f from L to D it may not be possible to extend f to a map F from M to D. That is if

$$0 \longrightarrow L \xrightarrow{\psi} M$$

is exact, then

$$\mathrm{Hom}_R(M, D) \xrightarrow{\psi'} \mathrm{Hom}_R(L, D) \longrightarrow 0$$

is not necessarily exact. Hence we have the following dual version of Theorem 6.10

Theorem 6.14 *Let D, L, M and N be R-modules. If*

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is exact, then the associated sequence

$$0 \longrightarrow \mathrm{Hom}_R(N, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(M, D) \xrightarrow{\psi'} \mathrm{Hom}_R(L, D) \tag{6.2}$$

is exact.
A homomorphism $f : L \to D$ lifts to a homomorphism $F : M \to D$ if and only $f \in \mathrm{Hom}_R(L, D)$ is in the image of ψ'. In general $\psi' : \mathrm{Hom}_R(M, D) \to \mathrm{Hom}_R(L, D)$ need not be surjective; the map ψ' is surjective if and only if every homomorphism from L to D lifts to a homomorphism from M to D, in which case the sequence (6.2) can be extended to a short exact sequence.
*The sequence (**??**) is exact for all R-modules D if and only if the sequence*

$$L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is exact.

Remark 6.14.1 *The second statement of 6.11 shows that the sequence*

$$0 \longrightarrow \mathrm{Hom}_R(N, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(M, D) \xrightarrow{\psi'} \mathrm{Hom}_R(L, D) \longrightarrow 0$$

is exact if

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is a split exact sequence. In fact in this case, the sequence

$$0 \longrightarrow \mathrm{Hom}_R(N, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(M, D) \xrightarrow{\psi'} \mathrm{Hom}_R(L, D) \longrightarrow 0$$

is also a split exact sequence of abelian groups for every R-module D.

Conversely, if

$$0 \longrightarrow \mathrm{Hom}_R(N, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(M, D) \xrightarrow{\psi'} \mathrm{Hom}_R(L, D) \longrightarrow 0$$

is exact for every R-module D, then

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is a split short exact sequence, which then implies that if the Hom *sequence is exact for every D, then in fact it is split exact for every D.*

Proposition 6.15 *Let Q be an R-module. Then the following are equivalent:*

1. *For any R-modules L, M and N, if*

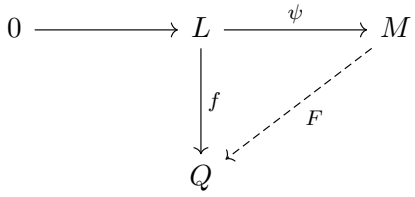
$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is a short exact sequence, then

$$0 \longrightarrow \mathrm{Hom}_R(N,Q) \xrightarrow{\varphi'} \mathrm{Hom}_R(M,Q) \xrightarrow{\psi'} \mathrm{Hom}_R(L,Q) \longrightarrow 0$$

is also a short exact sequence.

2. *For any R-modules L and M, if $0 \longrightarrow L \xrightarrow{\psi} M$ is exact, then every R-module homomorphism from L into Q lifts to an R-module homomorphism of M into Q, i.e., given $f \in \mathrm{Hom}_R(L,Q)$, there is a lift $F \in \mathrm{Hom}_R(M,Q)$ making the following digram commutate:*



3. *If Q is a submodule of the R-module M then Q is a direct summand of M, i.e., every short exact sequence*

$$0 \longrightarrow Q \longrightarrow M \longrightarrow N \longrightarrow 0$$

splits.

Definition 6.16 (Injective Modules) *An R-module Q is called **injective** if it satisfies any of the equivalent conditions of Proposition 6.15.*

If D is fixed, then given any R-module X we have an associated abelian group $\mathrm{Hom}_R(X,D)$. Furtherm an R-module homomorphism $\alpha : X \to Y$ induces an abelian group homomorphism $\alpha' : \mathrm{Hom}_R(Y,D) \to \mathrm{Hom}_R(X,D)$, defined by $\alpha'(f) = f \circ \alpha$, Put another way, the map $\mathrm{Hom}_R(D, \_)$ is a contravariant functor from the category of R-modules to the category of abelian groups. By applying this functor to the terms in the exact sequence

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

produces an exact sequence

$$0 \longrightarrow \mathrm{Hom}_R(N,Q) \xrightarrow{\varphi'} \mathrm{Hom}_R(M,Q) \xrightarrow{\psi'} \mathrm{Hom}_R(L,Q)$$

This is referred to by saying that $\mathrm{Hom}_R(\_, D)$ is a *left exact (contravariant) functor.*

Corollary 6.16.1 *If D is an R-module, then the functor $\mathrm{Hom}_R(\_, D)$ from the category of R-modules to the category of abelian groups is left exact. It is exact if and only if D is an injective R-module.*

. Next, we give a characterization of injective R-modules. Recall a $\mathbb{Z}$-module A is divisible if $A = nA$ for all nonzero integers n.

Proposition 6.17 *Let Q be an R-module*

1. *(Baer's Criterion) The module Q is injective if and only if for every left ideal I of R any R-module homomorphism $g : I \to Q$ can be extended to an R-module homomorphism $G : R \to Q$.*
2. *If R is a P.I.D. then Q is injective if and only if $rQ = Q$ for every nonzero $r \in R$. In particular, $\mathbb{Z}$-module is injective if and only if it is divisible. When R is a P.I.D., quotient modules of injective R-modules are again injective.*

Proof:

1. If Q is injective and $g : I \to Q$ is an R-module homomorphism from the nonzero ideal I of R into Q, then g can be extended to an R-module homomorphism from R into Q by Proposition 6.15, condition 2 applied to the exact sequence $0 \to I \to R$, which proves the "only if" part.

   Suppose conversely that every homomorphism $g : I \to Q$ can be lifted to a homomorphism $G : R \to Q$. To show that Q is injective, we must show that if $0 \to L \to M$ is exact, and $f : L \to Q$ is an R-module homomorphism, then there is a lift $F : M \to Q$ extending f. If $\mathcal{S}$ is the collection (f', L') of lifts $f' : L' \to Q$ of f to a submodule L' of M containing L, then the ordering $(f', L') \leq (f'', L'')$ if $L' \subset L''$ and $f'' = f'$ on L' is a partial order on $\mathcal{S}$. Since $\mathcal{S} \neq \emptyset$, by Zorn's Lemma there is a maxima element (F, M') in $\mathcal{S}$ ("union" of the chains is an upper bound). The map $F : M' \to Q$ is a lift of f and it suffices to show that $M' = M$. Suppose that there is some element $m \in M$ not contained in M' and let $I = \{r \in R \,|\, rm \in M'\}$. It is easy to check that I is a left ideal in R, and the map $g : I \to Q$ defined by $g(x) = F(xm)$ is an R-module homomorphism from I to Q. By hypothesis, there is a lift $G : R \to Q$ of g. Consider the submodule $M' + Rm$ of M, and define the map $F' : M' : Rm \to Q$ by $F'(m' + rm) = F(m') + G(r)$. If $m_1 + r_1 m = m_2 + r_2 m$, then $(r_1 - r_2)m = m_2 - m_1$ shows that $r_1 - r_2 \in I$, so that

   $$G(r_1 - r_2) = g(r_1 - r_2) = F((r_2 - r_2)m) = F(m_2 - m_1)$$

   so $F(m_1) + G(r_1) = F(m_2) + G(r_2)$. Hence F' is well defined and is an R-module homomorphism extending f to $M' + Rm$ which contradicts the maximality of M'. So it must be the case that $M' = M$.

2. Suppose R is a Principal Ideal Domain. Any nonzero ideal I of R is of the form $I = (r)$ for some nonzero element $r \in R$. An R-module homomorphism $f : I \to Q$ is completely determined by the image $f(r) = q$ in Q. This homomorphism can be extended to a homomorphism $F : R \to Q$ if and only if there is an element q' in Q with $F(1) = q'$ satisfying $q = (f_r) = F(r) = rq'$. It follows that Baer's criterion for Q is satisfied if and only if $rQ = Q$, which proves the first two statement in (2). The final statement follows since a quotient of a module Q with $rQ = Q$ for all $r \neq 0$ in R has the same property.

□

Corollary 6.17.1 *Every $\mathbb{Z}$-module is a submodule of an injective $\mathbb{Z}$-module.*

Proof: Let M be a $\mathbb{Z}$-module and let A be any set of $\mathbb{Z}$-module generators of M. Let $\mathcal{F} = F(A)$ be the free $\mathbb{Z}$-module on the set A. Then there is a surjective $\mathbb{Z}$-module homomorphism from $\mathcal{F}$ to M and if $\mathcal{K}$ denotes the kernel of this homomorphism then $\mathcal{K}$ is a $\mathbb{Z}$-submodule of $\mathcal{F}$ and we can identify $M = \mathcal{F}/\mathcal{K}$. Let $\mathcal{Q}$ be the free

$\mathbb{Q}$-module on the set A. Then $\mathcal{Q}$ is a direct sum of a number of copies of $\mathbb{Q}$, as is divisible, hence injective $\mathbb{Z}$-module containing $\mathcal{F}$. Then $\mathcal{K}$ is also a $\mathbb{Z}$-submodule of $\mathcal{Q}$, so the quotient $\mathcal{Q}/\mathcal{K}$ is injective. Since $M = \mathcal{F}/\mathcal{K} \subseteq \mathcal{Q}/\mathcal{K}$, it follows that M is contained in an injective $\mathbb{Z}$-module. □

Theorem 6.18 *Let R be a ring with 1 and let M be an R-module. Then M is contained in an injective R-module.*

6.4 Flat Modules and $D \otimes_R \_$

We now consider the behaviour of the extensions

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

of R-modules with respect to tensor products.
Suppose that D is a right R-module. For any homomorphism $f : X \to Y$ of left R-modules we obtain a homomorphism $1 \otimes f : D \otimes_R X \to D \otimes_R Y$ of abelian groups. If in addition D is an (S,R)-bimodule, and D is given the standard (R,R)-bimodule structure, then $1 \otimes f$ is a homomorphism of left S-modules. Put another way,

$$D \otimes_R \_ : X \longrightarrow D \otimes_R X$$

is a covariant functor from the category of left R-modules to the category of abelian group. In a similar way, if D is a left R-module, then $\_ \otimes_R D$ is a covariant functor from the category of right R-modules to the category of abelian groups.

Theorem 6.19 *Suppose that D is a right R-module and that L, M and N are left R-modules. If*

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is exact then the associated sequence of abelian groups

$$D \otimes_R L \xrightarrow{1\otimes\psi} D \otimes_R M \xrightarrow{1\otimes\varphi} D \otimes_R N \longrightarrow 0 \tag{6.3}$$

is exact.
If D is an (S,R)-bimodule, then the sequence (6.3) is an exact sequence of left S-modules. In particular, if $S = R$ is a commutative ring, then sequence (6.3) is an exact sequence of R-modules with respect to the standard R-module structures. The map $1 \otimes \varphi$ is not in general injective, i.e., the sequence (6.3) cannot in general be extended to a short exact sequence.
The sequence (6.3) is exact for all right R-modules D if and only if

$$L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is exact.

Proof: For the first statement it remains to prove the exactness of sequence (6.3) at $D \otimes_R M$. Since $\varphi \circ \psi = 0$, we have

$$(1 \otimes \varphi)\left(\sum_{d_i} \otimes \psi(l_i)\right) = \sum d_i \otimes (\varphi \circ \psi(l_i)) = 0$$

and it follows that image $(1 \otimes \psi) \subseteq \ker(1 \otimes \varphi)$. In particular, there is a natural projection $\pi : (D \otimes_R M)/\operatorname{image}(1 \otimes \psi) \to (D \otimes_R M)/\ker(1 \otimes \varphi) = D \otimes_R N$. The composite of the two projection homomorphism

$$D \otimes_R M \longrightarrow (D \otimes_R M)/\operatorname{image}(1 \otimes \psi) \xrightarrow{\pi} D \otimes_R N$$

is the quotient of $D \otimes_R M$ by $\ker(1 \otimes \varphi)$, as is just the map $1 \otimes \varphi$. We shall show that π is an isomorphism, which will show that the kernel of $1 \otimes \varphi$ is just the kernel of the first projection above, i..e, $\operatorname{image}(1 \otimes \psi) = \ker(1 \otimes \varphi)$. To see that π is an isomorphism, we define an inverse map. First define $\pi' : D \times N \to (D \otimes_R M)/\operatorname{image}(1 \otimes \psi)$ by $\pi'((d, n)) = d \otimes m$ for any $m \in M$ with $\varphi(m) = n$. Note that this is well-defined: any other element $M' \in M$ mapping to n differs from m by an element in $\ker \varphi = \operatorname{image} \psi$, i.e., $m' = m + \psi(l)$ for some $l \in L$, and $d \otimes \psi(l) \in \operatorname{image}(1 \otimes \psi)$. It is easy to check that π' is a balanced map, so induces a homomorphism $\tilde{\pi} : D \times N \to (D \otimes_R M)/\operatorname{image}(1 \otimes \psi)$ with $\tilde{\pi}(d \otimes n) = d \otimes m$. Then $\tilde{\pi} \circ \pi(d \otimes m) = \tilde{\pi}(d \otimes \varphi(m)) = d \otimes m$ shows that $\tilde{\pi} \circ \pi = 1$. Similarly $\pi \circ \tilde{\pi} = 1$, so π and $\tilde{\pi}$ are inverse isomorphism. Note that the injectivity of ψ was not required for the proof.
Now suppose sequence (6.3) is exact for every right R-module D. On general $R \otimes_R X \cong X$ for any left R-module X. Taking $D = R$, the exactness of the sequence

$$L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

follows. □

Proposition 6.20 *Let A be a right R-module, then the following are equivalent:*

1. *For any left R-modules L, M and N, if*

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is a short exact sequence, then

$$0 \longrightarrow A \otimes_R L \xrightarrow{1 \otimes \psi} A \otimes_R M \xrightarrow{1 \otimes \varphi} A \otimes_R N \longrightarrow 0$$

is also a short exact sequence.

2. *For any left R-modules L and M, if $0 \longrightarrow L \xrightarrow{\psi} M$ is an exact sequence of left R-modules (i.e., $\psi : L \to M$ is injective) then*

$$0 \longrightarrow A \otimes_R L \xrightarrow{1 \otimes \psi} A \otimes_R M$$

is an exact sequence of abelian groups (i.e., $1 \otimes \psi : A \otimes_R L \to A \otimes_R M$ is injective).

Remark 6.20.1 *Since tensor products commute with direct sums, it follows that if*

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is a split short exact sequence then

$$0 \longrightarrow D \otimes_R L \xrightarrow{1\otimes\psi} D \otimes_R M \xrightarrow{1\otimes\varphi} D \otimes_R N \longrightarrow 0$$

is a split short exact sequence

Definition 6.21 (Flat Modules) *A right R-moduleA is called* ***flat*** *if it satisfies either of the two equivalent conditions of Proposition 6.20.*

Corollary 6.21.1 *If D is a right R-module, then the functor $D \otimes_R \_$ from the category of left R-modules to the category of abelian groups is right exact. IF D is an (S,R)-bimodule, then $D \otimes_R \_$ is a right exact functor from the category of left R-modules to the category of let S-modules. The functor is exact if and only if D is flat R-module.*

Corollary 6.21.2 *Free modules are flat; more generally, projective modules are flat.*

Proof: To show that the free R-module F is flat it suffices to show that for any injective map $\psi : L \to M$ of R-modules$L$ and M the induced map $1 \otimes \psi : F \otimes_R L \to F \otimes_R M$ is also injective. Suppose first that $F \cong \mathbb{R}^n$ is finitely generated free module. In this case $F \otimes_R L = R^n \otimes_R L \cong L^n$ since $R \otimes_R L \cong L$ and tensor products commute with direct sums. Similarly $F \otimes_R M \cong M^n$. And under these isomorphism, the map $1 \otimes \psi : F \otimes_R L \to F \otimes_R M$ is just the natural map of L^n to M^n induced by the inclusion ψ in each component. In particular, $1 \otimes \psi$ is injective and it follows that any finitely generated free module is flat.

Suppose now that F is an arbitrary free module and that the element $\sum f_i \otimes l_i \otimes F \otimes_R L$ is mapped to 0 by $1 \otimes \psi$. This means that the element $\sum(f_i, \psi(l_i))$ can be written as a sum of generators in the free group on $F \times M$. Since this sum of elements is finite, all of the first coordinates of the resulting equation lies in some finitely generated free submodule F' of F. Then this equation implies that $\sum f_i \otimes l_i \in F' \otimes_R L$ is mapped to 0 in $F' \otimes_R M$. Since F' is finitely generated free module, the injective we proved above shows that $\sum f_i \otimes l_i$ is 0 in $F' \otimes_R L$ and so also in $F \otimes_R L$. It follows that $1 \otimes \psi$ is injective and hence that F is flat.

Suppose now that P is a projective module. Then P is a direct summand of free module F, say $F = P \oplus P'$. If $\psi : L \to M$ is injective, then $1 \otimes \psi : F \otimes_R L \to F \otimes_R M$ is also injective by what we have shown above. Since $F = P \oplus P'$ and tensor products commute with direct sums, this shows

$$1 \otimes \psi : (P \otimes_R L) \oplus (P' \otimes_R L) \to (P \otimes_R M) \oplus (P' \otimes_R M)$$

is injective. Hence $1 \otimes \psi : P \otimes_R L \to P \otimes_R M$ is injective, proving that P is flat. □

Theorem 6.22 (Adjoint Associativity) *Let R and S be rings, let A be a right R-module, let B be an (R,S)-bimodule and let C be a right S-module. Then there is an isomorphism of abelian groups:*

$$\mathrm{Hom}_S(A \otimes_R B, C) \cong \mathrm{Hom}_R(A, \mathrm{Hom}_S(B, C))$$

(the homomorphism groups are right module homomorphism) If $R = S$ is commutative, this is an isomorphism of R-modules with the standard R-module structures.

Proof: Suppose $\varphi : A \otimes_R B \to C$ is a homomorphism. For any fixed $a \in A$ define the map $\Phi(a)$ from B to C by $\phi(a)(b) = \varphi(a \otimes b)$. It is easy to check that $\Phi(a)$ is a homomorphism of right S-module and that the map Φ from A to $\mathrm{Hom}_s(B, C)$ given by a mapping a to $\Phi(a)$ is a homomorphism of right R-modules. Then $f(\varphi) = \Phi$ defines a group homomorphism from $\mathrm{Hom}_S(A \otimes_R B, C)$ to $\mathrm{Hom}_R(A, \mathrm{Hom}_S(B, C))$.
Conversely, suppose $\Phi : A \to \mathrm{Hom}_S(B, C)$ is a homomorphism. The map from $A \times B$ to C defined by mapping (a, b) to $\Phi(a)(c)$ is an R-balanced map, so induces a homomorphism φ from $A \otimes_R B$ to C. Then $g(\Phi) = \varphi$ defines a group homomorphism inverse to f. □

Corollary 6.22.1 *If R is a commutative then the tensor product of two projective R-modules is projective.*

Proof: Let P_1 and P_2 be projective modules. Then $\mathrm{Hom}_R(P_2, \_)$ is an exact functor from the category of R-modules to the category of R-modules. Then the composition $\mathrm{Hom}_R(P_1, \mathrm{Hom}_R(P_2, \_))$ is an exact functor. By Theorem 6.22, this means that $\mathrm{Hom}_R(P_1 \otimes_R P_2, \_)$ is an exact functor on R-modules. Hence $P_1 \otimes P_2$ is projective. □

6.5 Summary

Each of the functors $\mathrm{Hom}_R(A, \_)$, $\mathrm{Hom}_R(\_, A)$, and $A \otimes_R \_$, map left R-modules to abelian groups; the functor $\_ \otimes_R A$ maps right R-modules to abelian groups. When R is commutative all four functors map R-modules to R-modules.

1. Let A be a left R-module. The functor $\mathrm{Hom}_R(A, \_)$ is covariant and left exact; the module A is projective if and only if $\mathrm{Hom}_R(A, \_)$ is exact.

2. Let A be a left R-module. The functor $\mathrm{Hom}_R(\_, A)$ is contravariant and left exact; the module A is injective if and only if $\mathrm{Hom}_R(\_, A)$ is exact.

3. Let A be a right R-module. The functor $A \otimes_R \_$ is covariant and right exact; the module A is flat if and only if $A \otimes_R \_$ is exact.

4. Let A be a left R-module. The functor $\_ \otimes_R A$ is covariant and right exact; the module A is flat if and only if $\_ \otimes_R A$ is exact.

5. Projective modules are flat. The $\mathbb{Z}$-module $\mathbb{Q}/\mathbb{Z}$ is injective but not flat. The $\mathbb{Z}$-module $\mathbb{Z} \oplus \mathbb{Q}$ is flat but neither projective nor injective.

7 Vector Space

Lemma 7.1 *Assume the set $\mathcal{A} = \{v_1, v_2, \cdots, v_n\}$ spans the vector space $\mathbf{V}$ bu no proper subset of $\mathcal{A}$ spans V. Then $\mathcal{A}$ is a basis of $\mathbf{V}$. In particular, any finitely generated vector space over $\mathbb{F}$ is a free $\mathbb{F}$-module.*

Lemma 7.2 *Let $\mathbf{V}$ be a vector space over $\mathbb{F}$ and let $\mathbf{W}$ be a subspace of $\mathbf{V}$. Then $\mathbf{V}/\mathbf{W}$ is a vector space with $\dim \mathbf{V} = \dim \mathbf{W} + \dim \mathbf{V}/\mathbf{W}$.*
In particular, let $\varphi : \mathbf{V} \to \mathbf{U}$ be a linear transformation of vector spaces over $\mathbb{F}$. Then $\ker \varphi$ is subspace of $\mathbf{V}$, $\varphi(\mathbf{V})$ is a subspace of $\mathbf{U}$ and $\dim V = \dim \ker \varphi + \dim \varphi(\mathbf{V})$.

Definition 7.3 (Nonsingular) *An $m \times n$ matrix A is called **nonsingular** if $Ax = 0$ with $x \in \mathbb{F}^n$ implies $x = 0$. It is called **singular** if A is not nonsingular.*

Lemma 7.4 *If $\mathcal{B}$ is a basis of the n-dimensional space $\mathbf{V}$, the map $\varphi \mapsto [\varphi]_{\mathcal{B}}^{\mathcal{B}}$ is a ring and a vector space isomorphism of $\mathrm{Hom}_{\mathbb{F}}(\mathbf{V}, \mathbf{V})$ onto the space $M_n(\mathbb{F})$ of $n \times n$ matrices with coefficients in $\mathbb{F}$. In particular, $GL(\mathbf{V}) \cong GL_n(\mathbb{F})$ where $\dim \mathbf{V} = n$. If $\mathbb{F}$ is a finite field the order of the finite group $GL_n(\mathbb{F})$ is given by*

$$(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1})$$

where $|\mathbb{F}\| = q$.

Definition 7.5 (Similar) *Two $n \times n$ matrices A and B are said to be **similar** if there is an invertible $n \times n$ matrix P such that $P^{-1}AP = B$. Two linear transformations φ and ψ in $L(\mathbf{V})$ are said to be **similar** if there is an invertible linear transformation ξ from $\mathbf{V}$ to $\mathbf{V}$ such that $\xi^{-1}\varphi\xi = \psi$.*

Remark 7.5.1 *In other words, two matrices are similar if they are in the same orbit of $GL_n(\mathbb{F})$ acting on $M_n(\mathbb{F})$ by conjugation. And two linear transformations are similar if they are in the same orbit of $GL(\mathbf{V})$ acting on $L(\mathbf{V})$ by conjugation.*

Proposition 7.6 *Let $\mathbb{F}$ be a subfield of the field $\mathbb{K}$. If $\mathbf{W}$ is an m-dimensional vector space over $\mathbb{F}$ with basis $w_1, \cdots, w_m$, then $\mathbb{K} \otimes_{\mathbb{F}} \mathbf{W}$ is an m-dimensional vector space over $\mathbb{K}$ with basis $1 \otimes w_1, \cdots, 1 \otimes w_m$.*

Proposition 7.7 *Let $\mathbf{V}$ and $\mathbf{W}$ be finite dimensional vector spaces over the field $\mathbb{F}$ with basis $v_1, \cdots, v_n$ and $w_1, \cdots, w_m$ respectively. Then $\mathbf{V} \otimes_{\mathbb{F}} \mathbf{W}$ is a vector space over $\mathbb{F}$ of dimension nm with basis $v_i \otimes w_j$, $1 \leq i \leq n$ and $1 \leq j \leq m$.*

Definition 7.8 (Kronecker Product) *Let $A = (\alpha_{ij})$ and B be $r \times n$ and $s \times m$ matrices with coefficients from any commutative ring. The **Kronecker product** or **tensor product** of A and B, denoted $A \otimes B$, is the $rs \times nm$ matrix consisting of an $r \times n$ block matrix whose i, j block is the is the $s \times m$ matrix $a_{ij}B$.*

Proposition 7.9 *Let $\varphi : \mathbf{V} \to X$ and $\psi : \mathbf{W} \to Y$ be linear transformations of finite dimensional vector spaces. Then the Kronecker product of matrices representing φ and ψ is a matrix representation of $\varphi \otimes \psi$.*

Definition 7.10 (Dual) *For any vector space $\mathbf{V}$ over $\mathbb{F}$, let $\mathbf{V}^* = \mathrm{Hom}_{\mathbb{F}}(\mathbf{V}, \mathbb{F})$ be the space of linear transformation from $\mathbf{V}$ to $\mathbb{F}$, called the **dual space** of $\mathbf{V}$. Elements of $\mathbf{V}^*$ are called linear functionals.*

Proposition 7.11 *Assume φ is an n-multilinear alternating function on $\mathbf{V}$ that for some $v_1, \cdots, v_n$ and $w_1, \cdots, w_n \in \mathbf{V}$ and some $\alpha_{ij} \in \mathbb{R}$, we have*

$$\begin{aligned} w_1 &= \alpha_{11}v_1 + \alpha_{21}v_2 + \cdots + \alpha_{n1}v_n \\ &\vdots \\ w_n &= \alpha_{1n}v_1 + \alpha_{2n}v_2 + \cdots + \alpha_{nn}v_n. \end{aligned}$$

Then

$$\varphi(w_1, w_2, \cdots, w_n) = \sum_{\sigma \in S_n} \text{sign}\,(\sigma)\alpha_{\sigma(1)1}\alpha_{\sigma(2)2}\cdots\alpha_{\sigma(n)n}\varphi(v_1, \cdots, v_n).$$

7.1 Tensor Algebras

In this section R is any commutative ring with 1, and we assume the left and right actions of R on each R-module are the same.

Definition 7.12 *For each $k \geq 1$, define*

$$\mathcal{T}^k(M) = M \otimes_R M \otimes_R \cdots \otimes_R M \quad (k \text{ factors}),$$

*and set $\mathcal{T}^0(M) = R$. Th elements of $\mathcal{T}^k(M)$ are called **k-tensors**. Define*

$$\mathcal{T}(M) = R \oplus \mathcal{T}^1(M) \oplus \mathcal{T}^2(M) \oplus \mathcal{T}^3(M) \cdots = \bigoplus_{k=0}^{\infty} \mathcal{T}^k(M).$$

Every element of $\mathcal{T}(M)$ is a finite linear combination of k-tensors for various $k \geq 0$. We identify M with $\mathcal{T}^1(M)$, so that M is an R-submodule of $\mathcal{T}(M)$.

Theorem 7.13 *If M is any R-module over the commutative ring R, then*

1. *$\mathcal{T}(M)$ is an R-algebra containing M with multiplication defined by mapping*

$$(m_1 \otimes \cdots \otimes m_i)(m'_1 \otimes \cdots \otimes m'_j) = m_1 \otimes \cdots \otimes m_i \otimes m'_1 \otimes \cdots \otimes m'_j$$

   and extended to sums via the distributive laws. With respect to this multiplication $\mathcal{T}^i(M)\mathcal{T}^j(M) \subseteq \mathcal{T}^{i+j}(M)$.

2. *(Universal Property) If A is any R-algebra and $\varphi : M \to A$ is an R-module homomorphism, then there is a unique R-algebra homomorphism $\Phi : \mathcal{T}(M) \to A$ such that $\Phi|_M = \varphi$.*

Proof: The map

$$\underbrace{M \times M \cdots \times M}_{i \text{ factors}} \times \underbrace{M \times M \cdots \times M}_{j \text{ factors}} \to \mathcal{T}^{i+j}(M)$$

defined by

$$(m_1, \cdots ,_i , m'_1, \cdots , m'_j) \mapsto m_1 \otimes \cdots \otimes m_i \otimes m'_1 \otimes \cdots \otimes m'_j$$

is R-multilinear, so induces a bilinear map $\mathcal{T}^i(M) \times \mathcal{T}^j(M)$ to $\mathcal{T}^{i+j}(M)$ which is easily check to give a well-defined multiplication satisfying (1).

To prove (2), assume that $\varphi : M \to A$ is an R-algebra homomorphism. Then

$$(m_1, m_2, \cdots , m_k) \mapsto \varphi(m_1)\varphi(m_2) \cdots \varphi(m_k)$$

defines an R-multilinear map from $M \times \cdots \times M$ (k times) to A. This in turn induces a unique R-module homomorphism Φ from $\mathcal{T}^k(M)$ to A mapping $m_1 \otimes \cdots \otimes m_k$ to the element on the right hand side above. It is easy to check from the definition of multiplication in (1) that the resulting uniquely defined map $\Phi : \mathcal{T}(M) \to A$ is an R-algebra homomorphism. □

Definition 7.14 (Tensor Algebra) *The ring $\mathcal{T}(M)$ is called the **tensor algebra** of M.*

Proposition 7.15 *Let $\mathbf{V}$ be a finite dimensional vector space over the field $\mathbb{F}$ with basis $\mathcal{B} = \{v_1, \cdots , v_n\}$. Then the k-tensors*

$$v_{i_1} \otimes v_{i_n} \otimes \cdots \otimes v_{i_k} \quad \text{with } v_{i_j} \in \mathcal{B}$$

are a vector space basis of $\mathcal{T}^k(\mathbf{V})$ over F (when $k = 0$, the basis vector is the element $1 \in F$). In particular, $\dim_F(\mathcal{T}^k(\mathbf{V})) = n^k$.

Proof: Clear. □

Remark 7.15.1 *This in fact shows that $\mathcal{T}(\mathbf{V})$ may be regarded as the noncommutative polynomial algebra over F in the (noncommuting) variables $v_1, \cdots , v_n$. The analogous result also holds for finitely generated free modules over any commutative ring.*

Definition 7.16

- *A ring S is called a **graded ring** if it is the direct sum of additive subgroups: $S = S_0 \oplus S_1 \oplus S_2 \oplus \cdots$ such that $S_i S_j \subseteq S_{i+j}$ for all $i, j \geq 0$. The elements of S_k are said to be **homogeneous of degree** k, and S_k is called the **homogeneous component of** S **of degree** k.*
- *An ideal I of the graded ring S is called a **graded ideal** if $I = \bigoplus_{k=0}^{\infty}(I \cap S_k)$.*
- *A ring homomorphism $\varphi : S \to T$ between two graded rings is called a **homomorphism of graded rings** if it respects the grading structures on S and T, i.e., if $\varphi(S_k) \subseteq T_k$ for $k = 0, 1, 2, \cdots$.*

Remark 7.16.1 *Note that $S_0 S_0 \subseteq S_0$, which implies that S_0 is a subring of the graded ring S and then S is an S_0-module. If S_0 is in the center of S and it contains an identity of S, then S is an S_0-algebra. Note also that the ideal I is graded if whenever a sum $i_{k_1} + \cdots + i_{k_n}$ of homogeneous elements with distinct degrees $k_1, \cdots , k_n$ is in I then each of the individual summands $i_{k_1}, \cdots , i_{j_n}$ is itself in I.*

Example:
The polynomial ring $S = R[x_1, \cdots, x_n]$ in n variables over the commutative ring R is an example of a graded ring. Here $S_0 = R$ and the homogeneous component of degree k is the subgroup of all R-linear combinations of monomials of degree k.
The ideal I generated by $x_1, \cdots, x_n$ is a graded ideal: every polynomial with zero constant term may be written uniquely as a sum of homogeneous polynomials of degree $k > 1$, and each of these has zero constant term hence lies in I. More generally, an ideal is a graded ideal if and only if it can be generated by homogeneous polynomials.

Proposition 7.17 *Let S be a graded ring, let I be a graded ideal in S and let $I_k = I \cap S_k$ for all $k \geq 0$. Then S/I is naturally a graded ring whose homogeneous component of degree k is isomorphic to S_k/I_K.*

Proof: The map

$$S = \bigoplus_{k=0}^{\infty} S_k \longrightarrow \bigoplus_{k=0}^{\infty} (S_k/I_k)$$
$$(\cdots, s_k, \cdots) \longmapsto (\cdots, s_k \mod I_k, \cdots)$$

is surjective with kernel $I = \bigoplus_{k=0}^{\infty} I_k$ and defines an isomorphism of graded rings. □

7.2 Symmetric Algebras

Definition 7.18 (Symmetric Algebra) *The **symmetric algebra** of an R-module M is the R-algebra obtained by taking the quotient of the tensor algebra $\mathcal{T}(M)$ by the ideal $\mathcal{C}(M)$ generated by all elements of the form $m_1 \otimes m_2 - m_2 \otimes m_1$, for all $m_1, m_2 \in M$. The symmetric algebra $\mathcal{T}(M)/\mathcal{C}(M)$ is denoted by $\mathcal{S}(M)$.*

Remark 7.18.1 *The tensor algebra $\mathcal{T}(M)$ is generated as a ring by $R = \mathcal{T}^0(M)$ and $M = \mathcal{T}^1(M)$,, and these elements commute in the quotient ring $\mathcal{S}(M)$ by definition. It follows that the symmetric algebra $\mathcal{S}(M)$ is a commutative ring. The ideal $\mathcal{C}(M)$ is generated by the homogeneous tensors of degree 2 and it follows that $\mathcal{C}(M)$ is a graded ideal. Then by Proposition 7.17, the symmetric algebra is a graded ring whose homogeneous component of degree k is $\mathcal{S}^k(M) = \mathcal{T}^k(M)/\mathcal{C}^k(M)$. Since $\mathcal{C}(M)$ consists of k-tensors with $k \geq 2$, we have $\mathcal{C}(M) \cap M = 0$ and so the image of $M = \mathcal{T}^1(M)$ in $\mathcal{S}(M)$ is isomorphic to M. identifying M with its image, we see that $\mathcal{S}^1(M) = M$ and the symmetric algebra contains M. In a similar way $\mathcal{S}^0(M) + R$, so the symmetric algebra is also an R-algebra. The R-module $\mathcal{S}^K(M)$ is called the k^{th} **symmetric power** of M.*

Definition 7.19 (Symmetric Map) *A k-multilinear map $\varphi : M \times \cdots \times M \to N$ is said to be **symmetric** if $\varphi(m_1, \cdots, m_k) = \varphi(m_{\sigma(1)}, \cdots, m_{\sigma(k)}$ for all permutations σ of $1, 2, \cdots, k$.*

Theorem 7.20 *Le M be an R-module over the commutative ring R and let $\mathcal{S}(M)$ be its symmetric algebra.*

1. *The k^{th} symmetric power, $S^k(M)$, of M is equal to $M \otimes \cdots \otimes M$ (k factors) modulo the submodule generated by all elements of the form*

$$(m_1 \otimes m_2 \otimes \cdots \otimes m_k) - (m_{\sigma(1)} \otimes m_{\sigma(2)} \otimes \cdots \otimes m_{\sigma(k)})$$

for all $m_i \in M$ and all permutations σ in the symmetric group S_k.

2. *(Universal Property for Symmetric Multilinear Maps) If $\phi : M \times \cdots \times M \to N$ is a symmetric k-multilinear map over R then there is a unique R-module homomorphism $\phi' : S^k(M) \to N$ such that $\phi = \phi' \circ u$, where*

$$u : M \times \cdots \times M \to S^k(M)$$

is the map defined by

$$u(m_1, \ldots, m_k) = m_1 \otimes \cdots \otimes m_n \mod C(M).$$

3. *(Universal Property for maps to commutative R-algebras) If A is any commutative R-algebra and $\phi : M \to A$ is an R-module homomorphism, then there is a unique R-algebra homomorphism $\phi' : S(M) \to A$ such that $\phi M = \phi'$.*

Proof: The k-tensors $C^k(M)$ in the ideal $C(M)$ are finite sums of elements of the form

$$m_1 \otimes \cdots \otimes m_{i-1} \otimes (m_i \otimes m_{i+1} - m_{i+1} \otimes m_i) \otimes m_{i+2} \otimes \cdots \otimes m_k$$

with $m_1, \ldots, m_k \in M$ (where $k \geq 2$ and $1 \leq i < k$). This product gives a difference of two k-tensors which are equal except that two entries (in positions i and $i+1$) have been transposed, i.e., gives the element in (1) of the theorem corresponding to the transposition $(i\ i+1)$ in the symmetric group S_k. Conversely, since any permutation σ in S_k can be written as a product of such transpositions it is easy to see that every element in (1) can be written as a sum of elements of the form above. This gives (1).
The proofs of (2) and (3) are very similar to the proofs of the corresponding "asymmetric"' results, noting that $C^k(M)$ is contained in the kernel of any symmetric map from $T^k(M)$ to N by part (1). □

Corollary 7.20.1 *Let $\mathbf{V}$ be an n-dimensional vector space over the field F. Then $\mathcal{S}(\mathbf{V})$ is isomorphic as a graded F-algebra to the ring of polynomials in n variables over F (i.e., the isomorphism is also a vector space isomorphism from $\mathcal{S}^k(\mathbf{V})$ onto the space of all homogeneous polynomials of degree k). In particular,* $\dim_F(\mathcal{S}^k(\mathbf{V})) = \binom{k+n-1}{n-1}$.

Proof: Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis of $\mathbf{V}$. By Proposition 7.15 there is a bijection between a basis of $\mathcal{T}^k(\mathbf{V})$ and the set $\mathcal{B}^k$ of ordered k-tuples of elements from $\mathcal{B}$. Define two k-tuples in $\mathcal{B}^k$ to be equivalent if there is some permutation of the entries of one that gives the other – this is easily seen to be an equivalence relation on $\mathcal{B}^k$. Let $\mathcal{S}(\mathcal{B}^k)$ denote the corresponding set of equivalence classes. Any symmetric k-multilinear function from $\mathbf{V}^k$ to a vector space over F will be constant on all of the basis tensors whose corresponding k-tuples lie in the same equivalence class; conversely, any function from $S(B^k)$ can be uniquely extended to a symmetric k-multilinear function on V^k. It follows that the vector space over F with basis $\mathcal{S}(\mathcal{B}^k)$ satisfies the universal property of $\mathcal{S}^k(V)$, hence is isomorphic to $\mathcal{S}^k(\mathbf{V})$. Each equivalence class has a unique representative of the form $(u_1^{a_1}, u_2^{a_2}, \ldots, u_n^{a_n})$, where $u_i^{a_i}$ denotes the sequence $u_i, \cdots, u_i$ taken a_i times, each $a_i \geq 0$, and $a_1 + a_2 + \cdots + a_n = k$. Thus there is a bijection between the basis $\mathcal{S}^k(\mathcal{B})$ and the set $x_1^{a_1} \cdots x_n^{a_n}$ of monic monomials of degree k in the polynomial ring $F[x_1, \cdots, x_n]$. This bijection extends to an isomorphism of graded F-algebras, proving the first part of the corollary. The second part is a direct computation. □

7.3 Exterior Algebras

Definition 7.21 (Exterior Algebra) *The **exterior algebra** of an R-module M is the R-algebra contained by taking the quotient of the tensor algebra $\mathcal{T}(M)$ by the ideal $\mathcal{A}(M)$ generated by all elements of the form $m \otimes m$, for $m \in M$. The exterior algebra $\mathcal{T}(M)/\mathcal{A}(M)$ is denoted by $\bigwedge(M)$ and the image of $m_1 \otimes m_2 \otimes \cdots \otimes m_k$ is $\bigwedge(M)$ is denoted by $m_1 \wedge m_2 \wedge \cdots \wedge m_k$.*
The multiplication

$$(m_1 \wedge \cdots \wedge m_i) \wedge (m'_1 \wedge \cdots \wedge m'_j) = m_1 \wedge \cdots \wedge m_i \wedge m'_1 \wedge \cdots \wedge m'_j$$

*in the exterior algebra is called the **wedge / exterior product**.*

Remark 7.21.1 *Similarly to symmetric algebra, the ideal $\mathcal{A}(M)$ is generated by homogeneous elements hence is a graded ideal. Then by Proposition 7.17, the exterior algebra is graded, with k^{th} homogeneous component $\bigwedge^k(M) = \mathcal{T}^k(M)/\mathcal{A}^k(M)$. We can identify R with $\bigwedge^0(M)$ and M with $\bigwedge^1(M)$ and so consider M as an R-submodule of the R-algebra $\bigwedge(M)$. The R-module $\bigwedge^k(M)$ is called the k^{th} **exterior power** of M.*

On simple tensors, we have anticommutativity, that is

$$m \wedge m' = -m' \wedge m, \quad \forall m, m' \in M.$$

However, this may not extend to arbitrary products.

Theorem 7.22 *Let M be an R-module over the commutative ring R and let $\bigwedge(M)$ be its exterior algebra.*

1. *The k^{th} exterior power, $\bigwedge^k(M)$, of M is equal to $M \otimes \cdots \otimes M$ (k factors) modulo the submodule generated by all elements of the form*

   $$m_1 \otimes m_2 \otimes \cdots \otimes m_k \text{ where } m_i = m_j \text{ for some } i \neq j.$$

   In particular,

   $$m_1 \wedge m_2 \wedge \cdots \wedge m_k = 0 \text{ if } m_i = m_j \text{ for some } i \neq j.$$

2. *(Universal Property for Alternating Multilinear Maps) If $\phi : M \times \cdots \times M \to N$ is an alternating k-multilinear map then there is a unique R-module homomorphism $\varphi' : \bigwedge^k(M) \to N$ such that $\varphi = \varphi' \circ u$, where*

   $$u : M \times \cdots \times M \to \bigwedge^k(M)$$

   is the map defined by

   $$u(m_1, \ldots, m_k) = m_1 \wedge \cdots \wedge m_k.$$

Remark 7.22.1 *The exterior algebra also satisfies a universal property similar to (3) of Theorem 7.20, namely with respect to R-module homomorphisms from M to R-algebras A satisfying $a^2 = 0$ for all $a \in A$*

Proof: The k-tensors $\mathcal{A}^k(M)$ in the ideal $\mathcal{A}(M)$ are finite sums of elements of the form

$$m_1 \otimes \ldots \otimes m_{i-1} \otimes (m_i \otimes m_j) \otimes m_{i+2} \otimes \cdots \otimes m_k$$

with $m_1, \ldots, m_k \in M$ (where $k \geq 2$ and $1 \leq i < k$), which is a k-tensor with two equal entries (in positions i and $i+1$), so is of the form in (1). For the reverse inclusion, note that since

$$m' \otimes m = -m \otimes m' + (m+m') \otimes (m+m') - m \otimes m' = -m \otimes m' \mod A(M),$$

$$m' \otimes m = -m \otimes m' \mod \mathcal{A}(M),$$

interchanging any two consecutive entries and multiplying by -1 in a simple k-tensor gives an equivalent tensor modulo $\mathcal{A}^k(M)$. using such a sequence of interchanges and sign changes we can arrange for the equal entries m_i an dm_j of a simple tensor as in (1) to be adjacent, which gives an element of $\mathcal{A}^k(M)$. It follows that the generators in (1) are contained in $\mathcal{A}^k(M)$, which proves the first part of the theorem.

The proof for (2) follows from the corresponding result for the tensor algebra. □

Corollary 7.22.1 *Let* $\mathbf{V}$ *be a finite dimensional vector space over the field* $\mathbb{F}$ *with basis* $\mathcal{B} = \{v_1, \cdots, v_n\}$. *Then the vectors*

$$v_{i_1} \wedge v_{i_2} \wedge \cdots \wedge v_{i_k}, \quad \text{for } 1 \leq i_1 < i_2 < \cdots < i_k \leq n$$

are a basis of $\bigwedge^k(\mathbf{V})$, *and* $\bigwedge^k(\mathbf{V}) = 0$ *where* $k > n$ *(when* $k = 0$ *the basis vector is the element* $1 \in F$*). In particular,* $\dim_F(\bigwedge^k(\mathbf{V})) = \binom{n}{k}$.

