Lecture notes: Algebra II, HKU, Spring 2025

Jiang-Hua Lu

References:

- D. Dummit and R. Foote, “Abstract Algebra", Third Edition. Parts II, III, IV.
- Frederick M. Goodman, “Algebra: Abstract and Concrete", Edition 2.6, Chapters 6-10.

  http://homepage.math.uiowa.edu/ goodman/algebrabook.dir/algebrabook.html

About the lecture notes

These lecture notes are used to facilitate the lectures, and they are not sufficient to master the course contents without further reading and practices. Students are supposed to read the textbook/reference books **and** do enough exercises.

Contents

Chapter 1 | Rings and fields

1.1 Review: Definitions and Examples

§ 1. Rings and Fields

Definition 1.1.1. A ring is a set R together with two maps $R\times R \to R : (a, b) \mapsto a+b$ and $R \times R \to R : (a, b) \to ab$ such that

1) $(R, +)$ is an abelian group with 0 denoting its identity element;
2) $(ab)c = a(bc)$ for all $a, b, c \in R$;
3) $(a + b)c = ac + bc$ and $a(b + c) = ab + ac$ for all $a, b, c \in R$.

Unless otherwise specified, all rings in this course will be assumed to have an identity element 1 for the multiplication and that $0 \neq 1$. Most of the rings encountered in this course will be commutative.

Definition 1.1.2. A field is a non-zero commutative ring F such that for any $a \in F, a \neq 0$, there exists $a^{-1} \in F$ such that $aa^{-1} = 1$.

Definition 1.1.3. An element a in a ring R is called a *unit* if there exists $b \in R$ such that $ab = ba = 1$. The set of all units in a ring R is a group under the multiplication in R. A field is thus a non-zero commutative ring in which every non-zero element is a unit. Two elements a and b in a commutative ring are said to be *associates* if $a = ub$ for some unit u in R.

Definition 1.1.4. Let R be a ring. The *characteristic* of R, denoted by $\mathrm{char}(R)$, is the smallest positive integer n, if exists, such that $n \cdot 1 = 0$. When such an integer does not exist, we say that $\mathrm{char}(R) = 0$.

Definition 1.1.5. An element in a ring R is called a *zero divisor* if $a \neq 0$ and if there exists $b \in R\backslash\{0\}$ such that $ab = ba = 0$. A non-zero commutative ring R is called an *integral domain* if it has no zero divisor.

Lemma 1.1.6. *If R is an integral domain, then* $\mathrm{char}(R) = 0$ *or is a prime number.*

Proof. Suppose that R is an integral domain with $\mathrm{char}(R) = n > 0$. If n were not prime, say $n = n_1n_2$ with $1 < n_1, n_2 < n$, then $0 = n \cdot 1 = (n_1 \cdot 1)(n_2 \cdot 1)$. Since R is an integral domain, either $n_1 \cdot 1 = 0$ or $n_2 \cdot 1 = 0$, which is a contradiction. □

Definition 1.1.7. Let R be a commutative ring.

1) An *ideal* in R is a subset I of R closed under addition and such that $ab \in I$ whenever $a \in I$ and $b \in R$.

2) An ideal I is said to be *prime* if $I \neq R$ and if for any $a, b \in R$, if $ab \in I$, then $a \in I$ or $b \in I$.

3) An ideal I is said to be *maximal* if $I \neq R$ and if whenever M is an ideal in R such that $I \subset M \subset R$ then either $M = I$ or $M = R$.

4) For $a \in R$, the ideal of R generated by a is denoted by aR or (a), i.e.,

$$(a) = aR = \{ar : r \in R\},$$

and is called a *principal ideal* of R.

Exercise 1.1.8. If R is an integral domain, two non-zero elements $a, b \in R$ generate the same principal ideal if and only if $a = bu$ for some unit $u \in R$.

Lemma-Definition 1.1.9. Let R be a commutative ring and $I \subset R$ an ideal. For $x, y \in R$, define $x \equiv y \mod (I)$ if $x - y \in I$. Then "$\equiv \mod (I)$" is an equivalence relation. Let R/I be the set of equivalence classes with respect to "$\equiv \mod (I)$" and for $x \in R$ denote by $\bar{x}$ the equivalence class of x. Define two operations "$+$" and "$\cdot$" on R/I by

$$\bar{x} + \bar{y} = \overline{x + y}, \qquad \bar{x}\bar{y} = \overline{xy}, \qquad x, y \in R.$$

Then $(R/I, +, \cdot)$ is a commutative ring, called the quotient ring of R by I.

Lemma 1.1.10. *Let R be a commutative ring and $I \subset R$ an ideal, $I \neq R$.*

1) I is prime if and only if R/I is an integral domain;

2) I is maximal if and only if R/I is a field;

3) Maximal ideals are prime ideals;

4) The zero ideal $\{0\} \subset R$ is prime if and only if R is an integral domain, and $\{0\}$ is maximal if and only if R is a field.

Proof. Exercise. □

Lemma-Definition 1.1.11. Let R be an integral domain. The relation $\sim$ on $P(R) := \{(a, b) \in R : b \neq 0\}$ defined by $(a, b) \sim (c, d)$ if $ad = bc$ is an equivalence relation. Let $\mathrm{Frac}(R)$ be the set of all equivalence classes in $P(R)$, and for $(a, b) \in P(R)$, let $\frac{a}{b} \in \mathrm{Frac}(R)$ denote the equivalence class of (a, b). Define two operations, called *addition* and *multiplication*, on $\mathrm{Frac}(R)$ and denoted respectively as $+$ and $\cdot$, by

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}, \qquad (a, b),\ (c, d) \in P(R).$$

Then $(\mathrm{Frac}(R), +, \cdot)$ is a field with the zero element given by $\frac{0}{1}$ and 1 given by $\frac{1}{1}$, and the map

$$R \longrightarrow \mathrm{Frac}(R), \quad a \longmapsto \frac{a}{1}, \qquad a \in R,$$

is an injective ring homomorphism. The field $\mathrm{Frac}(R)$ is called the *quotient field* or the *field of fractions* of R.

Lemma 1.1.12. *(***Existence of maximal ideals***). For any commutative ring R and an ideal I of R such that $I \neq R$, there exists a maximal ideal of R containing I.*

Proof. The proof is a standard application of Zorn's Lemma. □

Corollary 1.1.13. *Every non-zero commutative ring that is not a field has a non-zero maximal ideal.*

Proof. Let R be a non-zero commutative ring that is not a field. By assumption, there exists $a \in R$, $a \neq 0$, and a is not a unit. Then the ideal aR is contained in some maximal ideal $\mathfrak{m}$ of R which is not the zero ideal. □

Corollary 1.1.14. *Every non-zero commutative ring R has a maximal ideal.*

Proof. If R is a field, then $\{0\}$ is a maximal ideal. If R is not a field, then R contains a non-zero maximal ideal by Corollary 1.1.13 □

Exercise 1.1.15. Classify all ideals, prime ideal, and maximal ideals of the ring $\mathbb{Z}/n\mathbb{Z}$, where $n \geq 2$ is any integer (Tutorial).

§ 2. Examples

Some main constructions of rings:

1. Sub-rings of a ring R generated by a subset $S \subset R$: this is the sub-ring of R consisting of all finite sums of finite products of elements in S.
2. Quotients by ideals;

Some main constructions of fields (more in the later part of the course):

1. Fraction fields of integral domains;
2. sub-fields generated by elements;
3. quotients by maximal ideals

Example 1.1.16. There are four (main) sources of rings:

• *Numbers*: the sets $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, consisting respectively of all integers, rational numbers, real numbers, and complex numbers, are commutative rings. All of them except $\mathbb{Z}$ are fields, and $\mathbb{Q} = \mathrm{Frac}(\mathbb{Z})$. These rings/fields have many interesting sub-rings or quotient rings. For example, for each integer $n \geq 1$, one has the quotient ring

$$\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$$

which is field if and only if n is a prime number. The set $\mathbb{Z}[\sqrt{-1}]$ of *Gaussian integers*, i.e., complex numbers of the form $a+b\sqrt{-1}$, where $a, b \in \mathbb{Z}$, is a sub-ring of $\mathbb{C}$, while the set $\mathbb{Q}[\sqrt{-1}]$ of *Gaussian rationals*, i.e., complex numbers of the form $a+b\sqrt{-1}$, where $a, b \in \mathbb{Q}$, is a sub-field of $\mathbb{C}$. Such rings/fields are studied in *Algebraic Number Theory*.

• *Polynomial rings*: If R is a commutative ring, one has the *polynomial ring* $R[x]$ *of one variable with coefficients in* R, namely[char42] ,

$$R[x] = \{a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n : \ n \in \mathbb{N},\ a_0, a_1, \ldots, a_n \in R\},$$

which is commutative ring with the usual addition and multiplication of "polynomials". Iterating the process, one has the polynomial ring

$$R[x_1, \ldots, x_m] = R[x_1][x_2] \cdots [x_m]$$

of m-variables over R. Polynomial rings and their quotient rings lie at the foundation of *Algebraic Geometry*.

• *Rings of functions*: If X is a set and R is a field, the set $\mathrm{Fun}(X; R)$ of maps from X to R is naturally a commutative ring under pointwise addition and multiplication. When $R = \mathbb{R}$ or $\mathbb{C}$ and when X has additional structures such as a *topology* or a *smooth structure*, one has the sub-ring of *continuous* or *differentiable functions* on X. Such rings are essential in *Topology* or *Differential Geometry*;

• *Matrix rings*: if R is a ring, the set of all $n \times n$ matrices with entries in R is naturally a (typically non-commutative) ring with standard matrix addition and multiplication. The group of units of such a matrix ring is an example of a *group scheme* or *algebraic group*.

◇

Example 1.1.17. [**Polynomial rings**]. One of the most important rings in algebra is the ring $R[x]$ of polynomials in x with coefficients in a commutative ring R. For

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \in R[x], \quad \text{where} \quad a_n \neq 0,$$

the *degree* of f is defined to be n and is denoted by $\deg(f)$. We will refer to a_0 as the *constant term* of f and to a_n as the *leading coefficient* of f. For $f = 0$, we define $\deg(f) = -\infty$. Assume that R is an integral domain. It is then easy to see that if $f, g \in R[x]$ are non-zero, then

$$\deg(fg) = \deg(f) + \deg(g). \tag{1.1}$$

◇

[char42] In these notes, $\mathbb{N}$ denotes the set of all non-negative integers, so it contains 0.

Lemma 1.1.18. *Let R be an integral domain. Then $R[x]$ is an integral domain; Consequently $R[x_1, \ldots, x_n]$ is an integral domain for every integer $n \geq 1$.*

Proof. The statement follows directly from (1.1). □

We now give a very important property of the polynomial rings.

Lemma 1.1.19. *Division Algorithm: let R be an integral domain. Let $f, g \in R[x]$ with $g \neq 0$ and suppose that the leading coefficient of g is a unit in R. Then there exist unique $q, r \in R[x]$ with $\deg(r) < \deg(g)$ such that $f = qg + r$.*

Proof. Let I be the ideal generated by g, and let

$$S = \{f - h \mid h \in I\}.$$

Let $r \in S$ be an element with the smallest degree. If $r = 0$, we are done. Assume that $r \neq 0$. If $\deg(r) \geq \deg(g) \geq 0$, say

$$r(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$
$$g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_k x^k$$

with $a_n, b_k \neq 0$ and $n \geq k$, let $q_1(x) = q(x) = a_n b_k^{-1} x^{n-k}$, which makes sense since b_k is a unit in R. Then

$$f(x) - q_1(x)g(x) = r(x) - a_n b_k^{-1} x^{n-k} g(x)$$

lies in S and has degree less than n. This is a contradiction. Hence $\deg(r) < \deg(g)$. For uniqueness, suppose that $q, q_1, r, r_1 \in R[x]$ are such that $f = qg + r = q_1 g + r_1$ and $\deg(r), \deg(r_1) < \deg(g)$. Then $(q - q_1)g = r_1 - r$. If $q - q_1 \neq 0$, then

$$\deg((q - q_1)g) \geq \deg(g) > \deg(r - r_1)$$

which is a contradiction. Hence q and r are unique. □

Example 1.1.20. Let R be an integral domain with $F = \mathrm{Frac}(R)$ as its field of fraction. Elements in the field of fraction

$$\mathrm{Frac}(R[x_1, \ldots, x_n]) = \mathrm{Frac}(F[x_1, \ldots, x_n])$$

of $R[x_1, \ldots, x_n]$ are called *rational functions in n variables* with coefficients in F. ⋄

Example 1.1.21. [Ring of formal power series] Let R be any commutative ring. A *formal power series* in one variable with coefficients in R is a formal infinite sum $\sum_{i=0}^{\infty} a_i x^i$, where $a_i \in R$ for all $i \geq 0$. The set of all formal power series is denoted by $R[[x]]$, and it is a commutative ring with addition

$$\sum_{i=0}^{\infty} a_i x^i + \sum_{i=0}^{\infty} b_i x^i = \sum_{i=0}^{\infty} (a_i + b_i) x^i$$

and multiplication by

$$\left(\sum_{i=0}^{\infty} a_i x^i\right)\left(\sum_{i=0}^{\infty} b_i x^i\right) = \sum_{i=0}^{\infty} c_i x^i,$$

where $c_i = \sum_{j=0}^{i} a_j b_{i-j}$ for $i \geq 0$. Every non-zero f in $R[[x]]$ can be uniquely written as $f = x^n g$, where $n \in \mathbb{N}$ and $g = \sum_i a_i x^i$ with $a_i \in R$ for every $i \geq 0$ and $a_0 \neq 0$, and n is called the *order of* f and denoted as $n = \mathrm{ord}(f)$. When R is an integral domain, it is easy to see that

$$\mathrm{ord}(fg) = \mathrm{ord}(f) + \mathrm{ord}(g), \quad \forall\ f, g \in R[[x]]\backslash\{0\}. \tag{1.2}$$

$\diamond$

Lemma 1.1.22. *If R is an integral domain, so is $R[[x]]$.*

Proof. We use the definition that an element $\sum_{i=0} a_i x^i = 0 \in R[[x]]$ if and only if $a_i = 0$ for all $i \geq 0$. Suppose that $f, g \in R[[x]]$ are such that $fg = 0$ and $f \neq 0$ and $g \neq 0$. Write $f = x^p f_1$ and $g = x^q g_1$, where the constant terns of f_1 and g_1 are not zero. Then by looking at the coefficient of the x^{p+q} term in $fg = 0$ we get a contradiction to the fact that R in an integral domain. Thus $f = 0$ or $g = 0$. $\square$

Exercise 1.1.23. Let K be a field and let $R = K[[x]]$. Consider the ideal $\mathfrak{m} = xR$. Show that every element in $R\backslash\mathfrak{m}$ is a unit. In other words, every $f = \sum_i a_i x^i \in K[[x]]$ with $a_0 \neq 0$ has an inverse.

Let's classify all ideals of $R = K[[x]]$ for a field K.

Lemma 1.1.24. *For a field K, every non-zero ideals in $R = K[[x]]$ is of the form $x^n R$ for a unique $n \in \mathbb{N}$. Moreover, there are precisely two prime ideals of R, namely, $\{0\}$ and xR, and xR is the unique maximal ideal.*

Proof. By Exercise 1.1.23, every non-zero element $f \in R$ is uniquely of the form $f = x^n g$, where $n = \mathrm{ord}(f)$ and g is a unit. Let I be a non-zero ideal of R, and let n be the smallest among all orders of non-zero elements in I and let $f_0 \in I$ with order n. Write $f_0 = x^n g_0$, where g_0 is a unit. Then $x^n = f_0 g_0^{-1} \in I$, so $x^n R \subset I$. Conversely, let $f \in I$ be arbitrary and write $f = x^m g$, where $m \in \mathbb{N}$, $m \geq n$, and g is a unit. Then $f = x^n x^{m-n} g_1 \in x^n R$. Thus $I = x^n R$. It is also clear that if $n, n' \in \mathbb{N}$ and $n \neq n'$, then $x^n R \neq x^{n'} R$. We thus conclude that every non-zero ideal of $R = K[[x]]$ is of the form $x^n R$ for a unique $n \in \mathbb{N}$. It is now clear that there are precisely two prime ideals of R, namely, $\{0\}$ and xR, and that xR is the unique maximal ideal. $\square$

Example 1.1.25. Let F be a field. The field of fractions of $F[[x]]$ is denoted as $F((x))$ and is called the *field of Laurent series* with coefficients in F. Its elements are of the form $\sum_{i=-n}^{\infty} a_i x^i$, where n is any integer, with the standard addition and multiplication.

1.2 Principal Ideal Domains and Unique Factorization Domains

§ 1. PIDs and Euclidean Domains

Definition 1.2.1. An integral domain R is called a *Principal Ideal Domain*, or a PID, if every ideal is generated by one element, i.e., if for every ideal I of R, there exist $x \in R$ such that $I = aR$.

Example 1.2.2. 1) A field K, having exactly two ideals, namely $\{0\}$ and R, is clearly a PID.

2) By Lemma 1.1.24, $K[[x]]$, where K is a field, is a PID;

3) In Algebra I, you have seen that $R = \mathbb{Z}$ is a PID. Indeed, for any ideal $I \subset \mathbb{Z}$, $I = n\mathbb{Z}$, where $n = \min\{|a| : a \in I\}$;

4) You may already know from Algebra I (if not, see Corollary 1.2.6 below) that for any field K, the polynomial ring $K[x]$ is a PID. Indeed, for any $I \subset K[x]$, $I = fK[x]$, where f has the smallest degree among all non-zero elements in I. ◇

The following inclusions illustrate where where PIDs belong in the realm of commutative rings:

$$\textit{Fields} \subseteq \textit{Euclidean Domains} \subseteq \textit{PIDs} \subseteq \textit{UFDs} \subseteq \textit{Integral Domains}$$
$$\subseteq \textit{Commutative Rings with Identity.}$$

The reason behind the four examples of PIDs in Example 1.2.2 is that they are all *Euclidean domains* and that every Euclidean domain is a PID.

Definition 1.2.3. An *Euclidean domain* is a pair (D, v), where D is an integral domain, and $v : D\backslash\{0\} \to \mathbb{N}$ a map such that

1) $v(ab) \geq \max\{v(a), v(b)\}$ for all $a, b \in D\backslash\{0\}$;

2) for all $a \in D\backslash\{0\}$ and $b \in D$, there exist $q, r \in D$ such that $b = aq + r$, where either $r = 0$ or $r \neq 0$ and $v(r) < v(a)$.

Example 1.2.4. 1) Every field K is trivially an Euclidean domain by defining $v(a) = 0$ for every $a \in K\backslash\{0\}$;

2) For any field K, $D = K[[x]]$ with $v(f) = \text{ord}(f)$ for $f \in K[[x]]\backslash\{0\}$ is an Euclidean domain;

3) $D = \mathbb{Z}$ with $v(a) = |a|$ for $a \in \mathbb{Z}\backslash\{0\}$ is an Euclidean domain;

4) For any field K, $D = K[x]$ with $v(f) = \deg(f)$ for $f \in K[x]\backslash\{0\}$ is an Euclidean domain by Lemma 1.1.19. ◇

Theorem 1.2.5. *Every Euclidean domain (D, v) is a PID.*

Proof. Assume that (D, v) is an Euclidean domain. Let I be any ideal of D. If $I = \{0\}$, then $I = 0R$ is principal. Assume that I is not zero. Choose $a \in I$, $a \neq 0$,

such that $v(a)$ is the smallest among all $b \in I$ such that $b \neq 0$. Since I is an ideal, $aD \subset I$. Assume that $b \in I$ is arbitrary. As (D, v) is an Euclidean domain, there exist $q, r \in D$ such that $b = aq + r$, and either $r = 0$ or $r \neq 0$ but $v(r) < v(a)$. By the choice of a, we must have $r = 0$. Thus $b = aq \in I$. This shows that $I = aD$ is principal. □

Corollary 1.2.6. *If K is a field, then $K[x]$ is a PID.*

Proof. This is because $K[x]$ is an Euclidean domain. □

Example 1.2.7. The ring $R = \mathbb{Z}[x]$ is an integral domain but not a PID. Indeed, by Lemma 1.1.18, $R = \mathbb{Z}[x]$ is an integral domain. Consider $I = 2R + xR$ which consists of all polynomials over $\mathbb{Z}$ whose constant term are even integers. If I is generated by one element $f(x)$. Then $2 = f(x)g(x)$ for some $g \in \mathbb{Z}[x]$, so $f(x) = \pm 2$ or $f(x) = \pm 1$, and in either case $I \neq fR$. Thus I is not principal.

Remark 1.2.8. If R is an integral domain but not necessarily a field, although the function $v(f) = \deg(f)$ for $f \in R[x]$, $f \neq 0$, satisfies condition 1) in the definition of an Euclidean domain, the Euclidean algorithm in Lemma 1.1.19 for $R[x]$ is weaker than what is required in 2) in the definition of an Euclidean domain, so $R[x]$ is not necessarily an Euclidean domain. Example 1.2.7 gives such an example. ◇

Exercise 1.2.9. 1) Consider the sub-ring $D = \mathbb{Z}[\sqrt{-1}]$ of $(\mathbb{C}, +, \cdot)$, where

$$\mathbb{Z}[\sqrt{-1}] = \{m + n\sqrt{-1} : \ m, n \in \mathbb{Z}\}.$$

Elements of $\mathbb{Z}[\sqrt{-1}]$ are called *Gauss integers*. Define $v(m + n\sqrt{-1}) = m^2 + n^2$ for $m + n\sqrt{-1} \in \mathbb{Z}[\sqrt{-1}]\backslash\{0\}$. Show that (D, v) is an Euclidean domain.

2) Consider the sub-ring $D = \mathbb{Z}[\sqrt{2}]$ of $(\mathbb{R}, +, \cdot)$, where

$$\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : \ a, b \in \mathbb{Z}\}$$

with the function $v : D\backslash\{0\} \to \mathbb{N}$ given by $v(a + b\sqrt{2}) = |a^2 - 2b^2|$. Show that (D, v) is an Euclidean domain.

§ 2. Two properties of a PID

Definition 1.2.10. Let R be an integral domain.

1) A non-zero non-unit $a \in R$ is said to be *irreducible* if whenever $a = bc$, either b or c is a unit. A non-zero and non-unit $a \in R$ is said to be *reducible* if it is not irreducible.

2) A non-zero non-unit $a \in R$ is said to be prime if aR is a prime ideal, or, equivalently, if $b, c \in R$ and $a|bc$, then $a|b$ or $a|c$.

Note that if $a \in R$ is irreducible, so is ua for every unit u of R.

Lemma 1.2.11. *Every prime element in an integral domain is irreducible.*

Proof. Assume that $a \in R\backslash\{0\}$ is not a unit and that a is prime. Suppose that $a = bc$. Then a divides bc, so a divides b or c. If a divides b, there exists $x \in R$ such that $b = ax$, so $a = axc$. Hence $xc = 1$, so c is a unit. Similarly, if a divides c, then b is a unit. Hence a is irreducible. □

We now state the first distinguished property of a PID (other than its defining property that every ideal is principal).

Proposition 1.2.12. *If R is a PID, then for every irreducible element $p \in R$, the ideal pR is maximal, so the quotient R/pR is a field.*

Proof. Let R be a PID and let $p \in R$ be irreducible. To show that pR is maximal, suppose that M is an ideal of R containing pR. Since R is a PID, $M = mR$ for some $m \in R$. Since $p \in pR \subset M$, $p = mx$ for some $x \in R$. Since p is irreducible, either m or x is a unit. If m is a unit, then $M = R$. If x is a unit, then $pR = mR = M$. Hence $M = R$ or pR. Thus pR is maximal, and R/pR is a field. □

We can reformulate the property of a PID in Proposition 1.2.12 as follows:

Proposition 1.2.13. *If R is a principal ideal domain, then prime elements in R are the same as irreducible elements, and a non-zero ideal I in R is prime if and only if it is maximal.*

Proof. The statements follow directly from the definitions and Proposition 1.2.12. □

Example 1.2.14. Consider $R = \mathbb{Z}[\sqrt{-5}]$. In this ring,

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

For $x = a + \sqrt{-5}b$ with $a, b \in \mathbb{Z}$, define $|x|^2 = a^2 + 5b^2$. Using $|\ ,\ |$, one can show that 2 is irreducible and that 2 does not divide either of $1 \pm \sqrt{-5}$. Thus $2 \in \mathbb{Z}[\sqrt{-5}]$ is irreducible but not prime. We also conclude that $\mathbb{Z}(\sqrt{-5})$ is not a PID.

Definition 1.2.15. Let R be an integral domain. An irreducible element $f(x) \in R[x]$ is called an *irreducible polynomial over R*, and we often denote $pR[x]$ by $\langle p \rangle$.

Corollary 1.2.16. *If F is a field and if $f(x)$ is an irreducible polynomial over F, the quotient $F[x]/\langle f(x) \rangle$ is a field.*

Proof. As $R = F[x]$ is a PID, the statement follows directly Proposition 1.2.12. □

Corollary 1.2.16 gives a an important tools to construct new fields from an old one. This fact will be emphasized when we study field extensions.

We now turn to a second property of PIDs.

Definition 1.2.17. An integral domain R is said to satisfy the *ascending chain condition for ideals* if for every increasing sequence

$$I_1 \subset I_2 \subset I_3 \subset \cdots \subset I_n \subset \cdots$$

of ideals in R, there exists $m \geq 1$ such that $I_n = I_m$ for all $n \geq 1$.

Lemma 1.2.18. *Every principal ideal domain R satisfies the ascending chain condition for ideals*

Proof. Given an increasing sequence $I_1 \subset I_2 \subset I_3 \subset \cdots \subset I_n \subset \cdots$ of ideals in R, consider $I = \cup_{n\geq 1} I_n$, which is an ideal of R. As R is a PID, there exists $a \in R$ such that $I = aR$. Since $a \in I$, there exists $m \geq 1$ such that $a \in I_m$. Then $I \subset I_m$ and thus $I = I_m$. It follows that $I_n = I_m$ for all $n \geq m$. □

Remark 1.2.19. A commutative ring satisfying the *ascending chain condition for ideals* are said to be *Noetherian*. PIDs are thus Noetherian.

§ 3. Definition of a UFD and two characterizing properties

Definition 1.2.20. An integral domain R is called a *unique factorization domain* (UFD), or *factorial ring*, if every non-zero element of R which is not a unit has a unique factorization into irreducibles, i.e., there exist irreducible elements $p_1, p_2, \ldots, p_r$ (not necessarily distinct) such that $a = p_1 p_2 \cdots p_r$, and if $a = q_1 q_2 \cdots q_s$ is another such factorization, then $r = s$ and after a permutation of the indices, p_i and q_i are associates for each index i.

The *Fundamental Theorem of Arithmetics* says that the ring $\mathbb{Z}$ is a UFD.

Definition 1.2.21. If R is a UFD and $a \in R$ is a non-zero non-unit, a factorization

$$a = p_1 p_2 \cdots p_n,$$

where each p_j is irreducible, is called a *prime factorization*, or a *factorization of a into primes*, or a *factorization of a into irreducibles.*

Lemma 1.2.22. *If R is a UFD, then every irreducible element is prime.*

Proof. Suppose that R is a UFD and let a be irreducible. Suppose that $b, c \in R$ and $a|bc$. Then $bc = ax$ for some $x \in R$. If b is a unit, then $a|c$, and if c is a unit, then $a|b$. Otherwise, write

$$b = p_1 \cdots p_n, \quad c = q_1 \cdots q_m, \quad x = r_1 \cdots r_t$$

as products of irreducibles. Then

$$bc = p_1 \cdots p_n q_1 \cdots q_m = ax = a r_1 \cdots r_t,$$

so the uniqueness of the factorizations implies that a and one of the p_i's or one of the q_j's are associates, so $a|b$ or $a|c$. □

We will now look at another property of UFDs.

Lemma 1.2.23. *Let R be an integral domain. If there exists a non-zero non-unit $a \in R$ which is not the product of irreducible elements in R, then there exists a chain*

$$aR \subset c_1R \subset c_2R \subset \cdots \subset c_nR \subset \cdots$$

of principal ideals in R with proper inclusion at each step.

Proof. Since a is not irreducible, we have $a = a_1b_1$, where neither a_1 nor b_1 is a unit. Consequently,

$$aR \subset a_1R, \quad aR \subset b_1R, \quad \text{and} \quad aR \neq a_1R, \quad aR \neq b_1R.$$

By the assumption on a, either a_1 or b_1 is not irreducible. Say a_1 is not irreducible. Then we can write $a_1 = a_2b_2$, where neither a_2 nor b_2 is a unit, so

$$a_1R \subset a_2R, \quad a_1R \subset b_2R, \quad \text{and} \quad a_1R \neq a_2R, \quad a_1R \neq b_2R.$$

Now $a = a_2b_2b_1$, and at least one of the three elements a_2, b_2, b_1 is not irreducible. Proceeding this way, we get a chain

$$aR \subset c_1R \subset c_2R \subset \cdots \subset c_nR \subset \cdots$$

of principal ideals in R with proper inclusion at each step. □

Lemma 1.2.23 motivates the following definition.

Definition 1.2.24. An integral domain is said to satisfy the *ascending chain condition for principal ideals* (ACCPI) if for every increasing sequence

$$I_1 \subset I_2 \cdots \subset I_n \subset \cdots$$

of principal ideals there exists m such that $I_n = I_m$ for all $n \geq m$.

Definition 1.2.25. If R is a UFD and $a \in R$, $a \neq 0$, define $l(a) = 0$ if a is a unit and $l(a) = n$ if $a = p_1p_2 \cdots p_n$ is a factorization of a as a product of irreducibles. The definition of a UFD implies that $l : R\backslash\{0\} \to \mathbb{N}$ is a well-defined function on R and $l(ab) = l(a) + l(b)$ for $a, b \neq 0$.

Theorem 1.2.26. *An integral domain R is a UFD if and only if it satisfies the ACCPI and every irreducible element in R is prime.*

Proof. Assume first that R is a UFD. By Lemma 1.2.22, every irreducible element in R is prime. Assume that

$$a_1R \subset a_2R \subset \cdots \subset a_nR \subset \cdots$$

is an increasing sequence of principal ideals in R. If $a_j = 0$ for every j, the sequence certainly terminates. Otherwise let $j \geq 1$ be the smallest j such that $a_j \neq 0$. Then the sequence of integers

$$(l(a_j),\ l(a_{j+1}), \ldots)$$

decreases so it must terminate at some m, i.e., $l(a_n) = l(a_m)$ for all $n \geq m$. Since $a_n | a_m$ for every $n \geq m$, a_n and a_m are associates for all $n \geq m$, i.e., $a_n R = a_m R$ for all $n \geq m$.

Conversely, if R satisfies the ACCPI, by Lemma 1.2.23, every non-zero non-unit element in R has a factorization as a product of irreducibles. For $a \in R$ non-zero and non-unit, let $m(a)$ be the smallest number of irreducible factors in such a product. Assuming that every irreducible element of R is prime, we use induction on $m(a)$ to prove uniqueness of the factorization in the sense of Definition 1.2.20.

If $m(a) = 1$, then a is irreducible, and uniqueness is clear. Assume that $m = m(a) > 1$ and that uniqueness of factorization holds for any $b \in R$, $b \neq 0$, with $m(b) < m$. Let $a = p_1 \cdots p_m = q_1 \cdots q_n$ be two factorizations of a into irreducible factors. Consider the irreducible element p_m. By the assumption on R, p_m is prime. As $p_m | q_1 \cdots q_n$, p_m divides q_j for some $1 \leq j \leq n$. By re-ordering the elements, we may assume that $j = n$, so $q_n = x p_m$ for some $x \in R$. As q_n is irreducible and p_m is not a unit, x must be a unit, so p_m and q_n are associates. Let

$$b' = (x^{-1} p_1) p_2 \cdots p_{m-1} = q_1 \cdots q_{n-1}.$$

Then $m(b') \leq m - 1$. By induction assumption, $n - 1 = m - 1$ and that, re-order the elements $q_1, \ldots, q_{n-1}$ if necessary, $x^{-1} p_1$ and q_1 are associates and p_j and q_j are associates for $j \geq 2$. □

§ 4. A PID is a UFD

We are now ready to prove the following

Theorem 1.2.27. *A PID is a UFD.*

Proof. Let R be a PID. By Lemma 1.2.18, R satisfies the ACCPI. By Proposition 1.2.13, every irreducible element of R is prime. By Theorem 1.2.26, R is a UFD. □

§ 5. Greatest common divisors in UFDs

Definition 1.2.28. Given an integral domain R and a non-empty set B of non-zero elements in R, a *great common divisor* (gcd) of B is an element $a \in R$ such that

1) $a | b$ for all $b \in B$;

2) If $a' | b$ for all $b \in B$, then $a' | a$.

Lemma 1.2.29. *In an integral domain R, gcds for a set of non-zero elements B, if exist, are unique up to associates.*

Proof. If both a and a' are gcds for B, then $a|a'$ and $a'|a$, so there exist $x, y \in \mathbb{R}$ such that $a = a'x$ and $a' = ay$. It follows that $a = axy$, so $xy = 1$. Thus x is a unit and a and a' are associates. □

With the understanding that gcds, when exist, are not necessarily unique but can only differ by multiplications by units, we still write $a = \gcd(B)$ if a is a gcd for B.

Lemma 1.2.30. *If R is a UFD, then gcd exists for any non-empty set $B \subset R\backslash\{0\}$.*

Proof. Let $D = \{x \in R : x \neq 0, x|b \text{ for every } b \in B\}$. Then $D \neq \emptyset$ as $1 \in D$. Recall the definition of the length function l on R. As B is non-empty, taking any $b \in B$, one has $l(x) \leq l(b)$ for all $x \in D$. Thus $\{l(x) : x \in D\}$ is a subset of $\mathbb{N} \cup \{0\}$ bounded from above. Let a be an element in D of maximal length. We will show that $a = \gcd(B)$.

As $a \in D$, we only need to show that $a'|a$ for any $a' \in D$. Let $a' \in D$. Let c be a common factor of a and a' that has maximal length among all the common factors of a and a', and write

$$a = cy, \quad a' = cy'.$$

We show that y' must be a unit, which would imply that $a'|a$. Suppose that y' is not a unit. Then y' has an irreducible factor p. As $l(cp) = l(c) + 1$ and c has the maximal length among all common factors of a and a', $cp \nmid a$. So $p \nmid y$. Let $b \in B$ be arbitrary, and write $b = az = cyz$. As $a'|b$, $cp|b$, so $cp|cyz$, which implies that $p|yz$. As p is prime and $p \nmid y$, $p|z$, so $ap|b$. As $b \in B$ is arbitrary, $ap \in D$. This contradicts to the fact that a has maximal length among all elements in D because $l(ap) = l(a) + 1$. Thus y' must be a unit, and hence $a'|a$. □

In a PID, gcds take particular forms. Recall that for a commutative ring R and a subset B of R, the ideal of R generated by B is the set of all finite sums of the form $r_1b_1 + \cdots + r_nb_n$, where $r_1, \ldots, r_n \in R$, $b_1, \ldots, b_n \in B$ and $n \geq 1$ is any integer.

Lemma 1.2.31. *Let R be a PID, and let $B \subset R\backslash\{0\}$. Then the gcds of B are precisely the generators of the ideal I_B of R generated by B. In particular,*

$$\gcd(B) = r_1b_1 + \cdots + r_nb_n$$

for some $r_1, \ldots, r_n \in R$ and $b_1, \ldots, b_n \in B$.

Proof. Let $I_B = aR$, where $a \in R$. Since every $b \in B$ is in I, we have $a|b$ for every $b \in B$, so a is a common divisor for all elements in B. Since $a \in I_B$, $a = r_1b_1 + \cdots + r_nb_n$ for some $r_1, \ldots, r_n \in R$ and $b_1, \ldots, b_n \in B$. Thus if a' is a common divisor for all elements in B, then $a'|b_i$ for each $i = 1, \ldots, n$, so $a'|a$. By definition, a is a gcd for B. Every other gcd divisor for B, being an associate of a, is thus also a generator of I_B. □

Corollary 1.2.32. *Let R be a PID. Let $\{b_1, b_2, \ldots, b_k\}$ be a finite set of non-zero elements such that* $\gcd(b_1, \ldots, b_k) = 1$. *Then*

$$R = b_1R + b_2R + \cdots + b_kR.$$

Remark 1.2.33. Lemma 1.2.31 does not hold for arbitrary UFD. For example, we will show for any field F, the polynomial ring $F[x, y]$ is a UFD. Let $B = \{x, y\}$. Then $\gcd(B) = 1$, but 1 does not line in the ideal generated by B.

The property of gcds of elements in PID, as stated in Lemma 1.2.31, will used in the *Smith normal form theorem* for matrices over PIDs.

§ 6. The Chinese Remainder Theorem for PIDs

We now give another application of the property of gcds of elements in PID stated in Lemma 1.2.31, namely the so-called Chinese Remainder Theorem.

We first introduce a terminology.

Definition 1.2.34. Two non-zero elements b_1 and b_2 in a UFD are said to be *co-prime or relatively prime* if $\gcd(b_1, b_2) = 1$.

Lemma 1.2.35. *Assume that b_1 and b_2 are co-prime, and suppose that $a \in R$ is such that $b_1|a$ and $b_2|a$. Then $(b_1b_2)|a$.*

Proof. Write $a = b_1x$ for $x \in R$, so $b_2|b_1x$. Let p be any prime element such that $p|b_2$ and let m be the highest power such that $p^m|b_2$. Since b_1 does not contains any power of p in its prime factorization, we have $p^m|x$. Thus $b_2|x$, so $(b_1b_2)|a$. □

Assume that R is a PID, and let $I \subset R$ be an ideal. Since R is a PID, there exists $a \in R$ such that $I = \langle a \rangle = aR$. Since a PID is a UFD, we can write a as a product of irreducibles: $a = up_1^{n_1}p_2^{n_2} \cdots p_k^{n_k}$, where $p_1, \ldots, p_k$ are primes that pairwise not associates of each other, and $n_1, \ldots, n_k$ positive integers. Let $q_j = p_j^{h_j}$ for $j = 1, \ldots, k$. Then the q_j's are pairwise co-prime.

Proposition 1.2.36. *Let R be a PID, and let $q_1, q_2, \ldots, q_k$ be elements in R that are pair-wise co-prime, i.e.* $\gcd(q_i, q_j) = 1$ *for all $i \neq j$. Then the map*

$$R/\langle q_1q_2 \cdots q_k \rangle \longrightarrow (R/\langle q_1 \rangle) \times (R/\langle q_2 \rangle) \times \cdots \times (R/\langle q_k \rangle)$$

given by $r + \langle q_1q_2 \cdots q_k \rangle \mapsto (r + \langle q_1 \rangle,\ r + \langle q_2 \rangle,\ \ldots,\ r + \langle q_k \rangle)$ is a ring isomorphism.

Proof. Note that it is enough to prove the statement for $k = 2$. By Lemma 1.2.31, if $\gcd(q_1, q_2) = 1$, then $\langle q_1 \rangle + \langle q_2 \rangle = R$. One then checks that the map

$$J : R/\langle q_1q_2 \rangle \longrightarrow (R/\langle q_1 \rangle) \times (R/\langle q_2 \rangle), \quad x + \langle q_1q_2 \rangle \longmapsto (x + \langle q_1 \rangle,\ x + \langle q_2 \rangle),$$

is an isomorphism of rings. □

The standard form of the Chinese Remainder Theorem says that for any two integers q_1, q_2 such that $(q_1, q_2) = 1$, and for any $r_1, r_2 \in \mathbb{Z}$, the equations

$$x \equiv r_1 \bmod (q_1), \quad x \equiv r_2 \bmod (q_2)$$

have a unique solution x mod q_1q_2. Indeed, as $\mathbb{Z} = q_1\mathbb{Z} + q_2\mathbb{Z}$, we can write

$$r_1 - r_2 = -aq_1 + bq_2$$

for some $a, b \in \mathbb{Z}$. One then sees that $x = r_1 + aq_1 = r_2 + bq_2$ is a solution, and the difference $x - x'$ between any two solutions x and x' is divisible by both q_1 and q_2 so is divisible by q_1q_2. In other words, any two solutions x and x' satisfies

$$x - x' \equiv \bmod (q_1q_2).$$

This is equivalent to saying that the above R-module homomorphism J is both surjective and injective.

Corollary 1.2.37. *Let R be a PID and let $p_1, p_2, \ldots, p_k$ be distinct primes in R. Let $n_1, n_2, \ldots, n_k$ be positive integers. Then*

$$R/\langle p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \rangle \cong (R/\langle p_1^{n_1} \rangle) \times (R/\langle p_2^{n_2} \rangle) \times \cdots \times (R/\langle p_k^{n_k} \rangle).$$

In the case of $\mathbb{Z}$, this says that the remainder of $n \in \mathbb{Z}$ modulo $p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$ is uniquely determined by its remainders modulo $p_j^{n_j}$ for each $j = 1, \ldots, k$.

1.3 Gauss' Lemma and polynomial rings over UFDs

Our goal in §1.3 is to prove that if R is a UFD, so is $R[x]$. This statement would imply that for any R that is a UFD, the multi-variable polynomial ring $R[x_1, \ldots, x_n]$ is a UFD for any integer $n \geq 1$, a fact that is of great importance in algebraic geometry.

Here is the idea: to start with, let R be an integral domain, and let F be the fraction field of R. Given a non-zero and non-unit $f(x) \in R[x]$, we can regard $f(x)$ as in $F[x]$. Since $F[x]$ is a PID and thus a UFD, we can then decompose f as a product of irreducible polynomials in $F[x]$. When R is a UFD, Gauss' Lemma tells us that there is a precise relation between irreducible polynomials in $F[x]$ and irreducible polynomials in $R[x]$.

We discuss Gauss' Lemma first.

§ 1. Gauss' Lemma on products of primitive elements in $R[x]$

For an integral domain R, we first need to know what all the units in $R[x]$ are. We have the following simple fact.

Exercise 1.3.1. If R is an integral domain, an element $f \in R[x]$ is a unit of $R[x]$ if and only if it is a unit of R.

Definition 1.3.2. Let R be a UFD. Given a non-zero polynomial $f \in R[x]$, define *a content* of f to be a gcd of the non-zero coefficients of f. We say that f is *primitive* if it has 1 as a content.

Note that the content of f is only defined up to multiplication by units.

Lemma 1.3.3. *Every non-zero $f(x) \in R[x]$ is a product*

$$f(x) = \gamma g(x)$$

where γ is a content of f, and $g(x) \in R[x]$ is primitive. Any other such product is of the form $f(x) = (\gamma u)(u^{-1}g(x))$ where $u \in R$ is a unit.

Proof. Let $g(x) = f(x)/\gamma$. Then $g(x) \in R[x]$ is primitive by the definition of γ. The uniqueness part of the statement follows from the fact that two greatest common divisors of any subset of $R\backslash\{0\}$ are associates. □

Theorem 1.3.4. *(Gauss' Lemma on products of primitive elements in $R[x]$) If f and g are primitive elements in $R[x]$, so is fg.*

Proof. **First proof.** Write

$$f = \sum_{i=0}^{n} a_i x^i, \quad g = \sum_{j=0}^{m} b_j x^j, \quad fg = \sum_{k=0}^{m+n} c_k x^k.$$

Let γ be a content of fg, and suppose that γ is not a unit. Let p be an irreducible factor of γ. As f and g are primitive, there exist a smallest $0 \leq i_0 \leq n$ such that $p \nmid a_{i_0}$ and a smallest j_0 such that $p \nmid b_{j_0}$. As R is a UFD, p is prime, so $p \nmid a_{i_0}b_{j_0}$. On the other hand, consider

$$c_{i_0+j_0} = \sum_{i \leq i_0} a_i b_{i_0+j_0-i} + a_{i_0}b_{j_0} + \sum_{j<j_0} a_{i_0+j_0-j}b_j.$$

By the assumptions on i_0 and j_0, p divides $\sum_{i\leq i_0} a_i b_{i_0+j_0-i}$ and $\sum_{j<j_0} a_{i_0+j_0-j}b_j$ and not $a_{i_0}b_{j_0}$. Thus $p \nmid c_{i_0+j_0}$, a contradiction. Thus γ must be a unit and hence fg is primitive.

