Alireza Salehi Golsefidy

Algebra: ring and field theory

Contents

Chapter 1

Lecture 1

In this lecture, we start with a *pseudo-historical note* on algebra. Next *ring* is defined and some examples are briefly mentioned. *Ring of polynomials* and *direct product of rings* are discussed. Then basic properties of ring operations are discussed. At the end, we define *subrings*, ring *homomorphism*, and ring *isomorphism*

1.1 Introduction: a pseudo-historical note

A large part of algebra has been developed to systematically study zeros of polynomials. The word *algebra* comes from the name of a book by *al-Khwarizmi*, a Persian mathematician, [1] where al-Khwarizmi essentially gave algorithms to find zeros of linear and quadratic equations. Khayyam, another Persian mathematician, made major advances in understanding of zeros of cubic equations. In the 16th century, Italian mathematicians came up with formulas for zeros of general cubic and quartic equations. The cubic case was solved by del Ferro, and Ferrari solved the quartic case.[2] In 1824, Abel proved that there is *no* solutions in radicals to a general polynomial equation of degree at least 5. In 1832, Galois used *symmetries* (group theory) of *system of numbers* of zeros of a polynomial to systematically study them, and he gave the precise condition under which solutions can be written using radicals (and the usual operations $+, -, \cdot, /$).

Another problem which had a great deal of influence on shaping modern algebra is Fermat's last conjecture: there are no *positive integers* x, y, z such that $x^n + y^n = z^n$ if n is an integer more than 2. As you can see this problem has two new directions:

1. it is a *multi-variable* equation,
2. it is a *Diophantine* equation. This means we are looking for *integer* solutions instead of complex or real solutions.

The first direction was important in the development of the *algebraic geometry*, and the second one was played a crucial role in the development of *algebraic number theory*.

[1] I am Persian, and so I have to start with this!

[2] In the book *A History of Algebra; from al-Khwarizmi to Emmy Noether*, by van der Waerden, you can read about the very interesting history of the solution of cubic equations by del Ferro, Tartaglia, and Cardano.

In this course, I often try to put what we learn in the perspective of these *pseudo-historical* remarks.

1.2 Rings: definition and basic examples.

As we mentioned earlier, our hidden agenda is to understand zeros of a polynomial. Say $p(x)$ is a polynomial with rational coefficients. We would like to *understand* properties of a zero $\alpha \in \mathbb{C}$ of $p(x)$. What exactly does understanding mean here? Whatever it means, we would expect to be able to do basic arithmetic with α: *add and multiply*, and find out if we are getting the same values or not. As we see later, this means we want to understand various properties of the *subring* of $\mathbb{C}$ that is *generated* by α.

Definition 1.2.1. 1. *A ring $(R, +, \cdot)$ is a set R with two binary operations: $+$ (addition) and $\cdot$ (multiplication) such that the following holds:*

- *(i) $(R, +)$ is an abelian group.*
- *(ii) (Associative) For every $a, b, c \in R$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.*
- *(iii) (Distributive) For every $a, b, c \in R$,*

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

2. *We say R is a* unital *ring if there is $1 \in R$ such that $1 \cdot a = a \cdot 1$ for every $a \in R$.*

3. *We say R is a* commutative *ring if $a \cdot b = b \cdot a$ for every $a, b \in R$.*

Basic examples.

The set $\mathbb{Z}$ of integers, the set $\mathbb{Q}$ of rational numbers, the set $\mathbb{R}$ of real numbers, and the set $\mathbb{C}$ of complex numbers are unital commutative rings.

Some non-examples.

The set of non-negative integers $\mathbb{Z}^{\geq 0}$ is not a ring as $(\mathbb{Z}^{\geq 0}, +)$ is not an abelian group.

The set of even integers $2\mathbb{Z}$ is a commutative ring, but it is not unital.

For an integer n more than 1, the set $\mathrm{M}_n(\mathbb{R})$ of n-by-n matrices with real entries is a unital ring, but it is not commutative. In fact, for every ring R and positive integer n, the set $\mathrm{M}_n(R)$ of n-by-n matrices with entries in R with the usual matrix addition and multiplication forms a ring. Moreover, if R is unital, then $\mathrm{M}_n(R)$ is also unital.

Ring of integers modulo n.

The set $\mathbb{Z}_n$ of integers modulo n is another important ring. Let us recall that the residue class $[a]_n$ of a modulo n consists all the integers of the form $nk + a$ where k is an integer. In group theory, you have learned that $\mathbb{Z}_n = \{[0]_n, \ldots, [n-1]_n\}$ can be identified with the quotient group $\mathbb{Z}/n\mathbb{Z}$, and the residue class $[a]_n$ of a modulo n

is precisely the coset $a + n\mathbb{Z}$ of the (normal) subgroup $n\mathbb{Z}$. Let us also recall that for every $a, a', b, b' \in \mathbb{Z}$ and positive integer n the following holds:

$$\left.\begin{array}{l} a \equiv a' \pmod{n} \\ b \equiv b' \pmod{n} \end{array}\right\} \Rightarrow aa' \equiv bb' \pmod{n}.$$

This implies that the following is a well-defined binary operator on $\mathbb{Z}_n$:

$$[a]_n \cdot [b]_n := [ab]_n$$

for every a and b in $\mathbb{Z}$. It is easy to check that $(\mathbb{Z}_n, +, \cdot)$ is a unital commutative ring.

Exercise 1.2.2. *Work out the details of why $\mathbb{Z}_n$ is a ring.*

Ring of Polynomials.

As we have mentioned earlier, polynomials play an indispensable role in algebra. Notice that we can and will work with polynomials with coefficients in an arbitrary ring R. The set of all polynomials with coefficients in a ring R and an indeterminant x is denoted by $R[x]$. Therefore

$$R[x] := \{a_n x^n + \cdots + a_0 |\, n \in \mathbb{Z}^{\geq 0}, a_0, \ldots, a_n \in R\}.$$

We sometimes write $\sum_{i=0}^n a_i x^i$ instead of $a_n x^n + \cdots + a_0$. In some arguments it is more convenient to write a polynomial as an infinite sum $\sum_{i=0}^\infty a_i x^i$ with an understanding that $a_{n+1} = a_{n+2} = \cdots = 0$ for some non-negative integer n. Based on our experience of working with polynomials with real or complex coefficients, we define the following operations:

$$\Big(\sum_{i=0}^\infty a_i x^i\Big) + \Big(\sum_{i=0}^\infty b_i x^i\Big) := \sum_{i=0}^\infty (a_i + b_i) x^i \qquad \text{(addition)}$$

$$\Big(\sum_{i=0}^\infty a_i x^i\Big)\Big(\sum_{i=0}^\infty b_i x^i\Big) := \sum_{n=0}^\infty \big(\sum_{i=0}^n a_i b_{n-i}\big) x^n \qquad \text{(multiplication)}$$

for every $\sum_{i=0}^\infty a_i x^i, \sum_{i=0}^\infty b_i x^i \in R[x]$. It is easy to see that $(R[x], +, \cdot)$ is a ring.

Example 1.2.3. *Compute $([2]_4 x + [1]_4)([2]_4 x^2 + [3]_4 x + [1]_4)$ in $\mathbb{Z}_4[x]$.*

Solution. We start the computation as if the coefficients were real numbers and use the distribution law. Moreover to simplify our notation, we drop the decoration $[\ ]_4$, but we remember that computation of coefficients should be done in $\mathbb{Z}_4$. Hence:

$$\begin{aligned} &([2]_4 x + [1]_4)([2]_4 x^2 + [3]_4 x + [1]_4) \\ &\qquad = (2 \cdot 2)x^3 + (2 \cdot 3 + 1 \cdot 2)x^2 + (2 \cdot 1 + 1 \cdot 3)x + (1 \cdot 1) \\ &\qquad = x + 1. \end{aligned}$$

□

Exercise 1.2.4. *1. Compute* $(x+1)^3$ *in* $\mathbb{Z}_3[x]$.

2. *Suppose* p *is prime. Compute* $(x+1)^p$ *in* $\mathbb{Z}_p[x]$.

   (Hint. By the binomial expansion the coefficient of x^i in $(x+1)^p$ is $\binom{p}{i}$. Argue why $\binom{p}{i}$ is zero in $\mathbb{Z}_p$ if $1 \leq i \leq p-1$.)

Warning. Prior to this course, you have viewed a polynomial $f \in R[x]$ as a function from R to R. There is, however, a subtle difference between polynomials and functions. For instance, $x, x^2, \ldots$ are distinct elements of $\mathbb{Z}_2[x]$, but all of them are the same functions from $\mathbb{Z}_2$ to $\mathbb{Z}_2$. Notice that *two polynomials* $\sum_{i=0}^{\infty} a_i x^i$ *and* $\sum_{i=0}^{\infty} b_i x^i$ *are equal if and only if* $a_i = b_i$ *for every non-negative integer* i.

Nevertheless, later we will see that viewing polynomials as functions is extremely useful.

Direct product of rings

Suppose $R_1, \ldots, R_n$ are rings. Then the set

$$R_1 \times \cdots \times R_n := \{(r_1, \ldots, r_n) |\, r_1 \in R_1, \ldots, r_n \in R_n\}$$

with operations

$$(r_1, \ldots, r_n) + (r'_1, \ldots, r'_n) := (r_1 + r'_1, \ldots, r_n + r'_n)$$
$$(r_1, \ldots, r_n) \cdot (r'_1, \ldots, r'_n) := (r_1 \cdot r'_1, \ldots, r_n \cdot r'_n)$$

is a ring, and it is called the *direct product* of R_i's. Notice the operations in the i-th component are done in R_i.

Example 1.2.5. *Compute* $(2,2) \cdot (3,3)$ *in* $\mathbb{Z}_5 \times \mathbb{Z}_6$.

Solution. We notice that $2 \cdot 3 = 1$ in $\mathbb{Z}_5$ and $2 \cdot 3 = 0$ in $\mathbb{Z}_6$. Hence we have $(2,2) \cdot (3,3) = (1,0)$ in $\mathbb{Z}_5 \times \mathbb{Z}_6$. □

1.3 Basic properties of operations in a ring.

Here we see that some basic computations hold in every ring, and a unital ring R has a unique identity, which is sometimes denoted by 1_R.

Lemma 1.3.1. *Suppose* R *is a ring and* 0 *is the neutral element of the abelian group* $(R, +)$. *Then for every* $a, b \in R$, *the following hold:*

1. $0 \cdot a = a \cdot 0 = 0$.
2. $(-a) \cdot b = -(a \cdot b) = a \cdot (-b)$.
3. $(-a) \cdot (-b) = a \cdot b$.

Proof. (1) Since $0 = 0 + 0$, we have $0 \cdot a = (0 + 0) \cdot a$ for every $a \in R$. Hence by the distribution law, we have

$$0 \cdot a = (0 \cdot a) + (0 \cdot a).$$

As $(R, +)$ is a group, we deduce that $0 = 0 \cdot a$. Similarly we have

$$a \cdot 0 = a \cdot (0 + 0) = (a \cdot 0) + (a \cdot 0), \text{ which implies that } 0 = a \cdot 0.$$

(2) To show $(-a) \cdot b = -(a \cdot b)$, we need to argue why $(a \cdot b) + ((-a) \cdot b) = 0$:

$$\begin{aligned}(a \cdot b) + ((-a) \cdot b) =& (a + (-a)) \cdot b && \text{(distribution law)}\\ =& 0 \cdot b \\ =& 0 && \text{(by the first part).}\end{aligned}$$

By a similar argument, we can deduce that $a \cdot (-b) = -(a \cdot b)$.

(3) Using the second part twice, we obtain the last part as follows:

$$(-a) \cdot (-b) = -(a \cdot (-b)) = -(-(a \cdot b)) = a \cdot b.$$

This finishes the proof. □

Lemma 1.3.2. *Suppose R is a unital ring. Then there is a unique element $1_R \in R$ such that*

$$1_R \cdot a = a \cdot 1_R = a \tag{1.1}$$

for every $a \in R$.

Proof. Suppose both 1 and $1'$ satisfy (1.1). Then

$$\begin{aligned}1 =& 1 \cdot 1' && \text{(as } 1' \text{ satisfies (1.1))}\\ =& 1' && \text{(as 1 satisfies (1.1)),}\end{aligned}$$

and the claim follows. □

Exercise 1.3.3. *Suppose $R_1, \ldots, R_n$ are unital rings. Show that $(1_{R_1}, \ldots, 1_{R_n})$ is the identity of $R_1 \times \cdots \times R_n$.*

1.4 Subring and homomorphism.

Whenever you learn a new structure, you should look for subsets that share the same properties (they are often called *sub-*), and more importantly *maps* that preserves those properties (they are often called *homomorphisms*).

Definition 1.4.1. *Suppose $(R, +, 0)$ is a ring. A subset S of R is called a* subring *of R if*

1. *$(S, +)$ is a subgroup of $(R, +)$.*
2. *S is closed under multiplication. This means that for every $a, b \in S$, we have $ab \in S$.*

Warning In your book, having an identity is part of the definition of a ring. As a result a subring of a ring R should contain the identity of R. In our course, we do not make that assumption for subrings.

Example 1.4.2. *$\mathbb{Z}$ is a subring of $\mathbb{Q}$. $\mathbb{Q}$ is a subring of $\mathbb{R}$. $\mathbb{R}$ is a subring of $\mathbb{C}$.*

Exercise 1.4.3.

1. *What is the smallest subring of $\mathbb{C}$ that contains $\mathbb{Q}$ and i?*
2. *What is the smallest subring of $\mathbb{C}$ that contains $\mathbb{Q}$ and $\sqrt{2}$?*
3. *What is the smallest subring of $\mathbb{C}$ that contains $\mathbb{Q}$ and $\sqrt[3]{2}$?*

Definition 1.4.4. *Suppose R_1 and R_2 are two rings. Then a function $f : R_1 \to R_2$ is called a ring* homomorphism *if for every $a, b \in R_1$*

1. $f(a+b) = f(a) + f(b)$,
2. $f(a \cdot b) = f(a) \cdot f(b)$.

Warning As it has been mentioned earlier, in your book, having an identity is part of the definition of a ring. As a result a ring homomorphism between two rings A and B should send 1_A to 1_B. In this course, we refer to the ring homomorphisms that send 1_A to 1_B as *unital ring homomorphisms*.

Example 1.4.5. *For every positive integer n, $c_n : \mathbb{Z} \to \mathbb{Z}_n, c_n(a) := [a]_n$ is a ring homomorphism.*

Chapter 2

Lecture 2

In this lecture, first we show the subring criterion and present important ring homomorphisms. Next we define the kernel and the image of a ring homomorphism. The third topic is on the group of units of a ring, and the definition of a field. As an important example, we find the group of units of the ring of integers modulo n. Finally we define zero-divisors and integral domains.

2.1 More on subrings and ring homomorphisms.

We start by defining a ring isomorphism.

Lemma 2.1.1. *Suppose $f : R_1 \to R_2$ is a bijective ring homomorphism. Then $f^{-1} : R_2 \to R_1$ is a ring homomorphism.*

Proof. Since f is a bijection, it is invertible and there is the function $f^{-1} : R_2 \to R_1$. For every $a, b \in R_2$, we have

$$\begin{aligned} f(f^{-1}(a) + f^{-1}(b)) =& f(f^{-1}(a)) + f(f^{-1}(b)) \\ =& a + b. \end{aligned}$$

Hence $f^{-1}(a + b) = f^{-1}(a) + f^{-1}(b)$. Similarly we have

$$\begin{aligned} f(f^{-1}(a) \cdot f^{-1}(b)) =& f(f^{-1}(a)) \cdot f(f^{-1}(b)) \\ =& a \cdot b. \end{aligned}$$

Hence $f^{-1}(a \cdot b) = f^{-1}(a) \cdot f^{-1}(b)$. The claim follows. □

Definition 2.1.2. *A bijective ring homomorphism is called a ring* isomorphism. *We say two rings are* isomorphic *if there is a ring isomorphism between them.*

As in group theory, two isomorphic rings are essentially the same with different *labelling*!

Let us start with *subgroup criterion* from group theory.

Lemma 2.1.3 (Subgroup criterion)**.** *Suppose $(G, \cdot)$ is a group and H is a non-empty subset. If for every $h, h' \in H$, we have $hh'^{-1} \in H$, then H is a subgroup.*

We can use the subgroup criterion in order to show the *subring criterion*.

Lemma 2.1.4 (Subring criterion)**.** *Suppose R is a ring and S is a non-empty subset of R. If for every $a, b \in S$, we have*

1. *$a - b \in S$, and*
2. *$a \cdot b \in S$,*

then S is a subring.

Proof. By the subgroup criterion, we deduce that $(S, +)$ is a subgroup of $(R, +)$. Since S is also closed under multiplication, we deduce that S is a subring. □

2.2 Kernel and image of a ring homomorphism.

A good application of the subring criterion is to show that the kernel of a ring homomorphism and its image are subrings. Let us recall from group theory that the kernel of a group homomorphism f between two abelian groups A_1 and A_2 is

$$\ker f := \{a \in A_1 | \ f(a_1) = 0\},$$

and $\ker f$ is a subgroup of A_1. We also have that the image of f is

$$\operatorname{Im} f := \{f(a) | \ a \in A_1\},$$

and it is a subgroup of A_2. Since a ring homomorphism f is also an additive group homomorphism, we deduce that $\ker f$ and $\operatorname{Im} f$ are subgroups of the domain of f and the codomain of f, respectively.

Lemma 2.2.1. *Suppose $f : R_1 \to R_2$ is a ring homomorphism. Then the kernel* $\ker f$ *of f is a subring of R_1 and the image* $\operatorname{Im} f$ *of is a subring of R_2. Moreover for every $a \in A$ and $x \in \ker f$, we have that ax and xa are in* $\ker f$.

Remark 2.2.2. *Notice that the* moreover *part of Lemma 2.2.1 is much stronger than saying* $\ker f$ *is closed under under multiplication. Later, when we are studying* ideals *we will come back to this extra property of kernels.*

Proof of Lemma 2.2.1. From group theory, we know that $\ker f$ and $\operatorname{Im} f$ are additive subgroups. It is enough to show that they are closed under multiplication. We show a stronger result for $\ker f$, and we will come back to this property when we define an ideal of a ring. For every $a \in \ker f$ and every $a' \in R_1$, we have

$$f(a \cdot a') = f(a) \cdot f(a') = 0 \cdot f(a') = 0, \text{ and so } a \cdot a' \in \ker f.$$

For every $b, b' \in \operatorname{Im} f$, there are $a, a' \in R_1$ such that $b = f(a)$ and $b' = f(a')$. Therefore

$$b \cdot b' = f(a) \cdot f(a') = f(a \cdot a') \in \operatorname{Im} f.$$

This completes the proof. □

Example 2.2.3. *Find the kernel of* $c_n : \mathbb{Z} \to \mathbb{Z}_n, c_n(a) := [a]_n$.

Solution. You have seen this in group theory: $a \in \ker c_n$ if and only if $c_n(a) = 0$. This means $a \in \ker c_n$ if and only if $[a]_n = [0]_n$. Hence $a \in \ker f$ if and only if a is a multiple of n. Therefore $\ker c_n = n\mathbb{Z}$. □

Example 2.2.4. *Notice that* $c_n : \mathbb{Z}[x] \to \mathbb{Z}_n[x], c_n(\sum_{i=0}^{\infty} a_i x^i) := \sum_{i=0}^{\infty} c_n(a_i)x^i$ *is a ring homomorphism. Find the kernel of* c_n.

Proof. Before we describe the kernel of c_n, let us point out that every ring homomorphism $f : A \to B$ can be extended to a ring homomorphism, which by the abuse of notation is also denoted by f, between $A[x]$ and $B[x]$: $f : A[x] \to B[x]$ such that $f(\sum_{i=0}^{\infty} a_i x^i) := \sum_{i=0}^{\infty} f(a_i)x^i$ (Justify for yourself why this is the case).

Now notice that $\sum_{i=0}^{\infty}$ is in the kernel of c_n if and only if for every i, a_i is in the kernel of c_n. Hence $\ker c_n = n\mathbb{Z}[x]$, which means it consists of polynomials that are multiple of n. □

2.3 A special ring homomorphism

Let's recall a notation from group theory before going back to ring theory. In group theory, you have learned that if $(G, \cdot)$ is a group and $g \in G$, then the cyclic group generated by g is

$$\{g^n |\ n \in \mathbb{Z}\},$$

and

$$e_g : \mathbb{Z} \to G, e_g(n) := g^n \tag{2.1}$$

is a group homomorphism. You have also learned that when we have an *abelian group* A, we often use the *additive* notation. The cyclic (additive) subgroup generated by $a \in A$ is

$$\{na |\ a \in \mathbb{Z}\},$$

where na is defined as follows: for a positive integer n we set

$$na := \underbrace{a + \cdots + a}_{n\text{-times}},$$

for a negative integer n, we set

$$na := \underbrace{(-a) + \cdots + (-a)}_{(-n)\text{-times}},$$

and for $n = 0$, $na = 0$. In the additive setting the group homomorphism e_g which is given in (2.1) is as follows:

$$e_a : \mathbb{Z} \to A, e_a(n) := na. \tag{2.2}$$

Since a ring $(R, +, \cdot)$ with addition $+$ is an abelian group, we can use the same notation as in group theory. This means for $n \in \mathbb{Z}$ and $a \in R$, we can talk about $na \in R$.

Warning. For a ring R, an integer n, and $a \in R$, na should not be confused with a ring multiplication $n \cdot a$. As it is explained above, this concept is borrowed from group theory. Notice that the ring multiplication is only defined for two elements of R, and it is not defined for an integer and an element of R.

Lemma 2.3.1. *Suppose R is a unital ring with the identity element 1_R. Then*

$$e : \mathbb{Z} \to R, \quad e(n) := n1_R$$

is a ring homomorphism.

Proof. From group theory, we know that e is an abelian group homomorphism. So it is enough to show that for every integers m and n we have $e(mn) = e(m) \cdot e(n)$. This is done by a case-by-case consideration, and is not particularly interesting!

Case 1. $m = 0$ or $n = 0$.

Proof of Case 1. By definition, $e(0) = 0$ (the first 0 is in $\mathbb{Z}$ and the second 0 is in R). By basics properties of ring operations (see Lemma 1.3.1), we have that $0 \cdot a = a \cdot 0 = 0$ for every $a \in R$. Therefore for $m = 0$, we have

$$e(mn) = e(0) = 0, \text{ and } e(m) \cdot e(n) = e(0) \cdot e(n) = 0 \cdot e(n) = 0,$$

and similarly for $n = 0$, we have

$$e(mn) = e(0) = 0, \text{ and } e(m) \cdot e(n) = e(m) \cdot e(0) = e(m) \cdot 0 = 0,$$

and the claim follows.

Case 2. $m, n > 0$.

Proof of Case 2. By definition, $e(mn) = 1_R + \cdots + 1_R$ where there are mn-many 1_Rs. On the other hand,

$$\begin{aligned}
e(m) \cdot e(n) =& (\underbrace{1_R + \cdots + 1_R}_{m\text{-times}}) \cdot (\underbrace{1_R + \cdots + 1_R}_{n\text{-times}}) \\
=& \underbrace{1_R \cdot 1_R + \cdots + 1_R \cdot 1_R}_{mn\text{-times}} && \text{(by the distribution law)} \\
=& \underbrace{1_R + \cdots + 1_R}_{mn\text{-times}} \\
=& e(mn).
\end{aligned}$$

This shows the claim in Case 2.

Case 3. $m > 0$ and $n < 0$.

Proof of Case 3. Since m is positive and n is negative, mn is negative. Hence $e(mn) = (-1_R) + \cdots + (-1_R)$ where there are $(-mn)$-many -1_Rs. On the other hand,

$$\begin{aligned}
e(m)\cdot e(n) =& \underbrace{(1_R+\cdots+1_R)}_{m\text{-times}}\cdot(\underbrace{(-1_R)+\cdots+(-1_R)}_{(-n)\text{-times}}) \\
=& \underbrace{1_R\cdot(-1_R)+\cdots+1_R\cdot(-1_R)}_{(-mn)\text{-times}} && \text{(by the distribution law)} \\
=& \underbrace{-(1_R\cdot 1_R)+\cdots+-(1_R\cdot 1_R)}_{(-mn)\text{-times}} && \text{(Lemma 1.3.1)} \\
=& \underbrace{(-1_R)+\cdots+(-1_R)}_{(-mn)\text{-times}} \\
=& e(mn).
\end{aligned}$$

This shows the claim in Case 3.

Case 4. $m < 0$ and $n > 0$.

This case is almost identical to Case 3.

Case 5. $m < 0$ and $n < 0$.

We leave this case as an **exercise**. □

2.4 The evaluation or the substitution map

As it has been already hinted to, polynomials can be viewed as functions. This means we can *evaluate* a polynomial. Next we make it more formal.

Proposition 2.4.1. *Suppose B is a commutative ring and A is a subring of B. Suppose $b \in B$. Then the* evaluation map

$$\phi_b : A[x] \to B, \quad \phi_b(f(x)) := f(b)$$

is a ring homomorphism.

Proof. We need to show that for every $f_1, f_2 \in A[x]$ we have

$$\begin{aligned}
\phi_b(f_1(x)+f_2(x)) =& \phi_b(f_1(x))+\phi_b(f_2(x)) && \text{and} \\
\phi_b(f_1(x)f_2(x)) =& \phi_b(f_1(x))\phi_b(f_2(x)).
\end{aligned}$$

Both are easy to be checked and we leave it as an exercise. □

Let's describe the image and the kernel of ϕ_b.

By the definition of kernel, the kernel of the evaluation map $\phi_b : A[x] \to B$ consists of polynomials that have b as a zero:

$$\ker \phi_b = \{p(x) \in A[x] | \, p(b) = 0\}.$$

This is an indication of how ring theory can help us to study zeros of polynomials.

The image of ϕ_b is

$$\operatorname{Im}\phi_b = \{p(b) |\; p(x) \in A[x]\} = \Big\{ \sum_{i=0}^{n} a_i b^i \mid n \in \mathbb{Z}^+, a_0, \dots, a_n \in A \Big\}.$$

In the next lecture we will show that the image of ϕ_b is the smallest subring of B that contains both A and b.

Chapter 3

Lecture 3

3.1 The evaluation or the substitution map

In the previous lecture we defined the evaluation map

$$\phi_b : A[x] \to B, \quad \phi_b(f(x)) := f(b)$$

where A is a subring of B and $b \in B$. We observed that

$$\ker \phi_b = \{p(x) \in A[x] \mid p(b) = 0\}.$$

Next we describe the image of ϕ_b.

Lemma 3.1.1. *Suppose A is a subring of a unital commutative ring B, and $b \in B$. Then the image of the evaluation map ϕ_b is the smallest subring of B that contains both A and b.*

Proof. Since ϕ_b is a ring homomorphism, its image is a subring. For every $a \in A$, $\phi_b(a) = a$, where a is viewed as the constant polynomial, and $\phi_b(x) = b$. Hence $\operatorname{Im} \phi_b$ is a subring of B which contains A and b.

Suppose C is a subring of B which contains A and b. Then for every $a_0, \ldots, a_n \in A$, we have

$$a_0 + a_1 b + \cdots + a_n b^n \in C$$

as C is closed under addition and multiplication. This implies that $\operatorname{Im} \phi_b$ is a subset of C. The claim follows. $\square$

Definition 3.1.2. *Suppose A is a subring of a unital commutative ring B, and $b \in B$. The smallest subring of B which contains A and b is denoted by $A[b]$.*

Warning. The notation $A[b]$ can be confusing because of its similarity with the ring of polynomials $A[x]$. You have to notice that $b \in B$ is not an indeterminant.

By Lemma 3.1.1, we have that $\operatorname{Im} \phi_b = A[b]$.

Exercise 3.1.3. *Earlier you have seen that the image $\mathbb{Q}[i]$ of $\phi_i : \mathbb{Q}[x] \to \mathbb{C}$ and the image $\mathbb{Q}[\sqrt{2}]$ of $\phi_{\sqrt{2}} : \mathbb{Q}[x] \to \mathbb{C}$ are given only using polynomials of degree at most 1. You have also observed that to get the entire $\mathbb{Q}[\sqrt[3]{2}]$, one can only use polynomials of degree at most 3. What do you think is the general rule?*

3.2 Units and fields

As it has been pointed out earlier, Khwarizmi was interested in solving degree 1 equations. Now we try to do same in a ring: suppose R is a ring and $a, b \in R$. Does the equation $ax = b$ have a solution in R? Over real numbers, such an equation has a solution as long as $a \neq 0$. In fact, if $a \neq 0$, then $x = a^{-1}b$ is the unique solution of $ax = b$. So the question is whether or not a has a multiplicative inverse.

Definition 3.2.1. *Suppose R is a unital ring. We say $a \in R$ is a* unit *if there is $a' \in R$ such that $a \cdot a' = a' \cdot a = 1_R$. The set of all units of R is denoted by $R^\times$.*

Lemma 3.2.2. *Suppose R is a unital commutative ring and $a \in R$ is a unit. Then there is a unique $a' \in R$ such that $a \cdot a' = 1_R$. (We call such an a' the* multiplicative inverse *(or simply the* inverse*) of a. The multiplicative inverse of a is denoted by a^{-1}.)*

Proof. Suppose $a \cdot a' = a \cdot a'' = 1_R$. We have to show that $a' = a''$. We have

$$
\begin{aligned}
a' =& a' \cdot 1_R = a' \cdot (a \cdot a'') \\
=& (a' \cdot a) \cdot a'' && \text{(by the associativity)} \\
=& (a \cdot a') \cdot a'' && \text{(by the commutativity)} \\
=& 1_R \cdot a'' = a''.
\end{aligned}
$$

□

Lemma 3.2.3. *Suppose R is a unital ring. Then $(R^\times, \cdot)$ is a group.*

Proof. We start by showing that $R^\times$ is closed under multiplication. Suppose $a, b \in R^\times$; then

$$(a \cdot b) \cdot (b^{-1} \cdot a^{-1}) = (b^{-1} \cdot a^{-1}) \cdot (a \cdot b) = 1_R. \qquad \text{(justify this!)}$$

Hence $a \cdot b \in R^\times$.

Next we show that $(R^\times, \cdot)$ has an identity. Notice since $1_R \cdot 1_R = 1_R$, $1_R \in R^\times$. As $1_R \cdot a = a \cdot 1_R = a$ for every $a \in R^\times$, we deduce that 1_R is the identity of $R^\times$.

Observe that we have the associativity of $\cdot$ for free as R is a ring.

Finally we show that every element of $R^\times$ has an inverse. Suppose $a \in R^\times$. Then $a \cdot a^{-1} = a^{-1} \cdot a = 1_R$. This implies that $a^{-1} \in R^\times$, which completes the proof. □

Example 3.2.4. $\mathbb{Q}^\times = \mathbb{Q} \setminus \{0\}$, $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$, *and* $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$.

Example 3.2.5. *Find $\mathbb{Z}^\times$.*

Proof. By the definition, $a \in \mathbb{Z}^\times$ if and only if $aa' = 1$ for some $a' \in \mathbb{Z}$. If $aa' = 1$, then $|a||a'| = 1$ and $|a|$ and $|a'|$ are two positive *integers*. Hence $|a|, |a'| \geq 1$ and $|a||a'| = 1$. This implies that $|a| = |a'| = 1$. Therefore $a = \pm 1$. As $(-1)(-1) = 1$ and $(1)(1) = 1$, we deduce that $\mathbb{Z}^\times = \{1, -1\}$. □

Example 3.2.6. *Find 2^{-1} in $\mathbb{Z}_3$.*

Proof. Notice that $[2]_3 \cdot [2]_3 = [1]_3$, and so $2^{-1} = 2$ in $\mathbb{Z}_3$. □

Warning. When we know that we are working with elements of $\mathbb{Z}_n$, we often write a instead of $[a]_n$. When we are asked to find the inverse of an apparently integer number a in $\mathbb{Z}_n$, we should not write $\frac{1}{a}$. We should find an integer a' such that

$$aa' \equiv 1 \pmod{n}.$$

Exercise 3.2.7. *Review your notes from either math 109 or math 100 a where the following property of the greatest common divisor of two integers is discussed. Suppose a and b are two non-zero integers. Then*

the equation $ax + by = c$ has an integer solution if and only if $\gcd(a, b)$ divides c.

This fact can be written in a compact form as $a\mathbb{Z} + b\mathbb{Z} = \gcd(a, b)\mathbb{Z}$. (See proposition 2.3.5 of your book.)

Using the above exercise, we can describe the group $\mathbb{Z}_n^\times$ of units of $\mathbb{Z}_n$.

Proposition 3.2.8. *Suppose n is a positive integer. Then*

$$\mathbb{Z}_n^\times = \{[a]_n | \ \gcd(a, n) = 1\}.$$

Proof. Notice that $[a]_n$ is a unit in $\mathbb{Z}_n$ if and only if for some $[x]_n \in \mathbb{Z}_n$ we have $[a]_n[x]_n = [1]_n$. This means the congruence equation $ax \equiv 1 \pmod{n}$ has a solution. This in turn means for some integers x and y we have $ax - 1 = ny$. So we are looking for as such that the following equation has an integer solution:

$$ax - ny = 1.$$

By the above exercise, this happens exactly when $\gcd(a, n) = 1$. The claim follows. □

Euler's phi function $\phi(n)$ is

$$|\{a \in \mathbb{Z} \mid 1 \leq a \leq n, \gcd(a, n) = 1\}|.$$

Hence by Proposition 3.2.8, we have that

$$|\mathbb{Z}_n^\times| = \phi(n).$$

As a corollary of this equation, we can deduce Euler's theorem.

Theorem 3.2.9 (Euler's theorem)**.** *Suppose n is a positive integer, and $\gcd(a, n) = 1$. Then*

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

Proof. In group theory, you have learned that if $(G, \cdot)$ is a finite group, then for every $g \in G$ we have

$$g^{|G|} = 1.$$

We apply this result for the group $\mathbb{Z}_n^\times$. When $\gcd(a,n)=1$, $[a]_n \in \mathbb{Z}_n^\times$. Therefore by the above discussion we have

$$[a]_n^{|\mathbb{Z}_n^\times|} = [a]_n^{\phi(n)} = [1]_n.$$

Hence

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

□

Definition 3.2.10. *A unital commutative ring F is called a* field *if $F^\times = F \setminus \{0\}$.*

Example 3.2.11. *$\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields, and $\mathbb{Z}$ is not a field.*

Corollary 3.2.12. *Suppose n is a positive integer. Then $\mathbb{Z}_n$ is a field if and only if n is prime.*

Proof. By Proposition 3.2.8, we have that $\mathbb{Z}_n$ is a field if and only if

$$\mathbb{Z}_n \setminus \{[0]_n\} = \{[a]_n | \ \gcd(a,n)=1\}.$$

This means $1 < n$ and every positive integer less than n is coprime with n. The claim follows. □

3.3 Zero-divisors and integral domains

Let's go back to a special case of linear equations: $ax = 0$. We know that over $\mathbb{C}$, 0 is the unique solution of this equation if $a \neq 0$. On the other hand, in $\mathbb{Z}_6$, we have $[2]_6[3]_6 = [0]_6$, which means $2x = 0$ has a non-zero solution in $\mathbb{Z}_6$. This brings us to the following definition.

Definition 3.3.1. *Suppose R is a commutative ring. We say $a \in R$ is a* zero-divisor *if $a \neq 0$ and $ab = 0$ for some non-zero $b \in R$. The set of zero divisors of R is denoted by $D(R)$.*

Definition 3.3.2. *A unital commutative ring D is called an* integral domain *if D has more than one element (alternatively we can say $0_D \neq 1_D$ (why?)) and D has no zero-divisors.*

Example 3.3.3. *$\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are integral domains, and $\mathbb{Z}_6$ is not an integral domain.*

Lemma 3.3.4. *Suppose R is a unital commutative ring. Then $R^\times \cap D(R) = \varnothing$.*

Proof. Suppose to the contrary that $a \in R^\times \cap D(R)$. Then for some $a' \in R \setminus \{0\}$ we have $a \cdot a' = 0$. Then

$$a^{-1} \cdot (a \cdot a') = a^{-1} \cdot 0 = 0.$$

On the other hand, we have

$$a^{-1} \cdot (a \cdot a') = (a^{-1} \cdot a) \cdot a' = 1_R \cdot a' = a'.$$

Hence $a' = 0$, which is a contradiction. □

Corollary 3.3.5. *Every field F is an integral domain.*

Proof. Since F is a field, $1_F \in F^\times = F \setminus \{0_F\}$. Hence $1_F \neq 0_F$. Next we want to show that F has no zero-divisors; that means we want to show $D(F) = \varnothing$. By Lemma 3.3.4, we have that $D(F) \cap F^\times = \varnothing$. Since F is a field, $F^\times = F \setminus \{0\}$. Altogether we deduce that $D(F) = \varnothing$, and the claim follows. □

Notice that the converse of Corollary 3.3.5 is not correct; for instance $\mathbb{Z}$ is an integral domain, but it is not a field. The converse statement, however, holds for finite integral domains. Before proving this result, let's show the *cancellation law* for integral domains.

Lemma 3.3.6 (Cancellation law)**.** *Suppose D is an integral domain. Then for every non-zero $a \in D$ and $b, c \in D$,*

$$ab = ac \quad \textit{implies} \quad b = c.$$

Proof. Since $ab = ac$, we have $a(b - c) = 0$. Since $a \neq 0$ and D does not have a zero-divisor, we deduce that $b - c = 0$, which means $b = c$. This completes the proof. □

Proposition 3.3.7. *Suppose D is a finite integral domain. Then D is a field.*

Proof. Since D is an integral domain, it is a unital commutative ring and $0_D \neq 1_D$. So it is enough to show that every non-zero element $a \in D$ is a unit. This means we have to show that for some $x \in D$ we have $ax = 1$. Let $\ell_a : D \to D, \ell_a(x) := ax$. With this choice of ℓ_a, it is enough to show that 1 is in the image of ℓ_a. We will show that ℓ_a is surjective. Notice that since D is a finite set, $\ell_a : D \to D$ is surjective if and only if it is injective. Therefore it is enough to prove that ℓ_a is injective. Notice that

$$\begin{aligned} \ell_a(b) = \ell_a(c) \Rightarrow & ab = ac \qquad\qquad \text{(By the cancellation law)} \\ \Rightarrow & b = c. \end{aligned}$$

Therefore ℓ_a is injective which finishes the proof. □

3.4 Characteristic of a unital ring

Definition 3.4.1. *Suppose R is a ring. Let*

$$N^+(R) := \{n \in \mathbb{Z}^+ |\, \textit{for every } a \in R, na = 0\}. \tag{3.1}$$

If $N^+(R)$ is empty, we say that the characteristic *of R is* zero*. If $N^+(R)$ is not empty, the* characteristic *of R is the minimum of $N^+(R)$. The characteristic of R is denoted by* $\operatorname{char}(R)$.

Notice that for every ring R we have that $\operatorname{char}(R)a = 0$ for every $a \in R$.

Let us recall that by Lemma 2.3.1 we have that

$$e : \mathbb{Z} \to R, e(n) := n1_R$$

is a ring homomorphism. The next lemma gives us a clear connection between the ring homomorphism e and the characteristic of R.

Lemma 3.4.2. *Let R be a unital ring and $e : \mathbb{Z} \to R, e(n) := n1_R$. For every unital ring R, we have $\ker e = \operatorname{char}(R)\mathbb{Z}$.*

Proof. From group theory, we know that every subgroup of $\mathbb{Z}$ is of the form $m\mathbb{Z}$ for some non-negative integer m. Since $\ker e$ is a subgroup of $\mathbb{Z}$, for some non-negative integer n_0 we have that $\ker e = n_0\mathbb{Z}$.

If $n_0 = 0$, then there is no positive integer n such that $n1_R = 0$. Hence $N^+(R)$ is empty where $N^+(R)$ is as in (3.1). Therefore $\operatorname{char}(R) = 0$. Thus in this case we have $\ker e = \operatorname{char}(R)\mathbb{Z}$.

Now suppose $n_0 \neq 0$. For every $n \in N^+(R)$, we have $n1_R = 0$ which implies that n is in $\ker e = n_0\mathbb{Z}$. Therefore

$$n \geq n_0 \qquad \text{if} \qquad n \in N^+(R). \tag{3.2}$$

On the other hand, for every $a \in R$, we have

$$\begin{aligned} n_0 a =& \underbrace{a + \cdots + a}_{n_0\text{-times}} \\ =& \underbrace{(1_R \cdot a) + \cdots + (1_R \cdot a)}_{n_0\text{-times}} \\ =& (\underbrace{1_R + \cdots + 1_R}_{n_0\text{-times}}) \cdot a = (n_0 1_R) \cdot a \qquad \text{(distribution)} \\ =& 0 \cdot a = 0 \end{aligned} \tag{3.3}$$

By (3.3), we deduce that

$$n_0 \in N(R). \tag{3.4}$$

By (3.2) and (3.4), we deduce that $n_0 = \min N^+(R) = \operatorname{char}(R)$, and the claim follows. ☐

Proposition 3.4.3. *Suppose D is an integral domain. Then $\operatorname{char}(D)$ is either 0 or a prime number.*

Proof. Suppose to the contrary that $\operatorname{char}(D)$ is neither 0 nor prime. Then either $\operatorname{char}(D)$ is either 1 or of the form ab where a and b are two integers more than 1.

If $\operatorname{char}(D) = 1$, then $1_D = 0_D$ which is a contradiction as D is an integral domain.

If $\operatorname{char}(D) = ab$ and a, b are integers more than 1, then by Lemma 3.4.2 we have $\ker e = ab\mathbb{Z}$. Hence $e(ab) = 0$, which implies that

$$e(a) \cdot e(b) = 0. \tag{3.5}$$

As D is an integral domain, by (3.5) we deduce that either $e(a) = 0$ or $e(b) = 0$. Hence either $a \in \ker e$ or $b \in \ker e$. Since $\ker e = ab\mathbb{Z}$ and a and b are integers more than 1, we get a contradiction. ☐

Chapter 4

Lecture 4

4.1 Defining fractions

In the previous lecture, we showed that *every field is an integral domain*, and we noticed that the converse does not hold in general: for instance $\mathbb{Z}$ is an integral domain but it is not a field. Today we will show every integral domain can be embedded into a field. Let's discuss this from the point of view of solving equations. Notice that in a field every linear equation of the form $ax = b$ has a (unique) solution if a is not zero. This property does not hold in an arbitrary integral domain. Let's say we start with an integral domain D and "add" all the zeros of the equations of the form $bx = a$ with $b \neq 0$ to D. What do we get? Let's look at the ring of integers $\mathbb{Z}$. In this case, we get $\{\frac{a}{b} | \, a, \in \mathbb{Z}, b \in \mathbb{Z} \setminus \{0\}\}$, which is the field $\mathbb{Q}$ of rational numbers. We use our understanding of rational numbers as our guide to create fractions for an arbitrary integral integral domain D. Every fraction is of the form $\frac{a}{b}$; so it is given by a pair of elements the *numerator* a and the *denominator* b. The numerator is arbitrary and the denominator is every *non-zero* element. The subtlety is that two different pairs might give us the same fractions. In the field of rational numbers we know that $\frac{a}{b} = \frac{c}{d}$ if and only if $ad = bc$. We use this to *identify* two different pairs together. Formally, we define a relation between the pairs, show that this is an equivalence relation, and use the corresponding equivalence relations to define fractions.

Suppose D is an integral domain. For (a, b) and (c, d) in $D \times (D \setminus \{0\})$, we say $(a, b) \sim (c, d)$ if $ad = bc$. Next we check that $\sim$ is an equivalence relation. Recall that a relation is an equivalence relation if it is *reflexive* (every element is "equal" to itself!), *symmetric* (if x is "equal" to y, then y is "equal" to x), and *transitive* (if x is "equal" to y and y is "equal" to z, then x is "equal" to z). This means we have to check the following:

1. *For every* $(a, b) \in D \times (D \setminus \{0\})$*, we have* $(a, b) \sim (a, b)$. This holds as $ab = ba$.

2. *For every* $(a, b), (c, d) \in D \times (D \setminus \{0\})$*, if* $(a, b) \sim (c, d)$*, then* $(c, d) \sim (a, b)$. This holds as $ad = bc$ implies that $cb = da$.

3. *For every* $(a, b), (c, d), (e, f) \in D \times (D \setminus \{0\})$*, if* $(a, b) \sim (c, d)$ *and* $(c, d) \sim (e, f)$*, then* $(a, b) \sim (e, f)$. The proof of this part is a bit more involved. Since

$(a, b) \sim (c, d)$, we have $ad = bc$, and $(c, d) \sim (e, f)$ implies that $cf = de$. Multiplying both sides of $ad = bc$ by f, and multiplying both sides of $cf = de$ by b, we obtain the following

$$adf = bcf, \text{ and } cfb = deb.$$

Hence $adf = deb$. As $d \neq 0$ and D is an integral domain, by the cancellation law, we have $af = eb$. Therefore

$$(a, b) \sim (e, f).$$

Notice that in the last item, we used the condition that D is an integral domain in a crucial way.

We let $\frac{a}{b}$ be the the *equivalence class* $[(a, b)]$, and let

$$Q(D) := \left\{ \frac{a}{b} \;\middle|\; (a, b) \in D \times (D \setminus \{0\}) \right\}.$$

4.2 Defining addition and multiplication of fractions

Next we will make define two binary operations on $Q(D)$. Again we imitate rational numbers, and we define

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

Whenever we are working with equivalence classes, we have to be extra careful. We need to check whether or not our definitions are independent of the choice of a representative from equivalence classes.

Let's make it more concrete by working with fractions. We are defining addition and multiplication of fractions in terms of their given numerator and denominator. A priori, it is not clear, why we end up getting the same result if we represent the same fractions with different numerators and denominators. That means we have to show that $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ and $\frac{c_1}{d_1} = \frac{c_2}{d_2}$ imply that

$$\frac{a_1 d_1 + b_1 c_1}{b_1 d_1} = \frac{a_2 d_2 + b_2 c_2}{b_2 d_2} \quad \text{and} \quad \frac{a_1 c_1}{b_1 d_1} = \frac{a_2 c_2}{b_2 d_2}.$$

We only discuss why the addition is well-defined. The well-definedness of the multiplication is much easier.

We have that $\frac{a_1 d_1 + b_1 c_1}{b_1 d_1} = \frac{a_2 d_2 + b_2 c_2}{b_2 d_2}$ if and only if

$$(a_1 d_1 + b_1 c_1)(b_2 d_2) = (a_2 d_2 + b_2 c_2)(b_1 d_1) \quad \Leftrightarrow \tag{4.1}$$

$$(a_1 b_2)(d_1 d_2) + (c_1 d_2)(b_1 b_2) = (a_2 b_1)(d_1 d_2) + (c_2 d_1)(b_1 b_2).$$

The second equality in (4.1) holds as we have $a_1 b_2 = a_2 b_1$ and $c_2 d_1 = c_1 d_2$ because of $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ and $\frac{c_1}{d_1} = \frac{c_2}{d_2}$.

4.3 Fractions form a field

I leave it to you to check that $(Q(D), +, \cdot)$ is a ring. Next we show that $Q(D)$ is a field by checking that every non-zero element of $Q(D)$ is a multiplicative inverse. Before showing this, let us show that $\frac{0}{1}$ is the zero of $Q(D)$ and $\frac{1}{1}$ is the identity of $Q(D)$: for every $\frac{a}{b} \in Q(D)$ we have

$$\frac{0}{1} + \frac{a}{b} = \frac{0 \cdot b + 1 \cdot a}{1 \cdot b} = \frac{a}{b}, \quad \text{and} \quad \frac{1}{1} \cdot \frac{a}{b} = \frac{1 \cdot a}{1 \cdot b} = \frac{a}{b}.$$

We also notice that for every non-zero a in D, we have

$$\frac{0}{1} = \frac{0}{a}, \quad \text{and} \quad \frac{1}{1} = \frac{a}{a}.$$

The first one holds as $0 \cdot a = 0 \cdot 1$ and the second one holds as $1 \cdot a = a \cdot 1$.

Suppose $\frac{a}{b}$ is not zero. Then $a \neq 0$. Hence $\frac{b}{a}$ is an element of $Q(D)$. We have that

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{a \cdot b}{b \cdot a} = \frac{1}{1},$$

which means that $\frac{a}{b}$ is a unit in $Q(D)$. Therefore $Q(D)$ is a field.

4.4 The universal property of the field of fractions

In this section, we show that $Q(D)$ is *the smallest* field that contains a copy of D. We have formulate this carefully. First we start by showing that $Q(D)$ has a *copy* of D; this means there is an injective ring homomorphism from D to $Q(D)$. This will be done similar to the way we view integers as fractions with denominator 1.

Lemma 4.4.1. *Suppose D is an integral domain. Let $i : D \to Q(D)$, $i(a) := \frac{a}{1}$. Then i is an injective ring homomorphism.*

Remark 4.4.2. *Suppose A and B are rings. We say A can be* embedded *in B or we say B has a* copy *of A if there is an injective ring homomorphism from A to B.*

Proof of Lemma 4.4.1. We have to show that $i(a) + i(b) = i(a + b)$ and $i(a) \cdot i(b) = i(a \cdot b)$ for every $a, b \in D$:

$$i(a) + i(b) = \frac{a}{1} + \frac{b}{1} = \frac{a \cdot 1 + 1 \cdot b}{1 \cdot 1} = \frac{a + b}{1} = i(a + b),$$

and

$$i(a) \cdot i(b) = \frac{a}{1} \cdot \frac{b}{1} = \frac{a \cdot b}{1 \cdot 1} = i(a \cdot b).$$

Next we show that i is injective:

$$i(a) = i(b) \Rightarrow \frac{a}{1} = \frac{b}{1} \Rightarrow a \cdot 1 = 1 \cdot b \Rightarrow a = b.$$

□

Next we show that if F is a field which contains a copy of D, then F contains a copy of $Q(D)$. In this sense, $Q(D)$ is the smallest field which contains a copy of D.

Theorem 4.4.3. *Suppose D is an integral domain and F is a field. Suppose $f : D \to F$ is an injective ring homomorphism. Then*

$$\widetilde{f} : Q(D) \to F, \quad \widetilde{f}\left(\frac{a}{b}\right) := f(a)f(b)^{-1}$$

is a well-defined injective ring homomorphism. Moreover the following is a commuting diagram

$$\begin{array}{ccc} D & \xrightarrow{\;i\;} & Q(D) \\ & {\scriptstyle f}\searrow & \downarrow{\scriptstyle \widetilde{f}} \\ & & F \end{array}$$

that means we have $\widetilde{f} \circ i = f$.

Proof. We start by showing that $\widetilde{f}$ is well-defined. Suppose $\frac{a_1}{b_1} = \frac{a_2}{b_2}$. Then $a_1b_2 = a_2b_1$ which implies that $f(a_1b_2) = f(a_2b_1)$. Since f is a ring homomorphism, we have

$$f(a_1)f(b_2) = f(a_2)f(b_1). \tag{4.2}$$

As f is injective and b_i's are not zero, we deduce that $f(b_i)$'s are not zero. As F is a field, $f(b_i)$'s are units in F. Therefore by (4.2), we have $f(a_1)f(b_1)^{-1} = f(a_2)f(b_2)^{-1}$. This implies that $\widetilde{f}$ is well-defined.

I leave it to you to check that $\widetilde{f}$ is a ring homomorphism. Next we show that $\widetilde{f}$ is injective. Let us recall an important result from group theory:

A group homomorphism is injective if and only if its kernel is trivial.

Based on the above mentioned result, to show that $\widetilde{f}$ is injective, it is enough to prove that the kernel of $\widetilde{f}$ is trivial:

$$0 = \widetilde{f}\left(\frac{a}{b}\right) = f(a)f(b)^{-1} \quad \Rightarrow \quad f(a) = 0 \quad \Rightarrow a = 0$$

where the last implication holds because f is injective.

Finally we prove that the given diagram is commutative. This means we have to show for every $a \in D$, we have $\widetilde{f}(i(a)) = f(a)$. By the definition of $\widetilde{f}$, we have to show $f(a)f(1)^{-1} = f(a)$. Hence we need to show that $f(1) = 1$. Notice that $f(1) = f(1 \cdot 1) = f(1) \cdot f(1)$. Since f is injective, $f(1) \neq 0$. As F is a field, $f(1)$ is a unit. Therefore $f(1) = f(1) \cdot f(1)$ implies that $f(1) = 1$, which finishes the proof. □

How can we use the *Universal Property of Field of Fractions*?

The universal property can be used to show that $Q(D)$ is isomorphic to a given ring F. We can use the following strategy to show $Q(D) \simeq F$:

1. Prove that F is a field.

2. Find an injective ring homomorphism $f : D \to F$.

3. Use the universal property of field of fractions to get the injective ring homomorphism
$$\widetilde{f} : Q(D) \to F, \quad \widetilde{f}\left(\frac{a}{b}\right) = f(a)f(b)^{-1}.$$

4. Show that every element of F is of the form $f(a)f(b)^{-1}$ for some $a, b \in D$.

The last step implies that $\widetilde{f}$ is surjective. By the third item, we know that $\widetilde{f}$ is injective. Hence $\widetilde{f}$ is a bijective ring homomorphism. This implies that $Q(D) \simeq F$.

In the next lecture, we use this strategy to show that $Q(\mathbb{Z}[i]) \simeq \mathbb{Q}[i]$.

Chapter 5

Lecture 5

5.1 Using the universal property of the field of fractions.

In the previous lecture we defined the field of fractions of an integral domain and proved its universal property. We also discussed a four step strategy of proving that the field of fractions of an integral domain is isomorphic to a given ring.

Example 5.1.1. *Prove that* $Q(\mathbb{Z}[i]) \simeq \mathbb{Q}[i]$.

Solution. **Step 1.** $\mathbb{Q}[i]$ *is a field.*

We have already seen how to show $\mathbb{Q}[i]$ is a subring of $\mathbb{C}$. So to show it is a field, it is enough to prove that every non-zero element of $\mathbb{Q}[i]$ is a unit. Let $a + bi \in \mathbb{Q}[i]$ be a non-zero element. Then we have

$$\frac{1}{a+bi} = \frac{a-bi}{(a+bi)(a-bi)} = \frac{a-bi}{a^2+b^2} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i.$$

Since $a, b \in \mathbb{Q}$, we have $\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2} \in \mathbb{Q}$. Hence $(a+bi)^{-1} \in \mathbb{Q}[i]$. Notice that $a+bi \neq 0, a-bi \neq 0$ and we are allowed to multiply the numerator and the denominator by $a - bi$.

Step 2. $f : \mathbb{Z}[i] \to \mathbb{Q}[i], f(z) := z$.

Then clearly f is an injective ring homomorphism.

Step 3. By the Universal Property of Field of Fractions,

$$\widetilde{f} : Q(\mathbb{Z}[i]) \to \mathbb{Q}[i], \quad \widetilde{f}\left(\frac{z_1}{z_2}\right) = f(z_1)f(z_2)^{-1}$$

is a well-defined injective ring homomorphism.

Step 4. $\widetilde{f}$ is surjective.

Suppose $a + bi \in \mathbb{Q}[i]$. Then by taking a common denominator for a and b we have that there are integers r, s and t such that

$$a + bi = \frac{r+si}{t} = f(r+si)f(t)^{-1}.$$

Therefore $\widetilde{f}$ is surjective.

By Steps 3 and 4, we have that $\widetilde{f}$ is an isomorphism. □

5.2 Ideals

In group theory (and linear algebra), you have seen the importance of kernel of homomorphisms. Next we find out exactly what subsets of a ring A can be the kernel of a ring homomorphism from A to another ring. We have already proved that if $f : A \to B$ is a ring homomorphism, then the kernel of f have the following properties:

1. For every $x, y \in \ker f$, $x - y \in \ker f$, and
2. For every $x \in \ker f$ and $a \in A$, then $ax \in \ker f$ and $xa \in \ker f$.

We will show that these conditions are enough to be the kernel of a ring homomorphism. This brings us to the definition of *ideals*.

It should be pointed out that this is not the historical route to the theory of ideals. The theory of ideals started in order to get the factorization property for more general rings than ring of integers. We will come back to this historical note later when we define prime ideals.

Definition 5.2.1. *Suppose A is a ring, and I is a non-empty subset. We say I is an* ideal *of A if*

1. For every $x, y \in I$, $x - y \in I$, and

2. For every $x \in I$ and $a \in A$, then $ax \in I$ and $xa \in I$.

When I is an ideal of A, we write $I \trianglelefteq A$ or $I \triangleleft A$.

So we have

Lemma 5.2.2. *For every ring homomorphism $f : A \to B$, we have that $\ker f$ is an ideal.*

Next we construct some ideals.

Lemma 5.2.3. *Suppose A is a unital commutative ring, and $x_1, \ldots, x_n \in A$. Then the smallest ideal of A which contains $x_1, \ldots, x_n$ is*

$$\{a_1x_1 + \cdots + a_nx_n \mid a_1, \ldots, a_n \in A\}. \tag{5.1}$$

We denote this ideal by $\langle x_1, \ldots, x_n\rangle$ and we call it the ideal generated by $x_1, \ldots, x_n$.

Proof. We start by showing that the set I given in (5.1) is an ideal and it contains x_i's. Suppose $y, y' \in I$; then

$$y = \sum_{i=1}^{n} a_ix_i \text{ and } y' = \sum_{i=1}^{n} a'_ix_i$$

for some a_i's and a'_i's in A. Hence

$$y - y' = (\sum_{i=1}^{n} a_ix_i) - (\sum_{i=1}^{n} a'_ix_i) = \sum_{i=1}^{n}(a_i - a'_i)x_i \in I.$$

For every $a \in A$, we have

$$ay = a(\sum_{i=1}^{n} a_i x_i) = \sum_{i=1}^{n} (aa_i)x_i \in I.$$

This shows that I is an ideal of A. For every i_0, we have

$$x_{i_0} = 0_A x_1 + \cdots + 0_A x_{i_0-1} + 1_A x_{i_0} + 0_A x_{i_0+1} + \cdots + 0_A x_n \in I,$$

which implies that x_i's are in I.

Next suppose J is an ideal of A which contains x_i's. Then for every $a_i \in A$ we have $a_i x_i \in A$, which in turn implies that

$$a_1 x_1 + \cdots + a_n x_n \in J.$$

Therefore $I \subseteq J$. This finishes the proof. □

We say an ideal I is a *principal ideal* if it is generated by one element. By Lemma 5.2.3, we have that in a unital commutative ring A the principal ideal generated by x is

$$\langle x \rangle = \{ax \mid a \in A\}.$$

We sometimes denote $\langle x \rangle$ by xA.

As in group theory, we will prove the isomorphism theorems. To get to that, we start by defining the quotient ring.

5.3 Quotient rings

Suppose I is an ideal of a ring A. Then for every $x, y \in I$, we have $x - y \in I$. Hence by the subgroup criterion, I is a subgroup of A. As A is abelian, I is a normal subgroup of A. Therefore the set A/I of all the cosets of I form an abelian group under the following operation

$$(x + I) + (y + I) := (x + y) + I.$$

Next we define a multiplication on A/I.

Lemma 5.3.1. *Suppose $I \trianglelefteq A$. The following is a well-defined operation on A/I*

$$(x + I) \cdot (y + I) := xy + I$$

for $x + I, y + I \in A/I$.

Proof. Suppose $x_1 + I = x_2 + I$ and $y_1 + I = y_2 + I$. Then $x_1 - x_2 \in I$ and $y_1 - y_2 \in I$. Here we are using a result from group theory which states that for two cosets $a + H$ and $a' + H$ we have

$$a + H = a' + H \text{ if and only if } a - a' \in H. \tag{5.2}$$

By (5.2), to show $x_1y_1 + I = x_2y_2 + I$ it is necessary and sufficient to show that

$$x_1y_1 - x_2y_2 \in I. \tag{5.3}$$

We show this by adding and subtracting a new term (this method is similar to how we find the formula for the derivative of product of two functions):

$$\begin{aligned} x_1y_1 - x_2y_2 =&(x_1y_1 - x_1y_2) + (x_1y_2 - x_2y_2) \\ =&x_1(y_1 - y_2) + (x_1 - x_2)y_2. \end{aligned} \tag{5.4}$$

Since $y_1 - y_2 \in I$ and $x_1 - x_2 \in I$, we have

$$x_1(y_1 - y_2), (x_1 - x_2)y_2 \in I. \tag{5.5}$$

By (5.4), (5.5), and the fact that I is closed under addition we deduce that $x_1y_1 - x_2y_2 \in I$. Hence $x_1y_1 + I = x_2y_2 + I$ which finishes the proof. □

Notice that Lemma 5.3.1 holds for non-commutative rings as well.

Proposition 5.3.2. *Suppose A is a ring and $I \lhd A$. Then*

1. *$(A/I, +, \cdot)$ is a ring where for every $x + I, y + I \in A/I$ we have*

$$(x + I) + (y + I) := (x + y) + I \quad \textit{and} \quad (x + I) \cdot (y + I) := xy + I.$$

2. *$p_I : A \to A/I, p_I(x) := x + I$ is a surjective ring homomorphism.*
3. *$\ker p_I = I$.*

Remark 5.3.3. *The ring A/I is called a quotient ring of A and p_I is called the natural quotient map.*

Proof of Proposition 5.3.2. Since all the operations are defined in terms of coset representatives, it is straightforward to check all the properties of rings and show that A/I is a ring. I leave this as an exercise.

Let's prove the second item:

$$p_I(x) + p_I(y) = (x + I) + (y + I) = (x + y) + I = p_I(x + y),$$

and

$$p_I(x) \cdot p_I(y) = (x + I) \cdot (y + I) = xy + I = p_I(xy).$$

Every element of A/I is of the form $x + I = p_I(x)$, which means that p_I is surjective.

Finally notice that

$$x \in \ker p_I \;\Leftrightarrow\; p_I(x) = 0 + I \;\Leftrightarrow\; x + I = 0 + I \;\Leftrightarrow\; x \in I,$$

and the claim follows. □

The following is a consequence of Proposition 5.3.2 and Lemma 5.2.2:

Corollary 5.3.4. *Suppose A is a ring and I is a subset of A. Then I is the kernel of a ring homomorphism from A to another ring if and only if I is an ideal.*

5.4 The first isomorphism theorem for rings

In this section, we prove the first isomorphism theorem for rings. Let's recall the group theoretic version of this theorem:

Theorem 5.4.1 (The 1st Isomorphism Theorem for Groups)**.** *Suppose $f : G \to G'$ is a group homomorphism. Then*

$$\overline{f} : G/\ker f \to \operatorname{Im} f, \quad \overline{f}(g \ker f) := f(g)$$

is a well-defined group isomorphism.

We use Theorem 5.4.1 to show the following:

Theorem 5.4.2. *Suppose $f : A \to A'$ is a ring homomorphism. Then*

$$\overline{f} : A/\ker f \to \operatorname{Im} f, \quad \overline{f}(a + \ker f) := f(a)$$

is a ring isomorphism.

Proof. Since f is an additive group homomorphism, by the first isomorphism theorem for groups we have that $\widetilde{f}$ is a well-defined group isomorphism. To finish the proof, it is enough to show that $\overline{f}$ preserves the multiplication:

$$\overline{f}(xy + \ker f) = f(xy) = f(x)f(y) = \overline{f}(x + \ker f)\overline{f}(y + \ker f),$$

for every $x, y \in A$. This finishes the proof. □

Example 5.4.3. *Suppose n is a positive integer. Then $\mathbb{Z}/n\mathbb{Z} \simeq \mathbb{Z}_n$.*

Proof. Let $c_n : \mathbb{Z} \to \mathbb{Z}_n$ be the residue map $c_n(x) := [x]_n$. Then c_n is surjective and

$$x \in \ker c_n \;\Leftrightarrow\; [x]_n = [0]_n \;\Leftrightarrow\; n|x \;\Leftrightarrow\; x \in n\mathbb{Z}.$$

By the first isomorphism theorem for rings, we have that

$$\overline{c}_n : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}_n, \quad \overline{c}_n(x + n\mathbb{Z}) = c_n(x)$$

is a ring isomorphism. □

A general strategy of using the first isomorphism theorem to show that a quotient ring A/I is isomorphic to a ring B is to start with a ring homomorphism $f : A \to C$ where B is a subring of C, and show that $\operatorname{Im} f = B$ and $\ker f = I$. This is what we did in the previous example and what we will do in the next example as well.

Example 5.4.4. *We have*

$$\mathbb{Q}[x]/\langle x^2 - 2\rangle \simeq \mathbb{Q}[\sqrt{2}],$$

and

$$\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}.$$

Proof. Let $\phi_{\sqrt{2}} : \mathbb{Q}[x] \to \mathbb{C}$ be the evaluation map $\phi_{\sqrt{2}}(f(x)) = f(\sqrt{2})$. Then by the first theorem for rings we have

$$\mathbb{Q}[x]/\ker\phi_{\sqrt{2}} \simeq \operatorname{Im}\phi_{\sqrt{2}}.$$

Recall that we have defined $\mathbb{Q}[\sqrt{2}]$ to be the image $\operatorname{Im}\phi_{\sqrt{2}}$ of $\phi_{\sqrt{2}}$.

Next we find the kernel $\ker\phi_{\sqrt{2}}$. Notice that $\sqrt{2}$ is a zero of $x^2 - 2$, and so $x^2 - 2$ is in $\ker\phi_{\sqrt{2}}$. Suppose $f(x) \in \ker\phi_{\sqrt{2}}$. By the long division, there are $q(x), r(x) \in \mathbb{Q}[x]$ such that

1. $f(x) = q(x)(x^2 - 2) + r(x)$, and
2. $\deg r < \deg(x^2 - 2)$.

Since $\deg r < 2$, there are $a, b \in \mathbb{Q}$ such that $r(x) = ax + b$. As $f(\sqrt{2}) = 0$, we deduce that

$$0 = f(\sqrt{2}) = q(\sqrt{2})\underbrace{((\sqrt{2})^2 - 2)}_{\text{is } 0} + (a\sqrt{2} + b).$$

Hence $a\sqrt{2} + b = 0$. If $a \neq 0$, then $\sqrt{2} = -b/a \in \mathbb{Q}$ which is a contradiction as $\sqrt{2}$ is irrational. Thus $a = 0$, which in turn implies that $b = 0$. This means $r(x) = 0$, and so $f(x) = q(x)(x^2 - 2) \in I$. Therefore $\ker\phi_{\sqrt{2}} = I$.

(We will continue in the next lecture.) □

Chapter 6

Lecture 6

6.1 An application of the first isomorphism theorem.

In the previous lecture, we were in the middle of the proof of the following result. We will be generalizing this result later in the course. We will be using similar techniques to describe the structure of $\mathbb{Q}[\alpha]$ where α is a zero of a polynomial.

Example 6.1.1. *We have*

$$\mathbb{Q}[x]/\langle x^2 - 2\rangle \simeq \mathbb{Q}[\sqrt{2}],$$

and

$$\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}.$$

Proof. We have already considered the evaluation map $\phi_{\sqrt{2}}$, used the first isomorphism theorem to show that

$$\mathbb{Q}[x]/\ker\phi_{\sqrt{2}} \simeq \operatorname{Im}\phi_{\sqrt{2}}.$$

Next we used the long division and proved that $\ker\phi_{\sqrt{2}} = \langle x^2 - 2\rangle$.

Next we want to show that $\mathbb{Q}[\sqrt{2}] = \{a_0 + a_1\sqrt{2} \mid a_0, a_1 \in \mathbb{Q}\}$. To show this we again use the long division.

Elements of $\mathbb{Q}[\sqrt{2}]$ are of the form $p(\sqrt{2})$ for some $p(x) \in \mathbb{Q}[x]$. By the long division, there are $q(x), r(x) \in \mathbb{Q}[x]$ such that

1. $p(x) = q(x)(x^2 - 2) + r(x)$, and
2. $\deg r < \deg(x^2 - 2)$.

Hence there are $a_0, a_1 \in \mathbb{Q}$ such that $r(x) = a_0 + a_1 x$. Therefore

$$p(\sqrt{2}) = q(\sqrt{2})(\sqrt{2}^2 - 2) + (a_0 + a_1\sqrt{2}) = a_0 + a_1\sqrt{2}.$$

This implies that $\mathbb{Q}[\sqrt{2}] = \{a_0 + a_1\sqrt{2} \mid a_0, a_1 \in \mathbb{Q}\}$, and the claim follows. □

As you can see in this examples, the long division plays an important role in understanding of polynomials. Next we want to see in what generality the long division holds.

6.2 Degree of polynomials

Suppose A is a unital commutative ring and

$$f(x) = a_0 + a_1x + \cdots + a_nx^n \in A[x] \quad \text{and} \quad a_n \neq 0.$$

Then we say a_nx^n is the *leading term* of f, and we write $\operatorname{Ld}(f) := a_nx^n$. The leading term contains two information: the *leading coefficient* a_n and the exponent n of x which is called the *degree* of f, and we write $\deg f = n$. We use the following convention for the zero polynomial:

$$\deg 0 = -\infty, \quad \text{and} \quad \operatorname{Ld}(0) := 0.$$

Example 6.2.1. *Find* $\deg((2x+1)(3x^2+1))$ *in* $\mathbb{Z}_6[x]$.

Solution. By the distribution property we have

$$(2x+1)(3x^2+1) = \underbrace{(2\cdot 3)}_{0 \text{ in } \mathbb{Z}_6} x^3 + 3x^2 + 2x + 1 = 3x^2 + 2x + 1.$$

Hence $\deg((2x+1)(3x^2+1)) = 2$. □

Notice that in the above example, $\deg(2x+1) = 1$ and $\deg(3x^2+1) = 2$. Hence sometimes,

$$\deg f \cdot g \neq \deg f + \deg g.$$

A closer examination of the above example reveals that existence of zero-divisors is responsible for the failure of the degree of the product formula. In fact, if *at least one of the leading coefficients of f or g is not a zero-divisor, then we have*

$$\deg f \cdot g = \deg f + \deg g.$$

Let's see the details.

Lemma 6.2.2. *Suppose A is a unital commutative ring, and $f(x), g(x) \in A[x]$.*

1. *Suppose the leading coefficient of f is a and the leading coefficient of g is b. If $ab \neq 0$, then $\operatorname{Ld}(fg) = \operatorname{Ld}(f)\operatorname{Ld}(g)$ and $\deg fg = \deg f + \deg g$.*

2. *Suppose that the leading coefficient of f is not a zero-divisor. Then*

$$\operatorname{Ld}(fg) = \operatorname{Ld}(f)\operatorname{Ld}(g) \quad \textit{and} \quad \deg fg = \deg f + \deg g; \tag{6.1}$$

   in particular, if D is an integral domain, then (6.1) holds.

Proof. (1) Suppose

$$f(x) = a_0 + a_1x + \cdots + a_nx^n,\ g(x) = b_0 + b_1x + \cdots + b_mx^m,$$

$a_n = a$, and $b_m = b$. Then

$$f(x)g(x) = a_nb_mx^{n+m} + \text{ terms of degree less than } m+n.$$

Hence if $a_n b_m$ is not zero, then $\mathrm{Ld}(fg) = a_n b_m x^{n+m}$. Notice that by the assumption we have $a_n b_m = ab \neq 0$. Therefore the claim follows as $\mathrm{Ld}(f) = ax^n$ and $\mathrm{Ld}(g) = bx^m$.

(2) Suppose g is not zero and its leading coefficient is b. Since the leading coefficient a of f is not a zero divisor, $ab \neq 0$. Therefore by part (1), the claim follows. If $g = 0$, then $fg = 0$. Hence $\deg fg = \deg g = -\infty$. As we are using the convention that $-\infty + n = -\infty$ for every $n \in \mathbb{Z}$, the claim follows in this case as well.

When D is an integral domain, the leading coefficient of a non-zero $f(x)$ is not a zero-divisor. Hence we get the claim. If $f = 0$, then $fg = 0$. Thus $\mathrm{Ld}(fg) = 0 = \mathrm{Ld}(f)\,\mathrm{Ld}(g)$ and $\deg fg = -\infty = -\infty + \deg g = \deg f + \deg g$, which finishes the proof. □

6.3 Zero-divisors and units of ring of polynomials

In this section, we use Lemma 6.2.2 to study the ring of polynomials of integral domains.

Lemma 6.3.1. *Suppose D is an integral domain. Then $D[x]$ is an integral domain.*

Proof. Since D is an integral domain, it is a unital commutative ring. Therefore $D[x]$ is a unital commutative ring. Since D is an integral domain, it is a non-trivial ring. As $D[x]$ has a copy of D (constant polynomials), $D[x]$ is a non-trivial ring. So it remains to show that $D[x]$ does not have a zero-divisor. Suppose $f(x)g(x) = 0$ for some $f, g \in D[x]$. Then $\deg fg = -\infty$, and so by Lemma 6.2.2 we have

$$-\infty = \deg f + \deg g.$$

Therefore not both of $\deg f$ and $\deg g$ can be integers, and at least one of them is $-\infty$. This means either $f = 0$ or $g = 0$. This means $D[x]$ does not have a zero-divisors. □

Lemma 6.3.2. *Suppose D is an integral domain. Then*

$$D[x]^\times = D^\times.$$

Proof. Suppose $u \in D^\times$. Therefore $u^{-1} \in D$ exists. Since $D[x]$ has a copy of D as the set of constant polynomials, we deduce that $u^{-1} \in D[x]$ (notice that $D[x]$ and D have the same identity). Hence $u \in D^\times$. This means $D^\times \subseteq D[x]^\times$.

Let's go to the more interesting part where the assumption that D is an integral domain is actually needed.

Suppose $f(x) \in D[x]^\times$. This means there is $g(x) \in D[x]$ such that $f(x)g(x) = 1$. By Lemma 6.2.2, we have that

$$\deg f + \deg g = \deg fg = \deg 1 = 0.$$

This, in particular, implies that f and g are not zero, and so their degrees are at least 0. Therefore $\deg f$ and $\deg g$ are two non-negative integers that add up to 0. Hence both of them are zeros. That means $f(x) = a \in D$, $g(x) = b \in D$, and $f(x)g(x) = ab$ is 1. This implies that $f(x) = a \in D^\times$, which finishes the proof. □

6.4 Long division

In this section, we will show the most general form of the long division for polynomials. Let's start with a quick overview of the long division for polynomials. Say we want to divide

$$f(x) = a_n x^n + \cdots + a_1 x + a_0$$

by

$$g(x) = b_m x^m + \cdots + b_1 x + b_0.$$

In the long division algorithm, first we look at the degrees. If $\deg f = n$ is smaller than $\deg g = m$, then we are done! In this case, the quotient is 0 and the remainder is $f(x)$. If $\deg f \geq \deg g$, then we look for a monomial cx^k to multiply by $\mathrm{Ld}(g)$ and end up getting $\mathrm{Ld}(f)$; that means $(cx^k)(b_m x^m) = a_n x^n$:

$$\begin{array}{r l} & \quad\; cx^k \\ \cline{2-2} b_m x^m + \cdots + b_0 & \big)\; a_n x^n + \cdots + a_0 \end{array}$$

This means that $k + m = n$ and $b_m a = a_n$. Since we assumed $n \geq m$, $n - m \geq 0$, and we can let $k := n - m$. The equation $b_m c = a_n$, however, does not necessarily have a solution in A. This equation has a solution in A if b_m is a unit. In this case, we see that the desired monomial is $(b_m^{-1} a_n) x^{n-m}$. After finding this monomial, we subtract $(b_m^{-1} a_n x^{n-m}) g(x)$ from $f(x)$, get a smaller degree polynomial and *continue this process*. This leads us to the following theorem.

Theorem 6.4.1 (Long Division For Polynomials)**.** *Suppose A is a unital commutative ring, $f(x), g(x) \in A[x]$ and the leading coefficient of $g(x)$ is a unit in A. Then there are unique $q(x) \in A[x]$ (quotient) and $r(x) \in A[x]$ (remainder) that satisfy the following properties:*

$$f(x) = g(x)q(x) + r(x) \quad \textit{and} \quad \deg r < \deg g. \tag{6.2}$$

(Whenever you see the phrase *and we continue this process*, it means that there is an *induction argument* in the formal proof.)

Proof. (The existence part) We proceed by the strong induction on $\deg f$. If $\deg f < \deg g$, then $q(x) = 0$ and $r(x) = f(x)$ satisfy (6.2). So we prove the strong induction step under the extra condition that $\deg f \geq \deg g$. Suppose $f(x) = \sum_{i=0}^{n} a_i x^i$, $g(x) = \sum_{i=0}^{m} b_i x^i$, $a_n \neq 0$, and $b_m \neq 0$. Then by the assumption b_m is a unit in A. Let

$$\overline{f}(x) := f(x) - (b_m^{-1} a_n) x^{n-m} g(x). \tag{6.3}$$

Then one can see that $\deg \overline{f} < \deg f$. Hence by the strong induction hypothesis, we can divide $\overline{f}$ by g and get a quotient $\overline{q}$ and a remainder r; this means we have

$$\overline{f}(x) = \overline{q}(x) g(x) + r(x) \quad \text{and} \quad \deg r < \deg g. \tag{6.4}$$

By (6.4) and (6.3), we obtain

$$f(x) = ((b_m^{-1} a_n) x^{n-m} + \overline{q}(x)) g(x) + r(x) \quad \text{and} \quad \deg r < \deg g.$$

Hence $q(x) := (b_m^{-1}a_n)x^{n-m} + \overline{q}(x)$ and $r(x)$ satisfy (6.2). This completes the proof of the existence part.

(The uniqueness part) Suppose q_1, r_1 and q_2, r_2 both satisfy (6.2). We have to prove that $q_1 = q_2$ and $r_1 = r_2$. As q_i, r_i satisfy (6.2). This means

$$f(x) = q_1(x)g(x) + r_1(x) = q_2(x)g(x) + r_2(x),$$

$$\deg r_1 < \deg g, \text{ and } \deg r_2 < \deg g.$$

Hence we have

$$(q_1(x) - q_2(x))g(x) = r_2(x) - r_1(x) \quad \text{and} \quad \deg(r_2 - r_1) < \deg g. \tag{6.5}$$

Since the leading coefficient of g is a unit, it is not a zero-divisor (see Lemma 3.3.4). Therefore by Lemma 6.2.2 and (6.5), we have

$$\deg(r_1 - r_2) = \deg((q_1 - q_2)g) = \deg(q_1 - q_2) + \deg g < \deg g.$$

Hence $\deg(q_1 - q_2) < 0$, which implies that $q_1 - q_2 = 0$. Thus by (6.5), we deduce that $r_1 = r_2$. Overall we showed that $q_1 = q_2$ and $r_1 = r_2$, which finishes the proof the uniqueness. □

Chapter 7

Lecture 7

7.1 The factor theorem and the generalized factor theorems

In the previous lecture we proved a general form of the long division for polynomials. We proved that if A is a unital commutative ring, we can divide $f(x)$ by $g(x)$ for $f, g \in A[x]$ and a quotient and a remainder if the leading coefficient of g is a unit in A. In particular, if A is a field, then the leading coefficient of every non-zero polynomial is a unit. Hence we can divide every polynomial by every non-zero polynomial.

The Factor Theorem is an important application of the long division for polynomials.

Theorem 7.1.1. *Suppose A is a unital commutative ring and $f(x) \in A[x]$. Then*

1. *for every $a \in A$, there is a unique $q(x) \in A[x]$ such that*

$$f(x) = (x - a)q(x) + f(a).$$

2. (The Factor Theorem) *We have that a is a zero of $f(x)$ if and only if there is $q(x) \in A[x]$ such that*

$$f(x) = (x - a)q(x).$$

Proof. (1) By the long division for polynomials, there are unique $q(x)$ and $r(x)$ with the following properties:

$$f(x) = (x - a)q(x) + r(x) \quad \text{and} \quad \deg r < \deg(x - a).$$

The second property implies that $r(x)$ is a constant, say $r(x) = c \in A$. Then we have $f(x) = (x - a)q(x) + c$. Evaluating both sides at $x = a$, we deduce that $c = f(a)$. Altogether, we obtain that $f(x) = (x - a)q(x) + f(a)$, which finishes the proof of the first part.

(2) Suppose a is a zero of f; then $f(a) = 0$. Therefore by part (1), we have that $f(x) = (x - a)q(x)$ for some $q(x) \in A[x]$.

To show the converse, we can evaluate both sides of $f(x) = (x - a)q(x)$ at $x = a$, and deduce that $f(a) = 0$. This finishes the proof. □

The factor theorem can be interpreted in terms of the evaluation map: for every $a \in A$ we have

$$\ker \phi_a = \langle x - a \rangle,$$

where $\phi_a : A[x] \to A, \phi_a(f(x)) := f(a)$.

Theorem 7.1.2. *Suppose D is an integral domain, $f(x) \in D[x]$, and $a_1, \dots, a_n$ are distinct elements of D. Then $a_1, \dots, a_n$ are zeros of $f(x)$ if and only if there is $q(x) \in D[x]$ such that*

$$f(x) = (x - a_1) \cdots (x - a_n) q(x).$$

Proof. We proceed by the induction on n. The base of induction $n = 1$ follows from the Factor Theorem. So we focus on the induction step. Suppose $a_1, \dots, a_{n+1}$ are distinct zeros of $f(x)$. Then by the induction hypothesis, there is $\overline{q}(x) \in D[x]$ such that

$$f(x) = (x - a_1) \cdots (x - a_n) \overline{q}(x). \tag{7.1}$$

Since a_{n+1} is a zero of $f(x)$, by (7.1) we deduce that

$$0 = (a_{n+1} - a_1) \cdots (a_{n+1} - a_n) \overline{q}(a_{n+1}). \tag{7.2}$$

Since a_j's are distinct, $a_{n+1} - a_i$'s are not zero. As D is an integral domain, it has no zero-divisor. Therefore by (7.2), we obtain that

$$\overline{q}(a_{n+1}) = 0.$$

Hence by the Factor Theorem, there is $q(x) \in D[x]$ such that

$$\overline{q}(x) = (x - a_{n+1}) q(x). \tag{7.3}$$

By (7.2) and (7.3), we obtain that

$$f(x) = (x - a_1) \cdots (x - a_n)(x - a_{n+1}) q(x).$$

This finishes the claim. □

Remark 7.1.3. *The Factor Theorem holds for every unital commutative ring, but the Generalized Factor Theorem is true only for integral domains.*

Exercise 7.1.4. *Give an example where the Generalized Factor Theorem fails.*

Corollary 7.1.5. *Suppose D is an integral domain and $f(x) \in D[x] \setminus \{0\}$. Then f does not have more than $\deg f$ distinct zeros in D.*

Proof. Suppose $a_1, \dots, a_m$ are distinct zeros of $f(x)$. Then by the generalized factor theorem there is $q(x) \in D[x]$ such that

$$f(x) = (x - a_1) \cdots (x - a_m) q(x). \tag{7.4}$$

Comparing the degrees of both sides of (7.4), we get

$$\deg f = m + \deg q.$$

Notice that since f is not zero, neither is q. Thus $\deg q \geq 0$. Hence $\deg f \geq m$, which finishes the proof. □

7.2 An application of the generalized factor theorem

In this section, we prove an interesting result in congruence arithmetic with the help of the generalized factor theorem. Later, we will prove a generalization of this result for all finite fields.

Theorem 7.2.1. *Suppose p is a prime number. Then*

$$x^p - x = x(x-1)\cdots(x-(p-1))$$

in $\mathbb{Z}_p[x]$.

Proof. By the Fermat's little theorem, for every $a \in \mathbb{Z}_p$, we have $a^p - a = 0$. This means $0, 1, \ldots, p-1$ are distinct zeros of $x^p - x$ in $\mathbb{Z}_p$. Since $\mathbb{Z}_p$ is an integral domain, we can employ the generalized factor theorem and deduce that there is $q(x) \in \mathbb{Z}_p[x]$ such that

$$x^p - x = x(x-1)\cdots(x-(p-1))q(x). \tag{7.5}$$

Comparing the degree of the both sides of (7.5), we obtain that $p = p + \deg q$. Hence $q(x) = c$ is a non-zero constant. Therefore

$$x^p - x = cx(x-1)\cdots(x-(p-1)). \tag{7.6}$$

Comparing the leading coefficients of (7.6), we deduce that $c = 1$. This implies that

$$x^p - x = x(x-1)\cdots(x-(p-1)),$$

and the claim follows. ☐

As a corollary of Theorem 7.2.1, we deduce Wilson's theorem.

Corollary 7.2.2. *Suppose p is prime. Then $(p-1)! \equiv -1 \pmod{p}$.*

Proof. By Theorem 7.2.1, we have

$$x^p - x = x(x-1)\cdots(x-(p-1)) \tag{7.7}$$

in $\mathbb{Z}_p[x]$. This means that all the coefficients of these polynomials are congruent modulo p. Let's compare the coefficients of x. The coefficient of x on the left hand side of (7.7) is -1, and the coefficient of x on the right hand side of (7.7) is $(-1)(-2)\cdots(-(p-1))$. Therefore

$$(-1)^{p-1}(p-1)! \equiv -1 \pmod{p}. \tag{7.8}$$

For $p = 2$, we have $(2-1)! \equiv -1 \pmod{2}$. So we can and will assume that $p \neq 2$. Therefore p is odd, which implies that $(-1)^{p-1} = 1$. By (7.8) and $(-1)^{p-1} = 1$, we obtain that

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

which finishes the proof of Wilson's theorem. ☐

We can use polynomial equations to deduce many interesting congruence relations. The next exercise is another such example.

Exercise 7.2.3. *Suppose p is an odd prime number. Use $(x-1)^p = x^p - 1$ in $\mathbb{Z}_p[x]$ and the cancellation law in $\mathbb{Z}_p[x]$, to deduce that*

$$\binom{p-1}{i} \equiv (-1)^i \pmod p$$

for every $0 \le i \le p-1$.

7.3 Ideals of ring of polynomials over a field

Let's go back to the zeros of polynomials. Suppose $\alpha \in \mathbb{C}$ is a zero of a polynomial. We would like to understand the ring structure of $\mathbb{Q}[\alpha]$. By the first isomorphism theorem, we have

$$\mathbb{Q}[x]/\ker\phi_\alpha \simeq \mathbb{Q}[\alpha]$$

where $\phi_\alpha : \mathbb{Q}[x] \to \mathbb{C}$ is the evaluation at α. To understand the ring structure of $\mathbb{Q}[\alpha]$, we need to study the ideals of $\mathbb{Q}[x]$.

Theorem 7.3.1. *Suppose F is a field. Then every ideal of $F[x]$ is principal.*

Proof. Suppose I is an ideal of $F[x]$. If I is the zero ideal, we are done. Suppose I is not zero, and choose $p_0(x) \in I$ such that

$$\deg p_0 = \min\{\deg p \mid p \in I \setminus \{0\}\};$$

$\deg p_0$ is the smallest among the degrees of non-zero polynomials of I. The next claim finishes the proof.

Claim. $I = \langle p_0 \rangle$.

Proof of Claim. Since p_0 is in I, $\langle p_0 \rangle \subseteq I$. Next we want to show that every element of I is in $\langle p_0 \rangle$. Suppose $f(x) \in I$. We have to show that $f(x)$ is a multiple of $p_0(x)$. Since F is a field every non-zero element of F is a unit. This implies that the leading coefficient of p_0 is a unit in F, and so we can use the long division and divide $f(x)$ by $p_0(x)$. Let $q(x)$ be the quotient and $r(x)$ be the remainder of $f(x)$ divided by $p_0(x)$: this means

$$f(x) = p_0(x)q(x) + r(x) \quad \text{and} \quad \deg r < \deg p_0. \tag{7.9}$$

Since $f(x), p_0(x) \in I$, $r(x) = f(x) - p_0(x)q(x) \in I$. As $r \in I$, $\deg r < \deg p_0$ and $\deg p_0$ is the smallest degree of non-zero polynomials of I, we obtain that $r(x) = 0$. Therefore $f(x) = p_0(x)q(x) \in \langle p_0(x) \rangle$. This completes the proof of the Claim. □

Definition 7.3.2. *Suppose D is an integral domain. We say D is a Principal Ideal Domain (PID) if every ideal of D is principal.*

Example 7.3.3. *The ring $\mathbb{Z}$ of integers and the ring $F[x]$ of polynomials over a field F are PIDs.*

Let's recall that the way we proved $\mathbb{Z}$ is a PID is by using a result from group theory which asserts that every subgroup of $\mathbb{Z}$ is of the form $n\mathbb{Z}$ for some integer n. This result, in part, was proved using the long division for integers. As we see, there is a common technique of using a long division to prove that $\mathbb{Z}$ and $F[x]$ are PIDs. This brings us to the definition of *Euclidean Domain*.

7.4 Euclidean Domain

In mathematics, we often find a common pattern, extract the essence of various proofs, and introduce a new object that has only the needed properties. The advantage of this process is that for new examples we can focus on only the needed properties.

Definition 7.4.1. *An integral domain D is called a* Euclidean domain *if there is a* norm function $N : D \to \mathbb{Z}^{\geq 0}$ *with the following properties:*

1. $N(d) = 0$ *if and only if* $d = 0$.
2. *For every* $a \in D$ *and* $b \in D \setminus \{0\}$*, there are* $q, r \in D$ *such that*
   - *(i)* $a = bq + r$*, and*
   - *(ii)* $N(r) < N(b)$.

In a Euclidean Domain, we have a form of a long division, and this help us prove that every Euclidean Domain is a PID.

Theorem 7.4.2. *Suppose D is a Euclidean Domain. Then D is a PID.*

Proof. Suppose I is an ideal of D. If I is zero, we are done. Suppose I is not zero. Choose $a_0 \in I$ such that $N(a_0)$ is the smallest among the norm of the non-zero elements of I:

$$N(a_0) = \min\{N(a) \mid a \in I \setminus \{0\}\}.$$

The following Claim finishes the proof.

Claim. $I = \langle a_0 \rangle$.

Proof of Claim. Since $a_0 \in I$, we have $\langle a_0 \rangle \subseteq I$. Next we show that every element of I is a multiple of a_0. For $a \in I$, by the main property of Euclidean Domains, there are $q, r \in D$ such that

$$a = a_0 q + r, \quad \text{and} \quad N(r) < N(a_0). \tag{7.10}$$

Since $a, a_0 \in I$, we have $r = a - a_0 q \in I$. As $r \in I$, $N(r) < N(a_0)$, and $N(a_0)$ is the smallest norm of non-zero elements of I, we obtain that $r = 0$. Therefore $a = a_0 q \in \langle a_0 \rangle$. This completes the proof of the Claim. □

Notice that because of the long division for integers, the function $N : \mathbb{Z} \to \mathbb{Z}^{\geq 0}, N(a) := |a|$ makes $\mathbb{Z}$ a Euclidean domain. Similarly the long division for polynomials and the function $N : F[x] \to \mathbb{Z}^{\geq 0}, N(f(x)) := 2^{\deg f}$ (with the convention that $2^{-\infty} = 0$) makes $F[x]$ a Euclidean domain when F is a field.

Next we use the concept of Euclidean Domain to prove that the Gaussian integers $\mathbb{Z}[i]$ is a PID. In the next lecture, we will prove:

Theorem 7.4.3. *The ring $\mathbb{Z}[i]$ of Gaussian integers is a Euclidean domain. Therefore $\mathbb{Z}[i]$ is a PID.*

Chapter 8

Lecture 8

8.1 Gaussian integers

In the other lecture, we defined Euclidean Domain and proved that every Euclidean domain is a PID. We have also pointed out that $\mathbb{Z}$ and $F[x]$, where F is a field, are Euclidean domains. Next we want to prove that the ring $\mathbb{Z}[i]$ of Gaussian integers is a Euclidean domain, and so it is a PID.

Theorem 8.1.1. *$\mathbb{Z}[i]$ is a Euclidean domain and a PID.*

Proof. To show $\mathbb{Z}[i]$ is a Euclidean domain, we have to find a *norm function* with the desired properties. Let

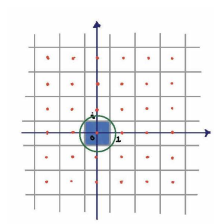
$$N : \mathbb{Z}[i] \to \mathbb{Z}^{\geq 0},\ N(z) := |z|^2,$$

where $|z|$ is the complex norm. Notice that for every integers a and b, we have $N(a + bi) = a^2 + b^2 \in \mathbb{Z}^{\geq 0}$. Next notice that for every complex number z, we have

$$N(z) = 0 \Leftrightarrow |z| = 0 \Leftrightarrow z = 0.$$

It is remained to show a type of division property for $\mathbb{Z}[i]$ with respect to the function N.

We start with the division in $\mathbb{C}$: for every $z \in \mathbb{Z}[i]$ and $w \in \mathbb{Z}[i] \setminus \{0\}$, consider $\frac{z}{w} \in \mathbb{C}$. Notice that the square tiling in the given figure implies that there is $q \in \mathbb{Z}[i]$ such that $\frac{z}{w} - q$ is in the central square. Therefore $\left|\frac{z}{w} - q\right| \leq \frac{\sqrt{2}}{2}$, and so the complex norm of $r := z - wq$ is at most $\frac{\sqrt{2}}{2}|q| < |q|$.



