Review of modern algebra

October 16, 2022

Contents

Preface

This note is written to help anyone who reviews undergraduate algebra. Thus, it is highly recommended to study undergraduate algebra at least once before reading this note, for most of details will not be introduced in this note. Also, some sections are not written yet, because I haven't studied them.

Part I

Structured objects

Chapter 1

Basic object theory

1.1 Structured objects in modern algebra

A structured object, or just an object, in short, is defined as a set X with operators on X such that the operators are 'compatible' and X satisfies some further properties.[1]

Definition 1.1.1 (Group)**.** A set G with an operator $\cdot$ on G is called a group if

(G1) (Compatibility) the operator is associative, i.e., $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in G$

and G satisfies the following further properties:

(G2) (The identity) There is an element e of G such that $e \cdot x = x \cdot e = x$ for all $x \in G$.

(G3) (The inverse) For each $g \in G$, there is an element $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$.

It is left as an exercise to explain why the identity of G and an inverse of a given element of G are unique.

Definition 1.1.2 (Ring)**.** A set R with two operators $+$ and $\times$ are, or to be precise, the tuple $(R, +, \times)$, is called a ring if

(R1) (Compatibility) Both $+$ and $\times$ are associative and $+$ and $\times$ are distributive

and R satisfies the following further property:

(R2) $(R, +)$ is an abelian group.

Remark. Assume a ring R has the multiplicative identity 1. (Check the uniqueness of the multiplicative identity.) If $1 = 0$ in a ring R, then $x = x \times 1 = x \times (1 + 0) = x \times (1 + 1) = x + x$ for all $x \in R$, so $x = 0$ for all $x \in R$.

Definition 1.1.3. (a) A ring with a commutative multiplication is called a commutative ring.

(b) A commutative ring R is called an integral domain (or a domain, in short), if $ab = 0$ implies $a = 0$ or $b = 0$.

(c) A ring with identity in which every nonzero element is a unit is called a division ring. Here, a nonzero element r in a ring R is called a unit if there is an element $u \in R$ such that $ru = ur = 1$.Check the uniqueness of the multiplicative inverse of a unit.

(d) A commutative division ring is called a field.

Definition 1.1.4 (R-module)**.** Assume that R is a ring with identity and a set M equips an addition and an R-scalar multiplication. The set M, together with R and all the operators, is called a left R-module if

[1] By operators we assume that they are 'compatible' and binary.

(M1) (Compatibility) All operators are pairwise compatible, e.g., associative, distributive,

and M satisfies the following further property:

(M2) $(M, +)$ is an abelian group.

Remark. If R is a division ring, then an R-module is called an R-vector space.

Definition 1.1.5 (R-algebra)**.** Assume that R is a ring with identity and a set A equips an addition, an R-scalar multiplication, and a multiplication. The set A, together with R and all the operators, is called an R-algebra if

(A1) (Compatibility) All operators are pairwise compatible, e.g., associative, distributive,

and A satisfies the following further properties:

(A2) A is a ring.

(A3) A is an R-module.

1.2 Subobjects

A subobject of an object X, or to be precise, a sub-□□ of a □□ X is a subset of X which is □□.

Example 1.2.1 (Subobject tests)**.** The subobject tests for groups, rings, R-modules, and R-algebras are listed in this example. Proving equivalences is left as an exercise.

(a) (Subgroup) Let G be a group. Then H is a subgroup of G if and only if $ab^{-1} \in H$ for all $a, b \in H$.

(b) (Subring) Let R be a ring. Then S is a subring of R if and only if S is a subgroup of R and S is closed under multiplication, i.e., $a - b, ab \in S$ for all $a, b \in S$.

(c) (R-submodule) Let R be a ring with identity and M be an R-module. Then N is an R-submodule of M if and only if N is closed under addition and R-scalar multiplication, i.e., $a + b, ca \in N$ for all $a, b \in N$ and $c \in R$.

(d) (R-subalgebra) Let R be a ring with identity and A be an R-algebra. Then B is an R-subalgebra of A if and only if B is both an R-submodule and a subring of A, i.e., B is closed under addition, R-scalar multiplication, and multiplication.

Proposition 1.2.2. If $Y_i \leq X$ for each i in the index set I, then the intersection of Y_i's is the largest subobject of X contained in each Y_i.

Definition 1.2.3 (Generated subobject)**.** Let X be an object □□ and E be a subset of X. The sub-□□ $\langle E \rangle$ generated by E is defined by the intersection of all sub-□□'s of X containing E.

Remark. One can easily check that the subobject of X generated by the subset E is the unique smallest subobject of X containing E.

A subobject generated by a single element is called a cyclic subobject.

Example 1.2.4 (Prime subfield)**.** Given a field F, the prime subfield E of F is the subfield of F generated by the identity. If $\text{char}(F) = 0$, then the prime subfield of F is isomorphic to $\mathbb{Q}$; if $\text{char}(F) = p > 0$, then the prime subfield of F is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ (or $\mathbb{F}_p$).

1.3 Quotient objects

Given a □□ X and its sub-□□ Y, we want the quotient X/Y a □□.

Example 1.3.1. (a) (For groups) Suppose N is a subgroup of G and we want to impose G/N a group structure, by defining

$$(aN)\cdot(bN)=(ab)N. \tag{1.1}$$

Assuming that the operation is well-defined, one can easily verify the group axioms. Hence, G/N, together with the operation defined in eq. (1.1), is a group if and only if the operation is well-defined.

The statement that the operation is well defined is equivalent to the following statement:

$$\text{If } aN=xN \text{ and } bN=yN, \text{ then } (ab)N=(xy)N. \text{ (Here, } a,b,x,y\in G.)$$

(1) Suppose the operation is well defined. Then $(ab)(xy)^{-1}\in N$, but we also have $(ab)(xy)^{-1}=aby^{-1}x^{-1}=ax^{-1}\cdot x\cdot by^{-1}\cdot x^{-1}$ and $x\cdot by^{-1}\cdot x^{-1}\in N$. Therefore, for any $g\in G$ and $n\in N$, we have $gng^{-1}\in N$.

(2) Suppose conversely that $gng^{-1}\in N$ for all $g\in G$ and $n\in N$. Whenever $aN=xN$ and $bN=yN$, because $ax^{-1},by^{-1}\in N$, we have $(ab)(xy)^{-1}=aby^{-1}x^{-1}=ax^{-1}\cdot(x\cdot by^{-1}\cdot x^{-1})\in N$, so $(ab)N=(xy)N$.

By (1) and (2), we can derive the following conclusion:

$$G/N \text{ is a group if and only if } gng^{-1}\in N \text{ for all } g\in G \text{ and } n\in N.$$

We call such subgroup N a normal subgroup of G.

(b) (For rings) Suppose I is a subring of R and we want to impose R/I a ring structure, by defining

$$(aI)+(bI)=(a+b)I \quad \text{and} \quad (aI)(bI)=(ab)I. \tag{1.2}$$

Assuming that the operations are well-defined, one can easily verify the ring axioms. Hence, R/I, together with the operations defined in equation eq. (1.2), is a ring if and only if the operations are well-defined. Since the addition is well-defined if and only if R/I is a group, the addition is well-defined if and only if I is a subgroup of R. We thus assume that I is a subgroup of R.

The statement that the multiplication is well defined is equivalent to the following statement:

$$\text{If } aI=xI \text{ and } bI=yI, \text{ then } (ab)I=(xy)I. \text{ (Here, } a,b,x,y\in R.)$$

Or equivalently,

$$\text{If } a-x\in I \text{ and } b-y\in I, \text{ then } ab-xy\in I.$$

Writing $a-x=s$ and $b-y=t$ $(s,t\in I)$, we can easily find that $ab-xy\in I$ if and only if $xt+sy\in I$. Hence, the multiplication is well defined if and only if $IR,RI\subset I$. To summarize,

$$R/I \text{ is a ring if and only if } I \text{ is a subring of } R \text{ and } RI,IR\subset I.$$

We call such subring I an ideal of R.

(c) (For R-modules) One can easily find that M/N is an R-module whenever N is an R-submodule of M.

(d) (For R-algebras) Suppose I is an R-subalgebra of an R-algebra A. If I is an ideal of the ring A, then the quotient ring A/I is well-defined; because I is an R-subalgebra of A, it is an R-submodule of A, so the quotient R-module A/I is also well-defined. Therefore, R/I is a well-defined R-algebra if and only if I is an R-subalgebra of A which is an ideal of the ring A.

1.4 Isomorphism theorems

In this section, we only consider groups, rings, R-modules, and R-algebras as objects, where R is a ring with identity. And we understand $Y \trianglelefteq X$ as the assumption that the quotient object X/Y is well-defined.

Before introducing isomorphism theorems, we first introduce a powerful lemma which is used to check well-definedness of maps defined on quotient objects.

Lemma 1.4.1. Suppose $\phi : X \to Y$ is a $\square\square$-homomorphism and $X' \trianglelefteq X$ and $Y' \trianglelefteq Y$. When trying to define a map $\overline{\phi} : X/X' \to Y/Y'$ by $\overline{\phi}\overline{x} := \overline{\phi x}\,(x \in X)$,

(a) the followings are equivalent:

(1) $\overline{\phi}$ is a well-defined $\square\square$-homomorphism.

(2) $\phi X' \leq Y'$, i.e., $X' \leq \phi^{-1}Y'$.

(b) Also, the followings are also equivalent:

(3) $\overline{\phi}$ is a well-defined $\square\square$-monomorphism.

(4) $X' = \phi^{-1}Y'$.

$$\begin{array}{ccc} X & \xrightarrow{\phi} & Y \\ {\scriptstyle \pi_1}\downarrow & & \downarrow{\scriptstyle \pi_2} \\ X/X' & \xrightarrow[\overline{\phi}]{} & Y/Y' \end{array}$$

Proof. We first prove the equivalence in (a). First, assume (1). Then, whenever $\overline{x} = \overline{y}$ we should have $\overline{\phi x} = \overline{\phi y}$, i.e., $\phi(xy^{-1}) \in Y'$. Hence, $\phi X' \leq Y'$, implying (2). Conversely, assume (2). If $\overline{x} = \overline{y}$ so that $xy^{-1} \in X'$, then $(\phi x)(\phi y)^{-1} \in Y'$ so $\overline{\phi x} = \overline{\phi y}$, implying (1).

We now prove (b). It is helpful to note that we need to further check injectivity. Assuming (3), (2) is automaticaly implied; if $x \in \phi^{-1}Y'$ so that $\overline{\phi}\overline{x} = \overline{\phi x} = \overline{0}$, because $\overline{\phi}$ is a monomorphism, we have $\overline{x} = \overline{0}$ and $x \in X'$, implying (4). Conversely, assuming (4), (1) is automaticaly implied; if $\overline{\phi}\overline{x} = \overline{0}$, then $\phi x \in Y'$, which implies $x \in \phi^{-1}Y' = X'$ and $\overline{x} = \overline{0}$, implying (3). □

The first isomorphism theorem is widely used.

Theorem 1.4.2 (First isomorphism theorem)**.** If $\phi : X \to Y$ is a $\square\square$-homomorphism and if one tries to define a map $\overline{\phi} : X/\mathit{ker}\,\phi \to \mathit{im}\,\phi$ by $\overline{\phi}\overline{x} = \phi x\,(x \in X)$, then $\overline{\phi}$ is a well-defined $\square\square$-isomorphism.

Proof. By Lemma 1.4.1, $\overline{\phi}$ is a well-defined monomorphism, since $\mathit{ker}\,\phi = \phi^{-1}0$. Surjectivity is clear by definition, so $\overline{\phi}$ is a $\square\square$-isomorphism. □

We now introduce the second isomorphism theorem, which is rarely applied. Before introducing the second isomorphism theorem, we introduce a temporal notation.

Notation. Let $A, B \leq X$. For a group X, we understand $A * B = AB$; for a ring, an R-module, and an R-algebra X, we understand $A * B = A + B$.

Proposition 1.4.3. If $Y \trianglelefteq X$ and $Z \leq X$, then

(a) the restriction of Y to Z is normal in Z, i.e., $Y \cap Z \trianglelefteq Z$.

(b) $Y \trianglelefteq Y * Z \leq X$.

Theorem 1.4.4 (Second isomorphism theorem)**.** Suppose $Y \trianglelefteq X$ and $Z \leq X$. The map

$$\overline{\jmath} : \frac{Z}{Y \cap Z} \to \frac{Y * Z}{Y}$$

defined by $\overline{\jmath}(\overline{z}) := \overline{z}\,(\overline{z} \in Z/(Y \cap Z))$ is a well-defined $\square\square$-isomorphism.

Proof. You should prove this isomorphism theorem for each object. □

Theorem 1.4.5 (Third isomorphism theorem)**.** Suppose $Y, Z \trianglelefteq X$ and $Z \leq Y$. If one tries to define a map

$$\phi : \frac{X}{Y} \to \frac{X/Z}{Y/Z}$$

by $\phi(\overline{x}) = \overline{\overline{x}}\,(\overline{x} \in X/Y)$ (be careful when interpreting overlines), then ϕ is a well-defined □□-isomorphism.

Proof. By hypothesis $Z \leq Y$, it is clear that $\overline{x} = \overline{y}$ in X/Y implies $xy^{-1} \in Y$ so that $\phi(x) = \phi(y)$. Both injectivity and surjectivity easily follows from definition, and it can also be easily checked that ϕ is a group homomorphism. (Check them.) Therefore, ϕ is a group isomorphism. □

Theorem 1.4.6 (Fourth isomorphism theorem; lattice isomorphism theorem)**.** Suppose $Y \trianglelefteq X$. Then the subobject lattice of X containing Y and the subobject lattice of X/Y are in bijection. Moreover, for example, when X is a group, indices, normality, inclusion, are also preserved.

Proof. You should prove this isomorphism theorem for each object. □

1.4.1 The fourth isomorphism theorem for groups and its proof

Theorem 1.4.7. Suppose G is a group and N is a normal subgroup of G. Then there is a bijection from the set of subgroups A of G which contain N onto the set of subgroups $\overline{A} = A/N$ of G/N. In particular, every subgroup of $\overline{G}$ is of the form A/N for some subgroup A of G containing N (namely, its preimage in G under the natural projection homomorphism from G to G/N). This bijection has the following properties: For all $A, B \leq G$ with $N \leq A$ and $N \leq B$,

(a) $A \leq B$ if and only if $\overline{A} \leq \overline{B}$.

(b) if $A \leq B$, then $[B : A] = [\overline{B} : \overline{A}]$.

(c) $\overline{\langle A, B\rangle} = \langle \overline{A}, \overline{B}\rangle$.

(d) $\overline{A \cap B} = \overline{A} \cap \overline{B}$.

(e) $A \trianglelefteq G$ if and only if $\overline{A} \trianglelefteq \overline{G}$.

Proof. Let $\pi : G \to G/N$ denote the canonical projection epimorphism. Becasue π is a group homomorphism, if A is a subgroup of G containing N, then $\pi(A) = A/N$ is a subgroup of G/N; conversely, if A/N is a subgroup of G/N, then its preimage $\pi^{-1}(G/N)$ is a subgroup of G containing N, because $N = \pi^{-1}(N/N) \subset \pi^{-1}(A/N)$. (Show that the images (and preimages) of subobjects under homomorphisms are objects.) Thus, the map π induces the map π_* from the collection of the subgroups of G containing N into the set of subgroups of G/N.

We now prove the injectivity of π_*. Assume A, B are subgroups of G containing N such that $\pi_*(A) = \pi_*(B)$. Because $A = \pi_*^{-1}(\pi_*(A))$ (why?), we easily obtain that $A = B$. Surjectivity is clear, because preimages of subgroups are subgroups of the domain. Therefore, π_* is a desired bijection.

We now prove remaining algebraic properties. Let A and B be subgroups of G containing N.

(a) Because π_* is order-preserving map, the result is clear.

(b) This result follows from the third isomorphism theorem.

(c) First, since $A, B \leq \langle A, B\rangle$, we have $\overline{A}, \overline{B} \leq \overline{\langle A, B\rangle}$ and $\langle \overline{A}, \overline{B}\rangle \leq \overline{\langle A, B\rangle}$. Conversely, because every element of $\overline{\langle A, B\rangle}$ turns out to be the coset by a word in $\langle A, B\rangle$, it is found that $\overline{\langle A, B\rangle} \leq \langle \overline{A}, \overline{B}\rangle$. (The converse could also be proved by applying that $\langle A, B\rangle$ is the unique smallest subgroup of G contined in any subgroup L of G contained in both A and B so $\overline{\langle A, B\rangle}$ is such subgroup of G/N contained in both $\overline{A}$ and $\overline{B}$.)

(d) If $\overline{x} \in \overline{A \cap B}$, then $x \in A \cap B$ so $\overline{x} \in \overline{A} \cap \overline{B}$. Converse can be similarly proved.

(e) It is almost clear that $\overline{A} \trianglelefteq \overline{G}$ if $A \trianglelefteq G$. (Check it.) Suppose conversely that $\overline{A} \trianglelefteq \overline{G}$ and assume $g \in G$ and $a \in A$. Because $\overline{gag^{-1}} \in \overline{A}$, we have $gag^{-1} \in A$, so $A \trianglelefteq G$. (Easy.)

This concludes the proof of the lattice isomorphism theorem for groups. As the name suggests, one can almost freely identify lattice diagrams if one is given as the diagram of a quotient. □

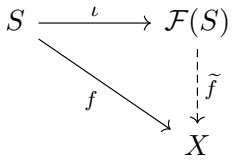
Chapter 2

Free objects

2.1 Remarks on free objects

Definition 2.1.1 (Free object)**.** For a nonempty set S, the pair $(\mathcal{F}(S), \iota : S \to \mathcal{F}(S))$ is called a free $\square\square$ generated by S if it satisfies the following universal property:

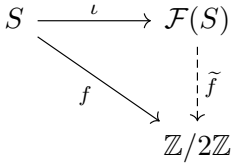
For any $\square\square$ X and a map $f : S \to X$, there is a unique $\square\square$-homomorphism $\widetilde{f} : \mathcal{F}(S) \to X$ such that $\widetilde{f} \circ \iota = f$.



Remark. Note that the map $\iota : S \to \mathcal{F}(S)$ in the definition of the free-$\square\square$ isn't mentioned as an injection. However, one can easily prove that ι is necessarily an injection.

Proposition 2.1.2. If $(\mathcal{F}(S), \iota)$ is a free-$\square\square$ generated by S, then ι is an injection, hence we may identify S as a subset of $\mathcal{F}(S)$.

Proof. Fix a point $x \in S$ and define a map $f : S \to \mathbb{Z}/2\mathbb{Z}$ by $f(x) = \overline{1}$ and $f(y) = \overline{0}$ for all $y \in S \setminus \{x\}$. The main idea of this proof is that the space $\mathbb{Z}/2\mathbb{Z}$ is a $\square\square$.



Since $f(x) \neq f(y)$ whenever $y \in X \setminus \{x\}$, we have $\iota x \neq \iota y$, as desired. □

Also, since free objects are defined in terms of a universal property, a free-$\square\square$ generated by S is unique up to $\square\square$-isomorphism.

Proposition 2.1.3. If $(\mathcal{F}_1, \iota_1)$ and $(\mathcal{F}_2, \iota_2)$ are free-$\square\square$ generated by S, then $\mathcal{F}_1 \approx_{\square\square} \mathcal{F}_2$.

Finally, as expected, the set of words with alphabets in S is a free-$\square\square$ generated by S.

Proposition 2.1.4. $\langle S \rangle$ is a free-$\square\square$ generated by S. Hence, we may identify $\langle S \rangle$ and $\mathcal{F}(S)$.

In particular, the above proposition implies the existence of a free-$\square\square$ generated by a nonempty set S.

Proposition 2.1.5. A $\square\square$ is a homomorphic image of a free-$\square\square$.

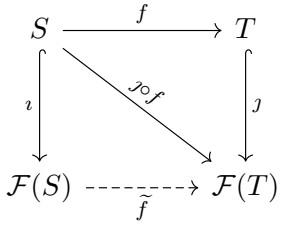
Proof. Given a $\square\square$ X generated by a set S, let $\mathcal{F}$ be the free-$\square\square$ generated by S and let $\imath : S \hookrightarrow \mathcal{F}$ be the natural embedding. Then there is a unique $\square\square$-homomorphism $\jmath_* : \mathcal{F} \to X$ extending the inclusion map $\jmath : S \hookrightarrow X$, which is necessarily a surjection. □

Remark. Indeed, $X \approx \mathcal{F}/ker\, \jmath_*$. In other words, X can be obtained from the free-$\square\square$ generated by X by declaring all elements in $ker\, \jmath_*$ zero.

One question may have been in one's mind as soon as we started this chapter:

$$\text{Are } \mathcal{F}(S) \text{ and } \mathcal{F}(T) \text{ isomorphic if } |S| = |T|?$$

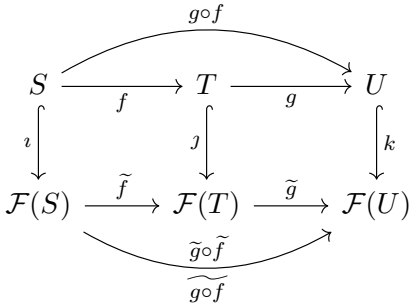
Observation 2.1.6. Let S, T be nonempty sets, $(\mathcal{F}(S), \imath)$ be the free-$\square\square$ generated by S, and $(\mathcal{F}(T), \jmath)$ be the free-$\square\square$ generated by T. Given a map $f : S \to T$, there is a unique $\square\square$-homomorphism $\widetilde{f} : \mathcal{F}(S) \to \mathcal{F}(T)$ such that $\widetilde{f} \circ \imath = \jmath \circ f$.



Given nonempty sets S, T, U, let the free $\square\square$ generated by these sets be denoted by $(\mathcal{F}(S), \imath)$, $(\mathcal{F}(T), \jmath)$, $(\mathcal{F}(U), k)$, respectively.

(a) By letting $S = T$ and $f : S \to T$ be the identity map and checking that the above diagram commutes if $\widetilde{f}$ is the identity map on $\mathcal{F}(S)$, it can be deduced that $\widetilde{id_S} = id_{\mathcal{F}(S)}$.

(b) If $f : S \to T$ and $g : T \to U$ are maps, then $\widetilde{g \circ f} = \widetilde{g} \circ \widetilde{f}$.



Now we can answer to the above question. Assume as in (b) and let $f : S \to T$ is a bijection. By letting $g = f^{-1}$, because $\widetilde{g} \circ \widetilde{f} = \widetilde{g \circ f} = id_{\mathcal{F}(S)}$ and $\widetilde{f} \circ \widetilde{g} = \widetilde{f \circ g} = id_{\mathcal{F}(T)}$, $\widetilde{f}$ denotes a $\square\square$-isomorphism of $\mathcal{F}(S)$ into $\mathcal{F}(T)$.

So far, we have found that the free objects generated by S and T are isomorphic if there is a bijection between S and T. Then how about its converse? In other words, is there a bijection between S and T if the free objects generated by S and T are isomorphic? For free groups and modules, this question will be answered when we study linear algebra over principal ideal domains.

Part II

Group theory

Chapter 3

Basic group theory

3.1 Quotient group

Imposing the quotient by a subgroup H of G gives a coset space, and the coset space is indeed a partition of G.

Theorem 3.1.1 (Lagrange's theorem)**.** Suppose G is a finite group and let H be a subgroup of G. Then $|H|$ divides $|G|$ and $|G| = |G/H||H|$.

Proof. Imposing a quotient by a subgroup gives a partition of the group. □

Remark. Clearly, partitions of a partition give a partition, from which some other results can be derived. For example, if $H \leq K \leq G$ and G/K and K/H are finite, then $[G : H] = [G : K][K : H]$.

Motivation of normal subgroups is introduced in the first part of this note. Thus, in the remaining of this section, further properties of normal subgroups are introduced.

We first start with kind of trivial statements, regarding the structures of quotient groups. Proving the following propositions is left as an exercise.

Proposition 3.1.2. Let G be an abelian group. Then every subgroup of G is naturally a normal subgroup and every quotient group is abelian. Furthermore, if G is cyclic, then every quotient group is cyclic.

Proposition 3.1.3 (Restriction of normality)**.** Suppose $H \leq K \leq G$. If H is a normal subgroup of G, then H is normal in K.

Proposition 3.1.4 (Normality of restriction)**.** Let K be a subgroup of a group G. If N is a normal subgroup of G, then the restriction of N to K, i.e., $N \cap K$, is normal in K.

Proposition 3.1.5. Let G be a group.

(a) The center $Z(G)$ of G is a normal subgroup of G.

(b) If $\phi \in \mathrm{Aut}(()G)$, then $\phi(Z(G)) = Z(G)$.

The following lemma is not about a normal subgroup, but it helps in some other situations.

Lemma 3.1.6. If H and K are finite subgroups of G, then

$$|HK| = \frac{|H||K|}{|H \cap K|}.$$

Proof. Note that $HK = \bigcup_{h \in H} hK$ so $|HK|$ is divisible by $|K|$. Given $a \in H$ and $b \in K$, assume $hk = ab$ for some $h \in H, k \in K$. Then $a^{-1}h = bk^{-1} \in H \cap K$ and vice versa, and there are precisely $|H \cap K|$ choices for (h, k). □

We now introduce a criterion for determining if the composite of two subgroups is a subgroup.

Proposition 3.1.7. Let H and K be subgroups of G. Then HK is a subgroup of G if and only if H and K commute, i.e., $HK = KH$. Hence, in particular, if $H \leq N_G(K)$, i.e., one normalizes the other, then HK is a subgroup of G.

Proof. Assume HK is a subgroup of G. We need to show that $KH \subset HK$ and $HK \subset KH$. Because HK is a subgroup of G, for any $hk \in HK$ with $h \in H$ and $k \in K$, we have $k^{-1}h^{-1} = (hk)^{-1} \in HK$, from which it is derived that $KH \subset HK$. We could also derive from the same equation that $HK \subset KH$, since $k^{-1}h^{-1} \in KH$.

Now assume conversely that $HK = KH$. Given $ab, hk \in HK$ with $a, h \in H$ and $b, k \in K$, we have $(ab)(hk)^{-1} = a(bk^{-1}h^{-1})$. Because $bk^{-1}h^{-1} \in KH = HK$, $bk^{-1}h^{-1} = uv$ for some $u \in H$ and $v \in K$, and $(ab)(hk)^{-1} = (au)v \in HK$, as desired. □

3.2 Some important propositions

3.2.1 Cauchy's group theorem

Theorem 3.2.1 (Cauchy's group theorem)**.** Let G be a finite group and p be a prime dividing $|G|$. Then there is an element of order p in G.

Proof. [1] Let S denite the set of p-tuples of elements of G the product of whose coordinates is 1, i.e.,

$$S := \{(x_1, x_2, \cdots, x_p) \in G^p : x_1 \cdot x_2 \cdot \cdots \cdot x_p = 1\}.$$

Step 1. It is easy to observe that $|S| = |G|^{p-1}$.

Step 2. It is also easy to observe that S is closed under (cyclic) permutations.

Step 3. Define the relation $\sim$ on S by letting $\alpha \sim \beta$ if and only if β is a cyclic permutation of α ($\alpha, \beta \in S$). Then it is easy to show that the relation $\sim$ is an equivalence relation on S.

Step 4. An equivalence class contains a single element if and only if a member of the class is of the form $(x, x, \cdots, x)$ for some $x \in G$. (Clearly, in this case, $x^p = 1$.)

Step 5. Hence, $|S| = k + pd$, where k is the number of equivalence classes with a unique member and d is the number of equivalence classes with p-distinct members.

Step 6. Since $k \geq 1$ and k is divisible by p, it is implied that there is a nonidentity element x such that $x^p = 1$, i.e., of order p. □

3.2.2 Conjugacy

Definition 3.2.2 (Conjugacy)**.** Let G be a group and let x, y be elements of G. The relation $\sim$ defined in G by $x \sim y$ if and only if $x = gyg^{-1}$ for some $g \in G$ is called the conjugacy relation, which is an equivalence relation on G. For an element $g \in G$, the map $\gamma_g : G \to G$ defined by $\gamma_g(x) = gxg^{-1}$ is called the conjugation (automorphism).

Suppose G acts on itself by conjugation. Then the stabilizer of $x \in G$ is the set $\{g \in G : \gamma_g(x) = x\}$ and the kernel of this action is the intersection of $\text{stab}_G(x)$ for $x \in X$. To be general, the stabilizer of $A \subset G$ is the intersection of $\text{stab}_G(x)$ for $x \in A$, which is also denoted by $C_G(A)$.

When conjugation is considerd to be acted on the power set of G, i.e., G acts on $\mathcal{P}(G)$ by conjugation, the stabilizer of $A \subset G$ is the set $\{g \in G : \gamma_g(A) = A\}$, and this set will be denoted by $N_G(A)$.

Problem 3.2.1. Check that $C_G(H) \leq N_G(H) \leq G$, whenever H is a subgroup of G.

3.2.3 Some automorphism groups

Theorem 3.2.3. $\text{Aut}(Z_n) \approx (\mathbb{Z}/n\mathbb{Z})^\times$.

Proof. Left as an exercise. □

[1] Reference: James Mckay (*Another proof of Cauchy's group theorem*, Amer. Math. Monthly, 66(1959), p.119).

3.2.4 Difficult order counting

Proposition 3.2.4. Let G be a cyclic group of order n and let m be a positive integer. Then the set

$$\{y \in G : y^m = 1\}$$

has (m, n)-distinct elements.

Proof 1. Suppose $(x^a)^m = 1$, where x is a generator of G and $0 \leq a < n$. Then $n|am$, so $n'|am'$, where $n = n'd$ and $m = m'd$ with $d = (m, n)$. Because $(m', n') = 1$, a is forced to be a multiple of n', i.e., $a \in \{0, n', 2n', \cdots, (d-1)n'\}$. Conversely, for any a in the preceeding set, $(x^a)^m = 1$, so the set in the proposition contains (m, n)-distinct elements. □

Proof 2. For convinience, let the set in the proposition be denited by H and write $|H| = a$.

(a) Writing $d = (m, n)$, there are integers r and s such that $d = mr + ns$.

(b) Since $H \leq G$, $H = \langle x^c \rangle$ for some integer c such that $ac = n$.

We first show that $a|d$; it is because $(x^c)^d = x^{cd} = x^{c(mr+ns)} = 1$. Conversely, since $(x^{n/d})^m = (x^{m/d})^n = 1$, we have $x^{n/d} \in H$; by Lagrange's theorem we fininally obtain that $d = |x^{n/d}|$ divides a. □

Chapter 4

Group action

4.1 Basic theory of group action

Definition 4.1.1 (Group action)**.** Let G be a group and A be a nonempty set. A function $\sigma : G \times A \to A$ is called a group action of G on A, if σ satisfies the following axioms: simply writing $\sigma(g, a) = g \cdot a$ for all $g \in G$ and $a \in A$,

(a) $1_G \cdot a = a$ for all $a \in A$.

(b) $x \cdot (y \cdot a) = xy \cdot a$ for all $x, y \in G$ and $a \in A$.

Definition 4.1.2. Suppose that a group G acts on a nonempty set A.

(a) (Stabilizer) Given $a \in A$, the subgroup $G_a := \{g \in G : g \cdot a = a\}$ of G is called the stabilizer of $a \in A$.

(b) (Orbit) Given $a \in A$, the subset $G \cdot a = \{g \cdot a : g \in G\}$ is called the orbit of a.

(c)

(a)

Observation 4.1.3. Suppose that a group G acts on a nonempty set A. Then the collection of all orbits in A form a partition of A. Hence, we have the following equation, when A is a finite set:

$$|A| = |A^G| + \sum_{i=1}^{r} |G \cdot a_i|,$$

where A^G is the collection of elements of A whose orbit has exactly one element and $a_1, \cdots, a_r$ are representatives of pairwise disjoint orbits with more than one elements.

4.2 Groups acting on quotient groups by left multiplication

4.3 Groups acting on itself by conjugation

4.4 Automorphism groups

4.5 Sylow's theorem

Chapter 5

Semidirect product of groups

5.1 Recognition theorem 1: Recognizing as a direct product

5.2 Recognition theorem 2: Recognizing as a semidirect product

Chapter 6

Free groups

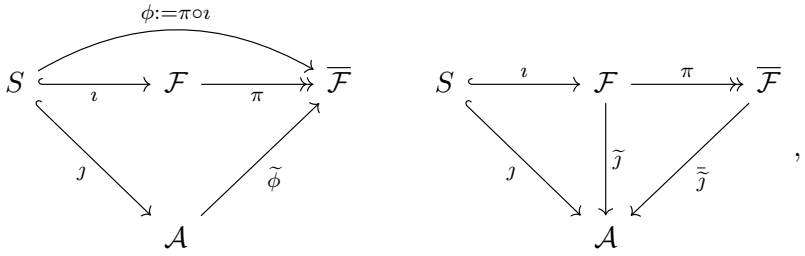
6.1 Remarks on free abelian groups

Definition 6.1.1 (Free abelian group)**.** A free $\mathbb{Z}$-module is called a free abelian group.

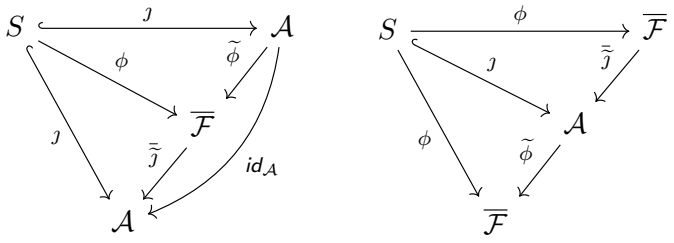
Note that the above definition is quite natural in the sense that an abelian group can be considered a $\mathbb{Z}$-module and vice versa. Other than the above somewhat formal definition, there is an intuitive method of considering free abelian groups. As noted in the chapter of free objects, any given structured object X can be considered the free object $\mathcal{F}(S)$ with some relations imposed. The free abelian group $\mathcal{F}(S)$ (here, S is a nonempty set) can be considered free groups "in which every pair of elements commute," and the normal subgroup playing such a role is exactly $[\mathcal{F}(S), \mathcal{F}(S)]$. Therefore, one might suggest that $\mathcal{F}(S)/[\mathcal{F}(S), \mathcal{F}(S)]$ is the free abelian group generated by S.

Theorem 6.1.2. Let $(\mathcal{F}_{\mathsf{gp}}(S), \imath)$ be the free group generated by a nonempty set S, and let $(\mathcal{F}_{\mathsf{ab}}(S), \jmath)$ be the free abelian group generated by S. Then $\mathcal{F}_{\mathsf{gp}}(S)/[\mathcal{F}_{\mathsf{gp}}(S), \mathcal{F}_{\mathsf{gp}}(S)] \approx \mathcal{F}_{\mathsf{ab}}(S)$.

Proof. For convinience, let $\mathcal{F} = \mathcal{F}_{\mathsf{gp}}$, $\mathcal{A} = \mathcal{F}_{\mathsf{ab}}$, and $\overline{\mathcal{F}} = \mathcal{F}/[\mathcal{F}, \mathcal{F}]$. Consider the following commutative diagrams:



where $\overline{\jmath_*}$ is a well-defined group homomorphism, since $[\mathcal{F}, \mathcal{F}] \leq \jmath_*^{-1}(\{0\})$. The above two commutative diagrams yields the following commutative diagrams.



From the left diagram, we can find that $\overline{\widetilde{\jmath}} \circ \widetilde{\phi} = id_{\mathcal{A}}$, by using a universal property of $\mathcal{A}$. From the right diagram, one can notice that $\phi = (\widetilde{\phi} \circ \widetilde{\jmath}) \circ \phi$, from which it can be deduced that $\overline{s} = (\widetilde{\phi} \circ \widetilde{\jmath})(\overline{s})$ whenever $s \in S$. Because $\overline{\mathcal{F}}$ is generated by $\overline{s}$ for $s \in S$ and $\widetilde{\phi} \circ \widetilde{\jmath}$ is a group homomorphism extending ϕ, we have $\widetilde{\phi} \circ \widetilde{\jmath} = id_{\overline{\mathcal{F}}}$. □

Chapter 7

Further group theory

7.1 Generated normal subgroups

Let G be a group and S be a subset of G. We have learned that the subgroup of G generated by S is the intersection of all subgroups of G containing S, which is the set of words with alphabets in $S \cup S^{-1}$.

Definition 7.1.1 (Generated normal subgroup)**.** Let G be a group and S be a subset of G. The normal subgroup (S) of G generated by S is defined as the intersection of all normal subgroups of G containing S.

Proposition 7.1.2. Let G be a group and S be a subset of G. Then

$$(S) = \left\langle gsg^{-1} : g \in G, s \in S \right\rangle.$$

Proof. Let $G_S = \left\langle gsg^{-1} : g \in G, s \in S \right\rangle$. We first show that G_S is a normal subgroup of G containing S. It is clear that G_S contains S. Given a word $w := (g_1 s_1 g_1^{-1}) \cdots (g_n s_n g_n^{-1}) \in G_S$ with $g_i \in G, s_i \in S \cup S^{-1}$ and $x \in G$, we have

$$xwx^{-1} = (xg_1 s_1 g_1^{-1} x^{-1}) \cdots (xg_n s_n g_n^{-1} x^{-1}) \in G_S.$$

Hence, G_S is a normal subgroup of G.

The proof is done if it is proved that (S) contains G_S, which is clear because (S) is a normal subgroup of G containing S. □

7.2 Chinese remainder theorem for groups

Theorem 7.2.1 (Chinese remainder theorem for groups (two normal subgroups))**.** Suppose M, N are normal subgroups of G such that $G = MN$. Then $G/(M \cap N) \approx (G/M) \times (G/N)$.

Proof. Define a map $\phi : G \to (G/M) \times (G/N)$ by $\phi(x) = (xM, xN)$ for $x \in G$. It is straightforward that ϕ is a group homomorphism with $M \cap N$ as the kernel. Thus, it remains to check surjectivity. Because $G = MN$, given $(aM, bN) \in (G/M) \times (G/N)$, we can write $a = m_a n_a$ and $b = m_b n_b$ for some $m_a, m_b \in M$ and $n_a, n_b \in N$. Using these elements, we also have $(aM, bN) = (n_a M, m_b N)$, and ϕ maps $n_a m_b$ to this tuple. □

7.3 Chains of groups

There are two widely used chains for groups. One is generally called the composition series, and the other will be called the solvability chain (or series) in this note.

We first introduce the composition series, after introducing its motivation.

Definition 7.3.1 (Maximal (normal) subgroup)**.** A subgroup M of G is called a maximal (normal) subgroup of G if

(a) $H < G$ ($H \lhd G$) and

(b) whenever $H \leq K \leq G$ ($H \leq K \trianglelefteq G$), either $K = H$ or $K = G$.

Definition 7.3.2 (Simple group)**.** A group is called a simple group if there are only two normal subgroups: the trivial subgroup and the group itself. In other words, a group is simple if and only if the trivial subgroup is the maximal normal subgroup.

From definition and our intuition, we first impose a straightforward result and its outstanding corollary, which states that every finite group has a *simple normal chain*, which is a composition series.

Proposition 7.3.3. Let G be a nontrivial group. Then the followings are equivalent:

(a) N is a maximal normal subgroup of G.

(b) $N \lhd G$ and G/N is simple.

Proof. Use the lattice isomorphism theorem. □

Definition 7.3.4 (Composition series)**.** Let G be a group, and suppose there are finitely many subgroups $G_1, \cdots, G_n$ such that

(a) $\{1_G\} = G_n \lhd G_{n-1} \lhd \cdots \lhd G_1 \lhd G_0 = G$ and

(b) each G_i/G_{i+1} $(i = 0, 1, \cdots, n-1)$ is a simple group.

To be short, a composition series is a simple normal chain.

Corollary 7.3.5. Every finite group has a composition series.

Proof. Since G is finite, G has a maximal normal subgroup G_1, and G_1 also has a maximal normal subgroup. By induction, we can establish a finite normal chain. Simplicity follows from normality. □

Now we study solvability chains of groups.

Definition 7.3.6 (Solvability chain)**.** Let G be a group, and suppose there are finitely many subgroups $G_1, \cdots, G_n$ such that

(a) $\{1_G\} = G_n \lhd G_{n-1} \lhd \cdots \lhd G_1 \lhd G_0 = G$ and

(b) each G_i/G_{i+1} $(i = 0, 1, \cdots, n-1)$ is an abelian group.

To be short, a solvability chain is an abelian normal chain. In addition, a group which has a solvability chain is called a solvable group.

As every finite group has a composition series, every abelian group (even if it is infinite) has a solvability chain.

Solvable groups behave well under subgroups and homomorphic images.

Proposition 7.3.7. Suppose G is a solvable group. Then subgroups of G are solvable, and homomorphic images of G are also solvable.[1]

Proof. To prove the first part of the proposition, let H be a subgroup of G and let $\{1_G\} = G_n \lhd G_{n-1} \lhd \cdots \lhd G_1 \lhd G_0 = G$ be a solvability chain of G. A good idea for the proof is to consider restrictions as follows:

$$\{1_G\} = G_n \cap H \lhd G_{n-1} \cap H \lhd \cdots \lhd G_1 \cap H \lhd G_0 \cap H = H.$$

For each $i = 0, 1, \cdots, n$, define $H_i = G_i \cap H$. It remains to show that H_i/H_{i+1} is abelian for $i = 0, 1, \cdots, n-1$, and for this, we decide to prove that H_i/H_{i+1} embeds into G_i/G_{i+1}. Because the natural

[1] Because the proposition holds when the words 'solvable' are replaced by 'abelian' or 'cyclic,' solvable groups can be seen as a generalization of abelian groups.

embedding $\iota_i : H_i \hookrightarrow G_i$ satisfies $\iota_i^{-1}(G_{i+1}) = H_{i+1}$, ι_i induces a group monomorphism $\overline{\iota_i} : H_i/H_{i+1} \hookrightarrow G_i/G_{i+1}$.

Before proving the second part of the proposition, note that a homomorphic image of G is isomorphisc to G/N for some normal subgroup N of G. Consider the canonical projection map $\pi : G \to G/N$ and the following chain:

$$\{\overline{1_G}\} = \pi(G_n) \triangleleft \pi(G_{n-1}) \triangleleft \cdots \triangleleft \pi(G_1) \triangleleft \pi(G_0) = \pi(G).$$

Letting $\pi_i : G_i \to G_i/N$ for each $i = 0, 1, \cdots, n$, because $G_{i+1} \leq \pi_i^{-1}(G_{i+1}/N)$, π_i induces the following group homomorphism:

$$\overline{\pi_i} : \frac{G_i}{G_{i+1}} \to \frac{G_i/N}{G_{i+1}/N}.$$

Because $\overline{\pi_i}$ is surjective, it is derived that $\overline{G_i}/\overline{G_{i+1}}$ is abelian for each $i = 0, 1, \cdots, n-1$, as desired. □

To review, the preceeding proposition states that solvable groups behave well under subgroups and homomorphic images (or equivalently, quotients by normal subgroups). The following proposition can be considered a converse of the preceeding proposition.

Proposition 7.3.8. Let G be a group and N be a normal subgroup of G. If both N and G/N are solvable, then G is also solvable.

Proof. Use the lattice isomorphism theorem and concatenate solvability chains. □

Another well-behaviour of solvable groups is given as the following proposition:

Proposition 7.3.9. If G_1 and G_2 are solvable groups, then so is $G_1 \times G_2$.

Proof. Find solvability chains of G_1 and G_2, and enlarge subgroups of $G_1 \times G_2$ term by term and step by step. Detailed proof is left as an exercise, because the proof is easy. □

We now introduce some results derived when composition series and solvability chain meet each other.

Proposition 7.3.10. (a) A finite abelian simple group is a cyclic group of a prime order.

(b) A finite solvable simple group is a cyclic group of a prime order.

(c) Thus, a finite nonabelian simple group is never solvable.

Proof. (a) easily follows, because every subgroup of an abelian group is a normal subgroup. To prove (b), note that a solvable simple group is necessarily abelian, and (a) forces such a group to be a cyclic group of a prime order. (c) follows easily when one assumes that there is a finite nonabelian simple group which is solvable; by (b), such a group is cyclic, which contradicts (b). □

We end this section with an alternative definition of 'finite' solvable groups.

Proposition 7.3.11. If G is a finite group, then G is solvable if and only if there are finitely many subgroups $G_1, \cdots, G_n$ of G such that

(a) $\{1_g\} = G_n \triangleleft G_{n-1} \triangleleft \cdots \triangleleft G_1 \triangleleft G_0 = G$ and

(b) G_i/G_{i+1} is a cyclic group of a prime order for each $i = 0, 1, \cdots, n-1$.

Proof. If part is clear, so it remains to prove only if part. Since G is finite, a composition series $\{1_g\} = G_n \triangleleft G_{n-1} \triangleleft \cdots \triangleleft G_1 \triangleleft G_0 = G$ of G exists. Because each G_i/G_{i+1} is finite, simple, and solvable, each quotient needs to be a cyclic group of a prime order, as desired. □

7.4 Commutator subgroups

Definition 7.4.1 (Commutator subgroup)**.** For a group G, write

$$[G,G] = \left\langle ghg^{-1}h^{-1} : g,h \in G \right\rangle$$

and call $[G,G]$ the commutator subgroup of G. Each element of G of the form $ghg^{-1}h^{-1}\,(g,h \in G)$ is called a commutator, and is also denoted by $[g,h]$.

Remark. A group G is an abelian group if and only if $[G,G] = \{1_G\}$.

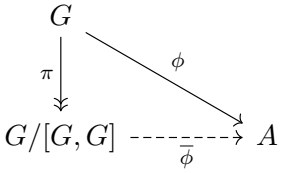
Proposition 7.4.2. Let G be a group.

(a) $[G,G] \trianglelefteq G$.

(b) $G/[G,G]$ is an abelian group.

(c) Assuming $N \trianglelefteq G$, G/N is an abelian group if and only if $[G,G] \leq N$. Hence, $G/[G,G]$ is the largest abelian quotient of G.

(d) If $\phi : G \to A$ is a homomorphism of a group G into an abelian group A, then ϕ factors through $[G,G]$, i.e., $[G,G] \leq \ker\phi$. Hence, ϕ induces a homomorphism $\overline{\phi} : G/[G,G] \to A$ of abelian groups such that $\overline{\phi} \circ \pi = \phi$.



Proof. Checking (a) and (b) is left as an exercise. To prove (c), note that G/N is an abelian group if and only if $[G/N, G/N] = \{1_{G/N}\}$, which is equivalent to $[G,G] \leq N$.

To prove (d), note that $[G,G] \leq \phi^{-1}(\{1_A\})$ so $\overline{\phi}$ is a well-defined group homomorphism; since $G/[G,G]$ is abelian, $\overline{\phi}$ is a homomorphism of abelian groups. □

Problem 7.4.1. Find a group G such that $Z(G/Z(G))$ is nontrivial.

Solution. Let $G = D_8$, where D_8 is the dihedral group of order 8. Then $Z(G) = \{1, r^2\}$ and $G/Z(G)$ is a group of order 4, which is abelian. Hence, the center of $G/Z(G)$ is nontrivial.

7.5 A bijection in a finite abelian group

Let A be a finite abelian group and let p be a prime number dividing the order of A. We will prove in this section that the subgroups of A of order p are in bijection with the subgroups of A of index p.

Part III

Ring theory

Chapter 8

Basic ring theory

8.1 Basic ideal theory

Proposition 8.1.1. Let R be a ring and assume A and B are ideals of R. Define two subsets $A+B$ and AB of R as follows:

(1) $A+B := \{x+y : x \in A,\, y \in B\}$.

(2) AB is defined as the collection of finite sums of elements of the form ab with $a \in A$ and $b \in B$.

Then both $A+B$ and AB are ideals of R. In fact, one can prove the following statements: For $I,\, J \trianglelefteq R$,

(a) $I+J$ is the smallest ideal of R containing I and J.

(b) $I \cap J$ is the largest ideal of R contained in $I \cap J$.

(c) IJ is an ideal of R contained in $I \cap J$.

(d) Assume that R is a commutative ring with the nonzero identity. If I and J are comaximal, i.e., $I+J=R$, then $IJ = I \cap J$. In general, if $I_1, I_2, \cdots, I_n$ are pairwise comaximal ideals of R, then $I_1 I_2 \cdots I_n = \bigcap_{i=1}^{n} I_i$.

Proof. Checking from (a) to (c) is left as an exercise. To prove (d), assume that R is a commutative ring with the nonzero identity. Because I and J are assumed to be comaximal, there are elements $a \in I$ and $b \in J$ such that $a+b=1$. Thus, if $x \in I \cap J$, then $x = 1x = (a+b)x = ax + xb \in IJ$, as desired. The generalized case will be explained when proving the Chinese remainder theorem. □

Proposition 8.1.2. Let R be a ring and A be a nonempty subset of R.

(a) RAR is an ideal of R.

(b) Assume that R contains the nonzero identity. Then RAR is the smallest ideal of R containing A, i.e., $(A) = RAR$.

Proof. (a) can be easily justified if one notes that $ras \times r'a's' = tasr' \times a' \times s' \in RAR$ for all $r, r', s, s' \in R$ and $a, a' \in A$. To prove (b), assume R is a ring with the nonzero identity. Clearly, RAR is an ideal of R containing A. If I is an ideal of R containing A, then RAR is contained in I by definition, so $(A) = RAR$. □

Observation 8.1.3. Let R be a commutative ring and let A and B be finitely generated ideals of R given by $A = (a_1, \cdots, a_m)$ and $B = (b_1, \cdots, b_n)$. One can easily check that $AB = (a_i b_j : 1 \le i \le m, 1 \le j \le n)$.

Proposition 8.1.4. Let R be a ring with the nonzero identity, and let I be an ideal of R.

(a) $I = R$ if and only if I contains a unit of R.

(b) Assume that R is commutative. Then R is a field if and only if its only ideals are 0 and R. Hence, a ring homomorphism from a field is either injective or trivial.

Definition 8.1.5. Let R be a ring.

(a) A proper ideal M of R is called a maximal ideal if I is an ideal of R containing M then $I = M$ or $I = R$.

(b) A proper ideal P of R is called a prime ideal if $ab \in P$ implies $a \in P$ or $b \in P$.

Proposition 8.1.6. Let R be a ring with the nonzero identity. Then R has a maximal ideal.

Proof. We try to apply Zorn's lemma to prove this statement.

Step 1. Setting a nonempty partially ordered subset.
Let X be a collection of all proper ideals of R. Then X is nonempty and partially ordered by set inclusion.

Step 2. Checking the upper bound axiom.
Let $\mathcal{C}$ be any ascending chain in X, and let U be the union of the members of $\mathcal{C}$. Our goal in this step is to prove that U is an upper bound of the the chain $\mathcal{C}$ in X, and for this it suffices to show that $U \in X$, i.e., U is a proper ideal of R.

Since U is a union of ideals in the chain $\mathcal{C}$, U is an ideal of R. If $U = R$, then U contains the identity, which implies that there is a member of $\mathcal{C}$ which contains the identity, a contradiction. Hence, $U \in X$ and U is an upper bound of $\mathcal{C}$.

Step 3. Deriving desired results.
Therefore, by Zorn's lemma, X has a maximal element M. And it is clear that a maximal element of X is a maximal ideal of R. □

Remark. Slightly modifying the proof, one can prove that every proper ideal of R is contained in a maximal ideal of R or that there is a maximal ideal of R containing $a \in R$ whenever a is a nonunit element of R.