Remark 7.22.2 *The statement is also true for any free R-module of rankn. In particular, if* $M \cong R^n$ *with R-module basis* $m_1, \cdots, m_n$, *then*

$$\bigwedge^n(M) = R(m_1 \wedge \cdots \wedge m_n)$$

if a free (rank 1) R-module with generator $m_1 \wedge \cdots \wedge m_n$ *and*

$$\bigwedge^{n+1}(M) = \bigwedge^{n+2}(M) = \cdots = 0.$$

7.4 Homomorphism of Tensor Algebras

If $\varphi : M \to N$ is any R-module homomorphism, then there is an induced map on the k^{th} tensor power:

$$\mathcal{T}^k(\varphi) : m_1 \otimes m_2 \otimes \cdots \otimes m_k \longmapsto \varphi(m_1) \otimes \varphi(m_2) \otimes \cdots \otimes \varphi(m_k).$$

It follows directly that this map sends generators of each of the homogeneous components of the ideal $\mathcal{C}(M)$ and $\mathcal{A}(M)$ to themselves. Thus φ induces R-module homomorphisms on the quotients:

$$\mathcal{S}^k(\varphi) : \mathcal{S}^k(M) \to \mathcal{S}^k(N), \quad \text{and} \quad \bigwedge^k(\varphi) : \bigwedge^k(M) \to \bigwedge^k(N).$$

Moreover, each of these three maps is a ring homomorphism (hence they are graded R-algebra homomorphism).

Proposition 7.23 *If φ is an endomorphism on a n-dimensional vector space $\mathbf{V}$, then $\bigwedge^n(\varphi)(w) = \det(\varphi)w$ for all $w \in \bigwedge^n(\mathbf{V})$.*

Proof: Since the space of n-linear alternating map is of dimension 1. □

The maps $\mathcal{S}^k(\varphi)$ and $\bigwedge^k(\varphi)$ induced from an injective map from M to N need not remain injective. So $\bigwedge^2(M)$ need not be a submodule of $\bigwedge^2(N)$ when M is a submodule of N. However, one can show that if M is an R-module direct summand of N, then $\mathcal{T}(M)$ (respectively, $\mathcal{S}(M)$ and $\bigwedge(M)$) is an R-subalgebra of $\mathcal{T}(N)$ (respectively, $\mathcal{S}(N)$ and $\bigwedge(N)$). When $R = F$ is a field, then every subspace M of N is a direct summand of N and so the corresponding algebra for M is a subalgebra of the algebra for N.

7.5 Symmetric and Alternating Tensors

We can also define symmetric and exterior algebras in terms of symmetric and alternating tensors, which identifies these algebras as subalgebras of the tensor algebra rather than as quotient algebras.

For any R-module M, there is a natural left group action of the symmetric group S_k on $M \times \cdots \times M$ (k factors) given by permuting the factors:

$$\sigma(m_1, m_2, \cdots, m_k) = (m_{\sigma^{-1}(1)}, m_{\sigma^{-1}(2)}, \cdots, m_{\sigma^{-1}(k)}), \quad \forall \sigma \in S_k.$$

So there is a well defined R-linear left group action of S_k on $\mathcal{T}^k(M)$ which is defined on simple tensors by

$$\sigma(m_1 \otimes m_2 \otimes \cdots \otimes m_k) = m_{\sigma^{-1}(1)} \otimes m_{\sigma^{-1}(2)} \otimes \cdots \otimes m_{\sigma^{-1}(k)}.$$

Definition 7.24 (Symmetric And Alternating Tensors)

- *An element $z \in \mathcal{T}^k(M)$ is called a **symmetric k-tensor** if $\sigma z = z$ for all σ in the symmetric group S_k.*
- *An element $z \in \mathcal{T}^k(M)$ is called an **alternating k-tensor** if $\sigma z = \operatorname{sgn}(\sigma)z$ for all σ in the symmetric group S_k.*

Proposition 7.25 *Let σ be an element in the symmetric group S_k, then*

- *for every $w \in \mathcal{S}^k(M)$, we have $\sigma w = w$;*
- *for every $w \in \bigwedge^k(M)$, we have $\sigma w = \operatorname{sign}(\sigma)w$.*

Remark 7.25.1 *This shows that S_k acts trivially on both the submodule of symmetric k-tensors and the quotient module $\mathcal{S}^k(M)$; Similarly, $S - k$ acts the same way on the submodule of alternating k-tensors as on $\bigwedge^k(M)$. When $k! = k! \cdot 1$ is a unit in R,we will show in the next theorem that these respective submodules and quotient modules are isomorphic.*

Definition 7.26 (Sym and Alt) *For any k-tensor $z \in \mathcal{T}^k(M)$, we define Sym and Alt from $\mathcal{T}^k$ to $\mathcal{T}^k$ by:*

$$Sym(z) = \sum_{\sigma \in S_k} \sigma z$$
$$Alt(z) = \sum_{\sigma \in S_k} \text{sign}(\sigma)\sigma z.$$

*The tensor $Sym(z)$ is called the **symmetrization of** z and $Alt(z)$ is called the **skew-symmetrization of** z).*

Proof: Note for any $z \in \mathcal{T}^k$, $\tau Sym(z) = Sym(z)$, and $\tau Alt(z) = \text{sign}(\tau)Alt(z)$ for all $\tau \in S_k$. It is also clear to note that

$$Sym(z) = k!z, \quad \text{for any symmetric tensor z}$$
$$Alt(z) = k!z,$$
$$, \quad \text{for any alternating tensor z}$$

□

Theorem 7.27 *Suppose $k!$ is a unit in the ring R and M is an R-module. Then*

- *The map $(1/k!)Sym$ induces an R-module isomorphism between the k^{th} symmetric power of M and the R-submodule of symmetric k-tensors.*
- *The map $(1/k!)Alt$ induces an R-module isomorphism between the k^{th} exterior power of M and the R-submodule of alternating k-tensors.*

Remark 7.27.1 *The maps $(1/k!)Sym$ and $(1/k!)Alt$ are projections onto the submodules of symmetric and anti-symmetric tensors, respectively. Hence if $k!$ is a unit in R, we have R-module direct sums*

$$\mathcal{T}^k(M) = \ker(\pi) \oplus \text{image}(\pi)$$

for $\pi = (1/k!)Sym$ or $\pi = (1/k!)Alt$. In the former case, the kernel consists of $\mathcal{C}^k(M)$ and the image is the collection of symmetric tensors; in the latter case, the kernel if $\mathcal{A}^k(M)$ and the image consists of the alternating tensors.

Remark 7.27.2 *If $k!$ is not invertible in R then in general we do not have such S_k-invariant direct sum decompositions so it is not in general possible to identify, for example, the k^{th} exterior power of M with the alternating k-tensors of M.*

8 Modules Over P.I.D

8.1 The Basic Theory

Definition 8.1 (Noetherian) *The left R-module M is said to be a **Noetherian R-module** or to satisfy the **ascending chain condition on submodules** if there are no infinite increasing chains of submodules, i.e., whenever*

$$M_1 \subseteq M_2 \subseteq M_2 \subseteq \cdots$$

*is an increasing chain of submodules of M, then there is a positive integer m such that for all $k \geq m$, $M_k = M_m$. The ring R is said to be **Noetherian** if it is Noetherian as a left module over itself, i.e., there are no infinite increasing chains of left ideals in R.*

Theorem 8.2 *Let R be a ring and let M be a left-module. Then the following are equivalent:*

1. *M is a Noetherian R-module.*
2. *Every nonempty set of submodules of M contains a maximal element under inclusion.*
3. *Every submodule of M is finite generated.*

Proof: $1 \Rightarrow 2$: Assume M is Noetherian and let Σ be any nonempty collection of submodules of M, choose any $M_1 \in \Sigma$. If M_1 is a maximal element of Σ, then (2) holds, so assume M_1 is not maximal. Then there is some $M_2 \in \Sigma$ such that $M_1 \subset M_2$. If M_2 is maximal in Σ, (2) holds, so we may assume there is an $M_3 \in \Sigma$ properly containing M_2. Proceeding in this way one sees that if (2) fails, we can produce an infinite strictly increasing chain of elements of Σ, contrary to (1).

$2 \Rightarrow 3$: Assume (2) holds and let N be any submodule of M. Let Σ be the collection of all finitely generated submodules of N. Since $\{0\} \in \Sigma$, this collection is nonempty. By (2), Σ contains a maximal element N'. If $N' \neq N$, let $x \in N - N'$. Since $N' \in \Sigma$, the submodule N' is finitely generated by assumption, hence also the submodule generated by N' and x is finite generated. This contradicts the maximality of N', so $N = N'$ is finitely generated.

$3 \Rightarrow 1$: Assume (3) holds and let $M_1 \subset M_2 \subset M_3 \cdots$ be a chain of submodules of M. Let

$$N = \bigcup_{i=1}^{\infty} M_i,$$

then N is a submodule of M. By (3), N is finitely generated, by $a_1, \cdots, a_n$. Since $a_i \in N$ for all i, each a_i lies in one of the submodules in the chain, say M_{j_i}, let $m = \max\{j_1, j_2, \cdots, j_n\}$. Then $a_i \in M_m$ for all i so the module they generate is contained in M_m, i.e., $N \subset M_n$. This implies $M_m = N = M_k$ for all $k \geq m$, which proves (1). □

Corollary 8.2.1 *If R is a P.I.D. then every nonempty set of ideals of R has a maximal element and R is a Noetherian ring.*

Proof: The every submodule of the P.I.D. R is a left R-ideal, hence finitely generated. □

Remark 8.2.1 *If M itself is a finitely generated R-module, submodules of M need not to be finitely generated, so the condition that M be a Noetherian R-module is in general stronger than the condition that M be a finitely generated R-module.*

Proposition 8.3 *Let R be an integral domain and let M be a free R-module of rank $n < \infty$. Then any $n+1$ elements of M are R-linear independent, i.e., for any $y_1, y_2, \cdots, y_{n+1} \in M$ there are elements $r_1, \cdots, r_{n+1} \in R$, not all zero, such that*

$$r_1 y_1 + r_2 y_2 + \cdots + r_{n+1} y_{n+1} = 0.$$

Proof: We embed R in its quotient field F (since R is an integral domain) and observe that since $M \cong R \oplus R \oplus \cdots \oplus R$ (n times) we obtain $M \subseteq F \oplus F \oplus \cdots \oplus F$. The latter is an n-dimensional vector space over F so any $n+1$ elements of M are F-linearly dependent. By clearing the denominators of the scalars (by multiplying through by the product of all the denominators, for example), we obtain an R-linear dependence relation among the $n+1$ elements of M. □

Definition 8.4 (Torsion and Annihilator) *If R is any integral domain and M is any R-module. Then we define the* ***torsion submodule of*** *M to be the set*

$$\mathrm{Tor}(M) = \{x \in M \,|\, rx = 0, \quad \textit{for some nonzero } r \in R\}.$$

If $\mathrm{Tor}(M) = 0$, then module M is said to be ***torsion free****.*
For any submodule N of M, the ***annihilator*** *of N is the ideal of R defined by*

$$\mathrm{Ann}(N) = \{r \in R \,|\, rn = 0, \ \textit{for all } n \in N\}.$$

Remark 8.4.1 *If N is not a submodule of the torsion submodule of M, then $\mathrm{Ann}(N) = (0)$. It is also easy to see that if N, L are submodules of M with $N \subset L$, then $\mathrm{Ann}(L) \subset \mathrm{Ann}(N)$. If R is a P.I.D., and $N \subset L \subset M$, with $\mathrm{Ann}(N) = (a)$ and $\mathrm{Ann}(L) = (b)$, then $a|b$; In particular, the annihilator of any element of x of N divides the annihilator of M.*

Definition 8.5 (Rank of Modules) *For any integral domain R, the* ***rank*** *of an R-module M is the maximum number of R-linear independent elements of M.*

Theorem 8.6 *Let R be a P.I.D., let M be a free R-module of finite rank n and let N be a submodule of M. Then*

1. *N is free of rank m, $m \leq n$ and*
2. *there exists a basis $y_1, y_2, \cdots, y_n$ of M so that $a_1 y_1, a_2 y_2, \cdots, a_m y_m$ is a basis of N where $a_1, a_2, \cdots, a_m$ are nonzero elements of R with the divisibility relations*

$$a_1 | a_2 | \cdots | a_m.$$

Proof: The theorem is trivial for $N = \{0\}$, so assume $N \neq \{0\}$. For each R-module homomorphism φ of M into R, the image $\varphi(N)$ of N is a submodule of R, i.e., an idael in R. Since R is a P.I.D. this ideal must be principal, say $\varphi(N) = (a_\varphi)$, for some $a_\varphi \in R$. Let

$$\Sigma = \{(a_\varphi) \,|\, \varphi \in \mathrm{Hom}_R(M, R)\}$$

be the collection of the principal ideals in R obtained in this way from the R-module homomorphisms of M into R. The collection Σ is certainly nonempty since taking φ to be the trivial homomorphism shows that $(0) \in \Sigma$. By Corollary 8.2.1, Σ has at least one maximal element i.e., there is at least one homomorphism ν of M to R so that the principal ideal $\nu(N) = (a_1)$ is not properly contained in any other element of Σ. Let $a_1 = a_1$ for this maximal element and let $y \in N$ be an element mapping to the generator a_1 under the homomorphism $\nu : \nu(y) = a_1$.
We now show the element a_1 is nonzero. Let $x_1, x_2, \ldots, x_n$ be any basis of the free module M and let $\pi_i \in \mathrm{Hom}_R(M, R)$ be the natural projection homomorphism onto the i-th coordinate with respect to this basis. Since $N \neq \{0\}$, there exists an i such that $\pi_i(N) \neq 0$, which in particular shows that Σ contains more than just the trivial ideal (0). Since (a_1) is a maximal element of Σ it follows that $a_1 \neq 0$.
We next show that this element a_1 divides $\varphi(y)$ for every $\varphi \in \mathrm{Hom}(M, R)$. To see this let d be a generator for the principal ideal generated by a_1 and $\varphi(y)$. Then d is a divisor of both a_1 and $\varphi(y)$ in R and $d = r_1 a_1 + r_2\varphi(y)$ for some $r_1, r_2 \in R$. Consider the homomorphism $\psi = r_1\nu + r_2\varphi$ from M to R. Then $\psi(y) = (r_1\nu + r_2\varphi)(y) = r_1 a_1 + r_2\varphi(y) = d$ so that $d \in \nu(N)$, hence also $(d) \subseteq \nu(N)$. But d is a divisor of a_1, so we also have $(a_1) \subseteq (d)$. Then $(a_1) \subseteq (d) \subseteq \nu(N)$ and by the maximality of (a_1) we must have equality: $(a_1) = (d) = \nu(N)$. In particular $(a_1) = (d)$ shows that $a_1 \mid \varphi(y)$ since d divides $\varphi(y)$.
If we apply this to the projection homomorphisms π_i we see that a_1 divides $\pi_i(y)$ for all i. Write $\pi_i(y) = a_1 b_i$ for some $b_i \in R$, $1 \leq i \leq n$ and define

$$y_1 = \sum_{i=1}^{n} b_i x_i.$$

Note that $a_1 y_1 = y$. Since $a_1 = \nu(y) = \nu(a_1 y_1) = a_1\nu(y_1)$ and a_1 is a nonzero element of the integral domain R this shows

$$\nu(y_1) = 1.$$

We now verify that this element y_1 can be taken as one element in a basis for M and that $a_1 y_1$ can be taken as one element in a basis for N, namely that we have

$$(a) \quad M = Ry_1 \oplus \ker\nu, \quad \text{and} \quad (b) \quad N = Ra_1 y_1 \oplus (N \cap \ker\nu).$$

To see (a) let x be an arbitrary element in M and write $x = \nu(x)y_1 + (x - \nu(x)y_1)$. Since

$$\nu(x - \nu(x)y_1) = \nu(x) - \nu(x)\nu(y_1) = \nu(x) - \nu(x) \cdot 1 = 0.$$

We see that $x - v(x)y_1$ is an element in the kernel of v, so $M = Ry_1 + \ker v$. To see that the sum is direct, suppose ry_1 is also an element in the kernel of v. Then $0 = v(ry_1) = rv(y_1) = r$ shows that this element is indeed 0.

For (b) observe that $v(x')$ is divisible by a_1 for every $x' \in N$ by the definition of a_1 as a generator for $v(N)$. If we write $v(x') = ba_1$ where $b \in R$, then the decomposition we used in (a) shows $x' = v(x')y_1 + (x' - v(x')y_1) =$

$ba_1y_1 + (x' - ba_1y_1)$ where the second summand is in the kernel of v and is an element of N. This shows that

$$N = Ra_1y_1 + (N \cap \ker \nu).$$

The fact that the sum in (b) is direct is a special case of the directness of the sum in (a).
We now prove part (1) of the theorem by induction on the rank, m, of N. If $m = 0$, then N is a torsion module, hence $N = 0$ since a free module is torsion free, so (1) holds trivially. Assume then that $m > 0$. Since the sum in (b) above is direct we see easily that $N \cap \ker \nu$ has rank $m - 1$. By induction $N \cap \ker \nu$ is then a free R-module of rank $m - 1$. Again by the directness of the sum in (b) we see that adjoining a_1y_1 to any basis of $N \cap \ker \nu$ gives a basis of N, so N is also free (of rank m), which proves (1).
Finally, we prove (2) by induction on n, the rank of M. Applying (1) to the submodule $\ker \nu$ shows that this submodule is free and because the sum in (a) is direct it is free of rank $n-1$. By the induction assumption applied to the module $\ker \nu$ (which plays the role of M) and its submodule $\ker \nu \cap N$ (which plays the role of N), we see that there is a basis $y_2, y_3, \ldots, y_n$ of $\ker \nu$ such that $a_2y_2, a_3y_3, \ldots, a_my_m$ is a basis of $N \cap \ker \nu$ for some elements $a_2, a_3, \ldots, a_m$ of R with $a_2 \mid a_3 \mid \cdots \mid a_m$. Since the sums (a) and (b) are direct, $y_1, y_2, \ldots, y_n$ is a basis of M and $a_1y_1, a_2y_2, \ldots, a_my_m$ is a basis of N. To complete the induction it remains to show that a_1 divides a_2. Define a homomorphism φ from M to R by defining $\varphi(y_1) = \varphi(y_2) = 1$ and $\varphi(y_i) = 0$, for all $i > 2$, on the basis for M. Then for this homomorphism φ we have $a_1 = \varphi(a_1y_1)$ so $a_1 \in \varphi(N)$ hence also $(a_1) \subseteq \varphi(N)$. By the maximality of (a_1) in Σ it follows that $(a_1) = \varphi(N)$. Since $a_2 = \varphi(a_2y_2) \in \varphi(N)$ we then have $a_2 \in (a_1)$ i.e., $a_1 \mid a_2$. This completes the proof of the theorem. □

Definition 8.7 (Cyclic Modules) *A left R-module C is called a **cyclic R-module** (for any ring R, not necessarily commutative nor with 1), if there is an element $x \in C$ such that $C = Rx$.*

Theorem 8.8 (Fundamental Theorem, Existence: Invariant Factor Form) *Let R be a P.I.D. and let M be a finitely generated R-module.*

1. *Then M is isomorphic to the direct sum of finitely many cyclic modules. More precisely*

$$M \cong R^r \oplus R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m)$$

for some integer $r \geq 0$ and nonzero elements $a_1, a_2, \cdots, a_m$ of R which are not units in R and which satisfy the divisibility relations

$$a_1|a_2|\cdots|a_m.$$

2. *M is torsion free if and only if M is free.*

3. *In the decomposition in* (1),

$$\mathrm{Tor}(M) \cong R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m).$$

In particular M is a torsion module if and only if $r = 0$ and in this case the annihilator of M is the ideal (a_m).

Remark 8.8.1 *The integer r in Theorem 8.8 is called the* ***free rank*** *of the* ***Betti number*** *of M and the elements $a_1, a_2, \cdots, a_m \in R$ are called the* ***invariant factors*** *of M.*

Theorem 8.9 (Fundamental Theorem, Existence: Elementary Divisor Form) *Let R be a P.I.D. and let M be a finitely generated R-module. Then M is the direct sum of a finite number of cyclic modules whose annihilator are either (0) of generated by powers of primes in R, i.e.,*

$$M \cong R^r \oplus R/(p_1^{\alpha_1}) \oplus R/(p_2^{\alpha_2}) \oplus \cdots \oplus R/(p_t^{\alpha_t})$$

Remark 8.9.1 *Let R be a P.I.D. and let M be a finitely generated R-module as in Theorem 8.9. The prime powers $p_1^{\alpha_1}, \cdots, p_t^{\alpha_t}$ (defined up to multiplication by units in R) are called the* ***elementary divisors*** *of M.*

Suppose M is a finitely generated torsion module over the Principal Ideal Domain R . If for the distinct primes $p_1, p_2, \cdots, p_n$, occurring in the decomposition in Theorem 8.9, we group together all the cyclic factors corresponding to the same prime p_i we see in particular that M can be written as a direct sum

$$M = N_1 \oplus N_2 \oplus \cdots \oplus N_n$$

where N_i consists of all the elements of M which are annihilated by some power of the prime p_i . This result holds also for modules over R which may not be finitely generated:

Theorem 8.10 (The primary Decomposition Theorem) *Let R be a P.I.D. and let M be a nonzero torsion R-module (not necessarily finitely generated) with nonzero annihilator a. suppose the factorization of a into distinct prime powers in R is*

$$a = up_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$$

and let $N_i = \{x \in M \,|\, p_i^{\alpha_i} x = 0\}$, $1 \leq i \leq n$. Then N_i is a submodule of M with annihilator $p_i^{\alpha_i}$ and is the submodule of M of all elements annihilated by some power of p_i. We have

$$M = N_1 \oplus N_2 \oplus \cdots \oplus N_n.$$

If M is finitely generated then each N_i is the direct sum of finitely many cyclic modules whose annihilators are divisors of $p_i^{\alpha_i}$

Proof: We only need to show for the case where M is infinitely generated over R. It is clear that N_i is a submodule of M with annihilator dividing $p_i^{\alpha_i}$, since R is a P.I.D., the ideals $(p_i^{\alpha_i})$ and $(p_j^{\alpha_j})$ are comaximal for $i \neq j$, so the direct sum decomposition of M can be shown easily. Using this direct sum decomposition, it is easy to see that the annihilator of N_i is precisely $p_i^{\alpha_i}$. □

Remark 8.10.1 *The submodule N_i in the theorem is called the* ***p_i-primary component of*** *M.*

Lemma 8.11 *Let R be a P.I.D. and let P be a prime in R. Let F denote the field $R/(p)$.*

1. *Let $M = R^r$. Then $M/pM \cong F^r$.*

2. *Let $M = R/(a)$, where a is a nonzero element of R. Then*

$$M/pM \cong \begin{cases} F & \text{if } p|a \text{ in } R \\ 0 & \text{if } p \nmid a \text{ in } R \end{cases}.$$

3. *Let $M = R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_k)$ where each a_i is divisible by p. Then $M/pM \cong F^k$.*

Proof:

1. There is a natural map from R^r to $(R/(p))^r$ defined by mapping $(\alpha_1, \cdots, \alpha_r)$ to $(\alpha_1 \mod (p), \cdots, \alpha_r \mod (p))$. This is clearly a surjective R-module homomorphism with kernel consisting of the r-tuples all of whose coordinates are divisible by p, i.e., pR^r, so $R^r/pR^r \cong (R/(p))^r$, which gives (1).

2. This follows from the Isomorphism Theorems: note first that $p(R/(a))$ is the image of the ideal (p) in the quotient $R/(a)$, hence is $(p) + (a)/(a)$. The ideal $(p) + (a)$ is generated by the greatest common divisor of p and a, hence is (p) is p divides a is $R = (1)$ otherwise. Hence $pM = (p)/(a)$ if p divides a and is $R/(a) = M$ otherwise. If p divides a, then $M/pM = (R/(a))/((p)/(a)) \cong R/(p)$, and if p does not divid a, then $M/pM = M/M = 0$, which gives (2).

3. (3) follows from (2).

□

Theorem 8.12 (Fundamental Theorem, Uniqueness) *Let R be a P.I.D.*

1. *Two finitely generated R-modules M_1, m_2 are isomorphic if and only if they have the same free rank and the same list of invariant factors.*

2. *Two finitely generated R-modules M_1 and M_2 are isomorphic if and only if they have the same free rank and the same list of elementary divisors.*

Corollary 8.12.1 *Let R be a P.I.D. and let M be a finitely generated R-module.*

1. *The elementary divisor of M are the prime power factors of the invariant factor of M.*

2. *The largest invariant of M is the product of the largest of the distinct prime powers among the elementary divisors of M, the next largest invariant factor is the product of the distinct prime powers among the remaining elementary divisors of M, and so on.*

Corollary 8.12.2 (The Fundamental Theorem of Finitely Generated Abelian Groups)

8.2 The Rational Canonical Form

Definition 8.13 (Minimal Polynomial) *The unique monic polynomial which generates the ideal* $\mathrm{Ann}(\mathbf{V})$ *in* $\mathbb{F}[x]$ *is called the* ***minimal polynomial of*** *T and will be denoted $m_T(x)$, where $\mathbf{V}$ here is since as a module over $\mathbb{F}[x]$ via the transformation T.*

Proposition 8.14 *The minimal polynomial $m_T(x)$ is the largest invariant factor of $\mathbf{V}$. All the invariant factors of $\mathbf{V}$ divide $m_T(x)$.*

Proof: Follows from the definition of minimal polynomials. □

Definition 8.15 (Rational Canonical Form) *A matrix is said to be in **rational canonical form** if it is the direct sum of companion matrices for monic polynomials $a_1(x), \cdots, a_m(x)$ of degree at least one with $a_1(x)|a_2(x)|\cdots|a_m(x)$. The polynomials $a_i(x)$ are called the **invariant factors** of the matrix. Such a matrix is also said to be a **block diagonal matrix** with blocks the companion matrices for the $a_i(x)$. A **rational canonical form** for a linear transformation T is a matrix representing T which is in rational canonical form.*

Theorem 8.16 (Rational Canonical Form for Linear Transformation) *Let $|V$ be a finite dimensional vector space over the field F and let T be a linear transformation of $\mathbf{V}$.*

1. *There is a basis for $\mathbf{V}$ with respect to which the matrix for T is in rational canonical form.*
2. *The rational canonical form for T is unique.*

Corollary 8.16.1 *Let S and T be linear transformations of $\mathbf{V}$. Then the following are equivalent:*

1. *S and T are similar linear transformations.*
2. *the $F[x]$-modules obtained from $\mathbf{V}$ via S and via T are isomorphic $F[x]$-modules.*
3. *S and T have the same rational canonical form.*

Corollary 8.16.2 (Rational Canonical Form for Matrices) *Let A be an $n \times n$ matrix over the field F.*

1. *The matrix A is similar to a matrix in rational canonical form.*
2. *The rational canonical form for A is unique.*
3. *Let B be another $n \times n$ matrix over the field F. Then A and B are similar if and only if A and B have the same rational canonical form.*

Corollary 8.16.3 *Let A and B be two $n \times n$ matrices over a field F and suppose F is a subfield of the field K.*

1. *The rational canonical form of A is the same whether it is computed over K or over F. The minimal and characteristic polynomials and the invariant factors of A are the same whether A is consider as a matrix over F or as a matrix over K.*
2. *The matrices A and B are similar over K if and only if they are similar over F, i.e., there exists an invertible $n \times n$ matrix P with entries from K such that $B = P^{-1}AP$ if and only if there exists an (in general different) invertible $n \times n$ matrix Q with entries from F such that $B = Q^{-1}AQ$.*

Theorem 8.17 (Smith Normal Form) *Let A be an $n \times n$ matrix over the field $\mathbb{F}$. Using the three elementary row and column operations, the $n \times n$ matrix $xI - A$ with entries from $F[x]$ can be put into the diagonal form*

$$\begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & a_1(x) & & & & \\ & & & & a_2(x) & & & \\ & & & & & \ddots & & \\ & & & & & & a_m(x) \end{pmatrix}$$

with monic nonzero elements $a_1(x), a_2(x), \cdots, a_m(x)$ of $F[x]$ with degrees at least one and satisfying

$$a_1(x)|a_2(x)|\cdots|a_m(x).$$

The elements $a_1(x), \cdots, a_m(x)$ are the invariant factors of A.

9 Field Extensions

9.1 Basic Definitions

Definition 9.1 (Field) *A **Field** is a set F with two binary operations on F, called addition $+$, and multiplication $\cdot$, such that the multiplicative and additive identity are distinct, $(F,+)$ and $(F\setminus\{0\},\cdot)$ are abelian groups, and satisfies the distribution property:*

$$\forall a,b,c\in F,\ (a+b)c=ab+bc \ \text{ and } \ a(b+c)=ab+ac.$$

Remark 9.1.1 *The smallest field must contain two elements, namely the multiplicative and additive identity.*

Remark 9.1.2 *We usually denote the additive inverse of an element $a\in F$ to be $-a$ and the multiplicative inverse of an element $a\in F\setminus\{0\}$ to be $\frac{1}{a}$ or a^{-1}. Note these are unique, as $(F,+)$ and $(F\setminus\{0\},\cdot)$ are abelian groups. In addition, if n is a positive integer, $a\in F$, then we denote*

$$na=\underbrace{a+a+\cdots+a}_{n\ a\text{'s}}.$$

If n is a negative integer, $a\in F$, then we denote

$$na=\underbrace{(-a)+\cdots(-a)}_{n\ (-a)\text{'s}}.$$

Example:

- $\mathbb{Q},\mathbb{R}$, and $\mathbb{C}$ with the usual addition and multiplication are fields.
- $(\mathbb{Z}/p\mathbb{Z},+,\cdot)$ is a field, where p is a prime.
- $F=\{0,2,4,6,8\}$ with addition modulo 10 and multiplication modulo 10 is a field, with multiplicative identity being 4.

Lemma 9.2 *$\mathbb{Z}_p[i]$, which is the ring of Gaussian Integers modulo p is a field if and only if p is a prime and $p=3$ mod 4.*

Proof: It is clear that $\mathbb{Z}_p[i]$ is a field if and only if every nonzero element has a multiplicative inverse (the other axioms clearly holds). Now let $a+bi\in\mathbb{Z}_p[i]$, then if there exists $c+di\in\mathbb{Z}_p[i]$, such that $(a+bi)(c+di)=1$. we have

$$ac-bd+(ad+bc)i=1$$

That is $ac-bd=1 \mod p$ and $ad+bc=0 \mod p$. Now if p is a composite number, with $nm=p$, $n,m>1$. Then consider $a=b=m$, then $m|ac-bd$, so the first equality has no solution. Hence if p is not a prime, then $\mathbb{Z}_p[i]$ is not a field.

Next if $p=4k+1$, then we know it can be written as a sum of two squares, so we have

$$p=a^2+b^2=(a+ib)(a-ib)$$

but then this shows that $(a+ib)$ and $(a-ib)$ are zero divisors, so $\mathbb{Z}_p[i]$ cannot be a field in this case.

Lastly, we just need to show that $\mathbb{Z}_p[i]$ is an integral domain for each $p = 4k+3$. Since $\mathbb{Z}_p[i]$ is finite (has p^2 elements), then it will be a field. Suppose $a+bi$ is a zero divisor of $\mathbb{Z}_p[i]$, then for some $c+di \neq 0$, we have

$$(a+bi)(c+di) = 0 \text{ in } \mathbb{Z}_p[i].$$

That is

$$\begin{aligned} ac - bd &= 0 \mod p \\ ad + bc &= 0 \mod p \end{aligned}$$

Since $a+bi$ is a zero divisor, then at least one of a, b is not 0 in $\mathbb{Z}_p$. WLOG let it be a. Then we have $c = bda^{-1} \mod p$. Hence

$$ad + b^2da^{-1} = 0 \mod p$$

Obviously d cannot be zero, as otherwise c would be 0. So we have

$$\begin{aligned} a + b^2a^{-1} = d^{-1} \cdot 0 = 0 &\mod p \\ b^2a^{-1} = -a &\mod p \\ b^2a^{-2} = -1 &\mod p \\ (ba^{-1})^2 = -1 &\mod p \end{aligned}$$

However, we know -1 is not a Quadratic residue of p if $p = 3 \mod 4$. Hence no such zero divisors $a+bi$ exists. And $\mathbb{Z}_p[i]$ is an integral domain, so it is a field. □

Proposition 9.3 (Properties of a field) *Suppose F is any field, then $\forall x, y, z \in F$*

1. $x + y = x + z \Rightarrow y = z$.
2. $x + y = x \Rightarrow y = 0$.
3. $x + y = 0 \Rightarrow y = -x$.
4. $-(-x) = x$.
5. $x \cdot y = x \cdot z \Rightarrow y = z$.
6. $x \cdot y = x \Rightarrow y = 1$.
7. $x \cdot y = 1 \Rightarrow y = \frac{1}{x}$.
8. $\frac{1}{\frac{1}{x}} = x$.
9. $0 \cdot x = 0$.

10. $x \cdot y = 0 \Rightarrow$ *either* $x = 0$ *or* $y = 0$.

11. $(-x) \cdot y = -(x \cdot y)$.

12. $(-x) \cdot (-y) = x \cdot y$.

Proof:

1. $y = 0 + y = [(-x) + x] + y = (-x) + (x + y) = (-x) + (x + z) = [(-x) + (x)] + z = 0 + z = z$.
2. consider $x + y = x = x + 0$.
3. consider $x + y = 0 = x + (-x)$.
4. $-(-x) + (-x) = 0 = (-x) + x$.
5. $y = 1 \cdot y = \frac{1}{x} \cdot x \cdot y = \frac{1}{x} \cdot x \cdot z = z$
6. consider $x \cdot y = x = x \cdot 1$
7. consider $x \cdot y = 1 = x \cdot \frac{1}{x}$.
8. $\frac{1}{\frac{1}{x}} \cdot \frac{1}{x} = 1 = \frac{1}{x} \cdot x$.
9. consider $0 \cdot x + x = 0 \cdot x + 1 \cdot x = (0 + 1) \cdot x = x$.
10. suppose $x \neq 0$, $y = 1 \cdot y = (\frac{1}{x} \cdot x) \cdot y = \frac{1}{x} \cdot (x \cdot y) = \frac{1}{x} \cdot 0 = 0$.
11. $(-x) \cdot y + x \cdot y = (-x + x) \cdot y = 0 \Rightarrow (-x) \cdot y = -(x \cdot y)$.
12. $(-x) \cdot (-y) = -(x \cdot -y) = -[(-y) \cdot x] = -[-(y \cdot x)] = -[-(xy)] = xy$.

□

Definition 9.4 (Subfield) *A* ***subfield*** K *of a field* F *is a subset of* F*, that is itself a field with respect to the field operations of* F.

Example:

- Any field is a subfield of itself.
- $(\mathbb{R}, +, \cdot)$ is a subfield of $(\mathbb{C}, +, \cdot)$, and $(\mathbb{Q}, +, \cdot)$ is a subfield of $\mathbb{R}$.
- $\mathbb{Z}/p\mathbb{Z}$ has no subfield besides itself. A subfield of $\mathbb{Z}/p\mathbb{Z}$ must also be a subgroup of $\mathbb{Z}/p\mathbb{Z}$ with respect to addition, so its order must divides p. Hence the only possible subfield of $\mathbb{Z}/p\mathbb{Z}$ are $\{0\}$ and $\mathbb{Z}/p\mathbb{Z}$, but $\{0\}$ is not a subfield, hence the only subfield of $\mathbb{Z}/p\mathbb{Z}$ is itself.

Proposition 9.5 (Criterion for Subfield) *Suppose* F *is a field, then* $K \subset F$ *is a subfield of* F *if and only if the following holds:*

1. K *is non-empty and contains the additive and multiplicative identity of* F.
2. K *is closed under field operations.*
3. K *is closed under inverse (both additive and multiplication).*

In particular, we just need to check that $0, 1 \in K$ *and* $\forall a, b \in K$*, we have*

1. $a - b \in K$;
2. $ab^{-1} \in K$.

Proof: Consider the language of groups. The equivalence is clear. □

Corollary 9.5.1 *Suppose* $\{K_\alpha\}$ *is a collection of subfields of* F*, that is for each* $\alpha \in \Lambda$*, we have* K_α *is a subfield of* F*. Then* $\bigcap K_\alpha$ *is a subfield of* F*.*

9.2 Characteristic of a Field

Definition 9.6 (Characteristic of a Field) *Suppose* 1 *is the multiplicative identity of a field* F*. Then the* ***characteristic of a field*** $\mathrm{Char}(F)$ *is the smallest positive integer* n *such that* $n \cdot 1 = 0$*. If no such positive integer exists, then we say the field* F *has a characteristic zero.*

Example:

1. $\mathrm{Char}(\mathbb{Q}) = \mathrm{Char}(\mathbb{R}) = \mathrm{Char}(\mathbb{C}) = 0$.
2. $\mathrm{Char}(\mathbb{Z}/2\mathbb{Z}) = 2$ and $\mathrm{Char}(\mathbb{Z}/p\mathbb{Z}) = p$ if p is a prime.

Lemma 9.7 *Suppose* $\mathrm{Char}(F) = n$*. If* $n \neq 0$*, then* $1, 2 \cdot 1, 3 \cdot 1, \cdots, n \cdot 1$ *are distinct elements of the field. If* $n = 0$*, then* $\{k \cdot 1 : k \in \mathbb{N}_{\geq 0}\}$ *are distinct elements of the field.*

Proof: Suppose $n = 0$, then if $p \cdot 1 = q \cdot 1$ for some distinct $p, q \in \mathbb{N}_{\geq 0}$, then $(p - q) \cdot 1 = 0$, which contradicts the definition of 0 being the characteristic of F. If $n \neq 0$, then and $\exists 1 \leq p, q \leq n$ such that $p \neq q$ and $p \cdot 1 = q \cdot 1$. Then WLOG, let $p \leq q$, so $(q - p) \cdot 1 = 0$. Thus, we have a contradiction, because clearly $q - p$ is a smaller positive integer than n. □

Corollary 9.7.1 *Any field with characteristic zero must be infinite.*

Proposition 9.8 *The characteristic of a field is either zero or prime.*

Proof: We know a field is always an integral domain. The characteristic of a integral domain is either zero or prime. Thus the characteristic of a field is either zero or prime. □

Lemma 9.9 *If a field* F *is of characteristic* p*, where* p *is a prime. Then for any* $a \in F$ *and* $n \in \mathbb{Z}$*, we have* $(np) \cdot a = 0$*. If a field* F *is of characteristic* 0*, then for any* $a \in F$*,* $n \in \mathbb{Z}$*,* $na = 0$ *iff* $a = 0$ *or* $n = 0$*.*

Proof: If $\mathrm{Char}(F) = p$, then $(np)a = n(p \cdot 1 \cdot a) = n(p \cdot 1) \cdot a = 0 \cdot a = 0$. Then the first assertion is clear. Next if $\mathrm{Char}(F) = 0$, and $n \cdot a = 0$, then $(n \cdot 1) \cdot a = 0$. Now since any field is an integral domain, we must have that either $n \cdot 1 = 0$ or $a = 0$. But as $\mathrm{Char}(F) = 0$, then we must have either $n = 0$ or $a = 0$. □

Proposition 9.10 *Suppose F is a field of characteristic p, where p is a prime. Then $K = \{x \in F : x^p = x\}$ is a subfield of F.*

Proof: Clear 0 and 1 are elements of K. Next, let $a, b \in K$, we show $(a-b) \in K$ and $ab^{-1} \in K$. Note if p is a odd prime, then

$$(a-b)^p = \sum_{i=0}^{p} \binom{p}{i} a^i (-b)^{p-i} = a - b$$

as $p|\binom{p}{i}$, then all the middle terms evaluates to 0 according to Lemma (9.9). So $a - b \in K$. Next for ab^{-1}, we have

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

Now if p is an even prime, then

$$(a-b)^2 = a^2 + b^2 + 2ab = a^2 + b^2 = a + b = a - b.$$

Since in a field of characteristic two, we have $b = -b$. And for ab^{-1} we have, $(ab^{-1})^2 = a^2 b^{-2} = a \cdot (b^2)^{-1} = ab^{-1}$. □

Lemma 9.11 *If K is a subfield of F, then K and F have the same characteristics.*

Proof: Since $1_F \in K$ (as K is a subfield of F) and by the closure property of a subfield, we have that for all $n \in \mathbb{N}_{\geq 0}$, $n \cdot 1_F \in K$. Thus by a moment of thought, it is clear that K and F have the same characteristics. □

Corollary 9.11.1 *Suppose* $\mathrm{Char}(F) = 0$*, then F contains a copy of natural numbers, namely the set $\{k \cdot 1 : k \in \mathbb{N}_{\geq 0}\}$. The smallest subfield of F is a copy of $\mathbb{Q}$ and any subfield K of F contains this copy of $\mathbb{Q}$ as well.*

Proof: Since $\{k \cdot 1 : k \in \mathbb{N}_{\geq 0}\}$ is isomorphic to $\mathbb{N}$, then its additive and multiplicative inverse is a copy of $\mathbb{Q}$, which must also be in F. Note any subfield K of F must contain 1, thus must also contain this isomorphic copy of $\mathbb{Q}$. □

Corollary 9.11.2 *Suppose* $\mathrm{Char}(F) = p$ *where p is a prime, then the set $\{n \cdot 1 : n = 1, 2, \cdots, p\}$ which is isomorphic to $\mathbb{Z}_p$ forms a subfield of F, and is the smallest subfield of F, that is any subfield K of F must contain a copy of $\mathbb{Z}_p$.*

Proof: By Lemma 9.11, a subfield of F must have characteristic p, hence at least p elements. So $\mathbb{Z}_p$ is the smallest subfield of F if it is a subfield.
We show $\mathbb{Z}_p = \{n \cdot 1 : n = 1, 2, \cdots, p\}$ is a subfield. It is clear that $(\mathbb{Z}_p, +)$ is an abelian group. It is also clear that $\mathbb{Z}_p$ is closed under multiplicaiton, in fact $(n \cdot 1)(m \cdot 1) = nm \cdot 1$. We show for every $n \cdot 1 \in \mathbb{Z}_p$, $n \cdot 1 \neq 0$, its multiplicative inverse is in $\mathbb{Z}_p$. Consider the inverse of $\bar{n}$ in $(\mathbb{Z}/p\mathbb{Z})^*$, given by $\bar{m}$ (we know it exists as $(\mathbb{Z}/p\mathbb{Z})^*$ is a group under multiplication). Then $(n \cdot 1)(m \cdot 1) = nm \cdot 1 = 1 \cdot 1 = 1$ as $nm = 1 \mod p$. Thus every element has a multiplicative inverse. So $\mathbb{Z}_p$ is a subfield.
Next, it is clear that any subfield of F must contain 1 hence $\mathbb{Z}_p$. □

Definition 9.12 (Prime Subfield) *Suppose F is a field, then the intersection of all subfields of F is the* ***prime subfield of*** *F.*

Remark 9.12.1 *We know the intersection of arbitrary collection of subfields of F is a subfield. In addition, since we taking the intersection of all subfields, then it is unique and smallest subfield of F. In fact, by Corollary 9.11.1 and 9.11.2 we know what they are based on the characteristic of the field.*

9.3 Extension Field

Definition 9.13 (Extension Field) *If K is a field containing the subfield F, then K is said to be an* ***extension field / extension*** *of F, denoted K/F (called K over F). The field F is sometimes called the* ***base field*** *of the extension.*

Remark 9.13.1 *If K/F is any extension of fields, then the multiplication defined in K makes K into a vector space over F. In particular every field F can be considered as a vector space over its prime field.*

Definition 9.14 (Degree Of Extension) *The* ***degree of a field extension*** *K/F, denoted $[K : F]$, is the dimension of K as a vector space over F ($\dim_F K$). The extension is said to be* ***finite*** *if $[K : F]$ is finite and is said to be* ***infinite*** *otherwise.*

Proposition 9.15 *Let $\varphi : F \to F'$ be a homomorphism of fields. Then φ is either identically 0 or is injective, so that the image of φ is either 0 or isomorphic to F.*

Proof: Since the only ideal of F is 0 or F, and only ideals of F can be the kernel of φ. □

Lemma 9.16 *Let E/F be a finite extension. Let $\phi : E \to E$, be a homomorphism that fixes F (F-linear map of vector spaces). Then ϕ is an isomorphism.*

Proof: ϕ is not trivial, then ϕ is injective. Then since $\phi : E \to E$ is also a map of F-vector spaces of same dimension, then ϕ is an isomorphism. □

Theorem 9.17 *Let F be a field and let $p(x) \in F[x]$ be an irreducible polynomial. Then there exists a field K containing an isomorphic copy of F in which $p(x)$ has a root. Identifying F with this isomorphic copy shows that there exists an extension of F in which $p(x)$ has a root.*

Proof: Consider the quotient

$$K = F[x]/(p(x))$$

of the polynomial ring $F[x]$ by the ideal generated by $p(x)$. Since by assumption $p(x)$ is an irreducible polynomial in the P.I.D. $F[x]$, then the ideal $(p(x))$ is a maximal ideal. Hence K is actually a field. The canonical projection π of $F[x]$ to the quotient $F[x]/(p(x))$ gives a homomorphism $\varphi = \pi|_F : F \to K$ which is not identically 0. Hence by

Proposition 9.15, $\varphi(F) \cong F$ is an isomorphic copy of F contained in K. We identify F with its isomorphic image in K and view F as a subfield of K. If $\bar{x} = \pi(x)$ denotes the image of x in the quotient K, then

$$\begin{aligned} p(\bar{x}) &= \overline{p(x)} \quad \text{(homomophism)} \\ &= p(x) \mod p(x) \\ &= 0 \end{aligned}$$

So K does indeed contain a root of the polynomial $p(x)$. Then K is an extension of F in which the polynomial $p(x)$ has a root. □

Theorem 9.18 *Let $p(x) \in F[x]$ be an irreducible polynomial of degree n over the field F and let K be the field $F[x]/(p(x))$. Let $\theta = x \mod (p(x)) \in K$. Then the elements*

$$1, \theta, \theta^2, \cdots, \theta^{n-1}$$

are a basis for K as a vector space over F, so the degree of the extension is n, i.e., $[K : F] = n$. Hence

$$K = \{a_0 + a_1\theta + a_2\theta^2 + \cdots + a_{n-1}\theta^{n-1} \,|\, a_0, a_1, \cdots, a_{n-1} \in F\}$$

consists of all polynomials of degree $< n$ in θ.