Since z, w, q are in $\mathbb{Z}[i]$, so is r. Altogether we obtain the existence of $q, r \in \mathbb{Z}[i]$ such that

$$z = qw + r, \quad \text{and} \quad N(r) < N(q).$$

This shows that the ring of Gaussian integers is a Euclidean domain. Earlier we have seen that every Euclidean domain is a PID, which finishes the proof. ☐

Exercise 8.1.2. *Let $\omega := \frac{-1}{2} + \frac{\sqrt{3}}{2}i$, and $\mathbb{Z}[\omega] := \{a + b\omega \mid a, b \in \mathbb{Z}\}$. Use a similar method as in the proof of Theorem 8.1.1 to show that $\mathbb{Z}[\omega]$ is a PID.*

8.2 Algebraic elements and minimal polynomials

Let's go back to zeros of polynomials.

Definition 8.2.1. 1. *We say $\alpha \in \mathbb{C}$ is an* algebraic *number if it is a zero of a polynomial $f(x) \in \mathbb{Q}[x]$.*

2. *More generally, when F is a subfield of another field E*[1]*, we say $\alpha \in E$* is algebraic over F *if α is a zero of a polynomial $f(x) \in F[x]$.*

3. *A complex number α is called* transcendental *if it is not algebraic.*

4. *Assuming E is a field extension of F, we say $\alpha \in E$ is* transcendental *over F if it is not algebraic over F.*

Example 8.2.2. *$\sqrt[3]{2}$ is an algebraic number, and there are interesting and not so easy results that the Euler number e and π are transcendental.*

One can easily see that, for E is a field extension of F, $\alpha \in E$ is algebraic over F if and only if the kernel $\ker \phi_\alpha$ of the evaluation map

$$\phi_\alpha : F[x] \to E, \quad \phi_\alpha(f(x)) := f(\alpha)$$

is non-zero. In this setting, our goal is to understand the structure of the ring $F[\alpha]$. So far we have seen many such examples: $\mathbb{Q}[i]$, $\mathbb{Q}[\sqrt{2}]$, etc. In the examples we have discussed, we described the elements of these rings as certain linear combinations, and proved that all of these rings are fields. We want to generalize these results.

Notice that by the first isomorphism theorem we have

$$F[x]/\ker \phi_\alpha \simeq F[\alpha]. \tag{8.1}$$

This means we need to investigate $\ker \phi_\alpha$. For instance we immediately deduce the following:

Corollary 8.2.3. *Suppose E is a field extension of F, and $\alpha \in E$ is transcendental over F. Then $F[\alpha] \simeq F[x]$.*

Proof. Since α is transcendental over F, $\ker \phi_\alpha = 0$. Therefore by (8.1), the claim follows. ☐

Next we use the fact that $F[x]$ is a PID to describe $\ker \phi_\alpha$ when $\alpha \in E$ is algebraic over F.

Theorem 8.2.4 (The minimal polynomial)**.** *Suppose E is a field extension of F, and $\alpha \in E$ is algebraic over F. Then the following statements hold.*

[1] In this case we say E is a *field extension* of F.

1. *There is a unique non-constant monic polynomial* $m_\alpha(x) \in F[x]$ *such that* $\ker \phi_\alpha = \langle m_\alpha(x) \rangle$. *(*$m_\alpha(x) \in F[x]$ *is called* the minimal polynomial of α over F.

2. *The minimal polynomial* $m_\alpha(x) \in F[x]$ *is a non-constant monic polynomial which cannot be written as a product of smaller degree polynomials in* $F[x]$.

Proof. (1) Since $F[x]$ is a PID, there is $f(x) \in F[x]$ which generates $\ker \phi_\alpha$. Since α is algebraic over F, $f(x)$ is not zero. We also know that non-zero constant functions are not in the kernel of ϕ_α. Hence $f(x)$ is not a constant polynomial. Suppose

$$f(x) = a_n x^n + \cdots + a_0$$

and $a_n \neq 0$. Then a_n is a unit in F (as F is a field). Let

$$\overline{f}(x) := a_n^{-1} f(x) = x^n + (a_n^{-1} a_{n-1}) x^{n-1} + \cdots + (a_n^{-1} a_0).$$

Since $\overline{f}(x) = a_n^{-1} f(x) \in \langle f(x) \rangle$ and $f(x) = a_n \overline{f}(x) \in \langle \overline{f} \rangle$, we deduce that

$$\langle \overline{f} \rangle = \langle f \rangle = \ker \phi_\alpha.$$

This shows the *existence* of a monic non-constant polynomial which generates $\ker \phi_\alpha$. Next we show the uniqueness of such a polynomial. It is clear that uniqueness is a special case of the following Claim.

Claim. *Suppose* f_1 *and* f_2 *are non-constant monic polynomials in* $F[x]$*, and* $\langle f_1 \rangle = \langle f_2 \rangle$*. Then* $f_1 = f_2$.

Proof of Claim. Since $\langle f_1 \rangle = \langle f_2 \rangle$, there are polynomials $q_1, q_2 \in F[x]$ such that $f_1 q_1 = f_2$ and $f_2 q_2 = f_1$. Comparing the degrees of the sides, we deduce that

$$\deg f_1 + \deg q_1 = \deg f_2 \quad \text{and} \quad \deg f_2 + \deg q_2 = \deg f_1. \tag{8.2}$$

Notice that since $f_i \neq 0$, so are q_i's. Therefore $\deg q_i \geq 0$. Hence by (8.2), we have $\deg f_1 \leq \deg f_2$ and $\deg f_2 \leq \deg f_1$. This implies that $\deg f_1 = \deg f_2$, and so q_i's are non-zero constants. Suppose $q_1(x) = c \in F^\times$. Then we have $cf_1 = f_2$. Comparing the leading coefficients of both sides, we obtain that $c = 1$. Therefore $f_1 = f_2$, and the claim follows. □

(2) Suppose to the contrary that $m_\alpha(x) = g(x)h(x)$ for some $g(x), h(x) \in F[x]$ with $\deg g, \deg h < \deg m_\alpha$. Then

$$\phi_\alpha(g)\phi_\alpha(h) = \phi_\alpha(m_\alpha) = 0$$

implies that either $g \in \ker \phi_\alpha$ or $h \in \ker \phi_\alpha$ (notice that F has no zero-divisors). As $\ker \phi_\alpha$ is generated by $m_\alpha(x)$, either $m_\alpha(x) | g(x)$ or $m_\alpha(x) | h(x)$. Since g and h are not zero, we deduce that either $\deg m_\alpha \leq \deg g$ or $\deg m_\alpha \leq \deg h$. This contradicts that $\deg g, \deg h < \deg m_\alpha$. □

Next we prove the converse of the second part of Theorem 8.2.4. This result will help us to actually find the minimal polynomial $m_\alpha(x)$ for some algebraic elements.

Theorem 8.2.5 (Characterization of minimal polynomials)**.** *Suppose E is a field extension of F, and $\alpha \in E$ is algebraic over F. Then a monic non-constant polynomial $p(x)$ in $F[x]$ is the minimal polynomial of α if and only if $p(\alpha) = 0$ and $p(x)$ cannot be written as a product of smaller degree polynomials in $F[x]$.*

Proof. Part (2) of Theorem 8.2.4 gives us ($\Rightarrow$), and so we focus on ($\Leftarrow$).

Since $p(\alpha) = 0$, $p(x)$ is in $\ker \phi_\alpha$. As $\ker \phi_\alpha$ is generated by the minimal polynomial m_α, we obtain that $p(x) = m_\alpha(x)q(x)$ for some $q(x) \in F[x]$. Since $p(x)$ cannot be written as a product of smaller degree polynomials in $F[x]$, we deduce that $\deg m_\alpha = \deg p$ and $q(x)$ is a non-zero constant polynomial. Suppose $q(x) = c \in F$. Then comparing the leading coefficients of both sides of $cm_\alpha(x) = p(x)$, it follows that $c = 1$. Thus $m_\alpha(x) = p(x)$, and the claim follows. □

It is useful to notice that $m_\alpha(x)$ has the smallest degree among non-zero polynomials in $F[x]$ that have α as a zero.

Proposition 8.2.6. *Suppose E is a field extension of F, and $\alpha \in E$ is algebraic over F. Then the following statements hold.*

1. *For $f(x) \in F[x]$, $f(\alpha) = 0$ if and only if $m_\alpha(x)|f(x)$ in $F[x]$.*
2. *Suppose α is a zero of a non-zero polynomial $p(x) \in F[x]$. If $\deg p \le \deg m_\alpha$, then there is a non-zero constant c such that $p(x) = cm_\alpha(x)$.*

Proof. (1) We have $f(\alpha) = 0 \Leftrightarrow f \in \ker \phi_\alpha = \langle m_\alpha(x)\rangle \Leftrightarrow m_\alpha(x)|f(x)$.

(2) Since $p(\alpha) = 0$, by the first part we have that $p(x)$ is a (non-zero) multiple of $m_\alpha(x)$; that means there is a non-zero polynomial $q(x) \in F[x]$ such that $p(x) = q(x)m_\alpha(x)$. As $\deg p \le \deg m_\alpha$, we deduce that

$$\deg m_\alpha \ge \deg p = \deg m_\alpha + \deg q, \text{ which implies that } \deg q = 0.$$

This means q is a non-zero constant, and the claim follows. □

8.3 Elements of quotients of ring of polynomials

Let's recall that one of our goals is to understand the ring structure of $F[\alpha]$ and describe its elements. By the discussions in the previous section, we have $F[\alpha] \simeq F[x]/\langle m_\alpha(x)\rangle$. The next result, which is based on the long division for polynomials, gives us a description of elements of the quotient ring of a ring of polynomials by a monic polynomial.

Proposition 8.3.1. *Suppose A is a unital commutative ring, and $p(x) \in A[x]$ is a monic polynomial of degree $n \ge 1$. Then every element of $A[x]$ can be uniquely written as*

$$a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + \langle p(x)\rangle$$

for some $a_0, \ldots, a_{n-1} \in A$.

Proof. **Existence.** For every $f(x) \in A[x]$, by the long division for polynomials there are unique $q(x) \in A[x]$ (the quotient) and $r(x) \in A[x]$ (the remainder) such that

1. $f(x) = q(x)p(x) + r(x)$, and
2. $\deg r < \deg p$.

The second item means that $r(x) = \sum_{i=0}^{n-1} a_i x^i$ for some $a_i \in A$. The first item implies that $f(x) - r(x) \in \langle p(x) \rangle$. Altogether we have

$$f(x) + \langle p(x) \rangle = \sum_{i=0}^{n-1} a_i x^i + \langle p(x) \rangle.$$

Uniqueness. Suppose $\sum_{i=0}^{n-1} a_i x^i + \langle p(x) \rangle = \sum_{i=0}^{n-1} a_i' x^i + \langle p(x) \rangle$. Then $h(x) := \sum_{i=0}^{n-1} a_i x^i - \sum_{i=0}^{n-1} a_i' x^i$ is a multiple of $p(x)$ and has degree at most $n-1$. As $\deg p = n$ and $p(x)$ is monic, the only multiple of $p(x)$ that has degree less than n is 0. Hence $\sum_{i=0}^{n-1} a_i x^i = \sum_{i=0}^{n-1} a_i' x^i$, which implies the uniqueness part. □

Chapter 9

Lecture 9

9.1 Elements of $F[\alpha]$

One of our main goals is to understand the ring structure of $\mathbb{Q}[\alpha]$ for an algebraic number α. In the previous lecture we showed that for a field extension E of F and $\alpha \in E$ that is algebraic over F, there is a unique monic non-constant polynomial $m_\alpha(x) \in F[x]$ such that

1. For every $f(x) \in F[x]$, $f(\alpha) = 0$ if and only if $m_\alpha(x)|f(x)$.

2. For a monic polynomial $p(x) \in F[x]$, we have that $p(x) = m_\alpha(x)$ if and only if $p(\alpha) = 0$ and $p(x)$ cannot be written as a product of smaller degree polynomials in $F[x]$.

3. $F[\alpha] \simeq F[x]/\langle m_\alpha(x)\rangle$.

The polynomial $m_\alpha(x) \in F[x]$ is called the minimal polynomial of α over F. [1] Because of the third property, we described elements of the quotient ring $F[x]/\langle p(x)\rangle$ where $p(x)$ is a polynomial of degree $n \geq 1$. Using the long division for polynomials, we proved that every element of this quotient ring can be *uniquely* written as $r(x) + \langle p(x)\rangle$ for some $r(x) \in F[x]$ with $\deg r \leq n-1$. Base on these results, we immediately get a fairly good description of elements of $F[\alpha]$.

Theorem 9.1.1. *Suppose E is a field extension of F, and $\alpha \in E$ is algebraic over F. Suppose the degree of the minimal polynomial $m_\alpha(x)$ of α over F is n. Then every element of $F[\alpha]$ can be uniquely written as*

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}$$

for some a_i's in F.

[1] A better notation for $m_\alpha(x)$ should include F as well, as the minimal polynomial of α only makes sense after we specify F. That is why in some texts $m_\alpha(x)$ is denoted by $m_{\alpha,F}(x)$. Here we assume that we know what F is from the context in which α is discussed.

Proof. By the first isomorphism theorem for rings, we know that

$$\overline{\phi}_\alpha : F[x]/\langle m_\alpha(x)\rangle \to F[\alpha], \quad \overline{\phi}_\alpha(f(x)+\langle m_\alpha(x)\rangle) := f(\alpha) \tag{9.1}$$

is an isomorphism. By Proposition 8.3.1, every element of $F[x]/\langle m_\alpha(x)\rangle$ can be uniquely written as $(\sum_{i=0}^{n-1} a_i x^i) + \langle m_\alpha(x)\rangle$ for some a_i's in F. Hence by (9.1), we obtain that every element of $F[\alpha]$ can be uniquely written as

$$\overline{\phi}_\alpha\Big(\sum_{i=0}^{n-1} a_i x^i) + \langle m_\alpha(x)\rangle\Big) = \sum_{i=0}^{n-1} a_i \alpha^i.$$

This completes the proof. □

Note that Theorem 9.1.1 is a generalization of many examples that we have discussed so far, e.g.

$$\mathbb{Q}[i] = \{a+bi \mid a,b\in\mathbb{Q}\} \quad \text{because} \quad m_{i,\mathbb{Q}}(x) = x^2+1,$$

and

$$\mathbb{Q}[\sqrt[3]{2}] = \{a_0 + a_1\sqrt[3]{2} + a_2\sqrt[3]{4} \mid a_0,a_1,a_2\in\mathbb{Q}\} \quad \text{because} \quad m_{\sqrt[3]{2},\mathbb{Q}}(x) = x^3-2.$$

9.2 Irreducible elements

By now it is clear that in order to understand the ring structure of $F[\alpha]$ for a given α which is algebraic over F, we have to figure out a way to find the minimal polynomial $m_\alpha(x)\in F[x]$. Theorem 8.2.5 gives us a key characterization of $m_\alpha(x)$ which brings us to the definition of irreducible elements.

Definition 9.2.1. *Suppose D is an integral domain. We say $d\in D$ is* irreducible *if*

1. $d\notin D^\times\cup\{0\}$*, and*
2. *If $d = ab$ for some $a,b\in D$, then either $a\in D^\times$ or $b\in D^\times$.*

For instance an integer n is irreducible in $\mathbb{Z}$ if $n=\pm p$ for some prime number p. Let me warn you that later we will define *prime elements* of an integral domain, and irreducible and prime elements do not always coincide!

Lemma 9.2.2. *Suppose F is a field. Then $p(x)\in F[x]$ is irreducible if and only if $p(x)$ is not constant and it cannot be written as a product of smaller degree polynomials in $F[x]$.*

Proof. ($\Rightarrow$) Since $f(x)$ is irreducible, $f(x)\notin F[x]^\times\cup\{0\}$. As $F[x]^\times = F^\times = F\setminus\{0\}$, we obtain that $f(x)$ is not constant. Now suppose to the contrary that $f(x) = g(x)h(x)$ and $\deg g, \deg h < \deg f$. This implies that $g(x)$ and $h(x)$ are not constant polynomials. On the other hand, since $f(x)$ is irreducible, $f(x) = g(x)h(x)$ implies that either $g\in F[x]^\times$ or $h\in F[x]^\times$. As $F[x]^\times = F^\times$, we deduce that either $\deg g = 0$ or $\deg h = 0$, which is a contradiction.

($\Leftarrow$) Suppose $f(x) = g(x)h(x)$. Since f cannot be written as a product of smaller degree polynomials in $F[x]$, we have that either $\deg g \geq \deg f$ or $\deg h \geq \deg f$. As $\deg f = \deg g + \deg h$, we deduce that either $\deg g = 0$ or $\deg h = 0$. That means either $g \in F \setminus \{0\}$ or $h \in F \setminus \{0\}$. Since F is a field, we obtain that either $g \in F^\times = F[x]^\times$ or $h \in F^\times = F[x]^\times$. This completes the proof. □

Now, some of the properties of minimal polynomials can be phrased in a more compact form.

Corollary 9.2.3 (Minimal polynomials and irreducibility)**.** *Suppose E is a field extension of F, $\alpha \in E$ is algebraic over F, and $p(x) \in F[x]$ is a monic polynomial. Then $p(x) = m_\alpha(x)$ if and only if $p(\alpha) = 0$ and $p(x)$ is irreducible.*

Proof. This is an immediate consequence of Theorem 8.2.5 and Lemma 9.2.2. □

This motivates us to answer the following questions:

1. Assuming that D is an integral domain or a PID, what can we say about ideals that are generated by irreducible elements and their quotient rings?

2. Can we come up with certain mechanisms to find out whether a given monic polynomial is irreducible?

We start by answering the first question. We have already pointed out that irreducible elements of the ring of integers are essentially prime numbers. Therefore for an irreducible element p of $\mathbb{Z}$ we have that $\mathbb{Z}/\langle p \rangle$ is a field. We will show that this result holds for every PID.

Let's begin by understanding when exactly two principal ideals are equal.

Lemma 9.2.4. *Suppose D is an integral domain, and $a, b \in D$. Then $\langle a \rangle = \langle b \rangle$ if and only if $a = bu$ for some unit u.*

Proof. We notice that $\langle a \rangle = \langle b \rangle$ if and only if $a \in \langle b \rangle$ and $b \in \langle a \rangle$. This means

$$\langle a \rangle = \langle b \rangle \Leftrightarrow \exists x, y \in D, a = bx \text{ and } b = ay. \tag{9.2}$$

($\Leftarrow$) If $a = bu$ for some unit u, then $b = au^{-1}$. Therefore by (9.2), we have $\langle a \rangle = \langle b \rangle$.

($\Rightarrow$) If $a = 0$, then $b \in \langle a \rangle$ implies that $b = 0$. Therefore $a = 1 \cdot b$ and there is nothing to prove.

Suppose $a \neq 0$, and $x, y \in D$ are as in (9.2). Then

$$a = bx = (ay)x = a(yx).$$

By the cancellation law, we deduce that $yx = 1$ (notice that D is an integral domain and $a \neq 0$, and so we are allowed to use the cancellation law). Hence x is a unit, which finishes the proof. □

Lemma 9.2.4 immediately gives us a description for units in terms of ideals.

Lemma 9.2.5. *Suppose A is a unital commutative ring, and $a \in A$. Then a is a unit if and only if $\langle a \rangle = A$.*

Proof. ($\Rightarrow$) Assuming that a is a unit, we have that $a' = (a'a^{-1})a \in \langle a \rangle$ for every $a' \in A$. This means that $A = \langle a \rangle$.

($\Leftarrow$) If $\langle a \rangle = A$, then $1 \in \langle a \rangle$, which implies that $1 = aa'$ for some $a' \in A$. Therefore a is a unit. □

Lemma 9.2.5 help us to describe fields in terms of their ideals.

Lemma 9.2.6. *Suppose F is a unital commutative ring. Then F is a field if and only if F has exactly two distinct ideals $\{0\}$ and F.*

Proof. ($\Rightarrow$) Since F is a field, F and $\{0\}$ are distinct. Now suppose I is a non-zero ideal of F. Then there is a non-zero element a in I. Since F is a field, a is a unit in F. Hence by Lemma 9.2.5

$$F = \langle a \rangle \subseteq I.$$

This means $I = F$.

($\Leftarrow$) Since F and $\{0\}$ are distinct, $0 \notin F^{\times}$. So it is enough to show that every non-zero element of F is a unit. Suppose $a \in F \setminus \{0\}$, and consider $\langle a \rangle$. As F is the only non-zero ideal of F, we have $F = \langle a \rangle = aF$. Hence by Lemma 9.2.5, a is a unit in F. This finishes the proof. □

We also notice that in a field F, there is no irreducible element as $F = F^{\times} \cup \{0\}$. So when we are studying irreducible elements, we can and will assume that the given integral domain is not a field.

Lemma 9.2.7. *Suppose D is an integral domain which is not a field. Then $a \in D$ is irreducible if and only if $\langle a \rangle$ is a maximal ideal among proper principal ideals.*

Let's begin by explaining various phrases in the statement of Lemma 9.2.7. Suppose Σ is a collection of subsets of a given set X, Then we say $A \in \Sigma$ is a *maximal* element of Σ if there is no element $B \in \Sigma$ that contains A as a proper subset. In mathematical language, it means

$$A \in \Sigma \text{ is maximal if and only if } \forall B \in \Sigma, A \subseteq B \Rightarrow B = A.$$

In Lemma 9.2.7, the collection Σ is $\{I \trianglelefteq D \mid I \text{ is principal}, I \neq D\}$. Altogether, we can rewrite Lemma 9.2.7 as follows.

Suppose D is an integral domain which is not a field, and $a \in D$. Then a is irreducible in D if and only if $\langle a \rangle \neq D$ and for every $b \in D$,

$$\langle a \rangle \subseteq \langle b \rangle \Rightarrow \text{ either } \langle a \rangle = \langle b \rangle \text{ or } \langle b \rangle = D.$$

Proof of Lemma 9.2.7. ($\Leftarrow$) Suppose a is irreducible in D and $\langle a\rangle \subseteq \langle b\rangle$. As a is irreducible, it is not a unit. Therefore by Lemma 9.2.5, $\langle a\rangle$ is a proper ideal.

As $a \in \langle b\rangle$, $a = bc$ for some $c \in D$. Since a is irreducible, either $b \in D^\times$ or $c \in D^\times$. If $b \in D^\times$, then by Lemma 9.2.5 we have $\langle b\rangle = D$. If $c \in D^\times$, then by Lemma 9.2.4, $\langle a\rangle = \langle b\rangle$.

($\Leftarrow$) Since $\langle a\rangle$ is a proper ideal, by Lemma 9.2.5 a is not a unit. Next we argue why $a \neq 0$.

Suppose to the contrary that $a = 0$. Then for every non-zero element $b \in D$, we have $\langle a\rangle \subsetneq \langle b\rangle$. Hence by the assumption $\langle b\rangle = D$. This together with Lemma 9.2.5 implies that b is a unit. This means every non-zero element of D is a unit, which implies that D is a field. This is a contradiction.

Next let's assume that $a = bc$ for some $b, c \in D$. Then $\langle a\rangle \subseteq \langle b\rangle$. By the assumption, we deduce that either $\langle a\rangle = \langle b\rangle$ or $\langle b\rangle = D$. Therefore by Lemma 9.2.4 and Lemma 9.2.5, we deduce that either there is $u \in D^\times$ such that $a = bu$ or $b \in D^\times$. In the former case, by the cancellation law, we have $c = u \in D^\times$ and in the latter case, $b \in D^\times$. This means a is irreducible. □

9.3 Maximal ideals and their quotient rings

Based on Lemma 9.2.7, we know that irreducibility is an information about principal ideals, and so we gain more information when D is a PID. If D is a PID and $a \in D$ is irreducible, then $\langle a\rangle$ is maximal among *all* proper ideals. This brings us to the definition of maximal ideals.

Definition 9.3.1. *Suppose A is a unital commutative ring and $I \trianglelefteq A$. We say I is a* maximal *ideal if it is maximal among* proper *ideals; that means*

$$\forall J \trianglelefteq A, I \subseteq J \Rightarrow \text{ either } J = I \text{ or } J = A.$$

So by Lemma 9.2.7 and Lemma 9.2.6, we immediately obtain the following:

Lemma 9.3.2. *Suppose D is a PID, and $a \in D$. Then*

1. *for $a \neq 0$, we have that $\langle a\rangle$ is a maximal ideal if and only if a is irreducible.*

2. *$\{0\}$ is a maximal ideal if and only if D is a field.*

The next Proposition gives us the key property of maximal ideals.

Proposition 9.3.3. *Suppose A is a unital commutative ring and $I \trianglelefteq A$. Then I is a maximal ideal if and only if A/I is a field.*

We start with the *corresponding* lemma which describes ideals of a quotient ring.

Lemma 9.3.4. *Suppose A is a unital commutative ring and $I \trianglelefteq A$. Then $\bar{J}$ is an ideal of A/I if and only if $\bar{J} = J/I$ for some $J \trianglelefteq A$ which contains I.*

Proof. (⇒) Suppose $\bar{J}$ is an ideal of A/I, and let

$$J := \{a \in A \mid a + I \in \bar{J}\}.$$

Then for every $a \in I$, we have $a + I = 0 + I \in \bar{J}$, and so $a \in J$. Therefore $I \subseteq J$. Next we show that J is an ideal of A. Suppose $a, a' \in J$. Then $a + I, a' + I \in \bar{J}$. As $\bar{J}$ is an ideal, we have $(a + I) - (a' + I) \in \bar{J}$. This implies that $(a - a') + I \in \bar{J}$, and so $a - a' \in J$. For $a \in J$, we have that $a + I \in \bar{J}$. Since $\bar{J}$ is an ideal of A/I, for every $b \in A$, we have that $(b + I)(a + I) \in \bar{J}$. Therefore $ba + I \in \bar{J}$. Hence $ba \in J$. Altogether we have that J is an ideal, it contains I, and

$$\bar{J} = \{a + I \mid a \in J\} = J/I.$$

(⇐) From group theory we know that J/I is a subgroup of A/I. Now suppose $a + I \in J/I$ and $b + I \in A/I$. Since J is an ideal of A and $a \in J$, we have that $ab \in J$. Hence $(a + I)(b + I) \in J/I$. Thus J/I is an ideal of A/I. □

Proof of Proposition 9.3.3. By Lemma 9.2.6, A/I is a field if and only if it has exactly two ideals I/I and A/I. By Lemma 9.3.4, every ideal of A/I is of the form J/I where J is an ideal of A which contains I. Hence A/I has exactly two ideals if and only if I and A are the only ideals of A which contain I and $I \neq A$. The latter happens exactly when I is a maximal ideal. This completes the proof. □

We immediately get the following corollary for PIDs.

Corollary 9.3.5. *Suppose that D is a PID and not a field, and $a \in D$. Then $D/\langle a \rangle$ is a field if and only if a is irreducible in D.*

Proof. By Proposition 9.3.3, $D/\langle a \rangle$ is a field if and only if $\langle a \rangle$ is a maximal ideal. By Lemma 9.3.2 and the assumption that D is not a field, $\langle a \rangle$ is a maximal ideal if and only if a is irreducible. This completes the proof. □

9.4 $F[\alpha]$ is a field!

Now we are well-prepared to prove the following:

Theorem 9.4.1. *Suppose E is a field extension of F, and $\alpha \in E$ is algebraic over F. Then $F[\alpha]$ is a field.*

Proof. We have already proved that $F[\alpha] \simeq F[x]/\langle m_{\alpha,F}(x) \rangle$ where $m_{\alpha,F}(x)$ is the minimal polynomial of α over F (see (9.1) and Theorem 8.2.4). We further showed that $m_{\alpha,F}(x)$ is irreducible in $F[x]$ (see Corollary 9.2.3), and $F[x]$ is a PID (see Theorem 7.3.1) which is not a field (see Lemma 6.3.2). Then by Corollary 9.3.5, we deduce that $F[x]/\langle m_{\alpha,F}(x) \rangle$ is a field. Thus $F[\alpha]$ is a field. □

As you can see in this proof, we do not show the existence of the multiplicative inverse of an element in a direct way. So for a given algebraic number α sometimes it is tricky to express the inverse of $p(\alpha)$ in terms of a linear combination of $1, \alpha, \alpha^2, \cdots$.

Exercise 9.4.2. *Suppose $\alpha \in \mathbb{C}$ is a zero of $x^3 - x + 1$. Express α^{-1}, $(\alpha + 1)^{-1}$, and $(\alpha^2 + 1)^{-1}$ in the form $a_0 + a_1\alpha + a_2\alpha^2$ for some $a_0, a_1, a_2 \in \mathbb{Q}$.*

Chapter 10

Lecture 10

We have seen that many properties of $F[\alpha]$ where α is algebraic over F depends on the minimal polynomial $m_{\alpha,F}(x)$ of α over F. So it is crucial to have a method of finding $m_{\alpha,F}(x)$. Let's recall that the key property of the minimal polynomial is the following:

$p(x) = m_{\alpha,F}(x)$ *if and only if* $p(\alpha) = 0$, $p(x)$ *is monic and irreducible in* $F[x]$.

We will prove a series of *irreducibility criteria* which help us find minimal polynomials of certain algebraic elements.

10.1 Irreducibility and zeros of polynomials

We start with pointing out a consequence of the factor theorem.

Lemma 10.1.1. *Suppose F is a field, and $f(x) \in F[x]$.*

1. *If* $\deg f = 1$*, then f is irreducible.*
2. *If* $\deg f \geq 2$ *and f has a zero in F, then f is not irreducible.*
3. *Suppose* $\deg f = 2$ *or* 3*. Then f is irreducible in $F[x]$ if and only if f does not have a zero in F.*

Proof. (1) Suppose $\deg f = 1$. Then clearly it is not constant. If $f = gh$, then $1 = \deg g + \deg h$, which implies that we cannot have $\deg g, \deg h < 1$. Therefore f is irreducible.

(2) Suppose $\deg f \geq 2$ and $f(a) = 0$ for some $a \in F$. Then by the factor theorem, there is $g(x) \in F[x]$ such that $f(x) = (x-a)g(x)$. Hence $\deg g = \deg f - 1 < \deg f$ and $\deg(x-a) < \deg f$. Therefore $f(x)$ is not irreducible in $F[x]$.

(3) Suppose $\deg f = 2$ or 3 and $f(x)$ is not irreducible. Then there are $g, h \in F[x]$ such that $f(x) = g(x)h(x)$ and $\deg g, \deg h < \deg f \leq 3$. These imply that $\deg g, \deg h \geq 1$ and $\deg f = \deg g + \deg h \leq 3$. Hence either $\deg g = 1$ or $\deg h = 1$. Without loss of generality, we can and will assume that $\deg g = 1$. Thus $g(x) = a_0 + a_1 x$ for some $a_0, a_1 \in F$ and $a_1 \neq 0$. Then $-a_0 a_1^{-1} \in F$ is a zero of $g(x)$, which implies that $f(x)$ has a zero in F. $\square$

Example 10.1.2. *1. $f(x) := x^3 - x + 1$ is irreducible in $\mathbb{Z}_3[x]$.*

2. $\mathbb{Z}_3[x]/\langle f(x)\rangle$ is a field of order 27.

Proof. (1) Since $\deg f = 3$, f is irreducible in $\mathbb{Z}_3[x]$ if and only if it does not have a zero in $\mathbb{Z}_3$. As we have seen earlier, by the Fermat's little theorem, $x^3 - x + 1$ does not have a zero in $\mathbb{Z}_3$, which finishes the proof of part one.

(2) Since $f(x)$ is irreducible and $\mathbb{Z}_3[x]$ is a PID, $\langle f(x)\rangle$ is a maximal ideal of $\mathbb{Z}_3[x]$ (see Lemma 9.3.2). Therefore $\mathbb{Z}_3[x]/\langle f(x)\rangle$ is a field (see Proposition 9.3.3). We have proved that every element of $\mathbb{Z}_3[x]/\langle f(x)\rangle$ can be uniquely written as $r(x)+\langle f(x)\rangle$ for a polynomial $r(x) \in \mathbb{Z}_3[x]$ with degree at most 2. Notice that there are 27 polynomials of degree at most 2 in $\mathbb{Z}_3[x]$. (see Proposition 8.3.1). □

Exercise 10.1.3. *Every odd degree polynomial in $\mathbb{R}[x]$ is not irreducible.*

10.2 Rational root criterion

Next we give an effective criterion for finding out whether or not a polynomial in $\mathbb{Z}[x]$ has a zero in $\mathbb{Q}$.

Proposition 10.2.1 (Rational root criterion)**.** *Suppose*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{Z}[x],$$

$a_0 \neq 0$, and $a_n \neq 0$. If $f(\frac{b}{c}) = 0$ for some $b, c \in \mathbb{Z}$ with $c \neq 0$ and $\gcd(b,c) = 1$, then

$$b|a_0 \quad \textit{and} \quad c|a_n.$$

(The denominator divides the leading coefficient and the numerator divides the constant term.)

Proof. Since $f(\frac{b}{c}) = 0$, multiplying both sides by c^n, we have

$$a_n b^n + a_{n-1}b^{n-1}c + \cdots + a_1 bc^{n-1} + a_0 c^n = 0. \tag{10.1}$$

This implies that

$$a_n b^n = -c(a_{n-1}b^{n-1} + \cdots + a_1 bc^{n-2} + a_0 c^{n-1}) \text{ is a multiple of } c.$$

Since $\gcd(b,c) = 1$ and $c|a_n b^n$, by Euclid's lemma, $c|a_n$. Similarly (10.1) implies that

$$a_0 c^n = -b(a_n b^{n-1} + a_{n-1}b^{n-2}c + \cdots + a_1 c^{n-1}) \text{ is a multiple of } b.$$

Therefore again by Euclid's lemma we deduce $b|a_0$. This finishes the proof. □

The rational root criterion has many implications. Here is one of them:

Corollary 10.2.2. *Suppose $f(x) \in \mathbb{Z}[x]$ is a monic polynomial. Then every rational zero of f is an integer which is the divisor of the constant term $f(0)$.*

Proof. Suppose $\frac{b}{c}$ is a zero of f and $\gcd(b,c)=1$. Then by the rational root criterion, c divides the leading coefficient which is 1. Hence $c=\pm 1$. This implies that $\frac{b}{c}=\pm b\in\mathbb{Z}$. Another application of the rational root criterion implies that b divides the constant term. This completes the proof. □

Example 10.2.3. *Suppose $f(x)=x^n+a_{n-1}x^{n-1}+\cdots+a_1x+1\in\mathbb{Z}[x]$. Prove that f has a rational zero if and only if either $f(1)=0$ or $f(-1)=0$.*

Proof. By Corollary 10.2.2, since f is a monic integer polynomial, every rational zero of f is integer and it is a divisor of the constant term which is 1. Hence a rational zero of f is either 1 or -1. This finishes the proof. □

10.3 Mod criterion: zeros

Though Corollary 10.2.2 theoretically gives us a relatively good algorithm for finding out the rational zeros of a monic integer polynomial, but from computational point of view it might be a daunting task to evaluate a polynomial of degree 20 at 2. On the other hand, finding what 2^n modulo 5 is actually easy! This means from computational point of view it is better to work with integers modulo a *small* positive integer. The following lemma shows us how we can employ this technique.

Lemma 10.3.1. *Suppose A and B are unital commutative rings, and $c:A\to B$ is a ring homomorphism. Then*

1. $c:A[x]\to B[x], c\Big(\sum_{i=0}^n a_ix^i\Big):=\sum_{i=0}c(a_i)x^i$ *is a ring homomorphism.*

2. *For $a\in A$ and $b\in B$, let*

$$\phi_a:A[x]\to A, \phi_a(f(x)):=f(a) \quad \textit{and} \quad \phi_b:B[x]\to B, \phi_b(g(x)):=g(b)$$

be the corresponding evaluation maps. Then for every $a\in A$ we have

$$c(\phi_a(f(x)))=\phi_{c(a)}(c(f(x))).$$

Proof. Both parts are easy to check and I leave the task of writing the details as an exercise. □

Lemma 10.3.1 immediately implies that if $f(x)\in A[x]$ has a zero in A, then $c(f)$ has a zero in B. The contrapositive of this statement is often used.

Suppose $c:A\to B$ is a ring homomorphism, and $f(x)\in A[x]$. If $c(f(x))$ does not have a zero in B, then $f(x)$ does not have a zero in A.

Here is one important example.

Lemma 10.3.2. *Suppose $f(x)\in\mathbb{Z}[x]$ is a monic polynomial. If $f(x)$ does not have a zero in $\mathbb{Z}_n$ for some positive integer n, then $f(x)$ does not have a zero in $\mathbb{Q}$.*

The common steps for proving statements of this type where we want to show certain property $\mathscr{P}$ passes from $\mathbb{Z}_n$ to $\mathbb{Q}$ are:

0. Look at the contrapositive, and start with $\mathbb{Q}$.
1. Show that we can pass to $\mathbb{Z}$.
2. Use the residue maps and pass to $\mathbb{Z}_n$.

Usually Step 1 is the hard step where we want to go from $\mathbb{Q}$ to $\mathbb{Z}$.

Proof of Lemma 10.3.2. Suppose $f(x)$ has a zero in $\mathbb{Q}$. Since $f(x) \in \mathbb{Z}[x]$ is monic, by Corollary 10.2.2 $f(x)$ has a zero $a \in \mathbb{Z}$. Then by Lemma 10.3.1, $c_n(a) := [a]_n$ is a zero of $c_n(f)$ where $c_n : \mathbb{Z} \to \mathbb{Z}_n$. (We simply say that a is a zero of $f(x)$ in $\mathbb{Z}_n$). This shows that the contrapositive of the claim holds, which finishes the proof. ☐

Lemma 10.3.2 in conjunction with Fermat's little theorem can become a very strong tool. Let's recall that Fermat's little theorem states

$$a^p = a \quad \text{for every } a \in \mathbb{Z}_p.$$

Hence $a^{p^2} = (a^p)^p = a^p = a$ for every $a \in \mathbb{Z}_p$. Therefore inductively we can show that the following holds:

For every positive integer n, prime p, and $a \in \mathbb{Z}_p$,

$$a^{p^n} = a. \tag{10.2}$$

By (10.2), we have that for non-negative integers $c_0, \ldots, c_n$, prime p, and $a \in \mathbb{Z}_p$ the following holds:

$$a^{c_n p^n + c_{n-1} p^{n-1} + \cdots + c_0} = a^{c_n + \cdots + c_0}.$$

This gives us a fast algorithm for finding large powers of elements in $\mathbb{Z}_p$. This makes it easier to evaluate (large degree) polynomials in $\mathbb{Z}_p$.

Example 10.3.3. *Suppose p is prime. Prove that $f(x) := x^{p^2} + px^{p^2-p} - x + (2p+1)$ does not have a rational zero.*

Proof. We will show that $f(x)$ does not have a zero in $\mathbb{Z}_p$. Notice that $f(x)$ modulo p is $x^{p^2} - x + 1$. Hence for every $a \in \mathbb{Z}_p$, we have

$$f(a) = a^{p^2} - a + 1 = 1, \tag{10.3}$$

where the last equality holds because of (10.2). By (10.3), $f(x)$ does not have a zero in $\mathbb{Z}_p$. By Lemma 10.3.2, we deduce that $f(x)$ does not have a zero in $\mathbb{Q}$. This finishes the proof. ☐

Chapter 11

Lecture 11

In the previous lecture, we showed that if F is a field, $f(x) \in F[x]$ is a polynomial with degree at least 2, and $f(x)$ has a zero in F, then f is not irreducible. We further showed that the converse holds if the degree of f is either 2 or 3. Then we proved the rational root criterion and use it to show that if $f(x) \in \mathbb{Z}[x]$ is a monic polynomial which does not have a zero in $\mathbb{Z}_n$ for some positive integer n, then f does not have a zero in $\mathbb{Q}$. We proved the the contrapositive by first passing from $\mathbb{Q}$ to $\mathbb{Z}$, and then from $\mathbb{Z}$ to $\mathbb{Z}_n$.

We can use the residue maps to find out if a polynomial $f(x) \in \mathbb{Z}[x]$ is irreducible or not.

Theorem 11.0.1 (mod-p criterion)**.** *Suppose $f(x) \in \mathbb{Z}[x]$ is a monic polynomial and p is a prime number. If $f(x)$ is irreducible in $\mathbb{Z}_p[x]$, then $f(x)$ is irreducible in $\mathbb{Q}[x]$.*

The proof of this theorem has many steps. The general strategy is the same as the one explained in Section 10.3. We prove the contrapositive statement, and it will be done by (1) going from $\mathbb{Q}$ to $\mathbb{Z}$ and (2) going from $\mathbb{Z}$ to $\mathbb{Z}_p$. The main difficulty is in the first step, where we need Gauss's Lemma.

11.1 Content of a polynomial with rational coefficients.

Before we go to the proof, we point out an important difference between being irreducible in $\mathbb{Q}[x]$ and being irreducible in $\mathbb{Z}[x]$.

Example 11.1.1. *$2x$ is irreducible in $\mathbb{Q}[x]$, but it is not irreducible in $\mathbb{Z}[x]$.*

Proof. By Lemma 10.1.1, we know that every degree 1 polynomial with coefficients in a field is irreducible. Therefore $2x$ is irreducible in $\mathbb{Q}[x]$. On the other hand, $2x$ is 2 times x and neither 2 nor x is a unit in $\mathbb{Z}[x]$ as $\mathbb{Z}[x]^\times = \mathbb{Z}^\times = \{1, -1\}$. □

In fact in general if the greatest common divisor of the coefficients of a non-constant integer polynomial $f(x)$ is not 1, then $f(x)$ cannot be irreducible in $\mathbb{Z}[x]$. This is the case as we can simply factor out the greatest common divisor of the coefficients of f

and write $f(x)$ as a product of two non-unit elements of $\mathbb{Z}[x]$. This brings us to the definition of the *content* of an integer polynomial.

Definition 11.1.2. *Suppose $f(x) := a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$ is a non-zero polynomial. The* content *of f is the greatest common divisor of the coefficients $a_0, \ldots, a_n$, and we denote it by $\alpha(f)$.* [1]

Example 11.1.3. 1. $\alpha(2x^2 - 6) = 2$ *and* $\alpha(2x^3 - 6x + 3) = 1$.

2. *The content of a monic integer polynomial is* 1.

Using properties of the greatest common divisors, one can prove the following basic properties of content of polynomials.

Lemma 11.1.4. *Let n be a positive integer, $c_n : \mathbb{Z}[x] \to \mathbb{Z}_n[x]$ be the modulo n residue map, $a \in \mathbb{Z} \setminus \{0\}$, and suppose $f(x), g(x) \in \mathbb{Z}[x]$ are two non-zero polynomials. Then*

1. $\alpha(af(x)) = |a|\alpha(f)$.
2. *If* $\alpha(f) = d$, *then* $\frac{1}{d}f(x) \in \mathbb{Z}[x]$ *and* $\alpha(\frac{1}{d}f(x)) = 1$.
3. $n|\alpha(f)$ *if and only if* $f \in \ker c_n$.

Proof. Part one follows from the fact that

$$\gcd(aa_0, \ldots, aa_m) = |a| \gcd(a_0, \ldots, a_m).$$

The second part is equivalent to the following property of the greatest common divisor:

$$\gcd(a_0, \ldots, a_m) = d \quad \text{implies} \quad \gcd\left(\frac{a_0}{d}, \ldots, \frac{a_m}{d}\right) = 1.$$

The last part is a consequence of the following statement:

$$n|a_0, \ldots, n|a_m \quad \text{if and only if} \quad n|\gcd(a_0, \ldots, a_m).$$

□

Definition 11.1.5. *We say $f(x) \in \mathbb{Z}[x]$ is a* primitive *polynomial if $\alpha(f) = 1$.*

Lemma 11.1.6. *For every $f \in \mathbb{Z}[x]$, there is a primitive polynomial $\overline{f}$ such that*

$$f(x) = \alpha(f)\overline{f}(x).$$

Proof. This is equivalent to part (2) of Lemma 11.1.4. □

Next, we extend the definition of *content* to polynomials in $\mathbb{Q}[x]$.

Lemma 11.1.7. *For every non-zero polynomial $f(x) \in \mathbb{Q}[x]$, there are unique positive rational number q and primitive polynomial $\overline{f}$ such that $f(x) = q\overline{f}(x)$. Moreover for $f(x) \in \mathbb{Z}[x]$, $q = \alpha(f)$.*

[1] The content of f is often denoted by $c(f)$, but we use the notation c_n for the residue map modulo n. So to avoid the possible confusion, we write $\alpha(f)$ for the content of f.

Proof. (Existence) After multiplying by the common denominator n of the coefficients of f, we get an integer polynomial $\widetilde{f}(x)$; that means $\widetilde{f}(x) := nf(x) \in \mathbb{Z}[x]$. By Lemma 11.1.6, there is a primitive polynomial $\overline{f}(x)$ such that $\widetilde{f}(x) = \alpha(\widetilde{f})\overline{f}(x)$. Overall we get

$$\alpha(\widetilde{f})\overline{f}(x) = nf(x) \quad \text{which implies that} \quad f(x) = \frac{\alpha(\widetilde{f})}{n}\overline{f}(x).$$

This completes proof of existence.

Lemma 11.1.6 implies that for $f(x) \in \mathbb{Z}[x]$, we have that $q = \alpha(x)$ satisfies the desired result.

(Uniqueness) Suppose $q_1, q_2 \in \mathbb{Q}$ are positive and $q_1\overline{f}_1(x) = q_2\overline{f}_2(x)$ for some primitive polynomials $\overline{f}_1(x)$ and $\overline{f}_2(x)$. Suppose $q_i = \frac{m_i}{n}$ for some positive integers m_1, m_2 and n. Then $m_1\overline{f}_1 = m_2\overline{f}_2$, which implies that

$$m_1 = \alpha(m_1\overline{f}_1) = \alpha(m_2\overline{f}_2) = m_2.$$

Hence $q_1 = q_2$. This in turn implies that $\overline{f}_1(x) = \overline{f}_2(x)$. The existence follows. □

Definition 11.1.8. *The unique rational number given in Lemma 11.1.7 is called the* content *of f, and it is denoted by $\alpha(f)$.*

Let's point out the Part (1) of Lemma 11.1.4 holds for polynomials in $\mathbb{Q}[x]$.

Lemma 11.1.9. *For every non-zero $f(x) \in \mathbb{Q}[x]$ and $a \in \mathbb{Q} \setminus \{0\}$, we have*

$$\alpha(af(x)) = |a|\alpha(f(x)).$$

Proof. By the definition of the content, there is a primitive polynomial $\overline{f}(x)$ such that $f(x) = \alpha(f)\overline{f}(x)$. Hence $af(x) = (a\alpha(f))\overline{f(x)}$. As $\pm\overline{f}(x)$ are primitive, we deduce that $\alpha(af(x)) = |a|\alpha(f)$, which finishes the proof. □

11.2 Gauss's lemma

Gauss's lemma is the critical result that help us pass from $\mathbb{Q}$ to $\mathbb{Z}$.

Lemma 11.2.1 (Gauss's lemma, version 1)**.** *If f and g are two primitive polynomials, then fg is also primitive.*

Proof. Suppose to the contrary that $\alpha(fg) \neq 1$. Then there is a prime p which divides $\alpha(fg)$. Hence $c_p(fg) = 0$ (by Part (3) of Lemma 11.1.4). Therefore $c_p(f)c_p(g) = 0$. Notice that as $\mathbb{Z}_p$ is an integral domain, so is $\mathbb{Z}_p[x]$. Therefore $c_p(f)c_p(g) = 0$ implies that either $c_p(f) = 0$ or $c_p(g) = 0$. Another application of Part (3) of Lemma 11.1.4 gives us that either $p|\alpha(f)$ or $p|\alpha(g)$. This contradicts the assumption that both f and g are primitive. □

Lemma 11.2.2 (Gauss's lemma, version 2)**.** *Suppose f and g are two non-zero polynomials in $\mathbb{Q}[x]$. Then*

$$\alpha(fg) = \alpha(f)\alpha(g).$$

Proof. By Lemma 11.1.7, there are primitive polynomials $\overline{f}$ and $\overline{g}$ such that

$$f(x) = \alpha(f)\overline{f}(x) \quad \text{and} \quad g(x) = \alpha(g)\overline{g}(x). \tag{11.1}$$

By (11.1), we have that

$$f(x)g(x) = \alpha(f)\alpha(g)\overline{f}\overline{g}. \tag{11.2}$$

Lemma 11.1.9 implies that

$$\begin{aligned}\alpha(f(x)g(x)) =& \alpha(\alpha(f)\alpha(g)\overline{f}(x)\overline{g}(x)) \\ =& \alpha(f)\alpha(g)\alpha(\overline{f}(x)\overline{g}(x)).\end{aligned} \tag{11.3}$$

By the first version of Gauss's lemma, $\alpha(\overline{f}(x)\overline{g}(x)) = 1$. Hence (11.3) implies that

$$\alpha(fg) = \alpha(f)\alpha(g).$$

This completes the proof. ☐

11.3 Factorization: going from rationals to integers.

The following is the main result of this section, which gives us Step 1 of proof of Theorem 11.0.1. This result says that having a non-trivial decomposition in over $\mathbb{Q}$, we can get a non-trivial decomposition over $\mathbb{Z}$.

Theorem 11.3.1. *Suppose $f(x)$ is a primitive polynomial and $f(x) = \prod_{i=1}^n g_i(x)$ for some $g_i \in \mathbb{Q}[x]$. Then there are primitive polynomials $\overline{g}_i(x)$ such that*

$$g_i(x) = \alpha(g_i)\overline{g}_i(x), \quad \prod_{i=1}^n \alpha(g_i) = 1, \quad \textit{and} \quad f(x) = \prod_{i=1}^n \overline{g}_i(x).$$

Proof. By the second version of Gauss's lemma, we have

$$\alpha(f) = \alpha\Big(\prod_{i=1}^n g_i\Big) = \prod_{i=1}^n \alpha(g_i) \quad \text{which implies that} \quad \prod_{i=1}^n \alpha(g_i) = 1. \tag{11.4}$$

The last implication holds as $\alpha(f) = 1$. Next notice that by the definition of the content, there are primitive polynomials $\overline{g}_i(x)$ such that $g_i(x) = \alpha(g_i)\overline{g}_i(x)$, and so

$$\prod_{i=1}^n \overline{g}_i(x) = \prod_{i=1}^n \Big(\alpha(g_i)^{-1} g_i(x)\Big) = \Big(\prod_{i=1}^n \alpha(g_i)\Big)^{-1} \prod_{i=1}^n g_i(x) = f(x).$$

This finishes the proof. ☐

We have already pointed out that a subtle difference between being irreducible in $\mathbb{Q}[x]$ and being irreducible in $\mathbb{Z}[x]$ is having a non-trivial content. By Theorem 11.3.1, we can show that this is the only thing that one needs to be worried about:

Corollary 11.3.2. *Suppose $f(x)$ is primitive and $\deg f \geq 1$. Then $f(x)$ is irreducible in $\mathbb{Z}[x]$ if and only if it is irreducible in $\mathbb{Q}[x]$.*

Proof. We prove the contrapositive of this statement. Suppose $f(x)$ is not irreducible in $\mathbb{Q}[x]$. As $\deg f \geq 1$, not being irreducible implies that $f(x) = g_1(x)g_2(x)$ for some smaller degree polynomials $g_1, g_2 \in \mathbb{Q}[x]$. By Theorem 11.3.1, there are primitive polynomials $\overline{g}_i$ such that

$$f(x) = \overline{g}_1(x)\overline{g}_2(x) \quad \text{and} \quad \deg \overline{g}_i = \deg g_i \geq 1. \tag{11.5}$$

By (11.5), we deduce that $f(x)$ is not irreducible in $\mathbb{Z}[x]$.

Now let's assume that $f(x)$ is not irreducible in $\mathbb{Z}[x]$. Since $\deg f \geq 1$, it is not a unit. Hence not being irreducible implies that there are non-unit polynomials $h_1, h_2 \in \mathbb{Z}[x]$ such that $f(x) = h_1(x)h_2(x)$. We claim that $\deg h_i \geq 1$. Suppose to the contrary that $\deg h_i = 0$. This means that $h_i(x) = c \in \mathbb{Z}$ and $c \neq \pm 1$ (as h_i is not a unit). This implies that $c|\alpha(f)$ which contradicts the assumption that f is primitive. Hence $\deg h_i \geq 1$, and so $f(x)$ is not irreducible in $\mathbb{Q}[x]$. □

11.4 Mod criterion: irreducibility

Now we are ready to prove the mod-p irreducibility criterion (Theorem 11.0.1). We show the following slightly stronger result.

Theorem 11.4.1. *Suppose $f(x) \in \mathbb{Q}[x]$ is primitive, p is prime which does not divide the leading coefficient of $f(x)$, and $c_p : \mathbb{Z}[x] \to \mathbb{Z}_p[x]$ is the modulo p residue map. If $c_p(f(x))$ is irreducible in $\mathbb{Z}_p[x]$, then $f(x)$ is irreducible in $\mathbb{Q}[x]$.*

Proof. As it has been mentioned earlier, we show the contrapositive of this statement. So suppose $f(x)$ is not irreducible in $\mathbb{Q}[x]$. Hence $f(x)$ is either a constant polynomial or it can be written as product of two smaller degree polynomials. Since $c_p(f)$ is irreducible in $\mathbb{Z}_p[x]$, $c_p(f)$ is not constant. Hence $f(x)$ cannot be constant either. Therefore there are non-constant polynomials $g_i(x) \in \mathbb{Q}[x]$ such that $f(x) = g_1(x)g_2(x)$. As $f(x)$ is primitive, by Theorem 11.3.1 there are non-constant primitive polynomials $\overline{g}_i$ such that

$$f(x) = \overline{g}_1(x)\overline{g}_2(x). \tag{11.6}$$

This equality implies that the leading coefficient of f is the product of the leading coefficients of $\overline{g}_i$'s. Since p does not divide the leading coefficient of f, we obtain that p does not divide the leading coefficient of $\overline{g}_i$'s. Hence

$$\deg c_p(\overline{g}_i) = \deg \overline{g}_i = \deg g_i \geq 1.$$

Another application of (11.6) implies that

$$c_p(f) = c_p(\overline{g}_1)c_p(\overline{g}_2),$$

which means that $c_p(f)$ can be written as a product of two smaller degree polynomials. As $\mathbb{Z}_p$ is a field, we deduce that $c_p(f)$ is not irreducible in $\mathbb{Z}_p[x]$. This completes the proof of the contrapositive statement. □

Chapter 12

Lecture 12

In the previous lecture, we proved many important results on irreducibility of integer polynomials in $\mathbb{Q}[x]$. We proved Gauss's lemma and used to show that a monic non-constant integer polynomial can be written as a product of two non-constant primitive polynomials if and only if it is not irreducible in $\mathbb{Q}[x]$. We used this result to show the mod-p irreducibility criterion.

12.1 An example on the mod irreducibility criterion.

Later we will show that for every prime p and $a \in \mathbb{Z}_p^\times$, $x^p - x + a$ is irreducible in $\mathbb{Z}_p$. This result in combination with the mod p irreducibility criteria can be quit helpful.

Example 12.1.1. *Prove that* $f(x) := x^7 - 7x^5 + 21x^3 + 14x^2 - 8x + 11$ *is irreducible in* $\mathbb{Q}[x]$.

Proof. Notice that $f(x)$ modulo 7 is $x^7 - x + 4$. By the mentioned result, this polynomial is irreducible in $\mathbb{Z}_7[x]$. We also notice that $f(x)$ is monic, it is primitive, and the leading coefficient is not a multiple of 7. Therefore by the mod-p irreducibility criteria, $f(x)$ is irreducible in $\mathbb{Q}[x]$. $\square$

For small degree and small primes p, one can go over all the polynomials and *cross out* all the multiples of smaller degree polynomials. This way we can get the list of all the irreducible polynomials of *small* degree in $\mathbb{Z}_p[x]$. Based on the mod-p irreducibility criteria and using the list of *small* degree irreducible polynomials of $\mathbb{Z}_p[x]$, we can find lots of irreducible polynomials in $\mathbb{Q}[x]$.

Example 12.1.2. *1. Prove that* $x^4 + x + 1$ *is irreducible in* $\mathbb{Z}_2[x]$.

2. Prove that $f(x) := 5x^4 + 2x^3 - 2020x^2 + 2021x + 1$ *is irreducible in* $\mathbb{Q}[x]$.

Proof. (1) For every $a \in \mathbb{Z}_2$ and every positive integer n, we have that $a^n = a$. Hence $a^4 + a + 1 = 1$ for every $a \in \mathbb{Z}_2$. This means this polynomial does not have a degree one factor. Hence it is enough to show that it does not have a degree 2 factor. There are exactly 2^2 degree 2 polynomials in $\mathbb{Z}_2[x]$. Let's the list of them:

$x^2, x^2+1, x^2+x, x^2+x+1$. Notice that the first three have zeros in $\mathbb{Z}_2$, and so they cannot possibly be a factor of x^4+x+1. Next we use the long division and divide x^4+x+1 by x^2+x+1. We deduce that $x^4+x+1 = (x^2+x+1)(x^2+x)+1$, and so the remainder is $1 \neq 0$. Hence x^4+x+1 does not have degree 1 or 2 factors. If x^4+x+1 is not irreducible in $\mathbb{Z}_2[x]$, then it can be written as a product of two non-constant polynomials. Since the degree of these factors should add up to 4, we get deduce that one of the factors should be of degree 1 or 2. This is a contradiction. Hence x^4+x+1 is irreducible in $\mathbb{Z}_2[x]$.

(2) Notice that $f(x)$ is primitive, the leading coefficient is odd, and $f(x)$ modulo 2 is x^4+x+1 which is irreducible in $\mathbb{Z}_2[x]$. Hence by the mod-p irreducibility criterion, $f(x)$ is irreducible in $\mathbb{Q}[x]$. □

12.2 Eisenstein's irreducibility criterion

One of the most elegant irreducibility criteria is due to Eisenstein.

Theorem 12.2.1. *Let $f(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$ and p be prime. Suppose*

$$p \nmid a_n,\ p|a_{n-1}, \ldots,\ p|a_0,\ \textit{and}\ \ p^2 \nmid a_0.$$

Then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Here we start our proof in a systematic manner, but we finish it by showing an ad-hoc result. One gets a better understanding of the final stage using the Unique Factorization property of the ring of polynomials with coefficients in a field. The Unique Factorization property will be proved later in the course.

Proof of Theorem 12.2.1. Suppose to the contrary that there are non-constant polynomials $g_1, g_2 \in \mathbb{Q}[x]$ such that $f(x) = g_1(x)g_2(x)$. Then there are primitive polynomials $\overline{g}_i(x)$ such that $g_i(x) = \alpha(g_i)\overline{g}_i(x)$ (see Lemma 11.1.7), and by the second version of Gauss's lemma $\alpha(f) = \alpha(g_1 g_2) = \alpha(g_1)\alpha(g_2)$. Altogether we obtain that

$$f(x) = \alpha(f)\ \overline{g}_1(x)\overline{g}_2(x). \tag{12.1}$$

Notice that $\mathrm{ld}(f) = \alpha(f)\,\mathrm{ld}(\overline{g}_1)\,\mathrm{ld}(\overline{g}_2)$ together with the assumption that p does not divide the leading coefficient a_n imply the p does not divide $\mathrm{ld}(\overline{g}_1)$ and $\mathrm{ld}(\overline{g}_2)$. Next we look at the equation 12.1 modulo p to obtain that $c_p(f) = c_p(\alpha(f))c_p(\overline{g}_1)c_p(\overline{g}_2)$. By the assumption on the divisibility of all the non-leading coefficients by p, we deduce that

$$c_p(a_n)x^n = c_p(\alpha(f))\ c_p(\overline{g}_1)c_p(\overline{g}_2). \tag{12.2}$$

Since p does not divide $\mathrm{ld}(\overline{g}_i)$, we have that $\deg(c_p(\overline{g}_i)) = \deg(\overline{g}_i) > 0$. Equation 12.2 takes us to the following lemma:

Lemma 12.2.2. *Suppose F is a field and $\overline{g}_1, \overline{g}_2 \in F[x]$ are two non-constant polynomials such that $\overline{g}_1(x)\overline{g}_2(x) = cx^n$ for some $c \in F^\times$. Then $\overline{g}_1(0) = \overline{g}_2(0) = 0$.*

Proof. Suppose to the contrary that $\overline{g}_1(0) \neq 0$. Set

$$\overline{g}_1(x) = b_r x^r + \cdots + b_1 x + b_0 \quad \text{and} \quad \overline{g}_2(x) = c_s x^s + \cdots + c_1 x + c_0,$$

where $b_i, c_j \in F$, $b_r, c_s \in F^\times$. The contrary assumption $\overline{g}_1(0) \neq 0$ implies that $b_0 \in F^\times$. Suppose s' is the smallest non-negative integer such that $c_{s'} \neq 0$. This means

$$c_{s'} \in F^\times \quad \text{and} \quad \overline{g}_2(x) = c_s x^s + \cdots + c_{s'} x^{s'}.$$

Consider the coefficient of $x^{s'}$ in $\overline{g}_1(x)\overline{g}_2(x)$. Since every term of $\overline{g}_2(x)$ is of degree at least s', we deduce that the coefficient of $x^{s'}$ in $\overline{g}_1(x)\overline{g}_2(x)$ is $b_0 c_{s'} \neq 0$. We also notice that $s' \leq s < s + r = n$; this implies that $\overline{g}_1(x)\overline{g}_2(x)$ has at least two non-zero terms and it cannot be equal cx^n. This is a contradiction. By symmetry, we obtain that $\overline{g}_2(0) = 0$, which completes proof of Lemma. □

By Lemma 12.2.2 and (12.2), we deduce that

$$c_p(\overline{g}_1)(0) = c_p(\overline{g}_2)(0) = 0.$$

This means $p|\overline{g}_1(0)$ and $p|\overline{g}_2(0)$. Hence

$$p^2|\overline{g}_1(0)\overline{g}_2(0).$$

On the other hand, $a_0 = f(0) = \alpha(f)\overline{g}_1(0)\overline{g}_2(0)$ is a multiple of $\overline{g}_1(0)\overline{g}_2(0)$. Hence $p^2|a_0$, which is a contradiction. This completes proof of Eisenstein's irreducibility criterion. □

Example 12.2.3. *Prove that $f(x) := \frac{5}{2}x^6 - \frac{4}{3}x^3 + 7x - \frac{3}{11}$ is irreducible in $\mathbb{Q}[x]$.*

Proof. First we find the content $\alpha(f)$ and the primitive form $\overline{f}$ of f. To find the content of a polynomial first we factor out a common denominator of the coefficients, and take the greatest common divisor of the numerators of the coefficients:

$$\frac{5}{2}x^6 - \frac{4}{3}x^3 + 7x - \frac{3}{11} = \frac{1}{66}((33 \times 5)x^6 - (22 \times 4)x^3 + (66 \times 7)x - (6 \times 3))$$

So the primitive form of $f(x)$ is

$$\overline{f}(x) = (33 \times 5)x^6 - (22 \times 4)x^3 + (66 \times 7)x - (6 \times 3)$$

Notice that since $\alpha(f)$ is a unit in $\mathbb{Q}$, $f(x)$ is irreducible in $\mathbb{Q}[x]$ if and only if $\overline{f}(x)$ is irreducible in $\mathbb{Q}[x]$. Next we check that we can apply Eisenstein's irreduciblity criterion for $p = 2$, and deduce that $\overline{f}$ is irreducible in $\mathbb{Q}[x]$:

$$2 \nmid (33 \times 5),\ 2|(22 \times 4),\ 2|(66 \times 7),\ 2|(6 \times 3), \text{and } 2^2 \nmid (6 \times 3),$$

and the claim follows. □

Next we discuss a *tricky* application of Eisenstein's irreducibility criterion. As you will see, the polynomial given in the next example at the first glance has nothing to do with Eisenstein's irreducibility criterion. After applying a useful trick, however, we will get a polynomial where the hypothesis of Eisenstein's criterion clearly hold.

Example 12.2.4. *Suppose p is prime. Then $f(x) := x^{p-1}+x^{p-2}+\cdots+1$ is irreducible in $\mathbb{Q}[x]$.*

Proof. Notice that

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

and so $f(x) = \frac{x^p-1}{x-1}$. Let $g(y) := f(y+1)$. Then

$$g(y) = \frac{(y+1)^p - 1}{y} = y^{p-1} + \binom{p}{p-1} y^{p-2} + \cdots + \binom{p}{1}.$$

Notice that $p \nmid 1,\ p | \binom{p}{i}$ for every integer $i \in [1, p-1],$ and $p^2 \nmid \binom{p}{1}$. Hence by Eisenstein's irreducibility criterion, we have that $g(y)$ is irreducible in $\mathbb{Q}[y]$. Finally notice that if $f(x) = f_1(x) f_2(x)$ for two non-constant polynomials f_1 and f_2 in $\mathbb{Q}[x]$, then $f(y+1) = f_1(y+1) f_2(y+1)$, which implies that $g(y)$ can be written as a product of two non-constant polynomials in $\mathbb{Q}[y]$. This contradicts the irreducibility of $g(y)$ in $\mathbb{Q}[y]$. □

12.3 Factorization: existence, and a chain condition

Let's go back to Lemma 12.2.2, and see what really we can say about the factors of x^n. Notice that x is an irreducible element of $F[x]$, and so all the irreducible factors of x^n are x. If $F[x]$ has the *Unique Factorization* property, then all the irreducible factors of $\overline{g}_i(x)$'s are x as well. This means $\overline{g}_i = c_i x^{n_i}$ for some $c_i \in F^{\times}$ and positive integer n_i.

Definition 12.3.1. *An integral domain D is called a* Unique Factorization Domain *(UFD) if every non-zero non-unit element of D can be written as a product of irreducible elements* (the existence part)*, and the irreducible factors are unique up to reordering and multiplying by a unit* (the uniqueness part).

Example 12.3.2. *The ring of integers is a UFD. Let's understand that the flexibility given in the uniqueness part are needed. In $\mathbb{Z}$, $2, 3, -2,$ and -3 are irreducible and $2 \times 3 = (-3) \times (-2)$. Hence for the uniqueness we have to allow a reordering of the factors and a possible multiplication by units.*

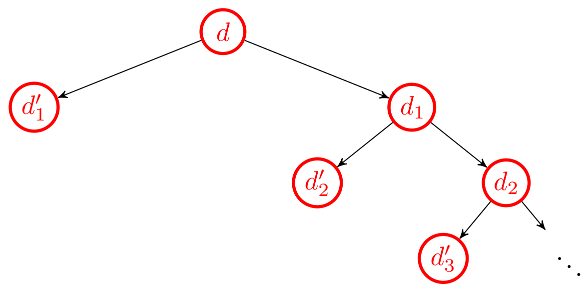
We start with investigating the *existence part* for an arbitrary integral domain D. Suppose $d \in D$ is a non-zero non-unit element. We would like to write d as a product of irreducible elements. We go through the following *pseudo-algorithm*:

1. If d is irreducible, we are done.

2. If d is not irreducible, then there are non-zero non-unit elements $d_1, d_1' \in D$ such that $d = d_1 d_1'$.

3. Repeat this process for each one of the factors.

If this process *terminates*, we end up writing d as a product of irreducible elements. Let's see what it means for this process to not terminate. We can visualize this process with a binary rooted tree, where all the vertices are labeled by non-zero non-units and label of each vertex is the product of its *children*.



Let's translate this to the language of ideals. Saying that d is a multiple of d_1 is equivalent to $\langle d\rangle \subseteq \langle d_1\rangle$. Recall that $\langle d\rangle = \langle d_1\rangle$ if and only if $d = ud_1$ for some $u \in D^\times$ (see Lemma 9.2.4). Hence $\langle d\rangle = \langle d_1\rangle$ if and only if $d_1 u = d_1 d'_1$. By the cancellation law and d'_1 not being a unit, we deduce that we have an infinite *ascending chain* of (principal) ideals:

$$\langle d\rangle \subsetneq \langle d_1\rangle \subsetneq \langle d_2\rangle \cdots .$$

This takes us to the definition of Noetherian rings.

Definition 12.3.3. *A ring A is called* Noetherian *if there is no infinite ascending chain of ideals. That means if $I_1 \subseteq I_2 \subseteq \cdots$ is an ascending chain of ideals of A, then for some positive integer n_0 we have $I_{n_0} = I_{n_0+1} = \cdots$.*

The above discussion on the existence of a factorization into irreducible elements immediately gives us the following result.

Proposition 12.3.4. *Suppose D is a Noetherian integral domain. Then every non-zero non-unit element of D can be written as a product of irreducible elements of D.*

Proposition 12.3.4 would not be a satisfactory result unless we have an effect way of determining whether or not an integral domain is Noetherian.

Lemma 12.3.5. *Suppose A is a unital commutative ring. Then A is Noetherian ring if and only if every ideal of A is finitely generated.*

(An ideal I is called *finitely generated* if there is a finite set $\{a_1, \ldots, a_n\}$ such that $I = \langle a_1, \ldots, a_n\rangle$ (see Lemma 5.2.3).)

Proof. ($\Rightarrow$) Suppose to the contrary that there is an ideal I which is not finitely generated. Inductively we define a sequence of elements $\{a_i\}_{i=1}^{\infty}$ of I such that

$$\langle a_1\rangle \subsetneq \langle a_1, a_2\rangle \subsetneq \cdots ,$$

which contradicts the assumption that A is Noetherian. Let a_1 be an element of I. Since I is not finitely generated, $\langle a_1\rangle$ is a proper subset of I. Hence there is $a_2 \in I \setminus \langle a_1\rangle$. Again, as I is not finitely generate, $\langle a_1, a_2\rangle$ is a proper ideal of I. Therefore there is $a_3 \in I \setminus \langle a_1, a_2\rangle$. We continue this process inductively, and the proof can be completed as it is explained above.

($\Leftarrow$) Suppose $I_1 \subseteq I_2 \subseteq \cdots$ is an ascending chain of ideals of A. Consider $I := \bigcup_{i=1}^{\infty} I_i$. Next we prove that I is an ideal of A.

For every $a, a' \in I$, there are positive integers i and i' such that $a \in I_i$ and $a' \in I_{i'}$. Without loss of generality we can and will assume that $i \le i'$, and so $I_i \subseteq I_{i'}$. Therefore $a, a' \in I_{i'}$. Hence $a - a' \in I_{i'}$, which implies that $a - a' \in I$.

For every $a \in I$, there is a positive integer i such that $a \in I_i$. Hence for every $r \in A$, we have that $ra \in I_i$, which implies that $ra \in I$. This completes the proof of the claim that I is an ideal.

Since I is an ideal, it is finitely generated. Hence there are $a_1, \ldots, a_n \in I$ such that $I = \langle a_1, \ldots, a_n\rangle$. Notice that $a_i \in I$ implies that $a_i \in I_{k_i}$ for some positive integer n_i. Suppose $m := \max\{k_1, \ldots, k_n\}$. Then I_m contains I_{k_i} for every i. Therefore $a_1, \ldots, a_n \in I_m$. This implies that

$$\langle a_1, \ldots, a_n\rangle \subseteq I_m.$$

Hence for every $j \ge m$, we have

$$I_j \subseteq \bigcup_{i=1}^{\infty} I_i = \langle a_1, \ldots, a_n\rangle \subseteq I_m \subseteq I_j. \tag{12.3}$$

By (12.3), we obtain that $I_m = I_j$ for every $j \ge m$. This means that A is Noetherian. □

We immediately deduce that a PID is Noetherian, and so every non-zero non-unit element can be factored into irreducible elements.

Corollary 12.3.6. *Suppose D is a PID. Then D is Noetherian and every non-zero non-unit element of D can be written as a product of irreducible elements.*

Proof. Since D is a PID, every ideal is principal. Hence every ideal is finitely generated. Therefore by Lemma 12.3.5, D is Noetherian. By Proposition 12.3.4, we obtain that every non-zero non-unit element of D can be written as a product of irreducible elements. □

Chapter 13

Lecture 13

In the previous lecture we said an integral domain is called a unique factorization domain if every non-zero non-unit element can written as a product of irreducible elements (the existence part) and the irreducible factors are unique up to reordering and multiplying by units (the uniqueness part). We showed that the existence part holds in a Noetherian integral domain (see Proposition 12.3.4 together with Lemma 12.3.5). Today we will investigate the uniqueness part.

13.1 Factorization: uniqueness, and prime elements.

Let's first formulate what the *uniqueness* precisely means: suppose $p_1, \ldots, p_m$ and $q_1, \ldots, q_n$ are irreducible elements of D. If

$$p_1 \cdots p_m = q_1 \cdots q_n, \tag{13.1}$$

then $p_1 = u_1 q_{i_1}$, $p_2 = u_2 q_{i_2}$, and so on, for some $u_i \in D^\times$ and a permutation $1 \mapsto i_1, \ldots, m \mapsto i_m$ of $1, 2, \ldots, n$; in particular $m = n$. This means we need to show if an irreducible element p divides a product of irreducible elements q_i's, then $p = uq_j$ for some unit u and some index j. This takes us to the definition of *prime* elements.

Definition 13.1.1. *Suppose D is an integral domain.*

1. *For $a, b \in D$, we say a divides b and write $a|b$ if there is $d \in D$ such that $b = ad$*

2. *A non-zero non-unit element p of D is called* prime *when for every $a, b \in D$, if $p|ab$, then either $p|a$ or $p|b$.*

Base on the above discussion, for uniqueness to hold, we need to have that every irreducible element is prime. Next we show this statement and its converse hold.

Proposition 13.1.2. *Let D be an integral domain. Suppose every non-zero non-unit element of D can be written as a product of irreducible elements. Then D is a UFD if and only if every irreducible element is prime.*

The formal proof has many little details that make the proof a bit hard to digest. The idea of proof, however, is rather simple. For that reason first I write an outline of the proof:

Outline of proof. ($\Rightarrow$) Suppose p is irreducible and $p|ab$. Then $ab = pd$ for some $d \in D$. We decompose a, b, and d into irreducible factors. We notice that p is an irreducible factor of the left hand side, and so by the uniqueness of irreducible factors, it should be an irreducible factor of either a or b. This means that either $p|a$ or $p|b$.

($\Leftarrow$) The existence part is given as an assumption. So we focus on the uniqueness part. Starting with $p_1 \cdots p_m = q_1 \cdots q_n$, using the assumption that p_1 is prime, we can find an index i_1 such that $p_1|q_{i_1}$. As q_{i_1} is irreducible, we can deduce that p_1 is q_{i_1} upto multiplying by a unit. Now we cancel out p_1 and continue by induction on the number of involved irreducible factors.

Proof. ($\Rightarrow$) Suppose p is an irreducible element. We have to show that p is prime. Since p is irreducible, it is not either zero or unit. Now suppose for $a, b \in D$, $p|ab$. Notice that if either $a = 0$ or $b = 0$, we are done as $p|0$. So without loss of generality we can and will assume that a and b are non-zero. By the assumption either a is a unit or it can be written as product of irreducible elements. A similar statement holds for b. Suppose $a = uq_1 \cdots q_m$ and $b = u'q_{m+1} \cdots q_n$ for irreducible elements $q_1, \ldots, q_n$ and units u, u'. Then $p|(uu' \prod_{i=1}^n q_i)$. This means there is $d \in D$ such that $pd = uu' \prod_{i=1}^n q_i$. Notice that the right hand side of this equation cannot be zero, and so $d \neq 0$. Therefore $d = u''\ell_1 \cdots \ell_k$ for some irreducible elements $\ell_1, \ldots, \ell_k$ and a unit u''. Hence

$$u''p\ell_1 \cdots \ell_k = uu'q_1 \cdots q_n. \tag{13.2}$$

Since p is not a unit, the right hand side of Equation 13.2 cannot be a unit. Therefore $n \geq 1$. As p and q_1 are irreducible, so are $u''p$ and $uu'q_1$. By the assumption the irreducible elements $u''p, \ell_1, \cdots, \ell_k$ are the same as $uu'q_1, \ldots, q_n$ upto reordering and multiplying by units. Hence there is a unit $\overline{u}$ and an index j such that

$$p = \overline{u}q_j. \tag{13.3}$$

Notice that, if $j \leq m$, then $\overline{u}q_j|a$, and if $j > m$, then $\overline{u}q_j|b$. Therefore by (13.3), we obtain that either $p|a$ or $p|b$. This shows that p is prime.

($\Leftarrow$) By the assumption every non-zero non-unit element can be written as a product of irreducible elements. So we focus on the uniqueness part. Suppose $p_1, \ldots, p_m$ and $q_1, \ldots, q_n$ are irreducible elements and

$$p_1 \cdots p_m = q_1 \cdots q_n. \tag{13.4}$$

We have to show that $m = n$, there is a reordering $i_1, \ldots, i_m$ of $1, \ldots, m$, and units u_j such that $p_j = u_j q_{i_j}$ for every j.

We proceed by induction on n. By (13.4), we have that p_1 divides $q_1 \cdots q_n$. Since every irreducible element is prime, p_1 is prime. Whenever a prime element divides product of certain elements, it should divide one of them. Hence there is an index i_1 such that $p_1|q_{i_1}$. This means $q_{i_1} = p_1 u_1$ for some $u_1 \in D$. Since q_{i_1} is irreducible, either p_1 is a unit or u_1 is a unit. As p_1 is irreducible, it is not a unit. Hence u_1 is a unit. Overall we showed that there are an index i_1 and a unit u_1 such that $q_{i_1} = u_1 p_1$. This implies that

$$p_1 \cdots p_m = u_1 p_1 q_1 \cdots q_{i_1-1} q_{i_1+1} \cdots q_n,$$

and so by the cancellation law, we obtain

$$p_2 \cdots p_m = u_1 q_1 \cdots q_{i_1-1} q_{i_1+1} \cdots q_n. \tag{13.5}$$

If $m = 1$, the left hand side is 1. Hence all the terms in the right hand side are units. This means $n = 1$, and we are done. For $m \geq 2$, the left hand side is not a unit, and so $n \neq 1$. Hence there is q_{j_0} factor in the right hand side of (13.5). Then $u_1 q_{j_0}$ is also irreducible. By the induction hypothesis, we deduce that $m - 1 = n - 1$, and there are a reordering $i_2, \ldots, i_m$ of $1, \ldots, i_1 - 1, i_1 + 1, \ldots, m$, and units u_j for every index $j \in [2, m]$ such that $q_{i_j} = u_j p_j$. This finishes the proof. □

13.2 Prime elements and prime ideals

In this section we investigate prime elements. We have seen that in an integral domain an element p is irreducible if and only if the ideal generated by p is maximal among proper principal ideals (see Lemma 9.2.7). As we want to understand the connection between prime and irreducible elements, we study properties of the principal ideals that generated by prime elements. By the definition, p is a prime element of an integral domain D if (1) p is not either zero or unit, and (2) for every $a, b \in D$, if $p|ab$, then either $p|a$ or $p|b$. We start with translating the concept of divisibility to the language of ideals.

Lemma 13.2.1. *Suppose D is an integral domain, and $a, b \in D$.*

1. *$a|b$ if and only if $b \in \langle a \rangle$ if and only if $\langle b \rangle \subseteq \langle a \rangle$.*
2. *$a|b$ and $b|a$ if and only if $a = bu$ for some unit u.*

Proof. We have that $a|b$ if and only if $b = ac$ for some $c \in D$. Since

$$\langle a \rangle = \{ar \mid r \in D\}$$

(see Lemma 5.2.3), the claim follows.

By the first part, we have $a|b$ and $b|a$ if and only if $\langle a \rangle = \langle b \rangle$. The latter happens if and only if $a = bu$ for some unit u (see Lemma 9.2.4). □

By Lemma 13.2.1, we have that $p \in D$ is prime if and only if (1) $\langle p \rangle$ is a non-zero proper ideal (see Lemma 9.2.5) and (2) if $ab \in \langle p \rangle$, then either $a \in \langle p \rangle$ or $b \in \langle p \rangle$. This takes us to the definition of prime ideals.

Definition 13.2.2. *Suppose A is a unital commutative ring and I is an ideal of A. we say I is a prime ideal if (1) I is proper (that means $I \neq A$), and (2) if $ab \in I$ for some $a, b \in D$, then either $a \in I$ or $b \in I$.*

Hence we immediately deduce the following interpretation of prime elements in the language of principal ideals:

Lemma 13.2.3. *Suppose D is an integral domain and $p \in D$. Then p is a prime element if and only if $p \neq 0$ and $\langle p \rangle$ is a prime ideal.*

We have seen that an ideal I in a unital commutative ring is maximal if and only if the quotient ring A/I is a field (see Proposition9.3.3). Next we understand when an ideal is prime in terms of the corresponding quotient ring.

Lemma 13.2.4. *Suppose A is a unital commutative ring and I is an ideal of A. Then I is a prime ideal if and only if A/I is an integral domain.*

Proof. ($\Rightarrow$) Since I is a proper ideal, A/I is a non-trivial ring. Next we show that A/I does not have a zero-divisor. Suppose $(a+I)(b+I) = 0+I$ for some $a, b \in A$. This means that $ab \in I$. As I is a prime ideal, either $a \in I$ or $b \in I$. From this we deduce that either $a+I = 0+I$ or $b+I = 0+I$. Hence A/I is an integral domain.

($\Leftarrow$) Since A/I is an integral domain, A/I is a non-trivial ring. Therefore I is a proper ideal. Now suppose $ab \in I$. Then $(a+I)(b+I) = 0+I$. Since A/I is an integral domain, we have that either $a+I = 0+I$ or $b+I = 0+I$. Hence either $a \in I$ or $b \in I$. Altogether, we deduce that I is a prime ideal. $\square$

We immediately obtain that every maximal ideal is prime.

Corollary 13.2.5. *Suppose A is a unital commutative ring and I is an ideal of A. If I is a maximal ideal, then I is a prime ideal.*

Proof. Suppose I is a maximal ideal. Then A/I is a field (see Proposition 9.3.3). Hence A/I is an integral domain, which implies that I is a prime ideal (by Lemma 13.2.4). $\square$

13.3 Prime vs irreducible

Next we investigate the connections between prime and irreducible elements. In view of Proposition 13.1.2, such a connection can help us prove that certain integral domains are UFD.

Lemma 13.3.1. *Suppose D is a PID. Then every irreducible element of D is prime.*

Proof. Suppose p is irreducible in D. Then by Lemma 9.3.2, $\langle p \rangle$ is a maximal ideal of D. Therefore by Corollary 13.2.5, $\langle p \rangle$ is a prime ideal. Since $p \neq 0$ (as p is irreducible) and $\langle p \rangle$ is a prime ideal, by Lemma 13.2.3 we deduce that p is a prime element. $\square$

The converse of Lemma 13.3.1 is true in every integral domain.

Lemma 13.3.2. *Suppose D is an integral domain and $p \in D$. If p is a prime element, then p is irreducible.*

Proof. Since p is prime, it is not either zero or unit. Hence to show it is irreducible, we have to argue why $p = ab$ implies that either a is a unit or b is a unit.

For $a, b \in D$ suppose $p = ab$. Since p is prime and $p|ab$, we deduce that either $p|a$ or $p|b$. This means that either $a = pa'$ for some $a' \in D$ or $b = pb'$ for some $b' \in D$. In the former case, we have

$$a = pa' = aba' \text{ which implies that } ba' = 1. \tag{13.6}$$

(Notice that since p is prime, it is not zero. Hence a and b are not zero, and so we are allowed to use the cancellation law.) By (13.6), we obtain that b is a unit. Similarly we can show that $b = pb'$ implies that a is a unit. Altogether we have that $p = ab$ implies that either a is a unit or b is unit. This completes this proof. □

An immediate consequence of the above lemmas is the following theorem.

Theorem 13.3.3. *Suppose D is a PID. Then*

1. *An element $d \in D$ is irreducible if and only if it is prime.*

2. *D is a UFD.*

Proof. Since D is an integral domain, by Lemma 13.3.2 every prime is irreducible. Since D is a PID, by Lemma 13.3.1 every irreducible is prime.

The existence part of being a UFD follows from Corollary 12.3.6. The Uniqueness part of being a UFD follows from the first part and Proposition 13.1.2. □

As a corollary we deduce the following:

Theorem 13.3.4. *The following rings are UFD: $\mathbb{Z}$, $F[x]$ where F is a field, $\mathbb{Z}[i]$, and $\mathbb{Z}[\omega]$ where $\omega := \frac{-1+\sqrt{-3}}{2}$.*

Proof. We have proved that all of these rings are Euclidean domains. This implies that they are PIDs. Hence they are UFDs. □

13.4 Some integral domains that are not UFD.

We have seen some interesting examples that are UFDs. Now we want to see that there are many interesting integral domains that are not UFDs.

Example 13.4.1. *The ring $\mathbb{Z}[\sqrt{-6}] := \{a + b\sqrt{-6} \mid a, b \in \mathbb{Z}\}$ is not a UFD.*

Proof. By Proposition 13.1.2, it is enough to find an irreducible element which is not a prime element. To show an element is irreducible, first we have to prove it is not a unit. Therefore we have to describe units of this ring. Let

$$N : \mathbb{Z}[\sqrt{-6}] \to \mathbb{Z}, \quad N(z) := |z|^2.$$

Notice that $N(z_1 z_2) = N(z_1)N(z_2)$ for every $z_1, z_2 \in \mathbb{Z}[\sqrt{-6}]$.

Claim 1. *$z \in \mathbb{Z}[\sqrt{-6}]$ is a unit if and only if $N(z) = 1$.*

Proof of Claim 1. (⇒) Since $z \in \mathbb{Z}[\sqrt{-6}]^\times$, there is $z' \in \mathbb{Z}[\sqrt{-6}]$ such that $zz' = 1$. Hence

$$N(zz') = 1 \text{ which implies that } N(z)N(z') = 1.$$

Therefore $N(z) \in \mathbb{Z}^\times = \{\pm 1\}$. Since $N(z)$ is non-negative, we deduce that $N(z) = 1$.

(⇐) Suppose $N(z) = 1$. and $x = a + b\sqrt{-6}$. Then

$$(a + b\sqrt{6})(a - b\sqrt{6}) = 1,$$

which implies that $x = a + b\sqrt{-6} \in \mathbb{Z}[\sqrt{-6}]^\times$ as $a - b\sqrt{-6} \in \mathbb{Z}[\sqrt{-6}]$. This completes the proof of Claim 1.

Claim 2. *$\sqrt{-6}$ is irreducible in $\mathbb{Z}[\sqrt{-6}]$.*

Proof of Claim 2. Since $N(\sqrt{-6}) = 6 \neq 1$, by Claim 1, $\sqrt{-6}$ is not a unit. Now suppose $\sqrt{-6} = xy$ for some $x, y \in \mathbb{Z}[\sqrt{-6}]$. Then

$$N(\sqrt{-6}) = N(xy) \text{ which implies that } 6 = N(x)N(y). \tag{13.7}$$

If neither x nor y are units, by Claim 1 and (13.7) we have that either $N(x) = 2$ or $N(y) = 2$. This means the next claim completes the proof of Claim 2.

Claim 3. *There is no $x \in \mathbb{Z}[\sqrt{-6}]$ such that $N(x) = 2$.*

Proof of Claim 3. Suppose $N(a + b\sqrt{-6}) = 2$ for some $a, b \in \mathbb{Z}$. Then

$$a^2 + 6b^2 = 2. \tag{13.8}$$

If $b \neq 0$, then $6b^2 \geq 6$. This implies that $a^2 + 6b^2 \geq 6$, which means (13.8) cannot hold. Hence $b = 0$, in which case (13.8) implies that $b^2 = 2$, which is not possible as $\sqrt{2}$ is irrational.

Claim 4. *$\sqrt{-6}$ is not prime in $\mathbb{Z}[\sqrt{-6}]$.*

Proof of Claim 4. Suppose to the contrary that $\sqrt{-6}$ is prime. Then $\sqrt{-6}|2 \times 3$ implies that either $\sqrt{-6}|2$ or $\sqrt{-6}|3$. This means there is $z \in \mathbb{Z}[\sqrt{-6}]$ such that either $z\sqrt{-6} = 2$ or $z\sqrt{-6} = 3$. Comparing the norms of both sides, we obtain that either $6N(z) = 4$ or $6N(z) = 9$. This is a contradiction as $6 \nmid 4$ and $6 \nmid 9$.

Altogether, we found an irreducible element which is not prime, and so $\mathbb{Z}[\sqrt{-6}]$ is not a UFD. □

Chapter 14

Lecture 14

We have proved that

$$\text{Euclidean Domain} \quad \Rightarrow \quad \text{PID} \quad \Rightarrow \quad \text{UFD}.$$

We have also showed a method to works with rings of the form $\mathbb{Z}[\alpha]$ where α is a zero of a monic integer quadratic polynomial. We argued how using a norm function sometimes we can find elements that are irreducible but not prime, and deduce that the given integral domain is not a UFD.

14.1 Ring of integer polynomials is a UFD.

Next we show that $\mathbb{Z}[x]$ is a UFD. Remember that this is not a PID as the ideal $\langle 2, x\rangle$ is not a principal ideal of $\mathbb{Z}[x]$.

Theorem 14.1.1. *The ring $\mathbb{Z}[x]$ is a UFD.*

There are three main ingredients in the proof:

1. $\mathbb{Z}$ is a UFD,
2. $\mathbb{Q}[x]$ is a UFD, and
3. Irreducibility of a polynomial in $\mathbb{Q}[x]$ is equivalent to the irreducibility of the primitive form the polynomial in $\mathbb{Z}[x]$ (Gauss's lemma).

Lemma 14.1.2. *Suppose $c \in \mathbb{Z}$. Then we have that*

1. *c is irreducible in $\mathbb{Z}$ if and only if it is irreducible in $\mathbb{Z}[x]$.*
2. *c is prime in $\mathbb{Z}$ if and only if it is prime in $\mathbb{Z}[x]$.*

Proof. (1) ($\Rightarrow$) Since c is irreducible in $\mathbb{Z}$, it is not 0 or ± 1. As $\mathbb{Z}[x]^{\times} = \{\pm 1\}$, we deduce that c is not zero or a unit in $\mathbb{Z}[x]$. Now suppose $c = f(x)g(x)$. Comparing the degrees of both sides, we deduce that $f(x) = a \in \mathbb{Z}$ and $g(x) = b \in \mathbb{Z}$. As c is irreducible in $\mathbb{Z}$, $c = ab$ implies that either $a = \pm 1$ or $b = \pm 1$. Therefore either $f(x)$ is a unit or $g(x)$ is a unit. This means c is irreducible in $\mathbb{Z}[x]$.

($\Leftarrow$) As c is irreducible in $\mathbb{Z}[x]$, c is not zero or ± 1. Hence c is a non-zero non-unit element of $\mathbb{Z}$. Suppose $c = ab$ for some $a, b \in \mathbb{Z}$. Then either $a \in \mathbb{Z}[x]^\times$ or $b \in \mathbb{Z}[x]^\times$. Since $\mathbb{Z}[x]^\times = \mathbb{Z}^\times$, we deduce that either $a \in \mathbb{Z}^\times$ or $b \in \mathbb{Z}^\times$. Hence c is irreducible in $\mathbb{Z}$.

(2) ($\Rightarrow$) Suppose $c|f(x)g(x)$ for some $f, g \in \mathbb{Z}[x]$. Then there is $q(x) \in \mathbb{Z}[x]$ such that $cq(x) = f(x)g(x)$. Hence $|c|\alpha(q) = \alpha(f)\alpha(g)$, which implies that $c|\alpha(f)\alpha(g)$. Since c is prime in $\mathbb{Z}$, we have that either $c|\alpha(f)$ or $c|\alpha(g)$. As $\alpha(f)|f(x)$ and $\alpha(g)|g(x)$ in $\mathbb{Z}[x]$, we deduce that either $c|f(x)$ or $c|g(x)$.

($\Leftarrow$) Suppose $c|ab$ for some integers a and b. Viewing a and b as constant polynomials, as c is prime in $\mathbb{Z}[x]$, we deduce that either $c|a$ or $c|b$ in $\mathbb{Z}[x]$. This means for some $f(x) \in \mathbb{Z}[x]$ we have that either $cf(x) = a$ or $cf(x) = b$. Comparing the degrees, we deduce that $f(x) \in \mathbb{Z}$. Hence $c|a$ or $c|b$ in $\mathbb{Z}$. This means c is prime in $\mathbb{Z}$. □

Next we show that the primitive form $\overline{f}(x)$ of a polynomial $f(x)$ in $\mathbb{Q}[x]$ captures the divisibility properties of $f(x)$ in $\mathbb{Q}[x]$.

Let's recall that for every non-zero polynomial $f(x) \in \mathbb{Q}[x]$, there is a unique primitive polynomial $\overline{f}(x) \in \mathbb{Z}[x]$ such that $f(x) = \alpha(f)\overline{f}(x)$ where $\alpha(f) \in \mathbb{Q}^\times$ is the content of f.

Proposition 14.1.3. *Suppose $f, g \in \mathbb{Q}[x]$ are two non-zero polynomials, and $\overline{f}(x), \overline{g}(x) \in \mathbb{Z}[x]$ are their primitive forms, respectively.*

1. *We have that $f \in \mathbb{Q}[x]^\times$ if and only if $\overline{f}(x) \in \mathbb{Z}[x]^\times$.*
2. *We have that $f|g$ in $\mathbb{Q}[x]$ if and only if $\overline{f}|\overline{g}$ in $\mathbb{Z}[x]$.*
3. *We have that f is irreducible in $\mathbb{Q}[x]$ if and only if $\overline{f}$ is irreducible in $\mathbb{Z}[x]$.*
4. *We have that f is prime in $\mathbb{Q}[x]$ if and only if $\overline{f}$ is prime in $\mathbb{Z}[x]$.*

Proof. (1) $f(x) \in \mathbb{Q}[x]^\times$ if and only if $f(x) = c \in \mathbb{Q}^\times$. The latter occurs if and only if $f(x) = \pm\alpha(f)$. Notice that $f(x) = \pm\alpha(f)$ precisely when $\overline{f}(x) = \pm 1$. Altogether we have that $f(x) \in \mathbb{Q}[x]^\times$ if and only if $\overline{f}(x) \in \mathbb{Z}[x]^\times$.