Properties of maximal ideals and prime ideals are given as follows:

Proposition 8.1.7. Let R be a ring with the nonzero identity, and suppose I is an ideal of R.

(a) I is a maximal ideal of R if and only if R/I is a field.

(b) I is a prime ideal of R if and only if R/I is an integral domain.

8.2 Field of fractions

Throughout this section, D is an integral domain.

Define a relation $\sim$ on $Y_D := D \times (D \setminus \{0\})$ by $(a, b) \sim (c, d)$ if and only if $ad - bc = 0$. (Explain why the relation is an equivalence relation on Y_D.) And let $[a/b]$ denote the equivalence class of (a, b) in $Q_D := Y_D/\sim$. Also, define the addition and the multiplication on $Y_D/\sim$ as follows:

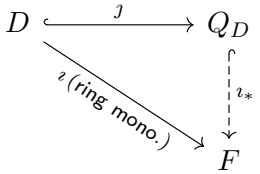
$$[a/b] + [c/d] := [(ad + bc)/bd], \quad [a/b] \times [c/d] := [ac/bd].$$

Theorem 8.2.1. Q_D is a field, and the map $\jmath : D \to Q_D$ defined by $\jmath(a) = [a/1]$ for $a \in D$ is a ring monomorphism. In fact, if D is a field, then $\jmath$ is a field isomorphism.

According to the preceeding theorem, we may identify D as a subset or a subring of Q_D. Hence, we may also write $[a/b] = [a/1][1/b] = \jmath(a)\jmath(b)^{-1} = ab^{-1}$ for all $a, b \in D$ with $b \neq 0$.

Together with the ring monomorphism $\jmath : D \hookrightarrow Q_D$, Q_D satisfies the following universal property:

Theorem 8.2.2 (A universal property of the field of fractions $(Q_D, \jmath : D \hookrightarrow Q_D)$)**.** For any monomorphism $\imath : D \hookrightarrow F$ of the integral domain D into a ring F, there is a unique field monomorphism $\imath_* : Q_D \hookrightarrow F$ such that $\imath_* \circ \jmath = \imath$.



Proof. Given an embedding $\imath$ of the integral domain D into a field F, define

$$\imath_*[a/b] := \imath(a)\imath(b)^{-1} \quad (a \in D, b \in D \setminus \{0\}).$$

Once it is proved that $\imath_*$ is a well-defined map, it can easily be seen that $\imath_*$ is a field homomorphism which is injective and that $\imath_* \circ \jmath = \imath$. □

8.3 Chinese remainder theorem for rings

Throughout this section, unless stated otherwise, all rings are assumed to be commutative and have the nonzero identity.

Remark. In this section, applying comaximality appropriately is essential, and the key propositions are as follows: For a commutative ring R with the nonzero identity and comaximal ideals A and B,

(a) $A \cap B = AB$.

(b) The map $\phi : R \to R/A \times R/B$ defined by $\phi(r) = (r + A, r + B)$ for $r \in R$ is a ring epimorphism. Thus, by the first isomorphism theorem, $R/AB = R/(A \cap B) \approx R/A \times R/B$.

The above two propositions will be a lot helpful not only when proving the Chinese remainder theorem but also when solving some relevant problems.

Theorem 8.3.1 (Chinese remainder theorem)**.** Let R be a commutative ring with the nonzero identity, and suppose that $A_1, \cdots, A_n$ be pairwise comaximal ideals of R. Then $A_1 \cdot \cdots \cdot A_n = \bigcap_{i=1}^n A_i$, so we have the following isomorphism of rings:

$$\frac{R}{A_1 \cdot \cdots \cdot A_n} \approx \frac{R}{A_1} \times \cdots \times \frac{R}{A_n}.$$

Proof. We prove the theorem by induction on n.

Step 1. Proof for $n = 2$.

When $n = 2$, since A_1 and A_2 are comaximal, $A_1A_2 = A_1 \cap A_2$ and there are elements $a \in A_1$ and $b \in A_2$ such that $a + b = 1$. Hence, the ring homomorphism $\phi : R \to R/A_1 \times R/A_2$ defined by $\phi(x) = (x + A_1, x + A_2)$ for $x \in R$ is surjective, since $\phi(xb + ya) = (x + A_1, y + A_2)$. The desired result follows from the first isomorphism theorem.

Step 2. Generalization.

What we want to show is the following two statements:

(a) $A_1 \cdot \cdots \cdot A_n = \bigcap_{i=1}^n A_i$.

(b) The ring homomorphism $\phi : R \to R/A_1 \times \cdots \times R/A_n$ defined by $\phi(x) = (x + A_1, \cdots, x + A_n)$ for $x \in R$ is surjective.

We prove (a) by induction; we assume the equation holds for $(n-1)$-pairwise comaximal ideals. For each $i = 1, \cdots, n-1$, let $a_i \in A_i$ and $b_i \in A_n$ be elements such that $a_i + b_i = 1$. Because

$$1 = (a_1 + b_1) \cdots (a_{n-1} + b_{n-1}) = a_1 \cdot \cdots \cdot a_{n-1} + \star$$

with $\star := 1 - (a_1 + b_1) \cdots (a_{n-1} + b_{n-1}) \in A_n$ and $a_1 \cdot \cdots \cdot a_{n-1} \in A_1 \cdot \cdots \cdot A_{n-1}$, we find that $A_1 \cdot \cdots \cdot A_{n-1}$ and A_n are comaximal. Therefore, $\bigcap_{i=1}^n A_i = A_1 \cdot \cdots \cdot A_n$, as desired.

To prove (b), it suffices to find $x_i \in R$ for each $i = 1, \cdots, n$ such that

$$x_i \equiv 1 \text{ mod } A_i \text{ and } x_i \equiv 0 \text{ mod } A_j \text{ whenever } j \neq i.$$

And for this, it suffices to find $x_i \in R$ for each i such that

$$x_i \equiv 1 \text{ mod } A_i \text{ and } x_i \equiv 0 \text{ mod } B_i,$$

where $B_i = \bigcap_{j \neq i} A_j$; such x_i indeed exists for each i, since A_i and B_i are comaximal as found in the preceeding paragraph. □

Example 8.3.2 (Ideals of product rings)**.** Suppose R and S are commutative rings with respective nonzero identities. We will justify that every ideal of $R \times S$ is of the form $I \times J$, where $I \trianglelefteq R$ and $J \trianglelefteq S$.

Suppose $A \trianglelefteq R \times S$, and let $\pi_1 : R \times S \to R$ and $\pi_2 : R \times S \to S$ be the natural projections.

$$\text{Goal: To prove that } A = \pi_1(A) \times \pi_2(A).$$

To prove the goal, it suffice to prove that $\pi_1(A) \times \pi_2(A) \subset A$. Choose a point $(p, q) \in \pi_1(A) \times \pi_2(A)$ and let $x \in R$ and $y \in S$ be elements such that $(p, y), (x, q) \in A$. Then it easily follows that $(p, q) = (1, 0)(p, y) + (0, 1)(x, q) \in A$, as desired.

8.4 Noetherian ring

Definition 8.4.1 (Noetherian ring)**.** Let R be a ring (not necessarily an integral domain). Then the followings are equivalent, and a ring satisfying any of the following property is called a Noetherian ring.

(a) (Finite condition) Every ideal of R is finitely generated.

(b) (Ascending chain condition) Every ascending chain of ideals of R is finite. To be precise, if $I_1, I_2, \cdots$ are ideals of R such that $I_1 \subset I_2 \subset \cdots$, then there is an integer $n \in \mathbb{N}$ such that $I_j = I - k$ for all $j, k \geq n$.

(c) (Maximal condition) Let S be any nonempty collection of ideals of R partially ordered by set inclusion. Then S contains a maximal member.

Remark. Every principal ideal domain is a Noetherian ring, since it satisfies the finite conditon.

Proof. We first prove that the finite condition implies the ascending chain condition. Let $I_1 \subset I_2 \subset \cdots$ be an ascending chain of ideals of R, and define

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

One can easily check that I is an ideal of R. By hypothesis, $I = (a_1, \cdots, a_k)$ for some $a_1, \cdots, a_k \in R$; because, for each i, $a_i \in I_j$ for some $j \in \mathbb{N}$, the ascending chain is finite.

We now prove that the ascending chain condition implies the maximal condition. Let S be a nonempty collection of ideals of R and partially order S by set inclusion. Choose a member I_1 of S; if I_1 is maximal, the proof is done. If I_1 is not maximal, there is another member I_2 of S strictly containing I_1. By induction, given a non-maximal member I_n of S, there is another member I_{n+1} of S strictly containing I_n. Because the ascending chain $I_1 \subset I_2 \subset \cdots$ is finite by hypothesis, there is an integer n such that I_n is maximal, which proves that S contains a maximal member. (Hence, in this case, we did not have to apply Zorn's lemma.)

Finally, we prove that the maximal condition implies the finite condition. Define the collection S of ideals of R by

$$S := \{J \trianglelefteq R : J \subset I \text{ and } J \text{ is finitely generated}\}.$$

By hypothesis, S contains a maximal member M. We will show that $I = M$ by contradiction. Assume $M \subsetneq I$. Then there is an element $x \in I \setminus M$, thus $M \subsetneq (M, x) \subset I$. Because (M, x) is also finitely generated, $(M, x) \in S$, so M is not a maximal member of S, a contradiction. □

Chapter 9

Types of integral domains

Throughout this chapter, all rings are assumed to be integral domains, unless stated otherwise. Also, unless stated otherwise, D denotes an integral domain.

9.1 Multiples and divisors

The idea of multiples and divisors are assumed to be considered in integral domains. Throughout this section, D denotes an integral domain.

Definition 9.1.1 (Multiple and divisor)**.** Let a and b be elements of D. If there is $c \in D$ such that $a = bc$, then a is called a multiple of b (and b is called a divisor of a). If $d \in D$ is a divisor of both a and b, then d is called a common divisor of a and b; if $m \in D$ is a multiple of both a and b, then m is called a common multiple of a and b.

Assume $a_1, \cdots, a_n \in D$. Then an element $d \in D$ is called a greatest common divisor of $a_1, \cdots, a_n$ if

(1) d is a common divisor of $a_1, \cdots, a_n$ and

(2) d is a multiple of every common divisor of $a_1, \cdots, a_n$.

Also, an element $l \in D$ is called a least common multiple of $a_1, \cdots, a_n$ if

(3) l is a common multiple of $a_1, \cdots, a_n$ and

(4) l is a divisor of every common multiple of $a_1, \cdots, a_n$.

We say a and b are relatively prime if (a) and (b) are comaximal, i.e., $(a, b) = (a) + (b) = D$.

To illustrate properties of multiple and divisor in terms of ideals, we define an equivalence relation $\sim$ on D as follows:

$$\text{For } a, b \in D,\ a \sim_\times b \text{ if and only if } a = ub \text{ for some } u \in D^\times.$$

Observation 9.1.2. Suppose $a, b \in D$.

(a) a is a divisor of b if and only if $(b) \trianglelefteq (a) \trianglelefteq R$. Thus, $a \sim_\times b$ if and only if $(a) = (b)$, and $(u) = D$ whenever u is a unit of D. Note that $(a) = (b)$ if and only if a divides b and b divides a.

(b) $a \sim_\times 0$ if and only if $a = 0$; assuming $a \in D^\times$, $a \sim_\times b$ if and only if $b \in D^\times$.

(c) Suppose $x_1, \cdots, x_n \in D$ and $a \sim_\times b$. If a is a common divisor (or a common multiple, respectively) of $x_1, \cdots, x_n$, then so is b.

Proposition 9.1.3. Suppose $a, b \in D$.

(a) $(a) + (b) = (a, b)$

(b) $(a)(b) = (ab)$.

Proposition 9.1.4. Suppose $x_1, \cdots, x_n \in D$. Then $d \in D$ is a greatest common divisor of $x_1, \cdots, x_n$ if and only if $(x_1, \cdots, x_n) \trianglelefteq (d) \trianglelefteq (d')$ whenever $d' \in D$ is a common divisor of $x_1, \cdots, c_n$; $m \in D$ is a least common multiple of $x_1, \cdots, x_n$ if and only if $(m') \trianglelefteq (m) \trianglelefteq (x_1) \cap \cdots \cap (x_n)$ whenever $m' \in D$ is a common multiple of $x_1, \cdots, x_n$. Also, a greatest common divisor and a least common multiple of $x_1, \cdots, x_n$ are unique up to multiplication by a unit in D, respectively.

We end this section with an observation on principal ideal domains.

Observation 9.1.5. Let D be a principal ideal domain and $a, b \in D$. If we let $(a) + (b) = (x)$, then x is a greatest common divisor of a and b, and vice versa. Similarly, if we let $(a) \cap (b) = (y)$, then y is a least common multiple of a and b, and vice versa.

9.2 Irreducible elements and prime elements

Definition 9.2.1. Let D be an integral domain.

(a) (Irreducible element) A nonzero and nonunit element $r \in D$ is called an irreducible element if r satisfies the following property:

$$\text{If } r = ab \text{ for some } a, b \in D, \text{ either } a \text{ or } b \text{ is a unit in } D.$$

(b) (Prime element) A nonzero and nonunit element $p \in D$ is called a prime element if p satisfies the following property:

$$\text{If } p|ab \text{ for some } a, b \in D, \text{ then } p|a \text{ or } p|b.$$

The above statement is equivalent to the following statement:

$$\text{If } ab \in (p) \text{ for some } a, b \in D, \text{ then } a \in (p) \text{ or } b \in (p).$$

Hence, a nonzero and nonunit element p of D is a prime element if and only if (p) is a prime ideal of D.

Observation 9.2.2. Suppose x, y are nonzero and nonunit elements of D and assume $x \sim_\times y$. Then y is an irreducible (a prime, respectively) element of D if x is an irreducible (a prime) element of D.

Proof. Write $y = ux$ for some unit u in D, and assume first that x is irreducible. Whenever $y = ab$ for some $a, b \in D$, because $x = u^{-1}ab$ and x is irreducible, $u^{-1}a$ or b is a unit in D, implying that a or b is a unit in D. Now assume that x is a prime element. Then y is clearly a prime element, since $(x) = (y)$. □

Proposition 9.2.3. A prime element is an irreducible element.

Proof. Let p be a prime element of the integral domain D, and write $p = ab$ with $a, b \in D$. Without loss of generality, we may assume that $a \in (p)$, i.e., $a = px$ for some $x \in D$; from $p = pxb$, we have $b \in D^\times$ as desired. □

Remark. Later in this chapter, it will be proved that a Euclidean domain is a principal ideal domain. Since any integral domain contains a maximal ideal, every Euclidean domain (or a principal ideal domain) contains a prime element and an irreducible element.

We end this section with a simple property satisfied in any integral domain.

Observation 9.2.4 (Factorization of elements of integral domains). Let D be an integral domain and r be a nonzero and nonunit element of D. Then there clearly exist elements $a_1, a_2 \in D$ such that $r = a_1 a_2$. If r is irreducible, then either a_1 or a_2 is a unit in D; if r is reducible, then a_1, a_2 can be chosen to be (nonzero and) nonunit. (What if not?)

9.3 Euclidean domain

Definition 9.3.1 (Size function and Euclidean domain)**.** Let D be an integral domain. Any function $N : D \to \mathbb{Z}^{>0}$ such that $N(0) = 0$ is called a size function on D. The integral domain D is called a Euclidean domain if it has a size funciton N on D satisfying the following property:

For any $a, b \in D$ with $b \neq 0$, there are $q, r \in D$ such that $a = qb + r$ with either $r = 0$ or $N(r) < N(b)$.

Example 9.3.2. Fields, the Gaussian integer ring $\mathbb{Z}[i]$ are Euclidean domains.

Theorem 9.3.3. Euclidean domains are principal ideal domains.

Proof. Let D be a Euclidean domain and I be an ideal of D. By the well-ordering principle of $\mathbb{N}$, there is a nonzero element α of I with the smallest value of a size function on D. If $x \in I$, there are elements $q, r \in D$ such that $x = q\alpha + r$ with either $r = 0$ or $N(r) < N(\alpha)$. Because $r = x - q\alpha$, r is an element of I, which forces $r = 0$ and $x = q\alpha$. Therefore, $I = (\alpha)$, proving that D is a principal ideal domain. □

9.4 Principal ideal domain

Definition 9.4.1 (Principal ideal domain)**.** An integral domain in which every ideal is principal is called a principal ideal domain.

Remark. Let D be a principal ideal domain and $a, b \in D$. Write $(a, b) = (x)$ and $(a) \cap (b) = (y)$. Then, x is a greatest common divisor of a and b, and y is a least common multiple of a and b, which exist uniquely up to multiplication by units in D, respectively.

Theorem 9.4.2. In principal ideal domains, nonzero prime ideals and nonzero maximal ideals coincide.

Proof. It suffices to prove that any nonzero prime ideal of a prinicpal ideal D is a maximal ideal of D. Let $P = (p)$ be a nonzero prime ideal of D, and suppose $P \leq I \trianglelefteq D$ with $I = (a)$ for some $a \in D$. Since $p \in (a)$, $p = ab$ for some $b \in D$. Because p is a prime element of D, we have $p|a$ or $p|b$, which, respectively, implies $I = P$ or $a \in D^\times$ so that $I = D$. Therefore, every nonzero prime ideal of D is a maximal ideal of D. □

We have observed that in any integral domain a prime element is an irreducible element and that nonzero prime ideals and nonzero maximal ideals coincide in principal ideal domains. The following theorem states some equivalences in principal ideal domains.

Theorem 9.4.3 (Equivalences in principal ideal domains)**.** Let D be a principal ideal domain and p be a nonzero element of D. Then the followings are equivalent:

(a) p is a prime element of D.

(b) p is an irreducible element of D.

(c) (p) is a prime ideal of D.

(d) (p) is a maximal ideal of D.

Proof. (a) and (c) are verified to be equivalent when we defined prime elements; we have proved that (c) and (d) are equivalent; we have proved that (a) implies (b). Thus, it remains to prove that (b) implies any other statement; here, we will show that (b) implies (d).

Suppose $(p) \leq I \trianglelefteq D$ and write $I = (a)$. We can write $p = ab$ for some $b \in D$, thus a or b is a unit in D, which, respectively, implies that $I = D$ or $I = P$, implying that (p) is a maximal ideal of D. □

9.5 Unique factorization domain

Definition 9.5.1 (Unique factorization domain)**.** An integral domain D is called a unique factorization domain if every nonzero and nonunit element r of D satisfies the following properties:

(a) r can be written as a finite product of irreducible elements of D.

(b) The decomposition in (a) is unique ip to multiplication by units; if $r = p_1p_2\cdots p_m = q_1q_2\cdots q_n$, where $m, n \in \mathbb{N}$ and p_i and q_j for $1 \leq i \leq m, 1 \leq j \leq n$ are irreducible elements of D, then $m = n$ and $p_i \sim_\times q_i$ for each i.

We have studied the following equivalences valid in principal ideal domains:

(a) p is a prime element of D.

(b) p is an irreducible element of D.

(c) (p) is a prime ideal of D.

(d) (p) is a maximal ideal of D.

Among them, (a) and (c) are equivalent by definition, (a) implies (b) and (d) implies (c) in any integral domain. In UFDs, a nonzero prime ideal is no longer necessarily a maximal ideal; howerver, an irreducible element is still a prime element.

Proposition 9.5.2. In a UFD, an irreducible element is a prime element.

Proof. Let r be an irreducible element of a UFD D and assume $r|ab$ for some nonzero elements a, b of D. Then r is an irreducible factor of ab and an irreducible factor of a or b. Therefore, $r|a$ or $r|b$ and r is a prime element. □

In the following observation, some obvious but helpful properties of unique factor domains are listed.

Observation 9.5.3. Let D be a UFD. Suppose that $a = up_1{}^{r_1}p_2{}^{r_2}\cdots p_n{}^{r_n} \in D$, where $u \in D^\times$, p_i is an irreducible element of D and $r_i \in \mathbb{N}$ for each i.

(a) If p is an irreducible element of D dividing a, then $a \sim_\times p_i$ for some i. Hence, if d is an element of D dividing a, then $d \sim_\times p_1{}^{f_1}p_2{}^{f_2}\cdots p_n{}^{f_n}$, where $0 \leq f_i \leq r_i$ for each i.

Suppose further that $b = vp_1{}^{s_1}p_2{}^{s_2}\cdots p_n{}^{s_n} \in D$, where $v \in D^\times$, p_i is an irreducible element of D and $s_i \in \mathbb{N}$ for each i.

(b) Letting $e_i = \min\{r_i, s_i\}$ and $f_i = \max\{r_i, s_i\}$ for each i, $p_1{}^{e_1}p_2{}^{e_2}\cdots p_n{}^{e_n}$ is a greatest common divisor of a and b, and $p_1{}^{f_1}p_2{}^{f_2}\cdots p_n{}^{f_n}$ is a least common multiple of a and b.

(c) Hence, if g is a greatest common divisor and l is a least common multiple of a and b, respectively, then $gl \sim_\times ab$. Also, if a and b are nonzero, then $(l/a, l/b) = (a/g, b/g) = D$.

In the remaining of this seciton, we will prove that a principal ideal domain is a UFD. In the proof, given a nonzero and nonunit element a from a principal ideal domain D, we should factorize r into irreducible elements. For this, we should investigate the existence of an irreducible element of D dividing a; for this, it suffices to prove the existence of a prime (or a maximal) ideal of D containing a, which is already proved in the preceeding chapter.

Theorem 9.5.4. Every principal ideal domain is a UFD.

Proof. Let D be a principal ideal domain and let a be a nonzero and nonunit element of D.

Step 1. Proving the existence part

Note that whenever $c \in D$ is nonzero and nonunit, there is a maximal ideal of D containing c. If a is reducible, find a maximal ideal (r_1) of D containing a and write $a = r_1a_1$; if a_1 is reducible, find an irreducible element r_2 of D dividing a_1 and write $a_1 = r_2a_2$. By induction, when a_n is reducible, let r_{n+1}

be an irreducible element of D dividing a_n and write $a_n = r_{n+1}a_{n+1}$. We will justify that such process terminates in finite steps so that a_n is irreducible for some $n \in \mathbb{N}$. Assume that a_n is not irreducible for all n. Because each r_n is nonunit, we have a properly ascending chain $(a) \subsetneq (a_1) \subsetneq (a_2) \subsetneq \cdots$ of ideals of D. Since a principal ideal domain is a Noetherian ring, the ascending chain is finite, i.e., $(a_n) = (a_{n+1}) = \cdots$ for some integer $n \in \mathbb{N}$, a contradiction. Therefore, whenever a is a nonzero and nonunit element of D, then one can find a factorization of a into irreducible elements of D.

Step 2. Proving the uniqueness part

The uniqueness part can be proved by induction. Assume that a has the following two factorizations into irreducible elements of D:

$$a = up_1 \cdots p_m \text{ and } a = vq_1 \cdots q_n,$$

where $u, v \in D^\times$ and p_i's and q_j's are irreducible elements of D for all i and j. (Without loss of generality, assume that $m \leq n$.) Since $p_1 | (q_1 \cdots q_n)$ and p_1 is a prime element of D, (after renumbering we can write) $p_1 | q_1$ so that $p_1 \sim_\times q_1$. (How?) By the same argument, we have (again, after renumbering) $p_i \sim_\times q_i$ for each i. Hence, if $m < n$, by the law of cancellation, $q_{m+1} \cdots q_n \sim_\times 1$, a contradiction. Therefore, $m = n$ and $p_i \sim_\times q_i$ for each i.

By Step 1 and Step 2, every principal ideal domain is a UFD. □

9.6 The Gaussian integer ring

9.6.1 Quotients of the Gaussian integer ring

Observation 9.6.1. If I is a nonzero ideal of $\mathbb{Z}[i]$, the quotient ring $\mathbb{Z}[i]/I$ is a finite ring. To be precise, letting $I = (\alpha)$, the order of $\mathbb{Z}[i]/I$ is at most $|\alpha|^2$.

Proof. Write $I = (\alpha)$ for some $\alpha \in \mathbb{Z}[i]$. Every element of $\mathbb{Z}[i]/I$ can be uniquely written in the form of $\overline{a + bi}\,(a, b \in \mathbb{Z})$ with $a \in \mathbb{Z}[i]$ and $a^2 + b^2 < |\alpha|^2$ by the division algorithm. Because there are finitely many Gaussian integers of modulus smaller than $|\alpha|$, the quotient ring $\mathbb{Z}[i]/I$ is finite. □

Proposition 9.6.2. Let p be a prime number.

(a) If $p \equiv 3$ mod 4, then $\mathbb{Z}[i]/(p)$ is isomorphic to the field of order p^2.

(b) If $p \equiv 1$ mod 4, then $p = \pi\overline{\pi}$ for some irreducible element $\pi \in \mathbb{Z}[i]$. Because (π) and $(\overline{\pi})$ are comaximal, by the Chinese remainder theorem,

$$\frac{\mathbb{Z}[i]}{(p)} \approx \frac{\mathbb{Z}[i]}{(\pi)} \times \frac{\mathbb{Z}[i]}{(\overline{\pi})}.$$

Proof. We first prove (a). Since p is a prime element of $\mathbb{Z}[i]$, the quotient ring is a field of characteristic p. Suppose that $a, b, x, y \in \mathbb{Z}$. Then $a + bi \equiv x + yi$ mod p if and only if $a \equiv x$ and $b \equiv y$ mod p, thus the order of the quotient ring is p^2.

We now prove (b). To show that (π) and $(\overline{\pi})$ are comaximal,it suffices to prove that $\overline{\pi} \neq (\pi)$; the maximality of (π) will prove the comaximality. Assuming $\overline{\pi} \in (\pi)$, we have $\alpha\overline{\pi} = \pi$ for some $\alpha \in \mathbb{Z}[i]$ with $|\alpha| = 1$. Writing $\pi = a + bi$ for some $a, b \in \mathbb{Z}$,

(1) When $\alpha = 1$, we have $a + bi = a - bi$ so that $b = 0$ and $p = a^2$.

(2) When $\alpha = -1$, we have $-a - bi = a - bi$ so that $a = 0$ and $p = b^2$.

(3) When $\alpha = i$, we have $-b + ai = a - bi$ so that $a = -b$ and $p = 2a^2$.

(4) When $\alpha = -i$, we hace $b - ai = a - bi$ so that $a = b$ and $p = 2a^2$.

In either of the above cases, p is not a prime number, a contradiction. Hence, (π) and $(\overline{\pi})$ are comaximal and the Chinese remainder theorem holds. □

Theorem 9.6.3. $|\mathbb{Z}[i]/(\alpha)| = |\alpha|^2$, where α is a nonzero element of $\mathbb{Z}[i]$.

Chapter 10

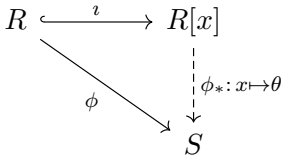
Polynomial rings

Throughout this chapter, we assume that R is a commutative ring with the nonzero identity and that D is an integral domain.

10.1 Basic properties

Definition 10.1.1 (Polynomial rings)**.** Let R be a commutative ring with the nonzero identity. The set $R[x]$ is defined as the collection of functions $f : \mathbb{Z}^{\geq 0} \to R$ with an integer $n \in \mathbb{Z}$ such that $f(k) = 0$ whenever $k \geq n$. Polynomial rings with multiple indeterminates are defined inductively: $R[x_1, \cdots, x_n] := R[x_1, \cdots, x_{n-1}][x_n]$. Imposing the usual operations on $R[x]$, $R[x]$ becomes a commutative ring with the nonzero identity.

Proposition 10.1.2 (A universal property of polynomial rings)**.** Let R be a commutative ring with the nonzero identity. Let $\phi : R \to S$ be a ring homomorphism and let θ be an element of S. Then there is a unique ring homomorphism $\phi_* : R[x] \to S$ which extends ϕ and maps x to θ.



Proof. If such ϕ_* exists, then it should be satisfied that

$$\phi_* \left(\sum_{r=0}^{n} a_r x^r \right) = \sum_{r=0}^{n} a_r \theta^r.$$

Checking details are left to readers. □

A simple proposition follows.

Proposition 10.1.3. Let I be an ideal of R and let (I) be the ideal of $R[x]$ generated by I, i.e., $(I) = R[x]I = I[x]$. Then $R[x]/(I) \approx (R/I)[x]$. In particular, if I is a prime ideal of R, then (I) is a prime ideal of $R[x]$.

Proof. Consider the projection map $f : R[x] \to (R/I)[x]$ defined by

$$f \left(\sum_{r=0}^{n} a_r x^r \right) = \sum_{r=0}^{n} \overline{a_r} x^r.$$

Checking details are left to readers. □

Another simple, but important, proposition follows.

Theorem 10.1.4 (Division algorithm on polynomial rings over fields)**.** Let F be a field and impose the following division algorithm on $F[x]$:

$$\text{Given } a(x), b(x) \in F[x] \text{ with } b(x) \neq 0, \text{ find } q(x), r(x) \in F[x] \text{ such that}$$

$$a(x) = q(x)b(x) + r(x),$$

$$\text{where either } r(x) = 0 \text{ or } \deg r(x) < \deg b(x).$$

(a) For each pair of $a(x)$ and $b(x)$, such $q(x)$ and $r(x)$ exist uniquely, respectively.

(b) Hence, $F[x]$ is a Euclidean domain.

Remark. By the uniqueness part, when E is a field extension of F and $a(x) = Q(x)b(x) + R(x)$ for some $Q(x), R(x) \in E[x]$ with $R(x) = 0$ or $\deg R(x) < \deg b(x)$, we have $Q(x) = q(x)$ and $R(x) = r(x)$.

Proof. We prove the assertion by induction on $\deg a(x)$. (By removing the leading term of $a(x)$, the case is reduced to the case which is assumed by the induction hypothesis, proving the existence part.) To prove the uniqueness part, assume $a(x) = q_1(x)b(x) + r_1(x) = q_2(x)b(x) + r_2(x)$ for some $q_i(x), r_i(x) \in F[x]$ with $r_i(x) = 0$ or $\deg r_i(x) < \deg b(x)$ for $i = 1, 2$. Because $r_1(x) - r_2(x) = (q_2(x) - q_1(x))b(x)$, considering the degrees of the polynomials, the uniqueness part can easily be explained. □

Corollary 10.1.5. Let F be a field and $f(x)$ be a nonzero and nonunit element of $F[x]$. Because $F[x]$ is a Euclidean domain, $(f(x))$ is a maximal ideal of $F[x]$ if and only if $f(x)$ is an irreducible element of $F[x]$, i.e., $f(x)$ is an irreducible polynomial. Therefore, $F[x]/(f(x))$ is a field if and only if $f(x)$ is irreducible.

Corollary 10.1.6. Let F be a field and $p(x)$ be a polynomial over F. By the lattice isomorphism theorem, the ideals of $F[x]/(p(x))$ and the ideals of $F[x]$ containing $(p(x))$ are in bijection. Also, an ideal of $F[x]$ containing $(p(x))$ is of the form $(a(x))$ for some $a(x) \in F[x]$ with $a(x)|p(x)$ and vice versa. Therefore, the ideals of $F[x]/(p(x))$ are of the form $ol(a(x))$ with $a(x)|p(x)$ and vice versa.

Problem 10.1.1. Let F be a field and $R = F[x, x^2y, x^3y^2, \cdots, x^{n+1}y^n, \cdots] \trianglelefteq F[x, y]$.

(a) Show that the field of fractions of R and $F[x, y]$ are the same.

(b) Explain why R contains an ideal which is not finitely generated.

Solution. (a) Clearly, the field Q_R of fractions of R is contained in the field Q of fractions of $F[x, y]$. Conversely, if $f(x, y) \in F[x, y]$, for some large integer N, we have $x^N f(x, y) \in R$, thus Q is contained in Q_R.

(b) Consider the following ascending chain $(x) \subsetneq (x, x^2y) \subsetneq (x, x^2y, x^3y^2) \subsetneq \cdots$ of ideals of R. If R does not contain an infinitely generated ideal, then R does not contain an infinite ascending chain of ideals, being a Noetherian ring.

Problem 10.1.2. Prove that $(x^i - y^j)$ is a prime ideal of $R[x, y]$, whenever i and j are relatively prime positive integers.

Solution. Note from $R[x, y] = R[x][y]$ that every polynomial in $R[x, y]$ differs by a polynomial in $(x^i - y^j)$ by a polynomial in $R[x, y]$ with degree in y less than j. In other words, given $a(x, y) \in R[x, y]$, there are $q(x, y), r(x, y) \in R[x, y]$ such that $a(x, y) = (x^i - y^j)q(x, y) + r(x, y)$ with $\deg_y r(x, y) < j$.

Define a ring homomorphism $f : R[x, y] \to R[s]$ extending the identity map on R by

$$f(x) = s^j, \quad f(y) = s^i.$$

Then $f(a(x, y)) = f(r(x, y))$, so

$$f(a(x, y)) = k_0(s^j) + s^i k_1(s^j) + \cdots + s^{(j-1)i} k_{j-1}(s^j),$$

where

$$r(x,y) = k_0(x) + k_1(x)y + \cdots + k_{j-1}(x)y^{j-1} \quad (k_0(x),\, k_1(x),\, \cdots,\, k_{j-1}(x) \in R[x]).$$

Because i and j are relatively prime, the above summation is a partition of $f(r(x,y))$ regarding degree of each monomial in s modulo j. In other words, all monomials in each summand $s^{mi}k_m(s^j)\,(m = 0, 1, \cdots, j-1)$ has the same degree modulo j, and any two monomials in two other summands have distinct degree modulo j. Therefore, $f(a(x,y)) = 0$ if and only if $k_0(x) = k_1(x) = \cdots = k_{j-1}(x) = 0$, i.e., $a(x,y) \in \ker f$ if and only if $a(x,y) \in (x^i - y^j)$. By the first isomorphism theorem, we have $R[x,y]/(x^i - y^j) \approx R[s]$, so $(x^i - y^j)$ is a prime ideal of $R[x,y]$.

10.2 Gauss's lemmas

Our goal in this section is to prove that $D[x]$ is a UFD if D is a UFD. Throughout this section, D is an integral domain and Q is the field of fractions of D.

Definition 10.2.1 (Content of a polynomial)**.** Let D be an integral domain and $f(x) = a_n x^n + \cdots + a_1 x + a_0$ be a polynomial over D. A greatest common divisor of $a_0, a_1, \cdots, a_n$ is called a content of f, and is denoted by $\mathit{cont}(f)$. If $\mathit{cont}(f)$ is a unit in D, then the polynomial f is called a primitive polynomial.

Some obvious observations:

Observation 10.2.2. (a) If $p(x) \in D[x]$ is a nonzero polynomial, then there is a primitive polynomial $p_1(x)$ over D such that $p(x) = \mathit{cont}(p) \cdot p_1(x)$.

Suppose $a, b \in D$ and $f(x), g(x) \in D[x]$ are nonzero and primitive polynomials.

(b) $\mathit{cont}(af) = a$.

(c) If $af(x) = bg(x)$, then $a \sim_\times b$.

Example 10.2.3. If $p(x) \in D[x]$ is a nonconstant irreducible polynomial, then $p(x)$ is primitive.

Theorem 10.2.4 (Gauss's lemma of primitivity)**.** Let D be a UFD and $f(x), g(x)$ be polynomials over D.

(a) $\mathit{cont}(fg) = \mathit{cont}(f) \cdot \mathit{cont}(g)$

(b) Hence, if $f(x)$ and $g(x)$ are primitive, then so is $f(x)g(x)$.

Proof. Since (b) follows from (a), it suffices to prove (a). Because $f(x) = \mathit{cont}(f) \cdot f_1(x)$ and $g(x) = \mathit{cont}(g) \cdot g_1(x)$ for some primitive polynomials $f_1(x), g_1(x) \in D[x]$, it suffices to prove that $f_1(x)g_1(x)$ is a primitive polynomial over D. For this, assume that $f_1(x)g_1(x)$ is not primitive and let p be a prime element of D dividing $\mathit{cont}(f_1 g_1)$. Consider the map $\phi : D[x] \to D[x]/(p)$ defined by

$$\phi\left(\sum_{r=0}^{n} a_r x^r\right) = \sum_{r=0}^{n} \overline{a_r} x^r \quad (a_r \in D \text{ for each } r).$$

Note that $D[x]/(p) \approx (D/pD)[x]$ is an integral domain and $\phi(f_1(x)g_1(x)) = 0$, and that ϕ is a ring homomorphism. Thus, without loss of generality, we may assume that $\phi(f_1(x)) = 0$, implying that p divides $f_1(x)$, which contradicts the primitivity of f_1. Therefore, $f_1(x)g_1(x)$ is primitive. □

Corollary 10.2.5. Let D be a UFD, and suppose that $p(x)$ is a nonzero polynomial over D. Then $p(x)$ is reducible in $D[x]$ if $p(x)$ is reducible in $Q[x]$. (To be precise, if $p(x) = A(x)B(x)$ for some $A(x), B(x) \in Q[x]$, then there are elements $s, t \in Q$ such that $a(x) := sA(x)$ and $b(x) := tB(x)$ belong to $D[x]$ and $p(x) = a(x)b(x)$. Hence, there exist $a(x), b(x) \in D[x]$ such that $p(x) = a(x)b(x)$ and $\deg a = \deg A$ and $\deg b = \deg B$.)

Proof. Because $p(x)$ is reducible over Q, we can write $p(x) = A(x)B(x)$ for some (nonconstant polynomials) $A(x), B(x) \in Q[x]$. By reducing fractions, we can write $A(x) = \frac{c}{m}a(x)$ and $B(x) = \frac{d}{n}b(x)$ for some primitive polynomials $a(x), b(x)$ over D and $c, d, m, n \in D$ with $m, n \neq 0$. Writing $p(x) = \textit{cont}(p) \cdot p_1(x)$ with $p_1(x) \in D[x]$ being primitive, we have $mn \cdot \textit{cont}(p) \cdot p_1(x) = cd \cdot a(x)b(x)$, hence $mn \cdot \textit{cont}(p) \sim_\times cd$, i.e., $p(x) \sim_\times \textit{cont}(p) \cdot a(x)b(x)$. This proves the desired statements. □

Theorem 10.2.6 (Gauss's lemma of irreducibility)**.** Let D be a UFD and Q be the field of fractions of D. Assume that $p(x)$ is a primitive polynomial over D. Then $p(x)$ is irreducible over D if and only if $p(x)$ is irreducible over Q.

Proof. We first assume that $p(x)$ is irreducible over D and let $a(x), b(x)$ be polynomials over Q such that $p(x) = a(x)b(x)$. By reducing fractions, we may write $a(x) = \frac{c}{m}a_0(x)$ and $b(x) = \frac{d}{n}b_0(x)$, where $c, d, m, n \in D \setminus \{0\}$ and $a_0(x), b_0(x)$ are primitive polynomials over D. Because $mn \cdot p(x) = cd \cdot a_0(x)b_0(x)$ and $p(x)$ is primitive, we have $mn \sim_\times cd$, so $p(x) \sim_\times a_0(x)b_0(x)$. Because $p(x)$ is irreducible over D, either $a_0(x)$ or $b_0(x)$ is a unit in $D[x]$, so either $a_0(x)$ or $b_0(x)$ is a constant polynomial. This proves that $p(x)$ is irreducible over Q.

We now assume that $p(x)$ is irreducible over Q and let $a(x), b(x)$ be polynomials over D such that $p(x) = a(x)b(x)$. Then, without loss of generality, $a(x) \in Q[x]^\times = Q$, thus $a(x)$ is a constant polynomial over D. Becasue $p(x)$ is primitive, we find that $a(x) \in D^\times$, priving that $p(x)$ is irreducible over D. □

Corollary 10.2.7. Let D be a UFD and $f(x), g(x)$ be primitive polynomials over D. Then $f(x) \sim_\times g(x)$ in $D[x]$ if and only if $f(x) \sim_\times g(x)$ in $Q[x]$.

Proof. It is clear that $f(x) \sim_\times g(x)$ in $Q[x]$ if $f(x) \sim_\times g(x)$ in $D[x]$. Assume conversely that $f(x) \sim_\times g(x)$ in $Q[x]$ so that we can write $f(x) = (b/a)g(x)$ for some $b/a \in Q[x]^\times = Q^\times$. Since both $f(x)$ and $g(x)$ are primitive and $af(x) = bg(x)$, we have $a \sim_\times b$, so $b/a \in D^\times$. Therefore, $f(x) \sim_\times g(x)$ in $D[x]$. □

Remark (Review of Gauss's lemmas). Let D be a UFD and Q be the field of fractions of D. Gauss's lemma of primitivity implies that any finite product of primitive polynomials over D is primitive. Gauss's lemma of irreduciblity implies the followings:

(a) A polynomial $p(x)$ over D is reducible over D if it is reducible over Q.

(b) Let $p(x)$ be a primitive polynomial over D. Then $p(x)$ is irreducible over D if and only if $p(x)$ is irreducible over Q.

Theorem 10.2.8. If D is a UFD, then so is $D[x]$.

Proof. Let $f(x)$ be a nonzero and nonunit polynomial over D.

Step 1. Proving the existence part

Let $f_1(x)$ be the polynomial over D such that $f(x) = \textit{cont}(f) \cdot f_1(x)$. The factorization of $\textit{cont}(f)$ cen be accomplished in D. To factorize $f_1(x)$ in $D[x]$, we first factorize $f_1(x)$ in a UFD $Q[x]$; write $f_1(x) = p_1(x) \cdot \cdots \cdot p_n(x)$ be the factorization of $f_1(x)$ in $Q[x]$. By reducing fractions, for each $i = 1, \cdots, n$, there is a primitive polynomial $q_i(x)$ over D and $a_i, b_i \in D \setminus \{0\}$ such that $p_i(x) = (b_i/a_i)q_i(x)$.

(1) Since each $p_i(x)$ is irreducible over Q, each $q_i(x)$ is also irreducible over Q, and so over D.

(2) Because $(a_1 \cdot \cdots \cdot a_n)f_1(x) = (b_1 \cdot \cdots \cdot b_n) \cdot q_1(x) \cdot \cdots \cdot q_n(x)$, we have $a_1 \cdot \cdots \cdot a_n \sim_\times b_1 \cdot \cdots \cdot b_n$.

Therefore, $f(x) = \textit{cont}(f) \cdot f_1(x) \sim_\times \textit{cont}(f) \cdot q_1(x) \cdot \cdots \cdot q_n(x)$ has a factorization into irreducible elements in $D[x]$.

Step 2. Proving the uniqueness part

Suppose that two factorization of $f(x)$ into irreducible elements in $D[x]$ are given as follows:

$$f(x) = (r_1 \cdot \cdots \cdot r_m) \cdot a_1(x) \cdot \cdots \cdot a_j(x) = (s_1 \cdot \cdots \cdot s_n) \cdot b_1(x) \cdot \cdots \cdot b_k(x),$$

where $r_1, \cdots, d_m, s_1, \cdots, s_n$ are irreducible elements of D and $a_1(x), \cdots, a_j(x), b_1(x), b_k(x)$ are irreducible polynomials in $D[x]$. (Being primitive, all nonconstant polynomial factors can be assumed to be primitive.)

In this case, $r_1 \cdots\cdots r_m$ and $s_1 \cdots\cdots s_n$ are the content of the polynomial $f(x)$, so they differ by multiplication by a unit; hence, $m = n$ and $r_i \sim_\times s$ for each i. Because $a_1(x), \cdots, a_j(x), b_1(x), \cdots, b_k(x)$ are primitive and irreducible over D, they are irreducible over Q; because $Q[x]$ is a UFD, $j = k$ and (up to renumbering) $a_i(x) \sim_\times b_i(x)$ for each i. This proves the uniqueness part. □

An additional problem, which is not essential when studying further theory.

Problem 10.2.1. Suppose that $f(x), g(x) \in D[x]$ are primitive. Explain that if $f(x) = g(x)h(x)$ for some $h(x) \in Q[x]$ then $h(x)$ is a polynomial over D.

Solution. By reducing fractions, we can write $h(x) = (b/a)h_0(x)$ for some $a, b \in D \setminus \{0\}$ and a primitive polynomial $h_0(x)$ over D. From $af(x) = bg(x)h_0(x)$, we obtain $a \sim_\times b$ so $u = b/a \in D^\times$. Therefore, $h(x) = uh_0(x) \in D[x]$, as desired.

Chapter 11

Further ring theory

11.1 The field of real numbers

In this section, we construct the field of real numbers. Keep in mind that we do not admit the existence of the field of real numbers yet.

Let $\mathcal{C}_{\mathbb{Q}}$ denote the collection of all Cauchy sequences of rational numbers and define operations as follows:

$$(a_n)_n + (b_n) := (a_n + b_n)_n, \quad c \cdot (a_n)_n := (ca_n)_n, \quad (a_n) \times (b_n) := (a_n b_n).$$

Checking well-definedness is left as an exercise. Then $\mathcal{C}_{\mathbb{Q}}$ is a $\mathbb{Q}$-algebra with the above operations and has the multiplicative identity $(1)_n$.

For a sequence $(a_n)_n$ of rational numbers and a rational number α, $(a_n)_n$ is said to converge to α if there is a rational number α with the following property:

Whenever $\epsilon > 0$, there is a positive integer N such that $n \geq N$ implies $|a_n - \alpha| < \epsilon$.

And let $\mathcal{M}$ denote the collection of all rational sequences which converges to 0.

Proposition 11.1.1. $\mathcal{M}$ is a maximal ideal of $\mathcal{C}_{\mathbb{Q}}$. Therefore, the quotient ring $\mathcal{C}_{\mathbb{Q}}/\mathcal{M}$ is a field containing an isomorphic copy of $\mathbb{Q}$.

Proof. It is easy to justify that $\mathcal{M}$ is an ideal of $\mathcal{C}_{\mathbb{Q}}$.

To show that $\mathcal{M}$ is a maximal ideal of $\mathcal{C}_{\mathbb{Q}}$, suppose that $(a_n)_n \in \mathcal{C}_{\mathbb{Q}} \setminus \mathcal{M}$. Set

$$x_n = \begin{cases} 10^{-n} & (a_n \neq 10^{-n}) \\ 0 & (a_n = 10^{-n}) \end{cases},$$

then $(x_n)_n \in \mathcal{M}$ and $(a_n - x_n)_n \in \mathcal{C}_{\mathbb{Q}}$ and $a_n - x_n \neq 0$ for all n. Because $((a_n - x_n)^{-1})_n \in \mathcal{C}_{\mathbb{Q}}$, $(1)_n = ((a_n - x_n)^{-1})_n \times (a_n - x_n)_n \in \mathcal{C}_{\mathbb{Q}}$.

Finally, it can be easily justified that $\mathcal{C}_{\mathbb{Q}}/M$ is a field containing an isomorphic copy of $\mathbb{Q}$ by considering the field embedding $\mu : \mathbb{Q} \hookrightarrow \mathcal{C}_{\mathbb{Q}}$ defined by $\mu(1) = \overline{(1)_n}$. □

Definition 11.1.2. (a) In the remaining of this section, we define $\mathbb{R} = \mathcal{C}_{\mathbb{Q}}/\mathcal{M}$.

(b) An element $\alpha \in \mathbb{R}$ is defined to be not less than 0 if there is a rational sequence $(a_n)_n$ such that $\alpha = \overline{(a_n)_n}$ and $a_n \geq 0$ for all n.

Proposition 11.1.3. Suppose that $\alpha, \beta \in \mathbb{R}$. Show the followings.

(a) Either $\alpha > \beta$ or $\alpha = \beta$ or $\alpha < \beta$, and not simultaneously.

(b) If $\alpha, \beta > 0$, then $\alpha + \beta, \alpha\beta > 0$.

Proof. Almost clear. □

Given $\alpha = \overline{(a_n)_n} \in \mathbb{R}$, let $|\alpha| := \overline{(|a_n|)_n}$. This induces the natural metric d on $\mathbb{R}$.

Theorem 11.1.4. $(\mathbb{R}, d)$ is a complete metric space.

Proof. Somebody proved the theorem. □

11.2 Limits and inverse limits

Part IV

Module theory

Chapter 12

Basic module theory

Throughout Part IV, when considering R-modules, R is assumed to be a ring with the nonzero identity, and it is assumed that the identity 1_R in R acts on R-modules trivially, i.e., whenever M is an R-module and $\cdot$ denotes the R-scalar multiplication on M, $1_R \cdot x = x$ for all $x \in M$.

12.1 Annihilation in submodules

Definition 12.1.1. Let M be an R-module.

(a) (Torsion) Given an element $x \in M$, if there is a nonzero element $r \in R$ such that $rx = 0$, then x is called an r-torsion element (or simply a torsion element) or said to be annihilated by r. Also, the collection M_{tor} of the torsion elements in M is called the torsion part of M. If $M = M_{\text{tor}}$, then M is called a torsion R-module; if $M_{\text{tor}} = \{0\}$, then M is called a torsion-free R-module.

(b) For a nonempty subset N of M, define

$$\mathit{ann}_R(N) := \{r \in R : rn = 0 \text{ for all } n \in N\}.$$

Also, for a nonempty subset S of R, define

$$\mathit{Ann}_M(S) := \{x \in M : sx = 0 \text{ for all } s \in S\}.$$

Observation 12.1.2. Let M be an R-module with a nonempty subset N, and let I be a nonempty subset of R.

(a) Assume that R is commutative. Then $\mathit{ann}_R(N)$ is an ideal of R, and $\mathit{Ann}_M(I)$ is an R-submodule of M.

(b) Assume that R may not be commutative. When $N \leq_R M$ and $I \trianglelefteq R$, then $\mathit{ann}_R(N)$ is an ideal of R and $\mathit{Ann}_M(I)$ is an R-submodule of M.

In either of the above case, $\mathit{ann}_R(N)$ is called the annihilator ideal of N in R and $\mathit{Ann}_M(I)$ is called the R-submodule of M annihilated by I. (Proving the above observations is left for the readers.) In most situations, when considering $\mathit{ann}_R(N)$ and $\mathit{Ann}_M(I)$, it is assumed that $N \leq_R M$ and $I \trianglelefteq R$.

Example 12.1.3. (a) If $r \in R^\times$, then $\mathit{Ann}_M(r) = 0$. In particular, every vector space is torsion-free.

(b) M is torsion-free if and only if $\mathit{ann}_R(x) = 0$ for all $x \in M \setminus \{0\}$.

(c) M_{tor} may not be an R-submodule of M. For example, consider the $\mathbb{Z}/6\mathbb{Z}$-module $\mathbb{Z}/6\mathbb{Z}$. Then its torsion part is $\{\overline{0}, \overline{2}, \overline{3}, \overline{4}\}$, which is not an $\mathbb{Z}/6\mathbb{Z}$-submodule of $\mathbb{Z}/6\mathbb{Z}$.

In fact, when R is an integral domain, then M_{tor} is an R-submodule of M.

Remarking that $G/[G,G]$ is a largest abelian quotient group for a group G, we can establish analogous propositions which coincide our intuition.

Proposition 12.1.4. Let D be an integral domain and let M be a D-module. Then M/M_{tor} is torsion-free.

Proof. Use the overline notation to denote the quotient by M_{tor}, and assume that $\overline{x} \in \overline{M}$ is a torsion element, i.e., there is a nonzero element $r \in R$ such that $r\overline{x} = \overline{0}$. Then $rx \in M_{\text{tor}}$, so x is a torsion element. Hence, $\overline{x} = \overline{0}$. □

Proposition 12.1.5. Let D be an integral domain and let N be a D-submodule of a D-module M. If M/N is torsion-free, then $M_{\text{tor}} \leq_D N$.