Proof: Division Algorithm. □

Corollary 9.18.1 *Let $p(x) \in F[x]$ be an irreducible polynomial of degree n over the field F and let K be the field $F[x]/(p(x))$. Let $a(\theta), b(\theta) \in K$ be two polynomials of degree $< n$ in θ. Then addition in K is defined simply by usual polynomial addition and multiplication in K is defined by*

$$a(\theta)b(\theta) = r(\theta)$$

where $r(x)$ is the remainder obtained after dividing the polynomial $a(x)b(x)$ by $p(x)$ in $F[x]$.

Remark 9.18.1 *We can find the inverse of an element in $F[x]/(p(x))$ using the Euclidean Algorithm. To find θ^{-1}, note if*

$$p(x) = p_n x^n + p_{n-1}x^{n-1} + \cdots + p_1 x + p_0$$

Then

$$\theta(p_n\theta^{n-1} + \cdots + p_1) = -p_0$$

So

$$\theta^{-1} = \frac{-1}{p_0}(p_n\theta^{n-1} + \cdots + p_1) \in K.$$

Definition 9.19 *Let K be an extension of the field F and let $\alpha, \beta, \cdots, \in K$ be a collection of elements of K. Then the smallest subfield of K containing both F and the elements $\alpha, \beta, \cdots$, denoted $F(\alpha, \beta, \cdots)$ is called the field generated by $\alpha, \beta, \cdots$ over F. If the field K is generated by a single element α over F, $K = F(\alpha)$, then K is said to be a **simple extension** of F, and the elements α is called a **primitive element** for the extension.*

Theorem 9.20 *Let F be a field and let $p(x) \in F[x]$ be an irreducible polynomial. Suppose K is an extension field of F containing a root α of $p(x)$. Let $F(\alpha)$ denote the subfield of K generated over F by α. Then*

$$F(\alpha) \cong F[x]/(p(x)).$$

Proof: There is a natural homomorphism:

$$\begin{aligned} \varphi : F[x] &\longrightarrow F(\alpha) \subseteq K \\ a(x) &\longmapsto a(\alpha) \end{aligned}$$

Since $p(\alpha) = 0$ by assumption, the element $p(x)$ is in the kernel of φ, so we obtain an induced homomorphism:

$$\varphi : F[x]/(p(x)) \to F(\alpha).$$

But since $p(x)$ is irreducible, the quotient on the left is a field, and φ is not the 0 map, hence φ is an isomorphism of the field on the left with its image. Since this image is then a subfield of $F(\alpha)$ containing F and containing α, by the definition of $F(\alpha)$, the map must be surjective. □

Remark 9.20.1 *This theorem says that any field over F in which $p(x)$ contains a root contains a subfield isomorphic to the extension of F constructed in the Theorem 9.17, and this field is (up to isomorphism) the smallest extension of F containing such a root.*
It is also clear that

$$F(\alpha) = \{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1} \,|\, a_0, a_1, \cdots, a_{n-1} \in F\} \subseteq K.$$

Theorem 9.21 *Let $\varphi : F \to F'$ be an isomorphism of fields. Let $p(x) \in F[x]$ be an irreducible polynomial and let $p'(x) \in F'[x]$ be the irreducible polynomial obtained by applying the map φ to the coefficients of $p(x)$. Let α be a root of $p(x)$ (in some extension of F) and let β be a root of $p'(x)$ (in some extension of F'). Then there is an isomorphism*

$$\begin{aligned} \sigma : F(\alpha) &\longrightarrow F'(\beta) \\ \alpha &\longmapsto \beta \end{aligned}$$

mapping α to β and extending φ, i.e., such that σ restricted to F is the isomorphism φ.

Proof: The isomorphism induces a natural isomorphism from $F[x]$ to $F'[x]$ which maps the maximal ideal $(p(x))$ to the maximal ideal $(p'(x))$. Taking the quotient by these ideals, we obtain an isomorphism of fields

$$F[x]/(p(x)) \longrightarrow F'[x]/(p'(x)).$$

The field on the left is isomorphic to $F(\alpha)$ and the field on the right is isomorphic to $F'(\beta)$. Composing these isomorphism, we obtain the isomorphism σ. It is clear that the restriction of this isomorphism to F is φ, completing the proof. □

Remark 9.21.1 *This extension theorem can be represented pictorially by the diagram:*

$$\begin{array}{lccc} \sigma: & F(\alpha) & \xrightarrow[\sim]{} & F'(\beta) \\ & \Big| & & \Big| \\ \varphi: & F & \xrightarrow{\sim} & F' \end{array}$$

Theorem 9.22 *Let K be an extension over F of degree n, then K is isomorphic to a subfield of the ring of $n \times n$ matrices over F. Hence the ring of $n \times n$ matrices over F contains an isomorphic copy of every extension of F of degree $\leq n$.*

Proof: For any $\alpha \in K$, we note that the map $T_\alpha : x \mapsto \alpha x$ is a F-linear isomorphism of K to itself. Then let B be a basis of K when considered as a F-vector space. Then we can correspond each α a $n \times n$ matrix given by $[T_\alpha]_B^B$. In particular, $[T_\alpha]_B^B$ is invertible. The last statement follows from block matrix multiplication. □

Corollary 9.22.1 *If $[T_\alpha]$ is the matrix representation for the map $T_\alpha : x \mapsto \alpha x$, then α is a root of the minimal polynomial of $[T_\alpha]$.*

9.4 Algebraic Extensions

Let F be a field and let K be an extension of F.

Definition 9.23 (Algebraic Extension) *The element $\alpha \in K/F$ is said to be **algebraic** over F if α is a root (in K) of some nonzero polynomial $f(x) \in F[x]$. If α is not algebraic over F (i.e., is not the root of any nonzero polynomial with coefficients in F) then α is said to be **transcendental** over F. The extension K/F is said to be **algebraic** if every element of K is algebraic over F.*

Remark 9.23.1 *Note that if α is algebraic over a field F then it is algebraic over any extension field L of F where L contains α.*

Remark 9.23.2 *Let K/F be a field extension, let $\alpha \in K$, then F is algebraic if and only if the following map has a non-zero kernel: $F[x] \to K$, $x \mapsto \alpha$. Then we can write* $\ker = (p(x))$*, since $F[x]$ is a P.I.D. this brings the definition of minimal polynomial.*

Lemma 9.24 *Suppose K/F is algebraic and R is a ring such that $F \subset R \subset K$, then R is a subfield of K.*

Proof: Suffices to show that for any $r \in R$, $r^{-1} \in R$. Since $r \in K$, then r is algebraic over F, so exists monic irreducible polynomial $p(x)$ with coefficients in F such that $p(r) = 0$. That is

$$\begin{aligned} r^n + a_{n-1}r^{n-1} + \cdots + a_1 r + a_0 &= 0 \\ r(r^{n-1} + a_{n-1}r^{n-2} + \cdots + a_1) &= -a_0 \end{aligned}$$

Hence the inverse of r which is given by

$$-\frac{1}{a_0}(r^{n-1} + a_{n-1}r^{n-2} + \cdots + a_1)$$

is also in R. □

Proposition 9.25 *Let α be algebraic over F. Then there is a unique monic irreducible polynomial $m_{\alpha,F}(x)$ which has α as a root. A polynomial $f(x) \in F[x]$ has α as a root if and only if $m_{\alpha,F}(x)$ divides $f(x)$ in $F[x]$.*

Proof: Let $g(x) \in F[x]$ be a polynomial of minimal degree having α as a root. Multiplying $g(x)$ by a constant, we may assume $g(x)$ is monic. Suppose $g(x)$ were reducible in $F[x]$, say $g(x) = a(x)b(x)$ with $a(x), b(x) \in F[x]$ both of degree smaller than the degree of $g(x)$. Then $g(\alpha) = a(\alpha)b(\alpha)$ in K, and since K is a field, either $a(\alpha) = 0$ or $b(\alpha) = 0$, contradicting the minimality of the degree of $g(x)$. It follows that $g(x)$ is a monic irreducible polynomial having α as a root.
Suppose now that $f(x) \in F[x]$ is any polynomial having α as a root. By the Euclidean Algorithm, we can easily show that $g(x)|f(x)$. In particular, this shows that $m_{a,F}(x) = g(x)$ is unique. □

Corollary 9.25.1 *If L/F is an extension of fields and α is algebraic over both F and L, then $m_{\alpha,L}(x)$ divides $m_{\alpha,F}(x)$ in $L[x]$.*

Definition 9.26 (Minimal Polynomial) *The polynomial $m_{\alpha,F}(x)$ (more simply denoted $m_\alpha(x)$ or $m(x)$) is called the **minimal polynomial** for α over F. The degree of $m_\alpha(x)$ is called the **degree** of α.*

Proposition 9.27 *Let α be algebraic over the field F and let $F(\alpha)$ be the field generated by α over F. Then*

$$F(\alpha) \cong F[x]/(m_\alpha(x))$$

so that in particular

$$[F(\alpha) : F] = \deg m_\alpha(x) = \deg \alpha.$$

Proposition 9.28 *The element α is algebraic over F if and only if the simple extension $F(\alpha)/F$ is finite. More precisely, if α is an element of an extension of degree n over F then α satisfies a polynomial of degree at most n over F and if α satisfies a polynomial of degree n over F then the degree of $F(\alpha)$ over F is at most n.*

Proof: If α is algebraic over F, then the degree of the extension $F(\alpha)/F$ is the degree of the minimal polynomial for α over F. Hence the extension is finite, of degree $\leq n$ if α satisfies a polynomial of degree n. Conversely, suppose α is an element of an extension of degree n over F. Then the $n+1$ elements of $F(\alpha)$ are linear dependent over F, say

$$b_0 + b_1\alpha + b_2\alpha^2 + \cdots + b_n\alpha^n = 0$$

with $b_0, b_1, \cdots, b_n \in F$ not all 0. Hence α is the root of a nonzero polynomial with coefficients in F (of degree $\leq n$), which proves α is algebraic over F. □

Corollary 9.28.1 *If the extension K/F is finite, then it is algebraic.*

Proof: If $\alpha \in K$, then the subfield $F(\alpha)$ is in particular a subspace of the vector space K over F. Hence $[F(\alpha) : F] \leq [K : F]$ and so α is algebraic over F. □

Definition 9.29 (Quadratic Extensions) *Extensions of degree 2 of a field F are called **quadratic extensions** of F. In particular, it is of the form $F(\sqrt{D})$, where D is an element of F which is not square in F.*

Theorem 9.30 *Let $F \subseteq K \subseteq L$ be fields. Then*

$$[L : F] = [L : K][K : F],$$

i.e., extension degrees are multiplicative, where if one side of the equation is infinite, the other side is also infinite.

Proof: Suppose first that $[L : K] = m$ and $[K : F] = n$ are finite. Let $\alpha_1, \cdots, \alpha_m$ be a basis for L over K and let $\beta_1, \cdots, \beta_n$ be a basis for K over F. Then every element of L can be written as a linear combination

$$s_1\alpha_1 + s_2\alpha_2 + \cdots + s_m\alpha_m$$

where $s_1, \cdots, s_m \in K$. Hence every element of L can be written as F-linear combinations of $\beta_1, \cdots, \beta_n$ by writing

$$s_i = t_{i1}\beta_1 + \cdots + t_{in}\beta_n, \quad i = 1, 2, \cdots, m.$$

Hence the mn elements $\alpha_i\beta_j$ spans L as a vector space over F. Next suppose now

$$\sum_{\substack{1\leq i\leq m\\ 1\leq j\leq n}} c_{ij}\alpha_i\beta_j = 0$$

with $c_{ij} \in F$. Then we see

$$\sum_{i=1}^{m} s_i\alpha_i = 0$$

where

$$s_i = \sum_{j=1}^{n} c_{ij}\beta_j.$$

Then it easily follows that all the $c_{ij} = 0$. Hence the $\alpha_i\beta_j$ are linearly independent over F, so forms a basis for L over F and $[L : F] = mn = [L : K][K : F]$.

Finally, if $[K : F]$ is infinite, then there are infinitely many elements of K, hence of L which are linear independent over F, so that $[L : F]$ is also infinite. Similarly, if $[L : F]$ is infinite, there are infinitely many element of L linearly independent over K, so also over F, and again $[L : F]$ is infinite. Lastly, if $[L : K]$ and $[K : F]$ are both finite, then $[L : F]$ cannot be infinite, so if $[L : F]$ is infinite implies at least one of $[L : K]$ or $[K : F]$ is infinite. □

Corollary 9.30.1 *Suppose L/F if a finite extension and let K be any subfield of L containing F, $F \subseteq K \subseteq L$, then $[K : F]$ divides $[L : F]$.*

Proposition 9.31 *Suppose $f(x)$ is an irreducible polynomial of degree n over a field F and $g(x) \in F[x]$. If $h(x)$ is an irreducible polynomial that divides $f(g(x))$, then the degree of h is divisible by n.*

Proof: Let α be a root of h. Then suffices to show that $n|[F(\alpha) : F]$. Now since $h(x)|f(g(x))$, then $f(g(\alpha)) = 0$, so $g(\alpha)$ is a root of f. Hence $[F(g(\alpha)) : F] = n$. Clearly, we also have $F(g(\alpha)) \subset F(\alpha)$. Then

$$[F(\alpha) : F] = [F(\alpha) : F(g(\alpha))][F(g(\alpha)) : F] = [F(\alpha) : F(g(\alpha))] \cdot n.$$

□

Definition 9.32 (Finitely Generated Extensions) *An extension K/F is **finitely generated** if there are elements $\alpha_1, \cdots, \alpha_k$ in K such that $K = F(\alpha_1, \alpha_2, \cdots, \alpha_k)$.*

Remark 9.32.1 *In fact we see that elements in $F(\alpha_1, \cdots, \alpha_n)$ are fractions of the form*

$$\frac{f(\alpha_1, \cdots, \alpha_n)}{g(\alpha_1, \cdots, \alpha_n)}$$

for polynomials f and g.

Lemma 9.33 *$F(\alpha, \beta) = (F(\alpha))(\beta)$, i.e., the field generated over F by α and β is the field generated by β over the field $F(\alpha)$ generated by α over F.*

Proof: This follows by the minimality of the fields in questions. The field $F(\alpha, \beta)$ contains F and α, hence contains the field $F(\alpha)$, and since it also contains β, we have the inclusion $(F(\alpha))(\beta) \subseteq F(\alpha, \beta)$. Since the field $(F(\alpha))(\beta)$ contains F, α and β, by the minimality of $F(\alpha, \beta)$, we have the reverse inclusion $F(\alpha, \beta) \subseteq (F(\alpha))(\beta)$. □

Corollary 9.33.1 *An extension generated by a finite number of algebraic elements is finite. More generally, if*

$$K = F(\alpha_1, \alpha_2, \cdots, \alpha_k),$$

let $F_0 = F$ and $F_{i+1} = F_i(\alpha_{i+1})$, $I = 0, 1, \cdots, k-1$. This gives a sequence of fields:

$$F = F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_k = K.$$

Then

$$[K : F] = [F_k : F_{k-1}][F_{k-1} : F_{k-2}] \cdots [F_1 : F_0].$$

In particular, if $\alpha_1, \cdots, \alpha_k$ are algebraic over F with degree $n_1, \cdots, n_k$, then $[K : F] \leq n_1 n_2 \cdots n_k$.

Proof: Induction. □

Theorem 9.34 *The extension K/F is finite if and only if K is generated by a finite number of algebraic elements over F. More precisely, a field generated over F by a finite number of algebraic elements of degrees*

$n_1, \cdots, n_k$ *is algebraic of degree* $\leq n_1 \cdots n_k$.

Proof: If K/F is finite of degree n, let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be a basis for K as a vector space over F. By Corollary 9.30.1, $[F(\alpha_i) : F]$ divides $[K : F] = n$ for $i = 1, 2, \cdots, n$, so that Proposition 9.28, we have that each α_i is algebraic over F. Since K is obviously generated over F by $\alpha_1, \cdots, \alpha_n$, we see that K is generated by a finite number of algebraic elements over F. The converse was Corollary 9.33.1. □

Corollary 9.34.1 *Suppose* α *and* β *are algebraic over* F. *Then* $\alpha \pm \beta$, $\alpha\beta$, α/β $(\beta \neq 0)$, *in particular* α^{-1} *for* $\alpha \neq 0$ *are all algebraic.*

Proof: All of these elements lie in the extension $F(\alpha, \beta)$, which is finite over F. Hence algebraic by Corollary 9.28.1. □

Corollary 9.34.2 *Let* L/F *be an arbitrary extension. Then the collection of elements of* L *that are algebraic over* F *form a subfield* K *of* L.

Proof: Follows from Corollary 9.34.1. □

Remark 9.34.1 *Using this corollary we can define the field of* ***algebraic numbers*** *to be the subfield of all elements in* $\mathbb{C}$ *that are algebraic over* $\mathbb{Q}$ *which is a subfield of* $\mathbb{C}/\mathbb{Q}$, *denoted* $\bar{\mathbb{Q}}$. *Note this is an infinite algebraic extension of* $\mathbb{Q}$, *as the elements* $\sqrt[n]{2}$ *are all elements of* $\bar{\mathbb{Q}}$. *Using a Cardinality argument, one can show that* $\bar{\mathbb{Q}}$ *is a proper subfield of* $\mathbb{C}$.

Theorem 9.35 *If* K *is algebraic over* F *and* L *is algebraic over* K, *then* L *is algebraic over* F.

Proof: Let α be any element of L. Then α is algebraic over K, so α satisfies some polynomial equation

$$a_n\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$$

where $a_0, \cdots, a_n \in K$. Consider the field $F(\alpha, a_0, a_1, \cdots, a_n)$ generated over F by α and the coefficients of this polynomial. Since K/F is algebraic, the elements $a_0, a_1, \cdots, a_n$ are algebraic over F, so the extension $F(a_0, a_1, \cdots, a_n)/F$ is finite. By the equation above, we see that α generates an extension of this field of degree at most n, since its minimal polynomial over this field is a divisor of the polynomial above. Therefore

$$[F(\alpha, a_0, \cdots, a_n) : F] = [F(\alpha, a_0, \cdots, a_n) : F(a_0, \cdots, a_n)][F(a_0, \cdots, a_n) : F]$$

is also finite and $F(\alpha, a_0, \cdots, a_n)/F$ is an algebraic extension. In particular, the elements α is algebraic over F, which proves L is algebraic over F. □

Definition 9.36 (Composite Field) *Let* K_1 *and* K_2 *be two subfields of a field* K. *Then the* ***composite field*** *of* K_1 *and* K_2, *denoted* K_1K_2, *is the smallest subfield of* K *containing both* K_1 *and* K_2. *Similarly the composite of any collection of subfields of* K *is the smallest subfield containing all the subfields.*

Lemma 9.37 *Let* F *be a field, then*

$$F(\alpha_1, \cdots, \alpha_n) = F(\alpha_1)F(\alpha_2)\cdots F(\alpha_n).$$

Proof: The left hand side is the smallest field containing F and elements $\alpha_1, \cdots, \alpha_n$. The right hand side is a field containing F and $\alpha_1, \cdots, \alpha_n$, hence we have one inclusion. For the other inclusions, we note that each $F(\alpha_i) \subset F(\alpha_1, \cdots, \alpha_n)$. Hence the field generated by these subfields should also lie in $F(\alpha_1, \cdots, \alpha_n)$. □

Proposition 9.38 *Let K_1 and K_2 be two finite extension of a field F contained in K. Then*

$$[K_1K_2 : F] \leq [K_1 : F][K_2 : F]$$

with equality if and only if an F-basis for one of the fields remains linearly independent over the other field. If $\alpha_1, \alpha_2, \cdots, \alpha_n$ and $\beta_1, \beta_2, \cdots, \beta_m$ are bases for K_1 and K_2 over F respectively, then the elements $\alpha_i\beta_j$ for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$ span K_1K_2 over F.

Proof: From $K_1K_2 = F(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m) = K_1(\beta_1, \cdots, \beta_m)$, $\beta_1, \cdots, \beta_m$ span K_1K_2 over K_1. Hence $[K_1K_2 : K_1] \leq m = [K_2 : F]$ with equality if and only if these elements are linearly independent over K_1. □

Corollary 9.38.1 *Suppose that $[K_1 : F] = n$, $[K_2 : F] = m$, where n and m are relatively prime. Then $[K_1K_2 : F] = [K_1 : F][K_2 : F] = nm$.*

Proof: In general the extension degree $[K_1K_2 : F]$ is divisible by both n and m since K_1 and K_2 are subfields of K_1K_2, hence is divisible by their least common multiple, which is nm in this case. We also have $[K_1K_2 : F] \leq nm$, so it must be the case that $[K_1K_2 : F]$ is precisely nm. □

Proposition 9.39 *Let E_1/F and E_2/F be field extensions. Assume $E_1, E_2 \subset K$ is contained in some ambient field K. If E_1/F is finite, then E_1E_2/E_2 is also finite.*

Proof: Let $E_1 = F(\alpha_1, \cdots, \alpha_n)$. Then we note that $E_1E_2 = E_2(\alpha_1, \cdots, \alpha_n)$, which is an finite extension over E_2. □

Proposition 9.40 *Let E_1/F and E_2/F be algebraic extensions. Assume $E_1, E_2 \subset K$ is contained in some ambient field K. Let E_1E_2 be the field generated by E_1 and E_2. Then E_1E_2/F is also algebraic.*

Proof: We observe that the elements in E_1E_2 are of the form

$$x = \frac{\sum_i a_i b_i}{\sum_i a_i' b_i'}$$

where the summation is finite and $a_i, a_i' \in E_1$, $b_i, b_i' \in E_2$. We note that the set of these elements indeed form a field and is actually equal to E_1E_2. Now since $x \in F(a_i, a_i', b_i, b_i')$ which is a finite extension, then x is algebraic. □

Proposition 9.41 *Let E_1/F and E_2/F be field extensions. Assume $E_1, E_2 \subset K$ is contained in some ambient field K. If E_1/F is algebraic, then E_1E_2/E_2 is also algebraic.*

Proof: Again we have

$$x = \frac{\sum_i a_i b_i}{\sum_i a_i' b_i'}$$

However, in this case, then we have $x \in E_2(a_i)$. Since each a_i is algebraic over F, then it is algebraic over E_2. Hence $E_2(a_i)/E_2$ is a finite extension. □

Proposition 9.42 *Let $\sigma : E \to E$ be an F-homomorphism, where E/F is algebraic, then σ is an isomorphism.*

Proof: Since $\sigma : E \to E$ is an F-homomorphism, so it is nontrivial. Then we know σ is injective. Suffices to show σ is surjective, i.e., $\sigma(E) = E$.

Let $\alpha \in E$ be arbitrary, we show α is in the image of σ. Since E/F is algebraic, then $m_{\alpha,F}(x)$ exists. Let

$$m_{\alpha,F}(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, \cdots, a_{n-1} \in F$. Then since σ fixes $a_0, \cdots, a_{n-1}$, then σ must map a root of $m_{\alpha,F}(x)$ in E to another root of $m_{\alpha,F}(x)$ in E. Then by induction, it is easy to see that the list of elements

$$\alpha, \sigma(\alpha), \sigma^2(\alpha), \cdots,$$

must all be roots of $m_{\alpha,F}(x)$. Since $m_{\alpha,F}(x)$ can have only finitely many distinct roots, then there exists $i < j \in \mathbb{N}_{\geq 0}$ such that $\sigma^i(\alpha) = \sigma^j(\alpha)$. Now by injectivity of σ, we must have $\sigma^{i-1}(\alpha) = \sigma^{j-1}(\alpha)$. Then doing this inductively, we would get $\alpha = \sigma^{j-i}(\alpha)$. Hence there exists $n \in \mathbb{N}$ such that $\sigma^n(\alpha) = \alpha$.

Lastly, we note that $\sigma^n(E) \subset \sigma(E)$. Hence $\alpha \in \sigma(E)$. Since $\alpha \in E$ is arbitrary, then σ is surjective, hence an isomorphism. □

Example: $\sigma : F(x) \to F(x)$, $x \mapsto x^2$ is injective but not surjective, where $F(x)$ is the field of fractions of $F[x]$. This shows that without the algebraic assumptions is necessary.

9.5 Classical Straight-edge and Compass Constructions

Assume we are given 0 and I, a base point and a unit interval. We are allowed to draw a circle with a point and existing radius (Compass), and connected two points with a straight line (Straight-edge).

Definition 9.43 (Constructable) *A real number (α length) is called **constructible** if it can be constructed by forming successive intersections of*

- *lines betweens constructed points;*
- *circles centered at constructed points with radius being a constructed line segment.*

Example: Given 0 and 1, $\sqrt{2}$ is constructable.

Definition 9.44 *Let $F \subset \mathbb{R}$ be a subfield, we consider $F \times F \subset \mathbb{R} \times \mathbb{R}$. An F-lines is a line in $\mathbb{R} \times \mathbb{R}$ through two F-points. An F-circle is a circle in $\mathbb{R} \times \mathbb{R}$ centered at an F-point with radius $r \in F$.*

Remark 9.44.1 *We note that an F-line must have equations of the form*

$$ax + by + c, \quad a, b, c \in F.$$

Also the equation of an F-circle is of the form

$$(x - a)^2 + (y - b)^2 = r^2 \;\; a, b \in F.$$

Lemma 9.45 *Let $L \neq L^1$ be two F-lines and $C \neq C'$ be two F-circles. Then*

- *$L \cap L' = \emptyset$ or consists of a single F-point.*
- *$L \cap C = \emptyset$ or consists of one or two points in $F(\sqrt{e})$-plane for some $e \in F$, $e > 0$.*
- *$C \cap C' = \emptyset$ or consists of one or two points in $F(\sqrt{e})$-plane for some $e \in F$, $e > 0$.*

Proposition 9.46 *If the element $\alpha \in \mathbb{R}$ is obtained from a field $F \subset \mathbb{R}$ by a series of compass and straightedge constructions then $[F(\alpha) : F] = 2^k$ for some integer $k \geq 0$.*

Lemma 9.47 *Assume we have 0 and I to start with.*

1. *Assume $c, d \in \mathbb{R}$ are constructible (given), then $c + d, c - d, -c, cd$ and $\frac{d}{c}$, $(c \neq 0)$ are constructible.*
2. *If $c > 0$ is constructible, then $\sqrt{c}$ is constructible.*

Theorem 9.48 *Assume we are given 0 and I.*

1. *The set of constructible numbers is a field in $\mathbb{R}$.*
2. *A number $\alpha \in \mathbb{R}$ is constructible if and only if it is contained in a subfield of $\mathbb{R}$ of the form*

$$\mathbb{Q}(\sqrt{a_1}, \cdots, \sqrt{a_n})$$

where $a_i \in \mathbb{Q}(\sqrt{a_1}, \cdots, \sqrt{a_{i-1}})$, $a_i > 0$.

Proof: (1) follows from the previous lemma. Then (2) follows from (1) and the previous lemma. □

Corollary 9.48.1 *None of the classical Greek Problems:*

- *Doubling the Cube: given a cube, we want to double its volume using Straightedge and compass construction (Construct $\sqrt[3]{2}$).*
- *Trisecting an Angle (Construct* sin *and* cos*).*
- *Squaring the circle (Construct $\sqrt{\pi}$).*

is possible.

9.6 Splitting Fields and Algebraic Closures

Definition 9.49 (Splitting Field) *The extension field K of F is called a **splitting field** for the polynomial $f(x) \in F[x]$ if $f(x)$ factors completely into linear factors in $K[x]$ and $f(x)$ does not factor completely into linear factors over any proper subfield of K containing F.*

Theorem 9.50 *For any field F, if $f(x) \in F[x]$, then there exists an extension K of F which is a splitting field for $f(x)$.*

Proof: We first show that there is an extension E of F over which $f(x)$ splits completely into linear factors by induction on the degree n of $f(x)$. If $n = 1$, then take $E = F$. Suppose now that $n > 1$. If the irreducible factors of $f(x)$ over F are all of degree 1, then F is the splitting field for $f(x)$ and we may take $E = F$. Otherwise, at least one of the irreducible factors, say $p(x)$ of $f(x)$ in $F[x]$ is of degree at least 2, then by Theorem 9.17, there is an extension E_1 of F containing a root α of $p(x)$. Over E_1, the polynomial $f(x)$ has the linear factor $x-\alpha$. The degree of the remaining factor $f_1(x)$ of $f(x)$ is $n-1$, so by induction, there is an extension E of E_1 containing all the roots of $f_1(x)$.

Now let K be the intersection of all the subfields of E containing F which also contain all the roots of $f(x)$. Then K is a field which is a splitting field for $f(x)$. □

Remark 9.50.1 *Later we will see that if $\overline{F}$ is the algebraic closure of F, and $f(x) \in F[x]$ can be written by*

$$f(x) = (x_1 - a_1)^{n_1}(x - a_2)^{n_2} \cdots (x - a_m)^{n_m} \quad \text{in } \overline{F}[x]$$

then the splitting field of $f(x)$ is the field extension E/F with $E = F(a_1, \cdots, a_m) \subset \overline{F}$.

Definition 9.51 (Normal Extension) *If K is an algebraic extension of F which is the splitting field over F for a collection of polynomials $f(x) \in F[x]$ then K is called a **normal extension** of F. Alternatively, we call K/F a splitting field of this collection if E is generated by all the roots of all polynomials in some $\bar{F}$.*

Proposition 9.52 *A splitting field of a polynomial of degree n over F is of degree at most $n!$ over F.*

Proof: Each time adjoining one root of $f(x)$ to F, it generates an extension of degree at most n, where n is the degree of $f(x)$. □

Definition 9.53 *A generator of the cyclic group of all n^{th} roots of unity is called a primitive n^{th} root of unity and is usually denoted by ξ_n; in particular, there are precisely $\varphi(n)$ primitive n^{th} roots of unity. The field $\mathbb{Q}(\xi_n)$ is called the **cyclotomic field** of n^{th} roots of unity.*

Remark 9.53.1 *If $n = p$ is a prime, then*

$$x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$$

and since $\xi_p \neq 1$, it follows that ξ_p is a root of the polynomial

$$\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1$$

which is irreducible. It follows that $\Phi_p(x)$ is the minimal polynomial of ξ_p over $\mathbb{Q}$, so that

$$[\mathbb{Q}(\xi_p) : \mathbb{Q}] = p - 1.$$

Theorem 9.54 *Let $\varphi : F \to F'$ be an isomorphism of fields. Let $f(x) \in F[x]$ be a polynomial and let $f'(x) \in F'[x]$ be the polynomial obtained by applying φ to the coefficients of $f(x)$. Let E be a splitting field for $f(x)$ over F and let E' be a splitting field for $f'(x)$ over F'. Then the isomorphism φ extends to an isomorphism $\sigma : E \to E'$, i.e., σ restricted to F is the isomorphism φ:*

$$\begin{array}{lcc} \sigma: & E & \xrightarrow{\sim} & E' \\ & \big| & & \big| \\ \varphi: & F & \xrightarrow{\sim} & F' \end{array}$$

Remark 9.54.1 *The result can be generalized to normal extensions using Zorn's Lemma.*

Proof: We proceed by induction on the degree n of $f(x)$. As in Theorem 9.21 F' induces a natural isomorphism between the polynomial rings $F[x]$ and $F'[x]$. In particular, if $f(x)$ and $f'(x)$ correspond to one another under this isomorphism then the irreducible factors of $f(x)$ in $F[x]$ correspond to the irreducible factors of $f'(x)$ in $F'[x]$.

If $f(x)$ has all its roots in F then $f(x)$ splits completely in $F[x]$ and $f'(x)$ splits completely in $F'[x]$ (with its linear factors being the images of the linear factors for $f(x)$). Hence $E = F$ and $E' = F'$, and in this case we may take $\sigma = \varphi$. This shows the result is true for $n = 1$ and in the case where all the irreducible factors of $f(x)$ have degree 1.

Assume now by induction that the theorem has been proved for any field F, isomorphism φ, and polynomial $f(x) \in F[x]$ of degree $< n$. Let $p(x)$ be an irreducible factor of $f(x)$ in $F[x]$ of degree at least 2 and let $p'(x)$ be the corresponding irreducible factor of $f'(x)$ in $F'[x]$. If $\alpha \in E$ is a root of $p(x)$ and $\beta \in E'$ is a root of $p'(x)$, then by Theorem 9.21 we can extend φ to an isomorphism $\sigma' : F(\alpha) \xrightarrow{\sim} F'(\beta)$:

$$\begin{array}{lccc} \sigma': & F(\alpha) & \xrightarrow{\sim} & F'(\beta) \\ & \downarrow & & \downarrow \\ \varphi: & F & \xrightarrow{\sim} & F'. \end{array}$$

Let $F_1 = F(\alpha)$, $F_1' = F'(\beta)$, so that we have the isomorphism $\sigma' : F_1 \xrightarrow{\sim} F_1'$. We have $f(x) = (x - \alpha)f_1(x)$ over F_1 where $f_1(x)$ has degree $n - 1$ and $f'(x) = (x - \beta)f_1'(x)$. The field E is a splitting field for $f_1(x)$ over F_1: all the roots of $f_1(x)$ are in E and if they were contained in any smaller extension L containing F_1, then, since F_1 contains α, L would also contain all the roots of $f(x)$, which would contradict the minimality of E as the splitting field of $f(x)$ over F. Similarly E' is a splitting field for $f_1'(x)$ over F_1'. Since the degrees of $f_1(x)$ and $f_1'(x)$ are less than n, by induction there exists a map $\sigma : E \xrightarrow{\sim} E'$ extending the isomorphism $\sigma' : F_1 \xrightarrow{\sim} F_1'$. This gives the extended diagram:

$$\begin{array}{lccc} \sigma: & E & \xrightarrow{\sim} & E' \\ & \downarrow & & \downarrow \\ \sigma': & F_1 & \xrightarrow{\sim} & F_1' \\ & \downarrow & & \downarrow \\ \varphi: & F & \xrightarrow{\sim} & F'. \end{array}$$

Then as the diagram indicates, σ restricted to F_1 is the isomorphism σ', so in particular σ restricted to F is σ' restricted to F, which is φ, showing that σ is an extension of φ, completing the proof. $\square$

Corollary 9.54.1 (Uniqueness of Splitting Fields) *Any two splitting fields for a polynomial $f(x) \in F[x]$ over a field F are isomorphic.*

Lemma 9.55 *Let $\{f_i(x)\}_{i\in I}$ be a collection of polynomials in $F[x]$. Let E_1/F be the splitting field in $\overline{F_1} \cong \overline{F}$ and let E_2/F be the splitting field in $\overline{F_2} \cong \overline{F}$. Then $E_1 \cong E_2$.*

Proof: We know $\tau : \overline{F_1} \to \overline{F_2}$ is an isomorphism fixing F. Then by Zorn's Lemma, we could extend to an $\tau : E_1 \to E_2$. We show τ is a bijection by showing $\tau(E_1) = E_2$. Note that $\tau(E_1) \subset E_2$ is a field where $\{f_i(x)\}_{i\in I}$ splits, we can find the factors using τ. Hence $\tau(E_1) = E_2$ by minimality of E_2 as a splitting field. $\square$

Proposition 9.56 *Let K be a finite extension of F, then K is a splitting field over F if and only if every irreducible polynomial in $F[x]$ that has a root in K splits completely in $K[x]$.*

Proof: $\Leftarrow$: if $K = F$, then the implication is trivially true. So assume $K \neq F$, then $K = F(\alpha_1, \alpha_2, \cdots, \alpha_n)$ for some $\alpha_1, \cdots, \alpha_n$ as the extension is finite. Let $m_{\alpha_i,F}$ be the corresponding minimal polynomial of α_i. Then

$$p(x) = \prod_{i=1}^{n} m_{\alpha_i,F}(x)$$

is a polynomial in F. Now $p(x)$ splits completely in K, since by assumption each $m_{\alpha_i,F}$ should split completely in K. Next, if $p(x)$ splits in some field, then it must contain $\alpha_1, \cdots, \alpha_n$. The definition of field generated by F and $\alpha_1, \cdots, \alpha_n$ implies that K is the smallest field such that $p(x)$ splits. Hence K is a splitting field.

$\Rightarrow$: Now suppose K is the splitting field of $g(x) \in F[x]$ and $p(x) \in F[x]$ is irreducible over F. Moreover, $p(x)$ has a root α in K, we show that $p(x)$ splits in $K[x]$.

If $p(x)$ is linear, then we done. Otherwise, suppose $p(x) = (x - \alpha)(x - \beta)\tilde{p}(x)$ over the splitting field E of $p(x)$, where $\alpha \in K$, $\tilde{p}(x) \in E[x]$ and $\beta \in E \setminus K$. Note there is a natural isomorphism between $F(\alpha)$ and $F(\beta)$, since $p(x)$ is irreducible. Therefore, we can extend this isomorphism naturally to an isomorphism between the splitting field of $g(x)$ over $F(\alpha)$ to an isomorphism between the splitting field of $g(x)$ over $F(\beta)$ by Theorem 9.54. If we let K' denote the splitting field of $g(x)$ over $F(\beta)$, then we have $K \cong K'$ and of course $[K' : F] = [K : F]$. Note that K' can be viewed as adjoining β to K and $\beta \in E \setminus K$ implies that

$$[K' : F] = [K' : K][K : F] > [K : F]$$

which is a contradiction.

Thus no root of $p(x)$ can be taken from $E \setminus K$ which implies $p(x)$ splits completely over $K[x]$. $\square$

Corollary 9.56.1 *Let K_1 and K_2 be finite extensions of F contained in the field K, and assume both are splitting fields over F, then $K_1 \cap K_2$ is a splitting field over F.*

Remark 9.56.1 *The result can easily be extended to normal extensions.*

Proof: $K_1 \cap K_2$ is clearly finite over F. Now suppose $p(x) \in F[x]$ is an irreducible polynomial that has a root in $K_1 \cap K_2$. Then it has a root in K_1 and K_2, so they split respectively in $K_1[x]$ and $K_2[x]$. Hence all the roots of p lies in $K_1 \cap K_2$, so $p(x)$ splits completely in $K_1 \cap K_2[x]$. □

Lemma 9.57 *Let K_1 and K_2 be finite extensions of F contained in the field K, and assume both are splitting fields over F, then K_1K_2 is a splitting field over F.*

Remark 9.57.1 *The result can easily be extended to normal extensions.*

Proof: Let K_i be a splitting field of $p_i(x)$ over F. Then we show K_1K_2 is a splitting field for $p(x) = p_1(x) \cdot p_2(x)$. Clearly, p splits in K_1K_2. Next by definition, K_i is the smallest field containing all roots of p_i. Then K_1K_2 is the smallest field containing all roots of p_1 and p_2 as desired. □

Definition 9.58 (Algebraic Closure) *A field K is said to be **algebraically closed** if every non-constant polynomial with coefficients in K has a root in K. The field $\bar{F}$ is called an **algebraic closure** of F if $\bar{F}$ is algebraic over F and if every polynomial $f(x) \in F[x]$ splits completely over $\bar{F}$, equivalently, one can show that $\bar{F}$ is the algebraic closure of F if $\bar{F}/F$ is an algebraic extension that is algebraically closed.*

Remark 9.58.1 *It is clear that if K is algebraically closed, every polynomial with coefficients in K splits completely, so $K = \bar{K}$ if and only if K is algebraically closed.*

Proposition 9.59 *Let $\bar{F}$ be an algebraic closure of F. Then $\bar{F}$ is algebraically closed.*

Proof: Let $f(x)$ be a polynomial in $\bar{F}[x]$ and let α be a root of $f(x)$. Then α generates an algebraic extension $\bar{F}(\alpha)$ of $\bar{F}$, and $\bar{F}$ is algebraic over F. Then $\bar{F}(\alpha)$ is algebraic over F so in particular its element α is algebraic over F. Let $m_{\alpha,F}$ be its minimal polynomial in F, then it must split completely in $\bar{F}$. Since $x - \alpha$ must be a factor, then $\alpha \in \bar{F}$, showing $\bar{F}$ is algebraically closed. □

Proposition 9.60 *For any field F there exists an algebraically closed field K containing F.*

Proof: For every nonconstant monic polynomial $f = f(x)$ with coefficients in F, let x_f denote an indeterminate and consider the polynomial ring $F[\ldots, x_f, \ldots]$ generated over F by the variables x_f. In this polynomial ring consider the ideal I generated by the polynomials $f(x_f)$. If this ideal is not proper, then 1 is an element of the ideal, hence we have a relation

$$g_1 f_1(x_{f_1}) + g_2 f_2(x_{f_2}) + \cdots + g_n f_n(x_{f_n}) = 1$$

where the g_i, $i = 1, 2, \ldots, n$, are polynomials in the x_f. For $i = 1, 2, \ldots, n$ let $x_f = x_i$ and let $x_{n+1}, \ldots, x_m$ be the remaining variables occurring in the polynomials g_j, $j = 1, 2, \ldots, n$. Then the relation above reads

$$g_1(x_1, x_2, \ldots, x_m) f_1(x_1) + \cdots + g_n(x_1, x_2, \ldots, x_m) f_n(x_n) = 1.$$

Let F' be a finite extension of F containing a root α_i of $f_i(x)$ for $i = 1, 2, \ldots, n$. Letting $x_i = \alpha_i$, $i = 1, 2, \ldots, n$ and setting $x_{n+1} = \cdots = x_m = 0$, say, in the polynomial equation above would imply that $0 = 1$ in F', clearly impossible.

Since the ideal I is a proper ideal, it is contained in a maximal ideal $\mathfrak{M}$ (this is where Zorn's Lemma is used). Then the quotient

$$K_1 = F[\ldots, x_f, \ldots]/\mathfrak{M}$$

is a field containing (an isomorphic copy of) F. Each of the polynomials f has a root in K_1 by construction, namely the image of x_f, since $f(x_f) \in I \subseteq \mathfrak{M}$. We have constructed a field K_1 in which every polynomial with coefficients from F has a root. Performing the same construction with K_1 instead of F gives a field K_2 containing K_1 in which all polynomials with coefficients from K_1 have a root. Continuing in this fashion we obtain a sequence of fields

$$F = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_j \subseteq K_{j+1} \subseteq \cdots$$

where every polynomial with coefficients in K_j has a root in K_{j+1}, $j = 0, 1, \ldots$. Let

$$K = \bigcup_{j \geq 0} K_j$$

be the union of these fields. Then K is clearly a field containing F. Since K is the union of the fields K_j, the coefficients of any polynomial $h(x)$ in $K[x]$ all lie in some field K_N for N sufficiently large. But then $h(x)$ has a root in K_{N+1}, so has a root in K. It follows that K is algebraically closed, completing the proof. □

Proposition 9.61 *Let K be an algebraically closed field and let F be a subfield of K. Then the collection of elements, denoted $\bar{F}$, of K that are algebraic over F is an algebraic closure of F. An algebraic closure of F is unique up to isomorphism.*

Proof: By definition, $\bar{F}$ is an algebraic extension of F. Every polynomial $f(x) \in F[x]$ splits completely over K into linear factors $x - \alpha$. But each α is a root of $f(x)$, so is algebraic over F, hence is an element of $\bar{F}$. It follows that the linear factors $x - \alpha$ have coefficients in $\bar{F}$, i.e., $f(x)$ splits completely in $\bar{F}[x]$ and $\bar{F}$ is an algebraic closure of F.