(2) ($\Rightarrow$) Since $f(x)|g(x)$ in $\mathbb{Q}[x]$, there is a polynomial $q(x) \in \mathbb{Q}[x]$ such that $g(x) = f(x)q(x)$. Let $\overline{q}(x)$ be the primitive form of $q(x)$. Then

$$\alpha(g)\overline{g}(x) = \alpha(f)\overline{f}(x)\alpha(q)\overline{q}(x). \tag{14.1}$$

By Gauss's lemma, we have

$$\alpha(g) = \alpha(f)\alpha(q) = \alpha(q). \tag{14.2}$$

By (14.1) and (14.2), we deduce that $\overline{g}(x) = \overline{f}(x)\overline{q}(x)$, which implies that $\overline{f}|\overline{g}$ in $\mathbb{Z}[x]$.

($\Leftarrow$) Since $\overline{f}|\overline{g}$ in $\mathbb{Z}[x]$, there is a polynomial $h(x) \in \mathbb{Z}[x]$ such that $\overline{g}(x) = \overline{f}(x)h(x)$. Hence

$$g(x) = \alpha(g)\overline{g}(x) = \alpha(g)\overline{f}(x)h(x) = \underbrace{(\alpha(g)\alpha(f)^{-1}h(x))}_{\text{is in } \mathbb{Q}[x]} f(x),$$

which implies that $f(x)|g(x)$ in $\mathbb{Q}[x]$.

(3) Since $f(x) = \alpha(f)\overline{f}(x)$ and $\alpha(f) \in \mathbb{Q}[x]^\times$, $f(x)$ is irreducible in $\mathbb{Q}[x]$ if and only if $\overline{f}(x)$ is irreducible in $\mathbb{Q}[x]$. By part (1) we can assume that $\deg f \geq 1$. By Corollary 11.3.2, we have that $\overline{f}(x)$ is irreducible in $\mathbb{Q}[x]$ precisely when it is irreducible in $\mathbb{Z}[x]$. This finishes the proof.

(4) ($\Rightarrow$) Suppose $\overline{f}|h_1(x)h_2(x)$ for some $h_1, h_2 \in \mathbb{Z}[x]$. This means that there is $q(x) \in \mathbb{Z}[x]$ such that $h_1(x)h_2(x) = \overline{f}(x)q(x) = (\alpha(f)^{-1}q(x))f(x)$. Hence $f(x)|h_1(x)h_2(x)$ in $\mathbb{Q}[x]$. Since f is prime in $\mathbb{Q}[x]$, we deduce that either $f(x)|h_1(x)$ in $\mathbb{Q}[x]$ or $f(x)|h_2(x)$ in $\mathbb{Q}[x]$. By part (2), we have that either $\overline{f}|\overline{h}_1$ in $\mathbb{Z}[x]$ or $\overline{f}|\overline{h}_2$ in $\mathbb{Z}[x]$. Notice that $\overline{h}_i|h_i$ in $\mathbb{Z}[x]$. Altogether we obtain that either $\overline{f}|h_1$ in $\mathbb{Z}[x]$ or $\overline{f}|h_2$. This means that $\overline{f}$ is prime in $\mathbb{Z}[x]$.

($\Leftarrow$) Suppose $f|g_1g_2$ for some $g_1, g_2 \in \mathbb{Q}[x]$. By part (2), we deduce that $\overline{f}$ divides the primitive form of g_1g_2 in $\mathbb{Z}[x]$. By Gauss's lemma, we have that the primitive form of g_1g_2 is the product of the primitive forms of g_1 and g_2. Hence $\overline{f}|\overline{g}_1\overline{g}_2$ in $\mathbb{Z}[x]$. Since $\overline{f}$ is prime in $\mathbb{Z}[x]$, either $\overline{f}|\overline{g}_1$ in $\mathbb{Z}[x]$ or $\overline{f}|\overline{g}_2$ in $\mathbb{Z}[x]$. Another application of part (2) implies that either $f|g_1$ in $\mathbb{Q}[x]$ or $f|g_2$ in $\mathbb{Q}[x]$. This means that f is prime in $\mathbb{Q}[x]$. □

Proof of Theorem 14.1.1. **Existence part.** Suppose $f(x) \in \mathbb{Z}[x]$ is a non-zero non-unit polynomial. We have to show that we can write $f(x)$ as a product of irreducible elements. If $f(x)$ is a constant function, then $f(x) = a \in \mathbb{Z}$. As $\mathbb{Z}$ is a UFD, a can be written as a product of irreducible elements of $\mathbb{Z}$. By Lemma 14.1.2, irreducible elements of $\mathbb{Z}$ are also irreducible in $\mathbb{Z}[x]$. Hence $f(x)$ can be written as a product of irreducible elements of $\mathbb{Z}[x]$.

Next we assume that $f(x)$ is not a constant polynomial and consider its primitive form $\overline{f}(x)$. Hence $f(x) = \alpha(f)\overline{f}(x)$, where $\alpha(f)$ is the content of f. Notice that $\alpha(f) \in \mathbb{Z}$ can be viewed as a constant polynomial, and so it can be written as a product of irreducible elements of $\mathbb{Z}[x]$ (unless it is 1). Next we view $\overline{f}(x)$ as a non-constant polynomial in $\mathbb{Q}[x]$. Since $\mathbb{Q}[x]$ is a UFD, $\overline{f}(x)$ can be written as a product of irreducible elements of $\mathbb{Q}[x]$. Say $p_i(x) \in \mathbb{Q}[x]$ are irreducible and $\overline{f}(x) = \prod_{i=1}^n p_i(x)$. Suppose $\overline{p}_i(x)$ is the primitive form of $p_i(x)$. By Theorem 11.3.1, we have

$$\overline{f}(x) = \prod_{i=1}^{n} \overline{p}_i(x). \tag{14.3}$$

By Proposition 14.1.3, part (3), we have that $\overline{p}_i$'s are irreducible in $\mathbb{Z}[x]$.

Altogether we end up getting a factorization of $f(x)$ into irreducible elements of $\mathbb{Z}[x]$.

Uniqueness part. By Proposition 13.1.2, it is sufficient to show that every irreducible element of $\mathbb{Z}[x]$ is prime. Suppose $f(x) \in \mathbb{Z}[x]$ is irreducible. The decomposition $f(x) = \alpha(f)\overline{f}(x)$ implies that either $f(x)$ is a constant polynomial or it is primitive and $f(x) = \overline{f}(x)$.

Case 1. $f(x) = a$ *is constant.*

By Lemma 14.1.2, part (1), a is irreducible in $\mathbb{Z}$. Since $\mathbb{Z}$ is a UFD, a is prime in $\mathbb{Z}$. Hence by Lemma 14.1.2, part (2), $f(x) = a$ is prime in $\mathbb{Z}[x]$.

Case 2. $f(x) = \overline{f}(x)$ *is primitive.*

Since $\overline{f}(x)$ is irreducible in $\mathbb{Z}[x]$, by Proposition 14.1.3 part (3), $f(x)$ is irreducible in $\mathbb{Q}[x]$. Since $\mathbb{Q}[x]$ is a UFD and $f(x)$ is irreducible in $\mathbb{Q}[x]$, $f(x)$ is prime in $\mathbb{Q}[x]$. By Proposition 14.1.3 part (4), $\overline{f}$ is prime in $\mathbb{Z}[x]$. This means $f(x) = \overline{f}(x)$ is prime in $\mathbb{Z}[x]$, which finishes the proof. □

Theorem 14.1.1 is a special case of the following theorem:

Theorem 14.1.4. *Suppose D is a UFD. Then $D[x]$ is a UFD.*

Going through the main ingredients of the above proof, we notice that we have to use the field of fractions $F := Q(D)$ of D. As $F[x]$ is a PID, we know that it is a UFD. So if we manage to define a *primitive* form of a non-zero polynomial $Q(D)[x]$ with properties as in Proposition 14.1.3, we can go through the above proof and show that Theorem 14.1.4 holds.

To define a primitive form of polynomials in $Q(D)[x]$, following the case of integer polynomials, we need to define *the greatest common divisor* of finitely many elements of a UFD D.

Proposition 14.1.5. *Suppose D is a UFD. Then for non-zero elements $a_1, \ldots, a_n$ there is $d \in D$ with the following properties:*

1. $d|a_1, \ldots, d|a_n$.
2. *If $d'|a_1, \ldots, d'|a_n$, then $d'|d$.*

If d_1 and d_2 satisfy the above properties, then $d_1 = ud_2$ for some $u \in D^{\times}$.

An element $d \in D$ which satisfies the above properties is called a greatest common divisor *of $a_1, \ldots, a_n$.*

Chapter 15

Lecture 15

In the previous lecture we proved that $\mathbb{Z}[x]$ is a UFD, and mentioned that in general $D[x]$ is a UFD if D is a UFD. We pointed out the missing ingredient in proving this general statement is a generalization of Gauss's lemma in the context of UFDs. In order to formulate this general form, we need to know what *greatest common divisor* mean in a UFD.

15.1 Valuations and greatest common divisors in a UFD

We prove the following result and use it to define a greatest common divisor of finitely many elements of a UFD.

Proposition 15.1.1. *Suppose D is a UFD. Then for non-zero elements $a_1, \ldots, a_n$ there is $d \in D$ with the following properties:*

1. $d|a_1, \ldots, d|a_n$.
2. *If $d'|a_1, \ldots, d'|a_n$, then $d'|d$.*

If d_1 and d_2 satisfy the above properties, then $d_1 = ud_2$ for some $u \in D^{\times}$.

An element $d \in D$ which satisfies the above properties is called a greatest common divisor *of $a_1, \ldots, a_n$.*

We start by recalling that in a UFD every non-zero non-unit element can be written as a product of irreducible factors and these irreducible factors are unique up to a *multiplication by a unit*. In order to avoid the need for multiplication by a unit, we fix a subset $\mathscr{P}_D$ of irreducible elements of D with the following properties:

1. Every element of $\mathscr{P}_D$ is irreducible.
2. For every irreducible element p of D, there is a unique element $\overline{p} \in \mathscr{P}_D$ such that $p = u\overline{p}$ for some unit u.

Let's recall that $p = u\overline{p}$ for some unit u precisely when $\langle p \rangle = \langle \overline{p} \rangle$. We also notice that in a UFD and element p is irreducible if and only if it is prime. The latter holds exactly when p is prime. An element p is prime if and only if $\langle p\rangle$ is a prime ideal. Altogether, we obtain that there is a bijection between $\mathscr{P}_D$ and the set of non-zero principal prime ideals of D. Notice that there are many choices for such a set. Here we fix one such set and many of the functions that will be defined later depend on this choice.

Since D is a UFD, for every $a \in D \setminus \{0\}$, there are unique $u_a \in D^\times$ and non-negative integers n_p such that

$$a = u_a \prod_{p\in \mathscr{P}_D} p^{n_p}.$$

We use the following functions to refer these values. Let $\sigma : D \setminus \{0\} \to D^\times$ and $v_p : D \setminus \{0\} \to \mathbb{Z}^{\geq 0}$ be such that for every $a \in D \setminus \{0\}$ the following holds

$$a = \sigma(a) \prod_{p\in \mathscr{P}_D} p^{v_p(a)}.$$

This means $v_p(a)$ is the power of p in the factorization of a with respect to the prime factors $\mathscr{P}_D$. Notice that every $a \in D \setminus \{0\}$ has only finitely many irreducible factors. This means only finitely many $v_p(a)$'s are non-zero for $p \in \mathscr{P}_D$. Therefore this product has finitely many terms (the rest are 1).

To understand the function σ better, let's go over the case of ring of integers. The classical convention in the definition of a prime number is slightly different from the way we have defined prime elements of $\mathbb{Z}$. The subtle difference is that in the classical setting a prime number must be *positive*, but in the modern language, say, -2 is also considered a prime element of the ring of integers. In a sense the classical convention factors integers with respect to

$$\mathscr{P}_{\mathbb{Z}} = \{p \in \mathbb{Z} \mid p \text{ is a positive prime element of the ring } \mathbb{Z}\}.$$

With this choice, $\sigma(a)$ is precisely the sign of a; that means it is 1 when a is positive, and it is -1 when a is negative. Because of this, even for an arbitrary UFD, we call $\sigma(a)$ the *sign* of a. Inspired with the case of $D = \mathbb{Z}$, we let

$$|a| := \sigma(a)^{-1}a = \prod_{p\in \mathscr{P}_D} p^{v_p(a)}.$$

For every $a \in D \setminus \{0\}$, $v_p(a)$ is called the *p-valuation* of a. Here are basic properties of these functions.

Proposition 15.1.2. *Suppose D is a UFD, $a, b \in D \setminus \{0\}$. Then*

1. *a) $\sigma(ab) = \sigma(a)\sigma(b)$.*
   b) $|ab| = |a||b|$.
   c) $v_p(ab) = v_p(a) + v_p(b)$ for every $p \in \mathscr{P}_D$.
2. *$a|b$ if and only if $v_p(a) \leq v_p(b)$ for every $p \in \mathscr{P}_D$.*
3. *There is $u \in D^\times$ such that $a = ub$ if and only if $v_p(a) = v_p(b)$ for every $p \in \mathscr{P}_D$.*

Proof. (1) By the factorization of a and b with respect to $\mathscr{P}_D$, we have

$$a = \sigma(a) \prod_{p \in \mathscr{P}_D} p^{v_p(a)}, \quad \text{and} \quad b = \sigma(b) \prod_{p \in \mathscr{P}_D} p^{v_p(b)}. \tag{15.1}$$

Multiplying equations given in (15.1), we deduce that

$$ab = (\sigma(a)\sigma(b)) \prod_{p \in \mathscr{P}_D} p^{v_p(a)+v_p(b)}.$$

Notice that since $\sigma(a)$ and $\sigma(b)$ are units, so is $\sigma(a)\sigma(b)$. Hence by the uniqueness of this factorization, we obtain that

$$\sigma(ab) = \sigma(a)\sigma(b) \quad \text{and} \quad v_p(ab) = v_p(a) + v_p(b) \tag{15.2}$$

for every $p \in \mathscr{P}_D$. Hence

$$|ab| = \sigma(ab)^{-1}(ab) = (\sigma(a)^{-1}a)(\sigma(b)^{-1}b) = |a||b|.$$

(2) ($\Rightarrow$) Suppose $a|b$. Then for $d \in D$, we have $b = ad$. Hence for every $p \in \mathscr{P}_D$ we have

$$v_p(b) = v_p(ad) = v_p(a) + v_p(d) \geq v_p(a).$$

($\Leftarrow$) We start with the prime factorizations of a and b (with respect to $\mathscr{P}_D$) $a = \sigma(a) \prod_{p \in \mathscr{P}_D} p^{v_p(a)}$ and $b = \sigma(b) \prod_{p \in \mathscr{P}_D} p^{v_p(b)}$. We want to write b as a multiple of a. This makes us to consider

$$d := \prod_{p \in \mathscr{P}_D} p^{v_p(b)-v_p(a)},$$

and notice that $d \in D$ as $v_p(b) \geq v_p(a)$ and $v_p(b) = v_p(a) = 0$ except for finitely many p's. Hence

$$\begin{aligned} b =& \sigma(b) \prod_{p \in \mathscr{P}_D} p^{v_p(b)} \\ =& \sigma(b) \prod_{p \in \mathscr{P}_D} p^{v_p(b)-v_p(a)} \prod_{p \in \mathscr{P}_D} p^{v_p(a)} \\ =& (\sigma(b) d \sigma(a)^{-1}) a. \end{aligned}$$

This implies that $a|b$ as $\sigma(a)$ is a unit.

(3) By part (2) we have that $v_p(a) = v_p(b)$ for every $p \in \mathscr{P}_D$ exactly when $a|b$ and $b|a$. By Lemma 13.2.1, we have that $a|b$ and $b|a$ holds if and only if $a = bu$ for some unit u. This completes the proof. □

Next we extend these functions to the group $Q(D)^{\times}$ of units of the field of fractions of D. This is needed as we have to work with the ring of polynomials $Q(D)[x]$ in order to show that $D[x]$ is a UFD.

Proposition 15.1.3 (Basic properties of valuations and the sign function)**.** *Suppose D is a UFD and $Q(D)$ is the field of fractions of D. Then*

1. *The following functions are well-defined group homomorphisms:*

$$\sigma : Q(D)^\times \to D^\times, \quad \sigma\left(\frac{a}{b}\right) := \sigma(a)\sigma(b)^{-1}.$$
$$v_p : F^\times \to \mathbb{Z}, \quad v_p\left(\frac{a}{b}\right) := v_p(a) - v_p(b).$$
$$|\cdot| : Q(D)^\times \to Q(D)^\times, \quad \left|\frac{a}{b}\right| := \frac{|a|}{|b|}.$$

2. $\ker|\cdot| = D^\times$ *and* $||q|| = |q|$ *for every* $q \in Q(D)^\times$.

3. *Let* $G(D)$ *be the image of* $|\cdot|$*. Then* $m : D^\times \times G(D) \to Q(D)^\times,\ m(u,q) := uq$ *is a group isomorphism.*

Proof. (1) Here we just check why these functions are well-defined. I leave to you to check why these maps are group homomorphisms.

Suppose $\frac{a}{b} = \frac{c}{d}$ for some $a, b, c, d \in D \setminus \{0\}$. Then $ad = bc$. Applying the functions σ, v_p and σ to the both sides of this equality, by Proposition 15.1.2, we obtain that

$$\sigma(a)\sigma(d) = \sigma(b)\sigma(c), \quad v_p(a) + v_p(d) = v_p(b) + v_p(c), \text{ and } \quad |a||d| = |b||c|$$

Therefore

$$\sigma(a)\sigma(b)^{-1} = \sigma(c)\sigma(d)^{-1}, \quad v_p(a) - v_p(b) = v_p(c) - v_p(d), \text{ and } \quad \frac{|a|}{|b|} = \frac{|b|}{|c|}.$$

This shows that the given functions are well-defined.

(2) $\frac{a}{b}$ is in the kernel of $|\cdot|$ if and only if $|\frac{a}{b}| = 1$. By part (1), the latter holds exactly when $\frac{|a|}{|b|} = 1$. This is equivalent to having $|a| = |b|$. By Proposition 15.1.2, $|a| = |b|$ holds if and only if $a = bu$ for some unit u. Altogether we have that

$$\frac{a}{b} \in \ker|\cdot| \Leftrightarrow \frac{a}{b} = \frac{u}{1}$$

for some $u \in D^\times$. Hence $\ker|\cdot| = D^\times$.

The other claim of Part (2) follows from the definition of $|\cdot|$.

(3) Since $Q(D)^\times$ is abelian, m is a group homomorphism. For every $q \in Q(D)^\times$, we have

$$q = \sigma(q)|q| = m(\sigma(q), |q|)$$

which implies that m is surjective.

Now suppose $(u, q) \in \ker m$. Then $q = u^{-1} \in D^\times \cap G(D)$. Then by Part (2), we have $q = |q| = |u^{-1}| = 1$. This means that $\ker m$ is trivial, and so m is injective. This finishes the proof. □

15.2 Greatest common divisor for UFDs

Using valuations, we can study common divisors of a finite set of non-zero elements of a UFD and prove Proposition 15.1.1.

Proof of Proposition 15.1.1. Suppose $a_1, \ldots, a_n$ are non-zero elements of a UFD D. By Proposition 15.1.2, $b \in D \setminus \{0\}$ is a common divisor of a_i's exactly when $v_p(b) \leq v_p(a_i)$ for every index i and every $p \in \mathscr{P}_D$. Hence we have

$$b|a_1, \ldots, b|a_n \Leftrightarrow v_p(b) \leq \min\{v_p(a_1), \ldots, v_p(a_n)\} \text{ for every } p \in \mathscr{P}_D. \tag{15.3}$$

Notice that $\min\{v_p(a_1), \ldots, v_p(a_n)\} = 0$ except for finitely many p's, and so

$$d := \prod_{p \in \mathscr{P}_D} p^{\min\{v_p(a_1), \ldots, v_p(a_n)\}}$$

is an element of $G(D) \cap D$. By (15.3), we deduce that

$$b|a_1, \ldots, b|a_n \Leftrightarrow b|d.$$

This shows the existence part of Proposition 15.1.1.

Now suppose d_1 and d_2 satisfy the mentioned properties in Proposition 15.1.1. This means d_i's are common divisors of $a_1, \ldots, a_n$, and every common divisor of $a_1, \ldots, a_n$ is a divisor of d_i's. Therefore $d_1|d_2$ and $d_2|d_1$. Hence by Lemma 13.2.1, there is a unit u such that $d_2 = ud_1$. As d_i's are in $G(D)$, we obtain that $m(1, d_2) = m(u, d_1)$ where m is the group isomorphism given in Part (3) of Proposition 15.1.2. Thus $d_1 = d_2$. This completes the proof. □

The greatest common divisor of $a_1, \ldots, a_n \in D \setminus \{0\}$ is the unique $d \in G(D)$ which is given by Proposition 15.1.1, and from the proof it is clear that

$$\gcd(a_1, \ldots, a_n) := \prod_{p \in \mathscr{P}_D} p^{\min(v_p(a_1), \ldots, v_p(a_n))}. \tag{15.4}$$

Notice that $\gcd$ depends on the choice of $\mathscr{P}_D$, but its value up to a multiplication by a unit is independent of the choice of $\mathscr{P}_D$. Now it is easy to get the following basic properties of the $\gcd$ function, and we leave it as an exercise.

Proposition 15.2.1. *In the above setting, suppose $a_1, \ldots, a_n \in D \setminus \{0\}$. Then*

1. *For every $c \in D \setminus \{0\}$, $\gcd(ca_1, \ldots, ca_n) = |c| \gcd(a_1, \ldots, a_n)$.*
2. *If $\gcd(a_1, \ldots, a_n) = d$, then $\frac{a_i}{d} \in D$ and $\gcd(\frac{a_1}{d}, \ldots, \frac{a_n}{d}) = 1$.*

15.3 Content of polynomials: UFD case

Now we are ready to define the content of $f(x) \in D[x]$ where D is a UFD.

Definition 15.3.1. *Suppose D is a UFD and $f(x) := a_nx^n + \cdots + a_1x + a_0 \in D[x]$ is a non-zero polynomial. The content of f is*

$$\alpha(f) := \gcd(a_n, a_{n-1}, \ldots, a_0),$$

where gcd *is defined as in* (15.4)*. We say $f(x) \in D[x]$ is* primitive *if $\alpha(f) = 1$.*

By Proposition 15.2.1, we deduce the following properties of the content function.

Lemma 15.3.2. *Suppose D is a UFD, $f, g \in D[x]$ are non-zero polynomials, and $a \in D \setminus \{0\}$. Then*

1. $\alpha(af) = |a|\alpha(f)$.
2. *If $\alpha(f) = d$, then $\frac{1}{d}f(x) \in D[x]$ and $\alpha(\frac{1}{d}f(x)) = 1$.*
3. *For $d \in D\setminus\{0\}$, $d|\alpha(f)$ if and only if $c_d(f) = 0$ where $c_d : D[x] \to (D/\langle d\rangle)[x]$ is the natural quotient map.*

By Part (2) of Lemma 15.3.2, every $f(x) \in D[x] \setminus \{0\}$ can be written as $\alpha(f)\overline{f}(x)$ and $\overline{f}(x)$ is a primitive polynomial.

Next we define the content of a non-zero polynomial $f(x) \in Q(D)[x]$ where $Q(D)$ is the field of fractions of D.

Lemma 15.3.3. *Suppose D is a UFD and $Q(D)$ is the field of fractions D. Then for every non-zero polynomial $f \in Q(D)[x]$ there are unique $q \in G(D)$ and primitive polynomial $\overline{f} \in D[x]$ such that $f(x) = q\overline{f}(x)$.*

Proof. (Existence) Suppose $f(x) = \sum_{i=0}^{n} \frac{a_i}{b_i}x^i$ for some $a_i, b_i \in D$. Let $d := \prod_{i=0}^{n} |b_i|$. Then $\widetilde{f}(x) := d\, f(x) \in D[x]$. Then by Lemma 15.3.2, $\widetilde{f}(x) = \alpha(\widetilde{f})\overline{f}(x)$ and $\overline{f}(x)$ is primitive. Hence we have that

$$f(x) = \frac{1}{d}\widetilde{f}(x) = \frac{\alpha(\widetilde{f})}{d}\overline{f}(x).$$

Notice that since $\alpha(\widetilde{f})$ and d are in the image of $|\cdot|$, $\frac{\alpha(\widetilde{f})}{d} \in G(D)$. This shows the existence part.

(Uniqueness) Suppose $q_1, q_2 \in G(D)$, $\overline{f}_1, \overline{f}_2 \in D[x]$ are primitive polynomials, and $q_1\overline{f}_1(x) = q_2\overline{f}_2(x)$. Suppose $q_i := \frac{c_i}{d_i}$ for $i = 1, 2$. Let $d := |d_1||d_2|$; then $dq_i \in D$. Hence $(dq_1)\overline{f}_1(x) = (dq_2)\overline{f}_2(x)$, which implies that

$$\alpha((dq_1)\overline{f}_1(x)) = \alpha((dq_2)\overline{f}_2(x)).$$

Therefore by Part (1) of Lemma 15.3.2, we have $|dq_1| = |dq_2|$. Since $d, q_i \in G(D)$, by Part (2) of Proposition 15.1.2 we have that $|dq_i| = dq_i$. Thus $dq_1 = dq_2$, which implies that $q_1 = q_2$. This in turn gives us that $\overline{f}_1 = \overline{f}_2$, and the uniqueness follows. □

The unique element $q \in G(D)$ given in Lemma 15.3.3 is called the *content* of $f(x)$ and it is denoted by $\alpha(f)$, and the primitive polynomial $\overline{f}(x)$ given in Lemma 15.3.3 is called the *primitive form* of $f(x)$.

15.4 Gauss's lemma for UFDs.

Having the definition of the content of a polynomial in $Q(D)[x]$, we can formulate and prove Gauss's lemma for UFDs.

Lemma 15.4.1. *Suppose D is a UFD, and $f, g \in D[x]$ are primitive. Then fg is primitive.*

Proof. Suppose to the contrary that fg is not primitive. Then there is $p \in \mathscr{P}_D$ which divides $\alpha(fg)$. This means all the coefficients of fg are in $\langle p \rangle$. Therefore $c_p(fg) = 0$ where $c_p : D[x] \to (D/\langle p \rangle)[x]$ is the natural quotient map. Notice that since D is a UFD and p is irreducible, p is a prime element of D. Hence $\langle p \rangle$ is a prime ideal. This implies that $D/\langle p \rangle$ is an integral domain. Thus $(D/\langle p \rangle)[x]$ is also an integral domain. Knowing that $c_p(f)c_p(g) = 0$ and $(D/\langle p \rangle)[x]$ is an integral domain, we obtain that either $c_p(f) = 0$ or $c_p(g) = 0$. This means either $p|\alpha(f)$ or $p|\alpha(g)$, which is a contradiction as $\alpha(f) = \alpha(g) = 1$. □

Lemma 15.4.2. *Suppose D is a UFD. Then for every $f, g \in Q(D)[x] \setminus \{0\}$ we have $\alpha(fg) = \alpha(f)\alpha(g)$.*

Proof. By the definition of the content, we have

$$f(x) = \alpha(f)\overline{f}(x) \quad \text{and} \quad g(x) = \alpha(g)\overline{g}(x) \tag{15.5}$$

and $\overline{f}(x)$ and $\overline{g}(x)$ are primitive polynomials. By (15.5), we obtain that

$$f(x)g(x) = (\alpha(f)\alpha(g))\overline{f}(x)\overline{g}(x). \tag{15.6}$$

By the first version of Gauss's lemma for UFDs, we have that $\overline{f}(x)\overline{g}(x)$ is primitive. Since $\alpha(f), \alpha(g) \in G(D)$, we have $\alpha(f)\alpha(g) \in G(D)$. By (15.6), $\alpha(f)\alpha(g) \in G(D)$, $\overline{f}(x)\overline{g}(x)$ being a primitive polynomial, and the definition of content of a polynomial, we have that $\alpha(fg) = \alpha(f)\alpha(g)$. This completes the proof. □

The following is an immediate consequence of the second version of Gauss's lemma for UFDs.

Corollary 15.4.3. *Let $\operatorname{prim} : Q(D)[x] \setminus \{0\} \to D[x] \setminus \{0\}, \operatorname{prim}(f)$ be the primitive form of f. Then*

$$\operatorname{prim}(fg) = \operatorname{prim}(f)\operatorname{prim}(g)$$

for every $f, g \in Q(D)[x] \setminus \{0\}$.

Proof. We have $f = \alpha(f)\operatorname{prim}(f)$, $g = \alpha(g)\operatorname{prim}(g)$, and $fg = \alpha(fg)\operatorname{prim}(fg)$. Hence by the second version of Gauss's lemma for UFDs, we obtain that

$$\operatorname{prim}(fg) = \operatorname{prim}(f)\operatorname{prim}(g).$$

This completes the proof. □

Now we have all the needed tools to redo the proof of why $\mathbb{Z}[x]$ is a UFD and obtain its generalization. I leave it to you to go over the proof and make sure all the arguments go through to prove the following theorem.

Theorem 15.4.4. *If D is a UFD, then $D[x]$ is a UFD.*

By induction, one can easily show the following.

Corollary 15.4.5. *If D is a UFD, then $D[x_1, \dots, x_n]$ is a UFD.*

In particular, we have that $\mathbb{Z}[x_1, \dots, x_n]$ and $F[x_1, \dots, x_n]$ where F is a field, are UFDs.

Chapter 16

Lecture 16

We have used the central problem of understanding zeros of polynomials as our point of reference in exploring algebra. So far we have worked under the assumption that we are given a field extension E of F that contains a zero α of $f(x) \in F[x]$ and among other things proved:

1. There is a unique polynomial $m_{\alpha,F}(x) \in F[x]$ with the following properties:

   a) α is a zero of $g(x) \in F[x]$ if and only if $m_{\alpha,F}(x)|g(x)$.

   b) $p(x) = m_{\alpha,F}(x)$ if and only if $p(x)$ is a monic irreducible element of $F[x]$ and $p(\alpha) = 0$.

2. $F[\alpha] \simeq F[x]/\langle m_{\alpha,F}(x)\rangle$.

3. $F[\alpha]$ is a field.

4. Every element of $F[\alpha]$ can be uniquely written as an F-linear combination of $1, \alpha, \ldots, \alpha^{n-1}$ where $n := \deg m_{\alpha,F}(x)$.

Next we want to answer the following questions:

1. For $f(x) \in F[x]$, can we find a field extension E of F that contains a zero of f? Is there a field extension that contains all the zeros of f?

2. Do we have a *canonical choice* for such a field extension? Can we talk about *the smallest* field extension that contains all the zeros of f?

16.1 Existence of a splitting field.

In this section we prove that every polynomial $f(x) \in F[x]$ can be decomposed to linear factors over a field extension.

Proposition 16.1.1. *Suppose F is a field and $f(x) \in F[x]$ is a non-constant polynomial. Then there are a field extension E of F and $\alpha_1, \ldots, \alpha_n$ such that*

1. $f(x) = a(x - \alpha_1) \cdots (x - \alpha_n)$, *where* $a = \mathrm{ld}(f)$ *is the leading coefficient of* f, *and*

2. $E = F[\alpha_1, \ldots, \alpha_n]$.

Here $F[\alpha_1, \ldots, \alpha_n]$ is the subring of E that is generated by F and α_i's. By adding α_i's one-by-one, we see that

$$F[\alpha_1, \ldots, \alpha_n] = (F[\alpha_1, \ldots, \alpha_{n-1}])[\alpha_n],$$

and so

$$F[\alpha_1, \ldots, \alpha_n] = \Big\{ \sum_{\mathbf{i}} c_{\mathbf{i}} \alpha_1^{i_1} \cdots \alpha_n^{i_n} \mid c_{\mathbf{i}} \in F, \mathbf{i} = (i_1, \ldots, i_n) \Big\}.$$

A field extension E of F which satisfies the properties of Proposition 16.1.1 is called a *splitting field of* $f(x)$ *over* F.

To prove this result, we start with finding a single linear factor in a field extension when f is irreducible.

Lemma 16.1.2. *Suppose* F *is a field and* $f(x) \in F[x]$ *is an irreducible polynomial. Then there are a field extension* E *of* F *and* $\alpha \in E$ *such that* $f(\alpha) = 0$ *and* $E = F[\alpha]$.

To find such a field extension, we make a backward argument. If $E = F[\alpha]$, then we have that

$$\theta : F[x]/\langle m_{\alpha,F}(x)\rangle \to E, \theta(g(x) + \langle m_{\alpha,F}(x)\rangle) := g(\alpha)$$

is an isomorphism. Notice that since $f(\alpha) = 0$, $m_{\alpha,F}(x)|f(x)$. As $f(x)$ is irreducible in $F[x]$ and $m_{\alpha,F}(x)|f(x)$, there is $c \in F^\times$ such that $f(x) = cm_{\alpha,F}(x)$. This implies that $\langle m_{\alpha,F}(x)\rangle = \langle f(x)\rangle$. Hence there is an isomorphism from $F[x]/\langle f(x)\rangle$ to E which sends $x + \langle f(x)\rangle$ to α. This shows us what we should choose for E and α.

Proof. Let $E := F[x]/\langle f\rangle$. Since F is a field, $F[x]$ is a PID. As $F[x]$ is a PID and $f \in F[x]$ is irreducible, $\langle f\rangle$ is a maximal ideal of $F[x]$. Therefore $F[x]/\langle f\rangle$ is a field.

Next we show that E is a field extension of F. Let $I := \langle f\rangle$, and $i : F \to E, i(c) := c + I$. It is easy to see that i is a ring homomorphism which sends 1_F to 1_E. Thus $\ker i$ is a proper ideal of F. Since 0 is the only proper ideal of a field, we obtain that $\ker i = 0$. This implies that i is injective. Hence E is a field extension of F.

Now we show that $\alpha := x + I \in E$ is a zero of f. In order to evaluate f at α, we have to view the coefficients of F as elements of E. This means we have to work with the copy of F in E. Suppose

$$f(x) = a_n x^n + \cdots + a_0.$$

Then

$$\begin{aligned} f(\alpha) =& i(a_n)\alpha^n + \cdots + i(a_0) \\ =& (a_n + I)(x + I)^n + \cdots + (a_0 + I) \\ =& (a_n x^n + \cdots + a_0) + I \\ =& f(x) + I = 0 + I. \end{aligned}$$

The last equality holds because $f(x) \in I$. Notice that $0 + I$ is the zero of E. Hence $f(\alpha) = 0$.

Finally every element of E is of the form

$$\Big(\sum_{j=0}^{m} b_j x^j\Big) + I = \sum_{j=0}^{m} i(b_j)\alpha^j \in F[\alpha].$$

Hence $E = F[\alpha]$. This completes the proof. □

Proof of Proposition 16.1.1. We proceed by the strong induction on $\deg f$. We start with the base of induction. Suppose $\deg f = 1$. Then $f(x) = ax + b = a(x + b/a)$. Then $\alpha := -b/a \in F$ is a zero of $f(x)$. Then $E := F$ and $\alpha \in E$ satisfy the properties mentioned in the statement of Proposition 16.1.1. This completes the proof of the base case.

To prove the strong induction step, we consider two cases.

Case 1. f is not irreducible in $F[x]$.

In this case, there are non-constant $g, h \in F[x]$ such that $f(x) = g(x)h(x)$. So $\deg g, \deg h < \deg f$. By the strong induction hypothesis, there are a field extension E_1 of F and $\alpha_1, \dots, \alpha_m \in E_1$ such that

$$g(x) = b(x - \alpha_1)\cdots(x - \alpha_m) \tag{16.1}$$

where $b = \mathrm{ld}(g)$ and

$$E_1 = F[\alpha_1, \dots, \alpha_m]. \tag{16.2}$$

Another application of the strong induction hypothesis implies that there are a field extension E of E_1 and $\beta_1, \dots, \beta_k \in E$ such that

$$h(x) = c(x - \beta_1)\cdots(x - \beta_k) \tag{16.3}$$

where $c = \mathrm{ld}(h)$ and

$$E = E_1[\beta_1, \dots, \beta_k]. \tag{16.4}$$

Altogether we obtain that

$$f(x) = g(x)h(x) = (bc)(x - \alpha_1)\cdots(x - \alpha_m)(x - \beta_1)\cdots(x - \beta_k),$$

and

$$E = (F[\alpha_1, \dots, \alpha_m])[\beta_1, \dots, \beta_k] = F[\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_k].$$

And the claim follows in this case.

Case 2. $f(x) \in F[x]$ is irreducible.

In this case, by Lemma 16.1.2, there are a field extension E_1 of F and $\alpha \in E_1$ such that

$$f(\alpha) = 0 \quad \text{and} \quad E_1 = F[\alpha].$$

By the factor theorem, there is $g(x) \in E_1[x]$ such that

$$f(x) = (x - \alpha)g(x).$$

Notice that $\deg g < \deg f$, and so by the strong induction hypothesis, there are a field extension E of E_1 and $\alpha_1, \ldots, \alpha_n \in E$ such that

$$g(x) = b(x - \alpha_1) \cdots (x - \alpha_n) \tag{16.5}$$

where $b = \mathrm{ld}(g)$ and

$$E_1 = F[\alpha_1, \ldots, \alpha_n]. \tag{16.6}$$

Altogether we have that

$$f(x) = (x - \alpha)g(x) = b(x - \alpha)(a - \alpha_1) \cdots (x - \alpha_n),$$

and

$$E = (F[\alpha])[\alpha_1, \ldots, \alpha_n] = F[\alpha, \alpha_1, \ldots, \alpha_n].$$

This completes the proof. □

16.2 Towards uniqueness of a splitting field.

In this section among other things we show that two splitting fields of $f(x)$ over F are isomorphic. The results of this section play an important role in Galois theory.

Similar to the proof of the existence part, we start with adding one zero of an irreducible factor. We formulate a result which is essentially proved in the discussion prior to the proof of Lemma 16.1.2.

Lemma 16.2.1. *Suppose F is a field and $f(x) \in F[x]$ is irreducible. Assume E is a field extension of E and $\alpha \in E$ such that*

$$f(\alpha) = 0 \quad \textit{and} \quad E = F[\alpha].$$

Then

$$\overline{\phi}_\alpha : F[x]/\langle f \rangle \to E, \ \overline{\phi}_\alpha(g(x) + \langle f \rangle) := g(\alpha)$$

is an isomorphism.

Proof. we have that

$$\overline{\phi}_\alpha : F[x]/\langle m_{\alpha,F}(x) \rangle \to E, \ \overline{\phi}_\alpha(g(x) + \langle m_{\alpha,F}(x) \rangle) := g(\alpha) \tag{16.7}$$

is an isomorphism. Notice that since $f(\alpha) = 0$, $m_{\alpha,F}(x) | f(x)$. As $f(x)$ is irreducible in $F[x]$ and $m_{\alpha,F}(x) | f(x)$, there is $c \in F^\times$ such that $f(x) = c m_{\alpha,F}(x)$. This implies that $\langle m_{\alpha,F}(x) \rangle = \langle f(x) \rangle$. Therefore the claim follows form (16.7). □

Lemma 16.2.1 can be viewed as a type of *uniqueness* result for such a field. In the next lemma, we strengthen this uniqueness result in a way which makes it more suitable for a later use in an inductive argument.

Roughly the next lemma says that if we have two copies of a field, let's call them F_1 and F_2, and an irreducible polynomial $f_1 \in F_1[x]$, then the copy of f_1 in $F_2[x]$, let's call it f_2, is irreducible, and after adding a zero α_1 of f_1 to F_1 and adding a zero α_2 of f_2 to F_2, we end up getting isomorphic fields.

Lemma 16.2.2. *Suppose F and F' are fields and $\theta : F \to F'$ is an isomorphism. Let $f(x) \in F[x]$ be an irreducible polynomial. Suppose E is a field extension of F, $\alpha \in E$, E' is a field extension of F', and $\alpha' \in E$ satisfy the following properties:*

1. *$f(\alpha) = 0$ and $E = F[\alpha]$.*
2. *$\theta(f)(\alpha') = 0$ and $E' = F'[\alpha']$.*

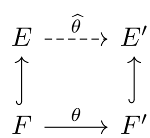
Then there is a unique isomorphism $\widehat{\theta} : E \to E'$ such that for every $a \in F$, $\widehat{\theta}(a) = \theta(a)$ and $\widehat{\theta}(\alpha) = \alpha'$.

Notice that the ring isomorphism $\theta : F \to F'$ can be extended to a ring isomorphism from $F[x]$ to $F'[x]$ that is also denoted by θ:

$$\theta\Big(\sum_{i=0}^{n} a_i x^i\Big) := \sum_{i=0}^{n} \theta(a_i) x^i.$$

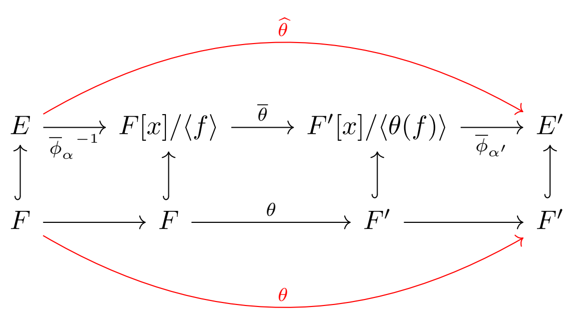
Roughly for $f \in F[x]$, $\theta(f)$ is the copy of f in $F'[x]$.

The conclusion of Lemma 16.2.2 is often captured in the following diagram as it is often better to *see* what we can prove. We say the following is a *commutative diagram*:



This means all directed paths in the diagram with the same start and endpoints lead to the same result.

Our proof can be summarized in the following diagram:



(16.8)

Going though the above diagram, we give the details of the proof.

Proof of Lemma 16.2.2. Lemma 16.2.1 gives us the first block in the diagram in (16.8). To understand the second block, we start with a ring homomorphism from the numerator of the left hand side to the right hand side. Let

$$\widetilde{\theta} : F[x] \to F'[x]/\langle \theta(f) \rangle, \;\; \widetilde{\theta}(g) := \theta(g) + \langle \theta(f) \rangle.$$

Notice that $\widetilde{\theta}$ is the composite of θ with the quotient map

$$p : F'[x] \to F'[x]/\langle \theta(f) \rangle,$$

and so $\widetilde{\theta}$ is a surjective ring homomorphism. By the first ring isomorphism we have that

$$\overline{\theta} : F[x]/\ker \widetilde{\theta} \to F'[x]/\langle \theta(f) \rangle, \;\; \overline{\theta}(g + \ker \widetilde{\theta}) := \theta(g) + \langle \theta(f) \rangle \tag{16.9}$$

is a ring isomorphism. We also have that $g \in \ker \widetilde{\theta}$ if and only if

$$\theta(g) \in \langle \theta(f) \rangle = \theta(\langle f \rangle),$$

and the latter holds precisely when $g \in \langle f \rangle$. This implies that $\ker \widetilde{\theta} = \langle f \rangle$. Hence by (16.9), we have that $\overline{\theta}$ is an isomorphism from $F[x]/\langle f \rangle$ to $F'[x]/\langle \theta(f) \rangle$. This gives us the middle block in the diagram given in (16.8). We also notice that since θ is an isomorphism and $f \in F[x]$ is irreducible, $\theta(f)$ is irreducible in $F'[x]$. As $\alpha' \in E'$ is a zero of $\theta(f)$, another application of Lemma 16.2.1 gives us the last block in the diagram given in (16.8). The composite of the ring isomorphisms in the first row give us an isomorphism $\widehat{\theta} : E \to E'$ and because the diagram in (16.8) is a commutative diagram, the claim follows. $\square$

Lemma 16.2.2 will be used to show that splitting fields of $f(x) \in F[x]$ are isomorphic.

Chapter 17

Lecture 17

In the previous lecture we proved the *existence of a splitting field* (see Proposition 16.1.1), and to work towards the uniqueness of splitting fields, we proved that adding zeros of an irreducible polynomial and its twin in another copy of the base field give us isomorphic fields (see Lemma 16.2.1).

From Lemma 16.2.1, we immediately obtain that adding two zeros of an irreducible polynomial to the base field give us two isomorphic fields.

Corollary 17.0.1. *Suppose F is a field and $f(x) \in F[x]$ is irreducible. Suppose E and E' are field extensions of F, $\alpha \in E$ and $\alpha' \in E'$ are zeros of $f(x)$. Then there is a ring isomorphism $\widehat{\theta} : F[\alpha] \to F'[\alpha']$ such that*

$$\widehat{\theta}(g(\alpha)) := g(\alpha')$$

for every $g(x) \in F[x]$.

Proof. By Lemma 16.2.2, there is a ring isomorphism $\widehat{\theta} : F[\alpha] \to F[\alpha']$ such that $\widehat{\theta}(c) = c$ for every $c \in F$, and $\theta(\alpha) = \alpha'$. Then, for every $g(x) = \sum_{i=0}^{n} c_i x^i \in F[x]$ we have

$$\widehat{\theta}(g(\alpha)) = \widehat{\theta}(\sum_{i=0}^{n} c_i \alpha^i) = \sum_{i=0}^{n} \widehat{\theta}(c_i)\widehat{\theta}(\alpha)^i = \sum_{i=0}^{n} c_i \alpha'^i = g(\alpha').$$

This completes the proof. □

Exercise 17.0.2. *Suppose E is a field extension of F and $\alpha, \alpha' \in E$ are algebraic over F. Suppose $g(\alpha) \mapsto g(\alpha')$ for every $g(x) \in F[x]$ is a well-defined map. Then $m_{\alpha,F}(x) = m_{\alpha',F}(x)$, and so they are zeros of a single irreducible polynomial in $F[x]$.*

17.1 Extension of isomorphisms to splitting fields.

Now we are ready to prove the uniqueness of splitting fields. The following theorem plays an important role in Galois theory and understanding *symmetries* of splitting fields.

Theorem 17.1.1. *Suppose F and F' are fields, and $\theta : F \to F'$ is a ring isomorphism. Let $f(x) \in F[x] \setminus F$. Suppose E is a splitting field of f over F, and E' is a splitting field of $\theta(f)$ over F'. Then θ can be extended to an isomorphism $\widehat{\theta} : E \to E'$. This means that for every $c \in F$, we have $\widehat{\theta}(c) = \theta(c)$.*

The conclusion of Theorem 17.1.1 can be captured in the following commutative diagram.

$$\begin{array}{ccc} E & \overset{\widehat{\theta}}{\dashrightarrow} & E' \\ \uparrow & & \uparrow \\ F & \xrightarrow{\theta} & F' \end{array}$$

A *dashed arrow* means that this function was not initially given, and having other functions, we can find this one in a way that results in obtaining a commutative diagram, and *a hooked arrow* means that it is a natural inclusion map.

Proof. We proceed by induction on $\deg f$. If $\deg f = 1$, then f has a zero in F, and $\theta(f)$ has a zero in F'. Therefore $E = F$ and $E' = F'$. Hence we can choose $\widehat{\theta} = \theta$.

To prove the induction step, we start by recalling what it means that E and E' are splitting fields. Since E is a splitting field of f over F, there are $\alpha_1, \ldots, \alpha_n \in E$ such that

$$E = F[\alpha_1, \ldots, \alpha_n] \quad \text{and} \quad f(x) = a(x - \alpha_1) \cdots (x - \alpha_n), \tag{17.1}$$

where $a = \mathrm{ld}(f)$. Similarly we have that there are $\alpha'_1, \ldots, \alpha'_n \in E'$ such that

$$E' = F'[\alpha'_1, \ldots, \alpha'_n] \quad \text{and} \quad \theta(f(x)) = a'(x - \alpha'_1) \cdots (x - \alpha'_n), \tag{17.2}$$

where $a' = \mathrm{ld}(\theta(f))$. Since α_1 is a zero of f, we have that $m_{\alpha_1,F}$ is an irreducible factor of f in $F[x]$. Therefore $\theta(m_{\alpha_1,F})$ is an irreducible factor of $\theta(f)$ in $F'[x]$. Since $x - \alpha'_i$'s are irreducible factors of $\theta(f)$ in $E'[x]$, $\theta(m_{\alpha_1,F})$ divides $\theta(f)$ in $E'[x]$ and $E'[x]$ is a UFD, we deduce that

$$\theta(m_{\alpha_1,F}) = (x - \alpha'_{i_1}) \cdots (x - \alpha'_{i_k}) \tag{17.3}$$

for some $i_1, \ldots, i_k$. After the rearranging the indexes, if needed, we can and will assume that $x - \alpha'_1$ is a factor of $\theta(m_{\alpha_1,F})$ which means α'_1 is a zero of $\theta(m_{\alpha_1,F})$.

Since $m_{\alpha_1,F}$ is irreducible in $F[x]$ and α'_1 is a zero of $\theta(m_{\alpha_1,F})$, by Lemma 16.2.2 there is ring isomorphism $\widehat{\theta}_1 : F[\alpha_1] \to F'[\alpha'_1]$ which is an extension of θ (this means the diagram in (17.4) is a commutative diagram), and $\widehat{\theta}_1(\alpha_1) = \alpha'_1$.

$$\begin{array}{ccc} F[\alpha_1] & \overset{\widehat{\theta}_1}{\dashrightarrow} & F'[\alpha'_1] \\ \uparrow & & \uparrow \\ F & \xrightarrow{\theta} & F' \end{array} \tag{17.4}$$

Notice that by the factor theorem, there is $g \in (F[\alpha_1])[x]$ such that

$$f(x) = (x - \alpha_1)g(x). \tag{17.5}$$

By (17.5) and (17.1), we deduce that

$$g(x) = a(x - \alpha_2) \cdots (x - \alpha_n). \tag{17.6}$$

Applying $\widehat{\theta}_1$ to the both sides of (17.5), we obtain that

$$\widehat{\theta}_1(f) = (x - \widehat{\theta}_1(\alpha_1))\widehat{\theta}_1(g). \tag{17.7}$$

Since $\widehat{\theta}_1(\alpha_1) = \alpha_1'$, by (17.2), it follows that

$$\widehat{\theta}_1(g) = a'(x - \alpha_2') \cdots (x - \alpha_n'). \tag{17.8}$$

By (17.6), after adding zeros of g to $F[\alpha_1]$

$$(F[\alpha_1])[\alpha_2, \ldots, \alpha_n] = F[\alpha_1, \ldots, \alpha_n]$$

we get E. Hence E is a splitting field of g over $F[\alpha_1]$. Similarly, by (17.8), after adding zeros of $\widehat{\theta}_1(g)$ to $F'[\alpha_1']$ we get E'. Therefore E' is a splitting field of $\widehat{\theta}_1(g)$ over $F'[\alpha_1']$. Since $\deg g < \deg f$, we can and will apply the induction hypothesis. By the induction hypothesis, we obtain a ring isomorphism $\widehat{\theta} : E \to E'$ which is an extension of $\widehat{\theta}_1$ (see the commutative diagram given in (17.9)).

$$\begin{array}{ccc} E & \overset{\widehat{\theta}}{\dashrightarrow} & E' \\ \uparrow & & \uparrow \\ F[\alpha_1] & \xrightarrow{\widehat{\theta}_1} & F'[\alpha_1'] \end{array} \tag{17.9}$$

By (17.4) and (17.9) (see the diagram in (17.10)),

$$\begin{array}{ccc} E & \xrightarrow{\widehat{\theta}} & E' \\ \uparrow & & \uparrow \\ F[\alpha_1] & \xrightarrow{\widehat{\theta}_1} & F'[\alpha_1'] \\ \uparrow & & \uparrow \\ F & \xrightarrow{\theta} & F' \end{array} \tag{17.10}$$

we deduce that $\widehat{\theta}$ is an extension of θ, which completes the proof. □

The idea of the above proof is easy:

1. Find an *irreducible* factor of f in $F[x]$, say $h(x)$.

2. Add a zero of h to F and a zero of $\theta(h)$ to F', and find $\widehat{\theta}_1 : F[\alpha_1] \to F'[\alpha_1']$.

3. View E as a splitting field of g and E' as a splitting field of $\widehat{\theta}_1(g)$. Use *induction hypothesis*.

Based on Theorem 17.1.1, we can prove the uniqueness of splitting fields up to an isomorphism.

Theorem 17.1.2. *Suppose F is a field, $f(x) \in F[x] \setminus F$, and E, E' are splitting fields of $f(x)$ over F. Then there is a ring isomorphism $\widehat{\theta} : E \to E'$ such that $\widehat{\theta}|_F = \mathrm{id}_F$; that means for every $c \in F$ we have that $\widehat{\theta}(c) = c$.*

Proof. Notice that $\mathrm{id}_F : F \to F$ is an isomorphism, and so by Theorem 17.1.1, there is a ring isomorphism $\widehat{\theta} : E \to E'$ which is an extension of id_F. This completes the proof. □

17.2 Two examples

In general giving a precise description of a splitting field of a polynomial is a very hard task. In this section, we learn two examples where to some extend we can describe a splitting of the given polynomial.

Example 17.2.1. *Let $\zeta_n := e^{2\pi i/n}$. Then $\mathbb{Q}[\zeta_n]$ is a splitting field of $x^n - 1$ over $\mathbb{Q}$.*

Proof. Notice that the multiplicative order of ζ_n is n. Hence $(\zeta_n^j)^n = 1$ for every integer j in $[0, n)$, and $1, \zeta_n, \ldots, \zeta_n^{n-1}$ are distinct. Therefore these are distinct zeros of $x^n - 1$. Thus by the generalized factor theorem, comparing the degrees and the leading coefficients, we obtain that

$$x^n - 1 = (x-1)(x-\zeta_n)\cdots(x-\zeta_n^{n-1}).$$

Hence $E := \mathbb{Q}[1, \zeta_n, \ldots, \zeta_n^{n-1}]$ is a splitting field of $x^n - 1$ over $\mathbb{Q}$. Notice that $\mathbb{Q}[\zeta_n] \subseteq E$. Since $\zeta_n^j \in \mathbb{Q}[\zeta_n]$ for every integer j, we have that $E \subseteq \mathbb{Q}[\zeta_n]$. The claim follows. □

Example 17.2.2. *Let $\zeta_n := e^{2\pi i/n}$. Then $\mathbb{Q}[\zeta_n, \sqrt[n]{2}]$ is a splitting field of $x^n - 2$ over $\mathbb{Q}$.*

Proof. Notice that $(\zeta_n^j \sqrt[n]{2})^n = 2$ for every integer j. Hence $\sqrt[n]{2}, \zeta_n \sqrt[n]{2}, \ldots, \zeta_n^{n-1} \sqrt[n]{2}$ are distinct zeros of $x^n - 2$. Therefore by the generalized factor theorem, comparing degrees and leading coefficients, we obtain that

$$x^n - 2 = (x - \sqrt[n]{2})(x - \zeta_n \sqrt[n]{2})\cdots(x - \zeta_n \sqrt[n]{2}^{n-1}).$$

Therefore $E := \mathbb{Q}[\sqrt[n]{2}, \zeta_n \sqrt[n]{2}, \ldots, \zeta_n^{n-1} \sqrt[n]{2}]$ is a splitting field of $x^n - 2$ over $\mathbb{Q}$. Notice that $\zeta_n := (\zeta_n \sqrt[n]{2})(\sqrt[n]{2})^{-1} \in E$. Hence $\mathbb{Q}[\sqrt[n]{2}, \zeta_n] \subseteq E$. We also have that $\zeta_n^j \sqrt[n]{2} \in \mathbb{Q}[\zeta_n, \sqrt[n]{2}]$ for every integer j. This implies that $E \subseteq \mathbb{Q}[\sqrt[n]{2}, \zeta_n]$, and the claim follows. □

Next we use splitting fields to study finite fields.

Chapter 18

Lecture 18

In the previous couple of lectures we proved the following results about splitting fields.

Theorem (Existence (See Proposition 16.1.1))**.** *Suppose F is a field and $f \in F[x] \setminus F$. Then there is a s splitting field E of f over F.*

Let's recall that E is called a *splitting field* of f over F if there are $\alpha_1, \ldots, \alpha_n \in E$ such that $f(x) = a(x - \alpha_1) \cdots (x - \alpha_n)$, for some $a \in F$, and $E = F[\alpha_1, \ldots, \alpha_n]$.

Theorem (Uniqueness (See Theorem 17.1.2))**.** *Suppose F is a field and $f \in F[x] \setminus F$, E, E' are splitting fields of f over F. Then there is $\widehat{\theta} : E \to E'$ such that for every $c \in F$, $\widehat{\theta}(c) = c$.*

For field extensions E and E' of F, we say a ring isomorphism $\widehat{\theta} : E \to E'$ is an F*-isomorphism* if $\widehat{\theta}(c) = c$ for every $c \in F$.

The Uniqueness result was proved using the following isomorphism extension theorem.

Theorem (Isomorphism extension (See Theorem 17.1.1))**.** *Suppose F and F' are fields, $\theta : F \to F'$ is an isomorphism, and $f(x) \in F[x] \setminus F$. Suppose E is a splitting field of f over F, and E' is a splitting field of $\theta(f)$ over F'. Then there is an isomorphism $\widehat{\theta} : E \to E'$ which is an extension of θ.*

Prove of Isomorphism Extension Theorem is based on the following result on sending a zero of an irreducible polynomial to another zero.

Theorem (Sending a zero to another (See Lemma 16.2.2))**.** *Suppose F and F' are fields, f is irreducible in $F[x]$. Suppose E is a field extension of which contains a zero α of f, and E' is a field extension of F' which contains a zero of $\theta(f)$. Then there is $\widehat{\theta} : F[\alpha] \to F'[\alpha']$ which is an extension of θ and $\widehat{\theta}(\alpha) = \alpha'$.*

Now we use these results to study finite fields.

18.1 Finite fields: uniqueness

Suppose F is a finite field. Then its characteristic is a prime number p.

Lemma 18.1.1 (Order of a finite field)**.** *Suppose F is a finite of characteristic p. Then $|F| = p^n$ for some positive integer n.*

Proof. Since F is a finite integral domain, p is prime. Suppose ℓ is a prime factor of $|F|$. Then by Cauchy's theorem from group theory, there is $a \in F$ such that the additive order of a is ℓ. Since $\mathrm{char}(F) = p$, $pa = 0$. This implies that the additive order ℓ of a divides p. As ℓ and p primes, we deduce that $\ell = p$. Hence the only prime factor of $|F|$ is p, which implies that $|F|$ is a power of p. This completes the proof. □

We have seen that $x^p - x = \prod_{a\in\mathbb{Z}_p}(x-a)$. Next we generalize this to any finite field. We start with the following lemma, which can be viewed as a generalization of Fermat's little theorem.

Lemma 18.1.2. *Suppose F is a finite field of order q. Then $a^q = a$ for every $a \in F$.*

Proof. If $a = 0$, then clearly we have that $a^q = a$. If $a \neq 0$, then a is a unit. Hence $a^{|F^\times|} = 1$ as we know that in every (multiplicative) group G we have $g^{|G|} = e$. Since F is a field, we have $|F^\times| = |F| - 1 = q - 1$. Therefore $a^{q-1} = 1$, which implies that $a^q = a$. This completes the proof. □

Theorem 18.1.3. *Suppose F is finite field of order q. Then*

$$x^q - x = \prod_{\alpha\in F}(x-\alpha)$$

in $F[x]$.

Proof. By Lemma 18.1.2, every $\alpha \in F$ is a zero of $x^q - x$. Hence by the generalized factor theorem, there is $g(x) \in F[x]$ such that

$$x^q - x = g(x)\prod_{\alpha\in F}(x-\alpha). \tag{18.1}$$

Comparing the degrees of both sides, we deduce that g is a non-zero constant. Subsequently comparing the leading coefficients of both sides of (18.1), we obtain that $g = 1$. The claim follows. □

Theorem 18.1.4 (Uniqueness)**.** *Suppose F is a finite field of order $q = p^n$ where p is a prime number. Then F is a splitting field of $x^q - x$ over $\mathbb{Z}_p$. In particular, if F and F' are two fields of order q, then they are isomorphic.*

Proof. By Lemma 18.1.1, we obtain that the characteristic of F is p. Hence $\mathbb{Z}_p$ can be viewed as a subfield of F. By Theorem 18.1.3, we have that $x^q - x$ can be factored as a product of degree one polynomials over F, and adding zeros of $x^q - x$ to $\mathbb{Z}_p$, we get the entire F. Hence F is a splitting field of $x^q - x$ over $\mathbb{Z}_p$.

If F and F' are fields of order q, then both of them are splitting fields of $x^q - x$ over $\mathbb{Z}_p$. Hence by Theorem 17.1.2, F and F' are isomorphic. This completes the proof. □

18.2 Finite fields: towards existence

We want to show the existence of a finite field of order $q = p^n$ where p is prime and n is a positive integer. By Theorem 18.1.4, we have to consider a splitting field E of $x^q - x$ over $\mathbb{Z}_p$ and show that it has q elements. So in this section, we let E be a splitting field of $x^q - x$ over $\mathbb{Z}_p$ and

$$F := \{\alpha \in E \mid \alpha^q = \alpha\}.$$

Lemma 18.2.1. *In the above setting, F is a field.*

Proof. To show F is a field, we prove that is closed under addition, multiplication, negation, and inversion.

Notice that since the characteristic of F is a prime number p, the Frobenius map $\sigma : E \to E, \sigma(a) := a^p$ is a ring homomorphism (see Problem 4 in Week 1 assignment). Therefore

$$\sigma^{(n)} : E \to E, \quad \sigma^{(n)}(a) = a^{p^n}$$

is also a ring homomorphism. Notice that F is the set of *fixed points* of $\sigma^{(n)}$; that means that

$$F = \{a \in E \mid \sigma^{(n)}(a) = a\}.$$

For every $\alpha, \beta \in F$, we have

$$\sigma^{(n)}(\alpha+\beta) = \sigma^{(n)}(\alpha)+\sigma^{(n)}(\beta) = \alpha+\beta \text{ and } \sigma^{(n)}(\alpha\cdot\beta) = \sigma^{(n)}(\alpha)\cdot\sigma^{(n)}(\beta) = \alpha\cdot\beta.$$

So $\alpha + \beta$ and $\alpha \cdot \beta$ are in F. Therefore F is closed under addition and multiplication. For $\alpha \in F$ we also have that

$$\sigma^{(n)}(-\alpha) = -\sigma^{(n)}(\alpha) = -\alpha,$$

and so $-\alpha \in F$. Suppose $\alpha \in F \setminus \{0\}$. Then $\alpha^{-1} \in E$, and

$$\sigma^{(n)}(\alpha^{-1}) = (\alpha^{-1})^q = (\alpha^q)^{-1} = \alpha^{-1},$$

which implies that $\alpha^{-1} \in F$. This completes the proof. □

Next we want to show that $|F| = q$, which completes the proof of the existence of a field of order q.

Corollary 18.2.2. *In the above setting, the order of F is the same of the number of distinct zeros of $x^q - x$ in E.*

Proof. Since E is a splitting field of $x^q - x$ over $\mathbb{Z}_p$, there are $\alpha_1, \ldots, \alpha_q \in E$ such that

$$x^q - x = \prod_{i=1}^{q}(x - \alpha_i).$$

Notice that $\alpha \in F$ if and only if α is a zero of $x^q - x$. Since E is an integral domain, we obtain that

$$F = \{\alpha_1, \ldots, \alpha_n\}.$$

The claim follows. □

By Corollary 18.2.2, we have that $|F| = q$ if and only if zeros of $x^q - x$ in its splitting field are distinct. So we need to find a mechanism to determine whether zeros of a polynomial in its splitting field are distinct.

18.3 Separability: having distinct zeros in a splitting field.

We need to come up with a technique of finding out whether or not $f(x)$ has a multiple zero. Recall that we say $a \in A$ is a *multiple zero* of f if $f(x) = (x-a)^2 g(x)$ for some $g(x) \in A[x]$. We use an idea from calculus: a polynomial $f(x) \in \mathbb{C}[x]$ has a multiple zero at z if and only if $f(z) = f'(z) = 0$. This means we need to define the derivative of a polynomial in $A[x]$ for an arbitrary unital commutative ring A.

Definition 18.3.1. *Suppose $f(x) := \sum_{i=0}^{\infty} a_i x^i \in A[x]$ where A is a unital commutative ring. We let*

$$f'(x) := \sum_{i=1}^{\infty} i a_i x^{i-1}, \tag{18.2}$$

and call it the derivative *of f.*

Sometimes it is useful to write the sum in (18.2) starting from 0

$$f'(x) = \sum_{i=0}^{\infty} i a_i x^{i-1}.$$

One can check that the following properties of ordinary derivatives still hold for polynomials in a general setting.

Lemma 18.3.2. *Suppose A is a unital commutative ring, $f, g \in A[x]$, and $a, b \in A$. Then the derivative of $af(x) + bg(x)$ is $af'(x) + bg'(x)$ and the product rule*

$$(fg)' = f'g + fg'$$

holds.

Proof. It is easy to check that $(af+bg)' = af' + bg'$. Here we only discuss the product rule. Suppose $f(x) = \sum_{i=0}^{\infty} a_i x^i$ and $g(x) = \sum_{j=0}^{\infty} b_j x^j$. Then the coefficient of x^k in fg is

$$c_k := \sum_{i+j=k, i,j \geq 0} a_i b_j.$$

Thus $(fg)' = \sum_{k=0}^{\infty} k c_k x^{k-1}$. Since $f'(x) = \sum_{i=0}^{\infty} i a_i x^{i-1}$ and $g'(x) = \sum_{j=0}^{\infty} j b_j x^{j-1}$, the coefficient of x^{k-1} in $f'g$ is

$$\sum_{i+j=k, i,j \geq 0} i a_i b_j$$

and the coefficient of x^{k-1} in fg' is

$$\sum_{i+j=k, i,j \geq 0} j a_j b_j.$$

Hence

$$f'g + fg' = \sum_{k=0}^{\infty} \Big(\sum_{i+j=k, i,j\geq 0} (i+j)a_i b_j \Big) x^{k-1} = \sum_{k=0}^{\infty} kc_k x^{k-1}.$$

The claim follows. □

Lemma 18.3.3. *Suppose A is a unital commutative ring and for $a \in A$ and $f, g \in A[x]$, we have $f(x) = (x-a)^2 g(x)$. Then $f(a) = f'(a) = 0$.*

Proof. Clearly $f(a) = 0$. By the product rule, we have that

$$f'(x) = (x-a)^2 g(x) + 2(x-a)g(x) = (x-a)((x-a)g'(x) + 2g(x)).$$

Hence $f'(a) = 0$. The claim follows. □

Proposition 18.3.4. *Suppose F is a field, $f(x) \in F[x] \setminus F$, and E is a splitting field of f over F. Then $f(x)$ does not have multiple zeros in E if and only if $\gcd(f, f') = 1$ in $F[x]$*[1].

Proof. ($\Rightarrow$) Suppose $\gcd(f, f') \neq 1$. Then there is a non-constant monic polynomial $q(x) \in F[x]$ which divides both $f(x)$ and $f'(x)$. Since E is a splitting field of f over F, there are $\alpha_1, \ldots, \alpha_n \in E$ such that

$$f(x) = a(x - \alpha_1) \cdots (x - \alpha_n),$$

for some $a \in F$. As $q(x)|f(x)$, $x - \alpha_i$'s are irreducible in $E[x]$, and $E[x]$ is a UFD, we have that

$$q(x) = (x - \alpha_{i_1}) \cdots (x - \alpha_{i_k})$$

for some $i_1, \ldots, i_k$. Since $q(x)|f'(x)$, we have that $f'(\alpha_{i_1}) = 0$. After rearranging the indexes, if necessary, we can and will assume that $i_1 = 1$. Thus $f'(\alpha_1) = 0$. By the product rule, we have that $f'(x)$ is equal to

$$a((x-\alpha_2)\cdots(x-\alpha_n)+(x-\alpha_1)(x-\alpha_3)\cdots(x-\alpha_n)+\cdots+(x-\alpha_1)\cdots(x-\alpha_{n-1})).$$

Hence

$$f'(\alpha_1) = a(\alpha_1 - \alpha_2) \cdots (\alpha_1 - \alpha_n).$$

Therefore $f'(\alpha_1) = 0$ implies that $\alpha_1 = \alpha_j$ for some index $j \geq 2$. This means f has multiple zeros.

($\Leftarrow$) Suppose $f(x) = (x - \alpha)^2 g(x)$. Then by Lemma 18.3.3, $f'(\alpha) = 0$. As $f'(x) \in F[x]$, we deduce that $m_{\alpha,F}(x)|f'(x)$ in $F[x]$. Similarly, since $f(\alpha) = 0$ and $f(x) \in F[x]$, we have $m_{\alpha,F}(x)|f(x)$. Therefore $m_{\alpha,F}(x)$ is a common divisor of f and f' in $F[x]$, which implies that $\gcd(f, f') \neq 1$. This completes the proof. □

[1]Here we are using the convention that the greatest common divisor of polynomials with coefficients in a field are monic.

18.4 Finite field: existence

Let's recall some of the notation and results from Section 18.2. Let $q = p^n$ where p is a prime and n is a positive integer. Let E be a splitting field of $x^q - x$ over $\mathbb{Z}_p$. Let

$$F := \{\alpha \in E \mid \alpha^q = \alpha\}.$$

By Lemma 18.2.1, F is a field, and by Corollary 18.2.2, the order of F is the number of distinct zeros of $x^q - x$ in E.

Lemma 18.4.1. *In the above setting, $|F| = q$.*

Proof. Since $|F|$ is the number of distinct zeros of $x^q - x$ in its splitting field, it is enough to show that $x^q - x$ does not have multiple zeros in its splitting fields. By Proposition 18.3.4, $f(x) := x^q - x$ does not have multiple zeros in E if and only if $\gcd(f, f') = 1$ in $F[x]$. Notice that $f'(x) = qx^{q-1} - 1 = -1$ in $F[x]$ as $\mathrm{char}(F) = p$. Hence $\gcd(f, f') = 1$, and the claim follows. □

Altogether, we have proved:

Theorem 18.4.2 (Existence)**.** *Suppose p is prime and n is positive integer. Then there is a finite field of order p^n.*

Theorem 18.4.3 (Construction)**.** *Finite field of order p^n is a splitting field of $x^{p^n} - x$ over $\mathbb{Z}_p$.*

We let $\mathbb{F}_{p^n}$ denote a finite field of order p^n. Notice that by Theorem 18.4.2, there is such a finite field, and by Theorem 18.1.4, $\mathbb{F}_{p^n}$ is unique up to an isomorphism.

Chapter 19

Lecture 19

19.1 Vector spaces over a field

Let's recall a couple of results that we have proved a while ago.

Proposition. *(See Proposition 8.3.1) Suppose F is a field and $f(x) \in F[x]$ is a polynomial of degree n. Then every element of $F[x]/\langle f\rangle$ can be uniquely written as*

$$c_0\overline{1} + c_1\overline{x} + \cdots + c_{n-1}\overline{x}^{n-1}$$

for some $c_0, \ldots, c_{n-1} \in F$ where $\overline{1} := 1 + \langle f\rangle$ and $\overline{x} := x + \langle f\rangle$.

Proposition. *(See Theorem 9.1.1) Suppose E is a field extension oof F, and $\alpha \in E$ is algebraic over F. Suppose $\deg m_{\alpha,F} = n$. Then every element of $F[\alpha]$ can be uniquely written as*

$$c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}$$

for some $c_0, \cdots, c_{n-1} \in F$.

In both of these statements, elements are *uniquely* written as an F-linear combination of certain elements. This is similar to the main property of a *basis* in a vector space. It brings us to the definition of a vector space over a field F.

Definition 19.1.1. *Suppose F is a field. We say V is a vector space over F if:*

(1) $(V, +)$ is an abelian group.

(2) There is a scalar multiplication $F \times V \to V$ and for every $c \in F$ and $v \in V$, the scalar multiplication of c by v is denoted by $c \cdot v$ (or simply cv). This scalar multiplication is supposed to have the following properties.

a) For every $c_1, c_2 \in F$ and $v \in V$,

$$(c_1 + c_2) \cdot v = c_1 \cdot v + c_2 \cdot v.$$

b) For every $c \in F$ and $v_1, v_2 \in V$,

$$c \cdot (v_1 + v_2) = c \cdot v_1 + c \cdot v_2.$$

c) For every $v \in V$, $1 \cdot v = v$.

Example 19.1.2. *Suppose F is a field and n is a positive integer. Then*

$$F^n := \underbrace{F \times \cdots \times F}_{n \text{ times}}$$

is a vector space with respect to the following scalar multiplication

$$c \cdot (a_1, \ldots, a_n) := (ca_1, \ldots, ca_n).$$

Another example which plays an important role in this course is the following.

Example 19.1.3. *Suppose A is a unital ring, F is a subfield of A, and $1_A = 1_F$. Then A is a vector space over F with respect to the following scalar multiplication:*

$$\forall c \in F, a \in A, \quad c \cdot a := ca$$

where ca is the multiplication in A.

Let's recall some basic terminologies in linear algebra.

Definition 19.1.4. *Suppose V is a vector space over a field F.*

1. *We say $v_1, \cdots, v_n \in V$ are F-*linearly independent *if, for $c_1, \ldots, c_n \in F$,*

$$c_1 \cdot v_1 + \cdots + c_n v_n = 0 \textit{ implies that } c_1 = \cdots = c_n = 0.$$

2. *If $v_1, \ldots, v_n \in V$ are not F-linearly independent, we say they are F-*linearly dependent.

3. *We say $\{v_1, \ldots, v_n\} \subseteq V$ is an F-*spanning set *if every element of V can be written as an F-linear combination of $v_1, \ldots, v_n$; that means for every $v \in V$ there are $c_1, \ldots, c_n \in F$ such that*

$$v = c_1 \cdot v_1 + \cdots + c_n \cdot v_n.$$

When $\{v_1, \ldots, v_n\}$ is an F-spanning set, we say $v_1, \ldots, v_n$ span V.

4. *We say $(v_1, \ldots, v_n)$ is an F-*basis of V *if $v_1, \ldots, v_n$ are F-linearly independent and $\{v_1, \ldots, v_n\}$ is an F-spanning set.*

Though a basis is formally an *ordered* set, we sometimes refer to a set as a basis if it is an F-spanning set and consists of F-linearly independent vectors.

19.2 Subspace and linear map

As always, when we learn a new math object, we should talk about its substructures and the maps that preserves its structure.

Definition 19.2.1. *Suppose V is vector space over a field F. We say $W \subseteq V$ is a* subspace *of V if W is closed under addition and scalar multiplication.*

Definition 19.2.2. *Suppose V_1 and V_2 are two vector spaces over a field F.*

1. *We say $f : V_1 \to V_2$ is F-linear if*

$$f(v + v') = f(v) + f(v') \quad \textit{and} \quad f(c \cdot v) = c \cdot f(v)$$

for every $c \in F$ and $v, v' \in V$; alternatively we can write $f(cv + v') = cf(v) + f(v')$.

2. *We say $f : V_1 \to V_2$ is an* isomorphism *of F-vector spaces if*
   a) *f is F-linear,*
   b) *f is bijective, and*
   c) *f^{-1} is F-linear.*

It is a good exercise to show that if f is F-linear and it is bijective, then f^{-1} is F-linear. So the last condition for being an F-vector space isomorphism is redundant.

Let's also point out that similar to Lemma 1.3.1, one can use the distribution properties and show that

$$0_F \cdot v = 0_V \quad \text{and} \quad c \cdot 0_V = 0_V$$

for every $v \in V$ and $c \in F$.

Lemma 19.2.3. *Suppose V is a vector space over a field F, and $v_1, \ldots, v_n \in V$. Then the smallest subspace of V which contains v_i's is*

$$\Big\{ \sum_{i=1}^{n} c_i v_i \mid c_i \in F \Big\}.$$

*(This is denoted by $\mathrm{Span}_F\{v_1, \ldots, v_n\}$ or $\mathrm{Span}_F(v_1, \ldots, v_n)$, and it is called either the F-*span *of v_i's, or the* subspace spanned *by $v_1, \ldots, v_n$).*

Proof. Suppose W is a subspace of V which contains v_i's. Since W is closed under scalar multiplication, we have $c_i v_i \in W$ for every $c_i \in F$. Since W is closed under addition, we deduce that $\sum_{i=1}^n c_i v_i \in W$. Hence $\mathrm{Span}_F(v_1, \ldots, v_n) \subseteq W$.

Next we show that $\mathrm{Span}(v_1, \ldots, v_n)$ is a subspace. Suppose $c \in F$ and w, w' are in $\mathrm{Span}_F(v_1, \ldots, v_n)$. Then $w = \sum_{i=1}^n c_i v_i$ and $w' = \sum_{i=1}^n c_i' v_i$ for some $c_i, c_i' \in F$. Hence

$$cw + w' = c \sum_{i=1}^{n} c_i v_i + \sum_{i=1}^{n} c_i' v_i = \sum_{i=1}^{n} (cc_i + c_i') v_i \in \mathrm{Span}_F(v_1, \ldots, v_n).$$

Therefore $\mathrm{Span}_F(v_1, \ldots, v_n)$ is a subspace.

Finally we notice that

$$v_i = 0 \cdot v_1 + \cdots + 0 \cdot v_{i-1} + 1 \cdot v_i + 0 \cdot v_{i+1} + \cdots + 0 \cdot v_n \in \mathrm{Span}_F(v_1, \ldots, v_n).$$

Altogether, we proved that $\mathrm{Span}_F(v_1, \ldots, v_n)$ is a subspace which contains v_i's and every other subspace that contains v_i' contains $\mathrm{Span}_F(v_1, \ldots, v_n)$ as a subset. This completes the proof. □

Next lemma shows the importance of Example 19.1.2.

Lemma 19.2.4. *Suppose V is a vector space over a field F, and $\mathfrak{B} := (v_1, \ldots, v_n)$ is an F-basis of V. Then*

1. *for every $v \in V$, there is a unique*

$$(c_1, \ldots, c_n) \in F^n$$

such that $v = c_1 v_1 + \cdots + c_n v_n$. We let $[v]_{\mathfrak{B}} := (c_1, \ldots, c_n)$.

2. *The map $V \to F^n, v \mapsto [v]_{\mathfrak{B}}$ is a vector space isomorphism.*

Proof. (1) Since $\mathfrak{B}$ spans V, every $v \in V$ can be written as an F-linear combination of v_i's; that means that there are c_i's in F such that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

Now we want to show the uniqueness. So suppose $\sum_{i=1}^n c_i v_i = \sum_{i=1}^n c_i' v_i$ for some $c_i, c_i' \in F$. Then

$$(c_1 - c_1')v_1 + \cdots + (c_n - c_n')v_n = 0. \tag{19.1}$$

As v_i's are F-linearly independent and $c_i - c_i' \in F$, by (19.1) we have that $c_i - c_i' = 0$ for every i. Hence

$$(c_1, \ldots, c_n) = (c_1', \ldots, c_n').$$

(2) By part (1), $v \mapsto [v]_{\mathfrak{B}}$ is well-defined and it is the inverse function of

$$F^n \to V, \ \ (c_1, \ldots, c_n) \mapsto \sum_{i=1}^n c_i v_i.$$

Hence $v \mapsto [v]_{\mathfrak{B}}$ is a bijection. Let $[v]_{\mathfrak{B}} = (a_1, \ldots, a_n)$ and $[v']_{\mathfrak{B}} = (a_1', \ldots, a_n')$. Then $v = \sum_{i=1}^n a_i v_i$ and $v' = \sum_{i=1}^n a_i' v_i$. Therefore for every $c, c' \in F$, we have $cv + cv' = \sum_{i=1}^n (ca_i + a' a_i')v_i$, which implies that

$$\begin{aligned}[cv + cv']_{\mathfrak{B}} =& (ca_1 + c'a_1', \ldots, ca_n + c'a_n') \\ =& c(a_1, \ldots, a_n) + c'(a_1', \ldots, a_n') \\ =& c[v]_{\mathfrak{B}} + c'[v']_{\mathfrak{B}},\end{aligned}$$

this completes the proof. □

19.3 Dimension of a vector space

The following theorem plays helps us define the dimension of a vector space and more.

Theorem 19.3.1. *Suppose V is a vector space over a field F. Suppose $\{v_1, \ldots, v_n\}$ is an F-spanning set, and $w_1, \ldots, w_m$ are F-linearly independent. Then $n \geq m$.*

Proof. Inductively we will find distinct indexes $i_1, \ldots, i_m$ such that for every integer k in $[0, m]$,

$$(\{v_1, \ldots, v_n\} \setminus \{v_{i_1}, \ldots, v_{i_k}\}) \cup \{w_1, \ldots, w_k\}$$

is an F-spanning set. We are substituting w_j for v_{i_j} in $\{v_1, \ldots, v_n\}$ and still spanning V.

Notice that finding these distinct indexes

$$1 \leq i_1, \ldots, i_m \leq n$$

implies that $m \leq n$, and the claim follows.

The base of induction ($k = 0$) follows from the assumption that $\{v_1, \ldots, v_n\}$ is an F-spanning set. Now we show the induction step. Suppose we have already found $i_1, \ldots, i_k$ such that

$$(\{v_1, \ldots, v_n\} \setminus \{v_{i_1}, \ldots, v_{i_k}\}) \cup \{w_1, \ldots, w_k\}$$

is an F-spanning set. To simplify our notation, after rearranging v_i's, we can and will assume that $i_1 = 1, \ldots, i_k = k$; and so

$$\mathrm{Span}_F(w_1, \ldots, w_k, v_{k+1}, \ldots, v_n) = V. \tag{19.2}$$

In particular, w_{k+1} can be written as an F-linear combination of $w_1, \ldots, w_k, v_{k+1}, \ldots, v_n$. Hence there are c_i's in F such that

$$w_{k+1} = c_1 w_1 + \cdots + c_k w_k + c_{k+1} v_{k+1} + \cdots + c_n v_n. \tag{19.3}$$

Claim. There exists $j \geq k+1$ such that $c_j \neq 0$.
Proof of Claim. If not, $w_{k+1} = \sum_{i=1}^{k} c_i w_i$. This contradicts the assumption that w_i's are F-linearly independent.

Without loss of generality, after rearranging v_l's, we can and will assume that $c_{k+1} \neq 0$.

Claim. $\mathrm{Span}_F(w_1, \ldots, w_{k+1}, v_{k+2}, \ldots, v_n) = V$.
Proof of Claim. Because of (19.2), to show the Claim it is sufficient to prove that v_{k+1} is in the F-span of $w_1, \ldots, w_{k+1}, v_{k+2}, \ldots, v_n$. By (19.3),

$$c_{k+1} v_{k+1} = -\sum_{i=1}^{k} c_i w_i + w_{k+1} - \sum_{i=k+2}^{n} c_i v_i.$$

Notice that since $c_{k+1} \neq 0$ and F is a field, c_{k+1}^{-1} exists. Hence

$$\begin{aligned} v_{k+1} &= -\sum_{i=1}^{k}(c_{k+1}^{-1}c_i)w_i + c_{k+1}^{-1}w_{k+1} - \sum_{i=k+2}^{n}(c_{k+1}^{-1}c_i)v_i \\ &\in \operatorname{Span}_F(w_1,\ldots,w_{k+1},v_{k+2},\ldots,v_n), \end{aligned}$$

and the claim follows. □

Theorem 19.3.2. *Suppose V is a vector space over a field F. Suppose V is the F-span of a finite set $\{v_1,\ldots,v_n\}$. Then*

1. *V has an F-basis which is a subset of $\{v_1,\ldots,v_n\}$.*
2. *If $\mathfrak{B} := (w_1,\ldots,w_m)$ and $\mathfrak{B}' := (w'_1,\ldots,w'_k)$ are two F-bases, then $m = k$.*

The size of a basis of V is called the *dimension* of V over F and we denote it by $\dim_F V$.

Proof of Theorem 19.3.2. (1) Suppose $\{v_{i_1},\ldots,v_{i_m}\}$ is a maximal subset of $\{v_1,\ldots,v_n\}$ that consists of F-linearly independent vectors. Then for every $j \notin \{i_1,\ldots,i_m\}$, the vectors $v_{i_1},\ldots,v_{i_m},v_j$ are F-linearly dependent. This means there are $c_1,\ldots,c_{m+1} \in F$ that are not all zero and

$$c_1v_{i_1} + \cdots + c_mv_{i_m} + c_{m+1}v_j = 0.$$

Since $v_{i_1},\ldots,v_{i_m}$ are F-linearly independent, $c_{m+1} \neq 0$. Hence c_{m+1}^{-1} exists (as F is a field). Therefore

$$\begin{aligned} v_j &= -(c_{m+1}^{-1}c_1)v_{i_1} - \cdots - (c_{m+1}^{-1}c_m)v_{i_m} \\ &\in \operatorname{Span}_F(v_{i_1},\ldots,v_{i_m}). \end{aligned} \tag{19.4}$$

Since (19.4) holds for every j not in $\{i_1,\ldots,i_m\}$, we deduce that

$$\operatorname{Span}_F(v_{i_1},\ldots,v_{i_m}) = \operatorname{Span}_F(v_1,\ldots,v_n) = V.$$

Hence $(v_{i_1},\ldots,v_{i_m})$ is an F-basis as it consists of F-linearly independent vectors and it is an F-spanning set.

(2) Since $\{w_1,\ldots,w_m\}$ is an F-spanning set and $w'_1,\ldots,w'_k$ are F-linearly independent, by Theorem 19.3.1 we have $k \leq m$. Similarly, since $\{w'_1,\ldots,w'_k\}$ is an F-spanning set and $w_1,\ldots,w_m$ are F-linearly independent, by Theorem 19.3.1 we have $m \leq k$. Altogether we get $m = k$, and this completes the proof. □

19.4 Quotient spaces

Similar to groups and rings, we want to define the quotient of a vector space. Suppose V is a vector space over a field F, and W is a subspace of V. Then in particular W is a (normal) subgroup of V. Hence we can consider the abelian group V/W.

Proposition 19.4.1. *Suppose V is a vector space over a field F, and W is a subspace of V. Then the following is a well-defined scalar multiplication*

$$F \times V/W \to V/W, \quad (c, v+W) \mapsto c \cdot (v+W) := cv + W.$$

Moreover V/W with its quotient abelian group structure and the above given scalar product is an F-vector space.

Proof. Let's start with arguing why $\cdot$ is a well-defined operation. So assuming $v_1+W = v_2+W$, we have to show that $cv_1 + W = cv_2 + W$ for every $c \in F$. Notice that $v_1+W = v_2+W$ implies that $v_1 - v_2 \in W$. As W is closed under scalar multiplication, we have that $c(v_1 - v_2) \in W$ for every c in F. Therefore $cv_1 - cv_2 \in W$, from which we deduce that $cv_1 + W = cv_2 + W$. This shows that $\cdot$ is a well-defined operation.

Next, we check why V/W is an F-vector space. For every $c \in F$ and $v_1, v_2 \in V$, we have

$$\begin{aligned} c \cdot ((v_1 + W) + (v_2 + W)) =& c \cdot ((v_1 + v_2) + W) \\ =& c(v_1 + v_2) + W \\ =& (cv_1 + cv_2) + W \\ =& (cv_1 + W) + (cv_2 + W) \\ =& c \cdot (v_1 + W) + c \cdot (v_2 + W). \end{aligned}$$

Similarly we can check that

$$(c_1 + c_2) \cdot (v + W) = c_1 \cdot (v + W) + c_2 \cdot (v + W)$$

for every $c_1, c_2 \in F$ and $v \in V$.

Finally we observe that

$$1 \cdot (v + W) = (1\ v) + W = v + W$$

for every $v \in V$. This completes the proof. □

Notice that the natural quotient map

$$p_W : V \to V/W, \ p_W(v) := v + W$$

is F-linear and $\ker p_W = W$.

Since by Lemma 19.2.4 an F-vector space of a given dimension is unique up to an isomorphism, we want to understand the dimension of V/W.

Proposition 19.4.2. *Suppose V is a vector space over F and W is a subspace of V. Then*

$$\dim_F W + \dim_F V/W = \dim_F V;$$

in particular if one of the sides is finite, then the other side is finite as well.

Proof. First notice that if $\dim_F V < \infty$, then there is a finite F-spanning set $\{v_1, \ldots, v_n\}$. Then by Theorem 19.3.1, every subset of W that consists of F-linearly independent vectors has cardinality at most n; in particular, $\dim_F W < \infty$. We also observe that $\{v_1 + W, \ldots, v_n + W\}$ is an F-spanning subset of V/W, and so by the first part of Theorem 19.3.2, we have $\dim_F V/W < \infty$. Hence from this point on, we can and will assume that

$$\dim_F W = m < \infty \quad \text{and} \quad \dim_F V/W = k < \infty.$$

Suppose $(w_1, \ldots, w_m)$ is an F-basis of W, and $(v_1 + W, \ldots, v_k + W)$ is an F-basis of V/W. We show that $(w_1, \ldots, w_m, v_1, \cdots, v_k)$ is an F-basis of V. We prove this in two steps. First we show this is an F-spanning set, and second we show that it consists of F-linearly independent vectors.

Step 1. $\operatorname{Span}_F(w_1, \ldots, w_m, v_1, \cdots, v_k) = V$.

Proof of Step 1. Let $W' := \operatorname{Span}_F(w_1, \ldots, w_m, v_1, \cdots, v_k)$. Then $W \subseteq W'$ as w_i's span W and they are in W'. Hence W'/W is a subspace of V/W. Since $v_i + W$'s are in W'/W and they span V/W, we deduce that $W'/W = V/W$. Therefore by the correspondence theorem for the subgroups of a quotient group, we have that $V = W'$. (We can avoid using the correspondence theorem and use the following argument: for every $v \in V$, knowing that $v + W \in W'/W$, we can deduce that there is $w' \in W$ such that $v + W = w' + W$. This means $v - w' = w$ for some $w \in W \subseteq W'$. Hence

$$v = w' + w \in W'.$$

Altogether we proved that every element v of V is in W. Therefore $V = W'$.)

Step 2. $w_1, \ldots, w_m, v_1, \cdots, v_k$ are F-linearly independent.

Proof of Step 2. Suppose

$$\sum_{i=1}^{m} c_i w_i + \sum_{j=1}^{k} c_{m+j} v_j = 0 \tag{19.5}$$

for some c_i's in F. Then

$$p_W\Big(\sum_{i=1}^{m} c_i w_i + \sum_{j=1}^{k} c_{m+j} v_j\Big) = 0,$$

where $p_W : V \to V/W, \quad p_W(v) := v + W$ is the natural quotient map. Since $W = \ker p_W$, we obtain that

$$\sum_{j=1}^{k} c_{m+j} \cdot p_W(v_j) = 0,$$

and so

$$c_{m+1} \cdot (v_1 + W) + \cdots + c_{m+k} \cdot (v_k + W) = 0. \tag{19.6}$$

As $v_j + W$'s are F-linearly independent, we deduce that $c_{m+1} = \cdots = c_{m+k} = 0$. Hence by (19.5), we obtain that

$$\sum_{i=1}^{m} c_i w_i = 0.$$

As w_i's are F-linearly independent, we have $c_1 = \cdots = c_m = 0$. Altogether we deduce that all the coefficients in (19.6) are zero. This completes the proof of the second step.

By Steps 1 and 2, we obtain that

$$(w_1, \ldots, w_m, v_1, \ldots, v_k)$$

is an F-basis. Hence

$$\dim_F V = m + k = \dim_F W + \dim_F V/W,$$

which completes the proof. □

19.5 The first isomorphism theorem for vector spaces

Similar to groups and rings, next we prove the first isomorphism theorem. Then we use this result to show the kernel-image theorem.

Theorem 19.5.1. *Suppose V_1 and V_2 are two F-vector spaces, and $f : V_1 \to V_2$ is an F-linear map. Then*

1. $\mathrm{Im}(f)$ *is a subspace of* V_2*, and* $\ker f$ *is a subspace of* V_1.
2. $\overline{f} : V_1/\ker f \to \mathrm{Im}\, f, \quad \overline{f}(v_1 + \ker f) := f(v_1)$ *is an isomorphism of F-vector spaces.*
3. $\dim_F(\ker f) + \dim_F(\mathrm{Im}\, f) = \dim_F V_1$.

Proof. (1) Since f is an additive group homomorphism, $\mathrm{Im}\, f$ is a subgroup of V_2 and $\ker f$ is a subgroup of V_1. So it is sufficient to prove that $\mathrm{Im}\, f$ and $\ker f$ are closed under scalar multiplication. Suppose $v_2 \in \mathrm{Im}\, f$. Then $v_2 = f(v_1)$ for some $v_1 \in V_1$. Hence for every $c \in F$, we have

$$cv_2 = cf(v_1) = f(cv_1) \in \mathrm{Im}\, f.$$

This shows that $\mathrm{Im}\, f$ is closed under scalar multiplication, and so it is a subspace of V_2.

Suppose $v_1 \in \ker f$ and $c \in F$. Then

$$f(cv_1) = cf(v_1) = c0 = 0,$$

which implies that $cv_1 \in \ker f$. Hence $\ker f$ is closed under scalar multiplication, which implies that $\ker f$ is a subspace of V_1.