Proof. Suppose that $x \in M_{\text{tor}}$ and r is a nonzero element of R such that $rx = 0$. Using the overline notation to denote the quotient by N, because $\overline{M}$ is torsion-free and $r\overline{x} = \overline{rx} = \overline{0}$, $\overline{x} = \overline{0}$, implying that $x \in N$. □

Assume that D is a PID. and M is a D-module. Given an element $d \in D$, we wish to understand M as a $D/(d)$-module. The following propositions deal with this situation, and these propositions are essential when proving the uniqueness part of the cyclic decomposition theorem in Section 15.2.

Lemma 12.1.6. Let D be a PID. and M be a D-module. Suppose that p is an irreducible (or a nonzero prime, equivalently) element of D annihilating M, i.e., $pM = 0$. Writing $\overline{D} = D/(p)$, then $\overline{D}$ is a field. Defining the $\overline{D}$-scalar multiplication on M by

$$\overline{a} \cdot x := a \cdot x \quad (a \in D, x \in M),$$

then M is a $\overline{D}$-vector space.

Proof. This is because the above scalar multiplication is well-defined. (Checking well-definedness is left as an exercise.) □

Remark. The above lemma extends to a general situation: If R is a commutative ring with an ideal (a) and M is an R-module annihilated by a, then M is an $R/(a)$-module if the $R/(a)$-scalar multiplication is defined by $\overline{s} \cdot x := sx$ for $\overline{s} \in R/(a)$ and $x \in M$.

Example 12.1.7. Let R be a commutative ring and M be an R-module, and let a be an element of R. Even if M may not be annihilated by a, $\mathit{Ann}_M(a)$ is ahhihilated by a. Thus, $\mathit{Ann}_M(a)$ can be considered an $R/(a)$-module. In particular, if (a) is a maximal ideal of R, then $\mathit{Ann}_M(a)$ can be considered an $R/(a)$-vector space, for $R/(a)$ is a field.

The following proposition states some coincidences between D-modules and $\overline{D}$-vector spaces, where D and $\overline{D}$ are as in the preceeding lemma.

Proposition 12.1.8. Let D be a PID., and let M and M' be D-modules. Suppose that p is an irreducible element of D annihilating M and M', and write $\overline{D} = D/(p)$.

(a) N is a D-submodule of M if and only if N is a $\overline{D}$-subspace of M.

(b) Suppose that S is a subset of M. Then the D-submodule of M generated by S and the $\overline{D}$-vector space generated by S coincide.

(c) Let $\phi : M \to M'$ be a map. Then ϕ is D-linear if and only if ϕ is $\overline{D}$-linear.

Proof. Each equivalence follows directly from the assumption that p annihilates M and M'. □

Remark. As in the preceeding lemma, these coincidences may extend to a general case.

Example 12.1.9. Suppose that D is a PID. and $p \in D$ is an irreducible element. Then $D/(p)$ is a field, so $D/(p) \not\approx D/(p) \oplus D/(p)$ as $D/(p)$-vector spaces. By (c) of the preceeding proposition, therefore, $D/(p)$ and $D/(p) \oplus D/(p)$ are not isomorphic as D-modules, too.

More preparation for proving the uniqueness part of the cyclic decomposition theorem will be stated in Section 15.2, as it should be.

12.2 Examples of submodules

Example 12.2.1 (Abelian groups and $\mathbb{Z}$-modules)**.** Given an abelian group A, the *natural* action of $\mathbb{Z}$ on A gives a $\mathbb{Z}$-scalar multiplication on A. With this action, the abelian group A can be considered a $\mathbb{Z}$-module. The converse understanding is clear, by forgetting $\mathbb{Z}$-scalar multiplications. Therefore, we can identify an abelian group as a $\mathbb{Z}$-module and vice versa.

In particular, suppose that A is an abelian group and let m be a positive integer such that $mx = 0 \in A$ for all $x \in A$. Then A can also be considered $\mathbb{Z}/m\mathbb{Z}$-module. In particular, if m is a prime number p, then A can be considered an $\mathbb{F}_p$-vector space.

Example 12.2.2 (Vector spaces over fields as modules over polynomial rings)**.** Let F be a field and V be an F-vector space and let T be a linear operator on V. Given a polynomial $f(t) \in F[t]$, define the action of $f(t)$ on the element v of V as follows:

$$f(t) \cdot v := f(T)(v).$$

Together with action, V can be considered an $F[t]$-module, whose scalar multiplication extends the scalar multiplication of the F-vector space V.

Now we discuss when an F-subspace of V is an $F[t]$-submodule of V. It is clear that an $F[t]$-submodule of V is an F-subspace of V. So let W be an F-subspace of V and try to find a condition under which W is an $F[t]$-submodule of V. If W is an $F[t]$-submodule of V, then W has to be closed under $F[t]$-linear combinations. Hence, in particular, W has to be T-invariant. Conversely, if W is an F-subspace ov V which is T-invariant, it can easily be explained that the $F[t]$-scalar multiplication defined above behaves well on W, i.e., $f(t) \cdot w \in W$ whenever $f(t) \in F[t]$ and $w \in W$. Therefore, an F-subspace W of V is an $F[t]$-submodule of V if and only if W is T-invariant.

We now study quotient modules. For this, assume that M is an R-module and N be an R-submodule of N. For the quotient M/N to be an R-module, M/N has to be an abelian group, which is automatically achieved. (How?) And it is easy to observe that the *natural* R-scalar multiplication on M/N behaves well. Therefore, M/N is an R-module whenever N is an R-submodule of M.

Example 12.2.3. Let F be a field and V be a finite dimensional vector space over F. Supppose that W is an F-subspace of V. Considering the F-linear map $\pi_N : M \to M/N$, by the dimension theorem, we have $\dim M = \dim \mathit{ker}(\pi_N) + \dim \mathit{im}(\pi_N) = \dim N + \dim(M/N)$.

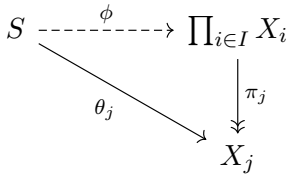
12.3 Direct sums of submodules

Throughout this section, we assume that I is a nonempty index set and that every set indexed by the elements of I is nonempty.

12.3.1 Direct products of sets

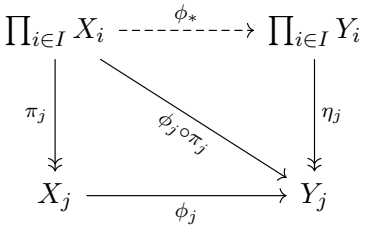
Definition 12.3.1 (Direct product)**.** Let I be a nonempty index set and suppose that X_i is a nonempty set for each $i \in I$. The product $\prod_{i\in I} X_i$ is defined as the collection of the function f from I into $\bigcup_{i\in I} X_i$ such that $f(i) \in X_i$ for each $i \in I$.

Proposition 12.3.2 (A universal property of direct products)**.** Let $X = \prod_{i\in I} X_i$ be the product of nonempty sets X_i for $i \in I$. For any nonempty set S and for any collection of functions $\theta_j : S \to X_j$ $(j \in I)$, there is a unique set map $\phi : S \to \prod_{i\in I} X_i$ such that $\pi_j \circ \phi = \theta_j$ for each $j \in I$.



By naturally defining operations on direct products, one can easily check that the direct product of $\square\square$'s is a $\square\square$. Here is an analogous universal property for the direct products of $\square\square$'s.

Proposition 12.3.3 (A universal property of dirct products of $\square\square$'s)**.** For each $i \in I$, assume that X_i and Y_i are $\square\square$ and let $\phi_i : X_i \to Y_i$ be a $\square\square$-homomorphism. Also, let $\pi_j : \prod_{i\in I} X_i \to X_j$ and $\eta_j : \prod_{i\in I} Y_i \to Y_j$ be the natural projections for $j \in I$. Then there is a unique $\square\square$-homomorphism $\phi_* : \prod_{i\in I} X_i \to \prod_{i\in I} Y_i$ such that $\eta_j \circ \phi_* = \phi_j \circ \pi_j$ for all $j \in I$.



12.3.2 External direct sum

While the idea of external direct sum can be extended to other structured objects (even for sets, as illustrated in the following definition), for convinience, we assume that the summands are either R-modules or R-algebras when considering their external direct sum.

Definition 12.3.4 (External direct sum)**.** Let I be a nonempty index set and X_i be a nonempty set for $i \in I$. The external direct sum $\bigoplus_{i\in I} X_i$ of X_i's are defined as the collection of the element x in the direct product of X_i's such that $x(i)$ is nonzero for all but finitely many i's. In other words,

$$\bigoplus_{i\in I} X_i := \left\{ x \in \prod_{i\in I} X_i : x(i) = 0 \text{ for all but finitely many } i \in I \right\}.$$

Several easy observations are given as follows:

Observation 12.3.5. (a) When the index set is finite, then the external direct sum and the direct product coincide.

(b) If X_i is a $\square\square$ for each $i \in I$, then $\bigoplus_{i\in I} X_i \leq_{\square\square} \prod_{i\in I} X_i$.

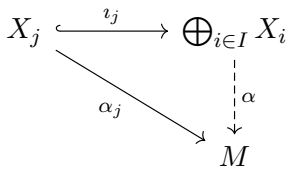
(c) Suppose that $\mathcal{B}_i$ is an F-basis of the F-vector space V_i for each $i \in I$. Then the union $\bigcup_{i\in I} \imath_i(\mathcal{B}_i) \subset \bigoplus_{i\in I} V_i$ is an F-basis of $\bigoplus_{i\in I} V_i$, where $\imath_i$ is the natural embedding.

The following observation is almost clear, but it will be a key observation when identifying external direct sums and internal direct sums.

Observation 12.3.6. Each element of $\bigoplus_{i\in I} X_i$ can be written as $\sum_{i\in I} \imath_i(x_i)$ with $x_i \in X_i$ for all $i \in I$, where $x_i = 0$ for all but finitely many i's. And such summation is uniquely determined.

Proof. Clearly, $x = \sum_{i\in I} \imath_i(x(i))$ is a desired sum. If $\sum_{i\in I} \imath_i(a_i) = \sum_{i\in I} \imath_i(b_i)$ with $a_i, b_i \in X_i$, then $\sum_{i\in I} \imath_i(a_i - b_i) = 0$, which forces $a_i = b_i$. □

Proposition 12.3.7 (A universal property of the external direct sum of R-modules)**.** Let X_i be an R-module for $i \in I$. For any R-module M and R-module homomorphisms $\alpha_j : X_j \to M$ for $j \in I$, there is a unique R-module homomorphism $\alpha : \bigoplus_{i\in I} X_i \to M$ such that $\alpha \circ \imath_j = \alpha_j$ for $j \in I$. (In general, this universal property is invalid for R-algebras.)



Proof. If such a *set map* $\alpha : \bigoplus_{i\in I} X_i \to M$ has to be defined, then it should be defined as follows: For each $x \in \bigoplus_{i\in I} X_i$,

$$\alpha(x) = \sum_{i\in I} \alpha_i(x(i)). \tag{12.1}$$

By the definition of the external direct sum, the summation in eq. (12.1) is a finite sum. For such *set map* α, if $x, y \in \bigoplus_{i\in I} X_i$ and $r \in R$, then

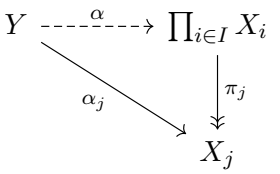
(1) $\alpha(x+y) = \sum_{i\in I} \alpha_i((x+y)(i)) = \sum_{i\in I}(\alpha_i(x(i)) + \alpha(y(i))) = \alpha(x) + \alpha(y)$ and

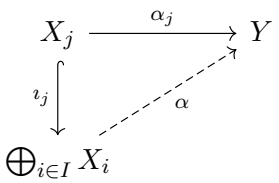
(2) $\alpha(rx) = \sum_{i\in I} \alpha_i((rx)(i)) = \sum_{i\in I} r\alpha_i(x(i)) = r\alpha(x)$.

Hence, α is a unique R-module homomorphism for which the above diagram commutes. It can be easily explained that such α fails to be an R-algebra homomorphism. □

Remark (Universal properties of direct products and external direct sums). (a) (A universal property of direct products) For any □□ Y and for every set of □□-homomorphisms $\{\alpha_j : Y \to X_j\}_{j\in I}$, there is a unique □□-homomorphism $\alpha : Y \to \prod_{i\in I} X_i$ such that $\pi_j \circ \alpha = \alpha_j$ for all $j \in I$.

(b) (A universal property of external direct sums) For any □□ Y and for every set of □□-homomorphisms $\{\alpha_j : X_j \to Y\}_{j\in I}$, there is a unique □□-homomorphism $\alpha : \bigoplus_{i\in I} X_i \to Y$ such that $\alpha \circ \imath_j = \alpha_j$ for all $j \in I$.

(a):


(b):


These two universal properties cannot be interchanged.

(1) We first justify that the direct product $\prod_{i\in I} X_i$ does not generally satisfy the universal property of the external direct sum in (b) by considering the following example: $I = \mathbb{N}$ with $X_j = Y = \mathbb{Z}$ and $\alpha_j = id_{\mathbb{Z}}$ for all $j \in \mathbb{Z}$.

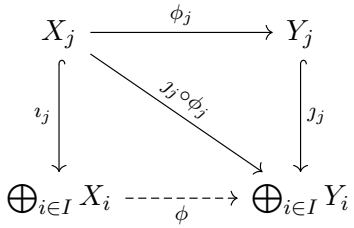
(2) We now check that the external direct sum $\bigoplus_{i\in I} X_i$ does not generally satisfy the universal property of the direct sum in (a). Suppose that the index set I is infinite and there is an element $y \in Y$ such that $\alpha_j(y) \neq 0$ for infinitely many $j \in I$. For such j, we have $(\pi_j \circ \alpha)(y) \neq 0$, implying that $\alpha(y)(j) \neq 0$ for infinitely many j's, a contradiction.

As an application of a universal property, we end this subsection with the following observation:

Observation 12.3.8. Let X_i and Y_i be R-modules for $i \in I$ and $\phi_i : X_i \to Y_i$ be an R-module homomorphism. And let $\imath_j : X_j \to \bigoplus_{i\in I} X_i$ and $\jmath_j : Y_j \to \bigoplus_{i\in I} Y_i$ be natural embeddings. Then there is a unique R-module homomorphism $\phi : \bigoplus_{i\in I} X_i \to \bigoplus_{i\in I} Y_i$ such that $\jmath_j \circ \phi = \phi_i \circ \imath_j$ for all $j \in I$. In fact, by the property of ϕ, the j-th component of ϕ satisfies the following identity:

$$\eta_j \circ \phi = \phi_j,$$

where $\eta_j : \bigoplus_{i\in I} Y_i \twoheadrightarrow Y_j$ is the natural projection, so there is no confusion of notation.



12.3.3 Internal direct sum

We studied the sum of F-linearly independent F-subspaces of an F-vector space. Such idea naturally extends to R-submodules, since a vector space is a module over a division ring.

Definition 12.3.9 (Internal direct sum of submodules)**.** Let M be an R-module and N_i be an R-submodule of M for each $i \in I$ with the following property: Given an element $x \in \sum_{i\in I} N_i$, for each $i \in I$, there is a unique element $x_i \in N_i$ such that

$$x_i \neq 0 \text{ only for finitely many } i \in I, \text{ and } x = \sum_{i\in I} x_i.$$

In this case, the sum of N_i's is called the internal direct sum of N_i's and denoted by $\bigoplus_{i\in I} N_i$.

Unlike the direct products and the external direct sums (where the latter are subobjects of the former), the internal direct sums are considered only for R-modules.

A simple observation which have been done in linear algebra is given as follows.

Observation 12.3.10. Let M be an R-module and N_i be an R-submodule of M for each $i \in I$. Then, the followings are equivalent:

(a) $M = \bigoplus_{i\in I} N_i$.

(b) Every element of M can be uniquely written as $\sum_{i\in I} x_i$, where only finitely many x_i's are nonzero.

(c) $M = \sum_{i\in I} N_i$ and the way of writing $0 \in M$ as $\sum_{i\in I} x_i$, where only finitely many x_i's are nonzero, is unique; the trivial method.

(d) $M = \sum_{i\in I} N_i$, and for each $i \in I$, $N_i \cap \sum_{j\in I\setminus\{i\}} N_j = 0$.

When proving their equivalence, note that (a) and (b) are equivalent by definition; (b) naturally implies (c) and (c) implies (b), which can be easily justified by the method of contradiction. To prove (d) when (c) is assumed, suppose $x \in N_i \cap \sum_{j\in I\setminus\{i\}} N_j$ and consider $0 = x - x$, which forces $x = 0$; to prove (c) when (d) is assumed, consider a nontrivial expression for $0 \in M$ and deduce a contradiction.

Observation 12.3.11. Let M be an R-module.

(a) If S is an R-linearly independent subset of M, then $\sum_{x\in S} Rx = \bigoplus_{x\in S} Rx$.

(b) In particular, if S is an R-basis of M, then $M = \bigoplus_{x\in S} Rx$.

Proposition 12.3.12. Let M be an R-module and assume that $N_i' \leq_R N_i \leq_R M$ for $i \in I$. If $\bigoplus_{i\in I} N_i' = \bigoplus_{i\in I} N_i$, then $N_i' = N_i$ for all $i \in I$.

Proof. The result is almost straightforward. It suffices to prove that $x_j \in N_j'$ for each $j \in I$ when $x_j \in N_j$ is given. Since $x_j = \imath_j(x_j) \in \bigoplus_{i\in I} N_i$ has the unique expression $x_j + \sum_{i\in I\setminus\{j\}} 0$ in $\bigoplus_{i\in I} N_i$, this expression should be valid in $\bigoplus_{i\in I} N_i'$, implying that $x_j \in N_j'$. □

12.3.4 Identifying direct sums

In this subsection, we prove that an external direct sum can be considered an internal direct sum and the converse consideration is also valid. When observing the definition of each direct sum, one can notice that finiteness is required for an element to be in the respective direct sum and that linear independency is also required. Hence, in particular when the index set I is finite so that one may assume that $I = \{1, \cdots, n\}$ for some $n \in \mathbb{N}$, we may identify the element $(x_1, \cdots, x_n) \in \bigoplus_{i\in I}^{\mathsf{ext}} N_i$ with the element $x_1 + \cdots + x_n \in \bigoplus_{i\in I}^{\mathsf{int}} N_i$.

In the following two observations, assume that M is an R-module and N_i is an R-submodule of M for each $i \in I$.

Observation 12.3.13. Suppose that $\sum_{i\in I} N_i = \bigoplus_{i\in I} N_i$. Define the map $\phi : \bigoplus_{i\in I}^{\text{int}} N_i \to \bigoplus_{i\in I}^{\text{ext}} N_i$ by

$$\left(\phi\left(\sum_{i\in I} x_i\right)\right)(j) = x_j \quad (j \in I, x_i \in I, \text{ and } x_i \neq 0 \text{ only for finitly many } i \in I).$$

Then ϕ is an R-module isomorphism.

Proof. The proof is technical. □

Observation 12.3.14. Consider the natural embedding $\imath_j : N_j \hookrightarrow \bigoplus_{i\in I}^{\text{ext}} N_i$. Then $\bigoplus_{i\in I}^{\text{ext}} N_i = \bigoplus_{i\in I}^{\text{int}} \imath_i(N_i)$. Hence, under the identification $N_i = \imath_i(N_i)$, we may also identify $\bigoplus_{i\in I}^{\text{ext}} N_i = \bigoplus_{i\in I}^{\text{int}} N_i$.

Proof. Clear. □

12.4 The Chinese remainder theorem for modules

In this section, we prove the Chinese remainder theorem for modules.

Theorem 12.4.1 (The Chinese remainder theorem for modules)**.** Let R be a commutative ring with the nonzero identity and assume that $I_1, \cdots, I_n$ are pairwise comaximal ideals of R. If M is an R-module, then $\bigcap_{i=1}^n I_i M = (I_1 \cdot \cdots \cdot I_n)M$ and

$$\frac{M}{(I_1 \cdot \cdots \cdot I_n)M} \approx \frac{M}{I_1 M} \times \cdots \times \frac{M}{I_n M}.$$

Before proving the theorem, we note the following observation, whose corresponding observation in ring thoery was necessary to prove the Chinese remainder theorem for rings.

Observation 12.4.2. Let R be a commutative ring with the nonzero identity and assume that I_1 and I_2 are comaximal ideals of R. We will show that $I_1M \cap I_2M = (I_1I_2)M$ if M is an R-module.

In fact, even if I_1 and I_2 are not comaximal, we have $(I_1I_2)M \leq_R I_1M \cap I_2M$. Using the comaximality of I_1 and I_2, there are elements $a \in I_1$ and $b \in I_2$ such that $a + b = 1 \in R$. If $x \in I_1M \cap I_2M$, then $x = 1x = (a+b)x = ax + bx$; because $ax \in (I_1I_2)M$ and $bx \in (I_2I_1)M = (I_1I_2)M$, so $x \in (I_1I_2)M$.

Proof of Theorem 12.4.1. As in the proof of the Chinese remainder theorem for rings, we prove the theorem by induction on n.

Step 1. Proof for $n = 2$

Consider the R-module homomorphism $\phi : M \to (M/I_1M) \times (M/I_2M)$ defined by

$$\phi(x) = (x + I_1M, x + I_2M) \text{ for } x \in M.$$

Using comaximality of I_1 and I_2, find $a \in I_1$ and $b \in I_2$ such that $a + b = 1 \in R$. Then, given $(u + I_1M, v + I_2M) \in (M/I_1M) \times (M/I_2M)$, we have $\phi(bu + av) = (u + I_1M, v + I_2M)$, so ϕ is surjective. The isomorphism now follows from the first isomorphism theorem.

Step 2. Generalization

What we want to show is the following two statements:

(a) $(I_1 \cdot \cdots \cdot I_n)M = \bigcap_{i=1}^n I_i M$.

(b) The ring homomorphism $\phi : R \to R/A_1 \times \cdots \times R/A_n$ defined by $\phi(x) = (x + I_1M, \cdots, x + I_nM)$ for $x \in R$ is surjective.

We prove (a) by induction; we assume the equation holds for $(n-1)$-pairwise comaximal ideals. For each $i \in 1, \cdots, n-1$, let $a_i \in I_i$ and $b_i \in I_n$ be elements such that $a_i + b_i = 1 \in R$. Because

$$1 = (a_1 + b_1) \cdots (a_{n-1} + b_{n-1}) = a_1 \cdot \cdots \cdot a_{n-1} + \star$$

with $\star = 1 - (a_1 + b_1) \cdots (a_{n-1} + b_{n-1}) \in I_n$ and $a_1 \cdot \cdots \cdot a_{n-1} \in I_1 \cdot \cdots \cdot I_{n-1}$, we find that $I_1 \cdot \cdots \cdot I_{n-1}$ and I_n are comaximal. Therefore, $\bigcap_{i=1}^n I_i M = (I_1 \cdot \cdots \cdot I_n)M$, as desired.

To prove (b), it suffices to find $x_i \in M$ for each $i = 1, \cdots, n$ such that

$$x_i \equiv 1 \bmod I_iM \text{ and } x_i \equiv 0 \bmod I_jM \text{ whenever } j \neq i.$$

And for this, it suffices to find $x_i \in M$ for each i such that

$$x_i \equiv 1 \bmod I_iM \text{ and } x_i \equiv 0 \bmod M_i,$$

where $M_i = \bigcap_{j \neq i} I_jM$; such x_i indeed exists for each i, since I_iM and M_i are comaximal as found in the preceeding paragraph. □

12.5 More on modules over polynomial rings over fields

In this section, we study some basic theory regarding finite dimensional F-vector spaces which are understood as $F[t]$-modules, where F is a field. In this section, we assume that F is a field and V is a finite dimensional F-vector space, and we understand V as an $F[t]$-module with a given F-linear operator T on V.

Before studying some properties of F-vector spaces as $F[t]$-modules, we first prove that a linear operator on an *algebraically closed* fields are triangularizable.

Theorem 12.5.1 (Triangularization)**.** Let E be an *algebraically closed* field and V be a finite dimensional E-vector space. If T is an E-linear operator on V, then there is an E-basis of V with respect to which the matrix representation of T is upper triangular, i.e., $[T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular.

Proof. We prove the theorem by induction on $dim_E V$. When $dim_E V = 1$, the result is clear. Assume the theorem holds for all E-vector spaces with $dim_E V < n$, and assume $dim_E V = n$. *Because E is algebraically closed*, the characteristic polynomial $\phi_T(t)$ of T has a root λ_1 in E; let v_1 be a nonzero eigenvector of T belonging to λ_1, and let $\mathcal{B}_1$ be an E-basis of V containing v_1. Then

$$[T]_{\mathcal{B}_1}^{\mathcal{B}_1} = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{array}\right),$$

and the lower-right $(n-1) \times (n-1)$ matrix is triangularizable by induction hypothesis. Therefore, there is an E-basis with respect to which the matrix representation of T is upper triangular. □

12.5.1 Annihilator ideals

Theorem 12.5.2 (Annihilator ideal)**.** The annihiltor ideal I_T of $F[t]$ with regard to T is defined as

$$I_T := \{f(t) \in F[t] : f(T) = O\}.$$

Remark. (1) Let $n = dim_F V$. Then $L(V, V)$ is a n^2-dimensional F-vector space, so

$$\{id_V, T, T^2, \cdots, T^{n^2}\}$$

is F-linearly dependent. This implies that there is a nonzero polynomial $f(t) \in F[t]$ such that $f(T) = O$, so I_T is nonzero.

(2) One can easily check that I_T is an ideal of $F[t]$: If $f(t), g(t) \in I_T$ and $r(t) \in F[t]$, then $f(t) - g(t), f(t)r(t) = r(t)f(t) \in I_T$. Also, since $F[t]$ is a Euclidean domain, $F[t]$ is a PID., so I_T can be generated by a unique monic polynomial $m_T(t) \in F[t]$, which is called the minimal polynomial of T. One can easily check that the F-dimension of $\{f(T) : f(t) \in F[t]\} \leq L(V, V)$ is the degree of the minimal polynomial of T.

Observation 12.5.3. Let $f(t)$ be a polynomial over F and S, T be an F-linear operator on V.

(a) If λ is an eigenvalue of T and v is an eigenvector of T belonging to λ, then v is an eigenvector of $f(T)$ belonging to $f(\lambda)$.

(b) If U is an automorphism of V, then $f(UTU^{-1}) = U \cdot f(T) \cdot U^{-1}$. Hence, $f(T) \sim f(S)$ if $T \sim S$.

In particular, from (b), it follows that $I_T = I_S$ (or equivalently, $m_T(t) = m_S(t)$) if $T \sim S$. In other words, annihiltor ideals are invariant under similarity transforms.

Proposition 12.5.4 (Cayley-Hamilton theorem)**.** Suppose that E is an *algebraically closed* field. If T is an E-linear operator on V, then $\phi_T(t) \in I_T$, where $\phi_T(t)$ is the characteristic polynomial of T. Hence, $m_T(t)|\phi_T(t)$.

Proof. Because E is algebraically closed, there is an E-basis $\mathcal{B}$ of V with respect to which the matrix representation of T is upper triangular. Letting $n = \dim_F V$ and writing $\phi_T(t) = (t-\lambda_1)\cdots(t-\lambda_n)$, the matrix representation of $T - \lambda_i id_V$ with respect to $\mathcal{B}$ is upper triangular with 0 on its (i,i)-entry. Hence, $\phi_T(T) = O$, so $\phi_T(t) \in I_T$. □

Problem 12.5.1. Suppose that $B \in \mathcal{M}_{m,m}(F)$, $C \in \mathcal{M}_{n,n}(F)$, and $D \in \mathcal{M}_{m,n}(F)$. And let $A = \left(\begin{array}{c|c} B & D \\ \hline O & C \end{array}\right)$. Prove the following statements:

(a) $\phi_A(t) = \phi_B(t)\phi_C(t)$.

(b) $m_A(t)$ is a common multiple of $m_B(t)$ and $m_C(t)$.

(c) If $D = O$, then $m_A(t)$ is the monic least common multiple of $m_B(t)$ and $m_C(t)$.

Solution. (a) is clear by the definition of characteristic polynomial. Because

$$f(A) = \begin{pmatrix} f(B) & * \\ O & f(C) \end{pmatrix}$$

whenever $f(t)$ is a polynomial over F, $m_A(t)$ is necessarily a multiple of $m_B(t)$ and $m_C(t)$. In particular, if $D = O$, then $f(A) = \begin{pmatrix} f(B) & O \\ O & f(C) \end{pmatrix}$, so the monic least common multiple of $m_B(t)$ and $m_C(t)$ annihilates A. Therefore, if $D = O$, then $m_A(t)$ is the monic least common multiple of $m_B(t)$ and $m_C(t)$.

Problem 12.5.2. If T is unipotent (in other words, $T - id_V$ is nilpotent) and there is a positive integer m such that T^m is the identity map on V, then T is the identity map on V.

Solution. By hypotheses, the minimal polynomial of T divides $(t-1)^r$ for some positive integer r and $t^m - 1 = (t-1)(t^{m-1} + \cdots + t + 1)$, hence $m_T(t) = t - 1$ and $T = id_V$.

Proposition 12.5.5. The monic irreducible divisors of $\phi_T(t)$ and $m_T(t)$ are the same.

Proof. It suffices to prove that the monic irreducible divisors of $\phi_T(t)$ are the monic irreducible divisors of $m_T(t)$. There is no problem when we assume that $T \in \mathcal{M}_{n,n}(F)$. Let $p(t) \in F[t]$ be an irreducible divisor of $m_T(t)$, and suppose that α is a root of $p(t)$ in the splitting field K of $p(t)$ over F. Then α is an eigenvalue of $T \in \mathcal{M}_{n,n}(K)$, so there is a nonzero vector $v \in K^n$ such that $Tv = \alpha v$, and we have $0 = m_T(T)v = m_T(\alpha)v$. Therefore, $m_T(\alpha) = 0$ and $p(t)$ divides $m_T(t)$, since $p(t)$ is the irreducible polynomial of α over F. □

The following proposition is valid when the base field is $\mathbb{R}$, whose algebraic closure is $\mathbb{C}$.

Proposition 12.5.6. Suppose that $A \in \mathcal{M}_{n,n}(\mathbb{R})$. Then the minimal polynomimal of A when A is considered a matrix over $\mathbb{R}$ or over $\mathbb{C}$ coincide.

Proof. Let $m(t) \in \mathbb{R}[t]$ denote the minimal polynomial of A when A is considered a matrix over $\mathbb{R}$; let $m_*(t) \in \mathbb{C}[t]$ denote the minimal polynomial of A when A is considered a matrix over $\mathbb{C}$. It suffices to show that $m_*(t) \in \mathbb{R}[t]$; if it is proved that $m_*(t) \in \mathbb{R}[t]$, then

(1) because $A \in \mathcal{M}_{n,n}(\mathbb{C})$ and $m(A) = O$, we have $m_*(t)|m(t)$, and

(2) because $m_*(t) \in \mathbb{R}[t]$ and $m_*(A) = O$, we have $m(t)|m_*(t)$

so that $m(t) = m_*(t)$. Because A is a matrix over $\mathbb{R}$, we have $\overline{A} = A$, thus $\overline{m_*}(A) = \overline{m_*}(\overline{A}) = \overline{m_*(A)} = O$, so $m_*(t)|\overline{m_*(t)}$. Because $\deg m_*(t) = \deg m(t)$, $m_*(t) \in \mathbb{R}[t]$, as desired. □

12.5.2 Subspaces which are invariant under linear operators

Example 12.5.7. Justifing the following statements is left as an exercise.

(a) If $\lambda \in F$ is an eigenvalue of T, then E_λ^T, the eigenspace of T belonging to λ is T-invariant.

(b) If $U, W \leq V$ are T-invariant, then both $U \cap W$ and $U + W$ are T-invariant subspaces of V.

(c) Given $v \in V$, the F-subspace $F[t]v = \{f(T)v : f(t) \in F[t]\}$ is the T-invariant subspace of V generated by $\{v, Tv, T^2v, \cdots\}$. (In fact, $F[t]v$ is the cyclic $F[t]$-submodule of V generated by $v \in V$.)

(d) If W is a T-invariant subspace of V and T is invertible, then $T|_W : W \to W$ is also invertible. Hence, W is T^{-1}-invariant and $T^{-1}|_W = (T|_W)^{-1}$.

The following example is separated from the preceeding example, due to its importance in further theory.

Example 12.5.8. Suppose $f(t) \in F[t]$.

(a) Then $\mathit{ker}\, f(T)$ and $\mathit{im}\, f(T)$ are T-invariant. Hence, in particular, $\mathit{ker} T$ and $\mathit{im} T$ are T-invariant.

Assume that W is a T-invariant subspace of V.

(b) W is $f(T)$-invariant, so $f(T|_W) = f(T)|_W$.

In particular, assume that $W = \mathit{ker}\, f(T)$.

(c) W is T-invariant, and $f(t)$ is a multiple of $m_{T|_W}(t)$, for $f(t)$ annihilates W, i.e., $f(T)w = 0$ for all $w \in W$.

Observation 12.5.9. Let W be a T-invariant subspace of V and let $\mathcal{B}$ be an F-basis of V which extends an F-basis $\mathcal{C}$ of W. Then the matrix representation of T with respect to $\mathcal{B}$ is block-diagonal:

$$[T]_{\mathcal{C}}^{\mathcal{B}} = \begin{pmatrix} [T|_W]_{\mathcal{C}}^{\mathcal{C}} & * \\ O & * \end{pmatrix}.$$

Thus, the characteristic polynomial of T_W plays an essential role when determining the characteristic polynomial of T; when the first quadrant of $[T]_{\mathcal{B}}^{\mathcal{B}}$ (or equivalently, the (1, 2)-block of $[T]_{\mathcal{B}}^{\mathcal{B}}$) is the zero matrix, the minimal polynomial of $T|_W$ plays an essential role when determining the minimal polynomial of T. For example, suppose that U, W are T-invariant subspaces of V such that $V = U \oplus W$. If $\mathcal{C}$ and $\mathcal{D}$ are F-bases of U and W, respectively, then $\mathcal{B} := \mathcal{C} \sqcup \mathcal{D}$ is an F-basis of V and

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} [T|_U]_{\mathcal{C}}^{\mathcal{C}} & O \\ O & [T|_W]_{\mathcal{D}}^{\mathcal{D}} \end{pmatrix},$$

so $\phi_T(t) = \phi_{T|_U}(t)\phi_{T|_W}(t)$ and $m_T(t) = \mathsf{lcm}\{\phi_{T|_U}(t), \phi_{T|_W}(t)\}$. In fact, this result naturally extends to the case where there are finitely many T-invariant subspaces of V whose direct sum is V: Letting $U_i\, (i = 1, \cdots, k)$ be T-invariant subspaces of V whose direct sum is V and $\mathcal{B}_i$ be an F-basis of U_i for each i, then

(1) $\mathcal{B} = \bigsqcup_{i=1}^k \mathcal{B}_i$ is an F-basis of V and

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \mathit{diag}\left([T_1]_{\mathcal{B}_1}^{\mathcal{B}_1}, \cdots, [T_k]_{\mathcal{B}_k}^{\mathcal{B}_k}\right),$$

where $T_i = T|_{U_i}$ for each i, and

(2) $\phi_T(t) = \phi_{T_1}(t) \cdots \phi_{T_k}(t)$ and $m_T(t) = \mathsf{lcm}\{m_{T_1}(t), \cdots, m_{T_k}(t)\}$.

12.5.3 Subspaces which are cyclic with respect to linear operators

Again, suppose that V is a finite dimensional vector space over the field F and write $n = \dim_F V$. We first observe the following definitions of T-cylcic (sub)spaces.

Definition 12.5.10. Let V be an n-dimensional vector space over the field F ($n < \infty$) and T be an F-linear operator on V.

(a) (T-cyclic space) V is called a T-cyclic space if there is a vector $v \in V$ such that $V = F[t]v$.

(b) (T-cyclic subspace) Suppose that W is a T-invariant subspace of V. Then W is called a T-cyclic subspace of V if W is a $T|_W$-cyclic space.

Observation 12.5.11. (a) By definition, it is clear that a T-cyclic space is a cyclic $F[t]$-module and vice versa. In fact, an $F[t]$-submodule W of V which is cyclic (with respect to T) is a T-cyclic subspace.

Assume that W is a cyclic $F[t]$-submodule of V. Then $W = F[t]v$ for some $v \in W$, so W is T-invariant, and this implies

$$F[t]v = \{f(T)v : f(t) \in F[t]\} = \{f(T|_W)v : f(t) \in F[t]\} \tag{12.2}$$

so W is $T|_W$-cyclic. Even it is assumed that W is a T-cyclic subspace of V, eq. (12.2) is valid, so W is a cyclic $F[t]$-submodule of V.

Therefore, a T-cyclic (sub)space can be considered a cyclic $F[t]$-(sub)module with respect to T, and vice versa.

(b) Cyclicity of a vector space is determined by the linear operator. If V is a vector space over a field F and W is a *finite dimensional* F-subspace of V, then there is a linear operator T of V for which W is T-cyclic, i.e., a cyclic $F[t]$-submodule of V. Since W is finite dimensional, there is an F-basis $\{v_1, \cdots, v_n\}$ of W, and there is a basis $\mathcal{B}$ of V extending $\{v_1, \cdots, v_n\}$. Defining the linear operator T of V by

$$Tu = \begin{cases} u & (\text{if } u \in \mathcal{B} \setminus \{v_1, \cdots, v_n\}) \\ v_{i+1} & (\text{if } u = v_i \text{ for some } i = 1, \cdots, n) \end{cases},$$

where $v_{n+1} = v_1$, then W is a cyclic $F[t]$-submodule of V, for $W = F[t]v_1$.

In this subsection, we discover some basic properties regarding cyclic $F[t]$-submodules of V; after studying cyclic decomposition theorem, we will discover further properties.

Example 12.5.12. Suppose that $V = F[t]v$ is T-cyclic and $g(t) \in F[t]$. If $g(T)v = 0$, then $g(t) \in \mathrm{ann}_{F[t]}(v) = \mathrm{ann}_{F[t]}V$, so $g(T)$ is the zero (F-linear) operator on V. It can be alternatively explained as follows: For any vector $x \in V$, there is a polynomial $a(t) \in F[t]$ such that $x = a(T)v$, so $g(T)x = g(T)a(T)v = a(T)g(T)v = 0$.

Lemma 12.5.13. Suppose that V is a T-cyclic space generated by $v \in V$, and let $d = \deg m_T(t)$.

(a) $\mathcal{B} := \{v, Tv, \cdots, T^{d-1}v\}$ is an F-basis of V. Hence, $\deg m_T(t) = \dim_F V$.

(b) $\phi_T(t) = m_T(t)$.

(The following converse of (b) will be proved in Proposition 15.3.7: If $\phi_T(t) = m_T(t)$, then V is T-cyclic.)

Proof. It suffices to prove that $\mathcal{B}$ is an F-basis of V, which follows easily from the minimality of $\deg m_T(t)$. Checking details are left as an exercise. □

By considering the natural F-basis for a T-cyclic space as given above, we can naturally introduce the companion matrix of a polynomial. In Lemma 12.5.13, if $m_T(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$, then

the matrix representation of T with respect to $\mathcal{B}$ is given as follows (this basis is called a T-cyclic basis of V):

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 0 & & & & & & -a_0 \\ 1 & 0 & & & & & -a_1 \\ & 1 & 0 & & & & -a_2 \\ & & \ddots & \ddots & & & \vdots \\ & & & \ddots & \ddots & & \vdots \\ & & & & 1 & 0 & -a_{n-2} \\ & & & & & 1 & -a_{n-1} \end{pmatrix}.$$

Remark that we proved $\phi_T(t) = m_T(t)$ when V is T-cyclic. Thus, we naturally consider the converse case; whether we can find a matrix over F whose minimal polynomial and characteristic polynomial are $\psi(t)$, where a nonconstant monic polynomial $\psi(t)$ over F is given.

Definition 12.5.14 (Companion matrix)**.** Given a nonconstant monic polynomial $\psi(t) = t^n + a_{n-1}^{n-1} + \cdots + a_1 t + a_0$ over F, define the companion matrix $C(\psi)$ of $\psi(t)$ by

$$C(\psi) = \begin{pmatrix} 0 & & & & & & -a_0 \\ 1 & 0 & & & & & -a_1 \\ & 1 & 0 & & & & -a_2 \\ & & \ddots & \ddots & & & \vdots \\ & & & \ddots & \ddots & & \vdots \\ & & & & 1 & 0 & -a_{n-2} \\ & & & & & 1 & -a_{n-1} \end{pmatrix}.$$

To state the result of the preceeding observation, $[T]_{\mathcal{B}}^{\mathcal{B}} = C(m_T)$ if $V = F[t]v$ is T-cyclic and $\mathcal{B}$ is a T-cyclic basis of V. In fact, this identity holds whenever $\psi(t)$ is a nonconstant monic polynomial over F.

Proposition 12.5.15. Suppose that a nonconstant monic polynomial $\psi(t)$ over F is given, and let $C = C(\psi)$ be the companion matrix of $\psi(t)$. Let $n = \textit{deg}\,\psi(t)$.

(a) $C^k e_1 = e_{1+k}$ whenever $0 \le k \le n-1$, so $F^n = F[t]e_1$ is C-cyclic.

(b) $m_C(t) = \psi(t) = \phi_C(t)$.

Proof. (a) follows easily from the given idenitity. Thus, $\phi_C(t) = m_C(t)$. To prove (b), it suffices to show that $\psi(t)$ annihilates e_1, which easily follows. □

Remark. Suppose that there is an F-basis $\mathcal{B} = \{v_i\}_{i=1}^n$ of V with respect to which T has the companion matrix of a monic polynomial $\psi(t) \in F[t]$ of degree n as the matrix representation, i.e., $[T]_{\mathcal{B}}^{\mathcal{B}} = C(\psi)$. Then $v_{1+k} = T^k v_1$ for $0 \le k \le n-1$, so V is a T-cyclic space generated by v_1, and $m_T(t) = \psi(t) = \phi_T(t)$. As a conclusion, if there is an F-basis with respect to which the matrix representation of the given linear operator on V is the companion matrix of $\psi(t) \in F[t]$, then V is T-cyclic and $\phi_T(t) = \psi(t) = m_T(t)$.

Notation. When W is a T-cyclic subspace of V generated by $w \in W$, the minimal polynomial of $T|_W$ will be denoted by $m_w(t)$ and will also be called the minimal polynomial of w. (Such abuse of notation is allowed, since W is generated by w and $F[t]$ is commutative.)

Observation 12.5.16. Given a nonzero vector w of V, let $W = F[t]w$. And let $d = \textit{deg}\,(m_w(t)) = \textit{dim}_F W$.

(a) Because W is a cyclic $F[t]$-submodule of V, $\mathcal{C} := \{w, Tw, \cdots, T^{d-1}w\}$ is an F-basis of V and $m_w(t) = \phi_{T|_W}(t)$.

(b) As observed earlier, $[T|_W]_{\mathcal{C}}^{\mathcal{C}} = C(m_w)$.

(c) Clearly, $m_w(t) | m_T(t)$.

In fact, (c) can be generalized as follows:

(d) If $\{v_1, \cdots, v_n\}$ is an F-basis of V, then $m_T(v) = \text{lcm}\{m_{v_1}(t), \cdots, m_{v_n}(t)\}$.

(d) can be justified as follows, letting $l(t) = \text{lcm}\{m_{v_1}(t), \cdots, m_{v_n}(t)\}$, it is clear that $l(t)|m_T(t)$ since each $m_{v_i}(t)$ divides $m_T(t)$; because $l(t)$ annihilates $F[t]v_i$ for all $i = 1, \cdots, n$, $l(t)$ is a multiple of $m_T(t)$, proving that $m_T(t) = l(t)$.

The following proposition will be essential when we apply the cyclic decomposition theorem on finite dimensional F-vector spaces.

Proposition 12.5.17. Suppose that $V = F[t]v$ is T-cyclic, and write $m_T(t) = d(t)e(t)$ for some monic polynomials with $\deg(d(t)) \geq 1$. And let $w = e(T)v$.

(a) $\ker d(T) = F[t]w$.

(b) $m_w(t) = d(t)$.

(c) Hence, $\dim_F \ker d(T) = \dim_F F[t]w = \deg d(t)$.

Proof. (a) Clearly, $F[t]w \subset \ker d(T)$. To show the converse inclusion, assume that $y \in \ker d(T)$ and write $y = f(T)v$ for some $f(t) \in F[t]$. Then $d(t)f(t)$ annihilates v, so $m_T(t)|d(t)f(t)$ and $e(t)|f(t)$. Thus, $y \in F[t]$, as desired.

(b) Note that $d(t)$ annihilates v so $m_w(t)|d(t)$. If $m_w(t)$ is a propoer divisor of $d(t)$, then $m_w(t)e(t)$ annihilates v but is a proper divisor of $m_T(t)$, a contradiction.

(c) It follows from (a) and (b). □

Chapter 13

Free modules

13.1 Remarks on free modules

Observation 13.1.1 (A free R-module has an R-basis). Let R be a ring with the nonzero identity, and assume that there is a free R-module $\mathcal{F}_R(S)$ genereated by a set S. Being a free object, we have $\mathcal{F}_R(S) = \langle S \rangle$, hence $\mathcal{F}_R(S)$ is the collection of all R-linear combinations of S. Moreover, since free objects are free from relations,

any two distinct R-linear combinations should be distinct.

In other words, if the two finite sums $\sum_{s\in S}(a_s \cdot s)$ and $\sum_{s\in S}(a_s \cdot s)$ (where $a_s, b_s \in R$ for each $s \in S$) are equal, then $a_s = b_s$ for each $s \in S$.

To sum up, if there is a free R-module generated by S, then S is an R-basis of $\mathcal{F}_R(S)$ and $\mathcal{F}_R(S) \approx \langle S \rangle \approx \bigoplus_{s\in S} R$.

Theorem 13.1.2 (Existence of free modules)**.** Let R be a ring with the nonzero identity and S be a set. Then $\bigoplus_{s\in S} R$ is a free R-module. Therefore, $\mathcal{F}_R(S) \approx \bigoplus_{s\in S} R$.

Proof. One needs to check if $\bigoplus_{s\in S} R$ satisfies a universal property of free R-modules. For this, it suffices to find an R-basis of $\bigoplus_{s\in S}$; one can easily verify that the collection $\{e_s : s \in S\}$ is an R-basis of $\bigoplus_{s\in S} R$, where $e_s \in \bigoplus_{s\in S} R$ is the element defined by $e_s(t) = \delta_{st}$ for all $s, t \in S$. Then, together with the injection $\imath : S \hookrightarrow \bigoplus_{s\in S} R$ defined by $\imath(s) = e_s$, one can easily check that $\bigoplus_{s\in S} R$ satisfies the universal property. □

If we state the preceeding theorem with its proof in other words, we can obtain the following result.

Theorem 13.1.3. The followings are equivalent:

(a) $\mathcal{F}$ is a free R-module.

(b) An R-module $\mathcal{F}$ has an R-basis.

Proof. The proof of [(a) implies (b)] is given in Observation 13.1.1, and the proof of [(b) implies of (a)] can easily be obtained by modifying the proof of Theorem 13.1.2: If $\mathcal{B}$ is an R-basis of an R-module $\mathcal{F}$, then $(\mathcal{F}, \imath)$ is a free R-module generated by $\mathcal{B}$, where $\imath : \mathcal{B} \hookrightarrow \mathcal{F}$ is the inclusion map. □

Chapter 14

Linear algebra over principal ideal domains - Part 1

Throughout this chapter, D denotes an integral domain and $Q = Q_D$ denotes the field of fractions of D.

14.1 Ranks of free modules

Idea: To start with Q^n to study D^n.

Example 14.1.1. We will explain that the $\mathbb{Z}$-modules $\mathbb{Z}^3$ and $\mathbb{Z}^2$ are not isomorphic. For this, it suffices to show that no pair of two elements of $\mathbb{Z}^3$ generate $\mathbb{Z}^3$. Assume that there is a pair of $x, y \in \mathbb{Z}^3$ generating $\mathbb{Z}^3$. Then, in particular, e_1, e_2, e_3 can be written as a $\mathbb{Z}$-linear combination of $\{x, y\}$. Hence, $\{x, y\} \subset \mathbb{Z}^3 \subset \mathbb{Q}^3$ generated the $\mathbb{Q}$-vector space $\mathbb{Q}^3$, implying that $dim_{\mathbb{Q}}\mathbb{Q}^3 \leq 2$, a contradiction. Therefore, $\mathbb{Z}^2 \not\approx \mathbb{Z}^3$.

Using the above idea, we first establish an essential theorem.

Theorem 14.1.2. Let I be a nonempty set and let $\mathcal{F}_D = \bigoplus_{i\in I} D$ and $\mathcal{F}_Q = \bigoplus_{i\in I} Q$. And identify $\mathcal{F}_D$ as a subset of $\mathcal{F}_Q$. And let S be a subset of $\mathcal{F}_D$.

(a) If S generates the (free) D-module $\mathcal{F}_D$, then S generates the Q-vector space $\mathcal{F}_Q$.

(b) S is a D-linearly independent subset of $\mathcal{F}_D$ if and only if S is a Q-linearly independent subset of $\mathcal{F}_Q$.

(c) Hence, if $\mathcal{B}$ is a D-basis of $\mathcal{F}_D$, then $\mathcal{B}$ is a Q-basis of $\mathcal{F}_Q$.

If M is a free D-module such that $M \approx \bigoplus_{i\in A} D$ for some nonempty set A and $\mathcal{B}$ is a D-basis of M, then $\mathcal{B}$ and A are in bijection.

Proof. (a) If S generates $\mathcal{F}_D$, then in particular, e_s is a D-linear combination of S for each $s \in S$. Since $\{e_s : s \in S\}$ is a Q-basis of $\mathcal{F}_Q$, it follows that S generates $\mathcal{F}_Q$.

(b) It is clear that S is D-linearly independent if it is Q-linearly independent. Suppose conversely that S is D-linearly independent but S may be Q-linearly dependent. By reducing fractions, we can find a nontrivial D-linear combination of $0 \in \mathcal{F}_D$ by elements in S, which contradicts the assumption.

(c) follows from (a) and (b).

Since M is a free D-module, there is a set A such that $M \approx \bigoplus_{a\in A} D$. Assume that A is nonempty, and let $\phi : M \xrightarrow{\approx} \bigoplus_{a\in A} D$ be a D-module isomorphism. Then $\phi(\mathcal{B})$ is a D-basis of $\bigoplus_{a\in A} D$; by (c) of Theorem 14.1.2, we have $|\mathcal{B}| = |\phi(\mathcal{B})| = |A|$. □

Corollary 14.1.3. Suppose that S and T are nonempty sets. Then S and T are in bijection if and only if the free D-modules generated by S and T are isomorphic.

Proof. It is already proved that the free D-modules generated by S and T are isomorphic if S and T are in bijection. Assume that the free D-modules generated by S and T are isomorphic. Then $\bigoplus_{s\in S} D \approx \bigoplus_{t\in T} D$(,so an isomorphic image of a D-basis of $\bigoplus_{s\in S} D$ is a D-basis of $\bigoplus_{t\in T} D$). Hence, S and T are in bijection. □

Remark. In accordance with the above corollary, we can define the rank of a free D-module, whose counterpart in linear algebra over fields is the dimension of a vector space over a field; when M is a free D-module and $M \approx \bigoplus_{i\in I} D$ and both $\mathcal{B}$ and $\mathcal{C}$ are D-bases of M, then $\mathcal{B}$, I, and $\mathcal{C}$ are in bijections.