The uniqueness (up to isomorphism) of the algebraic closure follows similarly from the proof of the uniquenss of splitting fields with an application of Zorn's Lemma. As we consider the set of all F homomorphism from a subset of C_1 to C_2 ordered by inclusion, where C_1, C_2 are two algebraic closure of F. Then if the maximal element has domain strictly less than C_1, we can extend the definition of the homomorphism by Theorem (9.21). To show, the surjectivity of the map constructed. Let $\sigma : C_1 \to C_2$, where σ is a F-homomorphism. Then let $\alpha \in C_2$, since C_2 is algebraic over F, exists $p(x) \in F[x]$ such that $p(\alpha) = 0$. Then since C_1 is algebraically closed,

$$p(x) = \prod_{i=1}^{n} (x - r_i) \quad \in C_1[x].$$

Apply σ, we get

$$p(x) = \prod_{i=1}^{n} (x - \sigma(r_i)) \quad \in C_2[x].$$

Letting $x = \alpha$, then LHS becomes 0. Since C_2 is a field, we must have $\alpha = \sigma(r_i)$ for some r_i. □

Remark 9.61.1 *An algebraic closure of F is largest algebraic extension over F, and is the smallest algebraically closed field containing F.*

Corollary 9.61.1 *Let E/F and K/E be algebraic extensions. Let L be an algebraic closure of K. Then any F-homomorphism $\tau : E \to L$, can be extended to an F-homomorphism $\tau : K \to L$. In particular, for any algebraically closed field L and an algebraic extension E/F (of the same characteristic), we can extend any homomorphism $F \to L$ to $E \to L$.*

Remark 9.61.2 *Pictorially we have the following:*

$$\begin{array}{ccc} K & \xrightarrow{\sigma} & L \\ \uparrow & & \uparrow \\ E & \xrightarrow{\sigma} & L \\ \uparrow & & \uparrow \\ F & \longrightarrow & L \end{array}$$

Proof: Zorn's Lemma. □

Corollary 9.61.2 *Let F be a field with an algebraic closure $\bar{F}$. Let $p(x) \in F[x]$ be irreducible with roots α and β in $\overline{F}$. We have a unique F-isomorphism*

$$\sigma : F(\alpha) \cong F(\beta),\ \alpha \mapsto \beta.$$

Corollary 9.61.3 *Let $\sigma : \bar{F} \to \bar{F}$ be a F-homomorphism. Then σ is an isomorphism.*

Corollary 9.61.4 *The field $\mathbb{C}$ contains an algebraic closure for any of its subfields. In particular, $\bar{\mathbb{Q}}$, the collection of complete numbers algebraic over $\mathbb{Q}$ is an algebraic closure of $\mathbb{Q}$.*

Theorem 9.62 *The following are equivalent:*

1. *E/F is a splitting field for a collection $\{f_i\}$ of polynomials in $F[x]$.*
2. *Any F-homomorphism $\bar{F} \to \bar{F}$ maps E isomorphically to E.*
3. *Any $\alpha \in E$ is a root of a polynomial $f(x) \in F[x]$ that splits in E.*

Proof: (1) → (2): let $\sigma : \bar{F} \to \bar{F}$ be an isomorphism. Then we know $E = F(\alpha_{11}, \alpha_{12}, \cdots, \alpha_{ij}, \cdots)$, where the $\alpha'_{ij}s$ are the root of $f_i(x)$. We show $\sigma(\alpha_{ij}) \in E$. Note that α_{ij} is a the root of $f_i(x) \in F[x]$. Then $\sigma(\alpha_{ij})$ is also a root of $f_i(x)$ (since it is an F-homomorphism, so would fix F). So $\sigma : E \to E$. Hence σ is an isomorphism by Proposition (9.42).

(2) → (3) Let $\alpha \in E$, we want to show that there is a polynomial $f(x) \in F[x]$ which splits over E with $f(\alpha) = 0$. Let $m_\alpha(x)$ be the minimal polynomial of α over F and assume β is another root of $m_\alpha(x)$. Then we have a natural

isomorphism $\sigma : F(\alpha) \to F(\beta)$ which can be naturally extended to an isomorphism $\bar{F} \to \bar{F}$ by corollary (9.69.1). Since $\sigma(E) = E$, this shows that $\beta \in E$.

(3) $\to$ (1). Take $\{m_{\alpha,F}(x)\}$ where $\alpha \in E$. □

Remark 9.62.1 *This also shows that E/F is a splitting field of all the minimal polynomials of its elements.*

Corollary 9.62.1 *Suppose E_1/F is normal and E_2/F is any extension. E_1, E_2 is contained in some ambient field. Then E_1E_2/E_2 is normal.*

Proof: Let $Q = \{m_{\alpha,F}(x)\}_{\alpha \in E_1}$ denote the set of minimal polynomials of elements of E_1 over F. Then we know E_1/F is an normal extension with respect to this collection Q. We show that E_1E_2/E_2 is an normal extension with respect to Q as well. First of all, for any polynomial in Q, Q splits in E_1E_2 since it splits in E_1. Since E_1 is generated by all the root of Q over F, then E_1 is the smallest field containing all the roots of Q and F. Now suppose K/E_2 is a field in which Q splits. Then K must contain all the roots of Q, and $F \subset E_2 \subset K$, hence $E_1 \subset K$. This shows $E_1E_2 \subset K$. Hence E_1E_2/E_2 is the splitting field for Q. □

Corollary 9.62.2 *Let E/F be a normal extension. Then for any intermediate field K, $F \subset K \subset E$, E/K is normal.*

Remark 9.62.2 *K/F may not be normal.*

Proof: E is the splitting field of a collection $\{f_i(x)\}$ of polynomials in $F[x] \subset K[x]$. The second statement follows from the previous theorem. □

9.7 Separable and Inseparable Extensions

Definition 9.63 (Separable) *A polynomial over F is called **separable** if it has no multiple roots (in any extension field, or simply in $\bar{F}$). A polynomial which is not separable is called **Inseparable**.*

Definition 9.64 (Derivative) *The **derivative** of the polynomial*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in F[x]$$

is defined to be the polynomial

$$D_x f(x) = n a_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2 x + a_1 \in F[x].$$

Proposition 9.65 *A polynomial $f(x)$ has a multiple root α if and only if α is also a root of $D_x f(x)$, i.e., $f(x)$ and $D_x f(x)$ are both divisible by the minimal polynomial for α. In particular, $f(x)$ is separable if and only if it is relatively prime to its derivative, that is $(f(x), D_x f(x)) = 1$.*

Proof: Suppose first that α is a multiple root of $f(x)$. Then over a splitting field,

$$f(x) = (x-\alpha)^n g(x)$$

for some integer $n \geq 2$ and some polynomial $g(x)$. Taking derivative, we obtain

$$D_x f(x) = n(x-\alpha)^{n-1} g(x) + (x-\alpha)^n D_x g(x)$$

which show that $D_x f(x)$ has α as a root.

Conversely, suppose that α is a root of both $f(x)$ and $D_x f(x)$. Then write

$$f(x) = (x-\alpha)h(x)$$

for some polynomial $h(x)$ and take the derivative:

$$D_x f(x) = h(x) + (x-\alpha)D_x h(x).$$

Since $D_x f(\alpha) = 0$ by assumption, substituting α into the last equation show that $h(\alpha) = 0$. Hence $h(x) = (x-\alpha)h_1(x)$ for some polynomial $h_1(x)$, and $f(x) = (x-\alpha)^2 h_1(x)$ shows that α is a multiple root of $f(x)$.

The second statement follows immediately. □

Corollary 9.65.1 *Every irreducible polynomial over a field of characteristic* 0 *is separable. A polynomial over such a field is separable if and only if it is the product of distinct irreducible polynomials.*

Proof: Suppose F is a field of characteristic 0 and $p(x) \in F[x]$ is irreducible of degree n. Then the derivative $D_x p(x)$ is a polynomial of degree $n-1$. Up to constant factors the only factors of $p(x)$ in $F[x]$ are 1 and $p(x)$, so $D_x p(x) \neq 0$ (since characteristic of the field is not a prime) must be relatively prime to $p(x)$. This shows that any irreducible polynomial over a field of characteristic 0 is separable. The second statement follows because distinct irreducibles never have zeros in common. □

Definition 9.66 (Frobenius Endomorphism) *Suppose F is a field of characteristic p, then the p^{th}-power map defined by $\varphi(a) = a^p$ is an injective field homomorphism from F to F, in particular it is called the* ***Frobenius endomorphism*** *of F.*

Remark 9.66.1 *The p^{th}-power map is injective since if $x^p - y^p = 0$, then $(x-y)^p = 0$, so $x = y$.*

Corollary 9.66.1 *Suppose that $\mathbb{F}$ is a finite field of characteristic p. Then every element of $\mathbb{F}$ is a p^{th} power in $\mathbb{F}$, i.e., $\mathbb{F} = \mathbb{F}^p$.*

Proposition 9.67 *Every irreducible polynomial over a finite field F is separable. A polynomial in $F[x]$ is separable if and only if it is the product of distinct irreducible polynomials in $F[x]$.*

Proof: Let $p(x) \in F[x]$ be an irreducible polynomial. If $p(x)$ were inseparable, then we have that $p(x) = q(x^p)$ for some polynomial $q(x) \in F[x]$, as we need $D_x p(x) = 0$ (all the other terms must vanish). Let

$$q(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0.$$

Then

$$\begin{aligned} p(x) = q(x^p) &= a_m (x^p)^m + a_{m-1}(x^p)^{m-1} + \cdots + a_1 x^p + a_0 \\ &= b_m^p (x^p)^m + \cdots + b_1^p x^p + b_0^p \\ &= (b_m x^m + \cdots + b_0)^p \quad (\mathrm{Char}(F) = p) \end{aligned}$$

which shows that $P(x)$ is the p^{th} power of a polynomial in $F[x]$, a contradiction to the irreducibility of $p(x)$. □

Definition 9.68 (Perfect) *A field K of characteristic p is called **perfect** if every element of K is an p^{th} power in K, i.e., $K = K^p$. Any field of characteristic 0 is also called perfect.*

Lemma 9.69 *If F is algebraically closed, then F is perfect.*

Proof: We only consider the case where $\mathrm{Char}\, F = p$. Then for any $\alpha \in F$, we show exists $\beta \in F$ such that $\beta^p = 0$. This is true since $x^p - \alpha = 0$ has a root. □

Corollary 9.69.1 *Every irreducible polynomial over a perfect field is irreducible.*

Proof: Similar to the proof of Proposition (9.67). □

Corollary 9.69.2 *Suppose $F \subset K$ are fields, and F is perfect. If $f(x) \in F[x]$ has no repeated irreducible factors in $F[x]$, then $f(x)$ has no repeated irreducible factors in $K[x]$.*

Proof: $f(x)$ is separable in F, hence it is also separable in K. Then it must have no repeated irreducible factors. □

Lemma 9.70 *Suppose K is a field of characteristic p which is not perfect, i.e., $K \neq K^p$. Then there exists an irreducible inseparable polynomial over K. In this case, K has an inseparable finite extension.*

Proof: Since $K \neq K^p$, let $c \in K$ be such that $c \notin K^p$. Consider $f(x) = x^p - c$, which has a root in the algebraic closure of K, denote this root by α. Then $x^p - c = x^p - \alpha^p = (x - \alpha)^p$ is inseparable. Also f is irreducible since if $g | f$, then $g = (x - \alpha)^q$, so $q\alpha \in K$ which implies $q = p$. Moreover, $F(\alpha)$ is an inseparable finite extension. □

Proposition 9.71 *Let $p(x)$ be an irreducible polynomial over a field F of characteristic p. Then there is a unique integer $k \geq 0$ and a unique irreducible separable polynomial $p_{sep}(x) \in F[x]$ such that*

$$p(x) = p_{sep}(x^{p^k}).$$

Proof: We have seen above that if $p(x)$ is an irreducible polynomial which is not separable, then its derivative $D_x p(x)$ is identically 0, so that $p(x) = p_1(x^p)$ for some polynomial $p_1(x)$. The polynomial $p_1(x)$ may or may not itself be separable. If not, then it too is a polynomial in x^p, $p_1(x) = p_2(x^p)$, so that $p(x)$ is a polynomial in x^{p^2} : $p(x) = p_2(x^{p^2})$. Continuing in this fashion we see that there is a uniquely defined power p^k of p such that $p(x) = p_k(x^{p_k})$ where $p_k(x)$ has nonzero derivative (this process terminates since the degree of p_k is strictly decreasing, and p_k is never a constant). It is clear that $p_k(x)$ is irreducible since any factorization of $p_k(x)$ would, after replacing x by x^{p^k} , immediately imply a factorization of the irreducible $p(x)$. It follows that $p_k(x)$ is separable. □

Remark 9.71.1 *Let $p(x)$ be an irreducible polynomial over a field of characteristic p. The degree of $p_{sep}(x)$ in the proposition is called the **separable degree** of $p(x)$, denote $\deg_s p(x)$. The integer p^k in the proposition is called the inseparable degree of $p(x)$, denoted $\deg_i p(x)$. Then*

$$\deg p(x) = \deg_s p(x) \deg_i p(x).$$

Note the proposition fails if the polynomial we considering is not irreducible, for example consider the polynomial $(x^{p^2} - t)(x^p - t)$ over $\mathbb{F}_p(t)$ which cannot be written in the form $f_{sep}(x^{p^k})$, yet it is inseparable since it is the product of two inseparable polynomials.

Example: we consider $\mathbb{F}_p(t)$ and $f(x) = x^p - t$. Then $f_{sep}(x) = x - t$. If we we consider $\mathbb{F}_p$ and $\alpha \in \mathbb{F}_p$ is nonzero. Then the polynomial $f(x) = x^p - x + \alpha$ is irreducible, the following is the proof:
Let r be a root of $f(x)$ in $\bar{F}$. Then $r + 1$ is also a root of f, since

$$f(r+1) = (r+1)^p - (r+1) + a = r^p - r + a = 0.$$

So $r + i$, $i \in \mathbb{F}_p$ are all the roots of f. Then if $f(x) = g(x)h(x)$ in $F[x]$ with g being degree $n \geq 1$. Then $g(x)$ is the product of linear factors $(x - r - i_k)$ in $\bar{F}[x]$, where $i_k \in F_p$. Then the coefficient of x^{n-1} term of g is $s - nr$ where s the sum of $i_k' s$ so $s \in \mathbb{F}_p$ and $n \in \mathbb{F}_p$. This shows that $r \in \mathbb{F}_p$ which is a contradiction, as f has no roots in $\mathbb{F}_p$. Lastly, the separable degree of f is clearly 0.

Definition 9.72 (Separable Field) *The field K is said to be **separable** over F if every element of K is the root of a separable polynomial over F (equivalently, the minimal polynomial of any element $\alpha \in K$ over F is separable, in this case, we will call α separable over F). A field which is not separable is **inseparable**.*

Remark 9.72.1 *We have the equivalence since if $f(x)$ is separable and has α as a root, then $m_{\alpha,F}(x) | f(x)$.*

Corollary 9.72.1 *Any finite field $\mathbb{F}_{p^n}/\mathbb{F}_p$ is always separable.*

Proof: Any α is the root of $x^{p^n} - x = \prod_{i=1}^{p^n}(x - r_i)$, where r_i are all the distinct elements of $\mathbb{F}_{p^n}$. □

Example: We claim that $x^2 - t$ is not a separable polynomial in $\mathbb{F}_2(t)(\sqrt{t})$, so $\mathbb{F}_2(t)(\sqrt{t})$ is not separable over $\mathbb{F}_2(t)$. This is the case since the minimal polynomial of $\sqrt{t}$,

$$x^2 - t = x^2 - (\sqrt{t})^2 = (x - \sqrt{t})^2.$$

More generally, if F is a field of characteristic 2, let $c \in F$ be such that $c \neq \alpha^2$ for any $\alpha \in F$. Then $x^2 - c$ is not separable, and $F(\sqrt{c})/F$ is not separable. Similarly, we can argue that $\mathbb{F}_p(t)$ is not perfect.

Corollary 9.72.2 *Let E/F be algebraic such that $\mathrm{Char}(F) = 0$ (or equivalently $\mathrm{Char}(E) = 0$). Then E/F is separable.*

Proof: Consider the minimal polynomial of any element of E over F, which is irreducible, hence separable by Lemma (9.65.1). □

Example: $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is separable, but $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not normal.

Corollary 9.72.3 *Every finite extension of a perfect field is separable, every algebraic extension of a perfect field is separable. In particular, every finite extension of either $\mathbb{Q}$ or a finite field is separable.*

Proof: Follows from Corollary (9.69.1). □

Proposition 9.73 *Let F be a field of characteristic $p > 0$. Let E/F be a finite extension such that the degree $[E : F]$ is prime to p. Then E/F is separable.*

Proof: Let $\alpha \in E$, we show that $m_{\alpha,F}(x)$ splits in E. Firstly, $m_{\alpha,F}(x)$ is irreducible in $F[x]$. Next $\deg m_{\alpha,F}(x) = [F(\alpha) : F]$ is divisible by $[E : F]$. Now we know that if $m_{\alpha,F}(x)$ is not separable, then $Dm_{\alpha,F}(x) = 0$, so

$$m_{\alpha,F}(x) = g(x^p)$$

for some polynomial g. However, this shows that $p | \deg m_{\alpha,F}(x)$, which contradicts the fact that $[E : F]$ and p are relatively prime. □

Corollary 9.73.1 *If F is characteristic $p > 0$ and E/F is inseparable. Then $p|[E : F]$.*

Definition 9.74 (Separable Degree of Extensions) *Let E/F be algebraic. Let S be the set of F-homomorphism $\sigma : E \to \bar{F}$. We define the **separable degree** of E/F as the cardinality of S and denoted by $[E : F]_S$.*

Remark 9.74.1 *$[E : F]_S$ is independent of the choice of $\bar{F}$, since we can always post-composing an isomorphism.*

Lemma 9.75 *Let $F(\alpha)/F$ be finite, then*

$$\#\text{ of distinct roots of } m_\alpha(x) = [F(\alpha) : F]_S \leq [F(\alpha) : F].$$

Proof: Let $m_\alpha(x)$ be the minimal polynomial of degree $n = [F(\alpha) : F]$. Let β be any root of $m_\alpha(x)$ in $\bar{F}$.Then we have

$$F(\alpha) \qquad \cong \qquad F[x]/m_\alpha(x) \qquad \to \qquad \bar{F}$$

$$\alpha \longrightarrow x \longrightarrow \beta$$

So the number of distinct roots of $m_\alpha(x) \leq [F(\alpha) : F]_S$. On the other hand, for any $\sigma : F(\alpha) \to \bar{F}$, $\sigma(\alpha)$ is a root of $m_\alpha(x)$. Therefore, the number of distinct roots of $m_\alpha(x) \geq [F(\alpha) : F]_S$. Hence$[F(\alpha) : F]_S \leq [F(\alpha) : F]$. □

Corollary 9.75.1 *Let α be algebraic over F. Then $[F(\alpha) : F]_S = [F(\alpha) : F]$ if and only if α is separable.*

Theorem 9.76 *Let E/F and K/E be algebraic field extensions. Let $S(E/F)$ denote the set of F-homomorphisms from E to $\bar{F}$. Let $S(K/E)$ denote the set of E-homomorphism from K to $\bar{F}$. Let $S(K/F)$ denote the set of F-homomorphisms from K to $\bar{F}$. (Here we assume $F \subset E \subset K \subset \bar{F}$, in particular, can take $\bar{F} = \bar{K}$). Then the is a bijection between*

$$S(K/F) \longleftrightarrow S(K/E) \times S(E/F).$$

In particular, we have

$$[K : F]_S = [K : E]_S[E : F]_S$$

If one of the side is infinite, then so must also be the other side.

Proof: We define the bijection explicitly. Let $\bar{F}$ be their common algebraic closure. For any $\tau \in S(K/F)$, that is τ is a F-homomorphism from K to $\bar{F}$, $\tau|_E$ is a F-homomorphism from E to $\bar{F}$, which we will denote by ϕ. Now, since $E \subset \bar{F}$ and $\bar{F}$ is algebraically closed. We can extend ϕ to a F-homomorphism $\bar{F} \to \bar{F}$, which we will call ϕ'. In particular, ϕ' is a bijection by Proposition (9.42). Hence the map

$$(\phi')^{-1}\tau$$

is a well defined F homomorphism from K to $\bar{F}$ which we will denote ψ. Now if $\alpha \in E$, then

$$\psi(\alpha) = (\phi')^{-1}\tau(\alpha) = (\phi')^{-1}\phi(\alpha) = \alpha$$

Hence ψ is an E-homomorphism. Then we define

$$S(K/F) \longrightarrow S(K/E) \times S(E/F), \quad \tau \longmapsto \psi \times \phi.$$

Moreover, we can find the inverse of this map, given ψ, ϕ, we first extend ϕ to $\phi' : \bar{F} \to \bar{F}$. Then define τ by $\phi' \circ \psi$. Hence we have established a bijection. □

Lemma 9.77 *If E/F is finite, then*

$$[E : F]_S \leq [E : F].$$

Proof: Let $E = F(\alpha_1, \cdots, \alpha_n)$. Then

$$F =: F_0 \subset F(\alpha_1) =: F_1 \subset F(\alpha_1, \alpha_2) =: F_2 \subset \cdots \subset F(\alpha_1, \cdots, \alpha_n) =: F_n.$$

Then $[E : F] = [F_n : F_{n-1}][F_{n-1} : F_{n-2}] \cdots [F_1 : F_0]$. Similarly, we also have

$$[E : F]_S = [F_n : F_{n-1}]_S [F_{n-1} : F_{n-2}]_S \cdots [F_1 : F_0]_S.$$

Since we have $[F_{i+1} : F_i]_S \leq [F_{i+1} : F_i]$, the lemma follows. □

Theorem 9.78 *Let E/F be finite. Then E/F is separable if and only if $[E : F]_S = [E : F]$.*

Proof: Assume $[E : F]_S = [E : F]$, we show $\alpha \in E$ is separable for any α. We have

$$F \subset F(\alpha) \subset E \Rightarrow [E : F(\alpha)]_S [F(\alpha) : F]_S = [E : F]_S = [E : F] = [E : F(\alpha)][F(\alpha) : F]$$

Thus by Lemma (9.77), it can only be the case that $[F(\alpha) : F]_S = [F(\alpha) : F]$, hence α is separable.

Next, we assume E/F is separable. Then

$$F =: F_0 \subset F(\alpha_1) =: F_1 \subset F(\alpha_1, \alpha_2) =: F_2 \subset \cdots \subset F(\alpha_1, \cdots, \alpha_n) =: F_n.$$

Then $[E : F] = [F_n : F_{n-1}] \cdots [F_1 : F_0]$ and $[E : F]_S = [F_n : F_{n-1}]_S \cdots [F_1 : F_0]_S$. Now since α_1 is separable over F, then $[F_1 : F_0] = [F_1 : F_0]_S$. Now α_{i+1} is separable over F, so it is separable over F_i, hence $[F_{i+1} : F_i] = [F_{i+1} : F_i]_S$. □

Corollary 9.78.1 *Let E/F be a field extension. Let $\alpha \in E$ be separable. Then $F(\alpha)/F$ is separable.*

Lemma 9.79 *Let K/E and E/F be algebraic (so K/F is algebraic), then K/F is separable if and only if both K/E and E/F are separable.*

Proof: $\Rightarrow$: We assume K/F is separable, then E/F is separable since $E \subset K$. We show K/E is separable. Let $\alpha \in K$. Then we have $m_{\alpha,E}(x) | m_{\alpha,F}(x)$ where $m_{\alpha,F}(x)$ is separable. Then $m_{\alpha,E}(x)$ is also separable.

$\Leftarrow$: Let $\alpha \in K$ and $m_{\alpha,E}(x)$ be the minimal polynomial of α over E. Suppose $m_{\alpha,E}(x)$ has coefficients $a_0, a_1, \cdots, a_{n-1}$, then consider

$$F \subset F(a_0) \subset F(a_0, a_1) \subset \cdots \subset F(a_0, \cdots, a_{n-1}) \subset F(a_0, \cdots, a_{n-1}, \alpha).$$

Since each extension is separable and finite, then we conclude

$$[F(a_0, \cdots, a_{n-1}, \alpha) : F]_S = [F(a_0, \cdots, a_{n-1}, \alpha) : F].$$

Hence α is separable over F, so K/F is separable. □

Corollary 9.79.1 *Let E/F be a field of extension. Let $\{\alpha_i\}_{i\in I}$ be a collection of separable elements in E. Then*

$$F(\{\alpha_i\}_{i\in I})/F$$

is separable.

Proof: This is true since any $\alpha \in F(\{\alpha_i\}_{i\in I})$ must be contained in some finite collection $F(\beta_1, \cdots, \beta_n)$ for $\beta_k \in \{\alpha_i\}_{i\in I}$. □

Proposition 9.80

1. *Let E_1/F and E_2/F be separable. Then E_1E_2/F is separable.*
2. *Let E_1/F be separable. Then for any E_2/F, E_1E_2/E_2 is separable.*

Proof:

1. We know $E_1E_2 = F(\{\alpha\}_{\alpha\in E_1\cup E_2})$. $\alpha \in E_1$ or $\alpha \in E_2$ is separable over F, then this follows from Corollary (9.79.1).
2. Let $E_1E_2 = E_2(\{\alpha\}_{\alpha\in E_1})$. Since $\alpha \in E_1$ is separable over F, then it is separable over E_2. hence E_1E_2 is separable.

□

Proposition 9.81 *Let E/F be separable and algebraic such that $[F(\alpha) : F] \leq n$ for any $\alpha \in E$ and a fixed constant n. Then E/F is finite.*

Proof: Let $\theta \in E$ be such that $F(\theta)$ is maximal among $\{F(\alpha) : \alpha \in E\}$, we claim $F(\theta) = E$. Assume the contrary, then we must have $\alpha \in E$, such that

$$F \subset F(\theta) \subsetneq F(\theta, \alpha) \subset E.$$

Then $F(\theta, \alpha)/F$ is algebraic, finitely generated, hence $F(\theta, \alpha)/F$ is finite and separable. By the Primitive Element Theorem (10.49), we have $F(\theta, \alpha) = F(\beta)$, which is a contradiction, since $\deg m_\beta > \deg m_\theta$. □

9.8 Cyclotomic Polynomials and Extension

Definition 9.82 (Cyclotomic Polynomial) *Let μ_n denote **the group of n^{th} roots of unity over** $\mathbb{Q}$. We know $\mathbb{Z}/n\mathbb{Z} \cong \mu_n$ as groups.*

*The **primitive n^{th} root of unity** are $\xi \in \mu_n$ such that $\xi^k \neq 1$ for $1 \leq k < n$. There are precisely $\varphi(n)$ of them.*

*Define the n^{th} **cyclotomic polynomial** $\Phi_n(x)$ to be the polynomial whose roots are the primitive n^{th} roots of unity:*

$$\Phi_n(x) = \prod_{\xi \text{ primitive} \in \mu_n} (x - \xi) = \prod_{\substack{1\leq a<n \\ (a,n)=1}} (x - \xi_n^a)$$

Lemma 9.83 *Suppose m and n are relatively prime, and ξ_m, ξ_n are primitive m, nth roots respectively. Then $\xi_m\xi_n$ is a primitive mn^{th} root.*

Proof: We know $\xi_n = e^{2\pi i\frac{t}{n}}$ and $\xi_m = e^{2\pi i\frac{q}{m}}$. Where $\gcd(t,n) = 1 = \gcd(q,m)$. Then $\xi_m\xi_n = e^{2\pi i\frac{tm+qn}{mn}}$. Then by elementary argument we can show $\gcd(tm+qn, mn) = 1$. □

Lemma 9.84 *Suppose ξ_n is a primitive nth root of unity, then if $d|n$, ξ_n^d is a primitive $(n/d)^{th}$ root of unity.*

Definition 9.85 (Cyclotomic Field) *We define the **Cyclotomic field** E be the splitting field of $x^n - 1$ over $\mathbb{Q}$. Then $E = \mathbb{Q}(\xi_n)$, where ξ_n is a primitive element of μ_n.*

The roots of the polynomial $x^n - 1$ are precisely the n^{th} roots of unity so we have the factorization

$$x^n - 1 = \prod_{\xi\in\mu_n}(x - \xi).$$

If we group together the factors $(x - \xi)$ where ξ is an element of order d in μ_n, we obtain

$$x^n - 1 = \prod_{d|n}\prod_{\substack{\xi\in\mu_d\\ \xi \text{ primitive}}}(x - \xi).$$

The inside product is $\Phi_d(x)$ by definition, we have the factorization

$$x^n - 1 = \prod_{d|n}\Phi_d(x). \tag{9.1}$$

By comparing degrees, this gives the identity:

Lemma 9.86 $n = \sum_{d|n}\varphi(d)$.

Equation 9.1 allows us to compute $\Phi_n(x)$ for any n recursively. Clearly $\Phi_1(x) = x - 1$ and $\Phi_2(x) = x + 1$. Then we have the following:

- $\Phi_p(x) = \frac{x^p-1}{\Phi_1(x)} = x^{p-1} + x^{p-2} + \cdots + x + 1$ for p being a prime;
- $\Phi_4(x) = x^2 + 1$;
- $\Phi_6(x) = x^2 - x + 1$;
- $\Phi_8(x) = x^4 + 1$;
- $\Phi_9(x) = x^6 + x^3 + 1$;
- $\Phi_{10}(x) = x^4 - x^3 + x^2 - x + 1$;
- $\Phi_{12}(x) = x^4 - x^2 + 1$.

Remark 9.86.1 *We note that $\mathbb{Q}(\xi_8)$ contains $\sqrt{2}$. This causes problems when we want to determining the splitting field of $x^8 - 2$.*

Lemma 9.87 *The cyclotomic polynomial $\Phi_n(x)$ is a monic polynomial in $\mathbb{Z}[x]$ of degree $\varphi(n)$.*

Proof: It is clear that $\Phi_n(x)$ is monic and has degree $\varphi(n)$ by definition. We must show that the coefficients lie in $\mathbb{Z}$. We use induction on n. The result is true for $n = 1$. Assume by induction that $\Phi_d(x) \in \mathbb{Z}[x]$ for all $1 \leq d < n$. Then $x^n - 1 = f(x)\Phi_n(x)$ where

$$f(x) = \prod_{\substack{d|n \\ d<n}} \Phi_d(x)$$

is monic and has coefficients in $\mathbb{Z}$. Since $f(x)$ clearly divides $x^n - 1$ in $F[x]$ where $F = \mathbb{Q}(\xi_n)$ is the field of n^{th} roots of unity and both $f(x)$ and $x^n - 1$ have coefficients in $\mathbb{Q}$, $f(x)$ divides $x^n - 1$ in $\mathbb{Q}$ by the Division Algorithm. By Gauss' Lemma, $f(x)$ divides $x^n - 1$ in $\mathbb{Z}[x]$, hence $\Phi_n(x) \in \mathbb{Z}[x]$. □

Theorem 9.88 *The cyclotomic polynomial $\Phi_n(x)$ is an irreducible monic polynomial in $\mathbb{Z}[x]$ of degree $\varphi(n)$.*

Proof: Suppose towards a contradiction,

$$\Phi_n(x) = f(x)g(x), \quad f(x), g(x) \in \mathbb{Z}[x] \text{ and is monic}$$

where we take $f(x)$ to be an irreducible factor of $\Phi_n(x)$. Let ξ be a primitive n^{th} root of 1 which is a root of $f(x)$(so then $f(x)$ is the minimal polynomial for ξ over $\mathbb{Q}$) and let p denote any prime not dividing n. Then ξ^p is again a primitive n^{th} root of 1, hence is a root of either $f(x)$ or $g(x)$.

Suppose $g(\xi^p) = 0$. Then ξ is a root of $g(x^p)$ and since $f(x)$ is the minimal polynomial for ξ, $f(x)$ must divide $g(x^p)$ in $\mathbb{Z}[x]$, say

$$g(x^p) = f(x)h(x), \quad h(x) \in \mathbb{Z}[x].$$

If we reduce this equation mod p, we obtain

$$\bar{g}(x^p) = \bar{f}(x)\bar{h}(x) \quad \text{in } \mathbb{F}_p[x].$$

Note

$$\bar{g}(x^p) = (\bar{g}(x))^p$$

then

$$(\bar{g}(x))^p = \bar{f}(x)\bar{h}(x)$$

in the Unique Factorization Domain $\mathbb{F}_p[x]$. It follows that $\bar{f}(x)$ and $\bar{g}(x)$ have a factor in common in $\mathbb{F}_p[x]$.

Now from $\Phi_n(x) = f(x)g(x)$, we see by reducing mod p that $\bar{\Phi}_n(x) = \bar{f}(x)\bar{g}(x)$, so $\bar{\Phi}_n(x) \in \mathbb{F}_p[x]$ has a multiple root. But then $x^n - 1$ would have a multiple root over $\mathbb{F}_p$ since it has $\bar{\Phi}_n(x)$ as a factor. This is a contradiction since all n roots of $x^n - 1$ are distinct over any field of characteristic not dividing n.

Hence ξ^p must be a root of $f(x)$. Since this applies to every root ξ of $f(x)$, it follows that ξ^a is a root of $f(x)$ for every integer a relatively prime to n (factorize a in terms of primes). But this means that every primitive n^{th} root of unity is a root of $f(x)$, i.e., $f(x) = \Phi_n(x)$, showing $\Phi_n(x)$ is irreducible. □

Corollary 9.88.1 *The degree over $\mathbb{Q}$ of the cyclotomic field of n^{th} roots of unity is $\varphi(n)$:*

$$[\mathbb{Q}(\xi_n) : \mathbb{Q}] = \varphi(n).$$

Proof: $\Phi_n(x)$ is the minimal polynomial for any primitive n^{th} root of unity ξ_n and is of degree $\varphi(n)$. □

Corollary 9.88.2 *If a field contains the n^{th} root of unity for n odd then it also contains the $2n$th root of unity.*

Proof: The field contains 1, hence $\mathbb{Q}$, so suffices to show that $\mathbb{Q}(\xi_{2n}) = \mathbb{Q}(\xi_n)$ when n is odd. Clearly $\mathbb{Q}(\xi_n) \subset \mathbb{Q}(\xi_{2n})$. $[\mathbb{Q}(\xi_{2n}) : \mathbb{Q}] = \varphi(2n) = \varphi(n) = [Q(\xi_n) : \mathbb{Q}]$, hence we have equality. □

10 Galois Theory

10.1 Basic Definitions

Let K be a field in this section.

Definition 10.1 (Automorphism) *An isomorphism σ of K with itself is called an **automorphism** of K. The collection of automorphisms of K is denoted $\mathrm{Aut}(K)$. If $\alpha \in K$ we shall write $\sigma\alpha$ for $\sigma(\alpha)$. An automorphism $\sigma \in \mathrm{Aut}(K)$ is said to fix an element $\alpha \in K$ if $\sigma\alpha = \alpha$. If F is a subset of K, then an automorphism σ is said to fix F if it fixes all the elements of F, i.e., $\sigma a = a$ for all $a \in F$.*

Remark 10.1.1 *Since any automorphism σ takes 1 to 1, then it follows that $\sigma a = a$ for all a in the prime field. In particular, $\mathbb{Q}$ and $\mathbb{F}_p$ have only the trivial automorphism.*

Definition 10.2 ($\mathrm{Aut}(K/F)$) *Let K/F be an extension of fields. Let $\mathrm{Aut}(K/F)$ be the collection of automorphisms of K which fix F.*

Remark 10.2.1 *If F is the prime subfield of K, then $\mathrm{Aut}(K) = \mathrm{Aut}(K/F)$.*

Proposition 10.3 *$\mathrm{Aut}(K)$ is a group under composition and $\mathrm{Aut}(K/F)$ is a subgroup of $\mathrm{Aut}(K)$.*

Proposition 10.4 *Let K/F be a field extension and let $\alpha \in K$ be algebraic over F. Then for any $\sigma \in \mathrm{Aut}(K/F)$, $\sigma\alpha$ is a root of the minimal polynomial for α over F, i.e., $\mathrm{Aut}(K/F)$ permutes the roots of irreducible polynomials. Equivalently, any polynomial with coefficients in F having α as a root also has $\sigma\alpha$ as a root.*

Proof: Suppose α satisfies the equation

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$$

Then it is clear that

$$(\sigma\alpha)^n + a_{n-1}(\sigma(\alpha))^{n-1} + \cdots + a_1(\sigma\alpha) + a_0 = 0$$

But this says precisely that $\sigma\alpha$ is a root of the same polynomial over F as α. □

In general, if K is generated over F by some collection of elements, then any automorphism $\sigma \in \mathrm{Aut}(K/F)$ is completely determined by what it does to the generators. If K/F is finite, then K is finitely generated over F by algebraic elements so by the proposition the number of automorphism of K fixing F is finite, i.e., $\mathrm{Aut}(K/F)$ is a finite group.

Theorem 10.5 *Let K be a field and $K(x)$ be the field of fractions of $K[x]$. Then any automorphism in $\mathrm{Aut}(K(x)/K)$ must be of the form $t \mapsto \frac{ax+b}{cx+d}$ where $a, b, c, d \in K$ and $ad - bc \neq 0$*

Proof: Firstly, we note that $t \mapsto \frac{ax+b}{cx+d}$ is an automorphism, since it is a field homomorphism and we can find its inverse. Clearly, it fixes K as well. So we just need to show any automorphism is of this form. Suffices to determine the image of x under the the automorphism. Suppose $x \mapsto y \in K(x)$, and $y = \frac{f(x)}{g(x)}$, where $f, g \in K[x]$ are coprime. Let $n = \max\{\deg f(x), g(x)\}$, then $[K(x) : K(y)] = n$, hence this map is not surjective if $n \geq 1$. This implies we must have f, g are at most degree 1 polynomials. □

Proposition 10.6 *Let K be an extension of the field F. Let $\varphi : K \to K'$ be an isomorphism of K with a field K' which maps F to the subfield F' of K'. Then the map $\sigma \mapsto \varphi\sigma\varphi^{-1}$ defines a group isomorphism $\mathrm{Aut}(K/F) \xrightarrow{\cong} \mathrm{Aut}(K'/F')$.*

Proposition 10.7 *Let $H \leq \mathrm{Aut}(K)$ be a subset of the group of automorphisms of K. Then the collection F of elements of K fixed by all the elements of H is a subfield of K.*

Proof: Let $h \in H$ and let $a, b \in F$. Then by definition $h(a) = a$, $h(b) = b$, so that $h(a \pm b) = h(a) \pm h(b) = a \pm b$ and $h(ab) = h(a)h(b) = ab$ and $h(a^{-1}) = h(a)^{-1} = a^{-1}$, so that F is closed, hence a subfield of K. □

Definition 10.8 (Fixed Field) *If H is a subgroup of the group of automorphisms of K, the subfield of K fixed by all the elements of H is called the **fixed field** of H and denoted by K^H.*

Lemma 10.9 *If K is generated over F by $\alpha_1, \cdots, \alpha_n$, then the automorphism σ of K fixing F is uniquely determined by $\sigma(\alpha_1), \cdots, \sigma(\alpha_n)$. In particular, an automorphism fixes K if and only if it fixes a set of generators of K.*

Lemma 10.10 *Let $G \leq \mathrm{Gal}(K/F)$ be a subgroup of the Galois group of the extension K/F. Suppose $\sigma_1, \cdots, \sigma_k$ are generators for G. Then the subfield E/F is fixed by G if and only if it is fixed by the generators $\sigma_1, \cdots, \sigma_k$.*

Proposition 10.11 *The association of groups to fields and fields to groups is inclusion reversing:*

- *If $F_1 \subset F_2 \subset K$ are two subfields of K, then $\mathrm{Aut}(K/F_2) \leq \mathrm{Aut}(K/F_1)$.*
- *If $H_1 \leq H_2 \leq \mathrm{Aut}(K)$ are two subgroups of automorphism with associated fixed fields F_1 and F_2, respectively, then $F_2 \subseteq F_1$.*

Definition 10.12 (Galois Extension) *Let K/F be a finite extension. Then K is said to be **Galois** over F and K/F is a **Galois extension** if $|\mathrm{Aut}(K/F)| = [K : F]$. If K/F is Galois, then the group of automorphism $\mathrm{Aut}(K/F)$ is called the **Galois group** of K/F, denoted $\mathrm{Gal}(K/F)$.*

Example: $[\mathbb{C} : \mathbb{R}] = 2 = |\mathrm{Aut}(\mathbb{C}/\mathbb{R})|$, since $\mathrm{Aut}(\mathbb{C}/\mathbb{R}) \cong \mathbb{Z}/2\mathbb{Z}$, with the generator σ which maps z to its complex conjugate.

Lemma 10.13 *Let E/F be finite. Then*

$$|\mathrm{Aut}(E/F)| \leq [E : F]_S$$

The equality holds if and only if E/F is normal.

Proof: Any F-homomorphism $\sigma : E \to E$ can be extend to $\sigma : E \to \bar{F}$. So $|\operatorname{Aut}(E/F)| \leq [E : F]_S$. Assume E/F is normal, then any F-homomorphism $\sigma : E \to \bar{F}$ maps $\sigma : E \to E$ isomorphically (since σ can be extended to $\sigma : \bar{F} \to \bar{F}$, which maps E to E, then consider σ^{-1}). This shows $[E : F]_S \leq |\operatorname{Aut}(E/F)|$, so $|\operatorname{Aut}(E/F)| = [E : F]_S$. □

Corollary 10.13.1 *If E/F is finite and $|\operatorname{Aut}(E/F)| = [E : F]_S$. Then any $\sigma : E \to \bar{F}$ must restricts to $\sigma : E \to E$, so E/F is normal.*

Theorem 10.14 *Let E/F be finite. Then E/F is Galois if and only if E/F is normal and separable.*

Proof: We have $|\operatorname{Aut}(E/F)| \leq [E : F]_S \leq [E : F]$. Then $|\operatorname{Aut}(E/F)| = [E : F]$ if and only if E/F is normal and separable. □

Corollary 10.14.1 *If K is the splitting field over F of a separable polynomial $f(x)$, then K/F is Galois.*

Corollary 10.14.2 *Suppose K/F is Galois, then any extension K/E is Galois with $F \subset E \subset K$. In particular, $|\operatorname{Aut}(K/E)| = [K : E]$.*

Proof: This is true since K/E is normal and separable. □

Definition 10.15 (Generalization Of Galois Extension) *An extension E/F is called **Galois** if it is algebraic, normal and separable. In this case $\operatorname{Aut}(E/F)$ is called the **Galois group** of the extension and is denoted by $\operatorname{Gal}(E/F)$.*

Example:

1. The extension $\mathbb{F}_{p^n}/\mathbb{F}_p$ is Galois, since it is the splitting field of $x^{p^n} - x$. Moreover $|\operatorname{Gal}(E/F)| = n$. We claim

$$\operatorname{Gal}(\mathbb{F}^{p^n}/\mathbb{F}_p) \cong \mathbb{Z}/n\mathbb{Z} = \langle Fr \rangle$$

   where Fr is the Frobenius morphism, $Fr : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$, given by $x \mapsto x^p$. Since $\mathbb{F}^{p^n}$ is finite, then this is an isomorphism By Lagrange's theorem, we have $Fr = id$ on $\mathbb{F}_p \to \mathbb{F}_p$. So we have $Fr \in \operatorname{Gal}(\mathbb{F}_{p^n}/F_p)$. We claim that Fr is of order n. Suffices to show Fr^k is nontrivial for $k = 1, \cdots, n-1$ and Fr^n is the identity map. Fr^n maps x to $x^{p^n} = x$, so Fr^n is the identity map. On the other hand, for any $k \in \{1, \cdots, n-1\}$, we show Fr^k is not the identity, i.e. there exists x such that $x^{p^k} \neq x$. This is true since a degree p^k polynomial has at most p^k roots.

2. The extension $\overline{\mathbb{Q}}/\mathbb{Q}$ is Galois, so we conclude that $|\operatorname{Aut}(\overline{\mathbb{Q}}/\mathbb{Q})| = \infty$ since the extension is not finite.

3. The extension $\mathbb{Q}(\sqrt{-1})/\mathbb{Q}$ is Galois with $\operatorname{Gal}(\mathbb{Q}(\sqrt{-1})/\mathbb{Q}) = \mathbb{Z}/2\mathbb{Z}$. Since this extension is separable (characteristic 0) and normal since it is the splitting field for $x^2 + 1 = 0$. The Galois groups is generated by the map $a + b\sqrt{-1} \mapsto a - b\sqrt{-1}$.

4. The extension $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is Galois with Galois group isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

5. Let $E/\mathbb{Q}$ be the splitting field of $x^3-2=(x-\sqrt[3]{2})(x-\sqrt[3]{2}\xi_3)(x-\sqrt[3]{2}\xi_3^2)$. Then E is Galois. We know $[E:Q]=6$. Let $G=\mathrm{Gal}(E/\mathbb{Q})$. Since $\sigma\in G$ must map a root of x^3-2 to a root of x^3-2, then $G\hookrightarrow S_3$. By a degree argument, we conclude $G\cong S_3$.