Second proof. Let γ be a content of fg, and suppose that γ is not a unit. Let p be an irreducible factor of γ. Let $R' = R/\langle p\rangle$. Since $p \in R$ is irreducible and R is a UFD, $p \in R$ is prime and so R' is an integral domain, and thus $R'[x]$ is an integral domain. Let $R \to R', r \mapsto \overline{r}$ be the projection and define the ring homomorphism

$$\phi:\ R[x] \to R'[x],\ \phi(a_0 + a_1x + \cdots + a_nx^n) = \overline{a}_0 + \overline{a}_1x + \cdots + \overline{a}_nx^n.$$

Since every coefficient of $fg \in R[x]$ is divisible by p, we have

$$\phi(f)\phi(g) = \phi(fg) = 0 \in R'[x].$$

Since $R'[x]$ is an integral domain, we have either $\phi(f) = 0$ or $\phi(g) = 0$, contradicting the fact that both f and g are primitive. Then r is irreducible. □

§ 2. Gauss' Lemma relating irreducible elements in $F[x]$ and $R[x]$

Let again R be a UFD, and let $F = \mathrm{Frac}(R)$ be the fraction field of R. Another version of Gauss' Lemma relates irreducibility polynomials over R with that of polynomials over F.

Clearly, $R[x] \subset F[x]$. By "clearing the denominators", we can write every $f(x) \in F[x]$ as $f(x) = \alpha g(x)$, where $\alpha \in F\backslash\{0\}$ and $g(x) \in R[x]$. We first give an example.

Example 1.3.5. Note that $\mathbb{Q}$ is the fraction field of $\mathbb{Z}$. In $\mathbb{Q}[x]$, one has

$$f(x) = 1 + \frac{1}{3}x + \frac{3}{4}x^5 \in \mathbb{Q}[x].$$

Clearing the denominator of f gives

$$f(x) = \frac{1}{12}\left(12 + 4x + 9x^5\right),$$

where $g(x) = 12 + 4x + 9x^5 \in \mathbb{Z}[x]$.

Lemma 1.3.6. *Any non-zero $f \in F[x]$ can be written as $f = \alpha g$, where $\alpha \in F$ and $g \in R[x]$ is primitive. If $f = \alpha' g'$ is another such expression, then there exists a unit $u \in R$ such that $g = ug'$ and $\alpha = \alpha' u^{-1}$. In particular, if $f(x) \in R[x]$, then $\alpha \in R$.*

Proof. Using a common denominator, one knows that there exists $\beta \in R$ such that $\beta f \in R[x]$. Let γ be a content of βf and write $\beta f = \gamma g$, where $g \in R[x]$ is primitive. Then $f = \alpha g$, where $\alpha = \frac{\gamma}{\beta}$.

Suppose that $f = \alpha' g'$ is another such expression. Then there exists $\delta \in R$, $\delta \neq 0$, such that $\delta\alpha \in R$ and $\delta\alpha' \in R$. Then $\delta f = \delta\alpha g = \delta\alpha' g' \in R[x]$. As $g, g' \in R[x]$ are primitive, both $\delta\alpha$ and $\delta\alpha'$ are contents of δf, so $\delta\alpha$ and $\delta\alpha'$ are associates, i.e., $\delta\alpha = u\delta\alpha'$ for a unit $u \in R$. It follows that $g = ug'$ and $\alpha = \alpha' u^{-1}$. □

Definition 1.3.7. For a non-constant $f \in F[x]$, if $f = \alpha g$, where $\alpha \in F\backslash\{0\}$ and $g \in R[x]$ is primitive, we call g the *primitive part* of f and write $\mathrm{pp}(f) = g \in R[x]$. Note that the primitive part of f is only well-defined up to multiplication by units of R.

Corollary 1.3.8. *(Gauss' Lemma relating irreducible elements in $F[x]$ and $R[x]$) A non-zero $f \in F[x]$ is irreducible in $F[x]$ if and only if its primitive part $g = \mathrm{pp}(f) \in R[x]$ is irreducible as an element in $R[x]$.*

Proof. The statement is equivalent to saying that $f \in F[x]$ is irreducible if and only if $\mathrm{pp}(f) \in R[x]$ is irreducible.

Assume first that $\mathrm{pp}(f) \in R[x]$ is reducible. Then

$$\mathrm{pp}(f) = k(x)h(x)$$

for some $k(x), h(x) \in R[x]$, neither of which is a constant polynomial defined by a unit of R. Since $\mathrm{pp}(f)$ is primitive, both $k, h \in R[x]$ have positive degrees. As $f(x) = \lambda k(x)h(x) \in F[x]$ for some $\lambda \in F\backslash\{0\}$, we see that $f(x) \in F[x]$ is reducible.

Conversely, assume that $f(x) \in F[x]$ is reducible. Then $f(x) = a(x)b(x)$ for some $a(x), b(x) \in F[x]$ with positive degrees. Write

$$a(x) = \alpha a_1(x) \quad \text{and} \quad b(x) = \beta b_1(x),$$

where $\alpha, \beta \in F$ and both $a_1(x), a_2(x) \in R[x]$ are primitive. Then $f(x) = \alpha\beta a_1(x)b_1(x)$. By Gauss' Lemma on primitive elements, $a_1(x)a_2(x) \in R[x]$ is primitive. Thus $\mathrm{pp}(f) = a_1(x)b_1(x) \in R[x]$. This shows that $\mathrm{pp}(f) \in R[x]$ is reducible. □

§ 3. Characterization of irreducible elements in $R[x]$

Assume again that R is a UFD and that F is the fraction field of R. We now use Gauss' Lemma to give a characterization of irreducible elements in $R[x]$ by treating them as elements in $F[x]$.

Corollary 1.3.9. *Irreducible elements of $R[x]$ are precisely of the two types: irreducible elements of R, and non-constant primitive elements of $R[x]$ that are irreducible in $F[x]$;*

Proof. Assume that $f \in R[x]$ is non-zero and non-unit.

If $\deg(f) = 0$, i.e., if f is a constant polynomial, then f is irreducible as an element in $R[x]$ if and only if f is irreducible as an element in R.

Assume that $\deg(f) > 0$. Writing $f = \gamma g$, where $\gamma \in R$ is a content of f and $g \in R[x]$ primitive. If f is irreducible in $R[x]$, then γ must be a unit of R, so f is primitive, and by Gauss' Lemma on irreducible elements, f is irreducible as an element in $F[x]$. Conversely, if $f \in R[x]$ is primitive and irreducible as an element in $F[x]$, then again by Gauss' Lemma on irreducible elements, f is irreducible as an element in $R[x]$. □

§ 4. If R is a UFD, so is $R[x]$

For elements in $R[x]$, we introduce a condition that is slightly weaker than that of irreducibility.

Definition 1.3.10. For an integral domain R and for a non-constant $g(x) \in R[x]$, by a *proper factorization* of g in $R[x]$, we mean a factorization $g(x) = h(x)k(x)$, where $h(x), k(x) \in R[x]$ and both have positive degrees.

Lemma 1.3.11. *For any integral domain R, if $g(x) \in R[x]$ is irreducible, then $g(x)$ has no proper factorization.*

Proof. Since polynomials in $R[x]$ with positive degrees are not units, the statement follows from the definition. □

Lemma 1.3.12. *Assume that R is a UFD and $g(x) \in R[x]$ is primitive. Then $g(x)$ is irreducible if and only if it has no proper factorization.*

Proof. If $g(x)$ is not irreducible, then $g(x) = k(x)h(x)$ for some $k(x), h(x) \in R[x]$ and neither $h(x)$ nor $k(x)$ is a unit, and since $g(x)$ is primitive, both $k(x)$ and $h(x)$ must have positive degrees, so g has a proper factorization. □

Corollary 1.3.13. *If R is a UFD, every non-zero non-unit $f(x) \in R[x]$ can be written as a product of irreducible elements in $R[x]$.*

Proof. We use induction on $\deg(f)$.

If $\deg(f) = 0$, then $f \in R$ is a constant which is non-zero and non-unit. Since R is a UFD, f is a product of irreducibles elements in R which are irreducible elements in $R[x]$ by Corollary 1.3.9.

Assume that $\deg(f) > 0$. Write $f(x) = \alpha g(x)$, where $\alpha \in R$ is a content of f and $g(x) \in R[x]$ is non-constant primitive. Write $\alpha = \alpha_1 \cdots \alpha_m$, where each α_j is an irreducible element in R which, by Corollary 1.3.9, is also an irreducible element in $R[x]$. If $g(x) \in R[x]$ is irreducible, we are done. If $g \in R[x]$ is not irreducible, then $g(x) = h(x)k(x)$ with $h(x), k(x) \in R[x]$ both having positive degrees. By induction, both $h(x)$ and $k(x)$ are products of irreducible elements in $R[x]$. Thus

$$f(x) = \alpha g(x) = \alpha_1 \cdots \alpha_m h(x)k(x)$$

is a product of irreducible elements in $R[x]$. □

The main theorem of this section is the following theorem.

Theorem 1.3.14. *If R is a UFD, so is $R[x]$.*

Proof. Let $f \in R[x]$ be non-zero and non-unit. By Corollary 1.3.13, f is a product of irreducible elements in $R[x]$. Suppose that

$$f = \alpha_1 \cdots \alpha_m f_1 \cdots f_n = \alpha'_1 \cdots \alpha'_l f'_1 \cdots f'_k,$$

where $\alpha_1, \ldots, \alpha_m, \alpha'_1, \ldots, \alpha'_l$ are irreducible elements in R, and $f_1, \ldots, f_n, f'_1, \ldots, f'_k$ are non-constant polynomials in $R[x]$ which are primitive and irreducible. By Gauss' Lemma on primitive elements, both $f_1 \cdots f_n \in R[x]$ and $f'_1 \cdots f'_k \in R[x]$ are primitive. By Lemma 1.3.6,

$$\alpha := \alpha_1 \cdots \alpha_m \quad \text{and} \quad \alpha' := \alpha'_1 \cdots \alpha'_l$$

are both contents of f, so they are associates in R. Since R is a UFD, $l = m$ and after a permutation, $\alpha_i = u_i \alpha'_i$ for some unit u_i of R for every $1 \leq i \leq l = m$. Let $u = u_1 u_2 \cdots u_n$ so that

$$f_1 \ldots f_n = u f'_1 \cdots f'_k.$$

By Gauss' Lemma on irreducible elements, $f_1, \cdots, f_n, f'_1, \ldots, f'_k$ are irreducible in $F[x]$. As $F[x]$ is a UFD, we know that $n = k$, and after a permutation $f_j = v_j f'_j$ for

some $v_j \in F\backslash\{0\}$ for all $1 \leq j \leq n = k$. By Lemma 1.3.6 again, v_j is a unit in R for each $1 \leq j \leq n = k$.

This finishes the proof that $R[x]$ is a UFD. □

Corollary 1.3.15. *If R is a UFD, so is $R[x_1, \ldots, x_n]$ for any $n \geq 1$.*

1.4 Testing irreducibility of polynomials over $\mathbb{Q}$

Recall from Corollary 1.2.16 that if F is a field, then

$$f(x) \in F[x] \text{ irreducible} \Longrightarrow F[x]/\langle f(x)\rangle \text{ is a field.}$$

The field $F[x]/\langle f(x)\rangle$ is an extension of F in the sense that it contains F as a subfield. In the case of $F = \mathbb{Q}$, irreducible polynomials in $\mathbb{Q}[x]$ will thus give rise to field extensions of $\mathbb{Q}$, which are extremely important in number theory. This is the reason for wanting to test irreducibility of polynomials in $\mathbb{Q}[x]$.

§ 1. Testing irreducibility of polynomials using Gauss' Lemma

Given a non-constant $f(x) \in \mathbb{Q}[x]$, we can always clear the denominator and write

$$f(x) = \alpha g(x),$$

where $\alpha \in \mathbb{Q}\backslash\{0\}$ and $g(x) \in \mathbb{Z}[x]$, and we don't need to assume that $g(x) \in \mathbb{Z}[x]$ is primitive. Since $f(x)$ and $g(x)$ are associates in $\mathbb{Q}[x]$, $f(x)$ is irreducible as an element in $\mathbb{Q}[x]$ if and only if $g(x)$ is irreducible as an element in $\mathbb{Q}[x]$. We thus have the following general problem.

Problem. Develop tests to see whether a given $f(x) \in \mathbb{Z}[x]$ is irreducible as an element in $\mathbb{Q}[x]$. For example, is

$$f(x) = x^5 + 5$$

irreducible over $\mathbb{Q}$?

Consider the general case: assume that R is a UFD and let $F = \mathrm{Frac}(R)$ be the fraction field of R. Given a non-constant polynomial $g(x) \in R[x]$, we regard $g(x)$ as in $F[x]$ and ask whether or not $g(x)$ is irreducible as an element in $F[x]$.

Recall that a proper factorization of $g(x) \in R[x]$ in $R[x]$ is of the form

$$g(x) = h(x)k(x),$$

where $h(x), k(x) \in R[x]$ and both have positive degrees.

Proposition 1.4.1. *(Gauss' Lemma on proper factorization) A non-constant polynomial $g(x) \in R[x]$ is irreducible as an element in $F[x]$ if and only if $g(x)$ has no proper factorization in $R[x]$.*

Proof. The statement is equivalent to $g(x) \in R[x]$ is reducible as an element in $F[x]$ if and only if $g(x)$ has a proper factorization in $R[x]$.

If $g(x)$ has a proper factorization in $R[x]$, then $g(x)$ is clear reducible as an element in $F[x]$.

Conversely, assume that $g(x) \in R[x]$ is reducible as an element in $F[x]$. Write

$$g(x) = \gamma h(x),$$

where $\gamma \in R$ is the content of $g(x)$ and $h(x)$ is a primitive part of $g(x)$. Since $g(x)$ is reducible as an element in $F[x]$, $h(x)$ is reducible as an element in $F[x]$. Since $h(x) \in R[x]$ is primitive, by Gauss' Lemma relating irreducible elements in $R[x]$ and $F[x]$, $h(x)$ is reducible as an element in $R[x]$. Thus $h(x)$ has a proper factorization in $R[x]$. It follows that $g(x)$ has a proper factorization in $R[x]$. □

In the next few subsections, we discuss a bag of tools for testing whether a given $g(x) \in \mathbb{Z}[x]$ has a proper factorization.

§ 2. The method of reduction modulo p

A basic technique for testing irreducibility in $\mathbb{Z}[x]$ is by reduction modulo a prime.

Let p be a prime number and let $\pi_p : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the projection. The induced ring homomorphism $\mathbb{Z}[x] \to \mathbb{Z}_p[x]$ is also denoted by π_p.

Lemma 1.4.2. *Let p be a prime number, and assume that $f \in \mathbb{Z}[x]$ is non-constant and leading coefficient not divisible by p. If $\pi_p(f) \in \mathbb{Z}_p[x]$ is irreducible, then f is irreducible in $\mathbb{Q}[x]$.*

Proof. We need to show that $f(x)$ has no proper factorization in $\mathbb{Z}[x]$.

Suppose that $f = gh$ for $g, h \in \mathbb{Z}[x]$ with $\deg(g) > 0$ and $\deg(h) > 0$. Then

$$\pi_p(f) = \pi_p(g)\pi_p(h) \in \mathbb{Z}_p[x].$$

The assumption on f implies that $\deg(\pi_p(f)) = \deg(f)$ and

$$\deg(\pi_p(g)) > 0 \quad \text{and} \quad \deg(\pi_p(h)) > 0,$$

contradicting irreducibility of $\pi_p(f)$. Thus $f(x)$ has no proper factorization in $\mathbb{Z}[x]$, so f is irreducible in $\mathbb{Q}[x]$. □

Example 1.4.3. Let $f(x) = 35x^3 + 3x^2 + 4x - 7$ and let $p = 2$. Then

$$\pi_2(f)(x) = x^3 + x^2 + 1 \in \mathbb{Z}_2[x].$$

We claim that $\pi_2(f)(x) = x^3 + x^2 + 1$ is irreducible in $\mathbb{Z}_2[x]$: if it were reducible, it must have a linear factor so must have a root in $\mathbb{Z}_2$. But neither 0 not 1 is a root. Thus $\pi_2(f)(x)$ is irreducible in $\mathbb{Z}_2[x]$. By Lemma 1.4.2, $f(x) = 35x^3 + 3x^2 + 4x - 7$ is irreducible over $\mathbb{Q}$.

§ 3. Eisenstein's Criteria

Eisenstein's criteria give a large class of irreducible polynomials over $\mathbb{Q}$.

Proposition 1.4.4. *(Eisenstein Criteria) Let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x]$.*

1) If there exists a prime number p such that p divides $a_0, a_1, \cdots, a_{n-1}$ but not a_n, and that p^2 does not divide a_0, then f is irreducible over $\mathbb{Q}$;

2) If there exists a prime number p such that p divides $a_1, a_2, \cdots, a_n$ but not a_0, and that p^2 does not divide a_n, then f is irreducible over $\mathbb{Q}$.

Proof. We will prove 1) only, the proof of 2) being similar.

Assume that f is reducible over $\mathbb{Q}$. By Gauss' Lemma, f has a proper factorization $f = gh$, where $g, h \in \mathbb{Z}[x]$ with $\deg(h) > 0$ and $\deg(g) > 0$. For $a \in \mathbb{Z}[x]$, set $\bar{a} = \pi_p(a) \in \mathbb{Z}_p[x]$. By the assumptions on f, we have $\bar{f} = \bar{a}_n x^n = \bar{g}\bar{h}$. By the following Lemma 1.4.5, $\bar{g} = bx^k$ and $\bar{h} = cx^l$ for some non-zero $b, c \in \mathbb{Z}/p\mathbb{Z}$ and $k + l = n$. Thus p divides all but one coefficients of g, and similarly for h. Since $\bar{g}\bar{h} = \bar{a}_n x^n \neq 0$, p divides all but the leading coefficients of both g and h. In particular, since we are assuming that both g and h have positive degrees, p divides the constant terms of both g and h, which is a contradiction to the assumption that p^2 does not divide a_0. □

Lemma 1.4.5. *Let R be an integral domain. Let $n \geq 0$ be an integer and $a \in R\backslash\{0\}$. If $g, h \in R[x]$ are such that $gh = ax^n$, then $g = bx^k$ and $h = cx^l$ for some $b, c \in R\backslash\{0\}$ and $k, l \in \mathbb{N}$ with $k + l = n$.*

Proof. Let $k = \deg(g)$ and $l = \deg(h)$. Let bx^{k_1} and cx^{l_1}, where $b, c \in R\backslash\{0\}$, are the monomial terms in g and h respectively that have the lowest degrees. If $k_1 \neq k$ or $l_1 \neq l$, then $bc = 0$ which contradicts the fact that R is an integral domain. Hence $k_1 = k$ and $l_1 = l$. Alternatively, since $R[x]$ is a UFD and since $x \in R[x]$ is irreducible, all the irreducible factors of g and h must be associates of x. Thus $g = bx^k$ and $h = cx^l$. □

Example 1.4.6. By Eisenstein's criterion, $x^4 + 14x^2 + 49x + 7$ is irreducible in $\mathbb{Z}[x]$ and in $\mathbb{Q}[x]$.

Example 1.4.7. Any polynomial of the form $f(x) = 2 + a_1x + a_2x^2 + \cdots + a_nx^n$ with $a_1, a_2, \cdots, a_{n-1}$ even and a_n odd is irreducible in $\mathbb{Q}[x]$ by the Eisenstein Criterion. For example, $f(x) = 2 + 4x - 6x^2 + 5x^3$ is irreducible in $\mathbb{Q}[x]$. For any integer $n \geq 1$ and prime number p,

$$f(x) = x^n + p$$

is irreducible over $\mathbb{Q}$.

§ 4. Change of variables

If $f(x) \in \mathbb{Q}[x]$, by setting $x = ay + b$, where $a, b \in \mathbb{Q}$ and $a \neq 0$, one get

$$g(y) = f(ay + b) \in \mathbb{Q}[y],$$

and $f(x) \in \mathbb{Q}[x]$ is irreducible if and only if $g(y) \in \mathbb{Q}[y]$ is irreducible. Here is an example.

Example 1.4.8. Let p be a prime, and consider the polynomial

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

Setting $y = x - 1$, one has

$$f(x) = f(y + 1) = \frac{(y + 1)^p - 1}{y} = y^{p-1} + C_1^p y^{p-2} + \cdots + p$$

which is irreducible in $\mathbb{Q}[x]$ by Eisenstein's Criterion. Since f is primitive, it is also irreducible over $\mathbb{Z}$.

§ 5. Rational Root Test

There is a necessary condition for an $f(x) \in \mathbb{Z}[x]$ to have a linear factor, stated as the following *Rational Root Test*, whose proof is straightforward.

Lemma 1.4.9. *If a polynomial*

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \in \mathbb{Z}[x]$$

has a rational root r/s, where r and s are integers that are relatively prime, then $s|a_n$ and $r|a_0$. In particular, if f is monic, i.e., if $a_n = 1$, then its only rational roots are integers.

To test whether a cubic or quadratic polynomial $f(x) \in \mathbb{Q}[x]$ is irreducible or not, If f is reducible in $\mathbb{Q}[x]$, it must have a linear factor, so it must have a rational root. Write $f(x) = \alpha g(x)$, where $\alpha \in \mathbb{Q}$, $\alpha \neq 0$, and $g(x) \in \mathbb{Z}[x]$. Then $g(x)$ must have a rational root. If the Rational Root Test applied to $g(x)$ says that $g(x)$ has no rational root, then we can conclude that $f(x)$ is irreducible over $\mathbb{Q}$.

Example 1.4.10. Consider $f(x) = x^3 + 5x + 2 \in \mathbb{Z}[x]$. Since it is monic, any rational root must be an integer r satisfying $r|2$. Since neither $r = 2$ nor $r = -2$ is a root of f, f has no rational root so f is irreducible over $\mathbb{Q}$. Since f is primitive, it is also irreducible over $\mathbb{Z}$.

1.5 More examples of PIDs

§ 1. Quadratic integers

A source of of Euclidean domains/PIDs/non-PIDs is the theory of *quadratic integers* (or *algebraic integers* in general) which is at the heart of *Algebraic Number Theory*: an element $x \in \mathbb{C}$ is called a *quadratic integer* if it satisfies the quadratic equation $x^2 + Bx + C = 0$ for some integers B and C.

Let $D \neq 0$ and $D \neq 1$ be a square free integer, i.e., D is not divisible by any n^2 for any integer $n \geq 2$ (for example $D = -1, 2, -5, 19$ etc.), and consider the sub-ring

$$\mathbb{Q}[\sqrt{D}] = \{a + b\sqrt{D} : \ a, b \in \mathbb{Q}\}$$

of $\mathbb{C}$ (it is a sub-ring of $\mathbb{R}$ if $D > 0$). As D is square free, one can show (do this!) that $\sqrt{D} \notin \mathbb{Q}$ and consequently $\mathbb{Q}[\sqrt{D}]$ is a sub-field of $\mathbb{C}$. It is easy to see that every quadratic integer is contained in $\mathbb{Q}[\sqrt{D}]$ for some $D \notin \{0, 1\}$ and square free. The set of quadratic integers contained $\mathbb{Q}[\sqrt{D}]$ is denoted by $\mathcal{O}_{\mathbb{Q}[\sqrt{D}]}$.

Fact: (try to prove this) $\mathcal{O}_{\mathbb{Q}[\sqrt{D}]}$ is the sub-ring of $\mathbb{C}$ consisting of all elements of the form $a + b\omega$, where $a, b \in \mathbb{Z}$ and

$$\omega = \begin{cases} \sqrt{D} & \text{if} \quad D = 2 \text{ or } 3 \bmod 4, \\ \frac{1}{2}(1 + \sqrt{D}) & \text{if} \quad D = 1 \bmod 4. \end{cases}$$

There has been a great deal of study (even still with open problems) on whether $\mathcal{O}_{\mathbb{Q}[\sqrt{D}]}$ is a PID, and when PID, whether it is a Euclidean domain, and when not a PID, how it fails to be a PID. Basically for $D < 0$, $\mathcal{O}_{\mathbb{Q}[\sqrt{D}]}$ is a PID precisely when $D = -1, -2, -3, -7, -11, -19, -43, -67, -163$, and among these the first five are Euclidean domains and the last four are not. The study of how $\mathcal{O}_{\mathbb{Q}[\sqrt{D}]}$ fails to be PIDs and exactly what they are (they are *Dedikind domains*) has been a corner stone in Algebra.

Chapter II *Factorization* of the book *Algebra* by M. Artin, 2004, contains many details on the ring $\mathbb{Z}[\sqrt{-1}]$ of Gauss integers and more generally on quadratic integers. See also the Wikipedia articles “Quadratic integer" for a quick discussion and references for further reading.

§ 2. Localization and local PIDs

To see more examples of PIDs, consider again the ring $R = K[[x]]$ of power series in x with coefficients in K, where K is a field. A distinguished property of $R = K[[x]]$ is that every ideal of R is of the form $x^n R$ for some integer $n \geq 0$, among which the ideal $\mathfrak{m} := xR$ is the unique maximal ideal. It is also easy to see that every $f \in R \backslash \mathfrak{m}$ is a unit (see Exercise 1.1.23).

Definition 1.5.1. A non-zero commutative ring R is said to be *local* if it has a unique maximal ideal.

Recall that for a commutative ring R, we denote by $R^\times$ the set of all unit of R.

Lemma 1.5.2. *A non-zero commutative ring R is local if and only if $R\backslash R^\times$ is an ideal, in which case $R\backslash R^\times$ is the unique maximal ideal of R.*

Proof. Assume first that R is local and let $\mathfrak{m}$ be the unique maximal ideal of R. As $\mathfrak{m} \neq R$ by definition, $\mathfrak{m}$ does not contain any unit, so $\mathfrak{m} \subset R\backslash R^\times$. Conversely, let $a \in R\backslash R^\times$ be arbitrary. By Lemma 1.1.12, there is a maximal ideal $\mathfrak{m}'$ of R containing a. As $\mathfrak{m}$ is the unique maximal ideal, $\mathfrak{m}' = \mathfrak{m}$, so $a \in \mathfrak{m}$. Thus $\mathfrak{m} = R\backslash R^\times$. □

Definition 1.5.3. A PID which is also a local ring is called a *local PID*.

Example 1.5.4. For any field K, the ring $K[[x]]$ of all formal power series in x with coefficients in K is a local PID.

Lemma 1.5.5. *Let R be an integral domain that is not a field. The the following are equivalent:*

1) R is a local PID;

2) there exists a non-unit $x \in R$ such that every non-zero element $a \in R$ is of the form $a = x^n u$ for some $n \in \mathbb{N}$ and some unit u in R.

When 1), or equivalently, 2) holds, x in 2) is unique up to multiplication by a unit, xR is the unique maximal ideal of R, and (R, v) is an Euclidean domain, where $v : R\backslash\{0\} \to \mathbb{N}$ is given by $v(x^n u) = n$ for any unit u.

Proof. Assuming first that R is a local PID. Let $\mathfrak{m}$ be the unique maximal ideal of R. As R is a PID, $\mathfrak{m} = xR$ for some non-unit $x \in R$. Let $a \in R$ and $a \neq 0$. If a is a unit, $a = x^0 a$. Assume now that a is not a unit. Then $a \in \mathfrak{m}$, so there exists $a_1 \in R$ such that $a = xa_1$. If a_1 is a unit, we are done. Otherwise, $a_1 = xa_2$ for some $a_2 \in R$, so $a = x^2 a_2$. If a_2 is a unit, we are done. Otherwise there exists $a_3 \in R$ such that $a_2 = xa_3$, so $a = x^3 a_3$. We must show that this process can not go on forever, i.e., there exists $n \geq 1$ such that $a = x^n a_n$, where a_n is a unit. Suppose not. Then we have a sequence of strictly increasing ideals

$$a_1 R \subset a_2 R \subset \cdots \subset a_n R \subset a_{n+1} R \subset \cdots .$$

By Lemma 1.2.18, R satisfies the ascending chain condition of ideals, one must then have $a_m R = I$ for all $m \geq n$, a contradiction. Thus 1) implies 2).

Assume 2). Then clearly the set of all non-units is the ideal xR, so R is local. Define $v : R\backslash\{0\} \to \mathbb{N}$ by $v(x^n u) = n$ for $n \in \mathbb{N}$ and u a unit. It is straightforward to check that (R, v) is an Euclidean domain: indeed, for $a = x^n u$ and $b = x^m v$, where u, v are units, one has $b = a(x^{m-n} v u^{-1})$ if $m \geq n$, and $b = a0 + b$ if $m < n$. Thus R is a PID.

Assume 1), or equivalently, 2) holds. Then x in 2), being a generator of the ideal that is the set of all non-units, is unique up to multiplication by a unit because R is an integral domain. □

Example 1.5.6. We now give more examples of local PIDs.

1) For a prime number p, let $\mathbb{Z}_{(p)}$ be the subset of $\mathbb{Q}$ given by

$$\mathbb{Z}_{(p)} = \{a/b : \ a, b \in \mathbb{Z}, \, b \neq 0, \, p \nmid b\}.$$

It is clear that $\mathbb{Z}_{(p)}$ is a sub-ring of $\mathbb{Q}$, so it is also an integral domain. Every $u = a/b$, with $a, b \in \mathbb{Z}\backslash\{0\}$ and $p \nmid a$ and $p \nmid b$, is a unit in $\mathbb{Z}_{(p)}$, so every non-zero element in $\mathbb{Z}_{(p)}$ is uniquely of the form $p^n u$, where u is a unit of $\mathbb{Z}_{(p)}$. By Lemma 1.5.5, $\mathbb{Z}_{(p)}$ is a local PID with the unique maximal ideal generated by $p \in \mathbb{Z}_{(p)}$.

2) Let K be a field and let

$$K[x]_{(x)} = \{f/g : \ f, \, g \in K[x], \, g(0) \neq 0\} \subset K(x).$$

Being a sub-ring of $K(x) = \mathrm{Frac}(K[x])$, $K[x]_{(x)}$ is an integral domain, and every non-zero element in $K[x]_{(x)}$ is of the form $x^n f/g$, where $f, g \in K[x]$ and $f(0) \neq 0$, $g(0) \neq 0$, so f/g is a unit in $K[x]_{(x)}$. It follows that $K[x]_{(x)}$ is a local PID with the unique maximal ideal generated by $x \in K[x]_{(x)}$.

The same arguments as in the previous two examples proves the following general way of constructing local PIDs from a UFD.

Lemma 1.5.7. *Let R be any UFD with fraction field F, and let $p \in R$ be a prime element. The sub-ring*

$$R_{(p)} \stackrel{\text{def}}{=} \left\{ p^n \frac{a}{b} : \ n \in \mathbb{N}, \, a, b \in R, b \neq 0, \, p \nmid a, \, p \nmid b \right\}$$

of F is a local PID with unique maximal ideal $pR_{(p)}$.

Remark 1.5.8. Another name for local PIDs that are not fields is *Discrete Valuation Rings (DVR)*, a topic in *Commutative Algebra*. Every DVR has a *completion* which is again a DVR. The completion of $K[x]_{(x)}$ is $K[[x]]$, which we have already seen to be a local PID; the completion of $\mathbb{Z}_{(p)}$ is the *ring of p-adic integers* is usually denoted by $\mathbb{Z}_p$. $\diamond$

The process in Lemma 1.5.7 from R to $R_{(p)}$ is an example of *localization*, a very powerful tool in algebra, which we now explain. Recall that rings considered in this course are always assumed to have the multiplicative identity 1.

Definition 1.5.9. Let R be any commutative ring. A subset D of $R\backslash\{0\}$ is said to be *multiplicatively closed* if $1 \in D$ and if $ab \in D$ for all $a, b \in D$.

Lemma-Definition 1.5.10. Let R be any commutative ring and let $D \subset R\backslash\{0\}$ be multiplicatively closed. For $(r_1, d_1), (r_2, d_2) \in R \times D$, define

$$(r_1, d_1) \sim (r_2, d_2) \quad \text{if} \quad d(r_1 d_2 - r_2 d_1) = 0 \quad \text{for some } \ d \in D.$$

Then $\sim$ is an equivalence relation on $R \times D$. For $(r, d) \in R, D$, denote by $\frac{r}{d}$ the equivalence class of (r, d), and let $D^{-1}R$ be the set of all equivalence classes in $R \times D$. Then $D^{-1}R$ is a ring with the operations

$$\frac{r_1}{d_1} + \frac{r_1}{d_1} = \frac{r_1 d_2 + r_2 d_1}{d_1 d_2}, \qquad \frac{r_1}{d_1} \cdot \frac{r_1}{d_1} = \frac{r_1 r_2}{d_1 d_2}, \qquad (r_1, d_1), (r_2, d_2) \in R \times D.$$

Moreover, the map

$$R \longrightarrow D^{-1}R, \;\; r \longmapsto \frac{r}{1},$$

is a ring homomorphism, which is injective if D has no zero divisor. The ring $D^{-1}R$ is called called the *localization of R at D*.

Proof. Direct check. □

Example 1.5.11. If R is integral domain and F its fractions field, then for any multiplicatively closed $D \subset R\backslash\{0\}$, one has the injective ring homomorphism

$$D^{-1}R \longrightarrow F, \;\; \frac{r}{d} \longmapsto \frac{r}{d},$$

whose image is the sub-ring of F generated by $D^{-1} = \{d^{-1} : d \in D\}$ and R. Note that as a sub-ring of F, the localization $D^{-1}R$ is also an integral domain.

Proof. Direct check. □

Let R be any integral domain and let D_0 be any non-empty subset of $R\backslash\{0\}$. Let D be the subset of R consisting of all finite products of elements in D_0. Then $D \subset R\backslash\{0\}$ and is multiplicatively closed. The localization $D^{-1}R$ is also called the localization of R at D_0.

For a better understanding of the ring $D^{-1}R$, we introduce some definition.

Definition 1.5.12. Let R and Q be any commutative rings and let $\phi : R \to Q$ be a ring homomorphism.

1) For any ideal I of R, the ideal $\phi(I)Q$ of Q is called *the extension of I to Q by ϕ*, and we write $I^e = \phi(I)Q \subset Q$;

2) For any ideal J of Q, the ideal $\phi^{-1}(J)$ of R is called *the contraction of J in R by ϕ*, and we write $J^c = \phi^{-1}(J) \subset R$. Note that when R is a sub-ring of Q and $\phi : R \to Q$ is the inclusion, the contraction of J in R is $J^c = J \cap R$.

Exercise 1.5.13. For any ring homomorphism $\phi : R \to Q$, prove that the following statements hold:

1) For any ideal I of R, one has $I \subset (I^e)^c$;

2) For any ideal J of Q, one has $(J^c)^e \subset J$.

Let R be any commutative ring and let $D \subset R\backslash\{0\}$ be multiplicatively closed. We now consider extensions and contractions of ideals with respect to the ring homomorphism $R \to D^{-1}R$. The following statements hold:

Lemma 1.5.14. *Let R be any commutative ring and let $D \subset R\backslash\{0\}$ be multiplicatively closed.*

1) For any ideal J of $D^{-1}R$, we have $J = (J^c)^e$. Consequently, every ideal of $D^{-1}R$ is the extension of some ideal of R, and distinct ideals of $D^{-1}R$ have distinct contractions in R;

2) For any ideal I of R, we have

$$(I^e)^c = \{r \in R : dr \in I \text{ for some } d \in D\}.$$

Moreover, $I^e = D^{-1}R$ if and only if $I \cap D \neq \emptyset$.

3) Extension and contraction give a bijection between prime ideals I of R such that $I \cap D = \emptyset$ and prime ideals of $D^{-1}R$.

Proof. Exercise. □

The most important example of localization is *localization at prime ideals.*

Lemma-Definition 1.5.15. Let R be an commutative ring and let $P \subset R$ be a prime ideal. Then $D = R\backslash P \subset R\backslash\{0\}$ is multiplicatively closed, and the localization $D^{-1}R$ is called the *localization of R at P* and is denoted as R_P.

Lemma 1.5.16. *Let R be an commutative ring and let $P \subset R$ be a prime ideal. Then by extension and contraction of ideals by the ring homomorphism $R \to R_P$, prime ideals of R_P are in bijection with prime ideals of R contained in P.*

Proof. Special case of Lemma 1.5.14. □

Proposition 1.5.17. *Let R be an commutative ring, let R_P be the localization of R at a prime ideal P, and let $\pi : R \to R_P, r \mapsto r/1$ be the natural ring homomorphism. Let P^e be the extension of P to R_P by π. Then*

1) R_P is a local ring with $P^e \subset R_P$ as its unique maximal ideal.

2) The map π induces an injective homomorphism from R/P to R_p/P^e, and R_p/P^e is isomorphic to the fraction field of the integral domain R/P.

Proof. Exercise. □

Remark 1.5.18. We have see that localization of a UFD at a *principal prime ideal* is a local PID. In general, localizations of UFD at arbitrary prime ideals are not necessarily PIDs. On the other hand, localization of any *Dedekind domain* at a prime ideal is a local PID. A PID is a Dedekind domain, but Dedekind domains are more general than PIDs. For example, for any square free integer $D \notin \{0, 1\}$, the ring $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ of quadratic integers in the quadratic field $\mathbb{Q}(\sqrt{D})$ may not always be a PID but is always a Dedekind domain. ◇

Chapter 2 | Finitely generated modules over PIDs

2.1 Smith Normal Forms of matrices with entries in PIDs

§ 1. Cauchy-Binet formula

We first assume that R is any commutative ring with 1.

For integers $m, n \geq 1$, let $M_{m,n}(R)$ be the set of all $(m \times n)$-matrices with entries in R. Then $M_{n,n}(R)$ is a (in general non-commutative) ring under matrix addition and matrix multiplication exactly the same way as when $R = \mathbb{R}$ or $\mathbb{C}$. Furthermore, one can define $\det(A) \in R$ for any $A \in M_{n,n}(R)$ using the same formula, namely, for $A = (a_{i,j}) \in M_{n,n}(R)$,

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)} \in R,$$

where S_n is the permutation matrix on n letters. For $i, j = 1, \ldots, n$, let $A_{i,j}$ be the $(n-1) \times (n-1)$ sub-matrix of A formed by deleting the ith row and and the jth column, and let

$$A^{\mathrm{co}} = \big((-1)^{i+j} \det(A_{j,i})\big)_{i,j=1,\ldots,n}$$

be the *co-factor matrix* of A. Then it follows from the definitions that

$$AA^{\mathrm{co}} = A^{\mathrm{co}} A = \det(A) I_n,$$

where I_n is the $n \times n$ identity matrix. It follows that $A \in M_{n,n}(R)$ has an inverse if and only if $\det(A)$ is a unit of R, and in such a case,

$$A^{-1} = (\det(A))^{-1} A^{\mathrm{co}}.$$

Let $GL(n, R)$ be the subset of $M_{n,n}(R)$ consisting of all the invertible matrices in $M_{n,n}(R)$, i.e.,

$$GL(n, R) = \{A \in M_{n,n}(R) : \det(A) \text{ is a unit in } R\}.$$

Then $GL(n, R)$ is a (typically non-abelian) group under matrix multiplication.

Notation 2.1.1. 1) For integer $m \geq 1$, let $[m] = \{1, \ldots, m\}$, and for $1 \leq k \leq m$, let $S_{[m],k}$ be the collection of all subsets of $[m]$ of size k;

2) For $A \in M_{m,n}(R)$, $1 \leq k \leq \min\{m, n\}$, and $I, J \in S_{[m],k}$, let $[A]_{I,J}$ be the determinant of the sub-matrix of A formed by the rows from I and columns from J, and call $[A]_{I,J}$ a $k \times k$ minor of A.

Lemma 2.1.2. *Let R be an integral domain. For any $A \in M_{m,n}(R), B \in M_{n,l}(R)$, $1 \leq k \leq \min\{m, l\}$, $I \in S_{[m],k}$, and $J \in S_{[l],k}$, one has*

$$[AB]_{I,J} = \sum_{K \in S_{[n],k}} [A]_{I,K}[B]_{K,J}. \tag{2.1}$$

Proof. One can prove (2.1) using the idea of R-modules, but we will prove it using linear algebra of vector spaces over the fraction field F of R.

Let first $1 \leq k \leq q$ be integers and consider the kth exterior power $\wedge^k F^p$, where F^p is regarded as a q-dimensional vector space over F. Let $e_1, \ldots, e_q$ be the standard basis of F^q as a vector space over F, i.e., e_j has coordinate 1 at the jth entry and 0 everywhere else. For $J = \{j_1, \ldots, j_k\} \in S_{[q],k}$ with $1 \leq j_1 < \ldots < j_k \leq q$, let $e_J = e_{j_1} \wedge \cdots \wedge e_{j_k} \in \wedge^k F^q$. Then

$$\{e_J : J \in S_{[q],k}\}$$

is a basis for $\wedge^k F^q$. Any $C \in M_{p,q}(F)$ then defines the F-linear map

$$T_C : \wedge^k F^q \longrightarrow \wedge^k F^p, \;\; v_1 \wedge \cdots \wedge v_k \longmapsto Cv_1 \wedge \cdots \wedge Cv_k.$$

Note that $\wedge^k F^p = \{0\}$ if $k > p$. It follows from the definition of the determinant that for any $J \in S_{[q],k}$, one has

$$T_C(e_J) = \sum_{K \in S_{[p],k}} [B]_{K,J} e_K \in \wedge^k F^p,$$

where by convention, $[B]_{K,J} = 0$ if $k > p$.

Assume now that A and B are as given and regard them as matrices with entries in F. For $1 \leq k \leq \min\{m, l\}$, we then have linear maps

$$T_A : \wedge^k(F^n) \longrightarrow \wedge^k(F^m), \quad T_B : \wedge^k(F^l) \longrightarrow \wedge^k(F^n), \quad T_{AB} : \wedge^k(F^l) \longrightarrow \wedge^k(F^m),$$

and we have $T_{AB} = T_A \circ T_B$. For any $J \in S_{[l],k}$ we then have

$$\begin{aligned} T_B(e_J) &= \sum_{K \in S_{[n],k}} [B]_{K,J} e_K, \\ T_A(T_B(e_J)) &= \sum_{K \in S_{[n],k}} [B]_{K,J} T_A(e_K) = \sum_{K \in S_{[n],k}, I \in S_{[m],k}} [B]_{K,J}[A]_{I,K} e_I \\ &= \sum_{I \in S_{[m],k}} \left(\sum_{K \in S_{[n],k}} [A]_{I,K}[B]_{K,J} \right) e_I. \end{aligned}$$

It follows that for every $I \in S_{[m],k}$, one has

$$[AB]_{I,J} = \sum_{K \in S_{[n],k}} [A]_{I,K}[B]_{K,J}.$$

□

Remark 2.1.3. Note that when $k = 1$, (2.1) is just the definition of the product matrix, and when $m = n = p$, (2.1) just says that $\det(AB) = \det(A)\det(B)$.

From now on we assume that R is a PID.

Definition 2.1.4. Let R be PID. For $A \in M_{m,n}(R)$ and an integer $1 \leq k \leq \min\{m, n\}$, let $I_k(A)$ be the ideal of R generated by the set of of all $k \times k$ minors of A, and let $m_k(A)$ be a generator of $I_k(A)$. Let $m_0(A) = 1$. Note that when $I_k(A) \neq 0$, $m_k(A)$ is a gcd of the set of all non-zero $k \times k$ minors of A. The integer

$$s(A) = \max\{0 \leq k \leq \min\{m, n\} : I_k(A) \neq \{0\}\}$$

is called the *rank of A*.

We now have the following corollary of the Cauchy-Binet formula:

Proposition 2.1.5. *Let R be a PID. For any $A \in M_{m,n}(R)$, and any $P \in GL(m, R)$ and $Q \in GL(n, R)$, one has*

$$I_k(PAQ) = I_k(A) \quad \text{and} \quad s(A) = s(PAQ)$$

for all $1 \leq k \leq \max(m, n)$. In particular, for every $1 \leq k \leq s(A)$, $m_k(A) = m_k(PAQ)$ up to multiplication by units.

Proof. By Lemma 2.1.2, for any $P \in GL(m, R)$,

$$I_k(PA) \subset I_k(A), \quad I_k(A) = I_k(P^{-1}PA) \subset I_k(PA),$$

so $I_k(PA) = I_k(A)$. The rest of the statements now follows from the definitions. □

§ 2. Statement of the Smith Normal Form Theorem

For integers $m, n \geq 1$, $1 \leq s \leq \min(m, n)$, and $d_1, \ldots, d_s \in R$, we set

$$\operatorname{diag}(d_1, d_2, \ldots, d_s, 0, \ldots, 0) = \begin{pmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & d_s & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

the $n \times m$ matrix with d_k the (k, k)-entry for $1 \leq k \leq s$ and all the other entries 0.