Definition 14.1.4. Given a free D-module M, if $\mathcal{B}$ is a D-basis of M, then $|\mathcal{B}|$ is called the rank of M, and we write $\mathit{rank}_D(M) = |\mathcal{B}|$.

Hence, to state (c) of Theorem 14.1.2, we can write as follows: $\mathit{rank}_D(\mathcal{F}_D) = \mathit{dim}_Q(\mathcal{F}_Q)$.

Corollary 14.1.5. Suppose that S and T are nonempty sets. Then S and T are in bijection if and only if the free groups generated by S and T are isomorphic.

Proof. Again, the free groups generated by S and T are isomorphic if S and T are in bijection. Assuming conversely and remarking that $\mathbb{Z}$ is an integral domain, we find that

$$\mathcal{F}_{\mathbb{Z}}(S) \approx \frac{\mathcal{F}_{\mathsf{gp}}(S)}{[\mathcal{F}_{\mathsf{gp}}(S), \mathcal{F}_{\mathsf{gp}}(S)]} \approx \frac{\mathcal{F}_{\mathsf{gp}}(T)}{[\mathcal{F}_{\mathsf{gp}}(T), \mathcal{F}_{\mathsf{gp}}(T)]} \approx \mathcal{F}_{\mathbb{Z}}(T).$$

By the preceeding corollary, S and T are in bijection. □

14.2 Finitely generated module over a principal ideal domain

Remark (Dimension theorem for vector spaces over fields). Suppose that F is a field and V, W are vector spaces over F. And let $L : V \to W$ be an F-linear map. Assuming that $\{w_i : i \in I\}$ is an F-basis of $\mathit{im}\,(L)$, let v_i be a vector in V for each $i \in I$ such that $L(v_i) = w_i$ and define $U = \sum_{i\in I} Fv_i$.

(a) $\{v_i : i \in I\}$ is an F-basis of U and $U \approx W$.

(b) $V = \mathit{ker}\,L \oplus U$.

In particular, if both V and W are finite dimensional, then $\mathit{dim}_F V = \mathit{dim}_F(\mathit{ker}\,L) + \mathit{dim}_F(\mathit{im}\,L)$.

In fact, the above dimension theorem extends to the following theorem.

Theorem 14.2.1 (Decomposition theorem for modules over integral domains)**.** Assume that D is an integral domain. Let M and N be D-modules and $\phi : M \to N$ be a D-module homomorphism *with a free image*, i.e., $\mathit{im}\,\phi$ is a free D-submodule of N. Then there is a free D-submodule $\mathcal{F}$ of M such that

$$M = \mathit{ker}\,\phi \oplus \mathcal{F} \quad \text{and} \quad \mathcal{F} \approx \mathit{im}\,\phi.$$

Proof. **Step 1. Setting a D-submodule**

Since $\mathit{im}\,\phi$ is a free D-module, there is a D-basis $\mathcal{B} = \{v_i : i \in I\}$ of $\mathit{im}\,\phi$. For each $i \in I$, choose $u_i \in M$ such that $\phi(u_i) = v_i$, and set $K = \sum_{i\in I} Du_i$.

Step 2. Justifying that the constructed submodule is isomorphic to the image of ϕ

Clearly, K is mapped onto $\mathit{im}\,\phi$ by the restriction of ϕ to K. Assuming that $\phi|_K(\sum_{i\in I} a_i u_i) = 0$, where the sum is finite, we have $a_i = 0$ for all $i \in I$, for $\sum_{i\in I} a_i v_i = 0$. Hence, $\phi|_K$ is a D-module isomorphism and $K \approx \mathit{im}\,\phi$.

Step 3. Deriving a desired result

Therefore, $M = K + \mathit{ker}\,\phi$. To show the sum is direct, assume that $x \in K \cap \mathit{ker}\,\phi$. Since $x \in K$, we may write $x = \sum_{i\in I} a_i u_i$, where the sum is finite; since $x \in \mathit{ker}\,\phi$, we have $0 = \phi(x) = \sum_{i\in I} a_i v_i$, implying that $a_i = 0$ for all $i \in I$. Thus, $K \cap \mathit{ker}\,\phi = 0$ and the sum $M = K + \mathit{ker}\,\phi$ is direct. □

Before studying further theory, we introduce some lemmas. Among the following lemmas, the first two lemmas could have been introduced in the preceeding sections, since they assume the ring to be an integral domain, rather than a PID.

Lemma 14.2.2. Free modules over integral domains are torsion-free.

Proof. Let D be an integral domain and M be a free D-module. And let $\{x_i : i \in I\}$ be a D-basis of M, and suppose that x is a torsion element of M which is annihilated by a nonzero scalar $r \in D$. Writing $x = \sum_{i\in I} a_i x_i$ (the sum is assumed to be finite), from $rx = 0$ we have $ra_i = 0$ for all $i \in I$. Since D is an integral domain and $r \neq 0$, we have $a_i = 0$ for all $i \in I$. Therefore, $x = 0$ and M is torsion-free. □

Lemma 14.2.3. Let D be an integral domain and M be a free D-module. If x is a nonzero element of M, then $D \approx Dx$ as D-modules.

Proof. Define the map $\alpha : D \to Dx$ by $\alpha(t) = tx$ for $t \in D$. One can easily check that α is a D-module epimorphism. If $\alpha(t) = 0$, because D is an integral domain, we have $t = 0$, hence α is a D-module isomorphism. □

In the remaining of this chapter, D is assumed to be a PID.

Lemma 14.2.4. Let D be a PID and M be a nonzero free D-module. If I is a nonzero ideal of D, then $I \approx D$ as D-modules. In other words, every ideal of the PID D is a free D-module of rank 1.

Proof. We can write $I = (a) = aD$ for some $a \in D\backslash\{0\}$. And define the map $\alpha : D \to I$ by $\alpha(x) = ax$ for $x \in D$. One can easily check that α is a D-module epimorphism. If $\alpha(x) = 0$, by the law of cancellation, we have $x = 0$. Therefore, α is a D-module isomorphism. □

The following theorem states a result what we have expected, but its proof is not simple.

Theorem 14.2.5. Let D be a PID, and let M be a *finitely generated* free D-module. Then a nonzero D-submodule N of M is also a free D-module, and $\mathit{rank}_D(N) \leq \mathit{rank}_D(M)$.

Proof. Let $r = \mathit{rank}_D(M)$ and $\mathcal{B} = \{x_1, \cdots, x_r\}$ is a D-basis of M. Since M is assumed to be finitely generated, we can prove the theorem by induction on r. (See in the following proof how the hypothesis that D is a PID is applied.)

Step 1. Proof for the case where $r = 1$

When $r = 1$, M is generated by a single element $x_1 \in M$, so $M = Dx_1$. Then a nonzero D-submodule N of M is generated by scalar multiples of x_1; because D is a PID, it can easily be explained that N is generated by one element of M. Thus, $N = D(cx_1)$ for some $c \in D \setminus \{0\}$. By Lemma 14.2.3, $N \approx M$, so N is a free D-module of rank 1.

Step 2. Proof by induction

Suppose that the theorem is valid for all free D-modules of rank less than r. Consider the D-module homomorphism $\pi_i : N \to Dx_i$ for $i = 1, \cdots, r$, where $\pi_i(\sum_{j=1}^{r} c_j x_j) = c_i x_i$ for $c_j \in D$. Since N is nonzero, $\pi_i(N)$ is nonzero for some index i; without loss of generality, assume $i = 1$.

(i) By Theorem 14.2.1, we have $N = \mathit{ker}\,\pi_1 \oplus \mathcal{F}$, where $\mathcal{F}$ is a D-submodule of N which is isomorphic to $\mathit{im}\,\pi_1 = Dx_1 \approx D$. Hence, $\mathcal{F}$ is a free D-module of rank 1.

(ii) $\mathit{ker}\,\pi_1$ is a D-submodule of $Dx_2 \oplus \cdots \oplus Dx_r \approx D^{r-1}$.By the induction hypothesis, $\mathit{ker}\,\pi_1$ is a free D-submodule of N and $\mathit{rank}_D(\mathit{ker}\,\pi_1) \leq r - 1$.

Therefore, N is a free D-submodule of M and $\mathit{rank}_D(N) \leq \mathit{rank}_D(M)$. □

Corollary 14.2.6. If D is a PID, then the submodules of a finitely generated D-module are also finitely generated.

Proof. Note that an object is a homomorphic image of a free object. Let M be a D-module generated by a finite set S, and consider the free D-module $\mathcal{F}_D(S)$ generated by S. If $\jmath : S \hookrightarrow M$ is the inclusion map, there is a unique D-module homomorphism $\tilde{\jmath} : \mathcal{F}_D(S) \to M$ extending $\jmath$, and $\tilde{\jmath}$ is clearly surjective. By Theorem 14.2.5, D-submodule $\tilde{\jmath}^{-1}(N)$ of $\mathcal{F}_D(S)$ is a free D-module of finite rank, hence its image $N = \tilde{\jmath}(\tilde{\jmath}^{-1}(N))$ is finitely generated. □

We introduce some application of the preceeding theory inlinear algebra over PID.s.

Example 14.2.7 (Dimension theorem)**.** When V and W are finite dimensional vector spaces over a field F and $L : V \to W$ is an F-linear map, then $\dim_F V = \dim_F \ker L + \dim_F \operatorname{im} L$. As V and W are free F-modules of finite ranks, we assume as follows: Let D be a PID and let M, N be free D-modules of finite ranks, and let $\phi : M \to N$ be a D-linear map. We will show that

$$\operatorname{rank}_D(M) = \operatorname{rank}_D(\ker\phi) + \operatorname{rank}_D(\operatorname{im}\phi).$$

Since W is a finitely generated free D-module and D is a PID, $\operatorname{im}\phi$ is a free D-submodule of W. Hence, we can apply Theorem 14.2.1 to write $M = \ker\phi \oplus \mathcal{F}$, where $\mathcal{F} \approx \operatorname{im}\phi$. Note that $\ker\phi$ is a free D-submodule of V, for V is a finitely generated free D-module. Therefore, we obtain that $\operatorname{rank}_D(M) = \operatorname{rank}_D(\ker\phi) + \operatorname{rank}_D(\operatorname{im}\phi)$.

Example 14.2.8 (Coincidence of the rank and the *corresponding* dimension)**.** Let D be a PID and let $\mathcal{F}_D = \bigoplus_{i\in I} D$ and $\mathcal{F}_Q = \bigoplus_{i\in I} Q$, where I is a nonempty finite set. Identify $\mathcal{F}_D$ as a subset of $\mathcal{F}_Q$ and let S be a nonempty subset of $\mathcal{F}_D$.

Observe that $\sum_{x\in S} Dx$ is a D-submodule of the finitely generated free D-module $\mathcal{F}_D$, so $\sum_{x\in S} Dx$ is a free D-module. Hence, a D-basis of $\sum_{x\in S} Dx$ is a Q-basis of $\sum_{x\in S} Qx$, implying that

$$\operatorname{rank}_D\left(\sum_{x\in S} Dx\right) = \dim_Q\left(\sum_{x\in S} Qx\right).$$

Using the above coincidence, one can prove the rank theorem for finite dimensional matrices over PID.s. Let A be an $m \times n$ matrix over D. Then $\operatorname{rank}_D C_D(A) = \dim_Q C_Q(A)$ and $\operatorname{rank}_D C(A^\star) = \dim_Q C_Q(A^\star)$(where the subscripts in the column spaces denotes the collection of scalars), where $\star = T$ or $\star = H$. Because $\dim_Q C_Q(A) = \dim_Q C_Q(A^\star)$, we obtain the rank theorem for finite dimensional matrices over PID.s.

Before proving another structure theorem, we introduce a lemma which characterizes a finitely generated free module over a PID.

Lemma 14.2.9. If D is a PID, then a finitely generated *torsion-free* D-module is *free*.

Proof. Let S be a finite set generating a torsion-free module M over D, and let $T = \{x_1, \cdots, x_s\}$ be a maximal D-linearly independent subset of S. Then $\langle T\rangle = \bigoplus_{i=1}^{s} Dx_i$ is a free D-module of rank s with a D-basis T.

Want to show: For each $x \in S$, there is a nonzero element $m_x \in D$ such that $m_x x \in \langle T\rangle$.

(The above statement is clearly valid if $x \in T$. If $x \in S \setminus T$, because $T \sqcup \{x\}$ is D-linearly dependent, $a_1x_1 + \cdots + a_sx_s + bx = 0$ implies $b \neq 0$.)

Since D is an integral domain and S is finite, the product m of m_x for $x \in S$ is nonzero, and $mx \in \langle T\rangle$ for all $x \in S$. Now, consider the map $\phi : M \to \langle T\rangle$ defined by $\phi(x) = mx$ for $x \in M$. (Because M is generated by S, the map ϕ is a well-defined D-module homomorphism.) *Because M is torsion-free*, $\ker\phi = 0$. Therefore, by the first isomorphism theorem, $M \approx \phi(M) = \langle T\rangle$, so M is free. □

Remark. It is clear that free modules over integral domains are torsion-free (and we already proved it). The preceeding proposition states that a finitely generated module over a PID is free if it is torsion-free.

Conclusion: For finitely generated modules over PID.s, being free and being torsion-free coincide.

Corollary 14.2.10 (Structure theorem for finitely generated modules over PID.s)**.** Let D be a PID and M be a finitely generated D-module. Then there is a free D-submodule $\mathcal{F}$ of M such that

$$M = M_{\text{tor}} \oplus \mathcal{F}.$$

Furthermore, if $T_1 \oplus \mathcal{F}_1 = M = T_2 \oplus \mathcal{F}_2$ for some torsion D-submodules T_1, T_2 and free D-submodules $\mathcal{F}_1, \mathcal{F}_2$ of M, then $T_1 = T_2$ and $\mathcal{F}_1 \approx \mathcal{F}_2$. We call $\mathcal{F}$ the free part of M and its rank the free rank of M.

Proof. Since M/M_{tor} is torsion-free and finitely generated, M/M_{tor} is a free D-module, so we naturally consider the natural projection $\pi : M \twoheadrightarrow M/M_{\text{tor}}$. By Theorem 14.2.1, we have $M = M_{\text{tor}} \oplus \mathcal{F}$ for some free D-submodule $\mathcal{F}$ of M such that $\mathcal{F} \approx M/M_{\text{tor}}$.

We now justify the uniqueness part. Suppose that $T_1 \oplus \mathcal{F}_1 = M = T_2 \oplus \mathcal{F}_2$ for some torsion D-submodules T_1, T_2 and free D-submodules $\mathcal{F}_1, \mathcal{F}_2$ of M. Because $M_{\text{tor}} = (T_i \oplus \mathcal{F}_i)_{\text{tor}} = (T_i)_{\text{tor}} \oplus (\mathcal{F}_i)_{\text{tor}} = T_i$ for $i = 1, 2$, we have $T_1 = T_2 = M_{\text{tor}}$. Also, because $M/M_{\text{tor}} = M/M_i \approx \mathcal{F}_i$, we have $\mathcal{F}_1 \approx \mathcal{F}_2$. □

Remark. The above structure theorem states the following:

(a) (Existence) If M is a finitely generated module over a PID, then M is the direct sum of a torsion submodule and a free submodule.

(b) (Uniqueness) Furthermore, if there are two such expressions $T_1 \oplus \mathcal{F}_1 = M = T_2 \oplus \mathcal{F}_2$ (T_i is a torsion submodule of M and $\mathcal{F}_i$ is a free submodule of M for $i = 1, 2$), then the torsion submodules T_1, T_2 are the torsion part of M and the free submodules $\mathcal{F}_1, \mathcal{F}_2$ are isomorphic as modules. In short, $T_1 = T_2 = M_{\text{tor}}$ and $\mathcal{F}_1 \approx \mathcal{F}_2 \approx M/M_{\text{tor}}$.

Later in this chapter (after studying some linear algebra over $\mathbb{Z}$), we will study the decomposition of the torsion part of a finitely generated modules over PID.s.

We end this section with an obvious observation.

Example 14.2.11. Let R be a ring with the nonzero identity and let M be an R-module. Suppose that N is an R-submodule of M, and assume that both N and M/N are finitely generated. Then $N = \langle x_1, \cdots, x_n \rangle$ and $\overline{M} = M/N = \langle \overline{y_1}, \cdots, \overline{y_j} \rangle$ for some $x_1, \cdots, x_n, y_1, \cdots, y_j \in M$, and one can easily verify that $M = \langle x_1, \cdots, x_n, y_1, \cdots, y_j \rangle$, i.e., M is finitely generated.

14.3 Linear algebra over the ring of integers

Even stronger than as mentioned in the preceeding section, we assume that D is a Euclidean domain.

Definition 14.3.1 (D-elementary operations)**.** Any of the following operations is called a D-elementary operation:

(E1) Interchanging two distinct rows or columns

(E2) Adding the r-scalar multiple of a row (or column, respectively) to another row (column). ($r \in D$)

(E3) Multiplying a row or a column by $u \in D^\times$. (Why u should be a unit in D?)

Theorem 14.3.2. Let D be a Euclidean domain and suppose that $A \in \mathcal{M}_{m,n}(D)$. Then there are matrices $U \in GL_m(D), V \in GL_n(D)$ and elements $d_1, \cdots, d_r$ ($r = \min\{m, n\}$) such that

(1) U and V are products of D-elementary matrices.

(2) $UAV = \textit{diag}\{d_1, \cdots, d_r, 0, \cdots, 0\} \in \mathcal{M}_{m,n}(D)$ and $d_1 | d_2 | \cdots | d_r$.

Proof. Idea: Permuting rows and columns until we have multiples of a diagonal entry on every upper-triangular entries.

Read page 328 of your second textbook. □

Even though the above theorem is theoretically essential, it is not practically essential.

Proposition 14.3.3. Use the notations in Theorem 14.3.2.

(a) d_1 is a greatest common divisor of all the mn-entries of A. (It is proved when proving Theorem 14.3.2.)

(b) When $m = n$, since U and V are invertible, we have $\textit{det}(A) \sim_\times d_1 \cdot \cdots \cdot d_n$.

Proposition 14.3.4. If D is a Euclidean domain and A is an invertible $n \times n$-matrix over D, then A is a product of D-elementary matrices.

Proof. Using the notation in Theorem 14.3.2, each d_i has to be a unit in D. □

Observation 14.3.5. We now consider Theorem 14.3.2 in the situation when we change bases. Let N and M be free D-modules of rank n and m, respectively, and let $\mathcal{B} = \{x_1, \cdots, x_n\}$ and $\mathcal{C} = \{y_1, \cdots, y_m\}$ be D-bases of N and M, respectively. Suppose further that $\phi : N \to M$ is a D-module homomorphism, and let $A = [\phi]_{\mathcal{C}}^{\mathcal{B}} \in \mathcal{M}_{m,n}(D)$. Let U, V $d_i\,(1 \leq i \leq r = \min\{m, n\})$ be as in Theorem 14.3.2. Since U and V are invertible, we may identify them as transition matrices; in other words, there are D-bases $\mathcal{B}' = \{z_1, \cdots, z_n\}$ of N and $\mathcal{C}' = \{w_1, \cdots, w_m\}$ of M such that

$$UAV = [\mathit{id}_{D^m}]_{\mathcal{C}'}^{\mathcal{C}} \cdot [\phi]_{\mathcal{C}}^{\mathcal{B}} \cdot [\mathit{id}_{D^n}]_{\mathcal{B}}^{\mathcal{B}'} = [\phi]_{\mathcal{C}'}^{\mathcal{B}'}.$$

Therefore, the D-module homomorphism ϕ can be understood as the following D-module homomorphism:

$$\phi(z_j) = \begin{cases} d_j w_j & \text{if } 1 \leq j \leq r = \min\{m, n\} \\ 0 & \text{if } m < n \text{ and } m < j \leq n \end{cases}.$$

Corollary 14.3.6. Let D be a Euclidean domain and M be a free D-module of finite rank. For a D-sumodule N of M, (because N is also a free D-module of finite rank) we can let $n = \mathit{rank}_D(N) \leq \mathit{rank}_D(M) = m$. Then there are m-elements $w_1, \cdots, w_m \in M$ and n-scalars $d_1, \cdots, d_n \in D$ such that

(1) $\{w_1, \cdots, w_m\}$ is a D-basis of M.

(2) $\{d_1 w_1, \cdots, d_n w_n\}$ is a D-basis of N.

(3) $d_1 | d_2 | \cdots | d_n$.

Furthermore, we have

$$\frac{M}{N} \approx \frac{D}{(d_1)} \oplus \cdots \oplus \frac{D}{(d_n)} \oplus \overbrace{D \oplus \cdots \oplus D}^{(m-n)-D\text{'s}}$$

Proof. Consider the injection $\imath : N \hookrightarrow M$ and find $U, V, \{d_i : 1 \leq i \leq n\}$ as in Observation 14.3.5. Because $\imath(N) = N$ and $\imath(z_j) = d_j w_j$ for $1 \leq j \leq n$, $\{w_1, \cdots, w_m\}$ is a D-basis of M and $\{d_1 w_1, \cdots, d_n w_n\}$ is a D-basis of N.

Since $M = \bigoplus_{i=1}^m Dw_i$ and $N = \bigoplus_{i=1}^n Dd_i w_i$, by considering the map $\pi : M \to D/(d_1) \oplus \cdots \oplus D/(d_n) \oplus \underbrace{D \oplus \cdots \oplus D}_{(m-n)-D\text{'s}}$ defined by

$$\phi(a_1 w_1 + \cdots + a_m w_m) = (\overline{a_1}, \cdots \overline{a_n}, a_{n+1}, \cdots, a_m),$$

it can easily be checked that ϕ is a well-defined D-module epimorphism sith $\mathit{ker}\phi = N$, from which the desired isomorphism is derived. □

Remark. Using this corollary, we can explain the existence part of the cyclic decomposition of a finitely generated module over a Euclidean domain. To be precise, if that D is a Euclidean domain and X be a finitely generated D-module, then there are elements $d_1, \cdots, d_n \in D$ such that

(1) $d_1 | \cdots | d_n$ and

(2) $X \approx D/(d_1) \oplus D/(d_n) \oplus D^k$ for some integer $k \geq 0$.

To justify the above assertion, one should remark that a finitely generated object is isomorphic to a homomorphic image of the free object generated by a finite set. In this case, a finitely generated D-module X is isomorphic to a quotient of a free D-module of a finite rank. Applying the above corollary to the quotient of the free D-module justifies the assertion.

In the remaining of this section, we practice finding a $\mathbb{Z}$-basis of a $\mathbb{Z}$-submodule of $\mathbb{Z}^r$, where r is a positive integer. Before studying some particular examples, we first investigate the following proposition.

Proposition 14.3.7. Let $\{x_1, \cdots, x_n\}$ be a subset of $\mathbb{Z}^n$ and let $N = \sum_{i=1}^n \mathbb{Z}x_i$. Then the followings are equivalent:

(a) $\mathit{rank}_{\mathbb{Z}} N = n$.

(b) $\mathit{det}(x_1, \cdots, x_n) \neq 0$.

(c) $[\mathbb{Z}^n : N] < \infty$.

Also, in any of the above case, $[\mathbb{Z}^n : N] = \mathit{det}(x_1, \cdots, x_n)$.

Proof. Let $\mathcal{B} = \{x_1, \cdots, x_n\}$ and $\mathcal{E} = \{e_1, \cdots, e_n\}$, and consider the natural inclusion $\imath : N \hookrightarrow \mathbb{Z}^n$ so that we have the matrix representation $A := [\imath]_{\mathcal{E}}^{\mathcal{B}} = (x_1, \cdots, x_n) \in \mathcal{M}_{n,n}(\mathbb{Z})$.

[(a)⇒(b)] Since $\mathcal{B}$ is a $\mathbb{Z}$-basis of N, $\mathcal{B}$ is a $\mathbb{Q}$-basis of the $\mathbb{Q}$-vector space $\sum_{i=1}^n \mathbb{Q}x_i = \mathbb{Q}^n$.

[(b)⇒(c)] By Theorem 14.3.2, there are matrices $U, V \in GL_n(\mathbb{Z})$ such that UAV is diagonal. Because $\mathit{det}(A) \neq 0$, each diagonal entry of UAV is nonzero. By Corollary 14.3.6, $[\mathbb{Z}^n : N] = |d_1 \cdot \cdots \cdot d_n| (= |\mathit{det}(A)|) < \infty$.

[(c)⇒(a)] Since the index of N in $\mathbb{Z}^n$ is finite, by the isomorphism suggested in Corollary 14.3.6, the free part of $\mathbb{Z}^n/N$ should be zero. Therefore, $\mathit{rank}_{\mathbb{Z}}(N) = n$. □

Example 14.3.8. Let M be the free $\mathbb{Z}$-module $\mathbb{Z}^2$ of rank 2, and let $N = \mathbb{Z}(8, 10)^T \oplus \mathbb{Z}(18, 24)^T$. (For convinience, let $x_1 = (8, 10)^T$ and $x_2 = (18, 24)^T$.) Our goal is to find a $\mathbb{Z}$-basis of N which is easy to deal with.

One strategy is to consider the natural embedding $\imath : N \hookrightarrow \mathbb{Z}^2$. Letting $\mathcal{B} = \{x_1, x_2\}$ and $\mathcal{E} = \{e_1, e_2\}$, we have $A := [\imath]_{\mathcal{E}}^{\mathcal{B}} = \begin{pmatrix} 8 & 18 \\ 10 & 24 \end{pmatrix}$. After $\mathbb{Z}$-elementary operations, we obtain the following diagonal matrix:

$$UAV = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix},$$

where $U = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$ and $V = \begin{pmatrix} -2 & 9 \\ 1 & -4 \end{pmatrix}$. Since U and V are obtained via $\mathbb{Z}$-elementary operations, they can be considered transition matrices, i.e., a matrix representation of the identity map over $\mathbb{Z}^2$. Identifying U and V as transition matrices $[\mathit{id}_{\mathbb{Z}^2}]_{\mathcal{C}'}^{\mathcal{E}}$ and $[\mathit{id}_{\mathbb{Z}^2}]_{\mathcal{B}}^{\mathcal{B}'}$, respectively, we find that

$$\mathcal{B}' = \left\{(2, 4)^T, (0, -6)^T\right\}, \quad \mathcal{C}' = \left\{(1, 2)^T, (0, -1)^T\right\}.$$

Hence, we obtained another $\mathbb{Z}$-basis $\mathcal{B}'$ of N. Also, we can find that $\mathbb{Z}^2/N \approx \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$, so $[\mathbb{Z}^2 : N] = 12 = |\mathit{det}\, A|$.

Example 14.3.9. In this example, we seek to find a 'simple' $\mathbb{Z}$-basis of $N = \mathbb{Z}(-4, 2)^T + \mathbb{Z}(6, 4)^T + \mathbb{Z}(10, 10)^T \leq \mathbb{Z}^2$. (Simply write $x_1 = (-4, 2)^T, x_2 = (6, 4)^T, x_3 = (10, 10)^T$.)

We develop another strategy to construct a $\mathbb{Z}$-linear map and some bases with regard to which matrix representation of the $\mathbb{Z}$-linear map is (x_1, x_2, x_3). Let $\mathcal{E} = \{e_1, e_2, e_3\}, \mathcal{F} = \{e_1, e_2\}$, and $\phi : \mathbb{Z}^3 \to \mathbb{Z}^2$ be the map defined by $\phi(e_i) = x_i$, we have $\mathit{im}\,\phi = N$ and $A := [\phi]_{\mathcal{F}}^{\mathcal{E}} = \begin{pmatrix} -4 & 6 & 10 \\ 2 & 4 & 10 \end{pmatrix}$. We can also find after $\mathbb{Z}$-elementary operations that

$$UAV = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix},$$

where $U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $V = \begin{pmatrix} 1 & -1 & 5 \\ 0 & -2 & 15 \\ 0 & 1 & -7 \end{pmatrix}$. Again, since U and V are obtained via D-elementary operations, they are invertible so they can be considered transition matrices; let $\mathcal{B}'$ and $\mathcal{C}'$ be $\mathbb{Z}$-bases of $\mathbb{Z}^3$

and $\mathbb{Z}^2$, respectively, such that $[id_{\mathbb{Z}^3}]_{\mathcal{F}}^{\mathcal{B}'} = V$ and $[id_{\mathbb{Z}^2}]_{\mathcal{C}'}^{\mathcal{F}} = U$. Since ϕ is onto N and $[\phi]_{\mathcal{C}'}^{\mathcal{B}'} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$, (writing $\mathcal{B}' = \{v_1, v_2, v_3\}$ and $\mathcal{C}' = \{u_1, u_2\}$) N is generated by $\{\phi v_1 = 2u_1 = (-4, 2)^T, \phi v_2 = 2u_2 = (2, 0)^T, \phi v_3 = 0\}$. And finally, it is clear that

$$N = \mathbb{Z}\begin{pmatrix} -4 \\ 2 \end{pmatrix} \oplus \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

14.4 Another definition of the rank

Again, let D be a PID. and M be a finitely generated D-module. We have studied that the rank of a finitely generated D-module is determined by the free part of the module. For a vector space over a field, its dimension and the maximal cardinality of linearly independent subset coincided. The following theorem ensures that such intuition for vector spaces over fields also holds for modules over PID.s.

Theorem 14.4.1. The maximal number of D-linearly independent elements of M is the rank of M.

Proof. Let m be the maximal number of D-linearly independent elements of M. By the definition of the rank of a finitey generated module over a PID., it is clear that $\textit{rank}_D(M) \leq m$. If $\mathcal{B} = \{x_1, \cdots, x_r, x_{r+1}\}$ is a subset of M which is D-linearly independent ($r = \textit{rank}_D(M)$), then $\mathcal{B}$ is a D-basis of the D-submodule $Dx_1 \oplus \cdots \oplus Dx_{r+1}$ of M, thus M has a rank $(r+1)$ free D-submodule of M, a contradiction. Therefore, $\textit{rank}_D(M) = m$. □

Some properties regarding the rank of a module over a PID. now follows. The following theorem states that a finitely generated D-module M has rank n if and only if M is nearly isomorphic to D^n.

Theorem 14.4.2. Let M be a finitely generated D-module of rank n, and let $\{x_1, \cdots, x_n\}$ is a D-linearly independent subset of M.

(a) Then $N := Dx_1 \oplus \cdots \oplus Dx_n \approx D^n$, and M/N is a torsion D-module.

(b) Conversly, suppose that M contains a free D-submodule N of rank n and M/N is a torsion D-module. Then $\textit{rank}_D(M) = n = \textit{rank}_D(N)$.

Therefore, $\textit{rank}_D(M) = n$ if and only if there is a free D-submodule N of M such that $\textit{rank}_D(N) = n$ and M/N is a torsion D-module.

Proof. The former implication is easy to prove. To show the latter implication, it suffices to show that $S \cup \{x_1, \cdots, x_n\}$ is never D-linearly independent whenever $S \subset M$, where $\{x_1, \cdots, x_n\}$ is a D-basis of N. Suppose $y_1, \cdots, y_k \in M$. Then, for each $1 \leq i \leq k$, there is a nonzero element r_i of D such that $r_i y_i \in N$, and such r_i's gives a nontrivial D-linear combination of 0 by $\{x_1, \cdots, x_n, y_1, \cdots, y_k\}$. □

The following two propositions are what we have desired.

Proposition 14.4.3. Let A and B be D-modules of finite ranks. Then $\textit{rank}_D(A \oplus B) = \textit{rank}_D(A) + \textit{rank}_D(B)$.

Proof. Writing $A = A_{\text{tor}} \oplus \mathcal{F}_A$ and $B = B_{\text{tor}} \oplus \mathcal{F}_B$ ($\mathcal{F}_A$ and $\mathcal{F}_B$ are free parts of A and B, respectively), we have $(A \oplus B)/(\mathcal{F}_A \oplus \mathcal{F}_B) \approx A/\mathcal{F}_A \oplus B/\mathcal{F}_B$, so $(A \oplus B)/(\mathcal{F}_A \oplus \mathcal{F}_B)$ is a torsion D-module. Therefore, $\textit{rank}_D(A \oplus B) = \textit{rank}_D(A) + \textit{rank}_D(B)$. □

Proposition 14.4.4. Let M be a D-module of a finite rank and let N be a D-submodule of M. Then $\textit{rank}_D(M/N) = \textit{rank}_D(M) - \textit{rank}_D(N)$.

Proof. Write $h = \textit{rank}_D(M/N)$. If $\{x_1, \cdots, x_n\}$ is a D-basis ofthe free part of N and $\overline{\mathcal{B}} := \{\overline{y_1}, \cdots, \overline{y_h}\}$ is a D-basis of the free part of $\overline{M} = M/N$, then $\mathcal{C} := \{x_1, \cdots, x_n, y_1, \cdots, y_h\}$ is a D-linearly independent subset of M. Also, $M/\langle\mathcal{C}\rangle \approx (M/N)/(\langle\mathcal{C}\rangle/N) = \overline{M}/\langle\overline{\mathcal{B}}\rangle$ is a torsion D-module. Therefore, $\textit{rank}_D(M) = \textit{rank}_D(N) + \textit{rank}_D(M/N)$. □

Chapter 15

Linear algebra over principal ideal domains - Part 2

In the last chapter, we proved that when D is a PID. and M is a finitely generated D-module, then there is a decomposition of M into a free D-submodule and a torsion D-submodule. Moreover, if $M = T_1 \oplus \mathcal{F}_1 = T_2 \oplus \mathcal{F}_2$ are such decompositions (T_i is a torsion D-submodule of M and $\mathcal{F}_i$ is a free D-submodule of M for $i = 1, 2$), then $T_1 = T_2 = M_{\text{tor}}$ and $\mathcal{F}_1 \approx M/M_{\text{tor}} \approx \mathcal{F}_2$. Since the structure of free D-modules is now clear, we are now interested in the structure of the torsion part. Therefore, throughout this section, D is, again, a PID., and we assume that M is a finitely generated *torsion* D-module.

15.1 Primary decomposition

We start our argument with the following easy, but important discovery.

Observation 15.1.1. If M is a finitely generated *torsion* module over the PID. D, then $\textit{ann}_D(M) \neq 0$. In fact, if $M = \langle x_1, \cdots, x_n \rangle$ and $\textit{ann}_D(x_i) = (m_i)$ for some nonzero element $m_i \in D$ for each integer $1 \leq i \leq n$, then

$$\textit{ann}_D(M) = (\text{lcm}\{m_1, \cdots, m_n\}).$$

Proof. Suppose that M is generated by $\{x_1, \cdots, x_n\} \subset M$. Since M is a *torsion* module, for each integer $1 \leq i \leq n$, there is a nonzero element $r_i \in D$ such that $r_i x_i = 0$. Since D is integral, the product r of r_i's is nonzero, thus $\textit{ann}_D(M) \neq 0$, for $r \in \textit{ann}_D(M)$.

We now justify the second assertion. Letting $\textit{ann}_D(M) = (a)$ for some (nonzero) element $a \in D$ and $m = \text{lcm}\{m_1, \cdots, m_n\}$, $m|a$, because a annihilates each x_i so that $m_i|a$. Conversely, $a|m$, because m annihilates any D-linear combination of $\{x_1, \cdots, x_n\}$. □

Example 15.1.2. We introduce some typical examples of finitely generated torsion modules over PID.s. Further theory on such objects will be discussed in Section 15.3.

(a) (Finite abelian groups) Suppose that A is an abelian group of finite order n. Considering A as a $\mathbb{Z}$-module, $n \in \textit{ann}_{\mathbb{Z}}(A)$, so we may write $\textit{ann}_{\mathbb{Z}}(A) = (m)$ for some nonzero integer m. Clearly, m divides n.

(b) (F-vector spaces as $F[t]$-modules) As we have done, we only consider finite dimensional vector spaces only; let V be a finite dimensional vector space over F and let T be an F-linear operator on V. Then the annihilator ideal I_T of T in $F[t]$ is nonzero and is generated by a unique monic polynomial over F, called the minimal polynomial of T. In fact, by the definition of the annihilator ideal of T, it is clear that $\textit{ann}_{F[t]}(V) = I_T$, so $\textit{ann}_{F[t](V)} = (m_T(t))$.

Notation. Given a finitely generated torsion D-module M, let $\mathcal{P}$ denote a complete set of representatives of irreducible elements in D modulo units. For example, if $D = \mathbb{Z}$, we set $\mathcal{P}$ be the set of all positive prime numbers; if $D = F[t]$, we set $\mathcal{P}$ be the set of all monic irreducible polynomials over F.

Also, given $p \in \mathcal{P}$, we call the following set

$$M(p) = \{x \in M : p^i x = 0 \text{ for some positive integer } i\} = \bigcup_{i=1}^{\infty} Ann_M(p^i)$$

the p-torsion submodule of M.

We now start preparation of proving the primary decomposition theorem.

Proposition 15.1.3. $M(p) \neq 0$ if and only if $Ann_M(p) \neq 0$.

Proof. It suffices to prove that $Ann_M(p) \neq 0$ if $M(p) \neq 0$. Assume that $Ann_M(p) = 0$ and let k be *the smallest* positive integer such that $Ann_M(p^k) \neq 0$ (such k exists, since $M(p) \neq 0$). If x is a *nonzero* element of $Ann_M(p^k)$, then $px \in Ann_M(p^{k-1})$, so $px = 0$ by the minimality of k. Then $x \in Ann_M(p)$, a contradiction. Therefore, $Ann_M(p) \neq 0$. □

Proposition 15.1.4. Let M be a finitely generated torsion D-module and p be an element of $\mathcal{P}$, and write $ann_D(M) = (m)$. If $p|m$, then $Ann_M(p) \neq 0$, so $M(p) \neq 0$. Hence, if $M(p) = 0$, then $p \nmid m$.

Proof. Suppose that p is an element of $\mathcal{P}$ dividing m. We may write $m = pk$ for some $k \in D$, and there is an element $x \in M$ such that $mx = 0$ but $kx \neq 0$; otherwise, since k annihilates M, we have $(k) = ann_D(M) = (p)$, a contradiction. Then $kx \in Ann_M(p) \leq M(p)$, proving the lemma. □

Remark. The converse of Proposition 15.1.4 will be proved in this section after proving the primary decomposition theorem.

Theorem 15.1.5 (Primary decomposition theorem)**.** Let D be a PID and M be a nonzero finitely generated torsion D-module. Write $ann_D(M) = (m)$ and let $m = {p_1}^{f_1} \cdots {p_k}^{f_k}$ be the factorization of m into irreducible factors, where $p_i \in \mathcal{P}$ and $f_i \geq 1$ for all $i = 1, \cdots, k$.

(a) $Ann_M(m) = M = Ann_M({p_1}^{f_1}) \oplus \cdots \oplus Ann_M({p_k}^{f_k})$.

(b) $ann_D(Ann_M({p_i}^{f_i})) = ({p_i}^{f_i})$ for all $i = 1, \cdots, k$.

Lemma 15.1.6 (Primary decomposition theorem)**.** In Theorem 15.1.5, write $m = ab$ for some $a, b \in D$, where $(a, b) = D$.

(a) $Ann_M(m) = M = Ann_M(a) \oplus Ann_M(b)$.

(b) $ann_D(Ann_M(a)) = (a)$ and $ann_D(Ann_M(b)) = (b)$.

Proof of Lemma 15.1.6. Since a and b are relatively prime, there are elements $s, t \in D$ such that $as + bt = 1$. Hence, for any $x \in M$, $x = 1x = b(tx) + a(sx) \in Ann_M(a) + Ann_M(b)$. Furthermore, if $x \in Ann_M(a) \cap Ann_M(b)$, then $ax = bx = 0$, so $x = 1x = s(ax) + t(bx) = 0$, so $M = Ann_M(a) \oplus Ann_M(b)$, proving (a).

To prove (b), note that it is clear that $a \in ann_D(Ann_M(a))$ and $b \in ann_D(Ann_M(b))$. Thus, letting $ann_D(Ann_M(a)) = (m_a)$ and $ann_D(Ann_M(b)) = (m_b)$ for some $m_a, m_b \in D$, we find that $m_a|a$ and $m_b|b$, so m_a and m_b are relatively prime. Thus,

if one can show that $ab = m \sim_\times m_a m_b$, then it follows that $(a) = (m_a)$ and $(b) = (m_b)$.

From (a), we have $ann_D(M) = (\text{lcm}\{m_a, m_b\}) = (m_a m_b)$, so $m \sim_t imes m_a m_b$, as desired. □

Proof of Theorem 15.1.5. Use the results of Lemma 15.1.6 inductively. □

Corollary 15.1.7. Use the notations in Theorem 15.1.5.

(a) If $p \in \mathcal{P}$ and $p \nmid m$, then $M(p) = 0$.

(b) $Ann_M({p_i}^{f_i}) = M(p_i)$ and $M = \bigoplus_{p \in \mathcal{P}} M(p)$.

(c) If $f_i \leq e_i$ for all $i = 1, \cdots, k$, then $Ann_M({p_i}^{f_i}) = Ann_M({p_i}^{e_i})$.

Proof. To prove (a), assume $x \in M(p)$. For each $i = 1, \cdots, k$, there is a unique element $x_i \in \mathit{Ann}_M({p_i}^{f_i})$ such that $x = x_1 + \cdots + x_k$. Since $p^l x = 0$ for some positive integer l, $p^l x_i = 0$ for each i; because $\gcd\{p^l, {p_i}^{f_i}\} \sim_\times 1$ for each i, we have $x_i = 0$ for each i, so $x = 0$.

To prove (b), it suffices to prove that $M(p_i) \leq \mathit{Ann}_M({p_i}^{f_i})$ for all $i = 1, \cdots, k$. Suppose that $x \in M(p_i)$ and write $x = x_1 + \cdots + x_k$ as in the proof of (a). If l is a positive integer such that $p_i^l x = 0$, we have $p_i^l x_j = 0$ for all $j = 1, \cdots, k$. In particular, if $j \neq i$, then $x_j = 0$, for p_i and p_j are relatively prime. Hence, $x = x_i \in \mathit{Ann}_M({p_i}^{f_i})$ and $M(p_i) = \mathit{Ann}_M({p_i}^{f_i})$ for each i. And we have $M = \bigoplus_{i=1}^k \mathit{Ann}_M({p_i}^{f_i}) = \bigoplus_{p \in \mathcal{P}} M(p)$ by (a).

(c) naturally follows from (b), since $M(p_i) = \mathit{Ann}_M({p_i}^{f_i}) \leq \mathit{Ann}_M({p_i}^{e_i}) \leq M(p_i)$. □

Remark (The weak Cauchy's theorem). (a) of the preceeding corollary and Proposition 15.1.4 directly yields the following equivalence, which is called the weak Cauchy's theorem:

> Suppose $p \in \mathcal{P}$. Then p divides m if and only if the p-torsion submodule of M is nonzero.

In short, the p-torsion submodule is nonzero if and only if p divides m. Moreover, by considering Proposition 15.1.3, we can establish the following equivalence:

(i) p divides m.

(ii) The p-torsion submodule of M is nonzero.

(iii) The D-submodule of M annihilated by p is nonzero.

In fact, Theorem 15.1.5 can be generalized to the following proposition, where there is a product of pairwise relatively prime elements which is divisible by m.

Proposition 15.1.8. Let D be a PID and M be a nonzero finitely generated torsion D-module. Write $\mathit{ann}_D(M) = (m)$ and suppose that $m | a_1 \cdots a_k$ for some nonzero and nonunit pairwise relatively prime elements $a_1, \cdots, a_k \in D$.

(a) $\mathit{Ann}_M(m) = M = \mathit{Ann}_M(a_1) \oplus \cdots \oplus \mathit{Ann}_M(a_k)$.

(b) Writing $\mathit{ann}_D(\mathit{Ann}_M(a_i)) = (m_i)$ for each $i = 1, \cdots, k$, we have $m_i | a_i$ for each i.

Proof. In fact, it suffices to prove the proposition for $k = 2$; the general case follows from (b) by induction on k.

Assume $k = 2$, and let $s, t \in D$ be elements such that $sa_1 + ta_2 = 1$. If $x \in M$, then $x = 1x = ta_2x + sa_1x \in \mathit{Ann}_M(a_1) + \mathit{Ann}_M(a_2)$, so $M = \mathit{Ann}_M(a_1) + \mathit{Ann}_M(a_2)$. To prove that the sum is direct, assume that $y \in \mathit{Ann}_M(a_1) \cap \mathit{Ann}_M(a_2)$. Then $y = 1y = sa_1y + ta_2y = 0 + 0 = 0$, so the internal sum is direct.

To show that $m_i | a_i$, it suffices to justify that $a_i \in (m_i) = \mathit{ann}_D(\mathit{Ann}_M(a_i))$, which is obvious. □

15.2 Cyclic decomposition

Before studying cyclic decompostion, we first observe the following result which is valid for modules over any rings with the ninzero identities.

Observation 15.2.1. Let R be a ring with the nonzero identity and M be an R-module. Given an element $x \in M$, define a map $\rho_x : R \to M$ by $\rho_x(r) = rx$ for $r \in R$.

(a) ρ_x is an R-module homomorphism with $\mathit{ker}\,\rho_x = \mathit{ann}_R(x)$ and $\mathit{im}\,\rho_x = Rx$. Hence, $R/\mathit{ann}_R(x)$ is a cyclic R-module which is isomorphic to Rx.

(b) Conversely, any cyclic R-submodule of M is isomorphic to $R/\mathit{ann}_R(v)$ for some $v \in M$. In fact, if N is the cyclic R-submodule of M generated by $v \in M$, by considering the R-module homomorphism ρ_v, we can deduce that $R/\mathit{ann}_R(v) \approx N$.

In particular, if D is a PID. and if we write $\mathit{ann}_D(x) = (a)$ for a given $x \in M$, then $Dx \approx D/\mathit{ann}_D(x) = D/(a)$.

In the preceeding section, we found that a finitely generated torsion D-module is an internal direct sum of D-submodules whose annihilator ideals are of the form (p^f) for some $p \in \mathcal{P}$ and a positive integer f. The cyclic decomposition states a decomposition of such D-modules. Hence, in this section, we assume further that $\mathit{ann}_D(M) = (p^f)$ for some $p \in \mathcal{P}$ and a positive integer f.

Before studying cyclic decomposition theorem, we first observe the following basic lemma.

Lemma 15.2.2. Suppose that M is a finitely generated torsion D-module, where D is a PID., and assume that $\mathit{ann}_D(M) = (p^f)$ for some $p \in \mathcal{P}$ and a positive integer f.

(a) If $x \in M$ is nonzero, then there is a positive integer r with $r \leq f$ such that $\mathit{ann}_D(x) = (p^r)$.

(b) If $M = Dx_1 + \cdots + Dx_h$ and $\mathit{ann}_D(x_i) = ({p_i}^{r_i})$ for each $i = 1, \cdots, h$, then $f = \max\{r_1, \cdots, r_h\}$.

Proof. (a) easily follows, since $\mathit{ann}_D(M) \trianglelefteq \mathit{ann}_D(x)$ so $\mathit{ann}_D(x) = (p^r)$ for some integer $0 \leq r \leq f$.

We now prove (b). For simplicity, let $m = \max\{r_1, \cdots, r_h\}$. Then p^m annihilates every D-linear combination of $\{x_1, \cdots, x_h\}$, so $p^m \in \mathit{ann}_D(M) = (p^f)$ and $f \leq m$. Conversely, since $\mathit{ann}_D(x_i) \trianglelefteq \mathit{ann}_D(M)$, we have $r_i \leq f$ for each $i = 1, \cdots, h$, so $m \leq f$. □

Theorem 15.2.3 (Cyclic decomposition theorem (Form I))**.** Let D be a PID. and M be a finitely generated torsion D-module, and write $\mathit{ann}_D(M) = (p^f)$ for some $p \in \mathcal{P}$ and $f \geq 1$. Then there exist

- a positive integer h
- $x_1, \cdots, x_h \in M$
- positive integers $r_1, \cdots, r_h$

satisfying the following properties:

(i) $M = Dx_1 \oplus \cdots \oplus Dx_h$.

(ii) For each $j = 1, \cdots, h$, $\mathit{ann}_D(x_j) = (p^{r_j})$ with $f = r_1 \geq \cdots \geq r_h \geq 1$.

Moreover, such integers h and $r_1, \cdots, r_h$ are unique, i.e., they are independent of the choice of $x_1, \cdots, x_h$.

Proof. We prove the existence part by induction on n, where n is the minimum number of the elements of a subset of M which generates M. If M is cyclic, the result is clear, so we may assume that M is not cyclic. Suppose that the result holds for any torsion D-modules generated by less than n-elements. And let $\{x_1, y_2, \cdots, y_n\}$ be a subset of M generating M and assume $\mathit{ann}_D(x_1) = (p^f)$. Now consider the quotient $\overline{M} = M/Dx_1$.

Step 1. Finding appropriate $x_2, \cdots, x_h \in M$

Note that $\overline{M}$ is a torsion D-module and $\overline{M} \neq 0$, since M is not cyclic. Because $\overline{M}$ is generated by $\{\overline{y_2}, \cdots, \overline{y_n}\}$, by induction hypothesis, there is a positive integer $h \geq 2$ and there are elements $\overline{\theta_2}, \overline{\theta_h} \in \overline{M}$ such that

$$\overline{M} = D\overline{\theta_2} \oplus \cdots \oplus D\overline{\theta_h}.$$

By Lemma 15.2.4, for each $j = 2, \cdots, h$, there is an element $x_j \in M$ such that

$$\overline{x_j} = \overline{\theta}_j \quad \text{and} \quad \mathit{ann}_D(x_j) = \mathit{ann}_D(\overline{\theta_j}).$$

Step 2. Proving the exsitence part

In this step, we will prove the following assertion:

$$M = Dx_1 \oplus \cdots \oplus Dx_h.$$

(1) Given $x \in M$, there are scalars $a_2, \cdots, a_h \in D$ such that $\overline{x} = a_2\overline{x_2} + \cdots + a_h\overline{x_h}$ (and such a_j is uniquely determined for each $j = 2, \cdots, h$), hence $x = a_1x_1 + a_2x_2 + \cdots + a_hx_h$ for some $a_1 \in D$.

(2) To show that the internal sum of Dx_i's is direct, assume that $a_1x_1 + a_2x_2 + \cdots + a_hx_h = 0$ for some scalars $a_1, a_2, \cdots, a_h \in D$. (Note that a_ix_i is a typical element of Dx_i for each $i = 1, \cdots, h$.) Then $a_2\overline{x_2} + \cdots + a_h\overline{x_h} = \overline{0}$, so $a_j\overline{x_j} = 0$ for each $j = 2, \cdots, h$. Since $ann_D(x_j) = ann_D(\overline{\theta_j})$ and $\overline{x_j} = \overline{\theta_j}$, we have $a_jx_j = 0$ $(j = 2, \cdots, h)$. It follows that $a_1x_1 = 0$, so the internal sum of Dx_i's is direct.

The proof of the uniqueness part will be given in the end of this section. □

Lemma 15.2.4. In the proof of Theorem 15.2.3, suppose that $\overline{\theta} \in \overline{M}$ $(\theta \in M)$. Then there is an element $y \in M$ such that $\overline{y} = \xi$ and $ann_D(y) = ann_D(\xi)$.

Proof. Before reading this proof, note that the proof is quite technical.