(2) By the first isomorphism theorem for groups, we have that

$$\overline{f} : V_1/\ker f \to \mathrm{Im}\, f, \quad \overline{f}(v_1 + \ker f) := f(v_1)$$

is a well-defined group isomorphism. So to show that $\overline{f}$ is an F-vector space isomorphism, it suffices to argue why $\overline{f}$ preserves the scalar multiplication. For every $c \in F$

and $v_1 \in V_1$, we have

$$\begin{aligned}\overline{f}(c \cdot (v_1 + \ker f)) =& \overline{f}(cv_1 + \ker f)\\ =& f(cv_1)\\ =& cf(v_1)\\ =& c \cdot \overline{f}(v_1 + \ker f),\end{aligned}$$

and part (2) follows.

(3) By Proposition 19.4.2 and the second part, we have

$$\dim_F(\operatorname{Im} f) = \dim_F(V_1/\ker f) = \dim_F V_1 - \dim_F(\ker f),$$

and the claim follows. □

Chapter 20

Lecture 20

We have proved basic properties of vector spaces over a field F. Here we will explore their implications in *field theory*.

20.1 Previous results in the language of linear algebra

We have motivated our digression to vector spaces over fields by considering the conclusions of Proposition 8.3.1 and Theorem 9.1.1. Here we rephrase those conclusions using terminologies from linear algebra.

Proposition 20.1.1. *Suppose F is a field and $f(x) \in F[x]$ is a polynomial of degree n, where n is a positive integer. Then $(\overline{1}, \overline{x}, \ldots, \overline{x}^{n-1})$ is an F-basis of $F[x]/\langle f\rangle$, where $\overline{x}^i := x^i + \langle f\rangle$ for every integer i in $[0, n-1]$. In particular,* $\dim_F F[x]/\langle f\rangle = \deg f$.

Proof. By Proposition 8.3.1, every element of $F[x]/\langle f\rangle$ can be uniquely written as

$$(c_0 + c_1 x + \cdots + c_{n-1}x^{n-1}) + \langle f\rangle = \sum_{i=0}^{n-1} c_i \overline{x}^i.$$

Hence the F-span of $\{\overline{1}, \overline{x}, \ldots, \overline{x}^{n-1}\}$ is $F[x]/\langle f\rangle$. Moreover if $\sum_{i=0}^{n-1} c_i\overline{x}^i = 0$, then because of the uniqueness the above expression we obtain that c_i's are 0. This implies that $\overline{1}, \overline{x}, \ldots, \overline{x}^{n-1}$ are F-linearly independent. The claim follows. □

Proposition 20.1.2. *Suppose E is a field extension of F, and $\alpha \in E$ is algebraic over F. Then $(1, \alpha, \ldots, \alpha^{n-1})$ is an F-basis of $F[\alpha]$ where $n = \deg m_{\alpha,F}$. In particular,* $\dim_F F[\alpha] = \deg m_{\alpha,F}$.

Proof. By Theorem 9.1.1, every element of $F[\alpha]$ can be uniquely written as

$$c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}$$

for some $c_0, \ldots, c_{n-1} \in F$ where $n = \deg m_{\alpha,F}$. Hence the F-span of $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is $F[\alpha]$. Moreover if $\sum_{i=0}^{n-1} c_i\alpha^i = 0$, then because of the uniqueness the above expression we obtain that c_i's are 0. This implies that $1, \alpha, \ldots, \alpha^{n-1}$ are F-linearly independent. The claim follows. □

20.2 Finite fields and vector spaces

Suppose F is a finite field and V is a vector space over F. If $\dim_F V = n$, then by Lemma 19.2.4, we have that $V \simeq F^n$, and so

$$|V| = |F|^{\dim_F V}. \tag{20.1}$$

This helps us get a strong condition for the tower of finite fields.

Proposition 20.2.1. *If $\mathbb{F}_{p^m}$ can be embedded into $\mathbb{F}_{p^n}$, then $m|n$.*

Proof. If $\mathbb{F}_{p^m}$ can be embedded into $\mathbb{F}_{p^n}$, we can view $\mathbb{F}_{p^n}$ as a vector space over $\mathbb{F}_{p^m}$. Since these are finite sets, $\dim_{\mathbb{F}_{p^m}} \mathbb{F}_{p^n} = d < \infty$. Hence by (20.1), we have

$$|\mathbb{F}_{p^n}| = |\mathbb{F}_{p^m}|^d, \quad \text{which implies that} \quad n = md.$$

This completes the proof. □

One can use the cardinality of the group of units of finite fields to prove the same result. Assuming that $\mathbb{F}_{p^m}$ can be embedded in $\mathbb{F}_{p^n}$, we deduce that the group of units of $\mathbb{F}_{p^m}$ can be embedded into the group of units of $\mathbb{F}_{p^n}$. Hence $p^m - 1 | p^n - 1$. From this one can show that $m|n$. As you can see the presented proof, which is based on linear algebra, is much more natural.

Exercise 20.2.2. *Suppose m and n are positive integers and $m|n$. Prove that $\mathbb{F}_{p^m}$ can be embedded into $\mathbb{F}_{p^n}$.*

20.3 Tower rule for field extensions

We have already seen in Proposition 20.2.1 how useful it is to think about a field extension E of F as an F-vector space.

Definition 20.3.1. *Suppose E is a field extension of F. Then we can view E as an F-vector space (see Example 19.1.3). The dimension $\dim_F E$ of E as an F-vector space is denoted by $[E : F]$ and it is called the* degree of this field extension.

Theorem 20.3.2 (Tower rule)**.** *Suppose L is a field extension of E, and E is a field extension of F. Then*

$$[L : F] = [L : E][E : F];$$

in particular, if one of the sides is finite, then the other side is finite as well.

$$\begin{array}{c} L \\ | \\ E \\ | \\ F \end{array} \tag{20.2}$$

We often use a diagram as in (20.2) to show field extensions. In this type of diagram, we connect two fields if one is a subfield of the other. The subfield is located lower than the larger field.

Proof of Theorem 20.3.2. If $[L : F] = n < \infty$, then there is a finite F-spanning set $\{v_1, \ldots, v_n\}$. Hence the L-span of $\{v_1, \ldots, v_n\}$ is also L, and so by Theorem 19.3.2, $[L : E] \le n$. And also, by Proposition 19.4.2, we have

$$[E : F] = \dim_F E \le \dim_F L < \infty.$$

Therefore from this point on, we can and will assume that

$$[L : E] = m < \infty \quad \text{and} \quad [E : F] = n < \infty.$$

Suppose $(\ell_1, \ldots, \ell_m)$ is an E-basis of L, and $(e_1, \ldots, e_n)$ is an F-basis of E. We prove that

$$\{\ell_i e_j \mid 1 \le i \le m, 1 \le j \le n\}$$

(with respect to some ordering) is an F-basis of L. We do this in two steps.

Step 1. $\mathrm{Span}_F(\ell_i e_j \mid 1 \le i \le m, 1 \le j \le n) = L$.

Proof of Step 1. Every $\ell \in L$ can be written as an E-linear combination of ℓ_i's. This means there are $x_i \in E$ such that

$$\ell = x_1\ell_1 + \cdots + x_m\ell_m. \tag{20.3}$$

Every element of E can be written as an F-linear combination of e_j's. Hence for every i, there are $y_{ij} \in F$ such that

$$x_i = y_{i1}e_1 + \cdots + y_{in}e_n. \tag{20.4}$$

By (20.3) and (20.4), we deduce that

$$\begin{aligned}\ell &= \sum_{i=1}^{m} x_i\ell_i = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} y_{ij}e_j\Big)\ell_i \\ &= \sum_{1\le i\le m, 1\le j\le n} y_{ij}\ \ell_i e_j.\end{aligned}$$

This means ℓ can be written as an F-linear combination of $\ell_i e_j$'s. This completes the proof of Step 1.

Step 2. $\ell_i e_j$'s are F-linearly independent.

Proof of Step 2. Suppose

$$\sum_{1\le i\le m, 1\le j\le n} y_{ij}\ \ell_i e_j = 0 \tag{20.5}$$

for some y_{ij}'s in F. Then by (20.5), we have

$$\sum_{i=1}^{m} \Big(\sum_{j=1}^{n} y_{ij}e_j\Big)\ell_i = 0. \tag{20.6}$$

Notice that for every i, $x_i := \sum_{j=1}^{n} \sum_{j=1}^{n} y_{ij}e_j$ is in E. Since ℓ_i's are E-linearly independent, by (20.6) we deduce that $x_i = 0$ for every i. Hence we have

$$\sum_{j=1}^{n} \sum_{j=1}^{n} y_{ij}e_j = 0 \tag{20.7}$$

for every index i. Since e_j's are F-linearly independent, by (20.7) we obtain that $y_{ij} = 0$ for every pair of indexes i and j. This completes the proof of the second step.

By Steps 1 and 2, we deduce that

$$[L : F] = |\{\ell_i e_j \mid 1 \le i \le m, 1 \le j \le n\}| = mn = [L : E][E : F],$$

which completes the proof. □

20.4 Some applications of the Tower Rule for field extensions

Here we mention some examples on how one can use the Tower Rule.

Example 20.4.1. *Suppose E is a field extension of $\mathbb{Q}$ such that $[E : \mathbb{Q}] = 2^n$ for some positive integer n. Then $x^3 - 2$ is irreducible in $E[x]$.*

Proof. Suppose to the contrary that $x^3 - 2$ is not irreducible in $E[x]$. Then by the irreducibility criterion for degree 2 and 3 polynomials (see 10.1.1), we deduce that there is $\alpha \in E$ which is a zero of $x^3 - 2$. Then $\mathbb{Q}[\alpha]$ is an *intermediate field*; that means we have the tower of field extensions given in (20.9). Hence by the Tower Rule (see Theorem 20.3.2), we have $[E : \mathbb{Q}] = [E : \mathbb{Q}[\alpha]][\mathbb{Q}[\alpha] : \mathbb{Q}]$. Therefore by Proposition 20.1.2 and our hypothesis, we obtain that

$$\begin{array}{c} E \\ | \\ \mathbb{Q}[\alpha] \\ | \\ \mathbb{Q} \end{array} \qquad (20.9)$$

$$2^n = [E : \mathbb{Q}[\alpha]] \deg m_{\alpha,\mathbb{Q}}. \qquad (20.8)$$

So we it is useful to find the minimal polynomial $m_{\alpha,\mathbb{Q}}$ of α over $\mathbb{Q}$. Notice that α is a zero of $x^3 - 2$, $x^3 - 2$ is monic, and by Eisenstein's irreducibility criterion (see Theorem 12.2.1), $x^3 - 2$ is irreducible in $\mathbb{Q}[x]$. Hence by Theorem 8.2.5, $m_{\alpha,\mathbb{Q}}(x) = x^3 - 2$. Thus by (20.8), 3 is a divisor of 2^n which is a contradiction. This completes the proof. □

Example 20.4.2. *Suppose E is a field extension of F and $\alpha \in E$ is algebraic over F. Suppose $[F[\alpha] : F]$ is odd. Then $F[\alpha] = F[\alpha^2]$.*

Proof. Notice that $F[\alpha]$ is a field extension of $F[\alpha^2]$. So we have the tower of field extensions given in (20.11). Hence by the Tower Rule (see Theorem 20.3.2), we have

$$[F[\alpha] : F] = [F[\alpha] : F[\alpha^2]][F[\alpha^2] : F].$$

$$\begin{array}{c} F[\alpha] \\ | \\ F[\alpha^2] \\ | \\ F \end{array} \qquad (20.11)$$

Therefore by Proposition 20.1.2 and our hypothesis, we obtain that

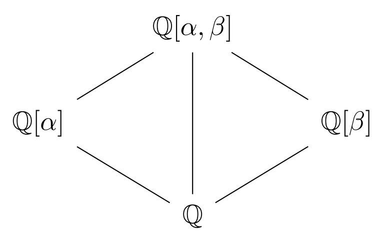
$$(\deg m_{\alpha,F[\alpha^2]})[F[\alpha^2] : F] \text{ is odd.} \qquad (20.10)$$

By (20.10), we have that $\deg m_{\alpha,F[\alpha^2]}$ is odd. Notice that α is a zero of $x^2 - \alpha^2 \in F[\alpha^2]$. Hence $\deg m_{\alpha,F[\alpha^2]} \le 2$. As the only positive odd integer less than or equal to 2 is 1, we have $\deg m_{\alpha,F[\alpha^2]} = 1$. This implies that $\alpha \in F[\alpha^2]$, and so $F[\alpha] \subseteq F[\alpha^2]$. The claim follows. □

Example 20.4.3. *Suppose $\alpha, \beta \in \mathbb{C}$ are algebraic over $\mathbb{Q}$. Let $f(x) := m_{\alpha,\mathbb{Q}}(x)$ and $g(x) := m_{\beta,\mathbb{Q}}(x)$. Then*

$$f \textit{ is irreducible in } (\mathbb{Q}[\beta])[x] \iff g \textit{ is irreducible in } (\mathbb{Q}[\alpha])[x].$$

Proof. ($\Rightarrow$) We will be using the right and the left legs of the diagram given in (20.12).



(20.12)

Going through the right leg of the diagram in (20.12), using the Tower Rule (see Theorem 20.3.2) and Proposition 20.1.2, we obtain that

$$\begin{aligned}[\mathbb{Q}[\alpha,\beta] : \mathbb{Q}] =& [\mathbb{Q}[\alpha,\beta] : \mathbb{Q}[\beta]][\mathbb{Q}[\beta] : \mathbb{Q}] \\ =& (\deg m_{\alpha,\mathbb{Q}[\beta]})(\deg m_{\beta,\mathbb{Q}}).\end{aligned} \tag{20.13}$$

Notice that since α is a zero of f, f is monic and irreducible in $(\mathbb{Q}[\beta])[x]$, by Theorem 8.2.5, we have $m_{\alpha,\mathbb{Q}[\beta]}(x) = f(x) = m_{\alpha,\mathbb{Q}}(x)$. Hence, by (20.13), we have that

$$[\mathbb{Q}[\alpha,\beta] : \mathbb{Q}] = (\deg m_{\alpha,\mathbb{Q}})(\deg m_{\beta,\mathbb{Q}}). \tag{20.14}$$

Going through the left leg of the diagram in (20.12), using the Tower Rule (see Theorem 20.3.2) and Proposition 20.1.2, we obtain that

$$\begin{aligned}[\mathbb{Q}[\alpha,\beta] : \mathbb{Q}] =& [\mathbb{Q}[\alpha,\beta] : \mathbb{Q}[\alpha]][\mathbb{Q}[\alpha] : \mathbb{Q}] \\ =& (\deg m_{\beta,\mathbb{Q}[\alpha]})(\deg m_{\alpha,\mathbb{Q}}).\end{aligned} \tag{20.15}$$

By (20.14) and (20.15), we deduce that

$$\deg m_{\beta,\mathbb{Q}[\alpha]} = \deg m_{\beta,\mathbb{Q}}. \tag{20.16}$$

Notice that since β is a zero of $m_{\beta,\mathbb{Q}} \in (\mathbb{Q}[\alpha])[x]$, we obtain that

$$m_{\beta,\mathbb{Q}[\alpha]} | \deg m_{\beta,\mathbb{Q}}. \tag{20.17}$$

By (20.16) and (20.17), we obtain that $g(x) = m_{\beta,\mathbb{Q}}(x) = m_{\beta,\mathbb{Q}[\alpha]}(x)$. This implies that $g(x)$ is irreducible in $(\mathbb{Q}[\alpha])[x]$.

($\Leftarrow$) This direction follows by a similar argument. □

Let's remark that if L is a field extension of E, E is a field extension of F, and $\alpha \in L$ is algebraic over F, then α is a zero of $m_{\alpha,F} \in E[x]$. Hence $m_{\alpha,E} | m_{\alpha,F}$; this is a generalization of (20.17).

20.5 Algebraic closure in a field extension

Suppose E is a field extension of F. The *algebraic closure of* F *in* E is the set of all the elements of E that are algebraic over F. Here we will prove that the algebraic closure of F in E is a subfield of E. We start with proving that a field extension of finite degree is an algebraic extension.

Lemma 20.5.1. *Suppose* E *is a field extension of* F *of finite degree. Then every* $\alpha \in E$ *is algebraic over* F.

We say a field extension E of F is an *algebraic extension* if every $\alpha \in E$ is algebraic. So we are proving that a field extension of finite degree is algebraic.

Proof of Lemma 20.5.1. Suppose $[E : F] = n$. Then by Theorem 19.3.1, every $n+1$ elements of E are F-linearly dependent. Hence $1, \alpha, \ldots, \alpha^n$ are F-linearly dependent. Thus there are $c_0, \ldots, c_n \in F$ that are not all zero and

$$c_0 + c_1\alpha + \cdots + c_n\alpha^n = 0.$$

This means that α is algebraic over F. □

Theorem 20.5.2 (Algebraic closure in a field extension)**.** *Suppose* E *is a field extension of* F*. Let*

$$K := \{\alpha \in E \mid \alpha \text{ is algebraic over } F\}.$$

Then K *is a field extension of* F.

Proof. Suppose $\alpha, \beta \in K$. Then by Proposition 20.1.2, we have $[F[\alpha] : F] = \deg m_{\alpha,F} < \infty$. Moreover as β is algebraic over $F[\alpha]$, we have $[F[\alpha, \beta] : F[\alpha]] < \infty$. Hence by the Tower Rule (see Theorem 20.3.2) we obtain that

$$[F[\alpha, \beta] : F] = [F[\alpha, \beta] : F[\alpha]][F[\alpha] : F] < \infty. \tag{20.18}$$

By Lemma 20.5.1 and (20.18), we deduce that $F[\alpha, \beta]$ is an algebraic extension of F; this means that $F[\alpha, \beta] \subseteq K$. This implies that $F \subseteq K$, $\alpha \pm \beta$ and $\alpha\beta$ are in K, and if $\beta \neq 0$, then $\alpha\beta^{-1}$ is in K, as well. Altogether, we deduce that K is field extension of F. This completes the proof. □

20.6 Tower of algebraic extensions

Here we show another application of the tower rule on algebraic extensions.

Proposition 20.6.1. *Suppose* E *is an algebraic extension field of* F *and* L *is an algebraic extension field of* E*. Then* L *is an algebraic extension field of* F.

Proof. Suppose $\alpha \in L$. Then α is algebraic over E. Suppose

$$m_{\alpha,E}(x) = x^n + e_{n-1}x^{n-1} + \cdots + e_0$$

is the minimal polynomial of α over E. Hence α is algebraic over $F[e_0, \ldots, e_{n-1}]$. Notice that since e_i's are algebraic over F, by the Tower rule by an argument similar to (20.18) inductively we can show that

$$[F[e_0, \ldots, e_{n-1}] : F] < \infty. \tag{20.19}$$

As α is algebraic over $F[e_0, \ldots, e_{n-1}]$,

$$[F[e_0, \ldots, e_{n-1}][\alpha] : F[e_0, \ldots, e_{n-1}]] < \infty. \tag{20.20}$$

Hence by the Tower Rule, (20.19), and (20.20), we have

$$[F[e_0, \ldots, e_{n-1}, \alpha] : F] < \infty.$$

Therefore by Lemma 20.5.1, α is algebraic over F. This completes the proof. □

20.7 Geometric constructions by ruler and compass

Let's recall some ancient Euclidean geometry problems. *Can we construct* $\sqrt[3]{2}, \pi$, *or angle* $20°$ *using ruler and compass?* Let's formulate it properly what it means to *construct* a number. We start with a unit segment. The end points are considered *constructed*. If two points are constructed, then the line which passes through them is considered constructed. The circles that are centered at one of these points and pass through the other are called constructed. The points of intersection of constructed circles and lines are considered constructed points. A number is called constructed if its absolute value is the distance of two constructed points. The following theorem gives us an excellent understanding of constructed points.

Theorem 20.7.1. *Suppose the initial points are* $(0, 0)$ *and* $(1, 0)$. *If* (α, β) *is a constructed point, then* $[\mathbb{Q}[\alpha] : \mathbb{Q}]$ *and* $[\mathbb{Q}[\beta] : \mathbb{Q}]$ *are powers of* 2.

Here only the main ideas will be presented. To get to the point (α, β), we have to construct finitely many lines and circles, and consider their intersection points. To find the coordinates of intersection points we end up solving degree 1 and degree 2 polynomials with coefficients that are in the ring generated by the coordinates of the constructed points that we have so far. This means there is a tower of field extensions

$$\mathbb{Q} =: F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n$$

such that $[F_{i+1} : F_i] = 2$ for every i and $\alpha \in F_n$. By the Tower Rule, $[F_n : \mathbb{Q}]$ is power of 2. Since $\mathbb{Q}[\alpha]$ is an intermediate subfield, by the Tower Rule we deduce that $[\mathbb{Q}[\alpha] : \mathbb{Q}]$ divides $[F_n : \mathbb{Q}]$. Hence $[\mathbb{Q}[\alpha] : \mathbb{Q}]$ is a power of 2.

Corollary 20.7.2. $\sqrt[3]{2}$ *and* π *cannot be constructed by ruler and compass.*

Proof. By Theorem 20.7.1, if α can be constructed by ruler and compass, then $[\mathbb{Q}[\alpha] : \mathbb{Q}]$ is a power of 2. In particular, α is algebraic. Hence π cannot be constructed (we do not prove this here, but it can be proved that π is not algebraic over $\mathbb{Q}$). Notice that $\deg m_{\sqrt[3]{2}, \mathbb{Q}} = 3$, and so $\sqrt[3]{2}$ cannot be constructed by ruler and compass. □

Exercise 20.7.3. *Show that* $\deg m_{\cos 20^\circ,\mathbb{Q}} = 3$*, and use this to deduce the angle* 20° *cannot be constructed by ruler and compass. (Hint:* $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$*.)*

By the above Exercise, we can deduce that there is no general method of dividing a given angle into three equal parts using only ruler and compass.

Chapter 21

Lecture 21

By now we have a basic understanding of vector spaces over a field and how it can help us study field extensions. We go back and further study splitting fields. Here we focus on the splitting field of $x^n - 1$ over $\mathbb{Q}$. Let us recall that by Example 17.2.1, we have that $\mathbb{Q}[\zeta_n]$ is a splitting field of $x^n - 1$ over $\mathbb{Q}$, where $\zeta_n := e^{2\pi i/n}$, and

$$x^n - 1 = (x-1)(x-\zeta_n)\cdots(x-\zeta_n^{n-1}). \tag{21.1}$$

This field has a historical significance, because of its role in the initial modern attempts towards proving Fermat's last conjecture. We want to answer a very basic question about this field: what is $[\mathbb{Q}[\zeta_n] : \mathbb{Q}]$? By Proposition 20.1.2, we have

$$[\mathbb{Q}[\zeta_n] : \mathbb{Q}] = \deg m_{\zeta_n,\mathbb{Q}}.$$

Hence we need to find the minimal polynomial of ζ_n over $\mathbb{Q}$.

21.1 Cyclotomic polynomials

In this section, we will arrive at the definition of the n-th cyclotomic polynomial. This will be done as we investigate the minimal polynomial of ζ_n over $\mathbb{Q}$.

Notice that since ζ_n is a zero of $x^n - 1$, $m_{\zeta_n,\mathbb{Q}}$ divides $x^n - 1$. Hence by (21.1) and the fact the $\mathbb{C}[x]$ is a UFD, we deduce that

$$m_{\alpha,\mathbb{Q}}(x) = (x-\zeta_n^{i_1})\cdots(x-\zeta_n^{i_m})$$

for some integers i_j's in $[1, n]$. As ζ_n is a zero of $m_{\zeta_n,\mathbb{Q}}$, without loss of generality, we can and will assume that $i_1 = 1$.

By Lemma 16.2.2, if E is a field extension of F and $\alpha, \alpha' \in E$ are two zeros of an irreducible polynomial $f \in F[x]$, then there is an F*-isomorphism* $\theta : F[\alpha] \to F[\alpha']$ such that $\theta(\alpha) = \alpha'$; an F-isomorphism is a ring isomorphism which F-linear. Applying this result for the two zeros ζ_n and $\zeta_n^{i_j}$ of the irreducible polynomial $m_{\zeta_n,\mathbb{Q}} \in \mathbb{Q}[x]$, we obtain a $\mathbb{Q}$-isomorphism $\theta_j : \mathbb{Q}[\zeta_n] \to \mathbb{Q}[\zeta_n^{i_j}]$ such that $\theta_j(\zeta_n) = \zeta_n^{i_j}$. Notice that, since θ_j is a ring isomorphism, the multiplicative order of ζ_n and $\theta_j(\zeta_n)$ are the same. As $o(g^k) = \frac{o(g)}{\gcd(o(g),k)}$, $o(\zeta_n) = n$, and $\theta_j(\zeta_n) = \zeta_n^{i_j}$, we deduce that $\gcd(n, i_j) = 1$ for every j. This takes us to the definition of the n-th cyclotomic polynomial.

Definition 21.1.1. *The n-th cyclotomic polynomial is*

$$\Phi_n(x) := \prod_{1\le i\le n, \gcd(i,n)=1} (x - \zeta_n^i).$$

In particular, Φ_n is a monic polynomial of degree $\phi(n)$, wherer ϕ is the Euler-phi function.

By the above discussion, we have that $m_{\zeta_n,\mathbb{Q}}(x)$ divides $\Phi_n(x)$ in $\mathbb{C}[x]$. We will prove that $m_{\zeta_n,\mathbb{Q}}(x) = \Phi_n(x)$.

21.2 Cyclotomic polynomials are integer polynomials

The following is a key property of cyclotomic polynomials that, among other things, help us prove cyclotomic polynomials are integer polynomials.

Theorem 21.2.1. *For every positive integer n, we have*

$$x^n - 1 = \prod_{d|n} \Phi_d(x). \tag{21.2}$$

Before we go to the details of the proof of Theorem 21.2.1, let us compare the degrees of the both sides of (21.2):

$$n = \sum_{d|n} \phi(d). \tag{21.3}$$

Proof of this formula is based on the partitioning of the set $[1..n] := \{1, \ldots, n\}$ in terms of the greatest common divisor of the elements with n. To be more precise, we let

$$C_{d,n} := \{i \in [1..n] \mid \gcd(i, n) = d\}. \tag{21.4}$$

Then $\{C_{d,n} \mid d|n\}$ is a partitioning of $[1..n]$. Moreover

$$i \in C_{d,n} \iff \gcd(i, n) = d \iff i = dj \text{ and } \gcd\left(j, \frac{n}{d}\right) = 1.$$

Hence

$$C_{d,n} = \{dj \mid j \in C_{1,\frac{n}{d}}\} \text{ which imlies that } |C_{d,n}| = |C_{1,\frac{n}{d}}| = \phi\Big(\frac{n}{d}\Big). \tag{21.5}$$

Therefore

$$n = |[1..n]| = \sum_{d|n} |C_{d,n}| = \sum_{d|n} \phi\Big(\frac{n}{d}\Big).$$

Finally we notice that as d ranges over all the positive divisors of n, so does $\frac{n}{d}$; that means $d \mapsto \frac{n}{d}$ is a bijection from the set of positive divisors of n to itself. Hence $\sum_{d|n} \phi(\frac{n}{d}) = \sum_{d|n} \phi(d)$, and (21.3) follows. We will be following the same steps to prove (21.2).

Proof of Theorem 21.2.1. Since $x^n - 1 = \prod_{i\in[1..n]}(x - \zeta_n^i)$ and $\{C_{d,n} \mid d|n\}$ is a partition of $[1..n]$, we have that

$$x^n - 1 = \prod_{d|n} \prod_{i\in C_{d,n}} (x - \zeta_n^i). \tag{21.6}$$

By (21.5), we have that

$$\prod_{i\in C_{d,n}} (x - \zeta_n^i) = \prod_{j\in C_{1,\frac{n}{d}}} (x - \zeta_n^{dj}). \tag{21.7}$$

Notice that $\zeta_n^d = e^{(2\pi i/n)d} = e^{2\pi i/(\frac{n}{d})} = \zeta_{\frac{n}{d}}$. So by (21.7), we deduce that

$$\prod_{i\in C_{d,n}} (x - \zeta_n^i) = \prod_{0\le j<\frac{n}{d}, \gcd(j,\frac{n}{d})=1} (x - \zeta_{\frac{n}{d}}^j) = \Phi_{\frac{n}{d}}(x). \tag{21.8}$$

By (21.6) and (21.8), we obtain

$$x^n - 1 = \prod_{d|n} \Phi_{\frac{n}{d}}(x). \tag{21.9}$$

As it is mentioned earlier, $d \mapsto \frac{n}{d}$ is a bijection from the set of positive divisors of n to itself. Hence by (21.9), (21.2) follows. □

Using we are ready to prove that cyclotomic polynomials are integer polynomials.

Corollary 21.2.2. *For every positive integer n, $\Phi_n(x) \in \mathbb{Z}[x]$.*

Proof. We proceed by strong induction on n. The base case is clear as $\Phi_1(x) = x - 1$. Next we prove the strong induction step. By the strong induction hypothesis, for every positive integer $m < n$, $\Phi_m(x) \in \mathbb{Z}[x]$. Hence

$$\Psi_n(x) := \prod_{d|n, d\ne n} \Phi_d(x) \in \mathbb{Z}[x], \tag{21.10}$$

and as Φ_d's are monic, $\Psi_n(x)$ is monic as well. By Theorem 21.2.1, we have

$$x^n - 1 = \Phi_n(x)\Psi_n(x). \tag{21.11}$$

As $x^n - 1, \Psi_n(x) \in \mathbb{Z}[x]$ and $\Psi_n(x)$ is monic, by the Long Division for elements in $\mathbb{Z}[x]$ (see Theorem 6.4.1) there are unique $q(x), r(x) \in \mathbb{Z}[x]$ such that

1. $x^n - 1 = q(x)\Psi_n(x) + r(x)$ and
2. $\deg r < \deg \Psi_n$.

Using the Long Division for elements in $\mathbb{C}[x]$, we see that the same q and r are the quotient and remainder of $x^n - 1$ divided by $\Psi_n(x)$ as elements of $\mathbb{C}[x]$. By (21.11), however, we have that the quotient and the remainder of $x^n - 1$ divided by $\Psi_n(x)$ as elements of $\mathbb{C}[x]$ are $\Phi_n(x)$ and 0, respectively. Hence $\Phi_n(x) = q(x) \in \mathbb{Z}[x]$, and the claim follows. □

Let us remark that the last part of the above argument implies the following:

Lemma 21.2.3. *Suppose A is a subring of a unital commutative ring B and $1_B \in A$. Suppose $f, g \in A[x]$ and $\mathrm{ld}(f) \in A^\times$. If $f|g$ in $B[x]$, then $f|g$ in $A[x]$.*

Use long division in $A[x]$ and $B[x]$ to prove this Lemma. I leave it to you to fill out the details.

21.3 Cyclotomic polynomials are irreducible

The main goal of this section is to prove the following:

Theorem 21.3.1. *For every integer n, $\Phi_n(x)$ is irreducible in $\mathbb{Q}[x]$.*

As before the general steps are proceeding by contradiction, going from $\mathbb{Q}$ to $\mathbb{Z}$, and using the residue maps modulo primes. The last step, however, will be more subtle than the other examples that we have done so far.

Proof of Theorem 21.3.1. Suppose to the contrary that $\Phi_n(x)$ is not irreducible in $\mathbb{Q}[x]$. Then $\Phi_n(x) = f(x)g(x)$ for some non-constant smaller degree polynomials f and g. Since $\Phi_n(x)$ is a monic integer polynomial, it is a primitive polynomial. By Gauss's lemma (see Corollary 15.4.3), we have $\Phi_n(x) = \overline{f}(x)\overline{g}(x)$ where $\overline{f}$ and $\overline{g}$ are primitive forms of f and g, respectively. Notice that ζ_n is a zero Φ_n, it is either a zero of $\overline{f}$ or a zero of $\overline{g}$. Without loss of generality, we can and will assume that ζ_n is a zero of $\overline{f}$. Every other zero ζ of $\overline{f}$ is a zero of $\Phi_n(x)$, and so the multiplicative order of ζ (as an element of $\mathbb{C}^\times$) is n.

(Here is where the magic is happening.)

If p is a prime which does not divide n and $\zeta \in \mathbb{C}^\times$ has multiplicative order n, then the multiplicative order of ζ^p is also n. Therefore ζ^p is a zero of $\Phi_n(x)$, and so it is a zero of either $\overline{f}$ or $\overline{g}$.

Claim 1. *If ζ is a zero of $\overline{f}$ and p is a prime which does not divide n, then ζ^p is a zero of $\overline{f}$ as well.*

Proof of Claim 1. Suppose to the contrary that $\overline{f}(\zeta^p) \neq 0$. Since $o(\zeta^p) = n$, $\Phi_n(\zeta^p) = 0$. As $\overline{f}(\zeta^p) \neq 0$ and $\Phi_n(\zeta^p) = \overline{f}(\zeta^p)\overline{g}(\zeta^p)$, we deduce that $\overline{g}(\zeta^p) = 0$. This means ζ is a common zero of $\overline{f}(x)$ and $\overline{g}(x^p)$. Thus $m_{\zeta,\mathbb{Q}}(x)$ is a common divisor of $\overline{f}(x)$ and $\overline{g}(x^p)$ in $\mathbb{Q}[x]$. Let $h(x)\mathbb{Z}[x]$ be the primitive form of $m_{\zeta,\mathbb{Q}}(x)$. By Gauss's lemma (see Corollary 15.4.3), $h(x)$ is a common divisor of $\overline{f}(x)$ and $\overline{g}(x^p)$ in $\mathbb{Z}[x]$. As $\overline{f}$ is monic, so is h. Therefore $c_p(h)$ is a common divisor of $c_p(\overline{f}(x))$ and $c_p(\overline{g}(x^p))$ where $c_p : \mathbb{Z}[x] \to \mathbb{Z}_p[x]$ is the residue map modulo p. Notice that h is a monic non-constant polynomial, so is $c_p(h)$.

(Here you see why we considered raising to power p at the first place.)

Since $\mathbb{Z}_p[x]$ is of characteristic p, by Fermat's little theorem we have

$$c_p(h(x))^p = c_p(h(x^p)). \tag{21.12}$$

To see this better, notice that in $\mathbb{Z}_p[x]$ we have

$$(\sum_{i=0}^{\infty} a_i x^i)^p = \sum_{i=0}^{\infty} a_i^p (x^i)^p = \sum_{i=0}^{\infty} a_i (x^p)^i.$$

So $c_p(h)$ is a non-constant common divisor of $c_p(\overline{f})$ and $c_p(\overline{g})^p$. Let $\ell(x)$ be a prime factor of $c_p(h)$. Then $\ell(x)$ divides $c_p(\overline{g})^p$, and so $\ell(x)$ divides $c_p(\overline{g})$ as $\mathbb{Z}_p[x]$ is a UFD. Therefore $\ell(x)^2$ divides

$$c_p(\overline{f})c_p(\overline{g}) = c_p(\overline{f}\overline{g}) = c_p(\Phi_n). \tag{21.13}$$

As $c_p(\Phi_n)$ divides $x^n - 1$ in $\mathbb{Z}_p[x]$, $\ell(x)^2$ divides $x^n - 1$ in $\mathbb{Z}_p[x]$. Hence $x^n - 1$ has multiple zeros in its splitting field over $\mathbb{Z}_p$. By Proposition 18.3.4, we deduce that $\gcd(x^n - 1, nx^{n-1}) \neq 1$. This is a contradiction as $p \nmid n$ and $x \nmid x^n - 1$. This completes the proof of Claim 1. □

Claim 2. Suppose i is a positive integer and $\gcd(i, n) = 1$. If ζ is a zero of $\overline{f}$, then ζ^i is a zero of $\overline{f}$.

Proof of Claim 2. We proceed by induction on the number k of prime factors of i. In the base case of $k = 0$, we have $i = 1$, and there is nothing to prove. Suppose $i = p_1 \cdots p_{k+1}$, where p_j's are primes that do not divide n. By the induction hypothesis $\zeta^{p_1 \cdots p_k}$ is a zero of $\overline{f}$. By Claim 1, we deduce that

$$(\zeta^{p_1 \cdots p_k})^{p_{k+1}} = \zeta^i$$

is a zero of $\overline{f}$. This completes the proof of Claim 2. □

By Claim 2, since ζ_n is a zero of $\overline{f}$, ζ_n^i is a zero of $\overline{f}$ if i is a positive integer and $\gcd(i, n) = 1$. This implies that $\Phi_n(x)$ divides $\overline{f}$, which is a contradiction as $\deg \overline{f} < \deg \Phi_n$. This completes the proof. □

21.4 The degree of cyclotomic extensions

Field $\mathbb{Q}[\zeta_n]$ is called a *cyclotomic extension*.

Theorem 21.4.1. *Suppose n is a positive integer and $\zeta_n := e^{2\pi i/n}$. Then the minimal polynomial of ζ_n over $\mathbb{Q}$ is $m_{\zeta_n,\mathbb{Q}}(x) = \Phi_n(x)$ and $[\mathbb{Q}[\zeta_n] : \mathbb{Q}] = \phi(n)$.*

Proof. We have that ζ_n is a zero of $\Phi_n(x)$, $\Phi_n(x)$ is a monic polynomial, and $\Phi_n(x)$ is irreducible in $\mathbb{Q}[x]$ (by Theorem 21.3.1). Hence by Theorem 8.2.5, we have that $m_{\zeta_n,\mathbb{Q}}(x) = \Phi_n(x)$. By Proposition 20.1.2, we have

$$[\mathbb{Q}[\zeta_n] : \mathbb{Q}] = \deg m_{\zeta_n,\mathbb{Q}} = \deg \Phi_n = \phi(n),$$

which completes the proof. □

Chapter 22

Lecture 22

22.1 The group of automorphism of a field extension.

Through out this course one of our main goals has been understanding zeros of polynomials. We proved the existence and the uniqueness (up to an isomorphism) of a smallest field which contains all the zeros of a given polynomial (a splitting field). A better understanding of splitting fields can help us to learn more about the zeros of polynomials. One of our main tools of characterizing (intricate) objects is their group of *symmetries*. The group of symmetries of a field extension E of F is defined as follows.

Definition 22.1.1. *For a field extension E of F, let*

$$\mathrm{Aut}_F(E) := \{\theta : E \to E \mid \theta \textit{ is a ring isomorphism, and } F\textit{-linear}\}.$$

An element of $\mathrm{Aut}_F(E)$ *is called an* F-automorphism*. An F-linear, ring homomorphism is called an* F-homomorphism*.*

One can easily see that $\mathrm{Aut}_F(E)$ is a group under composition. We would like to know how much $\mathrm{Aut}_F(E)$ tells us about the field extension.

Similar to the proof of the uniqueness of splitting fields, we need to work with two, possibly different, copies of the *base* field F, and with not necessarily surjective ring homomorphisms: we proved the *isomorphism extension theorem* in order to deduce the *uniqueness of splitting fields* up to an isomorphism. That is why we introduce the following notation.

Definition 22.1.2. *Suppose $\theta : F \to F'$ is a field isomorphism, E is a field extension of F, and L' is a field extension of F'. Then*

$$\mathrm{Emb}_\theta(E, L') := \{\widehat{\theta} : E \to L' \mid \widehat{\theta} \textit{ injective ring homomorphism and } \widehat{\theta}|_F = \theta\}$$

and an element of $\mathrm{Emb}_\theta(E, L')$ *is called an* θ-embedding*. An isomorphism which is an θ-embedding is called an* θ-isomorphism*, and the set of θ-isomorphisms is denoted by* $\mathrm{Iso}_\theta(E, L')$*. When $F' = F$ and $\theta = \mathrm{id}_F$, we write* $\mathrm{Emb}_F(E, L')$ *instead of* $\mathrm{Emb}_{\mathrm{id}_F}(E, L')$*. Instead of saying* id_F*-embedding, we say* F-embedding.

Notice that $\widehat{\theta}$ is in $\mathrm{Emb}_\theta(E, L')$ exactly when the following is a commutative diagram.

$$\begin{array}{ccc} E & \xrightarrow{\widehat{\theta}} & L' \\ \Big| & & \Big| \\ F & \xrightarrow{\theta} & F' \end{array}$$

Lemma 22.1.3. *If* $[E : F] < \infty$*, then* $\mathrm{Emb}_F(E, E) = \mathrm{Aut}_F(E)$.

Proof. Clearly $\mathrm{Aut}_F(E) \subseteq \mathrm{Emb}_F(E, E)$. Suppose $\theta \in \mathrm{Emb}_F(E, E)$. To show that θ is an F-automorphism, it suffices to argue why θ is surjective. By the first isomorphism theorem for vector spaces (see Theorem 19.5.1), we have

$$\dim_F \mathrm{Im}(\theta) + \dim_F \ker(\theta) = \dim_F E.$$

Since θ is a injective, $\ker \theta = 0$. Hence $\dim_F \mathrm{Im}(\theta) = \dim_F E$. Since $\mathrm{Im}_F(\theta)$ is a subspace of E and it has the same dimension as E, by Proposition 19.4.2 $\mathrm{Im}(\theta) = E$. This completes the proof. □

The following easy lemma is the corner stone of our understanding of the group of symmetries of algebraic field extensions.

Lemma 22.1.4. *Suppose* $\theta : F \to F'$ *is a field isomorphism,* E *is a field extension of* F*, and* L' *is a field extension of* F'*. Suppose* $f(x) \in F[x]$ *and* $\alpha \in E$ *is a zero of* f*. Then*

$$\text{for every } \widehat{\theta} \in \mathrm{Emb}_\theta(E, L'),\ \widehat{\theta}(\alpha) \text{ is a zero of } \theta(f).$$

In particular, if L *is a field extension of* F*, then*

$$\text{for every } F\text{-embedding } \widehat{\theta} : E \to L,\ \theta(\alpha) \text{ is a zero of } f.$$

Proof. Suppose $f(x) = \sum_{i=0}^n c_i x^i$. Then $\sum_{i=0}^n c_i \alpha^i = 0$. Therefore

$$0 = \widehat{\theta}\Big(\sum_{i=0}^n c_i \alpha^i\Big) = \sum_{i=0}^n \widehat{\theta}(c_i)\widehat{\theta}(\alpha)^i = \sum_{i=0} \theta(c_i)\widehat{\theta}(\alpha)^i = \theta(f)(\widehat{\theta}(\alpha)),$$

and the claim follows. □

22.2 Normal extensions

The following theorem and the ideas involved in its proof play an important role in our understanding of field extensions of finite degree.

Theorem 22.2.1. *Suppose* E *is a field extension of* F *and* $[E : F] < \infty$*. Then the following statements are equivalent.*

1. *There is* $f \in F[x]$ *such that* E *is a splitting field of* f *over* F*.*

2. *For every field extension L of E and $\theta \in \mathrm{Aut}_F(L)$, we have $\theta(E) = E$.*

3. *For every $\beta \in E$, $m_{\beta,F}(x) = (x - \beta_1)\cdots(x - \beta_m)$ for some $\beta_1, \ldots, \beta_m \in E$.*

Each one of these properties gives us a very different perspective of the given field extension.

1. The first property (in terms of splitting fields) is very *concrete* and one can construct many examples with it.

2. The second property gives us a relation between symmetries of field extensions of E over F and symmetries of E over F. It is quite surprising that a property about E and F tells us something about symmetries of every field extension of E.

3. In contrast with the second property, the third property is completely internal. It is all about E and F and no other additional information is involved.

The second and the third properties make sense even if the given field extension is not of finite degree. The first property, however, implies that the field extension is of finite degree. One can talk about a *splitting field of a family of polynomials*, replace the statement with this extended notation, and still get equivalent properties. This is a key result for understanding algebraic extensions of infinite degree. Here, however, we do not discuss infinite degree algebraic extensions.

Definition 22.2.2. *Suppose E is an algebraic extension of F. We say E is a* normal *field extension of F if the third property in Theorem 22.2.1 holds.*

Proof of Theorem 22.2.1. (1) ⇒ (2) Since E is a splitting field of f over F, there are $\alpha_1, \ldots, \alpha_n \in E$ such that

$$f(x) = \mathrm{ld}(f)\prod_{i=1}^{n}(x - \alpha_i) \quad \text{and} \quad E = F[\alpha_1, \ldots, \alpha_n].$$

Suppose L is a field extension of E and $\theta \in \mathrm{Aut}_F(L)$. Then by Lemma 22.1.4, $\theta(\alpha_i)$ is a zero of f in L. Since $\alpha_1, \ldots, \alpha_n$ are the only zeros of f in $E \subseteq L$, we obtain that

$$\theta(\alpha_i) \in \{\alpha_1, \ldots, \alpha_n\} \tag{22.1}$$

for every i. As θ is injective, form (22.1) we deduce that θ permutes elements of $\{\alpha_1, \ldots, \alpha_n\}$. Therefore

$$\theta(E) = \theta(F[\alpha_1, \ldots, \alpha_n]) = \theta(F)[\theta(\alpha_1), \ldots, \theta(\alpha_n)] = F[\alpha_1, \ldots, \alpha_n] = E.$$

(2) ⇒ (3) This is the most technical part of the proof. For every $\beta \in E$, we want to show that there are β_i's in E such that $m_{\beta,F}(x) = \prod_{i=1}^{m}(x - \beta_i)$. The second property is about the field extensions of E. Hence we need to work with field extensions of E that contain all the zeros β_i of $m_{\beta,F}$, say L is such a field. Since β and β_i

are zeros of the irreducible polynomial $m_{\beta,F}(x)$, by Lemma 16.2.2 there is an F-isomorphism $\theta_i : F[\beta] \to F[\beta_i]$ such that $\theta_i(\beta) = \beta_i$. *If we manage to extend θ_i to an F-automorphism $\widehat{\theta}_i$ of L*, then by hypothesis, $\widehat{\theta}_i(E) = E$, which implies that

$$\widehat{\theta}_i(\beta) = \theta_i(\beta) = \beta_i$$

is in E, and the claim follows. Hence we focus on extending θ_i to an F-isomorphism from L to itself. This reminds of the *isomorphism extension theorem* (see Theorem 17.1.1). By the isomorphism extension theorem, we can extend θ_i to an F-automorphism $\widehat{\theta}_i$ of L if L is a splitting field of a polynomial over F.

Altogether we have proved that the claim follows if we show the existence of a field L with the following properties.

1. L is a field extension of E.
2. There are β_i's in L such that $m_{\beta,F}(x) = \prod_{i=1}^{m}(x - \beta_i)$.
3. There is $f \in F[x]$ such that L is a splitting field of f over F.

Notice that the conditions (2) and (3) are satisfied by a splitting field of $m_{\beta,F}$ over F, but this field does not necessarily contain E as a subfield. The following is a common technique that is used to construct a field which is a splitting field of a polynomial over F and contains E as a subfield. [1]

Suppose $(\gamma_1, \ldots, \gamma_n)$ is an F-basis of E, and let

$$f(x) := m_{\beta,F}(x)m_{\gamma_1,F}(x)\cdots m_{\gamma_n,F}(x) \in F[x].$$

Suppose L is a splitting field of f over E. Clearly L satisfies the first and the second desired properties that are mentioned above. Next we show that L is a splitting field of f over F. Since L is a splitting field of f over E, there are β_i's and $\gamma_{i,j}$'s in L such that

$$m_{\beta,F}(x) = \prod_{i=1}^{m}(x - \beta_i) \quad \text{and} \quad m_{\gamma_i,F}(x) = \prod_{j=1}^{m_i}(x - \gamma_{i,j}), \tag{22.2}$$

and

$$L = E[\beta_1, \ldots, \beta_m, \gamma_{1,1}, \ldots, \gamma_{n,m_n}]. \tag{22.3}$$

Since $\gamma_i \in E \subseteq L$ is a zero of $m_{\gamma_i,F}$, $\gamma_i \in \{\gamma_{i,1}, \ldots, \gamma_{i,m_i}\}$. Hence without loss of generality we can and will assume that $\gamma_{i,1} = \gamma_i$ for every i.

Notice that (22.3) means that if a subfield of L contains E, β_i's and $\gamma_{i,j}$'s, then it is the entire L. On the other hand, as γ_i's form an F-basis of E, if a subfield of L contains F and γ_i's, then it contains E. Altogether we obtain that a subfield of L which contains F, β_i's, and $\gamma_{i,j}$'s is the entire L. This means

$$L = F[\beta_1, \ldots, \beta_m, \gamma_{1,1}, \ldots, \gamma_{n,m_n}]. \tag{22.4}$$

By (22.2) and (22.4), we deduce that L is a splitting field of f over F. This gives us a field L with the mentioned desired properties, and the claim follows.

[1] We will use this method to show the existence of a normal closure of a field extension.

(3) ⇒ (1) We use the same technique as in the proof of the previous step. Suppose $(\gamma_1, \ldots, \gamma_n)$ is an F-basis of E, and let

$$g(x) := m_{\gamma_1,F}(x) \cdots m_{\gamma_n,F}(x).$$

By hypothesis, $m_{\gamma_i,F}$ can be written as a product of degree one factors in $E[x]$. Hence there are α_i's in E such that

$$g(x) = (x - \alpha_1) \cdots (x - \alpha_d). \tag{22.5}$$

We also notice that γ_i's are zeros of g in E, and so

$$\gamma_i \in \{\alpha_1, \ldots, \alpha_d\} \tag{22.6}$$

for every i. Therefore

$$\begin{aligned} E = \mathrm{Span}_F(\gamma_1, \ldots, \gamma_n) \subseteq & F[\gamma_1, \ldots, \gamma_n] \\ \subseteq & F[\alpha_1, \ldots, \alpha_d] \subseteq E. \end{aligned}$$

Hence $E = F[\alpha_1, \ldots, \alpha_n]$, which together with (22.5) implies that E is a splitting field of g over F. This completes the proof. □

Chapter 23

Lecture 23

23.1 The group of automorphism of normal field extensions.

Using Theorem 22.2.1, we obtain the following result on the group of automorphisms.

Proposition 23.1.1. *Suppose E is a normal extension of F and $[E : F] < \infty$. Then*

1. *For every field extension L of E,*

$$r_{L,E} : \mathrm{Aut}_F(L) \to \mathrm{Aut}_F(E), \;\; r_{L,E}(\theta) := \theta|_E$$

is a well-defined group homomorphism. Moreover $\ker r_{L,E} = \mathrm{Aut}_E(L)$; in particular, $\mathrm{Aut}_E(L)$ is a normal subgroup of $\mathrm{Aut}_F(L)$.

2. *For every extension L of E which is a finite normal extension of F, $r_{L,E}$ is surjective and*

$$\mathrm{Aut}_F(L)/\,\mathrm{Aut}_E(L) \simeq \mathrm{Aut}_F(E).$$

Proof. (1) Since E is a finite normal extension of F, by Theorem 22.2.1 for every field extension L of E and every $\theta \in \mathrm{Aut}_F(L)$, $\theta(E) = E$. Hence $\theta|_E$ is an F-automorphism of E. Therefore $r_{L,E}$ is a well-defined map. It is easy to check that it is a group homomorphism.

Notice that $\theta \in \ker r_{L,E}$ if and only if $\theta|_E = \mathrm{id}_E$. Hence $\ker r_{L,E} = \mathrm{Aut}_E(L)$. From group theory, we know that kernel of a group homomorphism is a normal subgroup.

(2) Let's start by understanding what the surjectivity of $r_{L,E}$ means. It means that every $\overline{\theta} \in \mathrm{Aut}_F(E)$ can be extended to an F-automorphism of L. By the isomorphism extension theorem, $\overline{\theta}$ can be extended to an F-isomorphism from L to itself if L is a splitting field of a polynomial $f \in F[x]$. Let's explain why this is the case. If L is a splitting field of $f \in F[x]$ over F, then by $f = \theta(f)$ and $E = \theta(E)$, we observe that L is also a splitting field of $\theta(f)$ over $\theta(E)$. Therefore by the isomorphism extension theorem (see Theorem 17.1.1) we get the desired extension.

Since L is a finite normal extension of F, by Theorem 22.2.1 there is $f \in F[x]$ such that L is a splitting field of f over F. Hence as explained above by the isomorphism extension theorem, there is $\theta \in \mathrm{Aut}_F(L)$ such that $\theta|_E = \overline{\theta}$, and so $r_{L,E}$ is surjective.

By the first isomorphism theorem for groups, we have

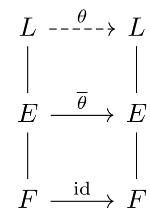
$$\mathrm{Aut}_F(L)/\ker r_{L,E} \simeq \mathrm{Im}\, r_{L,E},$$

and so

$$\mathrm{Aut}_F(L)/\mathrm{Aut}_E(L) \simeq \mathrm{Aut}_F(E).$$

This completes the proof. □

The following commutative diagram captures the surjectivity of $r_{L,E}$ when L is a finite normal extension of F. In this diagram, every row is an isomorphism, and the dashed arrow means that for a given $\overline{\theta}$, we can find θ that makes the diagram commutative.



23.2 Normal extensions and tower of fields

When we learn about a property of field extensions, we have to ask ourselves how it behaves in a tower of fields. For instance, by the Tower Rule, we know that for a tower of fields $F \subseteq E \subseteq L$, L is a finite extension of F if and only if L is a finite extension of E and E is a finite extension of F. We will see that normal extensions do *not* have such a nice behavior. We, however, start with a positive result.

Lemma 23.2.1. *Suppose $F \subseteq E \subseteq L$ is a tower of field extensions. Then the following holds.*

1. *For every $\beta \in L$, $m_{\beta,E}|m_{\beta,F}$ in $E[x]$.*
2. *If L is a normal extension of F, then L is a normal extension of E.*

Proof. (1) Since β is a zero of $m_{\beta,F}(x) \in E[x]$, by Proposition 8.2.6 we have that $m_{\beta,E}$ divides $m_{\beta,F}$ in $E[x]$.

(2) Since L is a normal extension of F, for every β, $m_{\beta,F}(x)$ can be written as a product of degree one factors in $L[x]$. By part one, $m_{\beta,E}$ divides $m_{\beta,F}$ in $E[x]$, and so $m_{\beta,E}$ divides $m_{\beta,F}$ in $L[x]$. Since $L[x]$ is a UFD, degree one polynomials are irreducible in $L[x]$, $m_{\beta,F}$ can be written as a product of degree one factors, and $m_{\beta,E}|m_{\beta,F}$ in $L[x]$, we obtain that $m_{\beta,E}$ can be written as product of degree one factors in $L[x]$. This means L is a normal extension of E, which completes the proof. □

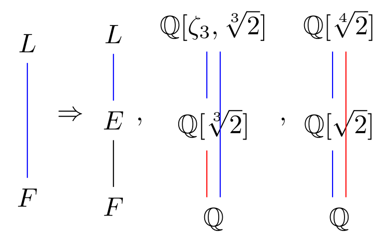
The following examples show us that the normal extension property cannot be deduced for other parts of a tower.

By Example 17.2.2 and Theorem 22.2.1, $\mathbb{Q}[\zeta_n, \sqrt[n]{2}]$ is a normal extension of $\mathbb{Q}$. We, however, claim that the intermediate field $\mathbb{Q}[\sqrt[n]{2}]$ is not a normal extension of $\mathbb{Q}$

if $n > 2$. By Eisenstein's criterion, $x^n - 2$ is irreducible in $\mathbb{Q}[x]$. As $\sqrt[n]{2}$ is a zero of $x^n - 2$, by Theorem 8.2.5 $m_{\sqrt[n]{2},\mathbb{Q}}(x) = x^n - 2$. This polynomial has at most two real zeros, and so not all of its zeros are in $\mathbb{Q}[\sqrt[n]{2}]$. Therefore $\mathbb{Q}[\sqrt[n]{2}]$ is not a normal extension of $\mathbb{Q}$. Notice that if $[E : F] = 2$, then for every $\alpha \in E \setminus F$ we have

$$1 < \deg m_{\alpha,F} = [F[\alpha] : F] \leq [E : F] = 2.$$

Hence for every $\alpha \in E$, we have $1 \leq \deg m_{\alpha,F} \leq 2$, and so all the zeros of $m_{\alpha,F}$ are in E. Therefore E is a normal extension of F. This implies that $\mathbb{Q}[\sqrt[4]{2}]$ is a normal extension of $\mathbb{Q}[\sqrt{2}]$ and $\mathbb{Q}[\sqrt{2}]$ is a normal extension of $\mathbb{Q}$, but as we showed above $\mathbb{Q}[\sqrt[4]{2}]$ is not a normal extension of $\mathbb{Q}$.



23.3 Normal closure of a field extension

Suppose E is a finite field extension of F. We prove the existence of a smallest field extension of E which is a normal extension of F.

Proposition 23.3.1. *Suppose E is a finite field extension of F. Then there is a field extension L of E such that the following holds:*

1. *L is a normal extension of F.*
2. *If L' is a field extension of E and L' is a normal extension of F, then there is an E-embedding $\theta : L \to L'$.*

In particular, if L_1 and L_2 satisfy the above properties, then there is an E-isomorphism $\theta : L_1 \to L_2$.

A field L which satisfies the properties mentioned in Proposition 23.3.1 is called a *normal closure of the field extension E of F.*

Proof. We use an identical technique as in the proof of Theorem 22.2.1 (going from (2) to (3)). Suppose $(\gamma_1, \ldots, \gamma_d)$ is an F-basis of E. Let L be a splitting field of

$$f(x) := m_{\gamma_1,F}(x) \cdots m_{\gamma_d,F}(x)$$

over E. Then there are $\gamma_{i,j}$'s in L such that for every i

$$m_{\gamma_i,F}(x) = \prod_{j=1}^{n_i}(x - \gamma_{i,j}) \quad \text{and} \quad L = E[\gamma_{1,1}, \ldots, \gamma_{d,n_d}]. \tag{23.1}$$

Since $\gamma_i \in E$ is a zero of $m_{\gamma_i,F}(x)$ and $E \subseteq L$, we can and will assume that $\gamma_{i,1} = \gamma_i$ for every i. Therefore

$$E = \operatorname{Span}_F(\gamma_1, \ldots, \gamma_d) \subseteq F[\gamma_1, \ldots, \gamma_n] \subseteq F[\gamma_{1,1}, \ldots, \gamma_{d,n_d}]. \tag{23.2}$$

By (23.2), we obtain that

$$L = E[\gamma_{1,1}, \ldots, \gamma_{d,n_d}] \subseteq F[\gamma_{1,1}, \ldots, \gamma_{d,n_d}] \subseteq L.$$

Hence $L = F[\gamma_{1,1}, \ldots, \gamma_{d,n_d}]$, and so by (17.2) we deduce that L is a splitting field of f over F. Hence by Theorem 22.2.1, L is a normal field extension of F.

Suppose L' is a field extension of E and L' is a normal extension of F. Then for every i, $\gamma_i \in L'$. Since L' is a normal extension of F, there are $\gamma'_{i,j}$'s in L' such that

$$m_{\gamma_i,F}(x) = \prod_{j=1}^{n_i}(x - \gamma'_{i,j}). \tag{23.3}$$

Then $L'' := E[\gamma'_{1,1}, \ldots, \gamma'_{d,n_d}] \subseteq L'$ is a splitting field of $f(x)$ over E. Therefore by the uniqueness of splitting fields (see Theorem 17.1.2) there is an E-isomorphism $\theta : L \to L''$. As L'' is a subfield of L', θ can be viewed as an element in $\mathrm{Emb}_E(L, L')$.

If L_1 and L_2 satisfy these conditions, then there are $\theta_1 \in \mathrm{Emb}_E(L_1, L_2)$ and $\theta_2 \in \mathrm{Emb}_E(L_2, L_1)$. As L_i's are finite field extensions of E, we deduce that θ_i's are isomorphisms. This completes the proof. □

23.4 Normal extension and extending embeddings

The following result on extending embeddings is a variant of the isomorphism extension theorem.

Proposition 23.4.1. *Suppose $F \subseteq E \subseteq L$ is a tower of fields, and L is a finite normal extension of F. Suppose $\theta \in \mathrm{Emb}_F(E, L)$. Then there is $\widehat{\theta} \in \mathrm{Aut}_F(L)$ such that $\widehat{\theta}|_E = \theta$.*

Proof. Since L is a finite normal extension of F, there is $f \in F[x]$ such that L is a splitting field of f over F. So there are α_i's in L such that

$$f(x) = \mathrm{ld}(f)\prod_{i=1}^{n}(x - \alpha_i) \quad \text{and} \quad L = F[\alpha_1, \ldots, \alpha_n].$$

Notice that since E and $\theta(E)$ contain F as a subfield, L can be viewed as a splitting field of f over E and also as a splitting field of $\theta(f) = f$ over $\theta(E)$. Thus by the isomorphism extension theorem, there is $\widehat{\theta} : L \to L$ such that $\widehat{\theta}|_E = \theta$, and this completes the proof. □

23.5 Group of automorphisms of a field extension

For every field extension E of F, by Lemma 22.1.4 and Lemma 16.2.2, the following is a bijection

$$\mathrm{Emb}_F(F[\alpha], E) \to \{\alpha' \in E \mid m_{\alpha,F}(\alpha') = 0\},\ \theta \mapsto \theta(\alpha). \tag{23.4}$$

Then by (23.4) we have

$$|\operatorname{Emb}_F(F[\alpha], E)| = \# \text{ of distinct zeros of } m_{\alpha,F} \text{ in } E \leq \deg m_{\alpha,F} = [F[\alpha] : F].$$

Suppose E is a finite normal extension of F and $E = F[\alpha]$. Then

$$|\operatorname{Aut}_F(E)| \leq [E : F]$$

and equality holds if $m_{\alpha,F}$ has distinct zeros in E. This takes us to the following questions.

1. What if E is not of the form $F[\alpha]$ for some α?

2. When can we be sure that $E = F[\alpha]$ for some α?

The following theorem addresses the first question (and more!).

Theorem 23.5.1. *Suppose $\theta : F \to F'$ is a field isomorphism, and $f(x) \in F[x]$. Suppose E is a splitting field of f over F, and E' is a splitting field of $\theta(f)$ over F'. Then*

$$|\operatorname{Iso}_\theta(E, E')| \leq [E : F].$$

Moreover the equality holds if irreducible factors of f in $F[x]$ do not have multiple zeros in E.

This is an extremely important result. Proof of this theorem has some similarities with the proof of the isomorphism extension theorem (see Theorem 17.1.1).

Proof of Theorem 23.5.1. We proceed by strong induction on $[E : F]$. If $[E : F] = 1$, then $E = F$ and $E' = F'$, and so $\operatorname{Iso}_\theta(E, E') = \operatorname{Iso}_\theta(F, F') = \{\theta\}$ has exactly 1 element, and equality holds.

Suppose $E \neq F$. Hence f has a zero $\alpha \in E$ which is not in F. Notice that for every $\widehat{\theta} \in \operatorname{Iso}_\theta(E, E')$, we have $\widehat{\theta}|_{F[\alpha]}$ is in $\operatorname{Emb}_\theta(F[\alpha], E')$. Notice that by Lemma 22.1.4

$$\operatorname{Emb}_\theta(F[\alpha], E') \to \{\alpha' \in E' \mid \theta(m_{\alpha,F})(\alpha') = 0\}, \quad \widehat{\theta} \mapsto \widehat{\theta}(\alpha) \tag{23.5}$$

is a well-defined function. Since a ring homomorphism $\theta_1 : F[\alpha] \to E'$ is uniquely determined by $\theta_1|_F$ and $\theta_1(\alpha)$, the function given in (23.5) is injective. If α' is a zero of $\theta(m_{\alpha,F})$ in E', then by Lemma 16.2.2 there is $\theta_1 \in \operatorname{Iso}_\theta(F[\alpha], F'[\alpha'])$, and so the function given in (23.5) is a bijection. Hence

$$\begin{aligned} |\operatorname{Emb}_\theta(F[\alpha], E')| =& \# \text{ of distinct zeros of } m_{\alpha,F} \text{ in } E \\ &\leq \deg m_{\alpha,F} = [F[\alpha] : F]. \end{aligned} \tag{23.6}$$

For every $\theta_1 \in \operatorname{Emb}_\theta(F[\alpha], E')$, notice that E is a splitting field of f over $F[\alpha]$, and E' is a splitting field of $\theta(f) = \theta_1(f)$ over $F'[\theta_1(\alpha)]$. Since $[E : F[\alpha]] < [E : F]$, by the strong induction hypothesis, we have

$$|\operatorname{Iso}_{\theta_1}(E, E')| \leq [E : F[\alpha]]. \tag{23.7}$$

Hence

$$
\begin{aligned}
|\operatorname{Iso}_\theta(E,E')| &= \sum_{\theta_1\in \operatorname{Emb}_\theta(F[\alpha],E')} |\operatorname{Iso}_{\theta_1}(E,E')| \\
&\le [E:F[\alpha]]|\operatorname{Emb}_\theta(F[\alpha],E')| && \text{(by (23.7))}\\
&\le [E:F[\alpha]][F[\alpha]:F] = [E:F]. && \text{(by (23.6))}
\end{aligned}
$$

To prove the *moreover* part, we go back through the above argument and show the equalities hold. If α is a zero of f, then $m_{\alpha,F}$ is an irreducible factor of f in $F[x]$. Then, by hypothesis, $m_{\alpha,F}$ has distinct zeros in E. Hence by Proposition 18.3.4, $\gcd(m_{\alpha,F}, m'_{\alpha,F}) = 1$. Thus $\gcd(\theta(m_{\alpha,F}), \theta(m_{\alpha,F})') = 1$, and so by Proposition 18.3.4, $\theta(m_{\alpha,F})$ has distinct zeros in E'. Therefore by (23.6), we have

$$|\operatorname{Emb}_\theta(F[\alpha],E')| = [F[\alpha]:F]. \tag{23.8}$$

As in the above argument, we want to use the strong induction hypothesis to obtain that $|\operatorname{Iso}_{\theta_1}(E,E')| = [E:F[\alpha]]$ for every $\theta_1 \in \operatorname{Emb}_\theta(F[\alpha],E')$. We have already pointed out that E is a splitting field of f over $F[\alpha]$ and E' is a splitting field of $\theta_1(f)$ over $F'[\theta_1(\alpha)]$. To use the strong induction hypothesis for θ_1, E, and E', it is enough to show that all the irreducible factors of f in $(F[\alpha])[x]$ do not have multiple zeros in E. Let $p(x)$ be a monic irreducible factor of f. Then there is $\beta\in E$ which is a zero of p. Hence by Theorem 8.2.5, $p(x) = m_{\beta,F[\alpha]}$. Hence by Lemma 23.2.1, $p(x)|m_{\beta,F}$ in $(F[\alpha])[x]$. Since $m_{\beta,F}$ is an irreducible factor of f in $F[x]$, by hypothesis it does not have multiple zeros in E. Hence its divisor p does not have multiple zeros in E, either. Therefore by the strong induction hypothesis, we have

$$|\operatorname{Iso}_{\theta_1}(E,E')| = [E:F[\alpha]]. \tag{23.9}$$

Hence

$$
\begin{aligned}
|\operatorname{Iso}_\theta(E,E')| &= \sum_{\theta_1\in \operatorname{Emb}_\theta(F[\alpha],E')} |\operatorname{Iso}_{\theta_1}(E,E')| \\
&= [E:F[\alpha]]|\operatorname{Emb}_\theta(F[\alpha],E')| && \text{(by (23.9))}\\
&= [E:F[\alpha]][F[\alpha]:F] = [E:F]. && \text{(by (23.8))}
\end{aligned}
$$

This completes the proof. ∎

Chapter 24

Lecture 24

24.1 Separable polynomials

To have a simpler formulation of Theorem 23.5.1, we define *separable polynomials* as follows.

Definition 24.1.1. *Suppose F is a field and $f \in F[x]$. We say f is* separable *(in $F[x]$) if its irreducible factors in $F[x]$ do not have multiple zeros in a splitting field of f over F.*

Let us make two remarks:

1. The way we defined *separability* of $f \in F[x]$ depends on both the polynomial f and the field F. For instance every polynomial $f \in F[x]$ is separable as an element of $E[x]$ where E is a splitting field of f over F (Notice that all the irreducible factors of f in $E[x]$ are of degree 1 and so they do not have multiple zeros). On the other hand, $x^p - t$ is irreducible in $\mathbb{F}_p(t)$ and it has multiple zeros in its splitting field. To show the latter you can use the fact that either the derivative of this polynomial is zero or $x^p - t = (x - \sqrt[p]{t})^p$.

2. If $p \in F[x]$ is irreducible, then by Proposition 18.3.4, p is separable in $F[x]$ if and only if $\gcd(p, p') = 1$ in $F[x]$. Notice that if E is a field extension of F and $\gcd(p, p') = 1$ in $F[x]$, then $\gcd(p, p') = 1$ in $E[x]$ as well. Hence for an irreducible polynomial $p \in F[x]$, separability only depends on the polynomial.

By the special case of Theorem 23.5.1 for $F = F'$ and $\theta := \mathrm{id}_F$, we obtain the following:

Theorem 24.1.2. *If E is a finite normal extension of F, then $|\operatorname{Aut}_F(E)| \leq [E : F]$.*

Theorem 24.1.3. *Suppose E is a splitting field of a separable polynomial $f \in F[x]$ over F. Then $|\operatorname{Aut}_F(E)| = [E : F]$.*

24.2 Separable and Galois extensions

We start by defining *separable field extensions*.

Definition 24.2.1. *Suppose E is an algebraic field extension of F. We say E is a separable extension of F if, for every $\alpha \in E$, $m_{\alpha,F}$ is a separable element of $F[x]$.*

Theorem 24.2.2. *Suppose E is a finite field extension of F. Then the following statements are equivalent.*

1. *E is a normal separable extension of F.*
2. *E is a splitting field of a separable $f \in F[x]$ over F.*
3. *$|\operatorname{Aut}_F(E)| = [E : F]$.*

Proof. (1) $\Rightarrow$ (2). Suppose $(\gamma_1, \ldots, \gamma_m)$ is an F-basis of E. Since E is a normal extension of F, there are $\gamma_{i,j} \in E$ such that $m_{\gamma_i,F}(x) = \prod_{j=1}^{n_i}(x - \gamma_{i,j})$. Let

$$f(x) := \prod_{i=1}^{m} m_{\gamma_i,F}(x) = \prod_{i,j}(x - \gamma_{i,j}).$$

We notice that $\gamma_i \in E$ is among $\{\gamma_{i,1}, \ldots, \gamma_{i,n_i}\}$. So

$$E \supseteq F[\gamma_{1,1}, \ldots, \gamma_{m,n_m}] \supseteq \operatorname{Span}_F(\gamma_1, \ldots, \gamma_m) = E.$$

Hence E is a splitting field of f over F. Since E is a separable extension of f, $m_{\gamma_i,F}$'s do not have multiple zeros in E. Hence f is separable in $F[x]$ (notice that that $m_{\gamma_i,F}(x)$'s are irreducible in $F[x]$).

(2) $\Rightarrow$ (3). It follows from Theorem 24.1.3.

(3) $\Rightarrow$ (1). For every $\alpha \in E$, we have

$$\begin{aligned} |\operatorname{Aut}_F(E)| &= \sum_{\theta \in \operatorname{Emb}_F(F[\alpha],E)} |\operatorname{Iso}_\theta(E,E)| \\ \text{(By Theorem 23.5.1)} \quad &\leq [E : F[\alpha]] |\operatorname{Emb}_F(F[\alpha], E)| \\ \text{(By (23.4))} \quad &= [E : F[\alpha]] \cdot (\#\text{of distinct zeros of } m_{\alpha,F} \text{ in } E) \end{aligned} \tag{24.1}$$

On the other hand, by hypothesis and the Tower Rule, we have

$$|\operatorname{Aut}_F(E)| = [E : F] = [E : F[\alpha]][F[\alpha] : F]. \tag{24.2}$$

Hence by (24.2), (24.1), and Proposition 20.1.2, we obtain that

$$\#\text{of distinct zeros of } m_{\alpha,F} \text{ in } E \geq [F[\alpha] : F] = \deg m_{\alpha,F}.$$

Therefore $m_{\alpha,F}$ has $\deg m_{\alpha,F}$ distinct zeros in E. Hence the following holds.

1. There are $\alpha_1, \ldots, \alpha_m \in E$ such that $m_{\alpha,F}(x) = \prod_{i=1}^{m}(x - \alpha_i)$.

2. $m_{\alpha,F}$ does not have multiple zeros in E. The first assertion implies that E is a normal extension of F, and form the second statement we deduce that E is a separable extension of F. This completes the proof.

□

Definition 24.2.3. *An algebraic field extension E of F is called a* Galois extension *if it is normal and separable.*

Galois extensions will be explored more later.

Chapter 25

Lecture 1

25.1 Review

Let us start by recalling some of the important results and terminologies that we have already mentioned on field theory and zeros of polynomials.

Splitting fields

Suppose F is a field and $f \in F[x] \setminus F$. Then there is an extension field E of F such that for some $\alpha_1, \ldots, \alpha_n \in E$ we have

1. $f(x) = \mathrm{ld}(f)(x - \alpha_1) \cdots (x - \alpha_n)$, and

2. $E = F[\alpha_1, \ldots, \alpha_n]$.

Such a field extension is called a splitting field of f over F (See Proposition 16.1.1). We refer to F as the *base field*.

Field extension

If F is a subfield of E, we say E is called an *extension field* of F, and the pair of fields is called a *field extension* and it is denoted by E/F. Admittedly this notation might be a bit confusing. One should understand from the context if E/F denotes a field extension or the quotient of the vector space E by the vector space F. The field F is called the *base* of this field extension.

Finite field extension

A field extension E/F is called a *finite extension* if it has a finite index $[E : F]$, where $[E : F] := \dim_F E$ (See Section 19.1).

Algebraic field extension

A field extension E/F is called an *algebraic extension* if every $\alpha \in E$ is algebraic over the base field F. Recall that α is called algebraic over F if α is a zero of a non-constant polynomial $f \in F[x]$.

Normal extension

A field extension E/F is called a *normal extension* if it is an algebraic extension and for every $\alpha \in E$ the minimal polynomial $m_{\alpha,F}$ of α over F decomposes into linear factors over E. This is equivalent to saying that E contains a copy of a splitting field of $m_{\alpha,F}$ over F.

Separable extension

A field extension E/F is called a *separable extension* if it is an algebraic extension and for every $\alpha \in E$, the minimal polynomial $m_{\alpha,F}$ of α over F is a separable polynomial of $F[x]$. Recall that an irreducible polynomial f of $F[x]$ is called a separable polynomial of $F[x]$ if it does not have multiple zeros in its splitting field over F. By Proposition 18.3.4, a polynomial $f \in F[x]$ does not have multiple zeros in its splitting field if and only if $\gcd(f, f') = 1$. A polynomial $f \in F[x]$ is called a *separable polynomial* of $F[x]$ if it is not constant and all of its irreducible factors are separable in $F[x]$.

Important results about finite normal extensions

Suppose E/F is a finite field extension. Then the following are equivalent statements (See Theorem 22.2.1):

1. E/F is a normal extension.
2. E is a splitting field of a polynomial $f \in F[x]$ over F.
3. If L/E is a field extension and $\theta \in \mathrm{Aut}_F(L)$, then $\theta(E) = E$.

If E/F is a finite normal extension, then

$$r_{L,E} : \mathrm{Aut}_F(L) \to \mathrm{Aut}_F(E), \quad r_{L,E}(\theta) := \theta|_E$$

is a well-defined group homomorphism and its kernel is $\mathrm{Aut}_E(L)$; in particular $\mathrm{Aut}_E(L)$ is a normal subgroup of $\mathrm{Aut}_F(L)$ and $\mathrm{Aut}_F(L)/\mathrm{Aut}_E(L)$ is isomorphic to a subgroup of $\mathrm{Aut}_F(E)$ (see Proposition 23.1.1).

If E/F is a finite normal extension, L/E is a finite extension, and L/F is a normal extension, then $r_{L,E}$ is surjective; this means every $\theta \in \mathrm{Aut}_F(E)$ can be extended to an F-automorphism $\widehat{\theta} \in \mathrm{Aut}_F(L)$. In this case, we have

$$\mathrm{Aut}_F(L)/\mathrm{Aut}_E(L) \simeq \mathrm{Aut}_F(E).$$

Important results about splitting fields

Suppose $\theta : F \to F'$ is a field isomorphism and $f \in F[x]$. Suppose E is a splitting field of f over F and E' is a splitting field of $\theta(f)$ over F'. Let

$$\mathrm{Iso}_\theta(E, E') := \{\widehat{\theta} : E \xrightarrow{\sim} E' \mid \widehat{\theta}|_F = \theta\}.$$

(the set of all possible isomorphisms that are extension of θ). Then

1. (Uniqueness) $\mathrm{Iso}_\theta(E, E') \neq \varnothing$ (see Theorem 17.1.2).
2. (Upper bound) $|\,\mathrm{Iso}_\theta(E, E')| \leq [E : F]$ (see Theorem 23.5.1).
3. (Separable case) $|\,\mathrm{Iso}_\theta(E, E')| = [E : F]$ if f is a separable polynomial in $F[x]$ (see Theorem 23.5.1).

Finite Galois extension

Suppose E/F is a finite field extension. Then the following statements are equivalent (see Theorem 24.2.2):

1. E/F is a splitting field of a separable polynomial f of $F[x]$ over F.
2. $|\,\mathrm{Aut}_F(E)| = [E : F]$.
3. E/F is a normal and separable extension.

A field extension E/F is called a *Galois extension* if it is a normal separable extension.

25.2 Symmetries and field extensions

In many parts of mathematics, we can classify objects based on their group of symmetries. We do the same in field theory. We will try to answer the following questions:

1. How much the group of symmetries of a field extension can tell us about the field extension?
2. What are the possible intermediate subfields of a field extension E/F?
3. How can we understand the group structure of $\mathrm{Aut}_F(E)$?

To get a better understanding of $\mathrm{Aut}_F(E)$ and answering the above questions, we work with the natural action of $\mathrm{Aut}_F(E)$ on E.

Recalling basics of group actions

Let us recall basics of *group actions*. We say a group G acts on a set X if there is a binary operation

$$G \times X \to X, (g, x) \mapsto g \cdot x$$

with the following properties:

1. for every $x \in X$, $1 \cdot x = x$.
2. for every $x \in X$, $g_1, g_2 \in G$, $g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x$.

Two important notion related with an action of a group G on a set X are G-*orbit* of x and *the stabilizer subgroup* of G with respect to x. The G-orbit of x is either denoted by $\mathscr{O}_x$ or $G \cdot x$ and it is

$$\mathscr{O}_x := \{g \cdot x \mid g \in G\}.$$

The stabilizer subgroup of G with respect to x is

$$G_x := \{g \in G \mid g \cdot x = x\}.$$

Orbit-Stabilizer Theorem states that

$$G/G_x \to G \cdot x, \quad gG_x \mapsto g \cdot x$$

is a bijection; in particular, if $\mathscr{O}_x$ is a finite orbit, then

$$|\mathscr{O}_x| = [G : G_x].$$

Orbits of symmetries of algebraic extensions

Symmetries of every object X naturally acts on X. Let's see this for the case of a field extension E/F. The group of symmetries of a field extension E/F is $\mathrm{Aut}_F(E)$. For $\alpha \in E$ and $\theta \in \mathrm{Aut}_F(E)$, we let

$$\theta \cdot \alpha := \theta(\alpha).$$

Then for every $\alpha \in E$, $\mathrm{id}_E \cdot \alpha = \mathrm{id}_E(\alpha) = \alpha$ and for every $\theta_1, \theta_2 \in \mathrm{Aut}_F(E)$ and $\alpha \in E$ we have

$$\theta_1 \cdot (\theta_2 \cdot \alpha) = \theta_1(\theta_2(\alpha)) = (\theta_1 \circ \theta_2)(\alpha) = (\theta_1 \circ \theta_2) \cdot \alpha.$$

Hence $\mathrm{Aut}_F(E)$ acts on E. The next lemma gives us some understanding of the orbits and the stabilizer subgroups of this group action.

Lemma 25.2.1. *Suppose E/F is an algebraic extension. Then for every $\alpha \in E$, the following statements hold.*

1. *The orbit $\mathscr{O}_\alpha$ of α under the action of $\mathrm{Aut}_F(E)$ is a subset of all the zeros of $m_{\alpha,F}(x)$ in E.*

2. *The stabilizer subgroup of* $\mathrm{Aut}_F(E)$ *with respect to* α *is* $\mathrm{Aut}_{F[\alpha]}(E)$. *In particular,*

$$[\mathrm{Aut}_F(E) : \mathrm{Aut}_{F[\alpha]}(E)] = |\mathscr{O}_\alpha|.$$

Proof. (1) Suppose $m_{\alpha,F}(x) = \sum_{i=0}^n a_i x^i$. Then $\sum_{i=0}^n a_i \alpha^i = 0$. Hence for every $\theta \in \mathrm{Aut}_F(E)$, we have

$$0 = \theta(\sum_{i=0}^n a_i \alpha^i) = \sum_{i=0}^n a_i \theta(\alpha)^i.$$

This means that $\theta(\alpha) \in E$ is a zero of $m_{\alpha,F}(x)$. Hence

$$\mathrm{Aut}_F(E) \cdot \alpha \subseteq \{\alpha' \in E \mid m_{\alpha,F}(\alpha') = 0\}.$$

(2) The element θ is in the stabilizer subgroup of $\mathrm{Aut}_F(E)$ with respect to α if and only if $\theta(\alpha) = \alpha$. Since $\theta|_F = \mathrm{id}_F$, we obtain that θ is in the stabilizer subgroup of $\mathrm{Aut}_F(E)$ with respect to α exactly when $\theta|_{F[\alpha]} = \mathrm{id}_{F[\alpha]}$. The claim follows. □

25.3 Finite Galois extensions and orbits of their symmetries

By Lemma 25.2.1, we have that

$$|\mathrm{Aut}_F(E) \cdot \alpha| \leq \deg m_{\alpha,F}.$$

Next we show that the equality holds precisely when E/F is a Galois extension. We prove this statement only for finite extensions.

Theorem 25.3.1. *Suppose E/F is a finite extension. Then E/F is a Galois extension if and only if for every $\alpha \in E$,*

$$|\mathrm{Aut}_F(E) \cdot \alpha| = \deg m_{\alpha,F}.$$

Moreover, if E/F is a finite Galois extension, then

$$m_{\alpha,F}(x) = \prod_{\alpha' \in \mathscr{O}_\alpha} (x - \alpha'),$$

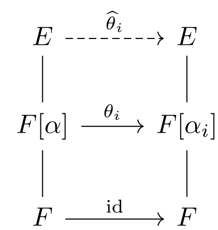
where $\mathscr{O}_\alpha := \mathrm{Aut}_F(E) \cdot \alpha$.

Proof. ($\Rightarrow$) Since E/F is a normal extension, there are $\alpha_1, \ldots, \alpha_n \in E$ such that

$$m_{\alpha,F}(x) = (x - \alpha_1) \cdots (x - \alpha_n). \tag{25.1}$$

Since α and α_i are zeros of an irreducible polynomial $m_{\alpha,F}$ in $F[x]$, by Lemma 16.2.2 there is an F-isomorphism $\theta_i : F[\alpha] \to F[\alpha_i]$ such that $\theta_i(\alpha) = \alpha_i$. Since E/F is a finite normal extension, by Proposition 23.4.1, there is $\widehat{\theta}_i \in \mathrm{Aut}_F(E)$ such that $\widehat{\theta}_i|_{F[\alpha]} = \theta_i$.

In particular, there is $\widehat{\theta}_i \in \mathrm{Aut}_F(E)$ such that $\widehat{\theta}_i(\alpha) = \alpha_i$. This implies that $\{\alpha_1, \ldots, \alpha_n\}$ is a subset of the orbit $\mathscr{O}_\alpha := \mathrm{Aut}_F(E) \cdot \alpha$. On the other hand, by Lemma 25.2.1 we have $\mathscr{O}_\alpha$ is a subset of $\{\alpha_1, \ldots, \alpha_n\}$. Altogether we obtain that $\mathscr{O}_\alpha = \{\alpha_1, \ldots, \alpha_n\}$. Since E/F is separable, $m_{\alpha,F}$ does not have multiple zeros in its splitting field. This means that α_i's are distinct. Therefore $|\mathscr{O}_\alpha| = n = \deg m_{\alpha,F}$.



($\Leftarrow$) Let $\mathscr{O}_\alpha := \{\alpha_1, \ldots, \alpha_n\}$. Inspired by the previous part, we want to show that $m_{\alpha,F}(x) = \prod_{i=1}^n (x - \alpha_i)$. Since α_i's are distinct zeros of $m_{\alpha,F}$, by the generalized factor theorem (see 7.1.2) there is $q(x) \in E[x]$ such that

$$m_{\alpha,F}(x) = q(x)(x - \alpha_1) \cdots (x - \alpha_n). \tag{25.2}$$

By hypothesis, $n = \deg m_{\alpha,F}$. Therefore comparing the degree and the leading coefficients of both sides of (25.2), we obtain that $q(x) = 1$. Thus

$$m_{\alpha,F}(x) = (x - \alpha_1) \cdots (x - \alpha_n). \tag{25.3}$$

By (25.3), we deduce that the minimal polynomial $m_{\alpha,F}$ can be decomposed into linear factors over E. Hence E/F is a normal extension. From (25.3), we also obtain that $m_{\alpha,F}$ has distinct zeros in its splitting field. Hence E/F is a separable extension. Altogether, we have E/F is a Galois extension, which completes the proof. □

25.4 Subgroups and intermediate subfields

In this section, we show that the set of fixed points $\mathrm{Fix}(G)$ of a subgroup G of the group of symmetries of a finite field extension E/F gives us an intermediate subfield and $E/\mathrm{Fix}(G)$ is a Galois extension. Later we will see that is indeed gives us a bijection between subgroups of $\mathrm{Aut}_F(E)$ and intermediate subfields if E/F is a finite Galois extension.

Suppose E/F is a field extension and G is a subgroup of $\mathrm{Aut}_F(E)$. Let

$$\mathrm{Fix}(G) := \{\alpha \in E \mid \forall \theta \in G, \theta(\alpha) = \alpha\}.$$

Lemma 25.4.1. *Suppose E/F is a field extension and G is a subgroup of $\mathrm{Aut}_F(E)$. Then $F \subseteq \mathrm{Fix}(G) \subseteq E$ is a chain of fields.*

Proof. Notice that for every $a \in F$ and every $\theta \in G \subseteq \mathrm{Aut}_F(E)$, we have $\theta(a) = a$. Hence $F \subseteq \mathrm{Fix}(G)$. Clearly $\mathrm{Fix}(G) \subseteq E$. Next we argue why $\mathrm{Fix}(G)$ is a subfield of E. For every $\alpha, \beta \in \mathrm{Fix}(G)$ and $\theta \in G$, we have

$$\theta(\alpha - \beta) = \theta(\alpha) - \theta(\beta) = \alpha - \beta.$$

Hence $\alpha - \beta \in \mathrm{Fix}(G)$. Similarly we have

$$\theta(\alpha\beta) = \theta(\alpha)\theta(\beta) = \alpha\beta,$$

and so $\alpha\beta \in \mathrm{Fix}(G)$. Therefore $\mathrm{Fix}(G)$ is a subring of E.

Suppose $\alpha \in \mathrm{Fix}(G)$ is a non-zero element. Since E is a field, $\alpha^{-1} \in E$. For every $\theta \in G$, we have

$$\theta(\alpha^{-1}) = \theta(\alpha)^{-1} = \alpha^{-1}.$$

Thus $\alpha \in \mathrm{Fix}(G)$. This shows that $\mathrm{Fix}(G)$ is a subfield of E, which completes the proof. □

Theorem 25.4.2. *Suppose E/F is a finite field extension and G is a finite subgroup of $\mathrm{Aut}_F(E)$. Then the following statements hold.*

1. *For every $\alpha \in E$, $m_{\alpha,\mathrm{Fix}(G)}(x) = \prod_{\alpha' \in G\cdot\alpha}(x - \alpha')$ where $G \cdot \alpha$ is the G-orbit of α.*
2. *The field extension $E/\mathrm{Fix}(G)$ is Galois.*

The first item is inspired by Theorem 25.3.1, but here we are saying that it is enough to only consider G-orbit instead of $\mathrm{Aut}_{\mathrm{Fix}(G)}(E)$-orbit. Later we will show that $G = \mathrm{Aut}_{\mathrm{Fix}(G)}(E)$.

Proof. Let $f_{\alpha,G}(x) := \prod_{\alpha' \in G\cdot\alpha}(x - \alpha')$. We have to show that $f_{\alpha,G} = m_{\alpha,\mathrm{Fix}(G)}$; in particular, we need to show that all the coefficients of $f_{\alpha,G}$ are fixed under the action of G. We extend the action of G to the ring of polynomials $E[x]$, and we need to show that $\theta(f_{\alpha,G}) = f_{\alpha,G}$ for every $\theta \in G$. For every $\theta \in G$, we have

$$\theta(f_{\alpha,G}) = \prod_{\alpha' \in G\cdot\alpha} (x - \theta(\alpha')). \tag{25.4}$$

Hence we need to understand what we get as we apply $\theta \in G$ to the elements of a G-orbit. For every $\alpha' \in G \cdot \alpha$, there is $\theta' \in G$ such that $\alpha' = \theta'(\alpha)$. Hence

$$\theta(\alpha') = \theta(\theta'(\alpha)) = \underbrace{(\theta \circ \theta')}_{\in G}(\alpha) \in G \cdot \alpha.$$

Thus $\alpha' \mapsto \theta(\alpha')$ induces a map from $G \cdot \alpha$ to $G \cdot \alpha$. Similarly $\alpha' \mapsto \theta^{-1}(\alpha')$ induces a map from $G \cdot \alpha$ to itself. As θ^{-1} is the inverse of θ, we deduce that θ simply permutes elements of the G-orbit $G \cdot \alpha$. Hence (25.4) implies that

$$\theta(f_{\alpha,G}) = \prod_{\alpha'' \in G\cdot\alpha} (x - \alpha'') = f_{\alpha,G}$$

for every $\theta \in G$. This means all the coefficients of $f_{\alpha,G}$ are fixed by all the elements of G. Hence

$$f_{\alpha,G} \in \mathrm{Fix}(G)[x]. \tag{25.5}$$

Next notice that α is in the G-orbit of α as $\alpha = \mathrm{id}_E(\alpha)$ and $\mathrm{id}_E \in G$. Hence α is a zero of $f_{\alpha,G}$. Therefore by (25.5), we have that

$$m_{\alpha,\mathrm{Fix}(G)} | f_{\alpha,G} \tag{25.6}$$

in $\mathrm{Fix}(G)[x]$ (see Proposition 8.2.6); in particular, we obtain

$$\deg m_{\alpha,\mathrm{Fix}(G)} \leq \deg f_{\alpha,G} = |G \cdot \alpha|. \tag{25.7}$$

On the other hand, by Lemma 25.2.1, we have

$$|\operatorname{Aut}_{\mathrm{Fix}(G)}(E) \cdot \alpha| \leq \deg m_{\alpha,\mathrm{Fix}(G)}. \tag{25.8}$$

Notice that for every $\theta \in G$, we have $\theta|_{\mathrm{Fix}(G)} = \mathrm{id}_{\mathrm{Fix}(G)}$, and so

$$G \subseteq \operatorname{Aut}_{\mathrm{Fix}(G)}(E). \tag{25.9}$$

By (25.7), (25.8), and (25.9), we obtain that

$$|\operatorname{Aut}_{\mathrm{Fix}(G)}(E) \cdot \alpha| \leq \deg m_{\alpha,\mathrm{Fix}(G)} \leq |G \cdot \alpha| \leq |\operatorname{Aut}_{\mathrm{Fix}(G)}(E) \cdot \alpha|.$$

Hence all these quantities are equal. In particular,

$$\deg m_{\alpha,\mathrm{Fix}(G)} = |\operatorname{Aut}_{\mathrm{Fix}(G)}(E) \cdot \alpha| \tag{25.10}$$

and

$$\deg m_{\alpha,\mathrm{Fix}(G)} = \deg f_{\alpha,G}. \tag{25.11}$$

By (25.10) and Theorem 25.3.1, we deduce that $E/\mathrm{Fix}(G)$ is a Galois extension. By (25.6) and (25.11), we obtain that

$$m_{\alpha,\mathrm{Fix}(G)} = f_{\alpha,G} = \prod_{\alpha' \in G \cdot \alpha} (x - \alpha')$$

which completes the proof. □

Chapter 26

Lecture 2

So far we have proved that if E/F is a finite field extension and G is a subgroup of $\mathrm{Aut}_F(E)$, then

$$\mathrm{Fix}(G) := \{\alpha \in E \mid \forall \theta \in G, \theta(\alpha) = \alpha\}$$

is an intermediate subfield of E/F and $E/\,\mathrm{Fix}(G)$ is a Galois extension.

26.1 Fixed points of a subgroup.

We want to show that $[E : \mathrm{Fix}(G)] = |G|$ and deduce that $\mathrm{Aut}_{\mathrm{Fix}(G)}(E) = G$. To show this, we start with the following lemma.

Lemma 26.1.1. *Suppose E is a field and G is a subgroup of $\mathrm{Aut}(E)$. For $\theta \in G$ and $\mathbf{c} := (c_1, \ldots, c_n) \in E^n$, let $\theta(\mathbf{c}) := (\theta(c_1), \ldots, \theta(c_n))$. Suppose V is a subspace of E^n which satisfies the following properties:*

1. *$V \neq \{0\}$ and*

2. *V is G-invariant; that mean for every $\theta \in G$ and $\mathbf{c} \in V$, $\theta(\mathbf{c}) \in V$.*

Then V has a non-zero G-fixed point; that means $V \cap (\mathrm{Fix}(G))^n \neq \{0\}$.