Theorem 2.1.6. *(***Smith Normal Form Theorem***) Let R be a PID. For any non-zero $A \in M_{m,n}(R)$, there exist $P \in GL(m,R)$ and $Q \in GL(n,R)$, an integer $1 \leq s \leq n$, and $d_1, \ldots, d_s \in R\backslash\{0\}$ with $d_1|d_2|\cdots|d_s$, such that*

$$PAQ = \mathrm{diag}(d_1, d_2, \ldots, d_s, 0, \ldots, 0). \tag{2.2}$$

Moreover, the integer s is unique and the elements $d_1, \ldots, d_s$ of R are unique up to up to associates.

By Proposition 2.1.5, the integer s in Theorem 2.1.6 is the rank of A.

Definition 2.1.7. The non-zero elements $d_1, \ldots, d_s$ of R in Theorem 2.1.6 are called the *invariant factors* of A. The diagonal matrix $\mathrm{diag}(d_1, d_2, \ldots, d_s, 0, \ldots, 0)$ is called the *Smith normal form* of A.

We will prove the Smith Normal Form Theorem in §§ 3. Assuming the theorem, and we see how to compute the Smith Normal form of a given A using its entries.

Proposition 2.1.8. *Suppose that $A \in M_{m,n}(R)$ is non-zero and that there exist an integer $1 \leq s \leq n$, elements $d_1, \ldots, d_s \in R\backslash\{0\}$ with $d_1|d_2|\cdots|d_s$, and matrices $P \in GL(m,R)$ and $Q \in GL(n,R)$ such that (2.2) holds. Then s is the rank of A, and $d_k = u_k m_k(A)/m_{k-1}(A)$ for $1 \leq k \leq s$, where $u_1, \ldots, u_s$ can be any units of R.*

Proof. Proposition 2.1.5, for each $1 \leq k \leq s$, we have $m_k = d_1 \cdots d_k$ up to associates. □

Proposition 2.1.8 tells us that one can find the Smith normal form of any matrix A by computing all of its minors and thus the elements $m_1(A), m_2(A), \ldots,$.

Exercise 2.1.9. Find the Smith normal form of the integral matrices $A = \mathrm{diag}(4,3)$ and $B = \mathrm{diag}(4,2,6,3)$. (Ans: $A' = \mathrm{diag}(1,12)$ and $B' = \mathrm{diag}(1,2,6,12)$).

Exercise 2.1.10. Find the Smith normal form of the matrix $A \in M_{2,3}(\mathbb{C}[x])$, where

$$A = \begin{pmatrix} x-1 & x^2-1 & 0 \\ 0 & x^3-1 & x+1 \end{pmatrix}.$$

(Ans: $\begin{pmatrix} 1 & 0 & 0 \\ 0 & x-1 & 0 \end{pmatrix}$).

§ 3. Proof of the Smith Normal Form Theorem

Throughout this sub-section, let R be a PID. To prepare for the proof of Theorem 2.1.6, we recall the *four* types of *elementary row operations* and *elementary column operations* one can perform on a matrix $A \in M_{m,n}(R)$.

Type I. Replace a row Row_i of A by that row plus a multiple of another row Row_j and leave all the other rows unchanged; This amounts to left multiplying A by

$P \in GL(m, R)$ which has 1's on the diagonal and 0 at all the other entries except at the (i, j)'s entry.

Type II. Replace some row Row_i by $\gamma \mathrm{Row}_i$, where γ is a unit in R;

Type III. Interchange two rows;

Type IV. (This is the one where you need to use the assumption that R is a PID.) If $\alpha \in R\backslash\{0\}$ and $\beta \in R$ are such that $\alpha \nmid \beta$, let $\delta = \gcd(\alpha, \beta)$. As R is a PID, there exist $s, t \in R$ such that $\delta = s\alpha + t\beta$. Then

$$\det \begin{pmatrix} s & t \\ -\beta/\delta & \alpha/\delta \end{pmatrix} = 1 \quad \text{and} \quad \begin{pmatrix} s & t \\ -\beta/\delta & \alpha/\delta \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \delta \\ 0 \end{pmatrix}.$$

Performing this operation to the i'th and the j'th entries in a given column replaces the original entries α and β respectively by δ and 0. (See §8.4 of Goodman's book for examples).

Elementary column operations are similarly defined.

Definition 2.1.11. We say that two matrices in $M_{m,n}(R)$ are *row-equivalent* if one is transformed into the other by a sequence of the above four types of elementary row operations; similarly, two matrices are *column-equivalent* if one is transformed into the other by a sequence of the four types elementary column operations. Two matrices are said to be *equivalent* if one is transformed into the other by a sequence of the four elementary row or column operations.

Proposition 2.1.8 suggests that we should first try perform row and column operations to A so that the new matrix A' will have $d_1 = m_1(A)$ as its $(1, 1)$-entry. With the example of $R = \mathbb{Z}$ in mind, we should first spot the "smallest" non-zero entry of A and bring it to the $(1, 1)$-entry. For this purpose, recall that a PID is a UFD, and for each non-zero element $a \in R$, we have defined the "length" $l(a)$ to be the number of prime factors in the prime decomposition of a. For $a, b \in R\backslash\{0\}$, we say that *$a$ is smaller than b* if $l(a) < l(b)$.

Lemma 2.1.12. *One has the following properties of the length function l on R: for $a, b \in R\backslash\{0\}$,*

i) $l(ab) \geq \max(l(a), l(b))$;

ii) $l(a) = l(b)$ if a and b are associates;

iii) If $l(a) \leq l(b)$ and a does not divide b, then any greatest common divisor c of a and b satisfies $l(c) < l(a)$.

Lemma 2.1.13. *Suppose that A has a nonzero $(1, 1)$-entry α.*

1) If there is an element β in the first row or column of A that is not divisible by α, then A is equivalent to a matrix with smaller $(1, 1)$-entry.

2) If α divides all entries in the first row and column, then A is equivalent to a matrix with $(1, 1)$-entry equal to α and all other entries in the first row and column equal to zero.

Proof. 1) If there is an element β in the first row or column of A that is not divisible by α, then by performing a row or column operations on A of Type IV, we get an equivalent matrix whose $(1,1)$ entry is a gcd of α and β which is smaller than α.

2) If α divides all entries in the first row and column, then by performing Type I row or column operations we get a matrix equivalent to A that has $(1,1)$-entry equal to α and all other entries in the first row and column equal to zero. □

We now give the main step in proving the Smith Normal Form Theorem.

Lemma 2.1.14. *For any non-zero $A \in M_{m,n}(R)$, there exists $B \in M_{m,n}(R)$ equivalent to A which is of the form*

$$B = \begin{pmatrix} d_1 & 0 \\ 0 & C \end{pmatrix}, \tag{2.3}$$

where $d_1 \in R\backslash\{0\}$, $C \in M_{m-1,n-1}(R)$, and d_1 divides all the non-zero entries of C.

Proof. Let $\mathcal{A}$ be the set of all matrices in $M_{m,n}(R)$ that are equivalent to A. Let

$$E = \{e \in R : e \text{ is a non-zero entry of some } A' \text{ in } \mathcal{A}\}.$$

Then $\{l(e) : e \in E\}$ is a subset of $\mathbb{N}$, so there exists $d_1 \in E$ such that $l(d_1) = \min\{l(e) : e \in E\}$. Let $A' \in \mathcal{A}$ be such that d_1 is an entry of A'. By switching the rows and columns of A', we may assume that d_1 is the $(1,1)$-entry of A'. By 1) of Lemma 2.1.13, d_1 must divide all the non-zero entries of the first row and the first column of A', for otherwise we would have a matrix in $\mathcal{A}$ with an entry smaller than d_1. By 2) of Lemma 2.1.13, we can find $B \in \mathcal{A}$ of the form in (2.3) where $d_1 \in R\backslash\{0\}$, and $C \in M_{m-1,n-1}(R)$.

We claim that d_1 must divide every non-zero entry of C. If not, let e be a non-zero entry of C, say at (i,j) with $i,j \geq 2$, such that $d_1 \nmid e$. By adding the ith row of B to the first row and leaving all the other rows unchanged (a Type I row operation), we get $B' \in \mathcal{A}$ whose $(1,j)$-entry is e. By Lemma 2.1.13, we get $B' \in \mathcal{A}$ whose $(1,1)$-entry is smaller than d_1, contradicting the choice of d_1. □

Proof of the Smith Norm Theorem (Theorem 2.1.6): We use induction on $m+n$. If $m+n=2$, i.e., $m=n=1$, there is nothing to prove. Assume now that $m+n>2$. Lemma 2.1.14 shows that there exist $P_1 \in GL(m,R)$, $Q_1 \in GL(n,R)$ and $C \in M_{m-1,n-1}$ such that $P_1AQ_1 = \begin{pmatrix} d_1 & 0 \\ 0 & C \end{pmatrix}$. If $C=0$, or if $m=1$ or $n=1$, we are done. Assume that $C \neq 0$ and $m \geq 2$ and $n \geq 2$. By induction assumption, there exist an integer $2 \leq s \leq \min(m,n)$, elements $d_2, \ldots, d_s \in R\backslash\{0\}$ with $d_2 | \cdots | d_s$, and matrices $P_2 \in GL(m-1,R), Q_2 \in GL(n-1,R)$ such that

$P_2CQ_2 = \mathrm{diag}(d_2, \ldots, d_s, 0, \ldots, 0)$. As d_1 divides every entry of C, we see that $d_1|d_2|\cdots|d_s$. Let

$$P = \begin{pmatrix} 1 & 0 \\ 0 & P_2 \end{pmatrix} P_1 \quad \text{and} \quad Q = Q_1 \begin{pmatrix} 1 & 0 \\ 0 & Q_2 \end{pmatrix}.$$

Then $P \in GL(m, R), Q \in GL(n, R)$ and $PAQ = \mathrm{diag}(d_1, \ldots, d_s, 0, \ldots, 0)$ as desired.

Combining with Proposition 2.1.8, this finishes the proof of Theorem 2.1.6.

Remark 2.1.15. One can give a more detailed description of the diagonal process, as in § 8.4 of Goodman's book.

2.2 Modules of commutative rings

§ 1. Modules, sub-modules, and module homomorphisms

Definition 2.2.1. 1) A left module over a ring R is an abelian group M together with a map $R \times M \to M, (r, m) \mapsto rm$, satisfying

$$\begin{aligned}(r_1 r_2)m &= r_1(r_2 m),\\ (r_1 + r_2)m &= r_1 m + r_2 m,\\ r(m_1 + m_2) &= rm_1 + rm_2,\end{aligned}$$

where $r, r_1, r_2 \in R$ and $m, m_1, m_2 \in M$. For $r \in R$, we also call the map $M \to M : m \mapsto rm$, the *action* of r on M.

2) When R has an identity element 1, a left R-module M is said to be unital if $1m = m$ for all $m \in M$.

3) Given a left R-module M, a sub-module of M is an abelian subgroup N of M such that $rn \in N$ for all $r \in R$ and $n \in N$.

Right R-modules are defined similarly.

Exercise 2.2.2. Let M be a left R-module. Show that for all $r \in R$ and $m \in M$,

$$0m = r0 = 0 \quad \text{and} \quad r(-m) = -(rm) = (-r)m,$$

If R has an identity element 1 and M is unital, then $(-1)m = -m$.

Throughout the course we will always assume that R has an identity element and R-modules are unital.

Example 2.2.3. Here are some examples.

1) A unital module over a field K is the same as a K-vector space.

2) R is itself a left R-module by left multiplication; the R-sub-modules of R are precisely the left ideals of R.

3) If R is a commutative ring with identity element, and $I \subset R$ an ideal, then R/I is naturally a unital R-module by $r(r_1 + I) = rr_1 + I$ for $r, r_1 \in R$.

4) Any abelian group A is a unital $\mathbb{Z}$-module.

5) Let V be a vector space over a field K and let $T \in \mathrm{End}_K(V)$, i.e., T is a K-linear operator on V. Then V is a unital $K[x]$-module via

$$\left(\sum_{i=0}^{n} a_i x^i\right) v := \sum_{i=0}^{n} a_i (T^i v),$$

and $K[x]$-sub-modules are precisely linear subspaces W of V that are invariant under T. Moreover, every $K[x]$-module structure on V arises this way by letting T be the action of $x \in K[x]$ on V.

6) If M is an R-module and $S \subset M$, then

$$RS := \{r_1 s_1 + \cdots + r_n s_n : n \in \mathbb{N}, r_i \in R, s_i \in S\}$$

is a sub-module of M called the *sub-module of M generated by S*.

Definition 2.2.4. The *direct sum* of R-modules $M_1, \ldots, M_n$, denoted as $M_1 \oplus \cdots \oplus M_n$, is the direct product abelian group $M_1 \times \cdots \times M_n$ equipped with the R-module structure given by

$$r(m_1, \ldots, m_n) = (rm_1, \ldots, rm_n), \quad r \in R,\, m_j \in M_j.$$

When $M_j = M$ for every $1 \leq j \leq n$, we denote the direct sum $M \oplus \cdots \oplus M$ (n copies) by M^n. In particular, one has the R-module R^n for each integer $n \geq 1$.

Definition 2.2.5. An R-module M is said to be *finitely generated* if there exists a finite subset S of M such that $RS = M$.

The goal of this chapter is to *classify* finitely generated modules of PIDs. The classification will be up to isomorphisms.

Definition 2.2.6. 1) Let M and N be two modules over an ring R. An R-module homomorphism is a homomorphism $\phi : M \to N$ of abelian groups such that

$$\phi(rm) = r\phi(m), \quad \forall r \in R,\, m \in M.$$

The set of all R-module homomorphisms from M to N is denoted by $\mathrm{Hom}_R(M, N)$. If $\phi \in \mathrm{Hom}_R(M, N)$ is bijective, it is called an R-module *isomorphism*. When such an isomorphism exists, we say that M and N are isomorphic as R-modules.

2) An R-module homomorphism from an R-module M to itself is called an *R-module endomorphism*, and the set of all R-module endomorphisms of M is denoted by $\mathrm{End}_R(M)$.

3) For $\phi \in \mathrm{Hom}_R(M, N)$, define the *kernel* and the *image* of ϕ to be

$$\ker(\phi) = \{m \in M : \phi(m) = 0\}, \quad \mathrm{im}(\phi) = \{\phi(m) : m \in M\}.$$

Then $\ker(\phi)$ is an R-sub-module of M, and $\mathrm{im}(\phi)$ is a R-sub-module of N.

Let M be an R-module and $N \subset M$ an R-sub-module. Form the quotient M/N of abelian groups and let $\pi : M \to M/N$ be the projection homomorphism of abelian groups.

Lemma 2.2.7. *The quotient M/N is an R-module via*

$$r(m + N) = rm + N, \quad r \in R,\, m \in M,$$

and $\pi : M \to M/N$ is an R-module morphism.

Proof. Straightforward. □

Theorem 2.2.8. *(Isomorphism Theorem) For any $\phi \in \mathrm{Hom}_R(M,N)$, the map*

$$\bar{\phi}: \ M/\ker(\phi) \longrightarrow \mathrm{im}(\phi), \quad m+\ker(\phi) \longmapsto \phi(m), \qquad m \in M,$$

is a well-defined R-morphism isomorphism.

Proof. Exercise (or read Theorem 8.2.5 of Goodman's notes). □

§ 2. Annihilators and torsion modules

Consider the abelian group $M = \mathbb{Z}/n\mathbb{Z}$ as a $\mathbb{Z}$-module. Then $nx = 0$ for all $x \in M$.

Lemma-Definition 2.2.9. Let R be any integral domain and M any R-module.

1) For $x \in M$, define $\mathrm{ann}(x) = \{r \in R : rx = 0\}$ and call it the *annihilator* of x. Then $\mathrm{ann}(x)$ is an ideal of R and $R/\mathrm{ann}(x) \cong Rx$. The set

$$\mathrm{ann}(M) = \cap_{x\in M}\mathrm{ann}(x) = \{r \in R : rx = 0 \ \forall\, x \in M\}$$

is an ideal of R, called the *annihilator ideal* of M.

2) An element $x \in M$ is called a *torsion element* if $\mathrm{ann}(x) \neq 0$. A module M is said to be *torsion free* if it has no non-zero torsion element.

3) Let M_{tor} be the set of all torsion elements in M. Then M_{tor} is an R-sub-module of M, called the *torsion sub-module* of M. The quotient module M/M_{tor} is torsion free. An R-modules is called a *torsion module* if $M = M_{\mathrm{tor}}$.

Proof. 1) is clear and nothing to prove for 2). For 3), if $x, y \in M_{\mathrm{tor}}$ and $r_1 \neq 0$ and $r_2 \neq 0$ are such that $r_1 \in \mathrm{ann}(x)$ and $r_2 \in \mathrm{ann}(y)$, then for any $s, t, \in R$, $r_1r_2(sx+ty) = 0$ and $r_1r_2 \neq 0$. This shows that $sx + ty \in M_{\mathrm{tor}}$. Thus M_{tor} is an R-sub-module of M. Let $x \in M$ and $x \notin M_{\mathrm{tor}}$. If $r \in R$ is such that $rx \in M_{\mathrm{tor}}$, then there exists $r' \in R\backslash\{0\}$ such that $r'rx = 0$. As $x \notin M_{\mathrm{tor}}$, $rr' = 0$. As R is an integral domain, $r = 0$. Thus M/M_{tor} is torsion free. □

Example 2.2.10. A finite abelian group is a finitely generated torsion module of $\mathbb{Z}$.

Example 2.2.11. Let K be any field and V a finite-dimensional K-vector space. For any $T \in \mathrm{End}_K(V)$, the $K[x]$-module V given by

$$fv = f(T)(v), \qquad f \in K[x],\, v \in V,$$

is a finitely generated torsion module of $K[x]$. Indeed, a basis of V as a vector space over K generates V as a K-module and thus also generates V as a $K[x]$-module. Let

$$f(x) = \det(T - x\mathrm{Id}_V) \in K[x]$$

be the characteristic polynomial of T. By Cayley - Hamilton theorem, $f(T) = 0$, so $f(T)(v) = 0$ for every $v \in V$. Thus V is a torsion module of $K[x]$.

§ 3. Cyclic modules over PIDs

Recall that all the commutative rings in this course are assumed to have identity 1.

Definition 2.2.12. A module M of a commutative ring R is said to be *cyclic* if it is generated by one element.

For an ideal I of R, the quotient R/I is a cyclic R-module with $a = 1 + I \in R/I$ as a generator. If I is non-zero and $I \neq R$, then R/I is not the zero module. The annihilator ideal of a is exactly I, so R/I is a torsion module if $I \neq 0$. The next lemma says that all cyclic modules of R are of the form R/I for an ideal I of R.

Lemma 2.2.13. *If M is a cyclic module of a commutative ring R with $a \in M$ as a generator, then the map*

$$\phi: \quad R/\mathrm{Ann}(a) \longrightarrow M, \quad r + \mathrm{Ann}(a) \longmapsto ra, \qquad r \in R,$$

is an R-module isomorphism.

Proof. Straightforward from the definitions. □

We now assume that R is a PID, and let $I \subset R$ be an ideal. What does R/I look like? Recall now the Chinese Remainder Theorem.

Theorem 2.2.14. *Let R be a PID and let $p_1, p_2, \ldots, p_k$ be distinct primes in R. Let $n_1, n_2, \ldots, n_k$ be positive integers. Then one has the R-module isomorphisms.*

$$R/\langle p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \rangle \cong (R/\langle p_1^{n_1} \rangle) \times (R/\langle p_2^{n_2} \rangle) \times \cdots \times (R/\langle p_k^{n_k} \rangle).$$

§ 4. Free modules

In this section, we again assume that R is a commutative ring with an identity element.

Definition 2.2.15. Let M be an R-module.

1) A subset S of M is said to be R-linearly independent if for any distinct elements $x_1, \ldots, x_n$ of M and any $r_1, \ldots, r_n \in R$,

$$r_1 x_1 + \cdots + r_n x_n = 0$$

implies that $r_1 = \cdots = r_n = 0$.

2) A *basis* of M is a linearly independent subset S of M that also spans M, i.e., such that $RS = M$.

3) M is said to be a *free* R-module if it has a basis.

Lemma 2.2.16. *If M is a free R-module with a basis $\{x_1, x_2, \ldots, x_n\}$, then the map*

$$R^n \longrightarrow M, \quad (r_1, r_2, \ldots, r_n) \longmapsto r_1 x_1 + x_2 x_2 + \cdots + r_n x_n, \qquad r_j \in R,$$

is an isomorphism of R-modules.

Proof. Straightforward. □

Proposition 2.2.17. *Let M be a non-zero finitely generated and free R-module. Then there is a unique integer $n \geq 1$ such that $M \cong R^n$. The integer n is called the rank of M.*

Proof. **Step 1.** Every basis of M must be finite: Let B be a basis and S a finite generating set. Then every element in S is a finite R-linear combination of elements in B, and since S is finite and generates M, a finite subset B_0 of B generates M. If $B \neq B_0$, then there exists $b \in B \backslash B_0$ and b is a R-linear combination of elements in B_0, contradicting the assumption that B is linearly independent. Thus $B = B_0$ is a finite set.

Step 2. Let $\{v_1, \ldots, v_n\}$ be a basis of M and $\{g_1, \ldots, g_m\}$ be a generating set. We prove that $m \geq n$. Let A be the $n \times m$ matrix with values in R and B an $m \times n$ matrix with values in R such that

$$(v_1, \ldots, v_n)A = (g_1, \ldots, g_m) \quad \text{and} \quad (g_1, \ldots, g_m)B = (v_1, \ldots, v_n).$$

Then $(v_1, \ldots, v_n)AB = (v_1, \ldots, v_n)$. Since $\{v_1, \ldots, v_n\}$ is a basis, $AB = \mathrm{I}_n$, the $n \times n$ identity matrix. If $n > m$, add $n - m$ column of 0's and $n - m$ rows of 0's to A and B respectively to get

$$(A, 0) \begin{pmatrix} B \\ 0 \end{pmatrix} = AB = I_n,$$

but $\det(A, 0) = 0$, a contradiction. Thus $n \leq m$. As any basis is in particular a generating set, we have shown that any two bases have the same cardinality. □

Note that we used the fact that determinants can be defined on square matrices with coefficients in a commutative ring R and that $\det(AB) = \det(A)\det(B)$ for any two square matrices of the same size.

Remark 2.2.18. We make an important remark that if F is a free R-module of rank n and if $(v_1, \ldots, v_n)$ is a basis for F, then for any $P \in GL(n, R)$,

$$(w_1, \ldots, w_n) := (v_1, \ldots, v_n)Q$$

is also a basis of F. Moreover any two basis of F are related this way.

Remark 2.2.19. Torsion free modules are not necessarily free. For example, any ideal I of R regarded as an R-module by multiplication is torsion free, but not necessarily free: consider the example of $R = K[x, y]$ where K is a field and let $I = \langle x, y \rangle$, the ideal of R generated by the two elements x and y. As an R-module, it is thus finitely generated. If it were free, the proof of Proposition 2.2.17 shows that it would be a free R-module of rank 1 or 2. If it has rank 1, let f be a basis. Then $\deg(f) > 0$ because $I \neq R$, and since $x \in I$ one must have $f = kx$ for some $k \in K$. Similarly one would have $f = k'y$ for $k' \in K$, contradiction. Suppose that I has rank

2 as an R-module and let $\{f, g\}$ be a basis. As $x, y \in I$, there exist $a, b, c, d \in R$ such that $x = af + bg$ and $y = cf + dg$. It follows from $xy - yx = 0$ that

$$y(af + bg) - x(cf + dg) = 0$$

so $ay = cx$ and $by = dx$. It follows that $x|a$ and $x|b$, so $x = x(a_1 f + b_1 g)$ for some $a_1, b_1 \in R$ and thus $1 = a_1 f + b_1 g \in I$, contradiction.

Example 2.2.20. Let R be a PID and let $A = (a_{ij}) \in M_{m,n}(R)$. Identifying

$$R^n \cong M_{n,1}(R) \quad \text{and} \quad R^m \cong M_{m,1}(R),$$

we have the R-module map

$$L_A : \mathbb{R}^n \longrightarrow R^m,\ L_A(x) = Ax.$$

Here we write an element in R^n as a column with entries in R. We then have the sub-modules

$$\mathrm{Ker}(L_A) \subset R^n \quad \text{and} \quad \mathrm{Im}(L_A) \subset R^m.$$

By definition, $\mathrm{Im}(L_A)$ is also the sub-module of R^n generated by the columns of A. Let $P \in GL(m, R)$ and $Q \in GL(n, R)$ be such that $A = PDQ^{-1}$, where

$$D = \mathrm{diag}(d_1, \ldots, d_s, 0, \ldots, 0)$$

is the Smith normal form of A. Let $\{p_1, \ldots, p_m\}$ be the columns of P and let $\{q_1, \ldots, q_n\}$ be the columns of Q. Then

1) $\{p_1, \ldots, p_m\}$ is a basis of R^m and $\{d_1 p_1, \ldots, d_s p_s\}$ is a basis of $\mathrm{Im}(L_A) \subset R^m$;

2) $\{q_1, \ldots, q_n\}$ is a basis of R^n and $\{q_{s+1}, \ldots, q_n\}$ is a basis of $\mathrm{Ker}(L_A) \subset R^n$.

Indeed, as both P and Q are invertible, $\{p_1, \ldots, p_m\}$ is a basis for R^m and $\{q_1, \ldots, q_n\}$ is a basis for R^n. Rewriting $A = PDQ^{-1}$ as

$$(Aq_1, \ldots, Aq_s, Aq_{s+1}, \ldots, Aq_n) = (d_1 p_1, \ldots, d_s p_s, 0, \ldots, 0), \tag{2.1}$$

it is clear that $\mathrm{Im}(L_A) \subset R^m$ is generated by $\{Aq_1 = d_1 p_1, \ldots, Aq_s = d_s p_s\}$. It is also clear that $\{d_1 p_1, d_2 p_2, \ldots, d_s p_s\}$ is a linearly independent subset of R^n.

Turning to the R-sub-module $\ker(L_A)$ of R^n, it also follows from (2.1) that $q_{s+1}, \ldots, q_n \in \ker(L_A)$. As $\{q_1, \ldots, q_n\}$ is a basis of R^n, $\{q_{s+1}, \ldots, q_n\}$ is a linearly independent subset of R^n. It remains to show that $\{q_{s+1}, \ldots, q_n\}$ generates $\ker(L_A)$. Note now that for $x \in R^n$, setting $y = Q^{-1}x$, then $Ax = 0$ if and only if $Dy = 0$, which is equivalent to $y = \begin{pmatrix} 0 \\ z \end{pmatrix}$ where $z \in R^{n-s}$ is arbitrary. We thus conclude that $\ker(L_A)$ is the R-sub-module of R^n generated by $q_{s+1}, \ldots, q_n$.

Example 2.2.21. The integer matrix $A = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ -4 & 4 \end{pmatrix} \in M_{3,2}(\mathbb{Z})$ defines a $\mathbb{Z}$-module map $L_A : \mathbb{Z}^2 \to \mathbb{Z}^3$. One has

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ -4 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix},$$

which we now rewrite as

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-1}.$$

Let p_1, p_2, p_3 be the three columns of $\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -4 & 0 & 1 \end{pmatrix}$. Then $\{p_1, p_2, p_3\}$ is a basis for $\mathbb{Z}^3$ and $\{p_1, 2p_2\}$ is a basis of $\mathrm{Im}(L_A)$. As A has rank 2, $\ker(L_A) = 0$.

2.3 Structure theorem on sub-modules of finite rank free modules over PIDs

Let again R be a PID. Recall that every free R-module of finite rank is isomorphic to R^n for some $n \geq 1$. We first explain why we want to look at sub-modules of R^n.

Recall that we are interested in finitely generated R-modules. In particular, recall that every finite abelian group is a finitely generated $\mathbb{Z}$-module, and that every K-vector space V together with a $T \in \mathrm{End}_K(V)$ is finitely generated $K[x]$-module. We want to understand these examples of finitely generated R-modules. The following Proposition 2.3.1 explains how finitely generated R-modules are related to sub-modules of R^n.

Proposition 2.3.1. *Let R be any commutative ring. An R-module M is finitely generated if and only if there exist an integer $n \geq 1$ and a sub-module N of R^n such that $M \cong R^n/N$.*

Proof. Assume first that $M \cong R^n/N$ for some integer $n \geq 1$ and some sub-module N of R. Let $\pi : R^n \to R^n/N$ be the natural projection and let $\phi : R^n/N \to M$ be an isomorphism. Let $\{e_1, \ldots, e_n\}$ be the standard basis of R^n. Then

$$\{\phi(\pi(e_1)), \ldots, \phi(\pi(e_n))\}$$

is a set of generators for M as an R-module. Thus M is finitely generated.

Conversely, assume that M is a finitely generated R-module and let $\{x_1, \ldots, x_n\}$ be a set of generators of M. Define the R-module homomorphism

$$\phi: \ R^n \longrightarrow M, \ \ \phi(r_1e_1 + \cdots + r_ne_n) = r_1x_1 + \cdots + r_nx_n.$$

Then ϕ is surjective, and one thus has the R-module isomorphism

$$[\phi]: \ R^n/\ker(\phi) \longrightarrow M.$$

□

To have a better understanding of sub-modules of free modules, let's ask the following question:

Question. Are sub-modules of a finite rank free module of any commutative ring R necessarily free?

The following example shows that the answer may be no if R is not a PID.

Example 2.3.2. Let R be any commutative ring, and consider the special case when $M = R$, i.e., R is regarded as an R-module. Then $N \subset R$ is an R-sub-module if and only of N is an ideal. The question then becomes when an ideal I of R is free as an R-module. Assume that I is a non-zero ideal of R which is free as an R-module. Then I has a basis B. On the other hand, for any subset B of R containing at least two non-zero elements b_1 and b_2, we have

$$b_2b_1 + (-b_1)b_2 = 0,$$

so B is not a linearly independent set. This shows that any linearly independent subset of R can contain only one element. In particular, any basis B of I can only contain one element, which is same as saying that I is a principal ideal. On the other hand, a principal ideal I of R is free of rank one. Thus an ideal of R, regarded as an R-module, is free if and only if I is principal. We thus conclude that if R is not a PID, a sub-module of a free R-module is not necessarily free.

Lemma 2.3.3. *Let R be a PID. If F is a free R-module of rank n, then any R-sub-module N of M can be generated by a subset with no more than n elements.*

Proof. We use induction on n. When $n = 1$, since R is assumed to be a PID, we have already seen in Example 2.3.2 that every sub-module of R, being an ideal of R, is free and has rank 1.

Let $n \geq 2$ and assume now that any sub-module of a free module of rank $n-1$ is generated by a subset of no more than $n-1$ elements. Let F be a free R-module of rank n with a basis $\{x_1, \ldots, x_n\}$, and let N be any sub-module of F. Consider the R-module homomorphism

$$\phi: \ F \longrightarrow R, \quad r_1x_1 + \cdots + r_nx_n \longmapsto r_nx_n.$$

Then $\phi(N) \subset R$, being a sub-module of R as an R-module, is an ideal of R. Since R is a PID, $\phi(N) = aR$ for some $a \in R$. Choose any $y \in N$ such that $\phi(y) = a$. Then for any $z \in N$, one has $\phi(z) = ra$ for some $r \in R$, and thus $z = z' + ry$ for some $z' \in N \cap \ker(\phi)$. This shows that

$$N = N \cap \ker(\phi) + Ry.$$

Note now that $N \cap \ker(\phi)$ is a sub-module of $\ker(\phi) = Rx_1 + \cdots + Rx_{n-1}$, which is a free R-module of rank $n-1$. By the induction assumption, there is a subset S of $N \cap \ker(\phi)$ with no more than $n-1$ elements such that S generates $N \cap \ker(\phi)$. If $y = 0$, S generates N. If $y \neq 0$, then $S \cup \{y\}$ generates N. □

Back to the case of a finitely generated R-module M, let again $\{x_1, \ldots, x_n\}$ be a set of generators for M, and consider the surjective homomorphism

$$\phi : \ R^n \longrightarrow M, \ \phi(r_1 e_1 + \cdots + r_n e_n) = r_1 x_1 + \cdots + r_n x_n.$$

We have seen that ϕ induces an R-module isomorphism

$$[\phi] : \ R^n / \ker(\phi) \longrightarrow M.$$

By Lemma 2.3.3, there exist $\{g_1, \ldots, g_m\} \subset R^n$ that generates $\ker(\phi)$. Regarding each g_j, for $j = 1, \ldots, m$, as a column with entries in R, we get a matrix

$$G = (g_1, g_2, \ldots, g_m) \in M_{n,m}(R)$$

whose columns generate $\ker(\phi)$. Consider the sequence of R-module homomorphisms

$$R^m \xrightarrow{L_G} R^n \xrightarrow{\phi} M \longrightarrow 0, \tag{2.2}$$

where $L_G : R^m \to R^n, L_G(x) = Gx$. The sequence in (2.2) is *exact* in the sense that ϕ is surjective and $\ker(\phi) = \mathrm{Im}(L_G)$. The exact sequence in (2.2) is called a *presentation* of the finitely generated module M.

We now prove the following *Structure Theorem for Sub-modules of Finite Rank Free Modules of PIDs.*

Theorem 2.3.4. *Let F be a free R-module of rank n and $N \subset F$ a sub-module.*

1) There exist a basis $\{v_1, \ldots, v_n\}$ of F, an integer $0 \leq s \leq n$, and non-zero elements $d_1, \ldots d_s \in R$ with $d_i | d_j$ for $1 \leq i \leq j \leq s$, such that $\{d_1 v_1, \ldots, d_s v_s\}$ is a basis of N.

2) The integer s is unique and the elements $d_1, \ldots, d_s$ are unique up to multiplication by units in R and they (i.e., s and the elements $d_1, \ldots, d_s$) are independent of the choices of the basis $\{v_1, \ldots, v_n\}$ of F.

Proof. By Lemma 2.3.3, N has a generating set of no more than n elements. Let $\{g_1, \ldots, g_s\}$ be a generating set for N of minimum cardinality, where $s \leq n$. Let $\{e_1, \ldots, e_n\}$ be a basis of F and write

$$(g_1, \ldots, g_s) = (e_1, \ldots, e_n)A \tag{2.3}$$

where $A \in M_{n,s}(R)$. Applying the Smith normal form theorem to A, one knows that there exist $P \in GL(n, R)$ and $Q \in GL(s, R)$ such that

$$PAQ = \begin{pmatrix} D \\ 0 \end{pmatrix},$$

where $D = \mathrm{diag}(d_1, \ldots, d_s)$, and $d_1, \ldots, d_s \in R$ with $d_i | d_j$ for $i \leq j$. Note that here we have s bigger than or equal to the rank of A. Rewrite (2.3) as

$$(g_1, \ldots, g_s)Q = (e_1, \ldots, e_n)P^{-1} \begin{pmatrix} D \\ 0 \end{pmatrix},$$

and set

$$(v_1, \ldots, v_n) = (e_1, \ldots, e_n)P^{-1} \quad \text{and} \quad (w_1, \ldots, w_s) = (g_1, \ldots, g_s)Q.$$

One then has

$$(w_1, \ldots, w_s) = (d_1 v_1, \ldots, d_s v_s).$$

As Q is invertible, $\{w_1, \ldots, w_s\}$ is a generating set of N. As s is the minimal cardinality among all the generating sets of N, $d_j \neq 0$ for any $1 \leq j \leq s$. It follows that $\{w_1, \ldots, w_s\}$ is a linearly independent set, and is thus a basis for N.

Note now that the integer s, being the minimal size of all the generating sets of N, is independent of the choices of the basis $\{v_1, \ldots, v_n\}$ of F. Suppose that $\{u_1, \ldots, u_n\}$ is another basis of F and $c_1, \ldots, c_s \in R$ are such that $c_i | c_j$ for $1 \leq i < j \leq s$ and that $\{c_1 u_1, \ldots, c_s u_s\}$ is a basis of N. Witting

$$(u_1, \ldots, u_n) = (v_1, \ldots, v_n)X^{-1} \quad \text{and} \quad (c_1 u_1, \ldots, c_s u_s) = (d_1 v_1, \ldots, d_s v_s)Y,$$

where $X \in GL(n, R)$ and $Y \in GL(s, R)$, one has

$$X \begin{pmatrix} D \\ 0 \end{pmatrix} Y = \begin{pmatrix} C \\ 0 \end{pmatrix}.$$

By Theorem 2.1.6 c_j and d_j are associates for each $1 \leq j \leq n$. □

Corollary 2.3.5. *A sub-module of a free module of rank n over a PID is again a free module of rank $r \leq n$.*

Definition 2.3.6. In the context of Theorem 2.3.4, the integer s is called the *rank* of N, and the elements $d_1, \ldots, d_s$ the *invariant factors* of N.

Definition 2.3.7. Given a sub-module N of a finite rank free R-module F, we say that a basis $\{f_1, \ldots, f_n\}$ of F is *N-admissible* if there exist non-zero elements $d_1, \ldots, d_s \in R$ with $d_1|d_2|\cdots|d_s$ such that $\{d_1 f_1, \ldots, d_s f_s\}$ is a basis for N.

By Theorem 2.3.4, when R is a PID, N-admissible bases exist for any sub-module N of a finite rank free module of R, and that s is unique and the elements $d_1, \ldots, d_s$ are unique up to multiplication by units.

Example 2.3.8. Recall from Example 2.2.20 that each $A = (a_{ij}) \in M_{m,n}(R)$ defines an R-module map $L_A : \mathbb{R}^n \to R^m$ by $L_A(x) = Ax$, where we regard elements in R^n as columns with entries in R. Let $P \in GL(m, R)$ and $Q \in GL(n, R)$ be such that $A = PDQ^{-1}$. Let $\{p_1, \ldots, p_m\}$ be the columns of P and let $\{q_1, \ldots, q_n\}$ be the columns of Q. Then $\{p_1, \ldots, p_m\}$ is an $\mathrm{Im}(L_A)$-admissible basis of R^m and $\{q_1, \ldots, q_n\}$ is a $\mathrm{Ker}(L_A)$-admissible basis of R^n. More precisely,

1) $\{p_1, \ldots, p_m\}$ is a basis of R^m and $\{d_1 p_1, \ldots, d_s p_s\}$ is a basis of $\mathrm{Im}(L_A) \subset R^m$;

2) $\{q_1, \ldots, q_n\}$ is a basis of R^n and $\{q_{s+1}, \ldots, q_n\}$ is a basis of $\mathrm{Ker}(L_A) \subset R^n$.

2.4 Structure theorems on finitely generated modules over PIDs

The following Theorem 2.4.1 is called the *Structure Theorem of Finitely Generated Modules over PIDs: Invariant Factor Form.*

Theorem 2.4.1. *Let R be a PID and M a finitely generated R-module. There exist non-units $d_1, \ldots, d_s \in R\backslash\{0\}$ with $d_1|d_2|\cdots|d_s$, and an integer $r \geq 0$ such that*

$$M \cong R/(d_1) \oplus \cdots \oplus R/(d_s) \oplus R^r. \tag{2.1}$$

If $c_1, \ldots, c_t \in R\backslash\{0\}$, non-units with $c_1|c_2|\cdots|c_s$, and an integer $r' \geq 0$ are such that

$$M \cong R/(c_1) \oplus \cdots \oplus R/(c_t) \oplus R^{r'},$$

then $s = t, r = r'$, and d_i and c_i are associates for each $1 \leq i \leq s = t$.

Proof of the existence part of Theorem 2.4.1. Let M be a finitely generated module over R, and let $\{x_1, \ldots, x_n\}$ be a set of generators of *minimal cardinality.* Consider the R-module R^n with the standard basis $\{e_1, \ldots, e_n\}$ and define an R-module homomorphism

$$\phi: \ R^n \longrightarrow M, \ \ \phi(r_1 e_1 + \cdots + r_n e_n) = r_1 x_1 + \cdots + r_n x_n.$$

Then ϕ is surjective, and one thus has the R-module isomorphism

$$[\phi]: \ R^n/\ker(\phi) \longrightarrow M.$$

Consider the sub-module $\ker(\phi)$ of the free module R^n. Let $d_1, \ldots, d_s$ be the invariant factors of $\ker(\phi)$. Using a $\ker(\phi)$-admissible basis for R^n whose existence is guaranteed by the Structure Theorem for sub-modules of finite rank free modules,

$$M \cong R^n/\ker(\phi) \cong R/(d_1) \oplus \cdots \oplus R/(d_s) \oplus R^{n-s}.$$

By definition, $d_j \neq 0$ for each j. If d_j is a unit for some j, then the module

$$R/(d_1) \oplus \cdots \oplus R/(d_s) \oplus R^{n-s}$$

is generated by less than n elements, contradicting the minimality of n. Thus d_j is a non-unit for each $1 \leq j \leq s$. This finishes the proof of the existence part of Theorem 2.4.1.

To prove the uniqueness part of Theorem 2.4.1, we describe the integers s, r, and $d_1, \ldots, d_s$ in the decomposition (2.1) of M in terms of M.

Notation 2.4.2. Let R be any commutative ring. If M is a finitely generated R-module, let $l(M)$ be the least number of generators needed to generate M (by definition, M is generated by *zero generators* iff $M = 0$). For $a \in R$, let aM be the sub-module aM of M given by

$$aM = \{ax : x \in M\} \subset M.$$

If $\{x_1, \ldots, x_l\} \subset M$ generates M, then $\{ax_1, \ldots, ax_l\} \subset aM$ generates aM, so aM is finitely generated. For integer $k \geq 0$, let

$$I_k(M) = \{a \in R : l(aM) \leq k\} \subset R.$$

Lemma 2.4.3. *Let R be any commutative ring and M a finitely generated R-module. Then for any $k \geq 0$, $I_k(M)$ is closed under multiplication by R, and one has $I_k(M) = R$ for all $k \geq l(M)$, and*

$$\mathrm{Ann}(M) = I_0(M) \subset I_1(M) \subset I_2(M) \subset \cdots \subset I_{l(M)}(M) = R. \tag{2.2}$$

Moreover, if $M \neq 0$, then $I_{l(M)-1}(M) \neq R$.

Proof. We have noticed that if $\{x_1, \ldots, x_l\} \subset M$ generates M, then for any $a \in R$, $\{ax_1, \ldots, ax_l\}$ generates aM. Applying this statement to any $a_1, a_2 \in R$, we see that

$$l(a_1 a_2 M) \leq l(a_2 M).$$

Thus $I_k(M)$ is closed under multiplication by R. It is also clear that $I_0(M) = \mathrm{Ann}(M)$, and

$$I_{k_1}(M) \subset I_{k_2}(M) \quad \text{if} \quad k_1 \leq k_2.$$

For $k \geq l(M)$, we have $1 \in I_k(M)$ so $I_k(M) = R$. Suppose that $M \neq 0$ so that $l(M) \geq 1$. If $I_{l(M)-1}(M) = R$, then $1 \in I_{l(M)-1}(M)$, so $l(M) \leq l(M) - 1$, a contradiction. Thus $I_{l(M)-1}(M) \neq R$ whenever M is non-zero. $\square$

Notation 2.4.4. Let R be a PID and M a finitely generated R-module. We denote by $r(M)$ the number of zero ideals in the sequence (2.2) of ideals of R. For $0 \leq k \leq l(M)$, let $\alpha_k(M) \in R$ be a gcd of $I_k(M)$. Note that $\alpha_k(M)$ is defined up to associates. By Lemma 2.4.3 and writing $l = l(M)$, $r = r(M)$, $\alpha_k = \alpha_k(M)$ for $k = r, r+1, \ldots, l-1$, one has

$$\{\alpha_0, \alpha_1, \ldots, \alpha_l\} = \{(\overbrace{0, \ldots, 0}^{r}, \alpha_r, \alpha_{r+1}, \ldots, \alpha_{l-1}, 1\},$$

and the elements $\alpha_r, \alpha_{r+1}, \ldots, \alpha_{l-1}$ are all non-zero non-units, and

$$\alpha_{l-1} | \cdots | \alpha_{r+1} | \alpha_r. \tag{2.3}$$

In terms of the sequence (2.2) of ideals of R associated to M, one has

$$\mathrm{Ann}(M) = \overbrace{I_0(M) \subset I_1(M) \subset \cdots \subset I_{r-1}(M)}^{\text{all } 0} \subset \overbrace{I_r(M) \subset \cdots \subset I_{l-1}(M)}^{\text{neither } 0, \text{ nor } R} \subset I_l(M) = R.$$

Definition 2.4.5. Let R be a PID. For a finitely generated R-module, the non-zero non-units $\alpha_r, \alpha_{r+1}, \ldots, \alpha_{l-1}$ in Notation 2.4.4 are called the *invariant factors of* M.