6. The extension $\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}$ is Galois. It is the splitting field of $(x^2-2)(x^3-3)$. Then we have $\mathrm{Gal}(\mathbb{Q}(\sqrt{2},\sqrt{3})\mathbb{Q})\hookrightarrow S_4$. However, the permutation $\sqrt{2}\to\sqrt{3}$, $-\sqrt{2}\to-\sqrt{2}$, $\sqrt{3}\mapsto\sqrt{2}$, $-\sqrt{3}\to-\sqrt{3}$ is not possible. In conclusion, we get $\mathrm{Gal}(\mathbb{Q}(\sqrt{2},\sqrt{3})\mathbb{Q})=\mathbb{Z}_2\times\mathbb{Z}_2$.

   We also know that $\mathbb{Q}(\sqrt{2},\sqrt{3})=\mathbb{Q}(\sqrt{2}+\sqrt{3})$. The minimal polynomial of $\sqrt{2}+\sqrt{3}$ is x^4-10x^2+1.

7. Let F be a field of Char 0. We consider the field $F(a,b,c,d,e,f)$, where a,b,c,d,e,f are in determinants. We consider the splitting field E of $ax^5+bx^4+cx^3+dx^2+ex+f$. Then with tools in later section, we can show $\mathrm{Gal}(E/F)\cong S_5$.

Definition 10.16 (Galois group) *If $f(x)$ is a separable polynomial over F, then the **Galois group** of $f(x)$ over F is the Galois group of the splitting field of $f(x)$ over F.*

Remark 10.16.1 *The inverse Galois Problem: for any finite G, is there a Galois extension $E/\mathbb{Q}$ such that $\mathrm{Gal}(E/\mathbb{Q})\cong G$. This is very difficult.*

Proposition 10.17 *Let $f(x)$ be a irreducible polynomial in $\mathbb{Q}[x]$ of $\deg p$ for some prime p. Let E be the splitting field of $f(x)$. Assume $f(x)$ has exactly two non-real roots. Then $\mathrm{Aut}(E/\mathbb{Q})\cong S_p$.*

Proof: Let $G=\mathrm{Aut}(E/\mathbb{Q})$. We know $G\hookrightarrow S_p$. We know $p|[E:\mathbb{Q}]$, since $[\mathbb{Q}(\alpha_1):\mathbb{Q}]=p$. So $p||G|$, and by Sylow's Theorem, there is a subgroup of order p or an element σ of order p, which can only be the p-cycle. On the other hand, there are two non-real roots of $f(x)$. Hence complex conjugate is an automorphism, and its representation in S_p is a transposition. Since a transposition and n-cycle generate S_n, then $G\cong S_p$. □

Example: Let $f(x)=x^5-4x+2$ over $\mathbb{Q}$. This is irreducible. Then by Calculus, we can get 3 real roots, which implies there are two non-real roots. This implies $\mathrm{Gal}(E/\mathbb{Q})\cong S_5$.

10.2 Character, Trace, Norm

Notation: let G be a group / monoid and F be a field. We denote by $F(G)$ the F-vector spaces of F-valued functions $G\to F$.

Definition 10.18 (Character) *A **character** χ of a group G with values in a field L is a homomorphism from G to the multiplicative group of L:*

$$\chi: G\to L^\times$$

i.e., $\chi(g_1g_2)=\chi(g_1)\chi(g_2)$ for all $g_1,g_2\in G$ and $\chi(g)$ is a nonzero element of L for all $g\in G$. Similarly we can define a character $\chi: M\to F^\times$ for a monoid M.

Remark 10.18.1 *For any field homomorphism $\sigma : F \to K$, we may view σ as a group homomorphism from $F^\times \to K^\times$, hence we obtain a character of $F^\times$ with values in the field K.*

Definition 10.19 (Linear Independence of Characters) *The characters $\chi_1, \chi_2, \cdots, \chi_n$ of G are said to be **linearly independent** over L if they have linearly independent as functions on G, i.e., there is no trivial relation*

$$a_1\chi_1 + a_2\chi_2 + \cdots + a_n\chi_n = 0 \text{ as a function}$$

*where $a_1, \cdots, a_n \in L$ are not all zero. The set of characters $\{\chi_i\}_{i\in I}$ of G are said to be **linearly independent** over L if any finite subset of $\{\chi_i\}_{i\in I}$ is linearly independent.*

Theorem 10.20 (Linear Independence of Characters) *If $\chi_1, \chi_2, \cdots, \chi_n$ are distinct characters of G with values in L then they are linearly independent over L.*

Proof: Suppose the characters were linearly dependent, among all the linear dependence relations (that is $a_1\chi_1 + a_2\chi_2 + \cdots + a_n\chi_n = 0$), choose one with the minimal number m of nonzero coefficients a_i. We may suppose (by renumbering), that the m nonzero coefficients are $a_1, \cdots, a_m$, then

$$a_1\chi_1 + a_2\chi_2 + \cdots + a_m\chi_m = 0.$$

Then for any $g \in G$, we have

$$a_1\chi_1(g) + a_2\chi_2(g) + \cdots + a_m\chi_m(g) = 0 \tag{10.1}$$

Let g_0 be an element with $\chi_1(g_0) \neq \chi_m(g_0)$ (since they are distinct, we can find such g_0), we need

$$a_1\chi_1(g_0g) + a_2\chi(g_0g) + \cdots + a_m\chi_m(g_0g) = 0,$$

so for all $g \in G$, we have

$$a_1\chi_1(g_0)\chi_1(g) + a_2\chi_2(g_0)\chi_2(g) + \cdots + a_m\chi_m(g_0)\chi_m(g) = 0.$$

But multiplying $\chi_m(g)$ on both sides of Equation (10.1), we have

$$a_1\chi_m(g_0)\chi_1(g) + \cdots + a_m\chi_m(g_0)\chi_m(g) = 0.$$

This gives

$$[\chi_m(g_0) - \chi_1(g_0)]a_1\chi_1(g) + [\chi_m(g_0) - \chi_2(g_0)]a_2\chi_2(g) + \cdots + [\chi_m(g_0) - \chi_{m-1}(g_0)]a_{m-1}\chi_{m-1}(g) = 0$$

which holds for all $g \in G$. But the first coefficient is nonzero and this is a relation with fewer nonzero coefficients, which is a contradiction. □

Corollary 10.20.1 *If $\{\chi_i\}_{i\in I}$ is a set of distinct characters of G with values in L, then they are linearly independently over L.*

Corollary 10.20.2 *Let M be a monoid and F be a field. Let $\{\chi_i\}$ be a collection of characters of M with coefficients in F, then they are linearly independent.*

Proposition 10.21 *Let $\alpha_1, \cdots, \alpha_n$ be distinct non-zero elements in a field K. Let $a_1, \cdots, a_n \in K$. Then if for any $m \in \mathbb{Z}_{\geq 0}$,*

$$a_1\alpha_1^m + a_2\alpha_2^m + \cdots + a_n\alpha_n^m = 0$$

then $a_i = 0$ for all i.

Proof: We consider characters

$$\chi_i : \mathbb{Z}_{\geq 0} \to K^\times, \quad m \mapsto \alpha_i^m.$$

Then we know $\{\chi_i\}$ are linearly independent. Therefore, we know $a_i = 0$ for all i by Corollary (10.20.2) □

Theorem 10.22 *Let $G = \{\sigma_1 = 1, \sigma_2, \cdots, \sigma_n\}$ be a subgroup of automorphisms of a field K and let $F = K^G$ be the fixed field. Then*

$$[K : F] = n = |G|$$

Proof: Suppose first that $n > [K : F]$ and let $\omega_1, \omega_2, \ldots, \omega_m$ be a basis for K over F (m = [K : F]). Then the system

$$\begin{aligned} \sigma_1(\omega_1)x_1 + \sigma_2(\omega_1)x_2 + \cdots + \sigma_n(\omega_1)x_n &= 0 \\ \vdots \qquad\qquad & \qquad \vdots \\ \sigma_1(\omega_m)x_1 + \sigma_2(\omega_m)x_2 + \cdots + \sigma_n(\omega_m)x_n &= 0 \end{aligned}$$

of m equations in n unknowns $x_1, x_2, \ldots, x_n$, has a nontrivial solution $\beta_1, \beta_2, \ldots, \beta_n \in K$ since by assumption there are more unknowns than equations.

Let $a_1, a_2, \ldots, a_m$ be m arbitrary elements of F. The field F is by definition fixed by $\sigma_1, \ldots, \sigma_n$ so each of these elements is fixed by every σ_i, i.e., $\sigma_i(a_j) = a_j$, $i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$. Multiplying the first equation above by a_1, the second by $a_2, \ldots$, the last by a_m then gives the system of equations

$$\begin{aligned} \sigma_1(a_1\omega_1)\beta_1 + \sigma_2(a_1\omega_1)\beta_2 + \cdots + \sigma_n(a_1\omega_1)\beta_n &= 0 \\ \vdots \qquad\qquad & \qquad \vdots \\ \sigma_1(a_m\omega_m)\beta_1 + \sigma_2(a_m\omega_m)\beta_2 + \cdots + \sigma_n(a_m\omega_m)\beta_n &= 0. \end{aligned}$$

Adding these equations we see that there are elements $\beta_1, \ldots, \beta_n \in K$, not all 0, satisfying

$$\sigma_1(a_1\omega_1 + a_2\omega_2 + \cdots + a_m\omega_m)\beta_1 + \cdots + \sigma_n(a_1\omega_1 + a_2\omega_2 + \cdots + a_m\omega_m)\beta_n = 0$$

for all choices of $a_1, \ldots, a_m$ in F. Since $\omega_1, \ldots, \omega_m$ is an F-basis for K, every $\alpha \in K$ is of the form $a_1\omega_1 + a_2\omega_2 + \cdots + a_m\omega_m$, so the previous equation means

$$\sigma_1(\alpha)\beta_1 + \cdots + \sigma_n(\alpha)\beta_n = 0$$

for all $\alpha \in K$. But this means the distinct automorphisms $\sigma_1, \ldots, \sigma_n$ are linearly dependent over K, contradicting Corollary **??**.

Suppose now that $n < [K : F]$. Then there are more than n F-linearly independent elements of K, say $\alpha_1, \ldots, \alpha_{n+1}$. The system:

$$\begin{aligned} \sigma_1(\alpha_1)x_1 + \sigma_1(\alpha_2)x_2 + \cdots + \sigma_1(\alpha_{n+1})x_{n+1} &= 0 \\ \vdots \qquad\qquad & \qquad \vdots \\ \sigma_n(\alpha_1)x_1 + \sigma_n(\alpha_2)x_2 + \cdots + \sigma_n(\alpha_{n+1})x_{n+1} &= 0 \end{aligned} \tag{1}$$

of n equations in $n+1$ unknowns $x_1, \ldots, x_{n+1}$ has a solution $\beta_1, \ldots, \beta_{n+1}$ in K where not all the β_i are 0. If all the elements of the solution $\beta_1, \ldots, \beta_{n+1}$ were elements of F then the first equation (recall $\sigma_1 = 1$ is the identity automorphism) would contradict the linear independence over F of $\alpha_1, \alpha_2, \ldots, \alpha_{n+1}$. Hence at least one $\beta_i, i = 1, 2, \ldots, n+1$, is not an element of F.

Among all the nontrivial solutions $(\beta_1, \ldots, \beta_{n+1})$ of the system (1) choose one with the minimal number r of nonzero β_i. By renumbering if necessary we may assume $\beta_1, \ldots, \beta_r$ are nonzero. Dividing the equations by β_r we may also assume $\beta_r = 1$. We have already seen that at least one of $\beta_1, \ldots, \beta_{r-1}$ is not an element of F (which shows in particular that $r > 1$), say $\beta_1 \notin F$. Then our system of equations reads

$$\begin{aligned} \sigma_1(\alpha_1)\beta_1 + \cdots + \sigma_1(\alpha_{r-1})\beta_{r-1} + \sigma_1(\alpha_r) &= 0 \\ \vdots \qquad\qquad & \qquad \vdots \\ \sigma_n(\alpha_1)\beta_1 + \cdots + \sigma_n(\alpha_{r-1})\beta_{r-1} + \sigma_n(\alpha_r) &= 0 \end{aligned} \tag{2}$$

Since $\beta_1 \notin F$, there is an automorphism $\sigma_{k_0} (k_0 \in \{1, 2, \ldots, n\})$ with $\sigma_{k_0}\beta_1 \neq \beta_1$. If we apply the automorphism σ_{k_0} to the equations in (6), we obtain the system of equations

$$\sigma_{k_0}\sigma_j(\alpha_1)\sigma_{k_0}(\beta_1) + \cdots + \sigma_{k_0}\sigma_j(\alpha_{r-1})\sigma_{k_0}(\beta_{r-1}) + \sigma_{k_0}\sigma_j(\alpha_r) = 0 \quad \text{for} \quad j = 1, 2, \ldots, n. \tag{3}$$

But the elements

$$\sigma_{k_0}\sigma_1, \sigma_{k_0}\sigma_2, \ldots, \sigma_{k_0}\sigma_n$$

are the same as the elements

$$\sigma_1, \sigma_2, \ldots, \sigma_n$$

in some order since these elements form a group. In other words, if we define the index i by $\sigma_{k_0}\sigma_j = \sigma_i$ then i and j both run over the set $\{1, 2, \ldots, n\}$. Hence the equations in (3) can be written

$$\sigma_i(\alpha_1)\sigma_{k_0}(\beta_1) + \cdots + \sigma_i(\alpha_{r-1})\sigma_{k_0}(\beta_{r-1}) + \sigma_i(\alpha_r) = 0. \tag{3'}$$

If we now subtract the equations in (3′) from those in (2) we obtain the system

$$\sigma_i(\alpha_1)[\beta_1 - \sigma_{k_0}(\beta_1)] + \cdots + \sigma_i(\alpha_{r-1})[\beta_{r-1} - \sigma_{k_0}(\beta_{r-1})] = 0$$

for $i = 1, 2, \ldots, n$. But this is a solution to the system of equations (1) with

$$x_1 = \beta_1 - \sigma_{k_0}(\beta_1) \neq 0$$

(by the choice of k_0), hence is nontrivial and has fewer than r nonzero x_i. This is a contradiction and completes

the proof. □

Corollary 10.22.1 *Let K/F be any finite extension. Then*

$$|\operatorname{Aut}(K/F)| \leq [K:F].$$

with equality if and only if F is the fixed field of $\operatorname{Aut}(K/F)$. Put another way, K/F is Galois if and only if F is the fixed field of $\operatorname{Aut}(K/F)$.

Proof: Let F_1 be the fixed field of $\operatorname{Aut}(K/F)$, so that

$$F \subseteq F_1 \subseteq K.$$

By Theorem 10.22, $[K:F_1] = |\operatorname{Aut}(K/F)|$. Hence $[K:F] = |\operatorname{Aut}(K/F)|[F_1:F]$ which proves the corollary. □

Corollary 10.22.2 *Let G be a finite subgroup of automorphisms of a field K and let F be the fixed field. Then every automorphism fixing F is contained in G, i.e., $\operatorname{Aut}(K/F) = G$, so that K/F is Galois, with Galois group G.*

Proof: By definition F is fixed by all the elements of G so we have $G \leq \operatorname{Aut}(K/F)$. Hence $|G| \leq |\operatorname{Aut}(K/F)|$. By Theorem 10.22, we have $|G| = [K:F]$, and by Corollary 10.22.1, $|\operatorname{Aut}(K/F)| \leq [K:F]$, this gives

$$[K:F] = |G| = |\operatorname{Aut}(K/F)|.$$

□

Corollary 10.22.3 *If $G_1 \neq G_2$ are distinct finite subgroups of automorphism of a field K, then their fixed fields are also distinct.*

Proof: Suppose F_1 is the fixed field of G_1 and F_2 is the fixed field of G_2. If $F_1 = F_2$, then by definition F_1 is fixed by G_2. By the previous corollary any automorphism fixing F_1 is contained in G_1, hence $G_2 \leq G_1$. Similarly, we can show $G_1 \leq G_2$, so $G_1 = G_2$ which is a contradiction. □

Definition 10.23 (Trace and Norm) *Let E/F be a finite extension. Let $\alpha \in E$, then we have a F-linear transformation*

$$T_\alpha : E \to E, \quad x \mapsto \alpha x.$$

*Then we define the **norm** of α,*

$$N^T_{E/F}(\alpha) = \det(T_\alpha)$$

*and we define the **trace** of α, by*

$$Tr^T_{E/F}(\alpha) = Tr(T_\alpha).$$

Remark 10.23.1 *Let $c_x(\alpha) = \det(xI - T_\alpha) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$, be the characteristic polynomial, then we know $N^T_{E/F}(\alpha) = (-1)^n a_0$ and $Tr^T_{E/F}(\alpha) = -a_{n-1}$.*

Definition 10.24 (Alternative Definition of Norm and Trace) *Let E/F be finite and separable. Let $\alpha \in E$, then we define the* ***norm***

$$N_{E/F}(\alpha) = \prod_{\sigma \in S(E/F)} \sigma(\alpha).$$

We define the ***trace*** *as*

$$Tr_{E/F}(\alpha) = \sum_{\sigma \in S(E/F)} \sigma(\alpha),$$

where $S(E/F)$ is the set of all F-homomorphism from E to $\bar{F}$.

Remark 10.24.1 *There is a way to define trace and norm for extensions E/F that is only finite but not separable. However, this is much more complicated and in application we will mostly assume that E/F is also separable. Hence we will assume in this section that E/F is always finite and separable.*

Example: We consider $E/F = \mathbb{C}/\mathbb{R}$. Then $S(\mathbb{C}/\mathbb{R}) = \{e, \bar{}\}$. Then for any $\alpha \in \mathbb{C}$, we have

$$N_{\mathbb{C}/\mathbb{R}}(\alpha) = e(\alpha) \cdot \bar{\alpha} = \alpha \cdot \bar{\alpha} = \|\alpha\|^2.$$

Similarly, we get

$$Tr_{\mathbb{C}/\mathbb{R}}(\alpha) = e(\alpha) + \bar{\alpha} = 2 \cdot \mathrm{Re}(\alpha).$$

Lemma 10.25 *Let $\tau : \bar{F} \to \bar{F}$ be a F-homomorphism. Then*

$$\tau(N_{E/F}(\alpha)) = N_{E/F}(\alpha), \quad \tau(Tr_{E/F}(\alpha)) = Tr_{E/F}(\alpha).$$

Proof: We prove the first equality as the second equality is proved in the same way. Since

$$\begin{aligned} N_{E/F}(\alpha) &= \prod_{\sigma \in S(E/F)} \sigma(\alpha) \\ \tau(N_{E/F}(\alpha)) &= \tau\left(\prod_{\sigma \in S(E/F)} \sigma(\alpha)\right) \\ &= \prod_{\sigma \in S(E/F)} \sigma(\alpha) \end{aligned}$$

Since $\{\sigma : E \to \bar{F}\} = \{\tau \circ \sigma : E \to \bar{F}\}$. □

Similarly, we can prove the following:

Lemma 10.26 *For any F-automorphism $\tau : E \to E$, we have*

$$N_{E/F}(\tau(\alpha)) = N_{E/F}(\alpha) \quad Tr_{E/F}(\tau(\alpha)) = Tr_{E/F}(\alpha).$$

Lemma 10.27 *Suppose E/F is finite and separable, then $N_{E/F}(\alpha) \in F$ and $Tr_{E/F}(\alpha) \in F$.*

Proof: Let K/F be the Galois closure of E/F. Note that K is normal, so any $\sigma : E \to \bar{F}$ must map E to K, that is $\sigma : E \to K$.

Now for any $\tau \in \mathrm{Gal}(K/F)$, we have

$$\tau(N_{E/F}(\alpha)) = N_{E/F}(\alpha)$$

So $N_{E/F}(\alpha)$ is fixed by $\mathrm{Gal}(K/F)$, so $N_{E/F}(\alpha) \in F$. Similarly, we can show $Tr_{E/F}(\alpha) \in F$. □

Corollary 10.27.1 *We have a multiplicative group homomorphism*

$$N_{E/F} : E^* \to F^*.$$

We have an additive group homomorphism

$$Tr_{E/F} : E \to F.$$

Proof: By definition, $N_{E/F}(\alpha) = \prod_{\sigma \in S(E/F)} \sigma(\alpha) \neq 0$ if $\alpha \neq 0$. We also know $N_{E/F}(\alpha) \in F$. From definition, it is clear that $N_{E/F}(\alpha \cdot \beta) = N_{E/F}(\alpha) \cdot N_{E/F}(\beta)$. Hence $N_{E/F}$ is a group homomorphism from E^* to F^*. Similarly, we can prove the second statement. □

Lemma 10.28 *Let E'/E and E/F be finite and separable with $n = [E' : E]$. Then for $\alpha \in E$, we have*

$$N_{E'/F}(\alpha) = N_{E'/E}(N_{E/F}(\alpha)) = (N_{E/F}(\alpha))^n.$$

Proof: Recall Theorem (9.76), we have the bijection

$$S(E'/E) \times S(E/F) \leftrightarrow S(E'/F), \quad (\tau, \sigma) \mapsto (\hat{\sigma} \circ \tau).$$

By direct computation, we have

$$\begin{aligned} N_{E'/E}(N_{E'/E}(\alpha)) &= N_{E/F}\left(\prod_{\tau \in S(E'/E)} \tau(\alpha)\right) \\ &= \prod_{\sigma \in S(E/F)} \sigma\left(\prod_{\tau \in S(E'/E)} \tau(\alpha)\right) \\ &= \prod_{\sigma \in S(E/F)} \prod_{\tau \in S(E'/E)} \hat{\sigma} \circ \tau(\alpha) \\ &= \prod_{\phi \in S(E'/F)} \phi(\alpha). \\ &= N_{E'/F}(\alpha) \end{aligned}$$

Next, notice that since $\alpha \in E$, we have

$$\prod_{\tau \in S(E'/E)} \tau(\alpha) = \alpha^n.$$

Then we have

$$N_{E'/E}(N_{E'/E}(\alpha)) = \prod_{\sigma \in S(E/F)} \sigma(\alpha^n) = N_{E/F}(\alpha^n) = (N_{E/F}(\alpha))^n.$$

□

Lemma 10.29 *Let E'/E and E/F be finite and separable with $n = [E' : E]$. Then for $\alpha \in E$, we have*

$$Tr_{E'/F}(\alpha) = Tr_{E'/E}(Tr_{E/F}(\alpha)) = n \cdot Tr_{E/F}(\alpha).$$

Lemma 10.30 *Let E/F be finite and separable. Let $\alpha \in E$ with the minimal polynomial $m_\alpha(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$. Then*

$$N_{F(\alpha)/F}(\alpha) = (-1)^n a_0 = N^T_{F(\alpha)/F}(\alpha)$$

and

$$Tr_{F(\alpha)/F)}(\alpha) = -a_{n-1} = Tr^T_{F(\alpha)/F}(\alpha).$$

Proof: Recall $N_{F(\alpha)/F}(\alpha) = \prod_{\sigma S(F(\alpha)/F)}$, where

$$S(F(\alpha)/F) = \{\sigma : F(\alpha) \to \bar{F}\}.$$

Since σ is completely determined by the image of α which is in one-to-one correspondence to $m_\alpha(x)$ over F. Then $N_{F(\alpha)/F}(\alpha)$ is the production of the roots of the minimal polynomial $m_\alpha(x)$, hence is equal to $(-1)^n a_0$.

Next, we consider $T_\alpha : F(\alpha) \to F(\alpha)$, then $c_x(\alpha) = \det(xI - T_\alpha)$ is of degree $\deg m_\alpha(x) = [F(\alpha) : F]$. Then since α satisfies $c_x(\alpha)$ and $c_x(\alpha)$ is monic with the same degree as $m_\alpha(x)$. Then the two polynomials are in fact equal. Consequently, we also have $N^T_{F(\alpha)/F}(\alpha = (-1)^n a_0$.

Similarly, we can prove the result for the trace part. □

Theorem 10.31 *Let E/F be finite and separable. Let $\alpha \in E$.*

$$N_{E/F}(\alpha) = N^T_{E/F}(\alpha), \quad Tr_{E/F}(\alpha) = Tr^T_{E/F}(\alpha).$$

Proof: We consider $F \subset F(\alpha) \subset E$. Let $n = [E : F(\alpha)]$. Then we know

$$N_{E/F}(\alpha) = (N_{F(\alpha)/F}(\alpha))^n.$$

Now let $\{x_i\}$ be an $F(\alpha)$-basis of E and let $\{y_j\}$ be an F-basis of $F(\alpha)$. Then $\{x_i y_i\}$ is an F-basis of E. Consider $T_\alpha : E \to E$, where we consider $E = x_1F(\alpha) \oplus x_2 \oplus F(\alpha) \oplus \cdots \oplus x_nF(\alpha)$. So as matrices, we have T_α is the diagonal block matrix whose every block is $T'_\alpha : F(\alpha) \to F(\alpha)$ and there are precisely $x_nF(\alpha)$ of them. Then by property of determinant, we have

$$N^T_{E/F}(\alpha) = \det(T'_\alpha)^n = (N_{F(\alpha)/F}(\alpha))^n.$$

Similarly, we can prove the for the trace part. □

10.3 The Fundamental Theorem of Galois Theory

Proposition 10.32 *Let E/F be Galois. Let $F \subset K \subset E$ be such that K/F is normal (equivalent to say that K/F is Galois). Then we have a short exact sequence of group*

$$1 \to \mathrm{Aut}(E/K) \to \mathrm{Aut}(E/F) \to \mathrm{Aut}(K/F) \to 1.$$

In particular, $\mathrm{Aut}(E/K)$ *is a normal subgroup of* $\mathrm{Aut}(E/F)$.

Proof: We define a homomorphism $\varphi : \mathrm{Aut}(E/F) \to \mathrm{Aut}(K/F)$ by restriction. That is for any

$$\sigma : E \to E \longmapsto \sigma : K \to K.$$

This is well-defined since K/F is a normal extension. We show $\mathrm{Aut}(E/F) \to \mathrm{Aut}(K/F)$ is surjective. Let $\sigma \in \mathrm{Aut}(K/F)$, we can extend $\sigma : E \to E$ by first extending to $\sigma : \bar{F} \to \bar{F}$, then consider the restriction.

The injection $\pi : \mathrm{Aut}(E/K) \to \mathrm{Aut}(E/F)$ is clearly injective. Now we show that the image of π is the kernel of φ. This is clear since any automorphism on E fixing K gets restricted to the identity automorphism on K fixing F. Moreover, if any automorphism on E fixing F gets restricted to the identity automorphism on K, then it must fix K originally. Hence $\mathrm{Aut}(E/K) \to \mathrm{Aut}(E/F) \to \mathrm{Aut}(K/F)$ is exact. □

Proposition 10.33 *Let E/F be Galois. Let $F \subset K \subset E$. Then K/F is normal if and only if* $\mathrm{Aut}(E/K) \hookrightarrow \mathrm{Aut}(E/F)$ *is normal.*

Proof: We have proven the forward direction, since if K/F is normal, then it is a kernel of a group homomorphism, hence normal.

On the other hand, we show that if K/F is not normal, then $\mathrm{Aut}(E/K) \subset \mathrm{Aut}(E/F)$ is not normal. We first define $\sigma \in \mathrm{Aut}(E/F)$ as follows: since K is not normal, then we can find $\sigma \in \mathrm{Aut}(E/F)$ such that $\sigma(\alpha) \notin K$ for some $\alpha \in K$ (otherwise K/F would be normal). We then find $\tau \in \mathrm{Aut}(E/K)$, let $m_{\sigma(\alpha),K}$ be the minimal polynomial of $\sigma(\alpha) \in E$ over K. Since $\sigma(\alpha) \notin E$, then $\deg(m_{\sigma(\alpha),k}) \geq 2$. Now since $\sigma(\alpha)$ is separable over F, hence K. We have another root β of $m_{\sigma(x),K}$. Then we have

$$\tau : K(\sigma(\alpha)) \to K(\beta), \quad \sigma(\alpha) \mapsto \beta.$$

Then we extend τ to $E \to E$, then $\tau \in \mathrm{Aut}(E/K)$.

We show that $\sigma^{-1}\tau\sigma \notin \mathrm{Aut}(E/K)$. Notice that

$$\sigma^{-1}\tau\sigma(\alpha) = \sigma^{-1}\tau(\sigma(\alpha)) = \sigma^{-1}(\beta) \neq \alpha$$

So $\sigma^{-1}\tau\sigma$ does not fix K. Hence $\mathrm{Aut}(E/K) \subset \mathrm{Aut}(E/F)$ is not normal. □

Definition 10.34 (Cyclic Extension) *Let E/F be Galois, then E/F is called* ***cyclic*** *(resp. abelian) extension, if* $\mathrm{Gal}(E/F)$ *is cyclic (resp. abelian).*

Corollary 10.34.1 *Let E/F be an cyclic (resp. abelian) Galois extension. Then for any $F \subset K \subset E$, K/F is a cyclic (resp. abelian) Galois extensions.*

Definition 10.35 (Conjugates) *Let K/F be a Galois extension. If $\alpha \in K$ the elements $\sigma\alpha$ for σ in $\mathrm{Gal}(K/F)$ are called the **Galois Conjugates** of α over F. If E is a subfield of K containing F, the field $\sigma(E)$ is called the **conjugate field** of E over F.*

To summarize what we have done, we have the following Fundamental Theorem of Galois Theory.

Theorem 10.36 (Fundamental Theorem of Galois Theory) *Let K/F be a finite Galois extension and set $G = \mathrm{Gal}(K/F)$. Then there is a bijection*

$$\left\{ \begin{array}{cc} \text{subfields } E & K \\ \text{of } K & \vert \\ \text{containing } F & E \\ & \vert \\ & F \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{cc} \text{subgroups } H & 1 \\ \text{of } G & \vert \\ & H \\ & \vert \\ & G \end{array} \right\}$$

given by the correspondences

$$E \longrightarrow \left\{ \begin{array}{c} \text{the elements of } G \\ \text{fixing } E \end{array} \right\}$$

$$\left\{ \begin{array}{c} \text{the fixed field} \\ \text{of } H \end{array} \right\} \longleftarrow H$$

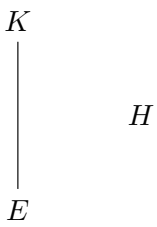
which are inverse to each other. Under this correspondence,

(1) (order reversing) If E_1, E_2 correspond to H_1, H_2, respectively, then $E_1 \subseteq E_2$ if and only if $H_2 \subseteq H_1$.

(2) $[K : E] = |H|$ and $[E : F] = |G : H|$, the index of H in G:

$$\begin{array}{ccc} K & & \\ \vert & \} & |H| \\ E & & \\ \vert & \} & |G : H| \\ F & & \end{array}$$

(3) K/E is always Galois, with Galois group $\mathrm{Gal}(K/E) = H$:



(4) E is Galois over F if and only if H is a normal subgroup in G. If this is the case, then the Galois group is isomorphic to the quotient group

$$\mathrm{Gal}(E/F) \cong G/H.$$

More generally, even if H is not necessarily normal in G, the isomorphisms of E (into a fixed algebraic closure of F containing K) which fix F are in one to one correspondence with the cosets $\{\sigma H\}$ of H in G.

(5) If E_1, E_2 correspond to H_1, H_2, respectively, then the intersection $E_1 \cap E_2$ corresponds to the group $\langle H_1, H_2\rangle$ generated by H_1 and H_2 and the composite field E_1E_2 corresponds to the intersection $H_1 \cap H_2$. Hence the lattice of subfields of K containing F and the lattice of the subgroups of G are "dual" (the lattice diagram of one is the lattice diagram for the other turned upside down).

Proof: The map $E \subset K \mapsto \mathrm{Aut}(K/E) \subset \mathrm{Aut}(K/F)$ is clearly injective. Since if $\mathrm{Aut}(K/E) = \mathrm{Aut}(K/E')$, as both extensions are Galois, then we must have $E = K^{\mathrm{Aut}(K/E)} = K^{\mathrm{Aut}(K/E')} = E'$. Now we see that this map is a bijection by a Cardinality argument, since by Corollary (10.22.3), the map $H \subset \mathrm{Aut}(K/F) \mapsto K^H$ is injective. We have already established order reversing inclusion.

If $E = K^H$, then K/E is Galois with $[K : E] = \mathrm{Gal}(K/E) = |H|$. Then by multiplicativity of degree of field extension, we have $[E : F] = [G : H]$.

We have also shown that E/F is Galois if and only if H is normal in G. Then from the short exact sequence, we conclude that $\mathrm{Gal}(E/F) \cong G/H$. The second part of (4) follows from the fact that all once we drop the restriction that E maps to E, then we can extend any automorphism from K to K to $\bar{F}$ to $\bar{F}$, then restrict to E.

Lastly, if $E_1 \cap E_2$ is the largest subfield of K contained in both E_1 and E_2, and $\langle H_1, H_2\rangle$ is the smallest subgroup of $\mathrm{Aut}(K/F)$ that fixes E_1 and E_2, hence by order reversing inclusion, we must have $E_1 \cap E_2 = K^{\langle H_1, H_2\rangle}$. Similarly, we have $E_1E_2 = K^{H_1 \cap H_2}$. □

Remark 10.36.1 *If K/F is not a finite Galois extension, this bijection fails. $E \subset K \mapsto \mathrm{Aut}(K/E)$ is always injective (can explicitly construct). However, the composition $H \mapsto K^H \to \mathrm{Aut}(K/K^H)$ may not be H any more; we can only conclude that it is a subgroup of $\mathrm{Aut}(K/F)$ containing H.*

Corollary 10.36.1 *If K/F is a finite Galois extension, then the number of subfield of K/F is finite.*

Corollary 10.36.2 *If K is a Galois extension of F of degree p^n for some prime p, and some $n \geq 1$. Then there are Galois extensions of F contained in K of degrees p and p^{n-1}.*

Proof: Since for any m, $0 \leq m \leq n$, there exists a normal subgroup of order p^m within $\mathrm{Gal}(K/F)$. $\square$

Example:

1. We consider $\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}$. Let $G = \{1, \sigma, \tau, \sigma\tau\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, where $\sigma(\sqrt{2}) = -\sqrt{2}$, $\tau(\sqrt{3}) = -\sqrt{3}$. Then
   - $\mathbb{Q}(\sqrt{2},\sqrt{3})^{1,\sigma} = \mathbb{Q}(\sqrt{3})$;
   - $\mathbb{Q}(\sqrt{2},\sqrt{3})^{1,\tau} = \mathbb{Q}(\sqrt{2})$.
   - $\mathbb{Q}(\sqrt{2},\sqrt{3})^{1,\sigma\tau} = \mathbb{Q}(\sqrt{6})$.
2. Let p be a prime, and let E be the splitting field of $x^p - 2$ over $\mathbb{Q}$. Then we can check that $E = \mathbb{Q}(\sqrt[p]{2}, \xi_p)$ and $[E : Q] = p(p-1)$. Now by Cauchy's Theorem, $\mathrm{Gal}(E/\mathbb{Q})$ has an element of order p, which implies that $\sigma(\sqrt[p]{2}) = \sqrt[p]{2}\xi_p^k$ for $k = 0, 1, \cdots, p-1$ (permuting the roots) have to define an automorphism. On the other hand, $\mathrm{Gal}(E/\mathbb{Q}(\sqrt[p]{2}))$ is generated by $\tau(\xi_p) \mapsto \xi_p$. Then $\mathrm{Gal}(E/\mathbb{Q}(\sqrt[p]{2}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$. Clearly, we have that the two groups are disjoint, hence we conclude that

   $$\mathrm{Gal}(E/\mathbb{Q}) \cong \mathbb{Z}/p\mathbb{Z} \rtimes (\mathbb{Z}/p\mathbb{Z})^\times.$$

   In fact, we have that this group is isomorphic to the group

   $$\left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mid \mid a, b \in \mathbb{F}_p,\ a \neq 0 \right\}.$$

Proposition 10.37 *Let E/F be a Galois extension and let K be an intermediate field. Let H be the subgroup of $\mathrm{Gal}(E/F)$ mapping K into iteself. Then H is the normalize of $\mathrm{Gal}(E/K)$ in $\mathrm{Gal}(E/F)$.*

Proof: We show two directions, namely $H \subset N(\mathrm{Gal}(E/K))$ and $H \supset N(\mathrm{Gal}(E/K))$. WLOG, we assume that the extension E/K is nontrivial, otherwise we see $\mathrm{Gal}(E/K)$ is trivial and H is just $\mathrm{Gal}(E/F)$, hence the statement clearly follows.

$H \subset N(\mathrm{Gal}(E/K))$: let $\sigma \in H$ and $\tau \in \mathrm{Gal}(E/K)$, we show $\sigma^{-1}\tau\sigma \in \mathrm{Gal}(E/K)$, so $\sigma \in N(\mathrm{Gal}(E/K))$. Since σ and σ^{-1} are both in H (H is a group), then they both map K onto K. $\tau : E \to E$ is a K-automorphism, which fix K by definition. Hence for any $k \in K$, we have

$$(\sigma^{-1} \circ \tau \circ \sigma)(k) = (\sigma^{-1}(\tau(\sigma(k)))) = \sigma^{-1}\sigma(k) = k$$

Also composition of field automorphisms is a field automorphism. Hence $\sigma^{-1}\tau\sigma$ is a field automorphism from E to E that fixes K, hence belongs to $\mathrm{Gal}(E/K)$. This shows $H \subset N(\mathrm{Gal}(E/K))$.

$N(\mathrm{Gal}(E/K)) \subset H$: suppose $\sigma \in \mathrm{Gal}(E/F)$ is a normalizer of $N(\mathrm{Gal}(E/K))$, we show $\sigma \in H$. Let $k \in K$ be arbitrary, then suppose $\sigma(k) = \alpha \notin K$. We consider the minimal polynomial $m_{\alpha,K}(x)$ of α over K, which is separable and splits in E since E/K is Galois. $\alpha \notin K$, then $m_{\alpha,K}(x)$ is degree at least 2, so let β be another root of $m_{\alpha,k}(x)$. Then we have an K-field isomorphism $\tau' : K(\alpha) \to K(\beta)$, induced by $\alpha \mapsto \beta$, since

$$K(\alpha) \cong K[x]/m_{\alpha,k}(x) \cong K(\beta).$$

Now τ' can be seen as an K-homomorphism from $K(\alpha)$ into $\bar{E}$, then we can extend τ' to an K-homomorphism from $\bar{E} \to \bar{E}$. Now taking restriction, since E/K is Galois (normal in particular), then $\tau'|_E$ if an K-homomorphism from E to E. Let $\tau'|_E = \tau$. Then by problem 3, τ is an isomorphism. Hence $\tau \in \text{Gal}(E/K)$.

Now since $\sigma \in N(\text{Gal}(E/K)$, then we must have

$$(\sigma^{-1} \circ \tau \circ \sigma)(k) = k.$$

However, if we compute directly, $\sigma(k) = \alpha$, $\tau(\alpha) = \beta$, and $\sigma^{-1}(\beta) \neq k$, since $\sigma(k) = \alpha \neq \beta$. So we have a contradiction. This implies that $\sigma(k) \in K$ for all $k \in K$, i.e., $\sigma \in H$.
□

10.4 Order of Finite Field

Definition 10.38 (Order) *Let F be a finite field, we define the **order of the field** to be the number of elements of F.*

Theorem 10.39 *Any finite field with characteristic p has p^n elements for some positive integer n. That is the order of a finite field must be of the form p^n.*

Proof: Let F be a finite field, and let K be its subfield. Then the vector space of F over K (The K-vector space) is of some finite dimension, say n, and there exists a basis $\alpha_1, \alpha_2, \cdots, \alpha_n$ of F over K. Since every element of F can be uniquely expressed as a linear combination of the basis element α_i's over K, i.e., for every $a \in F$,

$$a = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n,$$

where $c_1, \cdots, c_n \in K$. Now if we let K be the prime subfield of F, it has p elements. Thus F must have p^n elements. □

Corollary 10.39.1 *Suppose K is a subfield of F with q elements, then F has q^m elements, where $m = [F : K]$.*

Proof: It is clear that K has to be the prime subfield. □

Theorem 10.40 *Let p be any prime, and n be a natural number. Then there is a finite field of order p^n and it is unique up to isomorphism, which will be denoted by $\mathbb{F}_{p^n}$.*

Proof: Consider the splitting field F of the polynomial $x^{p^n} - x$ over $\mathbb{F}_p$. This polynomial has no multiple roots by the derivative test, since the derivative is $p^n x^{p^n-1} - 1 = -1 \neq 0$. Let $F' \subset F$ be the subset of all roots of $f(x)$, by direct verification, one can show that F' is a field, then $F' = F$.

Now let K be a finite field such that $|k| = p^n$. Then $K^\times \subset K$ is of order $p^n - 1$. So by Lagrange's theorem $x^{p^n-1} = 1$ for all $x \neq 0$, that is $x^{p^n} - x = 0$ for all $x = 0$. Then $F \subset K$ where F is the splitting field of $x^{p^n} - x$. Since $|F| = p^n$ and by the uniqueness of splitting field, we conclude the uniqueness of field of order p^n. □

Corollary 10.40.1 *The field $\mathbb{F}_{p^n}$ is Galois over $\mathbb{F}_p$, with cyclic Galois group of order n generated by the Frobenius automorphism*

$$\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p) = \langle \sigma_p \rangle \cong \mathbb{Z}/n\mathbb{Z}$$

where

$$\begin{aligned} \sigma_p : \mathbb{F}_{p^n} &\longrightarrow \mathbb{F}_{p^n} \\ \alpha &\longmapsto \alpha^p \end{aligned}$$

Proof: Clearly $\mathbb{F}_{p^n}$ is normal and separable, hence Galois. Then we know $|\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F})| = n$. We also know that the Frobenius homomorphism are automorphism in this case. Hence by a cardinality argument, we conclude that all the automorphism are given in this form. Moreover, this Galois group is cyclic. □

Proposition 10.41 *The subfields of $\mathbb{F}_{p^n}$ are all Galois over $\mathbb{F}_p$ and are in one to one correspondence with the divisors m of n. They are fields $\mathbb{F}_{p^m}$, the fixed fields of σ_p^m. That is, every subfield of $\mathbb{F}_{p^m}$ has order p^m, where m is a positive divisor of n. Conversely, if m is a divisor of n, then there exist exactly one subfield of $\mathbb{F}_{p^n}$ with p^m elements.*

Proof: Follows from the Fundamental Theorem of Galois Theory and preceeding results. □

If $\mathbb{F}_{p^{n_1}}$ and $\mathbb{F}_{p^{n_2}}$ are two finite fields, then there is a third finite field containing (an isomorphic copy of) them, namely $\mathbb{F}_{p^{n_1 n_2}}$. This gives us a partial ordering on these fields and allows us to think of their union. Since these give all the finite extensions of $\mathbb{F}_p$, we see that the union of $\mathbb{F}_{p^n}$ for all n is an algebraic closure of $\mathbb{F}_p$, unique up to isomorphism:

$$\overline{\mathbb{F}_p} = \bigcup_{n=1} \mathbb{F}_{p^n}.$$

Corollary 10.41.1 *The irreducible polynomial $x^4 + 1 \in \mathbb{Z}[x]$ is reducible modulo every prime p.*

Proof: Consider the polynomial $x^4 + 1$ over $\mathbb{F}_p[x]$ for some prime p. If $p = 2$, we have $x^4 + 1 = (x + 1)^4$ and the polynomial is reducible. Assume now that p is odd. Then $p^2 - 1$ is divisible by 8, hence $x^{p^2-1} - 1$ is divisible by $x^8 - 1$. Then we have the divisibilities

$$x^4 + 1 | x^8 - 1 | x^{p^2-1} - 1 | x^{p^2} - x$$

which shows that all the roots of $x^4 + 1$ are roots of $x^{p^2} - x$. Since the roots of $x^{p^2} - x$ are the fields $\mathbb{F}_{p^2}$, it follows that the extension generated by any root of $x^4 + 1$ is at most degree 2 over $\mathbb{F}_p$, which means that $x^4 + 1$ cannot be irreducible over $\mathbb{F}_p$. □

Proposition 10.42 *The finite field $\mathbb{F}_{p^n}$ is simple (a simple extension of $\mathbb{F}_p$). In particular, there exists an irreducible polynomial of degree n over $\mathbb{F}_p$ for every $n \geq 1$.*

Proof: The multiplicative group $\mathbb{F}_{p^n}^{\times}$ is a finite subgroup of the multiplicative group of a field. This is a cyclic group; if θ is any generator, then $\mathbb{F}_{p^n} = \mathbb{F}_p(\theta)$. □

Just like Cyclotomic polynomials, if we group together the factors $x - a$ of the polynomial $x^{p^n} - x$ according to the degree d of their minimal polynomials over $\mathbb{F}_p$ we obtain the following proposition.