Proof. For $\mathbf{c} \in E^n$, let $\ell(\mathbf{c})$ be the number of non-zero components of $\mathbf{c}$. Let

$$m := \min\{\ell(\mathbf{c}) \mid \mathbf{c} \in V \setminus \{0\}\}.$$

Notice that since $V \neq \{0\}$, m is a positive integer. Suppose $\mathbf{c}_0 \in V$ is such that $\ell(\mathbf{c}_0) = m$.

Notice that in this setting, if $\mathbf{x} \in V$ and $\ell(\mathbf{x}) < m$, then $\mathbf{x} = 0$. In what follows, we start with $\mathbf{c}_0$ and apply a series of algebraic manipulation which does not either increase the number of non-zero components or take us outside of V, and at the end we make sure that the number of non-zero components decreases at least by one. By the earlier remark, we deduce that the final vector should be zero. This will help us obtain the desired result. This technique is very useful and common.

After rearranging the components, if needed, we can and will assume that

$$\mathbf{c} = (\alpha_1, \ldots, \alpha_m, 0, \cdots, 0).$$

Since V is closed under scalar multiplication, $\alpha_1^{-1}\mathbf{c} \in V$. Let $\mathbf{c}' := \alpha_1^{-1}\mathbf{c}$. Then

$$\mathbf{c}' := (1, \beta_2, \ldots, \beta_m, 0, \ldots, 0).$$

Since V is G-invariant, for every $\theta \in G$ we have $\theta(\mathbf{c}') \in V$. As V is closed under subtraction, we obtain that $\mathbf{c}' - \theta(\mathbf{c}') \in V$. Notice that

$$\mathbf{c}' - \theta(\mathbf{c}') = (0, \beta_2 - \theta(\beta_2), \ldots, \beta_m - \theta(\beta_m), 0, \ldots, 0),$$

and so $\ell(\mathbf{c}' - \theta(\mathbf{c}')) < m$. Hence every $\theta \in G$, we have $\mathbf{c}' = \theta(\mathbf{c}')$. Therefore $\mathbf{c}' \in V \cap (\mathrm{Fix}(G))^n$, which completes the proof as $\mathbf{c}' \neq 0$. □

Proposition 26.1.2. *Suppose E is a field and G is a finite subgroup of $\mathrm{Aut}(E)$. Then $[E : \mathrm{Fix}(G)] \leq |G|$.*

Proof. Suppose $G = \{\sigma_1 = \mathrm{id}_E, \sigma_2, \ldots, \sigma_n\}$ and σ_i's are distinct. Suppose to the contrary that $[E : \mathrm{Fix}(G)] > n$. Hence there are $e_1, \ldots, e_{n+1} \in E$ that are $\mathrm{Fix}(G)$-linearly independent. Let

$$\mathbf{v}_i := (\sigma_1(e_i), \ldots, \sigma_n(e_i)) \in E^n$$

for $i \in [1..(n+1)]$. Since $\dim_E E^n = n$, $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ are E-linearly dependent. This means

$$V := \left\{\mathbf{c} := (c_1, \ldots, c_{n+1}) \in E^{n+1} \mid \sum_{i=1}^{n+1} c_i \mathbf{v}_i = 0\right\}$$

has a non-zero element.

Claim 1. *V is a subspace.*

Proof of Claim 1. We need to show that V is closed under subtraction and scalar multiplication. Suppose $\mathbf{c}, \mathbf{c}' \in V$. Then $\sum_{i=1}^{n+1} c_i \mathbf{v}_i = 0$ and $\sum_{i=1}^{n+1} c_i' \mathbf{v}_i = 0$. Hence $\sum_{i=1}^{n+1} (c_i - c_i')\mathbf{v}_i = 0$, which implies that $\mathbf{c} - \mathbf{c}' \in V$. For $e \in E$, we have $e(\sum_{i=1}^{n+1} c_i \mathbf{v}_i) = \sum_{i=1}^{n+1} (ec_i)\mathbf{v}_i = 0$. This means $e\mathbf{c} \in V$. Claim follows. □

Claim 2. *V is G-invariant.*

Proof of Claim 2. Suppose $\mathbf{c} \in V$. Then $\sum_{i=1}^{n+1} c_i \mathbf{v}_i = 0$. The j-th component of $\sum_{i=1}^{n+1} c_i \mathbf{v}_i$ is $\sum_{i=1}^{n+1} c_i \sigma_j(e_i)$. For every $\sigma \in G$, we have

$$\sum_{i=1}^{n+1} \sigma(c_i)\sigma(\sigma_j(e_i)) = 0. \tag{26.1}$$

Notice that

$$G = \{\sigma \circ \sigma_1, \ldots, \sigma \circ \sigma_n\}. \tag{26.2}$$

By (26.1) and (26.2), we deduce that $\sum_{i=1}^{n+1} \sigma(c_i)\sigma_{j'}(e_i) = 0$ for every j'. This implies that $\sum_{i=1}^{n+1} \sigma(c_i)\mathbf{v}_i = 0$. Thus $\sigma(\mathbf{c}) \in V$, which completes the proof of Claim 2. □

By Claim 1, Claim 2, and Lemma 26.1.1, there is a non-zero vector

$$(a_1, \ldots, a_{n+1}) \in V \cap (\mathrm{Fix}(G))^n. \tag{26.3}$$

This means that $\sum_{i=1}^{n+1} a_i \mathbf{v}_i = 0$. Since the first component of $\mathbf{v}_i$ is e_i, we obtain that $\sum_{i=1}^{n+1} a_i e_i = 0$. As a_i's are in $\mathrm{Fix}(G)$, we deduce that e_i's are $\mathrm{Fix}(G)$-linearly dependent. This is a contradiction, which completes the proof. □

The following is an immediate corollary of these results.

Theorem 26.1.3. *Suppose E is a field and G is a finite subgroup of $\mathrm{Aut}(E)$. Then $E/\,\mathrm{Fix}(G)$ is a Galois extension, $[E : \mathrm{Fix}(G)] = |G|$, and $\mathrm{Aut}_{\mathrm{Fix}(G)}(E) = G$.*

Proof. Let $F := \mathrm{Fix}(G)$. Then by Proposition 26.1.2, we have

$$[E : F] \le |G| < \infty. \tag{26.4}$$

We also notice that

$$G \subseteq \mathrm{Aut}_{\mathrm{Fix}(G)}(E) = \mathrm{Aut}_F(E). \tag{26.5}$$

By (26.4), (26.5), and Theorem 25.4.2, we deduce that $E/\,\mathrm{Fix}(G)$ is a Galois extension. Hence by Theorem 24.2.2, we have

$$[E : \mathrm{Fix}(G)] = |\,\mathrm{Aut}_{\mathrm{Fix}(G)}(E)|. \tag{26.6}$$

By (26.4), (26.5), and (26.6), we obtain

$$[E : \mathrm{Fix}(G)] = |\,\mathrm{Aut}_{\mathrm{Fix}(G)}(E)| \ge |G| \ge [E : \mathrm{Fix}(G)].$$

Hence $[E : \mathrm{Fix}(G)] = |G|$ and $|G| = |\,\mathrm{Aut}_{\mathrm{Fix}(G)}(E)|$. The latter together with (26.5) implies that $G = \mathrm{Aut}_{\mathrm{Fix}(G)}(E)$, which completes the proof. □

Corollary 26.1.4. *Suppose E/F is a finite field extension. Then E/F is a Galois extension if and only if $Fix(\mathrm{Aut}_F(E)) = F$.*

Proof. ($\Rightarrow$) Since E/F is a finite Galois extension, by Theorem 24.2.2

$$[E : F] = |\,\mathrm{Aut}_F(E)|. \tag{26.7}$$

Applying Theorem 26.1.3 for $G := \mathrm{Aut}_F(E)$, we obtain that

$$|\,\mathrm{Aut}_F(E)| = [E : \mathrm{Fix}(\mathrm{Aut}_F(E))]. \tag{26.8}$$

Notice that $F \subseteq \mathrm{Fix}(\mathrm{Aut}_F(E))$, and so by (26.7), (26.8), and the Tower Rule, we obtain that $[\mathrm{Fix}(\mathrm{Aut}_F(E)) : F] = 1$. Hence $\mathrm{Fix}(\mathrm{Aut}_F(E)) = F$.

($\Leftarrow$) Since $F = \mathrm{Fix}(\mathrm{Aut}_F(E))$, by Theorem 26.1.3, E/F is a Galois extension. This completes the proof. □

26.2 Fundamental Theorem of Galois Theory

In this section, we prove the Fundamental Theorem of Galois theory. In order to avoid a very long statement, we split this theorem into two parts. The first theorem gives us a concrete correspondence between the set

$$\mathrm{Int}(E/F) := \{K \mid K \text{ is an intermediate subfield of } E/F\}$$

of intermediate subfields of E/F and the set

$$\mathrm{Sub}(\mathrm{Aut}_F(E)) := \{H \mid H \leq \mathrm{Aut}_F(E)\}.$$

of all the subgroups of $\mathrm{Aut}_F(E)$.

Theorem 26.2.1 (Fundamental Theorem of Galois Theory: correspondence)**.** *Suppose E/F is a finite Galois extension. Then*

$$\Psi : \mathrm{Int}(E/F) \to \mathrm{Sub}(\mathrm{Aut}_F(E)), \quad \Psi(K) := \mathrm{Aut}_K(E) \qquad \text{, and}$$
$$\Phi : \mathrm{Sub}(\mathrm{Aut}_F(E)) \to \mathrm{Int}(E/F), \quad \Phi(G) := \mathrm{Fix}(G)$$

are inverse of each other.

Proof. We need to show that we have $\Psi(\Phi(G)) = G$ for every $G \leq \mathrm{Aut}_F(E)$ and $\Phi(\Psi(K)) = K$ for every intermediate subfield K.

Notice that, for a subgroup G, $\Psi(\Phi(G)) = \mathrm{Aut}_{\mathrm{Fix}(G)}(E)$, and by Theorem 26.1.3 the latter is G. Hence $\Psi(\Phi(G)) = G$.

For an intermediate subfield K, $\Phi(\Psi(K)) = \mathrm{Fix}(\mathrm{Aut}_K(E))$ and we want to show that this is K. By Corollary 26.1.4, we have $\mathrm{Fix}(\mathrm{Aut}_K(E)) = K$ exactly when E/K is a Galois extension. Hence the claim follows as soon as we show E/K is Galois for every intermediate subfield.

To show E/K is a Galois extension, we can argue that for every $\alpha \in E$, the minimal polynomial $m_{\alpha,K}$ of α over K can be decomposed as a product of degree one factors (*normal extension*) and all of its zeros are distinct (*separable extension*).

We start with our hypothesis that E/F is a Galois extension. Since E/F is Galois, the above properties hold for the minimal polynomial $m_{\alpha,F}$ of α over F. This means there are *distinct* elements $\alpha_1, \ldots, \alpha_n \in E$ such that

$$m_{\alpha,F}(x) = (x - \alpha_1) \cdots (x - \alpha_n). \tag{26.9}$$

Notice that $m_{\alpha,F}(x) \in F[x] \subseteq K[x]$ and α is a zero of $m_{\alpha,F}(x)$. Hence by Proposition 8.2.6, $m_{\alpha,K}(x)$ divides $m_{\alpha,F}(x)$ in $K[x]$. Therefore $m_{\alpha,K}(x)$ divides $m_{\alpha,F}(x)$ in $E[x]$. By (26.9), $x - \alpha_i$'s are all the irreducible factors of $m_{\alpha,F}(x)$ in $E[x]$. As $E[x]$ is a UFD, we obtain that

$$m_{\alpha,K}(x) = (x - \alpha_{i_1}) \cdots (x - \alpha_{i_m}) \tag{26.10}$$

for some integers $1 \leq i_1 < \cdots < i_m \leq n$. By (26.10), we deduce that $m_{\alpha,K}$ can be factored as a product of degree one factors in $E[x]$ and all of its zeros are distinct. Claim follows. □

Let's make a few remarks:

1. Properties of field extensions have the *tendency to stay at the surface*! We have seen that if E/F is a normal extension and K is an intermediate subfield of E/F, then E/K is normal extension, but the extension over the base field K/F is *not necessarily* a normal extension (see Section 23.2). In Theorem 26.2.1, we see that if E/F is a finite Galois extension and K is an intermediate subfield of E/F, then E/K is a finite Galois extension. In the second part of the Fundamental Theorem of Galois Theory, we give a group theoretic criterion for the extension of an intermediate subfield over the base field be a Galois extension.

2. Using the Tower Rule, we have seen that E/F is a finite extension if and only if both E/K and K/F are finite extensions for every intermediate subfield K of E/F. We will show that the same property holds for separable extensions.

3. Notice that $\mathrm{Int}(E/F)$ and $\mathrm{Sub}(\mathrm{Aut}_F(E))$ are *partially ordered* by inclusion, where E/F is a finite Galois extension. From this point of view, Φ and Ψ are *order reversing* bijections. That means in the setting of Theorem 26.2.1, if $H_1 \leq H_2$ are two subgroups of $\mathrm{Aut}_F(E)$, then $\Phi(H_2) \subseteq \Phi(H_1)$ (if α is fixed under the action of H_2, then it is also fixed under the action of H_1). Similarly, if $K_1 \subseteq K_2$ are two intermediate subfields of E/F, then $\Psi(K_2) \subseteq \Psi(K_1)$ (if restriction of an automorphism to K_2 is identity, then its restriction to K_1 is also identity).

Theorem 26.2.2 (Fundamental Theorem of Galois Theory: index and normal)**.** *Suppose E/F is a finite Galois extension. Let*

$$\Psi : \mathrm{Int}(E/F) \to \mathrm{Sub}(\mathrm{Aut}_F(E)), \quad \Psi(K) := \mathrm{Aut}_K(E) \qquad \text{, and}$$
$$\Phi : \mathrm{Sub}(\mathrm{Aut}_F(E)) \to \mathrm{Int}(E/F), \quad \Phi(G) := \mathrm{Fix}(G).$$

Then the following holds.

1. *For $K_1 \subseteq K_2$ in $\mathrm{Int}(E/F)$, $[\Psi(K_1) : \Psi(K_2)] = [K_2 : K_1]$. For $G_1 \subseteq G_2$ in $\mathrm{Sub}(\mathrm{Aut}_F(E))$, $[\Phi(G_1) : \Phi(G_2)] = [G_2 : G_1]$.*

2. *For every $K \in \mathrm{Int}(E/F)$, K/F is a normal extension if and only if $\Psi(K)$ is a normal subgroup of $\mathrm{Aut}_F(E)$.*

   For every $N \in \mathrm{Sub}(\mathrm{Aut}_F(E))$, N is a normal subgroup of $\mathrm{Aut}_F(E)$ if and only if $\Phi(N)/F$ is a normal extension.

3. *For an intermediate subfield K of E/F, K/F is a normal extension if and only if K/F is a Galois extension.*

Proof. 1. As E/K is a Galois extension, by Theorem 24.2.2 we obtain

$$|\Psi(K)| = |\mathrm{Aut}_K(E)| = [E : K]. \tag{26.11}$$

By (26.11) and the Tower Rule we deduce that

$$[\Psi(K_1) : \Psi(K_2)] = \frac{|\Psi(K_1)|}{|\Psi(K_2)|} = \frac{[E : K_1]}{[E : K_2]} = [K_2 : K_1]. \tag{26.12}$$

From (26.12) and the correspondence part of the fundamental theorem of Galois theory, we obtain

$$[\Phi(G_1) : \Phi(G_2)] = [\Psi(\Phi(G_2)) : \Psi(\Phi(G_1))] = [G_2 : G_1].$$

2. Suppose N is a normal subgroup of $\mathrm{Aut}_F(E)$. We want to show that $\mathrm{Fix}(N)/F$ is a normal extension. This means we have to show that, for every $\alpha \in \mathrm{Fix}(N)$, the minimal polynomial $m_{\alpha,F}(x)$ of α over F can be decomposed as a product of linear factors in $\mathrm{Fix}(N)[x]$.

Since E/F is a finite Galois extension, by Theorem 25.3.1 we have

$$m_{\alpha,F}(x) = \prod_{\alpha' \in \mathrm{Aut}_F(E)\cdot\alpha} (x - \alpha'). \tag{26.13}$$

By (26.13) and the above argument, we obtain that $\mathrm{Fix}(N)/F$ is a normal extension as soon as we show

$$\mathrm{Aut}_F(E)\cdot\alpha \subseteq \mathrm{Fix}(N). \tag{26.14}$$

The assertion of (26.14) holds if and only if $\sigma(\theta(\alpha)) = \theta(\alpha)$ for every $\theta \in \mathrm{Aut}_F(E)$ and $\sigma \in N$. Notice that $\theta^{-1} \circ \sigma \circ \theta \in N$ for every $\theta \in \mathrm{Aut}_F(E)$ and $\sigma \in N$ as N is a normal subgroup of $\mathrm{Aut}_F(E)$. Since $\alpha \in \mathrm{Fix}(N)$ and $\sigma^{-1} \circ \theta \circ \sigma \in N$ for every $\theta \in \mathrm{Aut}_F(E)$ and $\sigma \in N$, we obtain that $(\sigma^{-1} \circ \theta \circ \sigma)(\alpha) = \alpha$. Therefore $\sigma(\theta(\alpha)) = \theta(\alpha)$ for every $\theta \in \mathrm{Aut}_F(E)$ and $\sigma \in N$. Altogether we deduce that (26.14) holds, and so by (26.13), all the zeros of $m_{\alpha,F}$ are in $\mathrm{Fix}(N)$ and they are distinct. Hence $\mathrm{Fix}(N)/F$ is a Galois extension. This means we have proved that

$$N \trianglelefteq \mathrm{Aut}_F(E) \Rightarrow \Phi(N)/F \text{ is a Galois extension.} \tag{26.15}$$

Next suppose K/F is a normal extension. Then by Proposition 23.1.1, $\mathrm{Aut}_K(E)$ is a normal subgroup of $\mathrm{Aut}_F(E)$. This means $\Phi(K)$ is a normal subgroup of $\mathrm{Aut}_F(E)$.

One can finish the proof of the second step using the correspondence part of the fundamental theorem of Galois theory.

3. If K/F is a normal extension, then $\Psi(K)$ is a normal subgroup of $\mathrm{Aut}_F(E)$. Hence by (26.15), $\Phi(\Psi(K))/F$ is a Galois extension. Therefore by the correspondence part of the fundamental theorem of Galois theory, we obtain that K/F is a Galois extension. This completes the proof. □

Corollary 26.2.3. *If E/F is a Galois extension, then $\mathrm{Int}(E/F)$ is finite.*

Proof. By the correspondence part of the fundamental theorem of Galois theory, there is a bijection between $\mathrm{Int}(E/F)$ and $\mathrm{Sub}(\mathrm{Aut}_F(E))$. Since $\mathrm{Aut}_F(E)$ is a finite group, the claim follows. □

Next we extend the above corollary.

Proposition 26.2.4. *Suppose E/F is a finite separable extension. Then $\mathrm{Int}(E/F)$ is finite.*

Remark 26.2.5. *In fact we will prove that, if E/F is a finite separable extension and L/F is a normal closure of E/F, then L/F is a Galois extension.*

Proof of Proposition 26.2.4. Suppose $(\alpha_1, \ldots, \alpha_n)$ is an F-basis of E. Let L be a splitting field of $f(x) := \prod_{i=1}^n m_{\alpha_i,F}(x)$ over E. As in the proof of Proposition 23.3.1, L is a splitting field of f over F. For the sake of convenience, we go over the argument. Since L is a splitting field of f over E, there are $\alpha_{ij} \in L$ such that

$$m_{\alpha_i,F}(x) = (x - \alpha_{i1}) \cdots (x - \alpha_{im_i}),$$

and

$$L = E[\alpha_{11}, \ldots, \alpha_{nm_n}].$$

Since $E \subseteq L$, we can and will assume that $\alpha_{i1} = \alpha_i$. Then

$$F[\alpha_{11}, \ldots, \alpha_{nm_n}] \supseteq \mathrm{Span}_F(\alpha_1, \ldots, \alpha_n) = E.$$

Hence

$$F[\alpha_{11}, \ldots, \alpha_{nm_n}] \supseteq E[\alpha_{11}, \ldots, \alpha_{nm_n}] = L.$$

Therefore L is a splitting field of f over F.

Notice that f is a separable polynomial in $F[x]$ as E/F is a separable extension. Hence L/F is a splitting field of a separable polynomial over F, and so by Theorem 24.2.2, L/F is a finite Galois extension. Thus by Corollary 26.2.3, $\mathrm{Int}(L/F)$ is finite. As $\mathrm{Int}(E/F) \subseteq \mathrm{Int}(L/F)$, we obtain that $\mathrm{Int}(E/F)$ is finite. This completes the proof. □

Chapter 27

Lecture 3

27.1 Fundamental theorem of algebra

In this section, we use easy analysis and the fundamental theorem of Galois theory to prove the fundamental theorem of algebra.

Theorem 27.1.1. *Suppose $f \in \mathbb{C}[x] \setminus \mathbb{C}$. Then f has a zero in $\mathbb{C}$.*

Proof. Suppose to the contrary that there is $f \in \mathbb{C}[x] \setminus \mathbb{C}$ with no complex zeros. Let E be a splitting field of f over $\mathbb{C}$. Let $L/\mathbb{R}$ be a normal closure of $E/\mathbb{R}$. Notice that if $p(x) \in \mathbb{R}[x]$ is irreducible, then $p'(x)$ is a non-constant polynomial of lower degree, and so $\gcd(p, p') = 1$. Therefore, every non-constant polynomial in $\mathbb{R}[x]$ is separable. Hence $L/\mathbb{R}$ is a Galois extension. Let $G := \mathrm{Aut}_{\mathbb{R}}(L)$ and P be a Sylow 2-subgroup of G. Let $K := \mathrm{Fix}(P)$. Then by the fundamental theorem of Galois theory we have

$$[G : P] = [\mathrm{Fix}(P) : \mathrm{Fix}(G)] = [K : \mathbb{R}]. \tag{27.1}$$

By (27.1), we obtain that $[K : \mathbb{R}]$ is odd. Notice that for every $\alpha \in K$, by the Tower Rule $[\mathbb{R}[\alpha] : \mathbb{R}]$ divides $[K : \mathbb{R}]$, and so $[\mathbb{R}[\alpha] : \mathbb{R}]$ is odd. This implies that $\deg m_{\alpha,\mathbb{R}}(x)$ is odd for every $\alpha \in K$. Notice that if $p(x) \in \mathbb{R}[x]$ is a monic odd degree polynomial, then $\lim_{x\to\infty} p(x) = \infty$ and $\lim_{x\to-\infty} p(x) = -\infty$. Hence by the intermediate value theorem, p has a real zero. Therefore if $p(x) \in \mathbb{R}[x]$ is monic, irreducible, and of odd degree, then $\deg p = 1$. Altogether, we obtain that $\deg m_{\alpha,\mathbb{R}} = 1$ for every $\alpha \in K$. This means $K = \mathbb{R}$. Therefore $\mathrm{Fix}(P) = \mathrm{Fix}(G)$, and so by the fundamental theorem of Galois theory, we deduce that $G = P$. This means G is a finite 2-group. If $|G| \geq 4$, then by the first Sylow theorem there are subgroups $G_2 \subseteq G_1$ of G such that $[G : G_1] = 2$ and $[G : G_2] = 4$. Thus by the fundamental theorem of Galois theory, $[\mathrm{Fix}(G_1) : \mathbb{R}] = 2$. Hence for every $\alpha \in \mathrm{Fix}(G_1) \setminus \mathbb{R}$, we have that $\mathrm{Fix}(G_1) = \mathbb{R}[\alpha]$ and $m_{\alpha,\mathbb{R}}$ is a degree 2 irreducible element of $\mathbb{R}[x]$. Notice that such a polynomial has a complex zero α'. Since $[\mathbb{C} : \mathbb{R}] = 2$, $\mathbb{C} = \mathbb{R}[\alpha']$. Hence both $\mathbb{C}$ and $\mathrm{Fix}(G_1)$ are isomorphic to $\mathbb{R}[x]/\langle m_{\alpha,\mathbb{R}}\rangle$. Because $[G_1 : G_2]$, by the fundamental theorem of Galois theory $[\mathrm{Fix}(G_2) : \mathrm{Fix}(G_1)] = 2$. Since $\mathrm{Fix}(G_1) \simeq \mathbb{C}$, we obtain that $\mathbb{C}$ has an extension field E such that $[E : \mathbb{C}] = 2$. Hence for every $\alpha \in E \setminus \mathbb{C}$, $m_{\alpha,\mathbb{C}}$ is a monic quadratic irreducible element of $\mathbb{C}[x]$. Next we show that there is no such polynomial and get a contradiction.

For every $a, b \in \mathbb{C}$, $x^2 + ax + b = 0$ if and only if $(x + a/2)^2 = (a^2/4) - b$. Writing the right hand side in polar coordinates, we get the equation $(x + a/2)^2 = re^{i\theta}$ for some $r \in \mathbb{R}^{\geq 0}$ and $\theta \in \mathbb{R}$. Clearly $\pm\sqrt{r}e^{i\theta/2} - (a/2) \in \mathbb{C}$ satisfy this equation. Hence every monic degree complex polynomial has a complex zero. This completes the proof. $\square$

Going through the above proof, one can see that the crucial properties of $\mathbb{C}$ and $\mathbb{R}$ are the following items:

1. Every non-linear irreducible polynomial in $\mathbb{R}[x]$ is of even degree.
2. If $[E : \mathbb{R}] = 2$, then $E \simeq \mathbb{C}$.
3. Every degree 2 polynomial in $\mathbb{C}[x]$ has a zero in $\mathbb{C}$.

It is a good exercise to formulate a more general setting for which a similar statement holds.

27.2 Primitive Element Theorem

In Proposition 26.2.4, we proved that $\operatorname{Int}(E/F)$ is a finite set if E/F is a finite separable extension. Here we give a necessary and sufficient condition for $\operatorname{Int}(E/F)$ to be a finite set.

Theorem 27.2.1. *(Primitive Element Theorem: intermediate subfields) Suppose E/F is a finite extension. Then $\operatorname{Int}(E/F)$ is finite if and only if $E = F[\alpha]$ for some $\alpha \in E$.*

We say $\alpha \in E$ is a *primitive element* of a field extension E/F if the smallest subfield of E which contains F and α is E. In this case, we write $E = F(\alpha)$, and we say E/F is a simple extension.

Proof of Theorem 27.2.1. ($\Rightarrow$) We proceed by strong induction on $[E : F]$. Since $\operatorname{Int}(E/F)$ is a finite set, there is an intermediate subfield K of $\operatorname{Int}(E/F)$ such that there is no other intermediate subfield that is strictly between K and E. Alternatively we can say that there is $K \in \operatorname{Int}(E/F)$ which is maximal among proper subfields of E. Notice that $\operatorname{Int}(K/F)$ is a subset of $\operatorname{Int}(E/F)$, and so it is finite. As $[K : F] < [E : F]$, we can use the induction hypothesis for the extension K/F and deduce that $K = F[\beta]$ for some $\beta \in K$. Since K is maximal among proper subfields of E, for every $\gamma \in E \setminus K$ we have $E = K[\gamma]$. Altogether, we obtain that $E = F[\beta, \gamma]$.

If F is a finite field, then E is also a finite field as E/F is a finite extension. Since the multiplicative group of a finite field is cyclic, there is $\alpha \in E$ such that $E^\times = \langle \alpha \rangle$. Therefore $E = F[\alpha]$, and the claim follows. So without loss of generality we can and will assume that F is infinite.

Claim. *There is $c \in F$ such that $E = F[\beta + c\gamma]$.*

Proof of Claim. Consider the function

$$F \to \mathrm{Int}(E/F), \quad c \mapsto F[\beta + c\gamma].$$

Since F is infinite and $\mathrm{Int}(E/F)$ is finite, there are pairwise distinct elements $c_1, c_2, \ldots$ in F that are mapped to the same $L \in \mathrm{Int}(E/F)$. This means for every index i we have $L = F[\beta + c_i\gamma]$. Therefore

$$\beta + c_1\gamma \text{ and } \beta + c_2\gamma \in L \tag{27.2}$$

By (27.2), we obtain that $(\beta + c_1\gamma) - (\beta + c_2\gamma) \in L$. As $c_1 - c_2 \in F^\times$, we deduce that $\gamma \in L$. Because $c_1 \in F \subseteq L$ and $\gamma \in L$, by (27.2) we obtain that $\beta \in L$. Altogether, $F[\beta, \gamma]$ is a subset of L, which implies that $L = E$. This completes the proof of Claim. □

($\Leftarrow$) Suppose $E = F[\alpha]$. Consider the following function

$$g : \mathrm{Int}(E/F) \to E[x], \quad g(K) := m_{\alpha,K}(x).$$

We prove that (1) image of g is finite and (2) g is injective. Then one immediately obtains that the domain of g is finite, which is the desired result.

Claim 1. *For every $K \in \mathrm{Int}(E/F)$, $g(K)$ is a monic divisor of $m_{\alpha,F}(x)$ in $E[x]$. In particular, the image of g is finite.*

Proof of Claim 1. For every $K \in \mathrm{Int}(E/F)$, $m_{\alpha,F}(x) \in K[x]$ and α is a zero of $m_{\alpha,F}(x)$. Hence by Proposition 8.2.6, $m_{\alpha,K}$ divides $m_{\alpha,F}(x)$ in $K[x]$. Therefore $g(K)$ divides $m_{\alpha,F}(x)$ in $E[x]$. □

Claim 2. *g is injective.*

Proof of Claim 2. Before we go to the proof of injectivity, let's point out that for every $L \in \mathrm{Int}(E/F)$ we have $E = L[\alpha]$, and so

$$[E : L] = [L[\alpha] : L] = \deg m_{\alpha,L} = \deg g(L). \tag{27.3}$$

Now suppose $g(K_1) = g(K_2) = x^m + e_{m-1}x^{m-1} + \cdots + e_0$ for some K_1 and K_2 in $\mathrm{Int}(E/F)$. By (27.3), we obtain that

$$[E : K_j] = m \tag{27.4}$$

for $j = 1, 2$. we also notice that e_i's are in K_1 and K_2 as $g(K_j) = m_{\alpha,K_j}(x) \in K_j[x]$. Hence $K := F[e_0, \ldots, e_{m-1}]$ is a subfield of K_j's and

$$g(K_j) = x^m + e_{m-1}x^{m-1} + \cdots + e_0 \in K[x]. \tag{27.5}$$

Since α is a zero of $g(K_j)$, by (27.5) we obtain that $m_{\alpha,K}(x)$ divides $g(K_j)$ in $K[x]$. Hence by (27.3), we deduce that

$$[E : K] \leq m. \tag{27.6}$$

As K is a subfield of K_j, by the Tower Rule, (27.4) and (27.6), we obtain that $K = K_j$ for $j = 1, 2$. Hence $K_1 = K_2$, and this completes the proof. □

□

The following is an immediate consequence of Theorem 27.2.1.

Theorem 27.2.2 (Primitive Element Theorem: separable case)**.** *Suppose E/F is a finite separable extension. Then $E = F[\alpha]$ for some $\alpha \in E$.*

Proof. As E/F is a finite separable extension, by Proposition 26.2.4 $\mathrm{Int}(E/F)$ is finite. Hence by Theorem 27.2.1, $E = F[\alpha]$ for some $\alpha \in E$. This completes the proof. □

27.3 Separable closure of the base field of an algebraic extension

Primitive Element Theorem for finite separable extensions gives us an extra motivation to systematically study separability condition and investigate how it behaves in a tower of field extensions.

Let's recall that an algebraic extension E/F is separable if for every $\alpha \in E$, $m_{\alpha,F}(x)$ does not multiple zeros in its splitting field over F. The next lemma shows us exactly when an irreducible polynomial $f(x) \in F[x]$ is separable; in particular, we see that in certain sense irreducible polynomials are rarely not separable.

Lemma 27.3.1. *Suppose F is a field and $f(x) \in F[x]$ is irreducible. Then $f \in F[x]$ is separable if and only if $f' \neq 0$.*

Proof. An irreducible polynomial is separable if and only if it does not multiple zeros in its splitting field. By Lemma 18.3.3, the latter holds exactly when $\gcd(f, f') = 1$. Since f is irreducible, it is not divisible by a non-constant polynomial of degree smaller than $\deg f$. Hence either $\gcd(f, f')$ is a constant multiple of f or it is 1. Notice that $\gcd(f, f')$ is a constant multiple of f precisely when f divides f'. Since $\deg f'$ is smaller than $\deg f$, f divides f' if and only if $f' = 0$. Let's summarize what we have inferred so far:

1. $f \in F[x]$ is separable $\iff \gcd(f, f') = 1$.
2. $\gcd(f, f') = 1$ or $\gcd(f, f') \sim f$.
3. $\gcd(f, f') \sim f \iff f | f'$.
4. $f | f' \iff f' = 0$.

Hence $f \in F[x]$ is not separable if and only if $f' = 0$. This completes the proof. □

The following is an important corollary of Lemma 27.3.1.

Corollary 27.3.2. *Suppose F is a field of characteristic zero. Then for every non-constant polynomial $f \in F[x]$, $\deg f' = \deg f - 1$ and every non-constant polynomial in $F[x]$ is separable.*

Proof. Suppose $f(x) = \sum_{i=0}^{n} a_i x^i \in F[x]$ and $\deg f = n$. Then $f'(x) = \sum_{i=1}^{n} i a_i x^{i-1}$. Notice that since $\operatorname{char}(F) = 0$, $n1_F \neq 0$. As $\deg f = n$, $a_n \neq 0$. Therefore $(n1_F)(a_n) = na_n \neq 0$, which implies that $\deg f' = n - 1$.

If $p(x) \in F[x]$ is irreducible, then p is not a constant polynomial. Hence $\deg p' = \deg p - 1$; in particular $p' \neq 0$. Therefore by Lemma 27.3.1, $p \in F[x]$ is separable.

An arbitrary non-constant polynomial $f \in F[x]$ is separable if and only if all of its irreducible factors in $F[x]$ are separable. Since all the irreducible polynomials of $F[x]$ are separable we conclude that all non-constant polynomials of $F[x]$ are separable. This completes the proof. □

Corollary 27.3.3. *Suppose F is a field of characteristic zero. Then every algebraic extension E/F is separable.*

Proof. This immediately follows form Corollary 27.3.2. □

Next we see what happens when $\operatorname{char}(F) = p > 0$.

Proposition 27.3.4. *Suppose F is a field of characteristic $p > 0$ and $g(x) \in F[x]$ is irreducible. Then there are $g_{\text{sep}} \in F[x]$ and $k \in \mathbb{Z}^{\geq 0}$ with the following properties.*

1. *g_{sep} is irreducible and separable in $F[x]$.*

2. *$g(x) = g_{\text{sep}}(x^{p^k})$.*

Proof. If g is separable, then let $g_{\text{sep}} := g$ and $k = 0$, and claims hold. If g is not separable, then by Lemma 27.3.1, $g'(x) = 0$. Suppose $g(x) := \sum_{i=0}^{\infty} a_i x^i$. Then $g' = 0$ implies that for every positive integer i we have $ia_i = 0$. If the characteristic p does not divide i, then $(i1_F) \in F^{\times}$. Hence $(i1_F)a_i = ia_i = 0$ implies that $a_i = 0$. Therefore

$$g(x) = \sum_{i=0}^{\infty} a_{pi} x^{pi}. \tag{27.7}$$

Let $g_1(x) := \sum_{i=0}^{\infty} a_{pi} x^i$. Thus by (27.7), we conclude that $g(x) = g_1(x^p)$.

Claim. *g_1 is irreducible in $F[x]$.*

Proof of Claim. Suppose to the contrary that there are non-constant polynomials h_1 and h_2 in $F[x]$ such that $g_1(x) = h_1(x)h_2(x)$. Then $g(x) = g_1(x^p) = h_1(x^p)h_2(x^p)$ and $h_i(x^p)$'s are non-constant polynomials. This contradicts the hypothesis that g is irreducible in $F[x]$. □

Now we can repeat this argument for g_1. Formally we use strong induction on $\deg g$. The base of induction follows from the fact that every degree 1 polynomial in $F[x]$ is separable. So we focus on the strong induction step. We have already discussed the case where g is separable. Thus we can and will assume that g is not separable in $F[x]$. In this case, by the above argument, there is an irreducible polynomial g_1 such that $g(x) = g_1(x^p)$. In particular, we have $\deg g_1 = \frac{\deg g}{p} < \deg g$. Hence the strong induction hypothesis can be applied to g_1, and we obtain that there are a separable irreducible polynomial $g_{\rm sep}$ in $F[x]$ and a non-negative integer k such that

$$g_1(x) = g_{\rm sep}(x^{p^k}).$$

This implies that

$$g(x) = g_1(x^p) = g_{\rm sep}((x^p)^{p^k}) = g_{\rm sep}(x^{p^{k+1}}).$$

This completes the proof of the strong induction step. □

The polynomial $g_{\rm sep}$ is called *the separable form* of g. We use separable forms of minimal polynomials in an algebraic extension E/F in order to define the *separable closure* of the base field F in E.

Theorem 27.3.5. *Suppose E/F is an algebraic closure. Let*

$$E_{\rm sep} := \{\alpha \in E \mid m_{\alpha,F}(x) \text{ is separable in } F[x]\}.$$

Then the following statements hold.

1. *$E_{\rm sep} \in \mathrm{Int}(E/F)$ and $E_{\rm sep}/F$ is separable.*
2. *If $\mathrm{char}(F) = 0$, then $E_{\rm sep} = E$.*
3. *If $\mathrm{char}(F) = p > 0$, then for every $\alpha \in E$, $\alpha^{p^k} \in E_{\rm sep}$ for some non-negative integer k.*

Proof. 1. We need to show that, if $\alpha, \beta \in E_{\rm sep}$ and $\alpha \neq 0$, then $\alpha \pm \beta$, $\alpha\beta$, and α^{-1} are in $E_{\rm sep}$.

Since α and β are in $E_{\rm sep}$, $f(x) := m_{\alpha,F}(x)m_{\beta,F}(x)$ is separable in $F[x]$. Let L be a splitting field of f over $F[\alpha, \beta]$. This means that there are α_i's and β_j's in L such that

$$m_{\alpha,F}(x) = (x - \alpha_1)\cdots(x - \alpha_r), \quad m_{\beta,F}(x) = (x - \beta_1)\cdots(x - \beta_s),$$

and

$$L = (F[\alpha, \beta])[\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s].$$

Since α and β are in L, we can and will assume that $\alpha_1 = \alpha$ and $\beta_1 = \beta$. Hence

$$L = F[\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s].$$

This means L is generated by F and the zeros of f, and so L is a splitting field of f over F. Since $f \in F[x]$ is separable, by Theorem 24.2.2 we obtain that L/F is a Galois extension. Hence

$$\text{for every } \gamma \in L, m_{\gamma,F}(x) \text{ is a separable element of } F[x]. \tag{27.8}$$

As $F[\alpha, \beta] \subseteq L \cap E$, by (27.8) we infer that $F[\alpha, \beta] \subseteq E_{\text{sep}}$. In particular, $\alpha \pm \beta$, $\alpha\beta$, and α^{-1} are in E_{sep}, and $F \subseteq E_{\text{sep}}$. This implies that E_{sep} is an intermediate subfield of E/F. Clearly E_{sep}/F is a separable extension.

2. If $\text{char}(F) = 0$, then by Corollary 27.3.2 every polynomial in $F[x]$ is separable. Hence $E_{\text{sep}} = E$.

3. For every $\alpha \in E$, let $s_{\alpha,F}(x) \in F[x]$ be the separable form of $m_{\alpha,F}(x)$ (see Proposition 27.3.4); that means $s_{\alpha,F}(x) := (m_{\alpha,F})_{\text{sep}}$ is a separable irreducible element of $F[x]$ and for some non-negative integer k (this integer depends on α) we have

$$m_{\alpha,F}(x) = s_{\alpha,F}(x^{p^k}). \tag{27.9}$$

By (27.9), we conclude that α^{p^k} is a zero of $s_{\alpha,F}$. Since $s_{\alpha,F}$ is a monic irreducible element of $F[x]$, by Theorem 8.2.5 we infer that the minimal polynomial $m_{\alpha^{p^k},F}$ of α^{p^k} over F is $s_{\alpha,F}$. Hence $m_{\alpha^{p^k},F}(x)$ is a separable element of $F[x]$. Therefore α^{p^k} is in E_{sep}. This completes the proof. $\square$

Chapter 28

Lecture 4

28.1 Purely inseparable extensions

Motivated by Theorem 27.3.5, we define *purely inseparable extensions*. An algebraic extension E/F is called an *purely inseparable extension* if one (and so all) of the statements in the following proposition holds.

Proposition 28.1.1. *Suppose F is field of characteristic $p > 0$. Suppose E/F is an algebraic extension. Then the following statements are equivalent.*

1. *$E_{\mathrm{sep}} = F$; that means for every $\alpha \in E \setminus F$, $m_{\alpha,F}(x)$ is not separable in $F[x]$.*
2. *$E^{\times}/F^{\times}$ is p-torsion; that means for every $\alpha \in E$, there is a non-negative integer k such that $\alpha^{p^k} \in F$.*
3. *For every $\alpha \in E$, there are $a \in F$ and $k \in \mathbb{Z}^{\geq 0}$ such that $m_{\alpha,F}(x) = x^{p^k} - a$.*

Proof. (1) $\Rightarrow$ (2). By Theorem 27.3.5, for every $\alpha \in E$, there is a non-negative integer k such that $\alpha^{p^k} \in E_{\mathrm{sep}}$. Therefore the claim follows as $E_{\mathrm{sep}} = F$ by hypothesis.

(2) $\Rightarrow$ (3). If $\alpha = 0$, then $m_{\alpha,F}(x) = x$ and there is nothing to prove. So without loss of generality we can and will assume that $\alpha \in E^{\times}$. Since $E^{\times}/F^{\times}$ is p-torsion, there is a non-negative integer k such that $(\alpha F^{\times})^{p^k} = F^{\times}$. Then the order $o(\alpha F^{\times})$ of $\alpha F^{\times}$ in $E^{\times}/F^{\times}$ divides p^k. Hence $o(\alpha F^{\times}) = p^{k_0}$ for some non-negative integer k_0. Let $a := \alpha^{p^{k_0}}$. Notice that $a \in F$ and α is a zero of $x^{p^{k_0}} - a$. Therefore

$$m_{\alpha,F}(x) \text{ divides } x^{p^{k_0}} - a \text{ in } F[x]. \tag{28.1}$$

Notice that since $\mathrm{char}(F) = p$, we have

$$x^{p^{k_0}} - a = x^{p^{k_0}} - \alpha^{p^{k_0}} = (x - \alpha)^{p^{k_0}}. \tag{28.2}$$

By (28.1), (28.2), and the fact that $E[x]$ is a UFD, we conclude that

$$m_{\alpha,F}(x) = (x - \alpha)^m, \tag{28.3}$$

for some positive integer m which is at most p^{k_0}. Comparing the constant terms of both sides of (28.3) and using the fact that all the coefficients of $m_{\alpha,F}(x)$ are in F, we infer that $\alpha^m \in F$. Hence $(\alpha F^\times)^m = F^\times$, and so $o(\alpha F^\times)$ divides m. Therefore m is a multiple of p^{k_0}. As m is a positive integer, at most p^{k_0}, and a multiple of p^{k_0}, we conclude that $m = p^{k_0}$. This means

$$m_{\alpha,F}(x) = (x-\alpha)^{p_0^k} = x^{p^{k_0}} - \alpha^{p^{k_0}} = x^{p^{k_0}} - a,$$

which completes proof of this part.

(3) $\Rightarrow$ (1). For every $\alpha \in E \setminus F$, the minimal polynomial $m_{\alpha,F}$ is of degree at least 2. On the other hand, by hypothesis, $m_{\alpha,F}(x) = x^{p^k} - a$ for some non-negative integer k and $a \in F$. Since $\deg m_{\alpha,F}(x)$ is more than one, we conclude that k is positive. Hence the derivative $m'_{\alpha,F}$ of $m_{\alpha,F}$ is zero as the characteristic of F is p. Hence $m_{\alpha,F}$ is not separable in $F[x]$, which means $\alpha \notin E_{\text{sep}}$. This implies that $E_{\text{sep}} = F$, which completes the proof. $\square$

Combining Proposition 28.1.1 with Theorem 27.3.5, we conclude the following theorem.

Theorem 28.1.2. *Suppose E/F is an algebraic extension. Then*

1. *E/E_{sep} is purely inseparable, and*
2. *E_{sep}/F is separable.*

Based on the definition of E_{sep} we see that if K/F is separable and $K \in \text{Int}(E/F)$, then $K \subseteq E_{\text{sep}}$.

28.2 Block-Tower Phenomena for separable extensions

As we have mentioned it earlier, properties of field extensions have the *tendency of staying at the surface*. If a property can be *easily* deduced for the *bottom portion of a tower*, then we often can deduce that this property satisfies a *block-tower phenomena*! Here are a few examples and non-examples:

1. Finite extensions satisfy a block-tower phenomena: E/F is finite if and only if E/K and K/F are finite for every K in $\text{Int}(E/F)$.

2. Algebraic extensions satisfy a block-tower phenomena: E/F is algebraic if and only if E/K and K/F are algebraic for every K in $\text{Int}(E/F)$.

3. Normal extensions do not satisfy a block-tower phenomena: If E/F is a normal extension, then E/K is a normal extension, but K/F is not necessarily a normal extension.

4. Normal extensions do not satisfy a block-tower phenomena: If E/K and K/F are normal extensions for some K in $\text{Int}(E/F)$, K/F is not necessarily a normal extension.

5. Galois extensions do not satisfy a block-tower phenomena: If E/F is Galois, E/K is Galois for every K in $\mathrm{Int}(E/F)$, but K/F is not necessarily Galois. In fact, K/F is Galois if and only if $\mathrm{Aut}_K(E)$ is a normal subgroup of $\mathrm{Aut}_F(E)$.

Next we want to show that *separable extensions satisfy a block-tower phenomena*.

Theorem 28.2.1. *Suppose E/F is an algebraic extension and $K \in \mathrm{Int}(E/F)$. Then E/F is separable if and only if E/K and K/F are separable.*

Proof. ($\Rightarrow$) (The bottom portion) For every $\alpha \in K$, $m_{\alpha,F}(x)$ is separable in $F[x]$ as $\alpha \in E$ and E/F is separable. Hence K/F is separable.

(The top portion) We need to show that $m_{\alpha,K}(x)$ does not have multiple zeros in its splitting field, for every $\alpha \in E$. Our hypothesis implies that $m_{\alpha,F}(x)$ does not multiple zeros in its splitting field as E/F is separable. Similar to the proof of Theorem 27.2.1, we can argue that $m_{\alpha,K}(x)$ divides $m_{\alpha,F}(x)$ in $K[x]$. To see this, notice that α is a zero of $m_{\alpha,F}$ and $m_{\alpha,F}$ is in $K[x]$. As divisors of a polynomial with distinct linear factors cannot have multiple zeros, we conclude that $m_{\alpha,K}(x)$ is separable. Therefore E/K is separable.

($\Leftarrow$) To show the claim, it is (necessary and) sufficient to show that $E = E_{\mathrm{sep}}$. If $\mathrm{char}(F) = 0$, then polynomial is separable; so $E = E_{\mathrm{sep}}$. Hence without loss of generality we can and will assume that $\mathrm{char}(F) = p > 0$.

Since K/F is a separable extension, $K \subseteq E_{\mathrm{sep}}$.

Let's first see a less detailed argument: every $\alpha \in E$ is separable over K. Since $K \subseteq E_{\mathrm{sep}}$, we deduce that α is separable over E_{sep}. Since E/E_{sep} is purely inseparable, we conclude that $\alpha \in E_{\mathrm{sep}}$. Hence $E = E_{\mathrm{sep}}$.

Now let's have a more detailed almost identical argument: for every $\alpha \in E$, $m_{\alpha,K}(x)$ does not have multiple zeros in its splitting field over K. Since $K \subseteq E_{\mathrm{sep}}$, by a similar argument as above (see also 27.2.1) we have that $m_{\alpha,E_{\mathrm{sep}}}$ divides $m_{\alpha,K}$. Hence $m_{\alpha,E_{\mathrm{sep}}}$ does not have multiple zeros in its splitting field over E_{sep}. Hence by Lemma 27.3.1, $m'_{\alpha,E_{\mathrm{sep}}}(x) \neq 0$. Notice that since E/E_{sep} is purely inseparable, by Proposition 28.1.1 $m_{\alpha,E_{\mathrm{sep}}}(x) = x^{p^k} - a$ for some non-negative integer k and $a \in E_{\mathrm{sep}}$. Hence $m'_{\alpha,E_{\mathrm{sep}}}(x) = p^k x^{p^k-1}$ which is zero unless $k = 0$. In this case, $m_{\alpha,E_{\mathrm{sep}}}(x) = x - a$, which implies that

$$\alpha = a \in E_{\mathrm{sep}}.$$

This completes the proof. □

28.3 Solvability by radicals

Let's go back to zeros of polynomials and one of the main motivations of Galois for developing his theory. As it has been mentioning at the beginning of this lecture note, by works of many including Khwarizmi, Khayam, del Ferro, and Ferrari, we know that zeros of polynomials of degree at most 4 can be described in terms of the coefficients of the given polynomial using $+$, $-$, $\cdot$, $/$, and $\sqrt[n]{\cdot}$. Abel showed that there is no general formula $f(a_0, \ldots, a_4)$ in terms of the above operations such that $f(a_0, \ldots, a_4)$ is a zero of

$$x^5 + a_4 x^4 + \cdots + a_0.$$

Using *symmetries* of splitting fields, Galois gave a group theoretic condition which is necessary and sufficient for a polynomial $f \in F[x]$ to be *solvable by radicals* (at least when $\operatorname{char}(F) = 0$ and F has *enough* roots of unity). Our next goal is to carefully formulate Galois's solvability by radicals theorems and prove them.

Definition 28.3.1. *Suppose F is a field and $f \in F[x] \setminus F$.*

1. *We say E/F is a* radical extension *if there is a chain of fields*

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_m = E$$

   such that for every i, $F_{i+1} = F_i[\alpha_i]$ and $\alpha_i^{n_i} \in F_i$ for some positive integer n_i (it is customary to write $\alpha_i = \sqrt[n_i]{a_i}$ for some $a_i \in F_i$).

2. *We say f is* solvable by radicals *if there is a radical extension E/F such that f can be written as a product of degree 1 factors in $E[x]$; alternatively we can say if there is $L \in \operatorname{Int}(E/F)$ which is a splitting field of f over F.*

The building blocks of a radical extension are of the form $K[\sqrt[n]{a}]/K$. We have seen this type of extensions in Example 17.2.2. Kummer studied this type of extensions in connection with Fermat's last conjecture, and he observed the importance of existence of n-th roots of unity in the base field. That is why an extension of the form $K[\sqrt[n]{a}]/K$ where K contains n distinct n-th roots of unity is called a *Kummer extension*.

Proposition 28.3.2. *Suppose F is a field and there is $\zeta \in F$ that has multiplicative order n. Then*

1. *$F[\sqrt[n]{a}]/F$ is Galois, where $\sqrt[n]{a}$ is a zero of $x^n - a$.*

2. *$f_a : \operatorname{Aut}_F(F[\sqrt[n]{a}]) \to \{1, \zeta, \ldots, \zeta^{n-1}\}$, $f_a(\theta) := \frac{\theta(\sqrt[n]{a})}{\sqrt[n]{a}}$ is an injective group homomorphism. In particular, $\operatorname{Aut}_F(F[\sqrt[n]{a}])$ is cyclic.*

Proof. Since $o(\zeta) = n$, $1, \zeta, \ldots, \zeta^{n-1}$ are pairwise distinct, and $(\zeta^i)^n = 1$ for every integer i. Hence $(\zeta^i \sqrt[n]{a})^n = a$ for every integer i and $\sqrt[n]{a}, \zeta \sqrt[n]{a}, \ldots, \zeta^{n-1} \sqrt[n]{a}$ are pairwise distinct. Hence $F[\sqrt[n]{a}]$ is a splitting field of $x^n - a$ over F and $x^n - a$ is separable. Therefore by Theorem 24.2.2, $F[\sqrt[n]{a}]/F$ is a finite Galois extension. By Lemma 22.1.4, every $\theta \in \operatorname{Aut}_F(F[\sqrt[n]{a}])$ permutes zeros of every polynomial in $F[x]$. Hence θ permutes zeros of $x^n - a$, and so

$$\theta(\sqrt[n]{a}) \in \{\sqrt[n]{a}, \zeta \sqrt[n]{a}, \ldots, \zeta^{n-1} \sqrt[n]{a}\}.$$

This shows that f_a is a well-defined function. Next we show that f_a is a group homomorphism. For $\theta_1, \theta_2 \in \operatorname{Aut}_F(F[\sqrt[n]{a}])$, we have

$$\begin{aligned}\theta_1(\theta_2(\sqrt[n]{a})) =& \theta_1(\underbrace{f_a(\theta_2)}_{\in F} \sqrt[n]{a}) \\ =& f_a(\theta_2)\theta_1(\sqrt[n]{a}) \\ =& f_a(\theta_2) f_a(\theta_1) \sqrt[n]{a}.\end{aligned}$$

Hence $f_a(\theta_1 \circ \theta_2) = f_a(\theta_1)f_a(\theta_2)$, which means that f_a is a group homomorphism.

Notice that an element θ of $\mathrm{Aut}_F(F[\sqrt[n]{a}])$ is uniquely determined by $\theta(\sqrt[n]{a})$. If $f_a(\theta) = 1$, then $\theta(\sqrt[n]{a}) = \sqrt[n]{a}$. This implies that $\theta = \mathrm{id}$. Hence f_a is an injective group homomorphism.

As f_a is an injective group homomorphism, the domain of f_a is isomorphic to a subgroup of its codomain. Since the codomain of f_a is a cyclic group and subgroups of cyclic groups are cyclic, we conclude that the domain of f_a is cyclic. This completes the proof. □

Suppose E/F is a radical extension. Hence there is a chain of fields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_m = E$$

such that $F_{i+1} = F_i[\sqrt[n_i]{a_i}]$ for some $a_i \in F_i$ and positive integer n_i. In order to get Kummer extensions as building blocks of this chain, we need to add *enough* roots of unity to the base field. For every i, we need to have n_i distinct n_i-th roots of unity in F_i. Let $n := \mathrm{lcm}(n_1, \ldots, n_m)$. Then for every i, n/n_i is a positive integer, $a'_i := a_i^{n/n_i} \in F_i$. and $F_{i+1} = F_i[\sqrt[n]{a'_i}]$. So without loss of generality we can and will assume that all n_i's are equal to n.

If the base field contains an element ζ of multiplicative order n, then $F_i[\sqrt[n]{a_i}]/F_i$ is a Kummer extension for every index i. By Proposition 18.3.4, $x^n - 1$ has n distinct zeros in its splitting field over F if and only if $\gcd(x^n - 1, nx^{n-1}) = 1$. The latter holds exactly when $\mathrm{char}(F) \nmid n$. Next we show that this condition is enough in order to get an element of multiplicative order n in a splitting field of $x^n - 1$ over F.

Lemma 28.3.3. *Suppose F is a field, n is a positive integer, and $\mathrm{char}(F) \nmid n$. Let E be a splitting field of $x^n - 1$ over F. Then there is $\zeta \in E$ such that the multiplicative order of ζ is n and $E = F[\zeta]$.*

Proof. Consider the function $E^\times \to E^\times,\ \alpha \mapsto \alpha^n$. Since E is commutative, this is a group homomorphism. Let A_n be the kernel of this group homomorphism. Then C_n is a subgroup of $E^\times$ and

$$C_n := \{\alpha \in E^\times \mid \alpha^n - 1 = 0\}.$$

Notice that since $\mathrm{char}(F) \nmid n$, $\gcd(x^n - 1, nx^{n-1}) = 1$. Therefore by Proposition 18.3.4 $x^n - 1$ has distinct zeros in E. Hence $|C_n| = n$. We also notice that for every positive integer d, there are at most d elements α of the group C_n such that $\alpha^d = 1$. Therefore by a result from group theory, we conclude that C_n is a cyclic group. Let ζ be a generator of C_n. As $|C_n| = n$, the multiplicative order of ζ is n. Hence $1, \zeta, \ldots, \zeta^{n-1}$ are distinct zeros of $x^n - 1$. Therefore by the generalized factor theorem, comparing degrees, and leading coefficients, we obtain that

$$x^n - 1 = (x - 1)(x - \zeta)\cdots(x - \zeta^{n-1}).$$

This implies that $E = F[\zeta, \ldots, \zeta^{n-1}] = F[\zeta]$, which completes the proof. □

Starting with a radical extension E/F (with certain assumption on $\mathrm{char}(F)$), we can use Lemma 28.3.3 and find another radical extension E'/F such that $E \subseteq E'$ and all the *building blocks* of E'/F are Kummer extensions except possibly the first one!

Since E/F is a radical extension, there is a positive integer n and a chain of fields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_m = E$$

such that $F_{i+1} = F_i[\sqrt[n]{a_i}]$ for some $a_i \in F_i$. Let E' be a splitting field of $x^n - 1$ over E. Then by Lemma 28.3.3, there is $\zeta \in E'$ that has multiplicative order n. Let $E_i := F_i[\zeta]$. Then we have

$$F \subseteq E_0 \subseteq E_1 \subseteq \cdots \subseteq E_m =: E'$$

and $E_0 = F[\zeta]$ and $E_{i+1} = E_i[\sqrt[n]{a_i}]$ for every index i. Therefore for every i, E_{i+1}/E_i is a Kummer extension. It is remained to get a better understanding of the first block $F[\zeta]/F$.

Proposition 28.3.4. *Suppose F is a field, n is a positive integer, and $\mathrm{char}(F) \nmid n$. Let E be a splitting field of $x^n - 1$ over F. Then*

1. *$C_n := \{\alpha \in E \mid \alpha^n = 1\}$ is a cyclic group of order n.*
2. *the restriction map $r : \mathrm{Aut}_F(E) \to \mathrm{Aut}(C_n)$, $r(\theta) := \theta|_{C_n}$ is an injective group homomorphism. In particular, $\mathrm{Aut}_F(E)$ can be embedded into $\mathbb{Z}_n^\times$ and is abelian.*

Proof. The first part follows from Lemma 28.3.3. Every $\theta \in \mathrm{Aut}_F(E)$ permutes zeros of $x^n - 1$. Hence $\theta(C_n) = C_n$ for every $\theta \in \mathrm{Aut}_F(E)$. Therefore $\theta|_{C_n}$ is a group automorphism of C_n. This implies that r is a well-defined function. As the restriction of composite of two automorphisms is the same as the composite of restrictions of these automorphisms, we have that r is a group homomorphism. Next we notice that since E is generated by F and elements of C_n, an automorphism $\theta \in \mathrm{Aut}_F(E)$ is uniquely determined by its restriction on C_n. This means r is injective.

Finally from group theory we know that the group of automorphisms of a cyclic group of order n is isomorphic to $\mathbb{Z}_n^\times$; this follows from the fact that $o(g^m) = o(g)$ if and only if $\gcd(o(g), m) = 1$. □

We would like to relate our information on the symmetries of the building blocks of a radical extension to symmetries of the radical extension itself. As we have learn so far, the group of symmetries of an extension is *richest* when the extension is Galois. So Next we will replace E by an extension field E' of E such that E'/F is both radical and Galois. Then using the above chain of fields, we will prove an interesting property of $\mathrm{Aut}_F(E')$.

Chapter 29

Lecture 5

Let's recall that a field extension E/F is called a *radical extension* if there is a chain of subfields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_m = E$$

such that $F_{i+1} = F_i[\alpha_i]$ and $\alpha_i^{n_i} = a_i$ for some $n_i \in \mathbb{Z}$ and $a_i \in F_i$. We say $f \in F[x]$ is *solvable by radicals* over F if there is a radical extension E/F such that

$$f(x) = \mathrm{ld}(f)(x - \alpha_1) \cdots (x - \alpha_d)$$

for some $\alpha_1, \ldots, \alpha_d \in E$; alternatively we can say that there is $L \in \mathrm{Int}(E/F)$ which is a splitting field of f over F. Next we investigate radical extensions further.

29.1 Radical extensions

The main goal of this section is to prove the following theorem.

Theorem 29.1.1. *Suppose E/F is a radical extension. Then there is L/E such that L/F is a Galois radical extension.*

To prove this theorem we start by proving a few lemmas. In the first one, we show that radical extensions satisfy a block-tower phenomenon.

Lemma 29.1.2. *If L/E and E/F are radical extensions, then L/F is a radical extension.*

Proof. Since L/E is a radical extension, there is a chain of subfields

$$E =: L_0 \subseteq L_1 \subseteq \cdots \subseteq L_m := L$$

such that

$$L_{i+1} = L_i[\alpha_i] \text{ and } \alpha^{n_i} = a_i \in L_i \tag{29.1}$$

for $0 \leq i < m$. Since E/F is a radical extension, there is a chain of subfields

$$F =: E_0 \subseteq E_1 \subseteq \cdots \subseteq E_r := E$$

such that

$$E_{i+1} = E_i[\beta_i] \text{ and } \beta_i^{m_i} = b_i \in E_i \tag{29.2}$$

for $0 \le i < r$. Considering the chain of fields

$$F = E_0 \subseteq E_1 \subseteq \cdots \subseteq E_r = E = L_0 \subseteq L_1 \subseteq \cdots \subseteq L_m = L$$

by (29.1) and (29.2), we conclude that L/F is a radical extension. □

An important example of a radical extension which is also Galois is the following.

Lemma 29.1.3. *Suppose $a_1, \ldots, a_m \in F$, E is a splitting field of*

$$(x^n - a_1) \cdots (x^n - a_m)$$

over F, and $\operatorname{char}(F) = 0$. Then E/F is a Galois radical extension.

Proof. Since E is a splitting field of a polynomial in $F[x]$ over F, E/F is a normal extension (see Theorem 22.2.1). As $\operatorname{char}(F) = 0$, E/F is separable (see Theorem 27.3.5). Hence E/F is a Galois extension.

Since E is a splitting field of $(x^n - a_1) \cdots (x^n - a_m)$ over F, there are $\alpha_{i1}, \ldots, \alpha_{in}$ in E such that

$$x^n - a_i = (x - \alpha_{i1}) \cdots (x - \alpha_{in})$$

for every $1 \le i \le m$, and $E = F[\alpha_{11}, \ldots, \alpha_{mn}]$. Consider the following chain of fields

$$F \subseteq F[\alpha_{11}] \subseteq F[\alpha_{11}, \alpha_{12}] \subseteq \cdots \subseteq F[\alpha_{11}, \ldots, \alpha_{nm}] = E.$$

Notice that at every step we adding a zero of $x^n - a_i$ for some i. Hence E/F is a radical extension. □

The next lemma is of independent interest. We have seen instances of this lemma in other proofs.

Lemma 29.1.4. *Suppose E/F is a finite normal extension and $f \in F[x]$. Let L be a splitting field of f over E. Then L/F is a normal extension.*

Proof. Since L is a splitting field of f over E, there are $\alpha_1, \ldots, \alpha_d \in L$ such that

$$f(x) = \operatorname{ld}(f)(x - \alpha_1) \cdots (x - \alpha_d)$$

and $L = E[\alpha_1, \ldots, \alpha_d]$.

Since E/F is a finite normal extension, E is a splitting field of some monic $g \in F[x]$ over F (see Theorem 22.2.1). This means there are $\beta_1, \cdots, \beta_m \in E$ such that

$$g(x) = (x - \beta_1) \cdots (x - \beta_m)$$

and $E = F[\beta_1, \ldots, \beta_m]$. Hence

$$f(x)g(x) = \operatorname{ld}(f)(x - \alpha_1) \cdots (x - \alpha_d)(x - \beta_1) \cdots (x - \beta_m)$$

and $L = F[\beta_1, \ldots, \beta_m, \alpha_1, \ldots, \alpha_d]$. Therefore L is a splitting field of $f(x)g(x) \in F[x]$ over F. We conclude that L/F is a finite normal extension (see Theorem 22.2.1). □

Now we have all the needed tools to prove Theorem 29.1.1.

Proof of Theorem 29.1.1. We proceed by strong induction on $[E:F]$. If $[E:F]=1$, then $E=F$ and $L:=E$ satisfies the claim. Since E/F is a radical extension, there is a chain of fields

$$F=F_0 \subsetneq F_1 \subsetneq \cdots \subsetneq F_{m+1}=E$$

such that $F_{i+1}=F_i[\alpha_i]$ and $\alpha_i^{n_i}=a_i \in F_i$ for every i. Therefore F_m/F is a radical extension and $[F_m:F]<[E:F]$. Hence bu the induction hypothesis, there is a field extension K/F_m such that K/F is a Galois radical extension.

Next we would like to find a field extension L/K such that

1. L contains a copy of $E=F_m[\alpha_m]$, and
2. L/F is a Galois and radical extension.

Notice that $\alpha_m^{n_m}=a_m \in F_m \subseteq K$. So $x^{n_m}-a_m$ should have a zero in L. Since L/F and K/F are supposed to be Galois extensions, for every $\theta \in \mathrm{Aut}_F(K)$ there is $\widehat{\theta} \in \mathrm{Aut}_F(L)$ such that $\widehat{\theta}|_E=\theta$. Hence if $\alpha \in L$ is a zero of $x^{n_m}-a_m$, then $\widehat{\theta}(\alpha)$ is a zero of $x^{n_m}-\widehat{\theta}(a_m)=x^{n_m}-\theta(a_m)$. This suggests that we need to consider

$$g(x):=\prod_{\theta\in \mathrm{Aut}_F(K)}(x^{n_m}-\theta(a_m)).$$

Claim. *$g(x)$ is in $F[x]$.*

Proof of Claim. Every $\theta' \in \mathrm{Aut}_F(K)$ permutes factors of g, and so $\theta'(g)=g$ for every $\theta' \in \mathrm{Aut}_F(K)$. This means all the coefficients of g are fixed by $\mathrm{Aut}_F(K)$. Since K/F is a finite Galois extension, by the Fundamental Theorem of Galois Theory $\mathrm{Fix}(\mathrm{Aut}_F(K))=F$. The claim follows. □

Let L be a splitting field of g over K. As $\theta(a_m)$ is in K for every $\theta \in \mathrm{Aut}_F(K)$, by Lemma 29.1.3 L/K is a radical extension. Since K/F is a radical extension, by Lemma 29.1.2 we obtain that L/F is a radical extension. Since $g(x)\in F[x]$ (by the above Claim), K/F is Galois, and L is a splitting field of g over K, by Lemma 29.1.4 L/F is a Galois extension.

Finally we need to show that $E=F_m[\alpha_m]$ can be embedded into L. Since $\alpha_m^{n_m}-a_m=0$, m_{α_m,F_m} divides $x^{n_m}-a_m$. As $x^{n_m}-a_m$ divides $g(x)$ and $g(x)$ decomposes into linear factors over L, there is $\alpha\in L$ which is a zero of m_{α_m,F_m}. Since m_{α_m,F_m} is irreducible in $F_m[x]$, by Corollary 17.0.1 there is an F_m-isomorphism $\theta: F_m[\alpha_m]\to F_m[\alpha']$. Using the facts that $E=F_m[\alpha_m]$ and $F_m[\alpha']$ is a subfield of L, we obtain an F-embedding of E into L. This completes the proof. □

29.2 Solvable by radicals

Suppose F is a field of characteristic 0 and $f\in F[x]$ is solvable by radicals. Then there is a radical extension E/F and $K\in \mathrm{Int}(E/F)$ such that K is a splitting field f

over F. By Theorem 29.1.1, there is L/E such that L/F is Galois and radical. Hence there is a chain

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_{m+1} = L$$

such that $F_{i+1} = F_i[\alpha_i]$ and $\alpha_i^{n_i} = a_i \in F_i$. To make each block of the above tower of fields into a Kummer extension, we need to add enough roots of unity to the base field. Let $n := \mathrm{lcm}(n_1, \ldots, n_m)$ and let $\widehat{L}$ be a splitting field of $x^n - 1$ over L. Since $x^n - 1 \in F[x]$ and L/F is a finite Galois extension, by Lemma 29.1.4 we conclude that $\widehat{L}/F$ is a finite Galois extension. By Lemma 29.1.3, $\widehat{L}/L$ is a radical extension. As L/F is a radical extension as well, by Lemma 29.1.2 $\widehat{L}/F$ is a radical extension.

Since $\mathrm{char}(F) = 0$, by Proposition 28.3.4 there is $\zeta \in \widehat{L}$ such that $o(\zeta) = n$,

$$x^n - 1 = (x-1)(x-\zeta)\cdots(x-\zeta^{n-1}),$$

and $\widehat{L} = L[\zeta]$.

Let $E_i := F_i[\zeta]$. Then

$$F \subseteq E_0 \subseteq E_1 \subseteq \cdots \subseteq E_{m+1} = \widehat{L}, \tag{29.3}$$

$E_{i+1} = E_i[\alpha_i]$, $\alpha_i^{n_i} = a_i \in E_i$, and $E_0 = F[\zeta]$. Hence E_{i+1}/E_i is a Kummer extension; in particular, by Proposition 28.3.2, $\mathrm{Aut}_{E_i}(E_{i+1})$ is cyclic. By Proposition 28.3.4, E_0/F is a Galois extension and $\mathrm{Aut}_F(E_0)$ is abelian.

Applying Galois's bijection between intermediate subfields of $\widehat{L}/F$ and subgroups of $\mathrm{Aut}_F(\widehat{L})$ to the tower fields given in (29.3) we obtain the following chain of subgroups:

$$\mathrm{Aut}_F(\widehat{L}) \geq \mathrm{Aut}_{E_0}(\widehat{L}) \geq \cdots \geq \mathrm{Aut}_{E_{m+1}}(\widehat{L}) = 1. \tag{29.4}$$

For every i, E_{i+1}/E_i is a Galois extension. Hence by the Fundamental Theorem of Galois Theory, $\mathrm{Aut}_{E_{i+1}}(\widehat{L})$ is a normal subgroup of $\mathrm{Aut}_{E_i}(\widehat{L})$, and by Proposition 23.1.1 we obtain

$$\mathrm{Aut}_{E_i}(\widehat{L})/\mathrm{Aut}_{E_{i+1}}(\widehat{L}) \simeq \mathrm{Aut}_{E_i}(E_{i+1}). \tag{29.5}$$

Since E_{i+1}/E_i is a Kummer extension, $\mathrm{Aut}_{E_i}(E_{i+1})$ is cyclic. Notice that because E_0/F is a Galois extension and $\mathrm{Aut}_F(E_0)$ is abelian, we have that $\mathrm{Aut}_{E_0}(\widehat{L})$ is a normal subgroup of $\mathrm{Aut}_F(\widehat{L})$, and

$$\mathrm{Aut}_F(\widehat{L})/\mathrm{Aut}_{E_0}(\widehat{L}) \simeq \mathrm{Aut}_F(E_0) \tag{29.6}$$

is abelian. Let $G := \mathrm{Aut}_F(\widehat{L})$ and $G_i := \mathrm{Aut}_{E_i}(\widehat{L})$. Then by (29.4), (29.5), and (29.6), we obtain the following chain of subgroups

$$G \trianglerighteq G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_m \trianglerighteq G_{m+1} = 1, \tag{29.7}$$

and $G/G_0, G_0/G_1, \ldots, G_m/G_{m+1}$ are abelian.

All these fields and groups are not intrinsic of the polynomial f. These are all auxiliary tools to understand K/F where $K \in \mathrm{Int}(\widehat{L}/F)$ is a splitting field of f over F. Notice that $\mathrm{Aut}_K(\widehat{L})$ is a normal subgroup of $\mathrm{Aut}_F(\widehat{L})$ and

$$\mathrm{Aut}_F(\widehat{L})/\mathrm{Aut}_K(\widehat{L}) \simeq \mathrm{Aut}_F(K). \tag{29.8}$$

Let $N := \mathrm{Aut}_K(\widehat{L})$; so by (29.8) $N \trianglelefteq G$ and $\mathrm{Aut}_F(K) \simeq G/N$. Therefore by (29.4), we obtain the following chain of subgroups of $\mathrm{Aut}_F(K)$

$$\mathrm{Aut}_F(K) \simeq G/N \trianglerighteq p_N(G_0) \trianglerighteq p_N(G_1) \trianglerighteq \cdots \trianglerighteq p_N(G_m) \trianglerighteq p_N(G_{m+1}) = 1,$$

where $p_N : G \to G/N, p_N(g) := gN$ is the natural quotient map.

Lemma 29.2.1. *In the above setting, $p_N(G_i)/p_N(G_{i+1})$ is an abelian group.*

Proof. For every $g, g' \in G_i$, we have to show that

$$(p_N(g)p_N(G_{i+1}))(p_N(g')p_N(G_{i+1})) = (p_N(g')p_N(G_{i+1}))(p_N(g)p_N(G_{i+1})).$$

This is equivalent to

$$p_N(g')^{-1}p_N(g)^{-1}p_N(g')p_N(g) \in p_N(G_{i+1}). \tag{29.9}$$

Since p_N is a group homomorphism, (29.9) holds exactly when $p_N(g'^{-1}g^{-1}g'g) \in p_N(G_{i+1})$. So it suffices to show that $g'^{-1}g^{-1}g'g \in G_{i+1}$. Notice that G_i/G_{i+1} is abelian, and so $gg'G_{i+1} = g'gG_{i+1}$, which implies that $g'^{-1}g^{-1}g'g \in G_{i+1}$. The claim follows. □

This brings us to the definition of solvable groups.

Definition 29.2.2. *We say a group G is* solvable *if there is a chain of subgroups*

$$G := N_0 \trianglerighteq N_1 \trianglerighteq \cdots \trianglerighteq N_m = 1$$

such that N_i/N_{i+1} is abelian for every index i.

Altogether we have proved the following theorem of Galois.

Theorem 29.2.3 (Galois)**.** *Suppose F is a field of characteristic zero, and $f(x)$ in $F[x] \setminus F$. Let K be a splitting field of f over F. If f is solvable by radicals over F, then $\mathrm{Aut}_F(K)$ is solvable.*

From group theory we know that A_n is a non-abelian simple group if $n \geq 5$. Using this we show that A_n and S_n are not solvable if $n \geq 5$. Hence if $\mathrm{Aut}_F(K)$ is isomorphic to either A_n or S_n with $n \geq 5$ where K is a splitting field of f over F, then f is not solvable by radicals over F.

Chapter 30

Lecture 6

We proved Galois's theorem which asserts if F is a field of characteristic zero and $f \in F[x]$ is solvable by radicals over F, then $\mathrm{Aut}_F(K)$ is solvable where K is a splitting field of f over F. Here we go over some of the basic properties of solvable groups and use them to show that there are degree 5 polynomials in $\mathbb{Q}[x]$ that are not solvable by radicals over $\mathbb{Q}$.

30.1 Basics of solvable groups

In this section, we review some of the basic properties of solvable groups.

Proposition 30.1.1. *Suppose G is solvable. Then the following holds.*

1. *If $N \trianglelefteq G$, then G/N is solvable.*
2. *If $H \leq G$, then H is solvable.*
3. *Suppose $N \trianglelefteq G$. If N and G/N are solvable, then G is solvable.*

Proof. Since G is solvable, there is a chain of subgroups

$$G =: G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_m = 1$$

such that G_i/G_{i+1} is abelian for every integer $0 \leq i < m$.

(1) Let $p_N : G \to G/N, p_N(g) := gN$. Then by the virtue of the proof of Lemma 29.2.1, we have $p_N(G_i)/p_N(G_{i+1})$ is abelian, and

$$G/N =: p_N(G_0) \trianglerighteq p_N(G_1) \trianglerighteq \cdots \trianglerighteq p_N(G_m) = 1.$$

Hence G/N is solvable.

(2) For every i, consider $p : H \cap G_i \to G_i/G_{i+1}, p(h) := hG_{i+1}$. Then $h \in H \cap G_i$ is in the kernel of p if and only if $hG_{i+1} = G_{i+1}$. This holds exactly when h is in G_{i+1}. Hence $\ker p = (H \cap G_i) \cap G_{i+1} = H \cap G_{i+1}$. Therefore $H \cap G_{i+1}$ is a normal subgroup of $H \cap G_i$ and by the first isomorphism theorem for groups, $(H \cap G_i)/(H \cap G_{i+1})$ can

be embedded into G_i/G_{i+1}. In particular, $(H \cap G_i)/(H \cap G_{i+1})$ is abelian. Therefore considering the following chain of subgroups

$$H = (H \cap G_0) \trianglerighteq (H \cap G_1) \trianglerighteq \cdots \trianglerighteq (H \cap G_m) = 1$$

we deduce that H is solvable.

(3) Let's recall the following result form group theory:

Proposition 30.1.2. *Suppose G is a group and N is a normal subgroup of G. The the following is a bijection:*

$$\{H \mid N \leq H \leq G\} \to \mathrm{Sub}(G/N), \quad H \mapsto H/N.$$

Moreover $H_1/N \leq H_2/N$ if and only if $N \leq H_1 \leq H_2$, and if $N \leq H_1 \trianglelefteq H_2$, then $H_1/N \trianglelefteq H_2/N$ and $\frac{H_2/N}{H_1/N} \simeq \frac{H_1}{H_2}$.

Since G/N is solvable, by Proposition 30.1.2 there are $N \leq G_i \leq G$ such that

$$G/N = G_0/N \trianglerighteq G_1/N \trianglerighteq \cdots \trianglerighteq G_m/N = 1, \tag{30.1}$$

and $\frac{G_i/N}{G_{i+1}/N}$ is abelian for every i. Hence by Proposition 30.1.2, we have

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_m = N, \tag{30.2}$$

and $\frac{G_i}{G_{i+1}} \simeq \frac{G_i/N}{G_{i+1}/N}$ is abelian.

Since N is solvable, there is a chain of subgroups

$$N = N_0 \trianglerighteq N_1 \trianglerighteq \cdots \trianglerighteq N_k = 1 \tag{30.3}$$

such that N_i/N_{i+1} is abelian. Notice that the chain of subgroups given in (30.2) end where the chain of subgroups in (30.3) start. Hence we get the following chain of subgroups

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_m = N = N_0 \trianglerighteq N_1 \trianglerighteq \cdots \trianglerighteq N_k = 1,$$

and notice that the quotient of each subgroup by the next one is abelian. Hence G is solvable. This completes the proof. □

Next we show that a solvable simple group is a cyclic group of prime order.

Lemma 30.1.3. *If G is a solvable simple group, then G is a cyclic group of prime order.*

Proof. Since G is solvable, it has a proper normal subgroup N such that G/N is abelian. As G is simple, it does not have a non-trivial normal subgroup. Hence $N = 1$, which means G is abelian. Therefore every subgroup of G is normal. As G is simple, it does not have a normal subgroup. We conclude that G has exactly two subgroups $\{1\}$ and G; that means

$$\mathrm{Sub}(G) = \{\{1\}, G\}. \tag{30.4}$$

By (30.4), we deduce that $\langle g\rangle = G$ for every $g \in G \setminus \{1\}$. If there is $g \in G$ that has infinite order, then $\langle g^2\rangle \neq \langle g\rangle$. This contradicts (30.4). So every element g of G has finite order. Suppose p is a prime factor of $o(g)$ where $g \in G \setminus \{1\}$. By Cauchy's theorem, G has an element g_0 of order p. By (30.4), $G = \langle g_0\rangle$. This completes the proof. □

Corollary 30.1.4. *If $n \geq 5$, A_n and S_n are not solvable.*

Proof. If $n \geq 5$, A_n is a non-abelian simple group. Hence by Lemma 30.1.3, A_n is not solvable. Since A_n is a subgroup of S_n and A_n is not solvable, by Proposition 30.1.1 S_n is not solvable. This completes the proof. □

30.2 Galois groups and permutations

When E/F is a Galois extension, the group $\mathrm{Aut}_F(E)$ is called the *Galois group* of E/F, and it is denoted by $\mathrm{Gal}(E/F)$. For a polynomial $f \in F[x]$, the group of F-automorphism of a splitting field K of f over F is denoted by $\mathscr{G}_{f,F}$. The next proposition helps us get a concrete understanding of $\mathscr{G}_{f,F}$ as a subgroup of permutations of zeros of f in K. When $f \in F[x]$ is separable, K/F is a Galois extension and $\mathscr{G}_{f,F}$ is called the *Galois group* of f over F.

Proposition 30.2.1. *Suppose F is a field, $f \in F[x] \setminus F$, and K is a splitting field of f over F. Suppose*

$$X := \{\alpha_1, \ldots, \alpha_n\} \subseteq K$$

is such that $f(x) = \mathrm{ld}(f)\prod_{i=1}^n (x - \alpha_i)$. Let S_X be the symmetric group of X. Then

1. *The restriction $r : \mathrm{Aut}_F(K) \to S_X,\ r(\theta) := \theta|_X$ is a well-defined injective group homomorphism.*

2. *If K' is another splitting field of f over F, then there is an isomorphism*

$$c : \mathrm{Aut}_F(K) \to \mathrm{Aut}_F(K').$$

Proof. For every $\theta \in \mathrm{Aut}_F(K)$, $\theta(\alpha_i)$ is a zero of $\theta(f)$ (see Lemma 22.1.4). Since $f \in F[x]$, $\theta(\alpha_i)$ is a zero of f. As X is the set of all the zeros of f in K, we conclude that $\theta(X) \subseteq X$ for every $\theta \in \mathrm{Aut}_F(K)$. As X is a finite set and θ is injective, we deduce that $\theta|_X : X \to X$ is a bijection. Hence r is a well-defined function. Because the restriction of composite of two bijections is the composite of restrictions of those functions, we have that r is a group homomorphism. Since $K = F[\alpha_1, \ldots, \alpha_n]$, every $\theta \in \mathrm{Aut}_F(K)$ is uniquely determined by its values at $\alpha_1, \ldots, \alpha_n$. Hence $\theta|_X$ uniquely determines θ, which means that r is injective.

(2) Since K and K' are splitting fields of f over F, by Theorem 17.1.2 there is an F-isomorphism $\sigma : K \to K'$. Let

$$c : \mathrm{Aut}_F(K) \to \mathrm{Aut}_F(K'), \quad c(\theta) := \sigma \circ \theta \circ \sigma^{-1}.$$

Notice that composite of F-isomorphisms is an F-isomorphism. Hence $c(\theta)$ is indeed an element of $\mathrm{Aut}_F(K')$. For every $\theta_1, \theta_2 \in \mathrm{Aut}_F(K)$, we have

$$c(\theta_1 \circ \theta_2) = \sigma \circ (\theta_1 \circ \theta_2) \circ \sigma^{-1} = (\sigma \circ \theta_1 \circ \sigma^{-1}) \circ (\sigma \circ \theta_2 \circ \sigma^{-1}) = c(\theta_1) \circ c(\theta_2).$$

This means that c is a group homomorphism. It is easy to see that

$$c' : \mathrm{Aut}_F(K') \to \mathrm{Aut}_F(K), \quad c'(\theta') := \sigma^{-1} \circ \theta' \circ \sigma$$

is the inverse of c, and so c is a bijection. This completes the proof. □

We let $\mathscr{G}_{f,F}$ be the image of r. By Proposition 30.2.1, $\mathscr{G}_{f,F}$ is a subgroup of the symmetric group of $S_{\deg f}$ and it only depends on (f, F) up to a group isomorphism. By Galois's theorem (Theorem 29.2.3), for a field of characteristic zero and $f \in F[x]$, we have that

f is solvable by radicals over F $\Rightarrow$ $\mathscr{G}_{f,F}$ is solvable.

30.3 Examples of polynomials that are not solvable by radicals

In this section, we use Corollary 30.1.4 and Galois's theorem, to find polynomials $f \in \mathbb{Q}[x]$ that are not solvable by radicals over $\mathbb{Q}$. This is done based on the following lemma and a result from group theory.

Lemma 30.3.1. *Suppose p is prime, $f \in \mathbb{Q}[x]$ is irreducible of degree p, and f has 2 non-real complex zeros and $p-2$ real zeros. Then we can label zeros of f in a way that $\mathscr{G}_{f,F}$ contains $(1, 2, \ldots, p)$ and $(1, a)$ for some integer $a \in [2, p]$.*

Proof. Suppose $f(x) = \mathrm{ld}(f)(x - \alpha_1) \cdots (x - \alpha_p)$ for α_i's in $\mathbb{C}$. We further assumed that $\alpha_1 \in \mathbb{C} \setminus \mathbb{R}$. Let

$$K := \mathbb{Q}[\alpha_1, \ldots, \alpha_p].$$

Then K is a splitting field of f over $\mathbb{Q}$. Let's recall that $\mathscr{G}_{f,\mathbb{Q}}$ is the subgroup of permutations of $\{\alpha_1, \ldots, \alpha_p\}$ that are induced by elements of $\mathrm{Aut}_{\mathbb{Q}}(K)$. We will rearrange the zeros and identify the group of permutations of $\{\alpha_1, \ldots, \alpha_p\}$ with S_p. Notice that since $\mathrm{char}(\mathbb{Q}) = 0$, f is separable. Hence K is a splitting field of a separable polynomial over $\mathbb{Q}$, which implies that $K/\mathbb{Q}$ is a Galois extension (see Theorem 24.2.2). Therefore

$$|\mathscr{G}_{f,\mathbb{Q}}| = [K : \mathbb{Q}]. \tag{30.5}$$

Since f is irreducible in $\mathbb{Q}[x]$ and α_1 is a zero of f, $f(x) = \mathrm{ld}(f) m_{\alpha,F}(x)$ (see Proposition 8.2.6). Hence

$$\deg m_{\alpha_1,\mathbb{Q}} = p. \tag{30.6}$$

By the Tower Rule, we obtain that $[\mathbb{Q}[\alpha_1] : \mathbb{Q}]$ divides $[K : \mathbb{Q}]$. Therefore by (30.5), (30.6), and Proposition 20.1.2, we conclude that p divides $|\mathscr{G}_{f,\mathbb{Q}}|$. By Cauchy's theorem, there is $\sigma \in \mathscr{G}_{f,\mathbb{Q}}$ that has order p. Let's recall from group theory that if the cycle type of a permutation $\sigma \in S_p$ is $(n_1, \ldots, n_k)$, then the order of σ is $\mathrm{lcm}(n_1, \ldots, n_k)$ and

$n_1 + \ldots + n_k = p$. Hence $o(\sigma) = p$ for $\sigma \in S_p$ implies that σ is a cycle of length p. So σ is the cycle

$$(\alpha_1, \sigma(\alpha_1), \ldots, \sigma^{p-1}(\alpha_1)).$$

We relabel α_i's and let

$$\alpha_2 := \sigma(\alpha_1),\ \alpha_3 := \sigma(\alpha_2), \ldots,\ \alpha_p := \sigma(\alpha_{p-1}).$$

So σ is represented by $(1, 2, \ldots, p)$ in S_p.

Let $\tau : \mathbb{C} \to \mathbb{C}$ be the complex conjugation. Then $\tau \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{C})$. Since $K/\mathbb{Q}$ is a normal extension, by Proposition 23.1.1 $\tau|_K \in \mathrm{Aut}_{\mathbb{Q}}(K)$. Notice that $\mathrm{Fix}(\tau) = \mathbb{R}$. Hence $p - 2$ roots are fixed by τ and the other two non-real zeros are swapped by τ. We have already assumed that α_1 is one of the non-real zeros. Suppose α_a is the other non-real zero of f. Hence with respect to this labeling of α_i's, $\tau|_K$ is represented by $(1, a)$. This completes the proof. □

Exercise 30.3.2. *Suppose p is prime. Then $\langle (1, 2, \ldots, p), (1, a) \rangle = S_p$ for every integer a in $[2, p]$.*

By Exercise 30.3.2 and Lemma 30.3.1, we conclude the following.

Theorem 30.3.3. *If $f \in \mathbb{Q}[x]$ is irreducible, $\deg f = p$ is prime, f has 2 non-real complex zeros and $p - 2$ real zeros, then $\mathscr{G}_{f,\mathbb{Q}} \simeq S_p$. In particular, f is not solvable by radicals over $\mathbb{Q}$ if $p \geq 5$.*

Proof. The first part is an immediate consequence of Lemma 30.3.1 and Exercise 30.3.2. The second part can be deduced from the first part, Corollary 30.1.4, and Galois's solvability theorem (see Theorem 29.2.3). □

Exercise 30.3.4. *Show that $x^5 - 16x + 2$ is not solvable by radicals over $\mathbb{Q}$.*

Existence of a degree 5 polynomial which is not solvable by radicals implies that *there is no general formula in terms of $+, -, \cdot, /, \sqrt[n]{\cdot}$ to solve degree 5 equations.*

30.4 Finite solvable groups and prime order factors

Next we want to prove the converse of Galois's solvability theorem (see Theorem 29.2.3).

Theorem 30.4.1 (Galois)**.** *Suppose F is a field of characteristic zero and $f \in F[x]$. Then f is solvable by radicals over F if and only if $\mathscr{G}_{f,F}$ is solvable.*

Theorem 29.2.3 implies ($\Rightarrow$). So we focus on ($\Leftarrow$). We start with the following group theoretic lemma.

Lemma 30.4.2. *Suppose G is a finite solvable group. Then there is a chain of subgroups*

$$G =: G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_m = 1$$

such that G_i/G_{i+1}'s are cyclic groups of prime order.

Proof. Since G is solvable, there is a chain of a subgroups

$$G =: \widehat{G}_0 \trianglerighteq \widehat{G}_1 \trianglerighteq \cdots \trianglerighteq \widehat{G}_{d+1} = 1$$

such that $\widehat{G}_i/\widehat{G}_{i+1}$ is abelian. Since G is finite $\widehat{G}_i/\widehat{G}_{i+1}$ is a finite abelian group.

Claim. *If A is a finite abelian group, then there is a chain of subgroups*

$$A =: A_0 \trianglerighteq A_1 \trianglerighteq \cdots \trianglerighteq A_n = 1$$

such that A_i/A_{i+1} is cyclic group of prime order for every i.

Proof of Claim. We proceed by strong induction on $|A|$. If $|A| = 1$, there is nothing to prove. If $|A| > 1$, then $|A|$ has a prime factor p. By Cauchy's theorem, A has an element of order p. Since A is abelian, $\langle a\rangle$ is a normal subgroup and $A/\langle a\rangle$ is abelian. As $|A/\langle a\rangle| < |A|$, by the strong induction hypothesis, there is a chain of subgroups

$$\frac{A}{\langle a\rangle} =: \overline{A}_0 \trianglerighteq \overline{A}_1 \trianglerighteq \cdots \trianglerighteq \overline{A}_r = 1 \qquad (30.7)$$

such that $\overline{A}_i/\overline{A}_{i+1}$ is cyclic of prime order. By Proposition 30.1.2, for every i, $\overline{A}_i = \frac{A_i}{\langle a\rangle}$ for some subgroup A_i of A which contains $\langle a\rangle$. Hence by (30.7), we conclude that

$$A = A_0 \trianglerighteq A_1 \trianglerighteq \cdots \trianglerighteq A_r = \langle a\rangle. \qquad (30.8)$$

Notice that

$$\frac{\overline{A}_i}{\overline{A}_{i+1}} = \frac{A_i/\langle a\rangle}{A_{i+1}/\langle a\rangle} \simeq \frac{A_i}{A_{i+1}}.$$

Therefore A_i/A_{i+1} is a cyclic group of prime order for every i. Let $A_{r+1} = 1$ and notice that $A_r/A_{r+1} \simeq \langle a\rangle$ is a cyclic group of a prime order. Hence the chain of subgroups

$$A = A_0 \trianglerighteq A_1 \trianglerighteq \cdots \trianglerighteq A_{r+1} = 1$$

satisfies the desired property. This completes the proof. □

By the above claim, for every i, there is a chain of subgroups

$$\frac{\widehat{G}_i}{\widehat{G}_{i+1}} = A_{i1} \trianglerighteq A_{i2} \trianglerighteq \cdots \trianglerighteq A_{im_i} = 1$$

such that $\frac{A_{ij}}{A_{i(j+1)}}$ is a cyclic group of prime order for every j. By Proposition 30.1.2, there are subgroups G_{ij} of $\widehat{G}_i$ which contain $\widehat{G}_{i+1}$ and $A_{ij} = G_{ij}/\widehat{G}_{i+1}$. Hence

$$\frac{A_{ij}}{A_{i(j+1)}} = \frac{G_{ij}/\widehat{G}_{i+1}}{G_{i(j+1)}/\widehat{G}_{i+1}} \simeq \frac{G_{ij}}{G_{i(j+1)}}$$

are cyclic groups of prime order. Notice that $A_{im_i} = 1$ implies that $G_{im_i} = \widehat{G}_{i+1}$ for every i. Consider the following chain of subgroups

$$\begin{aligned} G = \widehat{G}_0 =& G_{01} \trianglerighteq \cdots \trianglerighteq G_{0m_0} = \\ \widehat{G}_1 =& G_{11} \trianglerighteq \cdots \trianglerighteq G_{1m_1} = \\ & \cdots \\ \widehat{G}_d =& G_{d1} \trianglerighteq \cdots \trianglerighteq G_{dm_d} = 1, \end{aligned}$$

and the claim follows. □

Next we will focus on finite Galois extensions E/F where $\mathrm{Gal}(E/F)$ is cyclic.

Chapter 31

Lecture 7

The main goal of this section is to prove Galois's solvability theorem (see Theorem 30.4.1). We will be working on ($\Leftarrow$).