Example 2.4.6. By definition, if $M = 0$, then $l(M) = 0$ and $r(M) = 0$. Take $M = R^n$, where $n \geq 1$. Then $l(M) = n$. If $\{e_1, \ldots, e_n\}$ is a basis of R^n, then for any non-zero $a \in R$, $\{ax_1, \ldots, ax_n\}$ is a basis for aM, so $l(aM) = n$. Thus

$$I_k(M) = \begin{cases} 0, & 0 \leq k \leq n-1, \\ R, & k = n. \end{cases}$$

Thus $r(R^n) = n$.

Example 2.4.7. Let R be a PID and let $M = R/(d)$, where $d \in R$ is non-zero and non-unit. Then $l(M) = 1$ and $I_0(M) = \mathrm{Ann}(M) = (d)$. In particular, $r(M) = 0$.

We need a some more preparation before proving the uniqueness part of Theorem 2.4.1. We make the following observation.

Lemma 2.4.8. *For any commutative ring R and any R-module M, if $a \in R$, then M/aM is naturally a module of the quotient ring $R/(a)$ via*

$$(r + aR)(x + aM) = rx + aM, \quad r \in R,\, x \in M. \tag{2.4}$$

In particular, if R is a PID and $p \in R$ is a prime element, then M/pM is a vector space over the field $R/(p)$.

Proof. Follows directly from the definitions. □

Lemma 2.4.9. *Let R be a PID and M a finitely generated R-module. Let p be a prime element in R. Then M/pM is a finite dimensional vector space over the field $R/(p)$ with dimension $d \leq l(M)$. In particular, if M is cyclic and $M/pM \neq 0$, then $M/pM \cong R/(p)$ as $R/(p)$-vector spaces.*

Proof. If $\{x_1, \ldots, x_k\}$ is a set of generators for M as an R-module, then

$$\{x_1 + pM, \ldots, x_k + pM\}$$

generates M/pM as a vector space over $R/(p)$. Thus M/pM is a finite dimensional vector space over $R/(p)$ with dimension $d \leq k$. In particular, $d \leq l(M)$. If M is cyclic and $M/pM \neq 0$, then $l(M) = 1$ and $d \neq 0$, so $d = 1$, and thus $M/pM \cong R/(p)$ as $R/(p)$-vector spaces. □

Example 2.4.10. Consider the case when R is a PID and $M = R/(d)$ as an R-module, where $d \in R$ is non-zero. Let $a \in R\backslash\{0\}$ and let $b = d/\gcd(a, d)$. Then we have the well-defined isomorphism

$$\pi : \quad R/(b) \longrightarrow aM, \quad x + (b) \longmapsto ax + (d), \qquad x \in R,$$

of R-modules. The proof is left as an exercise.

Theorem 2.4.11. *Suppose that R is a PID, $r \geq 0$ is an integer, and $d_1, \ldots, d_s \in R$ are non-zero non-units and $d_1|d_2|\cdots|d_s$. Consider*

$$M = R/(d_1) \oplus \cdots \oplus R/(d_s) \oplus R^r$$

as an R-module. Then $r + s = l(M)$, $r = r(M)$. Moreover, writing $l = l(M)$, then

$$(d_1, d_2, \ldots, d_s) = (\alpha_{l-1}(M), \alpha_{l-2}(M), \ldots, \alpha_r(M)).$$

Proof. We first show that $r + s = l(M)$. Since M can be generated by $r + s$ elements, we have $l(M) \leq r + s$. Since $d_1|d_2|\cdots|d_s$ and d_1 is not a unit, there exists a prime element $p \in R$ such that $p|d_1$. By Example 2.4.10 and Lemma 2.4.9,

$$M/pM \cong (R/(p))^{r+s}$$

as vector spaces over the field $R/(p)$. By Lemma 2.4.9, $r + s \leq l(M)$. Thus $r + s = l(M)$.

We now show that $r = r(M)$. If $r = 0$, we have $d_s \in \mathrm{Ann}(M)$, so $I_0(M) \neq 0$ and thus $r(M) = 0$. Assume now that $r > 0$. For any $a \in R$, $a \neq 0$, aM contains the free R-module aR^r which can not be generated by less than r elements. Thus $l(aM) \geq r$ for all $a \in R\backslash\{0\}$. Thus

$$I_0(M) = I_1(M) = \cdots = I_{r-1}(M) = 0. \tag{2.5}$$

Moreover, $d_sM = d_sR^r$ is free of rank r, so $l(d_sM) = r$. In particular, $d_s \in I_r(M)$, so $I_r(M) \neq 0$. By (2.5) and the definition of $r(M)$, we see that $r = r(M)$.

Let $1 \leq i \leq s$. It remains to show that d_i generates the ideal $I_{r+s-i}(M)$. Since

$$d_iM = d_iR/(d_{i+1}) \oplus \cdots \oplus d_iR/(d_s) \oplus d_iR^r,$$

we see that $l(d_iM) \leq r + s - i$, so $d_i \in I_{r+s-i}(M)$. Let $a \in R\backslash\{0\}$ be such that $a \in I_{r+s-i}(M)$, i.e., $l(aM) \leq r + s - i$. We need to show that $d_i|a$. For $1 \leq j \leq s$, let $g_j = \gcd(a, d_j)$ and let $b_j = d_j/g_j$. Then $b_1|b_2|\cdots|b_s$, and by Example 2.4.10,

$$aM \cong R/(b_1) \oplus \cdots \oplus R/(b_s) \oplus R^r.$$

If d_i does not divide a, then b_i is not a unit. Let i_0 be the largest index such that $1 \leq i_0 \leq i$ and b_i is not a unit. By the same arguments at the beginning of this proof,

$$l(aM) = r + s - i_0 + 1 \geq r + s - i + 1,$$

contradicting the assumption that $l(aM) \leq r + s - i$. Thus $d_i|a$. □

The uniqueness part of Theorem 2.4.1 is now a direct consequence of Theorem 2.4.11.

Recall now the Chinese Remainder Theorem, which says that if R is a PID and $d \in R\backslash\{0\}$ and $d = p_1^{n_1} \cdots p_k^{n_k}$ is a factorization of d into prime elements, then

$$R/(d) \cong R/(p_1^{n_1}) \oplus + \cdots \oplus R/(p_k^{n_k}),$$

Apply the Chinese Remainder Theorem to all the invariant factors $d_1, \ldots, d_s$ of a finitely generated R-module, we have the following *Structure theorem of finitely generated modules over PIDs: elementary divisor form.*

Theorem 2.4.12. *Let R be a PID. For any finitely generated module M over R, there exist prime elements $p_1, \ldots, p_k$ in R (not necessarily pair-wise non-associates), an integer $r \geq 0$, and positive integers $\{n_1, \ldots, n_k\}$, such that*

$$M \cong R^r \oplus R/(p_1^{n_1}) \oplus + \cdots \oplus R/(p_k^{n_k}).$$

If $G \cong R^{r'} \oplus (q_1^{m_1}) \oplus \cdots \oplus R/(q_l^{m_l})$, is another such isomorphism, then $r = r'$, $k = l$, and after re-ordering p_i and q_i are associates and $n_i = m_i$ for each $1 \leq i \leq k$.

Definition 2.4.13. The elements $p_i^{n_i} \in R$ in Theorem 2.4.12 are called the *elementary divisors* of M.

Remark 2.4.14. 1) The decomposition of M in the elementary divisor form is obtained from that in the invariant factor form by decomposing each invariant factor d_j as a product of prime elements and applying the Chinese Remainder Theorem.

2) If $q_1, \ldots, q_l$ is the set of pair-wise relatively prime elements in the set $\{p_1, \ldots, p_k\}$, the longest (in terms of the length function defined on any UFD) invariant factor d_s of M is the product the highest powers of for each q_j. ; the second longest invariant factor d_{s-1} is the the product of the remaining highest powers of the q_j's, etc.

Example 2.4.15. Consider the following abelian group of order $3 \times 9 \times 36 \times 40 \times 25$ as a finitely generated $\mathbb{Z}$-module

$$M = \mathbb{Z}/(3\mathbb{Z}) \oplus \mathbb{Z}/(9\mathbb{Z}) \oplus \mathbb{Z}/(36\mathbb{Z}) \oplus \mathbb{Z}/(40\mathbb{Z}) \oplus \mathbb{Z}/(25\mathbb{Z})$$

Its elementary divisor form is

$$M \cong \mathbb{Z}/(3\mathbb{Z}) \oplus \mathbb{Z}/(3^2\mathbb{Z}) \oplus \mathbb{Z}/(2^2\mathbb{Z}) \oplus \mathbb{Z}/(3^2\mathbb{Z}) \oplus \mathbb{Z}/(2^3\mathbb{Z}) \oplus \mathbb{Z}/(5\mathbb{Z}) \oplus \mathbb{Z}/(5^2\mathbb{Z}).$$

and its invariant factor form is

$$M \cong \mathbb{Z}/(3\mathbb{Z}) \oplus \mathbb{Z}/(2^2 3^2 5\mathbb{Z}) \oplus \mathbb{Z}/(2^3 3^2 5^2\mathbb{Z}).$$

In other words, the invariant factors of M are $\{3,\ 2^2 3^2 5,\ 2^3 3^2 5^2\}$ and the elementary divisors are $\{2^2,\ 2^3,\ 3,\ 3^2,\ 3^2,\ 5,\ 5^2\}$.

2.5 Applications

§ 1. Applications to finitely generated abelian groups

As $\mathbb{Z}$ is a PID and an abelian group is a $\mathbb{Z}$-module, we can apply to structure theorem for finitely generated modules over PID to get a classification of finitely generated abelian groups.

Theorem 2.5.1. Fundamental Theorem of Finite Abelian Groups. *Every finitely generated abelian group is isomorphic to the product of its torsion subgroup G_{tor} and the abelian group $\mathbb{Z}^r$ for a unique integer $r \geq 0$; If G is a finite abelian group, then there exist prime numbers $p_1, \ldots, p_k$, not necessarily pairwise distinct, and integers $n_1, \ldots, n_k$ such that*

$$G \cong (\mathbb{Z}/p_1^{n_1}\mathbb{Z}) \times \cdots \times \mathbb{Z}/(p_k^{n_k}\mathbb{Z}).$$

Moreover, if $q_1, \ldots, q_l$ are also prime numbers and if there are positive integers $m_1, \ldots, m_l$ such that

$$G \cong (\mathbb{Z}/q_1^{m_1}\mathbb{Z}) \times \cdots \times \mathbb{Z}/(q_l^{m_l}\mathbb{Z}),$$

then $l = k$ and after re-ordering one has $p_i = q_i$ and $n_i = m_i$ for each $1 \leq i \leq k$.

Example 2.5.2. For a prime number p and an integer n, every finite abelian group of order p^n is isomorphic to a unique one of the form

$$(\mathbb{Z}/p^{k_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p^{k_m}\mathbb{Z}),$$

where $1 \leq k_1 \leq \cdots \leq k_m \leq n$ and $k_1 + \cdots + k_m = n$. Thus the number of non-isomorphic classes of abelian groups of order p^n is equal to $\mathcal{P}(n)$, the number of partitions of n. For example

$$\mathcal{P}(1) = 1, \quad \mathcal{P}(2) = 2, \quad \mathcal{P}(3) = 3, \quad \mathcal{P}(4) = 5, \quad \mathcal{P}(5) = 7, \quad \mathcal{P}(6) = 11.$$

For example, the number of non-isomorphic classes of abelian groups of order $2^4 \times 5^5 = 50000$ is $\mathcal{P}(4)\mathcal{P}(5) = 35$.

§ 2. Applications to linear algebra: canonical forms of matrices

We look again at the example from linear algebra in Example 2.2.3. Fix a field K and let again V be an n-dimension vector space over K and $T \in \mathrm{End}_K(V)$. Fix a basis $\{v_1, \ldots, v_n\}$ of V. As the set $\{v_1, \ldots, v_n\}$ generates V as a K-module, it also generates V as a $K[x]$-module. Consequently, one has the surjective $K[x]$-module homomorphism

$$\begin{aligned}\phi: \ K[x]^n \longrightarrow V, \ \ \phi \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_n \end{pmatrix} &= f_1 \cdot v_1 + \cdots + f_n \cdot v_n \\ &= f_1(T)(v_1) + f_2(T)(v_2) + \cdots + f_n(T)(v_n).\end{aligned}$$

One thus has the $K[x]$-module isomorphism

$$V \cong K[x]^n / \ker(\phi).$$

We thus need to have a better understanding of the sub-module $\ker(\phi)$ of $K[x]^n$, and in particular its invariant factors. To this end, let $A = (a_{i,j}) \in M_{n,n}(K)$ be such that

$$T(v_1, \ldots, v_n) = (v_1, \ldots, v_n)A.$$

Then for each $j = 1, 2, \ldots, n$, we have $Tv_j = a_{1,j}v_1 + a_{2,j}v_2 + \cdots + a_{n,j}v_n$, which says that

$$-a_{1,j}v_1 - \cdots - a_{j-1,j}v_2 + (x - a_{j,j})v_j - a_{j+1,j}v_{j+1} - \cdots - a_{n,j} \cdot v_n = 0, \quad j = 1, 2, \ldots, n. \tag{2.1}$$

Introducing now the *characteristic matrix* of A to be

$$\mathrm{Ch}_A = xI_n - A = \begin{pmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & x - a_{22} & \cdots & -a_{2n} \\ & \cdots & & \\ -a_{n1} & -a_{n2} & \cdots & x - a_{nn} \end{pmatrix} \in M_{n,n}(K[x]).$$

Then all the columns of $xI_n - A$ are in $\ker(\phi)$.

Lemma 2.5.3. *The $K[x]$-sub-module $\ker(\phi)$ of $K[x]^n$ is generated by the columns of the characteristic matrix $xI_n - A$ of A.*

Proof. Denote by N the sub-module of $K[x]^n$ generated by the columns of $xI_n - A$. We have already seen that $N \subset \ker(\phi)$.

To show that $\ker(\phi) \subset N$, let K^n be the subset of $K[x]^n$ consisting of n-tuples of constant polynomials. We first show that

$$N + K^n = K[x]^n. \tag{2.2}$$

Indeed, for $1 \leq j \leq n$ and for any integer $k \geq 0$, let $(x^k)_j = (0, \ldots, x^k, \ldots, 0)^t \in K[x]^n$ which has 0 at every entry except at the jth where it is the constant polynomial x^k. Here recall that we write elements in $K[x]^n$ as columns, and the sup-script t stands for transpose. It is enough to show that

$$(x^k)_j \in N + K^n, \quad 1 \leq j \leq n,\ k \geq 1. \tag{2.3}$$

Note that (2.3) obviously holds for $k = 0$. Assume that $k \geq 1$. Using

$$a^k - b^k = (a-b)(a^{k-1} + a^{k-2}b + \cdots + ab^{k-2} + b^{k-1}),$$

we have $(xI_n)^k - A^k = (xI - A)B(x) \in M_{n,n}(K[x])$ for some $B(x) \in M_{n,n}(K[x])$. Denoting by $(A^k)_j$ and $B_j(x)$ the jth columns of the matrices A^k and $B(x)$ respectively, we see that $(x^k)_j = (A^k)_j + (xI - A)B_j(x) \in N + K^n$. This shows (2.3) for all j and k, so (2.2) holds.

Using now (2.2) to write an arbitrary $F \in K[x]^n$ as $F = \alpha + C$, where $\alpha \in N$ and $C \in K^n$. Then $\phi(F) = \phi(\alpha) + \phi(C) = \phi(C)$. As $\{v_1, \ldots, v_n\}$ is a basis of V, if $\phi(F) = \phi(C) = 0$, then $C = 0$, so $F = \alpha \in N$. This shows that $\ker(\phi) \subset N$. Thus $\ker(\phi) = N$. □

Recall that a non-zero polynomial $f \in K[x]$ is said to be monic if the leading coefficient of f is 1. Let $d_1, \ldots, d_n \in K[x]$, with each d_j monic and $d_1 | \cdots | d_n$, be the invariant factors of $xI_n - A \in M_{n,n}(K[x])$. Note that $\{v_1', \ldots, v_n'\}$ is another basis of V such that $T(v_1', \ldots, Tv_n') = (v_1', \ldots, v_n')A'$, then $xI_n - A' = B(xI_n - A)B^{-1}$ for some $B \in GL(n, K)$, $xI_n - A'$ has the same invariant factors as $xI_n - A$. Thus $d_1, d_2, \ldots, d_n$ are independent of the choices of the basis of V, and they are called the *invariant factors of* T. The prime powers in the unique factorizations of the d_j's into prime powers are called the *elementary divisors of* T.

Theorem 2.5.4. *Let $d_{k+1}, \ldots, d_n \in K[x]$ be the invariant factors of T that are not constant polynomials. Then, V as a $K[x]$-module in which $x \in K[x]$ acts as T, is isomorphic to*

$$R/\langle d_{k+1}\rangle \oplus R/\langle d_{k+2}\rangle \oplus \cdots \oplus R/\langle d_n\rangle.$$

Proof. As we have see in Lemma 2.5.3, the surjective $K[x]$-module homomorphism

$$\phi: \ K[x]^n \longrightarrow V, \ \ \phi \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_n \end{pmatrix} = f_1(T)(v_1) + \cdots + f_n(T)(v_n),$$

gives rise to the $K[x]$-module isomorphism

$$V \cong K[x]^n / \ker(\phi),$$

and we proved in Lemma 2.5.3 that $\ker(\phi)$ coincides with the $K[x]$-sub-module of $K[x]^n$ generated by the columns of the matrix $xI_n - A$. Theorem 2.5.4 now follows immediately from Example 2.3.8. □

Consider now the annihilator ideal $I = \mathrm{ann}(V)$ of the $K[x]$ module V where $x \in K[x]$ acts as T, i.e.,

$$I = \{f \in K[x] : f(T) = 0\}.$$

Definition 2.5.5. The *minimal polynomial* of T is defined to be monic polynomial $f \in K[x]$ that generates the ideal $I = \mathrm{ann}(V)$,

Lemma 2.5.6. *If $d_1, d_2, \ldots, d_n$ are the invariant factors of T, then the minimal polynomial of T is $d_n \in K[x]$.*

Proof. Direct consequence of Theorem 2.5.4. □

Example 2.5.7. Let $A = \begin{pmatrix} 1 & 6 & 0 \\ 0 & 2 & -1 \\ 1 & 0 & 3 \end{pmatrix}$, so

$$xI_3 - A = \begin{pmatrix} x-1 & -6 & 0 \\ 0 & x-2 & 1 \\ -1 & 0 & x-3 \end{pmatrix},$$

and let T_A be the linear map from $\mathbb{R}^3$ to R^3 given by $T_A(x) = Ax$. Let $m_j j$, $j = 1, 2, 3$, be the gcd of all the $j \times j$ minors of $xI_3 - A$. Easy to see that $m_1 = m_2 = 1$, and $m_3 = \det(xI_3 - A) = x((x-3)^2+2)$. Thus the invariant factors of T_A are $d_1 = d_2 = 1$ and $d_3 = x((x-3)^2 + 2)$, and its elementary divisors are x and $x^2 - 6x + 11$. If we regard T_A as a linear map from $\mathbb{C}^3$ to $\mathbb{C}^3$, then its elementary divisors are x, $x - 3 + \sqrt{2}i$ and $x - 3 - \sqrt{2}i$.

Example 2.5.8. (Example 8.6.11 of Goodman) For $A = \begin{pmatrix} -1 & 0 & 0 & 0 & 3 \\ 1 & 2 & 0 & -4 & 0 \\ 3 & 1 & 2 & -4 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 0 & 4 \end{pmatrix} \in$ $M_{5,5}(\mathbb{Q})$, and $T_A : \mathbb{R}^5 \to \mathbb{R}^5$ defined by $T_A(x) = Ax$, it is shown in Goodman's book that T_A has invariant factors $d_1 = d_2 = d_3 = 1$, $d_4 = x - 1$, and $d_5 = (x-2)^3(x-1)$. Its elementary divisors are then

$$x - 1, \quad (x-1), \quad (x-2)^3.$$

Lemma 2.5.9. *If $d_1, \ldots, d_n \in K[x]$ are the invariant factors of $xI_n - A \in M_{n,n}(K)$, where $A \in M_{n,n}(K)$, then*

$$\det(xI_n - A) = d_1 d_2 \cdots d_n.$$

Proof. Let $P, Q \in GL(n, K[x])$ be such that $xI_n - A = P\mathrm{diag}(d_1, \ldots, d_n)Q$. Then

$$\det(xI_n - A) = \det(P) d_1 d_2 \cdots d_n \det(Q) = d_1 d_2 \cdots d_n \det(P) \det(Q).$$

As both $\det(P)$ and $\det(Q)$ are units in $K[x]$, then are non-zero elements in K. Since both $\det(xI_n - A) \in K[x]$ and $d_1 d_2 \cdots d_n \in K[x]$ are monic, one has $\det(P)\det(Q) = 1$ and thus $\det(xI_n - A) = d_1 d_2 \cdots d_n$. □

We now look more closely at the $K[x]$-modules $V = K[x]/\langle f\rangle$ for $f \in K[x]$. First note that if $\deg(f) = n > 0$, then V is a K-vector space of dimension n. For $g \in K[x]$, denote by $\bar{g} \in V$ the image of g in V. Then it is easy to see that

$$v_0 = \bar{1},\ v_1 = \bar{x},\ \ldots,\ v_{n-1} = \overline{x^{n-1}}$$

is a basis, so $\dim_K V = n$. Assume that f is monic and $f = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$. Then

$$\overline{x^n} = -a_0 v_0 - a_1 v_1 - \cdots - a_{n-1} v_{n-1}.$$

Consider now the operator T on V which is defined to be the multiplication by $x \in K[x]$, i.e.,

$$T(\bar{g}) = x\bar{g} = \overline{xg}, \qquad g \in K[x].$$

Then

$$T(v_0, v_1, \ldots, v_{n-2}, v_{n-1}) = (v_0, v_1, \ldots, v_{n-2}, v_{n-1}) \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ & & \cdots & \cdots & \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}.$$

For $a = (a_0, a_1, \ldots, a_{n-1}) \in K^n$, the $n \times n$

$$R(a) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ & & \cdots & \cdots & \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix} \tag{2.4}$$

is called the *rational matrix* defined by a or the *companion matrix of* a. A matrix of the form $R(a)$ for some a is said to be *in rational form.*

Theorem 2.5.10. *Let V be a finite dimensional vector space over a field K and $T \in \mathrm{End}_K(V)$. Then there exists a direct sum decomposition $V = V_1 + \cdots + V_k$ such that for each $1 \le j \le k$, $T(V_j) \subset V_j$ and there exists a basis of V_j with respect to which T is in rational form.*

Equivalently, we have

Theorem 2.5.11. *Every square matrix with entries in a field K is similar to a block diagonal one for which each block is in rational form.*

Remark 2.5.12. It is trivial to check that

$$\det(\lambda I - R(a)) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

for the matrix $R(a)$ given in (2.4).

Consider now the $K[x]$-module $V = K[x]/\langle (x-a)^k \rangle$, where again K is a field and $a \in K$. Choose the basis

$$v_1 = \overline{(x-a)^{n-1}},\ \cdots,\ v_{n-1} = \overline{(x-a)},\ v_n = \overline{1}$$

of V. Then $T(v_1, \ldots, v_n) = (v_1, \ldots, v_n)J_n(a)$, where

$$J_n(a) = \begin{pmatrix} a & 1 & 0 & \cdots & 0 & 0 \\ 0 & a & 1 & \cdots & 0 & 0 \\ 0 & 0 & a & \cdots & 0 & 0 \\ & & \cdots & \cdots & & \\ 0 & 0 & 0 & \cdots & a & 1 \\ 0 & 0 & 0 & \cdots & 0 & a \end{pmatrix}.$$

We have thus proved the following theorem on Jordan canonical forms.

Theorem 2.5.13. *Let V be a finite dimensional vector space over an algebraically closed field K and $T \in \mathrm{End}_K(V)$. Then there exists a direct sum decomposition $V = V_1 + \cdots + V_k$ such that for each $1 \le j \le k$, $T(V_j) \subset V_j$ and there exists a basis of V_j with respect to which T is in Jordan canonical form.*

Equivalently, we have

Theorem 2.5.14. *Every square matrix with entries in an algebraically closed field K is similar to a block diagonal one for which each block is in Jordan canonical form.*

Note that the two theorems on rational canonical forms and on Jordan canonical forms respectively correspond to the structure theorem of finitely generated PIDs in invariant factor form and in elementary divisor form.

Note that the rational or Jordan canonical forms of a matrix A is computed by computing the Smith Normal Form of the characteristic matrix of A..

§ 3. Summary

Let R be a PID. For any integer $n \ge 1$ we write elements in R^n as columns $(r_1, \ldots, r_n)^T$, where $r_1, \ldots, r_n \in R$ and the subscript T stands for transpose. Let M be a finitely generated R-module. Here are the steps we have taken to understand M as an R-module.

Step 1. As M is finitely generated, there exist $v_1, \ldots, v_n \in M$ such that the map

$$\phi:\ \ R^n \longrightarrow M,\ \ (r_1, \ldots, r_n)^T \longmapsto r_1v_1 + \cdots + r_nv_n$$

is a surjective R-module homomorphism. Thus $\ker\phi \subset R^n$ is a sub-module of the free module R^n, and $M \cong R^n/\ker\phi$ as R-modules.

Step 2. Let $N = \ker\phi$, regarded as a sub-module of the free R-module R^n. We have proved in Lemma Lemma 2.3.3 that N is generated by m elements $a_1, \ldots, a_m$ in R^n with some $0 \leq m \leq n$. Let $A = (a_1, \ldots, a_m) \in M_{n\times m}(R)$, so that N is generated by the columns of A. We call A a *presentation of* N. Consider the R-module homomorphism

$$L_A: \ R^m \longrightarrow R^n, \ \ (r_1, \ldots, r_m)^T \longmapsto A(r_1, \ldots, r_m)^T.$$

Then $N = \mathrm{Im}(L_A)$. Applying the Smith Normal Form Theorem to A, we know that there exist $P \in GL(n, R)$ ad $Q \in GL(m, R)$ such

$$A = P\mathrm{diag}(d_1, \ldots, d_s, 0, \ldots, 0)Q,$$

where $s = \mathrm{rank}(A)$ and $d_1, \ldots, d_s \in R$ with $d_1|\cdots|d_s$ are the invariant factors of A. Let $p_1, \ldots, p_n$ be the columns of P, so that

$$A = (d_1p_1, \ldots, d_sp_s, 0, \ldots, 0)Q.$$

Then $\{p_1, \ldots, p_n\}$ is a basis of R^n as an R-module, and $N = \mathrm{Im}(L_A)$ is the free R-submodule spanned by $d_1p_1, \ldots, d_sp_s$. It follows that

$$M \cong R^n/N \cong R/(d_1) \oplus \cdots \oplus R/(d_s) \oplus R^{n-s}$$

as R-modules. We see that the crucial step is to find a matrix $A \in M_{n\times m}(R)$ whose columns generate $N = \ker\phi$ and then compute the invariant factors of A.

Example: In the linear algebra example in which we have a pair (V, T), where V is an n-dimensional vector space V over a field K and $T \in \mathrm{End}_K(V)$, we regard V as a $K[x]$-module such that

$$\lambda \cdot v = \lambda v \quad \text{ and } \quad x \cdot v = Tv$$

for $\lambda \in K$ and $v \in V$. We take any basis $v_1, \ldots, v_n$ of V over K which also generates V as a $K[x]$-module, so we have the surjective $K[x]$-module homomorphism

$$\phi: \ K[x]^n \longrightarrow V, \ \ (f_1, \ldots, f_n)^T \longmapsto f_1(T)v_1 + \cdots f_n(T)v_n.$$

We then prove (see Lemma Lemma 2.5.3) that $\ker\phi \subset K[x]$ is generated, as a $K[x]$-submodule, by the columns of the characteristic matrix $xI_n - A$, where $A \in M_{n\times n}(K)$ is such that

$$T(v_1, \ldots, v_n) = (v_1, \ldots, v_n)A.$$

Let $d_{k+1}, \ldots d_n \in K[x]$ be the non-constant invariant factors of $xI_n - A$. We then conclude that

$$V \cong K[x]/(d_{k+1}) \oplus \cdots \oplus K[x]/(d_n)$$

as $K[x]$-modules. Studying each $K[x]/(d)$, we are led to rational canonical forms and, in the case of $K = \mathbb{C}$, Jordan canonical forms of T.

Chapter 3 | Field extensions

3.1 Field extensions

§ 1. Motivations and definition of field extensions

Recall that a field F is called a *finite field* if it has only finitely many elements. Otherwise, F is called an *infinite field*.

Our immediate examples of fields are $\mathbb{F}_p := \mathbb{Z}_p$ for p prime, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. Clearly, $\mathbb{F}_p$ is finite for any p, while $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are not. Other examples of fields include number fields, p-adic fields, and function fields.

Why do we study fields? The reason is that fields appear in every area of mathematics. Here we give some examples.

- Linear algebra: here we have the notion of vector spaces over *any* field;
- Analysis: here we mostly use the field $\mathbb{R}$ of real numbers and the field $\mathbb{C}$ of complex numbers;
- Number theory: the field $\mathbb{Q}$ of rational numbers, and their finite extensions, called algebraic number fields, and p-adic fields, play prominent roles in number theory;
- Algebraic and arithmetic geometry: fields of rational functions on geometrical objects;
- Modern mathematical physics: all the fields above.

Regarding finite fields, we have the following immediate question.

Question 3.1.1. Are there finite fields other than $\mathbb{F}_p$ for prime numbers p? If the answer is yes, can we classify all finite fields?

One key tool for studying fields is the notion of field extensions, which is about one field sitting inside a bigger field. We will develop some concepts and some simple facts. At the end, we will see that we can classify all finite fields.

Definition 3.1.2. Let L be a field. If K is a subset of L containing 0 and 1 and is closed under addition, multiplication and taking minuses and inverses, then we say that K is a *sub-field* of L, and we call L an *extension* of K.

Exercise 3.1.3. 1) The intersection of any collection of sub-fields is again a sub-field.

2) If K is a field and R is a ring such that $\phi : K \to R$ is a ring homomorphism. Then $\ker \phi \subset K$, being an ideal, is either 0 or K, so if $\phi \neq 0$ then ϕ is injective.

When K and L are two fields and $\phi : K \to L$ is a ring homomorphism which is not identically 0, we will also called the injective map $\phi : K \to L$ an extension of the field L, and we identify K with its image in L.

Here is one of our fundamental constructions of field extensions:

Example 3.1.4. Let K be any field and let $p(x) \in K[x]$ be an irreducible polynomial. Since $K[x]$ is a PID, $p(x)$ is a prime element in $K[x]$, so the principal ideal $\langle p(x) \rangle$ in $K[x]$ is a prime ideal. By Proposition 1.2.13, $\langle p(x) \rangle$ is also a maximal ideal of $K[x]$. Thus

$$K[x]/\langle p(x) \rangle$$

is a field. Let $I : K \to K[x]$ be the inclusion map which maps $a \in K$ to the constant polynomial with constant term a, and let $\pi : K[x] \to K[x]/\langle p(x) \rangle$ be the projection map. Then

$$\phi = \pi \circ I : \quad K \longrightarrow K[x]/\langle p(x) \rangle,$$

is a non-zero ring homomorphism, and is thus injective. Hence $K[x]/\langle p(x) \rangle$ is a field extension of K. In particular, every irreducible polynomial in $\mathbb{Q}[x]$ defines a field extension of $\mathbb{Q}$.

In Example 3.1.4, since we regard K as sitting inside $L = K[x]/\langle p(x) \rangle$, the polynomial $p(x) \in K[x]$ can also be regarded as a polynomial in $L[x]$. Let $\bar{x} = \pi(x) \in L$.

Lemma 3.1.5. *The element $\bar{x} \in L$ is a root of $p(x)$ in L.*

The fact in Lemma 3.1.5 is one of the key reasons for our interests in the field extension L of K. Indeed, $p(x)$ may not have a root in K, but it now has a root in the larger field L.

Definition 3.1.6. Let K be a field. The *prime sub-field* of K is the intersection of all sub-fields of K.

Lemma 3.1.7. *Let K be a field. If K has characteristic p, then the prime sub-field is isomorphic to $\mathbb{F}_p$, so K is an extension of $\mathbb{F}_p$; If K is has characteristic 0, then the prime sub-field of K is isomorphic to $\mathbb{Q}$, so K is an extension of $\mathbb{Q}$.*

Thus you will know all fields if you understand all extensions of $\mathbb{F}_p$ and $\mathbb{Q}$.

§ 2. Degrees of field extensions

A very powerful idea is to regard an extension K of a field F as a vector space over F. For vector spaces over F, we have the notion of basis and dimension.

Definition 3.1.8. Let L be an extension of a field K. Regard L as a vector space of K. If L is an n dimensional vector space over K, we write $n = [L : K]$ and call it the *degree of L over K*. If n is finite, we say that L is a *finite extension* of K. Otherwise, we say that L is an infinite extension of K.

Clearly, $\mathbb{Q}$ is a sub-field of $\mathbb{R}$ and $\mathbb{C}$, and $\mathbb{R}$ is a sub-field of $\mathbb{C}$. Easy to see that $\mathbb{R}$ are $\mathbb{C}$ are infinite extensions of $\mathbb{Q}$, while $\mathbb{C}$ is a finite extension of $\mathbb{R}$ of degree 2. Indeed, $\mathbb{R}$ and $\mathbb{C}$ can not be a finite extension of $\mathbb{Q}$ because $\mathbb{Q}^n$ is countable while for any n while $\mathbb{R}$ and $\mathbb{C}$ are not.

Example 3.1.9. Let F be a field. Let

$$F(x) = \left\{ \frac{f(x)}{g(x)} : \ f, g \in F[x], g \neq 0 \right\},$$

which is the field of fractions of the polynomial ring $F[x]$, also called the field of rational function in x over F. Clearly, $F(x)$ is an infinite extension over F.

Turning to finite fields, we have

Lemma 3.1.10. *If K is a finite field, then $|K| = p^n$ for some prime number p and some integer n.*

Proof. Any field of characteristic zero must contain $\mathbb{Q}$ as a sub-field so must be an infinite field. Since K is a finite field, the characteristic of K is must be a prime number p. The dimension n of K as a vector space over $\mathbb{F}_p$ must be finite, so $|K| = p^n$. □

Thus the natural questions becomes the following: given a prime number p and a positive integer n, does there exist a field of order p^n? If yes, how many up to isomorphisms?

Proposition 3.1.11. *If $K \subset L$ and $L \subset M$ are finite extensions, then $K \subset M$ is a finite extension and*

$$[M : K] = [M : L][L : K].$$

Proof. Let $\{a_1, a_2, \cdots, a_n\}$ be the basis L over K and let $\{b_1, b_2, \cdots, b_m\}$ be a basis of M over L. Then $\{a_i b_j\}$ is basis of M over K. Proof straightforward. □

Remark 3.1.12. The identity $[M : K] = [M : L][L : K]$ even when either or both extensions $K \subset L$ and $L \subset M$ are infinite. In other words, if either $K \subset L$ or $L \subset M$ is infinite, then so is $K \subset M$. To prove this, one first prove that if $\{a_1, a_2, \cdots, a_n\} \subset L$ is linearly independent over K and let $\{b_1, b_2, \cdots, b_m\} \subset M$ linearly independent over L, then $\{a_i b_j\} \subset M$ is linearly independent over K.

§ 3. Simple field extensions

Given a field K, there are two situations in which we can construct field extensions of K: the first case is when K already lies in a field L as a sub-field, such as $\mathbb{Q} \subset \mathbb{C}$. The other case is when we do not have an L in which K lies. We look at the first case first.

Definition 3.1.13. Given a field extension $K \subset L$ and a subset S of L, denote by $K(S)$ the smallest sub-field of L containing K and S, i.e., $K(S)$ is the intersection of all the sub-fields of L that contain K and S. The field $K(S)$ is called the *sub-field of L generated by K and S*, or the *sub-field of L by adjoining S to K*.

In particular, when L is an extension of a field K, and $a \in L$, we have $K(a)$, the sub-field of L generated by K and a. Clearly, we have

$$K(a) = \left\{ \frac{f(a)}{g(a)} : \ f, g \in K[x], \ g(a) \neq 0 \right\} \subset L.$$

Definition 3.1.14. A field extension $K \subset L$ such that $L = K(a)$ for some $a \in L$ is said to be *simple*.

A simple observation is that

$$K(S \cup a) = K(S)(a)$$

for any subset S of L and $a \in L$. Thus when S has more than one element, for example, $S = \{a_1, a_2, \ldots, a_n)$, we have

$$K(a_1, \ldots, a_n) = K(a_1, \ldots, a_{n-1})(a_n).$$

This way we get sub-fields like $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{2} + \sqrt{3}, \sqrt{5})$ of $\mathbb{R}$. We would like to know what these fields look like and how to test whether the extension $K \subset K(a)$ is finite or not. Moreover, when finite, how to find the degree $[K(a) : K]$?

Definition 3.1.15. Let L be an extension of a field K. An element $a \in L$ is said to be *algebraic* over K if there exists a non-zero $f(x) \in K[x]$ such that $f(a) = 0$. If $a \in L$ is not algebraic over K, we say that a is *transcendental* over K.

For example, $\sqrt{2}$ is algebraic over $\mathbb{Q}$, and so are $\sqrt{5} + \sqrt{2}$ and $3^{\frac{1}{3}}$. But π and e are transcendental over $\mathbb{Q}$.

Lemma 3.1.16. *If $a \in L$ is transcendental over K, then*

$$E_a : \ K(x) \longrightarrow K(a), \ \ \frac{f(x)}{g(x)} \longmapsto \frac{f(a)}{g(a)}$$

is an isomorphism. In particular, $K(a)$ is an infinite extension of K.

Proof. Since a is transcendental over K, $g(a) \neq 0$ for all $g(x) \in K[x]$ and $g(x) \neq 0$. Thus E_a is a well-defined ring homomorphism. As E_a is not identically 0, it is injective. By definition, E_a is also surjective. Thus E_a is an isomorphism of fields. □

Example 3.1.17. One has $\mathbb{Q}(\pi) \cong \mathbb{Q}(e) \cong \mathbb{Q}(x)$.

Definition 3.1.18. By a *transcendental number* we mean a complex number that is transcendental over $\mathbb{Q}$.

All transcendental numbers give the isomorphic simple extensions of $\mathbb{Q}$, namely they are all isomorphic to $\mathbb{Q}(x)$.

The extensions $K(a)$ are much more interesting when $a \in L$ is algebraic over K.

Assume that $a \in L$ is algebraic over K. Then among the set of all $f \in K[x]$ such that $f(a) = 0$, there is one which is monic and of the smallest order. Indeed, let $I_a = \{f \in K[x] : f(a) = 0\}$. Then I_a is an ideal of $K[x]$, so it has a generator $g(x)$.

Definition 3.1.19. Let L be an extension of a field K and let $a \in L$ be algebraic. Any generator $g(x)$ of the ideal $I_a = \{f(x) \in K[x] : f(a) = 0\}$ is called a *minimal polynomial* of a. The unique minimal polynomial with leading coefficient 1 is called *the minimal polynomial* of a.

The following fact follows directly from the definitions.

Lemma 3.1.20. *Let L be an extension of a field K and let $a \in L$ be algebraic over K. Then a monic $p(x) \in K[x]$ is the minimal polynomial of a if and only if*
1) $p(a) = 0$;
2) $p(x)$ is irreducible.

Theorem 3.1.21. *Let L be an extension of a field K and let $a \in L$ be algebraic over K. Let $p(x)$ be the minimal polynomial of a. Then the evaluation map*

$$E_a : \quad K[x]/\langle p(x)\rangle \longrightarrow K(a),\ \overline{f(x)} \longmapsto f(a),$$

is an isomorphism of fields.

Proof. Clearly E_a is a non-zero ring homomorphism, so E_a is injective. We need to prove that E_a is surjective. Recall that

$$K(a) = \left\{ \frac{f(a)}{g(a)} :\ f, g \in K[x],\ g(a) \neq 0 \right\} \subset L.$$

Thus we only need to show that if $g \in K[x]$ is such that $g(a) \neq 0$ then $1/g(a) \in K[a]$. Since $p(x)$ is irreducible and $g \neq 0$, g and p have no common nontrivial divisors, so there exist $f(x), h(x) \in K[x]$ such that $f(x)g(x) + h(x)p(x) = 1$. It follows that $f(a)g(a) = 1$ so $1/g(a) = f(a) \in K[a]$. □

Note that

$$\text{Image}(E_a) = K[a] :\overset{\text{def}}{=} \{f(a) : f(x) \in K[x]\}$$

is the *sub-ring* of L generated by a over K. Consequently, we have

Corollary 3.1.22. *If L is a field extension of K and $a \in L$ is algebraic over K, then the sub-ring $K[a]$ coincides with the sub-field $K(a)$, and $K(a)$ is a finite extension of K. More specifically, $[K(a) : K] = \deg(p)$, where $p \in K[x]$ is the monomial polynomial of a in K.*

Remark 3.1.23. Because $K(a) \cong K[x]/\langle p\rangle$, the field $K(a)$ depends, up to isomorphisms, only on the minimal polynomial of a in K. In other words, if $a, b \in L$ have the same minimal polynomial p, then $K(a) \cong K(b) \cong K[x]/\langle p\rangle$.

We now summarize some equivalent properties of algebraic elements:

Proposition 3.1.24. *Let L be a field extension of K, and let $a \in L$. Then the following are equivalent:*

1) a is algebraic over K;

2) $K(a)$ is a finite extension of K;

3) there exists a sub-field L' of L containing K such that $a \in L'$ and L' is a finite extension of K.

Proof. Only need to show that 3) implies 1). If $a \in L' \subset L$ and L' is a finite extension of K of degree n, then the set $\{1, a, a^2, \cdots, a^n\}$ is linearly dependent over K, so a is algebraic over K. □

For the rest of this section, we look at some examples of simple algebraic extensions.

Example 3.1.25. Let $a = \sqrt[3]{2}$ the real cubic root of 2, and let $\alpha = \omega a$ and $\beta = \omega^2 a$, where $\omega = e^{2\pi i/3}$. Then all the three fields $\mathbb{Q}(a), \mathbb{Q}(\alpha)$ and $\mathbb{Q}(\beta)$ are isomorphic to the quotient field $\mathbb{Q}[x]/\langle x^3 - 2\rangle$, but as sub-fields of $\mathbb{C}$ they are not the same subsets. For example, $\mathbb{Q}(a) \subset \mathbb{R}$ while $\mathbb{Q}(\alpha)$ and $\mathbb{Q}(\beta)$ are not sub-fields of $\mathbb{R}$.

Exercise 3.1.26. Take some examples of non-zero elements in the field $\mathbb{Q}(\sqrt[3]{2})$ and express their inverses in terms of $\sqrt[3]{2}$.

Example 3.1.27. $\mathbb{Q}(\sqrt{2}+\sqrt{3})$: Let $\alpha = \sqrt{2}+\sqrt{3} \in \mathbb{R}$. It follows from $\alpha^2 = 5+2\sqrt{6}$ that $\alpha^4 - 10\alpha^2 + 1 = 0$. Let

$$p(x) = x^4 - 10x^2 + 1 \in \mathbb{Z}[x].$$

To check that $p(x)$ is irreducible over $\mathbb{Q}$, by Gauss' Lemma we only need to check that it has no proper factorizations in $\mathbb{Z}[x]$. By the Rational Root Test, since ± 1 is not a root, it has no linear factor. Suppose that

$$x^4 - 10x^2 + 1 = (x^2 + ax + b)(x^2 + cx + d).$$

Then $bd = 1$, $a + c = 0$, and $b + d + ac = -10$. Thus $b = d = 1$ or $b = d = -1$, so $a + c = 0$ and $ac = -12$ or -8, not possible. Thus $p \in \mathbb{Q}[x]$ is irreducible, so p is the minimal polynomial of α. We thus conclude that $\mathbb{Q}[\sqrt{2} + \sqrt{3}]$ has degree 4 over $\mathbb{Q}$.