Proposition 10.43 *The polynomial $x^{p^n} - x$ is precisely the product of all the distinct irreducible polynomials in $\mathbb{F}_p[x]$ of degree d where d runs through all divisors of n.*

Remark 10.43.1 *This proposition can be used to produce irreducible polynomials over $\mathbb{F}_p$ recursively.*

Remark 10.43.2 *Since the finite field $\mathbb{F}_{p^n}$ is unique up to isomorphism, the quotients of $\mathbb{F}_p[x]$ by any of the irreducible polynomials of degree n are all isomorphic, as we get two fields of the same order. So if $f_1(x)$ and $f_2(x)$ are irreducible of degree n, then $f_2(x)$ splits completely in the field $\mathbb{F}_{p^n} \cong \mathbb{F}_p[x]/(f_1(x))$. If we denote a root of $f_2(x)$ by α, then the isomorphism is given by*

$$\begin{aligned} \mathbb{F}_p[x]/(f_2(x)) &\cong \mathbb{F}_p[x]/(f_1(x)) \\ x &\mapsto \alpha \end{aligned}$$

The above Proposition also gives a way to produce irreducible polynomials over $\mathbb{F}_p$. For example, the irreducible quadratics over $\mathbb{F}_2$ are the divisors of

$$\frac{x^4 - x}{x(x-1)}$$

which can only be $x^2 + x + 1$. Similarly, the irreducible cubics over the field are the divisors of

$$\frac{x^8 - x}{x(x-1)} = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$$

which factors into $x^3 + x + 1$ and $x^3 + x^2 + 1$.

If we assume a result from elementary number theory, we can give a formula for the number of irreducible polynomials of degree n.

Definition 10.44 (Möbius μ-function) *We define the **Möbius μ-function** by*

$$\mu(n) = \begin{cases} 1 & n = 1 \\ 0 & \text{if } n \text{ is a square} \\ (-1)^r & \text{if } n \text{ has } r \text{ distinct prime factors.} \end{cases}$$

If now $f(n)$ is a function defined for all nonnegative integers n and $F(n)$ is defined by

$$F(n) = \sum_{d|n} f(d) \quad n = 1, 2, \cdots .$$

Then the Möbius inversionformula states that one can recover the function $f(n)$ from $F(n)$:

$$f(n) = \sum_{d|n} \mu(d) F(\frac{n}{d}), \quad n = 1, 2, \cdots .$$

Define $\psi(n)$ to be the number of irreducible polynomials of degree n in $\mathbb{F}_p[x]$. Then clearly

$$p^n = \sum_{d|n} d\psi(d).$$

Applying the Möbius inversion formula for $f(n) = n\psi(n)$, we obtain

$$n\psi(n) = \sum_{d|n} \mu(d) p^{n/d}$$

which gives us a formula for the number of irreducible polynomials of degree n over $\mathbb{F}_p$:

$$\psi(n) = \frac{1}{n} \sum_{d|n} \mu(d) p^{n/d}.$$

Corollary 10.44.1 *An algebraically closed field is infinite.*

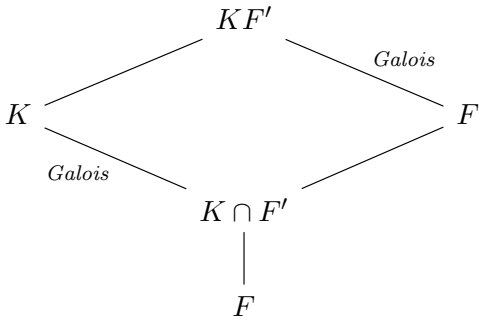
Proof: If the prime field is infinite, then we done. Otherwise, the prime field is $\mathbb{F}_p$ for some prime p. Since $\mathbb{F}_p$ has infinitely many irreducible polynomials by the above formula, we conclude that the algebraic closure of $\mathbb{F}_p$ is infinite. Hence the algebraic closed field which contains the algebraic closure is infinite. □

10.5 Composite Extension and Simple Extensions

Proposition 10.45 *Suppose K/F is a Galois extension and F'/F is any extension we know KF'/F' is a Galois extensions. Moreover, if K/F is finite, then Galois group of KF'/F' (finite) is given by*

$$\mathrm{Gal}(KF'/F') \cong \mathrm{Gal}(K/K \cap F')$$

isomorphic to a subgroup of $\mathrm{Gal}(K/F)$. *Pictorially,*



Proof: Since K/F is Galois, so the map

$$\varphi : \mathrm{Gal}(KF'/F') \longrightarrow \mathrm{Gal}(K/K \cap F')$$

$$\sigma \longmapsto \sigma|_K$$

is well defined. This is because K/F is normal, so any $\sigma \in \mathrm{Gal}(KF'/F')$ will map K to K and clearly fixes F', hence fixes F', $\sigma|_K$ fixes $K \cap F'$.

We show this map is injective. Let $\sigma \in \ker \tau$. Then $\tau|_K$ is the identity map. So $\sigma|_K : K \to K$ is the identity map. We also know σ fixes F', hence σ fixes both K and F' hence KF'. This shows σ is trivial.

We show this map is surjective. Let H be the image of the φ. We want to show $H = \mathrm{Gal}(K/K \cap F')$. We do this by showing that $K^H = K \cap F'$, then by finiteness of the Galois extension, we get $H = \mathrm{Gal}(K/K \cap F')$. Clearly $K \cap F' \subset K^H$. For the other direction, suppose there is some $\alpha \in K^H \setminus K \cap F' \subset K - K \cap F' = K \setminus F'$ that is fixed by H. Then $\alpha \in KF' - F'$ is fixed by H, which is a contradiction since α is not fixed by $\mathrm{Gal}(KF'/F')$ (α is not in F', and $\mathrm{Gal}(KF'/F')$ is a subgroup of H). □

Corollary 10.45.1 *Suppose K/F is a finite Galois Extensions and F'/F is any finite extensions. Then*

$$[KF' : F] = \frac{[K : F][F' : F]}{[K \cap F' : F]}.$$

Proof: This follows from the fact that $[KF' : F'] = [K : K \cap F']$, since the Galois group are isomorphic. So

$$[KF' : F] = [KF' : F'][F' : F] = [K : K \cap F'][F' : F] = \frac{[K : F]}{[K \cap F' : F]}[K : K \cap F'].$$

□

Example: The formula

$$[KF' : F] = \frac{[K : F][F' : F]}{[K \cap F' : F]}$$

fails if K/F is not Galois. Take $F = \mathbb{Q}$, and consider $K = \mathbb{Q}(\sqrt[3]{2})$ and $F' = \mathbb{Q}(\xi_3\sqrt[3]{2})$. Then $[KF' : \mathbb{Q}] = 6$, but

$$\frac{[K : F][F' : F]}{[K \cap F' : F]} = 9.$$

This shows neither K/F and F'/F is Galois.

Lemma 10.46 *Let G_1, G_2, G_3 be three finite groups, such that*

$$\begin{array}{ccc} & & G_2 \\ & & \downarrow \\ G_1 & \xrightarrow[surjective]{} & G_3 \end{array}$$

Then

$$G_1 \times_{G_3} G_2 = \frac{|G_1||G_2|}{|G_3|},$$

where

$$G_1 \times_{G_3} G_2 = \{(q_1, q_2) \in G_1 \times G_2 \,|\, \pi_2(g_1) = \pi_2(g_2)\}.$$

Proof: For any fixed $g_2 \in G_1$, we want to know the number of $g_1 \in G_2$ such that $\pi_2(g_2) = \pi_1(g_1)$, i.e., the number of $\{\pi_1^{-1}\pi_2(g_2)\}$. Since π_1 is surjective, then we have

$$\pi_1^{-1}\pi_2(g_2) = g_1 \cdot \ker \pi_1.$$

Then by first isomorphism theorem, we have $\#\pi_2^{-1}\pi_1(g_1) = \frac{|G_2|}{|G_3|}$, so the formula follows. □

Proposition 10.47 *Let K_1 and K_2 be finite Galois extensions of a field F with Galois group G_1 and G_2 respectively. Then*

1. *The intersection $K_1 \cap K_2$ is Galois over F. We have the diagram:*

$$\begin{array}{ccc} & & G_2 \\ & & \downarrow{\scriptstyle \pi_2} \\ G_1 & \xrightarrow{\pi_1} & \mathrm{Aut}(K_1 \cap K_2/F) \end{array}$$

   In particular, the image of π_1, π_2 are surjective.

2. *The composite K_1K_2 is Galois over F. The Galois group is isomorphic to the subgroup*

$$H = \{(\sigma, \tau) \,:\, \sigma|_{K_1 \cap K_2} = \tau|_{K_1 \cap K_2}\}$$

   of the direct product $\mathrm{Gal}(K_1/F) \times \mathrm{Gal}(K_2/F)$ consisting of elements whose restriction to the intersection $K_1 \cap K_2$ are equal. I.e., it is the fiber product

$$G_1 \times_{\mathrm{Aut}(K_1 \cap K_2/F)} G_2 = \{(g_1, g_2) \in G_1 \times G_2 \,|\, \pi_1(g_1) = \pi_2(g_2)\}.$$

   We have the following diagram:

$$\begin{array}{ccc} G & \longrightarrow & G_2 \\ \downarrow & & \downarrow{\scriptstyle \pi_2} \\ G_1 & \xrightarrow{\pi_1} & \mathrm{Aut}(K_1 \cap K_2/F) \end{array}$$

Proof: The first part is clear, for the second part, we know K_1K_2/F is Galois. The map

$$\begin{aligned} \varphi : \mathrm{Gal}(K_1K_2/F) &\longrightarrow \mathrm{Gal}(K_1/F) \times \mathrm{Gal}(K_2/F) \\ \sigma &\longmapsto (\sigma|_{K_1}, \sigma|_{K_2}) \end{aligned}$$

is clearly a homomorphism. The kernel consists of the elements σ which are trivial on both K_1 and K_2, hence trivial on the composite, so the map is injective. The image lies in the subgroup H, since

$$(\sigma|_{K_1})|_{K_1\cap K_2} = \sigma|_{K_1\cap K_2} = (\sigma|_{K_2})|_{K_1\cap K_2}.$$

The order of H can be computed by observing that for every $\sigma \in \mathrm{Gal}(K_1/F)$, there are $|\mathrm{Gal}(K_2/K_1\cap K_2)|$ elements $\tau \in \mathrm{Gal}(K_2/F)$ whose restrictions to $K_1 \cap K_2$ are $\sigma|_{K_1\cap K_2}$. Hence

$$\begin{aligned} |H| &= |\mathrm{Gal}(K_1/F)| \cdot |\mathrm{Gal}(K_2/K_1 \cap K_2)| \\ &= |\mathrm{Gal}(K_1/F)| \frac{|\mathrm{Gal}(K_2/F)|}{|\mathrm{Gal}(K_1 \cap K_2/F)|}. \end{aligned}$$

Then the order of H and $\mathrm{Gal}(K_1K_2/F)$ both equal to

$$\frac{[K_1 : F][K_2 : F]}{[K_1 \cap K_2 : F]}.$$

Hence the image of φ is precisely H, completing the proof. □

Corollary 10.47.1 *Let K_1 and K_2 be finite Galois extensions of a field F with $K_1 \cap K_2 = F$. Then*

$$\mathrm{Gal}(K_1K_2/F) \cong \mathrm{Gal}(K_1/F) \times \mathrm{Gal}(K_2/F).$$

Conversely, if K is finite Galois over F and $G = \mathrm{Gal}(K/F) = G_1 \times G_2$ is the direct product of two subgroups G_1 and G_2, then K is the composite of two Galois extensions $K_1 = K^{G_1}$ and $K_2 = K^{G_2}$ of F with $K_1 \cap K_2 = F$.

Proof: The first part is clear. For the second part, let K_1 be the fixed field of G_1 and let K_2 be the fixed field of G_2. Then $K_1 \cap K_2$ is the field corresponding to the subgroup G_1G_2, which is all of G in this case, so $K_1 \cap K_2 = F$. The composite K_1K_2 is the field corresponding to the subgroup $G_1 \cap G_2$, which is the identity here, so $K_1K_2 = K$, completing the proof. □

Corollary 10.47.2 *Let E/F be any finite separable extension. Then E is contained in an extension K which is Galois over F and is minimal in the sense that in a fixed algebraic closure of K any other Galois extension of F containing E contains K.*

Proof: There exists a Galois extension of F containing E, for example the composite of the splitting field of the minimal polynomial for a basis for E over F. Then the intersection of all the Galois extensions of F containing E is the field K. □

Remark 10.47.1 *The Galois Extension K of F in this corollary is called the **Galois Closure** of E over F.*

Corollary 10.47.3 *The composite field of K_1 and K_2 is an abelian Galois extension over F if K_1/F, K_2/F are abelian Galois extensions.*

Proposition 10.48 *Let K/F be a finite extension. Then $K = F(\theta)$ if and only if there exists only finitely many subfield of K containing F.*

Proof: Suppose first that $K = F(\theta)$ is simple. Let E be a subfield of K containing F. Let $f(x) \in F[x]$ be the minimal polynomial for θ over F and let $g(x) \in E[x]$ be the minimal polynomial for θ over E. Then $g(x)$ divides $f(x)$ in $E[x]$. Let E' be the field generated over F by the coefficients of $g(x)$. Then $E' \subseteq E$ and clearly the minimal polynomial for θ over E' is still $g(x)$. But then

$$[K : E] = \deg g(x) = [K : E']$$

implies that $E = E'$. It follows that the subfields of K containing F are the subfields generated by the coefficients of the monic factors of $f(x)$, hence there are finitely many such fields.

Suppose conversely that there are finitely many subfields of K containing F. If F is a finite field, then we have already seen that K is a simple extensions (multiplicative group is cyclic). Hence we may suppose F is infinite. It clearly sufficies to show that $F(\alpha, \beta)$ is generated by a single element since K is finitely generated over F. Consider the subfields

$$F(\alpha + c\beta), \quad c \in F.$$

Then since there are infinitely many choices for $c \in F$ and only finitely many such subfields, there exist c, c' in F, $c \neq c'$, with

$$F(\alpha + c\beta) = F(\alpha + c'\beta).$$

Then $\alpha + c\beta$ and $\alpha + c'\beta$ both lie in $F(\alpha + c\beta)$, so $\beta \in F(\alpha + c\beta)$, so $\alpha \in F(\alpha + c\beta)$. Therefore $F(\alpha, \beta) \subseteq F(\alpha + c\beta)$, so

$$F(\alpha, \beta) = F(\alpha + c\beta)$$

completing the proof. □

Remark 10.48.1 *As the proof of the proposition indicates, a primitive element for an extension can be obtained as a simple linear combination of the generators for the extension. In the case of Galois extensions it is only necessary to determine a linear combination which is not fixed by any nontrivial element of the Galois group since then by the Fundamental Theorem this linear combination could not lie in any proper subfield.*

Corollary 10.48.1 *Suppose K/F is a finite Galois extension. Then K is simple, i.e., $K = F(\theta)$.*

Proof: Since we have only finitely many intermediate fields. □

Corollary 10.48.2 *Let K/F be a finite Galois extension with Galois group G, then we know $K = F(\alpha)$ for some $\alpha \in K$. Let H be the subgroup of G, and $E = K^H$. We consider the polynomial*

$$f(x) = \prod_{g \in H} (x - g(\alpha)) = \sum_{i=0}^{n} a_i x^i \in E[x].$$

Then $E = F(\alpha_1, \cdots, \alpha_n)$.

Proof: Firstly, we note this polynomial f is fixed by H, hence its coefficients must lie in E. Next, $m_{\alpha,E}(x)$ divides f, hence by a degree argument, we conclude that $E = F(\alpha_1, \cdots, \alpha_n)$. □

Theorem 10.49 (The Primitive Element Theorem) *If K/F is finite and separable, then K/F is simple. In particular, any finite extension of fields of charactersitic 0 is simple.*

Proof: Proof 1: Let L be the Galois closure of K over F. Then any subfield of K containing F corresponds to a subgroup of the Galois group $\mathrm{Gal}(L/F)$ by the Fundamental Theorem. Since there are only finitely many such subgroups, the Proposition 10.48 shows that K/F is simple. The last statement follows since any finite extension of fields in characteristic 0 is separable.

Proof 2: If K is a finite field, then it must be simple. So we assume K is infinite. WLOG, we assume $K = F(\alpha, \beta)$. Let $\sigma_1, \cdots, \sigma_n$ be the distinct F-homomorphism from E to $\bar{F}$. We consider the polynomial

$$f(x) = \prod_{i \neq j} ((\sigma_i - \sigma_j)(\alpha) + (\sigma_i - \sigma_j)(\beta)x) \text{ in } \bar{F}(x).$$

Note that $(\sigma_i - \sigma_j)(\alpha)$ and $(\sigma_i - \sigma_j)(\beta)$ cannot both be zero, as otherwise we must have $\sigma_i = \sigma_j$ contradiction. So $f(x)$ is not a zero polynomial, hence not a zero function, which means that we can find $c \in F$ such that $f(c) \neq 0$ (F is infinite). Since $c \in F$, then $\sigma_i(c) = c$. Then

$$f(c) = \prod_{i \neq j} (\sigma_i - \sigma_j)(\alpha + c\beta) \neq 0.$$

So $\sigma_i(\alpha + c\beta) \neq \sigma_j(\alpha + c\beta)$ for any $i \neq j$. Now let $m(x)$ be the minimal polynomial of $\alpha + c\beta$ over F. Then σ_i permutes the roots of $m(x)$, since the coefficients of $m(x)$ is in F which is fixed by σ_i. Then $m(x)$ has at least n distinct roots, which implies

$$n \leq [F(\alpha + c\beta) : F]_S \leq [F(\alpha + c\beta) : F] \leq [F(\alpha, \beta) : F] = n.$$

So $\alpha + c\beta$ generates $F(\alpha, \beta)$. □

Corollary 10.49.1 *Let K/F be a finite Galois extension, then K is generated by α, where α is not fixed by any nontrivial elements of* $\mathrm{Gal}(K/F)$.

Proof: Suppose $F(\alpha)$ is a subfield of K, then α is fixed by some subgroup of $\mathrm{Gal}(K/F)$, which is a contradiction. □

10.6 Cyclotomic Extensions and Abelian Extensions Over $\mathbb{Q}$

Recall $(\mathbb{Z}/n\mathbb{Z})^* \cong \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$, where the Automorphism is understood as ring isomorphism.

Theorem 10.50 *The Galois group of the cyclotomic field $\mathbb{Q}(\xi_n)$ of n^{th} roots of unity is isomorphic to the multiplicative group $(\mathbb{Z}/n\mathbb{Z})^\times$. The isomorphism is given explicitly by the map*

$$\begin{aligned}(\mathbb{Z}/n\mathbb{Z})^* &\longrightarrow \mathrm{Gal}(\mathbb{Q}(\xi_n)/\mathbb{Q}) \\ (a \mod n) &\longmapsto \sigma_a\end{aligned}$$

where σ_a is the automorphism defined by

$$\sigma_a(\xi_n) = \xi_n^a.$$

Proof: Since for any $\sigma \in \mathrm{Aut}(\mathbb{Q}(\xi_n)/\mathbb{Q})$, it is completely determined by the image of the primitive n^{th} root, hence we conclude that $\mathrm{Aut}(\mathbb{Q}(\xi_n)/\mathbb{Q})$ is isomorphic to a subgroup of $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ hence a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. Next, by cardinality argument, as both have size $\varphi(n)$, we conclude that the two group are isomorphic. Then we also conclude that the map given in the Theorem has to be an isomorphism. □

Corollary 10.50.1 *Let $n = p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_k^{\alpha_k}$ be the decomposition of the positive integer n into distinct prime powers. The the cyclotomic fields $\mathbb{Q}(\xi_{p_i^{\alpha_i}})$, $i = 1, 2\cdots k$ intersects only in the field $\mathbb{Q}$ and their composite is the cyclotomic field $\mathbb{Q}(\xi_n)$, we have*

$$\mathrm{Gal}(\mathbb{Q}(\xi_n)/\mathbb{Q}) \cong \mathrm{Gal}(\mathbb{Q}(\xi_{p_i^{\alpha_i}}/\mathbb{Q}) \times \cdots \times \mathrm{Gal}(\mathbb{Q}_{p_k^{\alpha_k}}/\mathbb{Q}) \cong \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}).$$

Using Corollary 10.50.1, we can show that every abelian group appears as the Galois group of some extension of $\mathbb{Q}$, this is because $\mathrm{Gal}(\mathbb{Q}(\xi_n)/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^\times)$, and the rest follows from the fundamental theorem of finite abelian groups.

Corollary 10.50.2 *Let G be any finite abelian group. Then there is a subfield K of cyclotomic field with $\mathrm{Gal}(K/\mathbb{Q}) \cong G$.*

Proof: Suppose G is any finite abelian group. By the Fundamental Theorem of Abelian groups,

$$G \cong Z_{n_1} \times Z_{n_2} \times \cdots \times Z_{n_k}$$

for some integers $n_1, \cdots, n_k$. By Dirichelet's Theorem, we know there are infinitely many primes p with $p \cong 1 \mod m$, then there are distinct primes $p_1, p_2, \cdots, p_k$ such taht

$$\begin{aligned} p_1 &= 1 \mod n_1 \\ p_2 &= 1 \mod n_2 \\ &\vdots \\ p_k &= 1 \mod n_k \end{aligned}$$

and let $n = p_1p_2\cdots p_k$. By construction n_i divides $p_i - 1$ for $i = 1, 2, \cdots, k$, so the group Z_{p_i-1} (Cyclic) has a subgroup H_i of order $\frac{p_i-1}{n_i}$ for $i = 1, 2, \cdots, k$, and the quotient by this subgroup is cyclic of order n_i. Hence the quotient of $(\mathbb{Z}/n\mathbb{Z})^\times$ by $H = H_1 \times H_2 \times \cdots \times H_k$ is isomorphic to the group G, since

$$(\mathbb{Z}/n\mathbb{Z})^\times \cong (\mathbb{Z}/p_1\mathbb{Z})^\times \cdots \times (\mathbb{Z}/p_k\mathbb{Z})^\times$$

$$= Z_{p_1-1} \times \cdots \times Z_{p_k-1}.$$

We let $E = \mathbb{Q}(\xi_n)^H$, then since $(\mathbb{Z}/nZ)^*$ is abelian, $E/\mathbb{Q}$ is Galois. So we have $\mathrm{Gal}(E/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^\times/H \cong G$. $\square$

10.7 Normal Basis Theorem

Lemma 10.51 *Let E/F be a finite extension. Let $G = \mathrm{Aut}(E/F)$. Assume there is $\alpha \in E$ such that $S = \{g(\alpha) \mid g \in G\}$ is an F-basis (called the normal basis) of E. Then E/F is a Galois extension.*

Proof: We know $|G| \leq [E : F]$ by Corollary (10.22.1). By assumption, $[E : F] = |S| \leq |G|$. Then $|G| = [E : F]$, thus E/F is Galois. $\square$

Proposition 10.52 *Let E/F be a cyclic Galois extension. Then E/F admits a normal basis, that is there is some $\alpha \in E$ such that $\{g(\alpha) \mid g \in \mathrm{Gal}(E/F)\}$ is an F-basis of E.*

Proof: Let $G = \langle \sigma \rangle$ be of order n. We consider E as an $F[x]$-module that such $x \cdot a = \sigma(a)$ for $a \in E$. We claim the kernel of this action is $(x^n - 1)$. Clearly $x^n - 1$ is an element of the kernel. We show $\ker \subset (x^n - 1)$. Let $f(x) \in \ker$. Then $f(x) \cdot a = 0$ for any $a \in E$. That is $f(\sigma) \cdot a = \sum_{i=0}^{m} a_i \sigma^i(a) = 0$. Since by Linear independence of characters, we must have

$$a_0 + a_1\sigma + a_2\sigma^2 + \cdots + a_{n-1}\sigma^{n-1}$$

is not the zero function. This implies that $\deg f(x) \geq n$ or $f(x) = 0$. So we must have $\ker = (x^n - 1)$ by the Euclidean Algorithm over Euclidean Domains (F is a field, hence $F[x]$) is a Euclidean Domain).

Next we apply the structure theorem of finitely generated modules over $F[x]$ (PID):

$$E \cong F[x]/(f_1(x)) \oplus F[x]/(f_2(x)) \oplus \cdots \oplus F[x]/(f_k(x)).$$

such that $f_1(x)|f_2(x)|\cdots|f_k(x)|$. Notice $f_k(x) = \ker$. Then we must have

$$E \cong F[x]/(x^n - 1)$$

by dimension reasons (both dimension n). So we have some $\alpha \in E$, such that $\{\alpha, \sigma(\alpha), \cdots, \sigma^{n-1}(\alpha)\}$ is an F-basis of E. $\square$

Example: We consider $\mathbb{C}/\mathbb{R}$, $G = \langle , \sigma \rangle$. Then $\mathbb{C} = \mathbb{R}(i)$, but $G \cdot i = \{i, -i\}$ is not an $\mathbb{R}$-basis. So the naive choice $G(\alpha)$, where $E = F(\alpha)$ does not work.

Theorem 10.53 (Normal Basis Theorem) *Let E/F be a finite Galois extension with $\mathrm{Gal}(E/F) = G$. Then there exists $\alpha \in E$ such that $G \cdot \alpha = \{g(\alpha) \mid g \in G\}$ is a F-basis of E. In particular, we have an isomorphism of $F[G]$-model:*

$$E \cong F[G].$$

Proof: If F is finite, then E is finite as well. So E/F is cyclic, with $G = \langle Fr \rangle$. Then this directly follows from Proposition (10.52).

We assume F is infinite now. The idea is the following: we consider the equation

$$\begin{aligned} a_1 g_1 g_1(w) + a_2 g_1 g_2(w) + \cdots + a_n g_1 g_n(w) &= 0 \\ &\vdots \\ a_1 g_n g_1(w) + a_2 g_n g_2(w) + \cdots + a_n g_n g_n(w) &= 0 \end{aligned}$$

where $a_1, \cdots, a_n \in F$. Suppose we find $w \in E$ such that

$$(g_i g_j(w))_{ij}$$

is invertible, then by knowledge from Linear Algebra, we know $G \cdot w$ is a basis. Since G is a group, and every thing is invertible.

We know E/F is Galois, so E is simple with $E = F(\alpha)$ for some $\alpha \in E$. Let $m(x)$ be the minimal polynomial of $\alpha = \alpha_1$ in F. Then we have

$$m(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n)$$

We define

$$m_i(x) = \prod_{j \neq i} (x - \alpha_j).$$

We note $m_i(\alpha_i) \neq 0$. Normalize each m_i above so that $m_i(\alpha_i) = 1$. Since G permutes $\{m_i(x)\}$ (even after renormalization, $g \cdot \alpha_i$ is bijective), we can assume $g_i(m_1(x)) = m_i(x)$.

We consider the matrix

$$(g_i g_j(m_1(x))_{ij} = (g_i(m_j(x))) = (f_{ij}(x))_{ij}.$$

One can verify that $(m_{ij}(\alpha))_{ij}$ is a permutation matrix, since $m_1(\alpha) = 1$ and $m_k(\alpha) = 0$ for all $k \neq 1$. So $(f_{ij}(\alpha)) \neq 0$. This tells us that $\det(f_{ij}(x))$ is not a zero polynomial. Since F is infinite, then we can find $\beta \in F$ such that $\det(f_{ij}(\beta)) \neq 0$. Note

$$(f_{ij}(\beta)) = (g_i(m_j(\beta))_{ij}.$$

This is true since $\beta \in F$ is fixed by G. Then $(g_i g_j(\beta))_{ij}$ is invertible, since

$$\det(f_{ij}(\beta)) = \det(g_i(m_j(\beta)) = \det(g_i(g_j(m_1(\beta))) \neq 0.$$

So $G \cdot \beta$ is a basis. $\square$

10.8 Solvable and Radical Extensions

Theorem 10.54 (Hilbert's Theorem 90) *Let E/F be a cyclic Galois extension of* $\deg n$ *with the Galois group $G = \langle\sigma\rangle$. Let $\beta \in E$. Then $N_{E/F}(\beta) = 1$ if and only if there exists an element $\alpha \neq 0$ in E, such that $\beta = \frac{\alpha}{\sigma(\alpha)}$.*

Proof: $\Leftarrow$: Let $\beta = \frac{\alpha}{\sigma(\alpha)}$. Then

$$\begin{aligned} N_{E/F}(\beta) &= N_{E/F}(\alpha \cdot \sigma(\alpha)^{-1}) \\ &= \prod_{\tau \in S(E/F)} \tau(\alpha \cdot \sigma(\alpha)^{-1}) \\ &= \prod_{\tau \in G} \tau(\alpha \cdot \sigma(\alpha)^{-1}) \\ &= \prod_{\tau \in G} \tau(\alpha) \cdot \tau(\sigma(\alpha)^{-1}) \\ &= \prod_{\tau \in G} \tau(\alpha) \cdot ((\tau \circ \sigma)(\alpha))^{-1} \\ &= 1 \end{aligned}$$

since $\tau : E \to \bar{F}$ restricts to $\tau : E \to E$ as E/F is Galois.

$\Rightarrow$: now let $\beta \in E$ be such that $N_{E/F}(\beta) = 1 = \prod_{\tau \in G} \tau(\beta)$. Now we consider $\tau \in G$ as a character $\tau : E^* \to E^*$. So they are E-linearly independent. We define a function on E^* by

$$f = id + \beta\sigma + \beta\sigma(\beta)\sigma^2 + \beta\sigma(\beta)\sigma^2(\beta)\sigma^3 + \cdots + (\beta \cdots \sigma^{n-1}(\beta))\sigma^{n-1}.$$

We note that f is not the zero function by the Linear independence of Characters. So there is some $\theta \in E$ such that $f(\theta) = \alpha \neq 0$. Now we look at the properties of f. Let $x \in E$ we have

$$f(x) = x + \beta\sigma(x) + \beta\sigma(\beta)\sigma^2(x) + \cdots + (\beta \cdots \sigma^{n-2})\sigma^{n-1}(x).$$

Then

$$\sigma(f(x)) = \sigma(x) + \sigma(\beta)\sigma^2(x) + \sigma(\beta)\sigma^2(\beta)\sigma^3(x) + \cdots + \sigma(\beta)\sigma^2(\beta) \cdots \sigma^{n-1}(\beta)x.$$

Then $\beta\sigma(f(x)) = f(x)$, since $N_{E/F}(\beta) = 1$, we have $\beta\sigma(\beta)\sigma^2(\beta) \cdots \sigma^{n-1}(\beta) = 1$. Then since there exists θ such that $f(\theta) \neq 0$. Then $\beta = \frac{f(x)}{\sigma(f(x))}$. □

Example: We consider $\mathbb{C}/\mathbb{R}$. Then $G = \langle\sigma\rangle = \langle$complex conjugate$\rangle$. Let $i = \mathbb{C}$ be such that $N_{\mathbb{C}/\mathbb{R}}(i) = 1 = i \cdot \bar{i}$. Then by the Theorem, there exists α, such that $i = \frac{\alpha}{\sigma(\alpha)} = \frac{\alpha}{\bar{\alpha}}$. Note $\alpha = a + ai$ works. On the other hand, $N_{\mathbb{C}/\mathbb{R}}\left(\frac{\alpha}{\sigma(\alpha)}\right) = 1$.

Theorem 10.55 (Additive Version of Hilbert's Theorem 90) *Let E/F be a cyclic Galois extensions of* $\deg n$ *with $G = \mathrm{Gal}(E/F) = \langle\sigma\rangle$. Let $\beta = E$. Then $Tr_{E/F}(\beta) = 0$ if and only if there exists an element $\alpha \in E$, such that $\beta = \alpha - \sigma(\alpha)$.*

Proof: Let β be such that $Tr_{E/F}(\beta) = 0$ and G be of order n. By the linear independence of characters, since each $\sigma^i : E^* \to E^*$ is a character, then

$$Tr_{E/F} = \sigma + \sigma^2 + \cdots + \sigma^n$$

is not the zero function. In particular, there exists $\theta \in E^*$ such that $Tr_{E/F}(E) \neq 0$.
Next, we let

$$\alpha = \frac{1}{Tr_{E/F}(\theta)} \left[\beta\sigma(\theta) + (\beta + \sigma(\beta))\sigma^2(\theta) + \cdots + (\beta + \sigma(\beta) + \cdots + \sigma^{n-2}(\beta))\sigma^{n-1}(\theta)\right].$$

Then by direct computation, we have

$$\sigma(\alpha) = \frac{1}{Tr_{E/F}(\theta)} \left[\sigma(\beta)\sigma^2(\theta) + (\sigma(\beta) + \sigma^2(\beta))\sigma^3(\theta) + \cdots + (\sigma(\beta) + \cdots + \sigma^{n-1}(\beta))\sigma^n(\theta)\right].$$

So

$$\alpha - \sigma(\alpha) = \frac{1}{Tr_{E/F}(\theta)} \left[\beta\sigma(\theta) + \beta\sigma^2(\theta) + \cdots + \beta\sigma^{n-1}(\theta) - (\sigma(\beta) + \cdots + \sigma^{n-1}(\beta))\sigma^n(\theta)\right].$$

Since $Tr_{E/F}(\beta) = 0$, then

$$\sum_{i=1}^{n} \sigma^i(\beta) = 0 \Longrightarrow \sigma(\beta) + \cdots + \sigma^{n-1}(\beta) = -\beta.$$

Hence

$$\alpha - \sigma(\alpha) = \beta \cdot \frac{1}{Tr_{E/F}} \cdot \sum_{i=1}^{n} \sigma^i(\theta) = \beta \cdot \frac{Tr_{E/F}}{Tr_{E/F}} = \beta$$

as desired.

Conversely, if $\beta = \alpha - \sigma(\alpha)$, then by definition

$$\begin{aligned} Tr_{E/F}(\beta) &= \sum_{\tau \in G} \tau(\alpha - \sigma(\alpha)) \\ &= \sum_{\tau \in G} \tau(\alpha) - \tau(\sigma(\alpha)) \\ &= \sum_{\tau \in G} \tau(\alpha) - \sum_{\tau^* \in G} \tau^*(\alpha) \\ &= 0 \end{aligned}$$

□

Proposition 10.56 *Let F be a field of characteristic (can be zero) not dividing n which contains the n^{th} roots of unity. Then the extension $F(\sqrt[n]{a})/F$ is cyclic Galois extension of degree dividing n for any $a \in \mathbb{F}$.*

Proof: The extension $K = F(\sqrt[n]{a})$ is Galois over F since F contains the n^{th} roots of unity and $F(\sqrt[n]{a})$ is thus the splitting field for $x^n - a$. For any $\sigma \in \mathrm{Gal}(K/F)$, $\sigma(\sqrt[n]{a})$ is another root of this polynomial, hence $\sigma(\sqrt[n]{a}) = \xi_\sigma \sqrt[n]{a}$ for some n^{th} root of unity ξ_σ, this gives a map

$$\mathrm{Gal}(K/F) \longrightarrow \mu_n \cong \mathbb{Z}/n\mathbb{Z}$$

$$\sigma \longmapsto \xi_\sigma$$

where μ_n denotes the group of n^{th} roots of unity. Since F contains μ_n, every n^{th} root of unity is fixed by every element of $\mathrm{Gal}(K/F)$. Hence

$$\begin{aligned}\sigma\tau(\sqrt[n]{a}) &= \sigma(\xi_\tau \sqrt[n]{a}) \\ &= \xi_\tau \sigma(\sqrt[n]{a}) \\ &= \xi_\tau \xi_\sigma \sqrt[n]{a} = \xi_\sigma \xi_\tau \sqrt[n]{a}\end{aligned}$$

which shows that $\xi_{\sigma\tau} = \xi_\sigma \xi_\tau$, so the map above is a homomorphism. The kernel consists precisely of the automorphisms which fix $\sqrt[n]{a}$, which is the identity. This gives an injection of $\mathrm{Gal}(K/F)$ into the cyclic group μ_n of order n, hence the proposition follows. □

Proposition 10.57 *Any cyclic extension of degree n over a field F of characteristic not dividing n which contains the n^{th} roots of unity is of the form $F(\sqrt[n]{a})$ for some $a \in F$.*

Proof: Let $G = \mathrm{Gal}(E/F) = \langle\sigma\rangle$. Let $\xi_n \in F$. Then $N_{E/F}(\xi_n) = \sigma(\xi_n)\cdots\sigma^n(\xi_n) = \xi_n^n = 1$. Then there exists $\alpha \neq 0$ in E (by Theorem 10.54) such that $\xi_n = \frac{\alpha}{\sigma(\alpha)}$ or $\sigma(\alpha) = \xi_n^{-1}\alpha$. Then $\sigma(\alpha^n) = (\xi_n^{-1}\alpha)^n = \alpha^n$. Then $\alpha^n \in F$ by the fundamental theorem. So α is a root of the equation $x^n - a = 0$. Also $F \subset F(\alpha) \subset E$, we know $F(\alpha) = E^H$ for some $H \subset G$. Since any non-identity element of G does not fix α, because $\sigma(\alpha) = \xi_n^{-1}\alpha$. We conclude that $F(\alpha) = E$, i.e., $E = F(\sqrt[n]{a})$. □

Corollary 10.57.1 *Let F be a field containing n^{th} primitive roots of 1. Let E/F be a finite extension of degree d, such that $d|n$ and n does not divide the characteristic of F. Then E/F is cyclic if and only if $E = F(\sqrt[d]{a})$ for some $a \in F$.*

Remark 10.57.1 *A group G is said to have **exponent** n if $g^n = 1$ for every $g \in G$. Let F be a field of characteristic not dividing n which contains the n^{th} roots of unity. If we take elements $a_1, \cdots, a_k \in F^\times$ as in Proposition 10.56, then we can see the extension*

$$F(\sqrt[n]{a_1}, \sqrt[n]{a_2}, \cdots, \sqrt[n]{a_k})$$

is an abelian extension of F whose Galois group is of exponent n. Conversely, any abelian extension of exponent n is of the form.

Corollary 10.57.2 *Let $K = \mathbb{Q}(\sqrt[n]{a})$, where $a > 0$ and $a \in Q$. Suppose $[K : \mathbb{Q}] = n$ and let E be any subfield of K, with $[E : \mathbb{Q}] = d$. Then $E = \mathbb{Q}(\sqrt[d]{a})$.*

Remark 10.57.2 *This also shows under the hypothesis of the Corollary, if n is odd, then K has no nontrivial subfields which are Galois over $\mathbb{Q}$ and if n is even, then the only nontrivial subfield of K which is Galois over $\mathbb{Q}$ is $\mathbb{Q}(\sqrt{a})$.*

Proof: One way is to consider $F = \mathbb{Q}(\sqrt[n]{a}, \xi_n)$, and consider the fixed field of F. However, we show another method. We consider

$$N_{K/E}(\sqrt[n]{a}) = \prod_{\sigma \in S(K/E)} \sigma(\sqrt[n]{a}).$$

Note that $[K:E]$ is separable, hence $|S(K/E)| = [K:E]$ and any σ will map $\sqrt[n]{a}$ to $\xi_n^{i_\sigma}\sqrt[n]{a}$. Hence

$$N_{K/E}(\sqrt[n]{a}) = \xi_n^m(\sqrt[n]{\alpha})^{[K:E]}.$$

Since $N_{K/E} \in E$, as K/E is separable. Then we conclude that ξ_n^m is either 1 or -1. hence $N_{K/E} = \pm\sqrt[d]{\alpha}$. Then by a degree argument, we conclude that that $E = \mathbb{Q}(\sqrt[d]{a})$. □

Proposition 10.58 (Kummer Generators for Cyclic Extensions) *Let F be a field of characteristic not dividing n containing the n^{th} roots of unity and let K be a cyclic extension of degree d dividing n. We know $K = F(\sqrt[n]{a})$ for some nonzero $a \in F$. Then $K = F(\sqrt[n]{a}) = F(\sqrt[n]{b})$ if and only if $a = b^i c_1^n$ and $b = a^j c_2^n$ for some $c_1, c_2 \in F$ and positive integer i, j relatively prime to d.*

Proof: Firstly note that K is Galois over F, with cyclic Galois group. Then we let σ be a generator for the cyclic group. Then $\sigma(\sqrt[n]{\alpha}) = \xi\sqrt[n]{\alpha}$ for some primitive d^{th} roots of unity ξ. Then it is easy to see that if $K = F(\sqrt[n]{a}) = F(\sqrt[n]{b})$, then

$$\frac{\sigma(\sqrt[n]{a})}{\sqrt[n]{\alpha}} = \left(\frac{\sigma(\sqrt[n]{b})}{\sqrt[n]{b}}\right)^i$$

for some integer i relatively prime to d. However, this shows that σ fixes the expression

$$\frac{\sqrt[n]{a}}{(\sqrt[n]{b})^i}$$

so it must be in the fixed field of $\mathrm{Gal}(K/F)$ which is just F. Similarly, we have $\frac{\sqrt[n]{b}}{(\sqrt[n]{a})^j} \in F$. Thus we have $a = b^i c_1^n$ and $b = a^j c_2^n$ for some $c_1, c_2 \in F$ and some positive integers i, j relatively prime to d.

Conversely, suppose $a = b^i c_1^n$ and $b = a^j c_2^n$ for some $c_1, c_2 \in F$ and positive integer i, j relatively prime to d. Then we see that $F(\sqrt[n]{a}) = F(c_1\sqrt[n]{b^i}) \subset F(\sqrt[n]{b})$. Similarly, we have the other inclusion. Hence $F(\sqrt[n]{a}) = F(\sqrt[n]{b})$. □

Corollary 10.58.1 *Let p, q and r be primes in $\mathbb{Z}$ with $q \neq r$. Let $\sqrt[p]{q}$ denote any root of $x^p - q$ and let $\sqrt[p]{r}$ denote any root of $x^p - r$. Then $\mathbb{Q}(\sqrt[p]{q}) \neq \mathbb{Q}(\sqrt[p]{r})$.*

Proof: We prove by way of contradiction. Suppose $\mathbb{Q}(\sqrt[p]{q}) = \mathbb{Q}(\sqrt[p]{r})$, then we must have

$$\mathbb{Q}(\sqrt[p]{q}, \zeta_p) = \mathbb{Q}(\sqrt[p]{r}, \zeta_p)$$

where ζ_p is a primitive p^{th} roots of unity. Let $\mathbb{Q}_p$ denote the Cyclotomic field $\mathbb{Q}(\zeta_p)$. Then we have

$$\mathbb{Q}_p(\sqrt[p]{q}) = \mathbb{Q}_p(\sqrt[p]{r}).$$

We first show

$$\mathbb{Q}_p \cap Q(\sqrt[p]{q}) = \mathbb{Q}.$$

Since $\mathbb{Q}_p$ is the Cyclotomic field with prime order p, then we know $\mathbb{Q}_p/\mathbb{Q}$ is has an abelian Galois group of order $\varphi(p)$. Hence $Q(\sqrt[p]{q}) \cap \mathbb{Q}_p/Q$ being a subfield of $\mathbb{Q}_p/\mathbb{Q}$ is Galois as well (since its the subfield of an abelian Galois

extension). Now by Eisenstein's Criterion, we note that the polynomial $x^p - q = 0$ is irreducible over $\mathbb{Z}$, hence over $\mathbb{Q}$. This shows that $\mathbb{Q}(\sqrt[p]{q})$ has degree p. Then $[Q(\sqrt[p]{q}) \cap \mathbb{Q}_p : \mathbb{Q}]$ must divides both p and $\varphi(p)$. Since p is a prime and $\varphi(p) < p$ for $p \geq 2$, this shows $[Q(\sqrt[p]{q}) \cap \mathbb{Q}_p : \mathbb{Q}] = 1$, i.e.,

$$Q(\sqrt[p]{q}) \cap \mathbb{Q}_p = \mathbb{Q}.$$

Next, by the assumption $\mathbb{Q}(\sqrt[p]{q}) = \mathbb{Q}(\sqrt[p]{r})$, since $\mathbb{Q}_p$ contains all the p^{th} roots of unity, then we know that

$$\frac{\sqrt[p]{r}}{\left(\sqrt[p]{q}\right)^i} = \alpha \tag{10.2}$$

for some integer i and some $\alpha \in \mathbb{Q}_p$. Now by assumption $\mathbb{Q}(\sqrt[p]{q}) = \mathbb{Q}(\sqrt[p]{r})$, then the left hand side of (10.2) is in $\mathbb{Q}(\sqrt[p]{q})$, so $\alpha \in \mathbb{Q}(\sqrt[p]{q}) \cap \mathbb{Q}_p = \mathbb{Q}$. Raising both sides of (4.2) by the p^{th} power, we have

$$r = \left(\frac{m}{n}\right)^p q^i$$

for some $m, n \in \mathbb{Z}$ with $\gcd(m, n) = 1$. Then $nr = m^p q^i$. By Euclidean Lemma, $r|m$ since $r \neq q$. But then the right hand side is divisible by r^p and the left hand side only by r. This gives a contradiction. Hence we cannot have the two fields being equal.
□

Next, we assume that F is always a field with charactersitic 0. However, in practice we really just need $x^n - 1$ to be separable over F, but this n may vary depending on the problem, and becomes complicated for fields with characterstic non-zero.