Write $ann_D(\overline{\theta}) = (p^r)$, where it is clear that $r \leq f$. Choose any $z \in M$ such that $\overline{z} = \overline{\theta}$ and write $ann_D(z) = (p^s)$ (clearly, $r \leq s \leq f$). Because $p^r\overline{\theta} = \overline{0}$, $p^rz = bx_1$ for some $b \in D$; because $p^sz = 0$, $p^{s-r}bx_1 = 0$, so $p^f | p^{s-r}b$. Therefore, there is $c \in D$ such that $p^{s-r}b = cp^f$ and $b = cp^{f-s+r}$, so $p^r(z - cp^{f-s}x_1) = 0$.

Now define $y = z - cp^{f-s}x_1$. As observed in the preceeding paragraph, $p^ry = 0$, so $(p^r) \trianglelefteq ann_D(y)$. Conversely, (writing $ann_D(y) = (\alpha)$) because $\alpha y = 0$ and $\alpha\overline{y} = \alpha\overline{z}$, $\alpha \in ann_D(\overline{\theta}) = (p^r)$, proving that $ann_D(y) = ann_D(\overline{\theta})$. □

Combining the primary decomposition theorem, we obtain the following theorem.

Theorem 15.2.5 (Cyclic decomposition theorem (Form I-(i)))**.** Let D be a PID. and M be a finitely generated D-module, and write

$$ann_D(N) = (m) \quad \text{and} \quad m = {p_1}^{f_1} \cdots {p_k}^{f_k},$$

where each p_i belongs to $\mathcal{P}$ and $f_i \geq 1$. Then there are

- positive integers $h_1, \cdots, h_k$
- sets of module elements $\{x_{ij} \in M\}_{j=1}^{h_j}$ for each $i = 1, \cdots, k$
- sets of positive integers $\{r_{ij}\}_{j=1}^{h_j}$ for each $i = 1, \cdots, k$

satisfying the following properties:

(1) $M = \bigoplus_{i=1}^{k} \left(\bigoplus_{j=1}^{h_j} Dx_{ij} \right)$.

(2) $ann_D(x_{ij}) = (p_i^{r_{ij}})$ for each i and j.

(3) $f_i = r_{i1} \geq \cdots \geq r_{ih_i} \geq 1$ for each i.

Moreover, $\{h_i\}_i$ and $\{r_{ij}\}_{i,j}$ are uniquely determined and are independent of the choice of $\{x_{ij}\}_{i,j}$.

Proof. Clear. □

Combining the primary decomposition theorem together with Observation 15.2.1, we obtain the following theorem.

Theorem 15.2.6 (Cyclic decomposition theorem (Form I-(ii)))**.** Let D be a PID. and M be a finitely generated D-module, and write

$$ann_D(N) = (m) \quad \text{and} \quad m = {p_1}^{f_1} \cdots {p_k}^{f_k},$$

where each p_i belongs to $\mathcal{P}$ and $f_i \geq 1$. Then there are

- positive integers $h_1, \cdots, h_k$
- sets of positive integers $\{r_{ij}\}_{j=1}^{h_j}$ for each $i = 1, \cdots, k$

satisfying the following properties:

(1) $M \approx \bigoplus_{i=1}^{k} \left(\bigoplus_{j=1}^{h_j} D/({p_i}^{r_{ij}}) \right)$.

(2) $f_i = r_{i1} \geq \cdots \geq r_{ih_i} \geq 1$ for each i.

Moreover, $\{h_i\}_i$ and $\{r_{ij}\}_{i,j}$ are uniquely determined.

Proof. Clear. □

We call the data which consists of

$$\begin{aligned} m = {p_1}^{f_1} \cdots {p_k}^{f_k} &\quad \text{(the annihilator ideal of } M\text{)}, \\ \{h_i \in \mathbb{Z}^{>0} : 1 \leq i \leq k\} &\quad \text{(the number of cyclic summands in each primary summand)}, \\ \{r_{ij} \in \mathbb{Z}^{>0} : 1 \leq i \leq k, 1 \leq j \leq h_i\} &\quad \text{(the annihilator ideal of each cyclic summand)} \end{aligned}$$

the invariants of M. Clearly, the finitely generated D-module M is completely determined (up to isomorphism) by the invariants. And the uniqueness part of the cyclic decomposition theorem implies that any two finitely generated D-modules with distinct invariants are not isomorphic. This establishes the collection of the isomorphism types of finitely generated D-modules and the collection of invariants.

We are now concerened in proving the uniqueness part. For this, we introduce the following form of the cyclic decomposition theorem.

Theorem 15.2.7 (Cyclic decomposition theorem (Form II))**.** Let D be a PID. and M be a finitely generated D-module, and write $\mathit{ann}_D(M) = (m)$. Then there are nonzero and nonunit elements $d_1, \cdots, d_r \in D$ satisfying the following properties:

(1) $M \approx D/(d_1) \oplus \cdots \oplus D/(d_r)$.

(2) $d_1 | \cdots | d_r = m$.

Moreover, the positive integer r and the elements $d_1, \cdots, d_r \in D$ are determined uniquely (up to unit).

Proof. Consider the isomorphism given in Theorem 15.2.6 and apply the Chinese remainder theorem. (For example, if $m = p_1 p_2^3 p_3^4$ and

$$M \approx D/(p_1) \oplus D/(p_2^2) \oplus D/(p_2^2) \oplus D/(p_2^3) \oplus D/(p_3) \oplus D/(p_3^4),$$

enumerate the generators of ideals as follows

$$\begin{array}{ccc} & & p_1 \\ p_2^2 & p_2^2 & p_2^3 \\ & p_3 & p_3^4 \end{array}$$

to obtain $d_1 = p_2^2$, $d_2 = p_2^2 p_3$, and $d_3 = p_1 p_2^3 p_3^4 = m$. In fact,

$$\begin{aligned} M &\approx D/(p_1) \oplus D/(p_2^2) \oplus D/(p_2^2) \oplus D/(p_2^3) \oplus D/(p_3) \oplus D/(p_3^4) \\ &\approx D/(p_2^2) \oplus \left(D/(p_2^2) \oplus D/(p_3) \right) \oplus \left(D/(p_1) \oplus D/(p_2^3) \oplus D/(p_3^4) \right), \end{aligned}$$

and the Chinese remainder theorem implies that $D \approx D/(d_1) \oplus D/(d_2) \oplus D/(d_3)$.) □

Some more preparations are required. In the following propositions, assume that D is a PID. and $p \in \mathcal{P}$.

Observation 15.2.8. Let N be a D-module and suppose that $N_i \leq_D N$ for $i = 1, \cdots, r$ and

$$N = N_1 \oplus \cdots \oplus N_r.$$

We wish to show that a direct decomposition of N naturally inherits to the D-submodule of N annihilated by p; in other words, we wish to show that

$$\mathit{Ann}_N(p) = \mathit{Ann}_{N_1}(p) \oplus \cdots \oplus \mathit{Ann}_{N_r}(p).$$

The internal sum of $\mathit{Ann}_{N_i}(p)$ for all i is direct, since each $\mathit{Ann}_{N_i}(p)$ is a D-submodule of N_i. Also, it is clear that $\mathit{Ann}_{N_1}(p) + \cdots + \mathit{Ann}_{N_r}(p) \subset \mathit{Ann}_N(p)$. Thus, it remains to prove the converse inclusion; for this, write $x \in \mathit{Ann}_N(p)$ as $x = x_1 + \cdots + x_r$ with $x_i \in N_i$ for each i. Then $px_1 + \cdots + px_r = 0$, implying $x_i \in \mathit{Ann}_{N_i}(p)$ so that $\mathit{Ann}_N(p) \subset \mathit{Ann}_{N_1}(p) + \cdots + \mathit{Ann}_{N_r}(p)$.

Observation 15.2.9. If M is a D-module, by Theorem 15.2.7, there are nonzero and nonunit elements $d_1, \cdots, d_r \in D$ such that $M \approx D/(d_1) \oplus \cdots \oplus D/(d_r)$. In this observation, we study the structure of each M-submodule of the direct summand annihilated by p. In other words, we study the structure of each $Ann_{D/(d_i)}(p)$.

Given a nonzero element $d \in D$, consider the cyclic D-module $N = D/(d)$.

(a) Suppose $\overline{x} \in Ann_N(p)$ for some $x \in D$, where the overline notation is used to denote the quotient by (d). It is equivalent to $px \in (d)$, or equivalently, $d|px$.

  (i) Assume that p divides d and write $d = pb$ for some $b \in D$. Then $Ann_N(p) = bD/(pb)$, and $Ann_N(p)$ can be considered a $D/(p)$-vector space, for $Ann_N(p)$ is annihilated by p.

  (ii) Assume that p does not divide d. Then $d|x$, so $\overline{x} = \overline{0}$ and $Ann_N(p) = 0$.

(b) Under the situation of $p|d$, we wish to find the $D/(p)$-dimension of $Ann_N(p)$. Construct a map $\phi : D \to bD/(pb) = Ann_N(p)$ by $\phi(a) = \overline{ba}$ for $a \in D$. Because ϕ is a surjective D-module homomorphism with $ker\phi = (p)$, we have $Ann_N(p) = bD/(pb) \approx D/(p)$. Therefore, if p divides d, then $Ann_N(p)$ is a 1-dimensional $D/(p)$-vector space; if p does not divide d, then $Ann_N(p) = 0$.

Observation 15.2.10. Given a nonzero element $b \in D$, consider the (cyclic) D-module $N = D/(pb)$ and its D-submodule $pD/(pb)$. Construct a map $\psi : D \to pD/(pb)$ defined by $\psi(a) = \overline{pa}$, which is clearly a D-module homomorphism. Since ψ is surjective and $ker\psi = (b) \trianglelefteq D$, we have $pD/(pb) \approx D/(b)$. Then $pD/(pb) \approx D/(b)$.

Now we prove the uniqueness part of the cyclic decomposition theorem. In particular, we prove the uniqueness part of Theorem 15.2.7; this proves the uniqueness part of Theorem 15.2.6.

Proof of the uniqueness part of Theorem 15.2.7. We prove the uniqueness part by induction on the number of irreducible divisors of m, counting multiplicity. Suppose that there are two isomorphic types of M:

$$\begin{aligned} M &\approx D/(d_1) \oplus \cdots \oplus D/(d_r) \\ &\approx D/(c_1) \oplus \cdots \oplus D/(c_s) \end{aligned}$$

with $d_1|\cdots|d_r = m$ and $c_1|\cdots|c_s = m$, and all d_i's and c_j's being nonzero and nonunit. Choose an irreducible element $p \in \mathcal{P}$ such that $p|d_1$.

Step 1. Deducing $r = s$

By Observation 15.2.8, we have

$$Ann_M(p) \approx Ann_{D/(d_1)\oplus\cdots\oplus D/(d_r)}(p) = Ann_{D/(d_1)}(p) \oplus \cdots \oplus Ann_{D/(d_r)}(p).$$

Since $p|d_i$ for all i, writing $\overline{D} = D/(p)$, we have $r = dim_{\overline{D}}(Ann_M(p))$, thus r is the number of indices j for which $p|c_j$. So $r \leq s$; by symmetry, we have $s \leq r$. Thus, $r = s$ and $p|c_j$ for all j.

Step 2. Deducing $c_i \sim_\times d_i$ for all i

If m has (up to unit) one irreducible divisor, then $c_1 \sim_\times m \sim_\times d_1$, so the uniqueness is proved. To proceed the proof by induction, we consider the D-submodule pM of M. Writing $d_i = pd_i'$ and $c_i = dc_i'$ for each i, we have

$$pM \approx pD/(pd_1') \oplus \cdots \oplus pD/(pd_r') \approx D/(d_1') \oplus \cdots \oplus D/(d_r')$$

and

$$pM/ \approx D/(c_1') \oplus \cdots \oplus D/(c_r').$$

Then $d_r' = c_r'$ has a fewer irreducible divisors than $d_r = m = c_r$. Hence, by induction hypothesis, $d_i' \sim_\times c_i'$ for all i, thus $d_i \sim_\times c_i$ for all i. □

15.3 Applications of structure theorems

As assumed starting from the preceeding chapter, we assume that D is a PID. and M is a finitely generated D-module (by here, M need not be a *torsion* module). By the structure theorem, there is a decomposition of M into a torsion D-submodule and a free D-submodule. Moreover, if $M = T_1 \oplus \mathcal{F}_1 = T_2 \oplus \mathcal{F}_2$ are such decompositions (T_i is a torsion D-submodule of M and $\mathcal{F}_i$ is a free D-submodule of M for $i = 1, 2$), then $T_1 = M_{\text{tor}} = T_2$ and $\mathcal{F}_1 \approx M/M_{\text{tor}} \approx \mathcal{F}_2$. Since the free part of M has an obvious structure (being isomorphic to R^k for some nonnegative integer k), the structure of M_{tor} needs to be investigated to fully understand the structure of M. Thus, from now on, we assume that M is a finitely generated torsion D-module.

The first tool required to understand the structure of M is the primary decomposition theorem, which states that a finitely generated torsion module over a PID. can be understood as the internal direct sum of the p-torsion submodules for $p \in \mathcal{P}$; to be precise, if $\textit{ann}_M(D) = (m)$ with the factorization $m = {p_1}^{f_1} \cdots {p_k}^{f_k}$, then $M = \bigoplus_{i=k} \textit{Ann}_M({p_i}^{f_i}) = \bigoplus_{p \in \mathcal{P}} M(p)$. And the second tool (or together with the first tool) is the cyclic decomposition theorem, which determines the isomorphic type of the given torsion module. In fact, isomorphic torsion D-modules share the same invariants and non-isomorphic torsion D-modules have distinct invariants.

15.3.1 Finitely generated abelian groups

By the structure theorem, given a finitely generated abelian group A, we have $A = A_{\text{tor}} \oplus \mathcal{F}$ for some free $\mathbb{Z}$-submodule of A. Thus, we may assume that A is a torsion $\mathbb{Z}$-module.

Example 15.3.1. Let A be an abelian group of order $72 = 2^3 \cdot 3^2$. Letting $\textit{ann}_{\mathbb{Z}}(A) = (m)$ for some positive integer m, since $m|n$, $m = 2^a \cdot 3^b$, where a and b are integers such that $1 \leq a \leq 3$ and $1 \leq b \leq 2$.[1] By the primary decomposition theorem, $A = A(2) \oplus A(3)$ and $\textit{ann}_{\mathbb{Z}}(A(2)) = (2^a)$ and $\textit{ann}_{\mathbb{Z}}(A(3)) = (3^b)$. Using the cyclic decomposition theorem, the isomorphic types of $A(2)$ and $A(3)$ are given as follows:

$$\begin{aligned} A(2) &: Z_2 \oplus Z_2 \oplus Z_2\,(a=1),\ Z_2 \oplus Z_4\,(a=2),\ Z_8\,(a=3) \\ A(3) &: Z_3 \oplus Z_3\,(b=1),\ Z_9\,(b=2). \end{aligned}$$

Hence, there are 6 isomorphic types for A.

In fact, if A is an abelian group of order ${p_1}^{f_1} \cdots {p_k}^{f_k}$ with $p_1, \cdots, p_k$ being pairwise distinct positive prime numbers and $f_i \geq 1$ for all i, then $\textit{ann}_{\mathbb{Z}}(A) = (m)$ with $m = {p_1}^{e_1} \cdots {p_k}^{e_k}$ for some $1 \leq e_i \leq f_i$ for each i and $A = \bigoplus_{i=1}^{k} A(p_i)$ with $\textit{ann}_{\mathbb{Z}}(A(p_i)) = ({p_i}^{e_i})$ for each i. Because there are $\pi(f_i)$ isomorphic types for each $A(p_i)$, there are $\pi_(f_1) \cdots \pi(f_k)$ isomorphic types for A.

Proposition 15.3.2. Let A be a finite abelian group with $\textit{ann}_{\mathbb{Z}}(A) = (m)$ for some positive integer m. Then A is cyclic if and only if $|A| = m$.

Proof. Note that it is alwasy true that m divides the order of A. If A is cyclic, then there is an element x of A whose order is $|A|$, so $m = |A|$. Conversely, if $m = |A|$, then there is an element y of A annihiated by m but not by any proper divisor of m (otherwise, $\textit{ann}_{\mathbb{Z}}(A)$ would properly contain (m), a contradiction). Then $\langle y \rangle$ is a subgroup of A of order $|A|$, so A is cyclic. □

We now introduce an application of the above proposition in the field theory.

Proposition 15.3.3. Let F be a field and G be a finite subgroup of the multiplicative group $F^\times$. Then G is a cyclic group.

Proof. Let n be the order of G. Since G is abelian, by the cyclic decomposition theorem, there are positive integers $d_1, \cdots, d_k$ with $d_1 | \cdots | d_k$ such that $G \approx Z_{d_1} \times \cdots \times Z_{d_k}$ and $\textit{ann}_{\mathbb{Z}}(G) = (d_k)$. Hence, every element of G is a root of the polynomial $x^{n_k} - 1$, which has at most n_k-distinct roots. Therefore, $k = 1$ and G is cyclic. □

[1] Note that a, b cannot be zero, because there are elements of A of order 2 and 3, respectively.

15.3.2 Vector spaces over fields

Let V be an n-dimensional vector space over a field F and let T be an F-linear operator on V. Write

$$\phi_T(t) = p_1(t)^{e_1} \cdots p_k(t)^{e_k} \quad \text{and} \quad m_T(t) = p_1(t)^{f_1} \cdots p_k(t)^{f_k},$$

where $p_1(t), \cdots, p_k(t)$ are pairwise distinct monic irreducible polynomial over F and $1 \leq f_i \leq e_i$ for each $i = 1, \cdots, k$.

Applying the primary decomposition theorem on V

By the primary decomposition theorem, $V = \bigoplus_{i=1}^k \mathit{Ann}_V((p_i(t)^{f_i})) = \bigoplus_{i=1}^k \mathit{ker}(p_i(T)^{f_i})$. For notational convinience, let $W_i = \mathit{Ann}_V(p_i(t)^{f_i}) = \mathit{ker}(p_i(T)^{f_i})$ and $T_i = T|_{W_i}$ for each $i = 1, \cdots, k$. Then, whenever $\mathcal{B}_i$ is an F-basis of W_i and $\mathcal{B} = \bigsqcup_{i=1}^k \mathcal{B}_i$, then

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} [T_1]_{\mathcal{B}_1}^{\mathcal{B}_1} & & & \\ & [T_2]_{\mathcal{B}_2}^{\mathcal{B}_2} & & \\ & & \ddots & \\ & & & [T_k]_{\mathcal{B}_k}^{\mathcal{B}_k} \end{pmatrix}.$$

Hence, $m_T(t) = m_{T_1}(t) \cdots m_{T_k}(t)$ and $\phi_T(t) = \phi_{T_1}(t) \cdots \phi_{T_k}(t)$. The following results can also be deduced.

(i) $\mathit{ann}_{F[t]}(\mathit{Ann}_V(p_i(t)^{f_i})) = (p_i(t)^{f_i})$, so $m_{T_i}(t) = p_i(t)^{f_i}$. Also, $\phi_{T_i} = p_i(t)^{e_i}$, so $\mathit{dim}_F(W_i) = \mathit{deg}(p_i(t)^{e_i}) = e_i \mathit{deg}(p_i(t))$.

(ii) Whenever $g_i \geq f_i$, we have $\mathit{Ann}_V(p_i(t)^{f_i}) = V(p_i(t)) = \mathit{Ann}_V(p_i(t)^{g_i})$. Thus, in particular, $W_i = \mathit{ker}(p_i(T)^{e_i})$.

Before considering the cylcic decomposition, we deduce a criterion to determine if a given linear operator is diagonalizable.

Corollary 15.3.4 (Diagonalizability)**.** T is diagonalizable over F if and only if the minimal polynomial of T splits completely over F.

Proof. If T is diagonalizable over F and has eigenvalues $\lambda_1, \cdots, \lambda_k$, then $(t-\lambda_1) \cdots (t-\lambda_k)$ is the minimal polynomial of T. Conversely, if the minimal polynomial of T splits completely over F, then the primary decomposition of V with respect to T is the eigenspace decomposition of V, i.e., V is diagonalizable. □

Applying the cyclic decomposition theorem on each W_i

After determining the primary decomposition of V with respect to a given linear operator T, for each W_i, we can apply the first form of cyclic decomposition: For each primary summand W_i,

(i) there are T-invariant subspaces $U_{i1}, \cdots, U_{ih_i}$ of W such that $W_i = U_{i1} \oplus \cdots \oplus U_{ih_i}$.

(ii) Writing $\phi_{T|_{U_{ij}}}(t) = m_{T|_{U_{ij}}}(t) = p(t)^{r_{ij}}$ for each $j = 1, \cdots, h_i$, there are unique integer h_i and $r_{i1}, \cdots, r_{ih_i}$ such that $f_i = r_{i1} \geq \cdots \geq r_{ih_i} \geq 1$.

As illustrated in the preceeding section, the invariants

$$\{m_T(t), h_i, r_{ij} : 1 \leq i \leq k, 1 \leq j \leq h_i\}$$

determines the isomorphic types of V. In particular, since $\mathit{dim}_F W_i = \sum_{j=1}^{h_i} \mathit{dim}_F U_{ij} = \sum_{j=1}^{h_i} (r_{ij} \mathit{deg}\, p_i(t))$, the invariants determines the isomorphic types of V. In fact, the invariants determines the similarity class of the operator T.

Proposition 15.3.5. The invariants determine the similarity class of T.

Proof. As explained above, from the invariants we can restore the dimension of V. Considering a matrix representation of T restricted to the T-cyclic subspace with the minimal polynomial $p_i(t)^{r_{ij}}$ for each i and j, there is an F-basis $\mathcal{B}_{ij}$ of the T-cyclic subspace such that

$$[T|_{U_{ij}}]_{\mathcal{B}_{ij}}^{\mathcal{B}_{ij}} = C(p_i^{r_{ij}}).$$

This determines a matrix representation of T restricted to each primary summand up to similarity transform, hence determines a similarity class of T. □

Corollary 15.3.6. If it is given that the characteristic polynomial of T is $p_1(t)^{e_1} \cdots p_k(t)^{e_k}$, there are $\pi(e_1) \cdots \pi(e_k)$-distinct similarity classes of T.

In Lemma 12.5.13, it is proved that if V is T-cyclic then $\phi_T(t) = m_T(t)$. The statement and the proof of the converse is given as follows.

Proposition 15.3.7. Suppose that the characteristic polynomial and the minimal polynomial of T are the same. Then V is T-cyclic, i.e., V is a cyclic $F[t]$-module.

Proof. Using the notation of in Theorem 15.2.5, we have $e_i = f_i$ and $h_i = 1$ for all $i = 1, \cdots, k$. Therefore, by the Chinese remainder theorem

$$V \approx \frac{F[t]}{(p_1(t)^{f_1})} \oplus \cdots \oplus \frac{F[t]}{(p_k(t)^{f_k})} \approx \frac{F[t]}{(m_T(t))},$$

so V is a cyclic $F[t]$-submodule. □

Problem 15.3.1. Suppose that E/F be a field extension and $A, B \in \mathcal{M}_{n,n}(F)$. Show that $A \sim B$ over F if $A \sim B$ over E.

Solution. It suffices to show that the invariants computed over E and those computed over F are equal; then it is naturally deduced that the invariants of A and B computed over F are equal so A and B are similar over F. Let M be the rational canonical form of A computed over F, which is unique up to similarity transform. Since M can be considered a rational canonical form of A computed over E, by the uniqueness of the invariants, M is similar to the rational canonical form of A computed over E. This implies that the invariants of A computed over E and F are equal.

Remark. In fact, in the above solution, we proved a more general result: the invariants computed over a field and those computed over an extended field are identical. Transitivity of similarity is due to this general result and the uniqueness of cyclic decomposition. By the general result, it can easily be deduced that the minimal polynomial of a linear operator over a base field and the minimal polynomial computed over a subfield are equal.

15.3.3 Jordan canonical form

Assume that F is algebraically closed. Then we can write $m_T(t) = (t-\lambda_1)^{f_1} \cdots (t-\lambda_k)^{f_k}$ and $\phi_T(t) = (t-\lambda_1)^{e_1} \cdots (t-\lambda_k)^{e_k}$. Letting U_{ij} be the (i,j)-th T-cyclic subspace, define $N_{ij} = (T - \lambda_1 id_V)|_{U_{ij}}$. Then there is a T-cyclic basis $\mathcal{B}_{ij}$ such that

$$[N_{ij}]_{\mathcal{B}_{ij}}^{\mathcal{B}_{ij}} = \begin{pmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & 0 & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{pmatrix} \in \mathcal{M}_{r_{ij},r_{ij}}(F).$$

(By reversing $\mathcal{B}_{ij}$, we can obtain the transposed nilpotent matrix.) Hence,

$$[T_{U|_{ij}}]_{\mathcal{B}_{ij}}^{\mathcal{B}_{ij}} = \begin{pmatrix} \lambda_i & & & & \\ 1 & \lambda_i & & & \\ & 1 & \lambda_i & & \\ & & \ddots & \ddots & \\ & & & 1 & \lambda_i \end{pmatrix} \in \mathcal{M}_{r_{ij},r_{ij}}(F),$$

and gathering all the (reversed) bases, we obtain the Jordan canonical form of T.

Observation 15.3.8. Let $\mathcal{C}$ be the F-basis of V obtained by reversing an F-basis $\mathcal{B}$ of V. If $[T]_{\mathcal{B}}^{\mathcal{B}} = (a_{i,j})$, where $1 \leq i \leq n$ and $1 \leq j \leq n$, then

$$[T]_{\mathcal{C}}^{\mathcal{C}} = \begin{pmatrix} a_{n,n} & a_{n,n-1} & \cdots & a_{n,1} \\ a_{n-1,n} & a_{n-1,n-1} & \cdots & a_{n-1,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{1,n-1} & \cdots & a_{1,1} \end{pmatrix}.$$

In particular, applying this idea to a Jordan canonical form, we find that a Jordan canonical form is similar to its transposition over F; implying that $A \sim A^T$ over F.

Problem 15.3.2 (Exercise 12.2.15, *Abstract Algebra*, third edition)**.** Determine up to similarity all 2×2 matrices over $\mathbb{Q}$ and over $\mathbb{C}$, respectively.

Solution. Since A satisfies the polynomial $t^4 - 1 = (t-1)(t+1)(t^2+1)$ over $\mathbb{Q}$, the minimal polynomial of A over $\mathbb{Q}$ is among the following polynomials over $\mathbb{Q}$ (note that the degree of the minimal polynomial of A over $\mathbb{Q}$ is at most 2):

$$t-1, \quad t+1, \quad (t+1)(t-1), \quad t^2+1$$

The first three cases induce the rational canonical forms $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ over $\mathbb{Q}$, respectively, which are not of order 4. The last case induces the rational canonical form $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ over $\mathbb{Q}$ of order 4.

When A is understood as a matrix over $\mathbb{C}$, then the minimal polynomial of A over $\mathbb{C}$ divides $(t-1)(t+1)(t-i)(t+i)$. Among such cases, the only choices of the minimal polynomial $m(t)$ of A over $\mathbb{C}$ for which A is of order 4 are given as follows:

(i) $m(t) = t \pm i$, $A \sim \begin{pmatrix} \pm i & 0 \\ 0 & \pm i \end{pmatrix}$ (2 distinct similarity classes)

(ii) $m(t) = (t \pm i)(t \pm 1)$, $A \sim \begin{pmatrix} \pm i & 0 \\ 0 & \pm 1 \end{pmatrix}$ (4 distinct similarity classes)

Read also *The conjugacy classes of* $GL(2, \mathbb{F}_q)$, written by Harold Cooper.

Chapter 16

Further module theory

16.1 Tensor products of modules

16.2 Exact sequences, flat and projective modules

16.3 Some problems in module theory

Problem 16.3.1. Let p be a positive prime number and let n be a positive integer. And let $F : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ be the Frobenius map.

(a) Note that $F \in \mathrm{Aut}(\mathbb{F}_{p^n}/\mathbb{F}_p)$. Show that the order of F is n.

(b) Understand $\mathbb{F}_{p^n}$ as an n-dimensional vector space over $\mathbb{F}_p$. Find the rational canonical form and the Jordan canonical form of F.

Solution. (a) If α is a primitive element of $\mathbb{F}_{p^n}/\mathbb{F}_p$, then the order of $\alpha \in \mathbb{F}_{p^n}^\times$ is n.

(b) By (a), F is satisfied by the polynomial $f(t) = t^n - 1 \in \mathbb{F}_p[t]$. We wish to justify that $f(t)$ is the minimal polynomial of F over $\mathbb{F}_p$. Suppose that $f(t)$ is not the minimal polynomial of F over $\mathbb{F}_p$. Then, there is a nonconstant polynomial $g(t)$ of degree $k < n$ over $\mathbb{F}_p$ which satisfies F. In this case, we have

$$\sum_{i=0}^{k} a_i x^{p^i} = g(F)(x) = O\alpha = 0 \quad (a_i \in \mathbb{F}_p \text{ for } i = 0, 1, \cdots, k)$$

for all $x \in \mathbb{F}_{p^n}$. So, every element of $\mathbb{F}_{p^n}$ is a root of a polynomial of degree $p^k < p^n$, a contradiction.

Therefore, $f(t)$ is the minimal polynomial (and the characteristic polynomial) of F over $\mathbb{F}_p$. Hence, the rational canonical form of F over $\mathbb{F}_p$ is given by

$$\left(\begin{array}{ccccc|c} 0 & 0 & 0 & \cdots & 0 & 1 \\ \hline 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{array}\right).$$

To find the Jordan canonical form of F over $\mathbb{F}_p$, assume that n is not divisible by p. Then $f(t)$ is separable, so F is diagonalizable, with the eigenvalues being all n-th roots of unity. Suppose $n = p^a m$, where a and m are positive integers and p does not divide m. Then $f(t) = (t^m - 1)^{p^a}$, so all distinct roots of $f(t)$ are given as ζ_m^i for $i = 0, 1, \cdots, m-1$, where ζ_m is a primitive m-th root

of unity. In this case, the Jordan canonical form of F is *diag*$(J_0, J_1, \cdots, J_{m-1})$, where

$$J_i = \begin{pmatrix} \zeta_m^i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \zeta_m^i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \zeta_m^i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \zeta_m^i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \zeta_m^i \end{pmatrix} \in \mathcal{M}_{p^a,p^a}(\mathbb{F}_p)$$

for each $i = 0, 1, \cdots, m-1$.

Part V

Field theory

Chapter 17

Basic field theory

17.1 Field extensions

What you basically need to know: field extensions, algebraic elements, algebraic extensions.

Observation 17.1.1 (Basic observations on field compositions). For a field F and a set S, $F(S)$ is defined to be the smallest field containing $F \cup S$, i.e., $F(S)$ is the intersection of all fields containing $F \cup S$. (For composition of infinitely many fields, see Observation 18.2.8.)

Suppose that E and F are subfields of a field K. Then $E(F) = F(E)$, because both of them are the smallest subfield of K containing $E \cup F$. (Hence, it does not matter to write EF in place of $E(F)$.) Moreover,

$$EF = \left\{ \frac{\alpha'_1\beta'_1 + \cdots + \alpha'_n\beta'_n}{\alpha_1\beta_1 + \cdots + \alpha_m\beta_m} : m, n \in \mathbb{Z}^{>0}, \text{ and } \alpha_i, \alpha'_j \in E \text{ and } \beta_i, \beta'_j \in F \text{ for all } i, j \right\}. \tag{17.1}$$

(i) Since EF is the smallest subfield of K containing $E \cup F$, EF necessarily contains all fractional elements as illustrated in eq. (17.1).

(ii) Conversely, the collection suggested in eq. (17.1) is a field containing $E \cup F$.

By (i) and (ii), eq. (17.1) holds.

Now, assume that E/F is a field extension and $\alpha, \beta \in E$. Then there is an obvious equivalence:

$$(F(\alpha))(\beta) = F(\alpha, \beta) = (F(\beta))(\alpha) = F(\alpha)F(\beta).$$

The first three naturally coincides (the third also coincides by symmetry), because

(1) $(F(\alpha))(\beta)$ is the smallest field containing $F(\alpha) \cup \{\beta\}$, so it clearly contains $F(\alpha, \beta)$, the smallest field containing $F \cup (\alpha, \beta)$.

(2) $F(\alpha, \beta)$ is the smallest field containing $F \cup \{\alpha, \beta\}$, so it contains $(F(\alpha))(\beta)$, the smallest field containing $F(\alpha) \cup \{\beta\}$.

In short, the first two coincides since both are the smallest fields containing $F \cup \{\alpha, \beta\}$. To show that the fourth also coincides the first three, note that $F(\alpha)F(\beta)$ contains $F(\alpha) \cup F(\beta)$ and that $F(\alpha, \beta)$ contains $F \cup \{\alpha, \beta\}$ so that it contains $F(\alpha) \cup F(\beta)$.

Observation 17.1.2. Suppose that E/F is a field extension and $\alpha \in E$.

(a) $F(\alpha)$ is the smallest field containing $F \cup \{\alpha\}$, so it necessarily contains $Q_{F[\alpha]} = \{f(\alpha)/g(\alpha) : f(t), g(t) \in F[t], f(\alpha) \neq 0\}$. Conversely, $Q_{F[\alpha]}$ is a field containing F. Therefore, $F(\alpha) = Q_{F[\alpha]}$.

(b) Any element of the form $1/f(\alpha)$ with $f(t) \in F[t]$ and $f(\alpha) \neq 0$ if and only if $F(\alpha) = F[\alpha]$. (One implication is clear; because $F(\alpha) \geq F[\alpha]$, if the former condition is satisfied then $F(\alpha) = F[\alpha]$.)

Question 17.1.1. Given a field extension E/F with an elements $\alpha \in E$, what can be an equivalent condition for $F(\alpha) = F[\alpha]$?

Observation 17.1.3 (The minimal polynomial of an algebraic element). Suppose that E/F is a field extension and $\alpha \in E$ is algebraic over F. Let $\textit{ker}\mathcal{E}_\alpha = \{f(t) \in F[t] : \mathcal{E}_\alpha(f(t)) = 0\}$.

(1) Since $F[t]$ is a PID. and α is algebraic over F, there is a monic polynomial $m(t) \in F[t]$ which generates $\textit{ker}\mathcal{E}_\alpha$. Moreover, a monic generator of $\textit{ker}\mathcal{E}_\alpha$ is unique; if $m(t)$ and $n(t)$ are monic polynomials over F and each of them generates $\textit{ker}\mathcal{E}_\alpha$, then $(m(t)) = (n(t))$, so $m(t) \sim_\times n(t)$ and $m(t) = n(t)$.

Definition 17.1.4. The unique monic polynomial over F which generates $\textit{ker}\mathcal{E}_\alpha$ is called the minimal polynomial of α over F.

(2) Assume that $a(t)b(t) \in \textit{ker}\mathcal{E}_\alpha$ for some $a(t), b(t) \in F[t]$. Then $a(\alpha)b(\alpha) = 0$, so $a(t) \in \textit{ker}\mathcal{E}_\alpha$ or $b(t) \in \textit{ker}\mathcal{E}_\alpha$. This proves that $\textit{ker}\mathcal{E}_\alpha$ is a nonzero prime ideal of $F[t]$. Hence, the minimal polynomial of α over F is irreducible over F. Moreover, $\textit{ker}\mathcal{E}_\alpha$ is a maximal ideal of $F[t]$ and $F[t]/\textit{ker}\mathcal{E}_\alpha$ is a field, for $F[t]$ is a PID.

Suppose that E/F is a field extension with $\alpha \in E$, and assume that $f(\alpha) = 0$ for some nonzero monic polynomial $f(t)$ over F. Then the followings are equivalent:

(a) $f(t)$ is irreducible over F.

(b) $f(t)$ is the minimal polynomial of α over F.

(c) $f(t)$ is a polynomial of the least degree having a root α.

This equivalence follows from the observation that the minimal polynomial $m(t)$ of α over F is irreducible over F and that $f(t) = g(t)m(t)$ for some nonzero polynomial $g(t) \in F[t]$.

Observation 17.1.5 (An answer to Question 17.1.1). We will prove that $F(\alpha) = F[\alpha]$ if and only if α is algebraic over F.

If $F(\alpha) = F[\alpha]$, then $1/\alpha = u(\alpha)$ for some polynomial $u(t) \in F[t]$, thus α is a root of $tu(t) - 1 \in F[t]$ and α is algebraic over F. Assuming conversely, it suffices to show that $1/f(\alpha) \in F[\alpha]$ whenever $f(t) \in F[t]$ is a polynomial such that $f(\alpha) \neq 0$. Let $r(t) \in F[t]$ be a unique polynomial such that $\textit{deg}\, r(t) < \textit{deg}\, m(t)$, where $m(t)$ is the minimal polynomial of α over F. Since $r(t)$ and $m(t)$ are relatively prime, there are polynomials $a(t), b(t) \in F[t]$ such that $a(t)r(t) + b(t)m(t) = 1$, so $1/f(\alpha) = 1/r(\alpha) = a(\alpha)$.

The following theorem, called Kronecker's theorem, has been expected in Observation 17.1.3.

Theorem 17.1.6 (Kronecker's theorem)**.** Let F be a field and $f(t) \in F[t]$ be an irreducible polynomial. Then there is a field extension E/F such that E contains a root of $f(t)$.

Proof. Since $F[t]$ is a Euclidean domain and $f(t) \in F[t]$ is irreducible, the quotient ring $K := F[t]/(f(t))$ is a field. Our goal is to show that K contains an isomorphic copy of F and that K contains a root of $f(t)$.

First, consider the map $\imath : F \to K$ defined by $\imath(a) = \overline{a} = a + (f(t))$ for $a \in F$. One can easily check that $\imath$ is a field embedding, implying that K contains an isomorphic copy of F. Next, in the field $K = F[t]/(f(t))$, the element $\overline{t} = t + (f(t)) \in K$ satisfies

$$f^\imath(\overline{t}) = \overline{f(t)} = \overline{0},$$

so $\overline{t} \in K$ is a root of $f(t)$. □

Remark. In fact, $f(t)$ need not be irreducible, since we may replace $f(t)$ with its irreducible factor.

Question 17.1.2. Suppose that E/F is a field extension and $\alpha \in E$ is algebraic over F. Is $m_\alpha(t) \in F[t]$ of the form $m_\alpha(t) = ((t - \alpha_1) \cdots (t - \alpha_k))^m$? If so, can it be implied that $k = j$ and $m = n$, if we also have $m_\alpha(t) = ((t - \beta_1) \cdots (t - \beta_j))^n$? This question will be answered in Corollary 17.4.2.

We now study some preliminary properties regarding field extensions and algebraic extensions, in particular. Note that whenever E/F is a field extension, we can treat E as an F-vector space. As an application of this idea, we shortly prove that the order of any finite field is a prime power. If $E/\mathbb{F}_p$ is a finite field extension, where p is a positive prime number, then $E \approx \mathbb{F}_p^n$ as $\mathbb{F}_p$-vector space, implying $|E| = p^n$.

Proposition 17.1.7. Suppose that E/F is a field extension and $\alpha \in E$ is algebraic over F. Then $[F(\alpha) : F]$ is the degree of the minimal polynomial of α over F.

Proof. Since α is algebraic over F, we have $F(\alpha) = F[\alpha]$. Thus, if the degree of the minimal polynomial of α over F is n, we may conjecture that $\{1, \alpha, \cdots, \alpha^{n-1}\}$ is an F-basis of $F(\alpha)$, which is, in fact, true. □

Proposition 17.1.8. Finite field extensions are algebraic extensions.

Proof. Suppose that E/F is a finite field extension and let x be an element of E. Writinig $n = [E : F]$, then $\{1, x, \cdots, x^{n-1}, x^n\}$ is F-linearly dependent. □

Remark. Later, it will be proved that a finite field extension is a finite succesive simple algebraic extension.

Corollary 17.1.9. Let E/F be a field extension and let K be the set of all elements of E which are algebraic over F. Then K is a field. In particular, if $\alpha, \beta \in E$ are algebraic over F, then $\alpha \pm \beta$ and $\alpha\beta^{\pm 1}$ (assume $\beta \neq 0$ when taking -1) are also algebraic over F.

Proof. Given such α and β, $F(\alpha, \beta)/F$ is, clearly, a finite extension, so it is an algebraic extension. □

Applying the structure of the composition of two fields suggested in Observation 17.1.1 and Corollary 17.1.9, one can show that algebraic extensions shift to the composition of two fields. The statement and proof are given in Proposition 17.1.16.

Proposition 17.1.10. If K/E and E/F are field extensions, then $[K : F] = [K : E][E : F]$, even if either of the extensions is infinite.[1]

Proof. Let $\{\alpha_i : i \in I\}$ be an E-basis of K and $\{\beta_j : j \in J\}$ be an F-basis of E.

Goal: To show that $\mathcal{B} := \{\alpha_i\beta_j : i \in I, j \in J\}$ is an F-basis of K.

It is clear by hypothesis that $\mathcal{B}$ generates the F-vector space K. Suppose that a finite sum $\sum c_{i,j}\alpha_i\beta_j$ is zero, where $c_{i,j} \in F$ for all i and j. Gathering terms i by i, we have $\sum(\sum c_{i,j}\beta_j)\alpha_i = 0$, so $\sum c_{i,j}\beta_j = 0$ for each i, hence $c_{i,j} = 0$ for all i, j. □

The multiplicativity of extension degree is valid when an extension is given as a linear tower, but it may not be valid when the extension is not linear. In general, the 'sub'multiplicativity (not the multiplicativity) holds for finite extensions. And the following submultiplicativity implies that two field extensions are finite extensions if and only if the composition is a finite extension over the base field.

Proposition 17.1.11. If E/F is a field extension and both K/F and L/F are finite field extensions with $K, L \leq E$, then $[KL : F] \leq [K : F][L : F]$. The equality holds if and only if an F-basis of one of K and L is linearly independent over the other field. (See also Corollary 17.1.14.)

Proof. Let $\{x_1, \cdots, x_m\}$ be an F-basis of K and let $\{y_1, \cdots, y_n\}$ be an F-basis of L. Since $KL = L(x_1, \cdots, x_m)$, we obtain the inequality from

$$[KL : F] = [KL : L][L : F] \leq mn = [K : F][L : F].$$

As illustrated in the above inequality, the equaility holds if and only if $[KL : L] = [K : F]$ (or $[KL : K] = [L : F]$), which is equivalent to the case where $\{x_1, \cdots, x_m\}$ is L-linearly independent (or $\{y_1, \cdots, y_n\}$ is K-linearly independent). □

Corollary 17.1.12. If E/F is a field extension and both K and L are intermediate subfields such that $[K : F]$ and $[L : F]$ are relatively prime, then $[KL : F] = [K : F][L : F]$, hence an F-basis of one of K and L is linearly independent over the other field.

Proof. Remark that $[K : F]$ and $[L : F]$ divides $[KL : F]$. □

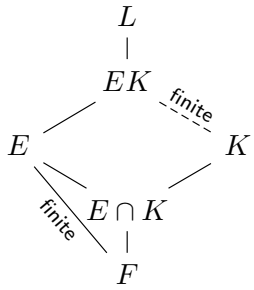
[1] We understand $[K : F] = [K : E][E : F]$ as an equality of cardinal numbers.

Theorem 17.1.13. A field extension E/F is a finite extension if and only if $E = F(\alpha_1, \cdots, \alpha_n)$ for some finitely many elements $\alpha_1, \cdots, \alpha_n \in E$ which are algebraic over F.

Proof. Assume first that E/F is a finite extension and let $\{\alpha_1, \cdots, \alpha_n\}$ be an F-basis of E. Then each α_i is algebraic over F and $E = F(\alpha_1, \cdots, \alpha_n)$.

Assume conversely that E is generated over F by finitely many elements in E which are algebraic over F. Then $[E : F]$ is not greater than the product of the degree of α_i over F for all i. □

Corollary 17.1.14. Suppose that $F \leq E \leq L$ and $F \leq K \leq L$ are field extensions. Then EK/K is a finite extension if E/F is a finite extension, even if K/F may be an infinite extension. (In Proposition 17.1.11, we proved that the EK/F is a finite extension if and only if both E/F and K/F are finite extensions.)

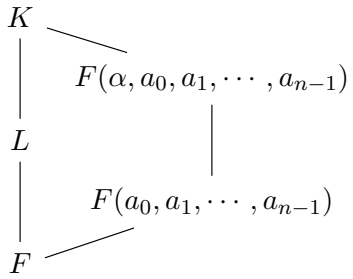


Proof. Let $\alpha_1, \cdots, \alpha_n$ be elements of E which are algebraic over F such that $E = F(\alpha_1, \cdots, \alpha_n)$. Then $EK = K(\alpha_1, \cdots, \alpha_n)$, so EK/K is a finite extension. □

Question 17.1.3 (The existence of a primitive element)**.** Given a finite field extension E/F, can we find an element $\alpha \in E$ such that $E = F(\alpha)$? (Such an element α is called a primitive elements of E over F.)

Theorem 17.1.15. Suppose that $F \leq L \leq K$ is a field extension. Then K/F is an algebraic extension if and only if both K/L and L/F are algebraic extensions.

Proof. It is clear that K/L and L/F are algebraic extensions if K/F is an algebraic extension. Assume conversely that voth K/L and L/F are algebraic extensions. Given an element $\alpha \in K$, write $m_{\alpha,K}(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$.



Then α is algebraic over $F(a_0, a_1, \cdots, a_{n-1})$ and $F(a_0, a_1, \cdots, a_{n-1})/F$ is a finite extension. Hence, $[F(\alpha, a_0, a_1, \cdots, a_{n-1}) : F] < \infty$, so α is algebraic over F. Therefore, K/F is an algebraic extension. □

Regarding algebraic extensions, there are some shifts of algebraic extensions to the composition of fields, as given in Corollary 17.1.14.

Proposition 17.1.16. Suppose that both E and K are intermediate subfield of L/F.

(a) Show that EK/K is an algebraic extension, if E/F is an algebraic extension.

(b) Show that EK/F is an algebraic extension if E/F and K/F are algebraic extensions.

Proof. Note that an element of EK is of the form

$$\frac{\alpha'_1\beta'_1 + \cdots + \alpha'_n\beta'_n}{\alpha_1\beta_1 + \cdots + \alpha_m\beta_m},$$

where $m, n \in \mathbb{Z}^{>0}$ and $\alpha_i, \alpha'_j \in E$ and $\beta_i, \beta'_j \in K$ for all integers $1 \leq i \leq m$ and $1 \leq j \leq n$. When proving (a), it suffices to show that $\alpha\beta$ is algebraic over K whenever $\alpha \in E$ and $\beta \in K$, which is clear. When proving (b), it suffices to show that $\alpha\beta$ is algebraic over F, which is justified in Corollary 17.1.9. □

17.2 Splitting fields

Definition 17.2.1 (Splitting field, normal extension)**.** Let F be a field and $\mathcal{R}$ be a collection of nonconstant polynomials over F. If all polynomials in $\mathcal{R}$ split completely over a field K containing F but not over any proper subfield of F, then K is called a splitting field for the polynomials in $\mathcal{R}$ over F and K/F is said to be a normal extension. In particular, if $\mathcal{R}$ is finite and $p(t) \in F[t]$ is the product of all polynomials in $\mathcal{R}$, then K is called a splitting field for $p(t)$ over F.

Proposition 17.2.2. If F is a field and $p(t)$ is a nonconstant polynomial over F, then there is a splitting field for $p(t)$ over F.

Proof. Using Kronecker's theorem inductively, we can find a field $K := F(\alpha_1, \cdots, \alpha_n)$, where $\alpha_1, \cdots, \alpha_n$ are pairwise distint roots of $p(t)$. It is clear that $p(t)$ splits completely over K; if $p(t)$ splits completely over an extended field, then the extension necessarily contains all rooots of $p(t)$, thus it contains (an isomorphic copy of) K. Therefore, K is a splitting field for $p(t)$ over F. □

Remark. The uniqueness part will be proved later in this chapter. See Theorem 17.4.5.

Example 17.2.3. Suppose that K/F is a finite extension. We will justify that the followings are equivalent:

(a) K/F is a normal extension.

(b) Every nonconstant polynomial over F with a root in K splits completely over K.

To prove that (a) implies (b), assume that K/F is the splitting field for $k(t) \in F[t]$ over F (we can think so, since K/F is a finite extension) and let $p(t) \in F[t]$ be a nonconstant polynomial with a root α in K. If β is another root of $p(t)$, by Theorem 17.4.1, there is an F-isomorphism $\sigma : F(\alpha) \to F(\beta)$ such that $\sigma(\alpha) = \beta$. Since the splitting field for $p(t)$ over $F(\alpha)$ is $K(\alpha) = K$ and the splitting field for $p(t)$ over $F(\beta)$ is $K(\beta)$ (why?), there is a field isomorphism $\widetilde{\sigma} : K \to K(\beta)$ extending σ. Since $\widetilde{\sigma}$ is an F-isomorphism, $\widetilde{\sigma}$ is an F-linear isomorphism, so $dim_F K = dim_F K(\beta)$, implying that $\beta \in K$.

We now show that (b) implies (a). Since K/F is a finite extension, we may write $K = F(\alpha_1, \cdots, \alpha_n)$, where α_i is algebraic over F for each $i = 1, \cdots, n$. Let $p(t)$ be the product of the minimal polynomials of α_i over F.

(i) By hypothesis, $p(t)$ splits completely over K. Hence, K is not smaller than the splitting field for $p(t)$ over F.

(ii) Conversely, the splitting field for $p(t)$ over F necessarily contains all roots of $p(t)$, so it contains K.

Therefore, K is the splitting field for $p(t)$ over F, so K/F is a normal extension.

Example 17.2.4. Let K_1/F and K_2/F be finite normal extensions. We will justify that K_1K_2 is a (finite) normal extension over F, and $K_1 \cap K_2$ is a (finite) normal extension over F.

Let $a(t), b(t) \in F[t]$ be nonconstant polynomials such that K_1 is the splitting field for $a(t)$ over F and K_2 is the splitting field for $b(t)$ over F (such setting is plausible, since K_1/F and K_2/F are finite extensions). We will show that K_1K_2 is the splitting field for $a(t)b(t)$ over F.

(i) Since K_1 contains all roots of $a(t)$ and K_2 contains all roots of $b(t)$, the composition K_1K_2 contains all roots of $a(t)b(t)$. Hence, K_1K_2 contains the splitting field for $a(t)b(t)$ over F.

(ii) Conversely, K_1K_2 is the smallest field containing F and the roots of $a(t)b(t)$, which are necessarily contained in the splitting field for $a(t)b(t)$ over F.

Therefore, K_1K_2 is the splitting field for $a(t)b(t)$ over F.

To show that $K_1 \cap K_2$ is a normal extension over F, we apply the result of the previous example. Let $p(t)$ be a polynomial over F with a root in $K_1 \cap K_2$. Because K_1/F and K_2/F are finite normal extensions, $p(t) \in F[t]$ splits completely over K_1 and K_2. This implies that all roots of $p(t)$ are in $K_1 \cap K_2$, so $K_1 \cap K_2$ is a finite normal extension over F.

Example 17.2.5. Let t be an indeterminate and $F = \mathbb{F}_p(t)$, where p is a positive prime number. And let $f(x) = x^p - t \in F[x]$ and α be a root of $f(x)$. Then $\alpha^p = t$ and $f(x) = x^p - \alpha^p = (x - \alpha)^p$, so the splitting field for $f(x)$ over F is $F(\alpha)$, which is a simple algebraic extension over F of degree p.