Example 3.1.28. [Quadratic fields] Let D be a positive integer which is not the square of any integer. Then $f(x) = x^2 - D$ is irreducible over $\mathbb{Q}$. The quotient field of $\mathbb{Q}[x]$ by the ideal generated by $f(x)$ is denoted by $\mathbb{Q}(\sqrt{D})$. As a set, we have

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

with addition and multiplication given respectively given by

$$\begin{aligned}(a + b\sqrt{D}) + (c + d\sqrt{D}) &= (a + c) + (b + d)\sqrt{D},\\ (a + b\sqrt{D})(c + d\sqrt{D}) &= (ac + bdD) + (ad + bc)\sqrt{D}.\end{aligned}$$

The inverse of $a + bD$ with $a \neq 0$ or $b \neq 0$ is given by

$$(a + b\sqrt{D})^{-1} = \frac{a - b\sqrt{D}}{a^2 + b^2 D}.$$

Examples of quadratic fields are $\mathbb{R}[i] = \mathbb{C}$, $\mathbb{Q}[i], \mathbb{Q}[\sqrt{2}]$ etc.

Example 3.1.29. The construction of quadratic field extensions in Example 3.1.28 work in a more general context: suppose that K be any field and $\alpha \in K$ has no square root in K, i.e., there exists no $a \in K$ such that $a^2 = \alpha$. Consider the polynomial

$$p(x) = x^2 - \alpha \in K[x].$$

Then $p(x)$ is irreducible in $K[x]$. Let $L = K[x]/\langle p(x)\rangle$, and denote $\bar{x} \in L$ by $\sqrt{\alpha}$. Then $L = K(\sqrt{\alpha})$ is a simple extension of K of degree 2, so we call L a *quadratic extension* of K.

As an example, consider $K = \mathbb{C}(y)$ and $\alpha = f(y) \in \mathbb{C}[x]$ with positive odd degree. Then α has no square root in K. Indeed, if α has a square root in K, then there exist $gy), h(y) \in \mathbb{C}[y]$ such that $f(y) = \frac{g(y)^2}{h(y)^2}$, so

$$h(y)^2 f(y) = g(y)^2$$

which is not possible because $h(y)^2 f(y)$ has odd degree while $g(y)^2$ has even degree. The quadratic extension $K[x]/\langle x^2 - f\rangle$ of $K = \mathbb{C}(y)$ will be denoted as $\mathbb{C}(y)(\sqrt{f(y)})$. When $f(y) \in \mathbb{C}[y]$ has order 3, these fields are closely related to *elliptic curves*.

We now turn to *cyclotomic Fields*. For an integer $n \geq 2$, let $\omega_n = e^{2\pi i/n} \in \mathbb{C}$. The sub-field $\mathbb{Q}(\omega_n)$ of $\mathbb{C}$ is called the n'th cyclotomic field. The minimal polynomial of ω_n over $\mathbb{Q}$, denoted by $\Phi_n(x)$, is called the n'th *cyclotomic polynomial*. Set

$$R^n = \{\omega^k : \ k \in [1, n-1], (k, n) = 1\}$$

and call elements in R_n *primitive n'th root sof unity.* Recall that the group of units of the ring $\mathbb{Z}/n\mathbb{Z}$ is $\{\overline{k} : k \in [1,n], (k,n) = 1\}$. Thus $\zeta \in \mathbb{C}$ is a primitive n'th root of unity if and only f $\zeta^n = 1$ and every n'th root of unity is an integer power of ζ.

We have the following basic fact on cyclotomic polynomials.

Theorem 3.1.30. *For $n \geq 1$, one has*

$$\Phi_n(x) = \prod_{1 \leq k \leq n,\, (k,n)=1} (x - \omega_n^k) \in \mathbb{Z}[x]. \tag{3.1}$$

Consequently, $[\mathbb{Q}(\omega_n) : \mathbb{Q}] = \deg(\Phi_n) = \phi(n)$, the number of integers between 1 and n that are relatively prime to n (the function ϕ of n is called Euler's ϕ function).

Proof. By abusing notation, denote the polynomial on the right hand side of (3.1) by $\Phi_n(x)$. We then need to show that for each $n \geq 1$, $\Phi_n(x) \in \mathbb{Z}[x]$ and is irreducible over $\mathbb{Q}$, equivalently, irreducible over $\mathbb{Z}$.

Let $n \geq 1$. Note first that

$$\begin{aligned} x^n - 1 &= \prod_{k=1}^{n} \left(x - e^{2\pi i \frac{k}{n}}\right) = \prod_{d|n} \prod_{k \in [1,n], \gcd(k,n)=d} \left(x - e^{2\pi i \frac{k}{n}}\right) \\ &= \prod_{d|n} \prod_{k \in [1,n], \gcd(k,n)=d} \left(x - e^{2\pi i \frac{k/d}{n/d}}\right) = \prod_{d|n} \Phi_{n/d}(x) = \prod_{d|n} \Phi_d(x). \end{aligned} \tag{3.2}$$

We now use induction on n to show that $\Phi_n(x) \in \mathbb{Z}[x]$. We have $\Phi_n(x) = x - 1$. Assume that $n \geq 2$ and that $\Phi_m(x) \in \mathbb{Z}[x]$ for $m \leq n-1$. Then

$$\Phi_n(x) = \frac{x^n - 1}{\prod_{d|n, d \in [1,n-1]} \Phi_d(x)}.$$

Using the fact that each Φ_d is monic and by the division algorithm, we see that $\Phi_n(x) \in \mathbb{Z}[x]$.

It remains to show that $\Phi_n(x)$ is irreducible over $\mathbb{Z}$. Suppose not. Then there exist monic $f(x), g(x) \in \mathbb{Z}[x]$, both having positive degree, and with $f(x)$ irreducible, such that $\Phi_n(x) = f(x)g(x)$. As $\deg(g) < \phi(n)$, there must be some $\zeta \in R_n$ such that $f(\zeta) = 0$.

Claim: For any $\zeta \in R_n$ such that $f(\zeta) = 0$, and for any prime number p not dividing n, one has $f(\zeta^p) = 0$.

Assume that we have proved the claim. Let $\zeta \in R_n$ be such that $f(\zeta) = 0$. Let $k \in [1, n-1]$ be such that $(k,n) = 1$, and let $k = p_1^{m_1} \cdots p_l^{m_l}$ be the prime decomposition of k. Then $f(\zeta^{p_1}) = 0$ by the claim. As $\zeta^{p_1} \in\in R_n$, we also have $f(\zeta^{p_1^2}) = 0$ so also $f(\zeta^{p_1^{m_1}}) = 0$. Continuing, we see that $f(\zeta^k) = 0$. As

$$\{\zeta^k : k \in [1, n-1], (k,n) = 1\} = R_n,$$

we see that every element in R_n is a root of f, contradicting to the assumption that $\deg(f) < \phi(n)$. Thus $\Phi_n(x) \in \mathbb{Z}[x]$ is irreducible.

We now prove the claim. Let $\zeta \in R_n$ be such that $f(\zeta) = 0$, and let p be a prime number not dividing n. As $\zeta^p \in R_n$, we have $\Phi_n(\zeta^p) = 0$, so either $f(\zeta^p) = 0$ or $g(\zeta^p) = 0$. Suppose that $g(\zeta^p) = 0$. Let $h(x) = g(x^p)$. Then $h(\zeta) = 0$. Since $f(x)$ is the minimal polynomial of ζ, we have

$$h(x) = f(x)a(x)$$

for some $a(x) \in \mathbb{Z}[x]$. Consider now the ring homomorphism

$$\mathbb{Z}[x] \longrightarrow \mathbb{F}_p[x], \;\; b(x) \longmapsto \overline{b}(x).$$

We then have $\overline{h}(x) = \overline{f}(x)\overline{a}(x) \in \mathbb{F}_p[x]$. On the other hand, by Fermat's Little Theorem which says $(\alpha + \beta)^p = \alpha^p + \beta^p$ in any ring of characteristic p, we have

$$\overline{h}(x) = (\overline{g}(x))^p.$$

Thus $(\overline{g}(x))^p = \overline{f}(x)\overline{a}(x) \in \mathbb{F}_p[x]$. Thus $\overline{f}(x)$ and $\overline{g}(x)$ have a common prime factor in $\mathbb{F}_p[x]$. On the other hand, we have

$$\overline{\Phi_n}(x) = \overline{f}(x)\overline{g}(x) \in \mathbb{F}_p[x],$$

so $\overline{\Phi_n}(x)$ has a multiple root in $\mathbb{F}_p[x]$, which implies that $x^n - 1$ has a multiple root in $\mathbb{F}_p[x]$. But as p is not a factor of n, $x^n - 1$ can not have a multiple root in $\mathbb{F}_p[x]$, for otherwise nx^{n-1} would have a non-zero root in $\mathbb{F}_p$. We have thus proved that $g(\zeta^p) \neq 0$. Thus $f(\zeta^p) = 0$. We have proved the claim. □

One can use (3.2) to inductively compute Φ_n. The first few are as follows:

$$\begin{aligned}
&\Phi_1(x) = x - 1, \quad \Phi_2(x) = x + 1, \quad \Phi_3(x) = x^2 + x + 1,\\
&\Phi_4(x) = x^2 + 1, \quad \Phi_5(x) = x^4 + x^3 + x^2 + x + 1, \quad \Phi_6(x) = x^2 - x - 1,\\
&\Phi_7(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1, \quad \Phi_8(x) = x^4 + 1,\\
&\Phi_9(x) = x^6 + x^3 + 1, \quad \Phi_{10}(x) = x^4 - x^3 + x^2 - x + 1.
\end{aligned}$$

$\Phi_{105}(x)$ is the first cyclotomic polynomial with a coefficient other than 0 and ± 1.

Note that for the special case when $n = p$ is a prime number, we have proved in Example 1.4.8 the polynomial $x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible over $\mathbb{Q}$, so

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

§ 4. Finite field extensions

We can adjoin many elements to a field to construct new fields. Recall that if L is an extension of K, for a subset S of L, $K(S)$ is defined to be the smallest sub-field of L containing K and S and is called the sub-field of L generated by S over K. If $S = \{a, b\}$, clearly

$$K(a, b) = K(a)(b)$$

as sub-fields of L. In general, if $S = \{a_1, a_2, \cdots, a_n\}$, we have a *tower of fields*

$$K \subset K(a_1) \subset K(a_1, a_2) \subset \cdots \subset K(a_1, a_2, \cdots, a_n).$$

By definition, we have

$$K(a_1, \cdots, a_n) = \left\{ \frac{f(a_1, \cdots, a_n)}{g(a_1, \cdots, a_n)} : \ f, g \in K[x_1, \ldots, x_n], g(a_1, \cdots, a_n) \neq 0 \right\}.$$

Define the sub-ring of L generated by $a_1, \ldots, a_n \in L$ over K as

$$K[a_1, \cdots, a_n] = \{f(a_1, \ldots, a_n) : f(x_1, \ldots, x_n) \in K[x_1, \ldots, x_n]\}.$$

Proposition 3.1.31. *Let L be an extension of K. If $a_1, a_2, \cdots a_n$ are all algebraic over K, then $K(a_1, a_2, \cdots, a_n)$ is a finite extension of K, and*

$$K(a_1, a_2, \cdots, a_n) = K[a_1, a_2, \cdots, a_n] \subset L.$$

Proof. Let $K_0 = K$ and for $1 \leq i \leq n$, let

$$K_i = K(a_1, \ldots, a_i) = K_{i-1}(a_i)$$

Then we have a tower of field extensions

$$K \subset K_1 \subset \cdots \subset K_n \subset L.$$

For $1 \leq i \leq n$, a_i, being algebraic over K, is also algebraic over K_{i-1}, so K_i is a finite extension of K_{i-1}. By the Tower Theorem, K_n is a finite extension over K. Moreover,

$$\begin{aligned} K_n &= K_{n-1}[a_n] = K_{n-2}[a_{n-1}][a_n] = K_{n-2}[a_{n-1}, a_n] = \cdots \\ &= K[a_1, \ldots, a_{n-1}, a_n]. \end{aligned}$$

□

We now give some important consequences of Proposition 3.1.31.

Lemma 3.1.32. *If L is a finite extension of K, then every element in L is algebraic over K.*

Proof. If L is an extension of K of degree n, then for any $a \in L$, the subset $\{1, a, a^2, \ldots, a^n\}$ must be linearly dependent over K, so there exist $\alpha_0, \alpha_1, \ldots, \alpha_n \in K$ such that $\alpha_0 + \alpha_1 a + \alpha_2 a^2 + \cdots + \alpha_n a^n = 0$. It follows that $f(a) = 0$ for $f(x) = \sum_{i=1}^{n} \alpha_i x^i \in K[x]$. Thus a is algebraic over K. □

Corollary 3.1.33. *An extension L of K is finite iff there exist $a_1, a_2, \cdots, a_n \in L$ which are algebraic over K such that $L = K(a_1, a_2, \cdots, a_n)$.*

Proof. If L is a finite extension of K, then L is an algebraic extension of K by Lemma 3.1.32. If $\{a_1, \ldots, a_n\}$ is a basis of L over K, then every a_j is algebraic over K and $L = K(a_1, a_2, \cdots, a_n)$. The converse holds by Proposition 3.1.31. □

We now give a very important class of examples of field extensions generated by finitely many algebraic elements. These examples will be studied in more details later.

Example 3.1.34. For any $f \in \mathbb{Q}[x]$, let $a_1, \ldots, a_n$ be all the roots of f in $\mathbb{C}$. Then

$$L = \mathbb{Q}[a_1, a_2, \ldots, a_n]$$

is a finite extension of $\mathbb{Q}$. The field f is called the *splitting field* of f in $\mathbb{C}$, and we will return to them in §3.2.

For algebraic elements $a_1, \ldots, a_n$ in L over K, we can compute the degree of $K[a_1, \ldots, a_n]$ over K by using the Tower Theorem.

Example 3.1.35. Consider $L = \mathbb{Q}(\sqrt{2}, \sqrt{3}) \subset \mathbb{C}$. Since both $\sqrt{2}$ and $\sqrt{3}$ are algebraic over $\mathbb{Q}$, L is a finite extension of $\mathbb{Q}$. What is the degree $[L : \mathbb{Q}]$?

We have $L = \mathbb{Q}(\sqrt{2})(\sqrt{3})$. Since the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$, we have $[L : \mathbb{Q}(\sqrt{2})] = 2$. Now what is the irreducible polynomial of $\sqrt{3}$ over $\mathbb{Q}(\sqrt{2})$? We know that $\sqrt{3}$ is a root of $x^2 - 3 = 0$. The polynomial $x^2 - 3$ is also irreducible over $\mathbb{Q}(\sqrt{2})$. Indeed, it has no zero in $\mathbb{Q}(\sqrt{2})$ (if it did, we would get a contradiction to the fact that $\sqrt{2}$ is irrational). Thus $[L : \mathbb{Q}(\sqrt{2})] = 2$. Thus $[L : \mathbb{Q}] = 4$.

Consider $K = \mathbb{Q}(\sqrt{2} + \sqrt{3})$. Clearly $K \subset L$. Let $\alpha = \sqrt{2} + \sqrt{3}$. It follows from

$$\begin{cases} \alpha = \sqrt{2} + \sqrt{3}, \\ \alpha^3 = 15\sqrt{2} + 5\sqrt{3} \end{cases}$$

that

$$\sqrt{2} = \frac{-5\alpha + \alpha^3}{10}, \qquad \sqrt{3} = \frac{15\alpha - \alpha^3}{10}.$$

Thus $L \subset K$, so $K = L$. We have already seen that the minimal polynomial of $\sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$ is $p(x) = x^4 - 10x^2 + 1$.

One in fact has the following remarkable fact, which we will prove later.

Theorem 3.1.36. *[Primitive Element Theorem] (see Lange's book, Theorem 2.5 on Page 272). Every finite extension of a field K of characteristic 0 is of the form $K(\alpha)$ for a single algebraic number over K.*

Corollary 3.1.37. *Let K be a field of characteristic 0. Then a field extension L of K is finite if and only if $L = K(a)$ for an algebraic element $a \in L$ over K.*

§ 5. Algebraic extensions and algebraically closed fields

Definition 3.1.38. A field extension $K \to L$ is said to be algebraic if every $a \in L$ is algebraic over K.

We can now re-formulate Lemma 3.1.32 as

Lemma 3.1.39. *Every finite field extension is algebraic.*

The converse is not true, as we will see very soon.

Corollary 3.1.40. *Let L be an extension of K and let $a \in L$. Then a is algebraic over K if and only if $K(a)$ is an algebraic extension of K.*

Proof. Assume that $a \in L$ is algebraic over K. Then $K(a)$ is a finite extension, so by Lemma 3.1.39, $K(a)$ is an algebraic extension of K. The converse is obvious. □

Example 3.1.41. Let $\alpha = \sqrt[5]{2}$. Then α is algebraic over $\mathbb{Q}$. Thus

$$\beta = \alpha^7 + 3\alpha^3 - 67$$

is algebraic over $\mathbb{Q}$. This is not obvious from the definition: it is not easy to find $f \in \mathbb{Q}[x]$ such that $f(\beta) = 0$.

Proposition 3.1.42. *If both $K \to L$ and $L \to M$ are algebraic field extensions, $K \to M$ is also an algebraic extension.*

Proof. Let $a \in M$ be arbitrary. As the extension $L \to M$ is algebraic, there exists $f(x) = \sum_{i=0}^{n} \alpha_i x^i \in L[x]$ such that $f(a) = 0$. Let

$$L' = K(\alpha_0, \alpha_1, \ldots, \alpha_n) \subset L.$$

As $K \to L$ is algebraic, every $\alpha_i \in L$ is algebraic over K. By Proposition 3.1.31, L' is a finite extension of K. Since a is algebraic over L', $L'(a)$ is a finite extension over L'. By the Tower Theorem, $L'(a)$ is a finite extension of K. By Lemma 3.1.32, $L'(a)$ is an algebraic extension over K. Thus $a \in L'(a)$ is algebraic over K. □

If L is a field extension of K, set

$$\overline{K}^L = \text{the set of all elements in } L \text{ that are algebraic over } K.$$

Proposition 3.1.43. *For any field extension L of K, the subset $\overline{K}^L$ of L is a sub-field of L.*

Proof. Assume that $a, b \in \overline{K}^L$ and $b \neq 0$. By Proposition 3.1.31, $K(a, b)$ is a finite extension over K. By Lemma 3.1.32, every element in $K(a, b)$ are algebraic over K. Then $a \pm b, ab, a/b \in \overline{K}^L$. □

Definition 3.1.44. Given a field extension $K \to L$, the sub-field $\overline{K}^L$ of L is called the *relative algebraic closure* of K in L.

We now turn to algebraically closed fields. Recall that the *Fundamental Theorem of Algebra* says that every polynomial $f(x) \in \mathbb{C}[x]$ has a root in $\mathbb{C}$.

Definition 3.1.45. A field L is said to be *algebraically closed* if every $f(x) \in L[x]$ has a root in L.

In this section, we show that every sub-field K of $\mathbb{C}$ has the so-called *algebraic closure* in $\mathbb{C}$ which is algebraically closed.

Theorem 3.1.46. *Assume that C is an algebraically closed field and $K \subset C$ is any sub-field. Then the relative algebraic closure $\overline{K}^C$ of K in C is algebraically closed.*

Proof. Let $L = \overline{K}^C \subset C$. Let

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + x^n \in L[x]$$

be arbitrary and non-zero. We need to show that f has a root in L. Since C is algebraically closed, f has a root α in C. Then $L(\alpha)$ is an algebraic extension of L. Since L is an algebraic extension of K, $L(\alpha)$ is an algebraic extension over K by Proposition 3.1.42. Thus α is algebraic over K. Hence $\alpha \in L$. □

Corollary 3.1.47. *The relative algebraic closure in $\mathbb{C}$ of any sub-field of $\mathbb{C}$ is algebraically closed.*

Definition 3.1.48. The relative closure of $\mathbb{Q}$ in $\mathbb{C}$ is denoted by $\overline{\mathbb{Q}}$, and simply called the algebraic closure of $\mathbb{Q}$.

Lemma 3.1.49. *The algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$ is countable and $[\overline{\mathbb{Q}} : \mathbb{Q}] = +\infty$.*

Proof. As $\mathbb{Q}$ is countable, $\mathbb{Q}[x]$ is countable. As every element in $\overline{\mathbb{Q}} \subset \mathbb{C}$ is a root of some $f \in \mathbb{Q}[x]$, $\overline{\mathbb{Q}}$ is countable. We know that for any integer n, $\mathbb{Q}(2^{1/n})$ has degree n over $\mathbb{Q}$ because the minimal polynomial of $2^{1/n}$ is $x^n - 2$. Thus $[\overline{\mathbb{Q}} : \mathbb{Q}] \geq n$ for every n, so $[\overline{\mathbb{Q}} : \mathbb{Q}] = +\infty$. □

In general, an *algebraic closure* of an arbitrary field K is an algebraic extension of K that is also algebraically closed. One can show that such a closure always exists and is unique up to isomorphisms. Our discussion above shows that if K already lies inside an algebraically closed field C, then the relative algebraic closure of K in C is already the algebraic closure we are looking for. For fields that a priori do not lie in any algebraically closed fields, the construction uses Zorn's Lemma and the construction of splitting fields that we will talk about in later sections.

The following lemma helps us to write down a lot of elements in $\overline{\mathbb{Q}}$.

Lemma 3.1.50. *If $\alpha \in \overline{\mathbb{Q}}$ and $\alpha > 0$, then for any integer $n \geq 1$, $\alpha^{1/n} \in \overline{\mathbb{Q}}$.*

Proof. Let $\beta = \alpha^{1/n}$. Since $\beta^n = \alpha$, β is algebraic over $L = \mathbb{Q}(\alpha)$, so $L(\beta)$ is an algebraic extension of L. Since L is an algebraic extension of $\mathbb{Q}$, by Proposition 3.1.42, $L(\beta)$ is an algebraic extension of $\mathbb{Q}$. Thus β is algebraic over $\mathbb{Q}$. Hence $\beta \in \overline{\mathbb{Q}}$. □

Example 3.1.51. One has

$$-2i\sqrt[3]{9-\sqrt{2}} + \frac{\sqrt{\sqrt{2}+\sqrt[3]{\sqrt{5}+3}}}{-3+i\sqrt{\sqrt{7}+2\sqrt[3]{\sqrt{5}+3}}} \in \overline{\mathbb{Q}}.$$

§ 6. Ruler-and-Compass constructions

In this section, we look at some questions that have been around since the ancient Greek time. We first formulate these questions into precise mathematical ones. Once this is done, it can be solved beautifully using the notion of degrees of field extensions.

Suppose that you are given a piece of (an infinite) blank paper, a pencil, a ruler (with no marks!), and a compass: the ruler can be used to draw straight lines, and the compass can be used to draw circles. Assume also that the ruler and the compass are also ideal in the sense that the ruler is infinitely long and the legs of the compass can be stretched as wide as you want. Start with any two distinct points on the paper and call them P_0 and P_1.

Definition 3.1.52. Given a set of points $S = \{P_0, P_1, P_2, \ldots, P_n\}$ on the paper,

- a (straight) line drawn on the paper is said to be constructable from S if it passes two distinct points in S;
- a circle drawn on the paper is said to be constructable from S if it is centered at a point in S and its radius is the distance between two distinct points in S;
- a point P on the paper is said to be *constructable from S* if P is the intersection of two lines, or one line and a circle, or two circles, which are constructable from S.

Definition 3.1.53. A point P on the paper is said to be *constructable* by a ruler and a compass if either $P = P_0$ or $P = P_1$, or if there exists a finite sequence $P_0, P_1, P_2, \ldots, P_n = P$ of points with $n \geq 2$ such that for each $1 \leq j \leq n$, P_{j+1} is constructable from the set $S_j = \{P_0, P_1, \ldots, P_j\}$.

Questions. Is every point on the paper constructable?

We first use P_0 as the origin and put coordinates for points on the paper so that we can identify the paper with $\mathbb{R}^2$.

Putting coordinates on the paper:

1. Draw the line connecting P_0 and P_1 and denote it by L_x;
2. Define the distance between P_0 and P_1 to be 1;
3. Construct the line through P_0 and perpendicular to L_x as follows: draw the circle C_0 with center at P_0 and passing through P_1; denote by P_{-1} the intersection of C_0 with the line L_x that is not P_1; draw the circle C_1 centered at P_1 and passing through P_{-1}; draw the circle C_{-1} centered at P_{-1} passing through P_1; let Q be either one of the two intersection points of C_1 and C_{-1}; let L_y be the line connecting Q and P_0.
4. Let P_2 be one of the intersections of L_y and C_0; declare P_0 to be the origin, L_x the x-axis, L_y the y-axis, and the coordinates of P_1 and P_2 to be $(1, 0)$ and $(0, 1)$ respectively. Each point on the paper now has coordinates (x, y), and we identified the paper with $\mathbb{R}^2$.

Definition 3.1.54. With the identification of the paper with $\mathbb{R}^2$, a point $(x, y) \in \mathbb{R}^2$ is said to be constructable if the corresponding point P on the paper is.

Lemma 3.1.55. *Let $S = \{P_0, P_1, \ldots, P_n\} \subset \mathbb{R}^2$, where $p_j = (x_j, y_j)$ for $j = 1, \ldots, n$, and let $P_{n+1} = (x_{n+1}, y_{n+1}) \in \mathbb{R}^2$ be a point constructable from S. Let $K = \mathbb{Q}(x_1, y_1, \ldots, x_n, y_n)$ and*

$$L = K(x_{n+1}, y_{n+1}) = \mathbb{Q}(x_1, y_1, \ldots, x_n, y_n,\ x_{n+1}, y_{n+1}).$$

Then $[L : K] = 1$ or 2.

Proof. The equation for any line constructable from S is of the form $ax + by + c = 0$ with $a, b, c \in K$, and the equation for a circle constructable from S is of the form $x^2 + y^2 + 2dx + 2ey + f = 0$, with $d, e, f \in K$.

Case 1: P_{n+1} is the intersection of two lines constructable from S. In this case, (x_{n+1}, y_{n+1}) is a solution to two linear equations with coefficients in K, so $x_{n+1} \in K$ and $y_{n+1} \in K$, and so $L = K$;

Case 2: P_{n+1} is the intersection of a line and a circle constructable from S. In this case, (x_{n+1}, y_{n+1}) satisfies

$$ax_{n+1} + by_{n+1} + c = 0, \quad x_{n+1}^2 + y_{n+1}^2 + 2dx_{n+1} + 2ey_{n+1} + f = 0$$

with $a, b, c, d, e, f \in K$. Assume, without loss of generality, that $b \neq 0$. Then we can solve y_{n+1} from the first equation and obtain a quadratic equation in x_{n+1}. It follows that $[K(x_{n+1}, y_{n+1}) : K(x_{n+1})] = 1$ and $[K(x_{n+1}) : K] = 1$ or 2 depending on whether or not the quadratic equation satisfied by x_{n+1} can be factored into the product of two linear factors. Thus $[L : K] = 1$ or 2;

Case 3: P_{n+1} is the intersection of two circles constructable from S. In this case, (x_{n+1}, y_{n+1}) satisfies

$$\begin{cases} x_{n+1}^2 + y_{n+1}^2 + 2dx_{n+1} + 2ey_{n+1} + f = 0, \\ x_{n+1}^2 + y_{n+1}^2 + 2d'x_{n+1} + 2e'y_{n+1} + f' = 0 \end{cases}$$

with coefficients in K. Subtracting the two equations, one gets a linear equation and a quadratic one, and the case reduces to the previous one. □

Theorem 3.1.56. *If $P = (x, y) \in \mathbb{R}^2$ is a constructable point, then $[\mathbb{Q}(x, y) : \mathbb{Q}] = 2^r$ for some integer $r \geq 0$, and consequently, $[\mathbb{Q}(x) : \mathbb{Q}]$ and $[\mathbb{Q}(y) : \mathbb{Q}]$ are both powers of 2.*

Proof. By Lemma 3.1.55, there exist sequence of field extensions

$$\mathbb{Q} = L_0 \subset L_1 \subset L_2 \subset \cdots \subset L_n$$

with $[L_j : L_{j-1}] = 1$ or 2 for each $j = 1, 2, \ldots, n$, such that $(x, y) \in L_n$. By the Tower theorem, $[L_n : \mathbb{Q}] = 2^m$ for some integer $m \geq 0$. As $\mathbb{Q} \subset \mathbb{Q}(x, y) \subset L_n$, the Tower Theorem says that $[\mathbb{Q}(x, y) : \mathbb{Q}]$ divides 2^m, so $[\mathbb{Q}(x, y) : \mathbb{Q}] = 2^r$ for some integer $r \geq 0$. It now follows from

$$\mathbb{Q} \subset \mathbb{Q}(x) \subset \mathbb{Q}(x, y) \quad \text{and} \quad \mathbb{Q} \subset \mathbb{Q}(y) \subset \mathbb{Q}(x, y)$$

and the Tower Theorem that both $[\mathbb{Q}(x) : \mathbb{Q}]$ and $[\mathbb{Q}(y) : \mathbb{Q}]$ are powers of 2. □

Note that if $(x, y) \in \mathbb{R}^2$ is constructable, both x and y are algebraic over $\mathbb{Q}$. This implies in particular that the set of all constructable points in $\mathbb{R}^2$ is countable.

Definition 3.1.57. n angle α is said to be constructable if there is a constructable point $P \neq (0, 0)$ on the half line L connecting $(0, 0)$ and P that has angle α with L_x.

Example 3.1.58. The points $(1/2, \pm\sqrt{3}/2)$, being the intersection points of the circles $x^2 + y^2 = 1$ and $(x-1)^2 + y^2 = 1$ are constructable using a ruler and compass, so we also say that the angle $\pi/3$ can be constructed using a ruler and a compass. The point $(\sqrt{3}/2, 1/2)$, being an intersection of the circles $x^2 + y^2 = 1$ and $x^2 + (y-1)^2 = 1$, is also constructable, so the we say that the angle $\pi/6$ can be *bisected* using a ruler and a compass.

The Greek mathematicians tried very hard to *trisect* $\pi/3$ using a ruler and a compass.

Definition 3.1.59. An angle α is said to be constructable if there is a constructable point $P \neq (0,0)$ on the half line L connecting $(0,0)$ and P that has angle α with L_x.

Theorem 3.1.60. *The angle $\pi/3$ can not be trisected using a ruler and a compass. More precisely, the angle of $\pi/9$ is not constructable.*

Proof. If there were a point P in L not equal to $(0,0)$ that is constructable, then the intersection of L with the circle $x^2+y^2=1$, which is the point $(\cos(\pi/9), \sin(\pi/9))$, would be constructable. Let $x = \cos(\pi/9)$ and $y = \sin(\pi/9)$. We would then have

$$[\mathbb{Q}(x,y):\mathbb{Q}] = [\mathbb{Q}(x,y):\mathbb{Q}(x)][\mathbb{Q}(x):\mathbb{Q}] = 2^r$$

for some non-negative integer r, which would imply that $[\mathbb{Q}(x):\mathbb{Q}]$ is a power of 2. By the triple angle formula

$$\cos(3\alpha) = 4(\cos\alpha)^3 - 3\cos\alpha,$$

so $\alpha = \cos(\pi/9)$ satisfies $4\alpha^3 - 3\alpha - 1/2 = 0$, or, equivalently, $x = 2\alpha$ satisfies

$$x^3 - 3x - 1 = 0.$$

It is easy to see that $f(x) = x^3 - 3x - 1$ is irreducible over $\mathbb{Q}$ (indeed, by Gauss' Lemma, it is enough to show that f is irreducible over $\mathbb{Z}$. If it were reducible over $\mathbb{Z}$, it would have a linear factor, which would need to be either $x-1$ or $x+1$, but as $f(1) = -3 \neq 0$ and $f(-1) = 1 \neq 0$, this is not possible). Thus $[\mathbb{Q}(x):\mathbb{Q}] = 3$. This shows that (x,y) is not constructable using a ruler and a compass. Thus no points on L other than $(0,0)$ is constructable. □

Exercise 3.1.61. Show that the point $(\sqrt[3]{2}, 0)$ is not constructable using a ruler and a compass. This statement is re-formulated as "one can not double a cube" using ruler-and-compass, i.e. making a new cube out of an older one with twice the volume.

Exercise 3.1.62. Say a line in $\mathbb{R}^2$ is constructable if it passes through two distinct constructable points. Assume that L is a constructable line and that P is a constructable point not on L. Show that both the line through P perpendicular through L and the line through P parallel to L are constructable.

Exercise 3.1.63. Construct a regular hexagon using a ruler and a compass.

Definition 3.1.64. We say a real number $a \in \mathbb{R}$ is constructable if its absolute value $|a|$ is the distance between two constructable points in $\mathbb{R}^2$. Similarly, a complex number $a+ib$ is constructable if $(a,b) \in \mathbb{R}^2$ is constructable.

The following theorem lists some further properties of constructable numbers, and we refer to other text books (for example M. Artin's Algebra for more details).

Theorem 3.1.65. *The following are true:*

1) The set of all constructable real numbers is a sub-field of $\mathbb{R}$.

2) A point $(x, y) \in \mathbb{R}^2$ *is constructable if and only if both* x *and* y *are constructable numbers.*

3) A real number x *is constructable iff* $x \in K_n$ *for a tower of real quadratic field extensions*

$$\mathbb{Q} = K_0 \subset K_1 \subset K_2 \subset \cdots \subset K_n \subset \mathbb{R},$$

where $[K_{j+1} : K] = 2$ *for all* $0 \leq j \leq n-1$*;*

4) A complex number $x + iy \in \mathbb{C} \cong \mathbb{R}^2$ *is constructable iff* $x + iy \in K_n$ *for a tower of complex quadratic field extensions*

$$\mathbb{Q}(i) = K_0 \subset K_1 \subset K_2 \subset \cdots \subset K_n \subset \mathbb{C},$$

where $[K_{j+1} : K] = 2$ *for all* $0 \leq j \leq n-1$.

5) The set of all constructable points in $\mathbb{R}^2 \cong \mathbb{C}$*, being a sub-field of* $\overline{\mathbb{Q}}$*, is countable.*

3.2 Splitting fields

§ 1. Definitions

We first recall some basic facts on roots of polynomials. Let L be any field and let $f \in L[x]$. If $a \in L$ is a root of f, i.e., $f(a) = 0$, then, by the Euclidean algorithm, there exists $g \in L[x]$ such that

$$f(x) = (x - a)g(x).$$

Thus f has a root in L if and only if f has a linear factor in $L[x]$. If $g(a) = 0$, factor out $(x - a)$ from $g(x)$ and continue. Then there exists an integer $m \geq 1$ and $h(x) \in L[x]$ such that

$$f(x) = (x - a)^m h(x)$$

and $h(a) \neq 0$. We call m the *multiplicity* of a as a root of f. When $m \geq 2$, we say that a is a *repeated root* of f.

We now explain the *derivative test* for repeated roots. Let $f = a_0 + a_1 x + \cdots + a_n x^n \in L[x]$, where $a_n \neq 0$, and assume that $n \geq 1$. Define $f' \in L[x]$ by

$$f' = a_1 + 2a_2 x + \cdots + n a_n x^{n-1} \in L[x].$$

and call it the derivative of f.

Lemma 3.2.1. *Let $f \in L[x]$ with $deg(f) \geq 2$. An element $a \in L$ is a repeated root of f if and only if $f(a) = 0$ and $f'(a) = 0$.*

Proof. Let m be the unique non-negative integer such that $f(x) = (x-a)^m h(x)$ and $h(a) \neq 0$. Then a is a root if and only if $m \geq 1$, and a is a repeated root if and only if $m \geq 2$. If a is a repeated root, then

$$f'(x) = m(x-a)^{m-1}h(x) + (x-a)^m h'(x),$$

so $f(a) = f'(a) = 0$. Conversely, if $f(a) = f'(a) = 0$, and if $m = 1$, then $f'(a) = h(a) \neq 0$, a contradiction, so $m \geq 2$. □

Note that if the characteristic of L is 0, and $f = a_0 + a_1 x + \cdots + a_n x^n \in L[x]$ with $a_n \neq 0$, then $na_n \neq 0$, so in particular $f' \neq 0$ and $\deg(f') = n-1$. When L has characteristic $p > 0$, it is possible that f is non-constant but $f' = 0$. For example, $f(x) = x^p \in \mathbb{F}_p[x]$ has $f' = 0$.

Lemma 3.2.2. *Assume that K is a field of characteristic 0 and $f \in K[x]$ is irreducible. Then f has no repeated roots in any field extension L of K.*

Proof. Suppose that L is an extension of K and that f has a root $a \in L$ of multiplicity at least 2. Then $f'(a) = 0$ as well. As K has characteristic 0, $f' \neq 0$, and since f is irreducible, f and f' are relatively prime. Thus there exist $\phi, \psi \in K[x]$ such that

$$\phi f + \psi f' = 1. \tag{3.1}$$

Regarding (3.1) as an identity in $L[x]$ and setting $x = a$, one gets $0 = 1$, a contradiction. Thus f can not have repeated roots in L. □

Note that the arguments in the proof of Lemma 3.2.2 do not hold if the characteristic of K is not 0.

Definition 3.2.3. Let L be a field. A polynomial $f(x) \in L[x]$ of degree $n \geq 1$ is said to *split completely* over L (or *split* over L) if f has exactly n roots (counting multiplicity) in L, i.e., if there exist $c_0, a_1, \cdots, a_n \in L$, not necessarily pointwise distinct, such that

$$f(x) = c_0(x-a_1)\cdots(x-a_n).$$

The following is an easy but crucial fact.

Lemma 3.2.4. *If L is an algebraically closed field, then every $f \in L[x]$ splits completely over L.*

Proof. Assume that $f \in L[x]$ has positive degree. Since L is algebraically closed, it has a root a in L, so $f = (x-a)^m h(x)$ for some $m \geq 1$ and $h(a) \neq 0$. As $\deg(h) < \deg(f)$, induction leads to the statement. □

Definition 3.2.5. [Splitting fields] Let K be a field and let $f \in K[x]$ with $n = \deg(f) \geq 1$. By a *splitting field of f over K* we mean an extension L of K such that f splits completely in L and that $L = K(a_1, a_2, \cdots, a_n)$, where $a_1, \ldots, a_n$ are the n roots of f in L (not necessarily pairwise distinct).

Remark 3.2.6. 1) As a way of checking our understanding of the definition, note that if $f \in K[x]$ has degree 1, then f always splits in K, so K itself is a splitting field of f over K.

2) We emphasize on the word "over K" in "splitting field of f over K". For example, the polynomial $f(x) = x^2 + 1$ can be regarded as in $\mathbb{Q}[x]$ or as in $\mathbb{R}[x]$. The splitting field of f over $\mathbb{Q}$ in $\mathbb{C}$ is $\mathbb{Q}(i)$, while the splitting field of f over $\mathbb{R}$ in $\mathbb{C}$ is $\mathbb{R}(i) = \mathbb{C}$.

3) By Proposition 3.1.31, a splitting field of any $f \in K[x]$ over K is a finite extension of K.

Note that it follows from the definition and Proposition 3.1.31 that any splitting field L of $f \in K[x]$ is a finite extension of K. The following Theorem 3.2.7 will be proved later.

Theorem 3.2.7. *Let K be a field and $f \in K[x]$ with positive degree.*

1) Splitting fields of f over K exist;

2) If $K \subset L$ and $K \subset L'$ are two splitting fields of f over K, then there exists a an isomorphism $\sigma : L \to L'$ such that $\sigma(a) = a$ for all $a \in K$.

§ 2. Examples of splitting fields

While we will prove the existence and uniqueness of splitting fields for any $f \in K[x]$ for any field K, we now look at the case when K is a sub-field of an algebraically closed field. For example, we look at $f(x) \in \mathbb{Q}[x]$ and regard $\mathbb{Q}$ as a sub-field of $\mathbb{C}$. In this case, the existence of splitting fields of any $f \in K[x]$ is easy to prove.

By Lemma 3.2.4, if $K \subset L$ and L is algebraically closed, then every $f \in K[x]$ with positive degree, when regarded as an element in $L[x]$, splits completely in $L[x]$. If $a_1, \ldots, a_n \in L$ are all the roots of f in L, then the sub-field $K(a_1, \ldots, a_n)$ of L is a splitting field of f over K. Thus the existence of splitting field for $f \in K[x]$ over K is proved.

Definition 3.2.8. Let L be algebraically closed field and K a sub-field of L. Let $f \in K[x]$ with $n = \deg(f) > 0$, and let $a_1, \ldots, a_n \in L$ the roots of f in L. The sub-field $K(a_1, \ldots, a_n)$ of L is a splitting field of f over K, also called *the splitting field of f over K in L*.

Example 3.2.9. Recall that for an integer $n \geq 2$, the n'th cyclotomic field is $\mathbb{Q}(\omega_n) \subset \mathbb{C}$, where $\omega_n = e^{2\pi i/n} \in \mathbb{C}$. As $\mathbb{Q}(\omega_n)$ contains all the roots ω_n^k for $0 \leq k \leq n-1$, it is the splitting field of the polynomial $f_n(x) = x^n - 1$ over $\mathbb{Q}$ in $\mathbb{C}$.

Example 3.2.10. The splitting field of $f(x) = x^3 - 2$ in $\mathbb{C}$ over $\mathbb{Q}$ in $\mathbb{C}$ is $\mathbb{Q}(\alpha, \beta, \gamma)$, where α, β and γ are the three complex roots of $x^3 - 2$. Let $\sqrt[3]{2}$ be the real root for $x^3 - 2$, and let $\omega = e^{\frac{2\pi i}{3}}$. Then the three roots for $x^3 - 2$ are $\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2}$. Thus the splitting field of $f(x) = x^3 - 2$ in $\mathbb{C}$ is

$$\mathbb{Q}(\sqrt[3]{2},\ \omega\sqrt[3]{2},\ \omega^2\sqrt[3]{2}) = \mathbb{Q}(\sqrt[3]{2},\ \omega).$$

To determine the degree of $\mathbb{Q}(\sqrt[3]{2}, \omega)$, we first note that since the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$ which is irreducible, the degree of $\mathbb{Q}(\sqrt[3]{2})$ over $\mathbb{Q}$ is 3. Now ω satisfies $x^2 + x + 1 = 0$ and the quadratic polynomial $x^2 + x + 1$ is irreducible over $\mathbb{Q}(\sqrt[3]{2})$ because it has no roots in $\mathbb{Q}(\sqrt[3]{2})$, the degree of $\mathbb{Q}(\sqrt[3]{2}, \omega)$ over $\mathbb{Q}(\sqrt[3]{2})$ is 2. Thus the degree of $\mathbb{Q}(\sqrt[3]{2}, \omega)$ over $\mathbb{Q}$ is 6.

Example 3.2.11. For $f(x) = x^4 - 4 = (x^2 - 2)(x^2 + 2) \in \mathbb{Q}[x]$, the splitting field of f over $\mathbb{Q}$ in $\mathbb{C}$ is

$$K = \mathbb{Q}(\pm\sqrt{2},\ \pm\sqrt{2}i) = \mathbb{Q}(\sqrt{2},\ i).$$

It is easy to see that $[K : \mathbb{Q}] = 4$.

Example 3.2.12. For $f(x) = x^4 + x^2 + 1 = (x^2 + x + 1)(x^2 - x + 1) \in \mathbb{Q}[x]$, the four roots of f in $\mathbb{C}$ are $\pm e^{\frac{\pi i}{3}}$ and $\pm e^{\frac{2\pi i}{3}}$, so the splitting field of f in $\mathbb{C}$ is

$$K = \mathbb{Q}\left(\pm e^{\frac{\pi i}{3}},\ \pm e^{\frac{2\pi i}{3}}\right) = \mathbb{Q}\left(e^{\frac{\pi i}{3}}\right).$$

As $\omega = e^{\pi i/3}$ satisfies $x^2 - x + 1 = 0$ and $x^2 - x + 1$ is irreducible over $\mathbb{Q}$, $[K : Q] = 2$. Note that $\mathbb{Q}\left(e^{\frac{\pi i}{3}}\right)$ is the 6'th cyclotomic extension of $\mathbb{Q}$.

For the rest of the section, let K be any sub-field of $\mathbb{C}$. We give a systematic analysis of splitting fields in $\mathbb{C}$ of quadratic or cubic polynomials $f \in K[x]$ over K. The case of quadratic polynomials is easy.