31.1 A tower of cyclic Galois extensions with roots of unity

Suppose F is a field of characteristic zero and $f \in F[x]$. Then

$$f \textit{ is solvable by radicals over } F \iff \mathscr{G}_{f,F} \textit{ is solvable.}$$

We have already proved ($\Rightarrow$). Suppose K is a splitting field of f over F and $\mathrm{Aut}_F(K)$ is solvable. Then by Lemma 30.4.2, there is a chain of subgroups

$$\mathrm{Aut}_F(K) =: G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_{m+1} := 1,$$

such that G_i/G_{i+1} is a cyclic group of prime order. Let $F_i := \mathrm{Fix}(G_i)$ for every i. Then by the fundamental theorem of Galois theory, K/F_i is a Galois group and $\mathrm{Aut}_{F_i}(K) = G_i$. Since G_{i+1} is a normal subgroup of G_i, by the fundamental theorem of Galois theory F_{i+1}/F_i is a Galois extension, and by Proposition 23.1.1 we obtain that

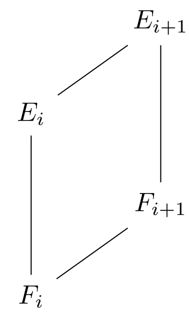
$$\mathrm{Aut}_{F_i}(F_{i+1}) \simeq \frac{\mathrm{Aut}_{F_i}(K)}{\mathrm{Aut}_{F_{i+1}}(K)} = \frac{G_i}{G_{i+1}}.$$

Hence $\mathrm{Aut}_{F_i}(F_{i+1})$ is a cyclic group of prime order. Following Kummer, we add enough roots of unity to the base field. Let $n := [K : F]$, and suppose L is a splitting field of $x^n - 1$ over K. Then by Lemma 29.1.4, L/F is a normal extension. Since $\mathrm{char}(F) = 0$, by Theorem 27.3.5 L/F is a separable extension. Hence L/F is a Galois extension. By Proposition 28.3.4, there is $\zeta \in L$ such that $o(\zeta) = n$, $x^n - 1 = \prod_{i=0}^{n-1}(x - \zeta^i)$, and $L = K[\zeta]$. Let $E_i := F_i[\zeta]$. Thus we have

$$F \subseteq E_0 \subseteq E_1 \subseteq \cdots \subseteq E_{m+1} = L. \tag{31.1}$$

Notice that $F_i[\zeta]$ is a splitting field of $x^n - 1$ over F_i as $x^n - 1 = \prod_{i=0}^{n-1}(x - \zeta^i)$. Hence E_i/F_i is a Galois extension (notice that since the characteristic of all these fields are zero, all these field extensions are separable). We also notice that since $x^n - 1 \in F_i[x]$,

F_{i+1}/F_i is a Galois extension, and E_{i+1} is a splitting field of $x^n - 1$ over F_{i+1}, by Lemma 29.1.4 E_{i+1}/F_i is a Galois extension. By Proposition 23.1.1, the restriction map from $\mathrm{Aut}_{F_i}(E_{i+1})$ to $\mathrm{Aut}_{F_i}(F_{i+1})$ is surjective and its kernel is $\mathrm{Aut}_{F_{i+1}}(E_{i+1})$. Notice that



$$\mathrm{Aut}_{F_{i+1}}(E_{i+1}) \cap \mathrm{Aut}_{E_i}(E_{i+1}) = \mathrm{id}\,.$$

Hence $\mathrm{Aut}_{E_i}(E_{i+1})$ can be embedded into $\mathrm{Aut}_{F_i}(F_{i+1})$. As subgroups of cyclic groups are cyclic, we conclude that $\mathrm{Aut}_{E_i}(E_{i+1})$ is cyclic.

Moreover $|\,\mathrm{Aut}_{E_i}(E_{i+1})|$ divides $|\,\mathrm{Aut}_{F_i}(F_{i+1})|$. As F_{i+1}/F_i is a Galois group, $|\,\mathrm{Aut}_{F_i}(F_{i+1})| = [F_{i+1} : F_i]$. Hence by the Tower Rule, $[F_{i+1} : F_i]$ divides $[K : F]$. Therefore, we conclude that $|\,\mathrm{Aut}_{E_i}(E_{i+1})|$ divides n. Altogether, we conclude the following lemma.

Lemma 31.1.1. *Suppose F is a field of characteristic zero and $f \in F[x]$. Suppose $\mathscr{G}_{f,F}$ is a solvable group of order n and K is a splitting field of f over F. Then there are a tower of fields*

$$F \subseteq E_0 \subseteq E_1 \subseteq \cdots \subseteq E_{m+1},$$

and $\zeta \in E_0$ such that $o(\zeta) = n$, $E_0 = F[\zeta]$, $E_{m+1} = K[\zeta]$, E_{i+1}/E_i is a Galois extension, $\mathrm{Aut}_{E_i}(E_{i+1})$ is cyclic, and $[E_{i+1} : E_i]$ divides n.

Next we will study the blocks of the tower given in Lemma 31.1.1. This will be done by proving Hilbert's theorem 90.

31.2 Hilbert's theorem 90

To formulate Hilbert's theorem 90, we need to define the norm function.

Definition 31.2.1. *Suppose E/F is finite Galois extension. Then for $\alpha \in E$, we let*

$$N_{E/F}(\alpha) := \prod_{\theta \in \mathrm{Aut}_F(E)} \theta(\alpha),$$

and we call it the norm of α.

Lemma 31.2.2. *Suppose E/F is a finite Galois extension. Then for every $\alpha \in E$, we have $N_{E/F}(\alpha) \in F$, and $N_{E/F} : E^\times \to F^\times$ is a group homomorphism.*

Proof. For every $\theta' \in \mathrm{Aut}_F(E)$, we have

$$\begin{aligned}
\theta'(N_{E/F}(\alpha)) =& \theta'\Big(\prod_{\theta \in \mathrm{Aut}_F(E)} \theta(\alpha) \Big) \\
=& \prod_{\theta \in \mathrm{Aut}_F(E)} (\theta' \circ \theta)(\alpha) \\
=& \prod_{\theta \in \mathrm{Aut}_F(E)} \theta(\alpha) = N_{E/F}(\alpha).
\end{aligned}$$

Hence $N_{E/F}(\alpha) \in \mathrm{Fix}(\mathrm{Aut}_F(E))$. As E/F is a finite Galois extension, $\mathrm{Fix}(\mathrm{Aut}_F(E)) = F$. Therefore $N_{E/F}(\alpha) \in F$.

For $\alpha_1, \alpha_2 \in E^\times$, we have

$$\begin{aligned} N_{E/F}(\alpha_1\alpha_2) = & \prod_{\theta \in \mathrm{Aut}_F(E)} \theta(\alpha_1\alpha_2) \\ = & \prod_{\theta \in \mathrm{Aut}_F(E)} \theta(\alpha_1) \cdot \prod_{\theta \in \mathrm{Aut}_F(E)} \theta(\alpha_2) \\ = & N_{E/F}(\alpha_1) N_{E/F}(\alpha_2). \end{aligned}$$

This completes the proof. □

Theorem 31.2.3 (Hilbert's theorem 90)**.** *Suppose E/F is a finite Galois extension, and $\mathrm{Aut}_F(E)$ is a cyclic group generated by σ. Then for $\alpha \in E$, we have $N_{E/F}(\alpha) = 1$ if and only if $\alpha = \frac{\sigma(\beta)}{\beta}$ for some $\beta \in E^\times$.*

We start with proof of ($\Leftarrow$). Suppose $[E : F] = n$. Then

$$\mathrm{Aut}_F(E) = \{\mathrm{id}, \sigma, \cdots, \sigma^{n-1}\}.$$

Hence

$$\begin{aligned} N_{E/F}(\sigma(\beta)) = & \sigma(\beta)\ \sigma(\sigma(\beta)) \cdots \sigma^{n-2}(\sigma(\beta))\ \sigma^{n-1}(\sigma(\beta)) \\ = & \sigma(\beta)\ \sigma^2(\beta) \cdots \sigma^{n-1}(\beta)\ \beta = N_{E/F}(\beta). \end{aligned}$$

Therefore by Lemma 31.2.2, we have

$$N_{E/F}\Big(\frac{\sigma(\beta)}{\beta}\Big) = \frac{N_{E/F}(\sigma(\beta))}{N_{E/F}(\beta)} = \frac{N_{E/F}(\beta)}{N_{E/F}(\beta)} = 1.$$

To show the other direction, we start with proving Dirichlet's independence of characters.

Theorem 31.2.4 (Dirichlet's independence of characters)**.** *Suppose E is a field, G is a group, and*

$$\chi_1, \ldots, \chi_n : G \to E^\times$$

are distinct group homomorphisms. Then χ_i's are E-linearly independent; that means if for some $e_i \in E$,

$$e_1\chi_1(g) + \cdots + e_n\chi_n(g) = 0 \textit{ for every } g \in G,$$

then $e_1 = \cdots = e_n = 0$.

Proof. Suppose to the contrary that χ_i's are E-linearly dependent, and consider

$$V := \{(e_1, \ldots, e_n) \in E^n \mid \sum_{i=1}^n e_i\chi_i = 0\}.$$

By $\sum_{i=1}^n e_i\chi_i = 0$, we mean $\sum_{i=1}^n e_i\chi_i(g) = 0$ for every $g \in G$. Notice that V is a subspace of E^n, and by the contrary assumption $V \neq 0$.

We also notice that, if $(e_1, \dots, e_n) \in V$, then for every $g_0, g \in G$, we have

$$\sum_{i=1}^n e_i\chi_i(g_0 g) = 0.$$

Therefore $\sum_{i=1}^n (e_i\chi_i(g_0))\chi_i(g) = 0$, which means $\sum_{i=1}^n (e_i\chi_i(g_0))\chi_i = 0$. Altogether, we have that for every $g_0 \in G$,

$$(e_1, \dots, e_n) \in V \text{ implies } (e_1\chi_1(g_0), \dots, e_n\chi_n(g_0)) \in V. \tag{31.2}$$

Let $\ell : E^n \to [0, n], \ell(e_1, \dots, e_n) := |\{i \mid e_i \neq 0\}|$. Let

$$m := \min\{\ell(v) \mid v \in V \setminus \{0\}\}. \tag{31.3}$$

Suppose $v_0 \in V$ and $\ell(v_0) = m$. After rearranging the χ_i's, we can and will assume that

$$v_0 = (e_1, \dots, e_m, 0, \dots, 0)$$

for some $e_i \in E^\times$. Notice that $m \neq 1$ as otherwise $\chi_1(g) = 0$ for every $g \in G$. This contradicts the fact that $\chi_1(g) \in E^\times$ for every $g \in G$. Since $\chi_1 \neq \chi_2$, there is $g_0 \in G$ such that $\chi_1(g_0) \neq \chi_2(g_0)$. By (31.2), we have

$$(e_1\chi_1(g_0), \cdots, e_m\chi_m(g_0), 0, \cdots, 0) \in V. \tag{31.4}$$

Since V is a subspace, by (31.4), we deduce that

$$(e_1\chi_1(g_0), \cdots, e_m\chi_m(g_0), 0, \cdots, 0) - \chi_1(g_0)(e_1, \dots, e_m, 0, \dots, 0) \in V.$$

This means that

$$w := (0, e_2(\chi_2(g_0) - \chi_1(g_0)), \dots, e_m(\chi_m(g_0) - \chi_1(g_0)), 0, \dots, 0) \in V. \tag{31.5}$$

By (31.5), we have $\ell(w) < m$. Thus by (31.3) and (31.5), we obtain that $w = 0$. Hence $e_2(\chi_2(g_0) - \chi_1(g_0)) = 0$, which is a contradiction as $e_2 \neq 0$ and $\chi_1(g_0) \neq \chi_2(g_0)$. This completes the proof. □

Proof of Hilbert's theorem 90. We have already proved ($\Rightarrow$). So we focus on ($\Leftarrow$). Suppose $N_{E/F}(\alpha) = 1$ for some $\alpha \in E$. We want to show that $\alpha = \frac{\sigma(\beta)}{\beta}$ for some $\beta \in E^\times$. Equivalently, we want to show that $\alpha^{-1}\sigma(\beta) = \beta$ for some $\beta \in E^\times$. Let

$$T_\alpha : E \to E, \; T_\alpha(e) := \alpha^{-1}\sigma(e).$$

This part is not needed in the proof, but it explains some hidden logic behind the argument. We notice that T_α is an F-linear function, and we would like to show $T_\alpha(\beta)$ for some non-zero element $\beta \in E$. This is equivalent to saying that 1 is eigenvalue of T_α. Later we will discuss that minimal polynomial of a linear map and show that eigenvalues are zeros of minimal polynomial.

We continue with understanding of T_α^i. Notice that

$$T_\alpha^2(e) = T_\alpha(\alpha^{-1}\sigma(e)) = \alpha^{-1}\sigma(\alpha^{-1}\sigma(e)) = (\alpha^{-1}\sigma(\alpha)^{-1})\,\sigma^2(e).$$

Hence

$$\begin{aligned} T_\alpha^3(e) =& T_\alpha((\alpha^{-1}\sigma(\alpha)^{-1})\,\sigma^2(e)) \\ =& \alpha^{-1}\sigma((\alpha^{-1}\sigma(\alpha)^{-1})\,\sigma^2(e)) \\ =& (\alpha\sigma(\alpha)\sigma^2(\alpha))^{-1}\,\sigma^3(e). \end{aligned}$$

Inductively we can show that

$$T_\alpha^i(e) = (\alpha\sigma(\alpha)\cdots\sigma^{i-1}(\alpha))^{-1}\sigma^i(e). \tag{31.6}$$

Substituting n for i in (31.6), we obtain

$$T_\alpha^n(e) = (\alpha\sigma(\alpha)\cdots\sigma^{n-1}(\alpha))^{-1}\sigma^n(e) = N_{E/F}(\alpha)^{-1}e = e;$$

this means

$$T_\alpha^n = \mathrm{id}\,. \tag{31.7}$$

Let's go back to what our goal is. We want to show that T_α has a non-zero fixed point. A common technique of finding a fixed point for a group action is looking at *the center of mass* of an orbit. Here $\{\mathrm{id}, T_\alpha, \ldots, T_\alpha^{n-1}\}$ is acting on E. So for every $e \in E$,

$$\frac{1}{n}(e + T_\alpha(e) + \cdots + T_\alpha^{n-1}(e))$$

showed be a fixed point of T_α. Since this action is linear, $e + T_\alpha(e) + \cdots + T_\alpha^{n-1}(e)$ should be a fixed point of T_α. After showing this, it would only remain to show that $e + T_\alpha(e) + \cdots + T_\alpha^{n-1}(e) \neq 0$ for some $e \in E$. For every $e \in E$, we have

$$\begin{aligned} T_\alpha(e + T_\alpha(e) + \cdots + T_\alpha^{n-1}(e)) =& T_\alpha(e) + \cdots + T_\alpha^{n-1}(e) + T_\alpha^n(e) \\ =& T_\alpha(e) + \cdots + T_\alpha^{n-1}(e) + e. \end{aligned} \tag{31.8}$$

Notice that by (31.6), we have $T_\alpha^i(e) = e_i\sigma^i(e)$ where $e_i := (\alpha\sigma(\alpha)\cdots\sigma^{i-1}(\alpha))^{-1} \in E$. Hence

$$e + T_\alpha(e) + \cdots + T_\alpha^{n-1}(e) = \mathrm{id}(e) + e_1\sigma(e) + \cdots + e_{n-1}\sigma^{n-1}(e). \tag{31.9}$$

Since $\mathrm{id}, \sigma, \ldots, \sigma^{n-1} : E^\times \to E^\times$ are distinct group homomorphisms, by Dirichlet's independence of characters theorem (see Theorem 31.2.4) $\mathrm{id}, \sigma, \ldots, \sigma^{n-1}$ are E-linearly independent. Thus for some $e \in E$,

$$\beta := \mathrm{id}(e) + e_1\sigma(e) + \cdots + e_{n-1}\sigma^{n-1}(e) \neq 0. \tag{31.10}$$

Then by (31.8), (31.9), and (31.10), we conclude that $T_\alpha(\beta) = \beta$ and $\beta \neq 0$. Therefore $\alpha^{-1}\sigma(\beta) = \beta$, which implies that $\alpha = \frac{\sigma(\beta)}{\beta}$. This completes the proof. □

31.3 Cyclic extensions with enough roots of unity

The main goal of this section is to understand cyclic extensions that are related with Galois's solvability theorem (see Lemma 31.1.1).

Theorem 31.3.1. *Suppose F is a field of characteristic zero and E/F is a finite Galois extension such that:*

1. $\mathrm{Aut}_F(E)$ *is a cyclic group of order n and it is generated by σ.*
2. *There is $\zeta \in F$ such that $o(\zeta) = n$.*

Then $E = F[\sqrt[n]{a}]$ for some $a \in F$, where $\sqrt[n]{a}$ is a zero of $x^n - a$. (This means E/F is a Kummer extension.)

Proof. Notice that

$$N_{E/F}(\zeta) = \zeta \cdot \sigma(\zeta) \cdots \sigma^{n-1}(\zeta) = \underbrace{\zeta \cdot \zeta \cdots \zeta}_{n \text{ times}} = \zeta^n = 1.$$

Hence by Hilbert's theorem 90, $\zeta = \frac{\sigma)\beta)}{\beta}$ for some $\beta \in E^\times$. Therefore $\sigma(\beta) = \zeta\beta$. This implies that $\sigma^2(\beta) = \sigma(\zeta\beta) = \zeta\sigma(\beta) = \zeta^2\beta$. Inductively we can show that $\sigma^i(\beta) = \zeta^i\beta$ for every positive integer i. Hence the $\mathrm{Aut}_F(E)$-orbit of β is

$$\mathscr{O}_\beta := \{\beta, \zeta\beta, \ldots, \zeta^{n-1}\beta\}.$$

As $o(\zeta) = n$, all these elements are distinct. Since E/F is a Galois extension, by Theorem 25.3.1

$$m_{\beta,F}(x) = \prod_{\beta' \in \mathscr{O}_\beta} (x - \beta') = \prod_{i=0}^{n-1} (x - \zeta^i\beta). \tag{31.11}$$

Notice that

$$N_{E/F}(\beta) = \beta \cdot \sigma(\beta) \cdots \sigma^{n-1}(\beta) = \beta \cdot \zeta\beta \cdots \zeta^{n-1}\beta = \zeta^{\frac{n(n-1)}{2}} \cdot \beta^n \tag{31.12}$$

We have $(\zeta^{\frac{n(n-1)}{2}})^2 = \zeta^{n(n-1)} = 1$. Therefore $\zeta^{\frac{n(n-1)}{2}} = \pm 1$. Hence by (31.11) and (31.12), we conclude that

$$\beta^n = \pm N_{E/F}(\beta) \in F.$$

Let $a := \beta^n \in F$. Then for every integer i, we have $(\zeta^i\beta)^n = \beta^n = a$. Therefore by the generalized factor theorem, comparing degrees and leading coefficients, we obtain that

$$x^n - a = (x - \beta)(x - \zeta\beta) \cdots (x - \zeta^{n-1}\beta). \tag{31.13}$$

By (31.11) and (31.13), we conclude that $m_{\beta,F}(x) = x^n - a$. Thus $[F[\beta] : F] = n$, and by the Tower Rule, we conclude that

$$E = F[\beta] = F[\sqrt[n]{a}],$$

which completes the proof. $\square$

31.4 Completing proof of Galois's solvability theorem

In this section, we complete the proof of Theorem 30.4.1, which states

f is solvable by radicals over F if and only if $\mathscr{G}_{f,F}$ is solvable

where F is a field of characteristic zero and $f \in F[x] \setminus F$. We have already proved ($\Leftarrow$) (see Theorem 29.2.3). So we focus on ($\Rightarrow$). Let K be a splitting field of f over F and $[K : F] = n$. Then by Lemma 31.1.1, there are a tower of fields

$$F \subseteq E_0 \subseteq \cdots \subseteq E_{m+1},$$

and $\zeta \in E_0$ such that $o(\zeta) = n$, $E_0 = F[\zeta]$, $E_{m+1} = K[\zeta]$, E_{i+1}/E_i is a Galois extension, $\mathrm{Aut}_{E_i}(E_{i+1})$ is cyclic, and $[E_{i+1} : E_i]$ divides n.

Hence by Theorem 31.3.1, $E_{i+1} = E_i[\sqrt[n_i]{a_i}]$ for some $a_i \in E_i$ and a divisor n_i of n. Therefore E_{m+1}/F is a radical extension. As K is a subfield of E_{m+1}, we conclude that f is solvable by radicals over F. This completes the proof.

Chapter 32

Lecture 8

We have been studying zeros of polynomials extensively. For a given polynomial $f \in F[x]$, we proved the existence of an algebraic extension E/F such that f can be decomposed into linear factors in $E[x]$. We can roughly say that *all* the zeros of f are in E. Next we want to see if there is an algebraic extension $\overline{F}/F$ which contains *all* the zeros of all non-constant polynomials over F.

32.1 Algebraically closed fields

By the fundamental theorem of algebra, all the zeros of a non-constant polynomial in $\mathbb{C}[x]$ are in $\mathbb{C}$. In this case, we do not need to go to an algebraic extension to find all the zeros of non-constant polynomials. This brings us to the definition of *algebraically closed fields*.

Definition 32.1.1. *We say a field F is algebraically closed if every $f \in F[x] \setminus F$ has a zero in F.*

Proposition 32.1.2. *Suppose F is a field. Then the following statements are equivalent.*

1. *For every $f \in F[x]$, there are $\alpha_1, \ldots, \alpha_n \in F$ such that*

$$f(x) = \mathrm{ld}(f)(x - \alpha_1) \cdots (x - \alpha_n). \tag{32.1}$$

2. *F is algebraically closed.*

3. *If E/F is algebraic extension, then $E = F$.*

Similar to many other properties of fields, we have two types of description: internal (based on properties of $F[x]$) and external (based on extensions of F).

Proof. (1) $\Rightarrow$ (2). For every non-constant polynomial $f \in F[x]$, there are α_i's in F such that (32.1) holds. Evaluating both sides of (32.1) at α_1, we conclude that $\alpha_1 \in F$ is a zero of F. So every non-constant polynomial in $F[x]$ has a zero in F. This means F is algebraically closed.

(2) ⇒ (3). Suppose E/F is an algebraic extension. To show that $E = F$, we argue that every element $\alpha \in E$ is in F. For every $\alpha \in E$, $m_{\alpha,F}$ is a non-constant polynomial in $F[x]$. Hence $m_{\alpha,F}$ has a zero $\alpha' \in F$ as F is algebraically zero. Therefore by the factor theorem (see Theorem 7.1.1), $m_{\alpha,F}(x) = (x - \alpha')g(x)$ for some $g(x) \in F[x]$. Since $m_{\alpha,F}(x)$ is irreducible in $F[x]$ (see Theorem 8.2.4) and it is monic, we obtain that $m_{\alpha,F}(x) = x - \alpha'$. Hence $\alpha = \alpha' \in F$. Therefore $E = F$.

(3) ⇒ (1). Suppose $f \in F[x] \setminus F$. Let E be a splitting field of f over F. Then E/F is an algebraic extension, and so by hypothesis $E = F$. This means there are $\alpha_1, \ldots, \alpha_n \in F$ such that (32.1) holds. This completes the proof. □

32.2 Zorn's lemma

Our next goal is to show the existence of an algebraic extension $\overline{F}/F$ such that $\overline{F}$ is algebraically closed. Such an extension is called an *algebraic closure* of F. Later we prove that there is a *unique* algebraic closure of F up to an F-isomorphism. Intuitively we *order* all the monic polynomials in $F[x] \setminus F$ and consider the *chain* of splitting fields of these polynomials each one over the *previous* one. Finally we take the union of this chain. This way we end up getting an algebraic field extension E/F which contains *all* the zeros of polynomials in $F[x] \setminus F$. One can show that every algebraic extension of E is E and conclude that E is an algebraic closure of F. We could make the above argument formal using induction if there are only *countably* many polynomials in $F[x]$. When there are uncountably many polynomials in $F[x] \setminus F$, it is not clear what we mean by *ordering* them and what *previous* can mean.

To have a formal *inductive* argument beyond countable sets that are naturally *ordered*, we need to use concepts, axioms, and results from set theory. In set theory, there are results like *transfinite induction* and *well-ordering theorem* that can be viewed as needed extensions for having inductive arguments beyond countable sets. These results are based on the following rather *intuitive* axiom from set theory.

Axiom of Choice. *Suppose I and X are non-empty sets and $\{X_i\}_{i\in I}$ is a collection of non-empty subsets of X; that means $i \mapsto X_i$ is a function from I to $P(X) \setminus \{\varnothing\}$ where $P(X)$ is the power set* [1] *of X. Then we can* choose one element from each *X_i; that means there is a function $f : I \to X$ such that $f(i) \in X_i$ for every $i \in I$.*

Similar to basic number theory, where one often uses the well-ordering principle of positive integers as a replacement for inductive arguments, in algebra we often use a result known as *Zorn's lemma* as a replacement for (well-ordering theorem and) transfinite induction. So next we introduce the needed concepts to formulate Zorn's lemma. Later we will provide a pseudo-reasoning on why Zorn's lemma is well-suited for many algebraic arguments.

We say a non-empty set Σ is a *partially ordered set* with a *partial order* $\preccurlyeq$ if $\preccurlyeq$ is a relation between *some* of the elements of Σ with the following properties:

1. (Reflexive) For every $a \in \Sigma$, $a \preccurlyeq a$.
2. (Anti-symmetric) For every $a, b \in \Sigma$, if $a \preccurlyeq b$ and $b \preccurlyeq a$, then $a = b$.

[1] The power set of X is the set of all the subsets of X.

3. (Transitive) For every $a, b, c \in \Sigma$, if $a \preccurlyeq b$ and $b \preccurlyeq c$, then $a \preccurlyeq c$.

A partially ordered set will be simply called a *poset*. The following two examples are our main sources of constructing many posets.

Example 32.2.1. *Suppose X is a set and $P(X)$ is its power set. Then $P(X)$ partially ordered by* inclusion $\subseteq$ *is a poset.*

Example 32.2.2. *Suppose Σ is a poset with partial order $\preccurlyeq$. Then $\preccurlyeq$ makes every non-empty subset of Σ into a poset.*

Definition 32.2.3 (Chain)**.** *A poset $\mathscr{C}$ with a partial order $\preccurlyeq$ is called a* chain *(or* totally ordered set*) if for every $a, b \in \mathscr{C}$ either $a \preccurlyeq b$ or $b \preccurlyeq a$.*

Example 32.2.4. *Suppose X is a non-empty set and $\{A_i\}_{i=1}^{\infty}$ is a sequence of subsets of X. If*

$$A_1 \subseteq A_2 \subseteq \cdots,$$

then $\{A_i \mid i \in \mathbb{Z}^+\}$ is a chain with respect to inclusion.

Definition 32.2.5 (Upper bound)**.** *Suppose Σ is a poset with a partial order $\preccurlyeq$ and A is a non-empty subset of Σ. Then we say $u \in \Sigma$ is an* upper bound *of A if for every $a \in A$ we have $a \preccurlyeq u$.*

Definition 32.2.6 (Maximal)**.** *Suppose Σ is a poset with a partial order $\preccurlyeq$. We say $m \in \Sigma$ is a* maximal *element of Σ if the only element a of Σ that satisfies $m \preccurlyeq a$ is m.*

The following is an important example in algebra.

Example 32.2.7. *Suppose A is a unital ring, and Σ is the set of all proper ideals of A. A maximal element of the poset Σ with respect to inclusion is a maximal ideal.*

Next we see an example of a chain with no maximal element.

Example 32.2.8. *Let $\mathscr{C} := \{[0, n] \mid n \in \mathbb{Z}^+\}$. Then $\mathscr{C}$ is a chain with respect to the inclusion, and $\mathscr{C}$ has no maximal elements.*

Maximum vs maximal. Let's remark that maximal elements of a poset Σ are not necessarily *maximum* in Σ; that means if $m \in \Sigma$ is maximal, we do not necessarily have that $a \preccurlyeq m$ for every $a \in A$. If Σ is a chain, however, then every maximal element of Σ is maximum in Σ.

Theorem 32.2.9 (Zorn's Lemma)**.** *Suppose Σ is a poset such that every subchain $\mathscr{C}$ of Σ has an upper bound in Σ. Then Σ has a maximal element.*

Assuming all the axioms of set theory besides the axiom of choice, we have that

axiom of choice is equivalent to *Zorn's lemma.*

Instead of proving this result, we will see how one can use Zorn's lemma and why it is so effective in algebra.

32.3 Maximal ideals

In this section, we use Zorn's lemma to prove the following important result in ring theory.

Theorem 32.3.1. *Suppose A is a unital commutative ring and I_0 is a proper ideal of A. Then there is $M \trianglelefteq A$ such that $I_0 \subseteq M$, and M is a maximal ideal of A.*

Proof. Let $\Sigma := \{J \trianglelefteq A \mid I_0 \subseteq J \text{ and } J \neq A\}$. Notice that $I_0 \in \Sigma$, and so Σ is a poset with inclusion as its partial order. To see why the poset Σ is helpful, we argue that the assertion of theorem follows from the existence of a maximal element in Σ. We do this by proving the following:

Step 1. *A maximal element M of Σ is a maximal ideal of A and $I_0 \subseteq M$.*

Proof of Step 1. Suppose J is a proper ideal of A and $M \subseteq J$. Then $I_0 \subseteq J$ as $I_0 \subseteq M$ and $M \subseteq J$. Hence $J \in \Sigma$. Since M is a maximal element of Σ, $J \in \Sigma$ and $M \subseteq J$, we conclude that $M = J$. Hence M is a maximal ideal of A. As M is in Σ, $I_0 \subseteq M$. This completes proof of the Claim.

To show Σ has a maximal element, we employ Zorn's lemma (see Theorem 32.2.9). Hence it is sufficient to show that every chain $\mathscr{C} \subseteq \Sigma$ has an upper bound.

Here is the key place. The following pseudo-argument shows why we often can find an upper bound for a chain consisting of *algebraic objects*, which in part shows effectiveness of Zorn's lemma in algebraic settings.

Suppose $\mathscr{C}$ is a chain consisting of *certain algebraic objects*. Then $\bigcup_{I \in \mathscr{C}} I$ satisfies *the same algebraic properties*. This is the case as the *algebraic objects* are often defined in terms of properties of operations and algebraic operations involve only *finitely many elements* of $\bigcup_{I \in \mathscr{C}} I$. Since every finitely many elements in a chain have a maximum, the given finitely many elements of $\bigcup_{I \in \mathscr{C}} I$ are in a single $I \in \mathscr{C}$. As every element of $\mathscr{C}$ has the *desired algebraic properties*, we obtain the same properties for $\bigcup_{I \in \mathscr{C}} I$.

The following step is an instance of the above pseudo-argument.

Step 2. *Suppose $\mathscr{C}$ is a chain of ideals of A. Then $J := \bigcup_{I \in \mathscr{C}} I$ is an ideal of A.*

Proof of Step 2. For every $x, x' \in J$, there are $I, I' \in \mathscr{C}$ such that $x \in I$ and $x' \in I'$. Since $\mathscr{C}$ is a chain, either $I \subseteq I'$ or $I' \subseteq I$. Without loss of generality we can and will assume that $I \subseteq I'$. Hence $x, x' \in I'$. As I' is an ideal of A, for every $a \in A$, we have $ax + x' \in I'$. Therefore $ax + x' \in J$ as $I' \subseteq J$. Thus J is an ideal of A. This completes the proof of Step 2.

In the next steps, we show that the ideal J given in Step 2 is an upper bound of $\mathscr{C}$ in Σ if $\mathscr{C} \subseteq \Sigma$ is a chain.

Step 3. *Suppose $\mathscr{C} \subseteq \Sigma$ is a chain. Then $J := \bigcup_{I \in \mathscr{C}} I$ is a proper ideal of A.*

Proof of Step 3. By Step 2, we know that J is an ideal of A. So if, on the contrary, J is not a proper ideal, then $1 \in J$. Hence $1 \in I$ for some $I \in \mathscr{C}$. Therefore we obtain that $I = A$ for some $I \in \mathscr{C}$. This contradicts the fact that $\mathscr{C}$ consists of *proper* ideals.

Step 4. *Suppose $\mathscr{C} \subseteq \Sigma$ is a chain. Then $J := \bigcup_{I \in \mathscr{C}} I$ is in Σ.*

Proof of Step 4. By Step 3, J is a proper ideal of A. It is remained to show that $I_0 \subseteq J$. Suppose $I \in \mathscr{C}$. Then $I_0 \subseteq I$ and $I \subseteq J$, and so $I_0 \subseteq I$. This completes the proof of Step 4.

Step 5. *Suppose $\mathscr{C} \subseteq \Sigma$ is a chain. Then $J := \bigcup_{I \in \mathscr{C}} I$ is an upper bound of $\mathscr{C}$.*

Proof of Step 5. By Step 4, $J \in \Sigma$. Notice that for every $I \in \mathscr{C}$, we have $I \subseteq J$. This implies Step 5.

By Step 5 and Zorn's lemma, we conclude that Σ has a maximal element M. This completes the proof in view of Step 1. □

32.4 Existence of algebraic closure: one zero of every polynomial

The main goal of this section is to show the following Proposition which plays an important role in proving the existence of an algebraic closure of a field.

Proposition 32.4.1. *Suppose F is a field. Then there is an algebraic extension K/F such that every non-constant polynomial in $F[x]$ has a zero in K.*

Let's start by recalling how we can find an extension which contains a zero of a given non-constant polynomial f. In this case, we take an irreducible factor p of f and consider $E := F[x]/\langle p\rangle$. Since p is irreducible, $\langle p\rangle$ is a maximal ideal. Hence E is a field and it has a copy of F as a subfield. Moreover $\alpha_f := x + \langle p\rangle \in E$ is a zero of f (as it is a zero of p and p divides f) (See Lemma 16.1.2).

To find an extension field which contains a zero of all non-constant polynomials in $F[x]$, we consider a ring of polynomials with infinitely many variables; one for each $f \in F[x] \setminus F$. Let

$$A := F[x_f \mid f \in F[x] \setminus F].$$

Similar to the case of one polynomial, we want to find a *maximal* ideal M of A such that $x_f + M \in A/M$ is a zero of f. If we find such a maximal ideal, then A/M is an extension field of F which contains a zero α_f of f for every $f \in F[x] \setminus F$. Notice that $x_f + M$ is a zero of f if and only if $f(x_f) + M = 0 + M$. This means $\alpha_f := x_f + M$ is a zero of f exactly when $f(x_f) \in M$. Therefore we would like to show the existence of a maximal ideal M of A which contains $f(x_f)$ for every $f \in F[x] \setminus F$. The latter is equivalent to saying that

$$I_0 \subseteq M \tag{32.2}$$

where I_0 is the ideal of A which is generated by $f(x_f)$'s. By Theorem 32.3.1, there is a maximal ideal M of A which contains I_0 as a subset exactly when I_0 is a proper ideal. So next we show that I_0 is a proper ideal of A.

Lemma 32.4.2. *Suppose F is a field and $A := F[x_f \mid f \in F[x] \setminus F]$ is the ring of polynomials with infinitely many variables; one for each $f \in F[x] \setminus F$. Let I_0 be the ideal of A which is generated by $\{f(x_f) \mid f \in F[x] \setminus F\}$. Then I_0 is a proper ideal.*

Proof. Suppose to the contrary that $I_0 = A$. Then $1 \in I_0$. This implies that there are $f_i \in F[x] \setminus F$ and $a_i \in A$ such that

$$1 = a_1 f_1(x_{f_1}) + a_2 f_2(x_{f_2}) + \cdots + a_n f_n(x_{f_n}). \tag{32.3}$$

Notice that only finitely many terms and variables are needed to express a_i's and Equation (32.3). To simplify our notation, let's rename the involved variables. We start by renaming x_{f_i}'s to x_i's, and the rest of the involved variables will be named

$x_{n+1}, \ldots, x_m$. This means we can and will view a_i's as polynomials in $F[x_1, \ldots, x_m]$ and (32.3) can be rewritten as

$$1 = a_1(x_1, \ldots, x_m) f_1(x_1) + \cdots + a_n(x_1, \ldots, x_m) f_n(x_n). \tag{32.4}$$

We want to show that (32.4) is not possible. Notice that if (32.4) holds, then for every field extension E/F and $\beta_1, \ldots, \beta_m \in E$ we should have

$$1 = a_1(\beta_1, \ldots, \beta_m) f_1(\beta_1) + \cdots + a_n(\beta_1, \ldots, \beta_m) f_n(\beta_n) \tag{32.5}$$

because of substituting β_i's for x_i's in both sides of (32.4) (the useful principle of viewing polynomials as functions). We get a contradiction if $f_i(\beta_i) = 0$ for every i. To this end, we let E be a splitting field of $\prod_{i=1}^n f_i(x)$ over F, and let $\beta_i \in E$ be a zero of f_i for $i \in [1..n]$. The rest of β_i's can be any element in E; say $\beta_{n+1} = \cdots = \beta_m := 0$. As we said earlier, by (32.5) we obtain $1 = 0$ which is a contradiction. This completes the proof. □

Now we can prove the following lemma.

Lemma 32.4.3. *Suppose F is a field. Then there is a field extension L/F such that every non-constant polynomial in $F[x]$ has a zero in L.*

Proof. Let $A := F[x_f \mid f \in F[x] \setminus F]$, and $I_0 := \langle f(x_f) \mid f \in F[x] \setminus F \rangle$. Then by Lemma 32.4.2, I_0 is a proper ideal of A. Hence by Theorem 32.3.1, there is a maximal ideal M of A which contains I_0. Let $L := A/M$ and $\alpha_f := x_f + M \in L$. Since M is a maximal ideal of A, A/M is a field. We also notice that non-zero constant polynomials are units in A, and so M does not contain any non-zero constant polynomial. This implies that $c \mapsto c + M$ is an embedding of F into L. Hence L is an extension field of F. For every $f \in F[x] \setminus F$, we have

$$f(\alpha_f) = f(x_f) + M = 0 + M,$$

where the last equation holds in view of $f(x_f) \in I_0 \subseteq M$. This completes the proof. □

Proof of Proposition 32.4.1. By Lemma 32.4.3, there is a field extension L/F which contains a zero of every non-constant polynomial of $F[x]$. This means for every $f \in F[x] \setminus F$, there is $\alpha_f \in L$ such that $f(\alpha_f) = 0$. Let $K \in \operatorname{Int}(L/F)$ be the algebraic closure of F in L. Notice that for every $f \in F[x] \setminus F$, $\alpha_f \in L$ is algebraic over F as it is a zero of f. Hence $\alpha_f \in K$ for every $f \in F[x] \setminus F$. Therefore K/F is an algebraic extension which contains a zero of every non-constant polynomial in $F[x]$. □

Chapter 33

Lecture 9

33.1 Existence of an algebraic closure

The main goal of this section is to prove the existence of an algebraic closure of a field. We say $\overline{\mathrm{F}}$ is an *algebraic closure* of F if $\overline{\mathrm{F}}\,/F$ is an algebraic extension and $\overline{\mathrm{F}}$ is algebraically closed.

Theorem 33.1.1. *Every field F has an algebraic closure* $\overline{\mathrm{F}}$.

Proof. Let $F_0 := F$. By applying Proposition 32.4.1, we obtain a tower of fields

$$F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots$$

such that for every i

1. F_{i+1}/F_i is an algebraic extension, and
2. every non-constant polynomial in $F_i[x]$ has a zero in F_{i+1}.

Let $\overline{\mathrm{F}} := \bigcup_{i=0}^{\infty}$. Using the rough principle of *union of a chain of algebraic objects should satisfy similar properties*, we expect $\overline{\mathrm{F}}$ to be a field

Step 1. $\overline{\mathrm{F}}$ *is a field.*

Proof of Step 1. For every $a, b \in \overline{\mathrm{F}}$, there are i and j such that $a \in F_i$ and $b \in F_j$. Let $k := \max(i, j)$ and notice that $F_i \cup F_j = F_k$. Hence for every $a, b \in \overline{\mathrm{F}}$ there is a positive integer k such that $a, b \in F_k$. We define the addition and multiplication of a and b as $a + b$ and $a \cdot b$ in F_k if $a, b \in F_k$. Notice that if a and b are in F_l as well, then either $a + b$ and $a \cdot b$ in F_k and F_l are the same as either F_k is a subfield of F_l or F_l is a subfield of F_k. Hence there are well-defined operations on $\overline{\mathrm{F}}$. For every $a, b, c \in \overline{\mathrm{F}}$, there are i, j, and k such that $a \in F_i$, $b \in F_j$, and $c \in F_k$. Hence $a, b, c \in F_r$ where $r := \max(i, j, k)$. As F_r is a field, we have that $a + 0 = 0 + a = a$, $a \cdot 1 = a$, $(a + b) + c = a + (b + c)$, $a + b = b + a$, $a \cdot b = b \cdot a$, $a \cdot (b + c) = a \cdot b + a \cdot c$, there is $-a \in F_r$ such that $a + (-a) = 0$, there is $a^{-1} \in F_r$ if $a \neq 0$. This completes the proof of Step 1. □

Step 2. $\overline{\mathrm{F}}\,/F$ *is an algebraic extension.*

Proof of Step 2. Suppose $\alpha \in \overline{\mathrm{F}}$. Then $\alpha \in F_i$ for some positive integer i. As $F_1/F_0, \ldots, F_i/F_{i-1}$ are algebraic extensions, F_i/F_0 is an algebraic extension (see Proposition 20.6.1). Hence α is algebraic over $F_0 = F$. □

Step 3. $\overline{\mathrm{F}}$ *is algebraically closed.*

Proof of Step 3. We want to show that every non-constant polynomial in $\overline{\mathrm{F}}[x]$ has a zero in $\overline{\mathrm{F}}$. Suppose

$$g(x) := a_m x^m + \cdots + a_1 x + a_0 \in \overline{\mathrm{F}}[x] \setminus \overline{\mathrm{F}}\,.$$

As $a_i \in \overline{\mathrm{F}}$, $a_i \in F_{n_i}$ for some positive integer n_i. Let n be the maximum of $n_1, \ldots, n_m$. Then $g(x) \in F_n[x] \setminus F_n$. Hence g has a zero in F_{n+1}. Hence g has a zero in $\overline{\mathrm{F}}$. This completes the proof. □

By Step 2 and Step 3, we deduce that $\overline{\mathrm{F}}$ is an algebraic closure of F. The claim follows. □

33.2 Isomorphism extension theorem for algebraic closures.

In this section, we prove an isomorphism extension theorem for algebraic closures and as a particular case, we obtain the uniqueness of algebraic closures up to an isomorphism.

Theorem 33.2.1 (The isomorphism extension for algebraic closures)**.** *Suppose F and F' are fields, $\theta : F \to F'$ is an isomorphism, and $\overline{\mathrm{F}}, \overline{\mathrm{F}}'$ are algebraically closures of F and F', respectively. Then there is an isomorphism $\widehat{\theta} : \overline{\mathrm{F}} \to \overline{\mathrm{F}}'$ such that $\widehat{\theta}|_F = \theta$.*

Roughly the isomorphism extension for algebraic closures hold because they can be viewed as the union of a family of splitting fields of polynomials in $F[x] \setminus F$ over F. We can construct the extension $\widehat{\theta}$ by gluing the F-isomorphism extensions for splitting fields. The gluing process, however, needs a set theoretic "permission" as there might be uncountably many splitting fields involved. Similar to the proof of the existence of algebraic closures, we use Zorn's lemma to avoid using an inductive argument.

Proof of Theorem 33.2.1. Let

$$\Sigma := \{(E, \phi) \mid E \in \mathrm{Int}(\overline{\mathrm{F}}/F) \text{ and } \phi : E \to \overline{\mathrm{F}}' \text{ such that } \phi|_F = \theta\}.$$

We say $(E_1, \phi_1) \preccurlyeq (E_2, \phi_2)$ if $E_1 \subseteq E_2$ and $\phi_2|_{E_1} = \phi_1$.

Step 1. $(\Sigma, \preccurlyeq)$ *is a poset.*

Proof of Step 1. (Reflexive) Clearly $(E, \phi) \preccurlyeq (E, \phi)$.

(Anti-symmetric) Suppose $(E_1, \phi_1) \preccurlyeq (E_2, \phi_2)$ and $(E_2, \phi_2) \preccurlyeq (E_1, \phi_1)$. Then $E_1 \subseteq E_2$ and $E_2 \subseteq E_1$, and so $E_1 = E_2$. As $\phi_2|_{E_1} = \phi_1$ and $E_1 = E_2$, $\phi_1 = \phi_2$.

(Transitive) Suppose $(E_1, \phi_1) \preccurlyeq (E_2, \phi_2)$ and $(E_2, \phi_2) \preccurlyeq (E_3, \phi_3)$. Then $E_1 \subseteq E_2$ and $E_2 \subseteq E_3$, and so $E_1 \subseteq E_3$. As ϕ_2 is an extension of ϕ_1 and ϕ_3 is an extension of ϕ_2, we obtain that ϕ_3 is an extension of ϕ_1. Thus $(E_1, \phi_1) \preccurlyeq (E_3, \phi_3)$. □

Next we use Zorn's lemma to show that Σ has a maximal element.

Step 2. *Suppose* $\mathscr{C} := \{(E_i, \phi_i) \mid i \in I\} \subseteq \Sigma$ *is a chain. Let* $E := \bigcup_{i\in I} E_i$ *and* $\phi : E \to \overline{\mathrm{F}}'$, $\phi(e) = \phi_i(e)$ *if* $e \in E_i$. *Then* $E \in \mathrm{Int}(\overline{\mathrm{F}}/F)$, ϕ *is well-defined,* ϕ *is an extension of* θ, *and* $(E_i, \phi_i) \preccurlyeq (E, \phi)$ *for every* $i \in I$.

Proof of Step 2. Suppose $a, b \in E \setminus \{0\}$. Then there are $i, j \in I$ such that $a \in E_i$ and $b \in E_j$. Since $\mathscr{C}$ is a chain, either $(E_i, \phi_i) \preccurlyeq (E_j, \phi_j)$ or $(E_j, \phi_j) \preccurlyeq (E_i, \phi_i)$. Without loss of generality, we can and will assume that $(E_i, \phi_i) \preccurlyeq (E_j, \phi_j)$. Hence $E_i \subseteq E_j$ and ϕ_j is an extension of ϕ_i. Therefore $a, b \in E_j$, which implies that $a \pm b, a \cdot b^{\pm 1} \in E_j$. Thus $a \pm b, a \cdot b^{\pm 1} \in E$ as $E_j \subseteq E$. Therefore $E \in \mathrm{Int}(\overline{\mathrm{F}}/F)$.

Next we show that ϕ is well-defined. Suppose $e \in E$ is in E_i and E_j for $i, j \in I$. Since $\mathscr{C}$ is a chain, without loss of generality we can and will assume that $(E_i, \phi_i) \preccurlyeq (E_j, \phi_j)$. Hence $E_i \subseteq E_j$ and ϕ_j is an extension of ϕ_i. Therefore $\phi_i(e) = \phi_j(e)$. This implies that ϕ is well-defined.

It remains to show that ϕ is an embedding which is an extension of θ. Suppose $a, b \in E$. As we proved in Step 1, $a, b \in E_i$ for some $i \in I$. Then $a + b, a \cdot b \in E_i$, and so

$$\phi(a+b) = \phi_i(a+b) = \phi_i(a) + \phi_i(b) = \phi(a) + \phi(b)$$
$$\phi(a \cdot b) = \phi_i(a \cdot b) = \phi_i(a) \cdot \phi_i(b) = \phi(a) \cdot \phi(b).$$

This means that ϕ is a ring homomorphism. For every $a \in F$ and every $i \in I$, $a \in E_i$, and so $\phi(a) = \phi_i(a) = \theta(a)$ as ϕ_i is an extension of θ. This implies that ϕ is an extension of θ. Since E is a field and $\phi \neq 0$, ϕ is injective. Therefore $\phi : E \to \overline{\mathrm{F}}'$ is an embedding and it is an extension of θ. Hence $(E, \phi) \in \Sigma$. For every $i \in I$, $E_i \subseteq E$ and $\phi|_{E_i} = \phi_i$. This means $(E_i, \phi_i) \preccurlyeq (E, \phi)$, which completes proof of Step 2. □

By Zorn's lemma and Step 2, Σ has a maximal element.

Step 3. *Suppose* $(L, \widehat{\theta})$ *is a maximal element of* Σ. *Then* $L = \overline{\mathrm{F}}$.

Proof of Step 3. Suppose to the contrary that $\alpha \in \overline{\mathrm{F}} \setminus L$. Since $\overline{\mathrm{F}}$ and $\overline{\mathrm{F}}'$ are algebraically closed, there are $\alpha_1, \ldots, \alpha_n \in \overline{\mathrm{F}}$ and $\alpha_1', \ldots, \alpha_n' \in \overline{\mathrm{F}}'$ such that

$$m_{\alpha,L}(x) = (x-\alpha_1)\cdots(x-\alpha_n) \text{ and } m_{\theta(\alpha),\theta(L)}(x) = (x-\alpha_1')\cdots(x-\alpha_n'). \quad (33.1)$$

Then $\widehat{L} := L[\alpha_1, \ldots, \alpha_n]$ is a splitting field of $m_{\alpha,L}$ over L and $\widehat{L}' := \theta(L)[\alpha_1', \ldots, \alpha_n']$ is a splitting field of $m_{\theta(\alpha),\theta(L)}$ over $\theta(L)$. Hence by the isomorphism extension for splitting fields (see Theorem 17.1.1), $\widehat{\theta}$ can be extended to an isomorphism $\widetilde{\theta} : \widehat{L} \to \widehat{L}'$. In particular, $\widetilde{\theta}$ is an extension of θ. Hence $(\widehat{L}, \widetilde{\theta}) \in \Sigma$ and $(L, \widehat{\theta}) \preccurlyeq (\widehat{L}, \widetilde{\theta})$. Since $(L, \widehat{\theta})$ is a maximal element of Σ, we conclude that $(L, \widehat{\theta}) = (\widehat{L}, \widetilde{\theta})$. This is a contradiction as $\alpha \in \widehat{L} \setminus L$. □

Step 4. *There is an* F*-isomorphism* $\widehat{\theta} : \overline{\mathrm{F}} \to \overline{\mathrm{F}}'$ *which is an extension of* θ.

Proof of Step 4. By Step 3, there is an embedding $\theta : \overline{\mathrm{F}} \to \overline{\mathrm{F}}'$ which is an extension of θ. Then $\widehat{\theta}(\overline{\mathrm{F}})$ is algebraically closed and $\widehat{\theta}(\overline{\mathrm{F}}) \in \mathrm{Int}(\overline{\mathrm{F}}'/F)$. Therefore $\overline{\mathrm{F}}'$ is an algebraic extension field of the algebraically closed field $\theta(\widehat{L})$. By Proposition 32.1.2, we deduce that $\theta(\widehat{L}) = \overline{\mathrm{F}}'$. This means $\widehat{\theta}$ is an isomorphism, which completes the proof. □

□

An immediate corollary of the isomorphism extension for algebraic closures is the uniqueness of algebraic closures up to an F-isomorphism.

Theorem 33.2.2 (Uniqueness of algebraic closures)**.** *Suppose F is a field and $\overline{\mathrm{F}}, \overline{\mathrm{F}}'$ are algebraic closures of F. Then there is an F-isomorphism $\widehat{\theta} : \overline{\mathrm{F}} \to \overline{\mathrm{F}}'$.*

Proof. Let $\theta := \mathrm{id}_F$ be the identity function from F to F. Then by the isomorphism extension for algebraic extensions, there is an isomorphism $\widehat{\theta} : \overline{\mathrm{F}} \to \overline{\mathrm{F}}'$ which is an extension of θ. This means $\widehat{\theta}$ is an F-isomorphsims, which completes the proof. □

33.3 Basic properties of algebraic closures

In this section, we show some of the basic properties of algebraic closures.

Definition 33.3.1. *We say a field F is a* perfect *field if either $\mathrm{char}(F) = 0$ or $\mathrm{char}(F) = p > 0$ and $F^p = F$; this means the Frobenius map $\phi : F \to F, \phi(a) := a^p$ is an isomorphism.*

Proposition 33.3.2. *Suppose F is a field and $\overline{\mathrm{F}}$ is an algebraic closure of F. Then*

1. *$\overline{\mathrm{F}}/F$ is a normal extension.*
2. *$\overline{\mathrm{F}}/F$ is a Galois extension if and only if F is a perfect field.*

Proof. (1) Since $\overline{\mathrm{F}}$ is an algebraic closure of F, $\overline{\mathrm{F}}/F$ is an algebraic extension. As $\overline{\mathrm{F}}$ is algebraically closed, for every $\alpha \in \overline{\mathrm{F}}$, $m_{\alpha,F}$ can written as a product of linear factors in $\overline{\mathrm{F}}[x]$. This implies part one.

(2) ($\Rightarrow$) If $\mathrm{char}(F) = 0$, there is nothing to prove. So without loss of generality we can and will assume that $\mathrm{char}(F) = p > 0$. As $\overline{\mathrm{F}}$ is algebraically closed, for every $a \in F$, there is $\alpha \in \overline{\mathrm{F}}$ which is a zero of $x^p - a$. This means

$$x^p - a = x^p - \alpha^p = (x - \alpha)^p.$$

Therefore $m_{\alpha,F}$ divides $(x - \alpha)^p$, which implies that

$$m_{\alpha,F}(x) = (x - \alpha)^i \tag{33.2}$$

for some positive integer i. As $\overline{\mathrm{F}}/F$ is a separable extension, $m_{\alpha,F}(x)$ does not have multiple zeros in its splitting field over F. Hence by (33.2), we obtain that $m_{\alpha,F}(x) = x - \alpha$. This means $\alpha \in F$, which implies that

$$a = \alpha^p \in F^p.$$

Since this is true for every $a \in F$, we deduce that $F = F^p$.

($\Leftarrow$) If $\operatorname{char}(\overline{\mathrm{F}}) = 0$, then by Theorem 27.3.5 $\overline{\mathrm{F}}/F$ is a separable extension. Hence together with part (1), we conclude that $\overline{\mathrm{F}}/F$ is a Galois extension. Suppose $\operatorname{char}(F) = p > 0$. Every $\alpha \in \overline{\mathrm{F}}$ is algebraic over F. By Proposition 27.3.4, there is a separable irreducible polynomial $s_{\alpha,F}(x) \in F[x]$ and a non-negative integer k such that $m_{\alpha,F}(x) = s_{\alpha,F}(x^{p^k})$. Suppose

$$s_{\alpha,F}(x) = x^m + c_{m-1}x^{m-1} + \cdots + c_0 \in F[x]. \tag{33.3}$$

Since $F^p = F$, there are d_i's in F such that $c_i = d_i^p$ for every i. Therefore by (33.3), we obtain

$$\begin{aligned} m_{\alpha,F}(x) =& x^{mp^k} + d_{m-1}^p x^{(m-1)p^k} + \cdots + d_0^p \\ =& (x^{mp^{k-1}} + d_{m-1}x^{(m-1)p^{k-1}} + \cdots + d_0)^p \qquad \text{if } k > 0. \end{aligned} \tag{33.4}$$

By (33.4), we reach to a contradiction as $m_{\alpha,F}(x) \in F[x]$ is irreducible and

$$x^{mp^{k-1}} + d_{m-1}x^{(m-1)p^{k-1}} + \cdots + d_0 \in F[x].$$

Hence $k = 0$, which means $m_{\alpha,F} = s_{\alpha,F}$. This implies that $m_{\alpha,F} \in F[x]$ is separable for every $\alpha \in \overline{\mathrm{F}}$. $\square$

Corollary 33.3.3. *A field F is perfect if and only if every algebraic extension E/F is separable.*

Proof. ($\Rightarrow$) Let $\overline{E}$ be an algebraic closure of E. Then $\overline{E}/F$ is an algebraic extension as $\overline{E}/E$ and E/F are algebraic extensions (see Proposition 20.6.1). As $\overline{E}$ is algebraically closed, we deduce that $\overline{E}$ is an algebraic closure of F. Therefore by Proposition 33.3.2, $\overline{E}/F$ is a Galois extension as F is perfect. In particular, $\overline{E}/F$ is a separable extension. As $E \in \operatorname{Int}(\overline{E}/F)$, we conclude that E/F is a separable extension.

($\Leftarrow$) Suppose $\overline{\mathrm{F}}$ is an algebraic closure of F. By hypothesis, $\overline{\mathrm{F}}/F$ is separable. By part (1) of Proposition 33.3.2, $\overline{\mathrm{F}}/F$ is a normal extension. Hence we conclude that $\overline{\mathrm{F}}/F$ is a Galois extension. Hence by part (2) of Proposition 33.3.2, we deduce that F is perfect. This completes the proof. $\square$

The next result shows that if add all the zeros of polynomials in $F[x]$ to F, we end up getting an algebraically closed field.

Proposition 33.3.4. *Suppose F is a field and $\overline{\mathrm{F}}$ is an algebraic closure of F. Let*

$$\operatorname{Int}_{\mathrm{f,n}}(\overline{\mathrm{F}}/F) := \{E \in \operatorname{Int}(\overline{\mathrm{F}}/F) \mid E/F \textit{ is a finite normal extension}\}.$$

Then $\overline{\mathrm{F}} = \bigcup_{E \in \operatorname{Int}_{\mathrm{f,n}}(\overline{\mathrm{F}}/F)} E$.

Proof. Every $\alpha \in \overline{\mathrm{F}}$ is algebraic over F. Since $\overline{\mathrm{F}}$ is algebraically closed,

$$m_{\alpha,F}(x) = (x - \alpha_1) \cdots (x - \alpha_n)$$

for some $\alpha_1 := \alpha, \alpha_2, \ldots, \alpha_n \in \overline{\mathrm{F}}$. Let $E := F[\alpha_1, \ldots, \alpha_n]$. Then E is a splitting field of $m_{\alpha,F}$ over F. Hence E/F is a finite normal extension (see Proposition 23.1.1). Thus $E \in \operatorname{Int}_{\mathrm{f,n}}(\overline{\mathrm{F}}/F)$. This completes the proof as $\alpha \in E$. $\square$

33.4 Group of automorphisms of algebraic closures

In this brief section, we briefly outline how $\mathrm{Aut}_F(\overline{\mathbb{F}})$ looks like where $\overline{\mathbb{F}}$ is an algebraic closure of F.

Lemma 33.4.1. *Suppose F is a field, $\overline{\mathbb{F}}$ is an algebraic closure of F, and $E \in \mathrm{Int}_{\mathrm{f,n}}(\overline{\mathbb{F}}/F)$. Let $r_E : \mathrm{Aut}_F(\overline{\mathbb{F}}) \to \mathrm{Aut}_F(E)$ be the restriction map $r_E(\theta) := \theta|_E$. Then r_E is a well-defined surjective group homomorphism and* $\ker r_E = \mathrm{Aut}_E(\overline{\mathbb{F}})$.

Proof. For every $\alpha \in E$ and $\widehat{\theta} \in \mathrm{Aut}_F(\overline{\mathbb{F}})$, $\widehat{\theta}(\alpha)$ is a zero of $m_{\alpha,F}$. Since E/F is a normal extension, $m_{\alpha,F}$ decomposes into linear factors in $E[x]$. Hence $\widehat{\theta}(\alpha) \in E$. This implies that r_E is well-defined. It is clear that r_E is a group homomorphism. Next we show that r_E is surjective. Suppose $\theta \in \mathrm{Aut}_F(E)$. Notice that $\overline{\mathbb{F}}$ is algebraically closed and $\overline{\mathbb{F}}/E$ is algebraic, and so $\overline{\mathbb{F}}$ is an algebraic closure of E. Therefore by the isomorphism extension for algebraically closed (see Theorem 33.2.1), there is an isomorphism $\widehat{\theta} : \overline{\mathbb{F}} \to \overline{\mathbb{F}}$ which is an extension of θ. Notice that θ is an F-isomorphism, so is $\widehat{\theta}$. Therefore $\widehat{\theta} \in \mathrm{Aut}_F(\overline{\mathbb{F}})$ and $\theta = r_E(\widehat{\theta})$. This shows that r_E is surjective. Finally suppose $\widehat{\theta} \in \ker r_E$. This means $\widehat{\theta}|_E = \mathrm{id}_E$, which implies that $\widehat{\theta} \in \mathrm{Aut}_E(\overline{\mathbb{F}})$. This completes the proof. ☐

If $E, E' \in Int_{\mathrm{f,n}}(\overline{\mathbb{F}}/F)$ and $E \subseteq E'$, then by a similar argument as in the proof of Lemma 33.4.1 we have that

$$r_{E',E} : \mathrm{Aut}_F(E') \to \mathrm{Aut}_F(E),\ r_{E',E}(\theta) := \theta|_E$$

is a well-defined group homomorphism. Notice that

$$r_E = r_{E',E} \circ r_{E'}. \tag{33.5}$$

Lemma 33.4.2. *Suppose F is a field, $\overline{\mathbb{F}}$ is an algebraic closure of F. Let*

$$r : \mathrm{Aut}_F(\overline{\mathbb{F}}) \to \prod_{E \in \mathrm{Int}_{\mathrm{f,n}}(\overline{\mathbb{F}}/F)} \mathrm{Aut}_F(E), \quad r(\theta) := (r_E(\theta)).$$

Then r is an injective group homomorphism.

Proof. Since r_E's are group homomorphisms, r is a group homomorphism. Suppose θ is in the kernel of r. Then for every $E \in \mathrm{Int}_{\mathrm{f,n}}(\overline{\mathbb{F}}/F)$, we have $\theta|_E = \mathrm{id}_E$. Therefore by Proposition 33.3.4, we have $\theta = \mathrm{id}$. This completes the proof. ☐

Next we want to understand the image of r. By (33.5), we have that the image of r is a subset of

$$G(\overline{\mathbb{F}}/F) := \{(\theta_E) \mid \forall E \subseteq E', E, E' \in \mathrm{Int}_{\mathrm{f,n}}(\overline{\mathbb{F}}/F), r_{E',E}(\theta_{E'}) = \theta_E\}.$$

In fact one can show that the image of r is $G(\overline{\mathbb{F}}/F)$, which implies the following result.

Theorem 33.4.3. *Suppose F is a field and $\overline{\mathbb{F}}$ is an algebraic closure of F. Then $r : \mathrm{Aut}_F(\overline{\mathbb{F}}) \to G(\overline{\mathbb{F}}/F)$ is an isomorphism.*

I leave the proof of Theorem 33.4.3 as an exercise. Here are the steps for proving the surjectivity of r. Suppose $(\theta_E) \in G(\overline{\mathrm{F}}/F)$. Let $\theta : \overline{\mathrm{F}} \to \overline{\mathrm{F}}$ be $\theta(\alpha) := \theta_E(\alpha)$ if $\alpha \in E$ and $E \in \mathrm{Int}_{\mathrm{f,n}}(\overline{\mathrm{F}}/F)$.

1. Show that θ is well-defined. To show this notice that if $E_1, E_2 \in \mathrm{Int}_{\mathrm{f,n}}(\overline{\mathrm{F}}/F)$, then there is $E_3 \in \mathrm{Int}_{\mathrm{f,n}}(\overline{\mathrm{F}}/F)$ such that $E_1, E_2 \subseteq E_3$.

2. Show that θ is a homomorphism, and then prove that θ is an F-automorphism.

3. Finally observe that $r(\theta) = (\theta_E)$.

Chapter 34

Lecture 10

34.1 Galois correspondence for algebraic closures

Suppose F is a field and $\overline{\mathrm{F}}$ is an algebraic closure of F. By Proposition 33.3.2, we have that $\overline{\mathrm{F}}/F$ is a Galois extension if and only if F is perfect. So we assume that F is perfect, and investigate how much Galois's correspondence between $\mathrm{Int}(\overline{\mathrm{F}}/F)$ and $\mathrm{Sub}(\mathrm{Aut}_F(\overline{\mathrm{F}}))$ hold. As in the fundamental theorem of Galois theory, we let

$$\Psi : \mathrm{Int}(\overline{\mathrm{F}}/F) \to \mathrm{Sub}(\mathrm{Aut}_F(\overline{\mathrm{F}})), \quad \Psi(E) := \mathrm{Aut}_E(\overline{\mathrm{F}}), \tag{34.1}$$

and

$$\Phi : \mathrm{Sub}(\mathrm{Aut}_F(\overline{\mathrm{F}})) \to \mathrm{Int}(\overline{\mathrm{F}}/F), \quad \Phi(H) := \mathrm{Fix}(H). \tag{34.2}$$

Proposition 34.1.1. *Suppose F is a field and $\overline{\mathrm{F}}$ is an algebraic closure of F. Then for every $E \in \mathrm{Int}(\overline{\mathrm{F}}/F)$, we have*

$$\mathrm{Fix}(\mathrm{Aut}_E(\overline{\mathrm{F}})) = E.$$

Proof. Clearly $E \subseteq \mathrm{Fix}(\mathrm{Aut}_E(\overline{\mathrm{F}}))$. Suppose to the contrary that α is in $\mathrm{Fix}(\mathrm{Aut}_E(\overline{\mathrm{F}}))$ but not in E. Hence $\deg m_{\alpha,E} \geq 2$. Suppose $\alpha' \in \overline{\mathrm{F}}$ is another zero of $m_{\alpha,E}$. Hence there is an E-isomorphism $\theta : E[\alpha] \to E[\alpha']$. Notice that $\overline{\mathrm{F}}$ is an algebraic closure of $E[\alpha]$ and $E[\alpha']$. Therefore by the isomorphism extension for algebraic closures (see Theorem 33.2.1), there is an isomorphism $\widehat{\theta} : \overline{\mathrm{F}} \to \overline{\mathrm{F}}$ which is an extension of θ. This implies that $\widehat{\theta} \in \mathrm{Aut}_E(\overline{\mathrm{F}})$ and $\widehat{\theta}(\alpha) = \alpha'$. Therefore α is not in $\mathrm{Fix}(\mathrm{Aut}_E(\overline{\mathrm{F}}))$, which is a contradiction. □

By Proposition 34.1.1, we have $\Phi \circ \Psi = \mathrm{id}$. Therefore Ψ and Φ are inverse of each other from $\mathrm{Int}(\overline{\mathrm{F}}/F)$ to the image of Ψ

$$\{\mathrm{Aut}_E(\overline{\mathrm{F}}) \mid E \in \mathrm{Int}(\overline{\mathrm{F}}/F)\}.$$

Later we will see that there are a lot of subgroups of $\mathrm{Aut}_F(\overline{\mathrm{F}})$ which are not in the image of Ψ. One can define a topology on $\mathrm{Aut}_F(\overline{\mathrm{F}})$ called the Krull topology and with respect to this topology the image of Ψ consists of all *closed* subgroups of $\mathrm{Aut}_F(\overline{\mathrm{F}})$. We do not go to the definition of the Krull topology here, but we refer to the image of Ψ as the closed subgroups of $\mathrm{Aut}_F(\overline{\mathrm{F}})$.

Proposition 34.1.2. *Let Ψ and Φ be as in (34.1) and (34.2). Then Ψ and Φ induce bijections between*

$$\{E \in \mathrm{Int}(\overline{\mathrm{F}}/F) \mid E/F \textit{is a normal extension}\}, \textit{ and}$$

$$\{N \in \mathrm{Sub}(\mathrm{Aut}_F(\overline{\mathrm{F}})) \mid N \textit{ is a normal closed subrgoup}\}.$$

Moreover if E/F is a normal extension, then

$$\frac{\mathrm{Aut}_F(\overline{\mathrm{F}})}{\mathrm{Aut}_E(\overline{\mathrm{F}})} \simeq \mathrm{Aut}_F(E).$$

Proof. Suppose E/F is a normal extension. We want to show that $\mathrm{Aut}_E(\overline{\mathrm{F}})$ is a normal subgroup of $\mathrm{Aut}_F(\overline{\mathrm{F}})$. For every $\theta \in \mathrm{Aut}_F(\overline{\mathrm{F}})$ and $\alpha \in E$, $\theta(\alpha)$ is another zero of $m_{\alpha,F}$. Since E/F is a normal extension, $\theta(\alpha)$ is in E. Therefore the restriction map

$$r_E : \mathrm{Aut}_F(\overline{\mathrm{F}}) \to \mathrm{Aut}_F(E),\ r_E(\theta) := \theta|_E$$

is a well-defined function. It is clear that r_E is a group homomorphism. By definition, we have that $\ker r_E = \mathrm{Aut}_E(\overline{\mathrm{F}})$. Therefore $\mathrm{Aut}_E(\overline{\mathrm{F}})$ is a normal subgroup of $\mathrm{Aut}_F(\overline{\mathrm{F}})$. Notice that $\overline{\mathrm{F}}$ is an algebraic closure of E as well. Hence by the isomorphism extension theorem for algebraic closures (see Theorem 33.2.1), every $\theta \in \mathrm{Aut}_F(E)$ can be extended to an element $\widehat{\theta} \in \mathrm{Aut}_F(\overline{\mathrm{F}})$. This means r_E is surjective. Thus by the first isomorphism theorem for groups, we have

$$\frac{\mathrm{Aut}_F(\overline{\mathrm{F}})}{\mathrm{Aut}_E(\overline{\mathrm{F}})} \simeq \mathrm{Aut}_F(E).$$

Next let's assume that N is a normal subgroup of $\mathrm{Aut}_F(\overline{\mathrm{F}})$. We want to show that $\mathrm{Fix}(N)/F$ is a normal extension. Suppose $\alpha \in \mathrm{Fix}(N)$. We want to show $m_{\alpha,F}$ can be decomposed into linear factors in $(\mathrm{Fix}(N))[x]$. As $\overline{\mathrm{F}}$ is algebraically closed, there are

$$\alpha_1 := \alpha, \alpha_2, \dots, \alpha_n \in \overline{\mathrm{F}}$$

such that

$$m_{\alpha,F}(x) = (x-\alpha_1)\cdots(x-\alpha_n).$$

Hence, for all i, by Lemma 16.2.2, there is an F-isomorphism $\theta_i : F[\alpha_1] \to F[\alpha_i]$ such that $\theta_i(\alpha_1) = \alpha_i$. Notice that $\overline{\mathrm{F}}$ is an algebraic closure of $F[\alpha_j]$ for every j. Hence by the isomorphism extension theorem for algebraic closures (see Theorem 33.2.1), there are $\widehat{\theta}_i : \overline{\mathrm{F}} \to \overline{\mathrm{F}}$ such that $\widehat{\theta}_i|_{F[\alpha_1]} = \theta_i$. In particular, $\widehat{\theta}_i \in \mathrm{Aut}_F(\overline{\mathrm{F}})$ and $\widehat{\theta}_i(\alpha_1) = \alpha_i$.

We want to use $\widehat{\theta}_i(\alpha_1) = \alpha_i$, $\alpha_1 \in \mathrm{Fix}(N)$, and $N \trianglelefteq \mathrm{Aut}_F(\overline{\mathrm{F}})$ to show $\alpha_i \in \mathrm{Fix}(N)$.

This means we want to show that $\sigma(\alpha_i) = \alpha_i$ for every $\sigma \in N$. The equality $\sigma(\alpha_i) = \alpha_i$ holds if and only if

$$\sigma(\widehat{\theta}_i(\alpha_1)) = \widehat{\theta}_i(\alpha_1). \tag{34.3}$$

Notice that (34.3) is equivalent to

$$(\widehat{\theta}_i^{-1} \circ \sigma \circ \widehat{\theta}_i)(\alpha_1) = \alpha_1. \tag{34.4}$$

Since N is a normal subgroup of $\mathrm{Aut}_F(\overline{\mathrm{F}})$ and $\sigma \in N$, we have $\widehat{\theta}_i^{-1} \circ \sigma \circ \widehat{\theta}_i \in N$. Therefore (34.4) holds as $\alpha_1 \in \mathrm{Fix}(N)$. This completes the proof. □

34.2 Statement of the cyclic case of Kummer theory and pairing

Inspired by Galois theory, one can believe that group theoretic properties of $\mathrm{Aut}_F(\overline{\mathrm{F}})$, viewed as *external* properties should have *internal* counterparts. In this section, we start our investigation of the following question.

Question 34.2.1. *What can we say about the finite cyclic quotients of* $\mathrm{Aut}_F(\overline{\mathrm{F}})$ *by its closed subgroups? How about the finite abelian quotients?*

We have proved that normal closed subgroups of $\mathrm{Aut}_F(\overline{\mathrm{F}})$ are of the form $\mathrm{Aut}_E(\overline{\mathrm{F}})$ where E/F is a normal extension. Moreover $\frac{\mathrm{Aut}_F(\overline{\mathrm{F}})}{\mathrm{Aut}_E(\overline{\mathrm{F}})} \simeq \mathrm{Aut}_F(E)$. So Question 34.2.1 can b rephrased as follows.

Question 34.2.2. *What are* internal *counterparts of finite cyclic or abelian extensions E/F (where $E \subseteq \overline{\mathrm{F}}$)?*

We have proved that if $\mathrm{Aut}_F(E)$ is a cyclic group of order n, $\mathrm{char}(F) = 0$, and F contains an element ζ of multiplicative order n, then E/F is a Kummer extension; this means $E = F[\sqrt[n]{a}]$ for some $a \in F$ where $\sqrt[n]{a} \in \overline{\mathrm{F}}$ is a zero of $x^n - a$ (see Theorem 31.3.1). Following Kummer, we will be working under the assumption that $\mathrm{char}(F) = 0$ and F has enough roots of unity.

Theorem 34.2.3 (Kummer theory: cyclic case)**.** *Suppose F is a field of characteristic zero and there is $\zeta \in F$ such that $o(\zeta) = n$. Let*

$$\mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F) := \{E \in \mathrm{Int}(\overline{\mathrm{F}}/F) \mid E/F \text{ is Galois and } \mathrm{Aut}_F(E) \hookrightarrow \mathbb{Z}_n\},$$

and

$$\mathrm{Cyc}(F^\times/F^{\times^n}) := \{\langle a(F^{\times^n})\rangle \mid a \in F^\times\}.$$

Then the following functions are inverse of each other

$$\Lambda : \mathrm{Cyc}(F^\times/F^{\times^n}) \to \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F), \;\; \Lambda(\langle a(F^{\times^n})\rangle) := F[\sqrt[n]{a}],$$

and

$$\Delta : \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F) \to \mathrm{Cyc}(F^\times/F^{\times^n}), \;\; \Delta(E) := \frac{E^{\times^n} \cap F^\times}{F^{\times^n}}.$$

Moreover $\langle a(F^{\times^n})\rangle \simeq \mathrm{Aut}_F(F[\sqrt[n]{a}])$ for every $a \in F^\times$, and $\mathrm{Aut}_F(E) \simeq \Delta(E)$ for every $E \in \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$.

Proof of Theorem 34.2.3 is based on the following function which is called a *Kummer pairing*. Let

$$f : \mathrm{Aut}_F(E) \times \Delta(E) \to M_n, \quad f(\sigma, \overline{a}) := \frac{\sigma(\alpha)}{\alpha}, \tag{34.5}$$

where $a := \alpha^n \in F^\times$ for some $\alpha \in E^\times$ and $\overline{a} := \alpha^n(F^{\times n})$ and $M_n := \{1, \zeta, \ldots, \zeta^{n-1}\}$.

Lemma 34.2.4. *Suppose F is a field of characteristic zero which contains an element ζ of order n. Suppose E/F is an algebraic extension and f is the Kummer pairing given in (34.5). Then f is well-defined.*

Proof. We need to address two issues: why $\frac{\sigma(\alpha)}{\alpha}$ is in M_n and why it only depends on $\overline{a} := \alpha^n(F^{\times n})$ and it is independent of the choice of α. Notice that α is a zero of $x^n - a$. Then $\alpha, \zeta\alpha, \ldots, \zeta^{n-1}\alpha$ are distinct zeros of $x^n - a$. By the generalized factor theorem, comparing degrees and leading coefficients, we obtain that

$$x^n - a = (x - \alpha)(x - \zeta\alpha) \cdots (x - \zeta^{n-1}\alpha). \tag{34.6}$$

As $x^n - a \in F[x]$, every $\sigma \in \mathrm{Aut}_F(E)$ permutes zeros of $x^n - a$. Hence $\sigma(\alpha) = \zeta^i\alpha$ for some integer i in $[0, n-1]$. This implies that $\frac{\sigma(\alpha)}{\alpha} \in M_n$.

If $\alpha_1^n(F^{\times n}) = \alpha_2^n(F^{\times n})$ for some $\alpha_1, \alpha_2 \in E^\times$ with $\alpha_i^n \in F$, then there are $a, c \in F^\times$ such that $\alpha_1^n = \alpha_2^n c^n = a \in F^\times$. Hence α_1 and $c\alpha_2$ are two zeros of $x^n - a$. By (34.6), we conclude that $\alpha_1 = \zeta^i c\alpha_2$ for some integer i in $[0, n-1]$. Therefore

$$\frac{\sigma(\alpha_1)}{\alpha_1} = \frac{\sigma(\zeta^i c\alpha_2)}{\zeta^i c\alpha_2} = \frac{\zeta^i c\sigma(\alpha_2)}{\zeta^i c\alpha_2} = \frac{\sigma(\alpha_2)}{\alpha_2}.$$

This completes the proof. □

Lemma 34.2.5. *Suppose F is a field of characteristic zero which contains an element ζ of order n. Suppose E/F is an algebraic extension and f is the Kummer pairing given in (34.5). Then f is a group homomorphism with respect to each component separately.*

Proof. First component. Suppose $\sigma_1, \sigma_2 \in \mathrm{Aut}_F(E)$ and $\overline{a} := a(F^{\times n})$ where $a = \alpha^n$ for some $\alpha \in E^\times$. Suppose

$$f(\sigma_1, \overline{a}) = \zeta^{i_1} \text{ and } f(\sigma_2, \overline{a}) = \zeta^{i_2}. \tag{34.7}$$

We want to show

$$f(\sigma_1 \circ \sigma_2, \overline{a}) = \zeta^{i_1} \cdot \zeta^{i_2} = \zeta^{i_1+i_2}. \tag{34.8}$$

By (34.7), we have $\sigma_1(\alpha) = \zeta^{i_1}\alpha$ and $\sigma_2(\alpha) = \zeta^{i_2}\alpha$. Hence

$$\sigma_1 \circ \sigma(\alpha) = \sigma_1(\zeta^{i_2}\alpha) = \zeta^{i_2}\sigma_1(\alpha) = \zeta^{i_2} \cdot \zeta^{i_1}\alpha = \zeta^{i_1+i_2}\alpha.$$

Therefore $\frac{\sigma_1 \circ \sigma_2(\alpha)}{\alpha} = \zeta^{i_1+i_2}$ which implies (34.8), and this shows that f is a group homomorphism with respect to the first component.

Second component. Suppose $\sigma \in \mathrm{Aut}_F(E)$ and, for $i = 1, 2$, $\overline{a}_i := a_i(F^{\times n})$ where $a_i = \alpha_i^n$ for some $\alpha_i \in E^\times$. Let

$$f(\sigma, \overline{a}_1) = \zeta^{i_1} \text{ and } f(\sigma, \overline{a}_2) = \zeta^{i_2}. \tag{34.9}$$

We want to show

$$f(\sigma, \overline{a}_1 \cdot \overline{a}_2) = \zeta^{i_1+i_2}. \tag{34.10}$$

By (34.9), we have $\sigma(\alpha_1) = \zeta^{i_1}\alpha_1$ and $\sigma(\alpha_2) = \zeta^{i_2}\alpha_2$. Hence

$$\sigma(\alpha_1\alpha_2) = \sigma(\alpha_1)\sigma(\alpha_2) = \zeta^{i_1}\alpha_1 \cdot \zeta^{i_2}\alpha_2 = \zeta^{i_1+i_2}\alpha_1\alpha_2.$$

This implies that $\frac{\sigma(\alpha_1\alpha_2)}{\alpha_1\alpha_2} = \zeta^{i_1+i_2}$ and (34.10) follows. This completes the proof. □

By Lemma 34.2.5, we conclude that for $\sigma_0 \in \mathrm{Aut}_F(E)$ and $\overline{a}_0 \in \Delta(E)$,

$$f_{\overline{a}_0} : \mathrm{Aut}_F(E) \to M_n, \quad f_{\overline{a}_0}(\sigma) := f(\sigma, \overline{a}_0) \tag{34.11}$$

and

$$f^{\sigma_0} : \Delta(E) \to M_n, \quad f^{\sigma_0}(\overline{a}) := f(\sigma_0, \overline{a}) \tag{34.12}$$

are group homomorphisms.

34.3 Functions in the cyclic case of Kummer theory

In this section, we prove that the functions Λ and Δ given in Theorem 34.2.3 are well-defined. To show $\Lambda(\langle a(F^{\times n})\rangle) := F[\sqrt[n]{a}]$ is a well-defined function, we need to address two issues: $F[\sqrt[n]{a}] \in \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$ and $F[\sqrt[n]{a}]$ depends only on the cyclic group $\langle a(F^{\times n})\rangle$ and it is independent of the choice of a. By Proposition 28.3.2, $F[\sqrt[n]{a}]$ is in $\mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$. We go over its prove. By (34.6) and having $\zeta \in F$, we conclude that $F[\sqrt[n]{a}]$ is a splitting field of $x^n - a$ over F. This implies that $F[\sqrt[n]{a}]/F$ is a normal extension. As $\mathrm{char}(F) = 0$, $F[\sqrt[n]{a}]/F$ is a separable extension. Hence $F[\sqrt[n]{a}]/F$ is a Galois extension. Let $\overline{a} := a(F^{\times n})$. By (34.11), $f_{\overline{a}}(\sigma) := \frac{\sigma(\sqrt[n]{a})}{\sqrt[n]{a}}$ is a well-defined group homomorphism.

Lemma 34.3.1. *Suppose F is a field of characteristic zero and it has an element ζ of order n. Suppose $a \in F$, $\sqrt[n]{a}$ is a zero of $x^n - a$ and $E := F[\sqrt[n]{a}]$. Let*

$$f_{\overline{a}} : \mathrm{Aut}_F(F[\sqrt[n]{a}]) \to M_n, f_{\overline{a}}(\sigma) := \frac{\sigma(\sqrt[n]{a})}{\sqrt[n]{a}},$$

where $\overline{a} := a(F^{\times n})$. Then $f_{\overline{a}}$ is a well-defined injective group homomorphism. In particular, $F[\sqrt[n]{a}] \in \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$.

Proof. By Lemma 34.2.4 and (34.11), $f_{\overline{a}}$ is a well-defined group homomorphism. Suppose $f_{\overline{a}}(\sigma) = 1$ for some $\sigma \in \mathrm{Aut}_F(E)$. Then $\sigma(\sqrt[n]{a}) = \sqrt[n]{a}$, which implies that $\sigma = \mathrm{id}$. This completes the proof. □

Lemma 34.3.2. *Suppose F is a field of characteristic zero and it has an element ζ of order n. Suppose $a_1, a_2 \in F^\times$ and $\langle a_1({F^\times}^n)\rangle = \langle a_2({F^\times}^n)\rangle$. Then*

$$F[\sqrt[n]{a_1}] = F[\sqrt[n]{a_2}].$$

Proof. Because $\langle a_1({F^\times}^n)\rangle = \langle a_2({F^\times}^n)\rangle$, $a_1 = a_2^i c^n$ for some $c \in F^\times$ and integer $i \in [0, n-1]$. Because $\zeta \in F$, $\sqrt[n]{a_1} \in F[\sqrt[n]{a_2}]$ which implies that $F[\sqrt[n]{a_1}] \subseteq F[\sqrt[n]{a_2}]$. By symmetry we have $F[\sqrt[n]{a_2}] \subseteq F[\sqrt[n]{a_1}]$. This completes the proof. □

By Lemma 34.3.1 and Lemma 36.2.2, we deduce that the function Λ given in Theorem 34.2.3 is well-defined.

Next we want to show that the function Δ given in Theorem 34.2.3 is well-defined. Here we need to address only one issue: why $\Delta(E)$ is a cyclic group if $E \in \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$.

Lemma 34.3.3. *Suppose F is a field of characteristic zero and it has an element of order n. Suppose $\overline{\mathrm{F}}$ is an algebraic closure of F and $E \in \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$. Suppose $\mathrm{Aut}_F(E) = \langle \sigma_0 \rangle$ and $f^{\sigma_0} : \Delta(E) \to M_n$ is given in (34.12). Then f^{σ_0} is an injective group homomorphism and $\Delta(E)$ is cyclic.*

Proof. By Lemma 34.2.5 and (34.12), f^{σ_0} is a group homomorphism. Suppose $f^{\sigma_0}(\overline{a}) = 1$, $\overline{a} = a({F^\times}^n)$, $a = \alpha^n$ for some $\alpha \in E^\times$. Then $f^{\sigma_0}(\overline{a}) = \frac{\sigma_0(\alpha)}{\alpha}$. Therefore $\sigma_0(\alpha) = \alpha$, which implies that $\alpha \in \mathrm{Fix}(\langle \sigma_0 \rangle)$. As E/F is Galois and $\mathrm{Aut}_F(E) = \langle \sigma_0 \rangle$, by Theorem 24.2.2 we have $F = \mathrm{Fix}(\langle \sigma_0 \rangle)$. Hence $\alpha \in F$, which implies that $a = \alpha^n \in {F^\times}^n$. Thus $\overline{a} = a({F^\times}^n) = \overline{1}$. We conclude that $\ker f^{\sigma_0} = \{\overline{1}\}$, and so f^{σ_0} is injective. Therefore $\Delta(E)$ can be embedded into the cyclic group M_n. As all the subgroups of a cyclic subgroup is cyclic, we deduce that $\Delta(E)$ is cyclic. This completes the proof. □

Chapter 35

Lecture 11

35.1 Kummer theory: the cyclic case

The main goal of this section is to prove Theorem 34.2.3. In Section 34.3, we have proved that

$$\Lambda : \mathrm{Cyc}(F^\times/F^{\times n}) \to \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F), \ \ \Lambda(\langle a(F^{\times n})\rangle) := F[\sqrt[n]{a}],$$

and

$$\Delta : \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F) \to \mathrm{Cyc}(F^\times/F^{\times n}), \ \ \Delta(E) := \frac{E^{\times n} \cap F^\times}{F^{\times n}}$$

are well-defined. Next we want to prove that $\Delta \circ \Lambda = \mathrm{id}$, which is equivalent to

$$\frac{F[\sqrt[n]{a}]^{\times n} \cap F^\times}{F^{\times n}} = \langle a(F^{\times n})\rangle. \tag{35.1}$$

As $(\sqrt[n]{a})^n \in F^{\times\times}$, $a(F^{\times n})$ is in the left hand side of (35.1). Hence

$$\frac{F[\sqrt[n]{a}]^{\times n} \cap F^\times}{F^{\times n}} \supseteq \langle a(F^{\times n})\rangle. \tag{35.2}$$

Suppose $\bar{b} := b(F^{\times n}) \in \Delta(F[\sqrt[n]{a}])$. We want to show that $\bar{b}$ is in $\langle a(F^{\times n})\rangle$. By definition of $\Delta(F[\sqrt[n]{a}])$, there is $\beta \in F[\sqrt[n]{a}]^\times$ such that $\beta^n(F^{\times n}) = b(F^{\times n})$. Hence by Lemma 36.2.2, $F[\beta] = F[\sqrt[n]{b}]$. Then $F[\sqrt[n]{b}] \in \mathrm{Int}(F[\sqrt[n]{a}]/F)$. Notice that by Lemma 34.3.1, $F[\sqrt[n]{b}], F[\sqrt[n]{a}] \in \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$ and

$$f_{\bar{b}} : \mathrm{Aut}_F(F[\sqrt[n]{b}]) \to M_n \quad \text{and} \quad f_{\bar{a}} : \mathrm{Aut}_F(F[\sqrt[n]{a}]) \to M_n \tag{35.3}$$

are injective group homomorphisms. On the there hand, by the fundamental theorem of Galois theory, if σ_0 generates $\mathrm{Aut}_F(F[\sqrt[n]{a}])$, then $\sigma_0|_{F[\sqrt[n]{b}]}$ generates $\mathrm{Aut}_F(F[\sqrt[n]{b}])$. Hence

$$\mathrm{Im}\, f_{\bar{b}} = \langle f_{\bar{b}}(\sigma_0|_{F[\sqrt[n]{b}]})\rangle \quad \text{and} \quad \mathrm{Im}\, f_{\bar{a}} = \langle f_{\bar{a}}(\sigma_0)\rangle, \tag{35.4}$$

and by (35.3), $|\,\mathrm{Im}\, f_{\bar{b}}| = |\langle\sigma_0|_{F[\sqrt[n]{b}]}\rangle|$ divides $|\,\mathrm{Im}\, f_{\bar{a}}| = |\langle\sigma_0\rangle|$. As $\mathrm{Im}\, f_{\bar{b}}$ and $\mathrm{Im}\, f_{\bar{a}}$ are subgroups of the cyclic group M_n and the order of $\mathrm{Im}\, f_{\bar{b}}$ divides the order of $\mathrm{Im}\, f_{\bar{a}}$, we conclude that $\operatorname{Im} f_{\overline{b}} \subseteq \operatorname{Im} f_{\overline{a}}$, which together with (35.4) implies that

$$f_{\overline{b}}(\sigma_0|_{F[\sqrt[n]{b}]}) = f_{\overline{a}}(\sigma_0)^i \tag{35.5}$$

for some non-negative integer i. By (35.5), we obtain that $\frac{\sigma_0(\sqrt[n]{b})}{\sqrt[n]{b}} = \left(\frac{\sigma_0(\sqrt[n]{a})}{\sqrt[n]{a}}\right)^i$. Hence $\sigma_0\left(\frac{\sqrt[n]{b}}{\sqrt[n]{a}^i}\right) = \frac{\sqrt[n]{b}}{\sqrt[n]{a}^i}$, which means that $\frac{\sqrt[n]{b}}{\sqrt[n]{a}^i} \in \operatorname{Fix}(\langle\sigma_0\rangle)$. As $F[\sqrt[n]{a}]/F$ is a Galois,

$$\operatorname{Fix}(\operatorname{Aut}_F(F[\sqrt[n]{a}]) = \operatorname{Fix}(\langle\sigma_0\rangle) = F.$$

Altogether, we conclude that $\sqrt[n]{b} = c\,\sqrt[n]{a}^i$ for some $c \in F^\times$. Therefore

$$\overline{b} = b(F^{\times n}) = a^i(F^{\times n}) \in \langle a(F^{\times n})\rangle. \tag{35.6}$$

By (35.6) and (35.2), we obtain that $\Delta \circ \Lambda = \mathrm{id}$.

Since $\Delta \circ \Lambda = \mathrm{id}$, Λ and Δ induce a bijection between the domain of Λ and the image of Λ. Hence to show Λ and Δ are inverse of each other it is enough to show that Λ is surjective.

For every $E \in \operatorname{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$, by Theorem 31.3.1, $E = F[\sqrt[n]{a}]$ for some $a \in F$, which means $E = \Lambda(\langle a(F^{\times n})\rangle)$. Therefore Λ is surjective, and so $\Lambda \circ \Delta = \mathrm{id}$.

Next we want to show that $\operatorname{Aut}_F(E) \simeq \Delta(E)$ for every $E \in \operatorname{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$. Suppose $\operatorname{Aut}_F(E) = \langle\sigma_0\rangle$ and $\Delta(E) = \langle a_0(F^{\times n})\rangle$. Then by Lemma 34.3.1 and Lemma 34.3.3,

$$f_{\overline{a}_0} : \operatorname{Aut}_F(F[\sqrt[n]{a_0}]) \to M_n \quad \text{and} \quad f^{\sigma_0} : \Delta(E) \to M_n \tag{35.7}$$

injective group homomorphisms, where $\overline{a}_0 := a_0(F^{\times n})$. Therefore

$$\begin{aligned} \Delta(E) \simeq& \langle f^{\sigma_0}(\overline{a}_0)\rangle = \langle f(\sigma_0, \overline{a}_0)\rangle \qquad \text{and} \\ \operatorname{Aut}_F(E) \simeq& \langle f_{\overline{a}_0}(\sigma_0)\rangle = \langle f(\sigma_0, \overline{a}_0)\rangle. \end{aligned} \tag{35.8}$$

By (35.8), we conclude that $\Delta(E) \simeq \operatorname{Aut}_F(E)$. Finally, we have

$$\operatorname{Aut}_F(F[\sqrt[n]{a}]) \simeq \Delta(F[\sqrt[n]{a}]) = \Delta \circ \Lambda(\langle a(F^{\times n})\rangle) = \langle a(F^{\times n})\rangle,$$

which completes the proof of Theorem 34.2.3.

35.2 Dual of abelian groups

Our next goal is to prove the abelian case of Kummer theory. To prove this result, we need to study *dual* of finite abelian groups and show further properties of the Kummer pairing. In this section, we introduce the dual of a finite abelian group, and prove some of its basic properties. It is worth pointing out that some of these properties can be proved using classification of finite abelian groups in a more straightforward fashion. Here, however, we present an approach that avoids using the classification of finite abelian groups.

Definition 35.2.1. *Suppose A is a finite abelian group. We let*

$$\widehat{A} := \mathrm{Hom}(A, S^1)$$

where $S^1 := \{z \in \mathbb{C} \mid |z| = 1\}$, and we call $\widehat{A}$ the dual *of A. Elements χ of $\widehat{A}$ are called* characters *of A.*

The dual $\widehat{A}$ of a finite abelian group A is a group itself under the *pointwise multiplication*.