17.3 Algebraic closures

Definition 17.3.1 (Algebraic closedness)**.** A field F is said to be algebraically closed if every nonconstant polynomial over F has a root in F.[2]

Observation 17.3.2. For a field F, the followings are equivalent:

(a) F is algebraically closed.

(b) If E/F is a field extension and $\alpha \in E$ is algebraic over F, then $\alpha \in F$.

(c) If E/F is an algebraic extension, then $E = F$.

(d) If E/F is a finite extension, then $E = F$.

Proof. (a)⇒(b): Since F is algebraically closed, the minimal polynomial $m_\alpha(t) \in F[t]$ of α over F splits completely over F, so $\alpha \in F$.
(b)⇒(c)⇒(d): Clear.
(d)⇒(a): Let $f(t) \in F[t]$ be a nonconstant polynomial and let α be a root of $f(t)$. Because $F(\alpha)/F$ is a finite extension, $\alpha \in F$. □

Definition 17.3.3 (Algebraic closure)**.** An algebraic extension $\overline{F}$ over F is called an algebraic closure of F if every nonconstant polynomial over F has a root in $\overline{F}$ (or equivalently, every nonconstant polynomial over F splits completely over $\overline{F}$).

Remark. By definition, a field F is algebraically closed if and only if F is an algebraic closure of F.

The following proposition solves a technical problem when defining algebraic closures.

Proposition 17.3.4. If $\overline{F}$ is an algebraic closure of the field F, then $\overline{F}$ is algebraically closed. Hence, by Observation 17.3.2, an algebraic closure of $\overline{F}$ is $\overline{F}$, i.e., there is no strictly larger algebraic extension over $\overline{F}$.

Proof. Let $f(t)$ be a nonconstant polynomial over $\overline{F}$ and let α be a root of $f(t)$. It suffices to show that $\alpha \in \overline{F}$. Since $\overline{F}(\alpha)/\overline{F}$ and $\overline{F}/F$ are algebraic extensions, $\overline{F}(\alpha)/F$ is an algebraic extension. Because the minimal polynomial of α over F splits completely over $\overline{F}$, $\alpha \in \overline{F}$, as desired. □

Theorem 17.3.5. An algebraic closure of a field exists.

Proof 1. We first introduce the proof due to Emil Artin.

Step 1: A remarkable setting.

Let F be a field. For every nonconstant polynomial $f = f(x) \in F[x]$, let x_f be an indeterminate and consider the polynomial ring $R := F[x_f : f \in F[u]]$. In this polynomial ring, consider the ideal I generated by the polynomials $f(x_f)$ for $f \in F[u]$.

We now prove that I is a proper ideal of R by contradiction; assume $I = R$. Then we have a relation

$$g_1 f_1(x_{f_1}) + \cdots + g_n f_n(x_{f_n}) = 1,$$

where $g_i \in R$ for $i = 1, \cdots, n$. For simplicity, let $x_i = x_{f_i}$ and let $x_{n+1}, \cdots, x_m$ be the remaining variables occurring in the polynomials g_j for $j = 1, \cdots, n$. Then the above relation reads

$$g_1(x_1, \cdots, x_m) f_1(x_1) + \cdots + g_n(x_1, \cdots, x_m) f_n(x_n) = 1.$$

By Kronecker's thoerem, for each $i = 1, \cdots, n$, there is a root α_i of f_i; letting $x_i = \alpha_i$ for each i, we have $0 = 1$, a contradiction.

Step 2: Deriving the result.

There is a maximal ideal M of R containing I. Then R/M is a field which contains an isomorphic copy of F. Moreover, the image of x_f in R/M is a root of a nonconstant polynomial $f(x) \in F[x]$, since

[2] Or equivalently, a field F is said to be algebraically closed if every nonconstant polynomial over F splits completely over F, i.e., the only irreducible polynomials over F are the polynomials of degree 1.

$f(x_f) \in I \subset M$. Therefore, every nonconstant polynomial over F has a root in $F_1 := R/M$. Continuing the above process on F_1 and so on, we obtain an ascending chain $F = F_0 \leq F_1 \leq F_2 \leq \cdots$.

Define $K := \bigcup_{n=0}^{\infty} F_n$. Since $(F_n)_{n=0}^{\infty}$ is monotonically increasing, K is a field. If $a(x) \in K[x]$, then $a(x) \in F_N[x]$ for some positive integer N, so $a(x)$ has a root in F_{N+1} and in K, proving that K is an algebraic closure of F. □

Proof 2. Invoking Zorn's lemma, we may prove the theorem.

Given a field F, let $\mathcal{X}$ be the collection of all algebraic extensions over F. Since F/F is algebraic, $\mathcal{X}$ is nonempty. Furthermore, $\mathcal{X}$ is partially ordered by set inlcusion (or by field extension).

Given a chain $\mathcal{C}$ in $\mathcal{X}$, define

$$K = \bigcup_{E \in \mathcal{C}} E.$$

It is clear that K is a field and K/F is algebraic, implying that $K \in \mathcal{X}$ and K is an upper bound of $\mathcal{C}$.

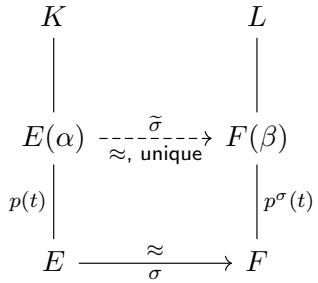
By Zorn's lemma, $\mathcal{X}$ has a maximal member L. Assume that L is not an algebraic closure of F. Then, there is a nonconstant polynomial with a root α such that $\alpha \notin L$. Because $L(\alpha)$ is a proper extension of L and $L(\alpha)$ is an algebraic extension over F, the maximality of L is disobeyed. □

In fact, an algebraic closure of a field is unique up to isomorphism. See Theorem 17.4.7.

17.4 Isomorphism extension theorems

Due to their importance, the below three isomorphism extension theorems are moved to this section, even though they could be proved in eralier sections.

Theorem 17.4.1 (For simple algebraic extensions)**.** Let K/E and L/F be field extensions and $\sigma : E \to F$ be a field isomorphism. If $p(t) \in E[t]$ is irreducible, then $p^\sigma(t) \in F[t]$ is also irreducible. Also, if $\alpha \in K$ is a root of $p(t)$ and $\beta \in L$ is a root of $p^\sigma(t)$, then there is a unique field isomorphism $\widetilde{\sigma} : E(\alpha) \to F(\beta)$ extending σ such that $\widetilde{\sigma}(\alpha) = \beta$.



Proof. It should be satisfied that $\widetilde{\sigma}(u(\alpha)) = u^\sigma(\beta)$ for all $u(t) \in E[t]$. □

Corollary 17.4.2 (An answer to Question 17.1.2)**.** Let K/F be a field extension and $\alpha \in K$ is algebraic over F. Then the minimal polynomial $p(t)$ of α is of the form

$$p(t) = ((t - \alpha_1) \cdots (t - \alpha_k))^m,$$

where $\alpha_1, \cdots, \alpha_k$ are in a splitting field for $p(t)$ over F. Moreover, if

$$p(t) = ((t - \beta_1) \cdots (t - \beta_j))^n,$$

where $\beta_1, \cdots, \beta_j$ are in a splitting field for $p(t)$ over F, then $k = j$ and $m = n$.

Proof. Suppose that α and β are roots of $p(t)$. Then there is a unique F-isomorphism $\mu : F(\alpha) \to F(\beta)$ such that $\mu(\alpha) = \beta$. Because $p(t) \in F[t]$, we have $p^\mu(t) = p(t)$, so the multiplicity of α is not greater than the multiplicity of β and vice versa, due to symmetry. □

Speaking of the isomorphism extension theorem for simple algebraic extensions, we introduce an equivalence regarding algebraic conjugates.

Definition 17.4.3 (Algebraic conjugates)**.** Let E/F and K/F be field extensions. Elements $\alpha \in E$ and $\beta \in K$ are called algebraic conjugates over F if their minimal polynomial over F are the same.

From Theorem 17.4.1, we can deduce the following equivalence:

Observation 17.4.4. Let α, β be elements in a field extension of F which are algebraic over F.

(a) α and β are algebraic conjugates over F if and only if there is an F-isomorphism $\sigma : F(\alpha) \to F(\beta)$ such that $\sigma(\alpha) = \beta$.

(b) Let $m(t)$ be the minimal polynomial of α over F. A root of $m(t)$ is an algebraic conjugate of α over F, and an algebraic conjugate of α over F is a root of $m(t)$.

Proof. (a) When α and β are algebraic conjugates over F, Theorem 17.4.1 proves the existence of a desired F-isomorphism. Assuming conversely, since $m_\alpha(t) \in F[t]$ is fixed by σ, we have $m_\alpha(\beta) = 0$, so $m_\beta(t) | m_\alpha(t)$. It naturally follows from symmetry that $m_\alpha(t) | m_\beta(t)$. Therefore, α and β are algebraic conjugates over F.

(b) By definition, an algebraic conjugate β of α over F is a root of $m(t)$. Let β be a root of $m(t)$ and assume that the minimal polynomial $f(t)$ of β over F is not $m(t)$. Then there is a nonconstant monic polynomial $g(t) \in F[t]$ such that $m(t) = f(t)g(t)$, and either $f(t)$ or $g(t)$ has α a root, a contradiction. Therefore, a root of the minimal polynomial of α over F is an algebraic conjugate of α over F. □

Remark. The above observation induces an equivalence relation regarding algebraic conjugates over a given field. Let F be a field and write $\alpha \sim \beta$ for elements α, β which are algebraic over F, whenever

$$\beta \text{ is an algebraic conjugate of } \alpha \text{ over } F.$$

Then $\sim$ denotes an equivalence relation on the field of elements algebraic over F. Hence, α and β are algebraically conjugate over F if and only if the minimal polynomials of α and β over F are the same.

Theorem 17.4.5 (For splitting fields)**.** Let $\sigma : E \to F$ be a field isomorphism and $f(t)$ be a nonconstant polynomial over E. Suppose that K is a splliting field for $f(t)$ over E and L is a splitting field for $f^\sigma(t)$ over F. Then there is a field isomorphism $\widetilde{\sigma} : K \to L$ extending σ.

$$\begin{array}{ccc} K & \overset{\widetilde{\sigma}}{\underset{\approx}{\dashrightarrow}} & L \\ {\scriptstyle f(t)}\Big| & & \Big|{\scriptstyle f^\sigma(t)} \\ E & \xrightarrow[\sigma]{\approx} & F \end{array}$$

Proof 1. We prove the theorem by applying Kronecker's theorem inductively. Let $a(t) \in E[t]$ be an irreducible factor of $f(t)$ and $\alpha_1 \in K$ be a root of $a(t)$. Then $a^\sigma(t) \in F[t]$ is an irreducible factor of $f^\sigma(t)$ and has a root $\beta_1 \in L$. By Theorem 17.4.1, there is a field isomorphism $\sigma_1 : E(\alpha_1) \to F(\beta_1)$ extending σ. Hence, there are polynomials $p_1(t) \in E(\alpha_1)[t]$ and $q_1(t) \in F(\beta_1)[t]$ such that

$$(t - \alpha_1)p_1(t) = f(t) = (t - \beta_1)q_1(t).$$

As we have done earlier, let $a_2(t) \in E(\alpha_1)[t]$ be an irreducible factor of $p_1(t)$ and $\alpha_2 \in K$ be a root of $a_2(t)$. Then $a_2^\sigma(t) \in F(\beta_1)[t]$ is also irreducible and has a root β_2, and there is a field isomorphism $\sigma_2 : E(\alpha_1)(\alpha_2) \to F(\beta_1)(\beta_2)$ extending σ_1. Since *deg* $f(t)$ is finite, this process will terminate and produces a field isomorphism $\widetilde{\sigma} : K \to L$. □

Proof 2. We prove the theorem by induction on *deg* $f(t)$. Note that the theorem is clear when *deg* $f(t) = 1$. Assuming that the theorem is valid for all nonconstant polynomials over E of degree less than n, suppose that *deg* $f(t) = n$. Let $\alpha \in K$ be a root of $f(t)$ and $\beta \in L$ be a root of $f^\sigma(t)$. By Theorem 17.4.1, there is a unique field isomorphism $\sigma_1 : E(\alpha) \to F(\beta)$ extending σ mapping α to β. Writing

$$(t - \alpha)p(t) = f(t) = (t - \beta)q(t)$$

for some $p(t) \in E(\alpha)[t]$ and $q(t) \in F(\beta)[t]$, we have $q(t) = p^{\sigma_1}(t)$. Thus, it remains to justify that K is a splitting field for $p(t)$ over $E(\alpha)$ and L is a splitting field for $q(t)$ over $F(\beta)$; it then follows from the induction hypothesis that there is a field isomorphism $\widetilde{\sigma} : K \to L$ extending σ_1 (thus, extending σ).

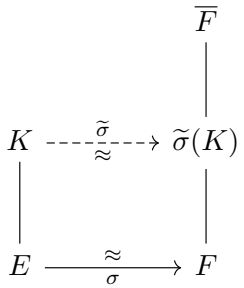
(i) It is clear that $p(t)$ splits completely over K.

(ii) If there is a proper subfield I of K containing $E(\alpha)$ over which $p(t)$ splits completely, then $f(t) = (t-\alpha)p(t)$ would split complitely over I, which contradicts the hypothesis that K is a splitting field for $f(t)$ over E. Therefore, there is no proper subfield of K over which $p(t)$ splits completely.

The same argument holds for L, as desired. □

Corollary 17.4.6. A splitting field for a nonconstant polynomial over a field is unique up to isomorphism.

Theorem 17.4.7 (For algebraic closures)**.** Let $\sigma : E \to F$ be a field isomorphism. If K/E is an algebraic extension, there is a field embedding $\widetilde{\sigma} : K \to \overline{F}$ extending σ.



Proof. We find a field embedding of K into $\overline{F}$ by applying Zorn's lemma.

Step 1. Setting a nonempty partially ordered set.

Set

$$\mathcal{X} := \left\{ (L, \tau) : \begin{array}{c} E \leq L \leq K \text{ and} \\ \tau : L \to K \text{ is a field embedding extending } \sigma \end{array} \right\}.$$

Then $(E, \sigma) \in \mathcal{X}$, so X is nonempty. And for $(L_1, \tau_1), (L_2, \tau_2) \in \mathcal{X}$, let $(L_1, \tau_1) \leq (L_2, \tau_2)$ if and only if

$$L_1 \leq L_2 \text{ and } \tau_2|_{L_1} = \tau_1.$$

Then this relation is a partial order on $\mathcal{X}$.

Step 2. Showing that every subchain has an upper bound.

Let $\mathcal{C}$ be an ascending chain of X. To find its upper bound in X, let

$$C := \bigcup_{(L,\tau)\in\mathcal{C}} L$$

and define a map $\tau_C : L \to \overline{F}$ by $\tau_C(x) := \tau_x(x)$, where $(L_x, \tau_x) \in \mathcal{C}$ is a member such that $x \in L_x$. Then $E \leq C \leq K$ and τ_C is a well-defined field embedding, so (C, τ_C) is an upper bound of $\mathcal{C}$ in X.

Step 3. Deriving the result.

By Zorn's lemma, there is a maximal element $(M, \mu) \in X$. We will justify that $M = K$ by contradiction. Assume $M \subsetneq K$. Then there is an element $\alpha \in K \setminus M$; let $p(t) \in M[t]$ be the minimal polynomial of α over M, and let $\beta \in \overline{F}$ be a root of $p^{\mu}(t) \in F[t]$. By Theorem 17.4.1, there is a field embedding $\widetilde{\mu} : M(\alpha) \to \overline{F}$ extenidng μ, so $(M, \mu) \subsetneq (M(\alpha), \widetilde{\mu})$, which contradicts the maximality of (M, μ). □

Corollary 17.4.8. An algebraic closure of a field is unique up to isomorphism.

Proof. Given a field F, let K and L be algebraic closures of F. For id_F, by Theorem 17.4.7, there is a field embedding $\mu : K \to L$. Since K is algebraically closed, so is $\mu(K)$. Because $L/\mu(K)$ is an algebraic extension, we have $\mu(K) = L$, as desired. □

17.5 Constructible numbers

Theorem 17.5.1 (Constructibility criterion I)**.** $\alpha \in \mathbb{R}$ is constructible if and only if there is a tower of quadratic extensions from $\mathbb{Q}$ whose head field contains α.

Proof. Assuming the existence of such a tower, it is almost clear that α is constructible. Conversely, if $\alpha \in \mathbb{R}$ is constructible, a construction of α by straight lines and circles naturally induces a tower of quadratic extensions with α being contained in the head field. □

Another constructibility criterion is introduced in Theorem 21.4.3.

Corollary 17.5.2. A real algebraic conjugate of a constructible real number is also constructible.

Proof. Let α be a constructible real number and β be a real algebraic conjugate of α. Then there is a unique $\mathbb{Q}$-isomorphism $\sigma : \mathbb{Q}(\alpha) \to \mathbb{Q}(\beta)$ mapping α to β. By the previous theorem, there is a tower of quadratic extensions

$$\mathbb{Q} < \mathbb{Q}(\sqrt{d_1}) < \cdots < \mathbb{Q}(\sqrt{d_1}, \cdots, \sqrt{d_n}) = K$$

where $d_1, \cdots, d_n \in \mathbb{Q}$ and K contains α. By isomorphism extension theorem, there is a field embedding $\widetilde{\sigma} : K \to \overline{\mathbb{Q}(\beta)} = \overline{\mathbb{Q}}$ extending σ. Then

$$\mathbb{Q} < \mathbb{Q}(\widetilde{\sigma}(\sqrt{d_1})) < \cdots < \mathbb{Q}(\widetilde{\sigma}(\sqrt{d_1}), \cdots, \widetilde{\sigma}(\sqrt{d_n})) = \widetilde{\sigma}(K)$$

is a tower of quadratic extensions and β belongs to $\widetilde{\sigma}(K)$. □

Corollary 17.5.3. If $\alpha \in \mathbb{R}$ is constructible, then the degree of α over $\mathbb{Q}$ is a power of 2.

Proof. Clear. □

Chapter 18

Separable extensions and normal extensions

18.1 Separable extensions

Definition 18.1.1 (Separability)**.** Given an algebraic field extension E/F, an element $\alpha \in E$ is said to be separable over F if its minimal polynomial over F is a separable polynomial. In particular, the extension E/F is said to be a separable extension if every element in E is separable over F. If every algebraic extension over a F is separable, then F is said to be perfect.

Example 18.1.2. (a) It is easy to check that an irreducible polynomial over a field of characteristic 0 is a separable polynomial. It will be proved in the next chapter that an irreducible polynomial over a finite field is also a separable polynomial. Hence, fields of characteristic 0 and finite fields are perfect fields.

(b) Algebraically closed fields are perfect fields, and algebraic extensions of perfect fields are perfect fields.

Some expereice from algebraic extensions helps studying separable extensions.

Question 18.1.1. (a) Are the sum and multiplication of two separable elements separable?

(b) Is a separable extension of a separable extension is a separable extension?

Our study begins with the degree which can measure the separability.

Observation 18.1.3. Suppose that E/F is a field extension and $\alpha \in E$ is algebraic over F. Then the minimal polynomial $m(t)$ of α over F is irreducible and

$$m(t) = ((t-\alpha_1)\cdots(t-\alpha_k))^n$$

for some pairwise distinct elements $\alpha_1, \cdots, \alpha_k \in \overline{F}$ and $n \in \mathbb{Z}^{>0}$. It is clear from the above expression that k is the number of distinct roots of the minimal polynomial of α over F and that k divides the extension degree of $F(\alpha)/F$.

(Without loss of generality, assume $\alpha_1 = \alpha$.) After choosing an index i, by Theorem 17.4.1, there is a unique F-embedding $\sigma : F(\alpha) \hookrightarrow \overline{F}$ such that $\sigma(\alpha) = \alpha_i$, so $k \leq |\mathrm{Emb}(F(\alpha)/F)|$.[1] Conversely, given $\sigma \in \mathrm{Emb}(F(\alpha)/F)$, since $m(t)$ is fixed by the action of σ, $\sigma(\alpha)$ is a root of $m(t)$; because every F-embedding of $F(\alpha)$ into $\overline{F}$ is determined by its action on α, we have $k \geq |\mathrm{Emb}(F(\alpha)/F)$. Therefore, k also denotes the number of all distinct F-embeddings of $F(\alpha)$ into $\overline{F}$.

Generalizing the above observation, we can define the separable degree of an algebraic extension as follows:

[1]When considering $\mathrm{Emb}(E/F)$ as a set, it is assumed that an algebraic closure $\overline{F}$ of F is given and we consider F-embeddings of E into $\overline{F}$. Still, its cardinality does not depend on the choice of an algebraic closure of $\overline{F}$.

Definition 18.1.4. Let E/F be an algebraic field extension, where $F \leq E \leq \overline{F}$ with $\overline{F}$ being the algebraic closure of F. The separable degree of E/F, denoted by $[E : F]_{\text{sep}}$, is defined by the cardinality of the set of F-embeddings of E into $\overline{F}$. In other words,

$$[E : F]_{\text{sep}} := |\{\tau : E \hookrightarrow \overline{F} : \tau \text{ is an } F\text{-embedding}\}|.$$

$$\begin{array}{ccc} \overline{F} & & \overline{F} \\ | & & | \\ E & \overset{\tau}{\underset{\approx}{\dashrightarrow}} & \tau(E) \\ | & & | \\ F & \underset{id_F}{\xrightarrow{=}} & F \end{array}$$

The above definition seems to be poorly defined, since the cardinality of the set of such F-embeddings seems to depend on the choice of an algebraic closure of F. (This is intuitively (and, in fact, logically) true, for an algebraic closure is unique up to isomorphism.) Moreover, one may wish to generalize the definition so that τ extends a field isomorphism, not just the idenity map on F.

Observation 18.1.5. Let E/F be an algebraic extension and let $\overline{F}$ and $\widetilde{F}$ be algebraic extensions of F, which contains E. We will show that there is a bijection

$$\{\tau : E \hookrightarrow \overline{F} : \tau \text{ is an } F\text{-embedding}\} \longleftrightarrow \{\mu : E \hookrightarrow \widetilde{F} : \mu \text{ is an } F\text{-embedding}\},$$

which explains that the above definition of separable degree does not depend on the choice of an algebraic closure of F.

Consider the following diagram, where an F-embedding $\tau : E \hookrightarrow \overline{F}$ is given.

$$\begin{array}{ccccc} \widetilde{F} & & \overset{\widetilde{id_F}}{\underset{\approx}{\longleftarrow}} & & \overline{F} \\ | & & & & | \\ \star & \overset{f(\tau)}{\underset{\approx}{\dashleftarrow}} & E & \overset{\tau}{\underset{\approx}{\longrightarrow}} & \tau(E) \\ | & & | & & | \\ F & \underset{id_F}{\xleftarrow{=}} & F & \underset{id_F}{\xrightarrow{=}} & F \end{array}$$

For the diagram to be commutative, it should be satisfied that $f(\tau) = \widetilde{id_F} \circ \tau$. This establishes a map

$$f : \{\tau : E \hookrightarrow \overline{F} : \tau \text{ is an } F\text{-embedding}\} \to \{\mu : E \hookrightarrow \widetilde{F} : \mu \text{ is an } F\text{-embedding}\},$$

which is a bijection; thus, we may write down $[E : F]_{\text{sep}} = |\text{Emb}(E/F)|$ without confusion.

Observation 18.1.6. As in the preceeding observation, assume that $F \leq E \leq \overline{F}$ is a tower of algebraic extensions, where $\overline{F}$ is an algebraic closure of F. And let $\sigma : F \hookrightarrow \overline{F}$ be a field embedding, i.e., $\sigma : F \to \sigma(F)$ is a field isomorphism. We will justify that

$$[E : F]_{\text{sep}} = |\{\mu : E \hookrightarrow \overline{F} : \tau \text{ is a field embedding and } \tau|_F = \sigma\}|$$

(the separable degree of a given algebraic extension does not depend on the base field isomorphism (embedding)) by showing that there is a bijection

$$\{\tau : E \hookrightarrow \overline{F} : \tau \text{ is an } F\text{-embedding}\} \longleftrightarrow \{\mu : E \hookrightarrow \overline{F} : \tau \text{ is a field embedding and } \tau|_F = \sigma\}.$$

Consider the following diagram when an F-embedding $\tau : E \hookrightarrow \overline{F}$ is given.

$$\begin{array}{ccccc} \overline{F} & & \overset{\widetilde{\sigma}}{\underset{\approx}{\longleftarrow}} & & \overline{F} \\ | & & & & | \\ \star & \overset{g(\tau)}{\underset{\approx}{\dashleftarrow}} & E & \overset{\tau}{\underset{\approx}{\longrightarrow}} & \tau(E) \\ | & & | & & | \\ \sigma(F) & \underset{\sigma}{\overset{\approx}{\longleftarrow}} & F & \underset{id_F}{\xrightarrow{=}} & F \end{array}$$

For the diagram to be commutative, it should be satisfied that $g(\tau) = \widetilde{\sigma} \circ \tau$. By defining g so, we have establised a map

$$g : \{\tau : E \hookrightarrow \overline{F} : \tau \text{ is an } F\text{-embedding}\} \to \{\mu : E \hookrightarrow \overline{F} : \tau \text{ is a field embedding and } \tau|_F = \sigma\}.$$

which is a bijection.

By the preceeding two observations, we may re-define the separable degree of an algebraic extension as follows.

Definition 18.1.7 (Separable degree)**.** Let E/F be an algebraic field extension and let $\overline{F}$ be the algebraic closure of F. And let $\sigma : F \hookrightarrow \overline{F}$ be a field embedding. Then the separable degree of E/F, denoted by $[E : F]_{\text{sep}}$, is defined by the cardinality of field embeddings from E into $\overline{F}$ whose restriction to F is σ. In other words,

$$[E : F]_{\text{sep}} := |\{\tau : E \hookrightarrow \overline{F} : \tau \text{ is a field embedding and } \tau|_F = \sigma\}|.$$

(As observed in the preceeding two observations, this definition is independent of the choice of an algebraic closure of F and a field embedding $\sigma : F \hookrightarrow \overline{F}$.)

Remark. Suppose that α is an element which is algebraic over F. Then α is separable over F if and only if its separable degree and extension degree are the same.

In the remaining of this section, we will investigate some properties regarding separable degrees and separability of algebraic field extensions, and the primitive element theorem for finite separable extensions.

Lemma 18.1.8. Suppose that K/E and E/F are algebraic field extensions.

(a) $[K : F]_{\text{sep}} = [K : E]_{\text{sep}}[E : F]_{\text{sep}}$.

(b) In particular, if E/F is a finite extension, then the separable degree of E/F divides the extension degree of E/F.

Proof. Fix an algebraic closure $\overline{F}$ of F containing K.

(a) Given $\sigma \in \text{Emb}(K/F)$, σ is an extension of $\sigma|_E \in \text{Emb}(E/F)$, so $[K : F]_{\text{sep}} \leq [K : E]_{\text{sep}}[E : F]_{\text{sep}}$. Conversely, there are $[E : F]_{\text{sep}}$-distinct F-embeddings of E into $\overline{F}$, thus there are at least $[K : E]_{\text{sep}}[E : F]_{\text{sep}}$-distinct F-embeddings of K into $\overline{F}$, so $[K : F]_{\text{sep}} \geq [K : E]_{\text{sep}}[E : F]_{\text{sep}}$.

(b) Assume that E/F is a finite extension and write $E = F(\alpha_1, \cdots, \alpha_n)$ for some elements $\alpha_1, \cdots, \alpha_n \in E$ which are algebraic over F. Considering the following tower of simple algebraic extensions:

$$F \leq F(\alpha_1) \leq F(\alpha_1, \alpha_2) \leq \cdots \leq F(\alpha_1, \cdots, \alpha_n) = E.$$

In each simple algebraic extension, the separable degree divides the extension degree. Since each degree is multiplicative, we find that $[E : F]_{\text{sep}}$ divides $[E : F]$.

This completes the proof. □

Proposition 18.1.9. Let E/F be a finite extension. Then E/F is a separable extension if and only if its separable degree and extension degree are the same.

Proof. Write $E = F(\alpha_1, \cdots, \alpha_r)$ for some $\alpha_1, \cdots, \alpha_r \in E$ which are algebraic over F, and consider a tower of simple algebraic extensions. If E/F is a separable extension, then each simple extension is separable. Hence, the separable degree and the extension degree of each simple extension are the same, and $[E : F] = [E : F]_{\text{sep}}$. Conversely, if $[E : F] = [E : F]_{\text{sep}}$ and $\alpha \in E$, then $[F(\alpha) : F] = [F(\alpha) : F]_{\text{sep}}$, implying that E/F is a sepatable extension. □

Remark. In particular, if $\alpha_1, \cdots, \alpha_r \in \overline{F}$, then $F(\alpha_1, \cdots, \alpha_r)/F$ is a separable extension if and only if $\alpha_1, \cdots, \alpha_r$ are separable over F.

Corollary 18.1.10. Suppose that K/E and E/F are finite field extensions. Then K/F is a separable extension if and only if both K/E and E/F are separable extensions.

Proof. It is clear that both K/E and E/F are separable extensions if K/F is a separable extension. Assume conversely that K/E and E/F are separable extensions. Since $[K : E] = [K : E]_{\text{sep}}$ and $[E : F] = [E : F]_{\text{sep}}$, we have $[K : F] = [K : F]_{\text{sep}}$, implying that K/F is a separable extension. □

In fact, Corollary 18.1.10 extends to the following proposition:

Proposition 18.1.11. Suppose that K/E and E/F are algebraic field extensions. Then K/F is a separable extension if and only if both K/E and E/F are separable extensions.

Proof. It is clear that both K/E and E/F are separable extensions if K/F is a separable extension. Assume conversely that K/E and E/F are separable extensions. Given $\alpha \in K$, there is the separable minimal polynomial $m(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \in E[t]$ of α over E, hence α is a separable element over $F(a_0, a_1, \cdots, a_{n-1})$. Since a_i is a separable element over F, we can easily find that $F(\alpha, a_0, a_1, \cdots, a_{n-1})/F$ is a separable extension by showing that its extension degree and separable degree are the same. Thus, α is a separable element over F, and K/F is a separable extension. □

Corollary 18.1.12. Given an algebraic extension E/F, let $S_{E/F}$ be the collection of all elements of E which are separable over F. Then $S_{E/F}$ is an intermediate subfield of E/F.

Proof. Given two elements $\alpha, \beta \in E$ which are separable over F, $F(\alpha, \beta)/F$ is a separable extension, so $\alpha \pm \beta, \alpha\beta^{\pm 1}$ are separable over F. □

Remark. Hence, if $\alpha, \beta \in E$ are algebraic over F, then so are $\alpha \pm \beta$ and $\alpha\beta^{\pm 1}$.

Proposition 18.1.13. Suppose that E_1/F and E_2/F are separable extensions, where E_1, E_2 are contained in $\overline{F}$. Then $(E_1E_2)/F$ is a separable extension.

Proof. Considering the form of the typical element of E_1E_2, it suffices to check that $\alpha\beta \in E_1E_2$ is separable over F, where $\alpha \in E_1$ and $\beta \in E_2$; now this is clear by Corollary 18.1.12. □

Theorem 18.1.14 (Primitive element theorem)**.** Suppose that E/F is a finite *separable* extension. Then there is an element $\alpha \in E$ such that $E = F(\alpha)$.

Proof. This proof is not well-motivated, yet.

Suppose that F is a finite field. Because E/F is a finite extension, E is a finite field, hence there is a primitive element for E/F. Thus, we may assume that F is an infinite field. It suffices to show that for any two elements $\beta, \gamma \in \overline{F}$ there is an element $\alpha \in \overline{F}$ such that $F(\gamma) = F(\alpha, \beta)$, for E/F is a finite extension. Let $\{\beta_1, \cdots, \beta_r\}$ be the set of roots of the minimal polynomial $f(t)$ of β over F, and let $\{\gamma_1, \cdots, \gamma_s\}$ be the set of roots of the minimal polynomial $g(t)$ of γ over F, where $\beta_1 = \beta$ and $\gamma_1 = \gamma$. Because F is infinite, there is an element $c \in F$ such that

$$c \neq \frac{\beta - \beta_i}{\gamma_j - \gamma} \quad (2 \leq i \leq r, 2 \leq j \leq s).$$

Then let $\alpha = \beta + c\gamma$; we will show that $F(\beta, \gamma) = F(\alpha)$. Consider the polynomial $h(t) = f(\alpha - ct) \in F(\alpha)[t]$.

(i) $h(\gamma) = f(\beta) = 0$, so the minimal polynomial $m(t)$ of γ over $F(\alpha)$ divides $h(t)$.

(ii) Since $c\gamma - c\gamma_j \neq \beta_i - \beta$ for all $2 \leq i \leq r$ and $2 \leq j \leq s$, we have $h(\gamma_j) = f(\beta + c\gamma - c\gamma_j) \neq 0$.

Because $m(t)$ also divides $g(t)$ and E/F is a separable extension, $m(t) = t - \gamma$. So, $\gamma \in F(\alpha)$ and $\beta = \alpha - c\gamma \in F(\alpha)$, hence $F(\beta, \gamma) = F(\alpha)$. □

18.2 Normal extensions

Before studying normal extensions, we first observe the following proposition, whose proof, in particular, provides us essential intuition regarding normal extensions.

Proposition 18.2.1. Suppose that E/F is an algebraic field extension and $\sigma : E \hookrightarrow E$ is an F-embedding. Then $\sigma(E) = E$, i.e., $\sigma \in \mathrm{Aut}(E/F)$.

Proof. If E/F is a finite extension, the proposition follows from the observation that σ is a vector space isomorphism between finite dimensional vector spaces over F. To prove the proposition in general settings, it suffices to show that $\alpha \in \sigma(E)$, where $\alpha \in E$. Given $\alpha \in E$, let $f(t)$ be the minimal polynomial of α over F and write

$$f(t) = ((t-\alpha_1)\cdots(t-\alpha_k) \times (t-\beta_1)\cdots(t-\beta_s))^r,$$

where $\alpha_1, \cdots, \alpha_k \in E$ and $\beta_1, \cdots, \beta_s \in \overline{F} \setminus E$ are pairwise distinct. Since $\sigma(E) \leq E$ and $f(t)$ is fixed by the action of σ, σ permutes $\{\alpha_1, \cdots, \alpha_k\}$, justifying that $\alpha \in \sigma(E)$. □

We defined a normal extension in Definition 17.2.1. Since we now know that the splitting field for a nonconstant polynomial $f(t)$ over a field F is the smallest field containing F and all roots $f(t)$ and that the splitting field for $f(t)$ over F is unique up to isomorphism, we emphasize the following general definition for splitting fields.

Definition 18.2.2 (Splitting field)**.** Let F be a field and R be a collection of nonconstant polynomials over F. If S is the collection of the roots of the polynomials in R, the splitting field for R over F is defined as the field $F(S)$. (Note that this definition coincides the old definition, where R is a finite collection.)

Example 18.2.3. If R is the collection of all nonconstant (monic) polynomials over a field F, then the splitting field for R over F is the algebraic closure of F.

Theorem 18.2.4 (Normal extension)**.** Suppose that E/F is an algebraic extension and assume $E \leq \overline{F}$. Then the followings are equivalent:

(a) There is a collection $\mathcal{R}$ of nonconstant polynomials over F for which E is the splitting field over F.

(b) If $\tau \in \mathrm{Emb}(E/F)$, then $\tau \in \mathrm{Aut}(E/F)$. In other words, $\mathrm{Emb}(E/F) = \mathrm{Aut}(E/F)$.[2]

(c) If $\alpha \in E$, then all roots of $m_{\alpha,F}(t)$ are in E. In other words, $m_{\alpha,F}(t)$ splits completely over E.

In either of the above cases, we call E/F a normal extension.

Proof. Assume (a) and let $\tau : E \hookrightarrow \overline{F}$ be an F-embedding. By the first proposition in this section, it suffices to verify that $\tau(E) \leq E$. If α is a root of a polynomial in $\mathcal{R}$, it can easily be checked that $\tau(\alpha)$ is a root of the same polynomial. Hence $\tau(E) \leq E$.

Assume (b), and let $m(t)$ be the minimal polynomial of $\alpha \in E$ over F. Given a root β of $m(t)$, let $\sigma : F(\alpha) \to F(\beta)$ be the unique F-isomorphism such that $\sigma(\alpha) = \beta$. We then can extend σ to $\widetilde{\sigma} : E \hookrightarrow \overline{F}$; by the assumption (b), $im\,\widetilde{\sigma} = E$, so $\beta \in E$, as desired.

Finally, assume (c). Then $E = F(\mathcal{R})$, where $\mathcal{R}$ is the collection of the minimal polynomial of $\alpha \in E$ over F for $\alpha \in E$. This proves that (a), (b), and (c) are equivalent. □

Example 18.2.5. (a) Suppose that E/F is an algebraic extension of degree 2. Then $E = F(\alpha)$ for some $\alpha \in E$, where the minimal polynomial $m(t)$ of α over F is of the form $t^2 + bt + c$ for some $b, c \in F$. The other root of $m(t)$ is given by $-b - \alpha \in E$, so E is the splitting field for $m(t)$ over F.

(b) Unlike algebraic and separable extensions, a normal extension of a normal extension may not be a normal extension. (An example: $\mathbb{Q} < \mathbb{Q}(\sqrt{2}) < \mathbb{Q}(\sqrt[4]{2})$.)

[2] Hence, in particular, a finite field extension E/F is a normal extension if and only if $[E : F]_{\text{sep}} = |\mathrm{Aut}(E/F)|$.

In fact, normal extensions seems to have some properties satisfied in group theory when extension towers are seen reversely. A counterpart of the following proposition in group theory is that if $H \leq K \leq G$ and H is a normal subgroup of G, then H is a normal subgroup of K.

Proposition 18.2.6. Suppose that $F \leq E \leq K$ and K/F is a normal extension. Then K/E is a normal extension.

Proof. Suppose that $\tau : K \hookrightarrow \overline{E}$ is an E-embedding. Since $F \leq E$ and we may assume that $\overline{E} = \overline{F}$, τ is an F-embedding of K into $\overline{F}$, so $\tau(K) = K$ and K/E is a normal extension. □

Proposition 18.2.7. Let E_1/F and E_2/F are algebraic extensions such that $E_1, E_2 \leq \overline{F}$.

(a) If $\tau : \overline{F} \hookrightarrow \overline{F}$ is a field embedding, then $\tau(E_1E_2) = \tau(E_1)\tau(E_2)$.

Assume further that E_1/F and E_2/F be normal extensions such that $E_1, E_2 \leq \overline{F}$.

(b) $(E_1E_2)/F$ is a normal extension.

(c) $(E_1 \cap E_2)/F$ is a normal extension.

Proof. (a) easily follows from the identity

$$\tau\left(\frac{\alpha_1'\beta_1' + \cdots + \alpha_n'\beta_n'}{\alpha_1\beta_1 + \cdots + \alpha_m\beta_m}\right) = \frac{\tau(\alpha_1')\tau(\beta_1') + \cdots + \tau(\alpha_n')\tau(\beta_n')}{\tau(\alpha_1)\tau(\beta_1) + \cdots + \tau(\alpha_m)\tau(\beta_m)},$$

where $\alpha_i, \alpha_j' \in E_1$ and $\beta_i, \beta_j' \in E_2$ for all integers $1 \leq i \leq m$ and $1 \leq j \leq n$.

To prove (b), let $\tau : E_1E_2 \hookrightarrow \overline{F}$ be an F-embedding. Then $\tau(E_1E_2) = \tau(E_1)\tau(E_2) = E_1E_2$, for E_1/F and E_2/F are normal extensions. This proves that $(E_1E_2)/F$ is a normal extension.

To prove (c), assume $\alpha \in E_1 \cap E_2$. Then all roots of the minimal polynomial $m(t)$ of α over F are in both E_1 and E_2, so $(E_1 \cap E_2)/F$ is a normal extension. □

We end this section with the notion of normal closures and introducing an algebraic extension which we call a Galois extension. But before this, we solve a technical problem.

Observation 18.2.8 (The composition of infinitely many fields). Let $\{F_\alpha : \alpha \in \mathcal{A}\}$ be a collection of fields which are contained in a larger field E. The composition F_0 of F_α for all $\alpha \in A$ (the smallest subfield of E containing $\bigcup_{\alpha\in\mathcal{A}} F_\alpha$) is defined as

$$K := \{x \in E : x \text{ belongs to a composition of } F_\alpha \text{ for finitely many values of } \alpha\}.$$

(In fact, F_0 axiomatically contains K, and K is a field containing $\bigcup_{\alpha\in\mathcal{A}} F_\alpha$.)

Theorem 18.2.9 (Normal closure)**.** Let E/F be a field extension and assume $E \leq \overline{F}$.

(a) The set $\{K : E \leq K \leq \overline{F}$ and K/F is a normal extension$\}$ has the least element K_0. In fact, K_0 is the composition of $\sigma(E)$ for $\sigma \in \mathrm{Emb}(E/F)$. We call K_0 the normal closure of E/F.

(b) In particular, if E/F is a separable extension, then so is K_0/F.

Proof. We first show that K_0 is the smallest normal extension over F containing E.

(i) Since $\mathrm{Emb}(E/F)$ contains the inclusion of E into $\overline{F}$, $E \leq K_0$.

(ii) Given an F-embedding $\tau : K_0 \hookrightarrow \overline{F}$, note that $\tau \circ \sigma \in \mathrm{Emb}(E/F)$ whenever $\sigma \in \mathrm{Emb}(E/F)$. In fact, τ permutes $\mathrm{Emb}(E/F)$ by left multiplication, so $\tau(K_0) = K_0$.

(iii) Finally, if K is an intermediate subfield of $\overline{F}/E$ which is normal over F, then K necessarily contains $\sigma(E)$ for $\sigma \in \mathrm{Emb}(E/F)$; if $\sigma \in \mathrm{Emb}(E/F)$, by extending σ to $\widetilde{\sigma} \in \mathrm{Emb}(K/F)$, we have $\sigma(E) \leq \widetilde{\sigma}(K) = K$.

We now assume that E/F is a separable extension. Then $(\sigma E)/F$ is a separable extension whenever $\sigma \in \mathrm{Emb}(E/F)$. If $x \in K_0$, then x belongs to the composition of some finitely many fields (σE)'s, and such finite composition is a separable extension over F. Therefore, when E/F is a separable extension, then the normal closure of E over F is also separable over F. □

Example 18.2.10. In this example, we find the normal closure of $E = \mathbb{Q}(\sqrt[3]{2})$ over $\mathbb{Q}$. For this, we first need to find all $\mathbb{Q}$-embeddings of E into $\overline{\mathbb{Q}}$; because $\sigma \in \mathrm{Emb}(E/\mathbb{Q})$ is completely determined by its action on $\sqrt[3]{2}$, it suffices to determine all possible values of $\sigma(\sqrt[3]{2})$. Let σ be a $\mathbb{Q}$-embedding of E into $\overline{\mathbb{Q}}$ and write $\alpha = \sqrt[3]{2}$. Then $(\sigma\alpha)^3 = \sigma(\alpha^3) = 2$, so $\sigma\alpha$ is a root of $t^3 - 2$. Hence, all possible values of $\sigma\alpha$ is α, $\alpha\zeta$, and $\alpha\zeta^2$, where $\zeta = \exp(2\pi i/3)$.

As indicated in the preceeding theorem, all we have to do to obtain the normal closure of E over $\mathbb{Q}$ is to adjoin all σE for all $\mathbb{Q}$-embeddings σ of E into $\overline{\mathbb{Q}}$, which are $E = \mathbb{Q}(\alpha)$, $\mathbb{Q}(\alpha\zeta)$, and $\mathbb{Q}(\alpha\zeta^2)$. Hence, the normal closure of E over $\mathbb{Q}$ is $\mathbb{Q}(\alpha, \alpha\zeta, \alpha\zeta^2) = \mathbb{Q}(\alpha, \zeta)$, which is the splitting field for $t^3 - 2$ over $\mathbb{Q}$. As indicated in part (b) of the preceeding theorem, because $E/\mathbb{Q}$ is separable (for $\mathrm{char}(\mathbb{Q}) = 0$), the extension $\mathbb{Q}(\alpha, \zeta)/\mathbb{Q}$ is also separable. (Hence, $\mathbb{Q}(\alpha, \zeta)/\mathbb{Q}$ is a (finite) Galois extension.)

In particular, (b) of Theorem 18.2.9 implies that we can construct a separable and normal extension whenever a separable field extension is given, by extending the extension field, if necessary.

Definition 18.2.11 (Galois extension)**.** A field extension which is separable and normal is called a Galois extension.

Chapter 19

Further field theory

19.1 Basic properties of finite fields

For a positive prime number p, it is conventional to denote a power of p by q. We already have studied in the previous chapter that if F is a finite field of characteristic p, then $|F| = q$ for some $n \in \mathbb{N}$. Even though we do not know the existence of a finite field of order q, we first investigate some properties that a finite field of order q should satisfy.

Proposition 19.1.1. Suppose that F is a finite field of order $q = p^n$.

(a) $\text{char}(F) = p$, so the prime subfield of F is isomorphic to $\mathbb{F}_p$. Hence, we may write $\mathbb{F}_p \leq F$.

(b) Every element of F is a root of $t^q - t \in \mathbb{F}_p[t]$.

Proof. Considering an additive group F and letting $\text{char}(F) = a$ for some positive prime number a, we must have $a = p$. (b) easily follows if we consider $F^\times$ □

Observation 19.1.2. The statement in (b) implies that every element of a finite field of order q (if such a field exists) is a root of the polynomial $t^q - t$ over $\mathbb{F}_p$. Since $(t^q - t)' = qt^{q-1} - 1 = -1 \neq 0$, the polynomial $t^q - t \in \mathbb{F}_p[t]$ is separable, so it has q-distinct roots. Thus, to find a finite field of order q, there is no other choice but to consider the collection of all roots of $t^q - t \in \mathbb{F}_p[t]$, and such collection is contained in the splitting field for $t^q - t$ over $\mathbb{F}_p$.

Theorem 19.1.3 (Existence and uniqueness of finite fields)**.** Let K be the splitting field for $t^q - t \in \mathbb{F}_p[t]$ over $\mathbb{F}_p$. If we let

$$F = \{\alpha \in K : \alpha^q = \alpha\},$$

then F is a finite field of oreder q. Hence, $F = K$. By Observation 19.1.2, a finite field of order q is unique up to isomorphism.

Proof. We already proved that $|F| = q$. Thus, it remains to show that F is a subfield of K; if it is done, then F is a field which consists of all roots of $t^q - t \in \mathbb{F}_p[t]$, so F is the smallest field containing all roots of $t^q - t$, i.e., $F = K$. (For example, if $\alpha, \beta \in F$, then $(\alpha + \beta)^q = \alpha^q + \beta^q = \alpha + \beta$ and $(\alpha\beta)^q = \alpha^q\beta^q = \alpha\beta$. Proving details are left as an exercise.) □

Remark. Note that a finite field of a prime order is a set-theoretically defined field, while a finit field whose order is a power of a prime number is uniquely defined up to isomorphism, for a field of order q is the splitting field for $t^q - t \in \mathbb{F}_p[t]$ over $\mathbb{F}_p$.

Remark that we have deduced by applying the cyclic decomposition theorem that a finite multiplicative subgroup of a field is a cyclic group.

Notation. In this chapter, for an element α which is algebraic over $\mathbb{F}_q$, the minimal polynomial of α over $\mathbb{F}_q$ will be denoted by $m_{\alpha,q}(t)$.

Proposition 19.1.4. (a) $\mathbb{F}_q = \mathbb{F}_p(\alpha)$ for some $\alpha \in \mathbb{F}_q$.

(b) Given a positive integer m, there is an irreducible polynomial over $\mathbb{F}_q$ whose degree is m.

In short, if E/F is a field extension and $|E| < \infty$, then there is a primitive element $\alpha \in E$ over F.

Proof. Writing $\mathbb{F}_q^\times = \langle \alpha \rangle$ for some $\alpha \in \mathbb{F}_q^\times$, we easily obtain (a). If $\beta \in \mathbb{F}_{q^m}$ is a generator of $\mathbb{F}_{q^m}$, then $m_{\alpha,q}(t) \in \mathbb{F}_q[t]$ is a desired polynomial, for $\mathbb{F}_{q^m} = \mathbb{F}_q(\beta)$. □

Regarding the separability of polynomials, we have studied that an irreducible polynomial over a field of characteristic 0 is separable. The same property holds for finite fields.

Proposition 19.1.5. An irreducible polynomial over a finite field is separable.

Proof. Let F be a finite field of order q and let $f(t) \in F[t]$ be an irreducible polynomial. Then $f(t) = m_{\alpha,q}(t)$ for some $\alpha \in \mathbb{F}_{q^k}$. Since $\alpha^{q^k} - \alpha = 0$, $f(t)$ divides $t^{q^k} - t \in \mathbb{F}_q[t]$, which is separable. Therefore, $f(t)$ is separable. □

Proposition 19.1.6. Suppose that r, n are positive integers. Then $\mathbb{F}_{p^r} \leq \mathbb{F}_{p^n}$ if and only if $r|n$.

Proof. If $\mathbb{F}_{p^r} \leq \mathbb{F}_{p^n}$, then $\mathbb{F}_{p^n}$ is a $\mathbb{F}_{p^r}$-vector space, so $r|n$. Conversely, if $r|n$, whenever $\alpha \in \mathbb{F}_{p^r}$, we have $\alpha^{p^n} = \alpha^{p^r p^{r \cdots p^r}} = (\cdots((\alpha^{p^r})^{p^r})\cdots)^{p^r} = \alpha$, so $\mathbb{F}_{p^r} \leq \mathbb{F}_{p^n}$. □

We end this section by introducing the algebraic closure of $\mathbb{F}_q$. Let $F/\mathbb{F}_q$ be an algebraic extension and suppose $\alpha \in F$. Then $F(\alpha) \approx \mathbb{F}_{p^n}$ for some $n \in \mathbb{N}$, so we may write $\alpha \in \mathbb{F}_{p^n}$.

Theorem 19.1.7. The algebraic closure of $\mathbb{F}_q$ is, up to isomorphism, $\bigcup_{n=1}^{\infty} \mathbb{F}_{q^n}$.

Proof. It is clear that $F := \bigcup_{n=1}^{\infty} \mathbb{F}_{q^n}$ is an algebraic extension over $\mathbb{F}_q$. It remains to show that an irreducible polynomial $f(t)$ over $\mathbb{F}_q[t]$ splits completely over F. If $n = \mathit{deg}\, f(t)$ and α is a root of $f(t)$ then $\alpha \in \mathbb{F}_{p^n} \subset F$, as desired. □

19.2 Some problems in field theory

Problem 19.2.1. Find all $\mathbb{Q}$-automorphisms of $\mathbb{R}$.

Solution. Suppose that $\sigma \in \mathrm{Aut}(\mathbb{R}/\mathbb{Q})$. Since σ is the identity on $\mathbb{Q}$, if σ is continuous, then $\sigma = \mathit{id}_{\mathbb{R}}$.

Suppose that $a \in \mathbb{R}$ is positive. Then $\sigma(a) = (\sigma(\sqrt{a}))^2 \geq 0$. Hence, if $a, b \in \mathbb{R}$ and $a < b$, then $\sigma a \leq \sigma b$, i.e., σ is monotonically increasing. Thus, in particular, if $a, b \in \mathbb{R}$ and $|a - b| < 1/n$ for some integer n, then $|\sigma a - \sigma b| \leq 1/n$, so σ is continuous whenever $\sigma \in \mathrm{Aut}(\mathbb{R}/\mathbb{Q})$. By continuity, $\sigma = \mathit{id}_{\mathbb{R}}$ and $\mathrm{Aut}(\mathbb{R}/\mathbb{Q}) = \{\mathit{id}_{\mathbb{R}}\}$.