Example 3.2.13. Assume that K is a sub-field of $\mathbb{C}$. It is well-known that roots of the polynomial $f(x) = x^2 + bx + c$ are given by

$$x = -\frac{1}{2}b \pm \frac{1}{2}\sqrt{b^2 - 4c}.$$

Thus the splitting field in $\mathbb{C}$ of $f(x) = x^2 + bx + c$ over K is $L = K(\sqrt{b^2 - 4c})$. Moreover, $L = K$ if $\sqrt{b^2 - 4c} \in K$. Otherwise, $[L : K] = 2$.

We now turn to cubic polynomials $f \in K[x]$. The following simple fact says that we can also assume that the coefficient of x^2 of f is zero.

Lemma 3.2.14. *Let K be a field of characteristic 0. For any*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in K[x],$$

replacing x by $x = z - \frac{a_{n-1}}{n}$ gives a degree n polynomial

$$\tilde{f}(z) = f(z - a_{n-1}) \in K[z]$$

whose coefficient of the term z^{n-1} is 0.

For a cubic $f(x) = x^3 + ax^2 + bx + c \in K[x]$, by setting $x = z - a/3$, we get $\tilde{f} = z^3 + pz + q = 0$, where

$$p = -\frac{a^2}{3} + b, \quad q = \frac{2a^3}{27} - \frac{ab}{3} + c. \tag{3.2}$$

The following fact on the roots in $\mathbb{C}$ of cubic polynomials was already known in the middle of the 16th century.

Lemma 3.2.15. *Let $a, b, c \in \mathbb{C}$, and let p, q be given as in (3.2). Then the three roots of $f(x) = x^3 + ax^2 + bx + c$ in $\mathbb{C}$ are*

$$\alpha_1 = -\frac{a}{3} + \beta_1 + \beta_2, \quad \alpha_2 = -\frac{a}{3} + \omega\beta_1 + \omega^2\beta_2, \quad \alpha_3 = -\frac{a}{3} + \omega^2\beta_1 + \omega\beta_2,$$

where $\omega = e^{2\pi i/3}$ and $\beta_1, \beta_2 \in \mathbb{C}$ are any cubic roots

$$\beta_1 = \sqrt[3]{\frac{1}{2}\left(-q + \sqrt{q^2 + \frac{4p^3}{27}}\right)}, \quad \beta_2 = \sqrt[3]{\frac{1}{2}\left(-q - \sqrt{q^2 + \frac{4p^3}{27}}\right)},$$

satisfying $\beta_1\beta_2 = -p/3$. Define

$$\Delta_f = (\alpha_1 - \alpha_2)^2(\alpha_2 - \alpha_3)^2(\alpha_3 - \alpha_1)^2, \tag{3.3}$$

where $\alpha_1, \alpha_2, \alpha_3$ are the three roots of the polynomial given above. Then

$$\Delta_f = -4p^3 - 27q^2 = a^2b^2 + 18abc - 4b^3 - 4a^3c - 27c^2 \in K. \tag{3.4}$$

Proof. By making the change of variables $z = x + \frac{a}{3}$, one first reduce the equation on x to the equation $z^3 + pz + q = 0$ on z with $p, q \in \mathbb{C}$ given in (3.2). Write $z = \beta_1 + \beta_2$ and imposing the condition $\beta_1\beta_3 = -p/3$, one gets

$$\beta_1^3 + \beta_2^3 = -q, \quad \beta_1^3\beta_2^3 = -p^3/27.$$

Solve for β_1 and β_2 from here. The claim on Δ_f is by a direct calculation. □

Definition 3.2.16. The element $b^2 - 4c \in K$ is called the discriminant of the quadratic polynomial $f(x) = x^2 + bx + c$, and the element $\Delta_f \in K$ given in (3.4), i.e.,

$$\Delta_f = -4p^3 - 27q^2 = a^2b^2 + 18abc - 4b^3 - 4a^3c - 27c^2 \in K,$$

is called the discriminant of the cubic polynomial $f(x) = x^3 + ax^2 + bx + c$.

In general, the discriminant of a monic polynomial $f(x) \in K[x]$ for any field K is an integral polynomial Δ_f in the coefficients of f. The discriminant of f gives information on the roots of f without knowing what the roots are. For example, $\Delta_f = 0$ if and only if f has a repeated root in some field extension of K. Read any textbook or online articles for further details on discriminants.

Returning now to the case of $f \in K[x]$ being an cubic polynomial, where K is any sub-field of $\mathbb{C}$, let $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}$ be the three roots of f in $\mathbb{C}$, and let

$$L = K(\alpha_1, \alpha_2, \alpha_3) \subset \mathbb{C}$$

be the splitting field of f in $\mathbb{C}$. We would like to determine the degree $[L : K]$. Note first that if $f(x) = x^3 + ax^2 + bx + c$, then $\alpha_1 + \alpha_2 + \alpha_3 = -a \in K$, so

$$L = K(\alpha_1, \alpha_2) = K(\alpha_2, \alpha_3) = K(\alpha_1, \alpha_3).$$

If f is reducible, then f must have a root in K, so $[L : K] = 1$ if all the three roots are in K and $[L : K] = 2$ if f has exactly one root with multiplicity one in K.

We can thus assume now that the cubic polynomial $f \in K[x]$ is irreducible. We can also assume that f is monic. By Lemma 3.2.2, the three roots $\alpha_1, \alpha_2, \alpha_3$ are pairwise distinct, so by (3.3), $\Delta_f \neq 0$.

Proposition 3.2.17. *Let K be a sub-field of $\mathbb{C}$, let $f = x^3 + ax^2 + bx + c \in K[x]$ be an irreducible cubic polynomial, and let L be the splitting field of f in $\mathbb{C}$. If Δ_f has a square root in K, then $[L : K] = 3$. Otherwise, $[L : K] = 6$.*

Proof. We may assume that $f(x) = x^3 + px + q$. Let $\alpha_1, \alpha_2, \alpha_3$ be the three roots of f in $\mathbb{C}$, and let

$$\delta = (\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)(\alpha_2 - \alpha_3) \in \mathbb{C}.$$

Then by definition $\delta^2 = \Delta_f$. Note that although the definition of δ depends on the ordering of the three roots, different choices of such orderings only change δ to possibly $-\delta$, so Δ_f has a square root in K if and only if $\delta \in K$. Using $\alpha_1+\alpha_2+\alpha_3 = 0$ and the definition of δ, one can show that

$$\alpha_2 = \frac{\delta + 2\alpha_1 p + 3q}{2(3\alpha_1^2 + p)}.$$

Thus $L = K(\alpha_1, \delta)$. Assume first that $\delta \in K$. Then $L = K(\alpha_1)$. Since f is the minimal polynomial of α_1, we have $[L : K] = 3$.

Assume now that $\delta \notin K$. Then

$$[L : K] = [K(\alpha_1)(\delta) : K(\alpha_1)][K(\alpha_1) : K] = 3[K(\alpha_1)(\delta) : K(\alpha_1)].$$

Thus $[L : K]$ is divisible by 3. Moreover, as δ satisfies the quadratic equation $\delta^2+\gamma = 0$, where $\gamma = 4p^3 + 27q^2 \in K \subset K(\alpha_1)$, $[L : K(\alpha_1)] = 1$ or 2 depending on whether

$\delta \in K(\alpha_1)$ or $\delta \notin K(\alpha_1)$, so $[L : K] \leq 6$. Similarly, as $\delta \notin K$ but $\delta^2 \in K$, $[K(\delta) : K] = 2$, so

$$[L : K] = [K(\delta)(\alpha_1) : K(\delta)][K(\delta) : K] = 2[K(\delta)(\alpha_1) : K(\delta)],$$

so $[L : K]$ is divisible by 2 as well. Thus $[L : K] = 6$. □

Example 3.2.18. The polynomial $f(x) = x^3 - 4x + 2 \in \mathbb{Q}[x]$ is irreducible by Eisenstein's criterion. We have

$$\Delta_f = -4(-4)^3 - 27 \times 4 = 4 \times 37,$$

which has no square root in $\mathbb{Q}$. Let L be the splitting field of f over $\mathbb{Q}$ in $\mathbb{C}$. Then $[L : \mathbb{Q}] = 6$.

§ 3. Existence of splitting fields

In this section, we prove the existence of splitting fields for an arbitrary polynomial over an arbitrary field K.

Recall first that if K is a sub-field of an algebraically closed field L, for example, $K = \mathbb{Q}$ and $L = \mathbb{C}$, then for any $f \in K[x]$ of positive degree, we rely on the following lemma to construct a splitting field of f over K.

Lemma 3.2.19. *If L is a field extension of K and $f(x) \in K[x]$ completely splits in $L[x]$ as*

$$f(x) = c(x - a_1)(x - a_2) \cdots (x - a_n),$$

where $c \in K$ and $a_1, a_2 \ldots, a_n \in L$. Then $K(a_1, a_2, \ldots, a_n)$ is a splitting field of f over K.

Note that in Lemma 3.2.19, we do not need L to be algebraically closed, and we only require that f completely split in $L[x]$. For any arbitrary K and arbitrary $f \in K[x]$, if we can always find an extension L of K over which f completely splits, then we have a splitting field of f by Lemma 3.2.19. But how to construct such an L? The first step is to find an L in which f has a root.

Lemma 3.2.20. *For any field K and $f \in K[x]$ with $n = \deg(f) \geq 1$, there exists a field extension L of K with $[L : K] \leq n$ such that f has a root in L, so f has a linear factor in $L[x]$.*

Proof. Let $p(x)$ be an irreducible factor of f and let $L = K[x]/\langle p(x)\rangle$. Then

$$[L : K] = \deg(p) \leq n.$$

Let a be the image of x in L. Then $p(a) = 0$ so $f(a) = 0$. By the division algorithm, there exists $f_1(x) \in L[x]$ such that $f(x) = (x - a)f_1(x) \in L[x]$. □

Lemma 3.2.21. *Suppose that K is a field, $f \in K[x]$ has degree $n \geq 2$, and L is an extension of K such that f has a root $a \in L$. Let $f_1(x) \in L[x]$ be such that*

$$f(x) = (x - a)f_1(x).$$

Then $f_1(x) \in K(a)[x]$. If f splits completely in $L[x]$, then so does f_1. If L is a splitting field of f over K, then L is a splitting field of f_1 over $K(a)$.

Proof. Without loss of generality, we may assume that f is monic. Write

$$f_1(x) = x^{n-1} + b_{n-2}x^{n-2} + \cdots + b_1x + b_0,$$

where $b_j \in L$ for each j. Using $f(x) = (x - a)f_1(x)$ and by comparing coefficients, one sees that $b_j \in K(a)$ for each j, so $f_1 \in K(a)[x]$.

Suppose now that f splits completely in $L[x]$. If $\{a, a_2, \ldots, a_n\}$ are all the roots of f in L counting multiplicity, it follows from

$$f(x) = (x - a)(x - a_2) \cdots (x - a_n)$$

that $f_1(x) = (x - a_2) \cdots (x - a_n)$, so f_1 splits completely over L. Moreover, if L is a splitting field of f over K, then

$$L = K(a, a_2, \ldots, a_n) = K(a)(a_2, \ldots, a_n),$$

so L is a splitting field of f_1 over $K(a)$. □

Exercise 3.2.22. Show if L is a splitting field of $f \in K[x]$, then for any sub-field L_1 of L containing K, L is also a splitting field of f over L_1.

We are now ready to prove the existence of splitting fields which is also called the Kronecker Theorem.

Theorem 3.2.23. *[Existence of splitting fields] For any field K and any $f(x) \in K[x]$ with $\deg(f) = n \geq 1$, there exists a splitting field L of f over K with $[L : K] \leq n!$.*

Proof. We prove by induction on n that there exists an extension L of K over which f completely splits. By Lemma 3.2.19, f has a splitting field over K.

If $n = 1$, then f is linear and K is already a splitting field of f over K. Assume now that $n > 1$. By Lemma 3.2.20, there exists an extension L_1 of K with $[L_1; K] \leq n$ and f has a root in L_1. Thus

$$f(x) = (x - a)f_1(x)$$

for some $a \in L_1$ and $f_1(x) \in L_1[x]$. Clearly $\deg(f_1) = n - 1$. By Lemma 3.2.21, $f_1(x) \in K(a)[x]$. By induction assumption, there exists an extension L of $K(a)$ over which f_1 splits completely. Regarding L as an extension of K, then f splits completely in $L[x]$. □

§ 4. Uniqueness of splitting fields

We now turn to uniqueness of splitting fields. Recall that an extension of a field K is defined to be a non-zero ring homomorphism $K \to L$, and any such homomorphism must be injective. When dealing with only one extension of K, we can identify K with its image in L so we regard K as a subset of L, and we say that $K \subset L$ is a field extension of K. Since we will be comparing different field extensions of K, we will need to specify the injective ring homomorphisms for some of the extensions of K. We make this more precise now.

Definition 3.2.24. Suppose that K is a field and $\varphi : K \to M$ is a field extension of K, so $\varphi(K) \subset M$. For $g = \sum_{i=0}^{n} c_i x^i \in K[x]$, let

$$\varphi(g) = \sum_{i=0}^{n} \varphi(c_i) x^i \in \varphi(K)[x] \subset M[x].$$

We say that $a \in M$ is a root of g if $\varphi(g)(a) = 0$, and we say that g completely splits in $M[x]$ if $\varphi(g)$ completely splits in $M[x]$. If $\psi : K \to L$ is another field extension, a ring homomorphism $\tilde{\varphi} : L \to M$ satisfying

$$\tilde{\varphi} \circ \psi = \varphi$$

is called a *K-homomorphism extending $\psi : K \to L$ and $\varphi : K \to M$*. When is it clear what the maps ψ and φ are, we simply call $\tilde{\varphi} : L \to M$ a K-homomorphism.

We now prove a crucial lemma that will lead to the uniqueness of splitting fields.

Lemma 3.2.25. *[Extension Lemma] Let K be any field, and let $K \subset L$ be a splitting field of a non-constant $f \in K[x]$ over K. If $\varphi : K \to M$ is another extension of K such that f splits completely over M, then there exists a K-homomorphism $\tilde{\varphi} : L \to M$.*

Proof. We assume that f is monic and we use induction on $n = \deg(f)$ to prove the existence of $\tilde{\varphi}$. For notational simplicity, we regard K as a subset of L.

If $n = 1$, then $L = K$ so there is nothing to prove. Assume now $n > 1$ and that Lemma 3.2.25 holds for any polynomial of degree less than n. Let $p \in K[x]$ be an irreducible factor of f and let $q \in K[x]$ be such that $f = pq$. As q, having degree less than n, can not have all the n roots of f in any field extension of K, $p(a) = 0$ for some root a of f in L. Similarly, there exists a root b of f in M which is also a root of p. As $p \in K[x]$ is irreducible, one has the field isomorphisms

$$\alpha : \ K(a) \longrightarrow K[x]/\langle p \rangle \quad \text{and} \quad \beta : \ \varphi(K)(b) \longrightarrow \varphi(K)[x]/\langle \varphi(p) \rangle.$$

Note that $\varphi : K \to \varphi(K)$ induces an isomorphism $K[x]/\langle p \rangle \to \varphi(K)/\langle \varphi(p) \rangle$ which we still denote as φ. One thus has the field extension $\beta^{-1} \circ \varphi \circ \alpha : K(a) \to M$ as the composition

$$K(a) \ \xrightarrow{\alpha} \ K[x]/\langle p \rangle \ \xrightarrow{\varphi} \ \varphi(K)[x]/\langle \varphi(p) \rangle \ \xrightarrow{\beta^{-1}} \ \varphi(K)(b) \subset M.$$

Let $f_1 \in L[x]$ be such that $f(x) = (x-a)f_1(x) \in L[x]$. By Lemma 3.2.21, $f_1 \in K(a)[x]$ and L is a splitting field of f_1 over $K(a)$. As $\deg(f_1) = n-1$, we can apply the induction assumption to L as a splitting field of f_1 over $K(a)$ and the field extension

$$\beta^{-1} \circ \varphi \circ \alpha : \ K(a) \longrightarrow M,$$

and we conclude that there exists an injective ring homomorphism $\tilde{\varphi} : L \to M$ such that $\tilde{\varphi}|_{K(a)} = \beta^{-1} \circ \varphi \circ \alpha)$. For $k \in K$, one then has

$$\tilde{\varphi}(k) = \beta^{-1}(\phi(\alpha(k))) = \beta^{-1}(\phi(k)) = \phi(k), \quad \forall\, k \in K.$$

In other words, $\tilde{\varphi}|_K = \varphi$. □

Lemma 3.2.26. *In the setting of Lemma 3.2.25, all K-homomorphisms $\tilde{\varphi} : L \to M$ extending $\varphi : K \to M$ map L to the same sub-field $\varphi(K)(\widetilde{R})$ of M, where $\widetilde{R}$ is the set of all roots of f in M.*

Proof. We may assume that f is monic. Suppose that $\tilde{\varphi} : L \to M$ is a K-homomorphism. Let $R = \{a_1, a_2, \ldots, a_n\}$ be the set of all roots of f in L counting multiplicity, so that

$$f(x) = (x-a_1)(x-a_2)\cdots(x-a_n) \in L[x].$$

Then $\varphi(f)(x) = (x-\varphi(a_1))(x-\varphi(a_2))\cdots(x-\varphi(a_n)) \in M[x]$, so f splits completely in $M[x]$, and $\widetilde{R} = \{\varphi(a_1), \ldots, \varphi(a_n)\}$ is the set of all roots of f in M. Furthermore, as $L = K(a_1, a_2, \ldots, a_2)$, one has

$$\tilde{\varphi}(L) = \varphi(K)(\varphi(a_1), \varphi(a_2), \ldots, \varphi(a_n)) = \varphi(K)(\widetilde{R}).$$

□

We can now prove the uniqueness of splitting fields.

Corollary 3.2.27. *Let K be a field and $f(x) \in K[x]$ with positive degree. If $\varphi_1 : K \to L$ and $\varphi_2 : K \to M$ are two splitting fields of f over K, then there exists a K-isomorphism $\Phi : L \to M$ extending φ_1 and φ_2.*

Proof. By Lemma 3.2.25, there exists a K-homomorphism $\Phi : L \to M$ such that $\Phi \circ \phi_1 = \phi_2$. If $a_1, \ldots, a_n$ are the roots of f in L, then $\Phi(a_1), \ldots, \Phi(a_n)$ are the roots of f in $\Phi(L) \subset M$, so f splits completely over the sub-field $\Phi(L)$ of M. As M is a splitting field of f, $\Phi(L) = M$ by Lemma 3.2.26. Thus $\Phi : L \to M$ is both injective and surjective, so Φ is an isomorphism. □

Corollary 3.2.28. *If $K \subset L$ is the splitting field of some $f \in K[x]$, and if two elements of L are roots of the same irreducible polynomial in $K[x]$, then there is an K-automorphism of L that takes one root to another.*

Proof. If $\alpha, \beta \in K[x]$ are both roots of an irreducible $p(x) \in K[x]$, then one has

$$K(\alpha) \longrightarrow K[x]\langle p\rangle \longrightarrow K(\beta) \subset L.$$

As L is also a splitting field of f over $K(\alpha)$, by the Extension Lemma, there exists a $K(\alpha)$-isomorphism from L to itself, i.e., a K-automorphism of L sending α to β. □

§ 5. Normal extensions

Definition 3.2.29. An algebraic field extension $K \subset L$ is said to be *normal* if every irreducible polynomial in $K[x]$ that has a root in L splits over L.

Note that K, as an extension of itself, is normal: if an irreducible polynomial p in $K[x]$ has a root in K, it must have degree 1 and thus splits over K.

The extension $L = \mathbb{Q}(\sqrt[3]{2})$ is not normal, because the irreducible polynomial $f(x) = x^3 - 2$ over $\mathbb{Q}$ has a root $\sqrt[3]{2}$ in L but does not split over L.

Lemma 3.2.30. *If a finite extension $K \subset L$ is normal, it must be the splitting field of some $f(x) \in K[x]$.*

Proof. Recall from Lemma 3.1.39 that a finite extension is algebraic. Let $a_1, \ldots, a_n$ be a basis of L over K. Then each a_i is algebraic over K. Let $p_i \in K[x]$ be the minimal polynomial of a_i over K for $i = 1, \ldots, n$. Since the extension $K \subset L$ is normal and p_i has root $a_i \in L$, p_i splits in $L[x]$ for each i. Let $f = p_1 \cdots p_n$ and let R be the set of all roots of f in L. Then f splits in $L[x]$. As $\{a_1, \ldots, a_n\} \subset R$, we have

$$L = K(a_1, \ldots, a_n) \subset K(R) \subset L,$$

so $L = K(R)$. By definition, L is a splitting field of f over K. □

The converse of Lemma 3.2.30 is also true. To prove that, we first prove a lemma.

Lemma 3.2.31. *Let K be any field and $f, g \in K[x]$ both of positive degrees. If L is a field extension of K such that fg splits completely in $L[x]$, then both f and g split completely in $L[x]$.*

Proof. Without loss of generality, we assume that both f and g are monic. As fg splits completely in $L[x]$, there are $a_1, a_2, \ldots, a_n \in L$ such that

$$f(x)g(x) = (x - a_1)(x - a_2) \cdots (x - a_n) \in L[x].$$

Note that $L[x]$ is a UFD. Write f and g as products of irreducible polynomials in $L[x]$. Then by the uniquess of factorizations, every irreducible factor of f or g must be linear. Thus both f and g split completely in $L[x]$. □

Theorem 3.2.32. *A finite field extension $K \subset L$ is normal if and only if L is the splitting field of some polynomial $f(x) \in K[x]$ over K.*

Proof. We only need to prove one direction: assume that L is a splitting field of some $f(x) \in K[x]$ with $\deg(f) \geq 1$. We show that L is normal. Let $p \in K[x]$ be irreducible and assume that p has a root $a \in L$. We need to show that p splits completely in $L[x]$.

Let $g = fp \in K[x]$, and let M be a splitting field of g over K. By Lemma 3.2.31, both f and p split completely in $M[x]$. By Lemma 3.2.25, M is a field extension of L. For notational simplicity, we identify L with its image in M so we have field inclusions $K \subset L \subset M$. As p also splits over M, we only need to show that all the roots of p in M lie in L. Let b be a root of p in M. We need to show that $b \in L$.

Consider the sub-field $K(a)$ of L, and note that we have K-isomorphisms

$$\varphi_1 : \ K(a) \longrightarrow K[t]/\langle p \rangle \quad \text{and} \quad \varphi_2 : \ K(b) \longrightarrow K[t]/\langle p \rangle,$$

and thus the K-isomorphism

$$\varphi = i \circ \varphi_2^{-1} \circ \varphi_1 : \ K(a) \longrightarrow M,$$

where $i : K(b) \to M$ is the inclusion. Note now that L is a splitting field of f over $K(a)$. We can thus apply Lemma 3.2.25 to L being a splitting field of $K(a)$ and the extension $\varphi : K(a) \to M$ to conclude that there exists a $K(a)$-homomorphism extending $\varphi : K(a) \to M$, i.e., $\tilde{\varphi} : L \to M$ such that

$$\tilde{\varphi}|_{K(a)} = \varphi : K(a) \longrightarrow M.$$

Thus $\tilde{\varphi}(k) = \varphi(k) = k$ for all $k \in K$ and $\tilde{\varphi}(a) = b$. It follows from $\tilde{\varphi}(k) = k$ and Lemma 3.2.26 that $\tilde{\varphi}(L) = L$. Thus $b = \tilde{\varphi}(a) \in L$. □

A second proof that a splitting field is normal. The proof of Theorem 3.2.32 above uses the crucial property of splitting field as stated in the Extension Lemma. We now give a proof which uses more directly the definition of splitting field. A reference for this proof is David Cox' book "Galois Theory", second edition Proposition 5.2.1.

Assume that L is a splitting field of some $f(x) \in K[x]$ with $\deg(f) \geq 1$. Let $p \in K[x]$ be irreducible and assume that p has a root $\beta \in L$. We need to show that p splits completely in $L[x]$. Write

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n),$$

where $\alpha_1, \ldots, \alpha_n \in L$ but not necessarily pairwise distinct. As $L = K[\alpha_1, \ldots, \alpha_n]$, there exists $h \in K[x_1, \ldots, x_n]$ such that $\beta = h(\alpha_1, \ldots, \alpha_n)$. Consider

$$s(x) = \prod_{\sigma \in S_n} (x - h(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(n)}) \in L[x].$$

Then $s(x)$ completely splits in $L[x]$ and has β as a root. The coefficients of $s(x)$ are all symmetric expressions in $\alpha_1, \ldots, \alpha_n$ with coefficients in K. By the fundamental

theorem on symmetric functions, the coefficients of $s(x)$ are then polynomials with coefficients in K in the elementary symmetric expressions of $\alpha_1, \ldots, \alpha_n$, which are, up to signs, the coefficients of f. Thus $s(x) \in K[x]$. As $p(x)$ is the minimal polynomial of β in $K[x]$ and as $s(\beta) = 0$, we have $p|s$ in $K[x]$. Since $s(x)$ completely spits over L, so does p. This shows that L is a normal extension of K.

§ 6. Finite Fields

Back to finite fields. We first state a distinctive feature of fields with positive characteristics.

Lemma 3.2.33. *For a prime number p and integer $1 \leq k < p$, the binary coefficient $\begin{pmatrix} p \\ k \end{pmatrix}$ is divisible by p. Consequently, if K is a field of characteristic p, then*

$$(a+b)^p = a^p + b^p, \qquad a, b \in K.$$

Proof. By definition, $\begin{pmatrix} p \\ k \end{pmatrix} k! = p(p-1)\cdots(p-k+1)$ is divisible by p, but p does not divide $k!$, so p must divide $\begin{pmatrix} p \\ k \end{pmatrix}$. □

If L is a field of characteristic $p > 0$, it then follows from Lemma 3.2.33 that the map $\sigma : L \to L$ defined by $\sigma(a) = a^p$ for all $a \in L$ is an injective ring homomorphism. If L is finite, then σ must also be surjective and thus an isomorphism from L to itself.

Definition 3.2.34. For a field L of characteristic $p > 0$, the injective ring homomorphism $\sigma : L \to L, \sigma(a) = a^p$ is called the *Frobenious homomorphism of L*.

It is also clear that the Frobenious homomorphism of a field L with characteristic $p > 0$ is surjective if and only if every element in L has a p'th root,

Example 3.2.35. It is easy to check that the Frobenious morphism on $L = \mathbb{F}_p(x)$ is not surjective. Indeed, $x \in \mathbb{F}_p(x)$ is not in the image of the Frobenious morphism.

We now restate the following fact in Lemma 3.1.20 which have been using over and over again.

Lemma 3.2.36. *If K is a field and if $f(x)$ is an irreducible polynomial over K of degree n, then $L = K[x]/\langle f(x)\rangle$ is a field extension of K with $[L : K] = n$. Moreover, if a is the image of x in L, then $L = K(a)$.*

Proof. Let $\phi : K[x] \to L = K[x]/\langle f(x)\rangle$ be the projection map. Then a basis of L over K is given by $\phi(1), \phi(x), \cdots, \phi(x^{n-1})$, so $[L : K] = n$. It is clear that $L = K(a)$. □

Given a prime number p and an integer n, Lemma 3.2.36 shows that if we can find an irreducible polynomial f of order n over $\mathbb{F}_p$, the field $\mathbb{F}_p[x]/\langle f(x)\rangle$ will have order p^n, and this will show the existence of such a field.

For small p and n, we find irreducible polynomials over $\mathbb{F}_p$ of degree n by hand.

Example 3.2.37. Let's try to find a field K of order $4 = 2^2$. We need to find a quadratic irreducible polynomial over $\mathbb{F}_2$. The 4 quadratic polynomials over $\mathbb{F}_2$ are

$$x^2, \quad x^2+1, \quad x^2+x, \quad x^2+x+1.$$

Out of these four, only x^2+x+1 is irreducible. Let $K = \mathbb{F}_2[x]/\langle x^2+x+1\rangle$ and denote by a the image of x in K. Then

$$K = \{0, 1, a, a+1\}$$

with $a^2 = a+1$. Let's now try to find a field of order $8 = 2^3$. Easy to see that there are only two irreducible cubic polynomials over $\mathbb{F}_2$ which are

$$f(x) = x^3+x+1, \quad g(x) = x^3+x^2+1.$$

Let $\mathbb{F}_8 = \mathbb{F}_2[x]/\langle f(x)\rangle$ and $\mathbb{F}'_8 = \mathbb{F}_2[x]/\langle g(x)\rangle$. Let a and b denote the images of x in $\mathbb{F}_8$ and $\mathbb{F}'_8$ respectively. Then

$$\mathbb{F}_8 = \{0, 1, a,\ a^2,\ a^3 = a+1,\ a^4 = a^2+a,\ a^5 = a^2+a+1,\ a^6 = a^2+1\},$$
$$\mathbb{F}'_8 = \{0, 1, b,\ b^2,\ b^3 = b^2+1,\ b^4 = b^2+b+1,\ b^5 = b+1,\ b^6 = b^2+b\}.$$

Note that $\mathbb{F}_8\backslash\{0\}$ and $\mathbb{F}'_8\backslash\{0\}$ are cyclic groups of order 7, so any elements in them is a generator. We see from above how a and b are generators. To construct a field of order 64, consider $f(x) = x^2+ax+1 \in \mathbb{F}_8[x]$. We have

$$f(0) = 1, \quad f(1) = a, \quad f(a) = 1, \quad f(a^2) = a^2, \quad f(a^3) = a,$$
$$f(a^4) = a^2, \quad f(a^5) = a^2+a+1, \quad f(a^6) = a^2.$$

Thus f has no roots in $\mathbb{F}_8$. Thus f is irreducible. Let

$$\mathbb{F}_{64} = \mathbb{F}_8[x]/\langle x^2+ax+1\rangle = \{\lambda_0+\lambda_1 c :\ \lambda_0, \lambda_1 \in \mathbb{F}_8,\ c^2 = ac+1\}.$$

Thus we have constructed a field $\mathbb{F}_{64}$ with 64 elements as $\mathbb{F}_{64} = \mathbb{F}_2(a, c)$.

We now show the existence and uniqueness of a finite field of order p^n for any prime number p and integer $n \geq 1$.

Lemma 3.2.38. *If K is a finite field of order q, then every element $a \in K$ satisfies*

$$x^q - x = 0.$$

Proof. The set $K - \{0\}$ under multiplication is an abelian group of order $q-1$, so $a^{q-1} = 1$ for every $a \in K, a \neq 0$. So $a^q = a$ for every $a \in K$. □

Theorem 3.2.39. *Let p be a prime number and let $n \geq 1$ be an integer. A finite field L has order p^n if and only if it is isomorphic to the splitting field of the polynomial $f(x) = x^{p^n} - x$ over $\mathbb{F}_p$.*

Proof. Assume first that L is a field of order p^n. Then the prime sub-field of L is $\mathbb{F}_p$, so we can regard L as an extension of $\mathbb{F}_p$. The polynomial $f(x) = x^{p^n} - x$ can have at most p^n roots in L. But by Lemma 3.2.38, every element in L is a root for f. Thus f has precisely p^n roots in L. Hence f splits in L. Again since every element in L is a root of f, L is a splitting field of f.

Conversely, let L be a splitting field of $f(x) = x^{p^n} - x$, and let R be the set of all the roots of f in L. Since the derivative of f is -1 which has no roots, all the roots of f in L are distinct, so R has exactly p^n elements. On the other hand, by Lemma 3.2.33, if $a, b \in R$, then

$$(a-b)^{p^n} = a^{p^n} + (-b)^{p^n} = a - b, \quad \text{and} \quad (ab^{-1})^{p^n} = ab^{-1} \text{ if } b \neq 0.$$

Thus R is a sub-field of L. Moreover, since $a^p = a$ for every $a \in \mathbb{F}_p \subset L$, R contains $\mathbb{F}_p$. Since L is the smallest sub-field of L containing $\mathbb{F}_p$ and R, we must have $R = L$, and hence L has p^n elements. □

By Theorem 3.2.39 and by the uniqueness of splitting fields, one has

Corollary 3.2.40. *For any prime number p and integer $n \geq 1$, any two fields of order p^n are isomorphic.*

Lemma 3.2.41. *For any prime number p and positive integers $d, n \geq 1$ such that $d|n$, then*

$$(x^{p^d} - x)|(x^{p^n} - x).$$

Proof. We will use the identity

$$z^{st} = 1 = (z^s - 1)((z^s)^{t-1} + \cdots + z^s + 1), \quad s, t \in \mathbb{Z}_{\geq 1}. \tag{3.5}$$

Write $n = dc$. By (3.5), $(p^d - 1)|(p^n - 1)$. By (3.5) again,

$$(x^{p^d-1} - 1)|(x^{p^n-1} - 1).$$

It follows that $(x^{p^d} - x)|(x^{p^n} - x)$. □

Corollary 3.2.42. *Let p be a prime number and $n \in \mathbb{Z}_{\geq 1}$. Then for each $d \in \mathbb{Z}_{\geq 1}$ such that $d|n$, there is one and exactly one sub-field of $\mathbb{F}_{p^n}$ with p^d elements. These are all sub-fields of $\mathbb{F}_{p^n}$.*

Proof. If F is any sub-field of $\mathbb{F}_{p^n}$, by the Tower Theorem, the degree d of F over $\mathbb{F}_p$ must divide n, and F has p^d elements. Let $d \in \mathbb{Z}_{\geq 1}$ such that $d|n$. It remains to show that there is one and only one sub-field of $\mathbb{F}_{p^n}$ with p^d elements.

Let $f_n = x^{p^n} - x$ and $f_d = x^{p^d} - x$. By Theorem 3.2.39, $\mathbb{F}_{p^n}$ is the splitting field of f_n over $\mathbb{F}_p$ and that all the roots of f_n in $\mathbb{F}_{p^n}$ are distinct. By Lemma 3.2.41, $f_d|f_n$, and the same proof of Theorem 3.2.39 shows that the set $R(f_d)$ of all roots of f_d in $\mathbb{F}_{p^n}$ is a sub-field of $\mathbb{F}_{p^n}$ with p^d elements. Any sub-field of $\mathbb{F}_{p^n}$ with p^d elements must lie in $R(f_d)$ so must be the same as $R(f_d)$. Thus there is one and only one sub-field of $\mathbb{F}_{p^n}$ with p^d elements. □

We now prove another important fact about finite fields.

Proposition 3.2.43. *Let K be any field. Then any finite subgroup G of the multiplicative group $K^\star = K\backslash\{0\}$ is cyclic. In particular, $K^\star$ is cyclic if K is finite.*

Proof. Let $|G| = p_1^{n_1} p_2^{n_2} \cdots p_l^{n_l}$ be the prime number decomposition of $|G|$. By the Fundamental Theorem on Finite Abelian Groups, we have

$$G = G(p_1) \times G(p_2) \times \cdots \times G(p_l),$$

where for each i, $G(p_i)$ is the product of cyclic groups whose orders are powers of p_i. We only need to show that each $G(p_j)$ is cyclic, for then G would be cyclic because the p_j's are relatively prime. Assume that there exists j such that $G(p_j)$ is not cyclic. Then there exists $r < n_j$ such that every $a \in G(p_j)$ satisfies $a^{p_j^r} = 1$. But $G(p_j)$ has $p_j^{n_j}$ elements, and the equation

$$x^{p_j^r} = 1$$

can have at most p^r solutions, so this is a contradiction. Hence every $G(p_j)$ is cyclic and that G is cyclic. □

Example 3.2.44. Consider $\mathbb{F}_5^\star = \{\bar{1}, \bar{2}, \bar{3}, \bar{4}\}$. Then $\bar{2}^2 = \bar{4}, \bar{2}^3 = \bar{3}, \bar{2}^4 = \bar{1}$, so $\bar{2}$ is a generator for $\mathbb{F}_5^\star$. Similarly, $\bar{3}$ is also a generator, but since $\bar{4}^2 = \bar{1}$, so $\bar{4}$ is not a generator. Similarly, $\bar{3}$ and $\bar{5}$ are generators for $\mathbb{F}_7^\star$ while $\bar{2}, \bar{4}, \bar{6}$ are not.

Recall that a field extension $K \subset L$ is said to be simple if there exists $a \in L$ such that $L = K(a)$.

Corollary 3.2.45. *Any finite extension of a finite field is simple.*

Proof. Let L be a finite extension of a finite field K. By Proposition 3.2.43, there exists $a \in L\backslash\{0\}$ such that every $b \in L\backslash\{0\}$ is a power of a. Hence $L = K(a)$. □

We now have the very beautiful fact on irreducible polynomials over $\mathbb{F}_p$.

Corollary 3.2.46. *For any prime number p and positive integer $n \geq 1$, irreducible polynomials over $\mathbb{F}_p$ of degree n exist. Moreover, the polynomial*

$$f(x) = x^{p^n} - x$$

is exactly the product of all monic irreducible polynomials in $\mathbb{F}_p[x]$ whose degrees divide n.

Proof. Let $\mathbb{F}_{p^n}$ be a field of order p^n. By Corollary 3.2.45, $\mathbb{F}_{p^n} = \mathbb{F}_p(a)$ for some $a \in \mathbb{F}_{p^n}$. Let $q(x)$ be the minimal polynomial of a over $\mathbb{F}_p$. Then $q(x)$ is irreducible over $\mathbb{F}_p$ and has degree n. Furthermore, if $q(x)$ is an arbitrary irreducible monic polynomial of degree n, the field $L = \mathbb{F}_p[x]/\langle q \rangle$ is a field extension of $\mathbb{F}_p$ with p^n elements, so the element $a = \bar{x} \in L$ has q as its minimal polynomial and satisfies $f(a) = 0$, so $q|f$.

Let now $f = q_1^{k_1} q_2^{k_2} \cdots q_l^{k_l}$ be the decomposition of $f \in \mathbb{F}_p[x]$ into irreducibles, where the q_j's are pairwise distinct and monic. Then the set roots of f in $\mathbb{F}_{p^n}$ is the union of the sets of roots of the q_j's in $\mathbb{F}_{p^n}$. Since the set of roots of f in $\mathbb{F}_{p^n}$ is exactly the whole $\mathbb{F}_{p^n}$, we see that $k_1 = \cdots = k_l = 1$, and each q_j, for $1 \leq j \leq l$, splits completely in $\mathbb{F}_{p^n}$. Let $1 \leq j \leq l$, and let $a \in \mathbb{F}_{p^n}$ be a root of q_j. Then q_j is the minimal polynomial of a over $\mathbb{F}_p$, so $\mathbb{F}_p(a)$ is a sub-field of $\mathbb{F}_{p^n}$, thus the degree of q_j divides n. Furthermore, suppose that $d|n$. By the discussion above, every irreducible monic polynomial q of degree d divides $x^{p^d} - x$ so by Lemma 3.2.41, q divides $x^{p^n} - x$. □

Example 3.2.47. In $\mathbb{F}_2[x]$, one has

$$\begin{aligned}
&x^2 - x = x(x-1),\\
&x^4 - x = x(x-1)(x^2+x+1),\\
&x^8 - x = x(x-1)(x^3+x+1)(x^3+x^2+1),\\
&x^{16} - x = x(x-1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1).
\end{aligned}$$

§ 7. Separable polynomials and perfect fields

Given a field extension $K \subset L$ and $f \in K[x]$, recall that a root $a \in L$ of f in L is a *repeated root* of f if $f(x) = (x-a)^2 g(x) \in L[x]$ for some $g(x) \in L[x]$.

Definition 3.2.48. For a field K, a polynomial $f(x) \in K[x]$ is said to be separable over K if it has no repeated roots in its splitting field over K.

Example 3.2.49. The polynomial $f(x) = x^3 - 2 \in \mathbb{Q}[x]$ is separable over $\mathbb{Q}$, but when regarded as a polynomial over $\mathbb{F}_3$ it is not separable because $2^3 = 2$ in $\mathbb{F}_3$, so $f(x) = (x-2)^3 \in \mathbb{F}_3[x]$.

Recall that if $f = c_0 + c_1 x + \cdots + c^n x^n \in K[x]$, the derivative of f is

$$f'(x) = c_1 + 2c_2 x + \cdots + nc_n x^{n-1} \in K[x].$$

In particular, if $a \in L$ is a repeated root of f in L, then $f(a) = f'(a) = 0$, and both $f(x)$ and $f'(x)$, as elements in $L[x]$, are divisible by $x - a$.

Lemma 3.2.50. *Let K be any field and let $f \in K[x]$ with positive degree. Then the following are equivalent:*

1) f is separable over K;

2) f and f' are relatively prime as elements in $K[x]$;

3) f has no repeated roots in any field extension L of K

Proof. Assuming 1) and assuming that f is monic and $\deg(f) = n > 0$, we now prove 2). Let L be a splitting field of f over K, and let $a_1, \ldots, a_n \in L$ be the n pairwise distinct roots of f in L. Then

$$f(x) = (x-a_1)(x-a_2)\cdots(x-a_n) \quad \text{and} \quad f'(x) = \sum_{i=1}^{n} (x-a_1)\cdots\widehat{(x-a_i)}\cdots(x-a_n).$$

If $g(x)$ is any divisor of $f(x)$ in $L[x]$ of positive degree and if $f(x) = g(x)h(x)$ where $h(x) \in K[x]$, then since $h(x)$ can not have more than n roots in L, there exists $1 \le i \le n$ such that $g(a_i) = 0$. But as $f'(a_i) \neq 0$, $g(x)$ does not divide $f'(x)$ in $L[x]$. This shows that f and f' are relatively prime in $L[x]$. An easy argument (see Exercise 3.2.51 below) shows that the greatest common divisor of any two polynomials in $K[x]$ is also their greatest common divisor over any field extension of K. Thus f and f' are relatively prime over K. This shows that 1) implies 2).

Assume now that 2) holds. Then there exist $p(x), q(x) \in K[x]$ such that

$$1 = p(x)f(x) + q(x)f'(x) \in K[x],$$

which, when viewed as an identify in $L[x]$ for any field extension $K \subset L$, implies that f and f' have no common roots in L. This shows that 2) implies 3). Clearly 3) implies 1). □

Exercise 3.2.51. Let $K \subset L$ be a field extension and let $f(x), g(x) \in K[x]$. Show that the greatest common divisor (with leading coefficient 1) of $f(x)$ and $g(x)$ in $L[x]$ is the same as the greatest common divisor of $f(x)$ and $g(x)$ in $K[x]$.

Lemma 3.2.52. *Let K be any field and let $p \in K[x]$ be irreducible. Then p is separable over K if and only if $p' \neq 0$.*

Proof. When $p \in K[x]$ is irreducible, $\deg(p) > 0$ by definition, and since p is irreducible, p and p' are relatively prime in $K[x]$ unless $p' = 0$. Lemma 3.2.52 now follows from Lemma 3.2.50. □

Example 3.2.53. Let $K = \mathbb{F}_2(t)$ be the field of rational functions in t with coefficients from $\mathbb{F}_2$. It is easy to prove that $f(x) = x^2 - t$ is irreducible, but f is not separable. Indeed, $f' = 2x = 0$. One can also see this in another way: let L be a splitting field of f over K, and let $\sqrt{t}$ be a root of f in L. Then

$$f(x) = x^2 - t = (x - \sqrt{t})^2.$$

We now turn to the notion of perfect fields.

Definition 3.2.54. A field K is said to be perfect if every irreducible polynomial in $K[x]$ is separable, i.e., if $p'(x) \neq 0$ whenever $p \in K[x]$ is irreducible.

Lemma 3.2.55. *A field K is perfect if and only if either it has characteristic 0 or if it has characteristic $p > 0$ but the Frobenious homomorphism $\sigma : K \to K, \sigma(a) = a^p$ is an isomorphism, or, equivalently, if every element $a \in K$ has a p'th root.*

Proof. An irreducible polynomial $f(x) \in K[x]$ has positive degree by definition. So if K has characteristic 0, it is clear that $f' \neq 0$ if f is irreducible.

Assume that K has characteristic $p > 0$. Suppose first that the Frobenious homomorphism σ is surjective, i.e., $K^p = K$. Let $f \in K[x]$ be irreducible. If $f'(x) = 0$, then it is easy to see that f is of the form $f(x) = \sum_{i=0}^{n} c_i x^{ip}$, where $c_i \in K$. Let $c_i = a_i^p$, where $a_i \in K$ for each i. Then $f(x) = (\sum_i a_i x^i)^p$, contradicting the assumption that f is irreducible. Thus $f' \neq 0$. This shows that if $K^p = K$, then K is perfect.