Lemma 35.2.2. *Suppose A is a finite abelian group, and $\widehat{A}$ is the dual of A. For $\chi_1, \chi_2 \in \widehat{A}$ and $a \in A$, we let $(\chi_1 \cdot \chi_2)(a) := \chi_1(a)\chi_2(a)$. Then $(\widehat{A}, \cdot)$ is an abelian group.*

Proof. First we show that $\chi_1 \cdot \chi_2$ is in $\widehat{A}$. For $a_1, a_2 \in A$, we have that

$$\begin{aligned}(\chi_1 \cdot \chi_2)(a_1a_2) =&\chi_1(a_1a_2)\chi_2(a_1a_2)\\ =&\chi_1(a_1)\chi_1(a_2)\chi_2(a_1)\chi_2(a_2)\\ =&(\chi_1(a_1)\chi_2(a_1))(\chi_1(a_2)\chi_2(a_2))\\ =&(\chi_1 \cdot \chi_2)(a_1)(\chi_1 \cdot \chi_2)(a_2).\end{aligned}$$

This means that $\chi_1 \cdot \chi_2 \in \widehat{A}$. Notice that for every $\chi_1, \chi_2 \in \widehat{A}$ and $a \in A$, we have

$$(\chi_1 \cdot \chi_2)(a) = \chi_1(a)\chi_2(a) = \chi_2(a)\chi_1(a) = (\chi_2 \cdot \chi_1)(a),$$

which means $\chi_1 \cdot \chi_2 = \chi_2 \cdot \chi_1$.

Let $\mathbb{1}_A$ be the trivial group homomorphism $\mathbb{1}_A(a) = 1$ for every $a \in A$. Then $(\mathbb{1}_A \cdot \chi)(a) = \mathbb{1}_A(a)\chi(a) = \chi(a)$ for every $a \in A$ and $\chi \in \widehat{A}$. Hence $\mathbb{1}_A \cdot \chi = \chi$ for every $\chi \in \widehat{A}$. As $(\widehat{A}, \cdot)$ is abelian, we conclude that $\mathbb{1}_A$ is the neutral element of $\widehat{A}$.

For every $\chi \in \widehat{A}$, let $\chi^{-1} : A \to S^1, (\chi^{-1})(a) := \chi(a)^{-1}$. Then it is easy to see that χ^{-1} is a group homomorphism, and so $\chi^{-1} \in \widehat{A}$. We also notice that $\chi^{-1} \cdot \chi = \chi \cdot \chi^{-1} = \mathbb{1}_A$. This completes the proof. □

By classification of finite abelian groups, we know that every such group is a direct product of finitely many circle groups. Using this result, we can show that $\widehat{A} \simeq A$ for every finite abelian group A. In this section, without using the classification of finite abelian groups, we prove a slightly weaker result. The following lemma plays an important role in the study of dual of abelian groups. It roughly says that $\widehat{A}$ has enough elements to *separate* elements of A: one can distinguish points of A using the *test functions* from $\widehat{A}$.

Lemma 35.2.3. *Suppose A is a finite abelian group. Then, for every $a \in A \setminus \{1\}$, there is $\chi \in \widehat{A}$ such that $\chi(a) \neq 1$.*

Proof. Suppose $o(a) = d$. Let $\zeta_d \in S^1$ be an element of order d (say $\zeta_d := e^{\frac{2\pi i}{d}}$). As $\langle a \rangle$ and $\langle \zeta_d \rangle$ are cyclic groups of order d, $\overline{\chi} : \langle a \rangle \to S^1, \overline{\chi}(a^i) := \zeta_d^i$ for every integer

i is an injective group homomorphism (and $\operatorname{Im}\overline{\chi}$) is the cyclic group $\langle\zeta_d\rangle$). We will show that $\overline{\chi}$ has an extension $\chi \in \widehat{A}$; in particular, $\chi(a) = \overline{\chi}(a) = \zeta_d \neq 1$. Let

$$\Sigma := \{(H, \chi) \mid H \leq A, a \in H, \chi \in \widehat{H}, \chi|_{\langle a\rangle} = \overline{\chi}\}.$$

Notice that $(\langle a\rangle, \overline{\chi}) \in \Sigma$, and so Σ is non-empty. Let's also point out that Σ is a finite set. This is the case because for every $\chi \in \widehat{A}$ and $a' \in A$,

$$\chi(a')^{|A|} = \chi(a'^{|A|}) = \chi(1) = 1,$$

and so $\chi(a')$ has at most $|A|$-many possibilities for every $a' \in A$. Our goal is to show that there is (A, χ) in Σ; this means we are looking for a *largest possible* element of in Σ. To make sense of this, we introduce the following partial order on Σ. For $(H_1, \chi_1), (H_2, \chi_2) \in \Sigma$, we say $(H_1, \chi_1) \preccurlyeq (H_2, \chi_2)$ if $H_1 \leq H_2$ and $\chi_2|_{H_2}$. It is easy to see that Σ is a poset with respect to $\preccurlyeq$. As Σ is a non-empty finite set, it has a maximal element. Suppose (H, χ) is a maximal element of Σ. We can finish the proof by showing that $H = A$.

Suppose to the contrary that there is $a' \in A \setminus H$. We want to extend χ to an element $\chi' \in \widehat{\langle H, a'\rangle}$ where $\langle H, a'\rangle$ is the group generated by a and H. Notice that if we manage to extend χ to a character of $\langle H, a'\rangle$, this would contradict the maximality of (H, χ) in Σ.

To extend χ to a character of $\langle H, a'\rangle$, we start by describing elements of $\langle H, a'\rangle$. Since A is abelian, every subgroup is normal. So we can consider the quotient group A/H. As H is in the kernel of the quotient map $p_H : A \to A/H$, we have that

$$p_H(\langle H, a'\rangle) = \langle p_H(a')\rangle = \langle a'H\rangle = \{a'^i H \mid i = 0, 1, \dots, d-1\} \tag{35.9}$$

where d is the order of $a'H$ in A/H. Because cosets of H in $\langle H, a'\rangle$ is a partition of $\langle H, a'\rangle$, we conclude that every element of $\langle H, a'\rangle$ can be written as $a'^i h$ for some integer i in $[0, d-1]$ and $h \in H$.

Next we notice that to find an extension χ' of χ in $\widehat{\langle H, a'\rangle}$, we only need to find a suitable choice for $\chi'(a')$. Since $p_H(a')^d = p_H(1)$, $a'^d = h_0$ for some $h_0 \in H$. Therefore for every possible extension χ' of χ, we have

$$\chi'(a')^d = \chi'(a'^d) = \chi'(h_0) = \chi(h_0). \tag{35.10}$$

Hence $\chi'(a')$ should be a zero of $x^d - \chi(h_0)$. Since h_0 is of finite order, $\chi(h_0)$ is a root of unity and so all the zeros of $x^d - \chi(h_0)$ are also roots of unity; in particular they are in S^1. Suppose $\zeta \neq 1$ is a zero of $x^d - \chi(h_0)$, and let

$$\chi'(a^i h) := \zeta^i \chi(h)$$

for every integer i and $h \in H$. We want to show that χ' is a well-defined extension of χ.

Suppose $a'^{i_1} h_1 = a'^{i_2} h_2$ for some integers i_1, i_2 and $h_1, h_2 \in H$. We have to show that $\zeta^{i_1}\chi(h_1) = \zeta^{i_2}\chi(h_2)$. Because $a'^{i_1} h_1 = a'^{i_2} h_2$, we have $a'^{i_1 - i_2} = h_2 h_1^{-1}$.

Therefore $p_H(a')^{i_1-i_2} = p_H(1)$ as $h_2h_1^{-1} \in H$. This implies that the order of $p_H(a')$ divides $i_1 - i_2$, and so $\frac{i_1-i_2}{d}$ is an integer. Hence

$$h_2h_1^{-1} = a'^{i_1-i_2} = (a'^d)^{\frac{i_1-i_2}{d}} = h_0^{\frac{i_1-i_2}{d}}. \tag{35.11}$$

By (35.11), we conclude that

$$\chi(h_2h_1^{-1}) = \chi(h_0^{\frac{i_1-i_2}{d}}) = \chi(h_0)^{\frac{i_1-i_2}{d}} = (\zeta^d)^{\frac{i_1-i_2}{d}} = \zeta^{i_1-i_2}.$$

Hence $\zeta^{i_1}\chi(h_1) = \zeta^{i_2}\chi(h_2)$, which implies that χ' is well-defined.

Next we check that χ' is a group homomorphism. Suppose $g, g' \in \langle a, H\rangle$. Then $g = a^i h$ and $g' = a^{i'}h'$ for some integers i, i' and $h, h' \in H$. Therefore

$$\chi'(gg') = \chi'((a^ih)(a^{i'}h')) = \chi'(a^{i+i'}hh') = \zeta^{i+i'}\chi(hh'), \tag{35.12}$$

and

$$\chi'(g)\chi'(g') = \chi'(a^ih)\chi'(a^{i'}h') = (\zeta^i\chi(h))(\zeta^{i'}\chi(h')) = \zeta^{i+i'}\chi(h)\chi(h'). \tag{35.13}$$

By (35.12) and (35.13), we conclude that χ' is a group homomorphism. Clearly $\chi'(h) = \chi(h)$ for every $h \in H$ and $\chi'(a) = \zeta \neq 1$. Altogether, we have $(\langle a, H\rangle, \chi') \in \Sigma$, $(H, \chi) \preccurlyeq (\langle a, H\rangle, \chi')$, and $H \neq \langle a, H\rangle$. This is a contradiction as (H, χ) is a maximal element of Σ. This completes the proof. □

Chapter 36

Lecture 12

36.1 Dual of abelian groups

Lemma 35.2.3 shows us that $\widehat{A}$ has *plenty* of elements to separate points of A. The next lemma says that the order of $\widehat{A}$ cannot be more than the order of A.

Lemma 36.1.1. *Suppose A is a finite abelian group and $\widehat{A}$ is its dual. Then $|\widehat{A}| \leq |A|$.*

Proof. We proceed by strong induction on $|A|$. Since A is a finite group, it has a maximal subgroup M. As A is abelian, every subgroup is normal and we can consider the quotient group A/M. Because of the bijection between subgroups of A/M and subgroups of A that contain M as a subgroup (see Proposition 30.1.2)and maximality of M, we deduce that A/M does not have a non-trivial proper subgroup. Hence A/M is a cyclic group of prime order (see Lemma 30.1.3); say $|A/M| = p$. Suppose $A/M = \langle aM \rangle$ and $a^p = b$; then $b \in M$. Every $\chi \in \widehat{A}$ is uniquely determined by $\chi|_M$ and $\chi(a)$. Notice that $\chi|_M \in \widehat{M}$. Suppose $\chi|M = \overline{\chi} \in \widehat{M}$; then

$$\chi(a)^p = \chi(a^p) = \chi(b) = \overline{\chi}(b).$$

Hence $\chi(a)$ is a zero of $x^p - \overline{\chi}(b)$. This means for a given $\overline{\chi}$, $\chi(a)$ has at most p possibilities. Altogether we conclude that

$$\begin{aligned}
|\widehat{A}| = \sum_{\chi \in \widehat{A}} 1 &= \sum_{\overline{\chi} \in \widehat{M}} \Big(\sum_{\zeta \text{ is a zero of } x^p - \overline{\chi}(b)} \sum_{\chi \in \widehat{A}, \chi|_M = \overline{\chi}, \chi(a) = \zeta} 1 \Big) \\
&\leq \sum_{\overline{\chi} \in \widehat{M}} \Big(\sum_{\zeta \text{ is a zero of } x^p - \overline{\chi}(b)} 1 \Big) \leq \sum_{\overline{\chi} \in \widehat{M}} p = p|\widehat{M}| \\
&\leq p|M| = |A|,
\end{aligned}$$

where the last inequality holds because of the strong induction hypothesis. This completes the proof. ☐

Now we are ready to prove our main result on dual of abelian groups.

Theorem 36.1.2. *Suppose A is a finite abelian group and $\widehat{A}$ is its dual. Then $|\widehat{A}| = |A|$ and*

$$\ell : A \to \widehat{\widehat{A}}, \quad (\ell(a))(\chi) := \chi(a)$$

is an isomorphism.

Recall that $\widehat{A}$ consists of functions on A, and $\ell(a)$ can be viewed as an evaluation at a map.

Proof. First we start by showing that ℓ is a well-defined function. For every $a \in A$, let $\ell_a : \widehat{A} \to S^1$ be the evaluation at a map; that means

$$\ell_a(\chi) := \chi(a).$$

We claim that ℓ_a is a group homomorphism and so it belongs to the dual of $\widehat{A}$. For every $\chi_1, \chi_2 \in \widehat{A}$, we have

$$\ell_a(\chi_1 \cdot \chi_2) = (\chi_1 \cdot \chi_2)(a) = \chi_1(a)\chi_2(a) = \ell_a(\chi_1)\ell_a(\chi_2),$$

and so $\ell_a \in \widehat{\widehat{A}}$. Next we show that $a \mapsto \ell_a$ is a group homomorphism from A to $\widehat{\widehat{A}}$. For every $a_1, a_2 \in A$ and $\chi \in \widehat{A}$, we have

$$\ell_{a_1 a_2}(\chi) = \chi(a_1 a_2) = \chi(a_1)\chi(a_2) = \ell_{a_1}(\chi)\ell_{a_2}(\chi) = (\ell_{a_1} \cdot \ell_{a_2})(\chi),$$

and so $\ell_{a_1 a_2} = \ell_{a_1} \cdot \ell_{a_2}$ for every $a_1, a_2 \in A$. This implies that

$$\ell : A \to \widehat{\widehat{A}}, \quad \ell(a) := \ell_a$$

is a group homomorphism. At this step, we show that ℓ is injective. To show this, it is necessary and sufficient that $\ell(a) \neq \mathbb{1}_{\widehat{A}}$ if $a \neq 1$. For $a \neq 1$, by Lemma 35.2.3, there is $\chi \in \widehat{A}$ such that $\chi(a) \neq 1$. Hence $\ell_a(\chi) \neq 1$, which means $(\ell(a))(\chi) \neq 1$. Thus $\ell(a) \neq \mathbb{1}_{\widehat{A}}$, which implies that ℓ is injective. In particular, we have

$$|A| \leq |\widehat{\widehat{A}}|. \tag{36.1}$$

On the other hand, by Lemma 36.1.1, we have

$$|\widehat{\widehat{A}}| \leq |\widehat{A}| \leq |A|. \tag{36.2}$$

By (36.1) and (36.2), we conclude that

$$|A| = |\widehat{A}| = |\widehat{\widehat{A}}|. \tag{36.3}$$

As $\ell : A \to \widehat{\widehat{A}}$ is an injective group homomorphism and $|A| = |\widehat{A}|$, we deduce that ℓ is an isomorphism. This completes the proof. □

We finish our discussion of dual of abelian groups by looking at the characters of an abelian group of exponent n.

Suppose A is a finite abelian group whose exponent divides n; that means $a^n = 1$ for every $a \in A$. Then for every $\chi \in \widehat{A}$ and $a \in A$, we have

$$\chi(a)^n = \chi(a^n) = \chi(1) = 1,$$

and so $\chi(a) = e^{\frac{2\pi k i}{n}}$ for some integer k. This means $\widehat{A}$ is the same as $\mathrm{Hom}(A, \langle \zeta_n \rangle)$ where $\zeta_n := e^{\frac{2\pi i}{n}}$. Therefore, if F is a field which contains an element ζ of order n, then for every abelian group A of exponent n the dual $\widehat{A}$ of A can be identified with $\mathrm{Hom}(A, M_n)$ where $M_n := \langle \zeta \rangle$.

Let's also observe that if A is a finite abelian group whose exponent divides n, then the exponent of $\widehat{A}$ is also a divisor of n. This holds because

$$(\chi^n)(a) = \chi(a)^n = \chi(a^n) = \chi(1) = 1$$

for every $a \in A$.

36.2 Statement of finite abelian case of Kummer theory

In this section, we state the finite abelian case of Kummer theory and prove that the involved functions are well-defined. In this section F is a field and $\overline{\mathrm{F}}$ is an algebraic closure of F. Let $\mathrm{Int}_{\mathrm{ab},n}(\overline{\mathrm{F}}/F)$ be the set of all $E \in \mathrm{Int}(\overline{\mathrm{F}}/F)$ such that E/F is a finite Galois extension and $\mathrm{Aut}_F(E)$ is an abelian group of exponent n, and

$$\mathrm{Sub}_{\mathrm{f}}(F^\times/{F^\times}^n) := \{\overline{A} \leq F^\times/{F^\times}^n \mid \overline{A} \text{ is finite}\}.$$

Theorem 36.2.1 (Kummer theory: abelian case)**.** *Suppose F is a field of characteristic zero which contains an element of order n. Suppose $\overline{\mathrm{F}}$ is an algebraic closure of F. Let*

$$\Lambda : \mathrm{Sub}_{\mathrm{f}}(F^\times/{F^\times}^n) \to \mathrm{Int}_{\mathrm{ab},n}(\overline{\mathrm{F}}/F), \quad \Lambda(\overline{A}) := F[\sqrt[n]{a_i} \mid a_i({F^\times}^n) \in \overline{A}]$$

where $\sqrt[n]{a_i} \in \overline{\mathrm{F}}$ is a zero of $x^n - a_i$, and

$$\Delta : \mathrm{Int}_{\mathrm{ab},n}(\overline{\mathrm{F}}/F) \to \mathrm{Sub}_{\mathrm{f}}(F^\times/{F^\times}^n), \quad \Delta(E) := ({E^\times}^n \cap F^\times)/{F^\times}^n.$$

Then Λ and Δ are inverse of each other. Moreover for every $E \in \mathrm{Int}_{\mathrm{ab},n}(\overline{\mathrm{F}}/F)$,

$$\mathrm{Aut}_F(E) \simeq \widehat{\Delta(E)}.$$

In the proof, we give a concrete isomorphism between $\mathrm{Aut}_F(E)$ and $\widehat{\Delta(E)}$. We start by showing that Λ and Δ are well-defined.

Lemma 36.2.2. *In the above setting, Λ is well-defined.*

Proof. To show Λ is well-defined, we have to show why $F[\sqrt[n]{a_i} \mid a_i({F^\times}^n) \in \overline{A}]$ is in $\mathrm{Int}_{\mathrm{ab},n}(\overline{\mathrm{F}}/F)$, why it does not depend on the choice of coset representatives a_i's and $\sqrt[n]{a_i}$.

Suppose $\overline{A} := \{a_1(F^{\times n}), \dots, a_m(F^{\times n})\}$ and $\sqrt[n]{a_i} \in \overline{\mathbb{F}}$ is a zero of $x^n - a_i$. Then $g(x) := \prod_{i=1}^m (x^n - a_i)$ can be written as

$$g(x) = \prod_{i=1}^{m} \prod_{j=0}^{n-1} (x - \zeta^j \sqrt[n]{a_i}).$$

Hence $F[\zeta^j \sqrt[n]{a_i} \mid 1 \le i \le m, 0 \le j < n]$ is a splitting field of g over F. Notice that

$$F[\zeta^j \sqrt[n]{a_i} \mid 1 \le i \le m, 0 \le j < n] = F[\sqrt[n]{a_1}, \dots, \sqrt[n]{a_m}]$$

as $\zeta \in F$. Hence $F[\sqrt[n]{a_1}, \dots, \sqrt[n]{a_m}]/F$ is a finite normal extension. Since $\mathrm{char}(F) = 0$, every algebraic extension of F is separable. Hence $F[\sqrt[n]{a_1}, \dots, \sqrt[n]{a_m}]/F$ is a finite Galois extension. Let $E := F[\sqrt[n]{a_1}, \dots, \sqrt[n]{a_m}]$ and $\overline{a}_i := a_i(F^{\times n})$. By the Kummer pairing and (34.11),

$$f_{\overline{A}} : \mathrm{Aut}_F(E) \to M_n \times \cdots \times M_n, \quad f_{\overline{A}}(\sigma) := (f_{\overline{a}_1}(\sigma), \dots, f_{\overline{a}_m}(\sigma))$$

is a group homomorphism.

Claim. $f_{\overline{A}}$ is injective.

Proof of Claim. Suppose $f_{\overline{A}}(\sigma) = (1, \dots, 1)$. Then, for every i, $f_{\overline{a}_i}(\sigma) = 1$, which means $\sigma(\sqrt[n]{a_i}) = \sqrt[n]{a_i}$. Since $E = F[\sqrt[n]{a_1}, \dots, \sqrt[n]{a_m}]$, we conclude that $\sigma = \mathrm{id}$. The claim follows. □

By the above Claim, we deduce that $\mathrm{Aut}_F(E)$ can be embedded in $M_n \times \cdots \times M_n$, and so $\mathrm{Aut}_F(E)$ is an abelian group of exponent n.

Finally, let's assume that $a_i(F^{\times n}) = a'_i(F^{\times n})$ for every i. We have to show that

$$F[\sqrt[n]{a_1}, \dots, \sqrt[n]{a_m}] = F[\sqrt[n]{a'_1}, \dots, \sqrt[n]{a'_m}].$$

Since $a_i(F^{\times n}) = a'_i(F^{\times n})$, $a'_i = a_i d_i^n$ for some $d_i \in F^{\times}$. Hence $\sqrt[n]{a'_i}$ and $\sqrt[n]{a_i} d_i$ are zeros of $x^n - a'_i$. As $\zeta \in F$, we obtain that $\sqrt[n]{a'_i} = c_i \sqrt[n]{a_i}$ for some $c_i \in F^{\times}$. Hence

$$F[\sqrt[n]{a'_i} \mid 1 \le i \le m] = F[c_i \sqrt[n]{a_i} \mid 1 \le i \le m] = F[\sqrt[n]{a_i} \mid 1 \le i \le m].$$

This completes the proof. □

Lemma 36.2.3. *In the above setting Δ is well-defined.*

Proof. We need to show that $\Delta(E)$ is a finite group. Suppose

$$\mathrm{Aut}_F(E) := \{\sigma_1, \dots, \sigma_k\}.$$

By the Kummer pairing and (34.12),

$$f^E : \Delta(E) \to M_n \times \cdots \times M_n, \quad f^E(\overline{a}) := (f^{\sigma_1}(\overline{a}), \dots, f^{\sigma_k}(\overline{a})).$$

Claim. f^E is injective.

Proof of Claim. Suppose $f^E(\overline{a}) = (1, \ldots, 1)$ where $\overline{a} = a(F^{\times n})$ and $a = \alpha^n$ for some $\alpha \in E^\times$. Then, for every i, $f^{\sigma_i}(\overline{a}) = 1$, which implies that $\frac{\sigma_i(\alpha)}{\alpha} = 1$. Therefore α is in $\mathrm{Fix}(\mathrm{Aut}_F(E))$. Since E/F is a finite Galois extension, $\mathrm{Fix}(\mathrm{Aut}_F(E)) = F$. Hence $\alpha \in F$. Therefore $\overline{a} = (\alpha^n)(F^{\times n}) = F^{\times n} = \overline{1} \in F^\times / F^{\times n}$. This completes the proof of injectivity of f^E. □

By the above Claim, $\Delta(E)$ can be embedded into $M_n \times \cdots \times M_n$, and so it is finite. This completes the proof. □

36.3 Kummer pairing is a perfect pairing

We have proved that Λ and Δ given in Theorem 36.2.1 are well-defined (see Lemma 36.2.2 and Lemma 36.2.3). The main step to show Λ and Δ are inverse is the following proposition.

Proposition 36.3.1. *Suppose $E \in \mathrm{Int}_{\mathrm{ab},n}(\overline{\mathrm{F}}/F)$ and let f be the Kummer pairing given in (34.5). Then f is a perfect pairing; that means*

$$\widehat{f} : \Delta(E) \to \widehat{\mathrm{Aut}_F(E)}, \quad (\widehat{f}(\overline{a})(\sigma) := f_{\overline{a}}(\sigma) = f(\sigma, \overline{a})$$

is an isomorphism.

Proof. We start by noticing that by (34.11), $f_{\overline{a}} \in \widehat{\mathrm{Aut}_F(E)}$. For every $\overline{a}_1, \overline{a}_2 \in \Delta(E)$,

$$\begin{aligned}\Big(\widehat{f}(\overline{a}_1\overline{a}_2)\Big)(\sigma) =& f(\sigma, \overline{a}_1\overline{a}_2)\\ =& f(\sigma, \overline{a}_1) f(\sigma, \overline{a}_2)\\ =& \Big(\widehat{f}(\overline{a}_1)\Big)(\sigma)\Big(\widehat{f}(\overline{a}_2)\Big)(\sigma),\end{aligned}$$

which means $\widehat{f}(\overline{a}_1\overline{a}_2) = \widehat{f}(\overline{a}_1) \cdot \widehat{f}(\overline{a}_2)$. Hence $\widehat{f}$ is a group homomorphism.

Next, we show that $\widehat{f}$ is injective. Suppose $\widehat{f}(\overline{a}) = \mathbb{1}$. Then for every $\sigma \in \mathrm{Aut}_F(E)$, $(\widehat{f}(\overline{a}))(\sigma) = 1$, which means $f(\sigma, \overline{a}) = 1$. Notice that $f(\sigma, \overline{a}) = \frac{\sigma(\alpha)}{\alpha}$ where $\alpha \in E^\times$ is a zero of $x^n - a$. Hence $\sigma(\alpha) = \alpha$ for every $\sigma \in \mathrm{Aut}_F(E)$. Therefore $\alpha \in \mathrm{Fix}(\mathrm{Aut}_F(E))$, which implies that $\alpha \in F$ as E/F is a finite Galois extension. Thus $\overline{a} = \alpha^n(F^{\times n}) = \overline{1}$. This implies that $\widehat{f}$ is injective.

Finally we want to show that $\widehat{f}$ is surjective. Suppose $\chi \in \widehat{\mathrm{Aut}_F(E)}$. This means $\chi : \mathrm{Aut}_F(E) \to M_n$ is a group homomorphism. Let $N := \ker \chi$. Then by the fundamental theorem of Galois theory, $K := \mathrm{Fix}(N) \in \mathrm{Int}(E/F)$, K/F is a Galois extension, $N = \mathrm{Aut}_K(E)$, and

$$\mathrm{Aut}_F(K) \simeq \frac{\mathrm{Aut}_F(E)}{\mathrm{Aut}_K(E)}. \tag{36.4}$$

As $\mathrm{Aut}_K(E) = \ker \chi$, by the first isomorphism theorem for groups and (36.4) we have

$$\mathrm{Aut}_F(K) \simeq \mathrm{Im}\, \chi \leq M_n. \tag{36.5}$$

By (36.5), $K \in \mathrm{Int}_{\mathbb{Z}_n}(\overline{\mathrm{F}}/F)$. Hence by the surjectivity of Λ from the cyclic case of Kummer theory (see Theorem 34.2.3), we have that $K = F[\sqrt[n]{b}]$ for some $\sqrt[n]{b} \in E$ and $b \in F$. Let $\overline{b} := b(F^{\times n}) \in \Delta(E)$. Then for every $\sigma \in \mathrm{Aut}_F(E)$, we have

$$(\widehat{f}(\overline{b}))(\sigma) = f(\sigma, \overline{b}) = f_{\overline{b}}(\sigma|_{F[\sqrt[n]{b}]}). \tag{36.6}$$

By (36.6), we conclude that $\mathrm{Im}\,\widehat{f}(\overline{b}) = \mathrm{Im}\, f_{\overline{b}}$, and so it is a cyclic subgroup of M_n. Moreover, by Lemma 34.3.1 and (36.5), we have

$$|\,\mathrm{Im}\,\widehat{f}(\overline{b})| = |\,\mathrm{Im}\, f_{\overline{b}}| = |\,\mathrm{Aut}_F(F[\sqrt[n]{b}])| = |\,\mathrm{Aut}_F(K)| = |\,\mathrm{Im}\,\chi|. \tag{36.7}$$

As $\mathrm{Im}\,\widehat{f}(\overline{b})$ and $\mathrm{Im}\,\chi$ are subgroups of the cyclic group M_n and they have the same order, we conclude that

$$\mathrm{Im}\,\widehat{f}(\overline{b}) = \mathrm{Im}\,\chi. \tag{36.8}$$

Since $\mathrm{Aut}_F(K)$ is cyclic and the restriction map from $\mathrm{Aut}_F(E)$ to $\mathrm{Aut}_F(K)$ is surjective, there is $\sigma_0 \in \mathrm{Aut}_F(E)$ such that $\mathrm{Aut}_F(K)$ is generated by $\sigma_0|_K$. Hence

$$\mathrm{Im}\,\widehat{f}(\overline{b}) = \mathrm{Im}\, f_{\overline{b}} = \Big\langle \frac{\sigma_0(\sqrt[n]{b})}{\sqrt[n]{b}} \Big\rangle. \tag{36.9}$$

We also notice that if $\sigma|_K = \sigma'|_K$, then $(\sigma^{-1} \circ \sigma \in \mathrm{Aut}_K(E)$. Because $\ker\chi = \mathrm{Aut}_K(E)$, we conclude that $\chi(\sigma) = \chi(\sigma')$. Hence

$$\mathrm{Im}\,\chi = \langle \chi(\sigma_0) \rangle. \tag{36.10}$$

By (36.8), (36.9), and (36.10), we conclude that

$$\chi(\sigma_0) = \Big(\frac{\sigma_0(\sqrt[n]{b})}{\sqrt[n]{b}} \Big)^i \tag{36.11}$$

for some integer i such that $\gcd(i, m) = 1$, where $m := |\,\mathrm{Im}\,\widehat{f}(b)|$. Hence by (36.11), we obtain that

$$\chi(\sigma_0) = f_{\overline{b}^i}(\sigma_0|_K). \tag{36.12}$$

For every $\sigma \in \mathrm{Aut}_F(E)$, $\sigma|_K = \sigma_0^j|_K$ for some integer j. Therefore $\sigma = \sigma_0^j \tau$ for some $\tau \in \mathrm{Aut}_K(E)$. Notice that $\ker\chi = \mathrm{Aut}_K(E)$, and so

$$\begin{aligned}
\chi(\sigma) =&\chi(\sigma_0^j \tau) = \chi(\sigma_0)^j \\
=&f_{\overline{b}^i}(\sigma_0|_K)^j = f_{\overline{b}^i}(\sigma_0^j|_K) \\
=&f_{\overline{b}^i}(\sigma|_K) = \Big(\widehat{f}(\overline{b}^i)\Big)(\sigma).
\end{aligned}$$

This means $\widehat{f}(\overline{b}^i) = \chi$. Therefore $\widehat{f}$ is surjective. This completes the proof that the Kummer pairing f is a perfect pairing. □

36.4 Perfect pairings

The main goal of this section is to prove the following property of perfect pairings. We formulate this result only for the Kummer pairing.

Proposition 36.4.1. *In the setting of Theorem 36.2.1, let* $A := \mathrm{Aut}_F(E)$ *and* $B := \Delta(E)$. *Let* $f : A \times B \to S^1$ *be the Kummer pairing given in (34.5). Let*

$$\widetilde{f} : A \to \widehat{B}, \quad (\widetilde{f}(a))(b) := f(a, b).$$

Then $\widetilde{f}$ *is an isomorphism.*

Proof. Notice that by (34.12), $\widetilde{f}(a) \in \widehat{B}$. For every $a_1, a_2 \in A$ and $b \in B$, we have

$$\begin{aligned}(\widetilde{f}(a_1a_2))(b) =& f(a_1a_2, b) = f(a_1, b)f(a_2, b)\\ =&(\widetilde{f}(a_1))(b)(\widetilde{f}(a_2))(b)\\ =&(\widetilde{f}(a_1) \cdot \widetilde{f}(a_2))(b),\end{aligned}$$

and so $\widetilde{f}(a_1a_2) = \widetilde{f}(a_1) \cdot \widetilde{f}(a_2)$. This means that $\widetilde{f} : A \to \widehat{B}$ is a group homomorphism.

Next we show that $\widetilde{f}$ is injective. Suppose to the contrary that there is $a \neq 1$ such that $\widetilde{f}(a) = \mathbb{1}$. Then for every $b \in B$, we have

$$f(a, b) = 1. \tag{36.13}$$

Since $a \neq 1$, by Lemma 35.2.3, there is $\chi_0 \in \widehat{A}$ such that

$$\chi_0(a) \neq 1. \tag{36.14}$$

Since by Proposition 36.3.1 $\widehat{f} : B \to \widehat{A}$ is surjective, there is $b_0 \in B$ such that $\widehat{f}(b_0) = \chi_0$. Hence

$$\chi_0(a) = (\widehat{f}(b_0))(a) = f(a, b_0). \tag{36.15}$$

By (36.13), (36.14), and (36.15), we get a contradiction. This shows that $\widetilde{f}$ is injective.

Since $\widehat{f} : B \to \widehat{A}$ is an isomorphism,

$$|B| = |\widehat{A}|. \tag{36.16}$$

By Theorem 36.1.2, we have that

$$|A| = |\widehat{A}| \quad \text{and} \quad |B| = |\widehat{B}|. \tag{36.17}$$

By (36.16) and (36.17), we deduce that $|A| = |\widehat{B}|$. As $\widetilde{f} : A \to \widehat{B}$ is injective and $|A| = |\widehat{B}|$, we conclude that $\widetilde{f}$ is a bijection, and so it is an isomorphism. This completes the proof. □

36.5 Proof of Kummer theory: abelian case

In this section, we finish the proof of Theorem 36.2.1. We start by proving that $\Lambda \circ \Delta = \mathrm{id}$. Suppose $\Delta(E) = \{a_1(F^{\times n}), \ldots, a_m(F^{\times n})\}$. We have to show that $E = F[\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}]$. Since $a_i(F^{\times n}) \in \Delta(E)$, there is $\alpha_i \in E$ such that $a_i(F^{\times n}) = \alpha_i^n(F^{\times n})$. Hence $a_i = \alpha_i^n c_i^n$ for some $c_i \in F^\times$. Therefore $\sqrt[n]{a_i} = (\zeta^j c_i)\alpha_i$ for some integer j. Because $\zeta^j \in F$, we conclude that $\sqrt[n]{a_i} \in E$. Therefore

$$F[\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}] \subseteq E. \tag{36.18}$$

Suppose $\sigma \in \mathrm{Aut}_{F[\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}]}(E)$. Then $\sigma(\sqrt[n]{a_i}) = \sqrt[n]{a_i}$ for every i. This implies that $f(\sigma, \overline{a}_i) = 1$ where $\overline{a}_i := a_i(F^{\times n})$. This means $\widetilde{f}(\sigma) = \mathbb{1}$. By Proposition 36.4.1, $\widetilde{f}$ is an isomorphism, and so $\sigma = \mathrm{id}$. Hence $\mathrm{Aut}_{F[\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}]}(E) = \{\mathrm{id}\}$. Therefore by the fundamental theorem of Galois theory, we have $E = F[\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}]$. This shows that $\Lambda \circ \Delta = \mathrm{id}$.

Next we show that $\Delta \circ \Lambda = \mathrm{id}$. Suppose

$$A := \{a_1(F^{\times n}), \ldots, a_m(F^{\times n})\} \le F^\times / F^{\times n}$$

and

$$E := \Lambda(A) = F[\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}].$$

We have to show that $\Delta(E) = A$.

For every i, $\sqrt[n]{a_i}^n \in E^{\times n} \cap F^\times$, and so $A \subseteq \Delta(E)$.

Suppose to the contrary that $A \subsetneq \Delta(E)$. Then by Theorem 36.1.2, $\widehat{(\frac{\Delta(E)}{A})} \neq 1$. So there is a non-trivial group homomorphism $\overline{\chi}_0 : \frac{\Delta(E)}{A} \to S^1$. Let

$$\chi_0 : \Delta(E) \to S^1, \ \chi_0(x) := \overline{\chi}(xA).$$

Then $A \subseteq \ker \chi_0$. Since by Proposition 36.4.1 $\widetilde{f} : \mathrm{Aut}_F(E) \to \widehat{\Delta(E)}$ is an isomorphism, there is $\sigma_0 \in \mathrm{Aut}_F(E)$ such that $\widetilde{f}(\sigma_0) = \chi_0$. Therefore, for every $\overline{a} \in \Delta(E)$, we have

$$f(\sigma_0, \overline{a}) = (\widetilde{f}(\sigma_0))(\overline{a}) = \chi_0(\overline{a}). \tag{36.19}$$

By (36.19) and $A \subseteq \ker \chi_0$, we conclude that $f(\sigma_0, \overline{a}_i) = 1$. This means

$$\sigma_0(\sqrt[n]{a_i}) = \sqrt[n]{a_i}$$

for every i. As $E = F[\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}]$ and $\sigma_0(\sqrt[n]{a_i}) = \sqrt[n]{a_i}$, we deduce that $\sigma_0 = \mathrm{id}$. Hence $\chi_0 = \widetilde{f}(\sigma_0) = \widetilde{f}(\mathrm{id}) = \mathbb{1}$, which is a contradiction. This completes the proof.

Chapter 37

Lecture 13

The common theme of studying zeros of polynomials has been our main source of motivation for introducing and understanding various parts of algebra. So far we have been mainly studying zeros of single variable polynomials. Next we want to move to a multivarible polynomials. Clearly the easiest case would be equations of degree one, also know as linear equations:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = b_1 \\ \vdots & & \vdots & & & & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = b_m. \end{array} \tag{37.1}$$

We have seen this type of equations in linear algebra, at least when a_{ij}'s and b_i's are complex numbers. In linear algebra, we use *vector spaces* and *linear transformations* to systematically study linear equations, and our main algorithmic source for solving a system of equations as in (37.1) is the *Gauss-Jordan elimination* process. Considering the elimination process only uses $+, -, \cdot$, and $/$, it is not surprising that the same process works for a system of linear equations over an arbitrary *field*. This method, however, does not necessarily work for an arbitrary (unital commutative) ring. For instance, consider the easiest case, where we have only one variable and one equation, over an arbitrary unital commutative ring A. This means we want to solve the equation $ax = b$ for some $a, b \in A$. We know that this equation has a solution if and only if b is in the ideal generated by a. If A is not an integral domain, then this equation might have more than one solution. So it is only natural to expect that the structure of ideals of A and other ring theoretic properties of A play a vital role in solving (37.1).

37.1 Module theory: basic examples

In this section, we introduce R-*modules* where R is a unital commutative ring. Modules are natural extensions of vector spaces from fields to rings.

Definition 37.1.1. *Suppose R is a unital commutative ring and M is an abelian group. We say M is an R-module if there is a* scalar multiplication $r \times M \to M, (r, m) \mapsto r \cdot m$ *with the following properties:*

1. $1 \cdot m = m$ *for every* $m \in M$*,*

2. $r_1 \cdot (r_2 \cdot m) = (r_1 r_2) \cdot m$ *for every* $r_1, r_2 \in R$ *and* $m \in M$*,*

3. $(r_1 + r_2) \cdot m = r_1 \cdot m + r_2 \cdot m$*, and* $r \cdot (m_1 + m_2) = r \cdot m_1 + r \cdot m_2$ *for every* $r, r_1, r_2 \in R$ *and* $m_1, m_2, m \in M$*.*

Here is a list of important examples of modules.

1. (Vector spaces) If F is a field, then V is an F-module exactly when V is an F-vector space.

2. (Free modules) Suppose R is a unital commutative ring. Then $R^n := R \times \cdots \times R$ is an R-module with respect to the following scalar multiplication:

$$r \cdot (r_1, \ldots, r_n) := (rr_1, \ldots, rr_n)$$

for every $r, r_1, \ldots, r_n \in R$.

3. (Direct product) Suppose $M_1, \ldots, M_n$ are R-modules. Then

$$\prod_{i=1}^{n} M_i := M_1 \times \cdots M_n$$

is an R-module with respect to the following scalar multiplication:

$$r \cdot (m_1, \ldots, m_n) := (r \cdot m_1, \ldots, r \cdot m_n).$$

4. (Ideals) Suppose R is a unital commutative ring and I is a subgroup of $(R, +)$. Then I is an R-module with respect the ring multiplication of R as its scalar multiplication if and only if I is an ideal of R.

5. (Ring extensions) Suppose R' is a unital commutative ring and R is a subring of R' such that $1_{R'} = 1_R$. Then R' is an R-module with respect to the scalar multiplication given by the ring multiplication of R'.

6. (Abelian groups) Suppose $(A, +)$ is an abelian group. Then A is a $\mathbb{Z}$-module with respect to the following scalar multiplication:

$$n \cdot a := \begin{cases} \underbrace{a + \cdots + a}_{n \text{ times}} & \text{if } n > 0, \\ \underbrace{(-a) + \cdots + (-a)}_{-n \text{ times}} & \text{if } n < 0, \\ 0 & \text{if } n = 0. \end{cases}$$

37.2 The first isomorphism theorem

As always, when we define a new object, we have to define its *substructures* and *homomorphisms*.

Definition 37.2.1. *Suppose M is an R-module. We say $N \subseteq M$ is a* submodule *of M if N is a subgroup of $(M, +)$ and it is closed under scalar multiplication (this means that $r \cdot n \in N$ for every $r \in R$ and $n \in N$).*

Notice that if N is a submodule of an R-module M, then N is an R-module with respect to the restriction of operations of M to N.

Definition 37.2.2. *Suppose M and M' are R-modules. Then $f : M \to M'$ is called an R-*module homomorphism *if*

$$f(m_1 + m_2) = f(m_1) + f(m_2) \quad \textit{and} \quad f(r \cdot m) = r \cdot f(m)$$

for every $r \in R$ and $m, m_1, m_2 \in M$.

Lemma 37.2.3. *Suppose $f : M \to M'$ is an R-module homomorphism. Then $\operatorname{Im} f$ is submodule of M and $\ker f$ is a submodule of M'.*

Proof. From group theory, we know that $\operatorname{Im} f$ and $\ker f$ are subgroups of M and M', respectively. So it is sufficient to show that $\operatorname{Im} f$ and $\ker f$ are closed under scalar multiplication.

Suppose $y \in \operatorname{Im} f$. Then $y = f(m)$ for some $m \in M$. Hence for every $r \in R$, we have

$$r \cdot y = r \cdot f(m) = f(r \cdot m) \in \operatorname{Im} f.$$

This shows that $\operatorname{Im} f$ is closed under scalar multiplication.

Next suppose $m \in \ker f$. Then $f(m) = 0$. We have to show that $r \cdot m \in \ker f$. This means we need to prove $f(r \cdot m) = 0$. As f is a module homomorphism, we have to show $r \cdot f(m) = 0$. Because $f(m) = 0$, we obtain the desired result by showing that $r \cdot 0 = 0$ for every $r \in R$. Notice that

$$r \cdot 0 = r \cdot (0 + 0) = r \cdot 0 + r \cdot 0,$$

and so $r \cdot 0 = 0$. This completes the proof. □

Similar to groups, rings, and vector spaces, we want to prove the first isomorphism theorem for modules. So we need to define quotient of modules.

Proposition 37.2.4. *Suppose M is an R-submodule and $N \subseteq M$ is a submodule. Then*

$$R \times M/N \to M/N, \quad r \cdot (m + N) := r \cdot m + N$$

is a well-defined and abelian group M/N with respect to this scalar multiplication is an R-module. Moreover $p_N : M \to M/N, p_N(m) := m + N$ is an R-module homomorphism, p_N is surjective, and $\ker p_N =$.

Proof. We start by showing that $r \cdot (m + N) := r \cdot m + N$ is well-defined. Suppose $m_1 + N = m_2 + N$. Then $m_1 - m_2 \in N$, and so for every $r \in R$, we have $r \cdot (m_1 - m_2) \in N$ as N is a submodule. This implies that

$$(r \cdot m_1) + (r \cdot (-m_2)) \in N. \tag{37.2}$$

Notice that

$$r \cdot (-m_2) + r \cdot m_2 = r \cdot ((-m_2) + m_2) = r \cdot 0 = 0,$$

which implies that $r \cdot (-m_2) = -r \cdot m_2$. Therefore by (37.2), we obtain that

$$r \cdot m_1 - r \cdot m_2 \in N.$$

Hence $r \cdot m_1 + N = r \cdot m_2 + N$. This shows that $\cdot$ is well-defined.

Since all the operations in M/N are defined in terms of coset representatives and operations of M, one can easily check that these operations satisfy properties of modules. Here we just check one of the distribution properties.

For $m_1, m_2 \in M$ and $r \in R$, we have

$$\begin{aligned} r \cdot ((m_1 + N) + (m_2 + N)) =& r \cdot ((m_1 + m_2) + N) \\ =& (r \cdot (m_1 + m_2)) + N \\ =& (r \cdot m_1 + r \cdot m_2) + N \\ =& (r \cdot m_1 + N) + (r \cdot m_2 + N) \\ =& r \cdot (m_1 + N) + r \cdot (m_2 + N). \end{aligned}$$

From group theory , we know that $p_N : M \to M/N$ is a surjective group homomorphism and $\ker p_N = N$. So it is enough to argue why p_N preserves scalar multiplication. For every $r \in R$ and $m \in M$, we have

$$p_N(r \cdot m) = r \cdot m + N = r \cdot (m + N) = r \cdot p_N(m),$$

which implies that p_N is an R-module homomorphism. This completes the proof. □

Now we are ready to prove the first isomorphism theorem for modules.

Theorem 37.2.5 (The first isomorphism theorem for modules)**.** *Suppose $f : M \to M'$ is an R-module homomorphism. Then*

$$\overline{f} : M \to \operatorname{Im} f, \quad \overline{f}(m) := f(m) + \ker f$$

is an R-module isomorphism.

Proof. From group theory, we know $\overline{f}$ is a group isomorphism. So it is enough to show that $\overline{f}$ preserves scalar multiplication. For every $r \in R$ and $m \in M$, we have

$$\begin{aligned} \overline{f}(r \cdot (m + \ker f)) =& \overline{f}(r \cdot m + \ker f) \\ =& f(r \cdot m) = r \cdot f(m) \\ =& r \cdot \overline{f}(m + \ker f). \end{aligned}$$

This completes the proof. □

37.3 Noetherian modules

Let's recall that if a vector space V over a field F is a span of finitely many vectors, then it is isomorphic to F^n for some integer n and every subspace of V is also finitely generated. In this section, we want to understand to what extent we can generalize these properties to modules. We start by defining finitely generated modules and proving that summation and intersection of two submodules is a submodule.

Lemma 37.3.1. *Suppose M is an R-module, $N_1, N_2 \subseteq M$ are submodules, and $m_1, \ldots, m_n \in M$. Then*

1. *$N_1 \cap N_2$ is a submodule.*
2. *$N_1 + N_2 := \{n_1 + n_2 \mid n_1 \in N_1, n_2 \in N_2\}$ is a submodule.*
3. *$\{r_1 \cdot m_1 + \cdots + r_n \cdot m_n \mid r_1, \ldots, r_n \in R\}$ is the smallest submodule of M which contains $m_1, \ldots, m_n$. We denote this submodule by $\langle m_1, \ldots, m_n\rangle$ or $Rm_1 + \cdots + Rm_n$.*

Proof. From group theory, we know that $N_1 \cap N_2$ and $N_1 + N_2$ are subgroups of M. So to show they are submodules of M, it is enough to argue why they are close under scalar multiplication.

Suppose $m \in N_1 \cap N_2$. Let $i = 1$ or 2. Then $m \in N_i$. As N_i is a submodule, $r \cdot m \in N_i$ for every $r \in R$. Hence $r \cdot m \in N_1 \cap N_2$. This implies that $N_1 \cap N_2$ is a submodule of M.

Next we want to show that $N_1 + N_2$ is close under scalar multiplication. Suppose $m \in N_1 + N_2$. Then $m = n_1 + n_2$ for some $n_1 \in N_1$ and $n_2 \in N_2$. Hence for every $r \in R$, we have

$$r \cdot m = r \cdot (n_1 + n_2) = r \cdot n_1 + r \cdot n_2. \tag{37.3}$$

Notice that $r \cdot n_i \in N_i$ for $i = 1, 2$ as N_i is a submodule. Therefore by (37.3), we conclude that $r \cdot m \in N_1 + N_2$. This implies that $N_1 + N_2$ is a submodule.

Let $N := \{r_1 \cdot m_1 + \cdots + r_n \cdot m_n \mid r_1, \ldots, r_n \in R\}$. Next we show that N is a submodule of M. To this end, we have to prove that N is close under taking difference and scalar multiplication. Suppose $x, x' \in N$. Then there are r_i's and r'_i's in R such that

$$x = r_1 \cdot m_1 + \cdots + r_n \cdot m_n \text{ and } x' = r'_1 \cdot m_1 + \cdots + r'_n \cdot m_n.$$

Hence

$$x - x' = (\sum_{i=1}^n r_i \cdot m_i) - (\sum_{i=1}^n r'_i m_i) = \sum_{i=1}^n (r_i \cdot m_i - r'_i \cdot m_i) = \sum_{i=1}^n (r_i - r'_i) \cdot m_i \in N,$$

and

$$r \cdot x = r \cdot \sum_{i=1}^n r_i m_i = \sum_{i=1}^n r \cdot (r_i \cdot m_i) = \sum_{i=1}^n (rr_i) \cdot m_i \in N$$

for every $r \in R$. This shows that N is a submodule. Next we show that $m_i \in N$ for every index i. To see this we notice that

$$m_i = 0 \cdot m_1 + \cdots + 0 \cdot m_{i-1} + 1 \cdot m_i + 0 \cdot m_{i+1} + \cdots + 0 \cdot m_n \in N.$$

To finish the proof, we show that if K is a submodule of M which contains m_i's, then $N \subseteq K$. Suppose K is a submodule of M and $m_i \in K$ for every index i. Since K is close under scalar multiplication and $m_i \in K$, we obtain that $r_i \cdot m_i \in K$ for every $r_i \in R$. As K is close under addition and $r_i \cdot m_i \in K$ for every index i, we deduce that $\sum_{i=1}^n r_i \cdot m_i \in K$. This implies that N is a subset of K, which completes the proof. □

We say a module M is *finitely generated* if $M = \langle m_1, \dots, m_n \rangle$ for some m_i's in M. We say M is a *cyclic* R-module if $M = \langle m \rangle$ for some $m \in M$.

Lemma 37.3.2. *An R-module M is cyclic if and only if $M \simeq R/I$ for some ideal I.*

Proof. ($\Rightarrow$) Suppose $M = \langle m \rangle$. This means every element of M is of the form $r \cdot m$ for some $r \in R$. Hence the function $f : R \to M, f(r) := r \cdot m$ is a surjective function. Next we show that f is an R-module homomorphism. For $r_1, r_2 \in R$, we have

$$f(r_1 + r_2) = (r_1 + r_2) \cdot m = r_1 \cdot m + r_2 \cdot m = f(r_1) + f(r_2)$$

and

$$r_1 \cdot f(r_2) = r_1 \cdot (r_2 \cdot m) = (r_1 r_2) \cdot m = f(r_1 r_2),$$

which implies that f is an R-module homomorphism. By the first isomorphism theorem for modules, we obtain that

$$M \simeq R/\ker f.$$

Notice that $\ker f$ is a submodule of R, and so $\ker f$ is an ideal of R. The claim follows.

($\Leftarrow$) Suppose I is an ideal of R. Then every element of R/I is of the form $r + I = r \cdot (1 + I)$, which implies that R/I is generated by $1 + I$ as an R-module. This means R/I is a cyclic R-module. This completes the proof. □

Next we show that a cyclic module is rarely a free module. This shows a contrast with the vector space case.

Lemma 37.3.3. *Suppose R is a ring and I is a proper non-trivial ideal of R. Then $R/I \not\simeq R^n$ for every non-negative integer n.*

Proof. Suppose to the contrary that there is an R-module isomorphism $f : R/I \to R^n$. Since $I \neq R$, $R/I \neq 0$. Hence $n \geq 1$. Suppose $r_0 \in I \setminus \{0\}$. Then for every $r \in R$

$$r_0 \cdot (r + I) = r_0 r + I = 0 + I.$$

Therefore $r_0 \cdot f(r + I) = f(r_0 \cdot (r + I)) = f(0 + I) = 0$ for every $r \in R$. Hence $r_0 \cdot \operatorname{Im} f = 0$. As f is surjective, $r_0 \cdot (1, \dots, 1) = (0, \dots, 0)$, which implies that $r_0 = 0$. This is a contradiction. □

Next we focus on finite generatedness of submodules of a finitely generated module. We consider R as an R-module. We have already pointed out that $I \subseteq R$ is a submodule if and only if I is an ideal of R. Hence I is generated by $x_1, \dots, x_n$ as an ideal if and only if it is generated by x_i's as an R-submodule. Hence all the submodules of R are finitely generated exactly when every ideal of R is finitely generated. By Lemma 12.3.5, all ideals of R are finitely generated precisely when R is Noetherian. This brings us to two points:

1. If R is not Noetherian, then R is a finitely generated R-module which has submodules that are not finitely generated.

2. Finite generatedness of submodules can be related to an ascending chain condition as it is the case for rings.

Definition 37.3.4. *Suppose M is an R-module. We say M is* Noetherian *if it does not have an infinite strictly ascending chain of submodules. This means if $M_1 \subseteq M_2 \subseteq \cdots$ are submodules of M, then $M_{n_0} = M_{n_0+1} = \cdots$ for some integer n_0.*

Notice that because R-submodules of R are ideals of R, R is a Noetherian ring if and only if R is a Noetherian R-module. Similar to rings, we have the following proposition.

Proposition 37.3.5. *Suppose M is an R-module. Then M is Noetherian if and only if every submodule of M is finitely generated.*

Proof. ($\Rightarrow$) Suppose to the contrary that M has a submodule which is not finitely generated. Inductively we define a strictly ascending chain of finitely generated submodules of M. Let $N_0 := \{0\}$. Since N is not finitely generated, $N \neq N_0$. Hence there is $m_1 \in N \setminus N_0$. Let $N_1 := \langle m_1 \rangle$. Since N is not finitely generated and N_1 is finitely generated, there is $m_2 \in N \setminus N_1$. Let $N_2 := \langle m_1, m_2 \rangle$. We repeat this argument. Suppose we have already found $m_1, \ldots, m_k \in N$ such that

$$N_1 \subsetneq N_2 \subsetneq \cdots \subsetneq N_k$$

where $N_i := \langle m_1, \ldots, m_i \rangle$. Since N is not finitely generated, there is $m_{k+1} \in N \setminus N_k$. Let

$$N_{k+1} := \langle m_1, \ldots, m_{k+1} \rangle.$$

Then $N_k \subsetneq N_{k+1} \subsetneq N$. Hence by induction there is a strictly ascending chain

$$N_0 \subsetneq N_1 \subsetneq \cdots$$

of submodules of N. This contradicts the hypothesis that M is Noetherian. $\square$

Chapter 38

Lecture 14

In this section, we continue with the study of Noetherian modules.

38.1 Noetherian modules

We continue proof of Proposition 37.3.5.

Proof of Proposition 37.3.5. ($\Leftarrow$) Suppose all submodules of M are finitely generated. Let $M_0 \subseteq M_1 \subseteq \cdots$ be an ascending chain of submodules of M. Let $N := \bigcup_{i=0}^{\infty} M_i$.

Claim. *N is a submodule of M.*

Proof of Claim. Suppose $n, n' \in N$. Then there are indexes i and j such that $n \in M_i$ and $n' \in M_j$. Without loss of generality we can and will assume that $i \leq j$. Then $M_i \subseteq M_j$, and so $n, n' \in M_j$. Hence $n - n' \in M_j$ and $rn \in M_j$ for every $r \in R$. Therefore $n - n', rn \in N$, which implies that N is a subgroup and closed under multiplication. The claim follows. □

By hypothesis and the above claim, $N = \langle n_1, \ldots, n_k \rangle$. As $n_i \in N$ and $N = \bigcup_{j=0}^{\infty} M_j$, there is an index r_i such that $n_i \in M_{r_i}$. Let $r := \max\{r_1, \ldots, r_k\}$. Then $M_{r_i} \subseteq M_r$ for every i. Hence $n_1, \ldots, n_r \in N$. Therefore

$$\langle n_1, \ldots, n_k \rangle \subseteq M_r. \tag{38.1}$$

By (38.1), we conclude that $\bigcup_{j=0}^{\infty} M_j \subseteq M_r$. Thus, for every $i \geq r$, we have both $M_i \subseteq M_r$ and $M_i \subseteq M_r$. Therefore

$$M_r = M_{r+1} = \cdots,$$

which completes the proof. □

The following proposition gives us a better understanding of Noetherian modules.

Proposition 38.1.1. *Suppose A, B, C are submodules of M and $B \subseteq A$. Then*

1. *A is Noetherian if and only if B and A/B are Noetherian.*

2. *If A and C are Noetherian, then $A+C$ is Noetherian.*

Proof. 1. ($\Rightarrow$) If A is Noetherian, every submodule of A is finitely generated. Hence every submodule of B is finitely generated as submodules of B are also submodules of A. Therefore by Proposition 37.3.5, B is Noetherian. To show A/B is Noetherian, we start by proving a correspondence result for submodules of A/B.

Claim. *The function $f(A') := A'/B$ is a bijection between submodules of A which contains B and submodules of A/B.*

Proof of Claim. It is easy to see that f is well-defined. Because of the correspondence theorem for groups (see Theorem 30.1.2), we obtain that f is injective. Finally we argue why this function is surjective. Suppose $\overline{A}'$ is a submodule of A/B; in particular, $\overline{A}'$ is a subgroup of A/B. Hence by the correspondence theorem for group quotients, $\overline{A}' = A'/B$ for some subgroup A' of A which contains B. Next we show that A' is a submodule of A. As A' is a subgroup, it is enough to show that A' is closed under scalar multiplication. For $a' \in A'$, $a'+B \in \overline{A}'$. Hence for every $r \in R$, we have $r \cdot (a'+B) \in \overline{A}'$. This implies that $ra'+B \in A'/B$, and so $ra' \in A'$. Therefore A' is a submodule. This completes proof of Claim. □

To show A/B is Noetherian, by Proposition 37.3.5 it is enough to show that every submodule $\overline{A}'$ of A/B is finitely generated. By the above Claim, $\overline{A}' = A'/B$ for some submodule A' of A (which contains B as a submodule). Since A is Noetherian, A' is finitely generated. So $A' = \{\sum_{i=1}^{n} r_i \cdot a'_i \mid r_i \in R\}$ for some $a'_i \in A'$. Then

$$\overline{A}' = A'/B = \{\sum_{i=1}^{n} r_i \cdot (a'_i + B) \mid r_i \in R\} = \langle a'_1 + B, \ldots, a'_n + B\rangle.$$

Hence $\overline{A}'$ is finitely generated. This shows that A/B is Noetherian.

($\Leftarrow$) To show A is Noetherian, it is enough to show every submodule N of A is finitely generated (see Proposition 37.3.5). Considering our hypothesis is on B and A/B, we use the projection of N to A/B and its intersection with B. Notice that since $p_B : A \to A/B$ is a module homomorphism, the restriction $p_B|_N$ of p_B to N is also a module homomorphism. Therefore $p_B(N)$ is a submodule of A/B. From group theory, we know that $p_B(N) = (N+B)/B$; here is how one can show this equality:

$$\begin{aligned} p_B(N) =& \{n + B \mid \forall n \in B\} \\ =& \{(n+b) + B \mid \forall n \in N, \forall b \in B\} = (N+B)/B. \end{aligned} \tag{38.2}$$

Hence $(N+B)/B$ is a submodule of A/B and $N \cap B$ is a submodule of B. Since A/B and B are Noetherian modules, by Proposition 37.3.5 we conclude that $(N+B)/B$ and $N \cap B$ are finitely generated modules. Hence there are $n_1, \ldots, n_k \in N$ and $n'_1, \ldots, n'_\ell \in N$ such that

$$p_B(N) = \langle n_1 + B, \ldots, n_k + B\rangle \quad \text{and} \quad N \cap B = \langle n'_1, \ldots, n'_\ell\rangle. \tag{38.3}$$

We want to show that N is generated by $n_1, \ldots, n_k, n'_1, \ldots, n'_\ell$. Let

$$N' := \langle n_1, \ldots, n_k, n'_1, \ldots, n'_\ell\rangle.$$

Claim. $N = N'$.

Proof of Claim. Since n_i's and n'_j's are elements of N, we have $N' \subseteq N$. Next we want to show that N is a subset of N'. Suppose $n \in N$. We have to write n as an R-linear combination of n_i's and n'_j's. First we do this for the projection $p_B(n)$ of n. Since $p_B(N)$ is generated by $n_i + B$'s, there are r_i's in R such that

$$n + B = r_1 \cdot (n_1 + B) + \cdots + r_k \cdot (n_k + B) = (\sum_{i=1}^{k} r_i \cdot n_i) + B. \tag{38.4}$$

By (38.4), we conclude that

$$n - \sum_{i=1}^{k} r_i \cdot n_i \in B. \tag{38.5}$$

As n and n_i's are in N, we also have that $n - \sum_{i=1}^{k} r_i \cdot n_i \in N$. Therefore by (38.5), we deduce that

$$n - \sum_{i=1}^{k} r_i \cdot n_i \in N \cap B. \tag{38.6}$$

Because $N \cap B$ is generated by n'_j's, (38.6) implies that there are r'_j's in R such that

$$n - \sum_{i=1}^{k} r_i \cdot n_i = r'_1 \cdot n'_1 + \cdots + r'_\ell \cdot n'_\ell. \tag{38.7}$$

By (38.7), we obtain that

$$n = \sum_{i=1}^{k} r_i \cdot n_i + \sum_{j=1}^{\ell} r'_j \cdot n'_j \in N',$$

which completes proof of the claim. □

The above claim implies that every submodule of A is finitely generated, and so by Proposition 37.3.5, A is Noetherian. This completes proof of the first part.

2. Let $p_C : M \to M/C$ be the natural quotient map. Then by (38.2), we have $p_C(A) = (A + C)/C$. Hence $p_C|_A : A \to (A + C)/C$ is a surjective module homomorphism. Therefore by the first isomorphism theorem, we conclude

$$\frac{A}{\ker(p_C|_A)} \simeq \frac{A + C}{C}. \tag{38.8}$$

Notice that $\ker(p_C|_A) = \ker p_C \cap A = C \cap A$. Hence by (38.8), we conclude (the second isomorphism theorem for modules)

$$\frac{A}{A \cap C} \simeq \frac{A + C}{C}. \tag{38.9}$$

As A is Noetherian, by the first part, $A/A \cap C$ is Noetherian. Hence by (38.9), we conclude that $(A+C)/C$ is Noetherian.

Since C and $(A+C)/C$ are Noetherian, by the first part, we deduce that $A+C$ is Noetherian. This completes the proof. □

The following theorem is an important consequence of Proposition 38.1.1. Roughly it says a *ring can be its own worst enemy!* More precisely, if R is a Noetherian R-module, then every finitely generated R-module is Noetherian.

Theorem 38.1.2. *Suppose R is Noetherian. Then every finitely generated R-module is Noetherian.*

We start by showing that every finitely generated R-module is a quotient of R^n for some positive integer n. We will be using this result later as well.

Lemma 38.1.3. *Suppose R is a unital commutative ring and M is an R-module. If $M = \langle m_1, \ldots, m_n \rangle$, then*

$$f : R^n \to M, \quad f(r_1, \ldots, r_n) := \sum_{i=1}^{n} r_i \cdot m_i$$

is a surjective R-module homomorphism.

Proof. Since M is generated by m_i's, every element of M is an R-linear combination of m_i's. Hence f is surjective. Next we show that f is an R-module homomorphism. For every r_i's, r_i''s, and r in R, we have

$$\begin{aligned} f(r_1, \ldots, r_n) + f(r_1', \ldots, r_n') =& \sum_{i=1}^{n} r_i \cdot m_i + \sum_{i=1}^{n} r_i' \cdot m_i = \sum_{i=1}^{n} (r_i \cdot m_i + r_i' \cdot m_i) \\ =& \sum_{i=1}^{n} (r_i + r_i') \cdot m_i = f(r_1 + r_1', \ldots, r_n + r_n') \\ =& f((r_1, \ldots, r_n) + (r_1', \ldots, r_n')), \end{aligned}$$

and

$$\begin{aligned} f(r \cdot (r_1, \ldots, r_n)) =& f(rr_1, \ldots, rr_n) = \sum_{i=1}^{n} (rr_i) \cdot m_i \\ =& \sum_{i=1}^{n} r \cdot (r_i \cdot m_i) = r \cdot \sum_{i=1}^{n} r_i \cdot m_i \\ =& r \cdot f(r_1, \ldots, r_n). \end{aligned}$$

This completes the proof. □

Proof of Theorem 38.1.2. Suppose $M = \langle m_1, \ldots, m_n \rangle$, and let

$$f : R^n \to M, \quad f(r_1, \ldots, r_n) := \sum_{i=1}^{n} r_i \cdot m_i.$$

Then by Lemma 38.1.3 and the first isomorphism theorem, we obtain that

$$M \simeq R^n / \ker f. \tag{38.10}$$

Hence by Proposition 38.1.1, to prove M is Noetherian, it is sufficient to show that R^n is Noetherian for every positive integer n.

For every integer in $[1..n]$, let

$$N_i := \{0\} \times \cdots \times \{0\} \times R \times \{0\} \times \cdots \times \{0\},$$

where the i-th term is R. Notice that $N_i \simeq R$ as an R-module. Hence N_i's are Noetherian R-modules. By applying the second part of Proposition 38.1.1 repeatedly we conclude that $N_1+\cdots+N_n$ is a Noetherian R-module. Since $R^n = N_1+\cdots+N_n$, we obtain that R^n is a Noetherian R-module. This completes the proof. □

Corollary 38.1.4. *Suppose D is a PID. Then every finitely generated D-module is Noetherian.*

Proof. Notice that since every ideal of a PID is principal, all ideals of D are finitely generated. Hence D is a Noetherian ring. Hence claim follows from Theorem 38.1.2. □

38.2 Finitely generated modules and cokernel of matrices

In this section, we use Theorem 38.1.2 to give a concrete connection between finitely generated modules of Noetherian rings and system of linear equations. To formulate our next result, we start by defining image and cokernel of a rectangular matrix $A \in \mathrm{M}_{n,m}(R)$.

Definition 38.2.1. *Suppose R is a ring and $A \in \mathrm{M}_{n,m}(R)$. Then the* image *of A is*

$$\mathrm{Im}\, A := \{A\mathbf{x} \mid \mathbf{x} \in R^m\} \subseteq R^n,$$

where we view R^n and R^m as the set of column vectors. The cokernel *of A is*

$$\mathrm{Coker}\, A := \frac{R^n}{\mathrm{Im}\, A}.$$

Notice that $\mathbf{x} \mapsto A\mathbf{x}$ is an R-module homomorphism from R^m to R^n, and so $\mathrm{Im}\, A$ is a submodule of R^n. Hence it makes sense to consider the quotient of R^n by $\mathrm{Im}\, A$ and $\mathrm{Coker}\, A$ is an R-module.

Proposition 38.2.2. *Suppose R is a Noetherian ring and M is a finitely generated R-module. Then there is $A_M \in \mathrm{M}_{n,m}(R)$ such that $M \simeq \mathrm{Coker}\, A_M$.*

Proof. Suppose $M = \langle m_1, \ldots, m_n\rangle$, and let

$$f : R^n \to M, \quad f(r_1, \ldots, r_n) := \sum_{i=1}^{n} r_i \cdot m_i.$$

Then by Lemma 38.1.3, f is a surjective R-module homomorphism. Therefore by the first isomorphism theorem, $M \simeq R^n/\ker f$. Since R is Noetherian and R^n is a finitely generated R-module, by Theorem 38.1.2 R^n is a Noetherian R-module. Hence by Proposition 37.3.5 all submodules of R^n are finitely generated. Thus $\ker f$ is a finitely generated R-module. Suppose

$$\ker f = R\mathbf{v}_1 + \cdots + R\mathbf{v}_m \tag{38.11}$$

for some $\mathbf{v}_i$'s in R^n. Then $\ker f$ is the image of

$$A_M := [\mathbf{v}_1 \cdots \mathbf{v}_m] \in \mathrm{M}_{n,m}(R).$$

Hence $M \simeq \operatorname{Coker} A_M$, which completes the proof. □

38.3 Reduced row/column operations, Smith normal form.

Proposition 38.2.2 indicates the importance of understanding image of a matrix. Notice that $\mathbb{F}$ is in the image of A if and only if $A\mathbf{x} = \mathbb{F}$ has a solution. This takes us back to solving system of linear equations. As we have mentioned at the beginning of Chapter 37, solving linear equations over a ring R is closely related to structure of ideals of R. From the point of view of structure of ideals, Euclidean domains are the easiest to consider after fields. For the rest of this section we assume that $R = D$ is a Euclidean domain. In linear algebra, we have learned how to use the Gauss-Jordan elimination process to solve system of linear equations with complex coefficients. In this section, we recall the reduced row and column operations and see how much *elimination* can be achieved over a Euclidean domain.

Let's recall that there are two types of *row reduction operations*. For two distinct integers $i, j \in [1..n]$,

$$\begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix} \xmapsto{E_{ij}(r)} \begin{pmatrix} \vdots \\ \mathbf{a}_i + r\mathbf{a}_j \\ \vdots \end{pmatrix}.$$

This operator multiplies the j-th row by r and adds it to the i-th row; it only changes the i-th row. The second type of operators, S_{ij} simply swaps the i-th and the j-th rows

$$\begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix} \xmapsto{S_{ij}} \begin{matrix} \\ i \\ \\ j \\ \\ \end{matrix} \begin{pmatrix} \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_i \\ \vdots \end{pmatrix}$$

Similarly there are two types of *column reduction operators*. For two distinct integers $i, j \in [1.m]$, $E'_{ij}(r)$ multiplies the j-th column by r and adds it to the i-th column; it only changes the i-th column, and S'_{ij} swaps the i-th and the j-th columns. Notice that

$$e_{ij}(r)\begin{pmatrix}\mathbf{a}_1\\ \vdots\\ \mathbf{a}_n\end{pmatrix}=\begin{pmatrix}\vdots\\ \mathbf{a}_i+r\mathbf{a}_j\\ \vdots\end{pmatrix},\quad\text{and}\quad s_{ij}\begin{pmatrix}\mathbf{a}_1\\ \vdots\\ \mathbf{a}_n\end{pmatrix}=\begin{matrix}\\ i\\ \\ j\\ \\ \end{matrix}\begin{pmatrix}\vdots\\ \mathbf{a}_j\\ \vdots\\ \mathbf{a}_i\\ \vdots\end{pmatrix},\tag{38.12}$$

where

$$e_{ij}(r):=\begin{matrix} & & i & & j & \\ & & & & & \\ i & \cdots & 1 & \cdots & r & \cdots\\ & & \vdots & & \vdots & \\ j & \cdots & 0 & \cdots & 1 & \cdots\\ & & \vdots & & \vdots & \end{matrix}\quad\text{and}\quad s_{ij}=\begin{matrix} & & i & & j & \\ & & \vdots & & \vdots & \\ i & \cdots & 0 & \cdots & 1 & \cdots\\ & & \vdots & & \vdots & \\ j & \cdots & 1 & \cdots & 0 & \cdots\\ & & \vdots & & \vdots & \end{matrix}.$$

Similarly we have

$$\begin{pmatrix}\mathbf{a}'_1 & \cdots & \mathbf{a}'_m\end{pmatrix}e_{ji}(r)=\begin{pmatrix}\cdots & \overset{i}{\mathbf{a}'_i+r\mathbf{a}'_j} & \cdots\end{pmatrix}\quad\text{and}\tag{38.13}$$

$$\begin{pmatrix}\mathbf{a}'_1 & \cdots & \mathbf{a}'_m\end{pmatrix}s_{ij}=\begin{pmatrix}\cdots & \overset{i}{\mathbf{a}'_j} & \cdots & \overset{j}{\mathbf{a}'_i} & \cdots\end{pmatrix}.$$

It is worth mentioning that $s_{ij}s_{ij}=I$ and for every $r,r'\in R$ and indexes i and j we have

$$e_{ij}(r)e_{ij}(r')=e_{ij}(r+r');$$

in particular, $e_{ij}(r)e_{ij}(-r)=I$ for every $r\in R$. Hence s_{ij} and $e_{ij}(r)$ are invertible matrices; that means they are units in $\mathrm{M}_n(R)$. The group of units of $\mathrm{M}_n(R)$ is denoted by $\mathrm{GL}_n(R)$. The matrices $e_{ij}(r)$'s are called *elementary matrices*, and the group generated by $e_{ij}(r)$'s is denoted by $\mathrm{E}_n(R)$. The group generated by $e_{ij}(r)$'s and s_{ij}'s is denoted by $\mathrm{E}_n^{\pm}(R)$. By (38.12) and (38.13), we obtain the following result.

Lemma 38.3.1. *Suppose R is a ring and $A, A'\in\mathrm{M}_{n,m}(R)$. We can go from A to A' by applying a series of row or column reduction operators if and only if $A'=UAV$ for some $U\in\mathrm{E}_n^{\pm}(R)$ and $V\in\mathrm{E}_m^{\pm}(R)$.*

This takes us to the defining the following relation between elements of $\mathrm{M}_{n,m}(R)$. For $A, A'\in\mathrm{M}_{n,m}(R)$, we say $A\sim A'$ if $A'=UAV$ for some $U\in\mathrm{E}_n^{\pm}(R)$ and $V\in\mathrm{E}_m^{\pm}(R)$. By Lemma 38.3.1, we have that $A\sim A'$ exactly when one can go from A to A' by applying a series of row and columns reduction operators.

Lemma 38.3.2. *In the above setting, $\sim$ is an equivalence relation.*

Before we get to the proof, let's point out that we can replace $\mathrm{E}_n^{\pm}(R)$ and $\mathrm{E}_m^{\pm}(R)$ with any other subgroups of $\mathrm{GL}_n(R)$ and $\mathrm{GL}_m(R)$, and still get an equivalence relation. One way of proving this is by showing that $(U, V) \cdot A := UAV^{-1}$ is a group action of $\mathrm{E}_n^{\pm}(R) \times \mathrm{E}_m^{\pm}(R)$ on $\mathrm{M}_{n,m}(R)$ and observing that $A \sim A'$ precisely when A and A' are in the same $\mathrm{E}_n^{\pm}(R) \times \mathrm{E}_m^{\pm}(R)$-orbit. Here we present a direct argument.

Proof. (Reflexive) Since $A = I_n A I_m$, $A \sim A$.

(Symmetric) Suppose $A \sim A'$. Then $A' = UAV$, which implies that

$$A = U^{-1} A' V^{-1}. \tag{38.14}$$

As $U \in \mathrm{E}_n^{\pm}(R)$ and $\mathrm{E}_n^{\pm}(R)$ is a group, we have that $U^{-1} \in \mathrm{E}_n^{\pm}(R)$. Similarly $V^{-1} \in \mathrm{E}_m^{\pm}(R)$. Therefore by (38.14), $A \sim A'$.

(Transitive) Suppose $A_1 \sim A_2$ and $A_2 \sim A_3$. Then there are $U, U' \in \mathrm{E}_n^{\pm}(R)$ and $V, V' \in \mathrm{E}_m^{\pm}(R)$ such that

$$A_2 = UA_1V \quad \text{and} \quad A_3 = U'A_2V'. \tag{38.15}$$

Thus $A_3 = (U'U)A_1(VV')$. Since $\mathrm{E}_n^{\pm}(R)$ and $\mathrm{E}_m^{\pm}(R)$ are groups, $UU' \in \mathrm{E}_n^{\pm}(R)$ and $VV' \in \mathrm{E}_m^{\pm}(R)$. Therefore $A_3 \sim A_1$, which completes the proof. ☐

In the reduction process, we are looking for a *simple form* of A in the same equivalence class with respect to $\sim$ as A. Here by *simple* we mean as sparse as possible; that means we are looking for a representative of the class of A with respect to $\sim$ which has the maximum possible zero entries.

Let's recall the elimination process over the field of complex numbers. If $A = 0$, we are done. If not, after swapping the needed rows and columns, we can and will assume that $a_{11} \neq 0$. Then a_{1j}'s and a_{i1}'s are multiples of a_{11}, and so after multiplying the first row by a suitable element (here it is $-\frac{a_{i1}}{a_{11}}$) and adding it to the i-th row, we get a matrix with zeros in its first column except at the $(1, 1)$ position. Similarly after multiplying the first column by a suitable element (here it is $-\frac{a_{1j}}{a_{11}}$) and adding it to the j-th column, we get a matrix of the form

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix} \tag{38.16}$$

Now we can repeat this process for B. Notice that the row and the column operations for B can be extended to operations for the entire matrix given in (38.16), and these operations do not change the first row and the first column. At the end, we get a matrix of the form $\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$, where $D = \mathrm{diag}(d_1, \ldots, d_r)$, for some non-zero d_i's. In the above process, we used the fact that in a *field if* $a_{11} \neq 0$*, then* $a_{11}x = a_{1j}$ *and* $a_{11}x = a_{i1}$ *have solutions*. In a Euclidean domain, this is false in general, but we can still divide a_{i1}'s and a_{1j}'s by a_{11}, and replace them by a remainder of this division. Repeating this process, by the virtue of the Euclid algorithm, we can get to the greatest common divisor of all the entries, and after swapping the needed rows and columns, we can assume that a_{11} divides all the entries of A. Now we can continue as we did for fields. Here is a precise statement and argument.

Proposition 38.3.3. *Suppose D is a Euclidean domain and $A \in \mathrm{M}_{n,m}(D)\setminus\{0\}$. Then there is $A' \in \mathrm{M}_{n,m}(D)$ such that $A' \sim A$, and $A' = \begin{pmatrix} d_1 & 0 \\ 0 & * \end{pmatrix}$ and $A' = d_1 A''$ for some $d_1 \in D$ and $A'' \in \mathrm{M}_{n,m}(D)$.*

Proof. Let's recall that since D is a Euclidean domain, it has a *norm function* $N : D \to \mathbb{Z}^{\geq} 0$ such that $N(d) = 0$ if and only if $d = 0$, and for every $a, b \in D \setminus \{0\}$ there are q and r in D such that $a = bq + r$ and $N(r) < N(b)$. The element q is called a *quotient* of a divided by b and the element r is called a *remainder* of a divided by b (see Section 7.4).

Let $n_0 := \min\{N(a'_{11}) \mid [a'_{ij}] \sim A \text{ and } a'_{11} \neq 0\}$. This means n_0 is the smallest norm of all the non-zero entries of all the matrices that we can reach to starting from A and using the row and the column reduction operations. Notice that by swapping the i-the row with the first row and the j-th column with the first column, we can move the (i,j)-entry to the $(1,1)$ position. Hence

$$n_0 := \{N(a'_{rs}) \mid [a'_{ij}] \sim A \text{ and } a'_{rs} \neq 0\}.$$

We will not be using this equation in the proof, but it gives us a better understanding of n_0.

Claim 1. *Suppose $A' \sim A$, $A' = [a'_{ij}]$, and $N(a'_{11}) = n_0$. Then $a'_{11}|a'_{i1}$ and $a'_{11}|a'_{1j}$ for every $i \in [1..n]$ and $j \in [1..m]$.*

Proof of Claim 1. Since D is a Euclidean domain, there are $q_j, r_j \in D$ such that $a'_{1j} = a'_{11}q_j + r_j$ and $N(r_j) < N(a'_{11})$. Then the $(1,j)$-entry of $E'_{j1}(-q_j)(A')$ is r_j. After swapping the j-th column and the first column of this matrix, we end up getting a matrix A'' whose $(1,1)$-entry is r_j and $A'' \sim A$. Therefore either $r_j = 0$ or $N(r_j) \geq n_0$. Since $N(r_j) < N(a'_{11}) = n_0$, we conclude that $r_j = 0$. This means that $a'_{11}|a'_{1j}$. By a similar argument, we can show that $a'_{11}|a'_{i1}$, and the claim follows. □

Now that every entry in the first row and the first column is a multiple of a'_{11}, we can change them to zero by applying suitable row and column reduction operations.

Claim 2. *There is $A'' \in \mathrm{M}_{n,m}(D)$ such that $A'' \sim A$, $A'' = \begin{pmatrix} d_1 & 0 \\ 0 & B \end{pmatrix}$, and $N(d_1) = n_0$.*

Proof of Claim 2. By Claim 1, there is $A' \in \mathrm{M}_{n,m}(D)$ such that $A' \sim A$, $A' = [a'_{ij}]$, and $a'_{11}|a'_{i1}$ and $a'_{11}|a'_{1j}$ for every i and j. Hence there are q_{i1}'s and q_{1j}'s in D such that $a'_{i1} = a'_{11}q_{i1}$ and $a'_{1j} = a'_{11}q_{1j}$. Hence after applying the row and the column operations $E'_{j1}(-q_{1j})$ and $E_{i1}(-q_{i1})$ one after another to A', we end up with a matrix $A'' \in \mathrm{M}_{n,m}(D)$ such that $A'' \sim A$ and $A'' = \begin{pmatrix} a'_{11} & 0 \\ 0 & B \end{pmatrix}$. Notice that

$$A'' = E'_{m1}(-q_{1m}) \circ \cdots \circ E'_{21}(-q_{12}) \circ E_{n1}(-q_{n1}) \circ \cdots \circ E_{21}(-q_{21})(A').$$

Since the $(1,1)$ entry stays the same, we still have that the norm of the $(1,1)$ entry is n_0. This implies the claim. □

Claim 3. *Suppose $A'' \sim A$, $A'' = \begin{pmatrix} d_1 & 0 \\ 0 & B \end{pmatrix}$, and $N(d_1) = n_0$. Then $d_1 | b_{ij}$ for all i and j, where $B = [b_{ij}]$.*

Proof of Claim 3. Suppose to the contrary that $d_1 \nmid b_{ij}$ for some i and j. Notice that b_{ij} is the $(i+1, j+1)$-entry of A''. Hence after adding the $(i+1)$-th row of A'' to its first row, we get a matrix which has

$$\begin{pmatrix} d_1 & b_{i1} & \cdots & b_{i(n-1)} \end{pmatrix}$$

as its first row. Since D is a Euclidean domain, there are $q, r \in D$ such that $b_{ij} = d_1 q + r$ and $N(r) < N(d_1) = n_0$. Because $d_1 \nmid b_{ij}$, $r \neq 0$. Multiplying the first column of this matrix by $-q$ and adding it to the $(j+1)$-th column, we end up getting r as the $(1, j+1)$-entry of the new matrix. After swapping the first and the $(j+1)$-th columns, we obtain a matrix A''' such that $A''' \sim A$ and its $(1,1)$-entry is r. This is a contradiction as $r \neq 0$ and $N(r) < n_0$. □

Claim 3 completes the proof. □

By repeated application of Proposition 38.3.3 for submatrices of A, we obtain a similar result as the case of matrices with entries in a field.

Theorem 38.3.4 (Smith Normal Form)**.** *Suppose D is a Euclidean domain, and $A \in \mathrm{M}_{n,m}(D) \setminus \{0\}$. Then $A \sim \begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix}$, where $T = \mathrm{diag}(d_1, \ldots, d_r)$ for some $d_1, \ldots, d_r \in D$ such that $d_1 | d_2 | \cdots | d_r$; this means $d_i | d_{i+1}$ for every $i \in [1..r)$.*

Matrix $\begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix}$ which is given in Theorem 38.3.4 is called a *Smith normal form* of A.

Proof of Theorem 38.3.4. We proceed by induction on $\min\{m, n\}$. By Proposition 38.3.3, there is $A' \in \mathrm{M}_{m,n}(D)$ such that $A' \sim A$ and $A' = \begin{pmatrix} d_1 & 0 \\ 0 & d_1 B' \end{pmatrix}$ for some B' in $\mathrm{M}_{(n-1),(m-1)}(D)$. From this, the base of induction immediately follows. So we focus on the induction step. If $B' = 0$, then we are done. If not, then by the induction hypothesis, there are $d'_2, \ldots, d'_r \in D$ such that $B' \sim \begin{pmatrix} \mathrm{diag}(d'_2, \ldots, d'_r) & 0 \\ 0 & 0 \end{pmatrix}$ and $d'_2 | d'_3 | \cdots | d'_r$. Now we make two comments. First, multiplication by d_1 commutes with row and column reduction operations. Second, the row and the column reduction operations that are applied to $d_1 B'$ can be extended to A', and these operations do not change the first row and the first column. This holds because the only non-zero entry of A' in the first row and the first column is in the $(1,1)$-position. Hence

$$A' \sim \begin{pmatrix} d_1 & 0 & 0 \\ 0 & \mathrm{diag}(d_1 d'_2, \ldots, d_1 d'_r) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and so $d_1, d_2 = d_1 d_2', \ldots, d_r := d_1 d_r'$ satisfy the desired properties. This completes the proof. □

Chapter 39

Lecture 15

In this section, first we give a classification of finitely generated modules over a Euclidean domain. This will be done using the fact that every finitely generated module over a Noetherian ring is isomorphic to the cokernel of a matrix (see Proposition 38.2.2) and existence of a Smith normal form over Euclidean domains (see Theorem 38.3.4). Next using this classification, we obtain a classification of finitely generated and finite abelian groups. Finally we further study linear transformations (over a field), and point out their connections with ring and module theory. Later these connections will be exploited to prove the existence of a rational canonical form of a matrix.

39.1 Finitely generated modules over a Euclidean Domain

By Proposition 38.2.2, we know that every finitely generated module over a Noetherian ring is isomorphic to the cokernel of a matrix. The next lemma shows that applying row and column reduction operations do not change the cokernel of a matrix up to an isomorphism. This result will enable us to use a Smith normal form to study the cokernel of a matrix over a Euclidean domain.

Lemma 39.1.1. *Suppose R is a unital commutative ring and $A, A' \in \mathrm{M}_{n,m}(R)$. Suppose there are $U \in \mathrm{GL}_n(R)$ and $V \in \mathrm{GL}_m(R)$ such that $A' = UAV$. Then $\mathrm{Coker}(A) \simeq \mathrm{Coker}(A')$. In particular, if $A \sim A'$, then $\mathrm{Coker}(A) \simeq \mathrm{Coker}(A')$ (recall that $\sim$ means that A' can be obtained from A by a series of row and column reduction operations; see Lemma 38.3.2)*

Proof. Note that $\mathrm{Im}\, A' = A'R^m = UAVR^m$. Since V is invertible, $VR^m = R^m$. Hence
$$\mathrm{Im}\, A' = UAR^m = U\, \mathrm{Im}\, A. \tag{39.1}$$
Because U is invertible, $\ell_U : R^n \to R^n, \ell_U(\mathbf{x}) := U\mathbf{x}$ is an R-module isomorphism. Let $f : R^n \to \mathrm{Coker}\, A'$, $f(\mathbf{x}) := \ell_U(\mathbf{x}) + \mathrm{Im}\, A'$. Since f is the composite of two surjective R-module homomorphisms ℓ_U and the natural quotient map $R^n \to R^n/\,\mathrm{Im}(A')$, it is surjective. Notice that $\mathbf{x} \in \ker f$ exactly when $\ell_U(\mathbf{x}) \in \mathrm{Im}\, A'$. Hence by (39.1), we obtain that the kernel of f is $\mathrm{Im}\, A$. Therefore by the first isomorphism theorem, $R^n/\,\mathrm{Im}\, A \simeq \mathrm{Coker}\, A'$. This means $\mathrm{Coker}\, A \simeq \mathrm{Coker}\, A$.

By Lemma 38.3.1, if $A \sim A'$, then there are $U \in E_n^{\pm}(R)$ and $V \in E_m^{\pm}(R)$, then $A' = UAV$. As $E_n^{\pm}(R) \subseteq \mathrm{GL}_n(R)$ and $E_m^{\pm}(R) \subseteq \mathrm{GL}_m(R)$, by the first part we conclude that $\mathrm{Coker}(A) \simeq \mathrm{Coker}(A')$. This completes the proof. □

Theorem 39.1.2. *Suppose D is a Euclidean domain and M is a finitely generated D-module. Then there are non-negative integer n and $d_1, \ldots, d_r \in D \setminus \{0\}$ such that $d_1|d_2|\cdots|d_r$ and*

$$M \simeq D^n \times \frac{D}{\langle d_1\rangle} \times \cdots \times \frac{D}{\langle d_r\rangle}.$$

Proof. Every Euclidean domain is a PID (see Theorem 7.4.2). By Corollary 12.3.6, every PID is Noetherian. Hence by Proposition 38.2.2, M is isomorphic to the cokernel of some $A_M \in \mathrm{M}_{k,m}(R)$. Suppose $\begin{pmatrix}\mathrm{diag}(d_1,\ldots,d_r) & 0\\ 0 & 0\end{pmatrix}$ is a Smith normal form of A_M; this means $d_1|d_2|\cdots|d_r$ and

$$A_M \sim \begin{pmatrix}\mathrm{diag}(d_1,\ldots,d_r) & 0\\ 0 & 0\end{pmatrix}$$

(see Theorem 38.3.4). By Lemma 39.1.1, we obtain that

$$M \simeq \mathrm{Coker}(A_M) \simeq \mathrm{Coker}\begin{pmatrix}\mathrm{diag}(d_1,\ldots,d_r) & 0\\ 0 & 0\end{pmatrix}. \tag{39.2}$$

Notice that the image of $\begin{pmatrix}\mathrm{diag}(d_1,\ldots,d_r) & 0\\ 0 & 0\end{pmatrix}$ is equal to

$$\langle d_1\rangle \times \cdots \times \langle d_r\rangle \times \{0\} \times \cdots \times \{0\}. \tag{39.3}$$

By (39.2) and (39.3), we conclude that

$$M \simeq \frac{D \times \cdots \times D \times D \times \cdots \times D}{\langle d_1\rangle \times \cdots \times \langle d_r\rangle \times \{0\} \times \cdots \times \{0\}} \simeq \frac{D}{\langle d_1\rangle} \times \cdots \times \frac{D}{\langle d_r\rangle} \times D^n.$$

This completes the proof. □

39.2 Finitely generated abelian groups

In this section, we classify finitely generated abelian groups.

Theorem 39.2.1. *Every finitely generated abelian group is isomorphic to*

$$\mathbb{Z}^n \times \mathbb{Z}_{d_1} \times \cdots \times \mathbb{Z}_{d_r}$$

for some positive integer n and d_i's such that $d_1|d_2|\cdots|d_r$.

Proof. Every finitely generated abelian group A is a finitely generated $\mathbb{Z}$-module. Since $\mathbb{Z}$ is a Euclidean domain, by Theorem 39.1.2 there are non-negative integer n and $d_1, \ldots, d_r \in \mathbb{Z} \setminus \{0\}$ such that $d_1 | cdots | d_r$ and

$$A \simeq \mathbb{Z}^n \times \frac{\mathbb{Z}}{\langle d_1 \rangle} \times \cdots \times \frac{\mathbb{Z}}{\langle d_r \rangle}. \tag{39.4}$$

Since $\langle d_i \rangle = \langle -d_i \rangle$ for every i, without loss of generality we can and will assume that d_i's are positive integers. Hence $A \simeq \mathbb{Z}^n \times \mathbb{Z}_{d_1} \times \cdots \times \mathbb{Z}_{d_r}$ as abelian groups. This completes the proof. □

Corollary 39.2.2 (Classification of finite abelian groups)**.** *Suppose A is a finite abelian group. Then $A \simeq \mathbb{Z}_{d_1} \times \cdots \times \mathbb{Z}_{d_r}$ for some positive integers d_i's such that $d_1 | \cdots | d_r$.*

Proof. By Theorem 39.2.1,

$$A \simeq \mathbb{Z}^n \times \mathbb{Z}_{d_1} \times \cdots \times \mathbb{Z}_{d_r}$$

for some positive integer n and d_i's such that $d_1 | d_2 | \cdots | d_r$. Since A is finite, we conclude that $n = 0$. This completes the proof. □

39.3 Linear transformations and matrices

In this section, we discuss the connection between linear transformations and matrices. Having an arsenal of results from ring and module theory in our disposal, we will prove the existence of a rational canonical from, the Cayley-Hamilton theorem, and further connections between the minimal and the characteristic polynomials of a linear transformation.

Suppose F is a field and V is a finite dimensional vector space over F. For a given F-basis $\mathfrak{B} := (v_1, \ldots, v_n)$ of V, we get an F-isomorphism between V and the set $\mathrm{M}_{n,1}(F)$ of column vectors. Hence

$$\ell_{\mathfrak{B}} : \mathrm{M}_{n,1}(F) \to V, \ \ell_{\mathfrak{B}} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} := \sum_{i=1}^{n} c_i v_i$$

is an F-isomorphism. For every $v \in V$, we let $[v]_{\mathfrak{B}} := \ell_{\mathfrak{B}}^{-1}(v)$. This means

$$[v]_{\mathfrak{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \quad \text{if} \quad v = c_1 v_1 + \cdots + c_n v_n.$$

Notice that

$$[v_i]_{\mathfrak{B}} = \mathbf{e}_i, \tag{39.5}$$

where $\mathbf{e}_i$ is an n-by-1 column vector with 1 in its i-th component and zero in every other. Sometimes we simply write F^n instead of $\mathrm{M}_{n,1}(F)$.