Problem 19.2.2. Let k be a field and let $t = P(x)/Q(x)$, where $P(x)$ and $Q(x)$ are relatively prime polynomials over k and $Q(x) \neq 0$. Then $k(x)$ is a simple extension over $k(t)$ obtained by adjoining t. Show that the extension degree of $k(x)/k(t)$ is $\max\{\mathit{deg}\, P(x), \mathit{deg}\, Q(x)\}$.

Solution. Because $k(x) = k(t)(x)$, it suffices to find the degree of the minimal polynomial of x over $k(t)$. Observe that the indeterminate x is satisfied by the polynomial $f(s) := P(s) - tQ(s) \in k(t)[s]$, and we will show that $f(s)$ is irreducible over $k(t)$.

In fact, since $f(s) \in k[t][s]$ and $\mathit{cont}(f) \sim_\times 1$, it suffices to show that $f(s)$ is irreducible over $k[t]$ (by Gauss's lemma), for the field of fractions of $k[t]$ is $k(t)$. Since $k[t][s] = k[t, s] = k[s][t]$, it also suffices to show that the polynomial $f(s)$ over $k[s]$ in t is irreducible over $k[s]$. The latter is clear, because $\mathit{deg}_t f(s) = 1$ and $P(s)$ and $Q(s)$ are relatively prime. Therefore, $[k(x) : k(s)] = \mathit{deg}_s f(s) = \max\{\mathit{deg}\, P(x), \mathit{deg}\, Q(x)\}$.

Problem 19.2.3 (Lüroth's theorem)**.** Show that $\mathrm{Aut}(k(x)/k) \approx PGL_2(k)$, where k is a field and x is an indeterminate.

Solution. Any k-automorphism of $k(x)$ is completely determined by its action on x. Write $\sigma(x) = f(x)/g(x)$ for some relatively prime polynomials $f(x)$, $g(x)$ over k such that $g(x) \neq 0$. Because σ fixes k, $k(\sigma(x)) = \sigma(k(x))$; because an automorphism is surjective, $\sigma(k(x)) = \sigma(x)$. Hence, $k(\sigma(x)) = k(x)$ and $[k(x) : k(\sigma(x))] = 1$. This implies that $\deg f(x)$ and $\deg g(x)$ are not greater than 1, with at least one of them being 1. Therefore, $\sigma(x) = (ax + b)/(cx + d)$ for some $a, b, c, d \in k$ and $ad - bc \neq 0$.

Conversely, assume that a k-embedding $\tau : k(x) \to k(x)$ defined by $\tau(x) = (ax + b)/(cx + d)$ with $a, b, c, d \in k$ and $ad - bc \neq 0$ is given. Because $\tau(k(x)) = k(\tau(x))$ and $[k(x) : k(\tau(x))] = \max\{\deg(ax + b), \deg(cx + d)\} = 1$, we have $\tau(k(x)) = k(x)$, so τ is a k-automorphism of $k(x)$.

So far, we have found a surjection $\rho : GL_2(k) \to \mathrm{Aut}(k(x)/k)$, defined by

$$\rho \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \sigma.$$

Because ρ is a group homomorphism, by the first isomorphism theorem, we have

$$GL_2(k)/\ker\rho \approx \mathrm{Aut}(k(x)/k),$$

where $\ker\rho = \{sI : s \in k^\times\} = Z(GL_2(k))$. Therefore, $\mathrm{Aut}(k(x)/k) \approx PGL_2(k)$.

Part VI

Galois theory

Chapter 20

Basic Galois theory

20.1 Basic observation regarding Galois extensions

Remark that an algebraic field extension which is both separable and normal is called a Galois extension. When E/F is a Galois extension, we write $\mathrm{Aut}(E/F) = \mathrm{Gal}(E/F)$; because E/F is a normal extension, we have $\mathrm{Aut}(E/F) = \mathrm{Gal}(E/F) = \mathrm{Emb}(E/F)$ and we call $\mathrm{Gal}(E/F)$ the Galois group of E/F (note that $\mathrm{Aut}(E/F)$ is a group with the multiplication being function composition.)

Remark. (a) (A review of Theorem 18.2.4) An algebraic field extension E/F is a normal extension if and only if $\mathrm{Emb}(E/F) = \mathrm{Aut}(E/F)$. Hence, in particular, if E/F is a finite extension, then E/F is a normal extension if and only if $[E : F]_{\mathsf{sep}} = |\mathrm{Aut}(E/F)|$.

(b) (Finite Galois extension) A finite field extension E/F is a Galois extension if and only if $[E : F] = |\mathrm{Aut}(E/F)|$. (It is left as an exercise to check the equivalence.)

(c) Except for Section 23.4, all Galois extensions are assumed to be a finite extension.

Note that if K/F is a Galois extension and E is an intermediate subfield of K/F, then K/E is a Galois extension and $\mathrm{Gal}(K/E) \leq \mathrm{Gal}(K/F)$. Conversely, given a subgroup H of $\mathrm{Gal}(K/F)$, we define the fixed field K^H of H in K by

$$K^H := \{x \in K : \sigma x = x \text{ for all } \sigma \in H\}.$$

It is easy to check that $F \leq K^H \leq K$.

In the following section, the fundamental theorem of finite Galois's extensions, also known as Galois's main theorem, is introduced, in which we are interested in a one-to-one bijection between the collection of the intermediate subfields of a finite Galois extension and the collection of the subgroups of the Galois group of the extension. The following propositions, which could be introduced before proving Galois's main theorem, are moved to this section, due to its generality.

Proposition 20.1.1. Suppose that E/F is a separable extension. If there is a positive integer n such that $[F(\alpha) : F] \leq n$ for all $\alpha \in E$, then (E/F is a finite extension and) $[E : F] \leq n$.

Proof. By assumption, there is an element $\beta \in E$ for which $[F(\beta) : F]$ is the greatest.

Goal: To show that $E = F(\beta)$.

In fact, our goal can easily be deduced from the primitive element theorem for finite separable extensions. If $F(\beta) < E$, then there is an element $\gamma \in E \setminus F(\beta)$ and $F(\beta) < F(\beta, \gamma)$; because $F(\beta, \gamma)/F$ is a finite separable extension, there is an element $\alpha \in F(\beta, \gamma)$ such that $F(\beta, \gamma) = F(\alpha)$, which contradicts the maximality of β. □

Lemma 20.1.2 (Artin's theorem)**.** Let K be a field and H be a finite subgroup of $\mathrm{Aut}(K)$. Then K/K^H is a Galois extension and $\mathrm{Gal}(K/K^H) = H$.[1]

[1] If H is an infinite subgroup of $\mathrm{Aut}(K)$, Artin's theorem is not valid, in general. Hence, the map $E \mapsto \mathrm{Gal}(K/E)$ is not a surjection, in general.

Proof. We first show that K/K^H is a Galois extension. For this, we show that for any $\alpha \in K$ the minimal polynomial $m(t)$ of α over K^H is separable and has all roots in K.

Let $\{\sigma_1, \cdots, \sigma_r\}$ be a maximal subset of H such that $\sigma_1\alpha, \cdots, \sigma_r\alpha$ are pairwise distinct. Letting $\tau_i = \sigma_1^{-1} \circ \sigma_i$ for each integer $i = 1, \cdots, r$, then $\tau_1\alpha, \cdots, \tau_r\alpha$ are pairwise distinct. If they are not maximally pairwise distinct, then there is another field automorphism τ_{r+1} of K such that $\tau_1\alpha, \cdots, \tau_r\alpha, \tau_{r+1}\alpha$ are pairwise distinct; then $\sigma_1\alpha, \cdots, \sigma_r\alpha, \sigma_{r+1}\alpha$ are also pairwise distinct, which contradicts the maximality of $\{\sigma_1, \cdots, \sigma_r\}$. Hence, $\{\tau_1 = \mathit{id}_K, \cdots, \tau_r\}$ is also a maximal subset of H such that $\tau_1\alpha, \cdots, \tau_r\alpha$ are pairwise distinct; thus, we may assume that $\sigma_1 = \mathit{id}_K$.

Define a polynomial

$$f(t) := (t - \sigma_1\alpha) \cdots (t - \sigma_r\alpha),$$

which is satisfied by α. Given $\tau \in H$, by the maximality of $\{\sigma_1, \cdots, \sigma_r\}$, we have $\tau \circ \sigma_i \in \{\sigma_1, \cdots, \sigma_r\}$ for all i; $\tau \in H$ *permutes* $\{\sigma_1, \cdots, \sigma_r\}$. Therefore, $f^\tau(t) = f(t)$ and $f(t) \in K^H[t]$, for evey coefficient of $f(t)$ is fixed by every field automorphism in H. Because $f(t)$ is a multiple of $m(t)$ and $f(t)$ is separable, $m(t)$ is separable and K/K^H is a separable extension. Moreover, a root of $m(t)$ is of the form $\sigma_i\alpha$ for some i, which is contained in K, so K/K^H is a normal extension. Because $[F(\alpha) : F] \leq r \leq |H| < \infty$ for all $\alpha \in K$, we conclude that K/K^H is a finite Galois extension with $[K : K^H] \leq |H|$.

We finally show that $\mathrm{Gal}(K/K^H) = H$. It is clear that $H \leq \mathrm{Gal}(K/K^H)$, thus it follows from

$$|H| \leq |\mathrm{Gal}(K/K^H)| = [K : K^H] \leq |H|$$

that $\mathrm{Gal}(K/K^H) = H$. □

20.2 Fundamental theorems of finite Galois extensions

Theorem 20.2.1 (Galois's theorem (Part I))**.** Let K/F be a finite Galois extension.

(a) There is an order-reversing bijection between the intermediate subfields of K/F and the subgroups of $\mathrm{Gal}(K/F)$, which maps an intermediate subfield E of K/F to the corresponding Galois group $\mathrm{Gal}(K/E)$ and a subgroup H of $\mathrm{Gal}(K/F)$ to the fixed field K^H.

(b) If E is an intermediate subfield of K/F, then $\mathrm{Emb}(E/F)$ is in bijection with $\mathrm{Gal}(K/F)/\mathrm{Gal}(K/E)$. Furthermore, E/F is a Galois extension if and only if $\mathrm{Gal}(K/E) \trianglelefteq \mathrm{Gal}(K/F)$. In particular, if E/F is a Galois extension, then

$$\mathrm{Gal}(E/F) \approx \frac{\mathrm{Gal}(K/F)}{\mathrm{Gal}(K/E)}.$$

Remark. Let K/F be a finite Galois extension. Suppose that $F \leq E_1 \leq E_2 \leq K$ and $H_1 \leq H_2 \leq \mathrm{Gal}(K/F)$. By order-reversing we mean that $\mathrm{Gal}(K/E_1) \geq \mathrm{Gal}(G/E_2)$ and $K^{H_1} \geq K^{H_2}$, which is easy to verify. Hence, the subfield lattice of a finite Galois extension and the *flipped* subgroup lattice of the Galois group are the same. Moreover, corresponding extension degrees and group indices coincide; for example, $[E_2 : E_1] = [K : E_1]/[K : E_2] = [\mathrm{Gal}(K/E_1) : \mathrm{Gal}(K/E_2)]$ and $[K^{H_1} : K^{H_2}] = [\mathrm{Gal}(K/K^{H_2}) : \mathrm{Gal}(K/K^{H_1})] = [H_2 : H_1]$. Finally, since an intermediate subfield E of K/F is a Galois extension over F if and only if E/F is a normal extension, E/F is a normal extension if and only if $\mathrm{Gal}(K/E)$ is a normal subgroup of $\mathrm{Gal}(K/F)$.

To prove Galois's theorem, we need the following proposition:

Proposition 20.2.2. Suppose that K/F is a finite Galois extension and $F \leq E \leq K$. Then $K^{\mathrm{Gal}(K/E)} = E$.

Proof. Write $L = K^{\mathrm{Gal}(K/E)}$. It is clear that $E \leq K^{\mathrm{Gal}(K/E)} = L$, and it follows that $\mathrm{Gal}(K/L) \leq \mathrm{Gal}(K/E)$. If $\sigma \in \mathrm{Gal}(K/E)$, then $\sigma x = x$ for all $x \in L$, thus $\sigma \in \mathrm{Gal}(K/L)$, i.e., $\mathrm{Gal}(K/E) \leq \mathrm{Gal}(K/L)$. Thus, $\mathrm{Gal}(K/E) = \mathrm{Gal}(K/L)$ and

$$[K : E] = |\mathrm{Gal}(K/E)| = |\mathrm{Gal}(K/L)| = [K : L]$$

implies that $L = E$. □

Remark. This also explains that the map $E \mapsto \mathrm{Gal}(K/E)$ is injective; if $\mathrm{Gal}(K/E_1) = \mathrm{Gal}(K/E_2)$, then $E_1 = K^{\mathrm{Gal}(K/E_1)} = K^{\mathrm{Gal}(K/E_2)} = E_2$.

Proof of (a) of Theorem 20.2.1. Clearly, the given bijection is order-reversing. If $F \leq E \leq K$, then $E \mapsto \mathrm{Gal}(K/E) \mapsto K^{\mathrm{Gal}(K/E)} = E$ by the preceeding proposition; if $H \leq \mathrm{Gal}(K/F)$, then $H \mapsto K^H \mapsto \mathrm{Gal}(K/K^H) = H$ by Artin's theorem. □

Proposition 20.2.3. Let K/F be a Galois extension (not necessarily finite) and suppose $\sigma \in \mathrm{Gal}(K/F)$. If $F \leq E \leq K$, then both K/E and $K/\sigma E$ are Galois extensions, and

$$\mathrm{Gal}(K/\sigma E) = \sigma \cdot \mathrm{Gal}(K/E) \cdot \sigma^{-1}.$$

Proof. It is clear that K/E and $K/\sigma E$ are Galois extensions. Define a map $\rho : \mathrm{Gal}(K/E) \to \mathrm{Gal}(K/\sigma E)$ by $\rho(\tau) = \sigma \circ \tau \circ \sigma^{-1}$. Then ρ is a well-defines group homomorphism with $\mathit{ker}\,\rho = \{\mathit{id}_K\}$. Also, given $\eta \in \mathrm{Gal}(K/\sigma E)$, clearly $\tau = \sigma^{-1} \circ \eta\sigma \in \mathrm{Gal}(K/E)$ and $\rho(\tau) = \eta$, so ρ is surjective. Therefore, $\mathrm{Gal}(K/\sigma E) = \mathit{im}\,\rho = \sigma \cdot \mathrm{Gal}(K/E) \cdot \sigma^{-1}$. □

Proof of (b) of Theorem 20.2.1. Because E/F is a separable extension, we have

$$|\mathrm{Emb}(E/F)| = [E : F] = \frac{[K : F]}{[K : E]} = [\mathrm{Gal}(K : F) : \mathrm{Gal}(K/E)],$$

so $\mathrm{Emb}(E/F)$ and $\mathrm{Gal}(K/F)/\mathrm{Gal}(K/E)$ are in bijection.

We now prove the normality part.

(i) Suppose that E/F is a normal extension (or equivalently, a Galois extension). Define a group homomorphism $\rho : \mathrm{Gal}(K/F) \to \mathrm{Gal}(E/F)$ by

$$\rho(\sigma) = \sigma|_E \quad (\sigma \in \mathrm{Gal}(K/F)).$$

It is clear that $\mathit{ker}\,\rho = \mathrm{Gal}(K/E)$. Given $\tau \in \mathrm{Gal}(E/F)$, there is an extension $\widetilde{\tau} : K \to \overline{F}$, where $\overline{F}$ is an algebraic closure of F containing K. The desired isomoprhism follows from the first isomoprhism theorem.

(ii) Assume that E/F is not a normal extension (or equivalently, not a Galois extension). Then there is an F-embedding $\sigma : E \hookrightarrow \overline{F}$ such that $\sigma E \neq E$. Then, $\mathrm{Gal}(K/E) \neq \mathrm{Gal}(K/\sigma E)$ by (a) of Theorem 20.2.1, while

$$\mathrm{Gal}(K/\sigma E) = \sigma \cdot \mathrm{Gal}(K/E) \cdot \sigma^{-1} = \mathrm{Gal}(K/E).$$

Therefore, if $\mathrm{Gal}(K/E)$ is a normal subgroup of $\mathrm{Gal}(K/F)$, then E/F is a normal (Galois) extension.

The desired isomoprhism under the condition that E/F is a normal extension follows from $\mathrm{Gal}(K/E) = \mathit{ker}\,\rho \trianglelefteq \mathrm{Gal}(K/F)$. □

The following further properties of Galois correspondence can be easily verified.

Theorem 20.2.4 (Galois's theorem (Part II))**.** Let K/F be a finite Galois extension and suppose E_1, E_2 are intermediate subfields of K/F. Write $H_1 = \mathrm{Gal}(K/E_1)$ and $H_2 = \mathrm{Gal}(K/E_2)$.

(a) $\mathrm{Gal}(K/E_1E_2) = H_1 \cap H_2$, i.e., $K^{H_1 \cap H_2} = E_1E_2$.

(b) $\mathrm{Gal}(K/(E_1 \cap E_2)) = \langle H_1, H_2 \rangle$, i.e., $K^{\langle H_1, H_2 \rangle} = E_1 \cap E_2$.

Proof. We first show that $K^{H_1 \cap H_2} = E_1E_2$. Since E_1E_2 contains E_1 and E_2, $\mathrm{Gal}(K/E_1E_2)$ is contained in H_1 and H_2, so $K^{H_1 \cap H_2} \leq K^{\mathrm{Gal}(K/E_1E_2)} = E_1E_2$. Conversely, since $H_1 \cap H_2$ is contained in H_1 and H_2, its fixed field $K^{H_1 \cap H_2}$ contains E_1 and E_2, hence $E_1E_2 \leq K^{H_1 \cap H_2}$.

We now prove the second correspondence. Since $E_1 \cap E_2$ is contained in E_1 and E_2, $\mathrm{Gal}(K/(E_1 \cap E_2))$ contains $H_1 \cup H_2$, hence $E_1 \cap E_2 \leq K^{\langle H_1, H_2 \rangle}$. Conversely, since $K^{\langle H_1, H_2 \rangle}$ is contained in K^{H_1} and K^{H_2}, $K^{\langle H_1, H_2 \rangle} \leq E_1 \cap E_2$. □

Theorems 20.2.1 and 20.2.4, together, are called Galois's main theorem. Sometimes, the following correspondence is also included in Galois's main theorem.

Proposition 20.2.5. Suppose that K/F and L/F are finite Galois extensions, where $K, L \leq \overline{F}$. Then KL/F is a finite Galois extension and $\mathrm{Gal}(KL/L) \approx \mathrm{Gal}(K/(K \cap L))$.

Proof. By assumption, it is clear that the extension KF/L is finite, separable, and normal, i.e., KF/L is a finite Galois extension. To show the isomorphism, consider the map $\rho : \mathrm{Gal}(KL/L) \to \mathrm{Gal}(K/(K \cap L))$ defined by $\rho(\sigma) = \sigma|_K$ for $\sigma \in \mathrm{Gal}(KL/L)$. Since K/F is a normal extension, $\sigma(K) = K$ for all $\sigma \in \mathrm{Gal}(KL/L)$, i.e., ρ is a well-defined group homomorphism. Also, $\ker\rho = \{id_{KL}\}$, so ρ is injective. Thus, it remains to show that ρ is surjective. The problem in this step is we could not apply the isomoprhism extension theorem to prove the surjectivity, for an extension may not be the identity map on L. Instead, we show that ρ is surjective by showing $im\,\rho = \mathrm{Gal}(K/(K \cap L))$. For this, it suffices to show $im\,\rho \geq \mathrm{Gal}(K/(K \cap L))$, or equivalently, $K^{im\,\rho} \leq K \cap L$. If $x \in K^{im\,\rho}$ and $\sigma \in \mathrm{Gal}(KL/L)$, then $\sigma(x) = \sigma|_K(x) = x$. Thus, $x \in (KL)^{\mathrm{Gal}(KL/L)} = L$, implying that $K^{im\,\rho} \leq K \cap L$, as desired. □

Remark. The following results can be considered corollaries of the above proposition: Suppose K/F and L/F are finite Galois extensions with $K, L \leq \overline{F}$.

(a) $[KL : F] = [KL : L][L : F] = \dfrac{[K : F][L : F]}{[K \cap L : F]}$.

(b) $[KL : L]$ divides $[K : F]$. In particular, $[KL : L] = [K : F]$ if and only if $K \cap L = F$.

Remark. In the above proof, one might think of the following proof with a gap when proving the surjectivity of ρ by constructing an extension $\widetilde{\tau} \in \mathrm{Gal}(KL/L)$ of $\tau \in \mathrm{Gal}(K/(K \cap L))$.

(1) Since L/F is a finite separable extension, by the primitive element theorem, there is an element $\gamma \in L$ such that $L = (K \cap L)(\gamma)$.

(2) Hence, $KL = K(\gamma)$. If $\gamma \in K$, there is nothing to prove, for $K = KL$ and $L \leq K$.

(3) Assume $\gamma \notin K$. Since γ is an algebraic conjugate of γ, by the isomoprhism extension theorem, there is an extension $\widetilde{\tau} : KL \to KL$ extending τ. It is easy to check that $\widetilde{\tau} \in \mathrm{Gal}(KL/L)$.

In the above proof, when one seeks to apply the isomophism extension theorem, one should consider the minimal polynomial $m(t)$ of γ over K and $m^\tau(t)$. Although $m(\gamma) = 0$ is clear, $m^\tau(t)$ may not be satisfied by γ, unless $m(t) \in (K \cap L)[t]$, where the field $K \cap L$ is fixed by τ. (We say this error is a 'gap,' because it is in fact true, as justified in Problem 20.3.1.)

The following correspondence will show some significance when computing the Galois group of a reducible polynomial over a field.

Proposition 20.2.6. Suppose that K/F and L/F are Galois extensions, where $K, L \leq \overline{F}$, and define the map $\rho : \mathrm{Gal}(KL/F) \to \mathrm{Gal}(K/F) \times \mathrm{Gal}(L/F)$ by $\rho(\sigma) = (\sigma|_K, \sigma|_L)$ for $\sigma \in \mathrm{Gal}(KL/F)$.

(a) ρ is a well-defined group monomorphism. Hence, $\mathrm{Gal}(KL/F)$ embeds into $\mathrm{Gal}(K/F) \times \mathrm{Gal}(L/F)$.

(b) Assume further that the field extensions K/F and L/F are finite. Then $|im\,\rho| = [KL : F] = [K : F][L : F] = |\mathrm{Gal}(K/F) \times \mathrm{Gal}(L/F)|$ if and only if $K \cap L = F$. In other words, ρ is a group isomophism if and only if $K \cap L = F$.

Observation 20.2.7. We review some properties regarding field compositions. Suppose that K, L are intermediate subfields of $\overline{F}/F$ so that the composition KL of K and L is well-defined.

(i) If K/F is a finite (algebraic, separable, normal) extension, then so is KL/L.

(ii) Hence, if K/F is a finite Galois extension, then so is KL/L. (In fact, if K/F is a Galois extension, then so is KL/L.) Furthermore, we have $\mathrm{Gal}(KL/L) \approx \mathrm{Gal}(K/(K \cap L))$. Hence, $[KL : L] = [K : K \cap L]$, and $[KL : L] = [K : F]$ if and only if $K \cap L = F$.

Note from Theorem 18.2.9 that given an algebraic field extension E/F there is the smallest field K containing E such that K/F is a normal extension, and that K/F is the smallest Galois extension if E/F is assumed to be a separable extension. Thus, the normal closure of a separable extension is often called a Galois closure.

One last remark:

Remark. Suppose that K/F is a Galois extension and $F \leq E \leq K$. Let $\sigma \in \mathrm{Gal}(K/F)$ and $\tau \in \mathrm{Emb}(E/F)$, where an algebraic closure $\overline{F}$ of F is given. By the isomophism extension theorem, there is an F-embedding $\widetilde{\tau} : K \hookrightarrow \overline{F}$ extending τ. Since K/F is a normal extension, $\widetilde{\tau}K = K$, so $\tau E \leq K$ and $\sigma \circ \tau : E \hookrightarrow K$ is a well-defined F-embedding of E into $\overline{F}$. Hence, an element of $\mathrm{Gal}(K/F)$ permutes the elements of $\mathrm{Emb}(E/F)$ by left multiplication.

20.3 Some problems regarding Galois's theorem

Problem 20.3.1. Suppose that K/F and L/F are finite Galois extensions with $K, L \leq \overline{F}$. As in the preceeding remark, write $L = (K \cap L)(\gamma)$. Show that the minimal polynomial of γ over K and over $K \cap L$ are the same.

Solution. Using the suggested notations, we have $KL = K(\gamma)$. Letting $m_1(t)$ and $m_2(t)$ be the minimal polynomial of γ over $K \cap L$ and over K, respectively, because $m_2(t)|m_1(t)$, it suffices to show that $[K : K \cap L] = [KL : K]$ (already justified); this implies $\mathit{deg}\, m_1(t) = \mathit{deg}\, m_2(t)$ and $m_1(t) = m_2(t)$, as desired.

Problem 20.3.2. Find the minimal polynomial of $1 + \sqrt[3]{2} + \sqrt[3]{4}$ over $\mathbb{Q}$.

Solution. Considering a Galois extension over $\mathbb{Q}$ containing $\alpha := 1 + \sqrt[3]{2} + \sqrt[3]{4}$, it is natural to suggest $E := \mathbb{Q}(\rho, \zeta)$, the splitting field for $t^3 - 2$ over $\mathbb{Q}$. (Here, $\rho = \sqrt[3]{2}$ and $\zeta = \exp(2\pi i/3)$.) Computing its Galois group, we find that $\mathrm{Gal}(E/\mathbb{Q}) = \langle \sigma, \tau \rangle$, where

$$\sigma : \begin{cases} \rho \mapsto \rho\zeta \\ \zeta \mapsto \zeta \end{cases}, \quad \tau : \begin{cases} \rho \mapsto \rho \\ \zeta \mapsto \zeta^{-1} \end{cases}.$$

If $f(t)$ is a nonconstant polynomial over $\mathbb{Q}$, then $f(t)$ is fixed by the action of automorphisms in $\mathrm{Gal}(E/\mathbb{Q})$ by Galois's theorem. Hence, if $f(\alpha) = 0$, then $\eta\alpha$ is also a root of $f(t)$, where $\eta \in \mathrm{Gal}(E/\mathbb{Q})$; this implies that the minimal polynomial $m(t)$ of α over $\mathbb{Q}$ is necessarily satisfied the following elements:

$$1 + \rho + \rho^2, \quad 1 + \rho\zeta + \rho^2\zeta^{-1}, \quad 1 + \rho\zeta^{-1} + \rho^2\zeta.$$

Because $\mathbb{Q}$ is of characteristic 0, $m(t)$ is separable. Hence, $m(t)$ is divisible by

$$(t - (1 + \rho + \rho^2))(t - (1 + \rho\zeta + \rho^2\zeta^{-1}))(t - (1 + \rho\zeta^{-1} + \rho^2\zeta)) = t^3 - 3t^2 - 3t - 1.$$

Since $t^3 - 3t^2 - 3t - 1$ is irreducible over $\mathbb{Q}$, we conclude that $m(t) = t^3 - 3t^2 - 3t - 1$.

Remark. In fact, when K/F is a finite Galois extension and $\alpha \in K$, then the minimal polynomial of α over F is the square-free part $p(t)$ of

$$\prod_{\sigma \in \mathrm{Gal}(K/F)} (t - \sigma\alpha).$$

Here's a justification. By Galois's theorem, the minimal polynomial $m(t)$ of α over F is satisfied by $\sigma\alpha$ for all $\sigma \in \mathrm{Gal}(K/F)$, so $m(t)$ is necessarily divisible by $p(t)$. On the other hand, because the roots of $p(t)$ are the whole pairwise distinct $\sigma\alpha$'s (even when counting multiplicity), we have $p(t) \in F[t]$ by Galois's theorem. This proves that $p(t)$ is the minimal polynomial of α over F.

Remark. Because a finite Galois extension is a finite separable extension, the primitive element theorem is applicable in this case. Hence, if K/F is a finite Galois extension, then there is an element $\alpha \in K$ such that $K = F(\alpha)$.

Problem 20.3.3. Let K/F be a Galois group of extension degree p^n, where p is a positive prime number and n is a positive integer. Show that, for each integer $1 \leq r \leq n$, that there is a subfield E_r of K/F such that E_r/F is a Galois extension of extension degree p^r.

Solution. Since $\mathrm{Gal}(K/F)$ is a group of order p^n, for each integer $1 \leq r \leq n$, there is a normal subgroup H_r of G such that $[G : H_r] = p^r$.[2] Therefore, the fixed field $E_r = K^{H_r}$ is a Galois extension over F of degree p^r.

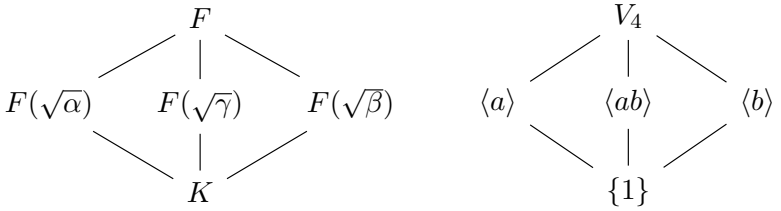
Problem 20.3.4 (Biquadratic extension)**.** Let F be a field such that $\mathrm{char}(F) \neq 2$.

(a) Suppose that $K = F(\sqrt{D_1}, \sqrt{D_2})$, where $D_1, D_2 \in F$ have the property that none of D_1, D_2 or $D_1 D_2$ is a square in F. Prove that K/F is a Galois extension with the Galois group isomorphic to the Klein 4-group.

(b) Conversely, suppose that K/F is a Galois extenion with the Galois group isomorphic to the Klein 4-group. Show that $K = F(\sqrt{D_1}, \sqrt{D_2})$, where $D_1, D_2 \in F$ have the property that none of D_1, D_2 or $D_1 D_2$ is a square in F.

Solution. Assume first that $K = F(\sqrt{D_1}, \sqrt{D_2})$, where $D_1, D_2 \in F$ have the property that none of D_1, D_2 or $D_1 D_2$ is a square in F. Then K/F is clearly a Galois extension with the extension degree at most 4. In fact, K is the splitting field for $(t^2 - D_1)(t^2 - D_2)$ over F, so (after identification) $\mathrm{Gal}(K/F) \leq \mathrm{Gal}(F(\sqrt{D_1})/F) \times \mathrm{Gal}(F(\sqrt{D_2})/F) \approx V_4$. Moreover, because none of D_1, D_2 or $D_1 D_2$ is a square in F, $F(\sqrt{D_1}) \cap F(\sqrt{D_2}) = F$, thus the Galois group of K/F is the Klein 4-group.

Conversely, assume that K/F is a Galois extension with the Galois group isomorphic to the Klein 4-group $\langle a, b : a^2, b^2, aba^{-1}b^{-1} \rangle$.



where $\alpha, \beta, \gamma \in F$ are not squares in F. Because a fixes α and b fixes β, $ab = a \circ b$ fixes $\alpha\beta$. Moreover, since $F(\sqrt{\alpha}) \neq F(\sqrt{\beta})$, $\alpha\beta$ is not a square in F. (Hence, we may let $\gamma = \alpha\beta$.) Therefore, $K = F(\sqrt{\alpha}, \sqrt{\beta})$.

20.4 Abelian extensions and solvable extensions

Definition 20.4.1. Let K/F be a Galois extension.

(a) (Abelian extension) K/F is called an abelian extension if $\mathrm{Gal}(K/F)$ is an abelian group.

(b) (Solvable extension) K/F is called a solvable extension if $\mathrm{Gal}(K/F)$ is a solvable group.

Throughout this section, in order for Galois's theorem to be valid, we assume that all extensions are finite extensions. We introduce some applications of Galois's theorem in some kinds of finite extensions.

Observation 20.4.2. If K/F is a finite cyclic(abelian) extension and $F \leq E \leq K$, then K/E and E/F are also cyclic(abelian) extensions. The converse is not true, in general.

Proposition 20.4.3. Suppose that $F \leq E \leq K$ and K/F is a finite Galois extension. If E/F and K/E are solvable extensions, then K/F is also a solvable extension.

Proof. $\mathrm{Gal}(E/F) \approx \mathrm{Gal}(K/F)/\mathrm{Gal}(K/E)$ and $\mathrm{Gal}(K/E)$ are solvable, so $\mathrm{Gal}(K/F)$ is solvable. □

Proposition 20.4.4. Suppose that $F \leq E \leq K$ and K/F is a solvable extension.

[2] This can be proved by deducing that the center of the Galois group is nontrivial and then applying the lattice isomorphism theorem for groups.

(a) K/E is a solvable extension.

(b) If E/F is a Galois extension, then E/F is a solvable extension.

Proof. (a) is clear, because a subgroup of a solvable group is solvable. (b) is clear, because $\mathrm{Gal}(E/F) \approx \mathrm{Gal}(K/F)/\mathrm{Gal}(K/E)$ is a quotient of a solvable group by its normal subgroup. □

Proposition 20.4.5. Suppose that K/F is a finite Galois extension. Then the followings are equivalent:

(a) K/F is a (finite) solvable extension.

(b) K/F has an abelian tower; there are fields $F_1, \cdots F_k$ such that

$$F = F_0 \leq F_1 \leq \cdots \leq F_{k-1} \leq F_k = K$$

and F_i/F_{i-1} is an abelian extension for $i = 1, \cdots, k$.

Proof. (This equivalence is almost clear by Galois's theorem.) Assume first that K/F is a finite solvable extension, and let $\{id_K\} = H_0 \trianglelefteq H_1 \trianglelefteq \cdots \trianglelefteq H_k = \mathrm{Gal}(K/F)$ be a 'solvability chain' of $\mathrm{Gal}(K/F)$. It is easy to check that $K = K^{H_0} \geq K^{H_1} \geq \cdots \geq K^{H_k} = F$ and $K^{H_{i-1}}/K^H$ is an abelian extension. Assuming that K/F has an abelian tower, Galois's theorem establishes a corresponding tower for $\mathrm{Gal}(K/F)$, and it is clear that $\mathrm{Gal}(F_i/F_{i-1}) \approx \mathrm{Gal}(K/F_{i-1})/\mathrm{Gal}(K/F_i)$ is abelian. □

Some topics regarding abelian extensions will be introduced later when studying cyclotomic extensions. In studying solvability of polynomial equations, we will justify that the equation is solvable by radicals if and only if its Galois group is a solvable group.

20.5 Galois groups of polynomials

Remark. Let F be a field and $f(t)$ be a nonconstant separable polynomial over F, and let $\alpha_1, \cdots, \alpha_n$ be the roots of $f(t)$, where $n = \mathit{deg}\, f(t)$. Let K be the splitting field for $f(t)$ over F.

(a) Every automorphism $\sigma \in \mathrm{Gal}(K/L)$ permutes the root of $f(t)$. Hence, the group action of $\mathrm{Gal}(K/L)$ on $\{\alpha_1, \cdots, \alpha_n\}$ (by left multiplication) affords a group embedding

$$\mathrm{Gal}(K/L) \hookrightarrow S_n.$$

In general, if $f(t) = f_1(t) \cdots f_k(t)$ is the factorization of $f(t)$ into irreducible polynomials in $F[t]$, then an automorphism $\sigma \in \mathrm{Gal}(K/L)$ permutes the roots of $f_i(t)$ for each $1 \leq i \leq k$, i.e., $\mathrm{Gal}(K/L)$ permutes the roots of the irreducible factors among themselves. Thus, the group action affords a group embedding

$$\mathrm{Gal}(K/L) \hookrightarrow S_{n_1} \times \cdots \times S_{n_d},$$

where $n_i = \mathit{deg}\, f_i(t)$ for each i.

(b) In particular, suppose that $f(t)$ is irreducible. Given any two roots α_i and α_j $(1 \leq i, j \leq n)$, by isomorphism extension theorem, id_F extends to an F-automorphism $\sigma : K \to K$ such that $\sigma\alpha_i = \alpha_j$ (how?). In other words, $\mathrm{Gal}(K/F)$ is transitive on the roots of each irreducible factor of $f(t)$.

(c) In (a), suppose, in particular, that $f(t) = g(t)h(t)$ for some nonconstant polynomial $g(t), h(t)$ over F. Then the splitting field K_f for $f(t)$ over F is the composition of the splitting field K_g and K_h for $g(t)$ and $h(t)$ over F, respectively. Thus, $G_f \hookrightarrow G_g \times G_h$ as explained in (a); here, $G_f \approx G_g \times G_h$ if and only if $K_g \cap K_h = F$.

Notation. Given a nonconstant polynomial $f(t)$ over F and its splitting field over F, $\mathrm{Gal}(K/F)$ is called the Galois group of $f(t)$ over F, and is denoted by $G_{f,F}$, or simply by G_f, if the base field F is understood. Also, unless stated otherwise, the splitting field for $f(t)$ over F is denoted by $K_{f,F}$, or shortly by K_f, if the base field F is understood.

Some basic propositions required to investigate Galois groups are given as the following two propositions.

Proposition 20.5.1. Let $f(t)$ be a nonconstant separable polynomial over F, and write $n = \deg f(t)$.

(a) G_f embeds into S_n. Furthermore, if $f(t) = f_1(t) \cdots f_k(t)$ is the factorization of $f(t)$ into irreducible polynomials over F, then G_f embeds into $S_{n_1} \times \cdots \times S_{n_k}$, where $n_i = \deg f_i(t)$ for each $i = 1, \cdots, k$.

(b) If $f(t)$ is irreducible over F, then n divides the order of G_f.

Proof. (a) is already proved in the beginning of this section. If $f(t)$ is irreducible and α is any root of $f(t)$, then the splitting field for $f(t)$ over F contains α, so n divides $|G_f|$. □

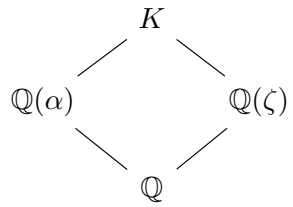
Proposition 20.5.2. Let $f(t)$ be a nonconstant separable polynomial over F. Then $f(t)$ is irreducible over F if and only if G_f acts transitively on the roots of $f(t)$.

Proof. Suppose that $f(t)$ is irreducible over F and let α and β be any roots of $f(t)$. By isomorphism extension theorem, there is an F-automorphism of K mapping α to β (why?), as desired.

Assume conversely that G_f acts on the roots of $f(t)$ transitively. If $f(t)$ is reducible, there are nonconstant polynomials $g(t), h(t) \in F[t]$ such that $f(t) = g(t)h(t)$. If α and β are roots of $g(t)$ and $h(t)$, respectively, there is an automorphism $\sigma \in G_f$ such that $\sigma\alpha = \beta$. So β is a root of $g(t)$, for $0 = \sigma(g(\alpha)) = g(\sigma\alpha) = g(\beta)$, which contradicts the separability of $f(t)$. □

From now on, throughout this section, F is assumed to be a perfect field, over which every irreducible polynomial is separable.

Example 20.5.3. Consider $f(t) = t^3 - 2 \in \mathbb{Q}[t]$. Its roots are $\alpha, \alpha\zeta, \alpha\zeta^2$, where $\alpha = \sqrt[3]{2}, \zeta = \exp(2\pi i/3)$. We then have the following (possibly incomplete) subfield lattice:



In fact, since $f(t)$ is irreducible over $\mathbb{Q}$, 3 divides $|G_f|$; because $[\mathbb{Q}(\zeta) : \mathbb{Q}] = 2$, 2 also divides $|G_f|$; because $G_f \hookrightarrow S_3$, we conclude that $G_f \approx S_3$.

Example 20.5.4. Let $f(t) = t^4 - 2 \in \mathbb{Q}[t]$, and write $\alpha = \sqrt[4]{2}$. Then the splitting field K for $f(t)$ over $\mathbb{Q}$ is $\mathbb{Q}(\alpha, i)$. Since $[K : \mathbb{Q}] = [K : \mathbb{Q}(\alpha)][\mathbb{Q}(\alpha) : \mathbb{Q}] = 8$ and $f(t)$ is irreducible over $\mathbb{Q}$, G_f is isomorphic to a transitive subgroup of S_4. Therefore, $G_f \approx D_8$, where D_8 is the dihedral group of order 8.

To investigate explicitly, note that $\mathbb{Q}(i)/\mathbb{Q}$ and $K/\mathbb{Q}(i)$ are separable extensions. Thus, there are two distinct $\mathbb{Q}$-embbedings of $\mathbb{Q}(i)$ into $\overline{\mathbb{Q}}$:

$$id_{\mathbb{Q}(i)}, \quad \gamma : i \mapsto -i.$$

Also, there are four distinct embeddings extending $id_{\mathbb{Q}(i)}$ and γ, respectively; they map α to either α or $-\alpha$ or αi or $-\alpha i$. Letting σ and τ be the $\mathbb{Q}$-automorphisms such that

$$\sigma(\alpha) = \alpha, \quad \sigma(i) = -i,$$
$$\tau(\alpha) = \alpha i, \quad \tau(i) = i,$$

we find that $G_f = \langle \sigma, \tau \mid \sigma^2 = \tau^4 = id_K, \sigma\tau\sigma = \tau^{-1} \rangle \approx D_8$.

Example 20.5.5. Let $f(t) = (t^2 - 2)(t^3 - 2) \in \mathbb{Q}[t]$ and write $a(t) = t^2 - 2$ and $b(t) = t^3 - 2$. Since $\sqrt{2} \notin K_b$, we have $K_a \cap K_b = \mathbb{Q}$, so $G_f \approx G_a \times G_b \approx Z_2 \times S_3$.

Example 20.5.6. Let $f(t) = (t^2 - 2)(t^2 - 3)(t^3 - 2) \in \mathbb{Q}[t]$. Then $K_f = \mathbb{Q}(\alpha, \beta, i)$, where $\alpha = \sqrt[6]{2}$ and $\beta = \sqrt{3}$. Letting $a(t) = t^2 - 2$ and $b(t) = (t^2 - 3)(t^3 - 2)$, we have $K_a = \mathbb{Q}(\sqrt{2})$ and $K_b = \mathbb{Q}(\sqrt[3]{2}, \sqrt{3}, i)$, so $K_a \cap K_b = \mathbb{Q}$ and $G_f \approx G_a \times G_b$. It is easy to check that $G_b \approx Z_2 \times S_3$, so $G_f \approx Z_2 \times Z_2 \times S_3 \approx V_4 \times S_3$.

Example 20.5.7. Let $f(t) = (t^2 - 5)(t^5 - 1) \in \mathbb{Q}[t]$. Since $\sqrt{5} \in \mathbb{Q}(\zeta_5)$, $K_f = \mathbb{Q}(\zeta_5)$, thus $G_f \approx (\mathbb{Z}/5\mathbb{Z})^\times \approx Z_4$.

Example 20.5.8. We will find a necessary and sufficient condition of an integer d for $\sqrt{d}$ being contained in $\mathbb{Q}(\zeta_5)$. In fact, $\sqrt{d} \in \mathbb{Q}(\zeta_5)$ if and only if $\mathbb{Q}(\sqrt{d})$ is a subfield of $\mathbb{Q}(\zeta_5)$ containing $\mathbb{Q}$. By Galois theorem, the only proper subfield of $\mathbb{Q}(\zeta_5)/\mathbb{Q}$ is $\mathbb{Q}(\zeta_5 + \zeta_5^{-1}) = \mathbb{Q}(\sqrt{5})$, so $\sqrt{d} \in \mathbb{Q}(\zeta_5)$ if and only if $d = 5k^2$ for an integer k. In particular, $\zeta_3 \notin \mathbb{Q}(\zeta_5)$, for otherwise, $\sqrt{-3} \in \mathbb{Q}(\zeta_5)$, which contradicts our result.

Example 20.5.9. Let $f(t) = (t^3 - 2)(t^3 - 3) \in \mathbb{Q}[t]$ and write $\alpha = \sqrt[3]{2}$ and $\beta = \sqrt[3]{3}$. Letting $a(t) = t^3 - 2$ and $b(t) = t^3 - 3$, we have

$$[K_f : \mathbb{Q}] = \frac{[K_a : \mathbb{Q}][K_b : \mathbb{Q}]}{[K_a \cap K_b : \mathbb{Q}]} = 18.$$

To determine the isomorphic type of G_f, note that G_f is not abelian. Also, since $G_f \hookrightarrow S_3 \times S_3$, there is no element in G_f or order 9. Thus, G_f is one among all possible nonabelian groups of order 18 with no element of order 9.

Chapter 21

Some Galois extensions

21.1 Galois extensions over finite fields

Throughout this section, p is a positive prime number and $q = p^n$ for some positive integer n.

Theorem 21.1.1. $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p) = \langle \sigma_p \rangle \approx \mu_n$, where $\sigma_p : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ is the $\mathbb{F}_p$-automorphism defined by $\sigma_p(x) = x^p$ for $x \in \mathbb{F}_{p^n}$.

Proof. Of course, one could directly show that $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$ and the cyclic group generated by σ_p are the same. Its rigorous justification, however, seems to have technical difficulty. Thus, we prove the thoerem by justifying that the subgroup $\langle \sigma_p \rangle$ of the Galois group has the same order of the Galois group.

Note that $|\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)| = [\mathbb{F}_{p^n} : \mathbb{F}_p] = n$ and $\sigma_p \in \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$. Since $\sigma_p^r(x) = x^{p^r}$ for all integer r and $x \in \mathbb{F}_{p^n}$, the order of σ_p is n, as desired. □

Observation 21.1.2. (a) Considering the subgroup lattice of μ_n, there is a unique subgroup of index d, where d is a positive divisor of n. By Galois's thoerem, it is equivalent to the statement that there is a unique intermediate subfield of $\mathbb{F}_{p^n}/\mathbb{F}_p$ whose extension degree over $\mathbb{F}_p$ is d. In fact, $\mathbb{F}_{p^d}$ is such a field, so $\mathbb{F}_{p^d}$ is a unique interediate subfield of $\mathbb{F}_{p^n}/\mathbb{F}_p$ whose extension degree over $\mathbb{F}_p$ is d. Because μ_n is abelian, $\mathbb{F}_{p^d}/\mathbb{F}_p$ is clearly a (finite) Galois extension, and we have

$$\mathrm{Gal}(\mathbb{F}_{p^d}/\mathbb{F}_p) \approx \frac{\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)}{\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_{p^d})}.$$

(b) Let x be an element of $\mathbb{F}_{p^n}$ and d be a positive divisor of n. By Galois's theorem, $x \in \mathbb{F}_{p^d}$ if and only if $\sigma_p^{n/d} x = x$. Writing $\mathbb{F}_{p^n} = \mathbb{F}_p(\alpha)$ for some $\alpha \in \mathbb{F}_{p^n}$, we can write $x = \alpha^k$ for some positive integer k.

Proposition 21.1.3. The polynomial $t^{p^n} - t$ is precisely the product of all the distinct irreducible polynomials over $\mathbb{F}_p$ of degree d, where d runs through all positive divisors of n.

Proof. □

We have observed that $t^p - t \in \mathbb{F}_p[t]$ is a *reducible* separable polynomial, whose roots are exactly the elements of $\mathbb{F}_p$. If one adds a nonzero constant to the polynomial, one can get a irreducible separable polynomial, as illustrated in the following proposition.

Proposition 21.1.4 (Artin-Schreier extension)**.** Let p be a positive prime number and a be a nonzero element of $\mathbb{F}_p$.

(a) The polynomial $f(t) = t^p - t + a \in \mathbb{F}_p[t]$ is irreducible and separable over $\mathbb{F}_p$.

(b) The splitting field K for $f(t)$ over $\mathbb{F}_p$ is $\mathbb{F}_{p^p}$. Writing $K = \mathbb{F}_p(\alpha)$, $\mathrm{Gal}(K/\mathbb{F}_p) = \langle \sigma_p \rangle$, where $\sigma_p : K \to K$ is the $\mathbb{F}_p$-automorphism defined by $\sigma_p \alpha = \alpha + 1$.

Proof. Because $f'(t) = -1 \neq 0$, $f(t)$ is separable. To show the irreducibility of $f(t)$, observe that $\beta + 1$ is a root of $f(t)$ if $\beta \in K$ is a root of $f(t)$. So, (after writing $\beta_i = \beta + (i-1)$ for each integer $1 \leq i \leq p$) we may write

$$f(t) = (t - \beta_1)(t - \beta_2) \cdots (t - \beta_p).$$

Letting $m_i(t)$ be the minimal polynomial of β_i over $\mathbb{F}_p$, we have $m_{i+k}(t) = m_i(t-k)$ for all allowed indices i, k. Hence, if $f(t)$ is not irreducible, then $\deg m_1(t) = \cdots = \deg m_p(t) < p$, so $\beta_i \in \mathbb{F}_p$ for all $1 \leq i \leq p$; then, $f(t) = t^p - t$ and $a = 0$, a contradiction. Therefore, when $a \in \mathbb{F}_p$ is nonzero, $f(t)$ is an irreducible and separable polynomial over $\mathbb{F}_p$.

Since $\deg f(t) = p$, the splitting field K for $f(t)$ over $\mathbb{F}_p$ is isomorphic to the finite field $\mathbb{F}_{p^p}$ and $\mathrm{Gal}(K/\mathbb{F}_p) \approx \mu_p$. Thus, we may write $K = \mathbb{F}_p(\alpha)$ for some $\alpha \in K$. To complete the proof, we need to justify that the *map* $\sigma_p : K \to K$ defined by $\sigma_p \alpha = \alpha + 1$ is an $\mathbb{F}_p$-automorphism of K of order p, which is easy to check. □

21.2 More on finite fields

21.3 Cyclotomic extensions

In this section, we study the polynomial $t^n - 1 \in F[t]$, where n is a positive integer and F is a fiven base field.

Definition 21.3.1 (n-th root of unity)**.** For a positive integer n, every member of the collection

$$\mu_n(\overline{F}) := \{\alpha \in \overline{F} : \alpha^n = 1\}$$

is called an n-th root of unity. Since $\mu_n(\overline{F})$ is a finite subgroup of the multiplicative group $F^\times$, $\mu_n(\overline{F})$ is a cyclic group. A generator of the finite cyclic group $\mu_n(\overline{F})$ is called a primitive n-th root of unity.

Example 21.3.2. $\mu_n(\overline{\mathbb{Q}}) = \{\exp(2\pi ik/n) : k \text{ is an integer such that } 0 \leq k \leq n-1\}$.

21.3.1 The splitting field for $t^n - 1$ over a finite field

Observation 21.3.3. Suppose that F is a field of characteristic $p > 0$ and let $n = qm$, where m is a positive integer relatviely prime to p.

(a) $\mu_p(\overline{F}) \approx \mathbb{F}_p$ and $\mu_q(\overline{F}) \approx \mathbb{F}_q$.

(b) We now show that $\mu_n(\overline{F}) = \mu_m(\overline{F})$ It is clear that $\mu_m(\overline{F}) \subset \mu_n(\overline{F})$. Suppose that $\alpha \in \mu_n(\overline{F})$. Then $(\alpha^m)^q = 1$ and $(\alpha^m - 1)^q = 0$, hence $\alpha \in \mu_m(\overline{F})$. Therefore, if $\mathrm{char}(F) = p > 0$ and when we consider $\mu_n(\overline{F})$, we may assume that $(n, p) = 1$.