Definition 10.59 (Solve By Radicals) *Assume F is characteristic 0.*

*An element α which is algebraic over F can be **expressed by radicals or solved for in terms of radicals** if α is an element of a field K which can be obtained by a succession of simple radical extensions*

$$F = K_0 \subset K_1 \subset \cdots \subset K_i \subset K_{i+1} \subset \cdots \subset K_s = K, \tag{10.3}$$

where $K_{i+1} = K_i(\sqrt[n_i]{a_i})$ for some $a_i \in K_i$. Here $\sqrt[n_i]{a_i}$ denotes some root of the polynomial $x^{n_i} - a_i$ and $x^{n_i} - a_i$ is separable.

*A polynomial $f(x) \in F[x]$ can be **solved by radicals** if all its roots can be solved for in terms of radicals.*

*A finite separable extension E/F is called **solvable by radicals** if there exists a finite extension K/E such that we have a tower of extension*

$$F = K_0 \subset K_1 \subset \cdots \subset K_i \subset K_{i+1} \subset \cdots \subset K_s = K,$$

where $K_{i+1} = K_i(\sqrt[n_i]{a_i})$ for some $a_i \in K_i$ such that $x^n - a$ is separable.

Remark 10.59.1 *Recall that a group G is called **solvable** if it admits a normal tower $G = G_0 \supset G_1 \supset \cdots \supset$*

$G_m = \{e\}$ such that G_i/G_{i+1} is abelian (or cyclic). Let E/F be a finite seperable extensions. Then E/F is called ***solvable*** *if* $\mathrm{Gal}(K/F)$ *is solvable where K is the Galois closure of E. Later we will show the two notion of "solvable extensions" coincides.*

Lemma 10.60 *Let E/F be a finite separable extensions. Let K/F be the Galois closure. Then E/F is solvable by radicals if and only if K/F is solvable by radicals.*

Proof: $\Leftarrow$: Assume K/F is solvable by radicals. Then we have a finite extension L/K such that

$$F = L_0 \subset L_1 \subset \cdots \subset L_n = L$$

satisfying the definition. Note L/E is also finite (adjoin finitely many roots), and the same tower shows that E/F is solvable by radicals.

$\Rightarrow$: Assume E/F is solvable by radicals, then we have a finite extension. L/E such that $F = L_0 \subset L_1 \subset \cdots \subset L_n \subset L$ such that $L_{i+1} = L_i(\sqrt[n_i]{a_i})$. Now consider $\tau : \bar{F} \to \bar{F}$, we can apply τ to the tower to obtain a tower

$$F = L_0 \subset \tau(L_1) \subset \cdots \subset \tau(L_n) = \tau(L) \supset \tau(E).$$

Then $\tau(E)$ is also solvable by radicals. Hence, $E\tau(E)$ is solvable by radicals, since we have

$$F = L_0 \subset L_1 \subset L_2 \subset \cdots \subset L_n = L \subset L\tau_1(L_1) \subset \cdots L\tau(L_0) \cdots L\tau(L)$$

where $L\tau(L_{i+1}) = L\tau(L_i)(\tau(\sqrt[n_i]{a_i}))$. Hence Note that

$$K = E\tau_1(E)\tau_2(E(\tau_1(E)) \cdots \tau_k(E\tau_1(E) \cdots \tau_{k_1}(E))$$

for some finite k and $\tau_i : \bar{F} \to \bar{F}$ an F-automorphism. This shows K/F is solvable by radicals. □

Recall the following: if

$$1 \to N \to G \to H \to 1$$

is a short exact sequence, then G is solvable if and only if N and H is solvable.

Theorem 10.61 *Let E/F be a finite separable extensions. Assume F contains all roots of unity. Then E/F is solvable by radicals if and only if the Galois closure of E/F is solvable.*

Proof: We can assume E/F is Galois, since we can replace E/F by its Galois closure using Lemma (10.60).

Assume E/F is solvable with $\mathrm{Gal}(E/F) = G$. Then since G is solvable, we have

$$\{e\} = H_0 \subset H_1 \subset H_2 \subset \cdots \subset H_n = G$$

such that H_{i+1}/H_i is cyclic. Thus we have a tower of field

$$E = E_0 \supset E_1 \supset \cdots \supset E_n = F,$$

where $E_i = E^{H_i}$. Note that we have a short exact sequence

$$1 \to \underbrace{\mathrm{Gal}(E/E_i)}_{H_i} \to \underbrace{\mathrm{Gal}(E/E_{i+1})}_{H_{i+1}} \to \underbrace{\mathrm{Gal}(E_i/E_{i+1})}_{H_{i+1}/H_i} \to 1$$

So E_i/E_{i+1} is a cyclic Galois extension. The field E_{i+1} contains $|[E_i : E_{i+1}]|^{th}$ primitive root of 1. Then by Corollary (10.57.1), $E_i = E_{i+1}(\sqrt[n_{i+1}]{a_{i+1}})$ where $a_{i+1} \in E_{i+1}$. So the extension is solvable by radicals.

Assume E/F is solvable by radicals. Then we have a finite extension K/E such that

$$F = K_0 \subset K_1 \subset K_2 \subset \cdots \subset K_n = K$$

such that $K_{i+1} = K_i(\sqrt[n_i]{a_i})$ for some $a_i \in K_i$. We can assume K/F is Galois (As otherwise, we can mimic the proof of Lemma (3.32), to extend the tower). Then the corresponding tower of group

$$\mathrm{Gal}(K/F) \supset \mathrm{Gal}(K/K_1) \supset \cdots \supset \mathrm{Gal}(K/K_n) = \{e\}.$$

Note that the extension K_{i+1}/K_i is normal since we have all roots of unity. So we have a short exact sequence

$$1 \to \mathrm{Gal}(K/K_{i+1}) \to \mathrm{Gal}(K/K_i) \to \mathrm{Gal}(K_{i+1}/K_i) \to 1.$$

We know $\mathrm{Gal}(K_{i+1}/K_i)$ is cyclic, so $\mathrm{Gal}(K/F)$ is solvable. Since E/F is Galois, we also have a short exact sequence,

$$1 \to \mathrm{Gal}(K/E) \to \mathrm{Gal}(K/F) \to \mathrm{Gal}(E/F) \to 1.$$

So $\mathrm{Gal}(E/F)$ is solvable. □

Lemma 10.62 *Let E/F be a finite separable extension. Let ξ_n be a primitive with root of 1. Then E/F is solvable by radicals if and only if $E(\xi_n)/F(\xi_n)$ is solvable by radicals.*

Proof: Suppose E/F is solvable by radicals, let K/E be finite such that

$$F = K_0 \subset K_1 \subset \cdots \subset K_i \subset K_{i+1} \subset \cdots \subset K_s = K,$$

satsifies the definition. Then

$$F(\xi_n) = K_0F(\xi_n) \subset K_1F(\xi_n) \subset \cdots \subset K_sF(\xi_n) = E(\xi_n)$$

shows $E(\xi_n)/F(\xi_n)$ is solvable by radicals. Conversely, suppose

$$F(\xi_n) = G_0 \subset \cdots \subset G_t = G,$$

where $G/E(\xi_n)$ is a finite extension satisfying the definition of $E(\xi_n)/F(\xi_n)$ is being solvable by radicals. Then we note that $F(\xi_n)$ is obtained by adjoining a primitive root of 1 to F, in particular $x^n - 1$ is separable (primitive

roots means the roots of unity is precise order n). Hence we have

$$F \subset F(\xi_n) = G_0 \subset \cdots \subset G_t = G,$$

and clearly G/E is still finite. So E/F is solvable by radicals. □

Theorem 10.63 *Let E/F be a finite separable extension. Then E/F is solvable by radicals if and only if the Galois closure of E/F is solvable.*

Proof: Let n be a large enough integer and ξ_n be a primtive n^{th} roots of 1. We know E/F is solvable by radicals if and only if $E(\xi_n)/F(\xi_n)$ is solvable by radicals. We show this happens if and only if the Galois closure of $E(\xi_n)/F(\xi_n)$ is solvable.

Suppose the Galois closure of E/F is solvable. We can assume that E/F is Galois. Then we know $E(\xi_n)/F(\xi_n)$ is also Galois (normal and separable), moreover

$$\mathrm{Gal}(E(\xi_n)/F(\xi_n)) \cong \mathrm{Gal}(E/E \cap F(\xi_n)).$$

Note that we have a short exact sequence of groups

$$1 \to \mathrm{Gal}(E/E \cap F(\xi_n)) \to \mathrm{Gal}(E/F) \to \mathrm{Gal}(E \cap F(\xi_n)/F) \to 1.$$

Then we see $\mathrm{Gal}(E/F)$ solvable implies $\mathrm{Gal}(E(\xi_n)/F(\xi_n))$ is solvable.

We also have a short exact sequence

$$1 \to \mathrm{Gal}(F(\xi_n)/E \cap F(\xi_n)) \to \mathrm{Gal}(F(\xi_n)/F) \to \mathrm{Gal}(E \cap F(\xi_n)/F) \to 1.$$

$\mathrm{Gal}(F(\xi_n)/F)$ is abelian, hence solvable. Then $\mathrm{Gal}(E \cap F(\xi_n)/F)$ is always solvable. Now suppose $\mathrm{Gal}(E(\xi_n)/F(\xi_n))$ is solvable. This shows $\mathrm{Gal}(E/F)$ is solvable. □

Corollary 10.63.1 *If α is contained in a radical extension K as shown in (10.3), then α is contained in a root extension which is Galois over F and we can find a tower such that L_{i+1}/L_i is cyclic.*

Corollary 10.63.2 *The polynomial $f(x)$ can be solved by radicals if and only if its Galois group is a solvable group.*

Next, we consider general fields with characteristic p where p is a prime.

Proposition 10.64 (Artin-Schreier Extension) *Let F be a field of charactersitic p and let K be a cyclic extension of F of degree p. Then $K = F(\alpha)$ where α is a root of the polynomial $x^p - x - a$ for some $a \in F$.*

Proof: Since $Tr_{K/F}(-1) = (-1) \times p = 0$. Then by Theorem (10.55), there exists $\alpha \in K$ such that $-1 = \alpha - \sigma\alpha$, where σ is the generator of $\mathrm{Gal}(K/F)$. We note that $\sigma(\alpha^p - \alpha) = \sigma(\alpha)^p - \sigma(\alpha) = (\alpha + 1)^p - (\alpha + 1) = \alpha^p - \alpha$. Hence we conclude that $\alpha^p - \alpha = a \in F$. Then α solves $x^p - x - a$ which is an irreducible polynomial as we saw in a previous example. Hence $K = F(\alpha)$ by degree reason (both degree p). □

Proposition 10.65 *Let α be a root of $x^p - x - a$ for some $a \in F$. Then $F(\alpha) = F$ or $F(\alpha)/F$ is a cyclic Galois extension of degree p.*

Proof: We know if $a = 0$, then $F(\alpha) = F$. Otherwise, $\alpha + i$ is a root of $x^p - x - a$ for $i \in \mathbb{F}_p$. We know $x^p - x - a$ is irreducible. Hence $F(\alpha)$ is the splitting field of a separable polynomial. Thus $F(\alpha)/F$ is Galois of order p, hence the Galois group must be cyclic. □

Definition 10.66 (Solvable By Radicals) *Let F be a field of* Char $p > 0$. *Let E/F be a finite and separable. Then E/F is called* ***solvable by radical*** *if there is a finite extensions K/E such that*

$$F = K_0 \subset K_1 \subset \cdots \subset K_n = K$$

such that $k_{i+1} = k_i(\sqrt[n_i]{a_i})$ where $\gcd(n_i, p) = 1$ *or $k_{i+1} = k_i(\alpha)$ where α is a root of $x^p - x - a$.*

Theorem 10.67 *Let E/F be a finite separable extension with characteristic $p > 0$. Then E/F is solvable by radicals if and only if the Galois closure of E/F is solvable.*

Proof: We may assume that F contains all roots of 1. And we may assume E/F is Galois.

$\Leftarrow$: Assume $G = \mathrm{Gal}(E/F)$ is solvable. That is

$$\{e\} = H_0 \subset H_1 \subset \cdots \subset H_n = G$$

such that H_{i+1}/H_i is cyclic of prime orders. Then if $|H_{i+1}/H_i| = q$ for $q \neq p$, then we know $E_i = E_{i+1}(\sqrt[q]{a_i})$. If $|H_{i+1}/H_i| = p$, then we have $E_i = E_{i+1}(\alpha)$ for some α satisfying $x^p - x - a = 0$ ($a \in E_{i+1}$).
$\Rightarrow$: Assume E/F is solvable by radicals. Then we assume we have a tower

$$F = E_0 \subset E_1 \subset \cdots \subset E_n = E$$

such that $E_{i+1} = E_i(\sqrt[n_i]{a_i})$ for $\gcd(n_i, p) = 1$, $a_i \in E_i$ or $E_{i+1} = E_i(\alpha)$ for α satisfying $x^p - x - a = 0$, $a \in E_i$. So we have a tower of groups

$$\mathrm{Gal}(E/F) \supset \mathrm{Gal}(E/E_1) \supset \cdots \supset \mathrm{Gal}(E/E_n) = \{e\}$$

such that

$$\mathrm{Gal}(E/E_i)/\mathrm{Gal}(E/E_{i+1}) \cong \mathrm{Gal}(E_{i+1}/E_i)$$

which is cyclic. □

10.9 Galois Groups of Polynomials

In general, if the factorization of $f(x)$ into irreducibles is $f(x) = f_1(x) \cdots f_k(x)$ where $f_i(x)$ has degree n_i, $i = 1, 2, \cdots, k$, then since the Galois group permutes the roots of the irreducible factors among themselves we have

$$\mathrm{Gal}(K/F) \leq S_{n_1} \times \cdots \times S_{n_k}$$

where K/F is the Galois group of f.

Definition 10.68 (Elementary Symmetric Function) *Let $x_1, x_2, \cdots, x_n$ be indeterminates over a field F. The **elementary symmetric functions** $s_1, s_2, \cdots, s_n$ (or $e_1, \cdots, e_n$) are defined by*

$$\begin{aligned} s_1 &= x_1 + x_2 + \cdots + x_n \\ s_2 &= x_1x_2 + x_1x_3 + \cdots + x_2x_3 + x_2x_4 + \cdots + x_{n-1}x_n \\ &\vdots \\ s_n &= x_1x_2 \cdots x_n \end{aligned}$$

i.e., the i^{th} symmetric function s_i of $x_1, x_2, \cdots, x_n$ is the sum of all products of the $x_j's$ taken i at a time.

Remark 10.68.1 *There are complete symmetric functions.*

Definition 10.69 (General Polynomial) *The **general polynomial of degree** n is the polynomial*

$$(x - x_1)(x - x_2) \cdots (x - x_n)$$

whose roots are the indeterminates $x_1, x_2, \cdots, x_n$.

Remark 10.69.1 *Note the coefficients of the general polynomial are given by the elementary symmetric functions, that is*

$$(x - x_1)(x - x_2) \cdots (x - x_n) = x^n - s_1x^{n-1} + s_2x^{n-2} + \cdots + (-1)^n s_n.$$

For any field F, suppose $x_1, \cdots, x_n$ are distinct. Then the extension $F(x_1, x_2, \cdots, x_n)$ is then a Galois extension of the field $F(s_1, s_2, \cdots, s_n)$ since it is the splitting field of the general polynomial of degree n.

Proposition 10.70 *The fixed field of the symmetric group S_n acting on the field of rational functions in n variables $F(x_1, \cdots, x_n)$ is the field or rational functions in the elementary symmetric function $F(s_1, s_2, \cdots, s_n)$. Moreover, $F(x_1, \cdots, x_n)/F(s_1, \cdots, s_n)$ is Galois with Galois group S_n.*

Proof: We note that $F(x_1, \cdots, x_n)$ is the splitting field of

$$(x - x_1)(x - x_2) \cdots (x - x_n) \in F(s_1, \cdots, s_n)[x].$$

Then $[F(x_1, \cdots, x_n) : F(s_1, \cdots, s_n)] \leq n!$. Next let $L = F(x_1, \cdots, x_n)^{S_n}$. Then since $s_1, \cdots, s_n \in L$. We know $F(s_1, \cdots, s_n) \subset L$. But $[F(x_1, \cdots, x_n) : L] = n!$. Then $[F(x_1, \cdots, x_n) : F(s_1, \cdots, s_n)] \geq n!$. This shows $L = F(s_1, \cdots, s_n)$. □

Remark 10.70.1 *This result says that if there are no relations among the coefficients of a polynomial of degree n (which is what we mean when we say the s_i are indeterminates above) then the Galois group of this polynomial over the field generated by its coefficients is the full symmetric group S_n.*

Corollary 10.70.1 *The general equation of degree n cannot be solved by radicals for $n \geq 5$.*

Proof: For $n \geq 5$ the group S_n is not solvable. □

Corollary 10.70.2 *The elementary symmetric functions is algebraically independent.*

Example: Let $E/\mathbb{Q}$ be the splitting field of $x^5 - 4x + 2$. We know $\mathrm{Gal}(E/\mathbb{Q}) \cong S_5$. Recall S_5 is not solvable. This means we cannot write every root in terms of repeated radicals and addition, multiplication, division and subtraction. We have shown that if $K/\mathbb{Q}$ is solvable, then so is $\sigma(K)/\mathbb{Q}$, where $\sigma : \bar{\mathbb{Q}} \to \bar{\mathbb{Q}}$. This shows that write any root of $x^5 - 4x + 2$ in terms of radicals, because the action of S^5 on the roots is transitive.

Definition 10.71 *A rational function $f(x_1, x_2, \cdots, x_n)$ is called **symmetric** if it is not changed by any permutation of the variables $x_1, x_2, \cdots, x_n$.*

Corollary 10.71.1 (Fundamental Theorem on Symmetric Functions) *Any symmetric functions in the variables $x_1, x_2, \cdots, x_n$ is a rational function in the elementary symmetric functions $s_1, s_2, \cdots, s_n$.*

Proof: A symmetric function lies in the fixed field of S_n, hence is a rational function in $s_1, \cdots, s_n$. □

Corollary 10.71.2 *There is an $f(x_1, \cdots, x_n) \in F(x_1, \cdots, x_n)$ such that $\{S_n \cdot f(x_1, \cdots, x_n)\}$ is a basis of $F(x_1, \cdots, x_n)$ over $F(s_1, \cdots, s_n)$.*

Proof: This follows from the Normal Basis Theorem (10.53). □

Definition 10.72 (Discriminant) *Define the **discriminant** D of $x_1, x_2, \cdots, x_n$ by the formula*

$$D = \prod_{i<j} (x_i - x_j)^2.$$

Define the discriminant of a polynomial to be the discriminant of the roots of the polynomial. And define the root of the discriminant D to be

$$\sqrt{D} = \prod_{i<j} (x_i - x_j).$$

Remark 10.72.1 *Observe that $D = 0$ for a polynomial if and only if the polynomial is not separable. Over a perfect field (e.g., $\mathbb{Q}$ or a finite field), this implies $f(x)$ is reducible since every irreducible polynomial over a perfect field is separable.*

Remark 10.72.2 *Suppose $f(x) \in F[x]$ is a degree n polynomial such that its Galois group is isomorphic to S_n, then its discriminant D is in F, since it is fixed by every element of S_n.*

Proposition 10.73 *If $\mathrm{char}(F) \neq 2$, then the permutation $\sigma \in S_n$ is an element of A_n if and only if it fixes the square root of the discriminant D.*

Proof: Since if $\mathrm{sgn}(\sigma) = -1$, i.e, $\sigma \in S_n \setminus A_n$, then $\sigma(\sqrt{D}) = -\sqrt{D}$. $\sigma(\sqrt{D}) = \sqrt{D}$ if $\mathrm{sgn}(\sigma) = 1$. □

Proposition 10.74 *The Galois group of $f(x) \in F[x]$ is a subgroup of A_n if and only if $\sqrt{D} \in F$.*

Proof: The Galois group is contained in A_n if and only if every element of the Galois group fixes

$$\sqrt{D} = \prod_{i<j}(\alpha_i - \alpha_j)$$

i.e., if and only if $\sqrt{D} \in F$. □

Proposition 10.75 *Let F be a field, and $p(x)$ be a separable irreducible polynomial in F. Then the Galois group G of p act transitively on the roots of F.*

Proof: Let α be a root of $p(x)$, then G maps α to another root of $q(x)$, let $\beta_1, \cdots, \beta_n$ be all the image of α under the elements of G, we show $n = \deg p(x)$. Suppose not, then

$$q(x) = \prod_{i=1}^{n}(x - \beta_i)$$

is a polynomial fixed by G, hence is a polynomial in F. Note that α is a root of $q(x)$ contradicting that $p(x)$ is an irreducible polynomial. □

Next we want to determine the Galois group of certain polynomials.

10.9.1 Polynomials of Degree 2

Consider the polynomial $x^2 + ax + b$ with roots α, β. The discriminant D for this polynomial is $(\alpha - \beta)^2$, so

$$\begin{aligned} D &= (\alpha - \beta)^2 \\ &= (\alpha + \beta)^2 - 4\alpha\beta \\ &= s_1^2 - 4s_2 \\ &= (-a)^2 - 4(b) = a^2 - 4b. \end{aligned}$$

The polynomial is separable if and only if $a^2 - 4b \neq 0$. The Galois group $G = \mathrm{Gal}(E/\mathbb{Q})$ is a subgroup of S_2 and is trivial if and only if $a^2 - 4b$ is a rational square, i.e., $x^2 + ax + b$ is reducible.

10.9.2 Polynomials of Degree 3

Suppose the cubic polynomial is given by

$$f(x) = x^3 + ax^2 + bx + c.$$

Then we know $G = \mathrm{Gal}(E/\mathbb{Q}) \subset S_3$. Suppose $f(x)$ is reducible, then we are back to the case $x^2 + a'x + b'$, which we know is either trivial or S_2. If $f(x)$ is irreducible. Then we claim that either $G \cong A_3$ or $G \cong S_3$.

In the irreducible case, we have $E = \mathbb{Q}(\alpha_1, \alpha_2, \alpha_3)$, where $\alpha_1, \alpha_2, \alpha_3 \notin \mathbb{Q}$ are roots of f. In this case, we always have some $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$ such that $\sigma(\alpha_i) = \alpha_j$ for any i, j. This is because $3 | \mathrm{Gal}(E/\mathbb{Q})$, so G must contain a 3-cycle. Hence the action $G \curvearrowright \{\alpha_1, \alpha_2, \alpha_3\}$ is transitive, in particular, $G \cong A_3$ or $G \cong S_3$. Next, we can determine whether we have S_3 or A_3 using determinants.

We make the substitution $x = y - a/3$ the polynomial becomes

$$g(y) = y^3 + oy + q \tag{10.4}$$

where

$$p = \frac{1}{3}(3b - a^2), \quad q = \frac{1}{27}(2a^3 - 9ab + 27c).$$

The splitting fields for these two polynomials are the same since their roots differ by the constant $a/3 \in \mathbb{Q}$ and since the formula for the discriminant involves the differences of roots, we see that these two polynomials also have the same discriminant.

Let the roots of the polynomial in Equation 10.4 be α, β and γ. Then

$$g(y) = (y - \alpha)(y - \beta)(y - \gamma)$$

so

$$D_y g(y) = (y - \alpha)(y - \beta) + (y - \alpha)(y - \gamma) + (y - \beta)(y - \gamma).$$

Then

$$\begin{aligned} D_y g(\alpha) &= (\alpha - \beta)(\alpha - \gamma) \\ D_y g(\beta) &= (\beta - \alpha)(\beta - \gamma) \\ D_y g(\gamma) &= (\gamma - \alpha)(\gamma - \beta). \end{aligned}$$

Taking the product we see that

$$D = [(\alpha - \beta)(\alpha - \gamma)(\beta - \gamma)]^2 = -D_y g(\alpha) D_y g(\beta) D_y g(\gamma).$$

Since $D_y g(y) = 3y^2 + p$, we can quickly conclude that

$$D = -4p^3 - 27q^2 = a^2b^2 - 4b^3 - 4a^3c - 27c^2 + 18abc.$$

Then $G = S_3$ if $\sqrt{D} \notin \mathbb{Q}$ and $G = A_3$ if $\sqrt{D} \in \mathbb{Q}$ (of course provided f is irreducible).

10.9.3 Polynomial $f(t) = x^n - a$

We study the Galois group of the polynomial $x^n - a$ over $\mathbb{Q}$.

Proposition 10.76 *Let G be the Galois group of $x^n - a$. Then G is isomorphic to a subgroup of $\mathbb{Z}/n\mathbb{Z} \rtimes (\mathbb{Z}/n\mathbb{Z})^\times$, where after identifying $(\mathbb{Z}/n\mathbb{Z})^\times$ with $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$, then multiplication is given by $(g, \sigma) \cdot (h, \tau) = (g\sigma(h), \sigma\tau)$.*

Proof: All roots of $x^n - a$ are

$$\{ \sqrt[n]{a}\xi_n^m \mid 1 \leq m \leq n\}.$$

So the splitting field is $\mathbb{Q}(\xi_n, \sqrt[n]{a})$ and the Galois group is determined by its action on $\sqrt[n]{a}, \xi_n$. We have the following possibilities:

$$\begin{aligned} \sqrt[n]{a} &\longmapsto \xi_n^m \sqrt[n]{a}, \quad 0 \leq m \leq n-1 \\ \xi_n &\longmapsto \xi_n^k \quad \gcd(k, n) = 1,\ 0 \leq a \leq n \end{aligned}$$

One may check this is precisely the group $\mathbb{Z}/n\mathbb{Z} \rtimes (\mathbb{Z}/n\mathbb{Z})^\times$ if we temporarily disregard the condition for field homomorphisms. □

Corollary 10.76.1 *$|G| = [\mathbb{Q}(\xi_n, \sqrt[n]{a}) : \mathbb{Q}]$ divides $n \cdot \varphi(n)$.*

Example:

1. Let E be the splitting field of $x^8 - 2$. We compute its Galois subgroup. Since $\xi_8 = \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}$, then $\xi_8 + \xi_8^{-1} = \sqrt{2}$. We know

   $$\mathbb{Q}(\xi_8, \sqrt[8]{2}) \supset \mathbb{Q}(\xi_8) \supset \mathbb{Q}.$$

   But since $8 \cdot \varphi(8) = 32$, and

   $$[\mathbb{Q}(\xi_8, \sqrt[8]{2}) : \mathbb{Q}] = [\mathbb{Q}(\xi_8, \sqrt[8]{2}) : \mathbb{Q}(\xi_8)][\mathbb{Q}(\xi_8) : \mathbb{Q}] \leq 4 \times 4.$$

   We conclude that $\mathrm{Gal}(E/\mathbb{Q})$ is a proper subgroup of $\mathbb{Z}/8Z \times (\mathbb{Z}/8Z)^\times$.

   We can compute the degree of $\mathrm{Gal}(E/\mathbb{Q})$ explicitly. Let z be a real root of $x^8 - 2$, then $\mathbb{Q}(z)/\mathbb{Q}$ is of degree 8 ($x^8 - 2$ is irreducible over $\mathbb{Z}[x]$, hence $\mathbb{Q}[x]$). Then since $\mathbb{Q}(z) \subset \mathbb{R}$, $\mathbb{Q}(z, i)/\mathbb{Q}(z)$ is of degree 2. So $\mathbb{Q}(z, i)/\mathbb{Q}$ is degree 16. As $\mathbb{Q}(z, i) \subset E$, then we must have that $E/\mathbb{Q}$ is degree 16.

   Then it follows that we must have the following two automorphism of $\mathbb{Q}(\xi_8, \sqrt[8]{2})/\mathbb{Q}$ (by cardinality reason),

   $$\begin{aligned} \sigma &: z \mapsto \xi_8 z, \quad i \mapsto i \\ \tau &: z \mapsto z, \quad i \mapsto -i. \end{aligned}$$

   In fact, we can conclude that $\langle \sigma, z\rangle \subset \mathrm{Gal}(E/\mathbb{Q}) \subset \mathbb{Z}/8\mathbb{Z} \rtimes (\mathbb{Z}/8\mathbb{Z})^*$, as one can show that $\langle \sigma, \tau\rangle$ has order 16, so it is precisely $\mathrm{Gal}(E/\mathbb{Q})$.

Proposition 10.77 *Let E be the splitting field of $x^n - a$ where $x^n - a$ is irreducible. Then $[E : \mathbb{Q}] = n\varphi(n)$ or $\frac{1}{2}n\varphi(n)$.*

Proof: Let $K = \mathbb{Q}(\xi_n)$, where n is a root of unity. Then we know $K/\mathbb{Q}$ is a Galois extension. Then by Corollary (10.45.1), we have that

$$[E : \mathbb{Q}] = \frac{[K : \mathbb{Q}][\mathbb{Q}(\sqrt[n]{a}) : \mathbb{Q}]}{[K \cap \mathbb{Q}(\sqrt[n]{a}) : \mathbb{Q}]}.$$

We know $[K : \mathbb{Q}] = \varphi(n)$, $[\mathbb{Q}(\sqrt[n]{a} : \mathbb{Q}] = n$, since $x^n - a$ is irreducible. Now $K \cap \mathbb{Q}(\sqrt[n]{\alpha})$ is a subfield of $\mathbb{Q}(\xi_n)/\mathbb{Q}$, which has an abelian Galois group. Thus $K \cap \mathbb{Q}(\sqrt[n]{\alpha})$ is Galois over $\mathbb{Q}$. Then by the remark of Corollary (10.57.2), $[K \cap \mathbb{Q}(\sqrt[n]{a}) : \mathbb{Q}]$ can only be 1 or 2. Hence $[E : \mathbb{Q}] = n\varphi(n)$ or $\frac{1}{2}n\varphi(n)$. □

10.9.4 Polynomial of Degree 4

Let the quartic polynomial be

$$f(x) = x^4 + ax^3 + bx^2 + cx + d$$

which under the substitution $x = y - \frac{a}{4}$ becomes the quartic

$$g(y) = y^4 + py^2 + qy + r$$

with

$$\begin{aligned} p &= \frac{1}{8}(-3a^2 + 8b) \\ q &= \frac{1}{8}(a^3 - 4ab + 8c) \\ r &= \frac{1}{256}(-3a^4 + 16a^2b - 64ac + 256d). \end{aligned}$$

Let the roots of $g(y)$ be $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and let G denote the Galois group for the splitting field of $g(y)$, then note that it is also the Galois group for the splitting field of $f(x)$.

Suppose first that $g(y)$ is reducible. If $g(y)$ splits into a linear and a cubic, then G is the Galois group of the cubic, which we know exactly how it looks like. Suppose then that $g(y)$ splits into two irreducible quadratics. Then the splitting field is the extension $F(\sqrt{D_1}, \sqrt{D_2})$ where D_1 and D_2 are the discriminants of the two quadratics. If D_1 and D_2 do not differ by a square factor then this extension is a **biquadratic extensions** and G is isomorphic to the Klein 4-subgroup of S_4. If $D - 1$ is a square times D_2, then this extensions is a quadratic extension and G is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

So it remains to consider the situation when $g(y)$ is irreducible. Then the Galois group is transitive on the roots. The only transitive subgroups of S_4 are the following:

- S_4;
- A_4;
- $D_8 = \{1, (1324), (12)(34), (1423), (13)(24), (14)(23), (12), (34)\}$ and its conjugates;

- $V = \{1, (12)(34), (13)(24), (14)(23)\}$
- $C = \{1, (1234), (13)(24), (1432)\}$ and its conjugates.

We remark thatD_8 is the dihedral group, a Sylow 2 - subgroup of S_4, with 3 (isomorphic) conjugate subgroups in S_4, V is the Klein 4 - subgroup of S_4, normal in S_4, and C is a cyclic group, with 3 (isomorphic) conjugates in S_4. Consider the elements

$$\theta_1 = (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)$$
$$\theta_2 = (\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)$$
$$\theta_3 = (\alpha_1 + \alpha_4)(\alpha_2 + \alpha_3)$$

in the splitting field for $g(y)$. These elements are permuted amongst themselves by the permutations in S_4. The stabilizer of θ_1 in S_4 is the dihedral group D_8. The stabilizers in S_4 of θ_2 and θ_3 are the conjugate dihedral subgroups of order 8. The subgroup of S_4 which stabilizes all three of these elements is the intersection of these subgroups, namely the Klein 4 group V.
Since S_4 merely permutes $\theta_1, \theta_2, \theta_3$ it follows that the elementary symmetric func - tions in the θ's are fixed by all the elements of S_4, hence are in F. An elementary computation in symmetric functions shows that these elementary symmetric functions are $2p, p^2 - 4r$, and $-q^2$, which shows that $\theta_1, \theta_2, \theta_3$ are the roots of

$$h(x) = x^3 - 2px^2 + (p^2 - 4r)x + q^2$$

called the **resolvent cubic** for the quartic $g(y)$. Since

$$\begin{aligned}\theta_1 - \theta_2 &= \alpha_1\alpha_3 + \alpha_2\alpha_4 - \alpha_1\alpha_2 - \alpha_3\alpha_4 \\ &= -(\alpha_1 - \alpha_4)(\alpha_2 - \alpha_3)\end{aligned}$$

and similarly

$$\theta_1 - \theta_3 = -(\alpha_1 - \alpha_3)(\alpha_2 - \alpha_4)$$
$$\theta_2 - \theta_3 = -(\alpha_1 - \alpha_2)(\alpha_3 - \alpha_4)$$

we see that the discriminant of the resolvent cubic is the *same* as the discriminant of the quartic $g(y)$, hence also as the discriminant of the quartic $f(x)$. Using our formula for the discriminant of the cubic, we can easily compute the discriminant in terms of p, q, r:

$$D = 16p^4r - 4p^3q^2 - 128p^2r^2 + 144pq^2r - 27q^4 + 256r^3$$

from which one can give the formula for D in terms of a, b, c, d:

$$\begin{aligned}D = &-128b^2d^2 - 4a^3c^3 + 16b^4d - 4b^3c^2 - 27a^4d^2 + 18abc^3 \\ &+ 144a^2bd^2 - 192acd^2 + a^2b^2c^2 - 4a^2b^3d - 6a^2c^2d \\ &+ 144bc^2d + 256d^3 - 27c^4 - 80ab^2cd + 18a^3bcd.\end{aligned}$$

The splitting field for the resolvent cubic is a subfield of the splitting field of the quartic, so the Galois group of

the resolvent cubic is a quotient of G. Hence knowing the action of the Galois group on the roots of the resolvent cubic $h(x)$ gives information about the Galois group of $g(y)$. We have the following cases:

1. Suppose first that the resolvent cubic is irreducible. If D is not a square, then G is not contained in A_4 and the Galois group of the resolvent cubic is S_3, which implies that the degree of the splitting field for $g(y)$ is divisible by 6. The only possibility is then $G = S_4$.

2. If the resolvent cubic is irreducible and D is a square, then G is a subgroup of A_4 and 3 divides the order of G (the Galois group of the resolvent cubic is A_3). The only possibility is $G = A_4$.

3. We are left with the case where the resolvent cubic is reducible. The first possibility is that $h(x)$ has 3 roots in F (i.e., splits completely). Since each of the elements $\theta_1, \theta_2, \theta_3$ is in F, every element of G fixes all three of these elements, which means $G \subseteq V$. The only possibility is $G = V$.

4. If $h(x)$ splits into a linear and a quadratic, then precisely one of $\theta_1, \theta_2, \theta_3$ is in F, say θ_1. Then G stabilizes θ_1 but not θ_2 and θ_3, so we have $G \subseteq D_8$ and $G \nsubseteq V$. This leaves two possibilities: $G = D_8$ or $G = C$. One way to distinguish between these is to observe that $F(\sqrt{D})$ is the fixed field of the elements of G in A_4. For the two cases being considered, we have $D_8 \cap A_4 = V, C \cap A_4 = \{1, (13)(24)\}$. The first group is transitive on the roots of $g(y)$, the second is not. It follows that the first case occurs if and only if $g(y)$ is irreducible over $F(\sqrt{D})$. We may therefore determine G completely by factoring $g(y)$ in $F(\sqrt{D})$, and so completely determine the Galois group in all cases. (cf. the exercises following and in the next section, where it is shown that over $\mathbb{Q}$ the Galois group cannot be cyclic of degree 4 if D is not the sum of two squares — so in particular if $D < 0$.)

11 Transcendental Galois Theorem

11.1 Topological Group

Definition 11.1 (Topological Group) *A set G equipped with a group structure and topology structure is called a* ***topological group*** *if*

$$m : G \times G \to G$$
$$inv : G \times G,\ g \mapsto g^{-1}$$

are both continuous. Here $G \times G$ is equipped with the product topology.

Remark 11.1.1 *It is often assumed that G is Hausdorff.*

Example:

1. $(\mathbb{R}, +)$ is a topological group equipped with the standard topology on $\mathbb{R}$.
2. $(\mathbb{R}^*, \times)$ is a topological group equipped with the standard topology on $\mathbb{R}$. $(\mathbb{Q}^*, \times)$ is a subgroup of $(\mathbb{R}^*, \times)$. This subgroup is neither open or closed.
3. $GL_n(\mathbb{R})$ is a topology group equipped with the subspace topology of $\mathbb{R}^{n^2}$.
4. For any group G we can equipped G with the discrete topology. Then G is a topological group.

Lemma 11.2 *Let G be a topological group, then*

$$L_g : G \to G,\ \ h \mapsto gh$$

is a homeomorphism.

Proof: The multiplication map $m : G \times G \to G$ is continuous. So $m : \{g\} \times G \to G$ is continuous. So L_g is continuous. Then Lg^{-1} being the inverse of L_g is continuous in the same way. Hence L_g is a homeomorphism. □

Lemma 11.3 *Let H be a subgroup of G. If H is open, then H is also closed. On the other hand, if H is closed and of finite index, then it is open.*

Proof: We consider $G = \bigsqcup gH$. Then if H is open, gH is also open since L_g is a heomoemorphism. Then $G \setminus H = \bigsqcup_{gH \neq H} gH$ is open. So it is closed. Similarly, if H is closed with finite index, then $G \setminus H$ is the finite union of closed sets, hence H is open. □

Example: $\{0\} \subset (\mathbb{R}, +)$ is closed but not open.

Definition 11.4 (Basis of Topology) *A* ***base*** *of a topological space X is a collection $\mathcal{B}$ of open subsets such that any open $U \subset X$ is a union of some subcollection of sets in $\mathcal{B}$. A* ***neighbourhood base*** *$\mathcal{N}$ of a point $x \in X$ is a collection of open subsets of X containing $x \in X$ such that any open set $U \subset X$ containing x must*

also contain some $N \in \mathcal{N}$.

Lemma 11.5 *The topology of X is determined by its base $\mathcal{B}$. Let G be a topological group, then the topology of G is determined by its neighbourhood base of e_G (the identity of G).*

Proof: We prove the second statement. Recall $m : G \times G \to G$ is continuous and $Lg : G \to G$, $x \mapsto gx$ is a homeomorphism. Then

$$\{gN \mid N \in \mathcal{N}\}$$

is a collection of open sets. Then we define

$$\mathcal{B} = \{gN \mid g \in G,\, N \in \mathcal{N}\}.$$

We claim this is a base of G. Let $U \in G$ be open, and $g \in U$. We consider $g^{-1}U \subset G$ which is open. Then $e \in g^{-1}U$. So we must have $N \subset g^{-1}U$ for some $N \in \mathcal{N}$. Then $gN \subset U$. Then it is clear that $\mathcal{B}$ is a base of G. □

Proposition 11.6 *Let G be a topological group and $\mathcal{N}$ be a neighbourhood base of e. Then*

1. *for any $N_1, N_2 \in \mathcal{N}$, there exists $e_3 \in N_3 \in \mathcal{N}$ such that $N_3 \subset N_1 \cap N_2$.*
2. *For any $N \in \mathcal{N}$, there exists $N' \in \mathcal{N}$ such that $N'N' \subset N$.*
3. *For any $N \in \mathcal{N}$, there exists $N' \in \mathcal{N}$ such that $N' \subset N^{-1}$.*
4. *For any $N \in \mathcal{N}$ and $g \in G$, there exists $N' \in \mathcal{N}$ such that $N' \subset gNg^{-1}$.*

Proof:

1. This follows from the definition of basis.
2. Since $m : G \times G \to G$ is continuous. Then we know $m^{-1}(N)$ is open in $G \times G$ and moreover $(e, e) \in m^{-1}(N)$. By product topology, $\mathcal{N} \times \mathcal{N}$ is a neighbourhood base of (e, e). Then we can find $U_1 \in \mathcal{N}$, $U_2 \in \mathcal{N}$ such that $U_1 \times U_2 \subset m^{-1}(N)$. Then take $N' \in \mathcal{N}$ such that $N' \subset U_1 \cap U_2$, then clearly we see $N'N' \subset U_1U_2 \subset N$.
3. Since $inv : G \to G$ is continuous. Then we know N^{-1} is open and contains e. So take any neighbourhood base element of e such that $N' \subset N^{-1}$.
4. Since $x \mapsto gxg^{-1}$ is a homeomorphism, as it is the composition $g \mapsto gx \mapsto gxg^{-1}$. Then gNg^{-1} is an open set around e. Thus we can find an $N' \subset gNg^{-1}$.

□

Proposition 11.7 *Let G be a group and $\mathcal{N}$ be a collection of (non-empty) subsets of G satisfying (1), (2), (3), (4) of Proposition (11.6). Then there is a unique topology on G making G a topological group with $\mathcal{N}$ being the neighbourhood base of e.*

Proof: The uniqueness: $\mathcal{B} = \{gN \,|\, g \in G,\, N \in \mathcal{N}\}$ uniquely determines the topology of G. Since by Lemma (11.5), the topology is uniquely determined by $\mathcal{N}$.

Existence: we first show $m : G \times G \to G$ is continuous and $G \to G$, $x \mapsto x^{-1}$ is continuous. Let $U \subset G$ be open. We need to show $m^{-1}(U)$ is open. Let $g_1, g_2 \in G$ be such that $g_1 g_2 = g \in U$. We show $m^{-1}(U)$ contains an open subset O in $G \times G$ containing (g_1, g_2). It suffices to show $m^{-1}(U)$ contains $(g_1 N_1, g_2 N_2)$ for some $N_1, N_2 \in \mathcal{N}$. Since $g_1 g_2 \mathcal{N}$ is a neighbourhood base of G, we can find $N \in \mathcal{N}$ such that $g_1 g_2 N \subset U$. By (2), we can find $N' \subset \mathcal{N}$, such that $N'N' \subset N$. So $g_1 g_2 N' N' \subset g_1 g_2 N \subset N$. By (4), we can find $N_1 \in \mathcal{N}$ such that $N_1 \subset g_2 N' g_2^{-1}$. Then $g_1 N_1 g_2 N' \subset g_1 g_2 N' g_2^{-1} g_2 N' \subset U$. So we have $(g_1 N_1, g_2 N') \subset m^{-1}(U)$, i.e., multiplication on G is continuous.

We also show taking inverse of elements in g is continuous. Given open set $U \subset G$. Since we know multiplication is continuous on G, then L_g is a homeomorphism. So WLOG, we may let U be an open set containing 0. Then there exists $N' \in \mathcal{N}$ such that $N' \subset U^{-1}$ as desired. □

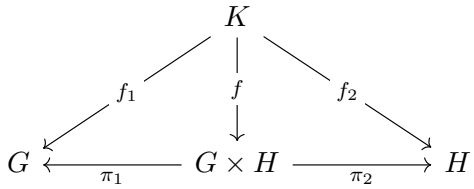
Next we consider the category of topological groups:

- **Object:** topological groups;
- **Morphisms:** continuous group homormophisms.

Lemma 11.8 *Let G and H be topological groups. Then $G \times H$ is a topological group with the product topology. In particular, $G \times H$ is the categorical product.*

Remark 11.8.1 *In other words, finite product exists in the category of topological groups. Moreover, it agrees with the product in the category of groups equipped with the natural product topology.*

Proof: We know $G \times H$ is a group under the categorical product. We equipped $G \times H$ with the the coarsest topology such that the projection maps π_1 and π_2 are continuous



To be precise, the open sets of $G \times H$ are given by finite intersections and arbitrary unions of

$$\{G \times U \,|\, U \text{ open}\} \quad \text{and} \quad \{V \times H \,|\, V \text{ open}\}$$

In other words, this is just the product topology. Then one can check that any morphism f_1 and f_2 from the topological group K to topological group G and H respectively induces a unique morphism (f_1, f_2) from K to $G \times H$. So $G \times H$ satisfies the universal property of products of topological groups.