Suppose $T : V \to V$ is an F-linear transformation. Then $T(v_i)$'s uniquely determine T as we have

$$T(v) = \sum_{i=1}^{n} c_i T(v_i), \quad \text{where} \quad \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = [v]_{\mathfrak{B}}.$$

Hence T is uniquely determined by n column vectors $[T(v_1)]_{\mathfrak{B}}, \ldots, [T(v_n)]_{\mathfrak{B}}$. We put these column vectors together and get an n-by-n matrix and denote it by $[T]_{\mathfrak{B}}$. This means

$$[T]_{\mathfrak{B}} = [t_{ij}] \quad \text{if} \quad T(v_j) = t_{1j}v_1 + \cdots + t_{nj}v_j \quad \text{for every } j \in [1..n].$$

Since the j-th column of $[T]_{\mathfrak{B}}$ is $[T(v_j)]_{\mathfrak{B}}$, we have $[T(v_j)]_{\mathfrak{B}} = [T]_{\mathfrak{B}}\mathbf{e}_j$, and so by (39.5), we obtain that

$$[T(v_j)]_{\mathfrak{B}} = [T]_{\mathfrak{B}}[v_j]_{\mathfrak{B}}. \tag{39.6}$$

From (39.6), we conclude that for every c_i's in F, we have

$$\begin{aligned} [T(\sum_{j=1}^{n} c_j v_j)]_{\mathfrak{B}} &= [\sum_{j=1}^{n} c_j T(v_j)]_{\mathfrak{B}} = \sum_{j=1}^{n} c_j [T(v_j)]_{\mathfrak{B}} \\ &= \sum_{j=1}^{n} c_j [T]_{\mathfrak{B}}[v_j]_{\mathfrak{B}} = [T]_{\mathfrak{B}}[\sum_{j=1}^{n} c_i v_i]_{\mathfrak{B}}. \end{aligned} \tag{39.7}$$

By (39.7), we deduce that

$$[T(v)]_{\mathfrak{B}} = [T]_{\mathfrak{B}}[v]_{\mathfrak{B}}$$

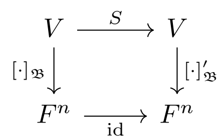
for every $v \in V$. This is equivalent to saying that the following is a commutative diagram.

$$\begin{array}{ccc} V & \xrightarrow{\;T\;} & V \\ {\scriptstyle [\cdot]_{\mathfrak{B}}}\big\downarrow & & \big\downarrow{\scriptstyle [\cdot]_{\mathfrak{B}}} \\ F^n & \xrightarrow[{[T]_{\mathfrak{B}}\times}]{} & F^n \end{array}$$

where the bottom row means multiplication by $[T]_{\mathfrak{B}}$, i.e. $\mathbf{x} \mapsto [T]_{\mathfrak{B}}\mathbf{x}$. We call $[T]_{\mathfrak{B}}$ the *matrix representation* of T with respect to the basis $\mathfrak{B}$.

From encoding point of view, the more sparse the matrix representation $[T]_{\mathfrak{B}}$ is the better. Here by sparse, we mean having many zero entries. A sparse matrix representation can help us from both computational and theoretical point of view. Later we will come back to the problem of finding a sparse matrix representation of a linear transformation. For now we want to understand the connection between different matrix representations of a given linear transformation.

Suppose $\mathfrak{B} := (v_1, \ldots, v_n)$ and $\mathfrak{B}' := (v'_1, \ldots, v'_n)$ are two F-bases of V. Since v_i's are linearly independent, there is an F-linear map $S : V \to V$ such that $S(v_i) := v'_i$ for every i. Because the F-span of v'_i's is V, S is surjective. As v'_i's are F-linearly independent, S is injective. Hence $S : V \to V$ is an isomorphism. Moreover since $[v_i]_{\mathfrak{B}} = [v'_i]_{\mathfrak{B}} = \mathbf{e}_i$ for every i, we have that the following is a commutative diagram.

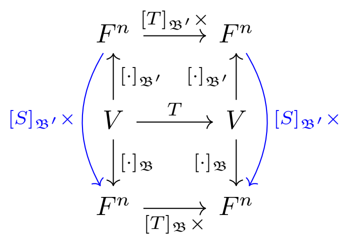


This means we have

$$[S(v)]_{\mathfrak{B}'} = [v]_{\mathfrak{B}} \tag{39.8}$$

for every $v \in V$. Therefore $[S]_{\mathfrak{B}'}[v]_{\mathfrak{B}'} = [v]_{\mathfrak{B}}$ for all $v \in V$.

For every linear transformation $T : V \to V$, the following is a commutative diagram.



Hence $[T]_{\mathfrak{B}'} = [S]_{\mathfrak{B}'}^{-1}[T]_{\mathfrak{B}}[S]_{\mathfrak{B}'}$. This means every two matrix representations of a given linear transformation are conjugate of each other.

Conversely, we show that every conjugate of $[T]_{\mathfrak{B}}$ is a matrix representation of T. Suppose $S := [s_{ij}] \in \mathrm{GL}_n(F)$, and let $v'_j := \sum_{i=1}^n s_{ij}v_j$. Then $\mathfrak{B}' := (v'_1, \ldots, v'_n)$ is an F-basis of V, and

$$[v'_j]_{\mathfrak{B}} = S[v'_j]_{\mathfrak{B}'}$$

for every j. Hence $[v]_{\mathfrak{B}} = S[v]_{\mathfrak{B}'}$ for all $v \in V$. Thus by the above discussion, $[T]_{\mathfrak{B}'} = S^{-1}[T]_{\mathfrak{B}}S$, and the converse follows.

Therefore the task of finding a simple matrix representation can be interpreted as a search for a sparse conjugate of a matrix.

A linear transformation $T : V \to V$ is also called an F-*endomorphism* of V, and the set of all endormorphisms of V is denoted by $\mathrm{End}_F(V)$. The following lemma shows that $\mathrm{End}_F(V)$ is a ring which is isomorphic to $\mathrm{M}_n(F)$.

Lemma 39.3.1. *Suppose F is a field and V is a vector space over F. Suppose $\mathfrak{B} := (v_1, \ldots, v_n)$ is an F-basis of V. Then*

$$\mathrm{End}_F(V) \to \mathrm{M}_n(F),\ T \mapsto [T]_{\mathfrak{B}}$$

is a bijection, and its inverse is given by

$$\mathrm{M}_n(F) \to \mathrm{End}_F(V),\ A \mapsto \lambda_A,$$

where $\lambda_A(v) := \ell_{\mathfrak{B}}(A[v]_{\mathfrak{B}})$. Moreover, for every $T_1, T_2 \in \mathrm{End}_F(V)$, we have

$$[T_1 + T_2]_{\mathfrak{B}} = [T_1]_{\mathfrak{B}} + [T_2]_{\mathfrak{B}} \quad \textit{and} \quad [T_1 \circ T_2]_{\mathfrak{B}} = [T_1]_{\mathfrak{B}}[T_2]_{\mathfrak{B}}.$$

In particular, $(\mathrm{End}_F(V), +, \circ)$ is a ring and it is isomorphic to $\mathrm{M}_n(F)$.

Proof. Since $\ell_{\mathfrak{B}}$, multiplication by A, and $[\cdot]_{\mathfrak{B}}$ are F-linear, $\lambda_A \in \mathrm{End}_F(V)$. Because $\ell_{\mathfrak{B}}$ and $[\cdot]_{\mathfrak{B}}$ are inverse of each other, for all $A \in \mathrm{M}_n(F)$ and $v \in V$, we have

$$[\lambda_A]_{\mathfrak{B}}[v]_{\mathfrak{B}} = [\lambda_A(v)]_{\mathfrak{B}} = A[v]_{\mathfrak{B}}.$$

Hence $[\lambda_A]_{\mathfrak{B}} = A$.

Similarly for all $T \in \mathrm{End}_F(V)$ and $v \in V$, we obtain

$$[\lambda_{[T]_{\mathfrak{B}}}(v)]_{\mathfrak{B}} = [T]_{\mathfrak{B}}[v]_{\mathfrak{B}} = [T(v)]_{\mathfrak{B}},$$

which implies that $\lambda_{[T]_{\mathfrak{B}}}(v) = T(v)$ for all $v \in V$. Hence $\lambda_{[T]_{\mathfrak{B}}} = T$. Altogether, we conclude that $T \mapsto [T]_{\mathfrak{B}}$ is a bijection and $A \mapsto \lambda_A$ is its inverse.

For every $v \in V$, we have

$$\begin{aligned}[T_1 + T_2]_{\mathfrak{B}}[v]_{\mathfrak{B}} =& [(T_1 + T_2)(v)]_{\mathfrak{B}} = [T_1(v) + T_2(v)]_{\mathfrak{B}} \\ =& [T_1(v)]_{\mathfrak{B}} + [T_2(v)]_{\mathfrak{B}} = [T_1]_{\mathfrak{B}}[v]_{\mathfrak{B}} + [T_2]_{\mathfrak{B}}[v]_{\mathfrak{B}} \\ =& ([T_1]_{\mathfrak{B}} + [T_2]_{\mathfrak{B}})[v]_{\mathfrak{B}},\end{aligned}$$

and so $[T_1 + T_2]_{\mathfrak{B}} = [T_1]_{\mathfrak{B}} + [T_2]_{\mathfrak{B}}$. Similarly we have

$$\begin{aligned}[T_1 \circ T_2]_{\mathfrak{B}}[v]_{\mathfrak{B}} =& [(T_1 \circ T_2)(v)]_{\mathfrak{B}} = [T_1(T_2(v))]_{\mathfrak{B}} \\ =& [T_1]_{\mathfrak{B}}[T_2(v)]_{\mathfrak{B}} = [T_1]_{\mathfrak{B}}([T_2]_{\mathfrak{B}}[v]_{\mathfrak{B}}) \\ =& ([T_1]_{\mathfrak{B}}[T_2]_{\mathfrak{B}})[v]_{\mathfrak{B}},\end{aligned}$$

which implies that $[T_1 \circ T_2]_{\mathfrak{B}} = [T_1]_{\mathfrak{B}}[T_2]_{\mathfrak{B}}$.

The rest of the claims immediately follow. □

39.4 Linear maps, evaluation map, and minimal polynomial

In the study of zeros of polynomials, we saw the importance of evaluation maps. In this section, we employ a similar idea to study linear transformations.

Suppose $T \in \mathrm{End}_F(V)$, and let $\phi_T : F[x] \to \mathrm{End}_F(V),\ \phi_T(f) := f(T)$ be the evaluation at T map. Since scalar transformations are in the center of $\mathrm{End}_F(V)$, the same argument as in the proof of Lemma 3.1.1 implies that ϕ_T is a ring homomorphism. The image of ϕ_T is denoted by $F[T]$. Notice that $F[T]$ is a commutative subring of $\mathrm{End}_F(V)$ which is isomorphic to $F[x]/\ker \phi_T$.

Notice that by Lemma 39.3.1, we have $\dim_F \mathrm{End}_F(V) = \dim_F \mathrm{M}_n(F) = n^2$. Hence $\dim_F F[T] \le n^2 < \infty$. Therefore $\dim_F(\frac{F[x]}{\ker \phi_T}) < \infty$, which implies that $\ker \phi_T \ne 0$. Because scalar endomorphisms are in $F[T]$, $\dim_F F[T] \ge 1$. Thus $\ker \phi_T$ is a proper ideal. Since $F[x]$ is a PID and $\ker \phi_T$ is a proper non-zero ideal, there is a unique monic non-constant polynomial $m_{T,F} \in F[x]$ such that

$$\ker \phi_T = \langle m_{T,F} \rangle.$$

The polynomial $m_{T,F}$ is called the *minimal polynomial* of T. Altogether we have proved the following.

Lemma 39.4.1. *Suppose F is a field and V is a finite dimensional F-vector space. For an F-linear map $T : V \to V$, there is a unique non-constant monic polynomial $m_{T,F} \in F[x]$ such that for every $p(x) \in F[x]$, we have $p(T) = 0$ if and only if $m_{T,F}(x)|p(x)$.*

For a given basis $\mathfrak{B} := (v_1, \ldots, v_n)$, by Lemma 39.3.1, $\mathrm{M}_n(F) \to \mathrm{End}_F(V) A \mapsto \lambda_A$ is a ring isomorphism. Let $m_{A,F}(x) := m_{\lambda_A,F}(x)$ and call it a *minimal polynomial* of A. A priori it is not clear why $m_{A,T}$ does not depend on the choice of basis in the definition of $A \mapsto \lambda_A$. The next lemma, however, shows that $m_{A,F}$ only depends on A.

Lemma 39.4.2. *Suppose $A \in \mathrm{M}_n(F)$, and let $\phi_A : F[x] \to \mathrm{M}_n(F), \phi_A(f) := f(A)$. Then the following statements hold.*

1. *The minimal polynomial $m_{A,F}$ is the unique monic generator of $\ker \phi_A$; in particular, it only depends on A.*
2. *The image $F[A]$ of ϕ_A is isomorphic to $F[\lambda_A]$ where $\lambda_A \in \mathrm{End}_F(F^n)$ is given by $\lambda_A(\mathbf{x}) := A\mathbf{x}$.*
3. *If $S \in \mathrm{GL}_n(F)$, then $m_{S^{-1}AS,F} = m_{A,F}$.*

Proof. 1. Since $\mathrm{M}_n(F) \to \mathrm{End}_F(F^n), A \mapsto \lambda_A$ is a ring isomorphism, for every $f \in F[x]$ we have

$$\lambda_{f(A)} = f(\lambda_A).$$

Hence $f \in \ker \phi_A$ if and only if $f \in \ker \phi_{\lambda_A}$. Part 1 follows.

2. By the first isomorphism theorem, $F[A] \simeq F[x]/\ker \phi_A$. By part 1, $\ker \phi_A = \ker \phi_{\lambda_A}$, and so by another application of the first isomorphism theorem we conclude that

$$F[x]/\ker \phi_A = F[x]/\ker \phi_{\lambda_A} \simeq F[\lambda_A].$$

This implies the second part.

3. Since both A and $S^{-1}AS$ are matrix representations of λ_A, we deduce that

$$m_{A,F} = m_{\lambda_A,F} = m_{S^{-1}AS,F},$$

which completes the proof. □

Chapter 40

Lecture 16

In this section, we use the classification of finitely generated modules over a Euclidean domain to find a fairly sparse matrix representation called a *rational canonical form* of a linear transformation $T : V \to V$ where V is a finite dimensional vector space over F. Along the way, we also prove the Cayley-Hamilton theorem and other properties of the characteristic and minimal polynomials of T.

40.1 Linear maps, evaluation map, module structure

As we have seen in Section 39.4, for a given linear transformation $T : V \to V$, we can consider the evaluation map $\phi_T : F[x] \to \mathrm{End}_F(V)$, and ϕ_T is a ring homomorphism. Via the ring homomorphism ϕ_T, we can view V as an $F[x]$-module. Here is a precise formulation of this result.

Lemma 40.1.1. *Suppose F is a field and V is a finite dimensional vector space over F. Let $T : V \to V$ is an F-linear map. For $f \in F[x]$ and $v \in V$, let*

$$f \cdot v := \phi_T(f)v = f(T)(v). \tag{40.1}$$

Then V is an $F[x]$-module with respect to the scalar multiplication given by (40.1)

Proof. It is straightforward to check that the scalar multiplication given in (40.1) satisfies all the properties of modules. Here we go through only one of the distribution properties. Suppose $f_1, f_2 \in F[x]$ and $v \in V$. Then

$$\begin{aligned}(f_1 + f_2) \cdot v =& \phi_T(f_1 + f_2)(v) = (\phi_T(f_1) + \phi_T(f_2))(v) \\ =& \phi_T(f_1)(v) + \phi_T(f_2)(v) = f_1 \cdot v + f_2 \cdot v.\end{aligned}$$

□

Notice that $F[x]$-module structure of V and its F-vector space structure are compatible. This is the case because constant polynomials are sent to scalar linear transformations: for every $c \in F$ and $v \in V$, we have $\phi_T(c)(v) = cv$ where cv is the scalar product of c by v as an F-vector space. By this remark, we deduce that, if

$\mathfrak{B} = (v_1, \ldots, v_n)$ is an F-basis of V, then v_i's generate V as an $F[x]$-module as well. Hence V is a finitely generated $F[x]$-module. Since $F[x]$ is a Euclidean domain, by Theorem 39.1.2 there are a non-negative integer k and $d_1, \ldots, d_r \in F[x] \setminus \{0\}$ such that

$$V \simeq F[x]^k \times \frac{F[x]}{\langle d_1 \rangle} \times \cdots \times \frac{F[x]}{\langle d_r \rangle} \tag{40.2}$$

as an $F[x]$-module and $d_1 | \cdots | d_r$. Since V is a finite dimensional vector space over F and $F[x]$ is an infinite dimensional vector space over F, by (40.2) we conclude that $k = 0$. Thus

$$V \simeq \frac{F[x]}{\langle d_1 \rangle} \times \cdots \times \frac{F[x]}{\langle d_r \rangle}. \tag{40.3}$$

Notice that if d_i is a non-zero constant, then $\langle d_i \rangle = F[x]$ and $\frac{F[x]}{\langle d_i \rangle} = 0$, and so such a factor can be dropped. Let's also observed that multiplying d_i's by a non-zero constant does not change the ideals $\langle d_i \rangle$'s. Hence without loss of generality we can and will assume that d_i's are monic non-constant polynomials. Let's summarize what we have obtained so far.

Lemma 40.1.2. *Suppose F is a field and V is a finite dimensional F-vector space. Suppose $T : V \to V$ is an F-linear map. Then V can be viewed as an $F[x]$-module with the scalar multiplication given in (40.1), and there are non-constant monic polynomials $d_1, \ldots, d_r \in F[x]$ such that $d_1 | \cdots | d_r$ and*

$$V \simeq \frac{F[x]}{\langle d_1 \rangle} \times \cdots \times \frac{F[x]}{\langle d_r \rangle}.$$

40.2 Reduction to the cyclic case.

In this section, we see how the module structure of V given in Lemma 40.1.2 can help us get a matrix representation of T that is of block form and its non-zero blocks can only be on the diagonal. Moreover the associated blocks are related to the case where the vector space is isomorphic to $F[x]/\langle d \rangle$ for some d as an $F[x]$-module. Notice that by Lemma 37.3.2, an $F[x]$-module is isomorphic to $F[x]/\langle d \rangle$ exactly when it is a cyclic $F[x]$-module. Altogether, we are reducing the problem of finding a sparse matrix representation of T to the case where V is a cyclic $F[x]$-module.

Suppose d_i's are given as in Lemma 40.1.2, and

$$\theta : \frac{F[x]}{\langle d_1 \rangle} \times \cdots \times \frac{F[x]}{\langle d_r \rangle} \to V$$

is an $F[x]$-module isomorphism. Let $M := \frac{F[x]}{\langle d_1 \rangle} \times \cdots \times \frac{F[x]}{\langle d_r \rangle}$. Then for all $m \in M$ we have

$$T(\theta(m)) = x \cdot \theta(m) = \theta(x \cdot m).$$

This means that the following is a commutative diagram.

$$\begin{array}{ccc} M & \xrightarrow{x\times} & M \\ {\scriptstyle\theta}\big\downarrow & & \big\downarrow{\scriptstyle\theta} \\ V & \xrightarrow[T]{} & V \end{array}$$

Hence if $\mathfrak{B} := (m_1, \ldots, m_n)$ is an F-basis of M, then $\theta(\mathfrak{B}) := (\theta(m_1), \ldots, \theta(m_n))$ is an F-basis of V and $[x\times]_{\mathfrak{B}} = [T]_{\theta(\mathfrak{B})}$. Roughly this diagram implies that after identifying M with V, $x\times$ gets identified with T. Hence $x\times$ and T share matrix representations. Here is a more formal argument of why the last equality holds: suppose $[x\times]_{\mathfrak{B}} = [r_{ij}]$. Then $x \cdot m_j = \sum_{i=1}^n r_{ij} m_i$ for all j. As θ is an $F[x]$-module homomorphism, we obtain that $\theta(x \cdot m_j) = x \cdot \theta(m_j) = T(\theta(m_j))$. Therefore

$$T(\theta(m_j)) = \theta(x \cdot m_j) = \theta(\sum_{i=1}^n r_{ij} m_i) = \sum_{i=1}^n r_{ij}\theta(m_i),$$

which implies that $[T]_{\theta(\mathfrak{B})} = [r_{ij}]$.

Since every matrix representation of $x\times$ is a matrix representation of T (and vice versa), we will find a sparse matrix representation of $x\times$.

Let $M_i := 0 \times \cdots \times \frac{F[x]}{\langle d_i \rangle} \times \cdots \times 0$ for all i. Notice that M_i's are $F[x]$-submodules of M, $M = M_1 + \cdots + M_r$, and if $v_i \in M_i \setminus \{0\}$, then v_i's are linearly independent. Hence if $\mathfrak{B}_i$ is an F-basis of M_i, then $\mathfrak{B} := \mathfrak{B}_1 \cup \cdots \cup \mathfrak{B}_r$ (with this ordering) is an F-basis of M. Because M_i is an $F[x]$-submodule, for every $m \in \mathfrak{B}_i$ the element $x \cdot m$ is an F-linear combination of elements of $\mathfrak{B}_i$. Hence $[x\times]_{\mathfrak{B}}$ is a matrix of the following form:

$$\begin{pmatrix} [x\times]_{\mathfrak{B}_1} & 0 & \cdot & 0 \\ 0 & [x\times]_{\mathfrak{B}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [x\times]_{\mathfrak{B}_r} \end{pmatrix} \tag{40.4}$$

Thus to make this matrix representation more sparse, we should focus on $x\times$ for a cyclic $F[x]$-module of the form $F[x]/\langle d \rangle$ where d is a non-constant monic polynomial.

40.3 Cyclic case, companion matrix, and rational canonical form

In this section, we find a nice matrix representation of $x\times$ for a cyclic $F[x]$-module of the form $F[x]/\langle d(x) \rangle$ where $d(x) := x^m + a_{m-1}x^{m-1} + \cdots + a_0 \in F[x]$. We also find its minimal polynomial.

By Proposition 20.1.1, $\mathfrak{B} := (\overline{1}, \overline{x}, \ldots, \overline{x}^{m-1})$ is an F-basis of $\frac{F[x]}{\langle d(x) \rangle}$ where $\overline{x}^i = x^i + \langle d \rangle$ for every non-negative integer i. Notice that

$$\overline{1} \xmapsto{x\times} \overline{x} \xmapsto{x\times} \overline{x}^2 \xmapsto{x\times} \cdots \xmapsto{x\times} \overline{x}^{m-1} \xmapsto{x\times} \overline{x}^m. \tag{40.5}$$

For every $i \in [0..(m-1)]$, we have $[\overline{x}^i]_{\mathfrak{B}} = \mathbf{e}_{i+1}$ where $(\mathbf{e}_1, \ldots, \mathbf{e}_m)$ is the standard basis of F^m. We also notice that

$$\overline{x}^m + a_{m-1}\overline{x}^{m-1} + \cdots + a_0 = 0$$

which implies that

$$[\overline{x}^m]_{\mathfrak{B}} = \Big[-\sum_{i=0}^{m-1} a_i \overline{x}^i\Big]_{\mathfrak{B}} = \sum_{i=0}^{m-1}(-a_i)\mathbf{e}_{i+1} = \begin{pmatrix} -a_0 \\ \vdots \\ -a_{m-1} \end{pmatrix}. \tag{40.6}$$

By (40.5) and (40.6), we obtain that

$$[x\times]_{\mathfrak{B}} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -a_{m-2} \\ 0 & 0 & 0 & \cdots & 1 & -a_{m-1} \end{pmatrix}. \tag{40.7}$$

Let $C(d)$ be the matrix given in (40.7). We call $C(d)$ the *companion matrix* of the monic polynomial $d(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_0$.

To find the minimal polynomial of $x\times : \frac{F[x]}{\langle d\rangle} \to \frac{F[x]}{\langle d\rangle}$, we need to find the kernel of the evaluation map $\phi_{x\times}$. Notice that $(x\times)^i = x^i\times$ for every non-negative integer i. Hence for every polynomial $f(x) \in F[x]$, we have $\phi_{x\times}(f) = f(x)\times$. This means for every $g + \langle d\rangle \in \frac{F[x]}{\langle d\rangle}$, we have

$$\phi_{x\times}(f)(g + \langle d\rangle) = f(x)g(x) + \langle d\rangle.$$

Hence $f \in \ker \phi_{x\times}$ if and only if $d|fg$ for every $g \in F[x]$. Hence $\ker \phi_{x\times} = \langle d\rangle$. Because d is a monic polynomial which generates $\ker \phi_{x\times}$, we conclude that the minimal polynomial of $x\times$ is d. Therefore we obtain the following result.

Lemma 40.3.1. *Suppose $d \in F[x]$ is a monic non-constant polynomial. Then the minimal polynomial of the companion matrix $C(d)$ is $d(x)$.*

By (40.4) and (40.7), we obtain the following result.

Theorem 40.3.2 (Rational canonical form)**.** *Suppose V is a finite dimensional vector space over a field F, and $T : V \to V$ is an F-linear map. Then T has a matrix representation of the form*

$$\begin{pmatrix} C(d_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C(d_r) \end{pmatrix} \tag{40.8}$$

for some monic non-constant polynomials d_i's in $F[x]$ such that $d_1|\cdots|d_r$.

The matrix given in (40.8) is called a *rational canonical form* of T.

40.4 The Cayley-Hamilton Theorem

In this section, we define the characteristic polynomial of a linear map, give a connection with a rational canonical form of a linear map, and prove the Cayley-Hamilton Theorem.

Definition 40.4.1. *Suppose R is a unital commutative ring and $A \in \mathrm{M}_n(R)$. The characteristic polynomial of A is $f_A(x) := \det(xI - A)$ where $xI - A \in \mathrm{M}_n(R[x])$. Suppose F is a field, V is a finite dimensional vector space over F, and $\mathfrak{B}$ is an F-basis of V. The characteristic polynomial of an F-linear map $T : V \to V$ with respect to the basis $\mathfrak{B}$ is $f_{[T]_{\mathfrak{B}}}(x)$.* [1]

Lemma 40.4.2. *Suppose R is a unital commutative ring, F is a field and V is a finite dimensional vector space over F.*

1. *If $A, A' \in \mathrm{M}_n(R)$ are conjugate of each other, then $f_A(x) = f_{A'}(x)$.*
2. *If $\mathfrak{B}$ and $\mathfrak{B}'$ are two F-bases of V, then the characteristic polynomials of T with respect to $\mathfrak{B}$ and $\mathfrak{B}'$ are the same.*

Proof. (1) Since A and A' are conjugate of each other, there is $S \in \mathrm{GL}_n(R)$ such that $A' = SAS^{-1}$. Because determinant is multiplicative we have

$$\begin{aligned}\det(A') &= \det(SAS^{-1}) = \det(S)\det(A)\det(S^{-1})\\ &= \det(S)\det(S^{-1})\det(A) = \det(SS^{-1})\det(A)\\ &= \det(I)\det(A) = \det(A).\end{aligned}$$

(2) Since $[T]_{\mathfrak{B}}$ and $[T]_{\mathfrak{B}'}$ are conjugate of each other, by part (1), $f_{[T]_{\mathfrak{B}}} = f_{[T]_{\mathfrak{B}'}}$. This completes the proof. □

By the second part of Lemma 40.4.2, we deduce that the characteristic polynomial of a linear map does not depend on the choice of a basis of the vector space. The characteristic polynomial of a linear map $T : V \to V$ is denoted by f_T.

Lemma 40.4.3. *Suppose $d(x) \in F[x]$ is a monic polynomial and $C(d)$ is its the companion matrix. Then the characteristic polynomial of $C(d)$ is d.*

Proof. We proceed by induction on the degree of d. If $\deg d = 1$, then $d(x) = x + a$ and $C(d) = [-a]$. So $f_{C(d)}(x) = \det([x + a]) = x + a$. Next we prove the induction

[1] Look at the HW assignment week 8 for the definition and properties of determinant.

step. We expand the determinant of $xI - C(d)$ with respect to the first column. Hence

$$\det\begin{pmatrix} x & 0 & \cdots & a_0 \\ -1 & x & \cdots & a_1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x+a_{m-1}\end{pmatrix} = x\det\begin{pmatrix} x & 0 & \cdots & a_1 \\ -1 & x & \cdots & a_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x+a_{m-1}\end{pmatrix} + \det\begin{pmatrix} 0 & 0 & \cdots & a_0 \\ -1 & x & \cdots & a_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x+a_{m-1}\end{pmatrix}$$

We use the induction hypothesis for the first term, and expand the determinant of the second term with respect to the first row. We obtain the following.

$$\begin{aligned} f_{C(d)}(x) =& x(x^{m-1} + a_{m-1}x^{m-2} + \cdots + a_1) \\ &+(-1)^{1+(m-1)}a_0 \det\begin{pmatrix} -1 & x & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1\end{pmatrix} \\ =& x(x^{m-1} + a_{m-1}x^{m-2} + \cdots + a_1) + (-1)^{m+m-2}a_0 \\ =& x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 = d(x). \end{aligned}$$

This completes the proof. □

Theorem 40.4.4. *Suppose $d_1, \ldots, d_r \in F[x]$ are monic polynomials, $d_1|\cdots|d_r$, and $\operatorname{diag}(C(d_1), \ldots, C(d_r))$ is a rational canonical form of a linear map $T : V \to V$. Then following statements hold.*

1. $m_{T,F}(x) = d_r(x)$.
2. $f_T(x) = d_1(x)\cdots d_r(x)$.
3. (Caley-Hamilton Theorem) $f_T(T) = 0$.
4. *Suppose $p(x) \in F[x]$ is irreducible. Then $p(x)|f_T(x)$ if and only if $p(x)|m_{T,F}(x)$.*

Proof. Let $A := \operatorname{diag}(C(d_1), \ldots, C(d_r))$. Then

$$f_T(x) = f_A(x) \quad \text{and} \quad m_{T,F}(x) = m_{A,F}(x).$$

For every polynomial $f \in F[x]$, we have

$$f(A) = \operatorname{diag}(f(C(d_1)), \ldots, f(C(d_r))).$$

Hence $f(A) = 0$ if and only if $f(C(d_i)) = 0$ for every i. By Lemma 40.3.1, $m_{C(d_i),F}(x) = d_i(x)$. Therefore $f(A) = 0$ exactly when $d_i(x)|f(x)$ for all i. Since

$d_1|\cdots|d_r$, we obtain that $f(A) = 0$ precisely when $d_r|f$. Thus $m_{A,F}(x) = d_r(x)$. Next we want to find the characteristic polynomial of A:

$$\begin{aligned} f_A(x) &= \det(\operatorname{diag}(xI - C(d_1), \dots, xI - C(d_r))) \\ &= \prod_{i=1}^{r} \det(xI - C(d_i)) = \prod_{i=1}^{r} f_{C(d_i)}(x) \qquad \text{(by Lemma 40.4.3)} \\ &= \prod_{i=1}^{r} d_i(x). \end{aligned}$$

Since $m_{T,F}(x) = d_r(x)$ and $f_T(x) = \prod_{i=1}^{r} d_i(x)$, $m_{T,F}(x)|f_T(x)$. Hence

$$f_T(T) = 0.$$

Suppose p is an irreducible factor of $f_T(x)$. Since $f_T(x) = \prod_{i=1}^{r} d_i(x)$, $p|f_T$, and p is prime, p divides d_i for some i. Since $d_i|d_r$ for all i, we conclude that $p|d_r$. As $m_{T,F}(x) = d_r(x)$, we obtain that $p|m_{T,f}$. This completes the proof. $\square$

Chapter 41

Lecture 17

In our study of zeros of multivariable polynomials, we investigated system of linear equations over unital commutative rings. We developed module theory and gave an algorithm for solving system of linear equations over a Euclidean domain. In the next step, we consider two related directions to extend our exploration:

1. Studying single variable polynomials over an arbitrary unital commutative ring.
2. Studying multivariable polynomials over a field.

Since $F[x_1, \ldots, x_n]$ is essentially $(F[x_1, \ldots, x_{n-1}])[x_n]$, it is not surprising that any result related to the first task has implications towards the second one.

Suppose A is a unital commutative ring. We would like to understand ring theoretic properties of $A[x]$. Here are some results that we have proved earlier:

1. If D is an integral domain, then $D[x]$ is an integral domain. (See Lemma 6.3.1)
2. If F is a field, then $F[x]$ is a PID. (See Theorem 7.3.1)
3. If D is a UFD, then $D[x]$ is a UFD. (See Theorem 15.4.4)

One of the key results which is instrumental in proving these results is the long division (see Theorem 6.4.1). Let's recall the long division algorithm.

Suppose A is a unital commutative ring and $g \in A[x]$ such that $\langle \mathrm{ld}(g) \rangle = A$ where $\mathrm{ld}(g)$ is the leading coefficient of g. Then for every $f \in A[x]$ there is a unique $r \in A[x]$ such that $f + \langle g \rangle = r + \langle g \rangle$ and $\deg r < \deg g$.

So the long division algorithm gives us a (canonical) simple coset representative of $f + \langle g \rangle$ if *the leading coefficient of g generates the entire ring of coefficients.*

The existence part of the long division is derived from the following algorithm.

Step 1. If $\deg f < \deg g$, we are done. Return $r := f$.

Step 2. If $\deg f \geq \deg g$, there is a monomial cx^k such that $\mathrm{Ld}(f) = \mathrm{Ld}(g)(cx^k)$. Let $f_{\text{new}} \leftarrow f_{\text{old}} - (cx^k)g$.

Here are a few comments about the above algorithm:

1. It terminates as $\deg f_{\text{new}}$ is less than $\deg f_{\text{old}}$.

2. At every step, we do not leave the coset $f + \langle g \rangle$ since $f_{\text{new}} + \langle g \rangle = f_{\text{old}} + \langle g \rangle$.

3. the key reason that it actually works is because $\mathrm{Ld}(f) \in \langle \mathrm{Ld}(g) \rangle$ unless $\deg f$ is less than $\deg g$.

There are a few ways to generalize long division. But the general idea stays the same: considering a *grading* on the given ring and *clearing* the leading term in the process.

41.1 Generalized long division

In this section, we present some generalizations of long division.

Lemma 41.1.1. *Suppose A is a unital commutative ring and $g_1, \ldots, g_n \in A[x]$ such that*

$$\langle \mathrm{ld}(g_1), \ldots, \mathrm{ld}(g_n) \rangle = A. \tag{41.1}$$

Then for every $f \in A[x]$, there is $r \in A[x]$ such that

$$f + \langle g_1, \ldots, g_n \rangle = r + \langle g_1, \ldots, g_n \rangle \quad \textit{and} \quad \deg r < \max(\deg g_1, \ldots, \deg g_n).$$

Proof. We follow the same algorithm as in the case when $n = 1$.

Step 1. If $\deg f < \max(\deg g_1, \ldots, \deg g_n)$, we are done. Return $r := f$.

Step 2. If $\deg f \geq \max(\deg g_1, \ldots, \deg g_n)$, then

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(g_1), \ldots, \mathrm{Ld}(g_k) \rangle. \tag{41.2}$$

Here is why (41.2) holds. By (41.1)

$$\mathrm{ld}(f) = a_1 \, \mathrm{ld}(g_1) + \cdots + a_n \, \mathrm{ld}(g_n)$$

for some a_i's in A. Therefore we obtain that

$$\mathrm{Ld}(f) = (a_1 x^{\deg f - \deg g_1}) \, \mathrm{Ld}(g_1) + \cdots + (a_n x^{\deg f - \deg g_n}) \, \mathrm{Ld}(g_n). \tag{41.3}$$

Suppose a_i's are as in (41.3), and let

$$f_{\text{new}} \leftarrow f_{\text{old}} - ((a_1 x^{\deg f - \deg g_1}) g_1 + \cdots + (a_n x^{\deg f - \deg g_n}) g_n).$$

Notice that $\deg f_{\text{new}} < \deg f_{\text{old}}$ and

$$f_{\text{new}} + \langle g_1, \ldots, g_n \rangle = f_{\text{old}} + \langle g_1, \ldots, g_n \rangle.$$

Hence this algorithm terminates and returns a desired r. $\square$

A closer look at the proof of Lemma 41.1.1 shows that the key properties to have are

$$\mathrm{ld}(f) \in \langle \mathrm{ld}(g_1), \ldots, \mathrm{ld}(g_n) \rangle \tag{41.4}$$

and

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(g_1), \ldots, \mathrm{Ld}(g_n) \rangle. \tag{41.5}$$

The following results are good examples of how one can use the conditions given in (41.4) and (41.5).

Lemma 41.1.2. *Suppose A is a untial commutative ring, $I \trianglelefteq A[x]$, and $g_1, \ldots, g_n \in I$ are such that*

$$\mathrm{ld}(f) \in \langle \mathrm{ld}(g_1), \ldots, \mathrm{ld}(g_n) \rangle \tag{41.6}$$

for all $f \in I$. Then for every $f \in I$, there is $r \in I$ such that

$$f + \langle g_1, \ldots, g_n \rangle = r + \langle g_1, \ldots, g_n \rangle \quad \text{and} \quad \deg r < \max(\deg g_1, \ldots, \deg g_n).$$

Lemma 41.1.3. *Suppose A is a untial commutative ring, $I \trianglelefteq A[x]$, and $g_1, \ldots, g_n \in I$ are such that*

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(g_1), \ldots, \mathrm{Ld}(g_n) \rangle. \tag{41.7}$$

Then $I = \langle g_1, \ldots, g_n \rangle$.

Proof of Lemma 41.1.2. We follow an almost identical line of argument as in the proof of Lemma 41.1.1. Consider the following algorithm:

Step 1. If $\deg f < \max(\deg g_1, \ldots, \deg g_n)$, we are done. Return $r := f$.

Step 2. If $\deg f \geq \max(\deg g_1, \ldots, \deg g_n)$, then

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(g_1), \ldots, \mathrm{Ld}(g_k) \rangle. \tag{41.8}$$

Here is why (41.8) holds. By (41.6)

$$\mathrm{ld}(f) = a_1 \,\mathrm{ld}(g_1) + \cdots + a_n \,\mathrm{ld}(g_n)$$

for some a_i's in A. Therefore we obtain that

$$\mathrm{Ld}(f) = (a_1 x^{\deg f - \deg g_1}) \,\mathrm{Ld}(g_1) + \cdots + (a_n x^{\deg f - \deg g_n}) \,\mathrm{Ld}(g_n). \tag{41.9}$$

Suppose a_i's are as in (41.9), and let

$$f_{\text{new}} \leftarrow f_{\text{old}} - ((a_1 x^{\deg f - \deg g_1}) g_1 + \cdots + (a_n x^{\deg f - \deg g_n}) g_n).$$

Notice that $\deg f_{\text{new}} < \deg f_{\text{old}}$ and

$$f_{\text{new}} + \langle g_1, \ldots, g_n \rangle = f_{\text{old}} + \langle g_1, \ldots, g_n \rangle.$$

Moreover, since f_{old} and g_i's are in I, so is f_{new}. Hence this algorithm terminates and returns a desired r. □

Proof of Lemma 41.1.3. Since g_i's are in I, $\langle g_1, \ldots, g_n\rangle \subseteq I$. Next by strong induction on degree, we show that every $f \in I$ is in $\langle g_1, \ldots, g_n\rangle$. This will be done essentially following a similar generalized long division algorithm. By hypothesis, $\mathrm{Ld}(f)$ is in the ideal generated by $\mathrm{Ld}(g_i)$'s. Hence there are q_i's in $A[x]$ such that

$$\mathrm{Ld}(f) = q_1 \,\mathrm{Ld}(g_1) + \cdots + q_n \,\mathrm{Ld}(g_n). \tag{41.10}$$

Suppose $\deg f = m$, $\deg g_i = m_i$, and $q_i(x) = c_i x^{m-m_i} + \overline{q}_i$ where $\overline{q}_i \in A[x]$ does not have a monomial of degree $(m - m_i)$. Notice that if $m < m_i$, then $c_i = 0$. Then none of the terms of

$$\overline{q}_1 \,\mathrm{Ld}(g_1) + \cdots + \overline{q}_n \,\mathrm{Ld}(g_n)$$

is of degree m. Therefore by (41.10), we obtain

$$\mathrm{Ld}(f) = (c_1 x^{m-m_1}) \,\mathrm{Ld}(g_1) + \cdots + (c_n x^{m-m_n}) \,\mathrm{Ld}(g_n). \tag{41.11}$$

Let

$$\overline{f}(x) := f(x) - ((c_1 x^{m-m_1}) g_1 + \cdots + (c_n x^{m-m_n}) g_n). \tag{41.12}$$

Then because of (41.11), $\deg \overline{f} < \deg f$. Moreover since f and g_i's are in I, so is $\overline{f}$. Hence by the strong induction hypothesis, $\overline{f} \in \langle g_1, \ldots, g_n\rangle$. Thus by (41.12), we conclude that $f \in \langle g_1, \ldots, g_n\rangle$. This completes the proof. □

By Lemma 41.1.3, we deduce that to generate an ideal I of $A[x]$, it is enough to generate all the leading terms. Next we focus more on leading coefficients.

Lemma 41.1.4. *Suppose A is a unital commutative ring and I is an ideal of $A[x]$. Let*

$$\mathrm{ld}(I) := \{\mathrm{ld}(f) \mid f \in I \setminus \{0\}\} \cup \{0\},$$

and

$$\mathrm{ld}_m(I) := \{\mathrm{ld}(f) \mid f \in I, \deg f = m\} \cup \{0\}$$

for every non-negative integer m. Then $\mathrm{ld}(I)$ and $\mathrm{ld}_m(I)$ are ideals of A.

Proof. Suppose $c_1, c_2 \in \mathrm{ld}(I)$. Then there are $f_1, f_2 \in I$ such that $\mathrm{Ld}(f_i) = c_i x^{m_i}$ for some non-negative integers m_i's. Without loss of generality, we can and will assume that $m_1 \le m_2$. Then, if $c_1 + c_2 \neq 0$, then

$$\mathrm{Ld}(x^{m_2-m_1} f_1 + f_2) = (c_1 + c_2) x^{m_2}$$

Hence $c_1 + c_2$ is the leading coefficient of $x^{m_2-m_1} f_1 + f_2$. Notice that $x^{m_2-m_1} f_1 + f_2$ is in I as f_i's are in I. Therefore $c_1 + c_2$ is either 0 or it is the leading coefficient of an element of I. Hence in either case $c_1 + c_2 \in \mathrm{ld}(I)$.

For every $a \in A$, either $ac_1 = 0$ or $\mathrm{ld}(af_1) = ac_1$. In either case, $ac_1 \in \mathrm{ld}(I)$. Altogether, we obtain that $\mathrm{ld}(I)$ is an ideal of A.

Next we show that $\mathrm{ld}_m(I)$ is an ideal. We start by discussing why it is closed under addition. If $c_1, c_2 \in \mathrm{ld}_m(I)$, then there are $f_1, f_2 \in I$ such that $\mathrm{Ld}(f_i) = c_i x^m$ for $i = 1, 2$. Hence either $c_1 + c_2 = 0$ or $\mathrm{Ld}(f_1 + f_2) = (c_1 + c_2) x^m$. In either case, $c_1 + c_2 \in \mathrm{ld}_m(I)$. For $a \in A$, either $ac_1 = 0$ or $\mathrm{Ld}(af_1) = ac_1 x^m$. Therefore in either case, $ac_1 \in \mathrm{ld}_m(I)$. Since $\mathrm{ld}_m(I)$ is closed under addition and multiplication by elements of A, we conclude that $\mathrm{ld}_m(I)$ is an ideal of A. This completes the proof. □

41.2 Hilbert's basis theorem

We have proved many results connecting (generating) an ideal of $A[x]$ with ideals of A. So we are well-placed to prove that if every ideal of A is finitely generated, then all ideals of $A[x]$ are finitely generated. Recall that all ideals of a ring R are finitely generated if and only if R is Noetherian.

Theorem 41.2.1 (Hilbert's basis theorem)**.** *Suppose A is a Noetherian unital commutative ring. Then $A[x]$ is Noetherian.*

Proof. By Lemma 12.3.5, we know that $A[x]$ is Noetherian if and only if every ideal of $A[x]$ is finitely generated. Suppose I is an ideal of $A[x]$. By Lemma 41.1.3, it is enough to show that there are $g_1, \ldots, g_n \in I$ such that for all $f \in I$,

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(g_1), \ldots, \mathrm{Ld}(g_n) \rangle.$$

Let $\mathrm{Ld}(I)$ be the ideal generated by $\mathrm{Ld}(f)$'s as f ranges in I.

Claim. *There are $g_1, \ldots, g_n \in I$ such that $\mathrm{Ld}(I) = \langle \mathrm{Ld}(g_1), \ldots, \mathrm{Ld}(g_n) \rangle$.*

Proof of Claim. Since A is Noetherian, $\mathrm{ld}(I)$ is a finitely generated ideal. Hence there are $f_1, \ldots, f_m \in I$ such that

$$\mathrm{ld}(I) = \langle \mathrm{ld}(f_1), \ldots, \mathrm{ld}(f_m) \rangle. \tag{41.13}$$

By (41.9), if $f \in I$ and $\deg f \geq \max(\deg f_1, \ldots, \deg f_m)$, then

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(f_1), \ldots, \mathrm{Ld}(f_m) \rangle. \tag{41.14}$$

Next we want to generate all the leading terms $\mathrm{Ld}(f)$ of polynomials $f \in I$ with $\deg f$ less than $k := \max(\deg f_1, \ldots, \deg f_m)$. By Lemma 41.1.4, $\mathrm{ld}_i(I)$'s are ideals of A for every integer i in $[0, k)$. Since A is Noetherian, $\mathrm{ld}_i(I)$ is a finitely generated ideal. Hence for every index i, there are $f_{i1}, \ldots, f_{in_i} \in I$ such that $\deg f_{ij} = i$ for every index j and

$$\mathrm{ld}_i(I) = \langle \mathrm{ld}(f_{i1}), \ldots, \mathrm{ld}(f_{in_i}) \rangle. \tag{41.15}$$

Next we show that if $f \in I$ and $\deg f = i$ for some non-negative integer $i < k$, then

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(f_{i1}), \ldots, \mathrm{Ld}(f_{in_i}) \rangle. \tag{41.16}$$

Notice that $\mathrm{ld}(f) \in \mathrm{ld}_i(f)$, and so by (41.15), there are $a_1, \ldots, a_{n_i} \in A$ such that

$$\mathrm{ld}(f) = a_1 \, \mathrm{ld}(f_{i1}) + \cdots + a_{n_i} \, \mathrm{ld}(f_{in_i}). \tag{41.17}$$

Because of (41.17), we conclude that

$$\begin{aligned} \mathrm{Ld}(f) &= \mathrm{ld}(f) x^i = (a_1 \, \mathrm{ld}(f_{i1}) + \cdots + a_{n_i} \, \mathrm{ld}(f_{in_i})) x^i \\ &= a_1 \, \mathrm{Ld}(f_{i1}) + \cdots + a_{n_i} \, \mathrm{Ld}(f_{in_i}). \end{aligned} \tag{41.18}$$

By (41.18), we deduce that (41.16) holds. By (41.14) and (41.16), we obtain that for all $f \in I$,

$$\mathrm{Ld}(f) \in \langle \mathrm{Ld}(f_r), \mathrm{Ld}(f_{ij}) \mid 1 \leq r \leq m, 0 \leq i < k, 1 \leq j \leq n_i \rangle.$$

This implies that

$$\mathrm{Ld}(I) = \langle \mathrm{Ld}(f_r), \mathrm{Ld}(f_{ij}) \mid 1 \leq r \leq m, 0 \leq i < k, 1 \leq j \leq n_i \rangle$$

as f_r's and f_{ij}'s are in I. This completes the proof of Claim. □

By above Claim and Lemma 41.1.3, we conclude that I is a finitely generated ideal. This completes the proof of Hilbert's basis theorem. □

41.3 Finitely generated rings and algebras

Hilbert's basis theorem implies that many of the rings that we have been working with are Noetherian. Here is an immediate corollary of Hilbert's basis theorem.

Corollary 41.3.1. *Suppose A is a Noetherian unital commutative ring. Then $A[x_1, \ldots, x_n]$ is Noetherian.*

Proof. We proceed by induction on n. The case of $n = 0$ follows from the hypothesis. So we focus on the induction step. By the induction hypothesis, $A[x_1, \ldots, x_n]$ is Noetherian. Hence by Hilbert's basis theorem, $(A[x_1, \ldots, x_n])[x_{n+1}]$ is Noetherian. This completes the proof. □

Corollary 41.3.2. *The rings $\mathbb{Z}[x_1, \ldots, x_n]$ and $F[x_1, \ldots, x_n]$, where F is a field, are Noetherian.*

Proof. This follows from Corollary 41.3.1 and the fact that the ring of integers and fields are Noetherian. □

Corollary 41.3.2 implies that every finitely generated ring or finitely generated F-algebra is Noetherian.

Definition 41.3.3. *(1) We say a unital commutative ring A is finitely generated if there are $a_1, \ldots, a_n \in A$ such that the smallest subring of A which contains a_i's is A.*

(2) Suppose F is field. We say a unital commutative ring A is called a finitely generated F-algebra if F is a subring of A, $1_F = 1_A$, and there are $a_1, \ldots, a_n$ in A such that the smallest subring of A which contains F and a_i's is A. In this case we write $A = F[a_1, \ldots, a_n]$.

The following easy lemma implies that every finitely generated ring is a quotient of the ring of polynomials $\mathbb{Z}[x_1, \ldots, x_n]$ for some positive integer n and every finitely generated F-algebra is a quotient of the ring of polynomials $F[x_1, \ldots, x_n]$ for some positive integer n.

Lemma 41.3.4. *1. Suppose A is generated by $a_1, \ldots, a_n$ as a ring. Then the following is a surjective homomorphism $\mathbb{Z}[x_1, \ldots, x_n] \to A$,*

$$\sum_{i_1, \ldots, i_n} m_{i_1, \ldots, i_n} x_1^{i_1} \cdots x_n^{i_n} \mapsto \sum_{i_1, \ldots, i_n} e(m_{i_1, \ldots, i_n}) a_1^{i_1} \cdots a_n^{i_n}$$

where $e : \mathbb{Z} \to A, e(m) := m1_A$ (see Lemma 2.3.1)

2. *Suppose F is a field and A is a finitely generated F-algebra. Suppose $A = F[a_1, \ldots, a_n]$. Then the evaluation map*

$$\phi_{a_1,\ldots,a_n} : F[x_1, \ldots, x_n] \to A, \phi_{a_1,\ldots,a_n}(f) := f(a_1, \ldots, a_n)$$

is a surjective ring homomorphism.

Proof. We leave the proof as an exercise. □

We obtain the following important consequence of Hilbert's basis theorem.

Theorem 41.3.5. *Suppose A is either a finitely generated ring or a finitely generated F-algebra, where F is a field. Then A is Noetherian.*

Proof. By Lemma 41.3.4, A is isomorphic to either $\mathbb{Z}[x_1, \ldots, x_n]/I$ or $F[x_1, \ldots, x_n]/I$ for some integer n and ideal I. Since a quotient of a Noetherian ring is Noetherian, by Corollary 41.3.2 we obtain that A is Noetherian. This completes the proof. □

Chapter 42

Lectures 18 and 19

In this section, we want to study zeros of multivariable polynomials with coefficients in a field. As we have mentioned in Section 38.3, a system of linear equations can be solved using the GAuss-Jordan elimination process. In the higher degree setting, we will use an argument inspired by the elimination process to find one solution of the given system of polynomial equations if it has a solution.

42.1 Set of common zeros and vanishing polynomials

In the single variable case, we have seen the importance of viewing polynomials as functions and exploring the connection between zeros of polynomials and ideals of ring of polynomials. We start by observing a similar connection in a multivariable setting.

For a subset S of $F[x_1, \ldots, x_n]$, let

$$Z(S) := \{\mathbf{a} \in F^n \mid f(\mathbf{a}) = 0 \text{ for all } f \in S\}.$$

For a subset V of F^n, let

$$I(V) := \{f \in F[x_1, \ldots, x_n] \mid f(\mathbf{v}) = 0 \text{ for all } \mathbf{v} \in V\}.$$

Alternatively we can say that $f \in I(V)$ if and only if the restriction $f|_V$ of f to V is zero where we are viewing f as a function from F^n to F. For $V \subseteq F^n$, let $\mathrm{Fun}(V, F)$ be the set of functions from V to F. Then it is easy to see that $\mathrm{Fun}(V, F)$ is a ring with respect to pointwise addition and multiplication; that means $(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ and $(f \cdot g)(\mathbf{x}) := f(\mathbf{x})g(\mathbf{x})$ for every $\mathbf{x} \in V$. By means of evaluation, we get a function

$$F[x_1, \ldots, x_n] \to \mathrm{Fun}(V, F), \quad f \mapsto f|_V. \tag{42.1}$$

It is easy to see that this is a ring homomorphism. From the definition $I(V)$, we have that $I(V)$ is the kernel of the ring homomorphism given in (42.1). The next lemma gives us basic properties of Z and I.

Lemma 42.1.1. *Suppose F is a field.*

1. *Functions I and Z are inclusion-reversing. That means for every $V_1 \subseteq V_2 \subseteq F^n$, $I(V_2) \subseteq I(V_1)$, and for every $S_1 \subseteq S_2 \subseteq F[x_1, \ldots, x_n]$, $Z(S_2) \subseteq Z(S_1)$.*
2. *For every $V \subseteq F^n$, $I(V)$ is an ideal of $F[x_1, \ldots, x_n]$, and $V \subseteq Z(I(V))$.*
3. *For every $S \subseteq F[x_1, \ldots, x_n]$, $\langle D \rangle \subseteq I(Z(S))$.*
4. *For every $S \subseteq F[x_1, \ldots, x_n]$, $Z(S) = Z(\langle S \rangle)$ and there are finitely many polynomials $g_1, \ldots, g_m$ such that*

$$Z(S) = Z(\{g_1, \ldots, g_m\}).$$

Proof. (1) If $f \in I(V_2)$, then $f|_{V_2} = 0$. Since $V_1 \subseteq V_2$, we conclude that $f|_{V_1} = 0$. Therefore $f \in I(V_1)$. This shows that $I(V_2) \subseteq I(V_1)$.

Suppose $\mathbf{v} \in Z(S_2)$. Then for every $f \in S_2$, $f(\mathbf{v}) = 0$. As $S_1 \subseteq S_2$, for every $f \in S_1$ we have $f(\mathbf{v}) = 0$. Hence $\mathbf{v} \in Z(S_1)$. This means $Z(S_2) \subseteq Z(S_2)$.

(2) For all $\mathbf{v} \in V$ and $f \in I(V)$, $f(\mathbf{v}) = 0$. Hence $\mathbf{v}$ is a common zero of elements of $I(V)$. This means $\mathbf{v} \in Z(I(V))$. Hence $V \subseteq Z(I(V))$.

We have already mentioned that $I(V)$ is the kernel of the ring homomorphism

$$F[x_1, \ldots, x_n] \to \text{Fun}(V, F), \quad f \mapsto f|_V.$$

Hence $I(V)$ is an ideal.

(3) For all $f \in S$ and $\mathbf{v} \in Z(S)$, $f(\mathbf{v}) = 0$. Hence $f|_{Z(S)} = 0$. Therefore $S \subseteq I(Z(S))$. Because $I(Z(S))$ is an ideal, we obtain that $\langle S \rangle I(Z(S))$.

(4) Since $\langle S \rangle \subseteq I(Z(S))$, for all $\mathbf{v} \in Z(S)$ and $f \in \langle S \rangle$ we have $f(\mathbf{v}) = 0$. Hence $\mathbf{v} \in Z(\langle S \rangle)$. Hence $Z(S) \subseteq Z(\langle S \rangle)$. Because Z is an order-reversing map and $S \subseteq \langle S \rangle$, we conclude that $Z(\langle S \rangle) \subseteq Z(S)$. Altogether, we obtain that

$$Z(S) = Z(\langle S \rangle).$$

Finally by Hilbert's basis theorem, every ideal of $F[x_1, \ldots, x_n]$ is finitely generated. Therefore there are $g_1, \ldots, g_n \in F[x_1, \ldots, x_n]$ such that $\langle S \rangle = \langle g_1, \ldots, g_n \rangle$. Hence

$$Z(S) = Z(\langle S \rangle) = Z(\langle g_1, \ldots, g_n \rangle) = Z(\{g_1, \ldots, g_n\}).$$

This completes the proof. □

42.2 Our general approach for finding a solution

Solving a system of polynomial equations

$$\begin{aligned} g_1(x_1, \ldots, x_n) =& 0 \\ \vdots& \\ g_m(x_1, \ldots, x_n) =& 0 \end{aligned} \tag{42.2}$$

is the same as understanding $Z(I)$ where I is the ideal generated by g_i's. As we mentioned earlier, we use an approach inspired by the elimination process to find one element of $Z(I)$ if possible. Eliminating variable x_n roughly means:

1. Deriving equations from (42.2) where x_n does not appear.

2. Finding x_n in terms of $x_1, \ldots, x_{n-1}$.

Notice that the first item essentially means considering $I \cap F[x_1, \ldots, x_{n-1}]$. Next for a given

$$(a_1, \ldots, a_{n-1}) \in Z(I \cap F[x_1, \ldots, x_{n-1}]),$$

we have to find $x_n = a_n$ which satisfies (42.2). This can be interpreted as follows.

Consider the projection

$$\pi : Z(I) \to Z(I \cap F[x_1, \ldots, x_{n-1}]), \quad \pi(\alpha_1, \ldots, \alpha_n) := (\alpha_1, \ldots, \alpha_{n-1}).$$

The above approach suggests that π is a surjective map; finding a point $(a_1, \ldots, a_{n-1})$ in the codomain, we look for $(a_1, \ldots, a_n)$ in the preimage π^{-1} of $(a_1, \ldots, a_{n-1})$. In general, it is, however, not true that π is surjective. To avoid this problem, instead of projecting to a *standard* $(n-1)$-dimensional subspace, we will carefully choose a projection to an $(n-1)$-dimensional subspace. To find out, how the needed projection can be chosen, we start with the standard projection π and find out when π is not surjective. Next we make a linear change of coordinates to avoid the bad cases, and make sure that π is surjective.

We also notice that one cannot expect that even a single variable polynomial to have a zero in F unless F is algebraically closed. So we will be assuming that the following holds:

Our standing assumption. Suppose F is an algebraically closed field and I is an ideal of $F[x_1, \ldots, x_n]$. Suppose $\mathbf{a} := (a_1, \ldots, a_{n-1})$ is in $Z(I \cap F[x_1, \ldots, x_{n-1}])$ and $\pi^{-1}(\mathbf{a}) = \varnothing$ where

$$\pi : Z(I) \to Z(I \cap F[x_1, \ldots, x_{n-1}]),\ \pi(\alpha_1, \ldots, \alpha_n) := (\alpha_1, \ldots, \alpha_{n-1}).$$

Notice that $\pi^{-1}(\mathbf{a}) = \varnothing$ if and only if $Z(\phi_{\mathbf{a}}(I)) = \varnothing$ where

$$\phi_{\mathbf{a}}(f) = f(a_1, \ldots, a_{n-1}, x_n) \in F[x_n].$$

Since the map $\phi_{\mathbf{a}}$ of evaluation at $\mathbf{a}$ is surjective, $\phi_{\mathbf{a}}(I)$ is an ideal of $F[x_n]$. Because $F[x_n]$ is a PID, $\phi_{\mathbf{a}}(I) = \langle p_{\mathbf{a}}(x_n) \rangle$ for some polynomial $p_{\mathbf{a}}(x_n) \in F[x_n]$. Hence $Z(\phi_{\mathbf{a}}(I))$ is the same as the set of zeros of $p_{\mathbf{a}}(x_n)$ in F. Since F is algebraically closed, every non-unit polynomial $F[x_n]$ has a zero in F. Therefore $Z(\phi_{\mathbf{a}}(I)) = \varnothing$ exactly when $p_{\mathbf{a}}(x_n)$ is a non-zero constant polynomial. Thus $\phi_{\mathbf{a}}(I) = \varnothing$ precisely when $1 \in \phi_{\mathbf{a}}(I)$. Hence our standing assumption implies that there is $f \in I$ such that $\phi_{\mathbf{a}}(f) = 1$. Let $A := F[x_1, \ldots, x_{n-1}]$, view $F[x_1, \ldots, x_n]$ as $A[x_n]$, and write

$$f := f_d x_n^d + \cdots + f_1 x_n + f_0$$

for some f_i's in A. Then $\phi_{\mathbf{a}}(f) = 1$ implies that

$$1 = f_d(\mathbf{a}) x_n^d + \cdots + f_1(\mathbf{a}) x_n + f_0(\mathbf{a}).$$

Hence we obtain

$$f_d(\mathbf{a}) = f_{d-1}(\mathbf{a}) = \cdots = f_1(\mathbf{a}) = 0, \text{ and } f_0(\mathbf{a}) = 1. \tag{42.3}$$

For every $g \in I$, we will look for elements of the form $rf + sg$ in A where $r, s \in A[x_n]$. Notice that for every such element, we have $rf + sg \in I \cap A$, and so $(rf + sg)(\mathbf{a}) = 0$. This brings us to the following question.

Question 42.2.1. *Suppose R is a unital commutative ring. For $f, g \in R[x]$, how can we find elements in $\langle f, g\rangle \cap R$?*

We will be studying Question 42.2.1 in the next section.

42.3 Resultant of two polynomials

Let's start investigating Question 42.2.1 with the case where $R = F$ is a field. In this case, $F[x]$ is a PID and $\langle f, g\rangle = \langle \gcd(f, g)\rangle$. So $\langle f, g\rangle \cap F \neq 0$ exactly when $\gcd(f, g) = 1$. Using Euclid's algorithm, we can find $r, s \in F[x]$ such that

$$\gcd(f, g) = rf + sg. \tag{42.4}$$

From computational point of view, we are interested in polynomials r and s that satisfy (42.4) and have *small* degrees.

Lemma 42.3.1. *Suppose F is a field and $f, g \in F[x]$. Then there are $r, s \in F[x]$ such that*

$$\gcd(f, g) = rf + sg, \quad \deg r < \deg g, \text{ and } \deg s < \deg f.$$

Proof. Since $F[x]$ is a PID, there are $\widetilde{r}, \widetilde{s} \in F[x]$ such that

$$\gcd(f, g) = \widetilde{r}f + \widetilde{s}g. \tag{42.5}$$

by the long division for elements of $F[x]$, there are $q_1, q_2, r, s \in F[x]$ such that

$$\widetilde{r} = q_1 g + r, \ \deg r < \deg g, \text{ and } \widetilde{s} = q_2 f + s, \ \deg s < \deg f. \tag{42.6}$$

By (42.5) and (42.6), we conclude that

$$\begin{aligned} \gcd(f, g) =&(q_1 g + r)f + (q_2 f + s)g \\ =&gf(q_1 + q_2) + rf + sg. \end{aligned} \tag{42.7}$$

By (42.7), we obtain that

$$gf(q_1 + q_2) = \gcd(f, g) - rf - sg,$$

which implies that

$$\begin{aligned} \deg g + \deg f + \deg(q_1 + q_2) \leq& \max(\deg(\gcd(f, g)), \deg r + \deg f, \deg s + \deg g) \\ \leq& \deg f + \deg g - 1. \end{aligned}$$

Hence $\deg(q_1 + q_2) \leq -1$. Therefore $q_1 + q_2 = 0$. Thus by (42.7), we obtain $\gcd(f, g) = rf + sg$. This completes the proof as $\deg r < \deg g$ and $\deg s < \deg f$ by (42.6). □

Inspired by Lemma 42.3.1, we consider only pairs (r, s) in $R[x] \times R[x]$ such that $\deg r < \deg g$ and $\deg s < \deg f$. Let

$$P_i := \{h \in R[x] \mid \deg h \leq i\},$$

and consider the following map

$$\ell : P_{m-1} \times P_{d-1} \to P_{d+m-1}, \quad \ell(r, s) := rf + sg, \tag{42.8}$$

where $d := \deg f$ and $m := \deg g$. Notice that ℓ is an R-module homomorphism. Moreover since every element of P_i can be uniquely written as

$$a_0 + a_1 x + \cdots + a_i x^i$$

for some a_i's in R, ℓ can be represented by the following matrix:

$$R(f,g) := \begin{pmatrix} f_0 & & & & g_0 & & & \\ f_1 & f_0 & & & g_1 & g_0 & & \\ \vdots & f_1 & \ddots & & \vdots & g_1 & \ddots & \\ f_{d-1} & \vdots & \ddots & f_0 & g_{m-1} & \vdots & \ddots & g_0 \\ f_d & f_{d-1} & & f_1 & g_m & g_{m-1} & & g_1 \\ & f_d & \ddots & \vdots & & g_m & \ddots & \vdots \\ & & \ddots & f_{d-1} & & & \ddots & g_{m-1} \\ & & & f_d & & & & g_m \end{pmatrix} \in \mathrm{M}_{d+m}(R).$$

where $f = f_d x^d + \cdots + f_1 x + f_0$, $g = g_m x^m + \cdots + g_1 x + g_0$, $f_d \neq 0$, $g_m \neq 0$, there are m columns of f_i's and d columns of g_i's. Notice that

$$\begin{pmatrix} 1 & x & \cdots & x^{d+m-1} \end{pmatrix} R(f,g) = \begin{pmatrix} f & \cdots (x^{m-1} f) & g & \cdots & (x^{d-1} g) \end{pmatrix}. \tag{42.9}$$

Multiplying both sides of (42.9) by the adjoint $\operatorname{adj}(R(f,g))$ of $R(f,g)$, we obtain that

$$\det(R(f,g)) \begin{pmatrix} 1 & x & \cdots & x^{d+m-1} \end{pmatrix} \in \mathrm{M}_{m+d,a}(\langle f, g \rangle). \tag{42.10}$$

The determinant $\det(R(f,g))$ of $R(f,g)$ is called the *resultant* of f and g, and it is denoted by $r(f,g)$. Altogether we conclude.

Lemma 42.3.2. *Suppose R is a unital commutative ring and $f, g \in R[x]$. Then*

$$r(f,g) \in \langle f, g \rangle \cap R.$$

Proof. By (42.10), $r(f,g) \in \langle f, g \rangle$. Since $R(f,g) \in \mathrm{M}_{d+m}(R)$ and $r(f,g)$ is the determinant of $\det(R(f,g))$, we have that $r(f,g) \in R$. This completes the proof. □

42.4 What happens if π is not surjective

In this section, we work under our standing assumption and find a strong condition on I. Let's recall that under the standing assumption, by (42.3) there is $f \in A[x_n]$ where $A := F[x_1, \ldots, x_{n-1}]$ such that $f = \sum_{i=0}^{d} f_i x_n^i$ and

$$f_d(\mathbf{a}) = \cdots = f_1(\mathbf{a}) = 0 \quad \text{and} \quad f_0(\mathbf{a}) = 0.$$

For every $g \in I$, by Lemma 42.3.2 we obtain

$$r(f, g) \in I \cap F[x_1, \ldots, x_n]. \tag{42.11}$$

Since $\mathbf{a} \in Z(I \cap F[x_1, \ldots, x_n]$, by (42.11) we conclude

$$(r(f, g))(\mathbf{a}) = 0 \tag{42.12}$$

for every $g \in I$. This means $\det(R(f, g)(\mathbf{a})) = 0$. Notice that by (42.3), $R(f, g)(\mathbf{a})$ is equal to

$$\begin{pmatrix}
f_0(\mathbf{a}) & & & & g_0(\mathbf{a}) & & & \\
f_1(\mathbf{a}) & f_0(\mathbf{a}) & & & g_1(\mathbf{a}) & g_0(\mathbf{a}) & & \\
\vdots & f_1(\mathbf{a}) & \ddots & & \vdots & g_1(\mathbf{a}) & \ddots & \\
f_{d-1}(\mathbf{a}) & \vdots & \ddots & f_0(\mathbf{a}) & g_{m-1}(\mathbf{a}) & \vdots & \ddots & g_0(\mathbf{a}) \\
f_d(\mathbf{a}) & f_{d-1}(\mathbf{a}) & & f_1(\mathbf{a}) & g_m(\mathbf{a}) & g_{m-1}(\mathbf{a}) & & g_1(\mathbf{a}) \\
 & f_d(\mathbf{a}) & \ddots & \vdots & & g_m(\mathbf{a}) & \ddots & \vdots \\
 & & \ddots & f_{d-1}(\mathbf{a}) & & & \ddots & g_{m-1}(\mathbf{a}) \\
 & & & f_d(\mathbf{a}) & & & & g_m(\mathbf{a})
\end{pmatrix}$$

is equal to

$$\begin{pmatrix}
1 & & & & g_0(\mathbf{a}) & & & \\
 & 1 & & & g_1(\mathbf{a}) & g_0(\mathbf{a}) & & \\
 & & \ddots & & & g_1(\mathbf{a}) & \ddots & \\
 & & & 1 & g_{m-1}(\mathbf{a}) & \vdots & \ddots & g_0(\mathbf{a}) \\
 & & & & g_m(\mathbf{a}) & g_{m-1}(\mathbf{a}) & & g_1(\mathbf{a}) \\
 & & & & & g_m(\mathbf{a}) & \ddots & \vdots \\
 & & & & & & & g_{m-1}(\mathbf{a}) \\
 & & & & & & & g_m(\mathbf{a})
\end{pmatrix}.$$

Hence $R(f, g)(\mathbf{a})$ is an upper-triangular matrix with diagonal entries equal to

$$\underbrace{1, \ldots, 1}_{m\text{-times}}, \underbrace{g_m(\mathbf{a}), \ldots, g_m(\mathbf{a})}_{d\text{-times}}.$$

Therefore $r(f, g)(\mathbf{a}) = \det(R(f, g)(\mathbf{a})) = g_m(\mathbf{a})^d$. Since $r(f, g)(\mathbf{a}) = 0$, we deduce that $g_m(\mathbf{a}) = 0$ for every $g \in I$. Let's summarize what we have proved.

Lemma 42.4.1. *Suppose F is an algebraically closed field, I is an ideal of $F[x_1,\ldots,x_n]$, $\mathbf{a}\in Z(I\cap F[x_1,\ldots,x_{n-1}]$, and $\pi^{-1}(\mathbf{a})=\varnothing$ where*

$$\pi: Z(I)\to Z(I\cap F[x_1,\ldots,x_{n-1}]), \pi(a_1,\ldots,a_n):=(a_1,\ldots,a_{n-1}).$$

Then for every $g=\sum_{i=0}^m g_i x_n^i\in I$ with $g_i\in F[x_1,\ldots,x_{n-1}]$ and $g_m\neq 0$, we have $g_m(\mathbf{a})=0$.

42.5 Finding a suitable linear change of coordinates

In this section, we show that after a linear change of coordinates, we can find $g=g_m x_n^m+\cdots g_1x_n+g_0$ in I such that g_m is a non-zero constant. By this result and Lemma 42.4.1, we conclude that π is surjective. This in turn shows that $Z(I)$ is not empty.

Lemma 42.5.1. *Suppose F is an infinite field, and $h\in F[x_1,\ldots,x_n]\setminus F$.*

1. *For every $\vec{\alpha}:=(\alpha_1,\ldots,\alpha_{n-1})\in F^{n-1}$, $\theta_{\vec{\alpha}}:F[x_1,\ldots,x_n]\to F[x_1,\ldots,x_n]$,*

$$\theta_{\vec{\alpha}}(f):=f(x_1+\alpha_1x_n,\ldots,x_{n-1}+\alpha_{n-1}x_n,x_n)$$

   is an F-isomorphism.

2. *There is $\vec{\alpha}\in F^{n-1}$ such that*

$$\theta_{\vec{\alpha}}(f)=cx_n^m+h_{m-1}x^{m-1}+\cdots+h_0$$

   for some h_i's in $F[x_1,\ldots,x_{n-1}]$ and $c\in F^{\times}$.

Proof. (1) Notice that for $\vec{\alpha},\vec{\beta}\in F^{n-1}$,

$$\begin{aligned}\theta_{\vec{\alpha}}\circ\theta_{\vec{\beta}}(f)=&\theta_{\vec{\beta}}(x_1+\alpha_1x_n,\ldots,x_{n-1}+\alpha_{n-1}x_n,x_n)\\=&f(x_1+\alpha_1x_n+\beta_1x_n,\ldots,x_{n-1}+\alpha_{n-1}x_n+\beta_{n-1}x_n,x_n)\\=&\theta_{\vec{\alpha}+\vec{\beta}}(f).\end{aligned}$$

In particular, $\theta_{\vec{\alpha}}\circ\theta_{-\vec{\alpha}}=\theta_{-\vec{\alpha}}\circ\theta_{\vec{\alpha}}=\mathrm{id}$, which implies that $\theta_{\vec{\alpha}}$ is bijective. It is easy to check that $\theta_{\vec{\alpha}}$ is an F-linear ring homomorphism.

(2) Since $\theta_{\vec{\alpha}}$ is an F-linear map and every polynomial is an F-linear combination of monomials, we start with understanding $\theta_{\vec{\alpha}}(\mathbf{x}^{\mathbf{i}})$ where $\mathbf{x}^{\mathbf{i}}:=x_1^{i_1}\cdots x_n^{i_n}$. We have

$$\begin{aligned}\theta_{\vec{\alpha}}(\mathbf{x}^{\mathbf{i}})=&\prod_{j=1}^{n-1}(x_j+\alpha_jx_n)^{i_j}\cdot x_n^{i_n}\\=&\Big(\prod_{j=1}^{n-1}(\alpha_j)^{i_j}\Big)x_n^m+r_{m-1}x_n^{m-1}+\cdots+r_0,\end{aligned}\tag{42.13}$$

where $m=i_1+\cdots+i_n$ and r_i's are in $F[x_1,\ldots,x_{n-1}]$. The summation $i_1+\cdots+i_n$ is called the *total degree* of the monomial $\mathbf{x}^{\mathbf{i}}$ and it is denoted by either $\deg\mathbf{x}^{\mathbf{i}}$ or $\|\mathbf{i}\|_1$. By

(42.13), the coefficient of $x_n^{\|\mathbf{i}\|_1}$ in $\theta_{\vec{\alpha}}(\mathbf{x}^{\mathbf{i}})$ is $\phi_{\vec{\alpha}'}(\mathbf{x}^{\mathbf{i}})$ where $\vec{\alpha}' := (\alpha_1, \ldots, \alpha_{n-1}, 1)$; this means it is $\mathbf{x}^{\mathbf{i}}$ evaluated at $\vec{\alpha}'$. Another consequence of (42.13) is that the degree of $\theta_{\vec{\alpha}'}(\mathbf{x}^{\mathbf{i}})$ as a polynomial in terms of the variable x_n is at most $\|\mathbf{i}\|_1$.

Suppose $h = \sum_{\mathbf{i}} c_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}$. Then

$$\begin{aligned}\theta_{\vec{\alpha}}(h) &= \sum_{\mathbf{i}} c_{\mathbf{i}} \theta_{\vec{\alpha}'}(\mathbf{x}^{\mathbf{i}}) \\ &= \sum_{\mathbf{i}} c_{\mathbf{i}} (\phi_{\vec{\alpha}'}(\mathbf{x}^{\mathbf{i}}) x_n^{\|\mathbf{i}\|_1} + p_{\mathbf{i}}(x_n)), \end{aligned} \tag{42.14}$$

where $p_{\mathbf{i}}(x_n) \in (F[x_1, \ldots, x_{n-1}])[x_n]$ is a polynomial of degree at most $\|\mathbf{i}\|_1 - 1$. The total degree of h is defined to be

$$\max\{\|\mathbf{i}\| \mid c_{\mathbf{i}} \neq 0\},$$

and it is denoted by $\deg h$. Suppose $m := \deg h$, and let

$$\overline{h} := \sum_{\|\mathbf{i}\|_1 = m} c_{\mathbf{i}} \mathbf{x}^{\mathbf{i}};$$

this means $\overline{h}$ is the sum of all the monomials of h that have the maximum possible total degree. Then by (42.14), we conclude that

$$\begin{aligned}\theta_{\vec{\alpha}}(h) = &\sum_{\|\mathbf{i}\|_1 = m} c_{\mathbf{i}} \phi_{\vec{\alpha}'}(\mathbf{x}^{\mathbf{i}}) x_n^m + p(x_n) \\ =&\phi_{\vec{\alpha}'}(\sum_{\|\mathbf{i}\|_1 = m} c_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}) x_n^m + p(x_n) \\ =&\overline{h}(\vec{\alpha}') x_n^m + p(x_n), \end{aligned} \tag{42.15}$$

where $p(x_n) \in (F[x_1, \ldots, x_{n-1}])[x_n]$ is a polynomial of degree at most $m - 1$. Next we will find $\vec{\alpha}$ such that $\overline{h}(\vec{\alpha}') \neq 0$. By this result and (42.15), we will finish the proof of Lemma.

Claim. If F is an infinite field and $\overline{h}' \in F[x_1, \ldots, x_k] \setminus \{0\}$, then $\overline{h}'(\vec{\alpha}) \neq 0$ for some $\vec{\alpha} \in F^k$.

Proof of Claim. We proceed by induction on the number k of variables. By Corollary 7.1.5, a single variable polynomial of degree m in $F[x]$ has at most m distinct zeros in F. Since F is an infinite field, $\overline{h}'(\alpha) \neq 0$ for some $\alpha \in F$. Next we show the induction step. Suppose

$$\overline{h}' = h_m x_k^m + \cdots + h_0$$

for some h_i's in $F[x_1, \ldots, x_{k-1}]$ and $h_m \neq 0$. By the induction hypothesis,

$$h_m(\alpha_1, \ldots, \alpha_{k-1}) \neq 0$$

for some α_i's in F. Hence

$$\overline{h}'(\alpha_1, \ldots, \alpha_{k-1}, x_k) = h_m(\alpha_1, \ldots, \alpha_{k-1}) x_k^m + \cdots + h_0(\alpha_1, \ldots, \alpha_{k-1})$$

is a non-zero single variable polynomial. Therefore by the single variable case,

$$\overline{h}'(\alpha_1, \ldots, \alpha_{k-1}, \alpha_k) \neq 0$$

for some $\alpha_k \in F$. This completes proof of the Claim. ☐

To finish proof of Lemma 42.5.1, we let

$$\overline{h}'(x_1, \ldots, x_{n-1}) := \overline{h}(x_1, \ldots, x_{n-1}, 1).$$

Notice that since all the monomials of $\overline{h}$ have the same total degree, $\overline{h}'$ is a non-zero polynomial. Hence by the above Claim, $\overline{h}'(\vec{\alpha}) \neq 0$ for some $\vec{\alpha} \in F^{n-1}$. Therefore $\overline{h}(\vec{\alpha}') \neq 0$. Thus by (42.15), we deduce that $\theta_{\vec{\alpha}}(h)$ has the desired form. This completes the proof. ☐

It is worth pointing out that $\theta_{\vec{\alpha}}(f)$ is evaluating f at

$$\begin{pmatrix} 1 & & & -\alpha_1 \\ & \ddots & & \vdots \\ & & 1 & -\alpha_{n-1} \\ & & & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$$

This is another hidden relation of our argument for higher degree polynomial equations and the reduced row-column process.

42.6 Hilbert's Nullstellensatz

In this section, we prove Hilbert's Nullstellensatz which gives us the necessary and sufficient condition for a system of polynomial equations to have a zero over an algebraically closed field.

Theorem 42.6.1 (Hilbert's Nnullstellensatz, v.1)**.** *Suppose F is an algebraically closed field, and I is an ideal of $F[x_1, \ldots, x_n]$. Then $Z(I) \neq \varnothing$ if and only if $1 \notin I$.*

Proof. ($\Rightarrow$) If $1 \in I$, then $Z(I) = \varnothing$.

($\Leftarrow$) We proceed by induction on the number n of variables. For the base of induction, we notice that $F[x_1]$ is a PID. Hence $I = \langle p(x_1) \rangle$ for some polynomial p. Since $1 \notin I$, p is not a nonn-zero constant polynomial. As F is algebraically closed, p has a zero in F. Therefore $Z(I) \neq \varnothing$. Next we prove the induction step. If $I = 0$, then $Z(I) = F^n$, and there is nothing to prove. So we can and will assume that there is a non-constant polynomial h in I. By Lemma 42.5.1, there is $\vec{\alpha} \in F^{n-1}$ such that

$$\theta_{\vec{\alpha}}(h) = c x_n^m + h_{m-1} x_n^{m-1} + \cdots + h_0 \tag{42.16}$$

for some h_i's in $F[x_1, \ldots, x_{n-1}]$ and $c \in F^{\times}$. Since $\theta_{\vec{\alpha}}$ is an isomorphism (see Lemma 42.5.1), $\theta_{\vec{\alpha}}(I)$ is an ideal of $F[x_1, \ldots, x_n]$. By (42.16) and Lemma 42.5.1, $\pi : Z(\theta_{\vec{\alpha}}(I)) \to Z(\theta_{\vec{\alpha}}(I) \cap F[x_1, \ldots, x_{n-1}])$,

$$\pi(a_1, \ldots, a_n) = (a_1, \ldots, a_{n-1}).$$

is surjective. Since $1 \notin I$, $1 \notin \theta_{\vec{\alpha}}(I)$. Therefore $1 \notin \theta_{\vec{\alpha}}(I \cap F[x_1, \ldots, x_{n-1}])$. Hence by the induction hypothesis, $Z(\theta_{\vec{\alpha}}(I) \cap F[x_1, \ldots, x_{n-1}]) \neq \varnothing$. Since π is surjective, $Z(\theta_{\vec{\alpha}}(I)) \neq \varnothing$. Notice that

$$\begin{aligned}(a_1, \ldots, a_n) \in Z(\theta_{\vec{\alpha}}(I)) &\Leftrightarrow \forall f \in I, \theta_{\vec{\alpha}}(f)(a_1, \ldots, a_n) = 0\\ &\Leftrightarrow \forall f \in I, f(a_1 + \alpha_1 a_n, \ldots, a_{n-1} + \alpha_{n-1} a_n, a_n) = 0\\ &\Leftrightarrow (a_1 + \alpha_1 a_n, \ldots, a_{n-1} + \alpha_{n-1} a_n, a_n) \in Z(I).\end{aligned}$$

Therefore $Z(I) \neq \varnothing$. This completes the proof. □

Hilbert's Nullstellensatz has many implications. For instance we can classify all the maximal ideals of $F[x_1, \ldots, x_n]$ when F is algebraically closed.

Theorem 42.6.2 (Hilbert's Nullstellensatz, v.2)**.** *Suppose F is an algebraically closed field. Then the following is a bijection:*

$$F^n \to \operatorname{Max}(F[x_1, \ldots, x_n]), \quad \mathbf{a} \mapsto I(\{\mathbf{a}\}),$$

where $\operatorname{Max}(\cdot)$ is the set of all the maximal ideals. Moreover

$$I(a_1, \ldots, a_n) = \langle x_1 - a_1, \ldots, x_n - a_n \rangle.$$

Proof. For $\mathbf{a} \in F^n$, let $\phi_{\mathbf{a}} : F[x_1, \ldots, x_n] \to F$ be the map of evaluation at $\mathbf{a}$. Then by the first isomorphism theorem,

$$F[x_1, \ldots, x_n] / \ker \phi_{\mathbf{a}} \simeq F.$$

Hence $\ker \phi_{\mathbf{a}} \in \operatorname{Max}(F[x_1, \ldots, x_n])$. Notice that $\ker \phi_{\mathbf{a}} = I(\{\mathbf{a}\})$. Therefore the given map is well-defined.

Next we show injectivity. Notice that $x_i - a_i$'s are in $I(\{\mathbf{a}\})$ if $\mathbf{a} = (a_1, \ldots, a_n)$. Hence $Z(I(\{\mathbf{a}\})) = \{\mathbf{a}\}$. Therefore $I(\{\mathbf{a}\}) = I(\{\mathbf{a}'\})$ implies that $\mathbf{a} = \mathbf{a}'$, and we deduce the injectivity.

Finally we show the surjectivity. Suppose $M \in \operatorname{Max}(F[x_1, \ldots, x_n])$. Then by Hilbert's Nullstellensatz, version 1, there is $\mathbf{a} \in Z(M)$. Hence $M \subseteq I(\{\mathbf{a}\})$. Because M is a maximal ideal, $M \subseteq I(\{\mathbf{a}\})$, and $I(\{\mathbf{a}\})$ is a proper ideal, we conclude that $M = I(\{\mathbf{a}\})$. We obtain the surjectivity.

Since $x_i - a_i$'s are in $I(\{\mathbf{a}\})$ where $\mathbf{a} := (a_1, \ldots, a_n)$, we have that

$$\langle x_1 - a_1, \ldots, x_n - a_n \rangle \subseteq I(\{\mathbf{a}\}). \tag{42.17}$$

Notice that $\operatorname{Ld}(x_i - a_i) = x_i$. Hence every non-constant monomial is in the ideal generated by $\operatorname{Ld}(x_i - a_i)$'s. Since $I(\{\mathbf{a}\})$ is a proper ideal, for every $f \in I \setminus \{0\}$, $\operatorname{Ld}(f)$ is a non-constant monomial. Therefore

$$\operatorname{Ld}(f) \in \langle \operatorname{Ld}(x_1 - a_1), \ldots, \operatorname{Ld}(x_n - a_n) \rangle \tag{42.18}$$

for every $f \in I(\{\mathbf{a}\})$. By (42.17), (42.18), and Lemma 41.1.3, we deduce that

$$\langle x_1 - a_1, \ldots, x_n - a_n \rangle = I(\{\mathbf{a}\}).$$

This completes the proof. □

Using the second version of Hilbert's Nullstellensatz, we can understand points of $Z(I)$ in terms of maximal ideals.

Theorem 42.6.3 (Hilbert's Nullstellensatz, v.3)**.** *Suppose F is an algebraically closed field and I is an ideal of $F[\mathbf{x}] := F[x_1, \ldots, x_n]$. Then the following maps are bijections:*

$$Z(I) \to \{M \in \mathrm{Max}(F[\mathbf{x}]) \mid I \subseteq M\}, \quad \mathbf{a} \mapsto I(\{\mathbf{a}\}), \tag{42.19}$$

$$\{M \in \mathrm{Max}(F[\mathbf{x}]) \mid I \subseteq M\} \to \mathrm{Max}\left(\frac{F[\mathbf{x}]}{I}\right), \quad M \mapsto \frac{M}{I}, \tag{42.20}$$

and

$$Z(I) \to \mathrm{Max}\left(\frac{F[\mathbf{x}]}{I}\right), \quad \mathbf{a} \mapsto \frac{I(\{\mathbf{a}\})}{I}. \tag{42.21}$$

Proof. For every $\mathbf{a} \in Z(I)$, we have that $I \subseteq I(\mathbf{a})$. By the second version of Hilbert's Nullstellensatz, $I(\{\mathbf{a}\})$ is in $\mathrm{Max}(F[\mathbf{x}])$. Therefore the map given in (42.19) is well-defined. By the second version of Hilbert's Nullstellensatz, $\mathbf{a} \mapsto I(\{\mathbf{a}\})$ is injective. Next we show surjectivity of the map given in (42.19). Suppose $M \in \mathrm{Max}(F[\mathbf{x}])$ and $I \subseteq M$. Then by the second version of Hilbert's Nullstellensatz, $M = I(\{\mathbf{a}\})$ for some $\mathbf{a} \in F^n$. So $I \subseteq I(\{\mathbf{a}\})$, which means $\mathbf{a} \in Z(I)$. This shows that the map given in (42.19) is surjective.

By Lemma 9.3.4, we know that the following is a bijection

$$\{J \trianglelefteq F[\mathbf{x}] \mid I \subseteq J\} \to \left\{\overline{J} \trianglelefteq \frac{F[\mathbf{x}]}{I}\right\}, \quad J \mapsto \frac{J}{I}. \tag{42.22}$$

Since this bijection preserves inclusion, we deduce that it sends maximal ideals to maximal ideals. Hence the function given in (42.20) is a bijection.

The map given in (42.21) is the composite of the bijections given in (42.19) and (42.20). Hence it is a bijection. This completes the proof. □

Another version of Hilbert's Nullstellensatz gives us a complete understanding of $I(Z(I))$. This means that it answers the following question:

What are the polynomials that vanish on the common zeros of elements of I?

Notice that if $f^n \in I$ for some positive integer n, then $f^n|_{Z(I)} = 0$. Hence $f|_{Z(I)} = 0$, which means that

$$\{f \in F[\mathbf{x}] \mid f^n \in I \text{for some positive integer } n\} \subseteq I(Z(I)). \tag{42.23}$$

The left hand side of the inclusion given in (42.23) is called the *radical* of I and it is denoted by $\sqrt{I}$. Our final version of Hilbert's Nullstellensatz states the equality in (42.23) holds.

Theorem 42.6.4 (Hilbert's Nullstellensatz, v.4)**.** *Suppose F is an algebraically closed field and I is an ideal of $F[x_1, \ldots, x_n]$. Then $I(Z(I)) = \sqrt{I}$.*

Proof. By (42.23), we have that $\sqrt{I} \subseteq I(Z(I))$. Now assume that $f \notin \sqrt{I}$. We want to show that $f \notin I(Z(I))$. This means we would like to show that there is $\mathbf{a} \in F^n$

such that for every $g \in I$, $g(\mathbf{a}) = 0$ and $f(\mathbf{a}) \neq 0$. Since an element of F is non-zero if and only if it is a unit, we are looking for $\mathbf{a} \in F^n$ and $b \in F$ such that

$$g(\mathbf{a}) = 0 \text{ for every } g \in I \text{ and } f(\mathbf{a})b = 1. \tag{42.24}$$

Existence of $(\mathbf{a}, b) \in F^n \times F$ so that (42.24) holds is equivalent saying that the following system of polynomial equations

$$g(\mathbf{x}) = 0 \text{ for every } g \in I \text{ and } f(\mathbf{x})y - 1 = 0 \tag{42.25}$$

have a common zero in F^{n+1}. By the first version of. Hilbert's Nullstellensatz, the system of equations given in (42.25) has a common zero in F^{n+1} if and only if the ideal J generated by I and $f(\mathbf{x})y - 1$ in $(F[\mathbf{x}])[y]$ is a proper ideal. Suppose to the contrary that $1 \in J$. This implies that

$$1 = f_1 r_1(y) + \cdots + f_k r_k(y) + (fy - 1)s(y) \tag{42.26}$$

for some f_i's in $F[\mathbf{x}]$ and $r_1, \ldots, r_k, s \in (F[\mathbf{x}])[y]$. Consider $\frac{1}{f}$ in the field $F(x_1, \ldots, x_n)$ of fractions of $F[x_1, \ldots, x_n]$, and evaluate r_i's and s at $\frac{1}{f}$. This means we consider the map of evaluation at $\frac{1}{f}$

$$\phi_{\frac{1}{f}} : (F[\mathbf{x}])[y] \to F(\mathbf{x}), \phi_{\frac{1}{f}}(r) := r\Big(\frac{1}{f}\Big),$$

and apply it to the both sides of (42.26). This way, we obtain

$$1 = f_1 r_1\Big(\frac{1}{f}\Big) + \cdots + f_k r_k\Big(\frac{1}{f}\Big). \tag{42.27}$$

Notice that $r_i(\frac{1}{f}) = \frac{p_i}{f^{d_i}}$ for some $p_i \in F[\mathbf{x}]$ and $d_i \in \mathbb{Z}^+$. Let $N := \max(d_1, \ldots, d_k)$. Then multiplying both sides of (42.27) by f^N, we obtain that

$$f^N = f_1 q_1 + \cdots + f_k q_k$$

for some q_i's in $F[\mathbf{x}]$. Therefore f^N is in the ideal generated by f_i's. Hence $f^N \in I$ as f_i's are in I. This implies that $f \in \sqrt{I}$, which is a contradiction. This completes the proof. □

Notice that if P is a prime ideal, then $\sqrt{P} = P$. This is the case as for a prime ideal P, $f^n \in P$ for some positive integer n implies that $f \in P$. Hence by the fourth version of Hilbert's Nullstellensatz, we have

$$I(Z(P)) = P$$

for every prime ideal P of $F[\mathbf{x}]$ when F is an algebraically closed field F.

Here is an interesting consequence of Hilbert's Nullstellensatz in ring theory.

Proposition 42.6.5. *Suppose F is an algebraically closed field. Then for every prime ideal P of $F[x_1, \ldots, x_n]$, we have*

$$P = \bigcap_{M \in \mathrm{Max}(F[\mathbf{x}]), P \subseteq M} M.$$

Proof. By the fourth version of Hilbert's Nullstellensatz, we have $P = I(Z(P))$. Hence

$$P = I(Z(P)) = \bigcap_{\mathbf{a}\in Z(P)} I(\{\mathbf{a}\}) = \bigcap_{M\in \mathrm{Max}(F[\mathbf{x}]), P\subseteq M} M$$

where the last equality holds because of the third version of Hilbert's Nullstellensatz. This completes the proof. □

42.7 Final remarks

Let's finish by mentioning that not surprisingly zeros of multivariable polynomials are much more complicated than their single variable counterpart. In the single variable case, $Z(I)$ is either a finite set or F. Many of the interesting surfaces or manifolds that we know can be viewed as the set of real or complex zeros certain system of polynomial equations. Here are some examples:

1. Sphere: $x^2 + y^2 + z^2 = 1$.
2. Special linear group: $\mathrm{SL}_n(\mathbb{C}) := \{A \in \mathrm{M}_n(\mathbb{C}) \mid \det A = 1\}$.
3. A curve related to Fermat's last conjecture: $X^n + Y^n = 1$.
4. Equation related to Markoff triples: $x^2 + y^2 + z^2 = 3xyz$.
5. Elliptic curves: $y^2 = x^3 + ax + b$.

Thinking about these equations from geometric point of view gives us the needed language to ask many interesting questions, helps us use results and tools from geometry and topology, and brings us many unexpected connections.