Observation 21.3.4. Let F be a field of characteristic $p > 0$ and n be a positive integer which is relatively prime to p. Let ζ be a primitive n-th root of unity.

(a) $F(\zeta)$ is the splitting field for $t^n - 1 \in F[t]$ over F. Since $t^n - 1$ is a separable polynomial over F, $F(\zeta)/F$ is a finite Galois extension and $\mu_n(\overline{F}) = \{1, \zeta, \cdots, \zeta^{n-1}\}$.

(b) Since $\sigma \in \mathrm{Aut}(F(\zeta))$ fixes 1, σ permutes $\mu_n(\overline{F})$, i.e., $\sigma(\mu_n(\overline{F})) = \mu_n(\overline{F})$. Therefore, $\sigma\mu$ is also a primitive n-th root of unity, hence $\sigma\mu = \mu^k$ for some integer which is relatively prime to n.

Our first goal is to compute $\mathrm{Gal}(F(\zeta)/F)$ when $\mathrm{char}(F) = p > 0$.

Theorem 21.3.5. Suppose that F is a field of characteristic $p > 0$ and let n be a positive integer relatively prime to p, and let ζ be a primitive n-th root of unity. Then $\mathrm{Gal}(F(\zeta)/F)$ embeds into $(\mathbb{Z}/n\mathbb{Z})^\times$, so $F(\zeta)/F$ is an abelian extension.

Proof. Let $\phi : \mathrm{Gal}(F(\zeta)/F) \to (\mathbb{Z}/n\mathbb{Z})^\times$ be the map defiend by

$$\phi(\sigma) = \overline{k} \quad (\sigma \in \mathrm{Gal}(F(\zeta)/F))$$

where k is an integer such that $\sigma(\zeta) = \zeta^k$. Since $\sigma \in \mathrm{Gal}(F(\zeta)/F)$ is a field automorphism of $F(\zeta)$, $\sigma(\zeta)$ is a primitive n-th root of unity. Hence, $(n,k) = 1$ and $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$. It is easy to check that ϕ is an injective group homomorphism, as desired. □

Remark. Unlike the splitting field for $t^n - 1$ over $\mathbb{Q}$ which will be studied in the following subsection, $\mathrm{Gal}(F(\zeta)/F)$ need not be isomorphic to $(\mathbb{Z}/n\mathbb{Z})^\times$. As an example, consider $\mathbb{F}_7$ and the splitting field K for $t^3 - 1$ over $\mathbb{F}_7$. Since $2 \in \mathbb{F}_7$ is a primitive third root of unity (in fact, $\mu_3(\overline{\mathbb{F}_7}) = \{1, 2, 4\} \subset \mathbb{F}_7$), we have $K = \mathbb{F}_7$ and $\mathrm{Gal}(K/\mathbb{F}_7) = \{id_{\mathbb{F}_7}\}$.

21.3.2 The splitting field for $t^n - 1$ over $\mathbb{Q}$

Definition 21.3.6. Let $\zeta_n \in \overline{\mathbb{Q}}$ be a primitive n-th root of unity.

(a) (Cyclotomic polynomial) The minimal polynomial $\Phi_n(t) \in \mathbb{Q}[t]$ of ζ_n is called the n-th cyclotomic polynomial.

(b) (Cyclotomic extension) The splitting field $\mathbb{Q}(\zeta_n)$ for $t^n - 1$ over $\mathbb{Q}$ is often called the n-th cyclotomic field. If $\mathbb{Q} \leq E \leq \mathbb{Q}(\zeta_n)$ for some positive integer n, we call E an cyclotomic extension over $\mathbb{Q}$.

Remark. In the above definition, the n-th cyclotomic polynomial is defined as the minimal polynomial of a primitive n-th root of unity over $\mathbb{Q}$. This definition of $\Phi_n(t)$ seems to depend on the choice of a primitive n-th root of unity.

Goal: To show that $\Phi_n(t)$ is the product of $t - \zeta$, where ζ runs through all primitive n-th roots of unity.

Fix a primitive n-th root ζ of unity and suppose $\Phi_n(t)$ is the minimal polynomial of ζ over $\mathbb{Q}$. If β is a root of $\Phi_n(t)$, then there is a $\mathbb{Q}$-isomorphism from $\mathbb{Q}(\beta)$ into $\mathbb{Q}(\zeta)$ mapping β to ζ. Since $\mathbb{Q}(\zeta)$ is the splitting field for $\Phi_n(t)$ over $\mathbb{Q}$, $\mathbb{Q}(\beta) = \mathbb{Q}(\zeta)$ and β is a primitive n-th root of unity. In other words, every root of $\Phi_n(t)$ is a primitive n-th root of unity. Conversely, if γ is a primitive n-th root of unity, there is a $\mathbb{Q}$-isomorphism from $\mathbb{Q}(\zeta)$ to $\mathbb{Q}(\gamma)$ mapping ζ to γ, so γ is an algebraic conjugate of ζ and is a root of $\Phi_n(t)$. Since $\mathrm{char}(\mathbb{Q}) = 0$, $\Phi_n(t)$ is separable, which completes the proof. To be precise, whenever n is a positive integer,

$$\Phi_n(t) = \prod_{\substack{1 \leq k \leq n \\ (n,k)=1}} \left(t - \exp\left(i\frac{2\pi k}{n} \right) \right).$$

Because $deg\, \Phi_n(t)$ is the number of primitive n-th roots of unity, $deg\, \Phi_n(t)$ is the number of positive integers not greater than n which are relatively prime to n. Therefore, $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \phi(n)$, where ϕ is Euler's ϕ-function.

Example 21.3.7. Let p be a positive prime number. Since $t^p - 1 = (t-1)(t^{p-1} + t^{p-2} + \cdots + t + 1)$ is a separable polynomial with roots $1, \zeta_p, \zeta_p^2, \cdots, \zeta_p^{p-1}$, ζ_p satisfies the polynomial $t^{p-1} + t^{p-2} + \cdots + t + 1$. Moreover, letting $t - 1 = s$ and applying Eisenstein's criterion, we can deduce that $t^{p-1} + t^{p-2} + \cdots + t + 1$ is an irreducible polynomial over $\mathbb{Q}$. Therefore, $\Phi_p(t) = t^{p-1} + t^{p-2} + \cdots + t + 1$ whenever p is a positive prime number.

Observation 21.3.8. By considering the orders of n-th roots of unity, we have

$$t^n - 1 = \prod_{\zeta \in \mu_n(\overline{\mathbb{Q}})} (t - \zeta) = \prod_{0<d|n} \prod_{\mathrm{ord}(\zeta)=d} (t - \zeta) = \prod_{0<d|n} \Phi_d(t).$$

Also, $\phi(n) = \sum \phi(d)$ with d running through all positive divisors of n.

Combining all preceeding observations and applying Gauss's lemma, we obtain the following theorems.

Theorem 21.3.9. For each positive integer n, $\Phi_n(t)$ is an irreducible polynomial over $\mathbb{Z}$ of degree $\phi(n)$.

Proof. It suffices to prove that $\Phi_n(t) \in \mathbb{Z}[t]$ for all $n \in \mathbb{N}$. Assume that $\Phi_n(t)$ is a polynomial over $\mathbb{Z}$ for all positive integers $n < N$. Note that $t^N - 1 = \Phi_N(t) \times \prod_{0<d|N, d\neq N} \Phi_d(t)$ and $\Phi_N(t) \in \mathbb{Q}[t]$. By Gauss's lemma (see Problem 10.2.1) we have $\Phi_N(t) \in \mathbb{Z}[t]$, for $t^N - 1$ and $\prod_{0<d|N, d\neq N} \Phi_d(t)$ are primitive polynomials over $\mathbb{Z}$. □

Theorem 21.3.10. The n-th cylcotomic field $\mathbb{Q}(\zeta_n)$ is, in fact, the splitting field for $\Phi_n(t)$ over $\mathbb{Q}$. The extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is a finite Galois extension of extension degree $\phi(n)$, and its Galois group is isomorphic to $(\mathbb{Z}/n\mathbb{Z})^\times$.

Proof. It suffices to prove that $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \approx (\mathbb{Z}/n\mathbb{Z})^\times$. In fact, any automorphism $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ is determined by its action on ζ_n, and the only possible return for ζ_n is ζ_n^k with $1 \leq k \leq n$ with $(n, k) = 1$. Thus, defining the map $\rho : \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \to (\mathbb{Z}/n\mathbb{Z})^\times$ by $\rho(\sigma) = \overline{k}$, it easily turns out that ρ is a group isomorphism. □

When studying Galois's theorem, we have observed how we could treat the Galois group of a composition field. Using such method, we can establish an isomorphism type of $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ where ζ_n is a primitive n-th root of unity.

Observation 21.3.11. In this example, assume $F = \mathbb{Q}$ and let ζ_k denote a primitive k-th root of unity. Assume further that m, n are relatively prime positive integers.

(a) $\zeta_m\zeta_n$ is a primitive mn-th root of unity. Hence, $\mathbb{Q}(\zeta_m) \cdot \mathbb{Q}(\zeta_n) = \mathbb{Q}(\zeta_{mn})$.

(b) $\mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n) = \mathbb{Q}$.

(c) If d is a positive divisor of n, then ζ_n^d is a primitive (n/d)-th root of unity.

Proof. To prove (a), note from $(m, n) = 1$ that $(\zeta_m\zeta_n)^m = \zeta_n^m$ is a primitive n-th root of unity and that there are integers a, b such that $na + mb = 1$. The former observation implies $\mathbb{Q}(\zeta_m\zeta_n)$ contains ζ_n (and ζ_m for a similar reason), and the latter observation implies $\zeta_m^a\zeta_n^b = \zeta_{mn}$. Therefore, $\mathbb{Q}(\zeta_m\zeta_n)$ contains $\zeta_{mn} = \zeta_m^a\zeta_n^b$, so $\zeta_m\zeta_n$ is a primitive mn-root of unity.

To prove (b), note from (a) that

$$[\mathbb{Q}(\zeta_{mn}) : \mathbb{Q}] = \frac{[\mathbb{Q}(\zeta_m) : \mathbb{Q}][\mathbb{Q}(\zeta_n) : \mathbb{Q}]}{[\mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n) : \mathbb{Q}]}.$$

Since $[\mathbb{Q}(\zeta_{mn}) : \mathbb{Q}] = \phi(mn) = \phi(m)\phi(n) = [\mathbb{Q}(\zeta_m) : \mathbb{Q}][\mathbb{Q}(\zeta_n) : \mathbb{Q}]$, we have $\mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n) = \mathbb{Q}$.

Checking (c) is easy. □

Proposition 21.3.12 (Chinese remainder theorem for cyclotomic fields)**.** Let $n = {p_1}^{a_1} \cdots {p_k}^{a_k}$ be the factorization of a positive integer to prime numbers. (Assume that $p_1, \cdots, p_k$ are pairwise distinct positive prime numbers and $a_1, \cdots, a_k$ are positive integers.)

(a) If s, t are relatively prime positive divisors of n, then $\mathbb{Q}(\zeta_s) \cdot \mathbb{Q}(\zeta_t) = \mathbb{Q}(\zeta_{st})$ and $\mathbb{Q}(\zeta_s) \cap \mathbb{Q}(\zeta_t) = \mathbb{Q}$.

(b) $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \approx \mathrm{Gal}(\mathbb{Q}(\zeta_{{p_1}^{a_1}})/\mathbb{Q}) \times \cdots \times \mathrm{Gal}(\mathbb{Q}(\zeta_{{p_k}^{a_k}})/\mathbb{Q})$.

Proof. (a) follows directly from the preceeding example; it remains to prove (b). Let $s = {p_1}^{a_1}$ and $t = n/s$. Because $\mathbb{Q}(\zeta_n)$ is the composition of $\mathbb{Q}(\zeta_s)$ and $\mathbb{Q}(\zeta_t)$, there is a group monomorphism $\rho : \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \hookrightarrow \mathrm{Gal}(\mathbb{Q}(\zeta_{{p_1}^{a_1}})/\mathbb{Q}) \times \mathrm{Gal}(\mathbb{Q}(\zeta_t)/\mathbb{Q})$; because $\mathbb{Q}(\zeta_s) \cap \mathbb{Q}(\zeta_t) = \mathbb{Q}$, ρ is a group isomorphism. Proceeding the proof inductively, we can obtain a desired isomorphism. □

Observation 21.3.13 (Subfield lattice of $\mathbb{Q}(\zeta_p)$). Let p be a prime number. We will show that every intermediate subfield of $\mathbb{Q}(\zeta_p)/\mathbb{Q}$ has a primitive element over $\mathbb{Q}$ and deliver a formula to find a primitive elemtent.

By Galois's theorem, an intermediate subfield E of $\mathbb{Q}(\zeta_p)/\mathbb{Q}$ and a subgroup H of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$ correspond bijectively. Because p is a prime number, $[\mathbb{Q}(\zeta_p):\mathbb{Q}] = \phi(p) = p-1$, so $\{\zeta_p, \zeta_p^2, \cdots, \zeta_p^{p-1}\}$ is a $\mathbb{Q}$-basis of $\mathbb{Q}(\zeta_p)$. Hence, the element

$$\alpha := \sum_{\sigma \in H} \sigma\alpha$$

is a (finite) sum of basis members. Thus, if $\tau \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$ and $\tau\alpha = \alpha$, then $\tau\zeta_p = \sigma\zeta_p$ for some $\sigma \in H$, hence $\tau = \sigma \in H$. This implies that $\mathbb{Q}(\alpha) \geq \mathbb{Q}(\zeta_p)^H$. Conversely, since $\mathbb{Q}(\alpha)$ is fixed by every automorphism in H, we have $\mathbb{Q}(\alpha) \leq \mathbb{Q}(\zeta_p)^H$. Therefore, $\mathbb{Q}(\zeta_p)^H = \mathbb{Q}(\alpha)$.

In particular, suppose that $E/\mathbb{Q}$ is an intermediate subfield of $\mathbb{Q}(\zeta_p)\mathbb{Q}$ of degree 2 over $\mathbb{Q}$, where p is an odd prime. (Such field E exists uniquely, because there is a unique subgroup $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) \approx Z_{p-1}$ of index 2.) Then

$$E = \begin{cases} \mathbb{Q}(\sqrt{p}) & (\text{if } p \equiv 1 \bmod 4) \\ \mathbb{Q}(\sqrt{-p}) & (\text{if } p \equiv 3 \bmod 4) \end{cases} \tag{21.1}$$

The proof of the above equation is given in Appendix A.2.

Example 21.3.14. We will justify that $\sqrt[3]{2}$ is not contained in any cyclotomic field of $\mathbb{Q}$. If $\sqrt[3]{2} \in \mathbb{Q}(\zeta_n)$ for some $n \in \mathbb{N}$, then the normal closure of $\mathbb{Q}(\sqrt[3]{2})$ over $\mathbb{Q}$ is contained in $\mathbb{Q}(\zeta_n)$. Though, the Galois closure of $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is nonabelian, a contradiction.

21.4 Applications of cyclotomic extensions

21.4.1 Cyclotomic extensions and abelian extensions

In this subsection, we study a matching between finite abelian groups and finite abelian extensions. To be precise,

(I) Given a finite abelian group G, there is a cyclotomic extension E of $\mathbb{Q}$ such that $\mathrm{Gal}(E/\mathbb{Q}) \approx G$.

(II) (Kronecker-Weber's theorem) Every finite abelian extension E over $\mathbb{Q}$ is a cyclotomic extension of $\mathbb{Q}$.

In this note, the proof of Kronecker-Weber's theorem will not be introduced, but only the proof of (I) will be introduced.

Sketch. Write $G \approx Z_{m_1} \times \cdots Z_{m_k}$, where $m_1 | \cdots | m_k$. We wish to find a positive integer n such that $\mathbb{Q} \leq E \leq \mathbb{Q}(\zeta_n)$ such $\mathrm{Gal}(E/\mathbb{Q}) \approx G$. (Note that $E/\mathbb{Q}$ is a Galois extension when such n exists, for $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is an abelian extension.) If $n = {p_1}^{a_1} \cdots {p_k}^{a_k}$ is the factorization of n into pairwise distinct positive prime numbers, then

$$\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \approx (\mathbb{Z}/{p_1}^{a_1}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/{p_k}^{a_k}\mathbb{Z})^\times.$$

If one assumes $a_i = 1$ for all $1 \leq i \leq k$ for easy computation, we have $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \approx Z_{p_1-1} \times \cdots \times Z_{p_k-1}$. Since $\mathrm{Gal}(E/\mathbb{Q}) \approx \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})/\mathrm{Gal}(\mathbb{Q}(\zeta_n)/E)$, if, for each $1 \leq i \leq k$, one can find a prime number p_i such that $m_i | (p_i - 1)$, then the proof proceeds as follows. For each $1 \leq i \leq k$, let $h_i = (q_i - 1)/m_i$ and find a subgroup H_i of Z_{p_i-1} of order h_i. Then there is a subgroup A of $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ such that $A \approx H_1 \times \cdots \times H_k$. If E is the fixed field of A in $\mathbb{Q}(\zeta_n)$, then

$$\mathrm{Gal}(E/\mathbb{Q}) \approx \frac{\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})}{\mathrm{Gal}(\mathbb{Q}(\zeta_n)/E)} \approx \frac{Z_{p_1-1} \times \cdots \times Z_{p_k-1}}{H_1 \times \cdots \times H_k} \approx \prod_{i=1}^{k} \frac{Z_{p_i-1}}{H_i} \approx \prod_{i=1}^{k} Z_{m_i} \approx G.$$

To complete the proof of (I), the following lemma should be proved, whose proof is given in Appendix A.1:

Lemma 21.4.1. Given a positive integer m, there are infinitely many prime numbers modulo m.

21.4.2 Constructibility

We first find the condition of n for which a regular n-gon can be constructed.

Theorem 21.4.2 (Constructibility of a regular n-gon)**.** Suppose that n is an integer greater than or equal to 3. Then a regular n-gon (with the length of a side 1) is constructible if and only if $n = 2^k p_1 \cdots p_l$, where $k, l \geq 0$ and $p_1, \cdots, p_l$ are pairwise distinct Fermat primes.[1]

Proof. Remark that the constructibility of a regular n-gon coincides the constructibility of

$$\gamma := \cos\left(\frac{2\pi}{n}\right) = \frac{\zeta_n + \zeta_n^{-1}}{2},$$

where $\zeta_n = \exp(2\pi i/n)$. Since $t^2 - 2\gamma t + 1$ is satisfied by $\zeta_n \in \mathbb{C} \setminus \mathbb{R}$, $[\mathbb{Q}(\zeta_n) : \mathbb{Q}(\gamma)] = 2$.

Assume first that a regular n-gon is constructible, i.e., γ is constructible. Then $[\mathbb{Q}(\gamma) : \mathbb{Q}]$ is a power of 2, so $[\mathbb{Q}(\zeta_n) : \mathbb{Q}]$ is also a power of 2. This forces n is the product of a power of 2 and pairwise distinct Fermat primes.

Assuming conversely, we find that $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ is a 2-group. Therefore, (writing $|\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})| = 2^m$) there is a subgroup H_r of $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ of order 2^r for each integer $0 \leq r \leq m$ such that

$$\{id_{\mathbb{Q}(\zeta_n)}\} = H_0 < H_1 < \cdots < H_{m-1} < H_m = \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}).$$

By Galois's theorem, we have

$$\mathbb{Q}(\zeta_n) = \mathbb{Q}(\zeta_n)^{H_0} > \mathbb{Q}(\zeta_n)^{H_1} > \cdots > \mathbb{Q}(\zeta_n)^{H_{m-1}} > \mathbb{Q}(\zeta_n)^{H_m} = \mathbb{Q},$$

hence $\gamma \in \mathbb{Q}(\zeta_n)$ is constructible, as desired. □

We end this section with another constructibility criterion. In Theorem 17.5.1, we proved that a real number α is constructible if and only if there is a tower of quadratic extensions from $\mathbb{Q}$ whose head field contains α. In the following theorem in which we consider the normal(Galois) closure of $\mathbb{Q}(\alpha)$, we do not have to consider its subfields but only have to know the extension degree of the Galois closure.

Theorem 21.4.3 (Constructibility criterion II)**.** Let α be a real algebraic number and K be the normal(Galois) closure of $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$. Then α is constructible if and only if $[K : \mathbb{Q}]$ is a power of 2.

Proof. Assume first that α is constructible and let $\alpha_1, \cdots, \alpha_n$ be all algebraic conjugates of α (and let $\alpha_1 = \alpha$). Because

$$\mathbb{Q} \leq \mathbb{Q}(\alpha_1) \leq \cdots \leq \mathbb{Q}(\alpha_1, \cdots, \alpha_n) = K$$

and α_i is constructible for $1 \leq i \leq n$, each subextension degree is a power of 2, as desired.

Assume conversely that $[K : \mathbb{Q}] = 2^r$ for some nonnegative integer r. For each integer $0 \leq i \leq r$, there is a subgroup H_r of $\mathrm{Gal}(K/\mathbb{Q})$ of index 2^i. Then

$$\mathbb{Q} = K^{H_0} < K^{H_1} < \cdots < K^{H_r} = K$$

is a desired tower of quadratic extensions from $\mathbb{Q}$. □

[1] A prime number of the form $2^a + 1$ for some positive integer a is called a Fermat prime.

Chapter 22

Solving polynomial equations

22.1 Symmetric polynomials and discriminants

Remark. When F is a field such that $\mathrm{char}(F) \neq 2$, the roots of the quadratic equation $x^2 + ax + b = 0\,(a, b \in F)$ are given by

$$x = \frac{-a \pm \sqrt{a^2 - 4b}}{2}.$$

We will consider above formula as a function in the coefficients of the polynomial t^2+at+b, i.e., a function of the coefficients of the polynomial.

Observation 22.1.1. Let F be a field and $s_1, \cdots, s_n$ be pairwise distinct indeterminates with $n \in \mathbb{Z}^{>0}$. Then the polynomial

$$G(t) := t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n \in F(s_1, \cdots, s_n)[t]$$

is called the general polynomial of degree n. (Here, the field $F(s_1, \cdots, s_n)$ is transcendental over F.) Writing $E = F(s_1, \cdots, s_n)$ (and fixing an algebraic closure $\overline{E}$ of E) and

$$G(t) = (t - x_1) \cdots (t - x_n) \quad (x_1, \cdots, x_n \in \overline{E}),$$

we can find the formula for s_i in $x_1, \cdots, x_n$ for each integer $i = 1, \cdots, n$. The indeterminates $s_1, \cdots, s_n$ are called the elementary symmetric polynomials in $x_1, \cdots, x_n$. (The extension $E(x_1, \cdots, x_n)/E$ is a finite Galois extension.)

For a clear argument, we start from $x_1, \cdots, x_n$, rather than from the elementary symmetric polynomials in $x_1, \cdots, x_n$.

Definition 22.1.2 (General polynomial)**.** Let F be a field and $x_1, \cdots, x_n$ be pairwise distinct indeterminates $(n \in \mathbb{Z}^{>0})$. Define

$$s_1 = \sum_{1 \leq i \leq n} x_i, \quad s_2 = \sum_{1 \leq i < j \leq n} x_i x_j, \quad \cdots, \quad s_n = x_1 \cdots x_n.$$

Then s_i is called the i-th elementary symmetric polynomial in $x_1, \cdots, x_n$ for each integer $1 \leq i \leq n$. Under the above definition, letting $E = F(s_1, \cdots, s_n)$ (which is transcendental over F), the polynomial

$$G(t) := (t - x_1) \cdots (t - x_n) = t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n \in E[t]$$

is a separable polynomial over E and is called the general polynomial over F of degree n. Hence, the splitting field for $G(t)$ over E is clearly $E(x_1, \cdots, x_n)$, which is a finite Galois extension over F.

Our first goal is to find the Galois group of $G(t)$ over E, which clearly embeds into S_n. Note that S_n acts on $\{x_1, \cdots, x_n\}$ by permutation, i.e., given $\sigma \in S_n$, $\sigma(x_i) = x_{\sigma(i)}$ for all i. This action naturally extends to the group action of S_n by left multiplication on $F[x_1, \cdots, x_n]$ and $F(x_1, \cdots, x_n)$. Hence, the latter group action affords a group embedding $S_n \hookrightarrow \mathrm{Aut}(K)$, where $K = E(x_1, \cdots, x_n)$. (How can S_K be reduced to $\mathrm{Aut}(K)$?) Because the action fixes the elements of E, the above group embedding reduces to $S_n \hookrightarrow \mathrm{Aut}(K/E) = \mathrm{Gal}(K/E)$. Therefore, the Galois group of $G(t)$ over E is, up to isomorphism, S_n.

We summarize the above observation as the following theorem:

Theorem 22.1.3. Let F be a field and $x_1, \cdots, x_n$ be pairwise distinct indeterminates for $n \geq 1$, and let $E = F(s_1, \cdots, s_n)$ and $K = E(x_1, \cdots, x_n)$. Then K is the splitting field for $G(t)$ over E and K/E is a finite Galois extension with the Galois group S_n, up to isomorphism. (Hence, we may identify $\mathrm{Gal}(K/E) = S_n$.)

Remark. (a) Because $G(t)$ is separable and $\mathrm{Gal}(K/E)$ acts on the roots of $G(t)$ transitively, $G(t)$ is an irreducible polynomial over E.

(b) By Galois's theorem, $F(x_1, \cdots, x_n)^{S_n} = F(s_1, \cdots, s_n)$. In fact, $F[x_1, \cdots, x_n]^{S_n} = F[s_1, \cdots, s_n]$. To justify the latter identity (which cannot be direcly deduced from Galois's theorem), it suffices to prove $F[x_1, \cdots, x_n]^{S_n} \subset F[s_1, \cdots, s_n]$: If $u \in F[x_1, \cdots, x_n]^{S_n}$, then $u \in F(s_1, \cdots, s_n) \cap F[x_1, \cdots, x_n]$, so $u \in F[s_1, \cdots, s_n]$.

(c) For any positive integer $n > 1$, A_n is defined set-theoretically to be the collection of all even permutations in S_n. Thus, there is a unique subgroup of $\mathrm{Gal}(K/E)$ of index 2, so there is no confusion to identify such a subgroup with A_n.

Our next goal is to find $F(x_1, \cdots, x_n)^{A_n}$. Define E and K as we have defined in this section. By Artin's theorem (or by Galois's theorem), we have $\mathrm{Gal}(K/K^{A_n}) = A_n$, thus $[K^{A_n} : E] = 2$, i.e., K^{A_n} is a quadratic extension over E.

Definition 22.1.4 (Discriminant)**.** Let $\alpha_1, \cdots, \alpha_n$ be the roots of $f(t) \in F[t]$ (where $n = \deg f(t) \geq 2$). Define

$$\delta = \delta_f = \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j), \quad \Delta = \Delta_f = \delta^2.$$

Both δ and Δ are called the discriminant of $f(t)$.[1]

Observation 22.1.5 (Computation of discriminants). Remark a formula for the determinant of a Vandermonde matrix:

$$\det \begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{pmatrix} = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

Hence, letting $V = (x_i^{j-1})_{\substack{1 \leq i \leq n \\ q \leq j \leq n}}$, $\Delta_G = \det(V^T V)$. Here, $V^T V = (z_{i+j-2})_{i,j}$, where $z_k = z_1^k + \cdots + z_n^k$ for $k \geq 0$.

By computing the determinant for $n = 3$, we have

$$\Delta_G = -4s_2^3 - 27s_3^2 - 4s_1^3 s_3 + s_1^2 s_2^2 + 18 s_1 s_2 s_3.$$

In fact, the above result reduces to a simpler formula when one reduces the t^{n-1}-term of $G(t)$ by shifting $G(t)$; if one obtains $b(t) = t^3 + pt + q$ by shifting $G(t)$, then

$$\Delta_G = \Delta_b = -4p^3 - 27q^3.$$

We postpone to compute the discriminant of the general polynomial of degree 4 later in this note.

Remarking that δ is a square root of δ which is defined up to sign, we investigate when δ is fixed by a permutaion of the roots.

Proposition 22.1.6. Throughout this proposition, $G(t)$ stands for the general polynomial in $F(s_1, \cdots, s_n)$ for $n \geq 2$. Let $f(t)$ be a separable polynomial over F of degree d and identify $G_f \leq S_d$.

(a) $\Delta_f \in F$. In particular, $\Delta_G \in F(s_1, \cdots, s_n)$.

[1] Note that δ is defined up to sign. In this note, we mean Δ when speaking of a discriminant.

(b) $\sigma\delta_f = \delta_f$ whenever $\sigma \in G_f$ is even and $\sigma\delta_f = -\delta_f$ whenever $\sigma \in G_f$ is odd. Hence, if K is the splitting field for $f(t)$ over F and char$(F) \neq 2$, then $K^{G_f \cap A_d} = F(\delta_f)$. In particular, if char$(F) \neq 2$, then $F(x_1, \cdots, x_n)^{A_n} = F(s_1, \cdots, s_n)(\delta_G)$.

Proof. In proving (a), the splitting field K for $f(t)$ over F is a finite Galois extension over F. Because Δ_f is fixed by every automorphism in $\mathrm{Gal}(K/E)$, by Galois's theorem, $\Delta_f \in K^{\mathrm{Gal}(K/F)} = F$.

The first part of (b) easily follows from the definition of δ_f, so $\delta_f \in K^{G_f \cap A_d}$. Since $f(t)$ is assumed to be separable, when char$(F) \neq 2$, we have $\delta_f \neq -\delta_f$. Thus, if $\sigma \in G_f \cap A_d$ fixes δ_f, then σ is necessarily even, so $\mathrm{Gal}(K/F(\delta_f)) \leq G_f \cap A_d$, i.e., $F(\delta_f) \geq K^{G_f \cap A_d}$. □

Corollary 22.1.7. Suppose that F is a field such that char$(F) \neq 2$ and $f(t)$ is a separable polynomial over F (where $n = \deg f(t) \geq 2$). Let K be the splitting field for $f(t)$ over F and identify $G_f \leq S_n$. Then $G_f \leq A_n$ if and only if $\delta_f \in F$.

Proof. $G_f \cap A_n = G_f$ if and only if $F(\delta_f) = K^{G_f \cap A_n} = K^{G_f} = F$. □

This concludes our study on general polynomials and their discriminants (for fields with characteristic not being 2). We end this section with categorizing the Galois group of a cubic polynomial over a field with the characteristic not being 2.

Observation 22.1.8. Let F be a field such that char$(F) \neq 2$ and $f(t) = t^3 + pt^2 + qt + r$ be a separable polynomial over F. (Separability is always ensured when the base field is perfect, and we generally consider perfect fields.) After shifting $f(t)$, redefine $f(t) = t^3 + at + b$. (The roots of the former and the latter $f(t)$ differ by $p/3$, respectively.)

(a) Suppose that $f(t)$ is reducible.

  (i) If $f(t)$ splits completely over F, then all root of $f(t)$ are in F, so $G_f = \{id_F\}$.

  (ii) If $f(t)$ does not split completely over F, then $f(t)$ is a product of a linear factor over F and an irreducible quadratic factor over F. Hence, $G_f \approx Z_2$.

(b) Suppose that $f(t)$ is irreducible. Then $\deg f(t)$ divides the order of the Galois group of $f(t)$. (And since $f(t)$ is assumed to be separable, G_f is transitive on the roots of $f(t)$.) Remark that the discriminant Δ of $f(t)$ is given by $\Delta = -4a^3 - 27b^2$.

  (iii) $\delta \in F$ if and only if $G_f \leq A_3$. In this case, $G_f \approx A_3$.

  (iv) $\delta \notin F$ if and only if $G_f \not\leq A_3$. In this case, $G_f \approx S_3$.

Example 22.1.9. In this example, we compute the Galois group of the polynomial

$$f(t) = t^3 + t + 1 \in F[t]$$

with the base field F varies among $\mathbb{Q}$, $\mathbb{Q}(\sqrt{-31})$, $\mathbb{F}_3$, and $\mathbb{F}_7$. Note that $\Delta = \Delta_f = -31$.

(a) Because $f(t)$ is irreducible over $\mathbb{Q}$ and $\delta = \delta_f = \sqrt{-31} \notin \mathbb{Q}$, $G_{f,\mathbb{Q}} \approx S_3$.

(b) One can verify that $f(t)$ is irreducible over $\mathbb{Q}(\sqrt{-31})$ by checking if $f(t)$ is irreducible over $\mathbb{Z}[\sqrt{-31}]$, whose field of fractions is $\mathbb{Q}(\sqrt{-31})$. Since $\delta = \sqrt{-31} \in \mathbb{Q}(\sqrt{-31})$, we have $G_{f,\mathbb{Q}(\sqrt{-31})} \approx A_3$.

(c) Since $f(t) = (t-1)(t^2 + t + 2)$ and $t^2 + t + 2 \in \mathbb{F}_3[t]$ is irreducible, $G_{f,\mathbb{F}_3} \approx Z_2$.

(d) Since $f(t)$ is irreducible over $\mathbb{F}_7$ and $\Delta = -31 = 4$ is a square in $\mathbb{F}_7$, $G_{f,\mathbb{F}_7} \approx A_3$.

22.2 Cyclic extensions

Starting from this section, we study the splitting field for $t^n - a \in F[t]$ over F under some conditions, due to some technical problems.

Observation 22.2.1. Let F be a field. Then $t^n - a \in F[t]$ is separable whenever $a \neq 0$ and $\mathrm{char}(F)$ does not divide n.

Note that we are mainly interested in (finite) Galois extensions and the condition that $\mathrm{char}(F)$ does not divide n (together with $a \in F \setminus \{0\}$) is the separability condition. Hence, we are interested in the splitting field for $t^n - a \in F[t]$ over F under the following assumption:

Assumption: F is a field and $\mathrm{char}(F)$ does not divide n.

In particular, when $\mathrm{char}(F) = 0$, then $t^n - a$ is separable whenever $n \in \mathbb{N}$ and $a \neq 0$.

Proposition 22.2.2. Let F be a field and n be a positive integer which is not divisible by $\mathrm{char}(F)$, and assume that F contains an n-th root of unity. If $a \in F$ and α is a root of $t^n - a \in F[t]$, then

(a) $F(\alpha)$ is the splitting field for $t^n - a$ over F.

(b) $F(\alpha)/F$ is a cyclic extension and $\mathrm{Gal}(F(\alpha)/F)$ embeds into Z_n.

Proof. The proof of (b) would be sufficient. Without loss of generality, we may assume $a \neq 0$, for $a = 0$ forces $F(\alpha) = F$. Since $t^n - a$ is a separable polynomial over F, $F(\alpha)/F$ is a finite Galois extension, and an automorphism $\sigma \in \mathrm{Gal}(F(\alpha)/F)$ maps α to $\alpha\zeta^k$, where ζ is a primitive n-th root of unity and k is an integer. This induces the map $\rho : \mathrm{Gal}(F(\alpha)/F) \to \mathbb{Z}/n\mathbb{Z}$ defined by $\rho\sigma = \overline{k}$, which is a group monomorphism. □

Corollary 22.2.3. Let F be a field and n be a positive integer which is not divisible by $\mathrm{char}(F)$. If $a \in F$ and $f(t) = t^n - a \in F[t]$, then G_f is a solvable group.

Proof. The roots of $f(t)$ are $\alpha, \alpha\zeta, \cdots, \alpha\zeta^{n-1}$, where α is a root of $f(t)$ and ζ is a primitive n-th root of unity. (Remark that the separability of $f(t)$ is ensured by the assumption on $\mathrm{char}(F)$ and n.) Consider the tower $F \leq F(\zeta) \leq F(\alpha, \zeta)$, and observe that bith $F(\zeta)/F$ and $F(\alpha, \zeta)/F(\zeta)$ are (finite) abelian extensions. It now follows that G_f is a solvable group. □

We now prove the following converse of (b) in Proposition 22.2.2.

Proposition 22.2.4. Suppose that n is a positive integer and F is a field whose characteristic does not divide n. Assume that F contains a primitive n-th root of unity. If K/F is a finite cyclic extension of degree n, then there is an element $\alpha \in K$ such that $\alpha^n \in F$ and $K = F(\alpha)$ (in short, $K = F(\sqrt[n]{a})$ for some $a \in F$).

Proof. See Observation 23.2.2. □

Remark. We summarize the preceeding equivalence as follows:

> Suppose that n is a positive integer and F is a field whose characteristic doen not divide n, and assume further that F contains a primitive n-th root of unity.
>
> (a) Let α be a root of $t^n - a \in F[t]$. Then $F(\alpha)$ is the splitting field for $t^n - a$ over F, and $F(\alpha)/F$ is a cyclic extension of degree dividing n.
>
> (b) Conversely, if K/F is a cyclic extension of degree n, then there is an element $\alpha \in K$ such that $K = F(\alpha)$ and $\alpha^n \in F$. (Hence, K/F is a radical extension.)

22.3 Radical extensions

We introduce another type of finite field extension, called the radical extension, and we define the solvability of a nonconstant polynomial over a field in terms of a radical extension.

Definition 22.3.1. Let F be a field

(a) (Radical extension) Let E/F be a finite field extension with a 'radical tower' given as follows:

$$F \leq (\alpha_1) \leq F(\alpha_1, \alpha_2) \leq \cdots \leq F(\alpha_1, \cdots, \alpha_k) = E,$$

where ${\alpha_1}^{n_1} \in F$ and ${\alpha_i}^{n_i} \in F(\alpha_1, \cdots, \alpha_{i-1})\,(2 \leq i \leq k)$ for some positive integers $n_1, \cdots, n_k$. Then E/F is called an $\{n_i\}$-radical extension, or just a radical extension, in short.

(b) (Solvability of a nonconstant polynomial) Let $f(t)$ be a nonconstant polynomial over F and let K be the splitting field for $f(t)$ over F. If there is a radical extension E/F such that $F \leq K \leq E$, then $f(t)$ is said to be solvable by radicals.

Remark. Indeed, given a nonconstant polynomial $f(t)$ over a field F, all roots of $f(t)$ can be written in terms of elementary operations and radicals if and only if the splitting field for $f(t)$ over F is contained in a radical extension over F (i.e., $f(t)$ is solvable by radicals).

Observation 22.3.2. Let E/F be an $\{n_i\}$-radical extension, and let K be the normal closure of E over F. (Here, K/F need not be a Galois extension.) Then K/F is also an $\{n_i\}$-radical extension.

Proof. Write a radical tower of E/F as

$$F \leq F(\alpha_1) \leq \cdots \leq F(\alpha_1, \cdots, \alpha_n) = E.$$

Remark that K is the composition of σE, where σ runs through $\mathrm{Emb}(E/F)$. For simplicity, write $\mathcal{A} = \{\alpha_1, \cdots, \alpha_n\}$ and $\mathrm{Emb}(E/F) = \{id_E = \sigma_1, \cdots, \sigma_s\}$. Then $K = F(\sigma_1\mathcal{A}, \cdots, \sigma_s\mathcal{A})$, so by adjoining $\sigma_i\alpha_j$ for each i and j one by one, one can establish a radical tower for K/F. □

Assumption: By the end of this chapter, for simplicity, we assume that all fields are of characteristic 0.

Our goal in this section is to prove the following equivalence:

Theorem 22.3.3 (Solvability of a polynomial)**.** Let F be a field (of characteristic 0) and $f(t)$ be a nonconstant polynomial over F. Then $f(t)$ is solvable by radicals if and only if G_f is a solvable group.

Proof of 'if' part. Assume that G_f is a solvable group and let K be the splitting field for $f(t)$ over F. To construct a tower for a field extension with each adjacent extension being cyclic, we make use of a property of finite solvable groups. Let $H_1, \cdots, H_k$ be subgroups of G_f such that

$$G_f = H_0 \triangleright H_1 \triangleright \cdots \triangleright H_{k-1} \triangleright H_k = \{id_K\},$$

where H_{i-1}/H_i is a cyclic group of a prime order p_i for $i = 1, \cdots, k$. By Galois's theorem, we have a tower

$$F = K_0 < K_1 < \cdots < K_{k-1} < K_k = K,$$

where K_i/K_{i-1} is a cyclic extension of degree p_i for $i = 1, \cdots, k$. To let each adjacent extension be radical, we adjoin an appropriate primitive root of unity. Write $n = |G_f|$ and let ζ be a primitive n-th root of unity and consider the following tower:

$$F \leq F(\zeta) = K_0(\zeta) < \cdots < K_{k-1}(\zeta) < K_k(\zeta) = K(\zeta).$$

Because $\mathrm{Gal}(K_i(\zeta)/K_{i-1}(\zeta)) \approx \mathrm{Gal}(K_i/(K_i \cap K_{i-1}(\zeta))) \hookrightarrow \mathrm{Gal}(K_i/K_{i-1})$ and each base field contains an appropriate primitive root of unity, $K_i(\zeta)/K_{i-1}(\zeta)$ is a radical extension $(i = 1, \cdots, k)$. Because $F(\zeta)/F$ is also radical, $K(\zeta)/F$ is a radical extension. Therefore, $f(t)$ is solvable by radicals. □

Proof of 'only if' part. Assume that $f(t)$ is sovable by radicals. Let E be the splitting field for $f(t)$ over F, L be an $\{n_i\}$-radical extension over F containing E, and K be the normal closure of L over F. Then K/F is a finite Galois extension which is $\{n_i\}$-radical. Hence, there are elements $\alpha_1, \cdots, \alpha_k \in K$ such that

$$F \leq F(\alpha_1) \leq \cdots \leq F(\alpha_1, \cdots, \alpha_k) = K$$

where ${\alpha_1}^{n_1} \in F$ and ${\alpha_i}^{n_i} \in F(\alpha_1, \cdots, \alpha_{i-1})\,(2 \leq i \leq k)$.

To let each adjacent extension be cyclic, we adjoin an appropriate primitive root of unity. Let $n = \text{lcm}\{n_1, \cdots, n_k\}$ and ζ be a primitive n-th root of unity. Adjoining ζ, we have

$$F \leq F(\zeta) \leq F(\zeta)(\alpha_1) \leq \cdots \leq F(\zeta)(\alpha_1, \cdots, \alpha_k) = K(\zeta).$$

(i) Note that K/F is a finite Galois extension, so K is the splitting field for some nonconstant polynomial $h(t)$ over F. Then, $K(\zeta)$ is the splitting field for $(t^n - 1)h(t)$ over F, so $K(\zeta)/F$ is also a finite Galois extension.

(ii) Observe that $F(\zeta)/F$ is an abelian extension with the Galois group isomorphic to a subgroup of Z_n and the other adjacent extensions are cyclic extensions.

Hence, the Galois extension $K(\zeta)/F$ is a solvable extension. Because K/F is a Galois extension, K/F is also a solvable extension. □

Corollary 22.3.4. General polynomials of degree at least 5 are insolvable.

Proof. S_n is insolvable if and only if $n \geq 5$. □

Proposition 22.3.5. Let p be a prime number and $f(t) \in \mathbb{Q}[t]$ be an irreducible polynomial of degree p. If $f(t)$ has only two nonreal complex roots, then $G_f \approx S_p$.

Proof. Since $f(t)$ is irreducible, p divides the order of G_f and G_f contains an element of order p. Identifying $G_f \leq S_p$, such an element is a p-cycle. On the other hand, because there are only two nonreal complex roots, G_f contains a transposition. Therefore, $G_f = S_p$, for G_f contains a p-cycle and a transposition. □

Example 22.3.6. The Galois group of $t^5 - 9t + 3 \in \mathbb{Q}[t]$ is S_5.

22.4 Cubic and quartic polynomials

Chapter 23

Further Galois theory

23.1 Character theory

Definition 23.1.1. Let G be a group and L be a field. A linear character χ of G with values in L is a group homomorphism from G into the multiplicative group $L^\times$. We say a collection of characters $\{\chi_1, \cdots, \chi_r\}$ of G with values in L is L-linearly indepenedent whenever there is no nontrivial relation

$$a_1\chi_1 + \cdots + a_r\chi_r = 0 \quad (a_1, \cdots, a_r \in L).$$

Proposition 23.1.2. Let $\{\chi_1, \cdots, \chi_d\}$ be a collection of pairwise distinct characters of a group G with values in a field L. Then $\{\chi_1, \cdots, \chi_d\}$ is L-linearly indepenedent.

Proof. Assume that $\{\chi_1, \cdots, \chi_d\}$ is L-linearly depenedent, and among all linear dependence relataions, we choose one with the minimal number m of nonzero coefficients a_i. (By renumbering, we may write $a_1\chi_1 + \cdots + a_m\chi_m = 0$ with $a_1, \cdots, a_m \in L^\times$.) Let g_0 be an element of G such that $\chi_1(g_0) \neq \chi_d(g_0)$. From

$$\begin{cases} a_1\chi_1(g) + \cdots + a_{d-1}\chi_{d-1}(g) + a_d\chi_d(g) = 0 \\ a_1\chi_1(g_0g) + \cdots + a_{d-1}\chi_{d-1}(g_0g) + a_d\chi_d(g_0g) = 0 \end{cases}$$

we have

$$\sum_{i=1}^{d-1} a_i(\chi_d(g_0) - \chi_i(g_0))\chi_i(g) = 0.$$

Since the first term is nonzero, we obtained another linear dependence with fewer nonzero coefficients. This contradicts the minimality of m, hence $\{\chi_1, \cdots, \chi_d\}$ is L-linearly indepenedent. □

23.2 Lagrange resolvent

In this section, we assume the followings:

(i) F is a field containing a primitive n-th root of unity, where n is not divisible by char(F).

(ii) K/F is a cyclic extension of degree n.

And let σ be a generator of $\mathrm{Gal}(K/F)$.

Definition 23.2.1 (Lagrange resolvent)**.** For $\alpha \in K$ and any n-th root of unity ζ, define the Lagrange resolvent $\mathcal{L}_\sigma(\alpha, \zeta) \in K$ by

$$\mathcal{L}_\sigma(\alpha, \zeta) = \alpha + \zeta\sigma(\alpha) + \zeta^2\sigma^2(\alpha) + \cdots + \zeta^{n-1}\sigma^{n-1}(\alpha).$$

Observation 23.2.2. Let ζ be any n-th root of unity.

(a) By Proposition 23.1.2, $\{id_K, \sigma, \cdots, \sigma^{n-1}\}$ is K-linearly independent. Hence, in particular, there is an element $\alpha \in K$ such that $\mathcal{L}_\sigma(\alpha, \zeta) \neq 0$.

(b) One can easily find that $\sigma^k \mathcal{L} = \zeta^{-k}\mathcal{L}$ for all integer k. Thus, when ζ is a primitive n-th root of unity and α is given as in (a), $id_K \in \mathrm{Gal}(K/F)$ is the unique automorphism fixing $\mathcal{L}$. Hence, $\mathcal{L}$ is contained in K but not in proper subfield of K containing F. This implies that $K = F(\mathcal{L})$.

(c) Furthermore, since $\sigma\mathcal{L} = \zeta^{-1}\mathcal{L}$, we have $\sigma(\mathcal{L}^n) = (\zeta^{-1}\mathcal{L})^n = \mathcal{L}^n$. By Galois's theorem, we have $\mathcal{L}^n \in F$.

To sum up, if ζ is a primitive n-th root of unity and α is an element of K such that $\mathcal{L}_\sigma(\alpha, \zeta) \neq 0$, then $K = F(\mathcal{L})$ and $\mathcal{L}^n$ belongs to F.

23.3 Norm and trace

Definition 23.3.1. Let E/F be a finite separable extension and write $\mathrm{Emb}(E/F) = \{\sigma_1, \cdots, \sigma_n\}$. The norm map $N_{E/F} : E \to F$ and the trace map $tr_{E/F} : E \to F$ is defined by

$$N_{E/F}(\alpha) = \prod_{i=1}^{n} \sigma_i(\alpha) \quad \text{and} \quad tr_{E/F}(\alpha) = \sum_{i=1}^{n} \sigma_i(\alpha) \quad (\alpha \in E).$$

Remark. Let K be the normal closure of E over F and suppose $\tau \in \mathrm{Gal}(K/E)$. Given $\sigma \in \mathrm{Emb}(E/F)$, its extension to K is an F-embedding of K into $\overline{F}$, so $\sigma(E) \leq K$ and $\tau \circ \sigma$ is an F-embedding of E into $\overline{F}$. Therefore, an automorphism in $\mathrm{Gal}(K/F)$ permutes $\mathrm{Emb}(E/F)$ by left multiplication. This proves that $\tau(N_{E/F}(\alpha)) = N_{E/F}(\alpha)$ and $\tau(tr_{E/F}(\alpha)) = tr_{E/F}(\alpha)$, i.e., $N_{E/F}(\alpha), tr_{E/F}(\alpha) \in F$.

Observation 23.3.2. One can easily find the norm and the trace of an element α which is separable over a field F by computing its minimal polynomial over F. To be precise, if α is separable over F, then its minimal polynomial $m(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ over F satisfies

$$m(t) = (t - \sigma_1\alpha) \cdots (t - \sigma_n\alpha),$$

where $n = [F(\alpha) : F]_{\mathsf{sep}} = [F(\alpha) : F]$, thus $N_{F(\alpha)/F}(\alpha) = (-1)^n a_0$ and $tr_{F(\alpha)/F}(\alpha) = -a_{n-1}$.

Proposition 23.3.3. Suppose that K/F is a finite separable extension and $F \leq E \leq K$. Then

$$N_{K/F} = N_{E/F} \circ N_{K/E} \quad \text{and} \quad tr_{K/F} = tr_{E/F} \circ tr_{K/E}.$$

Proof. Let L be the normal(Galois) closure of K over F, and write

$$G = \mathrm{Gal}(L/F), \quad H = \mathrm{Gal}(L/E), \quad I = \mathrm{Gal}(L/K).$$

Then, for $\alpha \in K$,

$$N_{K/E}(\alpha) = \prod_{\sigma \in \mathrm{Emb}(K/E)} \sigma\alpha = \prod_{\sigma I \in H/I} \sigma\alpha, \quad tr_{K/E}(\alpha) = \sum_{\sigma \in \mathrm{Emb}(K/E)} \sigma\alpha = \sum_{\sigma I \in H/I} \sigma\alpha.$$

(It suffices to check that the computation of the last (finite) product(sum) is well-defined. If $\sigma I = \tau I$ for some $\sigma, \tau \in H$, then $\tau^{-1}\sigma \in I$ so $\sigma\alpha = \tau\alpha$ whenever $\alpha \in K$.) Letting $n = N_{K/E}(\alpha)$ and $t = tr_{K/E}(\alpha)$, we have

$$N_{E/F}(n) = \prod_{\tau H \in G/H} \tau n = \prod_{\sigma\tau I \in G/I} (\sigma\tau\alpha), \quad tr_{E/F}(t) = \sum_{\tau H \in G/G} \tau t = \sum_{\sigma\tau I \in G/I} (\sigma\tau\alpha).$$

Such $\sigma\tau$'s form G/I and they are F-embeddings of K into $\overline{F}$, so $N_{K/F} = N_{E/F} \circ N_{K/E}$ and $tr_{K/F} = tr_{E/F} \circ tr_{K/E}$, as desired. □

23.4 Infinite Galois extensions

Krull topology

Appendix A

Proof of some propositions

A.1 Prime numbers congruent to 1 modulo an integer

See Lemma 21.4.1 for the statement.

Proof. □

A.2 Quadratic cyclotomic extension

See eq. (21.1) in page 113 for the statement.

Proof. □