Assuming now that K is perfect, we need to show that $K^p = K$. Suppose not. Then there exists $a \in K$ such that $a \neq b^p$ for any $b \in K$. Consider the polynomial $f(x) = x^p - a \in K[x]$. Clearly $f'(x) = 0$. If we can show that $f(x)$ is irreducible over K, then we would get a contradiction to the assumption that K is perfect. Suppose that $f(x)$ is reducible over K. Then there exist $g(x), h(x) \in K[x]$, both monic and of positive degree, such that $f(x) = g(x)h(x)$. Now the splitting field L of $f(x)$ must contain at least one root b of f. It then follows from $b^p = a$ that

$$f(x) = x^p - b^p = (x - b)^p \in L[x].$$

The identify $f = gh$ in $L[x]$ then implies that $g(x) = (x-b)^m$ for some $1 \leq m \leq p-1$. Since $g(x) \in K[x]$, the coefficient $-mb$ of x^{m-1} in $(x-b)^m$ must be in K. As $m \in K$ is invertible, it follows that $b \in K$, contradicting the assumption that $a \notin K^p$. This finishes the proof that if K is perfect, one must then have $K^p = K$. □

As a direct consequence of Lemma 3.2.55, we have

Proposition 3.2.56. *Any field with characteristic 0 is perfect, and any finite field is perfect.*

§ 8. Separable extensions and the Primitive Element Theorem

Definition 3.2.57. An algebraic extension $K \subset L$ is said to be *separable* if for every $a \in L$, the minimal polynomial of a over K is separable over K.

The following is thus a consequence of the definitions.

Lemma 3.2.58. *Every algebraic extension of a perfect field is separable.*

The following Theorem 3.2.59 is referred to as *The Primitive Element Theorem*, or the *Theorem of Primitive Elements*.

Theorem 3.2.59. *A finite separable extension $K \subset L$ must be simple.*

Proof. If K is finite, recall that Corollary 3.2.45 says that every finite extension of K is simple, so we are done. We thus assume that K is infinite.

We first assume that $L = K(\alpha, \beta)$ and we show that there exists $\gamma \in L$ such that $L = K(\gamma)$. In fact, we show that there are infinitely many $\lambda \in K$ such that $L = K(\alpha + \lambda\beta)$.

Let $p(x), q(x) \in K[x]$ be respectively the minimal polynomials of α and β over K, and let L' be an extension of L such that both $p(x)$ and $q(x)$ completely split over L' (for example, let L_1 be a splitting field of $p(x)$ over L, and regarding $q(x)$ as in $L_1[x]$, let L' be a splitting field of $g(x) \in L_1[x]$ over L_1. Then $L \subset L'$ and both $p(x)$ and $q(x)$ completely split over L'.) Let $n = \deg(q(x))$. If $n = 1$, we have $\beta \in K$ and $L = K(\alpha)$, so there is nothing to prove, so we assume that $n \geq 2$. By the separability assumption and by Lemma 3.2.50, $q(x)$ has no repeated roots in L'. Thus there exists at least one $\beta' \in L'$ such that $q(\beta') = 0$ and $\beta' \neq \beta$. Let

$$S = \{(\alpha' - \alpha)/(\beta - \beta') : \alpha', \beta' \in L', \beta' \neq \beta, p(\alpha') = 0,\, q(\beta') = 0\} \subset L'.$$

Since K is infinite, there exists $\lambda \in K$ such that $\lambda \notin S$. Choose $\lambda \in K$ and $\lambda \notin S$, let $\gamma = \alpha + \lambda\beta \in L$, and let $r_\lambda(x) \in K(\gamma)[x]$ be the minimal polynomial of β over $K(\gamma)$. Our goal is to prove that $\deg(r_\lambda(x)) = 1$, which would imply that $\beta \in K(\gamma)$ and thus also $\alpha \in K(\gamma)$.

As $\alpha = \gamma - \lambda\beta$, one has $p(\gamma - \lambda\beta) = 0$. Let

$$h(x) = p(\gamma - \lambda x) \in K(\gamma)[x].$$

Then $\beta \in L$ is a root of $h(x)$. The minimal polynomial $r_\lambda(x)$ of β over $K(\gamma)$ is thus a common divisor of both $q(x) \in K[x] \subset K(\gamma)[x]$ and $h(x) \in K(\gamma)[x]$. Since $q(x)$ completely splits over L' and has no repeated roots in L', $r_\lambda(x)$ completely splits over L' and has no repeated roots in L'. If $\deg(r_\lambda(x)) \geq 2$, we then have a root β' of $r_\lambda(x)$ in L' such that $\beta' \neq \beta$. Then β' is a common root for both $h(x)$ and $q(x)$, i.e., $p(\gamma - \lambda\beta') = 0$ and $q(\beta') = 0$. Let $\alpha' = \gamma - \lambda\beta' = \alpha + \lambda(\beta - \beta')$. We then get

$$\lambda = \frac{\alpha' - \alpha}{\beta - \beta'} \in S,$$

contradicting the assumption that $\lambda \notin S$. Thus $\deg(r_\lambda(x)) = 1$, which implies that $\beta \in K(\gamma)$ and $\alpha = \gamma - \lambda\beta \in K(\gamma)$. Thus $L = K(\alpha, \beta) = K(\gamma)$.

For an arbitrary finite separable extension $K \subset L$, write $L = K(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Using what we have just proved, we see that there exists $\gamma_1 \in L$ such that

$$L = K(\alpha_1, \alpha_2)(\alpha_3, \ldots, \alpha_n) = K(\gamma_1)(\alpha_3, \ldots, \alpha_n) = K(\gamma_1, \alpha_3, \ldots, \alpha_n),$$

so by induction $L = K(\gamma)$ for some $\gamma \in L$. □

Corollary 3.2.60. *If K is a perfect field, then very finite extension of K is simple.*

Corollary 3.2.61. *If K has characteristic 0 or is a finite field, then very finite extension of K is simple.*

The following example, taken from Cox's book *Galois Theory*, Second Edition, Example 5.4.4, says that the Primitive Element Theorem does not hold if we take away the separable condition on the finite extension.

Example 3.2.62. Let F be a field with characteristic $p > 0$, let $K = F(t, u)$, and consider the splitting field L over K of

$$f(x) = (x^p - t)(x^p - u) \in K[x].$$

We first make a simple observation on the extension $K \subset L$.

Claim. Suppose that $\gamma \in L$ is such that $\gamma \notin K$ but $\gamma^p = a \in K$. Then the minimal polynomial of γ in $K[x]$ is $x^p - a \in K[x]$.

Proof of Claim. We need to prove that $x^p - a \in K[x]$ is irreducible. Suppose not. Then there exist $g(x), h(x) \in K[x]$, both with positive degrees, such that $x^p - a = g(x)h(x) \in K[x]$. As elements in $L[x]$, we then have $(x - \gamma)^p = g(x)h(x)$, so $g(x) = (x - \gamma)^n$ for some $0 < n < p$ as an element in $L[x]$. Since the coefficient of $(x - \gamma)^n$ of x^{n-1} is $-n\gamma$ which is not in K by assumption, we get a contradiction to the fact that $g(x) \in K[x]$. Thus $x^p - a \in K[x]$ is irreducible.

Let now $\alpha \in L$ be a root of $x^p = t$ and $\beta \in L$ be a root of $x^p - u$. One then has

$$f(x) = (x^p - \alpha^p)(x^p - \beta^p) = (x - \alpha)^p(x - \beta)^p \in L[x].$$

Thus α, β are the only roots of f in L, so $L = K(\alpha, \beta)$. By the claim, $x^p - t \in K[x]$ is irreducible. Hence $|K(\alpha) : K| = p$. As $\beta \notin K(\alpha)$, a fact that be proved easily, we see that $x^p - u$ is irreducible over $K(\alpha)$, so $|L : K(\alpha)| = p$. We thus conclude that

$$|L : K| = p^2.$$

We now look at $K(\gamma) \subset L$ for arbitrary $\gamma \in L$. As $L = K(\alpha, \beta)$, there exist $a_{ij} \in K$ such that $\gamma = \sum_{ij} a_{ij}\alpha^i\beta_j$, from which one gets

$$\gamma^p = \left(\sum_{ij} a_{ij}\alpha^i\beta_j\right)^p = \sum_{ij} a_{ij}^p\alpha^{ip}\beta^{jp} = \sum_{ij} a_{ij}^p t^i u^j \in K[x].$$

In other words, γ is a root of $x^p - \gamma^p \in K[x]$. If $\gamma \in K$, then $K(\gamma) = K$. Suppose that $\gamma \notin K$. By the claim,

$$x^p - \gamma^p = (x - \gamma)^p \in K[x]$$

is the irreducible polynomial of γ over K, so $|K(\gamma) : K| = p$ and thus $K(\gamma) \neq L$. This in particular implies that the extension $K \subset L$, although finite, has no primitive elements. The extension $K \subset L$ is *purely inseparable* in the sense that the irreducible polynomial of every $\gamma \in L \backslash K$ is inseparable (i.e., not separable).

Chapter 4 | Introduction to Galois theory

4.1 Basic concepts and the fundamental theorem of Galois theory

§ 1. Automorphism groups and roots of polynomials

Definition 4.1.1. 1) If L is a field, a ring isomorphism from L to itself is also called an *automorphism* of L. The set of all automorphisms of L is denoted by $\mathrm{Aut}(L)$.

2) If $K \subset L$ is a field extension, an element $\sigma \in \mathrm{Aut}(L)$ such that $\sigma(k) = k$ for all $k \in K$ is called a *K-automorphism* of L, and the set of all K-automorphisms of L is denoted by $\mathrm{Aut}_K(L)$.

Lemma 4.1.2. *For any field L, $\mathrm{Aut}(L)$ is a group, and for any field extension $K \subset L$, $\mathrm{Aut}_K(L)$ is a subgroup of $\mathrm{Aut}(L)$.*

Proof. The proof is straightforward using the definition of groups and subgroups. □

Definition 4.1.3. For a field extension $K \subset L$, the group $\mathrm{Aut}_K(L)$, sometimes also denoted as

$$\mathrm{Aut}_K(L) = \mathrm{Gal}_K(L) = \mathrm{Gal}(L/K) = \mathrm{Aut}(L/K),$$

is called the *Galois group of the extension $K \subset L$*. If $f(x) \in K[x]$, and if L is a splitting field of f over K, the group $\mathrm{Aut}_K(L)$ is called *the Galois group of $f(x)$ over K* and is denoted by $\mathrm{Gal}_K(f)$ or $\mathrm{Gal}(f/K)$.

Example 4.1.4. Consider $\mathbb{C}$ as a field extension over $\mathbb{R}$. Any $\sigma \in \mathrm{Aut}_{\mathbb{R}}(\mathbb{C})$ must be given by $\sigma(a + bi) = a + b\sigma(i)$ for $a, b \in \mathbb{R}$, so σ is determined by $\sigma(i) \in \mathbb{C}$. But the element $\sigma(i) \in \mathbb{C}$ must satisfy $\sigma(i)^2 = \sigma(i^2) = \sigma(-1) = -1$, so $\sigma(i) = i$ or $\sigma(i) = -i$. In the first case one has $\sigma = \mathrm{Id}_{\mathbb{C}}$, and in the second case σ is the complex conjugation $\sigma(a + bi) = a - bi$ for $a, b \in \mathbb{R}$. Thus $\mathrm{Aut}_{\mathbb{R}}(\mathbb{C}) \cong \mathbb{Z}_2$.

Example 4.1.5. We want to determine $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{R})$. We claim that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{R})$ is the trivial group. To see this, let $\sigma \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{R})$. Then for any $a \in \mathbb{R}$, we have $\sigma(a^2) = (\sigma(a))^2$, so $\sigma(a) > 0$ if $a > 0$. It follows that for all $a, b \in \mathbb{R}$, if $a < b$, the $\sigma(a) < \sigma(b)$. As σ fixes every rational number, if $a, b \in \mathbb{R}$ and $n \in \mathbb{Z}_{>0}$ are such that $-\frac{1}{n} < a - b < \frac{1}{n}$, then $-\frac{1}{n} < \sigma(a) - \sigma(b) < \frac{1}{n}$. We thus conclude that $\sigma : \mathbb{R} \to \mathbb{R}$

must be continuous. As σ is the identity map on $\mathbb{Q}$ which is dense in $\mathbb{R}$, σ must be the identity map of $\mathbb{R}$.

We now prove some basic lemmas on automorphism groups in relation to roots of polynomials which are fundamental for Galois theory.

Lemma 4.1.6. *Let $K \subset L$ be a field extension and let $\sigma \in \mathrm{Aut}_K(L)$. Then for any $f \in K[x]$, if $a \in L$ is a root of f in L, then $\sigma(a)$ is also a root of f in L.*

Proof. Write $f = b_0 + b_1 x + \cdots b_n x^n$ with $b_0, b_1, \ldots, b_n \in K$ and $b_n \neq 0$. Then

$$b_0 + b_1 a + \cdots + b_n a^n = 0.$$

Applying σ to both sides, one has

$$0 = \sigma(b_0 + b_1 a + \cdots + b_n a^n) = b_0\sigma(a) + b_1(\sigma(a))^2 + \cdots + b_n(\sigma(a))^n,$$

so $f(\sigma(a)) = 0$. □

In the following, for any finite set R, let $\mathrm{Perm}(R)$, called the *permutation group of R*, be the group of all bijections from R to R with the group operation being the composition.

Lemma 4.1.7. *(***Key lemma on automorphisms of field extensions and roots of polynomials***) Let $K \subset L$ be a field extension and let $f \in K[x]$. Let R_f be the set of all roots of f in L. Then we have a group homomorphism*

$$\Phi: \quad \mathrm{Aut}_K(L) \longrightarrow \mathrm{Perm}(R_f), \quad \Phi(\sigma)(a) = \sigma(a), \quad a \in R_f.$$

If $L = K(R_f)$ (for example, if L is a splitting field of f), then Φ is injective, and thus $|\mathrm{Aut}_K(L)| \leq n!$, where $n = \deg(f)$.

Proof. By Lemma 4.1.6, we have $\sigma(R_f) \subset R_f$ for every $\sigma \in \mathrm{Aut}_K(L)$. Let $\mathrm{Map}(R_f, R_f)$ be the set of all maps from R_f to R_f. We then have a map

$$\Phi: \quad \mathrm{Aut}_K(L) \longrightarrow \mathrm{Map}(R_f, R_f), \quad \Phi(\sigma)(a) = \sigma(a), \quad a \in R_f.$$

It is clear that $\Phi(\sigma_1\sigma_2) = \Phi(\sigma_1)\Phi(\sigma_2)$ for $\sigma_1, \sigma_2 \in \mathrm{Aut}_K(L)$. As $\Phi(\mathrm{Id}_L) = \mathrm{Id}_{R_f}$, for every $\sigma \in \mathrm{Aut}_K(L)$,

$$\Phi(\sigma)\Phi(\sigma^{-1}) = \Phi(\mathrm{Id}_L) = \mathrm{Id}_{R_f},$$

so $\Phi(\sigma) \in \mathrm{Map}(R_f, R_f)$ is a bijection, i.e., $\Phi(\sigma) \in \mathrm{Perm}(R_f)$. Hence we have a group homomorphism

$$\Phi: \quad \mathrm{Aut}_K(L) \longrightarrow \mathrm{Perm}(R_f), \quad \Phi(\sigma) = (\sigma|_{R_f} : R_f \to R_f).$$

Assume that $L = K(R_f)$. If $\sigma \in \mathrm{Aut}_K(L)$ is such that $\sigma|_{R_f} = \mathrm{Id}_{R_f}$, then $\sigma = \mathrm{Id}_L$, so the kernel of Φ is trivial, and thus Φ is injective. □

We now give some other consequences of Lemma 4.1.7.

Lemma 4.1.8. *If $K \subset L$ is a finite extension, then $\mathrm{Aut}_K(L)$ is a finite group.*

Proof. Let $\{a_1, \ldots, a_n\}$ be a basis of L over K), and for $i = 1, \ldots, n$, let $p_i(x) \in K[x]$ be the minimal polynomial of a_i over K. Let

$$f(x) = p_1(x)p_2(x)\cdots p_n(x) \in K[x],$$

and let R_f be the set of all roots of f in L. Since $\{a_1, a_2, \ldots, a_n\} \subset R_f$, we have $L = K(R_f)$. By Lemma 4.1.6, $\mathrm{Aut}_K(L)$ is in bijection with a subgroup of $\mathrm{Perm}(R_f)$, so $\mathrm{Aut}_K(L)$ is finite. □

So far we have been looking at how automorphisms of a field extension $K \subset L$ map roots of any $f \in K[x]$ in L to roots of the same f. We now how look at how we use roots of polynomials to construct automorphisms of field extensions. We first look at finite simple extensions.

Lemma 4.1.9. *(***Basic lemma on automorphism groups of finite simple extensions***) Let $L = K(a)$ be a simple finite extension of K, let $p(x) \in K[x]$ be the minimal polynomial of a over K, and let R_p be the set of all the roots of $p(x)$ in L. Then*

$$\phi_a : \ \mathrm{Aut}_K(L) \longrightarrow R_p, \ \ \sigma \longmapsto \sigma(a) \in R_p$$

is a bijection. Consequently,

1) $|\mathrm{Aut}_K(L)| = |R_p| \leq \deg(p(x)) = [L : K]$;

2) if p completely splits over L and has no repeated roots in L, then $|\mathrm{Aut}_K(L)| = |R_p| = \deg(p(x)) = [L : K]$.

Proof. If $\sigma, \sigma' \in \mathrm{Aut}_K(L)$ are such that $\sigma(a) = \sigma'(a)$, then, as L is generated by a over K, we have $\sigma = \sigma'$. Thus $\phi_a : \mathrm{Aut}_K(L) \to R_p$ is injective. To show that ϕ is surjective, note that $p(x)$ is the minimal polynomial of every $b \in R_p$ over K, so one has the *evaluation map*

$$e_b : \quad K[x]/\langle p\rangle \longrightarrow L, \quad f(x) + \langle p\rangle \longmapsto f(b).$$

The image of e_b is the sub-field $K(b)$ of L. As

$$\deg(p) = [L : K] = [L : K(b)][K(b) : K] = [L : K(b)]\deg(p),$$

we have $[L : K(b)] = 1$, so $L = K(b)$, and e_b is a field isomorphism. It follows that for every $b \in R$, one has

$$\sigma_{a,b} \stackrel{\mathrm{def}}{=} e_b \circ e_a^{-1} \in \mathrm{Aut}_K(L)$$

and $\sigma_{a,b}(a) = b$. This shows that ϕ_a is surjective. The two consequences are immediate. □

Remark 4.1.10. In the case of a simple algebraic extension $L = K(a)$ of K, Lemma 4.1.9 not only identifies $\mathrm{Aut}_K(L)$ with the set R of all roots of the minimal polynomial $p(x)$ of a, it also tells us how to identify the group structure of $\mathrm{Aut}_K(L)$. Indeed, for $\sigma_i \in \mathrm{Aut}_K(L)$, $i = 1, 2, 3$, one has

$$\sigma_1\sigma_2 = \sigma_3 \quad \text{ if and only if } \quad \sigma_1(\sigma_2(a)) = \sigma_3(a).$$

$\diamond$

If $K \subset L$ is a splitting field, recall from Corollary 3.2.28 that given any two elements of L that are roots of the same irreducible polynomial $p(x) \in K[x]$, one has an K-automorphism of L that takes one root to another. In other words, $\mathrm{Aut}_K(L)$ acts transitively on the set R_p of all roots of p in L. We formulate this fact in a more precise fashion in the next Lemma 4.1.11 and we recall its proof.

Lemma 4.1.11. *(***Construction Lemma of Automorphisms of Splitting Fields***)* *Let L be a splitting field of some polynomial in $K[x]$ over K. Let $\alpha \in L$, and let $p(x) \in K[x]$ be the irreducible polynomial of α in $K[x]$. Let R_p be the set of all roots of p in L. Then the map*

$$\phi : \ \mathrm{Aut}_K(L) \longrightarrow R_p, \ \ \sigma \longmapsto \sigma(\alpha)$$

is surjective and it induces a bijection

$$[\phi] : \ \mathrm{Aut}_K(L)/\mathrm{Aut}_{K(\alpha)}(L) \longrightarrow R_p.$$

Proof. Let $\beta \in R_p$ and consider the field isomorphisms

$$e_\alpha : \ K[x]/\langle p\rangle \longrightarrow K(\alpha) \quad \text{ and } \quad e_\beta : \ K[x]/\langle p\rangle \longrightarrow K(\beta).$$

One thus has the field embedding $i \circ e_\beta \circ e_\alpha^{-1} : K(\alpha) \to L$, where $i : K(\beta) \to L$ is the inclusion map. If L is a splitting field of $f(x) \in K[x]$ over K, then L is also a splitting field of $f(x) \in K(\alpha)[x]$ over $K(\alpha)$. By the Extension Lemma (see Lemma 3.2.25), there exists $\sigma \in \mathrm{Aut}_K(L)$ such that $\sigma(\alpha) = \beta$. This shows that ϕ is surjective. Clearly, for $\sigma, \sigma' \in \mathrm{Aut}_K(L)$, $\sigma(\alpha) = \sigma'(\alpha)$ if and only if $\sigma^{-1}\sigma' \in \mathrm{Aut}_{K(\alpha)}(L)$, so ϕ induces a bijection from $\mathrm{Aut}_K(L)/\mathrm{Aut}_{K(\alpha)}(L)$ to R_p. $\square$

Lemma 4.1.9 is enough for us to determine the automorphism groups of many extensions.

Example 4.1.12. If $K \subset L = K(a)$ is a simple algebraic extension such that the minimal polynomial $p(x)$ of a over K has only one root in L, namely a, then $\mathrm{Aut}_K(L) = \{e\}$ is the trivial group. For example, $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2})) = \{e\}$.

Example 4.1.13. Consider $\mathbb{Q} \subset L = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. We have seen in Example 3.1.35 that $L = \mathbb{Q}(\sqrt{2} + \sqrt{3})$ and that the minimal polynomial of $\sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$ is

$$p(x) = x^4 - 10x^2 + 1.$$

It is easy to see that $p(x)$ has 4 roots given as $\pm(\sqrt{2} \pm \sqrt{3})$ which are all in L (note that $\sqrt{3} - \sqrt{2} = 1/(\sqrt{3} + \sqrt{2})$). By Lemma 4.1.9, $|\mathrm{Aut}_{\mathbb{Q}}(L)| = 4$, and thus $\mathrm{Aut}_{\mathbb{Q}}(L) \cong \mathbb{Z}_4$ or $\mathbb{Z}_2 \times \mathbb{Z}_2$. To determine which one it is, let $\sigma \in \mathrm{Aut}_{\mathbb{Q}}(L)$ be a non-identity element and consider $\sigma(\sqrt{2})$ and $\sigma(\sqrt{3})$. By Lemma 4.1.6,

$$\sigma(\sqrt{2}) = \pm\sqrt{2}, \quad \sigma(\sqrt{3}) = \pm\sqrt{3}$$

and thus $\sigma^2(\sqrt{2}) = \sqrt{2}$ and $\sigma^2(\sqrt{3}) = \sqrt{3}$. It follows that $\sigma^2 = 1$. Thus any non-identity element in $\mathrm{Aut}_{\mathbb{Q}}(L)$ has order 2. We thus conclude that $\mathrm{Aut}_{\mathbb{Q}}(L) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. $\diamond$

Example 4.1.14. Recall now that for an integer $n \geq 1$, the n'th cyclotomic field is $\mathbb{Q}(\omega_n)$ where $\omega_n = e^{2\pi i/n}$, and it is the splitting field of the polynomial $f(x) = x^n - 1$ over $\mathbb{Q}$. The minimal polynomial of ω_n over $\mathbb{Q}$ is the n'th cyclotomic polynomial

$$\Phi_n = \prod_{1 \leq k \leq n,\, (k,n)=1} (x - \omega_n^k).$$

As Φ_n completely splits in $\mathbb{Q}(\omega_n)[x]$ and has no repeated roots, the group $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\omega_n))$ is a group of order

$$\phi(n) := \deg(\Phi_n) = |\{1 \leq k \leq n : (k, n) = 1\}|.$$

Moreover, an element $g \in G$ is uniquely determined by $g(\omega_n)$ which must be a root of Φ_n. For each $1 \leq k \leq n$ such that $(k, n) = 1$, since $\mathbb{Q}(\omega_n) = \mathbb{Q}(\omega_n^k)$, one has $\sigma_k \in G$ given by $\sigma_k(\omega_n) = \omega_n^k$. For two such elements σ_k and σ_l, since

$$\sigma_k(\sigma_l(\omega_n)) = \omega_n^{kl},$$

one has $\sigma_k\sigma_l = \sigma_j$, where $1 \leq j \leq n$, $(j, n) = 1$, and $kl = j \bmod n$. It is now clear that G is isomorphic to the group $(\mathbb{Z}/n\mathbb{Z})^\times$ of invertible elements in the commutative ring $\mathbb{Z}/n\mathbb{Z}$.

Let $n = p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$ be the prime factorization of n. By the Chinese Remainder Theorem,

$$\mathbb{Z}/n\mathbb{Z} \cong (\mathbb{Z}/p_1^{k_1}\mathbb{Z}) \times (\mathbb{Z}/p_2^{k_2}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_m^{k_m}\mathbb{Z}),$$

so

$$(\mathbb{Z}/n\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{k_1}\mathbb{Z})^\times \times (\mathbb{Z}/p_2^{k_2}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/p_m^{k_m}\mathbb{Z})^\times.$$

Consider now the group $(\mathbb{Z}/p^k\mathbb{Z})^\times$ for a prime number p and integer $k \geq 1$. The integers in $[1, p^k]$ that are not co-prime with p^k are precisely those that are multiples

of p. Thus $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is an abelian group of size $p^k - p^{k-1}$. It is a well-known fact, see Wikipedia article on "Multiplicative group of integers modulo n", that for each prime number $p \neq 2$, $(\mathbb{Z}/p^k\mathbb{Z})^\times \cong \mathbb{Z}/(p^k - p^{k-1})\mathbb{Z}$ is cyclic, while for $p = 2$, $(\mathbb{Z}/2^k\mathbb{Z})^\times$ is cyclic iff $k = 0, 1, 2$. For example, we have

$$(\mathbb{Z}/8\mathbb{Z})^\times \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

$\diamond$

Example 4.1.15. Let p be a prime number and $n \geq 1$ an integer. Consider the finite extension

$$\mathbb{F}_p \subset \mathbb{F}_{p^n}.$$

We have seen that the extension $\mathbb{F}_p \subset \mathbb{F}_{p^n}$ is simple: $\mathbb{F}_{p^n} = \mathbb{F}_p(a)$, where $a \in \mathbb{F}_{p^n}\backslash\{0\}$ is any generator of $\mathbb{F}_{p^n}\backslash\{0\}$ as a cyclic group, and that the minimal polynomial of a in $\mathbb{F}_p[x]$, being a factor of $f_n(x) = x^{p^n} - x \in \mathbb{F}_p[x]$, splits completely over $\mathbb{F}_{p^n}$ and has no repeated roots in $\mathbb{F}_{p^n}$. Thus the automorphism group $G = \mathrm{Aut}_{\mathbb{F}_P}(\mathbb{F}_{p^n})$ has order n. To determine G, note that we already have the Frobenius isomorphism $\sigma \in G$ given by

$$\sigma(a) = a^p, \quad a \in \mathbb{F}_{p^n}.$$

It is clear that for each $1 \leq j \leq n-1$, $\sigma^j(a) = a^{p^j}$, and $\sigma^n = \mathrm{Id}_{\mathbb{F}_{p^n}}$. It is also clear that for any $1 \leq j \leq n-1$, the map σ^j could not be the identity map on $\mathbb{F}_{p^n}$ because otherwise the polynomial $x^{p^j} - x \in \mathbb{F}_{p^n}[x]$ would have p^n roots which is not possible. Thus G contains the subgroup generated by the Frobenius isomorphism σ which has order n. Consequently,

$$G = \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p) = \{\sigma^j : 0 \leq j \leq n-1\}$$

is a cyclic group of order n, generated by the Frobenius isomorphism. $\diamond$

§ 2. Galois extensions and first examples

Definition 4.1.16. A finite field extension $K \subset L$ is called a *Galois extension* if

$$|\mathrm{Aut}_K(L)| = |L : K|.$$

When $K \subset L$ is a Galois extension, it is also common to denote $\mathrm{Aut}_K(L)$ by $\mathrm{Gal}(L/K)$ or $\mathrm{Gal}_K(L)$

We have seen that $\mathrm{Aut}_{\mathbb{R}}(\mathbb{C}) \cong \mathbb{Z}_2$ and that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}[\sqrt[3]{2}])$ is the trivial group. Thus $\mathbb{C}$ is a Galois extension of $\mathbb{R}$, while the extension $\mathbb{Q} \subset \mathbb{Q}(\sqrt[3]{2})$ is not Galois.

By Example 4.1.13, Example 4.1.14, and Example 4.1.15, we have the following examples of Galois extensions:

- The field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is a Galois of $\mathbb{Q}$ of degree 4, with $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}, \sqrt{3})) \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$;

- For any integer $n \geq 2$, the cyclotomic extension $C_n = \mathbb{Q}\left(e^{\frac{2\pi i}{n}}\right)$ of $\mathbb{Q}$ is Galois with $\mathrm{GaL}_{\mathbb{Q}}(C_n) \cong (\mathbb{Z}/n\mathbb{Z})^{\times}$, the group of units of the ring $\mathbb{Z}/n\mathbb{Z}$.
- For any prime number p and any integer $n \geq 2$, the finite field $\mathbb{F}_{p^n}$ is a Galois extension of $\mathbb{F}_p$ with $\mathrm{Gal}_{\mathbb{F}_P}(\mathbb{F}_{p^n}) \cong \mathbb{Z}/n\mathbb{Z}$.

§ 3. Artin's Theorem

Before turning to Artin's construction of Galois extensions, we first have the following observation.

Lemma-Definition 4.1.17. For any field L and any subgroup H of $\mathrm{Aut}(L)$, the subset

$$L^H \stackrel{\text{def}}{=} \{a \in L : \sigma(a) = a \ \forall \sigma \in H\}$$

of L is a sub-field of L, which is called the *fixed sub-field of* H.

Theorem 4.1.18. *(Artin's Theorem) For any field L and any finite group H of $\mathrm{Aut}(L)$, L is a Galois extension of L^H and*

$$\mathrm{Aut}_{L^H}(L) = H.$$

Proof. Let $\alpha \in L$ be arbitrary, and list the distinguished elements in the H-orbit of α by $H\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Consider

$$f(x) = \prod_{i=1}^{n}(x - \alpha_i) \in L[x].$$

Note that all the coefficients of f have symmetric expressions in $\alpha_1, \ldots, \alpha_n$ are thus all in L^H. Consequently, $f(x) \in L^H[x]$. As α is a root of f in L, we see that α is algebraic over L^H. Let $p \in L^H[x]$ be the irreducible polynomial of α in $L^H[x]$. Then $p|f$ in $L^H[x]$. As f has no repeated roots in L, p also has no repeated roots in L, i.e., p is separable. Moreover,

$$|L^H(\alpha) : L^H| = \deg(p) \leq \deg f = n = |H\alpha| \leq |H|.$$

We also see that L is both an algebraic and a separable extension of L^H.

Now choose $\alpha \in L$ such that $|L^H(\alpha) : L^H|$ is the largest. We now prove that $L^H(\alpha) = L$. Suppose that $L^H(\alpha) \neq L$. Choose $\beta \in L \backslash L^H(\alpha)$. Then $L^H(\alpha, \beta)$ is a finite separable extension of L^H. By the *Primitive Element Theorem*,

$$L^H(\alpha, \beta) = L^H(\gamma)$$

for some $\gamma \in L$, contradicting the assumption on α. Thus $L^H(\alpha) = L$. By Lemma 4.1.9, the *Basic lemma on automorphism groups of finite simple extensions*, we have

$$|\mathrm{Aut}_{L^H}(L)| \leq |L : L^H| \leq |H|.$$

On the other hand, $H \subset \mathrm{Aut}_{L^H}(L)$ by definition. One thus has $\mathrm{Aut}_{L^H}(L) = H$ and

$$|\mathrm{Aut}_{L^H}(L)| = |L : L^H|.$$

In other words, L is a Galois extension of L^H with Galois group H. □

Artin's Theorem has many consequences. We first give an example.

Example 4.1.19. For any field K and any integer $n \geq 1$, let $L = K(x_1, \ldots, x_n)$, the fraction field of the polynomial ring $K[x_1, \ldots, x_n]$. Let the symmetric group S_n act on L by

$$\sigma \cdot f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}), \quad \sigma \in S_n.$$

Thus gives an embedding of S_n as a subgroup of $\mathrm{Aut}(L)$. By Artin's Theorem, L is a Galois extension of L^{S_n} with Galois group isomorphic to S_n. Note that L^{S_n} is the sub-field of L consisting of all symmetric rational functions in n variables with coefficients in K.

Artin's Theorem also implies that for any subgroup $H \subset S_n$, H is the Galois group of the extension $L^H \subset L$. As every finite group is a subgroup of S_n for some n, we also conclude that every finite group is the Galois group for some field extension.

§ 4. Characterizations of finite Galois extensions

We have the following consequence of Artin's Theorem.

Corollary 4.1.20. *Let $K \subset L$ be a finite field extension and let $G = \mathrm{Aut}_K(L)$. Then $|G|$ divides $[L : K]$. Consequently, $|G| \leq [L : K]$. Moreover, $|G| = [L : K]$ if and only if $K = L^G$.*

Proof. Applying Artin's Theorem to $G = \mathrm{Aut}_K(L)$, we see that

$$|G| = [L : L^G].$$

By the Tower Theorem,

$$[L : K] = [L : L^G][L^G : K] = |G|[L^G : K],$$

so $|G|$ divides $[L : K]$. In particular, $|G| \leq [L : K]$, and $|G| = [L : K]$ if and only if $[L^G : K] = 1$ which is the same as $L^G = K$. □

We thus have two equivalent definitions for a finite extension $K \subset L$ to be Galois:
1) $|\mathrm{Aut}_K(L)| = |L : K|$;
2) $L^G = K$ for $G = \mathrm{Aut}_K(L)$.

We now give two more characterizations of finite Galois extensions.

For any finite extension $K \subset L$ and $G = \mathrm{Aut}_K(L)$, recall from Lemma 4.1.6 that for any $\alpha \in L$, if $p(x)$ is the minimal polynomial of α in $K[x]$, then for any $\sigma \in G$, $\sigma(\alpha)$ is also a root of p in L. The next Proposition 4.1.21 says that when $K \subset L$ is a Galois extension, every root of p in L is of this form, and that there are exactly $\deg(p)$ of them.

Proposition 4.1.21. *Let $K \subset L$ be a finite Galois extension and let $G = \mathrm{Aut}_K(L)$. Let $\alpha \in L$ and let $p(x) \in K[x]$ be the minimal polynomial of α in $K[x]$. Then $p(x)$ splits completely in $L[x]$ and with no repeated roots, and the set of all roots of p in L is precisely*

$$G\alpha = \{\sigma(\alpha) : \sigma \in G\}.$$

In particular, $|G\alpha| = \deg(p) = |K(\alpha) : K|$.

Proof. Write $G\alpha = \{\alpha, \alpha_2, \ldots, \alpha_r\}$, and let

$$q(x) = (x - \alpha)(x - \alpha_2) \cdots (x - \alpha_r) \in L[x]$$

Then all the coefficients of $q(x)$ have polynomial expressions in elements in $G\alpha$ which are invariant under permutations on $G\alpha$. Now each $\tau \in G$ permutes the elements in $G\alpha$. Thus all the coefficients of $q(x)$ are in $L^G = K$. Since $q(\alpha) = 0$, one has $p(x)|q(x)$. On the other hand, we also know that every element in $G\alpha$ is a root of $p(x)$ in L, so $q(x)|p(x)$ in $L[x]$. Thus $q(x) = p(x)$ as elements in $L[x]$ and thus also as elements in $K[x]$. This shows that p splits completely in $L[x]$, the set of all roots of p in L is $G\alpha$, and p has no repeated roots in L. Thus we also have

$$\deg(p) = |G\alpha| = [K(\alpha) : K].$$

□

Recall that an extension $\subset L$ is said to be normal and separable if for each $\alpha \in L$, the minimal polynomial p of α in $L[x]$ splits completely and has no repeated roots. Proposition 4.1.21 thus says that any finite Galois extension is normal and separable. We now prove that the converse is also true.

Theorem 4.1.22. *A finite extension $K \subset L$ is Galois if and only if it is normal and separable.*

Proof. Assume that $K \subset L$ is a finite normal and separable extension and let $G = \mathrm{Aut}_K(L)$. We need to show that $L^G = K$. Since $K \subset L^G$, we only need to show that $L^G \subset K$.

Assume that $\alpha \in L^G$ and let $p \in K[x]$ be the minimal polynomial of α over K. By assumption, p splits completely over L and has no repeated roots in L. By Theorem 3.2.32, L is a splitting field. By the *Construction Lemma of automorphisms of Splitting fields*, if p has a root β in L that is different from α, then there exists $\sigma \in G$ such that $\sigma(\alpha) = \beta$, contradicting the assumption that $\alpha \in L^G$. Thus α is the only root of p in L, so $\alpha \in K$. □

We now give yet another, but more practical, criterion for a finite extension to be a Galois extension. Recall that for a field K, a non-constant $f(x) \in K[x]$ is said to be separable if K has no repeated root in its splitting field over K.

Theorem 4.1.23. *A finite extension $K \subset L$ is Galois if and only if L is the splitting field of a separable polynomial over K.*

Proof. Assume that L is a finite Galois extension of K. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis of L over K, and for $j = 1, \ldots, n$, let $p_j(x) \in K[x]$ be the minimal polynomial of α_j in $K[x]$. By Proposition 4.1.21, each $p_j \in K[x]$ splits completely in $L[x]$ and has no repeated roots in L. It is possible that $p_i = p_j$ for $i \neq j$. List $\{p_i : i = 1, \ldots, n\}$ as $\{p_{i_1}, p_{i_2}, \ldots, p_{i_k}\}$ where $p_{i_j} \neq p_{i_{j'}}$ for $j \neq j'$, and let

$$f(x) = p_{i_1}(x)p_{i_2}(x)\cdots p_{i_k}(x) \in K[x].$$

As different irreducible polynomials in $K[x]$ can not have common roots in any field extension of K, we know that $f(x)$ splits completely in $L[x]$ and has no repeated roots in L. Thus $f \in K[x]$ is separable and L is a splitting field of f over K.

Assume now that L is the splitting field of a separable polynomial $f(x) \in K[x]$ over K. We prove $|G| = [L : K]$ by induction on $[L : K]$. If $[L : K] = 1$, there is nothing to prove. Assume that $[L : K] \geq 2$. Let $p(x) \in K[x]$ be an irreducible factor of f in $K[x]$. Then there must be root α of $f(x)$ in L which is also a root of $p(x)$. Let R_p be the set of all roots of p in L. Since f completely splits in L and has no repeated roots, the same holds for $p(x)$. Thus $|R_p| = \deg(p) = [K(\alpha) : K]$. By Lemma 4.1.11,

$$|G| = |\mathrm{Aut}_{K(\alpha)}(L)||R_p| = |\mathrm{Aut}_{K(\alpha)}(L)|[K(\alpha) : K].$$

Note that L is also a splitting field of f over $K(\alpha)$ and that f is also separable over $K(\alpha)$. By induction assumption, $|\mathrm{Aut}_{K(\alpha)}(L)| = [L : K(\alpha)]$. By the Tower Theorem,

$$|G| = [L : K(\alpha)]\,[K(\alpha) : K] = [L : K].$$

Thus L is a Galois extension of K. □

We now summarize the four characterizations we have proved for finite Galois extensions.

Theorem 4.1.24. *Let $K \subset L$ be a finite extension and let $G = \mathrm{Aut}_K(L)$. The following are equivalent:*

1) $K \subset L$ is Galois, i.e., $|G| = [L : K]$;

2) $K = L^G$;

3) The extension $K \subset L$ is normal and separable;

4) L is a splitting field over K of some separable polynomial $f(x) \in K[x]$.

As every algebraic extension of a perfect field is separable, by Theorem 4.1.22, we have

Theorem 4.1.25. *Let K be a perfect field. Then a finite extension $K \subset L$ is Galois if and only if L is a splitting field over K of some $f(x) \in K[x]$.*

We now use Theorem 4.1.25 to compute more examples.

Example 4.1.26. Consider the irreducible polynomial $f(x) = x^3 - 2 \in \mathbb{Q}[x]$, and L be its splitting field in $\mathbb{C}$. Then

$$L = \mathbb{Q}(\sqrt[3]{2}, \omega),$$

where $\omega = e^{2\pi i/3}$. By Theorem 4.1.25, $|\mathrm{Aut}_K(L)| = [L : \mathbb{Q}] = 6$. By Lemma 4.1.7, $\mathrm{Aut}_K(L)$ is a subgroup of S_3 which has 6 elements. Thus $\mathrm{Aut}_K(L) = S_3$.

Example 4.1.27. Consider $f(x) = x^5 - 4x + 2 \in \mathbb{Q}[x]$ and its splitting field L over Q. By Theorem 4.1.25, L is a Galois extension of $\mathbb{Q}$. As f is irreducible over $\mathbb{Q}$ by Eisenstein's criterion, f has no repeated roots in its splitting field L. Thus $\mathrm{Gal}_{\mathbb{Q}}(f)$ is a subgroup of S_5.

We now use calculus to determine the real roots of f: f has two critical points $\pm\sqrt[4]{\frac{4}{5}}$, and $f(\sqrt[4]{\frac{4}{5}}) < 0$ and $f(-\sqrt[4]{\frac{4}{5}}) > 0$. Since also $\lim_{x\to\infty} f(x) = \infty$ and $\lim_{x\to-\infty} f(x) = -\infty$, we know that f has three real roots (they are approximately -1.5185, 0.5085, and 1.2435). Thus f has also two complex roots (they are approximately $-0.1168 \pm 1.4385i$). The complex conjugation is one element of order 2 in $\mathrm{Gal}_{\mathbb{Q}}(f)$.

Take any root r of f and let $L_1 = \mathbb{Q}(r)$. Then L is an extension of L_1 and $[L_1 : \mathbb{Q}] = 5$. Thus

$$[L : \mathbb{Q}] = [L : L_1][L_1 : \mathbb{Q}] = 5[L : L_1].$$

As $|\mathrm{Gal}_{\mathbb{Q}}(f)| = [L : \mathbb{Q}]$ by Theorem 4.1.25, so $|\mathrm{Gal}_{\mathbb{Q}}(f)|$ is divisible by 5. By Cauchy's theorem on finite groups, there is an element of $\mathrm{Gal}_{\mathbb{Q}}(f)$ with order 5. It is an easy fact to show that if a subgroup G of S_5 contains an order 5 element f and an order 2 element t, then $G = S_n$ (keep conjugating t by f to produce more elements in G). Thus we conclude that $\mathrm{Gal}_{\mathbb{Q}}(f) \cong S_5$.

§ 5. The Galois Correspondence

The Fundamental Theorem of Galois Theory establishes a one-to-one correspondence between *intermediate fields* of a finite Galois extension $K \subset L$ and *subgroups* of the automorphism group $\mathrm{Aut}_K(L)$.

Definition 4.1.28. If $K \subset L$ is a field extension, any sub-field M of L containing K will be called an *intermediate field* and denoted as $K \subset M \subset L$.

Lemma 4.1.29. *Let $K \subset L$ be a field extension and let $G = \mathrm{Aut}_K(L)$.*

1) For any intermediate field $K \subset M \subset L$, $\mathrm{Aut}_M(L)$ is a subgroup of G;

2) For any subgroup H of G,

$$L^H \overset{\text{def}}{=} \{a \in L : \sigma(a) = a,\ \forall\, \sigma \in H\}$$

is an intermediate field of $K \subset L$, called the fixed field of H.