Lastly we show $G \times H$ is a topological group. We show the multiplication map

$$m : G \times H \times G \times H \longrightarrow G \times H$$

$$(g_1, h_1, g_2, h_2) \longmapsto (g_1 g_1, h_1 h_2)$$

is continuous. Suffices to show that projection is continuous. But then it follows from the continuity of multiplication on G and H respectively. Similarly, we can show inverse is continuous. □

Proposition 11.9 *The infinite product exists in the category of topological groups.*

Proof: For topological groups, G_i, $i \in I$. We define

$$\prod_{i \in I} G_i.$$

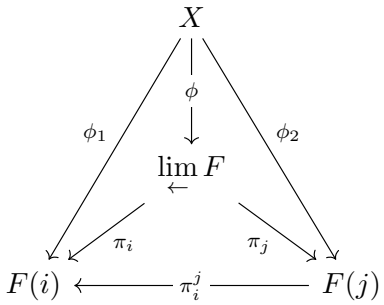
Then we check $\prod_{i \in I} G_i$ with the natural group and natural topological structure (product topology) is a topological group satisfying the universal property of products in the category of topological groups. □

Example: We consider the product $\prod_{i=1}^{\infty} \mathbb{Z}/2\mathbb{Z}$ in the category of topological groups. However, this space is still Hausdorff, since each component space is Hausdorff.

11.2 Inverse Limit

Definition 11.10 (Inverse Limit)

1. *A **direct poset** is a partially ordered set $(I, \leq)$ such that for any $i, j \in I$, there is $k \in I$ such that $i \leq k$, $j \leq k$. We want to consider $(I, \leq)$ as a category:*
   - *Objects: $i \in I$.*
   - *Morphisms: $\pi_i^j : j \to i$ if $i \leq j$.*
2. *Let $(I, \leq)$ be a directed poset considered as a category. Then a functor $F : I \to Grp$ is an **inverse system of groups**. Similarly, we can define an inverse system of topological space and topological groups.*
3. *The **inverse limit** of $F : I \to Grp$ is a group $\varprojlim F$ together with morphisms $\pi_i : \varprojlim F \to F(i)$ making the diagram commutes*



Moreover, suppose X is any group with morphism $\phi_i : X \to F(i)$, then there is a morphism $\phi : X \to \varprojlim F(i)$ making the diagram commutes. Similarly, we can define the inverse limit for the category of topological space and topological groups.

Example: Consider $(I, \leq) = (N, \leq)$. then we have an inverse system

$$0 \quad \leq \quad 1 \quad \leq \quad 2 \quad \leq \quad 3$$

$$\mathbb{Z} \longleftarrow \mathbb{Z} \longleftarrow \mathbb{Z} \longleftarrow \mathbb{Z} \longleftarrow$$

Now suppose each morphism between the group $\mathbb{Z}$ is just the identity map, then

$$\varprojlim \mathbb{Z} = \mathbb{Z}.$$

On the other hand, if each morphism between the group $\mathbb{Z}$ is the trivial map, then

$$\varprojlim \mathbb{Z} = \{e\}.$$

Example: We again consider $(I, \leq) = (\mathbb{N}, \leq)$. Then we have an inverse system

$$\{e\} \xleftarrow[\pi_1]{} \mathbb{Z}/p\mathbb{Z} \xleftarrow[\pi_2]{} \mathbb{Z}/p^2\mathbb{Z} \longleftarrow$$

Where π_i is the projection map.

Lemma 11.11 *The inverse limit exists in the category of groups, the category of topological spaces, and the category of topological groups.*

Proof: We only prove for the category of groups. Let $F : I \to Grp$, we first consider the infinite product

$$\prod_{i \in I} F(i).$$

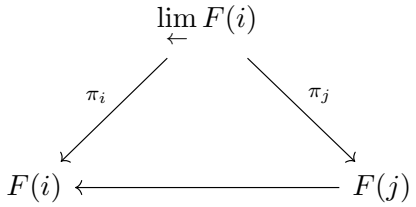
Then we claim

$$\varprojlim F = \varprojlim F(i) \cong \left\{ (g_i) \in \prod_{i \in I} F(i) \mid \pi_i^j(g_j) = g_i \right\}$$

where

$$\pi_i^j : F(j) \to F(i).$$

Then we check the following diagram commutes, where π_i is the projection onto the i^{th} coordinate.



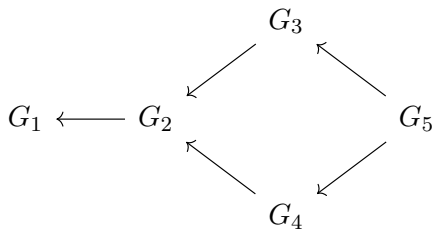
Since $\pi_j((g_n)) = g_j$ and $\pi_i((g_n)) = g_i$ and $\pi_i^j(g_j) = g_i$, then we see that the diagram commutes.
Similarly, suppose X is a group with morphisms ϕ_i mapping into $F(i)$ satisfying the compatibility criterion, then the map

$$\phi : X \to \varprojlim F(i), \quad X \mapsto (\phi_i(X))_{i \in I}$$

is the desired map for the universal property.

The same proof works for category of topological spaces, and the category of topological groups. □

Example: we consider the inverse system



Then

$$\lim_{\leftarrow} G_i = G_5 = \{(g_i) \in G_1 \times \cdots \times G_5 \,|\, \pi_i^j(g_i) = g_i\}.$$

But then (g_i) is completely determined by g_5 and any $g_5 \in G_5$ corresponds to an element in the set on the right. So the inverse limit is just G_5.

11.3 Infinite Galois Extension

Recall if E/F is finite and Galois. Then we have a bijection

$$\{F \subset K \subset E\} \longleftrightarrow \{H \subset \mathrm{Gal}(E/F)\}$$

by

$$\begin{aligned} K &\longrightarrow \mathrm{Gal}(E/K) \\ E^H &\longleftarrow H. \end{aligned}$$

However, this is not always correct for infinite Galois extension as we have the following example:
Example: Let $E/\mathbb{Q}$ be the splitting field of $\{x^2 - p \,|\, p \text{ is a prime}\}$. Then $E = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \cdots\}$. So $E/\mathbb{Q}$ is of countable and of infinite dimension over $\mathbb{Q}$. We have a spanning set $\{1, \sqrt{2}, \sqrt{3}, \cdots\}$. Note every element inside $G = \mathrm{Gal}(E/\mathbb{Q})$ must be of order 2, in particular, they are given by

$$\sqrt{p} \mapsto \pm\sqrt{p}.$$

Note that $\mathrm{Gal}(E/\mathbb{Q})$ is abelian as well. We consider G as $\mathbb{F}_2$-vector space. Then $\mathrm{Hom}_{\mathbb{F}_2}(G, \mathbb{F}_2) \cong \mathrm{Hom}(G, \mathbb{Z}/2\mathbb{Z})$ is uncountable. There are uncountable many $H \subset G$ such that $[G : H] = 2$. On the other hand, there are only countably many degree 2 extensions over $\mathbb{Q}$.

Now let E/F be a Galois extension. For any finite intermediate Galois extensions $F \subset K \subset E$, we have a short exact sequence

$$1 \to \mathrm{Gal}(E/K) \to \mathrm{Gal}(E/F) \to \mathrm{Gal}(K/F) \to 1.$$

Let I be the set of all finite intermediate Galois extensions ordered by inclusions ($\leq$).

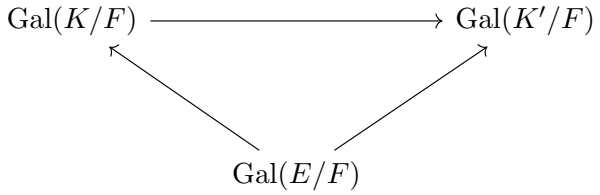
Lemma 11.12 *$(I, \leq)$ is a directed poset.*

Proof: It is clearly this is a poset. We show it is directed. Let $K_i \in I$, $K_2 \in I$. Then K_1K_2/F is finite and Galois. Then $K_1, K_2 \subset K_1K_2$ and $K_1K_2 \in I$. □

Now we have an inverse system of Galois groups ($K \to K'$ in I if $K' \subset K$) by

$$\mathrm{Gal}(K/F) \to \mathrm{Gal}(K'/F)$$

Where the morphism is just the restriction map. We note that the diagram commutes



where the morphism is again just taking the restriction. So it is natural to guess that $\mathrm{Gal}(E/F)$ is the inverse limit of this system. This is indeed the case, as we have Theorem (11.14).

Lemma 11.13 *We have $E = \bigcup_{K \in I} K$.*

Proof: we have $E \supset \bigcup_{k \in I} K$. Let $\alpha \in E$. Then we know α is algebraic. Then let K be the splitting field of α over F. Then K/F is finite and Galois. □

Theorem 11.14 *We have an isomorphism of groups*

$$\mathrm{Gal}(E/F) \cong \varprojlim_{K \in I} \mathrm{Gal}(K/F).$$

Proof: We have

$$\varprojlim_{K \in I} \mathrm{Gal}(K/F) \subset \prod_{K \in I} \mathrm{Gal}(K/F)$$

and

$$\varprojlim_{K \in I} \mathrm{Gal}(K/F) = \{(g_K) \in \prod_{K \in I} \mathrm{Gal}(K/F) \,|\, \pi(g_K) = g_{K'},\, \pi : \mathrm{Gal}(K/F) \to \mathrm{Gal}(K'/F)\}$$

We define

$$\mathrm{Gal}(E/F) \to \varprojlim_{K \in I} \mathrm{Gal}(K/F)$$

by $g \mapsto (g_K)$, where g_K is the restriction of g to K. This is a well-defined group homomorphism. We first show this is injective. Let $g \in \mathrm{Ker}$. Then $g_K = e_K \in \mathrm{Gal}(K/F)$. For any $\alpha \in E$, we have $\alpha \in K$ for some $K \in I$, this

shows $g(\alpha) = e$ for each $\alpha \in E$, i.e., g is the trivial map.

Next, we show surjectivity. Let $(g_K) \in \varprojlim_{K \in I} \mathrm{Gal}(K/F)$. We define $g \in \mathrm{Gal}(E/F)$ by $g(\alpha) = g_K(\alpha)$ if $\alpha \in K$. We check g is well-defined. Assume $\alpha \in K$ and $\alpha \in K'$. Then by definition, we have $g(\alpha) = g_K(\alpha)$ and $g(\alpha) = g'_K(\alpha)$. So we must show $g_K(\alpha) = g_{K'}(\alpha)$. Note that $K \cap K'$ is again finite and Galois over F with $\alpha \in K \cap K'$. Then we have

$$g_K(\alpha) = g_{K \cap K'}(\alpha) = g_{K'}(\alpha).$$

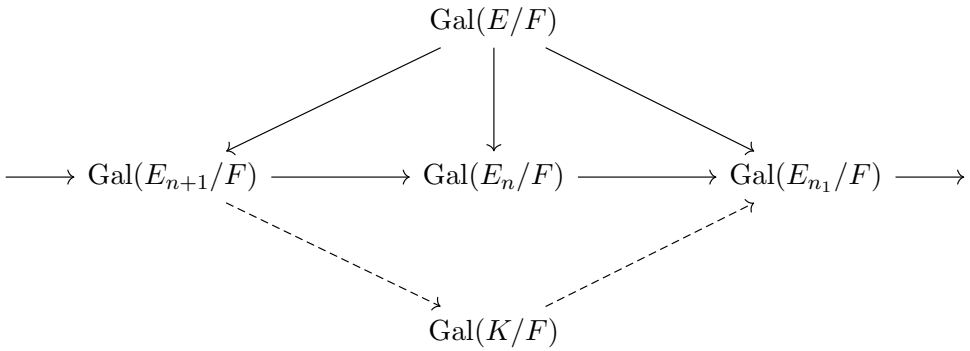
Where the equalities follows from the compatibility of morphisms in the inverse system. Hence $g \in \mathrm{Gal}(E/K)$ is well-defined. □

Remark 11.14.1 *If E/F is finite, then*

$$\mathrm{Gal}(E/F) = \varprojlim_{K \in I} \mathrm{Gal}(K/F)$$

Also holds.

Example: Let $F = \mathbb{Q}(x_1, x_2, \cdots)$ with $\mathbb{N}$-many variables. Let $E = \mathbb{Q}(\sqrt{x_1}, \sqrt{x_2}, \cdots)$. Then E/F is Galois. We claim $\mathrm{Gal}(E/F) \cong \prod_{i=1}^{\infty} \mathbb{Z}/2\mathbb{Z}$. We consider $E_n = F(\sqrt{x_1}, \cdots, \sqrt{x_n})$. Then E_n/F is finite Galois with $\mathrm{Gal}(E_n/F) = \prod_{i=1}^{n} \mathbb{Z}/2\mathbb{Z}$. Then we have an (sub) inverse system



Since the poset is directed, then

$$\begin{aligned} \mathrm{Gal}(E/F) &= \varprojlim_{K \in I} \mathrm{Gal}(K/F) \\ &= \varprojlim_{n} \mathrm{Gal}(E_n/F) \\ &= \prod_{i=1}^{\infty} \mathbb{Z}/2\mathbb{Z}. \end{aligned}$$

Note that

$$H = \bigoplus_{i=1}^{\infty} \mathbb{Z}/2\mathbb{Z} \subset \prod_{i=1}^{\infty} \mathbb{Z}/2\mathbb{Z}$$

where

$$\bigoplus_{i=1}^{\infty} \mathbb{Z}/2\mathbb{Z} = \{(g_i) \,|\, g_i \neq e \text{ for finitely many } i\}.$$

Then $E^{\mathrm{Gal}(E/F)} = F$, but we also have $E^H = F$.

Example: let p be a prime, we consider the cyclotomic extensions $\mathbb{Q}(\xi_{p^n})$. Then we have a tower of extensions

$$\mathbb{Q} \subset \mathbb{Q}(\xi_p) \subset \mathbb{Q}(\xi_{p^2}) \subset \cdots$$

Then we define $\mathbb{Q}(\xi_{p^\infty}) = \bigcup_{n=0}^{\infty} \mathbb{Q}(\xi_{p^n})$. We note $\mathbb{Q}(\xi_{p^\infty})$ is a Galois Extension over $\mathbb{Q}$. Moreover

$$\mathrm{Gal}(E/\mathbb{Q}) = \varprojlim \mathrm{Gal}(\mathbb{Q}(\xi_{p^n})/\mathbb{Q}) = \varprojlim (\mathbb{Z}/p^n\mathbb{Z})^*.$$

Then

$$\begin{aligned}\mathrm{Gal}(E/\mathbb{Q}) &= \left\{(a_n) \,|\, a_n \in (\mathbb{Z}/p^n\mathbb{Z})^*,\, a_n \equiv a_{n-1} \mod p^{n-1}\right\} \\ &= \left\{\sum_{i=0}^{\infty} b_i p^i \,|\, b_i \in \{0, 1, \cdots, p-1\},\, b_0 \neq 0\right\}\end{aligned}$$

Remark: we can also consider the inverse limit of rings

$$\mathbb{Z}/p\mathbb{Z} \longleftarrow \mathbb{Z}/p^2\mathbb{Z} \longleftarrow \mathbb{Z}/p^3\mathbb{Z} \longleftarrow$$

$$\varprojlim_n \mathbb{Z}/p^n\mathbb{Z} = \mathbb{Z}_p \supset \mathbb{Z}.$$

But in $\mathbb{Z}_p$, $1 - p$ is invertible.

$$\frac{1}{1-p} = \sum_{i=0}^{\infty} p^i \text{ in } \mathbb{Z}_p.$$

We consider the extensions $\mathbb{Q}(\xi_{2^\infty})/\mathbb{Q}$. Then

$$\mathrm{Gal}(\mathbb{Q}(\xi_{2^\infty})/\mathbb{Q}) \cong \varprojlim \mathrm{Gal}(\mathbb{Q}(\xi_{2^n})/\mathbb{Q}).$$

Then we consider the elements $\sigma_5, \sigma_{13} \in \mathrm{Gal}(\mathbb{Q}(\xi_{2^\infty})/\mathbb{Q})$. Where

$$\sigma_5(\xi_{2^n}) = \xi_{2^n}^5$$

this is well-defined, since 5 is in $(\mathbb{Z}/2^n\mathbb{Z})^*$ for any n. Similarly, we can define σ_{13}. Let $H = \langle\sigma_5\rangle$ and $K = \langle\sigma_{13}\rangle$. We claim that $H \cong \mathbb{Z} \cong K$. It is clear that K and H are cyclic. Then take n large enough, we see that K and H are not of finite order. We claim that $H \neq K$ in $\mathrm{Gal}(\mathbb{Q}(\xi_{2^\infty})/\mathbb{Q})$. This is the case since $\sigma_5 \neq \sigma_{13}$ and $\sigma_5 \neq \sigma_{13}^{-1}$. We define $H_n = \langle\sigma_5\rangle$ in $\mathrm{Gal}(\mathbb{Q}(\xi_{2^n})/\mathbb{Q})$ and define K_n in the similar way. We claim $H_n = K_n$ for any n, if this is the case, then

$$L_n = \mathbb{Q}(\xi_{2^n})^{H_n} = \mathbb{Q}(\xi_{2^n})^{K_n}.$$

Then we must have

$$L = \mathbb{Q}(\xi_{2^\infty})^H = \mathbb{Q}(\xi_{2^\infty})^K$$

even if $H \neq K$. Since if $x \in \mathbb{Q}(\xi_{2^\infty})^H$, then $x \in \mathbb{Q}(\xi_{2^\infty}) = \bigcup \mathbb{Q}(\xi_{2^n})$. So $x \in \mathbb{Q}(\xi_{2^n})$, and $x \in \mathbb{Q}(\xi_{2^n})^{H_n} = \mathbb{Q}(\xi_{2^n})^{K_n} \subset \mathbb{Q}(\xi_{2^\infty})^K$. Similarly, we have the other direction.

It remains to show that $H_n = K_n$. When $n = 1$,

$$H_1 = \langle \sigma_5 \rangle = \langle \sigma_1 \rangle = \langle \sigma_{13} \rangle = K_1.$$

Definition 11.15 (Krull Topology) *Let E/F be a Galois. We define the **Krull Topology** on $\mathrm{Gal}(E/F)$ by considering*

$$\mathrm{Gal}(E/F) \cong \varprojlim_{K \in I} \mathrm{Gal}(K/F)$$

In the category of topological group. Here $\mathrm{Gal}(K/F)$ is a discrete finite group.

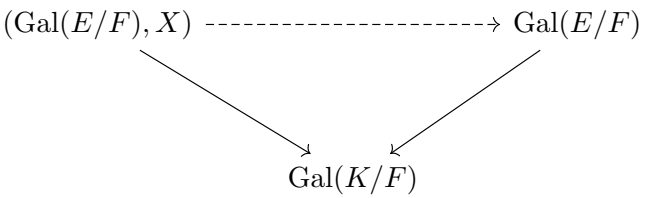
Lemma 11.16 *The Krull topology on $\mathrm{Gal}(E/F)$ is the coarsest topology such that all $\mathrm{Gal}(E/F) \to \mathrm{Gal}(K/F)$ are continuous.*

Proof: Recall that if X and Y are two topology on $\mathrm{Gal}(E/F)$, then Y is coarser than X if $Y \subset X$. Also recall the topology on $\mathrm{Gal}(E/F)$ is defined by the restriction of the product topology on $\prod_{k \in I} \mathrm{Gal}(K/F) \xrightarrow{\pi_k} \mathrm{Gal}(K/F)$. So the open sets are precisely unions of finite intersection of $\pi_k^{-1}(U_k)$. This is also the minimal requirepemnt for the maps $\mathrm{Gal}(E/F) \xrightarrow{\pi_k} \mathrm{Gal}(K/F)$ to be continuous. So this is the coarsest.

Alternatively, Let X, Y be two topology on $\mathrm{Gal}(E/F)$. We can say that Y is coarser than X if the identity map

$$(\mathrm{Gal}(E/F), X) \to (\mathrm{Gal}(E/F), Y)$$

is continuous. So if $(\mathrm{Gal}(E/F), Z)$ is the coarsest, then it is the terminal object (in the suitable category). Now let $(\mathrm{Gal}(E/F), X)$ be a topological group such that



there exists compatible continuous homomorphisms from $(\mathrm{Gal}(E/F), X)$ to $\mathrm{Gal}(K/F)$. Then it must factor through the inverse limit. Hence the topology on the inverse limit is coarsest. □

Definition 11.17 (Profinite) *A topological group is called **profinite** if it is the inverse limit of finite groups (with discrete topology).*

Lemma 11.18 *A topological group G is Hausdorff if and only if $\{e\}$ is closed.*

Proof: Suppose G is Hausdorff, then singletons are closed. On the other hand, we know G is Hausdorff if nad only if $\Delta_G \subset G \times G$ (the diagonal of G) is closed. We consider the continuous map

$$p : G \times G \to G \,:\, (x, y) \mapsto xy^{-1}.$$

Then if $\{e\}$ is closed $p^{-1}(\{e\}) = \Delta_G$ is closed. □

Proposition 11.19 $\mathrm{Gal}(E/F)$ *is compact, hence it is not discrete if* $\mathrm{Gal}(E/F)$ *is infinite.*

Proof: By Tychonoff theorem, the product $\prod_{K \in I} \mathrm{Gal}(K/F)$ is compact. We show $\mathrm{Gal}(E/F)$ is closed in the product topology, hence it will be compact. Let K/F be finite and Galois, we know there are only finitely many intermediate extensions $F \subset K' \subset K$; we denote this set by I_K. WLOG, we assume $|I_K| = 3$ and $I_K = \{F, L, K\}$. Then for any $g_K \in \mathrm{Gal}(K/F)$, we then write g_L for $g_K|_L$ for $F \subset L \subset K$. Then we define

$$V(g_K) = \bigcap_{L \in I_K} \pi_L^{-1}(g_L) = \{(g_L, g_K, \cdots)\}$$

which is closed as it is the arbitrary intersection of closed sets. Then we define

$$C_K = \bigcup_{g_K \in \mathrm{Gal}(K/F)} V(g_K)$$

which is the finite union of closed sets, which is closed. Lastly,we note that $\mathrm{Gal}(E/F) = \bigcap_{k \in I} C_k$ hence is closed. □

Theorem 11.20 $\mathrm{Gal}(E/F)$ *is Hausdorff, compact compact, totally disconnected.*

Proof: We show $\mathrm{Gal}(E/F)$ is Hausdorff by showing $\{e\}$ is closed.

$$\mathrm{Gal}(E/F) \cong \varprojlim_{k \in I} \mathrm{Gal}(K/F) \xrightarrow{\pi_K} \mathrm{Gal}(K/F)$$

Then we claim $\{e\} = \bigcap_{K \in I} \pi_K^{-1}\{e_K\} = (e_K)_K$, this would show $\{e\}$ being the arbitrary intersection of closed set is closed.

Compactness follows from Proposition (11.19).

To show totally disconnectedness, we know that the product of totally disconnected topological spaces are totally disconnected. Hence $\prod_K \mathrm{Gal}(K/F)$ is totally disconnected. We also know the subspace of a totally disconnected topological space is totally disconnected, hence $\mathrm{Gal}(E/F)$ is totally disconnected. □

11.4 Infinite Galois Correspondence

Definition 11.21 *Let E/F be Galois with the Galois group G. Then for any finite subset $S \subset E$, we define*

$$G(S) = \{g \in G \,|\, gs = s, \forall s \in S\}.$$

Lemma 11.22 *Let E/F be Galois with $\mathrm{Gal}(E/F) = G$. Let $F \subset K \subset E$ be such that K/F is Galois and finite. We have a natural projection map*

$$\pi : \mathrm{Gal}(E/F) \to \mathrm{Gal}(K/F)$$

by restriction. Then

$$\ker(\pi) = \mathrm{Gal}(E/K) = G(K) = G(S)$$

for some finite $S \subset K$.

Proof: Clearly $\mathrm{Gal}(E/K) = \ker(\pi)$, and $\mathrm{Gal}(E/K)$ is precisely the subgroup of $\mathrm{Gal}(E/F)$ that fixes K. Now since K/F is finite, then $K = F(\alpha_1, \cdots, \alpha_n)$. Hence $G(K) = G(\alpha_1, \cdots, \alpha_n)$. □

Lemma 11.23 *Let $S \subset E$ be finite. Then $G(S)$ is an open and closed subgroup of $G = \mathrm{Gal}(E/F)$. In particular, the stabilizer $G(x) = G_x$ for any $x \in E$ is both closed and open.*

Proof: Note that $G(S \cup S') = G(S) \cap G(S')$. So it suffices to show $G(y) = G_y$ is both open and closed, where $y \in E$. Let let K be the splitting field of $m_{y,F}(x)$. Then K/F is finite and Galois. Then we have

$$1 \longrightarrow \mathrm{Gal}(E/K) \longrightarrow \mathrm{Gal}(E/F) \xrightarrow{\pi} \mathrm{Gal}(K/F) \longrightarrow 1$$

Let $H = \mathrm{Gal}(K/F)$, $H_y = \{h \in H : hy = y\}$. Then we have $G_y = \pi^{-1}(H_y)$. Now H_y is open and closed, since $\mathrm{Gal}(K/F)$ is discrete and H_y is a subgroup of $\mathrm{Gal}(K/F)$. Then since π is continuous, $\pi^{-1}(H_y)$ is both open and closed. □

Corollary 11.23.1 *Let E/F be Galois. Let $F \subset K \subset E$ be such that K/F is finite and Galois. Then*

$$\mathrm{Gal}(E/K) \subset \mathrm{Gal}(E/F)$$

is both open and closed.

Proof: This is because $\mathrm{Gal}(E/K) = G(K) = G(S)$ for some finite subset S of K. □

Corollary 11.23.2 *Let E/F be Galois. Let $F \subset K \subset E$ be any intermediate extension. Then*

$$\mathrm{Gal}(E/K) \subset \mathrm{Gal}(E/F)$$

is closed.

Proof: We have $\mathrm{Gal}(E/K) = G(K) = \bigcap_{x \in K} G_x$ which is the arbitrary intersection of closed sets. Hence $\mathrm{Gal}(E/K)$ is closed. □

Theorem 11.24 *Let E/F be Galois and $\mathrm{Gal}(E/F) = G$. Then $\mathcal{B} = \{H = \mathrm{Gal}(E/K) \mid K/F \text{ is finite and Galois}\}$ forms a neighbourhood base of $e \in G$.*

Remark 11.24.1 *We note that elements of $\mathcal{B}$ are open and closed in $\mathrm{Gal}(E/F)$ by Corollary (11.23.1).*

Proof: Let $U \subset G$ be open and $e \in U$. Since the open sets in G are unions of finite intersections of $\pi_K^{-1}(U_K)$, where $U_K \subset \mathrm{Gal}(K/F)$. We can assume U is a finite intersection (basis element). We write

$$U = \bigcap_{K \in J \subset I} \pi_K^{-1}(U_K)$$

for some finite $J \subset I$, which is the set of all finite Galois extension over F.

since $e \in U$, we must have $e \in \pi_K^{-1}(U_K)$. Therefore, $\pi_K^{-1}(U_K) \supset \pi_K^{-1}(e_K) = \mathrm{Gal}(E/K)$. So we have

$$e \subset \bigcap_{K \in J \subset I} \mathrm{Gal}(E/K) \subset \bigcap_{K \in J \subset I} \pi_K^{-1}(U_k) = U.$$

Now let $L = \prod_{K \in J} K$. Then L being the finite product of Galois extensions is Galois. Then $\mathrm{Gal}(E/L) = \bigcap_{k \in J} \mathrm{Gal}(E/K)$ so

$$e \subset \mathrm{Gal}(E/L) = \bigcap_{k \in J} \mathrm{Gal}(E/K) \subset U.$$

This proves that $\mathcal{B}$ is a neighbourhood basis of $e \in G$. □

We now give another definition of the Krull topology.

Lemma 11.25 *Let E be a field. Let $G \subset \mathrm{Aut}(E)$. We consider the action map $G \times E \to E$. The stabilizer of each element of E is open in G if and only if $G \times E \to E$ is continuous, where E is discrete.*

Proof: Let $z \in E$. Then f is continuous if and only if $f^{-1}(z)$ is open. We have

$$f^{-1}(z) = \{(g, x) \mid gx = z\} = \bigcup_{x \in E} \{(g, x) \mid gx = z\}.$$

Then for fixed $x \in E$, we have either

$$\{(g, x) \mid gx = z\} = \emptyset$$

or $\{(g, x) \mid gx = z\}$ is hoemomorphic to the set $\{hg \mid gx = x\} \subset G$, where $hx = z$ for some $h \in G$. And clearly this set is just homeomorphic to $\{g \mid gz = z\}$. Then $f^{-1}(z)$ is open if $\{(g, z) \mid gz = z\}$ is open, which happens if the stabilizers $\{g \in G \mid gz = z\}$ are open. Conversely, suppose $G \times E$ is continuous, then the stabilizer G_x of an element $x \in E$ is contained in $f^{-1}(x)$ which is open. But since the second coordinate (E) is equipped with the discrete topology, this shows that G_x is open in G. □

Corollary 11.25.1 *Let E/F be Galois. We equip E with the discrete topology. Then $\mathrm{Gal}(E/F) \times E \to E$ is continuous. In particular, the Krull topology is the weakest one such that the action is continuous.*

Lemma 11.26 *Let G be a topological group with a subgroup H. Then $\bar{H}$ is a topological subgroup as well.*

Proof: Suffices to show that maps $m : \bar{H} \times \bar{H} \to \bar{H}$ and taking inverse $\bar{H} \to \bar{H}$ are well-defined. Since the inverse map : $G \to G$ is a homeomorphism, then it must map a closed set to a closed set. Hence it must map $\bar{H}$ to $\bar{H}$.

We show that $m : \bar{H} \times \bar{H} \to \bar{H}$. We know that $m : H \times H \to H$. Then $m^{-1}(\bar{H})$ is closed and contains $H \times H$. Then

$$m^{-1}(\bar{H}) \supset \overline{H \times H} = \bar{H} \times \bar{H}.$$

□

Proposition 11.27 *Let E/F be Galois with $G = \mathrm{Gal}(E/F)$. Let $H \subset G$ be a subgroup. Then*

$$\mathrm{Gal}(E/E^H) = \bar{H}.$$

Proof: We know $\mathrm{Gal}(E/E^H)$ is closed and contains H. Hence $\mathrm{Gal}(E/E^H) \supset \bar{H}$. Now let $g \in G \setminus \bar{H}$. we show that g doesn't fix E^H. Recall $\{g\,\mathrm{Gal}(E/K) \mid K/F \text{ is finite and Galois}\}$ forms a neighbourhood base of $g \in G$. Since $g \in G \setminus \bar{H}$, so there are open U such that $g \in U$ and $U \cap \bar{H} = \emptyset$. In particular, we can take a basis element, so there exists $\mathrm{Gal}(E/K)$ such that $g\,\mathrm{Gal}(E/K) \cap \bar{H} = \emptyset$, so $\mathrm{Gal}(E/K) \cap g^{-1}\bar{H} = \emptyset$. We consider the map

$$\mathrm{Gal}(E/K) \hookrightarrow G \xrightarrow{\pi} \mathrm{Gal}(K/F)$$

Then $e_K \notin \pi(g^{-1}\bar{H})$, so $\pi(g) \notin \pi(\bar{H})$. Then

$$K^{\pi(H)} = K^H \subset E^H$$

Now since K/F is finite and Galois and $\pi(g) \notin \pi(\bar{H}) \supset \pi(H)$. g cannot fix everything in K^H by finite Galois Correspondence, thus g does not fix E^H. This concludes the proof of the proposition. □

Theorem 11.28 (Galois Correspondence) *Let E/F be Galois. Then have an order reversing bijection between*

$$\{\textit{intermediate field}\, K,\ F \subset K \subset E\} \longleftrightarrow \{\textit{closed subgroup}\, H \subset \mathrm{Gal}(E/F)\}$$

by

$$K \longmapsto \mathrm{Gal}(E/K)$$
$$E^H \longleftarrow H$$

Proof: Order Reversing is clear. We show this is indeed a bijection. We know that for any intermediate field $K \subset E$, $\mathrm{Gal}(E/K)$ is closed by Corollary (11.23.2) and we know that the map $K \mapsto \mathrm{Gal}(E/K)$ is always injective. Now we show that the reverse map is injective. For a closed subgroup H, we know that $\mathrm{Gal}(E/E^H) = \overline{H} = H$ by Proposition (11.27), then if $H \neq H'$ are two closed subgroups, then $\mathrm{Gal}(E/E^H) \neq \mathrm{Gal}(E/E^{H'})$, hence $E^H \neq E^{H'}$. □

Theorem 11.29 *Let E be a field and $F = E^G$ for some $G \subset \mathrm{Aut}(E)$. Assume G is equipped with a topological group structure such that $G \times E \to E$ is continuous (E discrete) and G is compact. Then E/F is Galois with the Galois group isomorphic to G.*

Proof: Since $G \times E \to E$ is continuous, then the stabilizer of each element of E is open. The idea is to write E as the union of finite Galois extensions.

We first claim that the number of orbit Gx is finite. The image of the compact set $G \times \{x\}$ is compact. The only compact sets in E with discrete topology are finite sets. Hence Gx is finite.

Let $Gx = \{x_1, \cdots, x_n\}$. We define $N_x = \bigcap_{i=1}^{n} G_{x_i}$. This is an open subgroup since the action $G \times E \to E$ is continuous, so the stabilizers are continuous. One can verify that N_x is normal, since G acts on $\{x_1, \cdots, x_n\}$ and N_x is the kernel of this action. Note that G/N_x is a finite subgroup of S_n.

- We claim G/N_x acts $K_x = F(x_1, \cdots, x_n)$ with the fixed point subfield F, so $G/N_x \subset \mathrm{Aut}(K_x/F)$. This is because $K_x^G \subset E^G = F$. Then since everything is finite, we know K_x/F is Galois with the Galois group G/N_x.
- Now we see that $E = \bigcup_{x \in E} K_x$ is a union of finite Galois extensions over F, then E/F is Galois since it is normal and splitting.

Note that $\{K_x\}$ forms a directed system. This is because K_xK_y is a finite Galois extension, hence simple. So we must have $K_xK_y = K_z$ for some z. In particular, any finite Galois extension E/F must be equal to some K_x. We see that

$$\mathrm{Gal}(E/F) \cong \varprojlim_{x \in E} \mathrm{Gal}(K_x/F)$$

as topological groups. Next, one can verify that

- The natural map $G \to \mathrm{Gal}(E/F) = \mathrm{Aut}(E/F)$ is injective. Since $G \subset \mathrm{Aut}(E)$ and G fixes F.
- The natural map $G \to \mathrm{Gal}(E/F)$ is continuous. Since we know $G \times E \to E$ is continuous and $\mathrm{Gal}(E/F) \times E$ is continuous and $\mathrm{Gal}(E/F)$ has the coarsest topology for this to be continuous. Then $G \to \mathrm{Gal}(E/F)$ is continuous.
- The natural map $G \to \mathrm{Gal}(E/F)$ is closed. So the image is a closed subgroup. Recall $\mathrm{Gal}(E/F)$ is Hausdorff and compact in the Krull topology. Since G is compact, any closed subset U of G is compact. The image of the compact set U is compact in $\mathrm{Gal}(E/F)$. And compact set in the Hausdorff space $\mathrm{Gal}(E/F)$ must be closed.
- The natural map $G \to \mathrm{Gal}(E/F)$ is surjective. Since by Galois correspondence $\mathrm{Gal}(E/F) = \bar{G}$ and G is closed in $\mathrm{Gal}(E/F)$, so $\bar{G} = G$.

□

11.5 Transcendental Extensions and Inseparable extensions

Definition 11.30 (Transcendence Base) *A subset $\{a_1, a_2, \cdots, a_n\}$ of E is called **algebraically independent** over F if there is no nonzero polynomial $f(x_1, x_2, \cdots, x_n) \in F[x_1, x_2, \cdots, x_n]$ such that $f(a_1, a_2, \cdots, a_n) = 0$. An arbitrary subset S of E is called **algebraically independent over** F if every finite subset of S is algebraically independent. The elements of S are called **independent transcendentals** over F. A **transcendence base** for E/F is a maximal subset of E which is algebraically independent over F.*

Remark 11.30.1 *Note that if E/F is algebraic, the empty set is the only algebraically independent subset of E. In particular, elements of an algebraically independent set are necessarily transcendental. Moreover, one can easily check that $S \subset E$ is an algebraically independent set over F if and only if each $s \in S$ is transcendental over $F(S \setminus \{s\})$. One can also show that S is a transcendence base for E/F if and only if S is a set of algebraically independent transcendentals over F and E is algebraic over $F(S)$.*

Remark 11.30.2 *If S_1 and S_2 are transcendence bases for E/F, it is not necessarily the case that $F(S_1) = F(S_2)$.*

Theorem 11.31 *The extension E/F has a transcendence base and any two transcendence base of E/F have the same cardinality.*

Proof: The first statement follows from Zorn's Lemma. The proof for the second statement is similar to the proof for the Replacement Lemma of basis for a vector space. □

Definition 11.32 (Transcendence Degree) *The cardinality of a transcendence base for E/F is called the **transcendence degree** of E/F.*

Definition 11.33 (Purely Transcendental) *An extension E/F is called **purely transcendental** if it has a transcendence base S such that $E = F(S)$.*

Theorem 11.34 *Let x be transcendental over F.*

1. *(Lüroth) if $F \subseteq K \subseteq F(x)$. Then $K = F(y)$, for some $y \in F(x)$. In particular, every nontrivial extension of F contained in $F(x)$ is purely transcendental over F.*
2. *If $f(x), g(x)$ are nonzero relative prime polynomials in $F[x]$ which are not both constant, then*

$$[F(x) : F(y)] = [F(x) : F(f(x)/g(x))] = \max(\deg f(x), \deg g(x)).$$

Proof: We only prove the second statement. Since $E = F(x)$ is the field generated by x over F, and $F \subset F(y)$, the suffices to show that the minimal polynomial of x over $F(y)$ has degree n. WLOG, we may assume that $\deg f(x) \geq \deg g(x)$. As otherwise, $\frac{1}{y}$ is such that the degree of numerator is larger than degree of denominator. Since $\frac{1}{y} \in F(y)$ and $y \in F(\frac{1}{y})$, then we have $F(y) = F(\frac{1}{y})$, so we could work with $F(\frac{1}{y})$ instead if $\deg f(x) < \deg g(x)$.

Now let $h(t) \in F(y)[t]$ be given by

$$h(t) = f(t) - g(t)y.$$

As $\deg f(x) \geq \deg g(x)$, then $\deg h = n = \max(\deg f(x), \deg g(x))$. Moreover, $h(x) = f(x) - g(x)\frac{f(x)}{g(x)} = 0$, hence x is a root of the polynomial h. Next we show that $h(t)$ is irreducible in $F(y)[t]$, then this would imply that the minimal polynomial of x is degree n, since we know $m_{x,F(y)}(t)|h(t)$ in $F(y)[t]$, but if $h(t)$ is irreducible, this would mean $m_{x,F(y)}(t)$ is either a constant or a constant multiple of $h(t)$. The former case is clearly not possible, so we must have $m_{x,F(y)}(t) = ch(t)$ for some $c \in F(y)$, so the minimal polynomial is of degree n.

Since $F(y)$ is a field, then $F(y)[t]$ is a Euclidean domain, in particular a unique factorization domain. $h(t)$ has coefficient in $F[y]$, since $g(t), f(t)$ have coefficients in $F \subset F[y]$ and $y \in F[y]$, so we may apply Gauss Lemma applies. Note $F(y)$ is the field of fractions of $F[y]$. Hence $h(t)$ is reducible if and only if it is reducible in $F[y]$.

Now suppose $[F(y) : F]$ is finite, then

$$[E : F] = [E : F(y)][F(y) : F] \leq n[F(y) : F]$$

is also finite, contradicting the fact that x is transcendental over F. Hence we conclude that y is transcendental over F. So $F[y] \cong F[Y]$ where Y is an inter determinant and the isomorphism is explicitly given by $\sigma(y) \mapsto Y$. This isomorphism $\sigma(y)$ also induces an isomorphism between $F[y][t] \cong F[Y][t]$ which we will still denote by σ. We claim that $h(t)$ is irreducible in $F[y][t]$ if $\sigma(h(t))$ is irreducible in $F[Y][t]$. Suppose $h(t) = \alpha(t)\beta(t)$ is reducible in $F[y][t]$, then

$$\sigma(h(t)) = \sigma(\alpha(t))\sigma(\beta(t)) \quad \text{in} F[Y][t].$$

Since α, β are not constants, and $\sigma(\alpha), \sigma(\beta)$ are not constants. So $\sigma(h(t))$ is reducible in $F[Y][t]$. This proves that if $\sigma(h(t))$ is irreducible, then $h(t)$ is not irreducible.

We have $\sigma(h(t)) = f(t) - g(t)Y$. Now suppose $f(t) - g(t)Y = \alpha(t)\beta(t) \in F[Y][t]$ when considered as polynomials in t, then $f(t) - g(t)Y = \alpha(t)\beta(t)$ is also a factorization of $f(t) - g(t)Y$ when considered as polynomials in Y. Since $f(t) - g(t)Y$ is a degree 1 polynomial in Y, then WLOG we may assume that $\alpha(t)$ is a polynomial with Y degree being zero and coefficients in $F[t]$, that is $\alpha(t) \in F[t]$. In this way, $\beta(t)$ must be a degree one polynomial in terms of Y and coefficients in $F[t]$, so $\beta(t) = q(t) + p(t)Y$, where $q(t), p(t) \in F[t]$. Therefore

$$f(t) - g(t)Y = \alpha(t)q(t) + \alpha(t)p(t)Y.$$

By comparing coefficients of Y^0 and Y^1, we must have $\alpha(t)|f(t)$ and $\alpha(t)|g(t)$. Since f, g are coprime, then $\alpha(t) \in F$. This shows that $f(t) - g(t)Y$ is irreducible in $F[Y][t]$ when treated as a polynomial in t. Hence $h(t)$ is irreducible. □

If $a_1, a_2, \cdots, a_n$ are independent indeterminates over a field F, we may evaluate or **specialize** $a_1, \cdots, a_n$ at any elements of F, i.e., substitute values in F for the "variables" $a_1, \cdots, a_n$. If E is a Galois extension of $F(a_1, \cdots, a_n)$, then E is obtained as a splitting field of a polynomial whose coefficients lie in $F[a_1, \cdots, a_n]$. Any specialization of $a_1, \cdots, a_n$ into F maps this polynomial into one whose coefficient lie in F. The specialization of E is the splitting field of the resulting specialized polynomial.

Theorem 11.35 (Hilbert) *Let $x_1, x_2, \cdots, x_n$ be independent transcendental over $\mathbb{Q}$, let $E = \mathbb{Q}(x_1, \cdots, x_n)$ and let G be a finite group of automorphism of E with fixed field K. If K is a purely transcendental extension of $\mathbb{Q}$ with transcendence basis $a_1, a_2, \cdots, a_n$, then there are infinitely many specializations of $a_1, \cdots, a_n$ in $\mathbb{Q}$ such that E specializes to a Galois extension of $\mathbb{Q}$ with Galois group isomorphic to G.*

Corollary 11.35.1 *S_n is a Galois group over $\mathbb{Q}$, f or all n.*

Definition 11.36 (Purely Inseparable Extension) *An algebraic extension E/F is called **purely inseparable** if for each $\alpha \in E$, the minimal polynomial of α over F has one distinct root. This is equivalent to saying*

- *If $\alpha \in E$ is separable over F, then $\alpha \in F$.*
- *If $\alpha \in E$, then $\alpha^{p^n} \in F$ for some n, and $m_{\alpha,F}(x) = x^{p^n} - \alpha^{p^n}$.*

Proposition 11.37 *Let E/F be an algebraic extension. Then there is a unique field E_{set} with $F \subseteq E_{sep} \subseteq E$ such that E_{sep} is separable over F and E is purely inseparable over E_{sep}. The field E_{sep} is the set of elements of E which are separable over F.*

Proof: Let E_{sep} be the field generated by the set of all separable elements in E over F. Then one can see easily that E_{sep} is separable over F and E is purely inseparable over E_{sep} □

Remark 11.37.1 *The degree of E_{sep}/F is called the **separable degree** of E/F and the degree of E/E_{sep} is called the **inseparable degree** of E/F (often denoted as $[E : F]_s$ and $[E : F]_i$, respectively). We also note that $[E : F] = E_{sep}E_i$.*

Corollary 11.37.1 *Separable degrees and inseparable degrees are multiplicative.*

Proof: We know separable degree is multiplicative. Then note that $[E : F] = [E : F]_s$ *cdot*$[E : F]_i$ shows the same holds for inseparable degrees. □

Proposition 11.38 *If E_1 and E_2 are subfields of E which are both separable (or both purely inseparable) extensions of F, then their composite E_1E_2 is separable (purely inseparable, resp.) over F.*

Proof: Use separable and inseparable degree. In particular, they are maximal. □

Proposition 11.39 *If E is a finitely generated extensions of a perfect field F, then there is a transcendence base T of E/F such that E is a separable (algebraic) extensions $F(T)$.*

Remark 11.39.1 *Such transcendence base T is called a **separating transcendence base**.*