Proof. 1) follows from the definition of subgroups. For 2), by definition of $G = \mathrm{Aut}_K(L)$, one has $\sigma(a) = a$ for all $\sigma \in G$ and $a \in K$. Thus $K \subset L^H$. Let $a, b \in L^H$, Then for any $\sigma \in H$, $\sigma(a-b) = \sigma(a) - \sigma(b) = a - b$, and $\sigma(ab) = \sigma(a)\sigma(b) = ab$. Moreover, when $b \neq 0$, $1 = \sigma(1) = \sigma(bb^{-1}) = \sigma(b)\sigma(b^{-1})$, so $\sigma(b^{-1}) = \sigma(b)^{-1}$. Thus L^H is a sub-field of L. □

For a field extension $K \subset L$ and $G = \mathrm{Aut}_K(L)$, we then have two maps

$$\{\text{intermediate fields } \ K \subset M \subset L\} \ \underset{F}{\overset{\Gamma}{\rightleftarrows}} \ \{\text{subgroups of } \ G\},$$

where for an intermediate field $K \subset M \subset L$ and a subgroup $H \subset G$,

$$\Gamma(M) = \mathrm{Aut}_M(L) \quad \text{and} \quad F(H) = L^H = \{a \in L : \sigma(a) = a, \ \forall \sigma \in H\}. \tag{4.1}$$

How are the two maps Γ and F related? We first prove some easy properties.

Lemma 4.1.30. *Let $K \subset L$ be any field extension, and let $G = \mathrm{Aut}_K(L)$.*

1) Both F and Γ are inclusion reversing, i.e., $M_1 \subset M_2$ implies $\Gamma(M_1) \supset \Gamma(M_2)$ and $H_1 \subset H_2$ implies that $F(H_1) \supset F(H_2)$.

2) One has $M \subset F(\Gamma(M))$ and $H \subset \Gamma(F(H))$ for any intermediate field M of $K \subset L$ and any subgroup H of G;

3) $\Gamma \circ F \circ \Gamma = \Gamma$ and $F \circ \Gamma \circ F = F$.

Proof. 1) and 2) are straightforward and are left as exercise.

3) Let $K \subset M \subset L$ be any intermediate field. By 2), $M \subset F(\Gamma(M))$, so by 1), $\Gamma(M) \supset \Gamma(F(\Gamma(M))$. On the other hand, by taking $H = \Gamma(M)$, 2) also implies that $\Gamma(M) \subset \Gamma(F(\Gamma(M))$. Thus $\Gamma(M)) = \Gamma(F(\Gamma(M))$. Similarly, one shows that $F(\Gamma(F(H))) = F(H)$ for all subgroups H of G. □

Note that when $K \subset L$ is a finite extension, $\mathrm{Aut}_K(L)$ is a finite group, so by Artin's theorem, we have

$$H = \Gamma(F(H)), \quad \text{i.e.} \quad H = \mathrm{Aut}_{L^H}(L)$$

for any subgroup H of $\mathrm{Aut}_K(L)$. The Fundamental Theorem of Galois Theory states that when $K \subset L$ is a finite *Galois extension*, one also has

$$M = F(\Gamma(M)) \quad \text{i.e.} \quad M = L^{\mathrm{Aut}_M(L)} \tag{4.2}$$

for all intermediate fields $K \subset M \subset L$, which is the same as saying that $M \subset L$ is Galois. The correspondence between intermediate fields $K \subset M \subset L$ and subgroups H of the Galois group $\mathrm{Aut}_K(L)$ through the two maps Γ and F is called the *Galois correspondence*. One can thus study field extensions through group theory.

Example 4.1.31. Consider the case of $L = \mathbb{Q}(\sqrt[3]{2})$ as an extension of $\mathbb{Q}$ and $M = \mathbb{Q}$. We know that $\mathrm{Aut}_M(L) = G = \{e\}$ so $L^{\mathrm{Aut}_M(L)} = L$, which is bigger than M. Thus the first equality of (4.2) does not hold in this case. On the other hand, we have seen that L is NOT a Galois extension of $\mathbb{Q}$.

Theorem 4.1.32. *(***Fundamental Theorem of Galois Theory***) For a finite Galois extension $K \subset L$,*

(i) For any intermediate field M of $K \subset L$, the extension $M \subset L$ is Galois;

(ii) the two maps Γ and F are inverses of each other, i.e.,

$$M = L^{\mathrm{Aut}_M(L)} \quad \text{and} \quad H = \mathrm{Aut}_{L^H}(L),$$

for all intermediate fields $K \subset M \subset L$ and all subgroups H of $G = \mathrm{Aut}_K(L)$.

(iii) for an intermediate field $K \subset M \subset L$, one always has

$$|\mathrm{Aut}_K(L)/\mathrm{Aut}_M(L)| = \frac{|\mathrm{Aut}_K(L)|}{|\mathrm{Aut}_M(L)|} = \frac{[L:K]}{[L:M]} = [M:K], \tag{4.3}$$

and the following four statements are equivalent:

(a) $\mathrm{Aut}_M(L)$ is a normal subgroup of $\mathrm{Aut}_K(L)$;

(b) $\sigma(M) = M$ for all $\sigma \in \mathrm{Aut}_K(L)$;

(c) the extension $K \subset M$ is Galois;

(d) the extension $K \subset M$ is normal,

and in this case, one has the group isomorphism

$$\mathrm{Aut}_K(L)/\mathrm{Aut}_M(L) \longrightarrow \mathrm{Aut}_K(M) : \ \sigma\mathrm{Aut}_M(L) \longmapsto \sigma|_M.$$

Proof. (i) Let M be any intermediate field of $K \subset L$. By Theorem 4.1.24, L is the splitting field of a separable polynomial $f(x) \in K[x]$. Then L is also the splitting field of f over M by regarding f as in $M[x]$. Clearly f is also separable over M. Thus, by Theorem 4.1.24 again, L is a Galois extension of M.

(ii) Since G is finite, by Theorem 4.1.18, $H = \mathrm{Aut}_{L^H}(L)$ for every subgroup H of G. For any intermediate field of $K \subset L$, since $M \subset L$ is Galois by (i), one has $M = L^{\mathrm{Aut})_M(L)}$ by Theorem 4.1.24.

(iii) Fix any intermediate field $K \subset M \subset L$. Using the fact that both $K \subset L$ and $M \subset L$ are Galois, and by the Tower Theorem, we have

$$|\mathrm{Aut}_K(L)/\mathrm{Aut}_M(L)| = \frac{|\mathrm{Aut}_K(L)|}{|\mathrm{Aut}_M(L)|} = \frac{[L:K]}{[L:M]} = [M:K].$$

Let $\sigma \in G$ and $\tau \in \mathrm{Aut}_M(L)$ be arbitrary. Then $\sigma^{-1}\tau\sigma \in \mathrm{Aut}_M(L)$ is equivalent to $\tau(\sigma(a)) = \sigma(a)$ for all $a \in M$. Thus $\mathrm{Aut}_M(L)$ is a normal subgroup of G iff

$$\sigma(a) \in L^{\mathrm{Aut}_M(L)} = M, \quad \forall\, \sigma \in G,\, a \in M.$$

This shows that (a) and (b) are equivalent.

Assume (b) holds. We then have the group homomorphism

$$\phi: \ \mathrm{Aut}_K(L) \longrightarrow \mathrm{Aut}_K(M), \ \ \sigma \longmapsto \sigma|_M. \tag{4.4}$$

It is clear that $\ker\phi = \mathrm{Aut}_M(L)$, so ϕ induces an injective group homomorphism

$$[\phi]: \ \mathrm{Aut}_K(L)/\mathrm{Aut}_M(L) \longrightarrow \mathrm{Aut}_K(M).$$

On the other hand, by Corollary 4.1.20, we always have $|\mathrm{Aut}_K(M)| \leq [M:K]$, so (4.3) implies that

$$|\mathrm{Aut}_K(M)| = [M:K] = |\mathrm{Aut}_K(L)/\mathrm{Aut}_M(L)|.$$

Thus $[\phi]$ is an isomorphism and $K \subset M$ is Galois. This also shows that (b) implies (c).

Clearly (c) implies (d). Assume (d), i.e., $K \subset M$ is normal. By Theorem 3.2.32, M is a splitting field of some $g(x) \in K[x]$ over K. Let R_g be the set of all the roots of g in M, so that $M = K(R_g)$. By Lemma 4.1.7, one has $\sigma(R_g) = R_g$, thus $\sigma(M) = M$ for every $\sigma \in \mathrm{Aut}_K(L)$. This shows that (d) implies (b).

This finishes the proof of Theorem 4.1.32. □

Corollary 4.1.33. *A finite Galois extension $K \subset L$ has finitely many intermediate fields.*

Example 4.1.34. Recall from Example 4.1.27 that the Galois group of the splitting field L over $\mathbb{Q}$ of the polynomial $f(x) = x^5 - 4x + 2 \in \mathbb{Q}[x]$ is isomorphic to S_5. As there are 156 subgroups of S_5 (this fact can be found in many online articles), there are 156 fields K with $\mathbb{Q} \subset K \subset L$.

Corollary 4.1.33 is not true without the Galois condition. The following example is a continuation of Example 4.1.35 and is from Cox's book *Galois Theory*, Second Edition, Example 7.3.6.

Example 4.1.35. Let F be a field with characteristic $p > 0$, let $K = F(t,u)$, and consider the splitting field L over K of

$$f(x) = (x^p - t)(x^p - u) \in K[x].$$

Let $\alpha \in L$ be a root of $x^p = t$ and $\beta \in L$ be a root of $x^p - u$. We have seen in Example 3.2.62 that $L = K(\alpha, \beta)$, and

$$|L : K(\alpha)| = |K(\alpha) : K| = p.$$

Thus $|L:K| = p^2$ and $\{\alpha^i\beta^j : 0 \leq i,j \leq p-1\}$ is a basis of L over K. It is also shown in Example 3.2.62 that $|K(\gamma):K| = p$ for any $\gamma \in L\backslash K$.

For $\lambda \in K$, consider the intermediate field

$$K \subset K(\alpha + \lambda\beta) \subset L.$$

As $\alpha + \lambda\beta \notin K$, we have $|K(\alpha + \lambda\beta) : K| = p$. In particular, a basis of $K(\alpha + \lambda\beta)$ over K is given by $\{(\alpha + \lambda\beta)^n : n = 0, 1, 2, \ldots, p-1\}$. Since $\{\alpha^i\beta^j : 0 \leq i, j \leq p-1\}$ is a basis of L over K, if $\mu \in K$ and $\mu \neq \lambda$, then $\alpha + \mu\beta \notin K(\alpha + \lambda\beta)$. It follows that

$$K(\alpha + \lambda\beta) \neq K(\alpha + \mu\beta)$$

if $\lambda, \beta \in K$ and $\lambda \neq \mu$. Since K is an infinite field, we have thus obtained infinitely many pairwise distinct intermediate fields of $K \subset L$. Note, on the other hand, $\mathrm{Aut}_K(L) = \{e\}$. Indeed, for any $\sigma \in \mathrm{Aut}_K(L)$, as α is the only root of $x^p - t \in K[x]$ and β is the only root of $x^p = u \in K[x]$, we must have $\sigma(\alpha) = \alpha$ and $\sigma(\beta) = \beta$, so $\sigma = e$. Thus the Galois Correspondence fails miserably in this example.

The Galois Correspondence has more properties than stated in the Fundamental Theorem of Galois Theory, and we look at them now. We first recall some standard definitions.

Definition 4.1.36. 1) A *partially ordered set*, or a *poset*, is a non-empty set P with a binary relation $\preceq$ that is

(i) reflexive: $a \preceq a$ for all $a \in P$;

(ii) transitive: if $a \preceq b$ and $b \preceq c$ then $a \preceq c$;

(iii) antisymmetric: $a \preceq b$ and $b \preceq a$ imply that $a = b$.

2) Given two posets P_1 and P_2, a map $\phi : P_1 \to P_2$ is said to be *order preserving* if $a \preceq b$ implies $\phi(a) \preceq \phi(b)$ for all $a, b \in P_1$. Similarly, ϕ is said to be *order reserving* if $a \preceq b$ implies $\phi(b) \preceq \phi(a)$ for all $a, b \in P_1$.

Definition 4.1.37. 1) Given a poset $(P, \preceq)$ and $a, b \in P$, an *upper bound* of a and b is an element $c \in P$ such that $a \preceq c$ and $b \preceq c$; a *least upper bound* of a and b is an upper bound d of a and b such that $d \preceq c$ for every upper bound c of a and b. (note that least upper bound of a and b, if exists, is unique by antisymmetry of $\preceq$). Greatest lower bounds are similarly defined.

2) A lattice is a poset $(P, \preceq)$ such that each pair of elements $a, b \in P$ has a *least upper bound* denoted as $a \vee b$ and a greatest lower bound denoted as $a \wedge b$.

The following lemma is straightforward to prove.

Lemma 4.1.38. *If P_1 and P_2 are two lattices and if $\phi : P_1 \to P_2$ is order reversing bijection, then for all $a, b \in P_1$,*

$$\phi(a \vee b) = \phi(a) \wedge \phi(b) \quad \text{and} \quad \phi(a \wedge b) = \phi(a) \vee \phi(b).$$

Example 4.1.39. If G is a group, the set of all subgroups of G is a lattice, where $H \preceq K$ means $H \subset K$, and

$$H \vee K = \text{the subgroup of } G \text{ generated by } H \text{ and } K, \qquad H \wedge K = H \cap K.$$

Definition 4.1.40. Given a field L, the *compositum* of a collection $\{K_\alpha\}$ of sub-fields L, denoted as

$$\prod_\alpha K_\alpha,$$

is the smallest sub-field of L containing every K_α, or, equivalently, the intersection of all sub-fields of L that contain every K_α.

Let now $K \subset L$ be any field extension, and consider the set of all intermediate fields of $K \subset L$. For two intermediate fields M_1, M_2, define $M_1 \preceq M_2$ if $M_1 \subset M_2$, and define

$$M_1 \vee M_2 = M_1 M_2 \quad \text{and} \quad M_1 \wedge M_2 = M_1 \cap M_2.$$

Then it is straightforward to check that set of all intermediate fields of $K \subset L$ is a lattice under $\preceq$, which will be called the *lattice of intermediate fields* of $K \subset L$.

As a consequence of the Galois correspondence being order reserving, we have

Theorem 4.1.41. *For a finite Galois extension $K \subset L$, and for any intermediate fields $K \subset M_1 \subset L$, $K \subset M_2 \subset L$, and subgroups H_1, H_2 of $G = \mathrm{Aut}_K(L)$, one has*

$$\begin{aligned} &L^{H \vee H_2} = L^{H_1} \cap L^{H_2}, \quad L^{H_1 \cap H_2} = L^{H_1} \vee L^{H_2}, \\ &\mathrm{Aut}_{M_1 M_2}(L) = \mathrm{Aut}_{M_1}(L) \cap \mathrm{Aut}_{M_1}(L), \\ &\mathrm{Aut}_{M_1 \cap M_2}(L) = \mathrm{Aut}_{M_1}(L) \vee \mathrm{Aut}_{M_1}(L). \end{aligned}$$

We now recall another concept in group theory: if H is a subgroup of a group G, then for any $g \in G$,

$$gHg^{-1} = \{ghg^{-1} : h \in H\}$$

is also a subgroup of G, called a conjugate of H. According to this terminology, a subgroup H of G is normal iff it has no other conjugate other than itself.

Lemma-Definition 4.1.42. Given a field extension $K \subset L$ and an intermediate field $K \subset M \subset L$, for any $\sigma \in \mathrm{Aut}_K(L)$, $\sigma(M) = \{\sigma(a) : a \in M\}$ is an intermediate field with

$$\mathrm{Aut}_{\sigma(M)}(L) = \sigma \mathrm{Aut}_M(L) \sigma^{-1} \subset \mathrm{Aut}_K(L).$$

The intermediate field $K \subset \sigma(M) \subset L$ is called a *Galois conjugate* of M.

Proof. Proof is straightforward and is left an an exercise. □

Given a finite Galois extension $K \subset L$, Galois conjugates of intermediate fields then correspond to conjugates of subgroups under the Galois correspondence

$$\{\text{intermediate fields } K \subset M \subset L\} \longrightarrow \{\text{subgroups of } G = \mathrm{Aut}_K(L)\}.$$

Note that the Fundamental Theorem of Galois Theory says that an intermediate field M is Galois of K iff it has no Galois conjugates other than itself.

§ 6. Galois closures

Let $K \subset L$ be a finite Galois extension. By (iii)(b) of the Fundamental Theorem of Galois Theory, if $\{K \subset M_i \subset L\}_{i\in I}$ is any collection of intermediate fields such that $K \subset M_i$ is Galois for each $i \in I$, then the extension $K \subset \bigcap_{i\in I} M_i$ is Galois. This leads to the following definition.

Definition 4.1.43. Let $K \subset L$ be a finite Galois extension, and let $K \subset M \subset L$ be an intermediate field. The intersection, denoted by $\overline{M}^{L,\mathrm{Galois}}$, of all intermediate fields $K \subset M_i \subset L$ such that $M \subset M_i$, is called the *Galois closure of M over K in L*. Note that $\overline{M}^{L,\mathrm{Galois}}$ is the smallest intermediate field of $K \subset L$ containing M that is a Galois extension of K.

For any subgroup H of any group G, let $S_H = \bigcap gHg^{-1}$. It is straightforward to check that S_H is the largest normal subgroup of G that is contained in H.

Proposition 4.1.44. *Let $K \subset L$ be a finite Galois extension, and let $K \subset M \subset L$ be an intermediate field. Then*

$$\overline{M}^{L,\mathrm{Galois}} = \prod_{\sigma\in \mathrm{Aut}_K(L)} \sigma(M),$$

and, under the Galois Correspondence, $\overline{M}^{L,\mathrm{Galois}}$ corresponds to $S_{\mathrm{Aut}_M(L)}$, the largest normal subgroup of $G = \mathrm{Aut}_K(L)$ that is contained in $\mathrm{Aut}_M(L)$.

Proof. Set $C = \prod_{\sigma\in \mathrm{Aut}_K(L)} \sigma(M)$. Note that C is invariant under the action of every $\sigma \in \mathrm{Aut}_K(L)$, so $K \subset C$ is a Galois extension by the Fundamental Theorem of Galois Theory. On the other hand, by the Fundamental Theorem of Galois Theory again, any intermediate field M' containing M must contain all of its Galois conjugates and thus also C. Thus $C = \overline{M}^{L,\mathrm{Galois}}$. The statement that $\overline{M}^{L,\mathrm{Galois}}$ is the fixed field of the normal subgroup $S_{\mathrm{Aut}_M(L)}$ of $G = \mathrm{Aut}_K(L)$ then follows from Theorem 4.1.41. □

Exercise 4.1.45. 1) Let G be a group and $H \subset G$ a subgroup. Show that the subset $N_H = \{g \in G : gHg^{-1} = H\}$, called the *normalizer* of H in G, is a subgroup of G containing H as a normal subgroup.

2) For a finite Galois extension $K \subset L$ and an intermediate field $K \subset M \subset L$, show that the fixed field of $N_{\mathrm{Aut}_M(L)} \subset \mathrm{Aut}_K(L)$ is the smallest intermediate field of $K \subset K' \subset M$ such that $K' \subset M$ is a Galois extension.

For a finite extension $K \subset M$, so far we have only defined the Galois closure of M over K in an extension $M \subset L$ such that $K \subset L$ is Galois. We now show that the Galois closure is, up to isomorphism, independent of the extension $M \subset L$. We first make a definition and an observation.

Definition 4.1.46. Let $K \subset M$ be a finite extension. By a *Galois extension of* $K \subset M$, or a *Galois extension of* M *over* K, we mean a finite extension $M \subset L$ such that the induced extension $K \subset L$ is Galois.

Lemma 4.1.47. *A finite extension $K \subset M$ admits a Galois extension if and only if $K \subset M$ is separable.*

Proof. Let $M \subset L$ be a Galois extension of $K \subset M$. Then the minimal polynomial of every $\alpha \in M$ over K, being also the minimal polynomial of $\alpha \in L$ over K, is separable. Thus $K \subset M$ is separable.

Conversely, assume that $K \subset M$ is separable. By the Primitive Element Theorem, $M = K(\alpha)$ for some $\alpha \in M$. Let $p \in K[x]$ be the minimal polynomial of α, so p is separable by assumption. Let L be a splitting field of p over K. Then, using any root β of p in L, we have an embedding

$$M = K(\alpha) \cong K[x]/\langle p \rangle \cong K(\beta) \subset L.$$

As $p \in K[x]$ is separable, L is a Galois extension of K by Theorem 4.1.24. Then $K \subset M \subset L$ is as required. □

For a finite separable extension $K \subset M$, the construction of a Galois extension of $K \subset M$ in the proof of Lemma 4.1.47 gives rise to another characterization of the Galois closure of M over K in any Galois extension $M \subset L$ of $K \subset M$.

Lemma 4.1.48. *Let $K \subset M$ be a finite separable extension, and let $M \subset L$ be a Galois extension of M over K.*

1) There exists $\alpha \in L$ such that $M = K(\alpha)$;

2) Let $p(x) \in K[x]$ be the minimal polynomial of α in $K[x]$. Then $\overline{M}^{L,\mathrm{Galois}}$ is the splitting field of p in L, i.e.,

$$\overline{M}^{L,\mathrm{Galois}} = K(\alpha, \alpha_2, \ldots, \alpha_n),$$

where $\{\alpha, \alpha_2, \ldots, \alpha_n\}$ is the set of all roots of p in L.

Proof. 1) By the Primitive Element Theorem, there exists $\alpha \in M \subset L$ such that $M = K(\alpha)$.

2) As $K \subset M$ is separable, p has no repeated roots in any extension of K. Let L_1 be the splitting field of p in L, i.e.,

$$L_1 = K(\alpha, \alpha_2, \ldots, \alpha_n) = K(\alpha)K(\alpha_2)\cdots K(\alpha_n) \subset L,$$

where $\{\alpha, \alpha_2, \ldots, \alpha_n\}$ is the set of all roots of p in L, and recall that $K(\alpha)K(\alpha_2)\cdots K(\alpha_n)$ denotes the compositum of the sub-fields $K(\alpha), K(\alpha_2), \ldots, K(\alpha_n)$.

On the other hand, as $K \subset L$ is a Galois extension, we know from Proposition 4.1.21 that

$$\{\alpha, \alpha_2, \ldots, \alpha_n\} = G\alpha = \{g\alpha : g \in G\},$$

where $G = \mathrm{Aut}_K(L)$. Thus the collection $\{K(\alpha), K(\alpha_2), \ldots, K(\alpha_n)\}$ is precisely that of all Galois conjugates of $M = K(\alpha)$ in L. By Proposition 4.1.44, $L_1 = \overline{M}^{L,\mathrm{Galois}}$. □

When we need to deal with two Galois extensions of a given finite separable extension $K \subset M$, we denote one by $M \subset L_1$ and the second by $\varphi : M \to L_2$, i.e., ϕ_2 is a finite extension such that $\varphi|_K : K \to L_2$ is Galois.

Proposition 4.1.49. *Let $K \subset M$ be a finite separable extension. Let $M \subset L_1$ and $\varphi : M \to L_2$ be two Galois extensions of $K \subset M$. Then there exists an isomorphism*

$$\tilde{\varphi} : \ \overline{M}^{L_1,\mathrm{Galois}} \longrightarrow \overline{M}^{L_2,\mathrm{Galois}} \quad \textit{with} \quad \tilde{\varphi}|_M = \varphi : M \to L_2.$$

Proof. Let $M_2 = \varphi(M) \subset L_2$, $K_2 = \varphi(K) \subset M_2$, and

$$C_1 = \overline{M}^{L_1,\mathrm{Galois}} \subset L_1 \quad \text{and} \quad C_2 = \overline{M_2}^{L_2,\mathrm{Galois}} \subset L_2.$$

By Lemma 4.1.48, there exists $\alpha \in M \subset L_1$ such that $M = K(\alpha)$, and that C_1 is the splitting field in L_1 over K of the minimal polynomial $p(x) \in K[x]$ of α.

Consider $\varphi(p) \in K_2[x]$, which has $\varphi(\alpha) \in M_2 \subset L_2$ as a root in L_2. Since $K_2 \subset L_2$ is a normal extension, $\varphi(p)$ splits completely in L_2. Regarding p as in $M[x]$, then C_1 is also the splitting field in L_1 over M of $p \in M[x]$. As $\varphi(p) \in M_2[x]$ completely splits over L_2, Extension Lemma says that there exists an injective homomorphism $\tilde{\varphi} : C_1 \to L_2$ which extends φ and whose image is the splitting field in L_2 over M_2 of $\varphi(p)$, which, by Lemma 4.1.48, is C_2. □

§ 7. Examples of the Galois Correspondence

Example 4.1.50. Consider $L = \mathbb{Q}(\sqrt[3]{2}, \omega)$, where $\omega = e^{(2\pi i)/3}$, as the splitting field of $f(x) = x^3 - 2 \in \mathbb{Q}[x]$ over $|Q$. Being a splitting field over $\mathbb{Q}$ which is perfect, L is a Galois extension of $\mathbb{Q}$. We know that the Galois group $G = \mathrm{Aut}(\mathbb{Q})(L)$ has order $[L : \mathbb{Q}] = 6$, and since f has exactly three roots, namely

$$r_1 = \sqrt[3]{2}, \quad r_2 = \omega\sqrt[3]{2}, \quad r_3 = \omega^2\sqrt[3]{2},$$

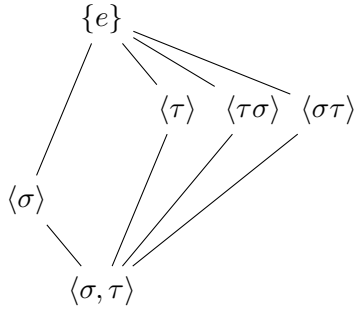
G can be identified with the permutation group S_3 of the three roots. Every $g \in G$ is determined by $g(\sqrt[3]{2}) \in L$ and $g(\omega) \in L$. Note that $g(\sqrt[3]{2})$ must be another root of f. On the other hand, since ω is a root of $g(x) = x^2 + x + 1$ and since the other root of g is ω^2, every $g \in G$ must satisfies $g(\omega) = \omega$ or $g(\omega) = \omega^2$. Define $\sigma, \tau \in G$ by

$$\sigma(\omega) = \omega, \quad \sigma(\sqrt[3]{2}) = \omega\sqrt[3]{2}, \quad \tau(\omega) = \omega^2, \quad \tau(\sqrt[3]{2}) = \sqrt[3]{2}.$$

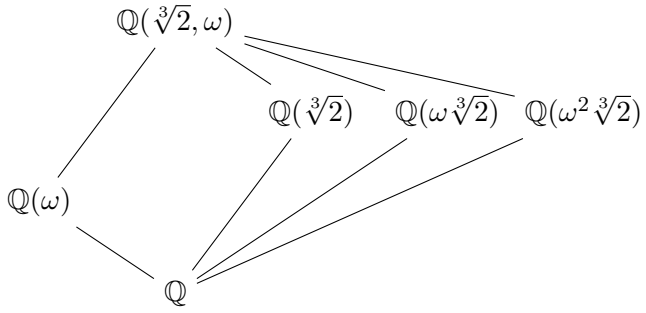
Then it is easy to see that $\sigma^3 = \tau^2 = \mathrm{Id}$. The group G is generated by σ and τ. Namely $G = \{e, \sigma, \sigma^2, \tau, \tau\sigma, \sigma\tau = \tau\sigma^2\}$, where

$$\sigma = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_2 & r_3 & r_1 \end{pmatrix},\ \tau = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_1 & r_3 & r_2 \end{pmatrix},$$
$$\sigma\tau = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_2 & r_1 & r_3 \end{pmatrix},\ \tau\sigma = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_3 & r_2 & r_1 \end{pmatrix}.$$

There are a total of 6 subgroups of $G \cong S_3$, corresponding to 6 sub-fields, given in the following diagrams:



,



Furthermore, the extensions $\mathbb{Q} \subset \mathbb{Q}$, $\mathbb{Q} \subset \mathbb{Q}(\omega)$, and $\mathbb{Q} \subset L = \mathbb{Q}(\omega, \sqrt[3]{2})$ are Galois, corresponding to the three normal subgroups $\{e\}$, $\{e, \sigma, \sigma^2\}$, and G, of G. The other three extensions

$$\mathbb{Q} \subset \mathbb{Q}(\sqrt[3]{2}), \quad \mathbb{Q} \subset \mathbb{Q}(\omega\sqrt[3]{2}), \quad \mathbb{Q} \subset \mathbb{Q}(\omega^2\sqrt[3]{2})$$

are not Galois, but they are all Galois congugates of each other, and their compositum is the whole L, so L is the Galois closure of each one of them.

Example 4.1.51. Let p be a prime number and $n \geq 1$ an integer. Consider the finite extension

$$\mathbb{F}_p \subset \mathbb{F}_{p^n}.$$

As $\mathbb{F}_{p^n}$ is the splitting field of the polynomial $f_n(x) = x^{p^n} = 1 \in \mathbb{F}_p[x]$ over $\mathbb{F}_p$ and that $\mathbb{F}_p$ is perfect, we know that the extension $\mathbb{F}_p \subset \mathbb{F}_{p^n}$ is Galois, a fact we had already seen in Example 4.1.15. We have also seen in Example 4.1.15 that the Galois group $G = \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p) = \mathrm{Aut}_{\mathbb{F}_P}(\mathbb{F}_{p^n})$ is the the cyclic group of order n generated by the Frobenius isomorphism σ.

By Lagrange's theorem, the order of every subgroup of G must divide n, and for every $m \in [1, n]$ and $m|n$, there exactly one subgroup of G of order m, namely the subgroup of G generated by σ^d, where $d = n/m$. Note that the fixed field of $\langle\sigma^d\rangle$ is precisely the sub-field $\mathbb{F}_{p^d}$ of $\mathbb{F}_{p^n}$. The Galois correspondence in this case coincides with Corollary 3.2.42 on the classification of sub-fields of $\mathbb{F}_{p^n}$.

We now turn to the cyclotomic extensions of $\mathbb{Q}$.

Example 4.1.52. Recall now that for an integer $n \geq 1$, the n'th cyclotomic field is $\mathbb{Q}(\omega_n)$ where $\omega_n = e^{2\pi i/n}$, and it is the splitting field of the polynomial $f(x) = x^n - 1$ over $\mathbb{Q}$. Thus $\mathbb{Q}(\omega_n)$ is a Galois extension of $\mathbb{Q}$, a fact we have proved already in Example 4.1.14. We have also seen in Example 4.1.14 that the Galois group $G = \mathrm{Gal}(\mathbb{Q}(\omega_n)/\mathbb{Q}) = \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\omega^n))$ is isomorphic to the group $(\mathbb{Z}/n\mathbb{Z})^\times$ of invertible elements in the commutative ring $\mathbb{Z}/n\mathbb{Z}$. More precisely, corresponding to each $k \in [1, n]$ with $(k, n) = 1$ we have $\sigma_k \in G$ given by $\sigma_k(\omega_n) = \omega_n^k$, and that the map

$$(\mathbb{Z}/n\mathbb{Z})^\times \longrightarrow G,\ \bar{k} \longmapsto \sigma_k,$$

is a group isomorphism. Thus, by the Galois Correspondence, sub-fields of $\mathbb{Q}(\omega_n)$ are in one-to-one correspondence with subgroups of $(\mathbb{Z}/n\mathbb{Z})^\times$.

As an example, consider $n = 10$ and denote $\mathbb{Z}/10\mathbb{Z} = \{\bar{j} : 0 \leq j \leq 10\}$. Then

$$(\mathbb{Z}/10\mathbb{Z})^\times = \{\bar{1},\ \bar{3},\ \bar{7},\ \bar{9}\}$$

so the Galois group $G = \mathrm{Gal}(\mathbb{Q}(\omega_{10})/\mathbb{Q})$ has 4 elements. As the order of the element $\bar{3}$ is 4, one has $G \cong \mathbb{Z}/4\mathbb{Z}$ as a group. Note that the element $\bar{7}$ also has order 4, and the element $\bar{9}$ has order 2. Thus other than the trivial subgroup $\{\mathrm{Id}\}$ and the whole group, G has one other subgroup which is of order 2, namely $H = \{\mathrm{Id}, \sigma_9\}$. To see what the intermediate field $F(H)$ is, we first note that the element ω_{10} satisfies $\omega_{10}^5 = -1$, so its minimal polynomial is

$$\Phi_{10} = \frac{x^5 + 1}{x + 1} = x^4 - x^3 + x^2 - x + 1.$$

Writing an arbitrary element of $\mathbb{Q}(\omega_{10})$ as $a = a_0 + a_1\omega_{10} + a_2\omega_{10}^2 + a_3\omega_{10}^3$ with $a_0, \ldots, a_3 \in \mathbb{Q}$, and using $\Phi_{10}(\omega_{10}) = 0$, we have

$$\begin{aligned}
&\sigma_9(\omega_{10}) = \omega_{10}^{-1} = -\omega_{10}^3 + \omega_{10}^2 - \omega_{10} + 1,\\
&\sigma_9(\omega_{10}^2) = \omega_{10}^{18} = \omega_{10}^{-2} = -\omega_{10}^3, \quad \sigma_9(\omega_{10}^3) = \omega_{10}^{27} = \omega_{10}^{-3} = -\omega_{10}^2.
\end{aligned}$$

It follows that

$$\sigma_9(a) = a_0 - a_1(\omega_{10}^3 - \omega_{10}^2 + \omega_{10} - 1) - a_2\omega_{10}^3 - a_3\omega_{10}^2.$$

Hence $\sigma_9(a) = a$ if and only if $a = a_0 + a_2(\omega_{10}^2 - \omega_{10}^3)$. Thus

$$F(H) = \mathbb{Q}(\omega_{10}^2 - \omega_{10}^3) = \mathbb{Q}(\cos(2\pi/5)).$$

Note that the element $\alpha = \omega_{10}^2 - \omega_{10}^3$ satisfies $\alpha^2 + \alpha - 1 = 0$, so indeed, $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 2$.

As another example, when p is a prime number, we know by Proposition 3.2.43 that $\mathrm{Gal}(\mathbb{Q}(\omega_p)/\mathbb{Q}) \cong \mathbb{Z}/(p-1)\mathbb{Z}$. For example, when $p = 7$, $\mathrm{Gal}(\mathbb{Q}(\omega_7)/\mathbb{Q}) \cong \mathbb{Z}/6\mathbb{Z}$.

Every non-trivial subgroup proper of $\mathbb{Z}/6\mathbb{Z}$ must have order 2 or 3. On the other hand, $\bar{2}, \bar{4} \in \mathbb{Z}/6\mathbb{Z}$ have order 3, and $\bar{3}$ has order 2, so we have two non-trivial proper subgroups

$$H_2 = \{\bar{0}, \bar{2}, \bar{4}\} \quad \text{and} \quad H_3 = \{\bar{0}, \bar{3}\}.$$

We leave it as an exercise to compute the intermediate fields M_2 and M_3 corresponding respectively to H_2 and H_3.

4.2 Solvability by radicals

§ 1. Radical extensions

Definition 4.2.1. (1) Let m be a positive integer. A field extension L of K is called a *pure extension of type* m if $L = K(a)$ for some $a \in L$ such that $a^m \in K$. A tower of fields

$$K = L_0 \subset L_1 \subset \cdots \subset L_n$$

is called a *radical tower* if each L_{j+1} is a pure extension of L_j. In this case, we call L_n a *radical extension* of L_0.

(2) Let K be a field. A polynomial $f(x) \in K[x]$ is said to be *solvable by radicals over* K if there exists a radical extension L over K in which f splits completely.

Example 4.2.2. For $f(x) = x^2 + bx + c \in \mathbb{C}[x]$, let

$$K = \mathbb{Q}(b, c), \quad L_1 = K(\sqrt{b^2 - 4c}).$$

Then L_1 is a pure extension over K of type 2, and L_1 is a splitting field of $f(x)$ over K. Therefore, $f(x)$ is solvable by radicals over K.

Example 4.2.3. For $f(x) = x^3 + px + q \in \mathbb{C}[x]$, recall that cubic formula says that the roots of $f(x)$ are

$$y + z, \quad \omega y + \omega^2 z, \quad \omega^2 y + \omega z,$$

where $\omega = e^{\frac{2\pi i}{3}}$, y is one solution of

$$y^3 = \frac{1}{2}\left(-q + \sqrt{q^2 + \frac{4p^3}{27}}\right),$$

and $yz = -p/3$. Let $K = \mathbb{Q}(p, q)$. Then $f \in K[x]$. Let

$$L_1 = K\left(\sqrt{q^2 + \frac{4p^3}{27}}\right), \quad L_2 = L_1(y), \quad L_3 = L_2(\omega),$$

so we have the tower of extensions

$$K \subset L_1 \subset L_2 \subset L_3. \tag{4.1}$$

Note that L_1 is a pure extension of K of type 2, L_2 is a pure extension of L_1 of type 3, and L_3 is a pure extension of L_2 of type 3. Thus (4.1) is a radical tower, and L_3 is a radical extension of K. Furthermore, f splits completely in L_3. Thus f is solvable by radicals over K.

Example 4.2.4. (**Roots of a quartic polynomial**) The following fact was also found in the middle of the 16th century. Let

$$f(x) = x^4 + qx^2 + rx + s \in \mathbb{C}[x]$$

be a reduced quartic polynomial, and regard f as a polynomial over $K = \mathbb{Q}[q, r, s]$. Assume that $r \neq 0$ because other wise can solve for x^2 first. We will look for k, l, and m in $\mathbb{C}$ such that

$$x^4 + qx^2 + rx + s = (x^2 + kx + l)(x^2 - kx + m).$$

If we can find k, l and m, then we use the formula for quadratic polynomials to find roots for f. Now the conditions for k, l and m are

$$\begin{cases} l + m - k^2 = q \\ k(m - l) = r \\ lm = s. \end{cases}$$

The first two equations give

$$\begin{cases} 2m = k^2 + q + r/k \\ 2l = k^2 + q - r/k. \end{cases}$$

Substituting these values of l and m into the third equation gives

$$k^6 + 2qk^4 + (q^2 - 4s)k^2 - r^2 = 0.$$

This is cubic in k^2. Solve k^2 using the cubic formula and then solve for l and m. Finally solve for x from the quadratic equations

$$x^2 + kx + l = 0 \quad \text{or} \quad x^2 - kx + m = 0.$$

Let $L_0 = K = \mathbb{Q}(q, r, s)$. Solving k^2 as a root of $x^3 + 2qx^2 + (q^2 - 4s)x - r^2 \in K[x]$ in $\mathbb{C}$, we know that there is a radical tower

$$L_0 = K \subset L_1 \subset L_2 \subset L_3$$

with $k^2 \in L_3$. Let $L_4 = L_3(k)$, so we have the radical tower

$$L_0 = K \subset L_1 \subset L_2 \subset L_3 \subset L_4.$$

Solving m and l from

$$2m = k^2 + q + \frac{r}{k}, \quad 2l = k^2 + q - \frac{r}{k},$$

we see that L_4 contains l and m. Now because

$$f(x) = (x^2 + kx + l)(x^2 - kx + m),$$

if we take $L_5 = L_4(\sqrt{k^2 - 4l})$ and $L_6 = L_5(\sqrt{k^2 - 4m})$, then we get a radical tower

$$L_0 = K \subset L_1 \subset L_2 \subset L_3 \subset L_4 \subset L_5 \subset L_6,$$

and f splits completely in L_6. Thus $f(x)$ is solvable by radicals over K.

Example 4.2.5. For example, for $f(x) = x^4 - 2x^2 + 8x - 3$, the above method gives

$$k^6 - 4k^4 + 16k^2 - 64 = 0 \quad \text{or} \quad (k^2 - 4)(k^4 + 16) = 0.$$

Take $k = \pm 2$. Then $l = -1$ and $m = 3$ if $k = 2$ and $l = 3$ and $m = -1$ if $k = -2$. Thus we get the four solutions to $f(x) = 0$:

$$-1 + i\sqrt{2}, \quad -1 - i\sqrt{2}, \quad 1 + i\sqrt{2}, \quad 1 - i\sqrt{2}.$$

We summarize the above discussions on quadratic, cubic, and quartic polynomials in the following proposition.

Proposition 4.2.6. *Every non-zero*

$$f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in \mathbb{C}[x]$$

of degree $n \leq 4$ is solvable by radicals over $K = \mathbb{Q}(a_0, \ldots, a_{n-1})$.

We now turn to polynomials of degree $n \geq 5$. Galois' theory will say that not every polynomial of degree 5 is solvable by radicals.

§ 2. Galois' great theorem

We recall the notion of solvable groups.

Definition 4.2.7. A group is G is solvable if there exists a finite sequence

$$\{e\} \subset G_n \subset G_{n-1} \subset \cdots \subset G_1 \subset G_0 = G$$

such that each G_j is a normal subgroup of G_{j-1} and that G_{j-1}/G_j is abelian. Such a series is also called a normal series.

Clearly an abelian group is solvable. One can prove the following from the definition.

Lemma 4.2.8. *A subgroup of a solvable group is solvable; a quotient of a solvable group is solvable.*

For an integer $n \geq 1$, let S_n be the permutation group on the set of n letters, and let A_n be the subgroup of S_n consisting of even permutations.

Theorem 4.2.9. *S_n is solvable for $n \leq 4$ and not solvable for $n \geq 5$.*

Proof. Since S_m is a subgroup of S_n for $m < n$, and since subgroups of a solvable group is solvable, it is enough to show that S_4 is solvable while S_5 is not.

For the group S_4, consider the following series

$$\{1\} \subset V \subset A_4 \subset S_4$$

where $V \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ consists of the four elements

$$1, \quad (12)(34), \quad (13)(24), \quad (14)(23).$$

It is a normal series. So S_4 is abelian.

To see that S_5 is not solvable, we only need to show that A_5 is not solvable, since every subgroup of a solvable group has to be solvable. It will be enough to show that A_5 has no non-trivial normal subgroup, i.e., A_5 is simple. We first note that if H is a normal subgroup of A_5, it will be a union of conjugacy classes. One first check that the class equation for A_5 is

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

Since no partial sum of the right hand side divides 60 except for 1 and 60, so $|H| = 1$ or 60. □

Theorem 4.2.10. *[Galois' Great Theorem] Let K be a field of characteristic 0. Then a non-zero $f(x) \in K[x]$ is solvable by radicals if and only if $\mathrm{Gal}_K(f)$ is a solvable group.*

Example 4.2.11. We have seen in Example 4.1.27 that the Galois group of $f(x) = x^5 - 4x + 2 \in \mathbb{Q}[x]$ over $\mathbb{Q}$ is S_5. As S_5 is not solvable, f is not solvable by radicals over $\mathbb{Q}$.

Example 4.2.12. Consider $f(x) = x^5 - 6x + 3 \in \mathbb{Q}[x]$ and let L be its splitting field over $\mathbb{Q}$ in $\mathbb{C}$. By Eisenstein's criterion, f is irreducible over $\mathbb{Q}$, so it has 5 distinct roots, so $[L : \mathbb{Q}]$ is divisible by 5, and the Galois group $G = \mathrm{Gal}_{\mathbb{Q}}(L)$ is a subgroup of S_5 of order divisible by 5. By Sylow's Theorem, G has a subgroup of order 5 and

thus contains an element of order 5, which must then be a 5-cycle. On the other hand,

$$f(-2) = -17, \quad f(0) = 3, \quad f(1) = -2, \quad f(2) = 23,$$

so f has at least real roots. If all the five roots of f are real, then $f'(x) = 5x^4 - 6$ would have at least three real roots which is not the case. Thus f has two complex roots which are conjugate to each other. Let τ be the complex conjugation on $\mathbb{C}$. Then τ restricts to an order two element in G. Again if a subgroup of S_5 contains a 5-cycle and a transposition, it must be all of S_5. We thus conclude that the Galois group of the polynomial $f(x) = x^5 - 6x + 3 \in \mathbb{Q}[x]$ over $\mathbb{Q}$ is S_5. Since S_5 is not solvable, f is not solvable by radicals over $\mathbb{Q}$.

The same arguments in the proofs of the two examples in Example 4.2.11 and Example 4.2.12 lead to the following general statement.

Lemma 4.2.13. *Assume that an irreducible quintic (i.e., order* 5*) polynomial* $f(x) \in \mathbb{Q}[x]$ *has three distinct real roots and two non-real roots in* $\mathbb{C}$*. Then the Galois group* $\mathrm{Gal}_{\mathbb{Q}}(f)$ *of* f *is isomorphic to* S_5*, and thus* f *is not solvable by radicals over* $\mathbb{Q}$*.*